JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/jacsau

Article

Sequence Tendency for the Interaction between Low-Complexity Intrinsically Disordered Proteins

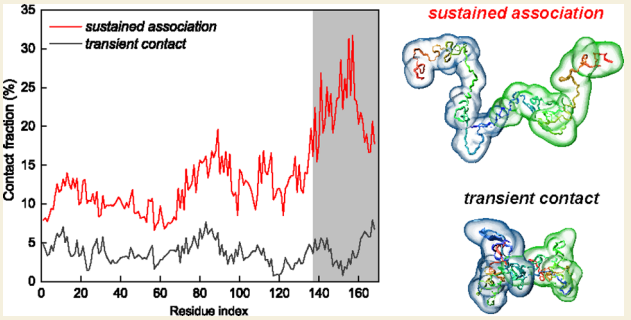
Moxin Zhang, Bin Xue, Qingtai Li, Rui Shi, Yi Cao, Wei Wang,* and Jingyuan Li*

Cite This: *JACS Au* 2023, 3, 93−104

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Reversible interaction between intrinsically disordered proteins (IDPs) is considered as the driving force for liquid−liquid phase separation (LLPS), while the detailed description of such a transient interaction process still remains a challenge. And the mechanisms underlying the behavior of IDP interaction, for example, the possible relationship with its inherent conformational fluctuations and sequence features, remain elusive. Here, we use atomistic molecular dynamics (MD) simulation to investigate the reversible association of the LAF-1 RGG domain, the IDP with ultra-low LLPS concentration (0.06 mM). We find that the duration of the association between two RGG domains is highly heterogeneous, and the sustained associations essentially dominate the IDP interaction. More interestingly, such sustained associations are mediated by a finite region, that is, the C-terminal region 138−168 (denoted as a contact-prone region). We noticed that such sequence tendency is attributed to the extended conformation of the RGG domain during its inherent conformational fluctuations. Hence, our results suggest that there is a certain region in this low-complexity IDP which can essentially dominate their interaction and should be also important to the LLPS. And the inherent conformational fluctuations are actually essential for the emergence of such a hot region of IDP interaction. The importance of this hot region to LLPS is verified by experiment.



KEYWORDS: *intrinsically disordered proteins, interaction hot region, molecular dynamics simulation, conformation fluctuation, liquid−liquid phase separation, sustained association*

INTRODUCTION

Protein−protein interactions (PPIs) are crucial determinants in the biological function of proteins.[1,2] The traditional paradigm of PPIs is about the structural proteins wherein PPIs exhibit high stability and commonly occur on the specific sites with certain sequence features.[3−7] Hence, such PPIs are determined by the conformation and sequence of structural proteins.[8,9] On the other hand, there are numerous proteins that lack a unique structure and tend to harbor low-complexity sequence domains (LCDs). They are denoted as intrinsically disordered proteins (IDPs).[10−12] Their interactions are directly related to many biological functions[13,14] such as mediating the assembly of biomolecule condensates[15−20] through liquid−liquid phase separation (LLPS). Sufficient interaction between IDPs should be demanded for the formation and maintenance of LLPS, especially for the IDPs with exceptionally low critical concentration.[16,21,22] On the other hand, the interaction between IDPs is considered to be transient and devoid of specific binding sites, which are distinct from their structural counterpart.[23−25] To date, it is still a big challenge to comprehensively and quantitatively characterize the IDP interaction. Moreover, IDPs continuously undergo inherent conformational fluctuations, whose impact on IDP interaction and the underlying sequence features also remain elusive.

There have been several experimental studies on the interaction between IDPs along with their conformation features.[15,21,23,25−35] For example, single-molecular Förster resonance energy transfer−fluorescence correlation spectroscopy (smFRET-FCS) experiments of Tau demonstrated that under LLPS condition, Tau adopted a rather extended conformation ensemble which facilitated protein interactions.[36] Moreover, there have been also studies combining experiments with bioinformatic analysis to discuss the interaction mode of intrinsically disordered LCDs with respect to their sequence characteristics. For example, a combination of the bioinformatics analysis with NMR experiment showed that phase separation of DDX4 disordered low-complexity domain was governed by cation−π interactions between

Received: July 25, 2022
Revised: December 16, 2022
Accepted: December 16, 2022
Published: December 30, 2022

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

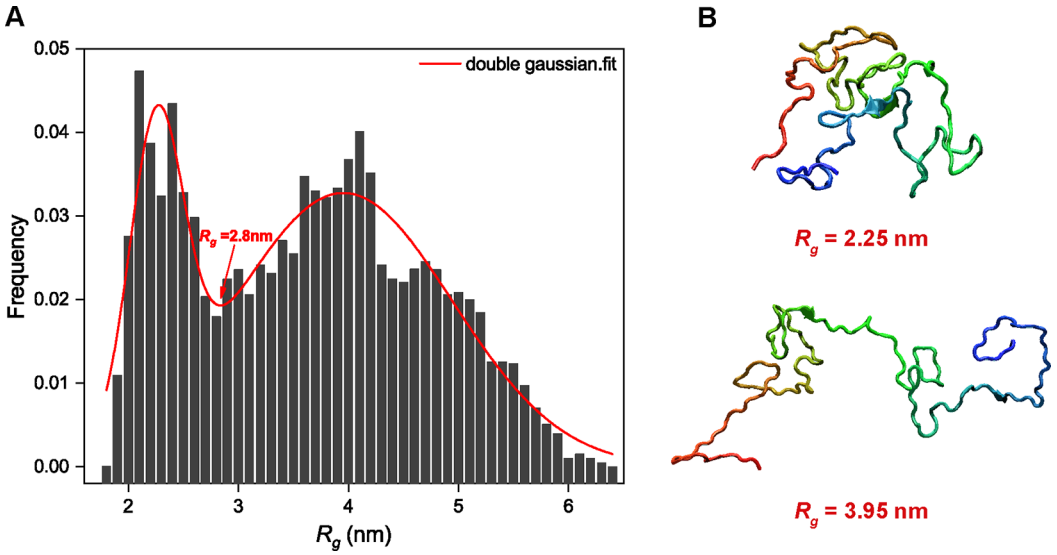


Figure 1. Conformational ensemble of LAF-1 RGG domain. (A) R_g distribution of the monomeric LAF-1 RGG domain. Double-Gaussian fitting curve is shown as the red line. Two Gaussian peaks are locate at R_g = 2.25 and 3.95 nm, and the crossover region is at around R_g = 2.80 nm. (B) Representative configurations of the LAF-1 RGG domain corresponding to two peaks. IDPs are colored with N- and C-terminal in red and blue, respectively.

repeatedly spaced RG/GR and FG/GF dipeptides.[37] Although much has been learned in these studies, the comprehensive descriptions about the reversible interaction of IDP as well as its relationship with the inherent conformational fluctuations and sequence features of IDP are still limited.

Molecular dynamics (MD) simulations have been widely used to decipher the behaviors of protein. The phase behaviors of IDP condensate and the relationship with IDP sequence are studied by the coarse-grained MD simulations and analytic theory,[38−45] for example, the sequence-dependent phase behavior of the synthetic proteins,[38,42] the impact of varying charge patterning on phase behavior of LAF-1 LCD and DDX4,[40,41] as well as the response of IDP phase diagram to the change of sequence distribution of charged residues.[39,43,45] Meanwhile, the conformational properties of various IDPs are widely studied by both atomic and coarse-grained MD simulations and combined with the discussion of sequence effect.[46−51] On the other hand, quantitative descriptions of the reversible IDP association and the relationship with conformational fluctuations and sequence features of IDP are still lacking due to the absence of a simulation with higher precision.

LAF-1 is recognized as a typical protein involved in LLPS[15,21] and drives the assembly of P granules, the membraneless organelle involved in *Caenorhabditis elegans* embryos.[15,21,52,53] It contains an N-terminal low-complexity domain, that is, the intrinsically disordered R/G-rich (RGG) domain[15] which undergoes considerable conformational fluctuation.[23,29,30] Previous studies have shown that RGG domain is important to the LLPS mediated by LAF-1, and the isolated RGG domain can sufficiently engender LLPS in vitro with the critical concentration as low as 0.06 mM.[40] In other words, LAF-1 RGG domain can serve as a representative IDP system.

