A biophysical model of molecular clusters: interplay of multivalency, membrane localization, steric hindrance, molecular flexibility and intracellular crowding

Aniruddha Chattarraj, Madeleine Youngstrom and Leslie M. Loew*

R.D. Berlin Center for Cell Analysis and Modeling

University of Connecticut School of Medicine

Farmington, CT 06030 USA

*correspondence: les@volt.uchc.edu

Abstract

Dynamic molecular clusters are assembled through weak multivalent interactions and are platforms for a variety of critical cellular functions, most importantly receptor-mediated signaling mechanisms. However, the molecular and cellular biophysical rules that control cluster kinetics, size and composition are not well understood. Modeling can offer insights, but molecular dynamics calculations would be excessively expensive because of the size of these systems and the > ms time scales for their formation; cellular reaction diffusion simulations, on the other hand, do not have the requisite molecular detail, failing to account for steric effects or molecular flexibility. To address this problem, we developed mesoscale kinetic Langevin dynamics models using the SpringSaLaD software. We used coarse grain models of a ternary molecular system based on the membrane-bound receptor nephrin, the adaptor protein Nck and the actin nucleation promoting factor NWASP. Kinetics of cluster formation and their steady state size distributions were analyzed as a function of molecular and cellular features using a reference system consisting of 36 molecules. The distribution at steady state favored stoichiometries that optimized binding interaction, but still was quite broad. A balance of enthalpy and entropy limited the number of molecules per cluster, with complete annealing into a single 36-mer complex being exceedingly rare. Domains close to binding sites sterically inhibited clustering much less than terminal domains because the latter effectively restrict access to the interior of nascent clusters. Increased flexibility of the interacting molecules diminished clustering by shielding binding sites within compact conformations. Membrane association of nephrin increased the cluster size distribution in a surface density-dependent manner. These properties provide insights into how molecular clusters can function to localize and amplify cell signaling.

Introduction

Dynamic molecular clusters can form through weak binding interactions between multivalent molecules. They are highly plastic structures with a distribution of stoichiometries and sizes; they are becoming increasingly recognized as molecular platforms to drive key cellular functions, especially receptor-mediated signaling [2-12]. We and others use the term “ensemble” [4, 6, 7, 12-14] to specifically convey the notion of dynamic composition and size. For example, the epidermal growth factor receptor (EGFR) dimerizes and develops kinase activity when it binds its ligand, resulting in multiple phosphorylated sites on the cytoplasmic domains; these in turn interact with multiple SH2 domains on other multivalent scaffold or adaptor proteins, which then recruit additional binding partners [15]. The result is the formation of multi-molecular dynamic ensembles with variable composition and size, but with robust and specific cell signaling functions. This general pattern applies to the entire class of receptor tyrosine kinases as well as other multi-molecular systems such as T-cell and B-cell receptors[16-19], focal adhesion complexes[20], or the post-synaptic densities of dendritic spines [14]. Formation of molecular ensembles is also a prerequisite for the phenomenon of liquid-liquid phase separation, which has become a major focus of cell biophysics research [21, 22]. And even cellular polymers such as F-actin and microtubules qualify as molecular ensembles because they display tremendous combinatorial complexity and plasticity, not

just because of the infinite number of possible polymer sizes, but also because of the plethora of binding proteins that can interact at any site along the polymer [23, 24].

The physical and chemical properties of molecular ensembles are not well understood, and are likely to be far from those of dilute molecules in solution. Multivalency underlies the formation of clusters. It serves to increase local concentration of biomolecules and trigger downstream signaling events. Multivalency can lead to the formation of large molecular clusters or polymers even when the individual binding affinities are weak – a form of cooperativity. Biophysical models of these systems may serve to address the interplay of valency and geometry by systematically varying the cellular and molecular features underlying these properties. Stochastic reaction-diffusion modeling at the cellular scale cannot fully account for steric effects or molecular flexibility, because molecules are typically modeled as infinitesimal points in space[25-27]. In principle, multimolecular/multistate interactions could be modeled with molecular dynamics simulations, but the large sizes of these systems and the need to simulate on the second timescale make such simulations computationally impractical. Such simulations might be handled through coarse-graining [28], where individual proteins are modeled as a collection of linked domains [29-31].

In this work, we explore the interplay of multivalency and spatial effects with SpringSaLaD [31], a modeling and simulation software application developed in our lab that bridges the scales between molecular dynamics and cellular modeling. SpringSaLaD uses a Langevin dynamics formulation for linking spherical domains (or “sites”) with stiff springs to transmit the random diffusion-derived forces impinging on each site. It uses a rule-based modeling[32, 33] scheme to build models based on defining the states of and interaction between individual sites within the molecules. Stochastic reaction-diffusion simulations are then run within a rectangular spatial domain. SpringSaLaD also has strong integrated analysis and visualization and a unique exact formulation to relate bimolecular macroscopic on-rates to the microscopic reaction probabilities.

We use SpringSaLaD to examine a prototypical multivalent system with a membrane anchored protein containing 3 phosphotyrosine (pTyr) sites, an adaptor protein containing one SH2 domain and 3 SH3 domains, and an effector protein consisting of 6 proline-rich motifs (PRMs)[34]. This model is inspired by the nephrin - Nck1 - NWASP system, which we have previously studied with a non-spatial simulation algorithm based on Flory-Stockmayer theory [4]. This system has also been extensively investigated experimentally by the Rosen lab [35, 36] and has been shown to exhibit liquid phase separation within physiological concentration ranges. However, as opposed to our earlier work [4], we do not explore the requirements for sol-gel transitions that underlie phase separation. Rather, we use SpringSaLaD to perform computational experiments to elucidate how structural features influence the size of molecular ensembles composed of a small number of these multivalent molecules.

We show how multivalency, membrane anchoring, steric interactions, molecular flexibility and the crowded cell environment all influence the size and distribution of molecular ensembles.

Methods

The SpringSaLaD simulation algorithm has been fully described[31], but will be briefly summarized here. Macromolecules are represented as a series of hard spherical domains or sites that are linked by stiff springs. The motions of molecules are governed by a Langevin dynamics formulation that uses a diffusion coefficient assigned by the modeler to each site to calculate a distribution of random forces applied to each sphere. The forces are transmitted vectorially to neighboring sites in the molecule through the stiff spring linkers. Some of the spheres can represent binding sites, where the assigned on-rates with binding partners, together with their respective diffusion coefficients are used to calculate a reaction radius and reaction probability within each a time step. The shorter the time step the higher the accuracy of the simulations. A new bond is represented, simply, as a new stiff spring linking the binding sites. Inputs of off-rates are directly used to determine the

