LETTER

doi:10.1038/nature10879

# Phase transitions in the assembly of multivalent signalling proteins

Pilong Li[1]*, Sudeep Banjade[1]*, Hui-Chun Cheng[1]*, Soyeon Kim[1], Baoyu Chen[1], Liang Guo[2], Marc Llaguno[3], Javoris V. Hollingsworth[4], David S. King[5], Salman F. Banani[1], Paul S. Russo[4], Qiu-Xing Jiang[3], B. Tracy Nixon[6] & Michael K. Rosen[1]

**Cells are organized on length scales ranging from ångström to micrometres. However, the mechanisms by which ångström-scale molecular properties are translated to micrometre-scale macroscopic properties are not well understood. Here we show that interactions between diverse synthetic, multivalent macromolecules (including multi-domain proteins and RNA) produce sharp liquid–liquid-demixing phase separations, generating micrometre-sized liquid droplets in aqueous solution. This macroscopic transition corresponds to a molecular transition between small complexes and large, dynamic supramolecular polymers. The concentrations needed for phase transition are directly related to the valency of the interacting species. In the case of the actin-regulatory protein called neural Wiskott–Aldrich syndrome protein (N-WASP) interacting with its established biological partners NCK and phosphorylated nephrin[1], the phase transition corresponds to a sharp increase in activity towards an actin nucleation factor, the Arp2/3 complex. The transition is governed by the degree of phosphorylation of nephrin, explaining how this property of the system can be controlled to regulatory effect by kinases. The widespread occurrence of multivalent systems suggests that phase transitions may be used to spatially organize and biochemically regulate information throughout biology.**

Covalent and non-covalent interactions between multivalent small molecules are central elements of classical polymer chemistry and physics and of supramolecular chemistry[2–4]. These fields have produced theories and experimental demonstrations of sharp transitions between small assemblies and macroscopic polymer gels (known as sol–gel transitions) as the degree of bonding increases. The transition point (critical point) depends on the physical properties of the monomeric species, including valency and affinity. The polymer can have a variety of physical forms, ranging from phase-separated liquid to crystalline solid. For non-covalent systems, phase separation can strongly influence the sol–gel transition by altering the degree of bonding[5,6]. In biology, interactions between multivalent entities are found in diverse processes, including extracellular carbohydrate–lectin binding, intracellular signalling, RNA metabolism and chromatin organization in the nucleus[7–10]. Biological multivalency has been studied most extensively in the context of extracellular ligands binding to cell surface receptors, where antibody–receptor[11] and carbohydrate–lectin[7] systems can assemble into crosslinked networks. These networks are typically precipitates[11,12], but liquid-like gels have also been described[13]. Multivalency has been less studied in the context of intracellular molecules, which often share the characteristics of high valency, modest affinity, and long, flexible connections between binding elements[14]. Here we asked whether these systems also undergo sharp transitions to polymers, and if so, what the macroscopic properties of the polymers are and how such transitions could be regulated and affect function.

Initially, we examined interactions between the SRC homology 3 (SH3) domain and its proline-rich motif (PRM) ligand, two widely observed modules that often appear in tandem arrays in signalling proteins[8,14]. We generated two classes of engineered proteins: one composed of repeats of a single SH3 domain ($SH3_m$, where $m = 1–5$), and the other composed of repeats of a PRM ligand ($PRM_n$, where $n = 1–5$) (dissociation constant ($K_d$) = 350 μM for the $SH3_1$–$PRM_1$ interaction; Supplementary Fig. 1). Initially, we mixed $SH3_4$ with $PRM_4$. At low concentrations, the solutions were clear; by contrast, at high concentrations, they were opalescent. Examination of these opalescent solutions using light microscopy showed the presence of numerous spherical droplets of approximately 1 μm to >50 μm in diameter that had phase-separated from the bulk solution (Fig. 1a and Supplementary Fig. 2). Smaller droplets tended to coalesce into larger droplets over time, which is consistent with liquid-like properties (Fig. 1b). When the proteins were mixed in a 1:1 ratio, both the droplet and bulk phases contained equal amounts of each molecule, but the proteins were concentrated by about 100-fold in the droplets relative to

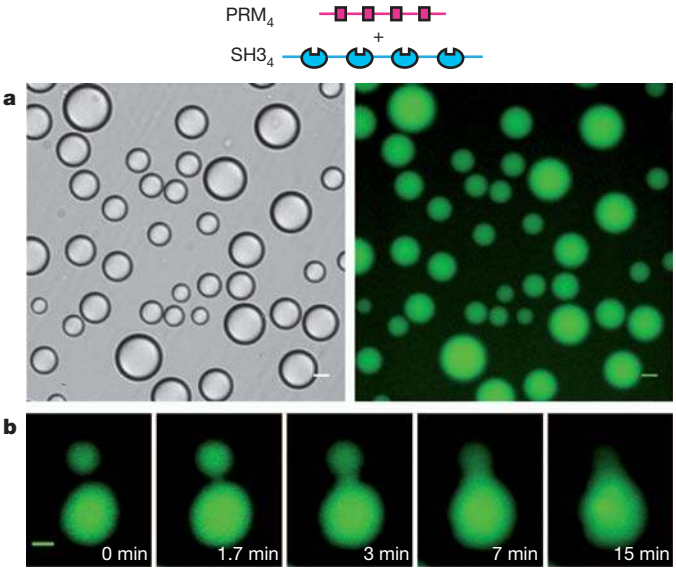



**Figure 1 | Macroscopic and microscopic phase transitions in multivalent SH3–PRM systems. a**, Liquid droplets observed by differential interference contrast microscopy (left) and wide-field fluorescence microscopy (right) when 300 μM $SH3_4$, 300 μM $PRM_4$ (both of which are module concentrations; molecule concentrations are 75 μM) and 0.5 μM OG–$SH3_4$ were mixed. Scale bars, 20 μm. **b**, Time-lapse imaging of merging droplets that were formed as in **a**. Scale bar, 10 μm.

[1]Department of Biochemistry and Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390-8812, USA. [2]BioCAT of IIT at the Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA. [3]Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9148, USA. [4]Department of Chemistry and Macromolecular Studies Group, Louisiana State University, Baton Rouge, Louisiana 70803, USA. [5]Howard Hughes Medical Institute Mass Spectrometry Laboratory and Department of Molecular & Cell Biology, University of California, Berkeley, California 94720, USA. [6]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.
*These authors contributed equally to this work.

the bulk phase (Fig. 1a, right; 116-fold for $SH3_5$ plus $PRM_5$; and 82-fold for $SH3_5$ plus an octameric dendrimer, PRM(N-WASP)$_8$). Analogous droplets were also observed for an unrelated $SH3_5$–ligand$_5$ pair and for the tetravalent RNA binding protein PTB interacting with an RNA oligonucleotide (Supplementary Fig. 3). Thus, liquid–liquid-demixing phase transitions may occur in many multivalent intracellular systems.

A large body of data indicates that the phase separation observed here is driven by the assembly of the multivalent proteins into large species, analogous to the behaviour observed in many small-molecule polymer systems[2,5] as well as in covalent protein crosslinking[15]. First, the phase boundary is strongly dependent on the valency of the interacting species (Fig. 2a and Supplementary Fig. 4). This observation is consistent with theory and our simulations, which indicate that a higher valency allows the formation of larger species at a lower fractional saturation of the binding modules[2,3,5,16] (Supplementary Information and Supplementary Figs 5–7).

Second, the phase transition can be blocked by a high-affinity monovalent ligand, PRM(H)$_1$ ($K_d = 10\,\mu M$ for $SH3_1$ binding to PRM(H)$_1$; Supplementary Fig. 8). Third, we used dynamic light scattering (DLS) and small-angle X-ray scattering (SAXS) to characterize the species that formed during titrations of PRM proteins into $SH3_5$ (Fig. 2b and Supplementary Figs 10–12). At concentrations below those needed for droplet formation, the addition of either $PRM_2$ or $PRM_4$ substantially increased the apparent hydrodynamic radius ($R_h$, for DLS) and the apparent radius of gyration ($R_g$, for SAXS), suggesting that oligomeric species formed. For equal total module concentrations, $PRM_4$ caused larger increases in these values than did $PRM_2$, consistent with predictions from polymer theory[2,3,5,16]. For higher concentrations of PRM proteins, we removed the droplets by centrifugation and found that increasing the concentration of $PRM_2$ or $PRM_4$ caused $R_h$ and $R_g$ to decrease in the bulk phase, suggesting that larger species partitioned selectively into the droplet phase. This behaviour resembles that observed in the sol–gel transitions of covalent polymers, in which, above the critical extent of reaction, the average size of oligomers in the sol phase decreases because larger species preferentially join the gel[2]. Titrations of $PRM_1$ or PRM(H)$_1$, which unlike $PRM_2$ and $PRM_4$ cannot generate polymers and did not cause phase separation, produced only small, saturable increases in $R_h$ and $R_g$. Taken together, these data indicate that phase separation is driven by the unique ability of multivalent $SH3_m$–$PRM_n$ interactions to create large assemblies.

Additional lines of evidence suggest that the multivalent proteins formed large polymers within the droplets, such that the phase transition probably coincides with a sol–gel transition. First, at the extremely high concentrations in the droplet phase, the fractional saturation of SH3 and PRM binding sites is estimated to be threefold to fivefold higher than that needed to induce the sol–gel transition[16] (Supplementary Materials, Cyclization).

Second, DLS analyses of the droplet phases created by mixing $SH3_m$ and $PRM_n$ showed multiphase intensity autocorrelation curves with a complex distribution of relaxation times spanning 0.2–20 ms (Supplementary Fig. 13) or longer (Fig. 2c). Some of these relaxation times nonlinearly scale with the square of the scattering angle ($q^2$) (Supplementary Fig. 13). The wide range of timescales, the presence of long-timescale processes and the $q^2$ independence of some of these processes are properties that are typical of polymer solutions in the semi-dilute range but are highly atypical of discrete molecular species[17].

Third, photobleaching experiments showed that the diffusion of the droplet constituents was slowed by about three orders of magnitude relative to free diffusion in water (data not shown). Furthermore, the photobleaching recovery rate correlated inversely with the monomer–monomer affinity and valency (Supplementary Fig. 14), suggesting that recovery represents reorganization of a polymer matrix. Small-molecule fluorophores and enhanced green fluorescent protein (eGFP) can enter, and move rapidly within, the droplets (Supplementary Fig. 15).

Last, cryo-electron microscopy images of $SH3_5$ plus $PRM_5$ solutions that were flash-frozen immediately after mixing showed numerous round, electron-dense objects with diameters of about 50–500 nm (Fig. 2d and Supplementary Fig. 16). These objects were not observed with high concentrations of either $SH3_5$ or $PRM_5$ alone or when the two components were mixed at concentrations below the droplet concentration. Although the objects were too dense for structures to be observed within them, their edges were highly irregular on the ~10-nm scale, suggesting that they contained large but disordered molecular species, consistent with a crosslinked gel. These characterizations of the droplet phase suggest that this phase contains large polymeric species and has undergone a sol–gel transition.

Taken together, our data suggest a model in which the association of multivalent proteins produces a macroscopic liquid–liquid phase separation, which is thermodynamically coupled to a molecular sol–gel transition within the droplet phase. The sharp transition between small complexes and polymers is consistent with condensation polymerization theory[2,3,6] and our own particle-based simulations (Supplementary Fig. 5). A theoretical analysis of these systems (Supplementary Information) indicates that the polymerization process is driven appreciably by the extremely high configurational entropy of the polymer[6,18].

The behaviour that we observed for these multivalent systems *in vitro* was mirrored in cells. The coexpression of mCherry–$SH3_5$ and eGFP–$PRM_5$ fusion proteins in HeLa cells resulted in the formation of approximately 0.5–2-μm diameter cytoplasmic puncta containing both fluorophores (Fig. 3a). Puncta were not observed in cells expressing either protein alone or in cells coexpressing mCherry–$SH3_5$ and eGFP–$PRM_3$, indicating that their formation depends on the interaction between the two high-valency molecules. The puncta did not stain with a large range of vesicle markers or a lipid dye, suggesting that they are phase-separated bodies rather than vesicular structures (Supplementary Fig. 18). Both mCherry and eGFP fluorescence of the bodies recovered within about 10 s of photobleaching (Fig. 3b), indicating that there was rapid exchange of both components with the

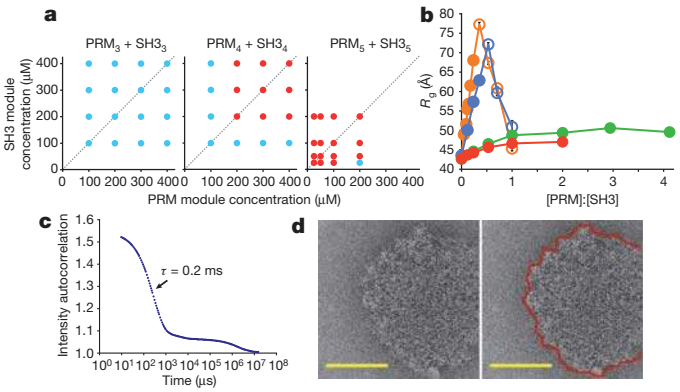



**Figure 2 | Multivalency drives phase separation and probably drives a sol–gel transition in the droplet phase.** **a**, Phase diagrams of multivalent SH3 and PRM proteins. The concentrations are in terms of the modules. The red circles indicate phase separation, and the blue circles indicate no phase separation. **b**, The $R_g$ values determined from SAXS data that were collected during titrations of PRM proteins into $SH3_5$. Closed circles indicate the absence of phase separation; open circles indicate data collected on the supernatant phase, which was separated from the droplets by centrifugation. The titrations used $PRM_4$ (orange), $PRM_2$ (blue), $PRM_1$ (green) and PRM(H)$_1$ (red). The error bars represent the s.d. calculated from five to ten independent measurements of intensity versus scattering angle ($q$). **c**, The intensity autocorrelation curve of light scattered at 90° from the pooled droplet phase of $SH3_5$ plus PRM(N-WASP)$_8$. $\tau$, the relaxation time constant of the most rapidly decaying phase. **d**, Cryo-electron microscopy image of a droplet formed by $SH3_5$ plus $PRM_5$ (identical image, left and right). The edge of the structure is outlined in red (right). Scale bars, 100 nm.

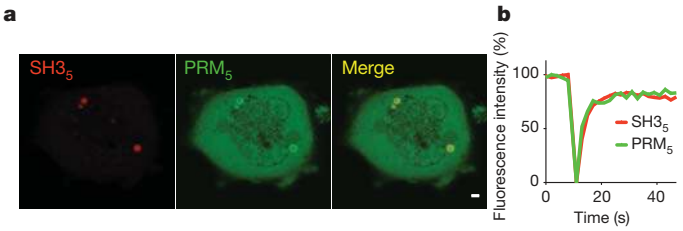



**Figure 3 | Coexpression of SH3$_5$ and PRM$_5$ in cells produces dynamic puncta. a**, mCherry–SH3$_5$ (left), eGFP–PRM$_5$ (centre) and an image overlay in a cell expressing both proteins (right). It should be noted that the non-uniform eGFP fluorescence in the puncta results from mCherry–eGFP fluorescence resonance energy transfer (FRET) rather than from differential localization of the proteins (Supplementary Fig. 17). Scale bar, 2 μm. **b**, Both mCherry and eGFP fluorescence recover rapidly after photobleaching.

surrounding cytoplasm and suggesting that the bodies had a dynamic liquid-like nature. Thus, interactions between multivalent proteins can produce phase-separated liquid droplets both *in vitro* and in cells.

The nephrin–NCK–N-WASP system is a natural, three-component interaction that can be used to investigate phase transitions that result from multivalent interactions, as well as the functional consequences of these transitions (Fig. 4a). In kidney podocytes, the transmembrane protein nephrin plays a central role in forming the glomerular filtration barrier, functioning partly through assembling cortical actin[1]. The cytoplasmic tail of nephrin contains three tyrosine phosphorylation (pTyr) sites, which can each bind the SH2 domain of NCK[1,19]. NCK contains three SH3 domains, which can bind the six PRMs in the proline-rich region of N-WASP[20]. N-WASP, in turn, stimulates the nucleation of actin filaments by the Arp2/3 complex. The multivalency of nephrin or NCK is necessary for proper actin assembly[19] and, together with the multivalency of N-WASP, has the potential to cause phase transitions.

The addition of NCK to an N-WASP construct (GBD-P-VCA, Supplementary Table 1; called N-WASP hereafter) caused droplet formation, as occurred in the model systems described above (Fig. 4b). The addition of a diphosphorylated (2pTyr) nephrin tail peptide dropped the phase boundary for both proteins by more than or equal to twofold (Fig. 4c), presumably because the effective valency of NCK increases when it is arrayed on nephrin. This effect was even more pronounced when nephrin–3pTyr peptide was added (to the same total pTyr concentration) (Fig. 4d). Thus, in cells, protein kinases could regulate phase transitions in this system (and the cooperative assembly of all three proteins) by controlling the degree of phosphorylation of nephrin and, consequently, by shifting the phase boundary from the micromolar (Fig. 4b) to the nanomolar (Fig. 4d) regime. It should be noted that, where measured, the cytoplasmic concentrations of WASP proteins and other actin-regulatory molecules were typically 1–10 μM, indicating that shifts in this range could be functionally important (Supplementary Fig. 19).

We next asked how droplet formation affects the ability of the nephrin–NCK–N-WASP system to stimulate Arp2/3-mediated actin assembly. We measured the half-time to completion ($t_{1/2}$) of pyrene–actin assembly reactions containing fixed concentrations of the Arp2/3 complex, NCK, nephrin–3pTyr peptide and N-WASP plus variable amounts of an N-WASP truncation mutant (N-WASPΔ) that contains the full proline-rich region and can assemble into polymers but cannot bind the Arp2/3 complex. In the absence of N-WASPΔ, 50 nM N-WASP produced only weak stimulation of the Arp2/3 complex (long $t_{1/2}$) (Fig. 4e). The addition of N-WASPΔ to concentrations between 0 and 750 nM had no effect on actin assembly. However, 1,000 nM N-WASPΔ sharply increased the activation of the Arp2/3 complex. The activity increased asymptotically as the concentration of N-WASPΔ was raised further. N-WASPΔ had no effect on an N-WASP protein that lacks the proline-rich region, suggesting that

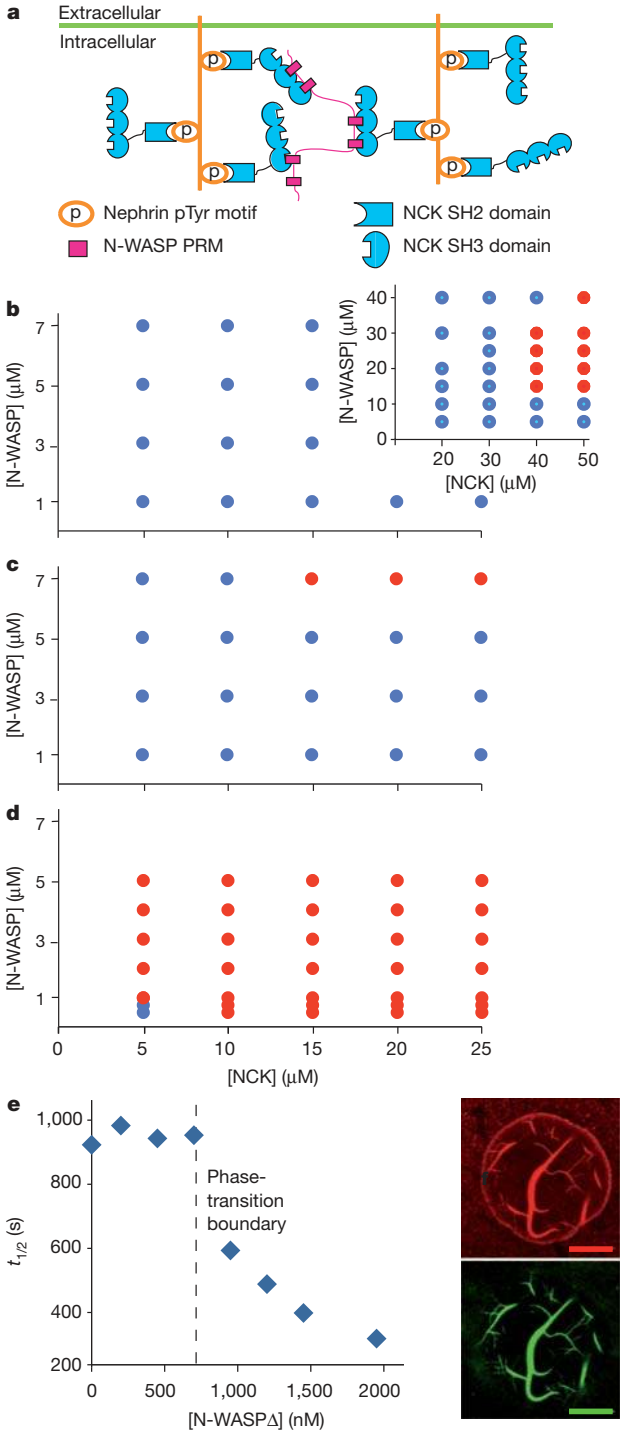



**Figure 4 | Phase transition correlates with biochemical activity transition in the nephrin–NCK–N-WASP system. a**, The interactions of nephrin, NCK and N-WASP. **b**–**d**, Phase diagrams of N-WASP and NCK alone (**b**, inset shows higher concentrations) or in the presence of 4.5 μM diphosphorylated nephrin tail peptides (**c**) or 3 μM triphosphorylated nephrin tail peptides (**d**). The red circles indicate phase separation, and the blue circles indicate no phase separation. **e**, The half-time to completion ($t_{1/2}$) of N-WASP-stimulated actin assembly by the Arp2/3 complex as a function of N-WASPΔ concentration. The vertical dashed line indicates the phase-separation boundary determined in separate assays without actin and the Arp2/3 complex. **f**, Rhodamine–actin (4 μM, 10% rhodamine-labelled), 300 nM Alexa-488–phalloidin and 10 nM Arp2/3 complex were added to droplets containing triphosphorylated nephrin, NCK and N-WASP, and the droplets were imaged by confocal microscopy; rhodamine (top) and Alexa 488 (bottom) are shown separately. Scale bars, 10 μm.

engagement by NCK is needed for the increase in activity (Supplementary Fig. 20).

