Protein compactness and interaction valency define the architecture of a biomolecular condensate across scales

Anton A. Polyansky[1*], Laura D. Gallego[2], Roman G. Efremov[3], Alwin Köhler[2] & Bojan Zagrovic[1*]

February 2022

[1]Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, Campus Vienna Biocenter 5, A-1030, Vienna, Austria

[2]Max Perutz Labs, University of Vienna and Medical University of Vienna, Vienna Biocenter Campus (VBC), Dr. Bohr-Gasse 9/3, A-1030 Vienna, Austria

[3]MM Shemyakin and Yu A Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya, 16/10, 117997 Moscow, Russia

To whom correspondence should be directed: *anton.polyansky@univie.ac.at; *bojan.zagrovic@univie.ac.at

Keywords: *liquid-liquid phase separation; biomolecular condensate; Large1; colloid; fractal cluster*

Abstract

Formation of biomolecular condensates via liquid-liquid phase separation (LLPS) contributes to the organization and function of living cells, but the general physicochemical principles involved are not fully understood. The present study provides a quantitative, multiscale framework that connects single- and multi-copy, microsecond molecular dynamics simulations with the experimentally observed, micrometer-scale LLPS behavior of an 80-aa N-terminal fragment of yeast Lge1 ($Lge1_{1-80}$), a regulator of histone H2B ubiquitination. Analysis of protein self-interactions modeled at all-atom resolution elucidates the impact of key residues on interaction valency and establishes a link between configurational entropy, valency and compactness of proteins inside a model condensate. An analytical formalism derived from first principles describes condensate structure across different length scales as a function of the interaction valency and compactness of its molecular components. In analogy to fractal cluster formation seen in biological and non-biological colloids, the formalism provides an atomistically resolved model of a micrometer-size condensate starting from the simulated ensemble of individual protein conformers. Finally, the simulation-derived fractal dimensions of the condensates of $Lge1_{1-80}$ and its mutants are shown to be consistent with their *in vitro* LLPS behavior. The presented framework provides a general, experimentally testable description of the multiscale features of biomolecular condensates and embeds their study in a wider context of self-organization in colloidal systems.

Introduction

Non-membrane-bound biomolecular condensates such as P-bodies, nucleoli and stress granules represent an important mode of subcellular organization and have been proposed to assemble via liquid-liquid phase separation (LLPS)[1-5]. Condensates have been reported to be involved in a number of key biological processes such as transcription, signaling and ribosome biogenesis[6 7], and have also been associated with different pathologies including neurodegenerative diseases, diabetes and cancer[8]. Despite their biological significance, the fundamental physicochemical principles behind the formation of biomolecular condensates, observed in the cellular environment for proteins[9 10 11], DNA[12], RNA[13] and their mixtures[14], are poorly understood. In particular, biomolecular condensates are typically enriched in intrinsically disordered proteins (IDPs), which are not only structurally heterogeneous, but also tend to establish transient intermolecular contacts[15]. A central challenge has been to provide a multiscale view of IDP LLPS, connecting the atomistic details of molecular interactions and conformational properties of disordered polypeptides in a crowded environment with the meso/macroscopic features of the LLPS droplets they build.

While an atomistic view of individual proteins in crowded environments is beyond the reach of even the most advanced high-resolution experimental techniques, molecular dynamics (MD) simulations have been developed precisely with this aim in mind[16]. Although they may be limited when it comes to sampling efficiency as compared to different coarse-grained approaches[17 18 19], only MD simulations can provide an accurate picture of protein structure, dynamics, thermodynamics and interactions, as well as of the specific effects of applied mutagenesis at the atomistic level. Recently, several groups have used atomistic simulations to model the dynamics of different peptides such as elastin-like peptide[20] or different fragments of NDDX4[21] in the droplet environment. Such simulations have also been coupled with coarse-grained approaches and experiments, such as in the case of the condensates of FUS and LAF-1 fragments[22] and have provided a detailed view of key interactions behind protein LLPS[23 24 25]. Despite these important advances, however, the study of biomolecular condensates at the atomistic resolution is still at its beginning and many questions, especially those relating to their most generalizable features, are only starting to be addressed[23 24 25]. These in particular concern the complex interplay between structure, dynamics and thermodynamics of biomolecules in crowded environments.

Equally importantly, it is critical to understand how the sub-nanometer- and nanometer-scale properties of biomolecules affect the condensate behavior at the level of complete phase-

separated droplets i.e. on the length scale of tens and hundreds of nanometers and beyond[26]. Polymer physics, in particular, has been widely used to understand the formation of biomolecular condensates[27 3 28 17 18 29 30 31] as this process shares similarities with phase transitions in simple polymer solutions. However, the compositionally complex biological condensates may display a rather different internal organization. In particular, in contrast to the dense, continuous phase observed upon macroscopic phase separation in simple polymers, many protein condensates tend to have a lower density and be enriched in water[10 31 32 33], making them closer in organization to typical colloids[34]. However, due largely to the highly complex nature of biomolecular interactions, the rich concepts and formalisms of colloid chemistry are only now beginning to enter the field of biomolecular condensates[37-39].

Importantly, starting with the seminal work of Forrest and Witten[35] it has been recognized that aggregates or clusters of colloidal particles typically exhibit fractal scaling, i.e. that the cluster mass scales with cluster size according to a non-integer power law with the so-called fractal dimension d_f as the exponent[36]. Here it should be emphasized that colloids generally build *statistical fractals* in which the scaling laws hold between the *average* values of mass and cluster size, in contrast with the regular geometric fractals where this relationship holds absolutely[37 38]. Overall, statistical fractal properties have been demonstrated under different conditions for many non-biological colloidal systems including among others silica, polystyrene and gold colloids[36]. Moreover, using mostly scattering techniques such as static and dynamic light scattering[36 39], but also confocal microscopy and scanning electron microscopy[40], the fractal nature has been associated with several biological colloidal systems, including different protein fibrils[41 42], whey protein isolates[43] and clusters of lysozyme[44] and amylin[40]. To the best of our knowledge, no biomolecular condensates have been analyzed from this perspective. Following the paradigm set by fractal colloidal systems[36], we derive here from first principles a general analytical framework for modelling statistical fractal cluster formation in the context of biomolecules. As the key result, we provide a quantitative connection between two fundamental descriptors of individual biomolecules in a condensed phase, interaction valency and compactness, and the fractal dimension d_f, which in turn enables one to describe the structural organization of the network across multiple length scales.

Here, we apply the above framework to the N-terminal 80-residue fragment of Lge1, a scaffolding protein required for chromatin ubiquitination during transcription[45]. $Lge1_{1\text{-}80}$ was shown to undergo robust LLPS *in vitro* and represents a *bona fide* LLPS system due to its sequence compositional bias (enriched in R, Y and G) (**Fig. 1a**) and a natural tendency to be

fully disordered[45]. We combine a detailed atomistic characterization of a model condensate containing 24 copies of $Lge1_{1-80}$, obtained via microsecond MD simulations (**Extended data 1a**), with our newly developed fractal formalism in order to propose an atomistic model of the $Lge1_{1-80}$ condensate extending to micrometer scale. To probe sequence determinants of $Lge1_{1-80}$ LLPS, we also investigate its all-R-to-K (R>K) mutant, where the ability to form condensates *in vitro* is modulated, and its all-Y-to-A (Y>A) mutant, which impairs LLPS *in vitro*[45] (**Fig. 1b**). Finally, we compare our theoretical predictions against the phase behavior of these different systems as studied experimentally via light microscopy. Our analysis provides a detailed, multiscale model of a biologically relevant condensate and establishes a general, experimentally testable framework for studying other similar systems.

Results

$Lge1_{1-80}$ LLPS critically depends on tyrosine residues To provide the context for computational analyses, we have first experimentally explored the phase diagrams of WT $Lge1_{1-80}$ and its R>K and Y>A mutants by light microscopy. A recombinant $Lge1_{1-80}$ peptide (**Extended data 1b**) undergoes LLPS already at a low C_{sat} protein concentration of 1 μM in a buffer containing 0.1 M NaCl (**Fig. 1c**). LLPS of the WT is robust and can be abolished only at a very high salt concentration (>2 M NaCl). Adding tyrosine amino acids to the solution, which could provide competing π-π interactions, has no significant impact on the WT LLPS in the concentration range tested, while another aromatic agent, imidazole, disrupts WT condensate formation at concentrations above 0.5 M. Mutating all arginine residues in WT $Lge1_{1-80}$ to lysines (R>K mutant) increases the C_{sat} about ten-fold (**Fig. 1c**). Also, the substitution of the arginine guanidium groups, prone to π-π interactions, by the lysine amino groups, which lack sp2 electrons and do not participate in π-π interactions, makes the mutant more sensitive to aromatic agents such as tyrosine (moderate effect) or imidazole (strong effect). At the same time, Y>A mutations abolish LLPS completely, suggesting that tyrosines are critical for LLPS of the WT and R>K (**Fig. 1c**). Interestingly, the Y>A mutant still displays macroscopic cluster formation at 40 μM or higher, but the observed structures lack a spherical droplet shape. Thus, the obtained phase diagrams for the WT and mutants highlight the importance of π-π interactions in LLPS as proposed earlier[46]. These diagrams are also in line with the observation that LLPS is critically related to the number of tyrosine and arginine residues in FUS and other phase-separating proteins[11 30].

To rationalize the biochemical effects of the mutations, we have calculated the interaction strength of selected pairwise contacts (Y-Y, R-Y, K-Y). To this end, the interaction free

energies (ΔG) for these pairs of residues were determined using all-atom Monte-Carlo simulations in different environments (see Methods). Interestingly, the energetic effect for Y-Y interactions is independent of the polarity of the environment (approx. -5 kcal/mol on average), whereas for R-Y and K-Y the interaction free energies display dependence on the dielectric permittivity of the medium (**Extended data 1c**), as shown for other aromatic sidechain analogs (F, W)[47] and nucleobases[48] interacting with these aromatic residues. R-Y and K-Y interaction free energies are lowest in the apolar environment (approx. -10 kcal/mol on average), intermediate in bulk water (approx. -7.5 kcal/mol in average) and highest at intermediate polarity, where they are comparable to Y-Y interactions. Importantly, only in the latter case are the R-Y interactions slightly more favorable than the K-Y ones (**Extended data 1c**). While this can potentially contribute to increasing of C_{sat} for the R>K mutant as compared to WT, the estimated thermodynamic effect is not too strong, especially if one considers that these interactions take place in an environment with largely water-like polarity. Therefore, the effect of R>K substitution on LLPS should be further explored in the context of protein-protein interactions. Moreover, while the Y>A substitution likely causes a strong thermodynamic effect due to the removal of all possible R-Y and Y-Y intermolecular contacts, the mutant can still form aggregates (**Fig. 1c**), with a different morphology as compared to the spherical, droplet-like condensates of WT and R>K.

π-π interactions may shape protein clustering in Lge1$_{1-80}$ condensates To better understand the effects of mutations on the organization of Lge1 condensates, we performed three 1 μs MD simulations with 24 copies each of Lge1$_{1-80}$ WT, Y>A, and R>K mutants in the mM concentration range and control simulations of the three proteins present as single copies (**Extended data 1a**). In the course of these simulations, proteins tend to form clusters and organize into mesh-like configurations (**Fig. 2a**), characterized by pronounced dynamics and heterogeneity. Specifically, the WT 24-copy system forms a single percolating cluster over the last 0.3 μs (**Fig. 2c**) with all copies of the peptide contributing to a mesh. On the other hand, the mutants do not form such an extensive interaction network. The large continuous WT cluster is enriched in π-π interactions (**Fig. 2a, b**) between the most abundant amino acids (R, Y, G), whereby R-Y contacts dominate (8% of all possible pairs). A closer comparison shows that R-Y contacts are similarly enriched in both intramolecular (within the protein molecule in the single-copy system) and intermolecular (between different protein molecules in the multi-copy system) contexts (**Extended data 2a, b**). Conversely, Y-Y and G-G contacts are enriched in both WT and R>K, but only in the intermolecular context (**Extended data 2a**). However,

the opposite is observed for R-R contacts for WT and Y>A (**Extended data 2b**). These results indicate a specific distribution of contacts which undergo significant rewiring when going from intramolecular interactions at low density to intermolecular interactions in crowded systems.