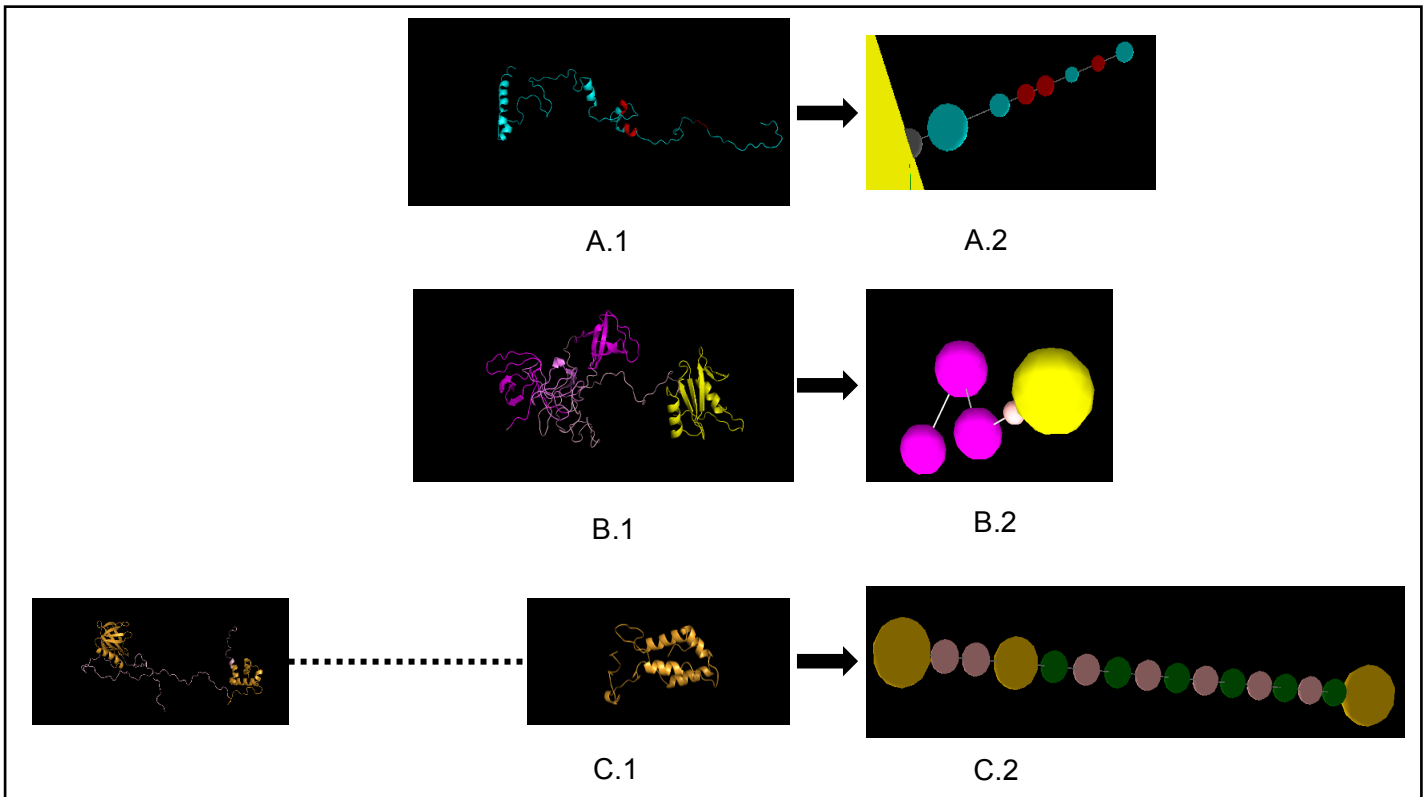


Figure 1: Construction of model components by coarse-graining the protein structures. (A) The 165 amino acid (aa) intracellular portion of the transmembrane protein nephrin is modelled with 8 sites, three of them (red) represent pY-containing binding sites which bind to SH2 domains; this molecule is tethered to the membrane (yellow surface) via an anchoring site (grey). (B) The 376 aa cytoplasmic protein Nck1 is modelled with three SH3 sites (magenta), one SH2 site (yellow) and one linker site (light pink). (C) The 505 aa cytoplasmic protein NWASP has a long intrinsically disordered region containing 6 proline rich motifs from 277-392 in its sequence (shown as a black dotted line in C.1). So this stretch is assumed to adopt a polyproline-II (PP-II) type helical structure with an average distance between two residues of 0.31 nm[1]; it is modelled with six binding sites (green) for SH3 domains and five structural sites (pink). The N-terminal (1-276) and C-terminal sequences (393-505) are modelled with four and one structural sites respectively, according to the predicted secondary structures; the C-terminal structural site corresponds to the VCA domain, which activates actin polymerization.

probability of dissociation during a time step. In this study, we are primarily interested in characterizing the size and composition of clusters at steady state. We do this by initiating simulations with a random spatial distribution of elongated monomers and allowing them to diffuse and react stochastically until the average cluster size fluctuates around a stable size. We statistically analyze 50 stochastic trajectories for each condition. The 50 simulations are run in parallel on the CCAM High Performance Compute Cluster (https://health.uconn.edu/high-performance-computing/resources/); a typical run with 36 molecules for 500ms requires 7 hours.

Molecule construction

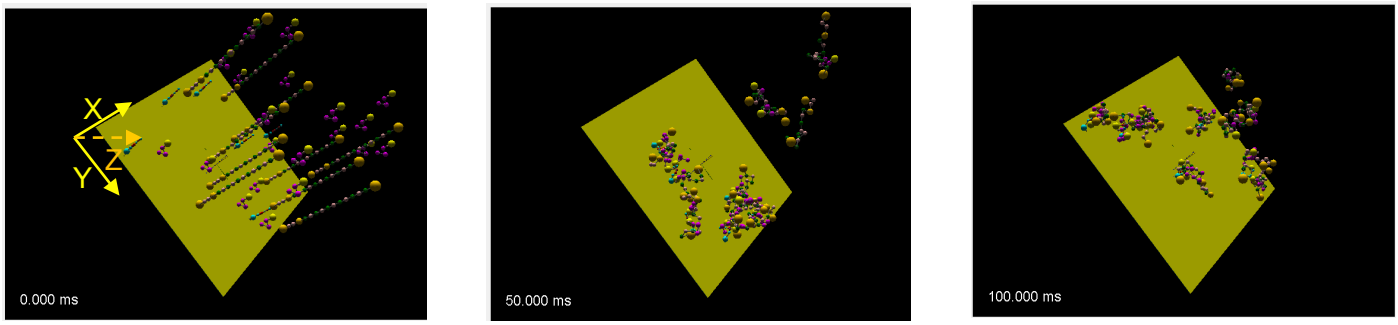
Our model has three molecular components – nephrin, Nck and NWASP. Each protein has multiple domains which take part in biochemical interactions. While some of the specific domain structures are solved experimentally, the full protein structures are not yet available in the literature. So we have used secondary structure prediction homology modeling with web-server platforms like RaptorX (http://raptorx.uchicago.edu) [37] and Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) [38] to generate approximate secondary structures from the amino acid (aa) sequences of interest. SpringSaLaD was then used to generate coarse grain models composed of multiple spheres linked with stiff spring linkers (Fig. 1). The relative sizes of the sites (Fig.S1) are determined by the predicted structures with the aid of a k-means clustering algorithm, *mol2sphere*[39].

Table 1: Parameters used in the model

[nephrin]	9.06 μM
[Nck]	33.22 μM
[NWASP]	16.61 μM
Kd, SH2 (Nck) – pY (nephrin)	1 μM (kon = 100 μM^{-1} s^{-1} ;koff = 100 s^{-1})
Kd, SH3 (Nck) – PRM (NWASP)	100 μM (kon = 100 μM^{-1} s^{-1}; koff = 10000 s^{-1})
D, nephrin_anchor_site	0.05 $\mu m^2/s$
D, nephrin_cytosolic_sites	1 $\mu m^2/s$
D, Nck_sites	2 $\mu m^2/s$
D, NWASP_sites	2 $\mu m^2/s$

Parameters

In SpringSaLaD, each site is assigned a diffusion coefficient (D). Because we are interested in the steady state cluster size distributions, as opposed to the kinetics of cluster formation, we could use somewhat smaller diffusion coefficients (Table 1) than realistic estimates; this permitted us to use longer time steps for our simulations, thus increasing computational throughput. Table 1 shows the D values assigned to the molecular sites. The high affinity binding of a pY site on nephrin to the SH2 domain on Nck is assigned a K_d of 1 μM[40]; the low affinity binding between a Nck SH3 domain and a PRM on NWASP is assigned a K_d of 100 μM[41]. Again, because the actual binding or dissociation rates do not affect the equilibrium cluster sizes, we chose values in Table 1 to optimize computational throughput. We checked some simulation results with both shorter timesteps and larger D values to assure that the steady state cluster characteristics were accurate when using our nominal values. All these input specifications can be found in the SpringSaLaD input files included in Supporting Information.



A B C

Figure 2: Clustering in a single trajectory of the "reference" system. (A) The system contains a total of 36 molecules (6 nephrin, 20 Nck, 10 NWASP) in a cubic reaction volume (X = Y = Z = 100 nm); XY surface represents the membrane where nephrin is anchored. At the beginning of a simulation (t = 0), 36 molecules are distributed randomly (nephrin onto the membrane, Nck and NWASP in the cytosol) in the reaction volume. (B, C) A pair of time points during the trajectory reveal the development of clusters. At steady state (t ~ 100 ms), most clusters reside in the vicinity of membrane. Molecular motions and cluster formation can be seen at the supplementary video file (Movie S1).

Results

Analysis of cluster size and composition

The baseline reference system (with which other configurations will be compared) consists of 36 molecules in a cubic domain 100nm on each side as described in Figs. 1 and 2. The proportion of nephrin:Nck:NWASP was 6:20:10, respectively and chosen to approximate the optimal stoichiometry for binding site interactions; that is, in the reference system there are 18 pTyr sites interacting with 20 SH2 sites and 60 SH3 sites interacting with 60 PRM sites. Typically, we analyzed outputs from 50 trajectories. We can display the steady state outputs of these simulations as histograms, as illustrated in Fig. 3. In Fig. 3A, the abscissa corresponds to the average cluster size at steady state from a single simulation; this is calculated as the number of molecules (36 for the reference system) divided by the number of clusters. Thus, if a simulation produces 2 clusters each containing 18 molecules and another simulation produced 2 clusters containing 1 and 35 molecules, they would each yield the same average cluster size of 18. Since the number of clusters and the number of molecules are both small integers, only certain average cluster sizes are mathematically possible as can be seen from the sparse distribution in the histogram (Fig.3A).To better examine how molecules are distributed over all possible cluster sizes, we can use an analysis displayed in Fig. 3B. Here, the abscissa is every possible cluster size from 1 to 36 molecules; the ordinate is the fraction of molecules populating these cluster sizes. For the scenario described above, a pair of 18mers would produce a single histogram entry at 18 on the abscissa with fraction of 1 on the ordinate; a configuration containing 1 monomer with 1 35mer would have an abscissa entry at 1 with a height of 0.028 and at 35 with a height of 0.972. For both of these histograms, we calculate weighted means, indicated by the dashed lines in Figs. 3A and B, that we term, respectively, Average Cluster Size (ACS) and Average Cluster Occupancy (ACO). For our 2 extreme scenarios, the ACS is 18 for both; the ACO for 2 18mers is 18 and for the combination of monomer and 35mer, ACO is 34.06. We can compute the ACS and ACO for each of the time points to generate averaged trajectories (Figs. 3C. and D.). The kinetics for both methods of assessing clustering are similar, reaching steady state by 100ms. To produce the histograms in Figs. 3A and B, we use this 100ms relaxation of the system as an interval over which to sample five time points