These data are consistent with the switch-like formation of a higher activity[21,22] (probably polymeric) form of nephrin–NCK–N-WASP when the total concentration of wild-type (active) plus truncated (inactive) N-WASP surpasses that needed for droplet formation. The sharp transition was followed by slower increases in activity as additional N-WASPΔ drew a greater percentage of the wild-type protein into the droplets/polymer. Consistent with these ideas, when mixtures containing N-WASP above its phase-separation concentration were stained with phalloidin, numerous bundles of actin filaments were observed within the droplets (Fig. 4f). In cells, nephrin is a transmembrane protein, and the system would not produce a three-dimensional polymer phase as observed here but rather its two-dimensional equivalent at the plasma membrane. Such interactions might contribute to the formation and/or the stability of the micrometre-scale clusters of nephrin and NCK with the associated actin tails, which are observed in cells when nephrin is crosslinked (and subsequently phosphorylated)[1]. Of the 28 known binding partners of the NCK SH2 domain, 14 are predicted or have been demonstrated to contain three or more pTyr sites[23,24], suggesting that analogous pTyr–NCK–N-WASP assembly may occur in many systems.

Sharp phase transitions occurring concomitantly with sol–gel transitions may be a general feature of multivalent systems in biology. Many 'cellular bodies'—that is, subcellular compartments that are compositionally distinct from the surrounding cytoplasm or nucleoplasm but that are not membrane bounded[25,26]—are enriched in multivalent proteins and nucleic acids[9,27]. These include promyelocytic leukaemia nuclear bodies, Cajal bodies, P bodies and P granules[27–29]. Moreover, P granules in the *Caenorhabditis elegans* embryo were recently shown to have liquid-like properties with many of the features that we observed here, including switch-like formation and greatly slowed diffusion of constituent molecules[30]. Many multivalent proteins, in general, can organize into puncta in the cytosol or at membranes (Supplementary Fig. 21). Within puncta, the physical properties of a polymer could impart micrometre-scale structural and dynamic organization and control the chemistry (for example, of catalysis, molecular interactions or structural rearrangements). Our findings provide a mechanism by which multivalent interactions could yield sharp transitions between physically and functionally distinct states, generating nonlinearity in signalling pathways, connecting disparate length scales in the cell, and perhaps contributing to the structure and function of cellular bodies and other two- and three-dimensional compartments.

## METHODS SUMMARY

**Material generation.** Details are provided in the Supplementary Information.

***In vitro* phase separation.** Samples were incubated for >12 h before scoring them for droplets using bright-field microscopy; when observed, the droplets formed immediately after mixing. For *in vitro* fluorescence recovery after photobleaching (FRAP) experiments, a 5-μm diameter spot was bleached in droplets of >20-μm diameter containing Oregon Green (OG)-labelled $SH3_4$ (OG–$SH3_4$) by using a 488-nm laser line. The mean intensity of the bleached spot was fit to a single exponential.

**DLS and SAXS.** For titrations monitored by DLS (DynaPro, Wyatt) and SAXS (Advanced Photon Source, Biophysics Collaborative Access Team (BioCAT) beamline), the samples contained 170 μM $SH3_5$ plus PRM proteins at PRM:SH3 module ratios of 0–5. Droplets were removed from all of the relevant samples by centrifugation (at 16,000*g* for 10 min) before analysis. For $SH3_5$ plus PRM(N-WASP)$_8$ and $SH3_5$ plus $PRM_5$ droplets, single-angle DLS data were collected on a DynaPro instrument, and multi-angle DLS data were collected on a custom-built apparatus.

**Cryo-electron microscopy.** The samples were blotted, frozen immediately after mixing and imaged under low-dose cryo conditions in a 2200FS FEG transmission electron microscope (JEOL), using a 2k × 2k slow-scan charge-coupled display (CCD) camera (Tietz).

**Cellular assays.** HeLa cells were imaged at 32 °C 24 h after transfection with vectors expressing mCherry–$SH3_5$ and/or eGFP–$PRM_5$ (or eGFP–$PRM_3$). Images were acquired on an LSM 510 confocal microscope (Zeiss). For FRAP experiments, individual puncta containing both mCherry–$SH3_5$ and eGFP–$PRM_5$ were bleached with a 488-nm laser line.

**Actin polymerization.** Actin (4 μM, 5% pyrene-labelled actin) was polymerized in the presence of 3 μM nephrin–3pTyr, 10 μM NCK, 50 nM N-WASP and 10 nM Arp2/3 complex plus increasing concentrations of N-WASPΔ in 150KMEI buffer (150 mM KCl, 1 mM $MgCl_2$, 1 mM EGTA and 10 mM imidazole, pH 7) as described previously[21]. Imaging was performed on an LSM 510 confocal microscope after adding rhodamine–actin (4 μM, 10% rhodamine-labelled actin), 300 nM Alexa 488–phalloidin and 10 nM Arp2/3 complex to droplets containing 3 μM nephrin–3pTyr, 10 μM NCK and 2 μM N-WASP.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 4 May 2010; accepted 20 January 2012.**
**Published online 7 March 2012.**

1. Jones, N. *et al.* Nck adaptor proteins link nephrin to the actin cytoskeleton of kidney podocytes. *Nature* **440,** 818–823 (2006).
2. Flory, P. J. *Principles of Polymer Chemistry* (Cornell Univ. Press, 1953).
3. Cohen, R. J. & Benedek, G. B. Equilibrium and kinetic theory of polymerization and the sol-gel transition. *J. Phys. Chem.* **86,** 3696–3714 (1982).
4. Lehn, J.-M. Supramolecular polymer chemistry—scope and perspectives. *Polym. Int.* **51,** 825–839 (2002).
5. Tanaka, F. *Polymer Physics: Applications to Molecular Association and Thermoreversible Gelation* (Cambridge Univ. Press, 2011).
6. Semenov, A. N. & Rubinstein, M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. *Macromolecules* **31,** 1373–1385 (1998).
7. Brewer, C. F., Miceli, M. C. & Baum, L. G. Clusters, bundles, arrays and lattices: novel mechanisms for lectin–saccharide-mediated cellular interactions. *Curr. Opin. Struct. Biol.* **12,** 616–623 (2002).
8. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300,** 445–452 (2003).
9. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nature Rev. Mol. Cell Biol.* **8,** 479–490 (2007).
10. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. Multivalent engagement of chromatin modifications by linked binding modules. *Nature Rev. Mol. Cell Biol.* **8,** 983–994 (2007).
11. Goldberg, R. A theory of antibody–antigen reactions. I. Theory for reactions of multivalent antigen with bivalent and univalent antibody. *J. Am. Chem. Soc.* **74,** 5715–5725 (1952).
12. Dam, T. K. *et al.* Thermodynamic, kinetic, and electron microscopy studies of concanavalin A and *Dioclea grandiflora* lectin cross-linked with synthetic divalent carbohydrates. *J. Biol. Chem.* **280,** 8640–8646 (2005).
13. Sisu, C. *et al.* The influence of ligand valency on aggregation mechanisms for inhibiting bacterial toxins. *Chembiochem* **10,** 329–337 (2009).
14. Jin, J. *et al.* Eukaryotic protein domains as functional units of cellular evolution. *Sci. Signal.* **2,** ra76 (2009).
15. Asherie, N. *et al.* Oligomerization and phase separation in globular protein solutions. *Biophys. Chem.* **75,** 213–227 (1998).
16. Stockmayer, W. H. Molecular distribution in condensation polymers. *J. Polym. Sci.* **9,** 69–71 (1952).
17. Li, J., Ngai, T. & Wu, C. The slow relaxation mode: from solutions to gel networks. *Polym. J.* **42,** 609–625 (2010).
18. Semenov, A., Charlot, A., Auzely-Velty, R. & Rinaudo, M. Rheological properties of binary associating polymers. *Rheol. Acta* **46,** 541–568 (2007).
19. Blasutig, I. M. *et al.* Phosphorylated YDXV motifs and Nck SH2/SH3 adaptors act cooperatively to induce actin reorganization. *Mol. Cell. Biol.* **28,** 2035–2046 (2008).
20. Rohatgi, R., Nollau, P., Ho, H. Y., Kirschner, M. W. & Mayer, B. J. Nck and phosphatidylinositol 4,5-bisphosphate synergistically activate actin polymerization through the N-WASP–Arp2/3 pathway. *J. Biol. Chem.* **276,** 26448–26452 (2001).
21. Padrick, S. B. *et al.* Hierarchical regulation of WASP/WAVE proteins. *Mol. Cell* **32,** 426–438 (2008).
22. Padrick, S. B. & Rosen, M. K. Physical mechanisms of signal integration by WASP family proteins. *Annu. Rev. Biochem.* **79,** 707–735 (2010).
23. Lettau, M., Pieper, J. & Janssen, O. Nck adaptor proteins: functional versatility in T cells. *Cell Commun. Signal.* **7,** 1 (2009).
24. Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* **31,** 3635–3641 (2003).
25. Matera, A. G., Izaguire-Sierra, M., Praveen, K. & Rajendra, T. K. Nuclear bodies: random aggregates of sticky proteins or crucibles of macromolecular assembly? *Dev. Cell* **17,** 639–647 (2009).
26. Buchan, J. R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* **36,** 932–941 (2009).
27. Parker, R. & Sheth, U. P bodies and the control of mRNA translation and degradation. *Mol. Cell* **25,** 635–646 (2007).
28. Matera, A. G. & Shpargel, K. B. Pumping RNA: nuclear bodybuilding along the RNP pipeline. *Curr. Opin. Cell Biol.* **18,** 317–324 (2006).
29. Bernardi, R. & Pandolfi, P. P. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nature Rev. Mol. Cell Biol.* **8,** 1006–1016 (2007).

30. Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324,** 1729–1732 (2009).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank J. Onuchic and S. Padrick for discussion of the theoretical aspects of this study, L. Rice for sharing his fluorescence microscope, M. Socolich for a gift of purified eGFP, K. Luby-Phelps and A. Bugde for advice on FRAP experiments, S. Padrick and L. Doolittle for help in purifying actin and the Arp2/3 complex and for sharing reagents, N. Grishin and S. Shi for help with database searches, K. Lynch for providing the PTB expression construct, D. Billadeau and T. Gomez for providing antibodies, A. Ramesh, W. Winkler and P.-L. Tsai for advice on RNA experiments, K. Roybal and C. Wülfing for sharing unpublished data, and J. Liu for help with cryo-electron tomography. This work was supported by the following: the Howard Hughes Medical Institute and grants from the National Institutes of Health (NIH) (R01-GM56322) and Welch Foundation (I–1544) to M.K.R., a Chilton Foundation Fellowship to H.-C.C., an NIH EUREKA award (R01-GM088745) to Q.-X.J., an NIH Cancer Biology T32 Training Grant to M.L., a National Science Foundation award (DMR-1005707) to P.S.R. and a Gates Millennium Fund award to J.V.H. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract number W-31-109-ENG-38. BioCAT is NIH-supported Research Center RR-08630.

**Author Contributions** M.K.R. oversaw the project, helped analyse all of the data and wrote the paper with assistance from all of the authors. P.L., H.-.C.C. and M.K.R. conceived of the project. P.L. developed and interpreted the theoretical and computational models, which promoted much of the experimentation. S.B. performed and analysed experiments on the nephrin–NCK–N-WASP system and performed monovalent competition studies. H.-.C.C. mapped and analysed the phase diagrams, and collected FRAP data, on the engineered model systems. S.K. performed and analysed the cellular experiments. S.B., B.C., L.G. and B.T.N. collected and/or analysed the SAXS data. S.B., M.L. and Q.-X. J. collected and/or analysed the electron microscopy data. S.B., J.V.H. and P.S.R. collected and/or analysed the multi-angle DLS data. H.-.C.C. and S.B. collected and analysed the single-angle DLS data. D.S.K. synthesized the octameric PRM dendrimer. S.F.B. analysed the cyclization in the sol–gel transition.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.K.R. (Michael.Rosen@utsouthwestern.edu).

## METHODS

**Measurements of protein stoichiometry and concentration in droplets.** All phase-separation assays were performed in 150KMEI buffer (150 mM KCl, 1 mM $MgCl_2$, 1 mM EGTA and 10 mM imidazole, pH 7), except where indicated. $SH3_4$ and $PRM_4$ (doped with 0.5% of either protein labelled with Oregon Green) were incubated at a 1:1 stoichiometry at various concentrations at room temperature overnight. The droplet phase (where formed) was collected by centrifugation, and the supernatant was removed and digested with proteinase K (Promega) overnight. The fluorescence intensity (excitation wavelength ($\lambda_{ex}$) = 488 nm, and emission wavelength ($\lambda_{em}$) = 512–521 nm) of the samples was converted to concentration using a standard curve of Oregon Green alone. The data revealed a 1:1 stoichiometry in all cases in both the droplet and the supernatant phases. In other systems that were mixed at a 1:1 module ratio, we calculated the concentrations from the absorbance at 280 nm using module-weighted extinction coefficients, assuming a 1:1 module stoichiometry in all phases.

**SAXS.** SAXS experiments were performed at the Biophysics Collaborative Access Team (BioCAT) undulator beamline 18-ID at the Advanced Photon Source, Argonne National Laboratory as described previously[31]. All samples were centrifuged for 10 min at 16,000*g* at 4 °C before mixing. In mixtures above the phase-transition concentration, solutions were centrifuged for an additional 10 min to remove the droplet phase before SAXS analysis. To minimize radiation damage, all samples were in 150KMEI buffer containing 5 mM TCEP and were flowed at 5 μl $s^{-1}$ during data collection[31]. Scattering intensity profiles over the $q$ range from 0.006 to 0.36 $Å^{-1}$ were calculated from radial averaging of the two-dimensional scattering patterns using macros written by the BioCAT staff for the software IGOR Pro (WaveMetrics). The scattering for each sample and the accompanying buffer were averaged from five to ten repeats, and the protein scattering was obtained by subtracting the buffer scattering. Guinier plots were used to calculate $R_g$ and to ascertain the absence of aggregates. Alternatively, $R_g$ and the distance distribution function, $P(r)$, were calculated using the program GNOM[32]. The apparent molecular weight was estimated[33] by integrating the Kratky plot to a maximum $q$ at 0.2 $Å^{-1}$.

**DLS titration.** In the titration experiments, scattering measurements were performed on a DynaPro instrument (Wyatt) at 30% laser power and 22 °C. Twenty repetitive data sets, each of 100 s duration, were averaged for each titration point. Autocorrelation data were analysed using the regularization method in Dynamics version 6.4.3 software (Proterion Corporation). All samples were centrifuged before and after mixing, as for the SAXS titration.

**DLS of $SH3_5$ plus $PRM(N\text{-}WASP)_8$ droplets.** To collect enough volume of the droplet phase (~25 μl), 850 μM $SH3_5$ was mixed with 850 μM $PRM(N\text{-}WASP)_8$ (module concentrations) to a final volume of 900 μl in 150KMEI buffer. The denser, droplet, phase was collected by centrifugation (16,000*g* for 10 min) and analysed using a DynaPro instrument at 22 °C and 20% laser power. Twenty repetitive data sets, each of 1,000 s duration, were averaged.

**Multi-angle DLS of $SH3_5$ plus $PRM_5$ droplet and supernatant phases.** $SH3_5$ (850 μM) was mixed with $PRM_5$ (850 μM) (module concentrations) to a final volume of 900 μl to collect enough volume of the droplet phase (~20 μl). The denser, droplet, phase was collected by centrifugation. Measurements were made using a custom-built apparatus equipped with an Innova 90 argon laser (Coherent) set to 514.5 nm. A wide-range photometer–preamplifier–discriminator (Pacific Precision Instruments) drove a pulse shaper (ALV), which fed an ALV-5000 digital autocorrelator (ALV). The temperature was maintained at 25 °C by a circulating water bath. For the supernatant phase, three 10-min runs were collected and averaged at each scattering angle between 30° and 110°. For the droplet phase, a single 30-min run was collected at each angle. The distributions of the decay rate and the average decay rate were determined by analysing the correlation functions using an inverse Laplace transformation in the program CONTIN[34].

**Cryo-electron microscopy.** Carbon-coated copper grids were glow-discharged in an EMS-100 unit with a 40 mA current for 90 s. $SH3_5$ (100 μM) was loaded onto the grid, followed by $PRM_5$ (100 μM) (module concentrations), giving a 4 μl final volume. The mixture was immediately blotted for 6.5 s in a Vitrobot Mark III (Gatan) before being plunged into liquid ethane. Mixtures of $SH3_5$ and $PRM_5$ below the critical concentrations (2.5 μM module concentration each), or $SH3_5$ or $PRM_5$ alone (250 μM module concentration) were prepared identically. Samples contained 150KMEI buffer. Frozen grids were stored in liquid nitrogen until electron microscopy analysis. Imaging was performed under low-dose cryo conditions in a 2200FS FEG transmission electron microscope (JEOL), using a 2k × 2k slow-scan charge-coupled display (CCD) camera (Tietz). The dose rate was adjusted to ~20 electrons $Å^{-2}$ $s^{-1}$; the defocus level varied from −1.0 to −3.0 μm.

**Cell culture and fluorescence microscopy.** HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% FBS. Cells were transfected using Effectene (QIAGEN) and incubated for 24 h at 37 °C. For Nile-Red staining, cells were incubated with Nile Red (at a final concentration of 0.5 μg $ml^{-1}$) for 15, 30 or 60 min before imaging. For immunofluorescence analysis, transfected cells were fixed with 4% paraformaldehyde in PBS and permeabilized with 0.1% Triton X-100 in PBS for 5 min. The antibodies used for staining were anti-EEA1, anti-GM130, anti-golgin-97, anti-VPS35, anti-CIMPR, anti-clathrin heavy chain, anti-LAMP1, anti-SNX1, anti-SNX2, anti-TGN46 (provided by D. Billadeau and T. Gomez) and anti-caveolin-1 antibodies. Secondary antibodies conjugated to Alexa Fluor 647 were obtained from Invitrogen. Images were collected on a DeltaVision system (Applied Precision). The mCherry–$SH3_5$ and eGFP–$PRM_5$ constructs were generated by cloning $SH3_5$ and $PRM_5$ into the vectors pmCherry-N1 and peGFP-N1, respectively. Images were processed using the program ImageJ (http://rsb.info.nih.gov/ij).

***In vitro* FRAP assays.** Fluorescence recovery after photobleaching (FRAP) experiments were performed on an LSM 510 confocal microscope (Zeiss) using a 488-nm laser line. The OG–$SH3_4$ or eGFP was bleached using ten iterative pulses (total time ~0.55 s) with full laser power. The droplets were >20 μm in diameter, and the bleached spots were 5 μm in diameter. The imaging power was 0.5% for OG–$SH3_4$ and 2% for eGFP. Images were processed using ImageJ. Background-corrected data were fit to a single exponential decay to yield the recovery time constant using GraphPad Prism 5 (GraphPad Software).


31. Chen, B. *et al.* ATP ground- and transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, σ54. *Structure* **15,** 429–440 (2007).
32. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25,** 495–503 (1992).
33. Fischer, H., Neto, M. O., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43,** 101–109 (2010).
34. Provencher, S. W. CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput. Phys. Commun.* **27,** 229–242 (1982).

doi:10.1038/nature10879

# SUPPLEMENTARY MATERIALS

## Simulations and Theory

We developed a particle simulation algorithm to model two-component multivalent-multivalent interactions. The model consists of two kinds of protein molecules (particles), each composed of multiple repeats of a distinct binding domain. The binding domains in the first molecule type have affinity for the binding domains in the second molecule type. The domains in a given molecule are implicitly flexibly connected, such that they function independently. The system is described by only 4 parameters, which represent local crowdedness and the probabilities of association, dissociation and diffusion. Even with these simple rules, under many conditions we observe assembly of the particles into large polymers over time (Movie S1). Polymerization occurs readily at high concentrations but not at low concentrations. The transition from a high probability to a low probability of forming polymer occurs over a narrow concentration range (Supplementary Fig. 5), indicating a sharp phase boundary. The boundary steadily shifts to lower concentrations as valency increases from 3 to 6 (Supplementary Fig. 6). Our simulations do not use real binding affinities or times, so the exact positions of the transitions here do not relate to specific physical parameters. But the valency dependence should hold qualitatively independent of these parameters. A theoretical analysis of these systems (see below and Supplementary Figs. 22, 23, 7) indicates that the polymerization process is driven primarily by the extremely high configurational entropy (i.e. the enormous number of bond configurations) of the polymer, as argued for analogous systems in polymer physics[1,2].

**Simulation of multivalent/multivalent interactions**

A rule-based stochastic simulation program (available upon request) was written in C to simulate interactions of multivalent molecules in a rectangular cuboid built of unit cubes with dimensions of 20 nm x 20 nm x 20 nm. Two sets of molecules are created: one contains m = 1-6 identical α domains fused together (implicitly through flexible linkers, to give molecules $\alpha_m$) and the other similarly contains n = 1-6 identical β domains (to give molecules $\beta_n$). Domains α and β can bind each other with 1:1 stoichiometry. For experiments with varying concentrations of molecules, the numbers of molecules are kept constant (phase diagram mapping in Supplementary Fig. 6) while adjusting the number of unit cubes accordingly or the numbers of unit cubes are kept constant while adjusting the number of molecules accordingly (equilibrium constant determination in Supplementary Fig. 22, see below for details).