Lge1$_{1-80}$ interaction valency is defined by its sequence composition The interaction valency, defined as the average number of partners per protein molecule, stabilizes in the course of simulations and reaches ~4 partners for WT on average over the last 0.3 µs (**Fig. 2d**), while it is approximately half of that value for the two mutants. Specifically, over the last 0.3 µs of simulation different WT copies have between 1 and 9 interaction partners (**Fig. 2e**). In contrast, multiple non-interacting protein copies are still observed for both mutants over the last part of the simulations and the maximum possible valency does not exceed 6 partners per protein (**Fig. 2e**). Thus, mutations with impaired LLPS, as detected experimentally, exhibit a lower valency of protein-protein interactions. These interactions of IDPs are characterized by a *dynamic binding mode* where the interacting sequence motifs ("stickers") and non-interacting sequence motifs ("spacers")[18] can be described only statistically and no well-defined, specific structural organization of protein complexes is detected (**Fig. 3a**). For WT, the typical binding mode over the last 0.3 µs corresponds to the average valency of 4, where different partners bind along the sequence via highly interacting sequence regions (**Fig. 3a, b**). In line with the valency decrease, both mutants display fewer prominent interaction motifs, whereby for the LLPS-disruptive Y>A mutant intermolecular binding is detected for the C-terminal part of the molecule only (**Fig. 3a**). Interestingly, the dynamic modes identified in the intramolecular context are clearly different for all modeled proteins (**Extended data 2c**), which is also observed at the level of pairwise contact statistics (**Extended data 2a, b**).

Alteration of Lge1$_{1-80}$ sequence impacts its conformational behavior The perturbation of intra- and inter-molecular interaction networks by mutations results in a different conformational behavior of the resulting protein variants. At the level of single molecules, extensive interactions involving R and Y residues result in a substantial compaction of the WT molecule with $\langle Rg_{MD} \rangle = 1.58 \pm 0.12$ nm (**Fig. 4a**). The effect is sequence specific and cannot be attributed to force-field artifacts since for the R>K and Y>A mutants the Rg_{MD} distributions cover a wider range and display higher average values (1.81 and 1.71 nm, respectively, **Fig. 4a**). This in all three cases is more compact as compared with the predictions for a random coil molecule of the same length $\langle Rg_{rc} \rangle = 2.50$ nm[49] (see Methods). On the other hand, the observed difference between WT and the two mutants is directly related to the extent of intramolecular interactions: the looser the interaction network (i.e. fewer long-range sequence

contacts), the larger the $<Rg>$ (see **Extended data 3a** for the illustration). In the crowded environment, WT again adopts a compact organization with almost the same $<Rg_{MD}>$ of 1.60 nm when averaged over all 24 copies (**Fig. 4b**), while the weakly self-interacting Y>A mutant unwinds toward more extended conformations ($<Rg_{MD}>$ = 1.97 nm). At the same time, R>K, being the loosest in the single molecule context, comes closer to WT values in the crowded environment ($<Rg_{MD}>$ = 1.67 nm, **Fig. 4b**). The latter can potentially be explained by the repulsive nature of K-K contacts, which are enriched in the single-molecule context and depleted in the crowded phase (**Extended data 2b**). Note here that, unlike K-K pairs, R-R pairs can engage in stacking interactions. These results highlight the fact that the conformational behavior of IDPs in the bulk or in the crowded phase displays a clear sequence-specific character and cannot easily be generalized. Moreover, WT and R>K, which are capable of LLPS, exhibit similar compactness in the crowed environment.

The effect of R and Y mutations on conformational dynamics can be quantified by estimating the configurational entropy change (ΔS_{conf}) via the MIST formalism (see Methods). ΔS_{conf} displays a reasonable convergence between individual replicas on the 1-µs time scale (<0.1 kJ/mol/K), especially in the case of the weakly self-interacting Y>A mutant (**Extended data 3b**). Interestingly, ΔS_{conf} decreases the least in the crowded environment for Y>A and the most for WT (**Fig. 4c**). Furthermore, there is a substantial increase in ΔS_{conf} associated with the LLPS disruptive Y>A mutation in both the single-molecule and the multi-copy contexts, in contrast with the R>K mutation, where a much weaker effect is seen (**Fig. 4d**). These results suggest that the crowding involves a conformational reorganization of the molecules toward decreasing the available free volume, i.e. increasing their compactness or volume fraction φ, defined as the ratio between the van-der-Waals and the hydrodynamic volume of a molecule. Importantly, the increase in φ upon crowding directly correlates with an unfavorable ΔS_{conf} (**Fig. 4e**, Pearson R is -0.6), an effect which can potentially be compensated for by the favorable enthalpy upon forming the extensive multivalent interaction network as observed for the WT. Conversely, ΔS_{conf} correlates poorly with the average number of bound partners (valency; Pearson R = -0.3) and there is no correlation observed between the compactness φ and the valency of interactions. Being mutually largely independent, these two characteristics of protein molecules – compactness and valency – are the key parameters describing the organization of the corresponding crowded phase, which generally reflect the "entropic" and the "enthalpic" contributions to self-organization, respectively.

Describing condensate architecture via a fractal scaling model To extrapolate the atomistic-level properties of the crowded protein phase to larger length scales, we modeled the assembly of LLPS droplets as an iterative, fractal process (**Fig. 5a**, see also **Appendix**). Colloid fractal models typically start with an ansatz capturing the power-law dependence between mass and size across different scales[50 36]. For conceptual clarity and to demonstrate where the power-law dependence comes from, our derivation starts from a simple physical picture of associating clusters and yields the known scaling relationship bottom-up. Thus, we assume that initially individual protein molecules, characterized by a given volume fraction φ, interact to form clusters of a given average valency n. In the next iteration, these clusters arrange into higher-order clusters, whereby the average valency of each cluster and the volume fraction occupied by the clusters formed in the previous iteration remain the same at all levels of organization. For instance, assuming that a single protein molecule binds four other proteins (n = 4), results in the second-iteration cluster consisting of 5 protein molecules (**Fig. 5a**). In the next iteration, 5-mers arrange into a new cluster with the same valency of 4, resulting in a larger cluster with 25 molecules. According to this model, in each ith iteration $i-1^{\text{th}}$ clusters occupy only φ of the apparent condensate volume (e.g. 2/3 as shown in **Fig. 5a**). This scenario results in a simple fractal formalism whose benefit is that it yields discrete distributions and exact solutions at each iteration (see **Appendix**). Thus, for each iteration i^{th}, the formalism returns the number of molecules (eq. 1), the apparent volume (V_i, eq. 4), the size (R_i, eq. 5), the mass (M_i, eq. 8) and the effective molar concentration of molecules in the condensate (C_i, eq. 6). While the above derivation for simplicity uses the constant volume fraction and valency at each iteration, we want to emphasize that in the case of structurally heterogeneous statistical fractal clusters, these parameters would be equivalent to their time- and ensemble-averaged counterparts.

To compare the model predictions with the actual condensation process, one can use the results obtained from all-atom MD simulations. Thus, over the last 0.3 μs of MD simulations, the WT crowded system displays a narrow distribution of sizes for the single molecule (**Fig. 4b**) and leads to a single percolating cluster (**Fig. 2d, Extended data 3c**) with an average radius *R*. Importantly, these two data points agree closely with the predictions of the analytical model when the average values for φ and n parameters obtained over the last 0.3 μs of MD simulations are used (**Extended data 3d**). Thus, the 24-mer observed in MD directly corresponds to the 3rd iteration of the model. Moreover, the slope (*A,* eq. 10) and the intercept (*B,* eq. 11) of the linear regression for the log *R* vs log *Mw* plots are also similar between the MD and the model (2.33 vs 2.42 and 1.13 vs. 1.11, respectively). Finally, the latter parameters

allow the exact calculation of the characteristic φ and n values (eq. 12 and 13; 0.638 and 3.76, respectively), which are very similar to those obtained directly from all-atom MD **(Extended data 3d)**. Notably, the characteristic valency of the interactions obtained from the *M(R)* scaling for the 1-mer and 24-mer MD simulations corresponds directly to the statistically defined valency obtained from MD data (see above).

The model also enables one to explore the space of φ and n parameters for a condensate with a fixed size and concentration (eq. 7, **Fig. 5b**). For instance, for a 1-µm condensate with a 1-mM apparent protein concentration, the corresponding φ and n values are generally in the range observed in MD for the WT crowded system. Generally, in order to keep the ratio of size-to-concentration fixed, the valency must increase non-linearly with decreasing protein compactness (**Fig. 5b,** dashed lines). Accordingly, perturbation of compactness and valency due to the two types of studied mutations can result in a decrease of the apparent concentration in condensates of fixed size. Importantly, the compactness of IDPs is not a stable parameter and is tunable by different factors (temperature, pH, ionic strength, etc) [51]. This, in turn, also suggests that IDP concentration inside condensates may also be adaptable and tunable. Thus, unwinding or compactization of an IDP due to any factor would change the apparent concentration and density in the condensates. The latter also suggests potential microphase transitions in LLPS droplets in analogy to that known for lipid membranes[52]. Finally, the complex topology (see below) of the condensate proposed by the model allows for the formation of droplets with a low apparent concentration.

Valency and compactness define fractal dimension and condensate topology The fractal model provides a direct relationship between size and mass of different clusters which directly capture the scaling behavior of the condensate matter. Thus, the slope of the line in the log R vs. log Mw plot (A) is equal to the fractal dimension d_f, which describes the topology of molecular clusters[50]. The fractal dimensions equaling exactly 1, 2 or 3 correspond to the objects exhibiting 1D, 2D or 3D organization, respectively, while systems with non-integer d_f have an intermediate dimensionality. Therefore, the closer d_f is to 3, the more prominently the condensate propagates in volume and the more compact it is, a fact that may be associated with the robustness of LLPS for a given agent. The proposed model facilitates a direct investigation of the scaling properties in condensates for a molecule with defined characteristic compactness and valency of interactions. Most importantly, the fractal dimension of a condensates is completely defined by φ and n (eq. 10), reflecting the predictive potential of the proposed model. For instance, the average φ and n parameters derived from MD simulations of the

crowded state for WT, R>K and Y>A proteins result in a different scaling behavior for these $Lge1_{1-80}$ variants, where dimensionality decreases respectively (**Fig. 5c**) and results in a different morphology of the corresponding condensates (**Fig. 6b**), as observed experimentally (**Fig. 1b, c**). Specifically, while the $Lge1_{1-80}$ WT condensate simulated here exhibits a d_f of 2.42 and is also experimentally found to undergo robust LLPS, the R>K mutant exhibits a lower d_f of 2.08 and undergoes LLPS under a more limited set of conditions. Even more extremely, d_f is 1.63 for the Y>A mutant, which does not undergo phase separation and can be thought of as an object between 1D and 2D.