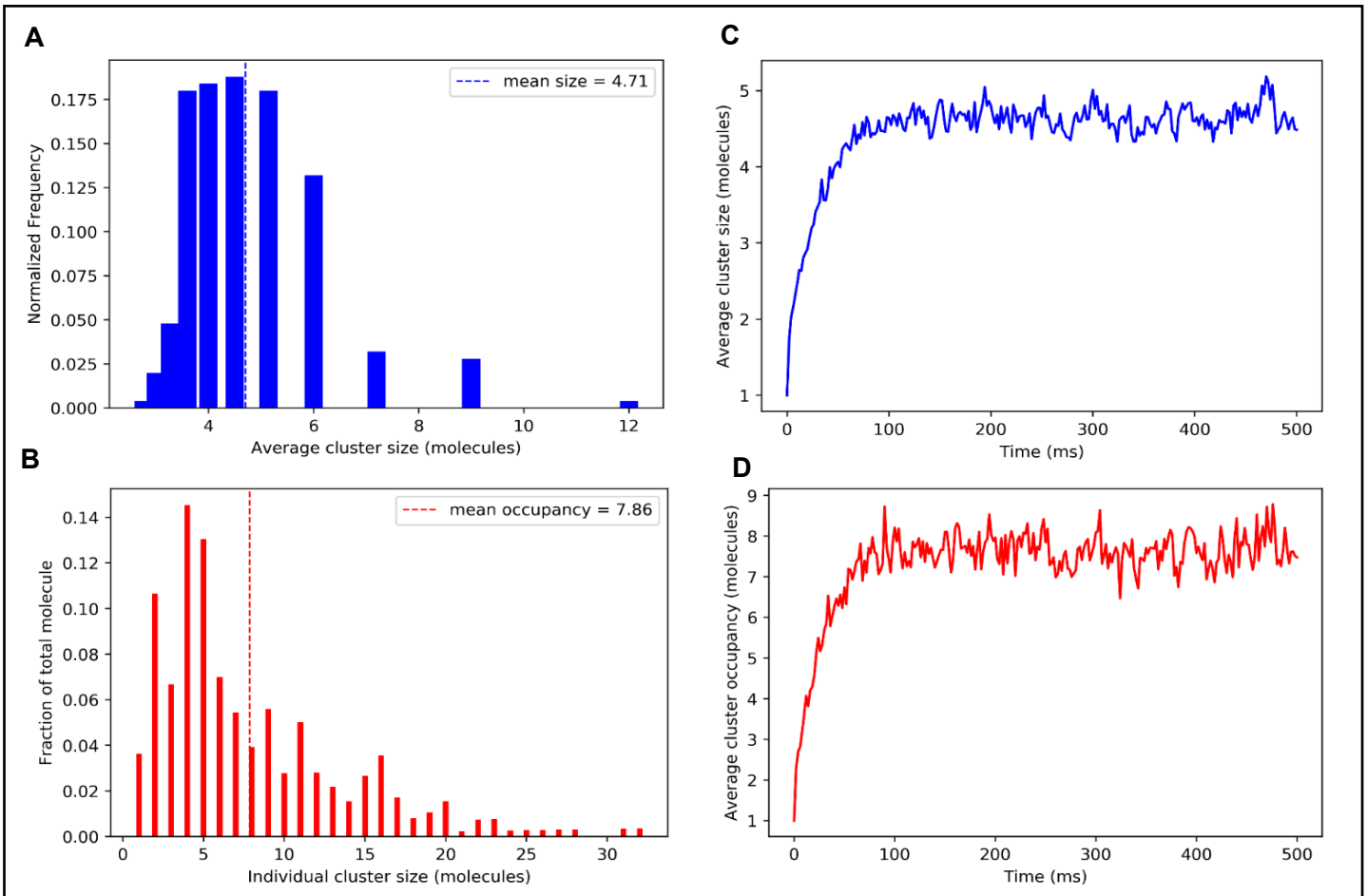


Figure 3: Quantification of molecular clustering in the reference system. A. Distribution of average cluster sizes in 250 steady state samples. B. Distribution of fraction of molecules occupying individual cluster sizes in 250 steady state samples. Steady states in A and B are from 50 trajectories, sampled at 5 equally spaced intervals from 100 - 500 ms. C., D. Corresponding time courses for the average of 50 trajectories.

(100 - 500 ms). As each time point corresponds to 50 runs, 250 independent data points are used to plot the steady state distributions in Fig.3A, 3B.

A surprising property is the insensitivity of the ACS and ACO to the system's size. When we increase or decrease the number of molecules, keeping the molecular concentrations the same by adjusting the size of the domain, cluster size distributions and means turn out to be similar (Fig. 4). This indicates that the small number of available monomers does not limit the size of the clusters. Rather, this similarity in cluster sizes suggests that the system is sufficiently large to be approximated by equilibrium thermodynamics. In particular, there appears to be a balance of enthalpy and entropy that governs the size distributions of clusters and does not allow monomers to condense into a single large cluster. With a larger number of molecules, the average cluster size displays a much tighter distribution (Fig. 4A, bottom) because the higher number of possible average cluster sizes reduces the stochasticity of each average for individual steady state points. However, remarkably, there is very little effect on the shape of the distribution of fractional occupancy (Fig. 4B). This suggests that the shape of the distributions in Fig. 4B correctly reflects the tendency of the system to favor certain cluster sizes, as will be further discussed below. Overall, this analysis also indicates that our reference

system with 36 molecules, is sufficiently large to provide a good representation of the steady state system behavior.

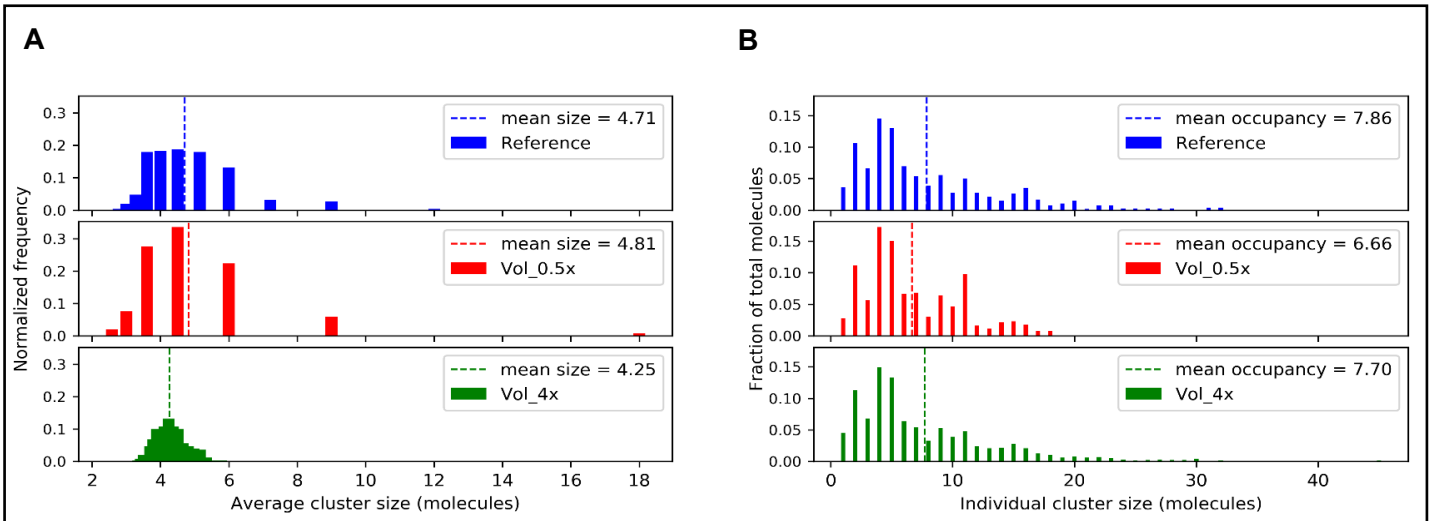


Figure 4: Clustering behavior is independent of system size. Steady state distribution of average cluster size (A) and molecule occupancy of individual cluster sizes (B), for 36 (blue), 18 (red) and 144 (green) molecules in each simulation. These results are from 250 steady state samples as in Fig. 3.

Effect of multivalency: Molecular clustering is a direct consequence of multivalent interactions between binding partners, but how this works in a specific system can be hard to predict. Here, we directly address this problem computationally. We manipulate the nephrin molecule to create three situations where the total number of binding sites are same, but the valency states are different (Fig.5A): 3 pY sites on each of 6 nephrin (3v, the "reference" system), 2 pY sites on each of 9 nephrin (2v), and 1 pY sites on each of 18 nephrin (1v). Nck and NWASP configurations are the same for all 3 scenarios. The histogram of fractional occupancy shifts to the left and, correspondingly, ACO decreases with reduced valence state (Fig. 5B). In considering this, it is important to appreciate that the strong interaction between the pY sites on nephrin and the single SH2 site on Nck assures that almost all the pY sites will be occupied with Nck in all 3 scenarios (as confirmed by the simulations results). Therefore, the effect of nephrin valency is actually to gather multiple Nck into a reduced local volume for interaction with the NWASP molecules. It should also be noted that this multivalency effect is over and above the effect of localizing the nephrin to the membrane, which actually becomes densely covered with nephrins for the 1v case, with 18 nephrins/ 100nm^2 (the effect of membrane localization is fully considered below).