Here, we use unbiased all-atom MD simulations to investigate the interaction between two LAF-1 RGG domains. In addition to the transient contact between IDPs with the duration of tens of nanoseconds, there are sustained contact events with considerably higher stability and the duration of hundreds-nanoseconds. Such sustained associations dominate the interaction between IDPs and, more strikingly, exhibit an intriguing sequence tendency. There is a finite region (i.e., C-terminal 138−168 residues, denoted as contact-prone region) that has much higher propensity (more than twice of the other regions) to interact with the other IDP. The dynamics of IDP association especially the emergence of such contact-prone region of IDPs are related to their inherent conformational fluctuations: IDPs with relatively compact conformation likely form transient contact, while the relatively extended conformation facilitates interaction through certain regions to realize stable association and the resulting sequence tendency. Besides, extended conformation can induce the entangling of two IDPs which can also accomplish sustained association. Moreover, our results suggest the conformational fluctuations of IDPs are essential for the emergence of such intriguing sequence tendency. The importance of contact-prone region to the LLPS mediated by the interaction of LAF-1 RGG domain is verified by the phase diagram experiment.

RESULTS AND DISCUSSION

Conformation of the LAF-1 RGG Domain

The sequence of the intrinsically disordered RGG domain of LAF-1 is shown in Figure S1. The 168-residues fragment RGG domain is conceived as the low-complexity domain (LCD) which is rich in arginine and glycine (59 and 24 in total, respectively) and sufficient to form liquid-like protein droplets in vitro.[15,21] We first study the conformation ensemble of LAF-1 RGG domain based on the trajectory of single-protein system. The radius of gyration (R_g) of monomeric RGG domain spans a broad range from 1.8 to 6.4 nm, suggesting the significant conformational change of RGG domain (Figure 1A). Moreover, we notice the distribution of R_g can be well fitted by double-Gaussian function with the peaks at R_g = 3.95 and 2.25 nm, and the crossover region locates at $R_g \sim$ 2.8 nm (see Table S1 for more detailed information). Double-Gaussian distribution of R_g indicates that the conformational ensemble of protein can be treated as two domains, that is,

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

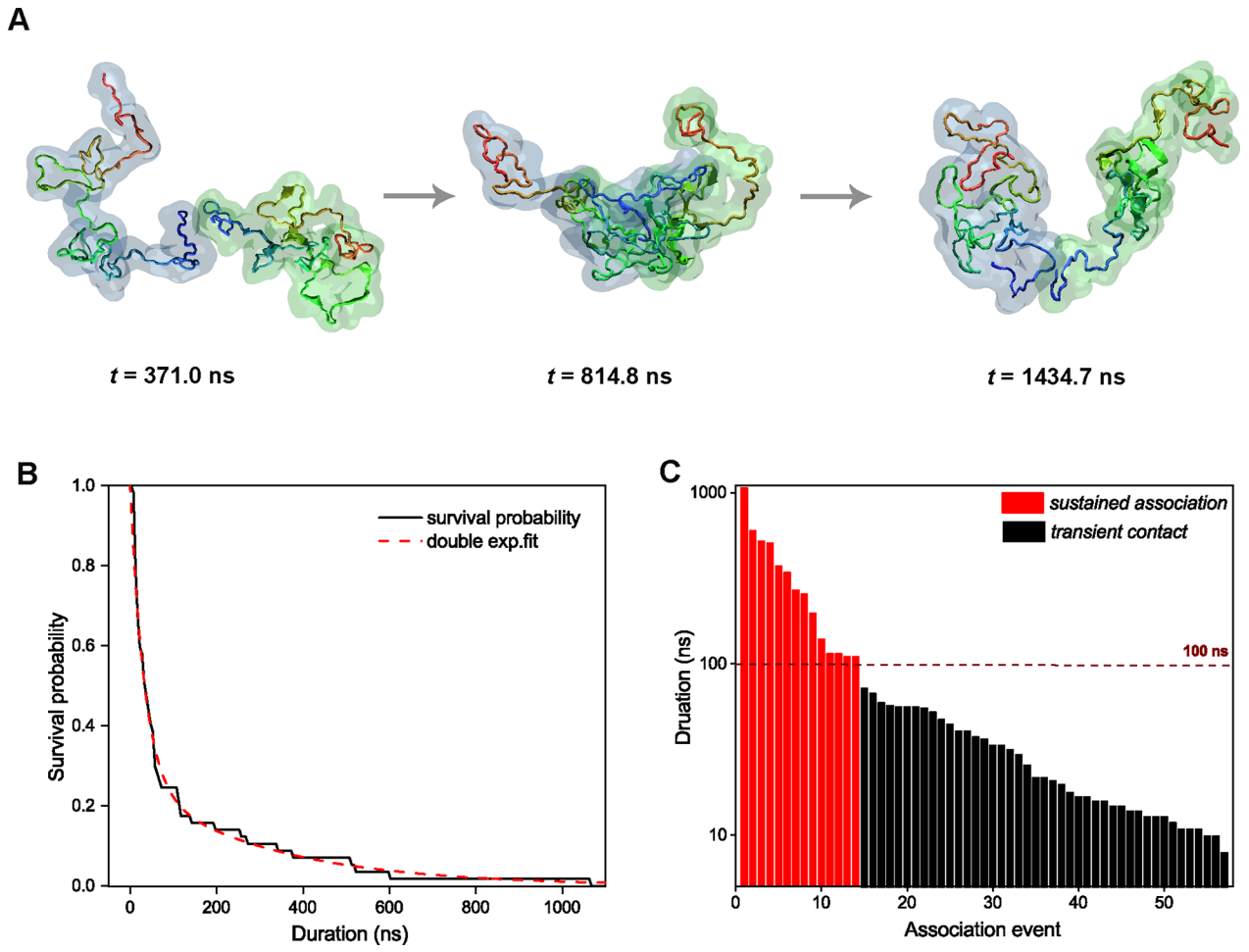


Figure 2. Reversible association between LAF-1 RGG domains. (A) Representative snapshots of a reversible association: the initial contact (at t = 371.0 ns) and the intermediate (at t = 814.8 ns) and final (at t = 1439.7 ns) stages. IDPs are colored with N- and C-terminal in red and blue, respectively. Two IDPs are highlighted by blue and green outlines separately. (B) Survival curve of the reversible association events. (C) Summary of association duration. A total of 57 association events are categorized into *sustained association* and *transient contact* with the threshold value of duration 100 ns.

compact conformation and stretched conformation, with the corresponding population of 30 and 70%. The representative conformations are shown in Figure 1B.

In addition, the asphericity (Asphe) is used to estimate the protein conformational anisotropy and defined as[54]

$$\text{Asphe} = 1 - 3(\langle I_2 \rangle / \langle I_1^2 \rangle)$$

where $I_1 = I_x + I_y + I_z$, $I_2 = I_xI_y + I_xI_z + I_zI_y$, and I_x, I_y, and I_z are the moments of inertia along three principal axes.[55] The contour map of protein conformation with respect to R_g and *Asphe* shows an apparent diagonal distribution (Figure S2), and the domain corresponding to the compact structure is well separated from other regions. In other words, R_g could largely describe the conformation ensemble of this IDP and identify the population of the compact configuration (state I). We notice that the region with the highest probability and lowest free energy distributes at around R_g = 2.25 nm and Asphe = 0.03, corresponding to the compact configuration. This structure is thus selected as the initial conformation of LAF-1 RGG domain for the subsequent simulation about the protein interaction. IDPs with a stretched structure, that is, R_g = 3.95 nm, considered as the free energy excited state, are also adopted as the initial conformation to comprehensively describe the behavior of IDP association (Figure S3).

Reversible Association between LAF-1 RGG Domains

The interaction between IDPs is considered to be essential for the formation and mediation of phase separation.[25,56−58] Because of the absence of a unique structure, the behavior of IDP interaction should be distinct from that of the structural proteins.[59] To characterize the behaviors of IDP association, we construct the system of two RGG domains (i.e., *double-protein system*). Two proteins are placed with an initial center-of-mass (COM) distance of 6 nm. 17 independent trajectories are obtained to evaluate the interaction between IDPs, and the accumulated simulation time is up to 18 μs.

RGG domains continuously undergo reversible association throughout the simulation; that is, they dissociate and subsequently diffuse before reapproaching each other (Figures S4 and S5). An effective association event is defined as the atom contact number >200 and two proteins keeping contact for more than 5 ns. A series of effective association events with varying durations are observed, and the most sustained association lasts up to 1069 ns. Figure 2A shows several representative structures of this sustained association, including the initial contact (at t = 371.0 ns) and the intermediate (t = 814.8 ns) and final (t = 1439.7 ns) stages of association.