Reconstructing the condensate architecture across scales The fractal formalism described here allows one to describe the structure of the condensate at any scale, starting just from the values of φ and n, determined at the atomistic resolution. In fact, there are algorithms in the colloid literature for reconstructing the geometry of a fractal cluster with a given fractal dimension d_f[53] [54] [55] (and prefactor k_f, which is equal 1 in the present model; see **Appendix** for derivation) , where d_f is uniquely determined by φ and n. Recently, Moran et al.[55] have proposed a robust and tunable algorithm, FracVAL, for modeling the formation of clusters consisting of polydisperse primary particles. We have used the value of d_f derived from our simulations of the condensates of $Lge1_{1-80}$ WT and its R>K and Y>A mutants in combination with the FracVAL algorithm to generate individual realizations of the respective condensate structure on the submicron length scale. In **Fig. 6a**, we demonstrate this procedure for WT $Lge1_{1-80}$: FracVAL produces cluster geometries using spherical particles with radii corresponding to the respective $<Rg_{MD}>$ values, which are then computationally replaced by the realistic protein conformations obtained from our simulations. In agreement with its micrometer-scale behavior observed in the experiment, the modelled WT condensate exhibits a densely packed fractal geometry. In contrast, the reconstructed Y>A cluster exhibits an extremely elongated, filamentous topology (**Fig. 6b**), consistent with the absence of well-defined LLPS droplets and/or presence of irregular aggregates seen experimentally (**Fig. 1c**). The R>K cluster, in turn, exhibits an intermediate topology. The clusters shown in **Fig. 6b** were generated using 1024 primary particles in all three cases: clearly, the Y>A cluster occupies a significantly larger volume as compared to the R>K cluster and especially the WT. While all three reconstructed systems exhibit holes and cavities of different sizes, it is evident how the Y>A system may not appear as spherical droplet-like in light microscopy experiments. Finally, note that the three reconstructions shown in **Fig. 6a** and **b** are individual snapshots of the local architecture; the

full fractal model entails an ensemble of such snapshots, all different, yet all still conforming to the same scaling pattern.

Discussion

We have here provided a multiscale description of an LLPS condensate extending from a detailed description of its constituent biomolecules at the atomistic level all the way to the level of complete, micrometer-sized droplets involving thousands of individual molecules. We have shown that mutations of $Lge1_{1-80}$ R and Y residues induce perturbations at the level of both intra- and intermolecular interaction networks and result in conformational effects that can be related to the phase behavior and the ability to form LLPS condensates as detected by light microscopy. Specifically, the characteristic descriptors of protein behavior in the crowded phase, valency and compactness (volume fraction), are shown to be sufficient to describe LLPS condensate assembly and structural organization across scales in the context of the proposed analytical fractal model. Importantly, the studied mutations substantially change either just the valency (R>K) or both the valency and compactness (Y>A) of the $Lge1_{1-80}$ peptide as shown in all-atom MD simulations in the microsecond range. These changes of valency and compactness then translate to different scaling behavior and, subsequently, topology of the condensates as suggested by the fractal model. Brangwynne and coworkers have recently successfully adopted the theoretical formalism of patchy colloids to capture the relative contributions of oligomerization, RNA binding and structural disorder in the formation of stress granules and P-bodies[56]. In their analysis, they emphasized the role of the valency of colloid particles as a key parameter defining specificity and tunability of condensate features. Furthermore, Collepardo-Guevara and coworkers have used a coarse-grained patchy-particle colloid model to study the determinants of condensate stability and have highlighted the role of valency, while also clearly demonstrating the general impact of volume fraction [57]. Our study now provides a general framework for assessing the importance of these two parameters on LLPS in biological systems.

Colloidal cluster formation is typically discussed in the context of two limiting regimes[39 58 36]. In diffusion-limited cluster aggregation (DLCA), the rate of cluster formation is determined by the time it takes for colloidal particles to encounter each other and every encounter leads to binding. In reaction-limited cluster aggregation (RLCA), particles need to overcome a repulsive barrier before binding and not every encounter is productive. Both regimes result in fractal behavior, with DLCA leading to looser structural organization and lower fractal dimensions and RLCA leading to more compact structural organization and higher fractal

dimensions. Computer simulations and scattering experiments show that d_f is ~1.8 for DLCA and ~2.1 for RLCA[39 36] . Importantly, Meakin and colleagues have shown that the two regimes of colloid cluster formation are universal and do not depend on the chemical nature of the underlying particles, making them an attractive paradigm for modelling the multiscale structure of biomolecular condensates as applied here[39]. However, for flexible, multivalent molecules like IDPs, the exact regime of cluster formation would depend on the values of φ and n and may be a tunable feature of the exact conditions. In **Fig. 6c**, we present the non-linear dependence of the fractal dimension d_f on φ and n, which still exhibits certain general trends. For example, high values of either compactness φ or valency n or both, such as in the case of WT $Lge1_{1\text{-}80}$, result in high values of d_f and may be more associated with the RLCA model, while low values of both parameters may more be associated with the DLCA model. A more quantitative analysis of the connection between the exact mechanism of cluster formation and the underlying parameters of compactness and valency will be a topic of future work.

The fractal formalism also provides an elegant framework for understanding some particular aspects related to the question of the concentration of biomolecules inside of phase-separated condensates. Namely, the measured concentrations of biomolecules inside of different condensates studied *in vitro* cover a large range extending to tens of mM and beyond[28 23 59]. However, some proteins form droplets with extremely low concentrations in the dense phase. For example, the measured binodals of LAF-1 indicated that the concentration of the protein inside the droplet is 86.5 μM, which corresponds to an average separation between molecular centers of mass of 27 nm[10]. Considering that the average *Rg* of LAF-1 is approximately 4 nm, it is not immediately clear how the molecules inside the droplet are organized in order to simultaneously establish intermolecular contacts and also form low-density droplets. Fractal organization, as proposed here, provides a simple resolution of this apparent conundrum. Namely, fractal systems are characterized by a remarkable property that their packing density is a function of the length-scale on which it is examined[37]. For example, the density of a Sierpinski gasket, which is self-similar on all length scales, *decreases* exponentially with the length scale, with the exponent of $d_f - d$, where d is the dimensionality of the space. Translated to the question of biomolecular condensates, fractal organization enables high local concentration of biomolecules at short length scales at all locations in space and, simultaneously, low global concentration at long length scales, micrometer-droplet scale, as also illustrated in **Fig. 5b**. Importantly, fractal systems are characterized by the existence of holes, i.e. unoccupied regions of space, the size of which covers a wide range of length scales[37].

Our observation that the µm-sized WT $Lge1_{1-80}$ droplets are fully permeable to dextrans in vitro, even up to Mw = 2000 kDa or *Rg* ~ 27 nm[60] (**Extended data 3e, f**), suggests that their organization must allow for large holes (~55 nm in diameter or more). This may be consistent with the fractal model suggesting that the dimensionality of WT condensates is below 3 (d_f = 2.42, **Fig. 6**). Finally, the multiscale nature of LLPS biomolecular condensates, as embodied in the statistical fractal model, also points to the possible formation of LLPS nanodroplets with sizes below the resolution limit of light microscopy. The latter suggests that LLPS can be even more widespread than previously suspected.

Here, it should be emphasized that the proposed fractal framework is a model, which may be applicable to varying degrees in different systems. In particular, it was shown that *in vitro* Lge1 forms a liquid-like core and the E3 ubiquitin ligase Bre1 forms an enzymatic outer shell[45]. Whereas such condensates can exhibit a diameter of 1 µm or more when grown in vitro, they are expected to be much smaller in cells (i.e. low nm range), capable of encapsulating a gene body that is occupied by a few nucleosomes[45]. Hence, the relevance of the proposed dynamic fractal model will need to be tested experimentally with respect to nm-sized core-shell condensates in cells. Moreover, future studies must include the spherical Bre1 shell as a boundary condition which presumably constrains the 3D-orientation of the Lge1 C-terminus and thereby impinges on the geometry of the Lge1 meshwork. Finally, the fractal model predicts a coexistence of differently sized clusters within a condensate, as reported recently[61], which have a characteristic scaling of mass with condensate size in the nm to µm range. It will be important to analyze to what extent these features are retained in more complex, biologically relevant contexts.

Overall, our results provide an atomistically realistic framework for understanding the role of valency and compactness of IDPs on condensate stability and topology across scales. This presents a potentially universal foundation for the rational design of LLPS agents with predefined condensate properties. Indeed, recent studies have demonstrated the possibility to tune condensate properties *in vitro* and *in vivo* by manipulating the aromatic residue content and molecular weight in IDPs[30] or the size of disordered linkers and valency in modular proteins[62]. Finally, it is our hope that this and similar efforts may help to critically embed the field of biomolecular condensates in the wider context of colloid chemistry. We expect that the powerful theoretical, computational and experimental tools of colloid chemistry could propel the study of biomolecular condensates to an even higher level of fundamental understanding.

Methods

Protein expression and purification All proteins were expressed in *Escherichia coli* BL21 CodonPlus (DE3) RIL cells. 6His-Lge1(1-80) -StrepII constructs were induced by addition of 0.5 mM isopropyl 1-thio-β-D-galactopyranoside at OD_{600} = 0.8 at 23 °C for 3 h and purified as published elsewhere[45] in a final elution buffer (10 mM Tris, 1 M NaCl, 1 mM TCEP, 1 M imidazole, 10 % vol/vol glycerol, pH 7.5) and stored at -80 °C. Protein quality was assessed by SDS-Page (4-12 % gel, MOPS buffer) and Coomassie staining.

Phase diagrams Different concentrations of purified 6His-Lge1(1-80)-StrepII proteins in a total volume of 100 µL were added to a bottom-clear 96-well plate (Greiner Bio-One) in a buffer containing 25 mM Tris, 100 mM NaCl, 100 mM imidazole, 1 % vol/vol glycerol and 1 mM DTT, pH 7.5. Varying concentrations of NaCl (100 mM to 3 M), imidazole (100 mM to 2.25 M) or tyrosine (0,25 mM to 2.5 mM) were analyzed as indicated in Figure 1. Plates with the protein mix were incubated for 10 min at 20 °C. Turbidity of the samples was measured at 450 nm in a Victor Nivo™ plate reader (Perkin Elmer) at 20 °C. Assessment of protein phase separation or aggregation was performed by applying a total volume of 20 µL of the sample to a pretreated[45] 16-well glass-bottom ChamberSLIP slide (Grace, BioLabs). DIC imaging was performed as previously described[45].

Dextran experiments Experiments were performed as published elsewhere[45]. Dextrans with a final concentration of 0.05 mg/ml were added to the samples containing 6His-Lge1 (1-80)-StrepII (2 µM) in a final buffer containing 25 mM Tris, 100 mM NaCl, 100 mM imidazole, 1% vol/vol glycerol and 1 mM DTT, pH 7.5. Samples were incubated for 15 min at 20 °C on pretreated 16-well glass-bottom ChamberSLIP slides prior to imaging.

Pairwise interaction free energy calculations Pairwise interaction free energies were calculated using all-atom Monte Carlo (MC) simulations according to the previously established framework[47]. MC simulations were carried out in TIP4P water[63], methanol, dimethylsulfoxide and chloroform for the sidechain analogs of tyrosine, arginine, and lysine using OPLS force field[64]. Initial structures of the molecules were optimized *in vacuo* using AM1 semiempirical molecular orbital method[65] and placed in rectangular boxes with explicit solvent. All calculations were performed with the BOSS 4.2 program[66] with periodic boundary conditions in the NPT ensemble at 298 K and 1 bar. Standard procedures were employed including Metropolis criterion and preferential sampling for the solutes[67]. The potential of mean force (pmf) computations for Y-Y, Y-K and Y-R pairs were performed by gradually moving the

solute molecules apart in steps of 0.05 Å along an axis defined by the particular atoms or centers of geometry, while both solute molecules were allowed to rotate around this axis. Non-bonded interactions were truncated with spherical cutoffs of 12 Å. The free energy changes for a particular pair were calculated in a series of consecutive MC simulations using statistical perturbation theory and double-wide sampling[67]. Each simulation consisted of 3×10^6 configurations used for equilibration, followed by 6×10^6 configurations used for averaging.