Returning to Fig. 4B, we can appreciate that even for different numbers of molecules (maintaining their initial concentrations), the shape of the distribution of fractional occupancies remains remarkably the same. While the histograms of fractional occupancy are quite broad, some cluster sizes are preferred over others for all 3 system sizes. This reveals a tendency to optimize the binding site occupancy in individual molecules by favoring certain stoichiometries. Specifically, it appears that clusters with 2, 4, 5, 9 or 11 molecules are preferred over their neighbors. For example, the 11mer almost always has a composition of nephrin$_2$Nck$_6$NWASP$_3$, despite the fact that other compositions are mathematically possible (Supplementary Table S1). This observed composition corresponds to what would be predicted for a perfect stoichiometric binding of all the binding sites. However, close examination of individual 11mers at a given steady state time point reveals that some binding sites are often unoccupied for the weaker binding between the SH3 sites on

Nck and the PRM sites on NWASP. What enforces the 11mer stability is the enhanced rebinding rates of these sites emerging from their proximity within the cluster.

Fig. 5B shows how the broad distribution of the reference trivalent nephrin system gives way to much sharper distributions when the nephrin valency is decreased. Strong preference for tetramers and pentamers are seen for the bivalent and monovalent, respectively, nephrin molecules. Again, optimal binding site stoichiometry underlies these preferences. The perfect binding site stoichiometry (nephrin: Nck: NWASP) for the bivalent system is 1:2:1 and for the monovalent nephrin, 2:2:1, accounting for the preference for tetramers and pentamers seen in Fig. 5B and Table S1. Interestingly, computational experiments that increase the affinities of the interactions for the trivalent nephrin (top panels of Fig. S2 A. and B.), also produce much more discrete histograms of fractional occupancy; stoichiometries that maximize multivalent binding are now highly favored (see also Table S1). Overall, this analysis indicates that at a fixed affinity, increasing multivalency leads to larger clusters, but with broader size distributions; higher affinity binding at a fixed valency, leads to larger

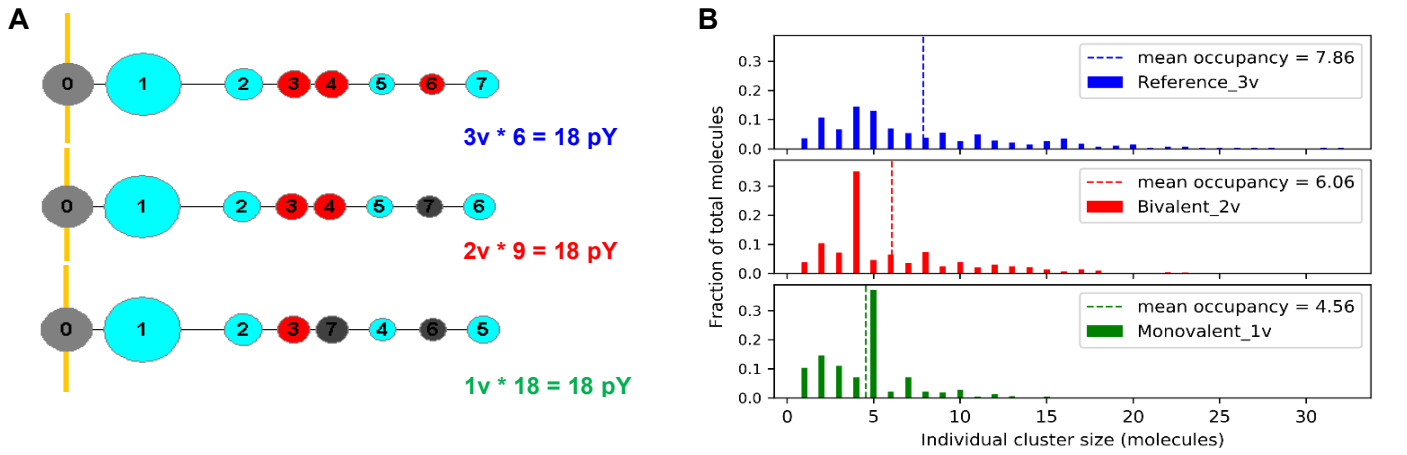


Figure 5: Multivalency produces larger clusters. (A) Three Nephrin constructs with differing valency; red sites represent pY (SH2 binding sites). The number of these tri, bi, and monovalent Nephrin molecules are adjusted (6, 9, 18 respectively) such that total number of pY-sites remains the same in all three cases. (B) Steady state distributions of individual cluster sizes (the dotted line represents the mean of the

clusters with more discrete cluster size distributions.

As noted above, the count ratio of each molecule in the reference system, 6:20:10, was chosen to optimize the possible number of interactions between, respectively, nephrin, Nck and NWASP. We asked how clustering might be affected by altering the ratios of these molecules, keeping the total number of molecules at 36 (Table. S2). As expected, the reference system produces the largest clusters. Interestingly, however, a system where the ratio is, respectively, 10:20:6, is practically as good, despite the poor match between available binding site partners. What seems to be more important than the best match is the availability of a high level of Nck, which serves as an “adaptor” to link multivalent nephrin to multivalent NWASP.

Effect of molecular structural features: Our model gives us a unique opportunity to probe the interplay of steric effects, proximity effects and protein flexibility on cluster formation. We first asked if the clustering propensity depends on the intramolecular distances between binding sites. We created two systems where the linker lengths within a molecule are elongated by 1 nm and 2 nm respectively compared with the reference

system (Fig. 6A). As the distance between sites is increased, there is a dramatic shift in the distribution of molecules into larger clusters and a corresponding increase in mean cluster occupancy (Fig. 6B). The probable reason for this behavior is that with longer linkers, binding sites in the interior of a large cluster remain more accessible to additional binding partners; i.e. steric hindrance within the interior of a cluster is reduced with longer distances between binding sites.

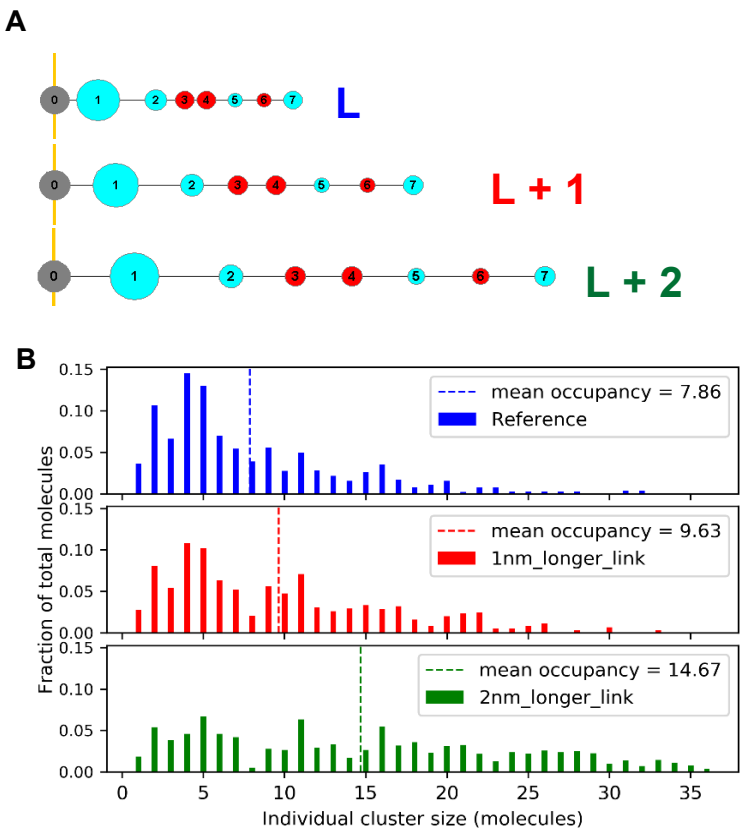


Figure 6: Effect of linker length. (A) The reference system compared to molecules where the linkers are increased by 1nm and 2nm. Only Nephrin is shown but the linker lengths were incremented in the same way in Nck and NWASp. (B) Steady state distributions of molecular occupancies (dotted line indicates the mean of the distributions).

While the analysis in Figure 6 is informative, in that it isolates the effect of distance between sites, the linkers in SpringSaLaD have 2 attributes that limit how they can represent the regions between binding sites: they are inflexible and they are sterically transparent. Therefore, to better represent the actual molecular structures, we included "structural" sites in all our molecules, as identified in Fig. 1.

These do not have any binding attributes, but they do exclude volume and serve as pivot points for conformational flexibility. As might be expected, completely removing the structural sites dramatically increases the steady state cluster size (Fig. S3). There are two possible underlying factors that could cause such dramatic shift in cluster size: steric effect and conformational flexibility. While steric effect should inhibit the clustering by hindering the approach of two binding partners, the effect of flexibility is not that intuitive.

To isolate the steric effect, we create three NWASP constructs (Fig. 7) with similar flexibilities, but different steric effects by varying the sizes (Fig. S4) of the structural sites (nephrin and Nck have the reference structures in all these simulations). Compared to the reference system (Fig. 7A), increasing the size of all structural sites (Fig. 7B) moderately decreases the clustering, while decreasing their sizes (Fig. 7C) moderately increases the cluster sizes. As would be expected, these changes reveal the influence of steric hindrance for binding interactions. Interestingly, removal of the peripheral structural sites from the reference NWASP molecule (Fig. 7D) dramatically increases the tendency to form larger clusters. This large effect of the peripheral binding sites can be attributed to their exclusion from the cross-linked interior of larger clusters. This exclusion has the synergistic effects of unfavorably lowering the entropy of larger clusters and also serving as a steric barrier for the binding of monomers to available free binding sites in the interior.