For a simulation experiment, the appropriate number of $\alpha_m$ and $\beta_n$ molecules are randomly distributed in the virtual space at t = 0. At t = 1, the program cycles through each non-empty unit cube and determines whether both free α domains and free β domains exist. When the number of free α domains (#free_α) is no fewer than the number of free β domains (#free_β), the program steps through each free β domain and allows it try to associate with free α domains one by one until it binds or all α domains are tried. The procedure is opposite for #free_α < #free_β. The probability ($p_{on}^{eff}$) to associate for each trial is:

$$p_{on}^{eff} = p_{on}^{int} \times (1 - p_{off}^{int}) \times \frac{20}{20 + \#\alpha + \#\beta}$$

in which $p_{on}^{int}$ (= 0.6 in the current study) is the intrinsic association probability, $p_{off}^{int}$ (= 0.1 in the current study) is the intrinsic dissociation probability and the last term is an arbitrary scaling factor to reflect the crowdedness (total number of domains α and β in the unit cube). Two or more molecules can become physically associated through newly formed α-β bonds, producing a multi-molecule particle, or interactions can occur within an existing particle. After association, each molecule and complex is given the opportunity to diffuse to one of the six neighboring cubes (randomly chosen) with a probability ($p_{diff}$):

$$p_{diff} = \frac{1}{\sqrt{\#\alpha_m + \#\beta_n}} \times \frac{1}{6}$$

where $\#\alpha_m$ and $\#\beta_n$ are the number of α domain containing molecules and the number of β domain containing molecules, respectively. If the targeted cube does not exist, i.e. the particle tries to move across one face of the virtual space, the particle remains in the current cube. At $t \geq 2$, in each non-empty unit cube the program determines whether both free α domains and free β domains exist and whether α-β bonds exist. If α-β bonds exist, they are given the opportunity to dissociate according to a dissociation probability ($p_{off}^{eff}$):

$$p_{off}^{eff} = p_{off}^{int} \times (1 - p_{on}^{int}).$$

If both free α and β domains exist, α-β bonds may form according to $p_{on}^{eff}$ as at the t = 1 step. After association and dissociation reactions are completed for all non-empty unit cubes, the resulting new particles will diffuse as in step one. The program is terminated

when the desired number of steps is reached. The coordinates and compositions of the final particles are recorded for further analysis.

We note that 1) the ratio of intrinsic association and dissociation probabilities determines binding affinity of the two domains; 2) virtual molecules are dimensionless and therefore there is no upper limit of how many molecules can exist in one unit cube; 3) cooperativity is not considered here since all free α/β pairs are treated identically for association reactions, i.e. inter-complex binding and intra-complex binding share the same $p_{on}^{eff}$; 4) the association/dissociation reactions are segregated from the diffusion reactions and therefore the time steps in the simulation cannot be translated into real time literally; 5) the outcome of each trajectory is probabilistic.

**Simulations recapitulate theoretical calculations in dilute conditions**

To better understand the driving force for polymerization we developed a theoretical description of multivalent-multivalent interactions. This also allowed us to examine the validity of our simulations by comparing binding affinities calculated from the trajectories with theoretical predictions. In both cases, we examined how affinities scaled with valency of the interacting species. In developing the theoretical description we examine in this section a simple reaction:

$$AB + AB \leftrightarrow A_2B_2 \qquad (1)$$

and then expand to the general case in the following sections. The product of Reaction (1), $A_2B_2$, can be reasonably sampled to determine the binary equilibrium constant during

the simulation trajectory in dilute conditions. We consider virtual molecules with matched valences (A = $\alpha_n$ and B = $\beta_n$) and at equal concentrations ([A] = [B]). Although individual α-β interactions may be weak, multiple α-β binding events cooperatively give rise to strong binding for an A/B pair[3], which ensures that heterodimer (AB) will be the major species in dilute conditions (below the phase transition—see below). For simplicity, we assume 100% saturation of the domains in both the reactant (AB) and the product ($A_2B_2$) and therefore the equilibrium constant is determined entirely by the balance of only two entropic terms: 1) translational entropy loss due to assimilation of two complexes into one, and 2) configuration/microstate entropy gain due to the additional ways of matching the complementary domains in the large complex ($A_2B_2$) versus two smaller ones (AB + AB). The first term is insensitive to valency (n). However as shown below, the second term increases strongly with valency, thus providing a large entropic driving force for the reaction for large n. Phrased differently, the equilibrium constant of Reaction (1) increases sharply with valency.

In order to calculate the equilibrium constant change with valency, we assume 1) molecules A and B are dimensionless, 2) the linkers between domains are infinitely flexible, and therefore 3) each α domain has free access to each β domain, and vice versa. In an AB heterodimer, there are n α domains and n β domains and the number of unique ways to match domains α and β is n!, i.e.

$$M_{AB} = n! \qquad (2)$$

In $A_2B_2$, there are 2n α domains and 2n β domains and the number of unique ways to match domains α and β is (2n)!. However some of the (2n)! configurations result in two

separate AB hetero-dimers. There are two ways to form two AB hetero-dimers out of 2 A molecules and 2 B molecules and each AB dimer again has n! microstates. Therefore the portion of configurations resulting in two separate AB dimers is 2*n!*n!, i.e.

$$M_{A2B2} = (2 \times n)! - 2 \times n!^2 = (2 \times n)! - 2 \times M_{AB}^2 \quad (3)$$

in which $M_{AB}$ is from (2). The first term in the right hand side, (2xn)!, is the total microstates and the second term, $2xM_{AB}^2$ or $2xn!^2$, represents the microstates resulting in two separate AB complexes. The microstate gain ($M^{gain}$) for Reaction (1) is therefore:

$$M^{gain} = \frac{M_{A2B2}}{M_{AB}^2} = \frac{(2 \times n)! - 2 \times n!^2}{n!^2} \quad (4)$$

where $M_{A2B2}$ and $M_{AB}$ are the combinatorial microstates for $A_2B_2$ and AB, respectively. The values of $M^{gain}$ for Reaction (1) are 4, 18, 68, 250 and 922 for valencies of 2, 3, 4, 5 and 6, respectively. Since $M^{gain} \equiv \exp(\Delta S/R)$ (= $\exp(-\Delta G/RT)$ under the assumptions here):

$$\frac{M^{gain}(n = j)}{M^{gain}(n = 2)} = \frac{K_A(n = j)}{K_A(n = 2)} \quad (5)$$

in which j is an integer no fewer than 2, the equilibrium association constants of valencies 2, 3, 4, 5 and 6 are predicted to be 1-, 4.5-, 17-, 62.5- and 230.5-fold higher than that of valency 2, respectively.

To evaluate how well our simulations recapitulate the theoretical predictions, we carried out simulation experiments for valencies 2, 3, 4, 5 and 6 in dilute conditions, i.e. at concentrations where only low-order complexes (A, B, AB, $A_2B_2$, $AB_2$, $A_2B$ etc.) are

significantly populated (data not shown). For each valency, trajectories for nine different concentrations were calculated up to 2,000,000 steps. Different concentrations of each valency were achieved via changing the number of A and B molecules (#A and #B, respectively) in a constant volume. The volumes for valencies 2, 3 and 4 were 64,000 unit cubes. The volumes for valencies 5 and 6 were 128,000 and 512,000 unit cubes, respectively. #A (and #B) at the lowest concentration were 400, 240, 80, 42 and 42 for valencies 2, 3, 4, 5 and 6, respectively. #A (and #B) at the highest concentration were 800, 400, 120, 50 and 50 for valencies 2, 3, 4, 5 and 6, respectively. The time steps required for the systems to reach equilibrium, varied with valency. For instance, for the highest concentration of valency 2, the system started with 1,600 complexes (800 A and 800 B), reached 1115 complexes at t = 1,000 and fluctuated between 1,083 and 1,175 until the end of the simulation (t = 2,000,000); for the highest concentration of valency 6, the system started with 100 complexes (50 A and 50 B), reached 50 complexes at t = 156,400 and fluctuated between 45 and 53 until the end of the simulation (t = 2,000,000). The concentrations of heterodimer (<AB>) and tetramer (<$A_2B_2$>) in each simulation were obtained by averaging [AB] and [$A_2B_2$] over the reaction trajectories after the systems reached equilibrium, for example for valency 6, data points at t = 500,000, 500,200, 500,400, … 1,999,800, 2,000,000 (7501 data points in total) were used. For each valency the nine trajectories yielded nearly identical equilibrium association constants (data not shown) and the mean of the nine values was used for further analysis. In Supplementary Figure 22, the increases in equilibrium association constant for valencies 2-6 versus valency 2 obtained from simulation ($K_A^{sim}$) and from theory ($K_A^{the}$) are plotted against each other. The slope, intercept and $r^2$ of the best fit line are very

close to 1.0, 0.0 and 1.0, respectively, indicating that the simulations recapitulate theoretical predictions very well in dilute conditions.

**Microstates of a generic oligomer**

In order to extend our description of multivalent interactions to higher-order complexes, we first developed an analytical expression for the number of microstates in generic oligomers (which ultimately will be related to the entropic change upon polymerization as for the m = 2 case above). For systems composed of m A molecules and m B molecules (m > 1), there are m*n α domains and the same number of β domains. As above, we assume 100% saturation of the domains. There are (m*n)! unique ways to match these domains. As for $A_2B_2$, some configurations result in a single $A_mB_m$ complex and the rest result in more than one complex. Therefore $M_{AmBm}$, the microstates of $A_mB_m$, equals (m*n)! minus the microstates that result in more than one sub-complexes. For example when m = 3:

$$M_{A3B3} = (3\times n)! - 3^2 \times M_{AB} \times M_{A2B2} - 3! \times M_{AB}^3 \qquad (6)$$

where $M_{AB}$ and $M_{A2B2}$ are from (2) and (3), respectively. As for Equation (3), the first term on the right hand side of Equation (6), (3xn)!, represents the total microstates; the second term, $3^2 x M_{AB} x M_{A2B2}$, represents the microstates that result in an AB complex and an $A_2B_2$ complex; the third term, $3! x M_{AB}^3$, represents the microstates that result in three separate AB complexes. Since each term except the first in $M_{AmBm}$ represents a different way of creating smaller complexes whose stoichiometries add to m, the number of terms

in $M_{AmBm}$ corresponds precisely to the total number of ways of expressing m as a sum of positive integers, i.e. the integer partition function in number theory[4]. The number of terms in the partition function cannot be described analytically, but can be determined iteratively for small m or approximated by the Rademacher equation for large m (reference [4]). While there is also no analytical description for the partition of an integer, it can be determined iteratively as well[5]. The actual terms of $M_{AmBm}$ are closely related to the partition of m. For each term in $M_{AmBm}$, m is partitioned into j unique positive integers, i(1)…i(j), and the value of the term is:

$$(C_{m}^{i(1)})^2 \times M_{Ai(1)Bi(1)} \times (C_{m-i(1)}^{i(2)})^2 \times M_{Ai(2)Bi(2)} \times \ldots \times (C_{m-i(1)-i(2)-\ldots i(j-2)}^{i(j-1)})^2 \times M_{Ai(j-1)Bi(j-1)} \times (C_{m-i(1)-i(2)-\ldots i(j-1)}^{i(j)})^2 \times M_{Ai(j)Bi(j)} = \prod_{k=1}^{j} (C_{m-\sum_{h=1}^{k-1} i(h)}^{i(k)})^2 \times M_{Ai(k)Bi(k)} \quad (7)$$

in which C represents the combination operation. If m is partitioned into j positive integers, i(1)…i(j), among which there are x unique numbers of more than one copy and the copy numbers for the x unique numbers are a(1)…a(x), Equation 7 will be modified to eliminate multiple-counting of the same microstates, yielding:

$$\frac{\prod_{k=1}^{j} (C_{m-\sum_{h=1}^{k-1} i(h)}^{i(k)})^2 \times M_{Ai(k)Bi(k)}}{a(1)! \times a(2)! \times \ldots \times a(x-1)! \times a(x)!} = \frac{\prod_{k=1}^{j} (C_{m-\sum_{h=1}^{k-1} i(h)}^{i(k)})^2 \times M_{Ai(k)Bi(k)}}{\prod_{k=1}^{k=x} a(k)!} \quad (8).$$

In this way, $M_{AmBm}$ can be iteratively determined based on all partitions of m, Equations (7-8), and $M_{AB}$ ….$M_{A(m-1)B(m-1)}$.

To find a concise, approximate expression to replace the cumbersome iterative calculation of microstates of a generic oligomer, we evaluated the ratio of $M_{AmBm}$ to

(m*n)! (i.e. the fraction of total microstates available to mA+mB molecules that yield an intact $A_mB_m$ complex) as a function of valency (n) and complex size (m) (Supplementary Fig. 23). For a valency of 2, $M_{AmBm}/(m*2)!$ monotonically decreases from 0.67 to less than 0.2 as complex size increases from 2 to 20 (Supplementary Fig. 23a). In contrast, the values of $M_{AmBm}/(m*n)!$ are much larger for valencies > 2, and moreover they increase with complex size (Supplementary Fig. 23b). This indicates that there is a fundamental difference between divalency and higher order valencies in terms of the entropic driving force to form large assemblies (see below). Importantly, results in Supplementary Figure 23 indicate that for n > 2 and large m, the fraction of configurations resulting in more than one complex is negligible, and the microstates producing a single complex can be approximated by (m*n)!, i.e.

$$M_{AmBm} \approx (m \times n)! \qquad (9)$$

**Valency-dependence of critical concentration**

The concise expression for microstates in the $A_mB_m$ complex allows us to analyze the thermodynamic properties of generic multivalent-multivalent interactions and understand how they are affected by valency. We focus in the simple oligomerization of m AB heterodimers to form an $A_mB_m$ complex:

$$mAB \leftrightarrow A_mB_m \qquad (10)$$

For simplicity, we again assume 100% saturation of the domains in both the reactant (AB) and the product ($A_mB_m$), and therefore the equilibrium properties of the reaction are

determined entirely by the balance between two opposing entropic terms: the increase in microstates of the polymer relative to smaller species and the loss of translational/rotational entropy upon complex formation. As in Equations 4 and 5 above, and using Equation 9 for $M_{AmBm}$, the entropy change due to the increase in microstates for Reaction (10) is:

$$\frac{\Delta S^{configuration}}{k_B} = \ln(M_{AmBm}) - m \times \ln(M_{AB}) \cong \ln((m \times n)!) - m \times \ln(n!) \cong m \times n \times \ln(m) \quad (11)$$

where $k_B$ is the Boltzmann constant and the last step is obtained through Stirling's approximation. $\Delta S^{configuration}$ is a positive quantity that is linearly dependent on n and a concave function of m (but independent of protein concentration) (Supplementary Fig. 7a). This theoretical treatment of the combinatorial entropy gain is similar to that for polymers with self-associating stickers by Semenov and Rubinstein[2] and that for polymers with binary associating stickers by Semenov et al.[1] The entropy change due to loss of rotational and translational freedom upon complex formation for Reaction (10) is:

$$\frac{\Delta S^{diffusion}}{k_B} = (m-1) \times \Delta S \quad (12)$$

where $\Delta S$ is the entropy loss due to assimilation of one AB dimer into the complex, whose magnitude is inversely dependent on the concentration of AB dimer, [AB][6,7]. Therefore, $\Delta S^{diffusion}$ is a negative quantity with magnitude linearly dependent on m and inversely dependent on [AB][6,7] (but independent of valency).

The configurational entropy change, $\Delta S^{configuration}$, is insensitive to [AB] but linearly dependent on valency (Equation 11 and horizontal green lines in Supplementary Fig. 7b).

The magnitude of diffusional entropy change, $-\Delta S^{diffusion}$, decreases with increasing reactant concentrations, [AB], but is insensitive to valency (Equation 12 and red line in Supplementary Fig. 7b). For reactant concentrations where $\Delta S^{configuration} > -\Delta S^{diffusion}$ the reaction will produce large oligomers, whereas it will not when the magnitudes are reversed. The point where the two terms are equal represents a critical concentration, at which polymer begins to form (dotted circles in Supplementary Fig. 7b). Because $\Delta S^{configuration}$ is linearly proportional to valency n (Supplementary Figs. 7a and 7b), the critical concentration is inversely proportional to valency (Supplementary Fig. 7c), as observed in our simulations (Supplementary Figs. 6 and 22).

**Concluding considerations**

The theoretical analysis above is based on the assumption of ideal molecules. In real systems, imperfect linkers and the physical size of proteins prevent a domain from having free access to any complementing domains in the same complex. In addition, domain occupancy cannot be 100% as assumed here. These conditions will modify the theoretical treatment of configuration entropy change presented here. Nevertheless, this analysis presents a significant entropic driving force for creation of large oligomers, and shows how this can lead qualitatively to the switch-like phase transitions and their dependence on valency that are observed experimentally. Further modifications of this theory will be necessary to explain more features of the system and describe it quantitatively.

## Cyclization in the Sol-Gel Transition

Consider a generic two-component system capable of undergoing condensation polymerization with monomers of configuration $\mathbf{A}_f$ and $\mathbf{B}_g$, where $f$ and $g$ designate the valencies of **A** and **B** modules, respectively, and where **A** modules bind to **B** modules in a 1:1 fashion. Systems of this configuration have the propensity to polymerize and yield a distribution of polymer complexes of different sizes, as well as to undergo a sol-gel transition above a critical degree of saturation[1]. The critical condition for sol-gel transition in this system has been derived by Stockmayer[2] as

$$p_A p_B \geq \frac{1}{(f-1)(g-1)}, \tag{1}$$

where the degrees of saturation $p_A$ and $p_B$, estimated by module affinity and concentrations, denote the probability that a given **A** or **B** module, respectively, is in the bound state. This derivation of the condition for gel formation assumes ideal, Caley-tree binding configurations throughout the system, where each additional bond in any given complex is accompanied by an addition of a monomeric unit to the complex. Thus, intra-complex binding, or cyclization, is prohibited in this analysis. The possibility of cyclization cannot be ignored in biological systems, where the effects of avidity are prevalently observed.

To account for the possibility of cyclic binding in our systems, we define the degrees of cyclization as

$\sigma_A \equiv$ (cyclic **A** modules)/(cyclic **A** modules + acyclic **A** modules); and

$\sigma_B \equiv$ (cyclic **B** modules)/(cyclic **B** modules + acyclic **B** modules), (2)

where $\sigma_A$ and $\sigma_B$ are the probability that a bound **A** or **B** module, respectively, participates in intra-complex binding. Allowing for the possibility of cyclization in our system, we redefine the critical condition in eq. (1) as

$$(1-\sigma_A)(1-\sigma_B)p_A p_B \geq \frac{1}{(f-1)(g-1)}, \quad (3)$$

where we note that only acyclic (ideal) bonds will contribute in the approach to the sol-gel transition point as $p_A$ and $p_B$ increase. Consider the simplifying case of equal module concentrations, $[\mathbf{A}]_{\text{total}} = [\mathbf{B}]_{\text{total}}$. For this case, $p_A = p_B \equiv p$; $\sigma_A = \sigma_B \equiv \sigma$; and eq. (3) reduces to

$$[(1-\sigma)p]^2 \geq \frac{1}{(f-1)(g-1)}. \quad (4)$$

Eq. (4) indicates that a system above the gel point with degree of saturation $p$ can tolerate some level non-ideal behavior in the form of cyclic bonds without being pushed below the sol-gel transition critical point. To determine this maximally tolerable degree of cyclization $\sigma_{tol}$, we equate the two sides of eq. (4) and solve for $\sigma$ to obtain

$$\sigma_{tol} = 1 - \frac{1}{p}\sqrt{\frac{1}{(f-1)(g-1)}}. \quad (5)$$

Eq. (5) indicates that systems at higher degrees of saturation or with higher valencies can tolerate higher degrees of cyclization before being pushed below the sol-gel transition critical point. The $SH3_5$+$PRM_5$ and $SH3_5$+$PRM(N\text{-}WASP)_8$ systems at degrees of saturation 81% (based on module concentration of 8 mM in the droplet phase and $K_D$ = 356 $\mu$M for $SH3_1$ + $PRM_1$) and 90% (based on module concentration of 13.1 mM in the droplet phase and $K_D$ = 147 $\mu$M for $SH3_5$ + $PRM(N\text{-}WASP)_1$) could tolerate up to 69% and 79%, respectively, of all bonds in the system being cyclic and still remain above the sol-gel transition critical point. In the absence of cyclization ($\sigma$ = 0), eq. (4) predicts $p_{critical}$ of $(1/(4x4))^{0.5} = 0.25$ and $(1/(5x7))^{0.5} = 0.17$, for the $SH3_5$+$PRM_5$ and $SH3_5$+$PRM(N\text{-}WASP)_8$ systems, respectively.

This analysis assumes that $p_A$ and $p_B$ can be determined based only on the known binding affinity of **A** for **B** modules and the module concentrations, treating each module as a monovalent molecule. It is assumed that the system does not exhibit any degree of negative cooperativity, where the values of $p_A$ and $p_B$ determined in this fashion would overestimate the actual degrees of saturation. Positive cooperativity (avidity) would result in underestimation of the actual degrees of saturation.