Molecular dynamics (MD) simulations The full-length structure of Lge1 was modeled *de novo* using Phyre2 web portal[68]. The disordered N-terminal 1-80 aa fragment of Lge1 ($Lge1_{1-80}$) in the modeled structure lacked any secondary structure and was used as an initial configuration in further Lge1 simulations. All-atom MD simulations were performed for WT $Lge1_{1-80}$ fragment as well as for its Y>A and R>K mutants. Single-protein-copy MD simulations were carried out in 9 x 9 x 9 nm^3 water boxes using 2 independent 1-µs replicas for each of the three $Lge1_{1-80}$ variants. The same initial configuration (see above) was used for all three. All systems had zero net charge and effective NaCl concentration of 0.1 M. Systems with 24 protein copies were simulated in 18 x 18 x 18 nm^3 (WT, Y>A) or 19 x 19 x 19 nm^3 (R>K) water boxes. Initial configurations for these simulations were generated as follows. Four different protein conformers were selected from the initial 100 ns parts of the two independent single-copy MD simulations (two conformations from each run) based on the criteria of having been the centers of the most highly populated clusters after clustering analysis performed using *cluster* utility (GROMACS) with the applied RMSD cut-off for backbone atoms of neighboring structures of 1.5 Å. The cells containing 4 copies were assembled manually and translated 6 times in different directions, resulting in a protein grid containing 24 protein copies. A total of 1 µs of MD statistics were collected for each large system. All MD simulations and the analysis were performed using GROMACS 5.1.4 package[69] and Amber99SB-ILDN force field[70]. After initial energy minimization, all systems were solvated in an explicit aqueous solvent using TIP4P-D[71] water model, which was optimized for simulating IDPs. The final NaCl concentration was 0.1 M (**Extended data 1a**). The solvated systems were again energy-minimized and subjected to an MD equilibration of 30000 steps using a 0.5-fs time step with position restraints applied to all protein atoms (restraining force constants $Fx = Fy = Fz = 1000$ kJ mol^{-1} nm^{-1}) and 250000 steps using a 1-fs time step without any restraints. Finally, production runs were carried out for all systems using a 2-fs time step. A twin-range (10/12 Å) spherical cut-off function was used to truncate van der Waals interactions. Electrostatic interactions were treated using the particle-mesh Ewald summation with a real space cutoff 12 and 1.2 Å grid with fourth-order spline

interpolation. MD simulations were carried out using 3D periodic boundary conditions in the isothermal−isobaric (NPT) ensemble with an isotropic pressure of 1.013 bar and a constant temperature of 310 K. The pressure and temperature were controlled using Nose-Hoover thermostat[72] and a Parrinello-Rahman barostat[73] with 0.5 and 10 ps relaxation parameters, respectively, and a compressibility of 4.5×10^{-5} bar^{-1} for the barostat. Protein and solvent molecules were coupled to both thermostat and barostat separately. Bond lengths were constrained using LINCS[74].

Radii of gyrations (*Rg*) of simulated proteins were calculated using GROMACS *gyrate* utility, respectively. The average number of interaction partners per protein and the detailed statistics of intermolecular contacts were evaluated using GROMACS *mindist* and *pairdist* utilities with an applied distance cutoff of 3.5 Å, respectively, while intramolecular contact maps were generated using *mdmat* utility. Protein structures were visualized using PyMol[75].

Theoretical estimate of Rg The estimation was done according to the scaling model used in the polymer theory to connect *Rg* and the chain length of a random coil as follows:

$$Rg = R_0 N^{\upsilon},$$

where N is the length of $Lge1_{1\text{-}80}$ (N = 80) and R_0 and υ are the empirical parameters refined for IDPs (R_0 = 0.254 nm; υ = 0.522)[49].

Cluster analysis The largest protein-protein interaction clusters in the 24-copy simulated system were identified using hierarchical clustering. For this purpose, minimum-distance matrices were calculated from each MD trajectory sampled at every 100 ps using GROMACS *mindist.* The clustering was done in MATLAB (R2009) using function *cluster* with applied distance cutoff of 3.5 Å.

Entropy calculations The configurational entropy was evaluated by applying the maximum information spanning tree (MIST) approximation[76] using the PARENT suite[77], a collection of programs for the computation-intensive estimation of configurational entropy by information theoretical approaches on parallel architectures. All MD trajectories were first converted from Cartesian to Bond-Angle-Torsion (BAT) coordinates. To assess the convergence of configurational entropy (S_{conf}), cumulative plots were generated for single copy systems using a 50-ns time step. Due to the relatively slow convergence of the entropy values (**Extended data 3b**), the final entropy calculations were done for the entire 1-µs trajectories. Note that the absolute S_{conf} values are negative and carry arbitrary units due to the exclusion from the calculations of the constant momentum part of the configurational entropy integrals and are

reported just to illustrate the convergence of the entropy values as a function of simulated time. However, upon subtraction of these absolute values (i. e. for single and crowded systems), the relative entropy (ΔS_{conf}) carries correct physical units and is equal to the total configurational entropy change between the two systems. The relative entropies of a protein in single-copy and multi-copy systems were averaged over all possible combinations of the two single-protein copies and the 24 crowded-system copies. To estimate the effect of mutations on configurational heterogeneity, the corresponding entropy differences were calculated as follows:

$$\Delta S_{conf}^{mut} = \left[\frac{S_{conf}^{mut}}{(3N_{mut} - 6)} - \frac{S_{conf}^{WT}}{(3N_{WT} - 6)}\right](3N_{mut} - 6)$$

where N_{mut} and N_{WT} represent the numbers of atoms in mutant and WT proteins, respectively. The final values of configurational entropy differences were multiplied by the temperature (T = 310 K) and converted to kcal/mol units.

Modeling and visualization of condensates topology To generate a visual representation of a condensate with a fractal dimension obtained by the self-propagation model (see **Appendix**), FracVAL algorithm was used[55]. FracVAL is a tunable algorithm for generation of fractal structures of aggregates of polydisperse primary particles, which preserves the predefined fractal dimension (d_f) and the fractal prefactor (k_f) to generate aggregates of desired size. The scaling law in this case can be defined as shown in eq. 14. The prefactor k_f is equal to 1 in the present model (see **Appendix**). The d_f values for generating cluster models in FracVAL were calculated according to eq. 10 using the averaged φ and n values over the last 0.3 μs of the 24-copy MD simulations, while $\langle Rg_{MD} \rangle$ values over the last 0.3 μs were taken as effective sizes of primary particles (detailed parameters are listed in **Extended data 4**). For visualization purpose the size of condensates generated by FracVAL was defined as 1024 molecules. FracVAL cluster models were transformed to all-atom resolution by using selected protein MD conformations with the respective *Rg* values and scripts specially written for this purpose. The obtained coarse-grained and all-atom structures were visualized using PyMol[75].

Appendix

Self-propagation model of condensate assembly The number of molecules in a cluster at iteration i can be calculated according to a simple geometrical progression:

$$(n+1)^{i-1} \quad (1)$$

where n is the valency of interactions or more generally – the coordination number. An apparent volume v of a single molecule, whereby atoms occupy the fraction φ of that volume, can be expressed as:

$$v = \frac{V_{mol}}{\varphi} \quad (2)$$

where V_{mol} is the molecular volume, which in turn is proportional to the molecular weight (M_W) with a factor κ ($\kappa = 1.21$ used for all calculations[78]):

$$V_{mol} = \kappa M_W \quad (3)$$

Thus, combination of (1) and (2) allows to estimate an apparent volume of condensate (V_i) at iteration i:

$$V_i = \frac{(n+1)^{i-1}}{\varphi^i} V_{mol} \quad (4)$$

Under the assumption of a spherical geometry, a characteristic size of the condensate at iteration i (R_i) can be derived from (4):

$$R_i = \sqrt[3]{\frac{3}{4} V_{mol} \frac{(n+1)^{i-1}}{\pi \varphi^i}} \quad (5)$$

An effective molar concentration (C_i) of molecules in a condensate at iteration i then can be estimated as:

$$C_i = \frac{\varphi^i}{V_{mol} N_a} 10^{-3} m^3 \quad (6)$$

where, N_a is the Avogadro number. A combination of (5) and (6) gives the volume fraction as a function of valency for a condensate with a given concentration (C) and size (R):

$$\varphi(n) = e^{1+\frac{\frac{\ln(V_{mol} N_a 10^3 m^{-3})}{\ln(\frac{4}{3}\pi R^3 C N_a 10^3 m^{-3})}}{\ln(n+1)}} \quad (7)$$

The above formalism allows to extract the characteristic parameters φ and n from the linear regression of an empirical log R vs. log M plot. First, the mass M_i of a condensate at iteration i is given as:

$$M_i = M_W (n+1)^{i-1} \quad (8)$$

A combination of (5) and (8) then gives analytical expressions for a slope (A) and an intercept (B) for the linear regression plot:

$$\log M = A \log R + B \quad (9)$$

$$A = \frac{3}{1 - \frac{\log \varphi}{\log(n+1)}} \quad (10)$$

$$B = \frac{\log \frac{4\pi}{3\kappa} - \log M_W \frac{\log \varphi}{\log(n+1)} + \log \varphi}{1 - \frac{\log \varphi}{\log(n+1)}} \quad (11)$$

Finally, a combination of (10) and (11) gives equations for φ and n using the slope (A) and the intercept (B) for the linear regression (9):

$$\varphi = e^{\frac{3B + (A-3)\log M_W}{A} - \log \frac{4\pi}{3\kappa}} \quad (12)$$

$$n = \varphi^{\frac{A}{A-3}} - 1 \quad (13)$$

Note that A, also known as the "fractal dimension" (d_f)[50] is the key parameter describing the condensate organization. The scaling law in this case is typically defined as:

$$N = k_f \left(\frac{R}{R_g}\right)^{d_f} \quad (14)$$

where N is the number of molecules in each cluster with the corresponding size R, while R_g is an effective size of an individual molecule, and k_f is the fractal prefactor. In the present model of condensate assembly, k_f is equal to 1, as proven below. Thus, according to (14) the number of molecules for a cluster obtained at iteration i can be calculated as:

$$N_i = k_f \left(\frac{R_i}{R_1}\right)^{d_f}$$

where d_f is a slope (A) for log R vs. log M linear regression and is defined in (10). A combination of (1) and (5) gives:

$$N_i = (n+1)^{i-1}$$

$$R_i = \sqrt[3]{\frac{3}{4} V_{mol} \frac{(n+1)^{i-1}}{\pi \varphi^i}}$$

$$R_1 = \sqrt[3]{\frac{3}{4} \frac{V_{mol}}{\pi \varphi}}$$

$$\frac{R_i}{R_1} = \sqrt[3]{\frac{(n+1)^{i-1}}{\varphi^{i-1}}} = \left(\frac{(n+1)^{i-1}}{\varphi^{i-1}}\right)^{\frac{1}{3}}$$

then simplification leads to:

$$(n+1)^{i-1} = k_f \left(\frac{(n+1)^{i-1}}{\varphi^{i-1}}\right)^{\frac{A}{3}}$$

$$k_f = \frac{\left(\varphi^{i-1}\right)^{\frac{A}{3}}}{\left((n+1)^{i-1}\right)^{\frac{A}{3}-1}}$$

and by using (13) to substitute $n+1$ one derives:

$$k_f = \frac{\left(\varphi^{i-1}\right)^{\frac{A}{3}}}{\left(\left(\varphi^{\frac{A}{A-3}}\right)^{i-1}\right)^{\frac{A}{3}-1}}$$

$$k_f = \frac{\left(\varphi^{i-1}\right)^{\frac{A}{3}}}{\left(\varphi^{i-1}\right)^{\frac{A-3}{3}\frac{A}{A-3}}}$$

$$k_f = \frac{\left(\varphi^{i-1}\right)^{\frac{A}{3}}}{\left(\varphi^{i-1}\right)^{\frac{A}{3}}}$$

$$k_f \equiv 1$$

References

1. Brangwynne, C.P. et al. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **324**, 1729-1732 (2009).
2. Lafontaine, D.L.J., Riback, J.A., Bascetin, R. & Brangwynne, C.P. The nucleolus as a multiphase liquid condensate. *Nature Reviews Molecular Cell Biology* **22**, 165-182 (2021).
3. Mitrea, D.M. & Kriwacki, R.W. Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling* **14**, 1 (2016).
4. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123-33 (2015).
5. Feng, Z., Chen, X., Wu, X. & Zhang, M. Formation of biological condensates via phase separation: Characteristics, analytical methods, and physiological implications. *J Biol Chem* **294**, 14823-14835 (2019).
6. Banani, S.F., Lee, H.O., Hyman, A.A. & Rosen, M.K. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
7. Boeynaems, S. et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**, 420-435 (2018).
8. Alberti, S. & Dormann, D. Liquid-Liquid Phase Separation in Disease. *Annu Rev Genet* **53**, 171-194 (2019).
9. Nott, T.J. et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell* **57**, 936-947 (2015).
10. Wei, M.T. et al. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat Chem* **9**, 1118-1125 (2017).
11. Wang, J. et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699 e16 (2018).
12. King, J.T. & Shakya, A. Phase separation of DNA: From past to present. *Biophysical Journal* **120**, 1139-1149 (2021).
13. Jain, A. & Vale, R.D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243-247 (2017).
14. Garcia-Jove Navarro, M. et al. RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nature Communications* **10**, 3230 (2019).