The 3 proteins that inspire this work all have intrinsically disordered region (IDR) in their sequences; the proline rich sequence in NWASP [42], the linker region between SH2 domain and first SH3 domain in Nck [36], and almost

all the intracellular portions of the nephrin sequence [43]. As a consequence of the flexibility of IDRs, the intramolecular distance between sites that flank them are more likely to vary in the course of interaction. To test the effect of flexibility, we created the NWASP molecules shown in Fig. 8 and simulated their clustering behavior with the reference structures of nephrin and Nck (Fig. 8A is the reference structure of NWASP).

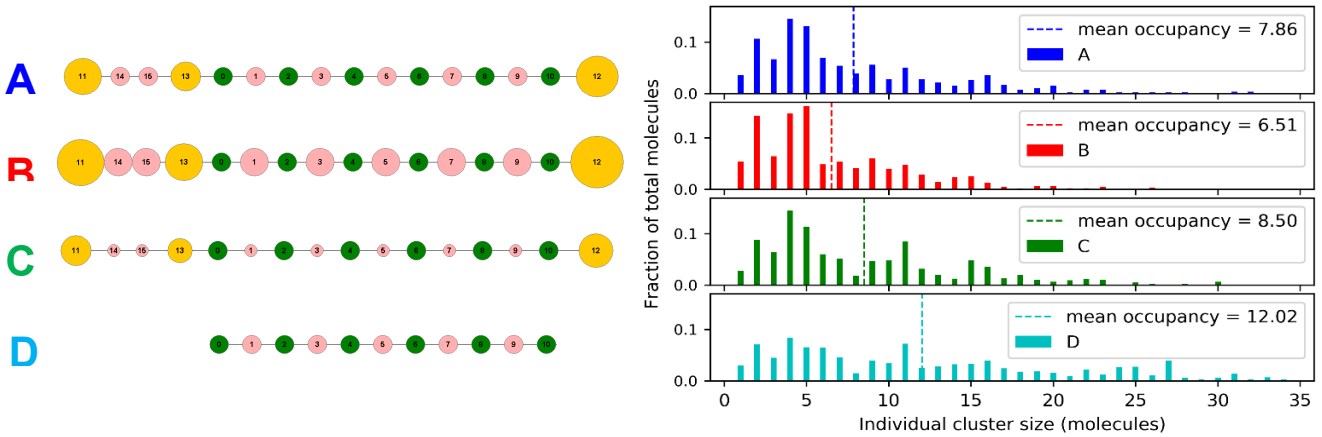


Figure 7: Steric effects. Sizes of NWASP structural sites are varied, keeping the sizes of PRM binding sites (green) unchanged. (A) Reference; (B) Larger structural sites; (C) Smaller structural sites; (D) NWASP without the peripheral structural sites. Corresponding steady state distributions of individual cluster sizes (dotted line indicates mean of the distribution) are shown on the right. Nephrin and Nck retain the reference structures in all 4 cases.

Interestingly, we see that the less flexible structure of Figs. 8B have the tendency to form very large clusters; structures with similar flexibilities but differently sized pivot sites (Fig. 8A vs. Fig. 8C) yield similar clustering behavior. This flexibility dependence could be explained by considering entropic effects. Since average cluster size is an equilibrium property, the loss of entropy would prevent the system from forming larger clusters, as discussed above. But the loss of entropy is not just due to the formation of multimers from monomers, but also because of the loss of conformational freedom resulting from the cross-linked multivalent binding. The more flexible the monomers, the greater the entropic loss upon cluster formation. So the steady state cluster size would go up with the less flexible initial structures, consistent with the trend shown in Fig. 8. Another way of explaining this would be the tendency of the less flexible molecules to populate relatively more extended conformations, where their binding sites remain more highly accessible. To examine this idea, we analyzed the distance between the outer binding sites in isolated NWASP molecules shown in Figs. 8 A, B and C. When fully stretched, the distance between these sites is 35nm. For each case, we averaged 25,000 independent conformations from 50 trajectories, and they show precisely this trend (Figs. 8 and S5). It is interesting that the smaller pivot sites in Fig. 8C actually produce a more compact average conformation than the reference structure of Fig. 8A, consistent with the smaller average cluster size in the former. Steric inhibition of binding would have predicted the opposite, suggesting that flexibility is a more important factor in limiting cluster size than steric hindrance. However, turning to Fig. 8D, we see that decreasing both steric hindrance and flexibility produces the most dramatic enhancement of clustering, over and above the effects of each alone (compare Figs. 7D and 8B).

Effect of membrane anchoring: Signaling cluster formation takes place in the spatial region near the plasma membrane; so membrane associated proteins are likely to play a key role in this process. In principle, membrane anchoring can promote clustering by producing by confining binding sites to a region of approximately one molecular length from the membrane; this effectively increases the local concentration

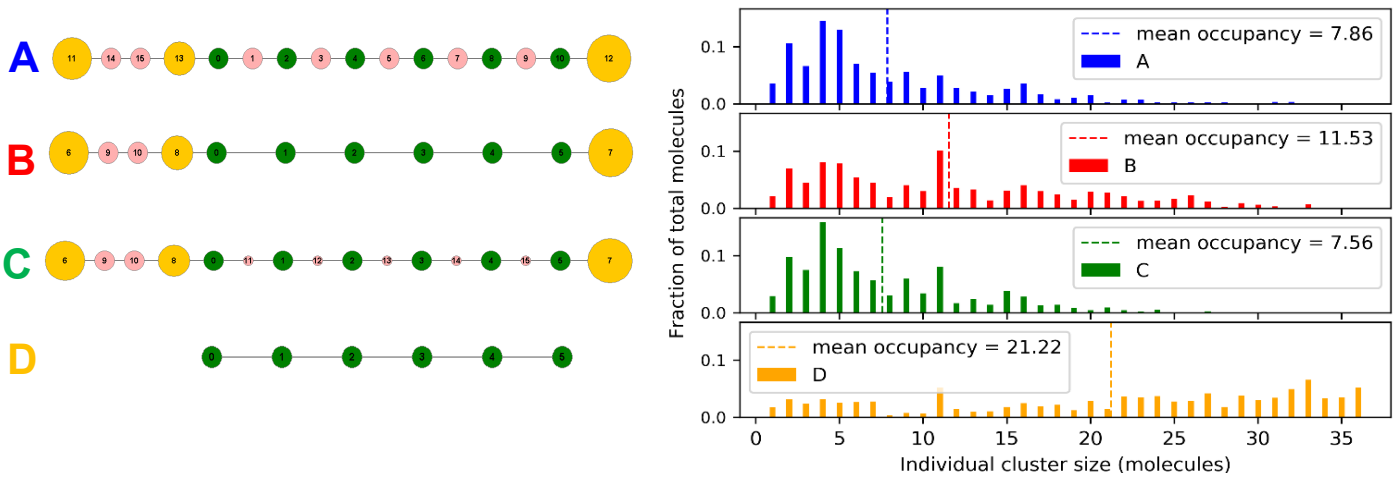


Figure 8: Decreased flexibility around binding sites promotes clustering. (A) Reference configuration, (B) pivot sites (pink) flanking the PRM sites (green) are removed; (C) Similar to reference system, except the sizes of pivot sites in between PRM are smaller; (D) NWASP with only PRM binding sites. Corresponding steady state distributions of individual cluster sizes (dotted line indicates the mean of the distribution) are shown on the right. Nephrin and Nck retain their reference structures in all cases.

compared to sites that are free within the bulk volume. On the other hand, the membrane might serve as a steric barrier preventing the binding partners to approach the membrane-tethered molecular binding sites. To understand the interplay of these 2 opposing effects, we created three nephrin constructs where length of the anchor linker is increased gradually (keeping the rest of the configurations same) to reduce the potential steric effect from the membrane; we also include simulations where nephrin is detached from the membrane and free to diffuse around the cytosol, which should demonstrate the role of spatial confinement or effective higher

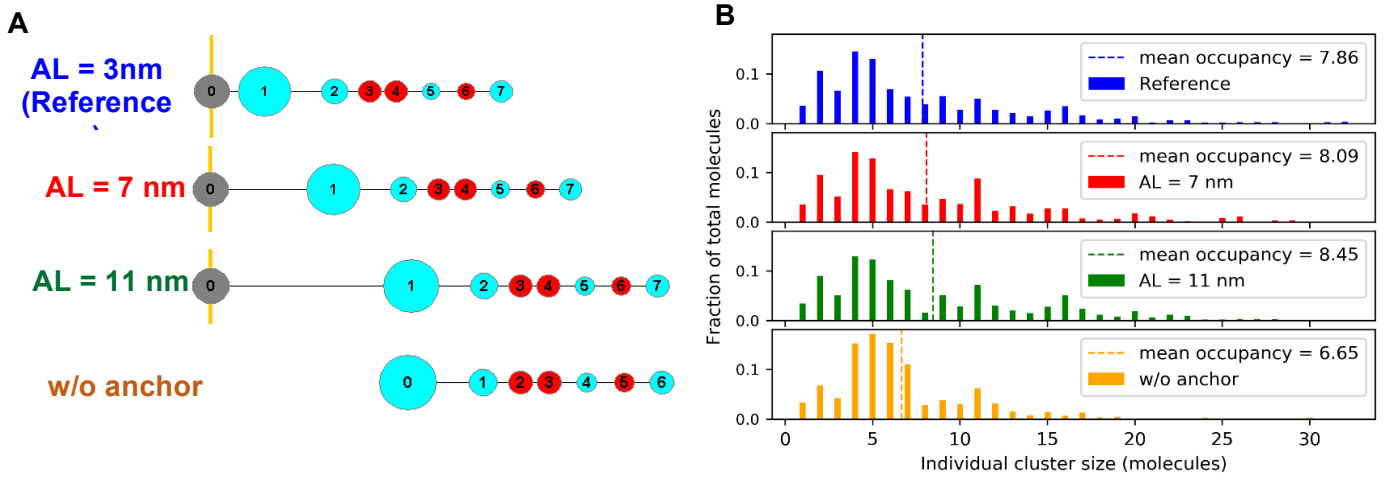


Figure 9: Membrane association increases clustering but can also exert steric inhibition. (A) Nephrin constructs with varying anchor length, AL (which is the linker that connects first structural site with the anchor site) and without anchor. (B) Steady state distributions of individual cluster sizes (dotted line indicates the mean of the distribution).

local concentration around the membrane (Fig. 9). There is a small increase in fractional occupancy as a function of anchor length, indicating the steric contribution from the membrane is minor. However, removing the membrane anchor results in a more significant reduction of mean cluster occupancy, indicating that membrane confinement can potentiate clustering. This effect is much more prominent when the same experiment is done with the binding sites only (BSO) system (Fig. S6).