## References

1. Flory, P. J. Principles of Polymer Chemistry. (Cornell University Press, 1953).

2. Stockmayer, W. H. Molecular Distribution in Condensation Polymers. J. Polymer Sci. 9, 69-71, (1952).

# Experimental Procedures

Supplementary Table 1. Constructs used in this study

| Construct name | Protein sequence | Organism/Note |
|---|---|---|
| $PRM_1$ | Acetyl-CGGSWGGSKKKKTAPTPPKRS | Derived from ABL1 (residues 606~618) |
| $PRM_2$ | GHMKGGSWGGSKKKKTAPTPPKRSGGSGGSGGSGGSKKKKTAPTPPKRSGGSGSENLYFQ | |
| $PRM_3$ | GHMKGGSWGGS(KKKKTAPTPPKRSGGSGGSGGSGGS)$_2$KKKKTAPTPPKRSGGSGSENLYFQ | |
| $PRM_4$ | GHMKGGSWGGS(KKKKTAPTPPKRSGGSGGSGGSGGS)$_3$KKKKTAPTPPKRSGGSGSENLYFQ | |
| $PRM_5$ | GHMKGGSWGGS(KKKKTAPTPPKRSGGSGGSGGSGGS)$_4$KKKKTAPTPPKRSGGSGSENLYFQ | |
| $PRM(H)_1$ | Acetyl-CGGSWGGSKKTPPPVPPRTTSK | Derived from Disks Large-associated protein 2 (residues 575~588) |
| $PRM(H)_2$ | GHMKGGSWGGSKKTPPPVPPRTTSKGGSGGSGGSGGSKKTPPPVPPRTTSKGGSGSENLYFQ | |
| $PRM(H)_3$ | GHMKGGSWGGS(KKTPPPVPPRTTSKGGSGGSGGSGGS)$_2$KKTPPPVPPRTTSKGGSGSENLYFQ | |
| $PRM(H)_4$ | GHMKGGSWGGS(KKTPPPVPPRTTSKGGSGGSGGSGGS)$_3$KKTPPPVPPRTTSKGGSGSENLYFQ | |
| $PRM(H)_5$ | GHMKGGSWGGS(KKTPPPVPPRTTSKGGSGGSGGSGGS)$_4$KKTPPPVPPRTTSKGGSGSKNLYFQ | |
| $PRM(N\text{-}WASP)_8$ | (((PPPPPPARGR-peg)$_2$-K-peg)$_2$-K-peg)$_2$-KGGWGC | Derived from human N-WASP (residues 298~309) |
| Robo $PRM_5$ | GHMKGGSWGGS(DDLPPPPVPPGGKGGSGGSGGSGGS)$_4$DDLPPPPVPPGGKGGSGSKNLYFQ | Derived from fly Robo (residue 1478~1487) |
| $SH3_1$ | GHMDLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPL | Derived from the human Nck1 second SH3 domain (residues 106~168, C139S) |
| $SH3_2$ | GHMDLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPLASGAGGSEGGGSEGGTSGATDLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPL | |
| $SH3_3$ | GHM(DLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPLASGAGGSEGGGSEGGTSGAT)$_2$DLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPL | |
| $SH3_4$ | GHMDLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPLASGAGGSEGGGSEGGTSGATHM(DLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPLASGAGGSEGGGSEGGTSGAT)$_2$DLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPL | |
| $SH3_5$ | GHM(DLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPLASGAGGSEGGGSEGGTSGAT)$_2$HM(DLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPLASGAGGSEGGGSEGGTSGAT)$_2$DLNMPAYVKFNYMAEREDELSLIKGTKVIVMEKSSDGWWRGSYNGQVGWFPSNYVTEEGDSPL | |
| srGAP1 $SH3_5$ | GHM(GSPEFAIAKFDYVGRSARELSFKKGASLLLYHRASEDWWEGRHNGIDGLVPHQYIVVASGAGGSEGGGSEGGTSGAT)$_4$GSPEFAIAKFDYVGRSARELSFKKGASLLLYHRASEDWWEGRHNGIDGLVPHQYIVVGGSGSKNL | Derived from mouse Slit-Robo Rho GTPase-activating protein 1 (residue 377~428) |

| | | |
|---|---|---|
| | YFQ | |
| N-WASP (GBD-P-VCA) | 193-501 | Rat |
| N-WASPΔ (B-Crib-P) | 183-239 + 273-396 | Rat |
| N-WASP B-P-VCA | 183-193 + 273-501 | Rat |
| N-WASPΔP (GBD-VCA) | 193-272-$(GGS)_2$- 396-501 | Rat |
| Nck | 1-377 | Human/C139S, C232A, C266S, C340S |
| Nephrin | 1174-1223 | Human/Y1183F, Y1210F |
| PTB | 1-531 | Human |
| | | |
| **RNA name** | **RNA sequence** | **Note** |
| PTB_ligand | $(UCUCUAAAAA)_5$ | Synthesized by Integrated DNA Technologies, Inc., Iowa |

**Protein expression and purification**

Most proteins (except for PTB) were expressed in *E. coli* strain BL21(DE3)T1$^{R}$ (Sigma-Aldrich). PTB was expressed in BL21-CodonPlus RP cells (Stratagene). Bacteria were grown at 37 ºC and induced at $OD_{600}$ ~0.8 with 1 mM IPTG at either 20 ºC (for SH3 proteins and N-WASP derivatives) for 16 hours or 18 ºC (for PRM and PTB proteins) for 16 hours. Cells were lysed by either sonication or homogenization (EmulsiFlex-C5, Avestin) and lysates were cleared by centrifugation at 50,000 g for 40 minutes at 4 °C. U-[$^{15}$N] labeled $SH3_1$ was expressed in M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source.

*PRM proteins (n=2-5) and PRM(H)$_5$ and Robo PRM$_5$*

PRM proteins have an N terminal Tev-cleavable MBP tag to drive expression and a C terminal Tev-cleavable $His_6$ tag for affinity purification. Proteins were first isolated through a Ni-NTA column (Novagen) and then loaded onto a Mono Q column. The proteins were collected in the flow through, concentrated, and cleaved with Tev protease. Cleaved products were further purified using a Mono S column (GE Healthcare), followed by a Superdex 75 column equilibrated in 150KMEI (150 mM KCl, 1 mM $MgCl_2$, 1 mM EGTA and 10 mM imidazole, pH 7). All assays were performed in 150KMEI buffer, except for the RNA experiments, which were performed in 20 mM Tris pH 7.5, 170 mM NaCl, 30 mM KCl, 1 mM $MgCl_2$, 1 mM ATP, 20 mM creatine phosphate, 20 mM DTT.

*SH3$_5$ and srGAP1 SH3$_5$*

$SH3_5$ proteins have an N-terminal Tev-cleavable MBP tag and a C-terminal Tev-cleavable $His_6$ tag for tandem affinity purification. Proteins were sequentially isolated on Ni-NTA and then amylose affinity columns (New England Biolabs). Eluate was concentrated and cleaved using Tev protease. Cleaved products were further purified using a Mono Q column (GE Healthcare), followed by a Superdex 200 column (GE Healthcare) equilibrated in 150KMEI.

*PTB*

To remove bacterial RNAs from PTB, the lysis buffer contained 1.5 M NaCl in 20 mM Tris pH 8.0, 20 mM Imidazole and 5% glycerol. $His_6$-PTB was purified by a Ni-NTA column, followed by a Superdex 200 column.

*N-WASP Constructs (N-WASP GBD-P-VCA, N-WASPΔ, N-WASPΔP, N-WASP BPVCA)*

All N-WASP constructs were expressed as $His_6$-tagged proteins. They were purified over Ni-NTA agarose, Mono Q/Source15 Q, Mono S/Source15 S and Superdex 75 or Superdex 200. TEV cleavage was done either before or after the anion exchange columns. All purified proteins were exchanged into 150KMEI before use in assays.

*Triply phosphorylated nephrin peptide*

Nephrin peptide was fused to a Tev-cleavable MBP tag and a PreScission protease-cleavable $His_6$ tag at the N and C termini, respectively. Ni-affinity purified protein was subject to Tev and PreScission protease cleavage. The cleaved product was further purified by anion exchange and size-exclusion chromatography. Concentrated nephrin peptide (> 1.0 mM) was phosphorylated using 8 $\mu$M Fyn tyrosine kinase (Invitrogen) in

50 mM HEPES (pH = 7.5), 10 mM DTT, 100 mM ATP and 20 mM $MgSO_4$ at 30 °C for > 24 hours. Phosphorylated peptide was isolated by anion exchange chromatography and exchanged into 150KMEI buffer via size-exclusion chromatography. More than 95% of the peptide was triply phosphorylated as determined by mass spectrometry.

*Oregon Green (OG) labeled $SH3_4$ and $PRM_4$*

To generate OG-labeled proteins, a cysteine was inserted immediately following the first glycine of the $SH3_4$ protein; the first lysine in $PRM_4$ proteins was replaced with a cysteine. Purification protocols for the cysteine containing proteins were the same as for wild type proteins. Labeling reactions were carried out in 25 mM Hepes (pH 7.0), 100 mM NaCl and 1 mM EDTA buffer in the presence of a 5- to 10-fold molar excess of Oregon Green maleimide (Invitrogen) at 4 °C overnight. Reactions were quenched with a large excess of β-mercaptoethanol. To remove free dye, the reaction mixture was purified over a HighTrap 1 ml Q XL column (for $SH3_4$) or a HighTrap 1 ml SP XL column (for $PRM_4$) (GE healthcare). Fractions containing labeled protein were dialyzed against 150KMEI buffer. Labeling efficiency was > 90 %, as determined spectroscopically by absorbance at 280 and 492 nm.

*$PRM(N\text{-}WASP)_8$*

The primary sequence of $PRM(N\text{-}WASP)_8$ is $(((PPPPPPARGR\text{-}peg)_2\text{-}K\text{-}peg)_2\text{-}K\text{-}peg)_2$-KGGWGC. The core is a second generation poly-lysine dendrimer conjugated to $PEG_2$-SUC, to provide a flexible linker between the core and PRM(N-WASP) peptides, through a one step synthesis from the precursor, N-(Fmoc-8-amino-3,6-dioxa-octyl) succinamic acid (Anaspec).

*Other Materials*

Nck was isolated as described by Padrick et al.[8] and all SH3 proteins were generated using similar protocols. $PRM_1$, $PRM(H)_1$, $PRM(N\text{-}WASP)_1$ and doubly phosphorylated human nephrin peptide were synthesized at University of Texas Southwestern Medical Center's Protein Chemistry Core. An initial sample of doubly phosphorylated nephrin peptide was a kind gift from Dr. Tony Pawson (Samuel Lunenfeld Research Institute). Bovine Arp2/3 complex rabbit skeletal muscle actin and pyrene actin[9], and rhodamine-actin were purified as described previously[10,11,12].

**NMR titration and data analysis**

The $^{1}H/^{15}N$ heteronuclear single quantum coherence (HSQC) spectra of U-[$^{15}N$] labeled $SH3_1$ in the presence of increasing concentration of PRM ligands were collected on a Varian 600 MHz spectrometer at 25 °C. The initial concentration of $SH3_1$ was 100 $\mu$M and its final concentration remained above 80 $\mu$M at the end point of the titration. HSQC spectra were processed and analyzed using NMRPipe[13] and NMRView[14], respectively. The binding of $SH3_1$ to PRM peptides was in the intermediate to fast exchange regime. Therefore chemical shifts of eleven or more resonances (11 for $PRM(N\text{-}WASP)_1$, 12 for $PRM_1$ and 16 for $PRM(H)_1$) were fit to a single site binding model to yield dissociation constants.

## Supplemental Movie Legend

Movie S1. **A movie showing large polymer formation.** Forty-five $\alpha_4$ and 45 $\beta_4$ react in 10 x 10 x10 unit cubes. Symbol sizes are proportional to complex sizes. Symbols are colored by the parameter $X = (\#\beta_4 - \#\alpha_4)/(\#\beta_4 + \#\alpha_4)$ to indicate relative stoichiometry in each complex. Red dots (X = 1.0) correspond to $\beta_4$ only, blue dots (X = -1.0) correspond to $\alpha_4$ only and complexes are colored in between (e.g. complexes with equal numbers of $\beta_4$ and $\alpha_4$ are green).

## Supplemental References


1. Semenov, A., Charlot, A., Auzely-Velty, R. & Rinaudo, M. Rheological properties of binary associating polymers. Rheologica Acta 46, 541-568, (2007).
2. Semenov, A. N. & Rubinstein, M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. Macromolecules 31, 1373-1385, (1998).
3. Jencks, W. P. On the attribution and additivity of binding energies. Proc Natl Acad Sci U S A 78, 4046-4050, (1981).
4. Andrews, G. E. & Eriksson, K. Integer Partitions (Paperback). (2004).
5. Zoghbi, A. & Stojmenovic, I. Fast Algorithms for Generating Integer Partitions. Intern. J. Computer Math. 70, 319-332, (1998).
6. Flory, P. J. Thermodynamics of High Polymer Solutions. Journal of Chemical Physics 9, 660, (1941).
7. Huggins, M. L. Solutions of Long Chain Compounds. Journal of Chemical Physics 9, 440, (1941).
8. Padrick, S. B. *et al*. Hierarchical regulation of WASP/WAVE proteins. Mol Cell 32, 426-438, (2008).
9. Leung, D. W., Morgan, D. M. & Rosen, M. K. Biochemical properties and inhibitors of (N-)WASP. Methods in Enzymology, Vol 406, Regulators and Effectors of Small Gtpases: Rho Family 406, 281-296, (2006).
10. Cooper, J. A. & Pollard, T. D. Methods to measure actin polymerization. Methods Enzymol 85 Pt B, 182-210, (1982).

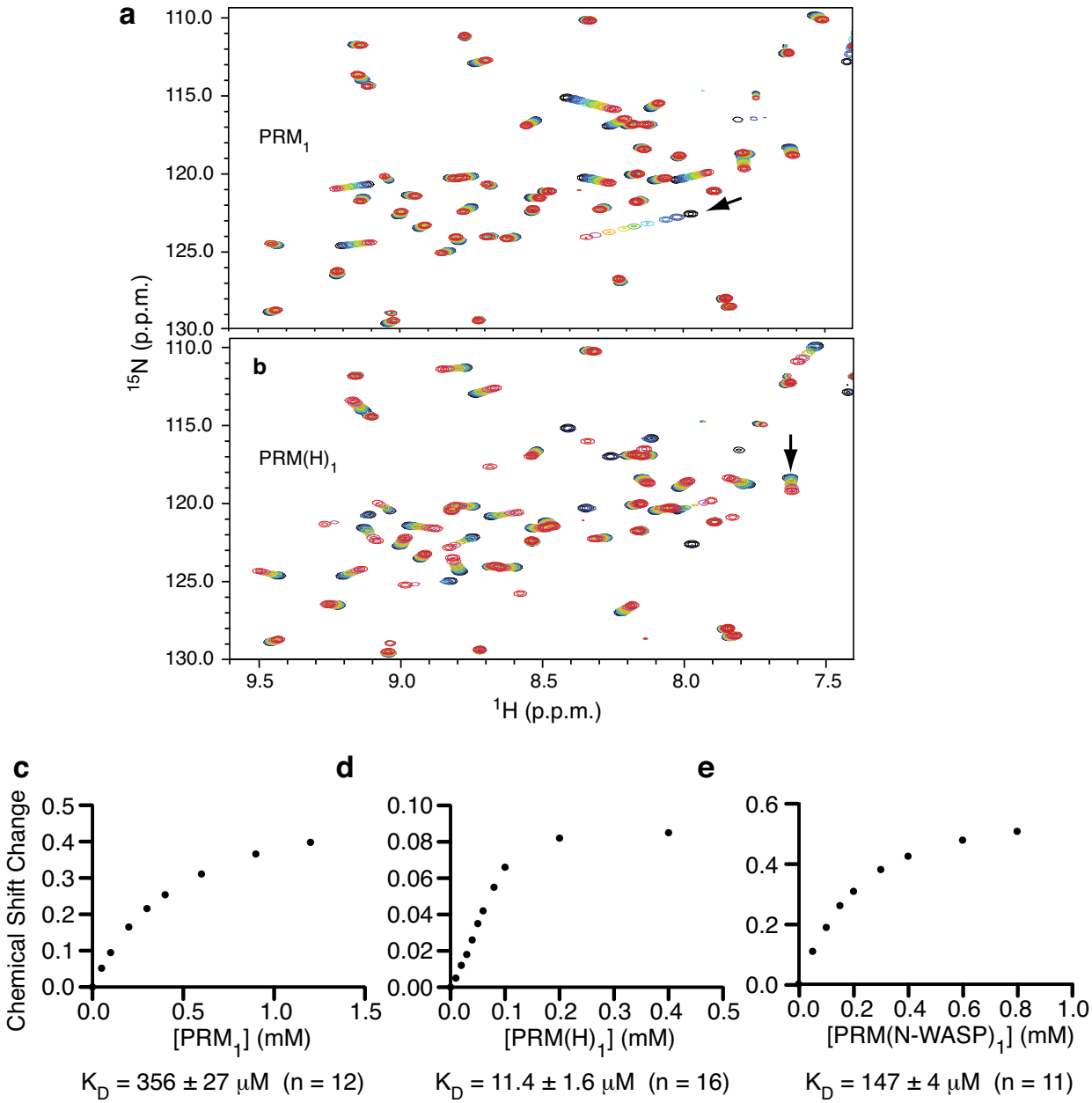
11. Higgs, H. N., Blanchoin, L. & Pollard, T. D. Influence of the C terminus of Wiskott-Aldrich syndrome protein (WASp) and the Arp2/3 complex on actin polymerization. Biochemistry 38, 15212-15222, (1999).

12. Otterbein, L. R., Graceffa, P. & Dominguez, R. The crystal structure of uncomplexed actin in the ADP state. Science 293, 708-711, (2001).

13. Delaglio, F. *et al*. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6, 277-293, (1995).

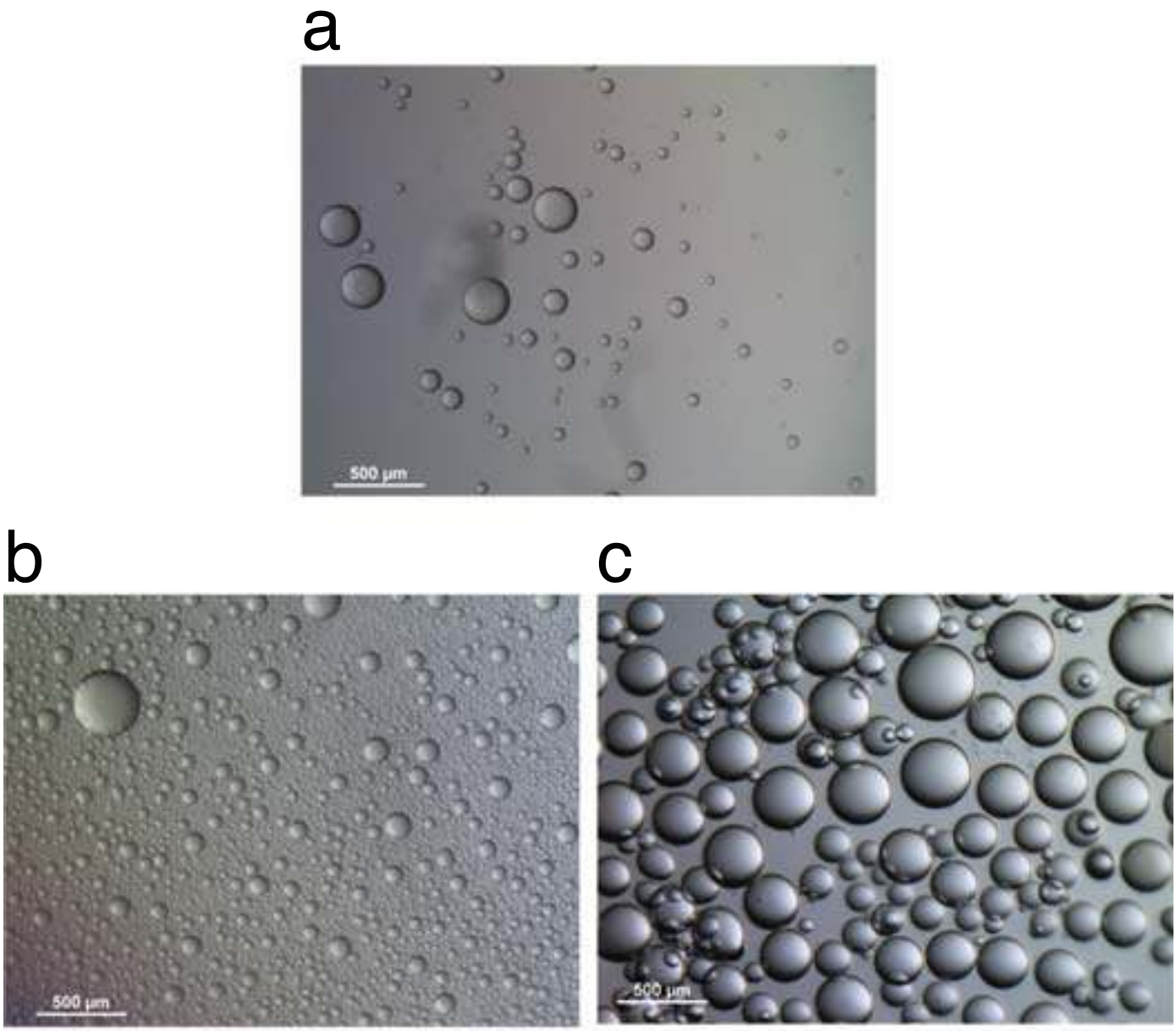
14. Johnson, B. A. & Blevins, R. A. Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data. Journal of Biomolecular Nmr 4, 603-614, (1994).

15. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Cryst. 25, 495-503, (1992).