15. Uversky, V.N. Recent Developments in the Field of Intrinsically Disordered Proteins: Intrinsic Disorder–Based Emergence in Cellular Biology in Light of the Physiological and Pathological Liquid–Liquid Phase Transitions. *Annual Review of Biophysics* **50**, 135-156 (2021).
16. Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H. & Shaw, D.E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics* **41**, 429-452 (2012).
17. Dignon, G.L., Zheng, W., Best, R.B., Kim, Y.C. & Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc Natl Acad Sci U S A* **115**, 9929-9934 (2018).
18. Martin, E.W. et al. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694-699 (2020).
19. Benayad, Z., von Bülow, S., Stelzl, L.S. & Hummer, G. Simulation of FUS Protein Condensates with an Adapted Coarse-Grained Model. *Journal of Chemical Theory and Computation* **17**, 525-537 (2021).
20. Rauscher, S. & Pomes, R. The liquid structure of elastin. *Elife* **6**(2017).
21. Paloni, M., Bailly, R., Ciandrini, L. & Barducci, A. Unraveling Molecular Interactions in Liquid-Liquid Phase Separation of Disordered Proteins by Atomistic Simulations. *J Phys Chem B* **124**, 9009-9016 (2020).
22. Zheng, W. et al. Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations. *J Phys Chem B* **124**, 11671-11679 (2020).
23. Ryan, V.H. et al. Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol Cell* **69**, 465-479 e7 (2018).
24. Murthy, A.C. et al. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat Struct Mol Biol* **26**, 637-648 (2019).
25. Conicella, A.E. et al. TDP-43 alpha-helical structure tunes liquid-liquid phase separation and function. *Proc Natl Acad Sci U S A* **117**, 5883-5894 (2020).
26. McSwiggen, D.T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes & development* **33**, 1619-1634 (2019).
27. Brangwynne, Clifford P., Tompa, P. & Pappu, Rohit V. Polymer physics of intracellular phase transitions. *Nature Physics* **11**, 899-904 (2015).

28. Brady, J.P. et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences* **114**, E8194-E8203 (2017).
29. Guillén-Boixet, J. et al. RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346-361.e17 (2020).
30. Dzuricky, M., Rogers, B.A., Shahid, A., Cremer, P.S. & Chilkoti, A. De novo engineering of intracellular condensates using artificial disordered proteins. *Nature Chemistry* **12**, 814-825 (2020).
31. Alberti, S. & Hyman, A.A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nature Reviews Molecular Cell Biology* **22**, 196-213 (2021).
32. Keating, C.D. & Pappu, R.V. Liquid–Liquid Phase Separation: A Widespread and Versatile Way to Organize Aqueous Solutions. *The Journal of Physical Chemistry Letters* **12**, 10994-10995 (2021).
33. Zaslavsky, B.Y. & Uversky, V.N. In Aqua Veritas: The Indispensable yet Mostly Ignored Role of Water in Phase Separation and Membrane-less Organelles. *Biochemistry* **57**, 2437-2451 (2018).
34. Slomkowski, S. et al. Terminology of polymers and polymerization processes in dispersed systems (IUPAC Recommendations 2011). *Pure and Applied Chemistry* **83**, 2229-2259 (2011).
35. Forrest, S.R. & Witten, T.A. Long-range correlations in smoke-particle aggregates. *Journal of Physics A: Mathematical and General* **12**, L109-L117 (1979).
36. Lazzari, S., Nicoud, L., Jaquet, B., Lattuada, M. & Morbidelli, M. Fractal-like structures in colloid science. *Adv Colloid Interface Sci* **235**, 1-13 (2016).
37. Stanley, H.E. Application of fractal concepts to polymer statistics and to anomalous transport in randomly porous media. *Journal of Statistical Physics* **36**, 843-860 (1984).
38. Havlin, S. & Ben-Avraham, D. Diffusion in disordered media. *Advances in Physics* **36**, 695-798 (1987).
39. Lin, M.Y. et al. Universality in colloid aggregation. *Nature* **339**, 360-362 (1989).
40. Khatun, S. et al. Fractal self-assembly and aggregation of human amylin. *Soft Matter* **16**, 3143-3153 (2020).
41. Knowles, T.P. et al. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science* **318**, 1900-3 (2007).

42. Nicoud, L., Lazzari, S., Balderas Barragán, D. & Morbidelli, M. Fragmentation of amyloid fibrils occurs in preferential positions depending on the environmental conditions. *J Phys Chem B* **119**, 4644-52 (2015).
43. Kharlamova, A., Nicolai, T. & Chassenieux, C. Gelation of whey protein fractal aggregates induced by the interplay between added HCl, CaCl2 and NaCl. *International Dairy Journal* **111**, 104824 (2020).
44. da Silva, M.A. & Arêas, E.P. Solvent-induced lysozyme gels: rheology, fractal analysis, and sol-gel kinetics. *J Colloid Interface Sci* **289**, 394-401 (2005).
45. Gallego, L.D. et al. Phase separation directs ubiquitination of gene-body nucleosomes. *Nature* **579**, 592-597 (2020).
46. Vernon, R.M. et al. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **7**(2018).
47. Polyansky, A.A., Volynsky, P.E., Arseniev, A.S. & Efremov, R.G. Adaptation of a membrane-active peptide to heterogeneous environment. I. Structural plasticity of the peptide. *J Phys Chem B* **113**, 1107-19 (2009).
48. de Ruiter, A., Polyansky, A.A. & Zagrovic, B. Dependence of Binding Free Energies between RNA Nucleobases and Protein Side Chains on Local Dielectric Properties. *J Chem Theory Comput* **13**, 4504-4513 (2017).
49. Bernado, P. & Blackledge, M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys J* **97**, 2839-45 (2009).
50. Carpineti, M. & Giglio, M. Spinodal-type dynamics in fractal aggregation of colloidal clusters. *Phys Rev Lett* **68**, 3327-3330 (1992).
51. Uversky, V.N. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J* **28**, 305-25 (2009).
52. Lewis, R.N.A.H. & McElhaney, R.N. Membrane lipid phase transitions and phase organization studied by Fourier transform infrared spectroscopy. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1828**, 2347-2358 (2013).
53. Thouy, R. & Jullien, R. A cluster-cluster aggregation model with tunable fractal dimension. *Journal of Physics A: Mathematical and General* **27**, 2953-2963 (1994).
54. Kätzel, U. et al. Dynamic Light Scattering for the Characterization of Polydisperse Fractal Systems: I. Simulation of the Diffusional Behavior. *Particle & Particle Systems Characterization* **25**, 9-18 (2008).

55. Morán, J., Fuentes, A., Liu, F. & Yon, J. FracVAL: An improved tunable algorithm of cluster–cluster aggregation for generation of fractal structures formed by polydisperse primary particles. *Computer Physics Communications* **239**, 225-237 (2019).
56. Sanders, D.W. et al. Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181**, 306-324.e28 (2020).
57. Espinosa, J.R. et al. Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components. *Proceedings of the National Academy of Sciences* **117**, 13238-13247 (2020).
58. Klein, R. et al. Theory of scattering from colloidal aggregates. in *Trends in Colloid and Interface Science IV* (eds Zulauf, M., Lindner, P. & Terech, P.) 161-168 (Steinkopff, Darmstadt, 1990).
59. McCall, P.M. et al. Quantitative phase microscopy enables precise and efficient determination of biomolecular condensate composition. *bioRxiv*, 2020.10.25.352823 (2020).
60. Armstrong, J.K., Wenby, R.B., Meiselman, H.J. & Fisher, T.C. The Hydrodynamic Radii of Macromolecules and Their Effect on Red Blood Cell Aggregation. *Biophysical Journal* **87**, 4259-4270 (2004).
61. Kar, M. et al. Phase separating RNA binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *bioRxiv*, 2022.02.03.478969 (2022).
62. Lasker, K. et al. A modular platform for engineering function of natural and synthetic biomolecular condensates. *bioRxiv*, 2021.02.03.429226 (2021).
63. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).
64. Jorgensen, W.L., Maxwell, D.S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **118**, 11225-11236 (1996).
65. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. & Stewart, J.J.P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society* **107**, 3902-3909 (1985).
66. Jorgensen, W.L. & Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J Comput Chem* **26**, 1689-700 (2005).

67. Jorgensen, W.L. & Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *The Journal of Chemical Physics* **83**, 3050-3054 (1985).
68. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-58 (2015).
69. Abraham, M.J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
70. Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-8 (2010).
71. Piana, S., Donchev, A.G., Robustelli, P. & Shaw, D.E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B* **119**, 5113-23 (2015).
72. Hoover, W.G. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A Gen Phys* **31**, 1695-1697 (1985).
73. Parrinello, M. & Rahman, A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *Journal of Applied Physics* **52**, 7182-7190 (1981).
74. Hess, B., Bekker, H., Berendsen, H.J.C. & Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472 (1997).
75. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. (2010).
76. King, B.M., Silver, N.W. & Tidor, B. Efficient Calculation of Molecular Configurational Entropies Using an Information Theoretic Approximation. *The Journal of Physical Chemistry B* **116**, 2891-2904 (2012).
77. Fleck, M., Polyansky, A.A. & Zagrovic, B. PARENT: A Parallel Software Suite for the Calculation of Configurational Entropy in Biomolecular Systems. *Journal of Chemical Theory and Computation* **12**, 2055-2065 (2016).
78. Harpaz, Y., Gerstein, M. & Chothia, C. Volume changes on protein folding. *Structure* **2**, 641-9 (1994).