To further probe the importance of spatial confinement at the membrane, we explore the effect of density of nephrin molecules at the membrane surface. In our reference configuration, we have a total of 36 molecules (6 nephrin, 20 Nck, 10 NWASP) in a cubic (X = Y = Z = 100 nm) reaction volume with nephrin anchored on the XY plane. We can change the membrane surface density of nephrin, while maintaining constant volume and monomer concentrations, by altering the X, Y, Z dimensions (Fig. 10A, 10B, and 10C). For a constant volume (concentration), when we increase the membrane density of nephrin, we see a correlated increase in average cluster size (Fig. 10D); the effect in cluster occupancy is much more prominent indicating the larger sizes of the clusters with higher membrane density (Fig. 10E). However, without the membrane, this change in aspect ratio does not have any effect of clustering dynamics (Fig. S7). So receptor density at

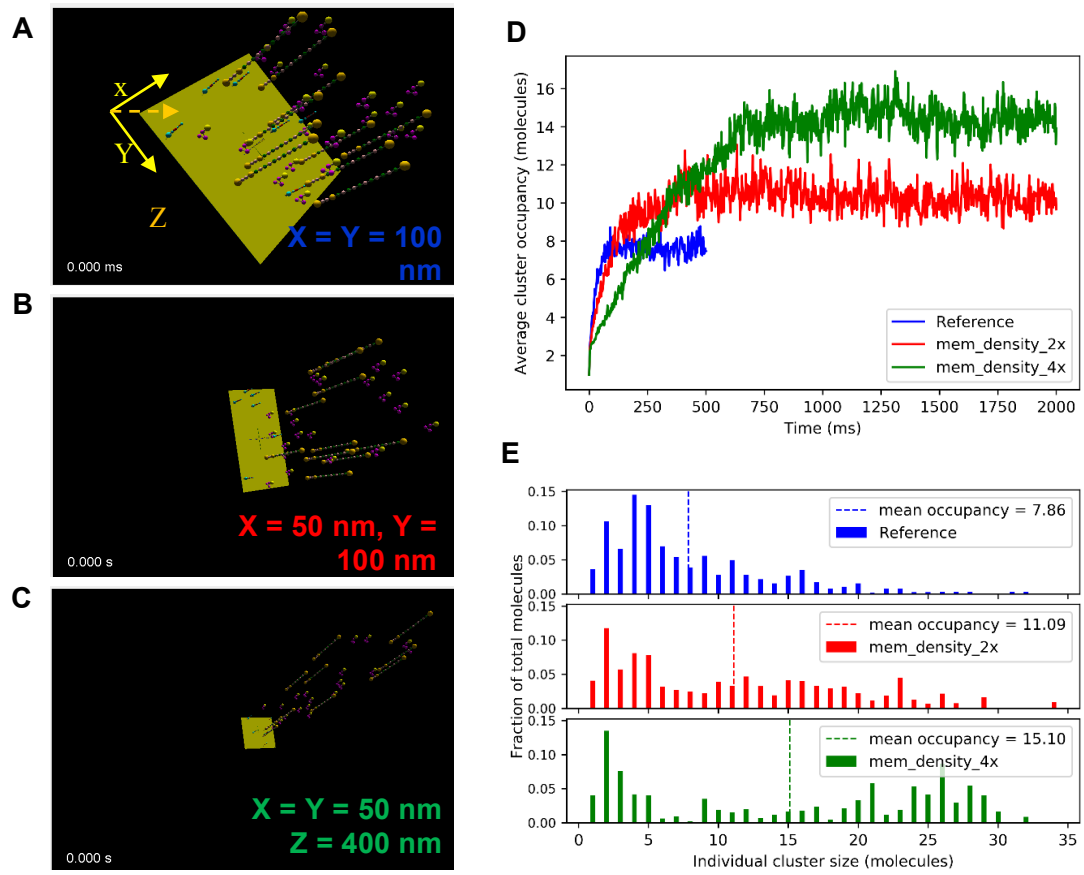


Figure 10: Increased nephrin membrane density promotes clustering. (A) Reference configuration in a cubic reaction volume (X = Y = Z = 100 nm); nephrin density = d (= 0.06 molecules.μm^{-2}). (B) X = 50 nm, Y = 100 nm, Z = 200 nm; nephrin density = 2*d. (C) X = Y = 50 nm, Z = 400 nm; nephrin density = 4*d. (D) Dynamics of average molecular occupancy; note that times to reach steady states are different for three. (E) Steady state distributions of molecular occupancies. For higher nephrin density cases (red and green), two steady state time points (100 realizations) are sampled, unlike the reference system (250 realizations). Total number of molecules are the same in all simulations.

the membrane potentiates clustering. Interestingly, however, the kinetics for approach to steady state is much slower for the higher density systems. Thus the local confinement afforded by membrane anchoring significantly increases the propensity for clustering while decreasing the rate of cluster growth.

Effect of molecular crowding: Cells contain a wide range of biomolecules with varying shapes and sizes; these molecules occupy physical volume and behave as obstacles to other diffusion driven processes. We next use our system to explore how molecular crowding might influence clustering. To address that, we create a crowded environment by adding inert molecules to our system. We employ 320 spherical crowders of radius = 5 nm (Fig.11B). We see a large increase in average cluster size and occupancy upon adding the crowder. The effect of molecular crowding is attributed to the excluded volume effect (repulsive steric process) which enhances the effective concentration (or thermodynamic activity) of the reactive molecules by reducing the available volume. Although a simple calculation shows that our spheres fill ~1/6 of the $10^6 nm^3$ reaction volume, the actual excluded volume might be better approximated by subtracting the packed volume of the spheres from the total volume; if we assume a cubic packing, this volume is 320 X $(2r)^3$, or ~1/3 of the $10^6 nm^3$ simulation volume. To test this effect of volume reduction, we modeled an additional case where 36 interacting molecules are put into a 2/3 of reference reaction volume with no crowders (Fig. 11C). Interestingly the increase in average cluster size or occupancy is less

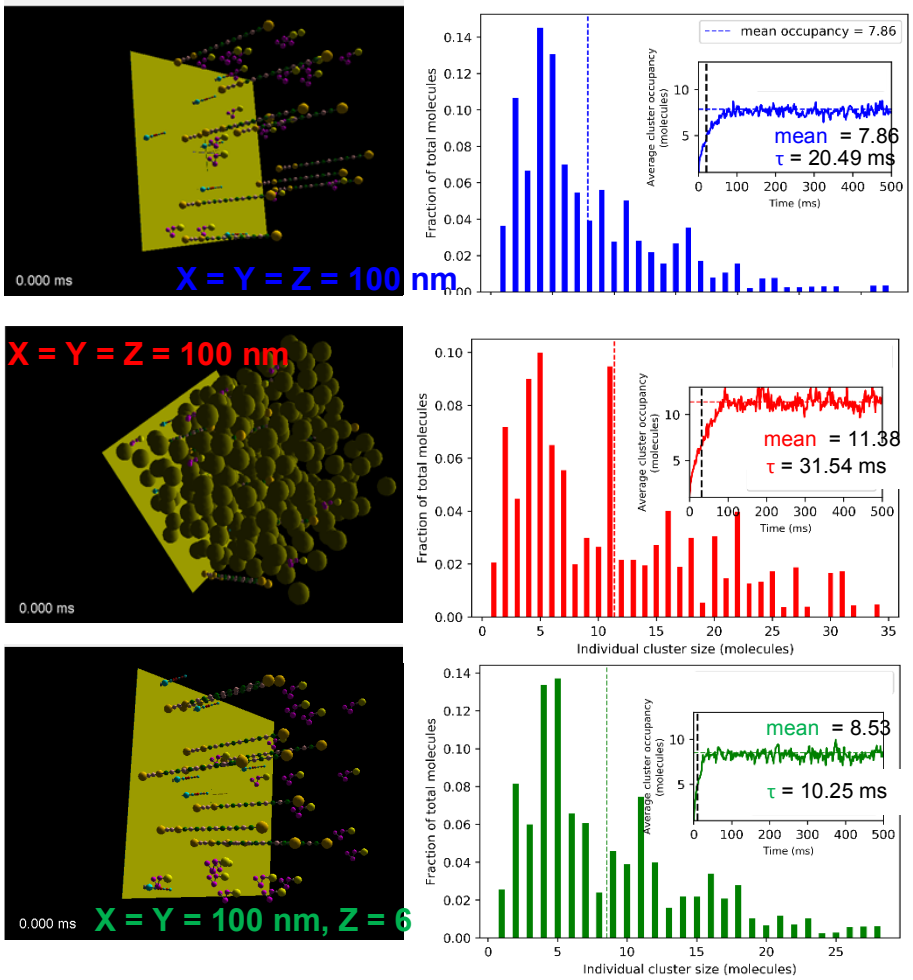


Figure 11: Effect of molecular crowding. (A) Reference system with 36 interacting molecules in a 100^3 nm^3 cubic reaction volume. (B) System with 320 inert crowder (radius = 5 nm) along with 36 interacting molecules in the reference reaction volume. The crowders would take up ~1/3 of the volume based on cubic packing (each sphere with radius r would occupy an effective cubic volume having a side length of 2r). (C) 36 interacting molecules in 2/3 of the reference volume, i.e., (100*100*67) nm^3. The Corresponding steady state distributions of molecular occupancies are shown in the right hand panels (inset shows the time course). Since the crowded system takes longer to reach steady state, four time points (*50 runs = 200 realizations) are sampled for the distribution.