A total of 57 effective association events are obtained. The survival probability of IDP association $S(t)$ (Figure 2B) is

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

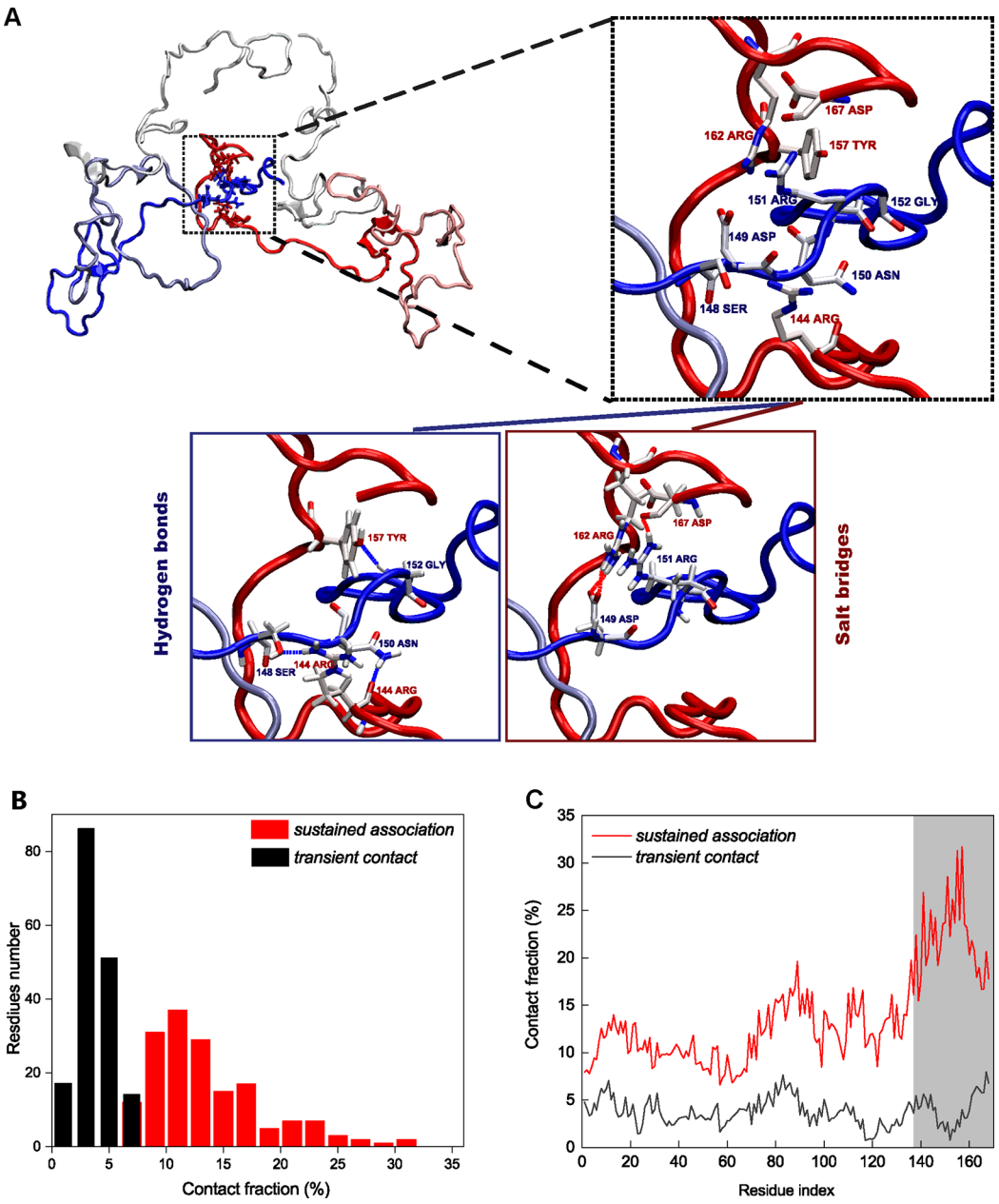


Figure 3. Interaction mode for two series of IDP associations. (A) Representative configuration of RGG domains in a *sustained association*. The IDPs are colored in white-to-red and white-to-blue (from N-terminus to C-terminus) separately. The enlarged drawing depicts the hydrogen bonds and salt bridges between two IDPs and the corresponding residues which largely distribute in the C-terminal region. (B) Distribution of the contact fraction of residues (i.e., the time fraction of being in contact with the other IDP). (C) Contact fraction of sequence profile in *sustained association* (red) and *transient contact* (black). The C-terminal region dominating the *sustained association* is highlighted by shaded area.

defined as the ratio of association events that can maintain up to time t (i.e., with a duration longer than t). We notice that the decay of survival probability can be described as a double-exponential function:

$$S(t) = A \times \exp(-t/\tau_1) + (1\ A) \times \exp(-t/\tau_2)$$

where A and (1−A) are the weighting factors for the fast and slow decay modes, and $\tau 1$ and $\tau 2$ are the time constants for the two decay modes. And the time constants $\tau 1 = 33.10 \pm 0.87$ ns and $\tau 2 = 316.56 \pm 10.60$ ns; partition coefficient $A = 0.745 \pm 0.009$. The double-exponential decay profile of the survival probability suggests there are two types of IDP association: most (i.e., A = 74.5%) are transient contacts with short duration, and the rest (25.5%) are stable associations with much longer duration. The duration of these 57 association events is summarized in Figure 2C. Clearly, the duration of these spontaneous association events is highly heterogeneous, whereby the longest association can reach 1 μs, and some transient contacts only last 8 ns. The weighting factor 1−A (25.5%) corresponds to 14 out of 57 association events, and the resulting threshold value of duration is 100 ns. In other words, 57 association events are thus categorized into two groups according to their duration, that is, sustained association (duration >100 ns) and transient contact (duration <100 ns). Moreover, the threshold value of duration is also consistent with the weighted mean of time constants based on double-exponential fitting (i.e.,τ = A$\tau 1$ + (1−A) $\tau 2$ = 105.4 ns). The evolutions of the atom contact number in several

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

representative trajectories are shown in Figure S5. The atom contact number in sustained association (upper panel) is considerably higher than in transient contacts (lower panel). It should be noted that although sustained association only accounts for 1/4 of association events, they contribute to three-quarter of accumulated association duration: 4736 ns out of 6020 ns. In other words, the sustained association dominates the IDP interaction throughout all simulations!

In the reversible IDP association, even though the *transient contact* more or less resembles the preconceived nonspecific contact between structural proteins, that is, the conventional protein−protein interactions (PPIs), the dominant long-lasting association is quite different. In other words, a considerable portion of reversible associations between IDPs may be stronger than the conventional PPIs of structural counterpart and is, therefore, conducive to the formation and maintenance of condensed phase. It should be noted that such stable association still remains reversible and does not lead to protein aggregation. To further study the mechanisms underlying the associations with different durations, especially the formation of the long-lasting association, two series of associations (i.e., *sustained association* and *transient contact*) are investigated separately. The alternative threshold values of duration (i.e., 80 and 120 ns) to discriminate *sustained association* and *transient contact* are also considered to evaluate the sensitivity of the analysis result to the choice of the threshold value (Figures S6 and S7).

Comparison of the IDP Interactions with Different Durations

Different types of interactions are considered, including hydrogen bond, salt bridge, pi−pi stacking, and cation−pi interaction[60,61] (Figure S8). There are 5.14 hydrogen bonds and 1.83 salt bridges between RGG domains, considerably more than pi-related interaction (i.e., 0.62 pi−pi stacking and 0.26 cation−pi interaction). The interaction between RGG domains can be largely attributed to the hydrogen bond and salt bridge, which is consistent with the fluorescence recovery after the photobleaching (FRAP) experiment of LAF-1 RGG domain.[40] The IDP interaction in *sustained association* is quite stable, and both hydrogen bond and salt bridge are commonly observed (89 and 57% of association duration, respectively). The average numbers of hydrogen bonds and salt bridges are 5.96 and 2.17, respectively. Figure 3A shows a representative configuration of *sustained association*, with 3 hydrogen bonds and 3 salt bridges. On the other hand, the interaction between IDPs in *transient contact* is much weaker. The occurrence of hydrogen bond and salt bridge is less frequent (55 and 19% of association duration, respectively). There are only 1.52 hydrogen bonds, corresponding to a fifth of the value in *sustained association* (5.96). Moreover, the average salt bridge number is only 0.34, only one-sixth of that in the *sustained association* (2.17).