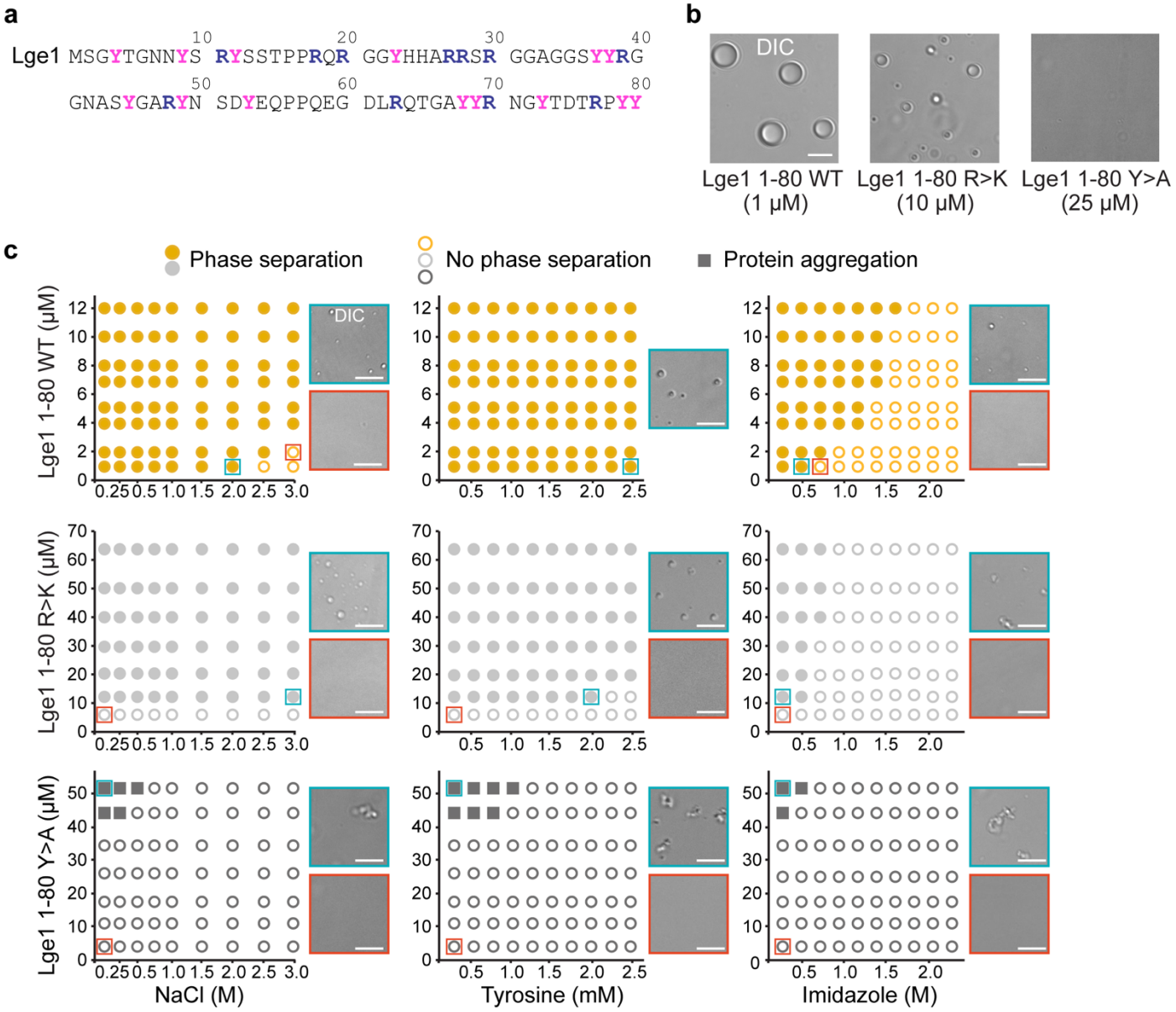
a
Lge1 MSGYTGNNYS RYSSTPPRQR GGYHHARRSR GGAGGSYYRG
GNASYGARYN SDYEQPPQEG DLRQTGAYYR NGYTDTRPYY
b
DIC
Lge1 1-80 WT (1 μM)
Lge1 1-80 R>K (10 μM)
Lge1 1-80 Y>A (25 μM)
c
Phase separation
No phase separation
Protein aggregation
Lge1 1-80 WT (μM)
Lge1 1-80 R>K (μM)
Lge1 1-80 Y>A (μM)
NaCl (M)
Tyrosine (mM)
Imidazole (M)

Figure 1

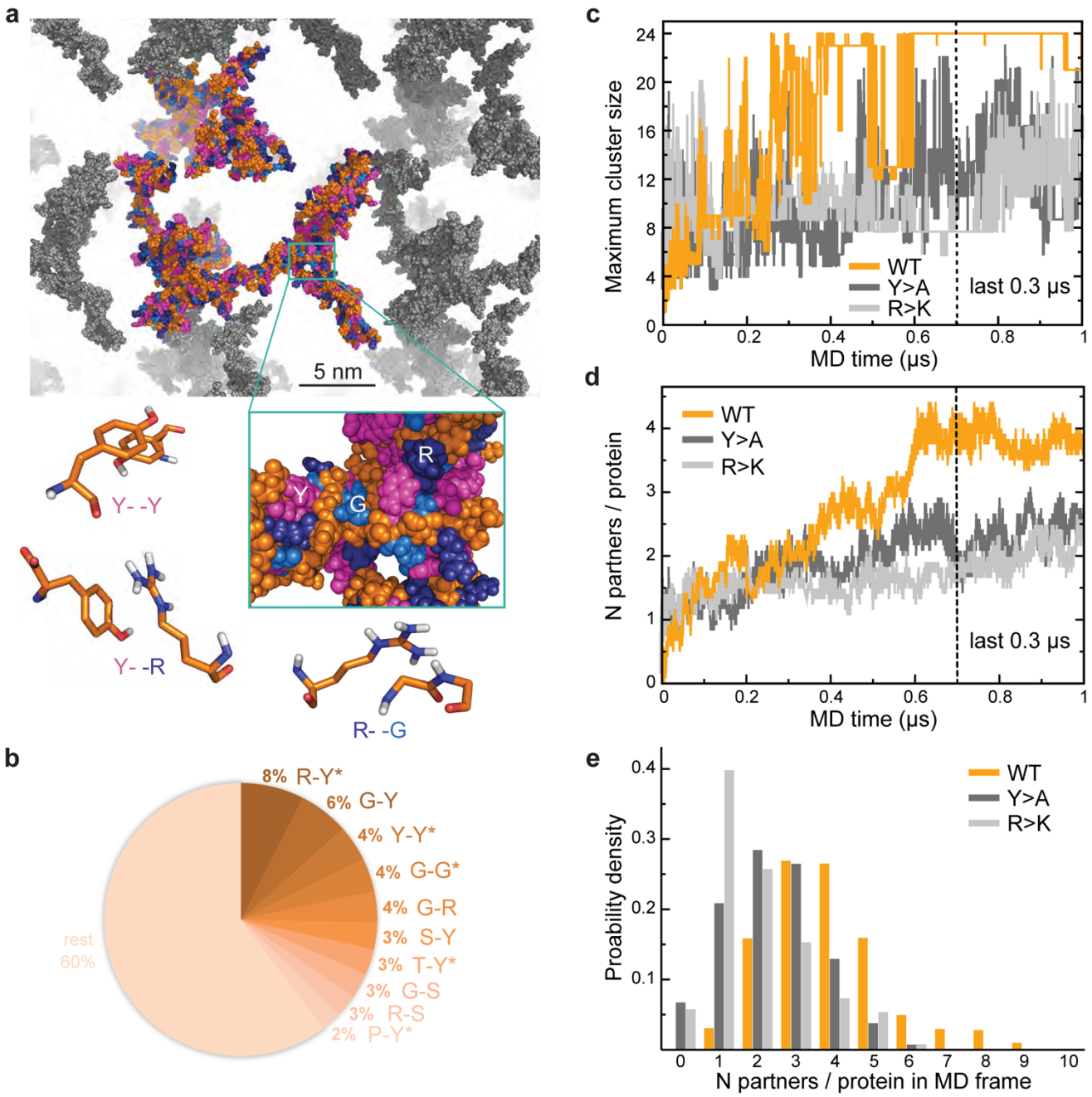
a
5 nm
R
Y
G
Y- -Y
Y- -R
R- -G
b
8% R-Y*
6% G-Y
4% Y-Y*
4% G-G*
4% G-R
3% S-Y
3% T-Y*
3% G-S
3% R-S
2% P-Y*
rest
60%
c
Maximum cluster size
MD time (μs)
WT
Y>A
R>K
last 0.3 μs
d
N partners / protein
MD time (μs)
WT
Y>A
R>K
last 0.3 μs
e
Proability density
N partners / protein in MD frame
WT
Y>A
R>K

Figure 2

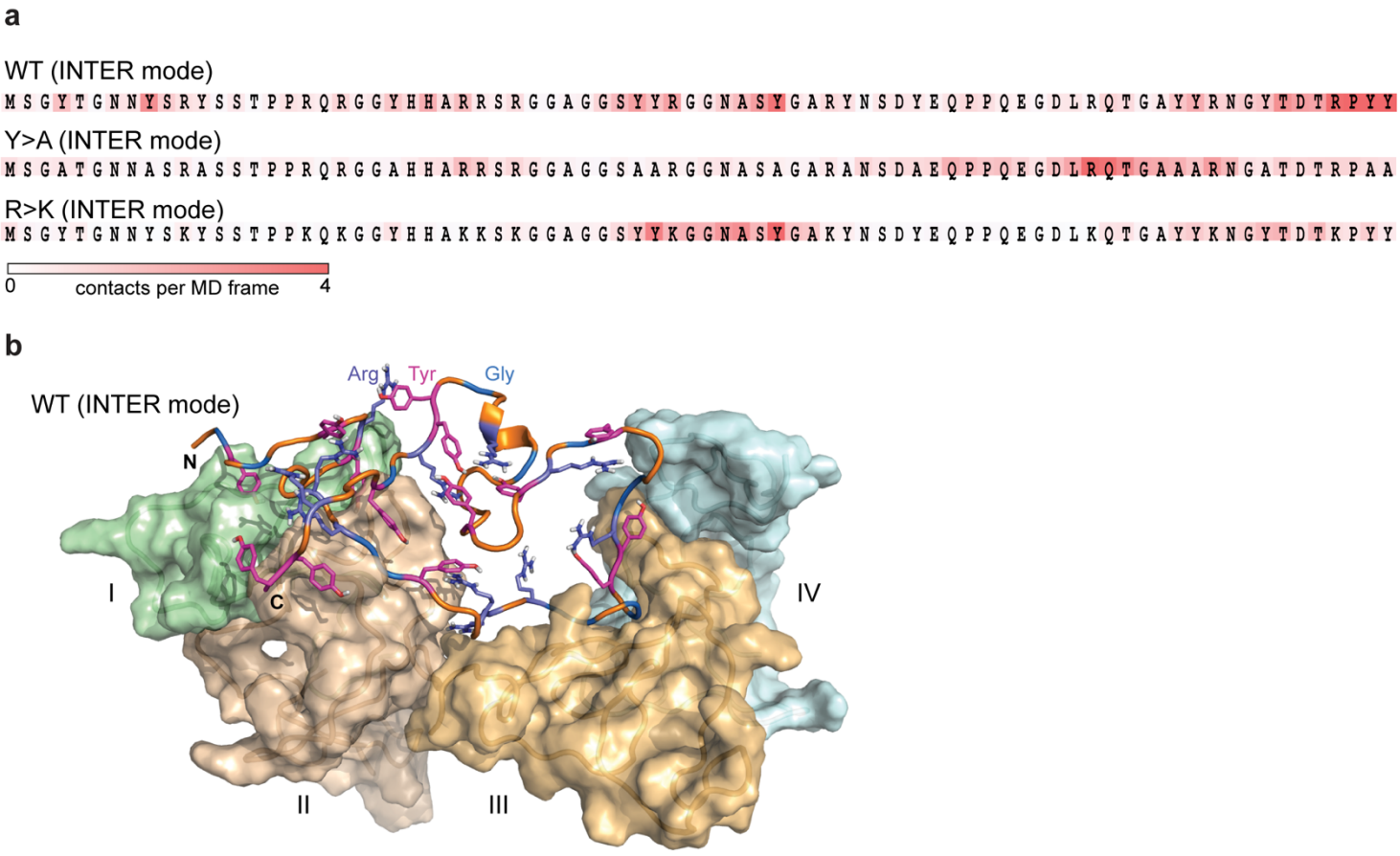
a
WT (INTER mode)
MSGYTGNNYSRYSSTPPRQRGGYHHARRSRGGAGGSYYRGGNASYGARYNSDYEQPPQEGDLRQTGAYYRNGYTDTRPYY
Y>A (INTER mode)
MSGATGNNASRASSTPPRQRGGAHHARRSRGGAGGSAARGGNASAGARANSDAEQPPQEGDLRQTGAAARNGATDTRPAA
R>K (INTER mode)
MSGYTGNNYSKYSSTPPKQKGGYHHAKKSKGGAGGSYYKGGNASYGAKYNSDYEQPPQEGDLKQTGAYYKNGYTDTKPYY
0
contacts per MD frame
4
b
WT (INTER mode)
Arg
Tyr
Gly
N
C
I
II
III
IV