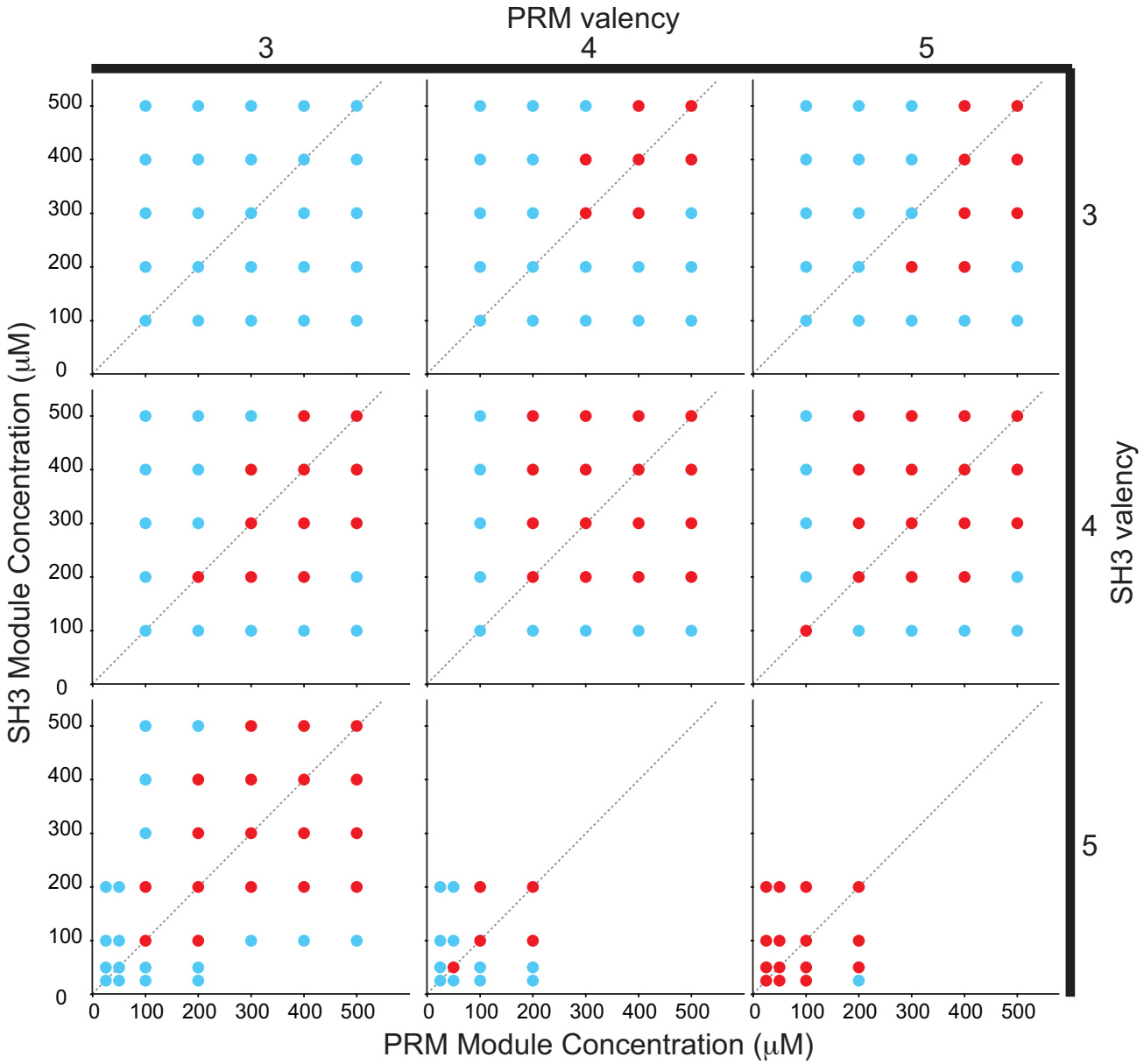



**Supplementary Figure 1. NMR titration to measure binding affinity of $SH3_1$ for $PRM_1$, $PRM(H)_1$ and $PRM(N\text{-}WASP)_1$.** Overlaid $^1H$-$^{15}N$ HSQC spectra of 100 $\mu$M $SH3_1$ in the presence of increasing concentrations of $PRM_1$ (a) or $PRM(H)_1$ (b). See Supplementary Table 1 for sequences of all SH3 ligands. (a) PRM concentrations are 0, 50, 100, 200, 300, 400, 600, 900 and 1200 $\mu$M from the black spectrum to the red. A binding curve derived from the resonance indicated by a black arrow in (a) is shown in (c). Binding isotherms derived from 12 resonances in the $PRM_1$ titration were fit separately to yield 12 dissociation constants. The mean $K_D$ and standard deviation are reported in (c). (b) $PRM(H)_1$ concentrations are 0, 10, 20, 30, 40, 50, 60, 100, and 400 $\mu$M from the black spectrum to the red. A representative binding curve derived from the resonance pointed by a black arrow in (b) is shown in (d) with two more data points ($PRM(H)_1$ = 80 and 200 $\mu$M). The affinity was determined by fitting 16 binding curves. The mean and standard deviation of the 16 calculated dissociation constants are reported in (d). Similar titration experiments were carried out for the binding reaction of $SH3_1$ with a proline-rich motif derived from N-WASP (spectra not shown). A representative binding curve is shown (e). The mean and standard deviation of the 11 calculated dissociation constants are reported in (e).

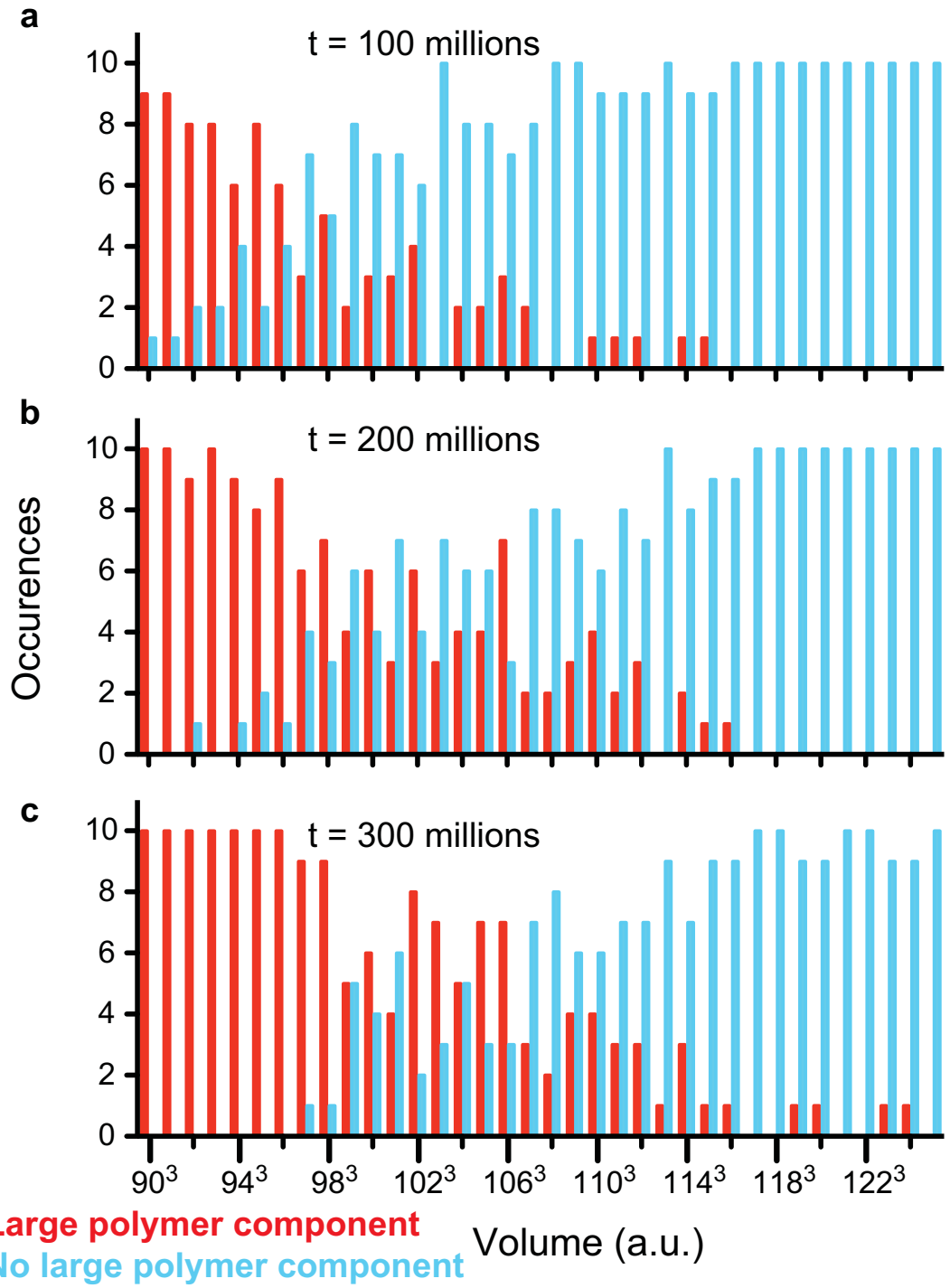
**Supplementary Figure 2. A wide range of droplet sizes is formed by the phase transition.** Liquid droplets observed by differential white light microscopy shortly after 300 $\mu$M of $SH3_4$ and 300 $\mu$M $PRM_4$ (module concentrations) are mixed.



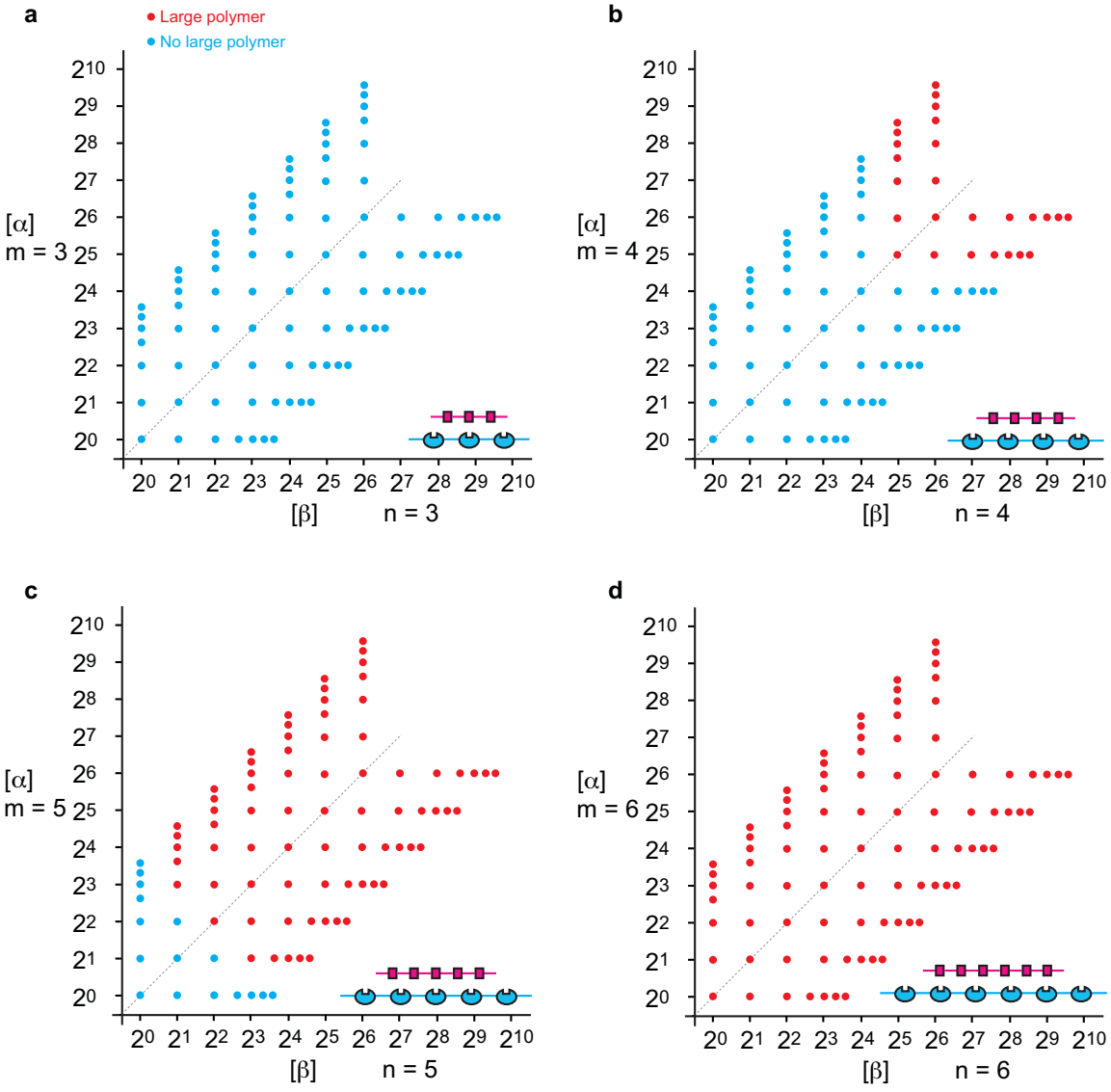

**Supplementary Figure 3. Liquid-liquid demixing can be observed in other multi-valent systems.** (a) Bright-field images of droplets formed from (a) 40 µM srGAP1 $SH3_5$ plus 40 µM Robo $PRM_5$ (module concentrations of 200 $\mu$M each), and (b), (c) 136 µM PTB mixed with 80 µM RNA ligand (see Supplementary Table 1). Images in (b) and (c) were taken immediately after PTB and RNA were mixed and 2 days later, respectively, illustrating coalescence of droplets over time. Experiments in (a) performed in 150KMEI buffer. Experiments in (b) and (c) performed in 20 mM Tris pH 7.5, 170 mM NaCl, 30 mM KCl, 1 mM $MgCl_2$, 1 mM ATP, 20 mM creatine phosphate, 20 mM DTT. Homology models of the srGAP1 SH3 + Robo PRM complex (used here) and Nck1 SH3(2) + ABL1 PRM complex (used elsewhere in this work) have only 25 % identity of surface exposed residues.



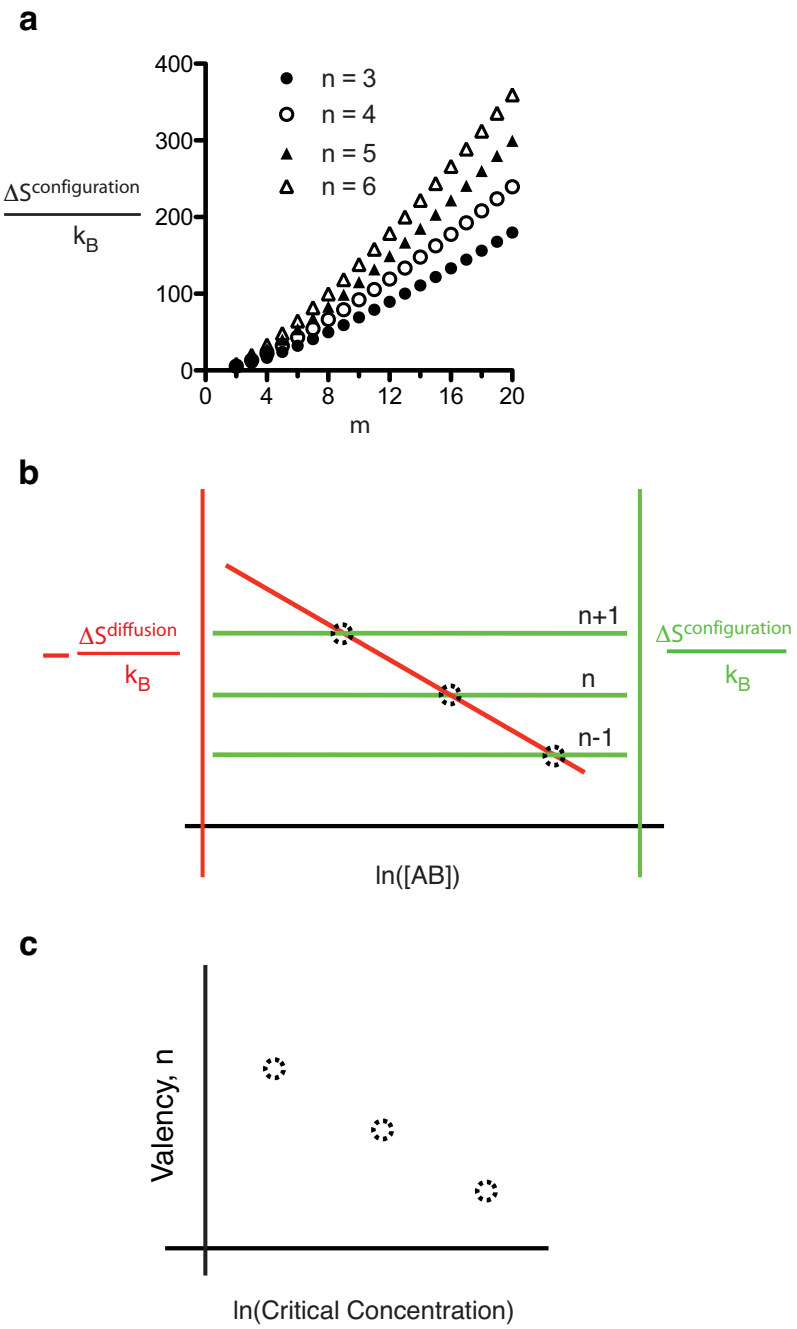

**Supplementary Figure 4. Increasing valency decreases the critical concentration.** Phase diagrams of multivalent SH3 and PRM proteins.



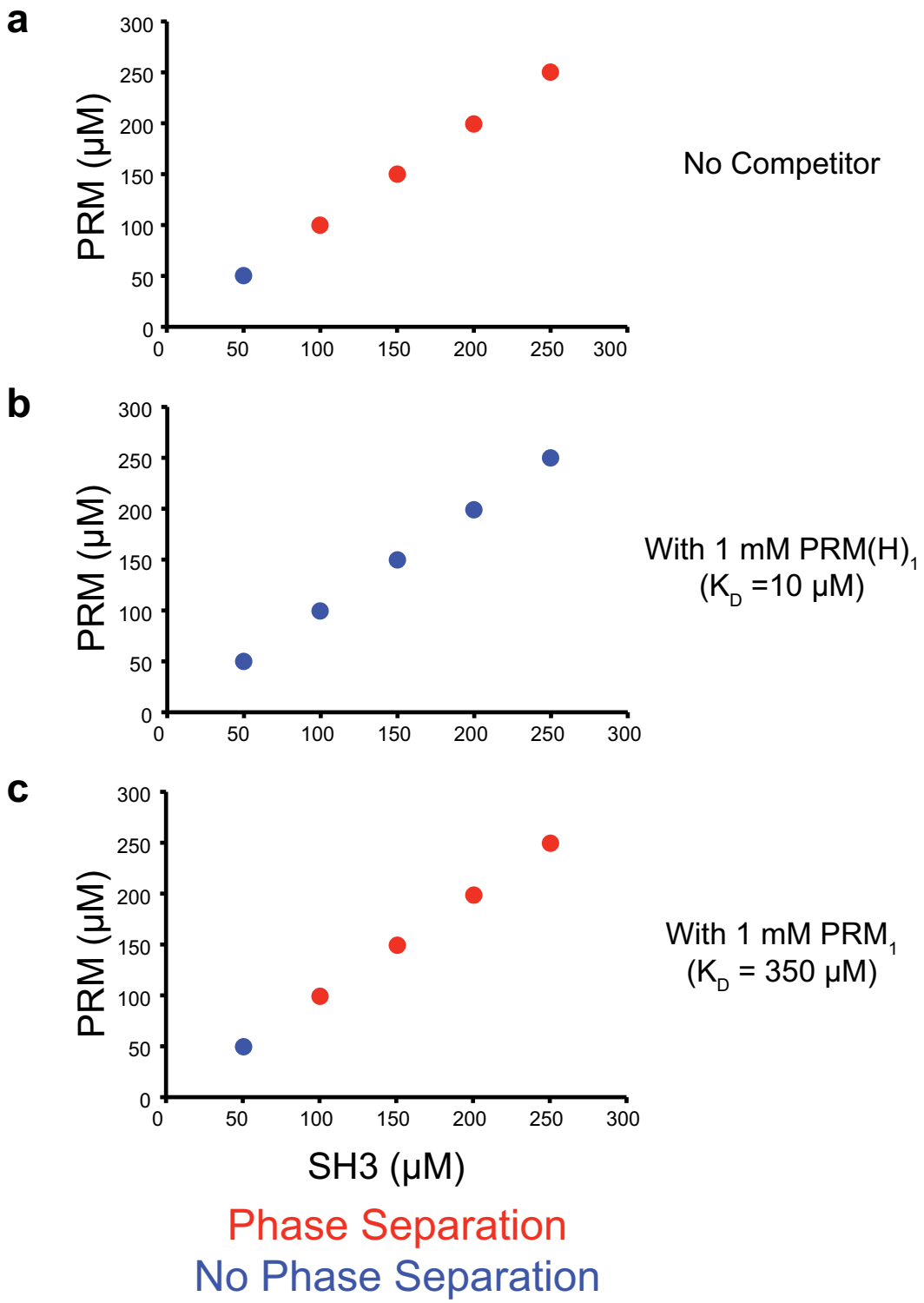

**Supplementary Figure 5. A narrow concentration range for probabilistic formation of large polymers.** Forty $\alpha_6$ and 40 $\beta_6$ react in varying volumes ($i^3$ unit cubes, i = 90 ~ 125). For each volume (i.e. each concentration), 10 independent, 300 million step reaction trajectories were calculated. In each trajectory we assessed whether a large complex, defined as containing 20 or more each of $\alpha_6$ and $\beta_6$, had formed. The appearance of large complex (red) or not (cyan) in the 10 trajectories is summarized at 100 (a), 200 (b), and 300 (c) million time steps. At high concentrations, large complex forms in all trajectories (left sides of (a-c)). At low concentrations, a large complex never forms (right sides of (a-c)). At intermediate concentrations large complex forms in some trajectories but not others. The concentration difference between the two deterministic regimes is ~2 fold ($117^3/91^3 \approx 2.1$, (b)), suggesting a narrow critical concentration regime under these conditions.



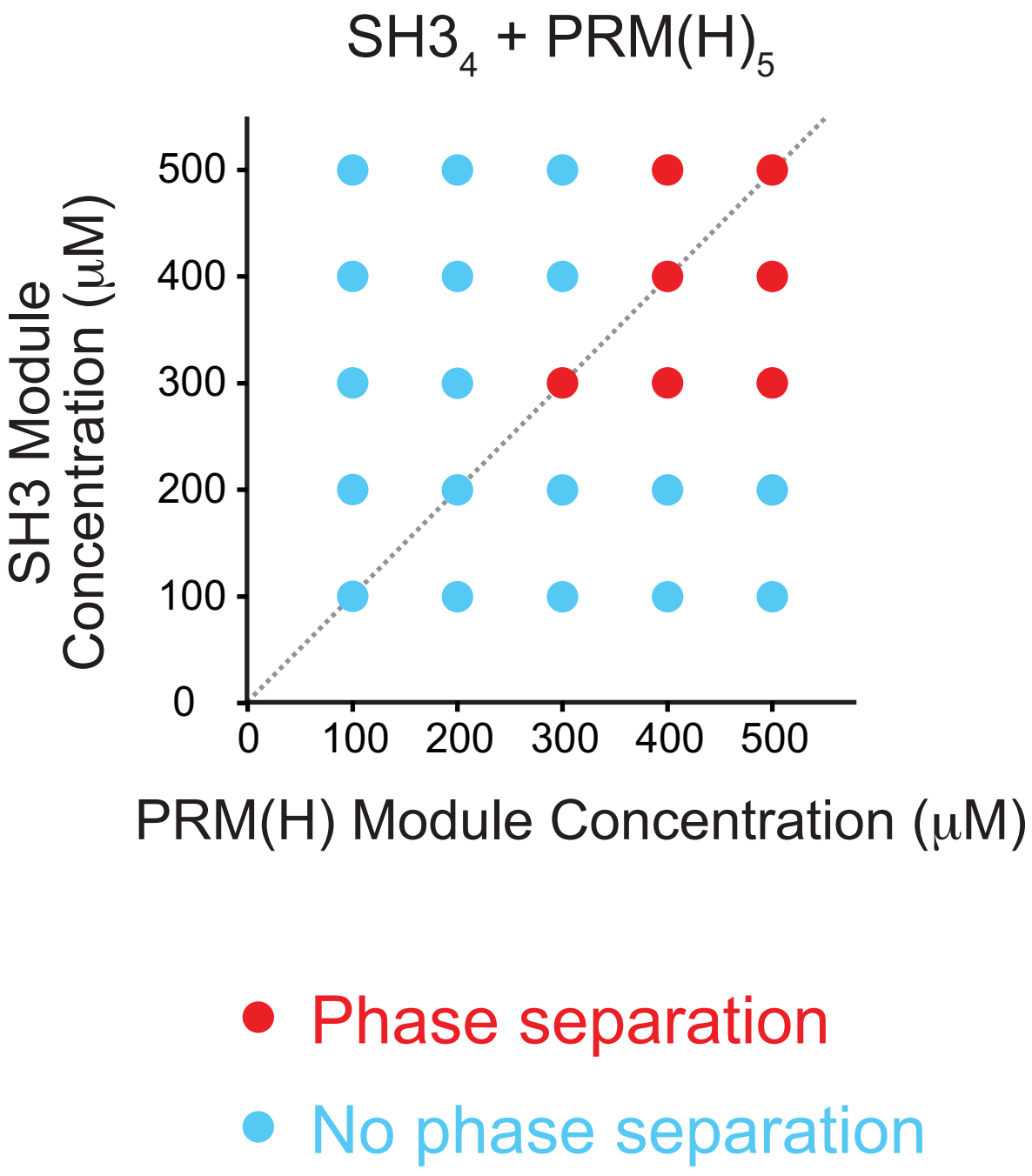

**Supplementary Figure 6. Simulated phase diagrams predict valency dependent phase transition.** A 30,000,000 time step trajectory was calculated for three hundred α domains in molecules of varying valencies (3-6 in a-d) reacting with 300, 600, 1200, 1800, 2400, 3000, or 3600 β domains in molecules of matched valencies in a spaces containing 15,625-1,000,000 unit cubes in 2-fold expansions. Large polymers, defined as ≥ 50 % of α domains in a single complex, formed at the end of some trajectories (red dots) but not at the others (blue dots). Results were plotted on and under the diagonal line in (a-d). Due to the symmetry of virtual α and virtual β in silico, mirror images of the abovementioned results were generated and plotted above the diagonal line in (a-d) to represent likely results for 300 β domains reacting with 600~3600 α domains.