Our results suggest that there are distinct interaction modes of IDPs with respect to duration time. In *transient contact*, the IDP interaction is rather weak and resembles the nonspecific interaction of structural proteins. On the other hand, there are considerably more hydrogen bonds and salt bridges between IDPs which are essential to maintain *sustained association*.

We further calculate the time fraction of the residues being in contact with the other IDP (Figure 3B) to characterize their propensities to participate in IDP association. The distributions of residue propensities in two series of associations are quite different. For *transient contact*, all residues share a rather low yet similar propensity to participate in IDP interaction: there are 154 out of 168 residues whose propensity is less than 6%, and the residues with a contact propensity larger than 8% are absent. As for *sustained association*, the residue propensities are higher and rather heterogeneous. For most residues (156 out of 168 residues), their propensities are larger than 8%. More strikingly, the propensity of some specific residues even exceeds 20%. In short, residues show a similarly limited tendency to be involved in the IDP interaction with short duration (i.e., *transient contact*), exhibiting the characteristic of random contact. On the other hand, there are residues with an extraordinary tendency to participate in the IDP interaction with long duration (i.e., *sustained association*).

The contact fraction of sequence profile is shown in Figure 3C. The tendency of a residue to participate in the IDP interaction with long duration (i.e., *sustained association)* is considerably heterogeneous. Interestingly, residues with high propensity (more than 20%) are distributed in the C-terminal region, especially in the region of 138−168. The average propensity of residues in the C-terminal region 138−168 is twice that of the other regions (i.e., 22% vs 12%). Although this region only accounts for 17% of the sequence, it contributes more than a third of hydrogen bonds (35%) and salt bridges (38%) (Figure S9). This region serves as the hot region of *sustained association* and is then denoted as the *contact-prone region*. In *transient contact*, the propensity of all residues participating in the IDP interaction keeps less than 8%. And the enhancement of propensity in the C-terminal region 138−168 is absent. The contributions of hydrogen bond (16%) and salt bridge (19%) of this region are similar to its sequence proportion (17%). Similar phenomena of the enrichment of residues' propensity in the C-terminal region can be observed in the *sustained association* based on the other threshold values of duration (i.e., 80 ns and 120 ns, Figures S6 and S7). Taken together, the IDP interaction in *sustained association* is considerably enhanced, and more strikingly, the enhancement of IDP interaction is accompanied by the emergence of sequence tendency!

In the C-terminal region 138−168, there are 6 positively charged residues (all ARG) and 6 negatively charged residues. The proportion of charged residues in this region is modestly higher than the overall proportion of whole protein (38% vs 26%). Similarly, the proportion of ARG which is considered to be important for the LLPS of RGG domain is mildly higher than the sequence proportion (19% vs 14%). Besides, the proportion of the polar residues, that is, SER, ASN, and GLN, in this region is also relatively higher than the overall proportion of whole protein (32% vs 21%). As for the aromatic TYR, its proportion (6.5%) is comparable to the sequence proportion (6.6%). The conformation analysis of IDP monomer based on *single-protein system* (Figure S1) shows that both termini tend to be exposed among which the C-terminal region is even more exposed than its N-terminal counterpart. Besides, relatively higher propensities of charged residues (e.g., ARG) and polar residues in the C-terminal region facilitate the formation of hydrogen bonds and salt bridges and the consequent *sustained association* between IDPs. Among these residues, ARG makes significant contribution to the salt bridge (100%) and hydrogen bond (55%). Thus, this region may serve as an essential component of LLPS engendered by the LAF-1 RGG domain. And the replacement of ARG may affect the IDP interaction, especially the

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

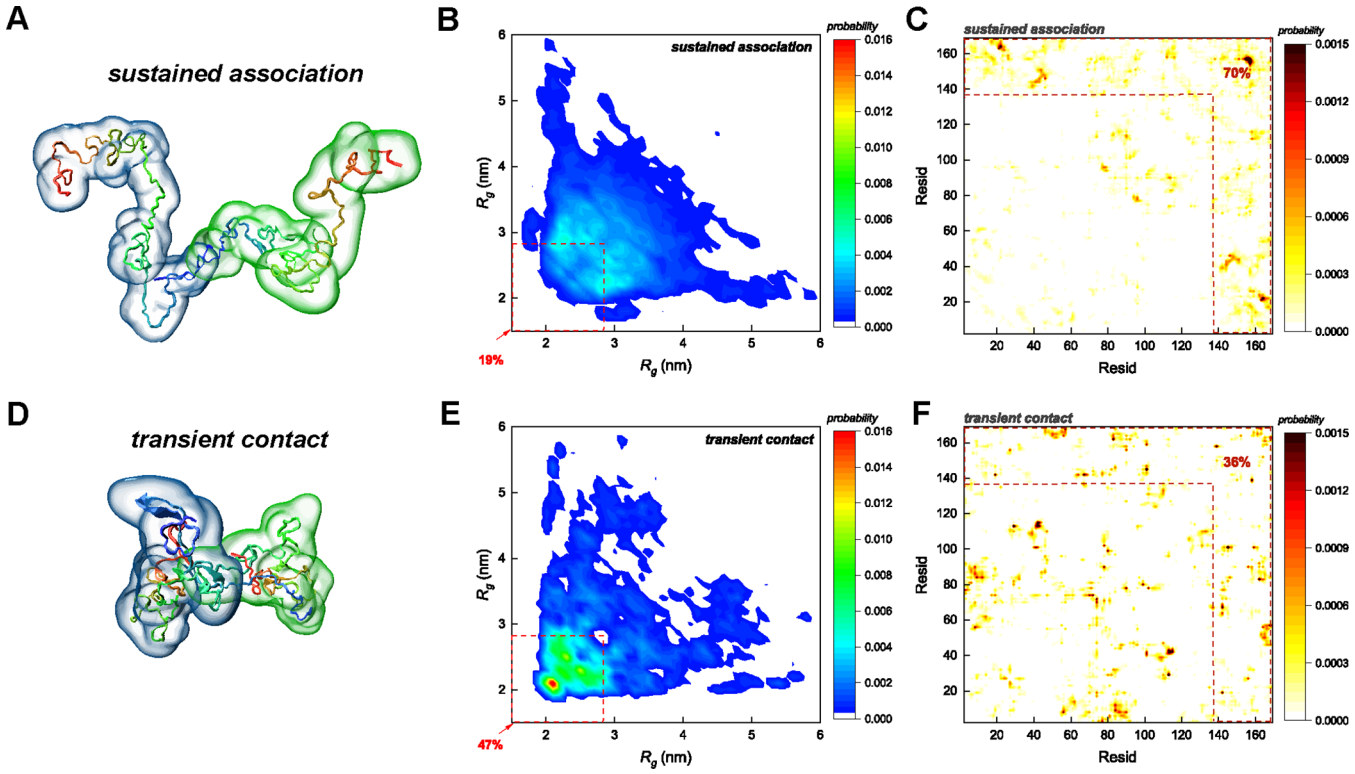


Figure 4. Conformation and contact matrix of RGG domain for two series of associations. Representative snapshots of two series of associations: *sustained association* (A); *transient contact* (D). IDPs' N- and C-terminals are colored in red and blue, respectively. And two IDPs are highlighted by blue and green outlines separately. Contour maps of the R_g of two proteins in *sustained association* (B) and *transient contact* (E). The red dotted box marks the region with R_g < 2.8 nm. Contact matrix for *sustained association* (C) and *transient contact* (F). The *contact-prone region* is marked by a red dashed line.

preference to the C-terminal region. It should be noted that the LAF-1 RGG domain we study is an IDR isolated from LAF-1 protein. To better understand the biological relevance and the interaction behavior of RGG domain as a part of LAF-1 protein, for example, the identified hot region, it is essential to discuss the possible impact of the presence of the attached folded domain of multidomain LAF-1 in a future work.