Figure 3

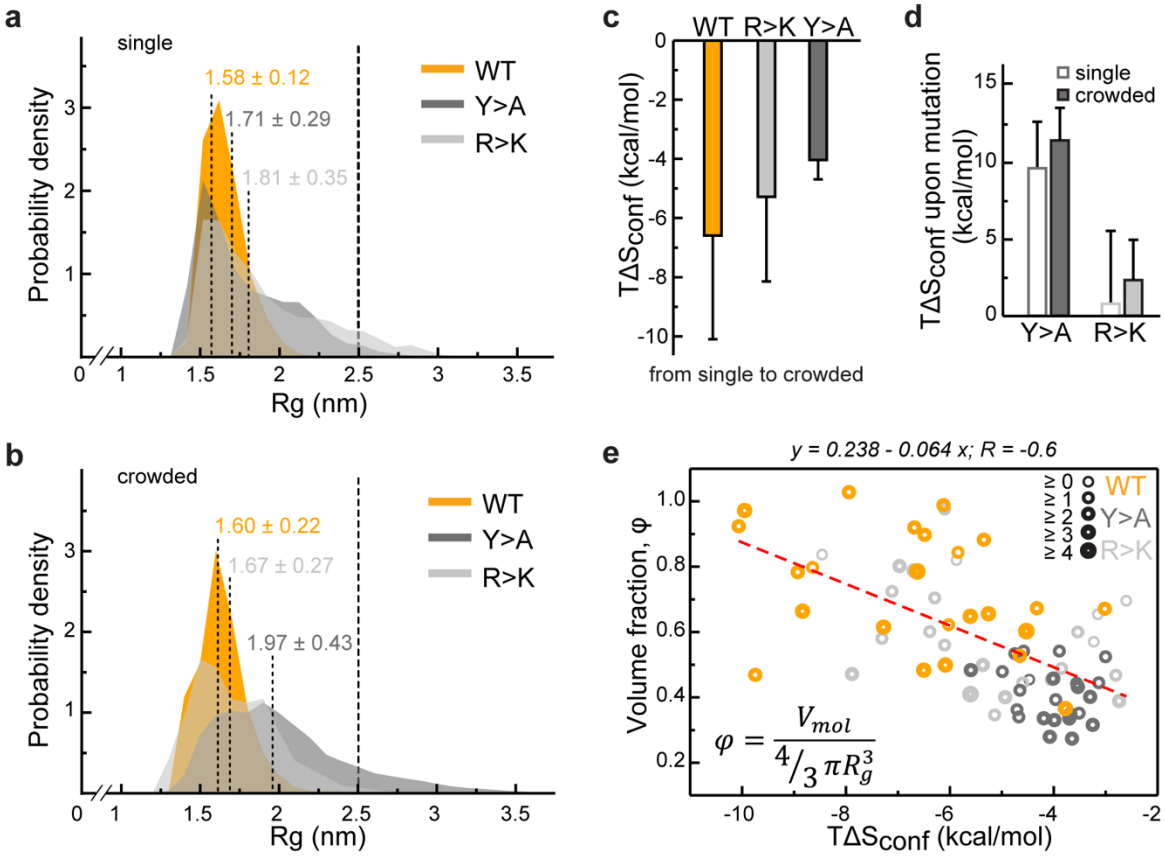
a
single
1.58 ± 0.12
1.71 ± 0.29
1.81 ± 0.35
WT
Y>A
R>K
Probability density
Rg (nm)
b
crowded
1.60 ± 0.22
1.67 ± 0.27
1.97 ± 0.43
WT
Y>A
R>K
Probability density
Rg (nm)
c
WT R>K Y>A
TΔSconf (kcal/mol)
from single to crowded
d
single
crowded
TΔSconf upon mutation (kcal/mol)
Y>A R>K
e
y = 0.238 - 0.064 x; R = -0.6
≥ 0 WT
≥ 1
≥ 2 Y>A
≥ 3
≥ 4 R>K
Volume fraction, φ
TΔSconf (kcal/mol)
$\varphi = \frac{V_{mol}}{4/3\,\pi R_g^3}$

Figure 4

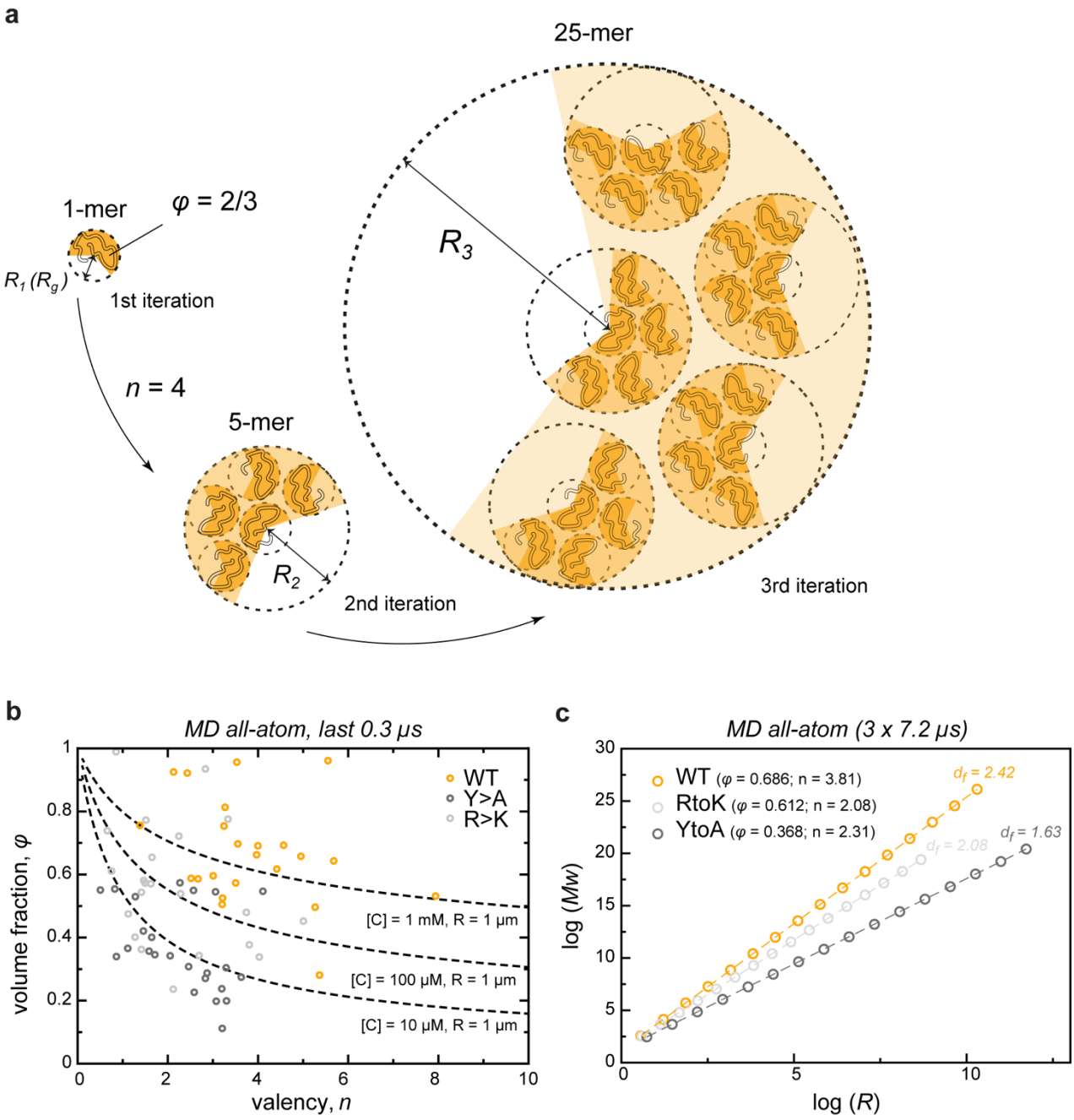
a
25-mer
1-mer
φ = 2/3
R1 (Rg)
1st iteration
n = 4
5-mer
R3
R2
2nd iteration
3rd iteration
b
MD all-atom, last 0.3 μs
WT
Y>A
R>K
[C] = 1 mM, R = 1 μm
[C] = 100 μM, R = 1 μm
[C] = 10 μM, R = 1 μm
volume fraction, φ
valency, n
c
MD all-atom (3 x 7.2 μs)
WT (φ = 0.686; n = 3.81)
RtoK (φ = 0.612; n = 2.08)
YtoA (φ = 0.368; n = 2.31)
df = 2.42
df = 2.08
df = 1.63
log (Mw)
log (R)

Figure 5

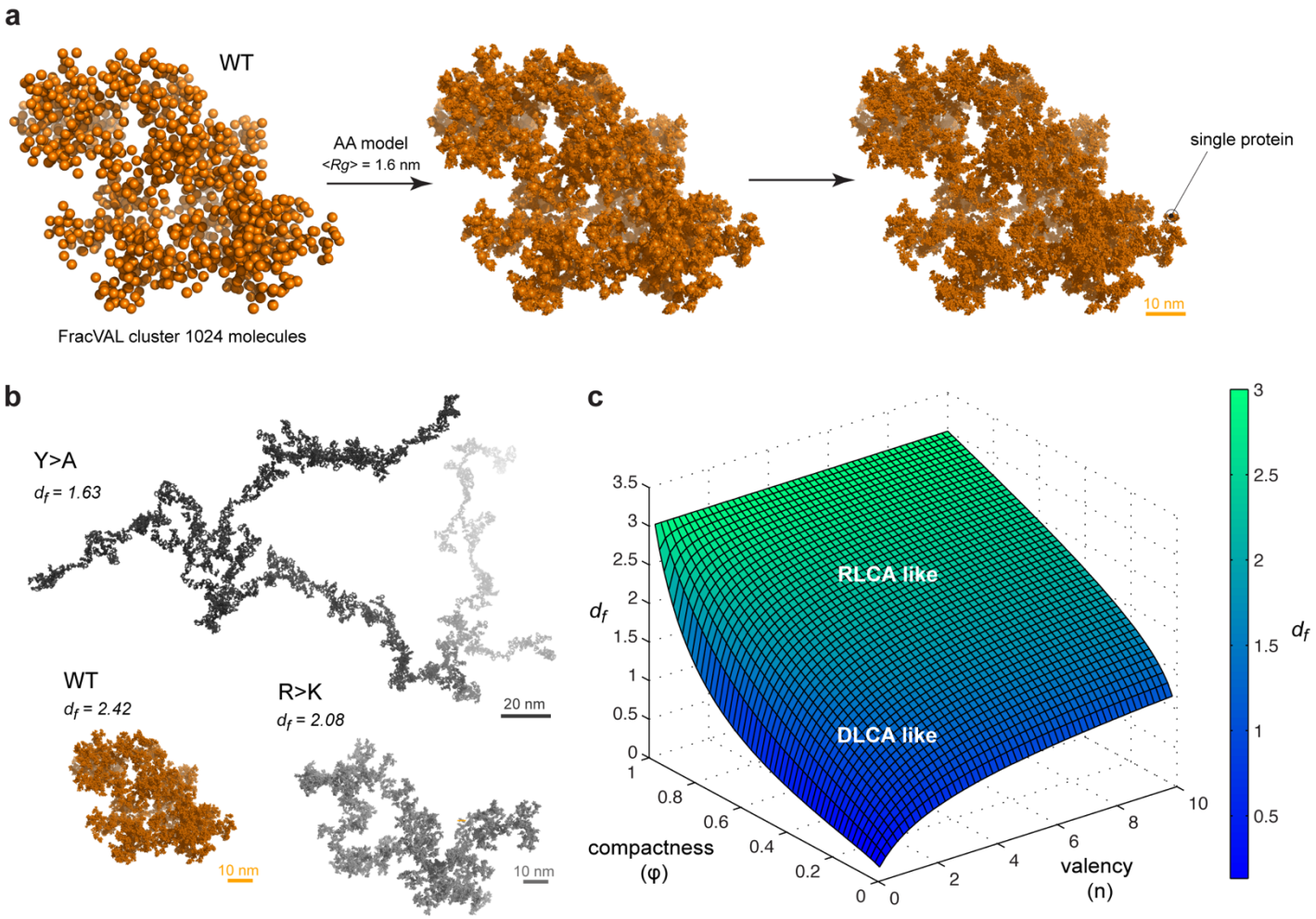
a
WT
AA model
<Rg> = 1.6 nm
single protein
10 nm
FracVAL cluster 1024 molecules
b
Y>A
d_f = 1.63
WT
d_f = 2.42
R>K
d_f = 2.08
20 nm
10 nm
10 nm
c
RLCA like
DLCA like
d_f
compactness
(φ)
valency
(n)