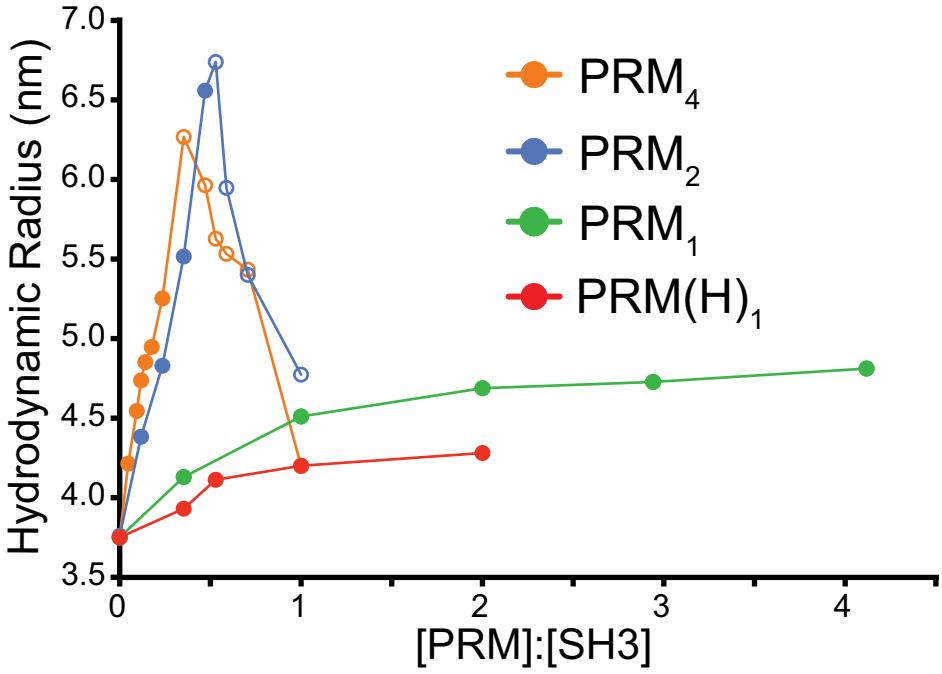

**Supplementary Figure 7. Theoretical predictions of the valency-dependence of critical concentration.** The reaction considered is m AB heterodimers oligomerizing into a single intact complex, $A_mB_m$, in which A is composed of n α domains and B is composed of n β domains. (a) Configurational entropy change of the reaction for n = 3-6 and m = 2-20 is calculated based on Equation (11). (b) Valency-dependence of the phase transition. Configuration entropy change of a specific reaction (green curves) is an intrinsic characteristic of the molecules involved. It is linearly dependent on valency of the molecules but independent of molecule concentration, [AB]. The diffusional entropy change, however, is independent of valency but inversely dependent on [AB]. The red and green curves cross at the critical concentration for oligomer formation for the corresponding reaction (m, n). With increasing valency (n), the critical concentration decreases (dotted circles). (c) The linear relationship between valency and critical concentration as predicted from (b).



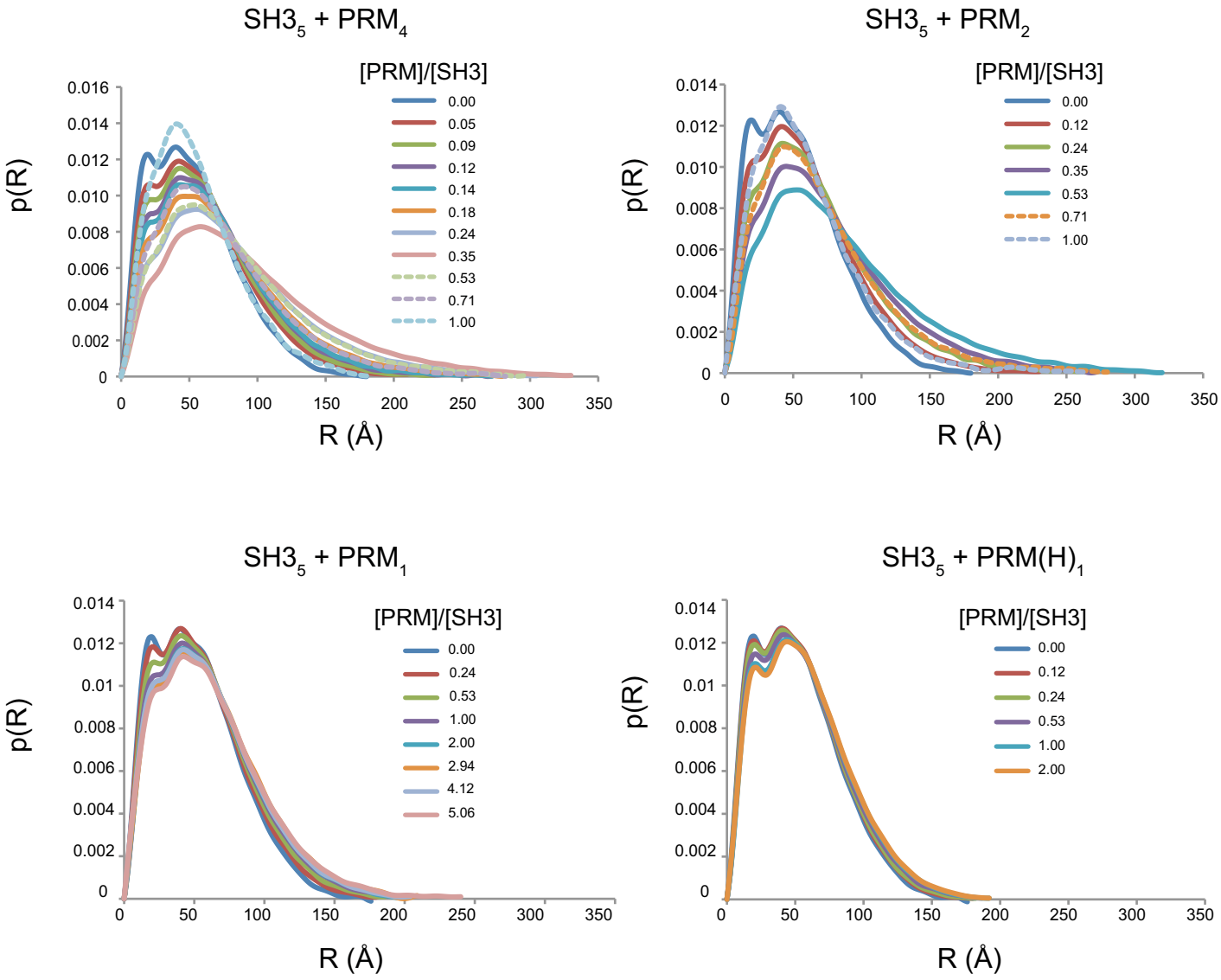

**Supplementary Figure 8. A high affinity monovalent competitor can inhibit droplet formation.** Phase diagrams for the $SH3_5$ + $PRM(N\text{-}WASP)_8$ system a. alone, b. in the presence of 1 mM high affinity monovalent competitor, $PRM(H)_1$, and c. in the presence of 1 mM low affinity monovalent competitor, $PRM_1$. Axes indicate module concentrations. The high affinity competitor can inhibit droplet formation (even though its pentavalent version, $PRM(H)_5$, can induce droplets, see Supplementary Figure 9), while the low affinity competitor cannot.



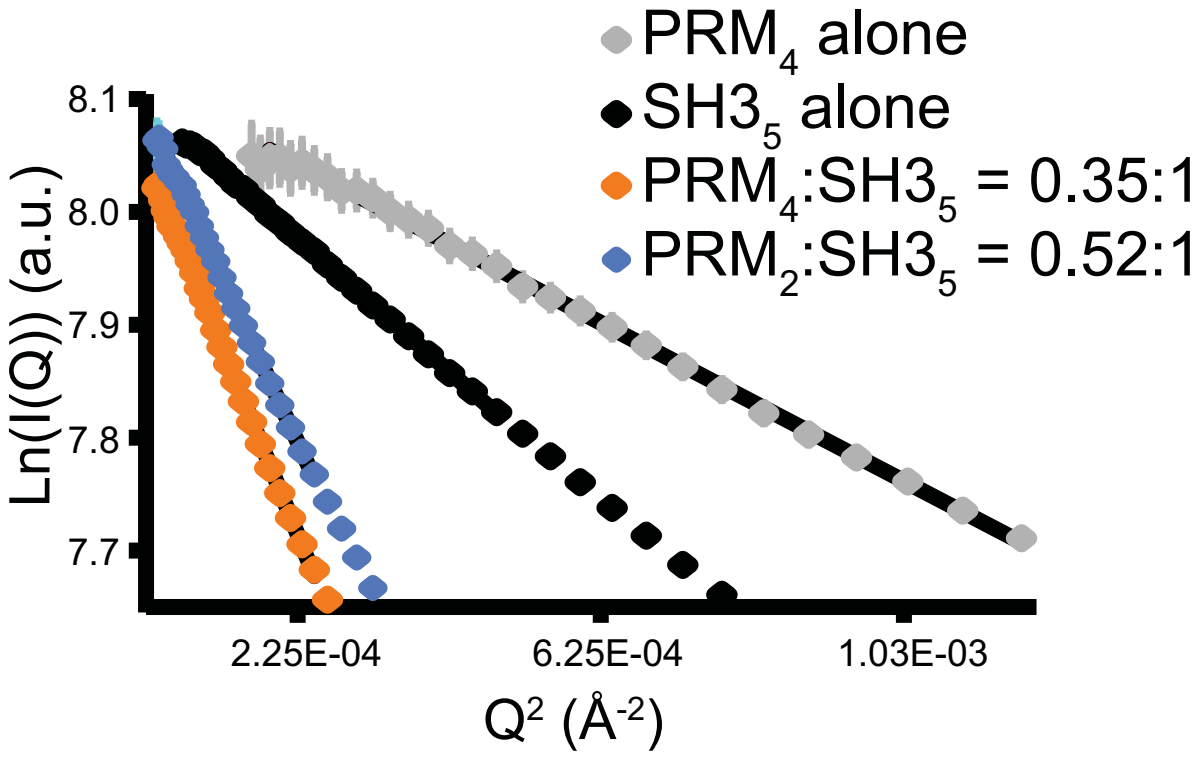

**Supplementary Figure 9. Multivalent PRM(H) can induce phase transitions.** Phase diagram for the $SH3_4$ + $PRM(H)_5$ system.



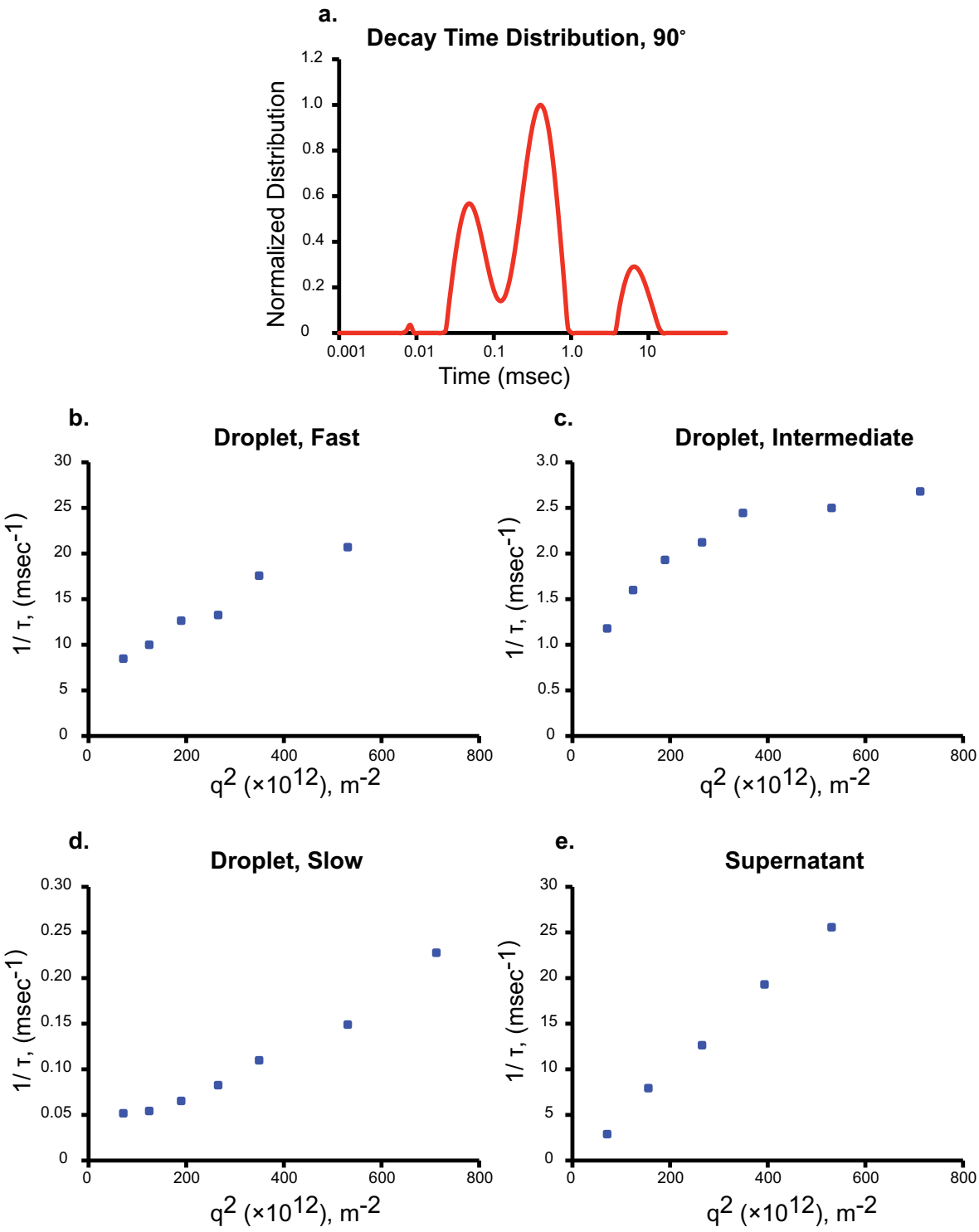

**Supplementary Figure 10. Hydrodynamic radius values determined from DLS data collected during titrations of PRM proteins into $SH3_5$.** Closed circles indicate the absence of phase separation; open circles indicate data collected on supernatant phase, separated from droplets by centrifugation. Titrations used $PRM_4$ (orange), $PRM_2$ (blue), $PRM_1$ (green), $PRM(H)_1$ (red).



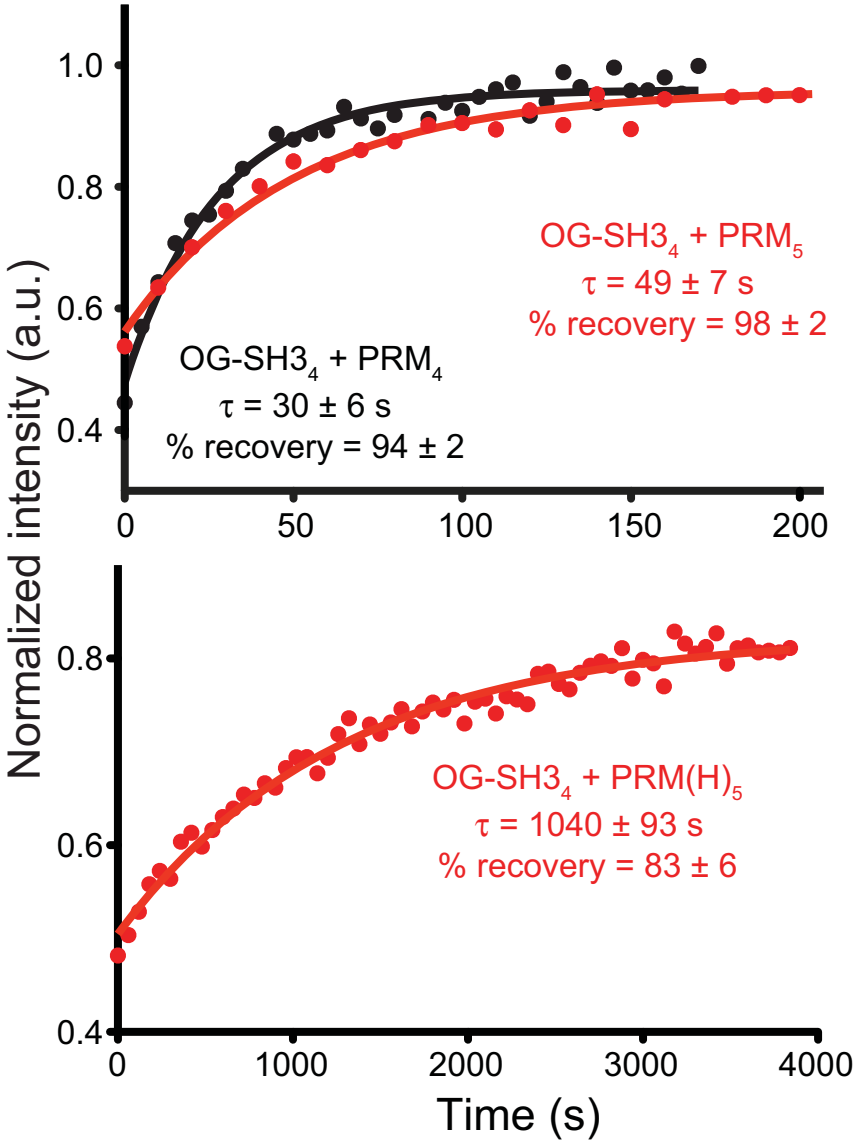

**Supplementary Figure 11. P(r) data.** The interatomic distribution functions [p(R)] for each titration sample in Fig. 2b were calculated using the program GNOM[15] and normalized to a total probability of one. Solid lines are used for samples before phase separation was reached (i.e. at concentrations below that needed for phase separation), while dashed lines are for samples that had phase separated, and show data collected on the supernatant phase, separated from droplets by centrifugation. For each sample, different maximum particle dimensions ($D_{max}$) were surveyed to choose a minimum $D_{max}$ (shown here) that allowed p(R) to naturally approach zero at $D_{max}$. The $R_g$ values calculated from GNOM as shown in Fig. 2b were not sensitive to the $D_{max}$ values surveyed.