pronounced as compared to the crowded system. So the extent of volume exclusion is higher than the sum of closely packed volume of the crowder. This is likely a consequence of the physical size of the interacting molecules, which increases the excluded volume still further. That is, the finite size of the sites within the 36 interacting molecules because the mass-center of each interacting molecule can't access a spherical volume corresponding to the sum of its own radius and that of the crowder (Fig. S8), and even with the wall of the chamber. This additional volume exclusion is reflected in the larger cluster formation in crowded environment (Fig. 11B vs. 11C).

Discussion

The aims of our study were to establish the biophysical factors that shape how weakly binding multivalent interactions control the formation of dynamic clusters, which have been also called molecular ensembles. We have developed a kinetic model that utilizes a coarse representation (Figs. 1 and S1) of the key molecular features in 3 interacting multivalent molecules. It is inspired by the nephrin/Nck/NWASP system that underlies the structural integrity of the kidney filtration unit [44, 45]. The multivalent binding domains from each of these molecules also served as the basis for a seminal study of microdroplet formation by liquid-liquid phase separation [35]. It is important to emphasize that our model is not intended to predict the ability of a system to phase separate. This would require a more macroscopic theory that also accounts for electrolyte interactions and the state of water in the 2 phases [22, 46]. However, cluster formation through weak multivalent interactions is a definite prerequisite for liquid droplet phases. More broadly, molecular ensembles can serve as functional platforms for key cellular signaling systems.

Our reference system consisted of 6 membrane-bound nephrin molecules, 20 cytosolic Nck molecules and 10 cytosolic NWASP molecules in a cubic reaction volume of 100nm on a side; one face of the cube represented a planar patch of membrane within which the nephrin molecules could diffuse (Fig. 2). This stoichiometry was close to the optimal for maximizing interaction between all the binding sites on these multivalent molecules. Based on stochastic simulations, we analyzed the kinetic approach to steady state, the distribution of cluster sizes at steady state and the distribution of molecular occupancies within each possible cluster size (Fig. 3). We found the reference system of 36 molecules was sufficiently large to simulate all the key features of the system (Fig. 4) because, perhaps somewhat surprisingly, the cluster size and cluster occupancy histograms were very similar for 18, 36 or 144 molecules. This gave us confidence that the reference system properties could serve as a good baseline for computational experiments that systematically probed for the effects of molecular and cellular structure on the clustering behavior. It also demonstrated that a balance of enthalpic and entropic factors limited the size distribution of clusters and prevented their annealing into a single large complex.

As would be expected, decreasing valency decreases the steady state cluster size distribution (Fig. 5). This computational experiment mirrors the trend in an in vitro experiment by the Rosen lab [35], in which nephrin constructs with 3, 2 and 1 phosphotyrosines required progressively higher concentrations of Nck and NWASP to produce phase separation. But, more subtly, the simulations show that there are preferred cluster sizes with specific monomer compositions that are dependent on valency, with lower valency sharpening the preference for clusters of optimal stoichiometric composition. On the other hand, the distributions are also sharpened to favor specific cluster sizes and compositions as the binding site affinities are increased, which increases the average cluster sizes (Fig. S2).

We deconvolved the influence of steric interactions and molecular flexibility on steady state cluster sizes by systematically altering NWASP structural features (Figs. 6, 7, 8). In SpringSaLaD, each spherical site excludes volume to represent steric effects and also serves as a pivot to impart molecular flexibility; the stiff spring links serve to maintain a fixed distance between the sites and also transmit forces. Decreasing flexibility by

removing sites that were not involved in binding produced a major increase in cluster sizes. We showed that this effect is not due to a decrease in local steric interference to binding, which might also be an effect of removing these structural sites (compare Fig. 8A and C). We believe this effect can be attributed to a loss in entropy during cluster formation, where more flexible monomers would lose more entropy than less flexible monomers. Another way of thinking about this is to consider that the less flexible molecules would tend to adopt more open conformations with more exposed binding sites. Indeed, an analysis of the distribution of distances between the terminal binding sites (Fig. S5) shows that removing the pivot sites results in a significantly stretched average conformation. This synergistic interplay of entropy and access to binding sites also emerges in our consideration of steric interactions.

We examined steric interactions by changing the sizes or eliminating the structural sites in NWASP, keeping the sizes of binding sites constant (Figs. 7 and S4). The most surprising conclusion was the dominance of the peripheral structural sites in NWASP; removing or shrinking these sites significantly increased cluster sizes (Fig. 7). Shrinking the structural sites that were located between binding sites without changing the peripheral sites (Fig. 8C) actually slightly shifted the cluster occupancy histogram to smaller sizes. Apparently, the crosslinking of binding sites favors exclusion of the structural sites from the interior of clusters, with larger structural sites having a greater propensity for exclusion. This self-organizing effect would lower the entropy and explain the greater steric effect for peripheral domains. As an alternate view, these large structural sites at the periphery of clusters would serve as a steric barrier to the recruitment of additional monomers to binding sites in the interior, thereby limiting cluster expansion. A biological implication is that these domains would be present at the exterior where they could participate in downstream signaling. In particular, the VCA domains of NWASP (within the right-most yellow structural site in Fig. 1C.2) would be assembled at the periphery of clusters; this is particularly opportune, because a pair of proximate VCA domains is required to recruit and activate ARP2/3, which in turn, nucleates branched actin polymerization [47, 48]. In general, such non-linearity is a hallmark of signaling systems.

Turning to cellular effects, our model shows increased steady state clustering effected by inert crowders (Fig. 11). We attribute this to an effective increase in binding site concentration due to the excluded volume of the spherical crowders. On the other hand, crowders significantly reduce the rate of cluster formation. A similar, but subtler consideration of effective concentration can explain how membrane anchoring of nephrin significantly promotes clustering (Fig. 9). This is because confining reactions to the membrane effectively reduces the available reaction volume once the initial complement of Nck and NWASP molecules are recruited from the bulk volume. Thus, annealing of small membrane-associated clusters into larger ones is favored because the binding sites are at an effectively higher concentration. This effect is further enhanced when the surface area of the membrane is decreased while keeping the bulk concentrations constant (Fig. 10): increasing the membrane density 4-fold doubles the average cluster size occupancy. Lipid rafts provide a biological platform to concentrate membrane proteins and are thought to play an especially important role in receptor-mediated signaling[49]. Our results suggest that clustering can serve as a positive feedback mechanism to amplify the ability of lipid rafts to localize membrane receptors and thereby amplify spatially encoded signals.

Our computational experiments have allowed us to gain insights into the biophysical features that control the formation of molecular clusters. But most importantly, they serve to suggest experiments that will both help to validate these ideas and manipulate downstream cellular responses.

Acknowledgments

The authors gratefully acknowledge useful discussions with Michael Blinov, James Schaff, Boris Slepchenko and Paul Michalski. The project is supported by the National Institute of General Medical Science through grants P41 GM103313.

References

[1] Adzhubei, A. A., Sternberg, M. J. E., and Makarov, A. A. (2013) Polyproline-II Helix in Proteins: Structure and Function, *Journal of Molecular Biology 425*, 2100-2132.

[2] Bray, D., Levin, M. D., and Morton-Firth, C. J. (1998) Receptor clustering as a cellular mechanism to control sensitivity, *Nature 393*, 85-88.

[3] Jilkine, A., Angenent, S. B., Wu, L. F., and Altschuler, S. J. (2011) A Density-Dependent Switch Drives Stochastic Clustering and Polarization of Signaling Molecules, *PLoS Comput Biol 7*, e1002271.

[4] Falkenberg, Cibele V., Blinov, Michael L., and Loew, Leslie M. (2013) Pleomorphic Ensembles: Formation of Large Clusters Composed of Weakly Interacting Multivalent Molecules, *Biophysical journal 105*, 2451-2460.

[5] Roob, E., 3rd, Trendel, N., Rein Ten Wolde, P., and Mugler, A. (2016) Cooperative Clustering Digitizes Biochemical Signaling and Enhances its Fidelity, *Biophysical journal 110*, 1661-1669.