Figure 6

Figure captions:

Figure 1 | Lge1$_{1-80}$ LLPS critically depends on tyrosine residues. (**a**) Sequence of Lge1$_{1-80}$. Arginines and tyrosines are highlighted with deep blue and magenta colors, respectively. (**b**) LLPS for Lge1$_{1-80}$ WT, R>K and Y>A in buffer with 200 mM NaCl. (**c**) Phase diagrams for the WT (*top row*), R>K (*middle row*), and Y>A (*bottom row*) Lge1$_{1-80}$ variants with protein concentration given on the *y*-axis and concentration of NaCl (*left panels*), tyrosine (*middle panels*) and imidazole (*right panels*) given on the *x*-axis. Representative micrographs of light microscopy analyses for the selected conditions. Scale bar, 5 µm.

Figure 2 | Direct assessment of interaction network for Lge1$_{1-80}$ variants by all-atom MD in 24-copy systems. (**a**) Exemplary MD-snapshot of the WT interaction network. Proteins in the simulation box are shown using the orange sphere atomic representation, whereby glycine, arginine and tyrosine residues are highlighted with sky blue, deep blue and magenta colors, respectively. Periodic images of the system in *x*-, *y*-, *z*- directions are shown in gray. (**b**) Pairwise contact statistics over the last 0.3 µs of MD simulations for the WT. Contacts enriched over the randomized background are marked with stars (see **Extended Data 2a** for exact enrichment values). (**c**) Time evolution of the largest detected protein interaction cluster size and (**d**) the average number of interaction partners per protein molecule, i.e. interaction valency, for the WT (*orange line*), R>K (*dark gray line*), and Y>A (*light gray line*) multi-copy systems. (**e**) Distributions of the number of interaction partners per protein over the last 0.3 µs of MD simulations. The color code is the same as in (**c**) and (**d**).

Figure 3 | Lge1$_{1-80}$ variants exhibit a *dynamic binding mode* in 24-copy systems. (**a**) Representative distributions of statistically defined interaction regions ("stickers") mapped onto the protein sequence. Protein sequences are colored according to the average contact statistics over the last 0.3 µs. Four interaction profiles of proteins having the number of partners corresponding to the average valency in the system (4 partners for the WT and 2 partners for both mutants) and displaying the highest mutual correlations were used for the determination of the representative mode. The representative modes in all three cases display the Pearson correlation coefficient $R > 0.6$ with the interaction profiles obtained by averaging over all 24 copies in each system. (**b**) A 3D model of the representative binding mode for WT: an MD snapshot at 1 µs is given for a protein copy (shown in cartoon and sticks representation; the color scheme is the same as in **Fig. 2a**), which is simultaneously interacting with four partners

(shown in surface representation in pale cyan, pale green, wheat and light orange colors, and indicated by Latin numbers, respectively).

Figure 4 | Alteration of Lge1$_{1\text{-}80}$ sequence impacts its conformational behavior. Distributions of radii of gyration (*Rg*) for Lge1$_{1\text{-}80}$ variants in (**a**) single-copy and (**b**) multi-copy systems. The last 0.3 μs of MD trajectories were used to collect *Rg* statistics for (**a**) 2 independent runs of the single-copy simulation and (**b**) all 24 protein copies in the multi-copy system. Average *Rg* values ($<Rg_{MD}>$) and the corresponding standard deviations over the last 0.3 μs are indicated. Theoretical Rg_{rc} value for an 80-aa disordered protein chain (see **Methods**) is shown with a vertical black dashed line. (**c**) Average changes in the configurational entropy (ΔS_{conf}) of a protein molecule for the transition from the single-molecule context (a diluted state) to the crowded environment. Entropy values are given in energy units ($T\Delta S_{conf}$, T = 310 K) and were obtained using complete 1 μs MD trajectories. Averaging was done for entropy differences in all possible combinations between two independent runs of single molecules and 24 protein copies in the crowded system. (**d**) Average changes in $T\Delta S_{conf}$ upon mutations (see **Methods**) in the single molecule context and in the crowded environment. Averaging was done for entropy differences over all possible combinations (2x2 and 24x24, respectively). (**e**) Correlation between relative single molecule configurational entropies and the corresponding average compactness (φ) values of all protein copies in the crowded system. The average valency of different protein copies is proportional to the thickness of circles. Entropy values, average compactness and average valency values were calculated over the complete 1μs MD trajectories.

Figure 5 | Describing condensate architecture via a fractal scaling model (**a**) Schematic representation of a self-propagation principle of condensate assembly. Interaction valency (n) and compactness (φ) of individual proteins determine the properties of protein clusters at different iterations (see **Appendix** for the corresponding formalism). (**b**) The parameter space for φ and n for a condensate with a fixed size and concentration described by the model (*dashed lines*). The φ and n values for each individual protein molecule in the crowded environment averaged over the last 0.3 μs of MD trajectories are shown with open circles. (**c**) Power-law dependence between mass and size of protein clusters at different iterations of the model with the applied valency and compactness corresponding to their average values over the 24 simulated protein copies and the last 0.3 μs of MD trajectories (indicated in the legend). Dashed

lines show linear regression for the log *R* vs. log *Mw* plot with the corresponding slope or fractal dimension (d_f) indicated above the lines.

Figure 6 | Reconstructing the condensate architecture on the submicron scale with atomistic resolution. (**a**) Transformation of a coarse-grained 1024 particle cluster obtained by FracVAL algorithm to an all-atom representation. The cluster was reconstructed using the fractal dimension (d_f) and the averaged *Rg* value derived from MD simulations of multi-copy WT system. (**b**) Representative 1024-protein clusters for $Lge1_{1\text{-}80}$ variants at all-atom resolution. (**c**) The non-linear dependence of the fractal dimension d_f on φ and n as given by the model formalism (see eq.10, **Appendix**). The surface is colored according to the corresponding d_f values (see the scale bar).

Extended data captions:

Extended data 1 | Experimental and modeling studies of the effect of tyrosine and arginine mutations in Lge1$_{1\text{-}80}$. (**a**). Protein systems studied with all-atom molecular dynamic simulations. For each system a detailed composition, effective molar and mass protein concentration, size of the simulated cubic box, simulation time and the number of replicas are indicated. (**b**) SDS-Page with Coomassie staining (4-12 % gel, MOPS buffer) illustrating the quality of *E. coli* expressed Lge1$_{1\text{-}80}$ variants (see Methods for the protocol). (**c**) Dependence of the pairwise interaction free energies (ΔG) of arginine (R), lysine (K), and tyrosine (Y) side-chain analogs on the polarity of the environment (ε). Results of all-atom Monte-Carlo simulations (see **Methods** for the details).

Extended data 2 | Inter- and intramolecular interactions of Lge1$_{1\text{-}80}$ variants. (a) Top 10 most enriched pairwise contacts between residues in different protein molecules in the multi-copy simulations (INTER, solid bars) for the WT *(left panels),* the R>K *(middle panels*), and the Y>A (*right panels*) Lge1$_{1\text{-}80}$ variants. INTER contacts are sorted according to their corresponding frequencies among all pairwise inter-residue contacts observed in multi-copy systems, which are indicated above the bars. The enrichments were calculated in comparison to the randomized background. A red dashed line indicates the enrichment value of 1. The corresponding pairwise contacts statistics in the single molecule context (INTRA) are shown with open bars. Contact statistics were collected over the last 0.3 μs of MD simulations for two independent single-molecule MD runs (INTRA) and all 24 proteins in the multi-copy MD runs (INTER). (**b**) Top 10 most enriched pairwise contacts in the single molecule context (INTRA, open bars). INTRA contacts are sorted according to their corresponding frequencies among all pairwise inter-residue contacts observed in the single-copy systems, which are indicated above the bars. The corresponding pairwise contact statistics for the multi-copy systems (INTRA) are shown with filled bars. (**c**) Distributions of statistically defined interaction regions ("stickers") along the protein sequence in the single-copy systems. Protein sequences are colored according to the average contact statistics over the last 0.3 μs in two independent MD runs.

Extended data 3 | Conformational behavior and condensate topology of Lge1$_{1\text{-}80}$. (**a**) MD conformations of WT (*top panels*), Y>A (*middle panels*), and R>K (*bottom panels*) variants

illustrating different levels of compactness with the corresponding inter-residue distance matrices averaged over the 100 ns of a single-copy MD trajectory (contact maps; scale: white – 0 nm; black – 1.5 nm). Proteins are shown in cartoon and sticks representation, and colored in orange (WT), dark gray (Y>A), and gray color (R>K) with tyrosine (Y), arginine (R), lysine (K), and glycine (G) residues colored according to the legend given below. (**b**) Convergence of configurational entropy (S_{conf}) of a protein molecule in the single-molecule context. Data for two independent MD runs are shown with solid and dashed lines, respectively. Cumulative plots were generated using MIST approximation (see **Methods**) and a 50-ns time step. (**c**) A 24 protein cluster for WT Lge1$_{1\text{-}80}$ fitted to an ellipsoid with radii corresponding to *x*, *y*, *z* components of *Rg*. Proteins are shown as spheres. (**d**) Power-law dependence between mass and size of the WT protein clusters over four iterations of the fractal model (open circles) with the applied MD-derived valency (n) and compactness (φ) values and MD clusters of 1 and 24 proteins (filled circles); the corresponding linear regression trends are shown with dashed and solid lines, respectively, with slope (A) and intercept (B) values indicated in the legend. The φ and n values used in the model and given in the legend correspond to the averages over 24 protein copies, while MD clusters were directly extracted from the slopes and the intercepts of the linear fits performed according to the equations given on the plot (eq. 9,12, 13 in **Appendix**). All averaging was done over the last 0.3 µs of the multi-copy MD trajectories. The size of the 24-protein MD cluster was estimated as the radius of the sphere with an equivalent apparent volume as the gyration ellipsoid. (**e**, **f**) Partition of dextran of different sizes into condensates formed by the WT Lge1$_{1\text{-}80.}$ Condensates were incubated with TRITC-labelled (final dextran concentration 0.05 mg/ml) for 15 min at 20ºC and imaged by DIC and fluorescence microscopy. Scale bar, 10 µm.

Extended data 4 | Parameters used in the self-propagation fractal model for different Lge1$_{1\text{-}80}$ variants. Radius of gyration (*Rg*)**,** valency (n) and compactness (φ) values averaged over the last 0.3 µs of MD for all 24 copies. Slope (A) and intercept (B) of the linear regression for the log *R* vs. log *Mw* plot are derived from the model (eq. 10, 11 in **Appendix**).