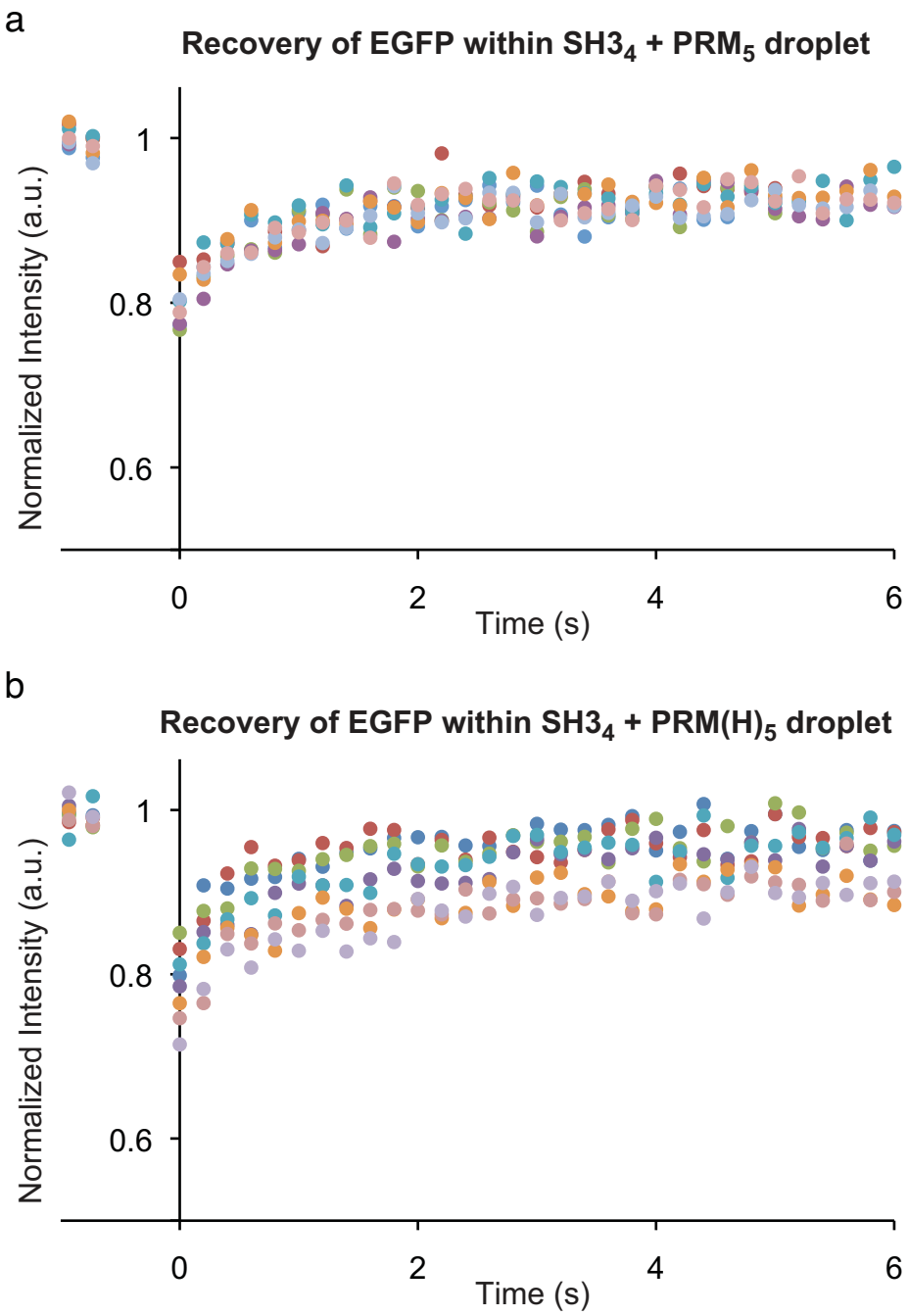

**Supplementary Figure 12. Guinier plots for selected SAXS data from Figure 2b.** Guinier plots for $PRM_4$ (gray), $SH3_5$ (black), a $PRM_4$:$SH3_5$ mixture (orange), and a $PRM_2$:$SH3_5$ mixture (blue).



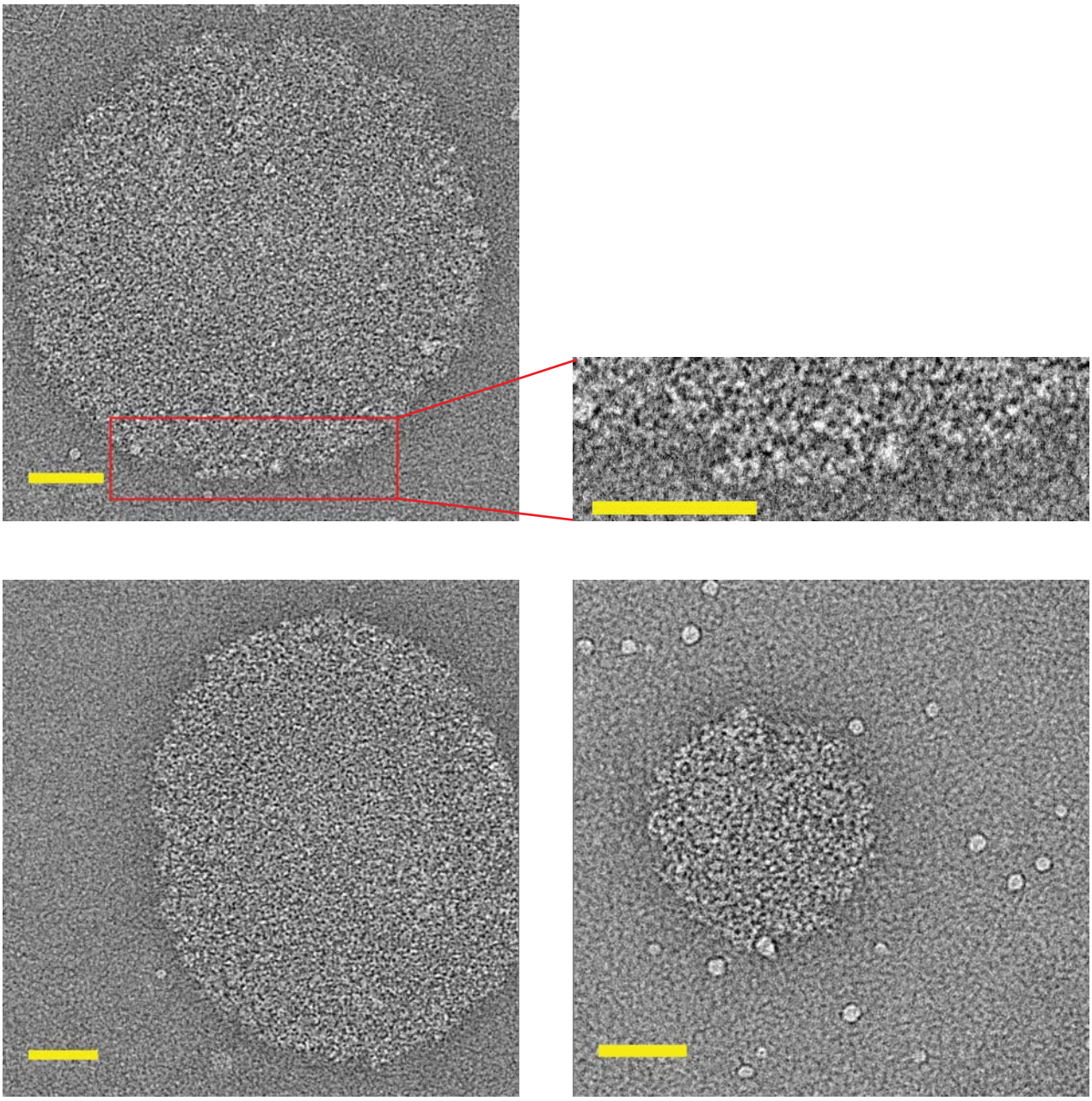

**Supplementary Figure 13. MA-DLS analyses of $SH3_5$+$PRM_5$ droplets and supernatant.** a. Decay time constant (gamma) distribution for the collected droplet phase formed by mixing $SH3_5$ and $PRM_5$, determined by CONTIN analysis. b-d. Variation in gamma with $q^2$ for the fast (b), intermediate (c) and slow (d) processes in the $SH3_5$+$PRM_5$ droplet phase. Note the strong non-linearity in panels c and d. e. The supernatant phase in the $SH3_5$+$PRM_5$ mixture has only a single process (not shown), whose gamma value varies linearly with $q^2$.



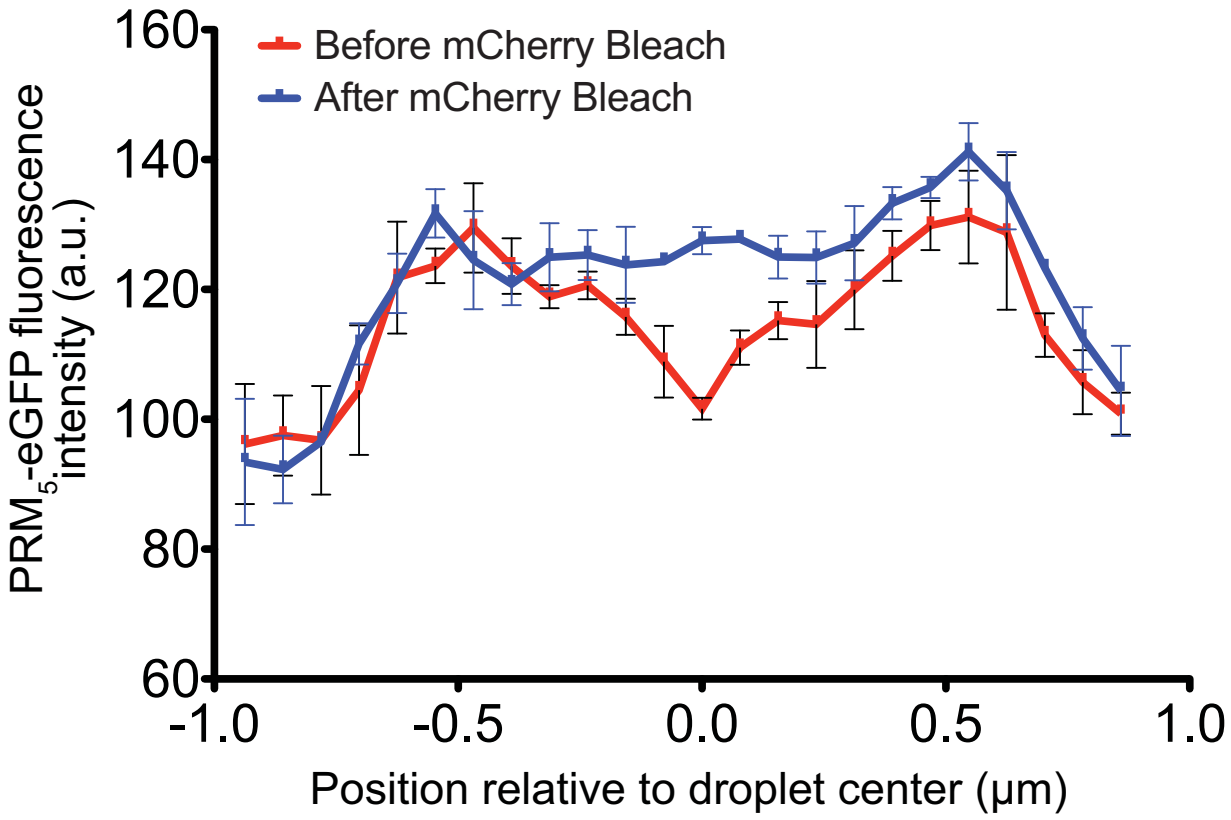

**Supplementary Figure 14. Diffusion of polymerized SH3$_4$ inversely correlates with the valency and affinity of its ligands.** Representative fluorescence recovery curves after photobleaching the center of droplets (<10% total diameter bleached) containing OG-SH3$_4$ plus PRM$_4$ (top, black), PRM$_5$ (top red) or PRM(H)$_5$ (bottom). For the interactions of SH3$_1$ with PRM$_1$ and PRM(H)$_1$, $K_D$ values are 356 $\mu$M and 11 $\mu$M, respectively (Supplementary Fig. 1). Lines show best fit to a single exponential. Time constant of exponential decay and percent total recovery are indicated for n ≥ 5 independent experiments.



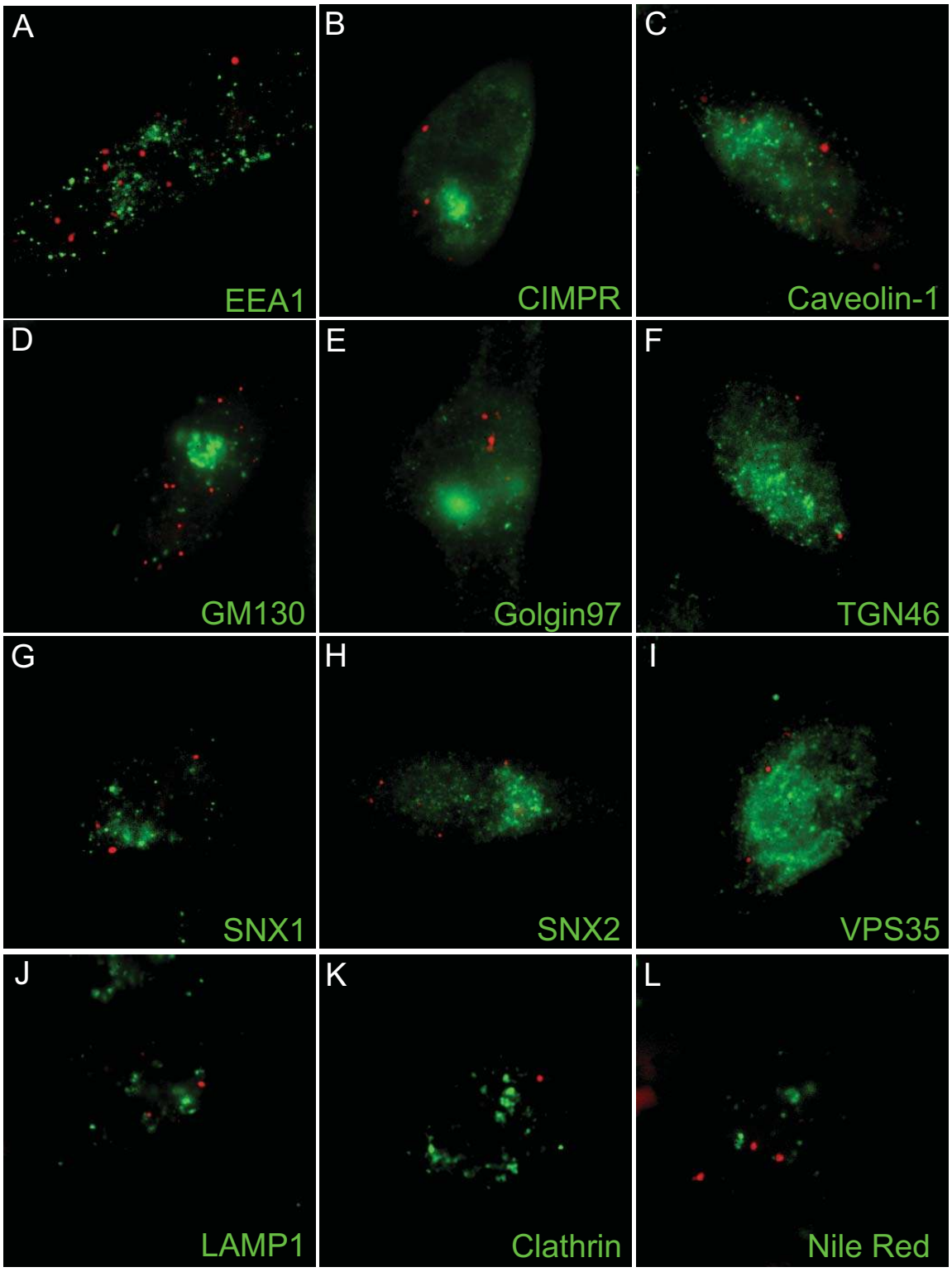

**Supplementary Figure 15. Diffusion of EGFP within droplets is very rapid.** EGFP was added to the indicated droplet-containing solution to the final concentration of 0.5 $\mu$M. Recovery kinetics of EGFP within the droplet after photobleaching are shown. The data show that EGFP can enter the droplet phase and diffuse within it much more rapidly than the polymer constituents (compare to Supplementary Fig. 14), indicating that macromolecules can move rapidly within the interstitial spaces of the droplet. Similar results were obtained for small molecule fluorophores (not shown).




**Supplementary Figure 16. Additional electron cryo-microscopic images of droplets formed by $SH3_5$ + $PRM_5$.**




**Supplementary Figure 17. Non-uniform eGFP fluorescence in mCherry-SH3$_5$/eGFP-PRM$_5$ puncta is due to mCherry-eGFP FRET.** Image shows eGFP fluorescence in a line-scan through the center of a cellular mCherry-SH3$_5$/eGFP-PRM$_5$ punctate structure before and after mCherry photobleaching. Error bars represent the standard error of the mean intensity for three closely spaced lines. The mCherry-SH3$_5$/eGFP-PRM$_5$ puncta have lower eGFP fluorescence at their centers than at their peripheries. Upon photobleaching of mCherry, the central green fluorescence increases appreciably, suggesting the initial non-uniformity is due to FRET between mCherry and eGFP. The different FRET properties at the periphery and center suggest that the bodies may be structurally inhomogeneous, for reasons we do not currently understand.




**Supplementary Figure 18. Cellular puncta formed by mCherry-SH35 and eGFP-PRM5 are not vesicular.** HeLa cells co-expressing mCherry-$SH3_5$ and eGFP-$PRM_5$ were stained with antibodies against various endocytic markers (identified with Cy5-labeled secondary antibodies) or co-expressing eGFP-$SH3_5$ and eGFP-$PRM_5$ were stained with a lipid dye (Nile Red) as indicated. In all panels $SH3_5$-$PRM_5$ puncta are colored red and the various stains are colored green. Punctate structures did not coincide with cytoplasmic vesicles or lipid. (A-C) Endocytic markers; (D-F) Golgi markers; (G-I) Retromer components; (J) lysosome marker; (K) clathrin; (L) Nile red lipid marker.

| Human Neutrophil[1] | | Human T cell (Roybal and Weulfing unpublished data) | |
|---|---|---|---|
| Protein | Global cytoplasmic concentration (μM) | Protein | Global cytoplasmic concentration (μM) |
| WASP | 9 | WASP | 4.0 |
| | | HS1 | 14.5 |

| Schizosaccharomyces pombe[2] | | Schizosaccharomyces pombe[3] | |
|---|---|---|---|
| Protein | Global cytoplasmic concentration (μM) | Protein | Global cytoplasmic concentration (μM) |
| **Actin patch proteins** | | **Endocytosis proteins** | |
| YFP-actin Act1p | 0.8 | Clathrin HC | 3.7 |
| Actin Act1p | 63.2 | Clathrin LC | 2.4 |
| Arp2 (Arp2p) | 2.9 | Sla2/Hip1R | 1.2 |
| Arp3 (Arp3p) | 4.1 | Pan1/Eps15 | 1.5 |
| ARPC1 (Arc1p/Sop2p) | 2.5 | WASp | 3.5 |
| ARPC3 (Arc3p/Arc21p) | 2.4 | Verprolin | 1.0 |
| ARPC5 (Arc5p/Arc16p) | 1.9 | Myosin-1 | 3.1 |
| Capping protein Acp2p | 1.2 | Arp2 | 2.9 |
| Fimbrin Fim1p | 5.3 | Arp3 | 4.2 |
| | | ARPC5 | 1.9 |
| | | Fimbrin | 5.4 |
| | | Capping protein | 1.2 |
| | | App1p | 1.2 |
| | | Coronin | 3.9 |
| | | Twinfilin | 2.1 |
| | | Actin (6%) | 2.0 |

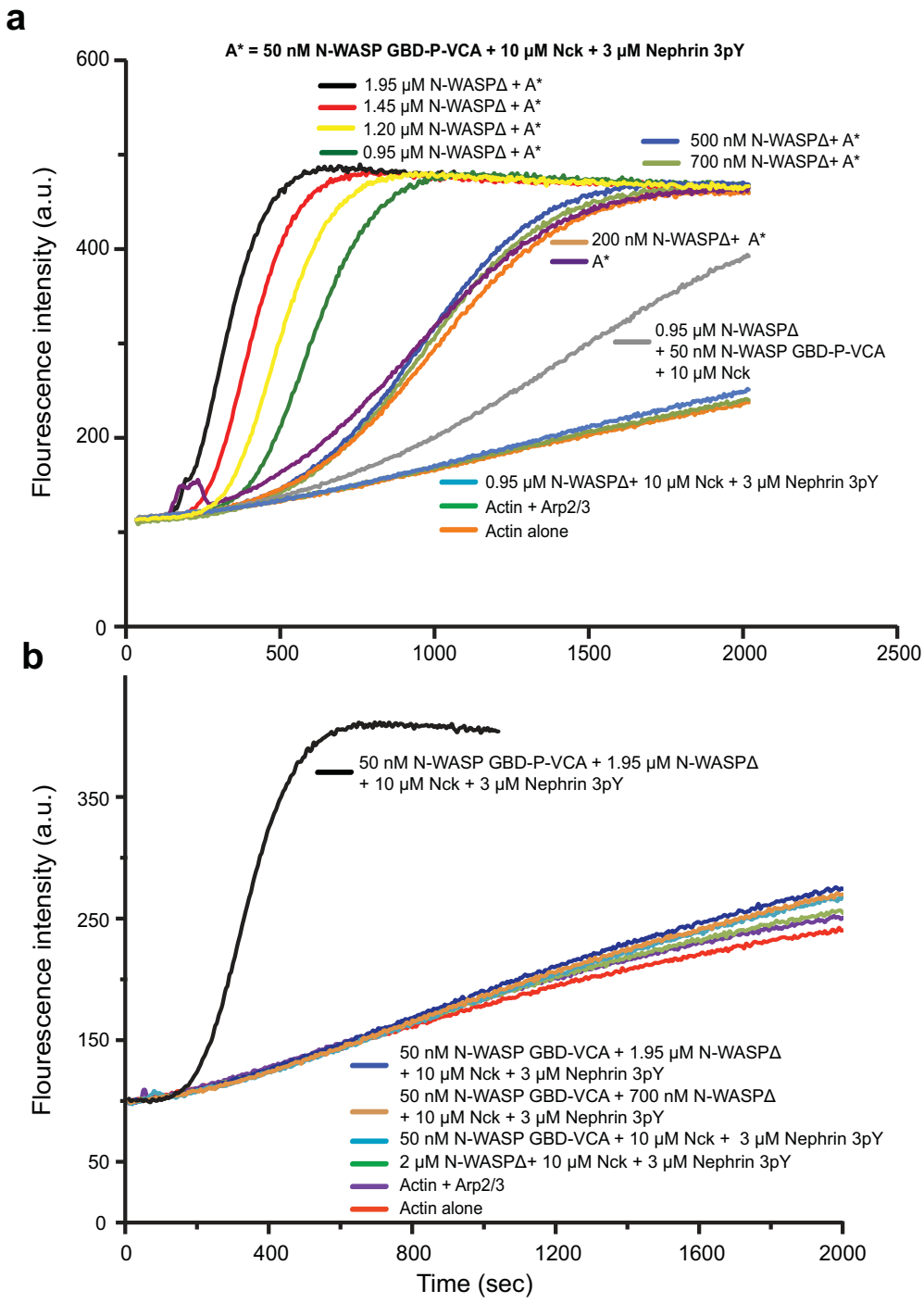
**References for Supplementary Figure 19**

1. Higgs, H. N. & Pollard, T. D. Activation by Cdc42 and PIP(2) of Wiskott-Aldrich syndrome protein (WASp) stimulates actin nucleation by Arp2/3 complex. J Cell Biol 150, 1311-1320, (2000).

2. Wu, J. Q. & Pollard, T. D. Counting cytokinesis proteins globally and locally in fission yeast. Science 310, 310-314, (2005).