As indicated by our results, *transient contact* between IDPs resembles nonspecific random contact between structural proteins, while the IDP interaction in *sustained association* is rather stable and exhibits an exceptional sequence tendency. Residues in the C-terminal region have a higher propensity to participate in the IDP interaction which may result from the abundance of charged and polar residues and the charge neutralization. In short, even though LAF-1 RGG domain is considered to harbor low-complexity sequence, their interaction can be largely attributed to a finite hot region (denoted as *contact-prone region*). The existence of such finite hot region with similar propensity can be also found in the simulation with stretched initial structures of IDPs (i.e., R_g = 3.95 nm, Figure S3). The behaviors of sustained IDP interaction, especially the emergence of *contact-prone region*, are distinct from the nonspecific contact of structural proteins.

Conformation of IDPs in the Associations with Different Durations

Figure 4 A,D shows the representative configurations of two series of associations. The N- and C-terminals of IDP are colored red and blue, respectively. The outlines with different colors are used to distinguish two proteins. In *sustained association*, both proteins adopt relatively expanded conformation, and the corresponding R_g are 4.38 and 3.47 nm (Figure 4A). Such expanded conformation should facilitate the formation of sustained association through the exposure of a certain region, that is, *contact-prone region*. As for *transient contact*, both proteins are rather compact, and the corresponding R_g is 1.98 and 2.09 nm (Figure 4D). The tendency of IDPs to associate through a certain region thus diminishes. Additional configurations of the association of IDPs with different compactness are shown in Figure S10.

We further analyze the protein conformation and contact matrix for the association events with different durations. The intra-chain contact matrices of RGG domains of two series of associations are shown in Figure S11. The contact matrix of *sustained association* is concentrated near the diagonal, corresponding to a stretched configuration. On the other hand, the contact matrix in *transient contact* exhibits a dispersed distribution with higher propensity in the region far away from the diagonal, indicating a rather compact structure with more intra-chain contact. The distributions of the conformation of RGG domain of two series of associations are also considered (Figure 4B, *sustained association*, Figure 4E, *transient contact*). As mentioned in the previous section, the LAF-1 RGG domain undergoes considerable conformational fluctuations and exhibits a double-Gaussian distribution of R_g with the crossover region at 2.80 nm. Thus, the conformational ensemble of IDP can be roughly categorized into *stretched conformation* (i.e., R_g > 2.8 nm) and *compact conformation* (i.e., R_g < 2.8 nm), and their lifetimes are estimated to be 12.69 and 3.34 ns, respectively. The symmetrized contour map of protein conformation in the association shows a two-dimensional

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

probability density map in terms of the R_g of two proteins. The distributions of protein conformation in two series of associations are quite different. For *sustained association*, the contour map shows a broad conformational distribution. In most cases (81%), at least one protein adopts *stretched conformation* (i.e., R_g > 2.8 nm). On the other hand, the conformation distribution corresponding to *transient contact* concentrates in *compact conformation* (i.e., R_g < 2.8 nm). For around half of the duration (47%), both IDPs adopt a *compact conformation*. Similar phenomena about the prevalence of *stretched conformation* in the *sustained association* can be observed in the analysis based on the alternative threshold values of duration (i.e., 80, 120 ns; Figures S6 and S7). The analysis of R_g distribution suggests that the IDP associations with different durations are also related to their conformations. Relatively extended conformation facilitates the exposure of various regions of proteins and their involvement in IDP interactions (see also the discussion about contact fraction of residues in Figure 3C). The C-terminal region which can mediate a stable IDP interaction can thus have a considerably enhanced tendency to IDP interaction. As for the compact conformation, the reduced exposure of protein suppresses the involvement of such regions to IDP interaction: their involvements are restraint by their exposures on protein surface. The sequence tendency should thus be negligible. Consequently, there are limited hydrogen bonds and salt bridges between IDPs which resembles nonspecific transient contact between structural proteins.

To validate the impacts of conformation fluctuation on the involvement of C-terminal region and the association duration, the symmetrized contact matrices of two series of associations are analyzed separately (Figure 4C,F). The contact distribution of *sustained association* is apparently concentrated in the C-terminal *contact-prone region* (31 residues) with an occupancy of 70%, much more than the other regions (the residue combination of the rest 137 residues only accounts for 30% of residue contacts). For comparison, the contact matrix of *transient contact* is rather homogeneous. The C-terminal region 138−168 only contributes 36% of residue contacts, close to the ratio of the residue combination with the C-terminal *contact-prone region* involved (33%). Taken together, conformational fluctuations, especially the extended conformation, serve as the prerequisite for sustained association by favoring the exposure and the involvement of the C-terminal *contact-prone region*. Our results about the existence of finite hot region and the relevance of conformational fluctuations could provide direct evidence for the perspective "stickers" in IDPs[62,63] as well as the possible mechanism underlying the emergence of such stickers.

To better illustrate the impact of conformational fluctuations on the formation of sustained association, we analyze the relationship of IDP conformation in the early stage of association and the duration throughout all association events. The average R_g of IDPs in the early stage, that is, 2 ns before and after the occurrence of contact, are calculated, and the larger one is considered. All 57 association events can thus be organized according to the IDP conformation in their early stage, that is, the association with the involvement of the IDP adopting *stretched conformation* (R_g > 2.8 nm) and the association with both IDPs in *compact conformation* (R_g < 2.8 nm). The correlation coefficient between the dichotomy index of the IDP conformation and the association type with respect to the duration (i.e., *sustained association* vs *transient contact*) are then calculated (Figure S12). The correlation coefficient for all association events is 0.31, suggesting a certain correlation between association duration and the involvement of IDP with extended conformation. More strikingly, IDPs with *stretched conformation* are found to be involved in all *sustained associations*. In other words, *stretched conformation* is the necessary condition of *sustained association*. On the other hand, IDPs with *stretched conformation* are also involved in several *transient contacts*. Hence, involvement of *stretched conformation* does not guarantee *sustained association*. Taken together, the association duration is highly affected by the IDP conformation in the early stage; *stretched conformation* serves as a necessary but not sufficient condition of *sustained association*. The stable IDP interaction in the *sustained association* exclusively results from the *stretched conformation*. Such conformation facilitates the exposure and subsequent interaction of the *contact-prone region*. The *stretched conformation* is kinetically more stable (its lifetime is longer than the compact counterpart, 12.69 ns vs 3.34 ns) and is conducive to the accomplishment of sufficient association. Besides, the expanded conformation is necessary for IDP to form entanglement which can also lead to sustained association (Figure S13). Hence, the expanded conformation in the early stage is indispensable for the sustained IDP association.

As mentioned above, *stretched conformation* is also found to be involved in the *transient contact*. In some cases, the exposed C-terminal region 138−168 is also involved in the interaction (Figure S10). However, the C-terminal region interacts with the N-terminal G-rich region of the partner IDP which is barely involved in the formation of hydrogen bond and salt bridge. In other words, the interaction of C-terminal region with the somewhat neutral region (i.e., G-rich region) of the partner IDP cannot accomplish the stable association. On the other hand, during the *sustained association*, the exposed C-terminal region tends to interact with more active regions, for example, the C-terminal region of the partner IDP. In addition, the analysis of the conformation of IDP in *sustained association* (Figure 4B) suggests such association may in turn stabilize stretched conformation.

As indicated by our results, the conformational fluctuations of IDPs are highly related to their interaction. The compact conformation tends to form transient random contact. By contrast, the expanded conformation facilitates association through the interaction hot region or the protein entanglement, both of which result in the association with sustained duration. Hence, such *stretched conformation* can serve as the necessary condition for *sustained association*. And such stable association may reciprocally stabilize *stretched conformation* (Figure 4B). As indicated by the previous study of FUS IDP,[33] IDP interaction can increase secondary structure content. The interaction of LAF-1 RGG domain modestly increases the content of secondary structure, including β-turn, β-sheet, and α-helix, from 11.1% (monomer) to 11.9% (during association), while the contents in two series of associations are almost the same: 11.9% (*sustained association*) and 11.8% (*transient contact*). It is well recognized that the IDP interaction is crucial to the formation and evolution of LLPS wherein the conformational fluctuations of IDPs are considered to extend their interaction range. Our results about LAF-1 RGG domain suggest that the impact of inherent conformational fluctuations of IDPs can be more profound, including the exposure of *contact-prone region* and the enhanced stability of IDP association as mediated by this region. Such *contact-prone*