[6] Suderman, R., and Deeds, E. J. (2013) Machines vs. Ensembles: Effective MAPK Signaling through Heterogeneous Sets of Protein Complexes, *PLoS Comput Biol 9*, e1003278.

[7] Mayer, B. J., Blinov, M. L., and Loew, L. M. (2009) Molecular machines or pleiomorphic ensembles: signaling complexes revisited, *J Biol 8*, 81.81-81.88.

[8] Hlavacek, W. S., Faeder, J. R., Blinov, M. L., Perelson, A. S., and Goldstein, B. (2003) The complexity of complexes in signal transduction, *Biotech Bioeng 84*, 783-794.

[9] Faeder, J. R., Blinov, M. L., Goldstein, B., and Hlavacek, W. S. (2005) Combinatorial complexity and dynamical restriction of network flows in signal transduction, *Syst Biol (Stevenage) 2*, 5-15.

[10] Pawson, T., and Nash, P. (2003) Assembly of Cell Regulatory Systems Through Protein Interaction Domains, *Science 300*, 445-452.

[11] Schlessinger, J. (2000) Cell Signaling by Receptor Tyrosine Kinases, *Cell 103*, 211-225.

[12] Sato, P. M., Yoganathan, K., Jung, J. H., and Peisajovich, S. G. (2014) The Robustness of a Signaling Complex to Domain Rearrangements Facilitates Network Evolution, *PLoS biology 12*, e1002012.

[13] Falkenberg, C. V., Carson, J. H., and Blinov, M. L. (2017) Multivalent Molecules as Modulators of RNA Granule Size and Composition, *Biophysical journal 113*, 235-245.

[14] McCann, James J., Choi, Ucheor B., and Bowen, Mark E. (2014) Reconstitution of Multivalent PDZ Domain Binding to the Scaffold Protein PSD-95 Reveals Ternary-Complex Specificity of Combinatorial Inhibition, *Structure 22*, 1458-1466.

[15] Jorissen, R. N., Walker, F., Pouliot, N., Garrett, T. P. J., Ward, C. W., and Burgess, A. W. (2003) Epidermal growth factor receptor: mechanisms of activation and signalling, *Experimental Cell Research 284*, 31-53.

[16] Huse, M. (2009) The T-cell-receptor signaling network, *Journal of Cell Science 122*, 1269-1273.

[17] Dustin, M. L., and Choudhuri, K. (2016) Signaling and Polarized Communication Across the T Cell Immunological Synapse, *Annu Rev Cell Dev Biol 32*, 303-325.

[18] Ketchum, C., Miller, H., Song, W., and Upadhyaya, A. Ligand Mobility Regulates B Cell Receptor Clustering and Signaling Activation, *Biophysical journal 106*, 26-36.

[19] Tolar, P., Sohn, H. W., Liu, W., and Pierce, S. K. (2009) The molecular assembly and organization of signaling active B-cell receptor oligomers, *Immunological Reviews 232*, 34-41.

[20] Zamir, E., and Geiger, B. (2001) Molecular complexity and dynamics of cell-matrix adhesions, *J Cell Sci 114*, 3583-3590.
[21] Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017) Biomolecular condensates: organizers of cellular biochemistry, *Nat Rev Mol Cell Biol 18*, 285-298.
[22] Shin, Y., and Brangwynne, C. P. (2017) Liquid phase condensation in cell physiology and disease, *Science 357*.
[23] Ditlev, J. A., Mayer, B. J., and Loew, L. M. (2013) There is More Than One Way to Model an Elephant. Experiment-Driven Modeling of the Actin Cytoskeleton, *Biophysical journal 104*, 520-532.
[24] Pollard, T. D., and Cooper, J. A. (2009) Actin, a central player in cell shape and movement, *Science 326*, 1208-1212.
[25] Andrews, S. S. (2017) Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface, *Bioinformatics 33*, 710-717.
[26] Blinov, M. L., Schaff, J. C., Vasilescu, D., Moraru, I. I., Bloom, J. E., and Loew, L. M. (2017) Compartmental and spatial rule-based modeling with Virtual Cell (VCell), *bioRxiv (and Biophys. J. in press)*.
[27] Czech, J., Dittrich, M., and Stiles, J. R. (2009) Rapid creation, Monte Carlo simulation, and visualization of realistic 3D cell models, *Methods Mol Biol 500*, 237-287.
[28] Saunders, M. G., and Voth, G. A. (2013) Coarse-graining methods for computational biology, *Annu Rev Biophys 42*, 73-93.
[29] Ibrahim, B., Henze, R., Gruenert, G., Egbert, M., Huwald, J., and Dittrich, P. (2013) Spatial Rule-Based Modeling: A Method and Its Application to the Human Mitotic Kinetochore, *Cells 2*, 506-544.
[30] Biedermann, J., Ullrich, A., Schöneberg, J., and Noé, F. (2015) ReaDDyMM: Fast Interacting Particle Reaction-Diffusion Simulations Using Graphical Processing Units, *Biophysical journal 108*, 457-461.
[31] Michalski, P. J., and Loew, L. M. (2016) SpringSaLaD: A Spatial, Particle-Based Biochemical Simulation Platform with Excluded Volume, *Biophysical journal 110*, 523-529.
[32] Chylek, L. A., Harris, L. A., Tung, C. S., Faeder, J. R., Lopez, C. F., and Hlavacek, W. S. (2014) Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems, *Wiley Interdiscip Rev Syst Biol Med 6*, 13-36.
[33] Hogg, J. S., Harris, L. A., Stover, L. J., Nair, N. S., and Faeder, J. R. (2014) Exact hybrid particle/population simulation of rule-based models of biochemical systems, *PLoS Comput Biol 10*, e1003544.
[34] Padrick, S. B., and Rosen, M. K. (2010) Physical Mechanisms of Signal Integration by WASP Family Proteins, *Annual review of biochemistry 79*, 707-735.
[35] Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q.-X., Nixon, B. T., and Rosen, M. K. (2012) Phase transitions in the assembly of multivalent signalling proteins, *Nature 483*, 336–340.
[36] Banjade, S., Wu, Q., Mittal, A., Peeples, W. B., Pappu, R. V., and Rosen, M. K. (2015) Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck, *Proc Natl Acad Sci U S A 112*, E6426-6435.
[37] Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012) Template-based protein structure modeling using the RaptorX web server, *Nature protocols 7*, 1511-1522.
[38] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modelling, prediction and analysis, *Nature protocols 10*, 845-858.

[39] Masison, J., Michalski, P. J., Loew, L. M., and Schuyler, A. D. (2018) mol2sphere: Spherical Decomposition of Multi-Domain Molecules for Visualization and Coarse Grained Spatial Modeling, *Bioinformatics*.
[40] Blasutig, I. M., New, L. A., Thanabalasuriar, A., Dayarathna, T. K., Goudreault, M., Quaggin, S. E., Li, S. S. C., Gruenheid, S., Jones, N., and Pawson, T. (2008) Phosphorylated YDXV Motifs and Nck SH2/SH3 Adaptors Act Cooperatively To Induce Actin Reorganization, *Molecular and Cellular Biology 28*, 2035-2046.
[41] Li, Shawn S.-C. (2005) Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction, *Biochemical Journal 390*, 641-653.
[42] Theillet, F.-X., Kalmar, L., Tompa, P., Han, K.-H., Selenko, P., Dunker, A. K., Daughdrill, G. W., and Uversky, V. N. (2013) The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins, *Intrinsically Disordered Proteins 1*, e24360.
[43] Pak, Chi W., Kosno, M., Holehouse, Alex S., Padrick, Shae B., Mittal, A., Ali, R., Yunus, Ali A., Liu, David R., Pappu, Rohit V., and Rosen, Michael K. (2016) Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein, *Molecular Cell 63*, 72-85.
[44] Schell, C., Baumhakl, L., Salou, S., Conzelmann, A.-C., Meyer, C., Helmstädter, M., Wrede, C., Grahammer, F., Eimer, S., Kerjaschki, D., Walz, G., Snapper, S., and Huber, T. B. (2013) N-WASP Is Required for Stabilization of Podocyte Foot Processes, *Journal of the American Society of Nephrology 24*, 713-721.
[45] New, L. A., Martin, C. E., Scott, R. P., Platt, M. J., Keyvani Chahi, A., Stringer, C. D., Lu, P., Samborska, B., Eremina, V., Takano, T., Simpson, J. A., Quaggin, S. E., and Jones, N. (2016) Nephrin Tyrosine Phosphorylation Is Required to Stabilize and Restore Podocyte Foot Process Architecture, *Journal of the American Society of Nephrology*.
[46] Harmon, T. S., Holehouse, A. S., Rosen, M. K., and Pappu, R. V. (2017) Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins, *eLife 6*, e30294.
[47] Ditlev, J. A., Michalski, P. J., Huber, G., Rivera, G. M., Mohler, W. A., Loew, L. M., and Mayer, B. J. (2012) Stoichiometry of Nck-dependent actin polymerization in living cells, *The Journal of Cell Biology 197*, 643-658.
[48] Padrick, S. B., Doolittle, L. K., Brautigam, C. A., King, D. S., and Rosen, M. K. (2011) Arp2/3 complex is bound and activated by two WASP proteins, *Proceedings of the National Academy of Sciences 108*, E472-E479.
[49] Simons, K., and Toomre, D. (2000) Lipid rafts and signal transduction, *Nat Rev Mol Cell Biol 1*, 31-39.