3. Sirotkin, V., Berro, J., Macmillan, K., Zhao, L. & Pollard, T. D. Quantitative analysis of the mechanism of endocytic actin patch assembly and disassembly in fission yeast. Mol Biol Cell 21, 2894-2904.

**Supplementary Figure 19. Measured cellular concentrations of actin regulatory proteins.** Cytoplasmic concentrations of WASP and actin regulatory proteins reported in the literature. Note that cytoplasmic concentration of WASP ranges from 3.5 to 9 $\mu$M. The WASP interacting protein, HS1, is as high as 14.5 $\mu$M in T cells.




**Supplementary Figure 20. Phase transition correlates to biochemical activity transition in the N-WASP/Nck/nephrin system.** (a) Raw data for actin assembly reaction data shown in Figure 4e. All reactions (except actin alone and actin + Arp2/3 complex) contained 4 $\mu$M actin (5% pyrene labeled), 10 nM Arp2/3 complex, 10 µM Nck, 3 µM nephrin 3pY, 50 nM N-WASP GBD-P-VCA plus the indicated concentrations of an N-WASP construct lacking the activity-bearing VCA region (N-WASPΔ). (b) Actin assembly reactions performed as in (a), except that N-WASP GBD-P-VCA was replaced by an N-WASP construct lacking the proline-rich region (GBD-VCA). Since GBD-VCA cannot bind Nck, and therefore does not incorporate into the nephrin 3pY/Nck/N-WASPΔ polymer, its activity is not enhanced even by 1.95 µM N-WASPΔ.

# Supplementary Figure 21

| ***S. pombe/S. cerevisiae*** | | | | |
|---|---|---|---|---|
| **Protein** | **Domain** | **Valency** | **Localization** | **Reference** |
| Vgl1 | KH | 13 | Stress Granules | [1] |
| SCP160 | KH | 7 | Stress Granules | [1] |
| Pab1 | RRM | 4 | Stress Granules | [2] |
| Hek2p | KH | 3 | Uncharacterized Cytoplasmic Puncta | [3] |
| EDE1 | EH | 3 | Uncharacterized Cytoplasmic Puncta | [4] |
| FRQ1 | EH | 3 | Uncharacterized Cytoplasmic Puncta | [4] |
| PAN1 | EH | 4 | Uncharacterized Cytoplasmic Puncta | [4] |
| ***C. elegans*** | | | | |
| **Protein** | **Domain** | **Valency** | **Localization** | **Reference** |
| Puf-5 | Pumilio | 7 | P-body like mRNP granules | [5] |
| UNC-75 | RRM | 3 | Nuclear Speckles | [6] |
| UNC-97 | LIM | 5 | Focal Adhesion | [7] |
| ***D. melanogaster*** | | | | |
| **Protein** | **Domain** | **Valency** | **Localization** | **Reference** |
| Tud | Tudor | 10 | Polar Granules | [8] |
| Pasilla | KH | 3 | Nuclear Puncta | [9] |
| Staufen | DSRM | 4 | P-bodies | [10] |
| ***M. musculus*** | | | | |
| **Protein** | **Domain** | **Valency** | **Localization** | **Reference** |
| hnRNP E1 | KH | 3 | Interchromatin Granules | [11] |
| hnRNP E2 | KH | 3 | Interchromatin Granules | [11] |
| hnRNP K | KH | 3 | Interchromatin Granules | [11] |
| KH type splicing regulatory factor | KH | 4 | Interchromatin Granules | [11] |
| APOBEC-1 | RRM | 3 | Interchromatin Granules | [11] |
| Rod1 | RRM | 4 | Interchromatin Granules | [11] |
| hnRNP F/H | RRM | 3 | Interchromatin Granules | [11] |
| hnRNP H' | RRM | 3 | Interchromatin Granules | [11] |
| hnRNP I | RRM | 4 | Interchromatin Granules | [11] |
| hnRNP L | RRM | 3 | Interchromatin Granules | [11] |
| hnRNP M | RRM | 3 | Interchromatin Granules | [11] |
| Siah binding protein | RRM | 3 | Interchromatin Granules | [11] |
| Splicing factor HCC1 | RRM | 3 | Interchromatin Granules | [11] |
| U2AF65 | RRM | 3 | Interchromatin Granules | [11] |
| Elav-like 1 | RRM | 3 | Interchromatin Granules | [11] |
| RNA binding protein HuR | RRM | 3 | Interchromatin Granules | [11] |

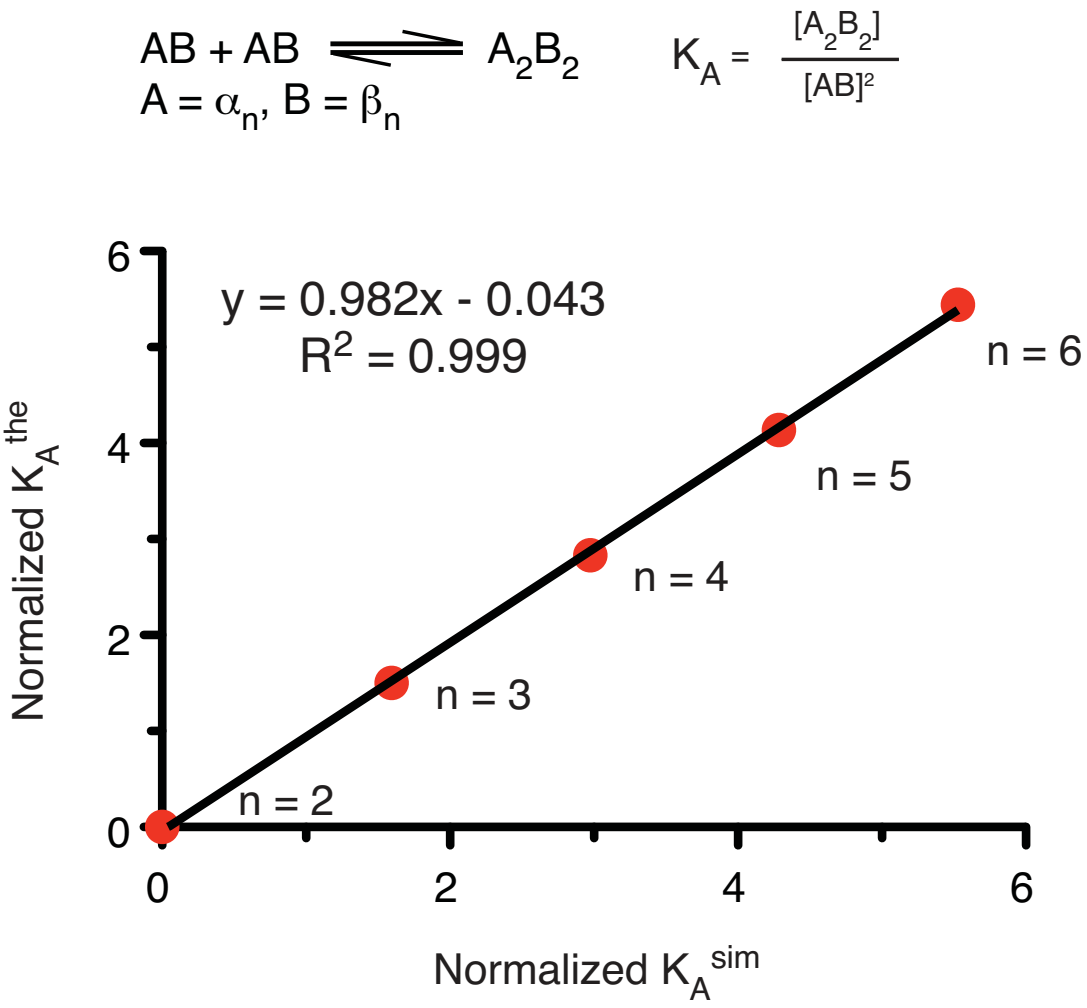
| | | | | |
|---|---|---|---|---|
| Rnpc2 | RRM | 3 | Interchromatin Granules | [11] |
| Zinc finger RNA binding protein, ZFR (KIAA1086) | ZnF | 3 | Interchromatin Granules | [11] |
| POZ domain protein FBI-1 | ZnF_C2H2 | 4 | Interchromatin Granules | [11] |
| POZ/zinc finger transcription factor, ODA-8 | ZnF_C2H2 | 5 | Interchromatin Granules | [11] |
| CPSF 30 kDa subunit | ZnF_C3H1 | 5 | Interchromatin Granules | [11] |
| KIAA0663 protein | ZnF_C3H1 | 3 | Interchromatin Granules | [11] |
| TDRD6 | Tudor | 7 | Nuages | [12] |
| TDRD7 | Tudor | 3 | Nuages | [12] |
| HSF1 | HSF | Oligomer | Stress Granule | [13] |
| ***H. sapiens*** | | | | |
| **Protein** | **Domain** | **Valency** | **Localization** | **Reference** |
| Microcephalin | BRCT | 3 | DNA-damage Foci | [14] |
| TopBP1 | BRCT | 7 | DNA-damage Foci | [15] |
| c-Myb | SANT | 3 | Nuclear Speckles | [16] |
| PTB-associated splicing factor | Poly-proline | >6 | Nuclear Speckles | [17] |
| Splicing Factor 3b | Poly-proline | >7 | Nuclear Speckles | [18] |
| SNP70 | Poly-proline | >7 | Nuclear Speckles | [19] |
| Mucin | Poly-proline | >10 | Mucin Granules | [20] |
| WIP | Poly-proline | >6 | Lytic Granules | [21] |
| IQGAP1 | IQ | 4 | hStaufen Granules | [22] |
| Leupaxin | LIM | 4 | Focal Adhesion | [23] |
| Zyxin | LIM | 3 | Focal Adhesion | [24] |
| HuD | RRM | 3 | P-Bodies | [25] |
| TIA1 | RRM | 3 | Stress Granules | [26] |
| PABP | RRM | 4 | Stress Granules | [27] |
| HuR | RRM | 3 | Stress Granules | [28] |
| PTB1 | RRM | 4 | Perinucleolar Compartment | [29] |
| CUG-BP | RRM | 3 | Perinucleolar Compartment | [30] |
| hnRNP Q | RRM | 3 | P-Bodies | [25] |
| Sp1 | ZnF | 3 | PML Bodies | [31] |
| PLZF | ZnF | 9 | PML Bodies | [31] |
| RNF17 | Tudor | 4 | Nuages | [32] |
| KSRP | KH | 4 | Perinucleolar Compartment | [33] |
| IMP1 | KH | 3 | RNP Granules | [34] |
| PCBP2 | KH | 3 | P-Bodies | [25] |

## References for Supplementary Figure 21

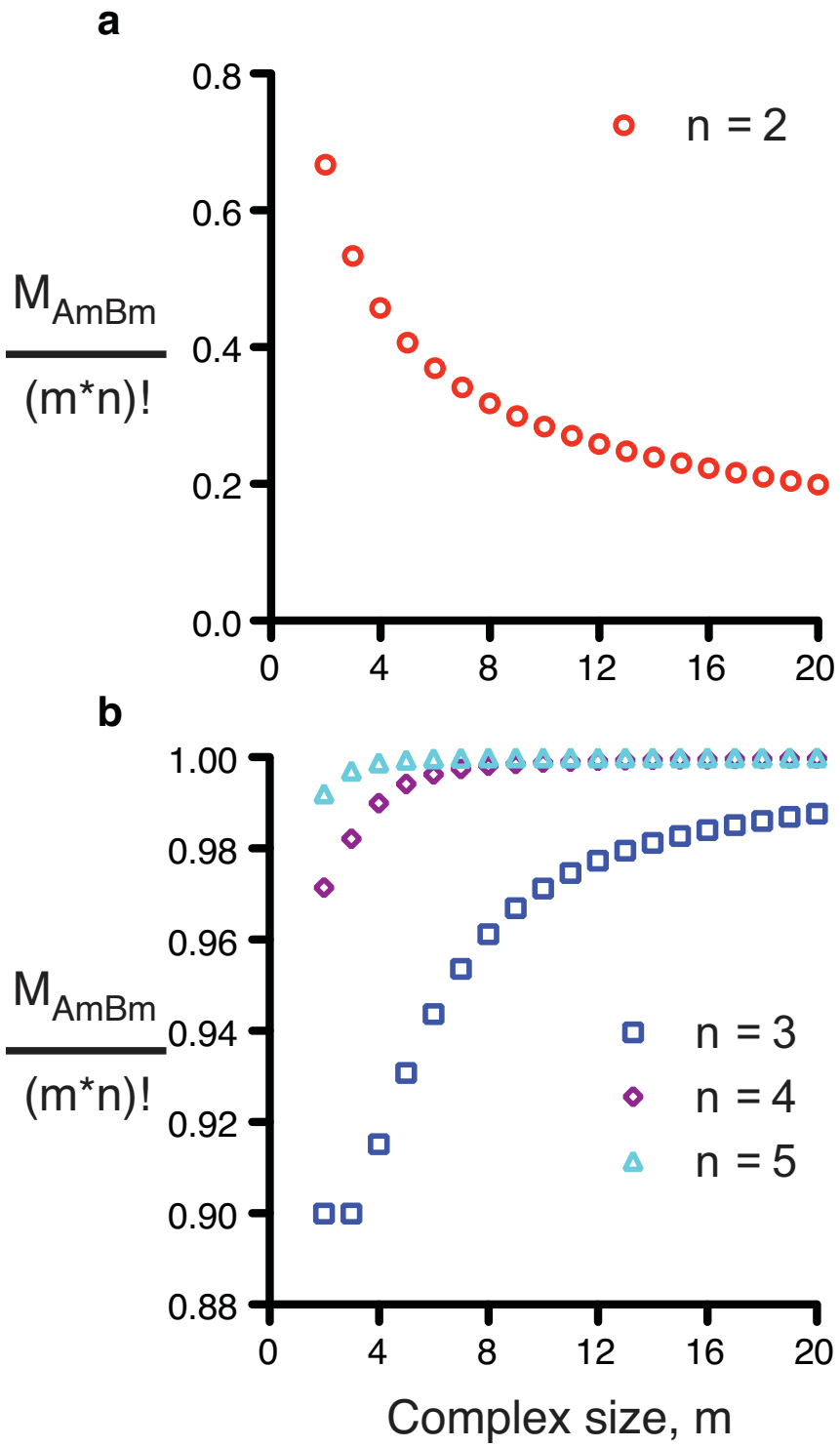

1. Wen, W.L., et al., Vgl1, a multi-KH domain protein, is a novel component of the fission yeast stress granules required for cell survival under thermal stress. Nucleic Acids Res, 2010.
2. Hoyle, N.P., et al., Stress-dependent relocalization of translationally primed mRNPs to cytoplasmic granules that are kinetically and spatially distinct from P-bodies. J Cell Biol, 2007. **179**(1): p. 65-74.
3. Noree, C., et al., Identification of novel filament-forming proteins in Saccharomyces cerevisiae and Drosophila melanogaster. J Cell Biol, 2010. **190**(4): p. 541-51.
4. Huh, W.K., et al., Global analysis of protein localization in budding yeast. Nature, 2003. **425**(6959): p. 686-91.
5. Noble, S.L., et al., Maternal mRNAs are regulated by diverse P body-related mRNP granules during early Caenorhabditis elegans development. J Cell Biol, 2008. **182**(3): p. 559-72.
6. Loria, P.M., et al., Two neuronal, nuclear-localized RNA binding proteins involved in synaptic transmission. Curr Biol, 2003. **13**(15): p. 1317-23.
7. Norman, K.R., et al., UNC-97/PINCH is involved in the assembly of integrin cell adhesion complexes in Caenorhabditis elegans body wall muscle. Dev Biol, 2007. **309**(1): p. 45-55.
8. Arkov, A.L., et al., The role of Tudor domains in germline development and polar granule architecture. Development, 2006. **133**(20): p. 4053-62.
9. Seshaiah, P., et al., pasilla, the Drosophila homologue of the human Nova-1 and Nova-2 proteins, is required for normal secretion in the salivary gland. Dev Biol, 2001. **239**(2): p. 309-22.
10. Barbee, S.A., et al., Staufen- and FMRP-containing neuronal RNPs are structurally and functionally related to somatic P bodies. Neuron, 2006. **52**(6): p. 997-1009.
11. Saitoh, N., et al., Proteomic analysis of interchromatin granule clusters. Mol Biol Cell, 2004. **15**(8): p. 3876-90.
12. Hosokawa, M., et al., Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: domain composition, intracellular localization, and function in male germ cells in mice. Dev Biol, 2007. **301**(1): p. 38-52.
13. Sarge, K.D., S.P. Murphy, and R.I. Morimoto, Activation of heat shock gene transcription by heat shock factor 1 involves oligomerization, acquisition of DNA-binding activity, and nuclear localization and can occur in the absence of stress. Mol Cell Biol, 1993. **13**(3): p. 1392-407.
14. Wood, J.L., et al., MCPH1 functions in an H2AX-dependent but MDC1-independent pathway in response to DNA damage. J Biol Chem, 2007. **282**(48): p. 35416-23.
15. Zhang, F., J. Wu, and X. Yu, Integrator3, a partner of single-stranded DNA-binding protein 1, participates in the DNA damage response. J Biol Chem, 2009. **284**(44): p. 30408-15.
16. Alm-Kristiansen, A.H., et al., FLASH acts as a co-activator of the transcription factor c-Myb and localizes to active RNA polymerase II foci. Oncogene, 2008. **27**(34): p. 4644-56.

17. Shav-Tal, Y., et al., Nuclear relocalization of the pre-mRNA splicing factor PSF during apoptosis involves hyperphosphorylation, masking of antigenic epitopes, and changes in protein interactions. Mol Biol Cell, 2001. **12**(8): p. 2328-40.
18. Terada, Y. and Y. Yasuda, Human immunodeficiency virus type 1 Vpr induces G2 checkpoint activation by interacting with the splicing factor SAP145. Mol Cell Biol, 2006. **26**(21): p. 8149-58.
19. Craggs, G., et al., A nuclear SH3 domain-binding protein that colocalizes with mRNA splicing factors and intermediate filament-containing perinuclear networks. J Biol Chem, 2001. **276**(32): p. 30552-60.
20. Verdugo, P., Mucin exocytosis. Am Rev Respir Dis, 1991. **144**(3 Pt 2): p. S33-7.
21. Krzewski, K., X. Chen, and J.L. Strominger, WIP is essential for lytic granule polarization and NK cell cytotoxicity. Proc Natl Acad Sci U S A, 2008. **105**(7): p. 2568-73.
22. Villace, P., R.M. Marion, and J. Ortin, The composition of Staufen-containing RNA granules from human cells indicates their role in the regulated transport and translation of messenger RNAs. Nucleic Acids Res, 2004. **32**(8): p. 2411-20.
23. Tanaka, T., et al., LIM domain-containing adaptor, leupaxin, localizes in focal adhesion and suppresses the integrin-induced tyrosine phosphorylation of paxillin. Cancer Sci, 2010. **101**(2): p. 363-8.
24. Krylyshkina, O., et al., Nanometer targeting of microtubules to focal adhesions. J Cell Biol, 2003. **161**(5): p. 853-9.
25. Kulkarni, M., S. Ozgur, and G. Stoecklin, On track with P-bodies. Biochem Soc Trans, 2010. **38**(Pt 1): p. 242-51.
26. Delestienne, N., et al., The splicing factor ASF/SF2 is associated with TIA-1-related/TIA-1-containing ribonucleoproteic complexes and contributes to post-transcriptional repression of gene expression. FEBS J, 2010. **277**(11): p. 2496-514.
27. Ralser, M., et al., An integrative approach to gain insights into the cellular function of human ataxin-2. J Mol Biol, 2005. **346**(1): p. 203-14.
28. Gallouzi, I.E., et al., HuR binding to cytoplasmic mRNA is perturbed by heat shock. Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3073-8.
29. Huang, S., et al., The dynamic organization of the perinucleolar compartment in the cell nucleus. J Cell Biol, 1997. **137**(5): p. 965-74.
30. Timchenko, L.T., et al., Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. Nucleic Acids Res, 1996. **24**(22): p. 4407-14.
31. Vallian, S., K.V. Chin, and K.S. Chang, The promyelocytic leukemia protein interacts with Sp1 and inhibits its transactivation of the epidermal growth factor receptor promoter. Mol Cell Biol, 1998. **18**(12): p. 7147-56.
32. Pan, J., et al., RNF17, a component of the mammalian germ cell nuage, is essential for spermiogenesis. Development, 2005. **132**(18): p. 4029-39.
33. Markovtsov, V., et al., Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol Cell Biol, 2000. **20**(20): p. 7463-79.
34. Nielsen, F.C., et al., Cytoplasmic trafficking of IGF-II mRNA-binding protein by conserved KH domains. J Cell Sci, 2002. **115**(Pt 10): p. 2087-97.

**Supplementary Figure 21. Proteins reported to show punctate, non-vesicular staining in cells.** Table includes multi-valent proteins that have been reported to localize to various punctate structures in cells. The list was generated by searching for proteins containing three or more of the listed domains on the SMART database (http://smart.embl-heidelberg.de/) and correlating the list with constituents of various punctate structures observed in literature. The search was limited to proteins containing valencies of three or more.




**Supplementary Figure 22. Simulations recapitulate theoretical predictions in dilute solution.** Y-axis shows the fold increase in microstates of the reaction of two AB heterodimers oligomerizing to $A_2B_2$ calculated from theoretical considerations for valencies of 2, 3, 4, 5 and 6 ($K_A^{the}$, normalized to the value for valency = 2). Y-axis shows the normalized $K_A$ values for each valency determined from simulation trajectories ($K_A^{sim}$). The two quantities are linearly related ($r^2 \sim 1.0$) with slope and intercept close to 1.0 and 0.0, respectively, indicating the theory and simulation recapitulate each other well.




**Supplementary Figure 23. Fraction of (m*n)! giving intact $A_mB_m$ complex.** For systems composed of m A molecules and m B molecules, each with valency n, the number of possible configurations is (m*n)!, among which some configurations result in more than one complex. The difference between the former and the latter is the possible configurations for an $A_mB_m$ complex, $M_{AmBm}$. The ratio of $M_{AmBm}$ to (m*n)! is plotted for valencies of (a) 2 (red circle); (b) 3 (blue square), 4 (purple diamond) and 5 (cyan triangle).