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

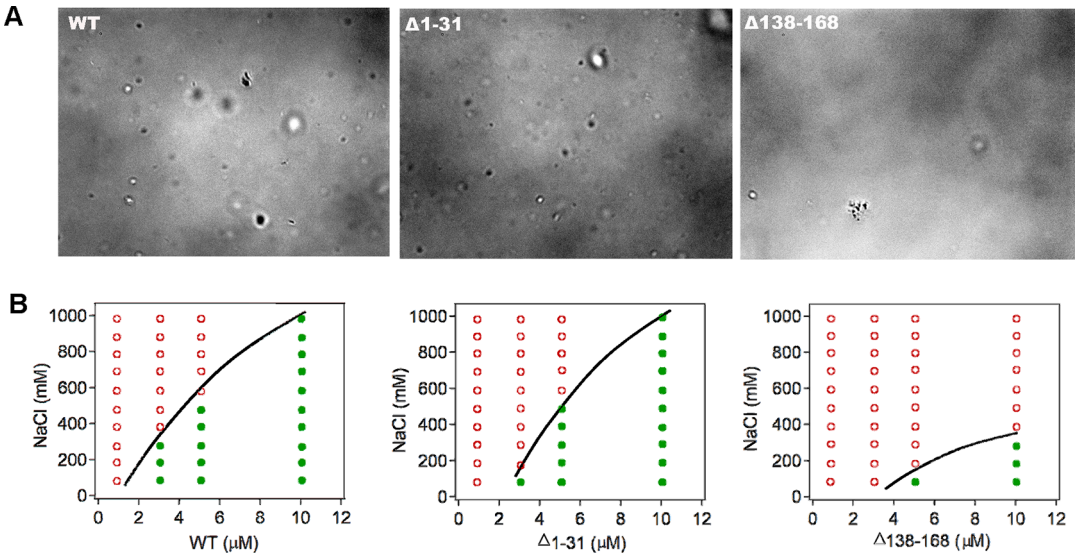


Figure 5. C-terminal region 138−168, which serves as the hot region for IDP interaction, is critical for liquid−liquid phase separation (LLPS). (A) DIC image of phase separated droplets formed by the LAF-1 RGG domain (WT) and the variants with deletion of residues 1−31 (Δ1−31) and 138−168 (Δ138−168). The experiment is performed at room temperature. (B) Corresponding phase diagrams as a function of salt and protein concentrations. Red circles indicate no phase separation; green dots indicate phase separation (line drawn to guide the eye).

region emerging from protein conformational fluctuations can significantly affect IDP interactions and may further facilitate the formation and stabilization of LLPS.

Experimental Verification

As described by our simulation results, the C-terminal region 138−168 of LAF-1 RGG domain serves as the hot region for sustained IDP association which essentially dominates the IDP interaction and may eventually affect LLPS. To validate the contribution of this region to the LLPS mediated by IDP interaction, we express and purify LAF-1 RGG domain (WT) and its variants with the deletion of the C-terminal region 138−168 (Δ138−168) for phase separation experiment. And we also construct the additional variant with N-terminal region 1-31 deleted (Δ1−31) for comparison (Figures 5, Figures S14 and S15). Experiments show that Δ1−31 can form micro-sized spherical phase-separated droplets, similar to WT. On the other hand, only a few of Δ138−168 droplets can be observed, indicating that the deletion of the C-terminal region 138−168 obviously weakens the phase-separation ability of RGG domain (Figure 5A). Moreover, we built a phase diagram by measuring whether phase separation droplets could be observed under different salt conditions (Figure 5B), which has been widely exploited to describe the protein phase-separation ability and the underlying intermolecular interactions.[15,16,28,29,56,57] WT and Δ1−31 share similar phase diagram and exhibit similar phase behavior, while the phase diagram of Δ138−168 is apparently different. For any given salt concentration, the critical protein concentration of Δ138−168 variant droplets is much higher than for WT and Δ1−31 variant (Figure 5B). At 100 mM salt, the critical protein concentration for both WT and Δ1−31 variant is 3 μM, while the critical concentration for Δ138-168 variant increases by 66% (5 μM). The difference is even more pronounced at 200 mM salt: the critical concentration of Δ138−168 variant significantly increases to 10 μM, more than three-fold of WT which remains 3 μM. Thus, the critical concentration of Δ138−168 variant should be much higher than WT and Δ1−31 variant in the physiological solution (i.e., 150 mM salt). Taken together, the deletion of residues 138−168 significantly decreases the phase-separation ability which is attributed to the reduced protein−protein interaction.[29,40,64] Considering the prevalence of hydrogen bond and salt bridge in IDP association, it is essential to simulate the IDP interaction in the solution with higher salt concentration in the near future. The phase separation experiment clearly illustrates that the C-terminal region 138−168 serves as the critical region to regulate the phase separation, while the N-terminal region 1−31 is largely absent for the phase separation. Both agree with our simulation results that C-terminal is important to the IDP interaction and the resulting LLPS, while N-terminal is rather neutral (Figures 3 and S1).

CONCLUSIONS

In this work, we study the reversible association between LAF-1 RGG domains and the impacts of their sequence and conformational characteristics. In addition to the contact events with a duration of tens of nanoseconds (i.e., *transient contact*) which more or less resemble nonspecific contact between structural proteins, there are association events with considerable stability and sustained duration up to micro-seconds (i.e., *sustained association*). IDPs form multiple hydrogen bonds and salt bridges in these long-lasting association events which essentially dominate the dynamics of IDP interaction. More interestingly, the propensity of residues participating in the IDP interaction is highly heterogeneous. And there is a finite region in this low-complexity IDP that exhibits an exceptional tendency (more than twice of the other regions) to participate in the protein interaction, that is, the C-terminal region 138−168, denoted as the *contact-prone region*. This region is rich in charged/polar residues and is overall neutral, which is conducive to the hot region for IDP interaction. In other words, the enhancement of IDP interaction is accompanied by the emergence of such sequence tendency. Previous studies suggested that LLPS was largely driven by the IDP sequence featured by the alternative arrangement of positively and negatively charged residues.[29,64]

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

Our findings about the *contact-prone region* in LAF-1 RGG domain illustrate that the abundance of charged residues and overall neutrality in a finite region (31 residues) are essential for such unique protein association with substantial stability. Hence, this finite region should largely affect the interaction of the whole IDP chain and even affect the behavior of LLPS. Its importance to the intermolecular interaction and the resulting LLPS has been confirmed by the phase-separation experiment of the variant Δ138−168. Moreover, the association of IDPs is entwined with their inherent conformational fluctuations. The expanded conformation serves as the necessary but not sufficient condition for *sustained association*: it is conducive to the emergence of *contact-prone region* which mediates such stable association. On the other hand, all residues share a similar tendency to participate in *transient contact* of relatively compact proteins.

In summary, our results identify a finite *contact-prone region* in LAF-1 RGG domain which exhibits an apparent tendency to the interaction with other IDPs and is essential to the *sustained association*. The emergence of such a hot region in the preconceived low-complexity IDP is attributed to its inherent conformational fluctuation. Thus, the seemingly random interaction of IDP is actually dominated by the stable associations mediated by finite regions. All these features are distinct from the transient interaction of structural proteins. There is growing evidence that the behavior of IDP monomer is related to its collective behavior.[43,65] Our findings about the IDP association will be helpful in better understanding the mechanism about the interaction as well as the formation and mediation of LLPS, especially for a series of IDPs with exceptionally low critical concentration. And it will be interesting to explore the existence of such a hot region of IDP interaction (i.e., *contact-prone region*) as well as the relevance of conformational fluctuations and key residues (e.g., arginine and tyrosine) in other IDPs, especially other RGG domains. Previous studies about the N-terminal DDX4 indicated that its LLPS could be attenuated by methylation of a few arginine residues in a finite sequence region.[37] Another study about the TDP-43 C-terminal domain showed that the ALS mutations within the 321−340 region could significantly disrupt phase separation.[66] The concept of a hot region of IDP interaction is helpful to understand the impacts of modification and mutation of these IDPs and to design mutations (e.g., the mutation of key residues like arginine and tyrosine in the hot region) to effectively regulate the LLPS behavior of other IDPs. Besides, the existence of a hot region provides new insights for the development of a characterization method for the behavior of IDP interaction. For example, such a region can be labeled with fluorescence dye; then the IDP interaction and the involvement of this region can be effectively characterized by the single molecular technique, including fluorescence recovery after photobleaching (FRAP) and fluorescence resonance energy transfer in a quantitative way.

METHODS

Conformation of Monomeric LAF-RGG Domain

LAF-1 is a member of the DDX3 family of RNA helicases.[15] It is the main component of the membraneless organelles P granules in *C. elegans* embryos.[15,21] LAF-1 contains an intrinsically disordered, low-complexity RGG domain composed of 168 residues at its N-terminal region. This domain is rich in arginine (R) and glycine (G) andconsidered as an essential ingredient for liquid−liquid phase separation of LAF-1.[15,52] Moreover, previous experimental studies have shown that the isolated RGG domain can undergo liquid−liquid phase separation in vitro.[15] The effective concentration in the condensed phase is very low (0.23−0.45 mM), suggesting that the LAF-1 RGG domain is very efficient to engender LLPS.[21] Therefore, the RGG domain has been selected for studying the interactions between IDPs and their relevance to the phase separation.

The LAF-1 RGG protein is constructed by Pymol[67−69] and then subjected to 100 ps relaxation simulation in vacuum. The resulting conformation is placed in a cubic box with the dimension of 16 nm and solvated by water molecules (i.e., *single-protein system*). Sodium and chloride ions are added to neutralize the systems and mimic physiological conditions (150 mM NaCl). The solvated system is subjected to 50,000 steps of energy minimization, followed by 10 ns *NVT* and 10 ns *NPT* simulation. A 2000 ns production run is performed to investigate the conformational ensemble of RGG domain.

Reversible Association between LAF-RGG Domains

In order to study the interaction between RGG domains, we further construct the system with two proteins (i.e., *double-protein system*) by adopting the most probable conformation obtained from the simulation of *single-protein system* (i.e., R_g = 2.25 nm). It should be noted that the IDP association in *double-protein* system is reversible wherein they undergo dissociation and subsequent diffusion before they reapproach each other (Figures S4A and S5). Hence, the interaction modes of IDP with diverse conformations are effectively considered in this work (Figure S4B). Two proteins are placed in a 16 nm × 16 nm × 16 nm box with an initial center-of-mass (COM) distance of 6 nm; the protein concentration is 16.9 mg/mL (∼0.8 mM), which is on the same order of magnitude (∼10^{-1} mM) as the RGG concentration in the droplets that condense in vitro.[21] Appropriate amounts of sodium and chloride ions are added to neutralize the system and mimic physiological conditions (150 mM NaCl). Equilibrium procedure is similar to the *single-protein system*. A total of 17 independent simulations are preformed to evaluate the interaction between IDPs. The sustained protein association events are observed in two independent simulations. These two simulations are thus performed for 1500 ns to describe the whole association event. The other 15 simulations are conducted for 1000 ns. To comprehensively describe the behavior of IDP association, additional simulations are conducted by considering the free energy excited state of the *single-protein system* (i.e., R_g = 3.95 nm). 5 independent simulations are performed, and the accumulated simulation time is 5 μs (Figure S3).

Computational Details

All molecular dynamics (MD) simulations are performed using GROMACS (version 2018.3)[70] molecular modeling package, and VMD[71] is used for trajectory visualization. The a99SB-disp force field[72] is applied for protein, and the TIP4P-DE[72] water model is adopted. The a99SB-disp force field with the TIP4P-DE water model is obtained by optimizing the torsion parameters and introducing small changes in the vdW interaction terms from the a99SB-ILDN force field[73] with the TIP4P-D water model.[74] The v-rescale thermostat[75] and the Parrinello-Rahman barostat[76,77] are adopted to control the temperature and pressure at 300 K and 1 bar, respectively. The particle-mesh Ewald method is used to calculate the long-range electrostatic interaction,[78] while the short-range electrostatic interaction and vdW interactions are calculated with a cutoff of 1.2 nm, and the periodic boundary conditions are applied in all three directions.[79] The covalent bonds with hydrogen atoms are constrained by the LINCS algorithm,[80] which allows a time step of 2 fs.

Protein Engineering, Expression, and Purification

The genes for LAF-1 RGG WT, LAF-1 RGG △1−31, LAF-1 RGG △70−100, and LAF-1 RGG △138−168 are custom synthesized codon-optimized for expression in Escherichia Coli (GenScript, China). The genes encoding the chimeras are constructed in pQE80L vectors using standard molecular biology techniques. The constructed plasmids are sequenced to confirm the correct protein sequence. The proteins are expressed in *E. coli* (BL21) and purified by Co^{2+}-affinity

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

chromatography. The proteins are dialyzed into the solution containing 500 mM NaCl and 20 mM Tris (pH 7.5) at room temperature to avoid the phase separation and lyophilized before use.

Microscopic Imaging

For microscopy experiments, protein aliquots are mixed with buffer (20 mM Tris, pH 7.5, 0−150 mM NaCl) to obtain solutions containing the desired protein and NaCl concentrations. Protein solutions containing 4 M urea are specially prepared to avoid the phase separation and used to determine the protein concentrations based on OD_{280nm} using a Nanodrop spectrophotometer (Thermo-Fisher). All the micrographs are obtained using an OLYMPUS-IX73 microscope (OLYMPUS, USA).

DATA AND MATERIAL AVAILABILITY

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supporting Information Additional data related to this paper may be requested from the authors.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.2c00414.

Sequence characteristics and conformational features of the LAF-1 RGG domain in the *single-protein system*, double-Gaussian fitting parameters of *Rg* distribution of the monomeric LAF-1 RGG domain, simulation with different initial structures of IDPs, conformation of proteins in the early stage of association, several reversible association events observed in the simulation, interaction mode and conformation of the RGG domain, different types of IDP interaction, phenomena of enrichment of interaction in the C-terminal region, snapshots of IDP associations with different protein compactness, intra-chain contact matrix for protein in *sustained association* and *transient contact*, correlation analysis of IDP conformation and association duration, sustained association formed by entanglement between two stretched proteins, Coomassie blue stained SDS-PAGE photograph for the proteins, and LLPS behavior of LAF-1 RGG WT and variants (PDF)

(PDF)

AUTHOR INFORMATION

Corresponding Authors

Wei Wang − *Collaborative Innovation Center of Advanced Microstructures, National Laboratory of Solid State Microstructure, Department of Physics, Nanjing University, Nanjing 210093, China;* orcid.org/0000-0001-5441-0302; Email: wangwei@nju.edu.cn

Jingyuan Li − *Zhejiang Province Key Laboratory of Quantum Technology and Device, School of Physics, Zhejiang University, Hangzhou 310058, China;* orcid.org/0000-0003-2926-1864; Email: jingyuanli@zju.edu.cn

Authors

Moxin Zhang − *Zhejiang Province Key Laboratory of Quantum Technology and Device, School of Physics, Zhejiang University, Hangzhou 310058, China*

Bin Xue − *Collaborative Innovation Center of Advanced Microstructures, National Laboratory of Solid State Microstructure, Department of Physics, Nanjing University, Nanjing 210093, China*

Qingtai Li − *Collaborative Innovation Center of Advanced Microstructures, National Laboratory of Solid State Microstructure, Department of Physics, Nanjing University, Nanjing 210093, China*

Rui Shi − *Zhejiang Province Key Laboratory of Quantum Technology and Device, School of Physics, Zhejiang University, Hangzhou 310058, China;* orcid.org/0000-0002-0411-3067

Yi Cao − *Collaborative Innovation Center of Advanced Microstructures, National Laboratory of Solid State Microstructure, Department of Physics, Nanjing University, Nanjing 210093, China;* orcid.org/0000-0003-1493-7868

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.2c00414

Author Contributions

M.Z., and B.X. contributed equally. J.L. and W.W. designed the study. M.Z. and J.L. performed MD simulations and analysis. B.X. and Q.L. performed the phase separation experiments and data analysis. All authors discussed the results and contributed to the writing of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Yanee Wutthinitikornkit, Shiyan Lin, and Yuchuan Zheng for helpful discussions. We are also grateful for the computational resources provided by the supercomputer TianHe-1A in Tianjin, China. Funding: This work was supported by the National Natural Science Foundation of China (NSFC) (Grant nos. 12175195 and 11874319), Natural Science Foundation of Jiangsu Province (No. BK20220120).

REFERENCES

(1) Stelzl, U.; et al. A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell* **2005**, *122*, 957−968.

(2) Rual, J.-F.; et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **2005**, *437*, 1173−1178.

(3) Zhang, C.; Freddolino, P. L.; Zhang, Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res.* **2017**, *45*, W291−W299.

(4) Rodrigues, C. H. M.; Pires, D. E. V.; Ascher, D. B. pdCSM-PPI: Using Graph-Based Signatures to Identify Protein-Protein Interaction Inhibitors. *J. Chem. Inf. Model.* **2021**, *61*, 5438−5445.

(5) Santos, R.; et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **2017**, *16*, 19−34.

(6) Scott, D. E.; et al. Small molecules, big targets: drug discovery faces the protein−protein interaction challenge. *Nat. Rev. Drug Discovery* **2016**, *15*, 533−550.

(7) Rickard, M. M.; et al. In-Cell Protein−Protein Contacts: Transient Interactions in the Crowd. *J. Phys. Chem. Lett.* **2019**, *10*, 5667−5673.

(8) Watkins, A. M.; Bonneau, R.; Arora, P. S. Side-Chain Conformational Preferences Govern Protein−Protein Interactions. *J. Am. Chem. Soc.* **2016**, *138*, 10386−10389.

(9) Wuo, M. G.; Mahon, A. B.; Arora, P. S. An Effective Strategy for Stabilizing Minimal Coiled Coil Mimetics. *J. Am. Chem. Soc.* **2015**, *137*, 11618−11621.

(10) Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **2014**, *83*, 553−584.
(11) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **2015**, *11*, 899−904.
(12) Uversky, V. N. Introduction to Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6557−6560.
(13) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18−29.
(14) Csizmok, V.; et al. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. *Chem. Rev.* **2016**, *116*, 6424−6462.
(15) Elbaum-Garfinkle, S.; et al. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 7189−7194.
(16) Molliex, A.; et al. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **2015**, *163*, 123−133.
(17) Riback, J. A.; et al. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **2017**, *168*, 1028−1040.
(18) Uversky, V. N.; et al. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *Febs Letters* **2015**, *589*, 15−22.
(19) Altmeyer, M.; et al. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun.* **2015**, *6*, 8088.
(20) Majumdar, A.; et al. Liquid−Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules. *J. Phys. Chem. Lett.* **2019**, *10*, 3929−3936.
(21) Wei, M. T.; et al. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.* **2017**, *9*, 1118−1125.
(22) Wegmann, S.; et al. Tau protein liquid-liquid phase separation can initiate tau aggregation. *Embo Journal* **2018**, *37*, No. e98049.
(23) Reichheld, S. E.; et al. Direct observation of structure and dynamics during phase separation of an elastomeric protein. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E4408−E4415.
(24) Bergeron-Sandoval, L. P.; Safaee, N.; Michnick, S. W. Mechanisms and Consequences of Macromolecular Phase Separation. *Cell* **2016**, *165*, 1067−1079.
(25) Murthy, A. C.; et al. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **2019**, *26*, 637−648.
(26) Patel, A.; et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162*, 1066−1077.
(27) Wang, J.; et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 688−699.
(28) Lin, Y.; et al. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **2015**, *60*, 208−219.
(29) Tsang, B.; et al. Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 4218−4227.
(30) Brady, J. P.; et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E8194−E8203.
(31) Martin, E. W.; et al. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694−699.
(32) Burke, K. A.; et al. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* **2015**, *60*, 231−241.
(33) Murray, D. T.; et al. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **2017**, *171*, 615−627.
(34) Borgia, A.; et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555*, 61−66.
(35) Wang, W. N. Recent advances in atomic molecular dynamics simulation of intrinsically disordered proteins. *Phys. Chem. Chem. Phys.* **2021**, *23*, 777−784.
(36) Wen, J.; et al. Conformational Expansion of Tau in Condensates Promotes Irreversible Aggregation. *J. Am. Chem. Soc.* **2021**, *143*, 13056−13064.
(37) Nott, T. J.; et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **2015**, *57*, 936−947.
(38) Dignon, G. L.; et al. Temperature-Controlled Liquid−Liquid Phase Separation of Disordered Proteins. *ACS Central Science* **2019**, *5*, 821−830.
(39) McCarty, J.; et al. Complete Phase Diagram for Liquid-Liquid Phase Separation of Intrinsically Disordered Proteins. *J. Phys. Chem. Lett.* **2019**, *10*, 1644−1652.
(40) Schuster, B. S.; et al. Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 11421−11431.
(41) Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles. *Phys. Rev. Lett.* **2016**, *117*, 178101.
(42) Das, S.; et al. Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters. *Phys. Chem. Chem. Phys.* **2018**, *20*, 28558−28574.
(43) Dignon, G. L.; et al. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Computational Biology* **2018**, *14*, No. e1005941.
(44) Benayad, Z.; et al. Simulation of FUS Protein Condensates with an Adapted Coarse-Grained Model. *J. Chem. Theory Comput.* **2021**, *17*, 525−537.
(45) Joseph, J. A.; et al. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science* **2021**, *1*, 732−743.
(46) Henriques, J.; et al. On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J. Mol. Biol.* **2018**, *430*, 2521−2539.
(47) Baul, U.; et al. Sequence Effects on Size, Shape, and Structural Heterogeneity in Intrinsically Disordered Proteins. *J. Phys. Chem.B* **2019**, *123*, 3462−3474.
(48) Shrestha, U. R.; Smith, J. C.; Petridis, L. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Communications Biology* **2021**, *4*, 243.
(49) Tesei, G.; et al. Accurate model of liquid−liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, No. e2111696118.
(50) Thomasen, F. E.; et al. Improving Martini 3 for Disordered and Multidomain Proteins. *J. Chem. Theory Comput.* **2022**, *18*, 2033−2041.
(51) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein−Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.
(52) Brangwynne, C. P.; et al. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **2009**, *324*, 1729−1732.
(53) Voronina, E.; et al. RNA Granules in Germ Cells. *Cold Spring Harbor Perspectives in Biology* **2011**, *3*, a002774.
(54) Kanchi, S.; et al. Molecular Dynamics Study of the Structure, Flexibility, and Hydrophilicity of PETIM Dendrimers: A Comparison with PAMAM Dendrimers. *J. Phys. Chem. B* **2015**, *119*, 12990−13001.

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104

(55) Rudnick, J.; Gaspari, G. The aspherity of random walks. *J. Phys.-Mathematical and General* **1986**, *19*, L191−L193.
(56) Lee, K.-H.; et al. C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* **2016**, *167*, 774−788.
(57) Yang, P. G.; et al. G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **2020**, *181*, 325−345.
(58) Boeynaems, S.; et al. Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties. *Proc Natl Acad Sci U S A* **2019**, *116*, 7889−7898.
(59) Mathieu, C.; Pappu, R. V.; Taylor, J. P. Beyond aggregation: Pathological phase transitions in neurodegenerative disease. *Science* **2020**, *370*, 56−60.
(60) Vernon, R. M.; et al. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **2018**, *7*, No. e31486.
(61) Das, S.; et al. Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 28795−28805.
(62) Lin, Y.-H.; et al. Assembly of model postsynaptic densities involves interactions auxiliary to stoichiometric binding. *Biophys. J.* **2022**, *121*, 157−171.
(63) Mittag, T.; Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol. Cell* **2022**, *82*, 2201−2214.
(64) Hofweber, M.; et al. Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* **2018**, *173*, 706−719.
(65) Amin, A. N.; et al. Analytical Theory for Sequence-Specific Binary Fuzzy Complexes of Charged Intrinsically Disordered Proteins. *J. Phys. Chem.B* **2020**, *124*, 6709−6720.
(66) Conicella, A. E.; et al. ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **2016**, *24*, 1537−1549.
(67) Schrodinger, L. J. S.*The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint* version 1.8; LLC: New York, NY, 2015.
(68) Schrödinger, L.*The PyMOL Molecular Graphics System* version 1.8, 2015.
(69) Schrodinger, L. J. V.. *The JyMOL molecular graphics development component*, 2015; Vol. *1*, p 8.
(70) Abraham, M. J.; et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.
(71) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics Modell.* **1996**, *14*, 33−38.
(72) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758−E4766.
(73) Lindorff-Larsen, K.; et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins-Structure Function and Bioinformatics* **2010**, *78*, 1950−1958.
(74) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
(75) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(76) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(77) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(78) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(79) Dolan, E. A.; et al. Simulations of membranes and other interfacial systems using P2(1) and pc periodic boundary conditions. *Biophys. J.* **2002**, *82*, 2317−2325.
(80) Hess, B.; et al. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

https://doi.org/10.1021/jacsau.2c00414
JACS Au 2023, 3, 93−104