PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

# Stoichiometric balance of protein copy numbers is measurable and functionally significant in a protein-protein interaction network for yeast endocytosis

**David O. Holland[1], Margaret E. Johnson[2]***

**1** Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America, **2** Department of Biophysics, Johns Hopkins University, Baltimore, Maryland, United States of America

* margaret.johnson@jhu.edu



OPEN ACCESS

**Citation:** Holland DO, Johnson ME (2018) Stoichiometric balance of protein copy numbers is measurable and functionally significant in a protein-protein interaction network for yeast endocytosis. PLoS Comput Biol 14(3): e1006022. https://doi.org/10.1371/journal.pcbi.1006022

**Editor:** Martin Meier-Schellersheim, National Institutes of Health, UNITED STATES

**Received:** October 16, 2017

**Accepted:** February 3, 2018

**Published:** March 8, 2018



**Data Availability Statement:** All data files and software are available from a public GitHub repository, https://github.com/mjohn218/StoichiometricBalance.

**Funding:** Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R00GM098371 to M.E.J. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Stoichiometric balance, or dosage balance, implies that proteins that are subunits of obligate complexes (e.g. the ribosome) should have copy numbers expressed to match their stoichiometry in that complex. Establishing balance (or imbalance) is an important tool for inferring subunit function and assembly bottlenecks. We show here that these correlations in protein copy numbers can extend beyond complex subunits to larger protein-protein interactions networks (PPIN) involving a range of reversible binding interactions. We develop a simple method for quantifying balance in any interface-resolved PPINs based on network structure and experimentally observed protein copy numbers. By analyzing such a network for the clathrin-mediated endocytosis (CME) system in yeast, we found that the real protein copy numbers were significantly more balanced in relation to their binding partners compared to randomly sampled sets of yeast copy numbers. The observed balance is not perfect, highlighting both under and overexpressed proteins. We evaluate the potential cost and benefits of imbalance using two criteria. First, a potential cost to imbalance is that 'leftover' proteins without remaining functional partners are free to misinteract. We systematically quantify how this misinteraction cost is most dangerous for strong-binding protein interactions and for network topologies observed in biological PPINs. Second, a more direct consequence of imbalance is that the formation of specific functional complexes depends on relative copy numbers. We therefore construct simple kinetic models of two sub-networks in the CME network to assess multi-protein assembly of the ARP2/3 complex and a minimal, nine-protein clathrin-coated vesicle forming module. We find that the observed, imperfectly balanced copy numbers are less effective than balanced copy numbers in producing fast and complete multi-protein assemblies. However, we speculate that strategic imbalance in the vesicle forming module allows cells to tune where endocytosis occurs, providing sensitive control over cargo uptake via clathrin-coated vesicles.

**Competing interests:** The authors have declared that no competing interests exist.


## Author summary

Protein copy numbers are often found to be stoichiometrically balanced for subunits of multi-protein complexes. Imbalance is believed to be deleterious because it lowers complex yield (the dosage balance hypothesis) and increases the risk of misinteractions, but imbalance may also provide unexplored functional benefits. We show here that the benefits of stoichiometric balance can extend to larger networks of interacting proteins. We develop a method to quantify to what degree protein networks are balanced, and apply it to two networks. We find that the clathrin-mediated endocytosis system in yeast is statistically balanced, but not perfectly so, and explore the consequences of imbalance in the form of misinteractions and endocytic function. We also show that biological networks are more robust to misinteractions than random networks when balanced, but are more sensitive to misinteractions under imbalance. This suggests evolutionary pressure for proteins to be balanced and that any conserved imbalance should occur for functional reasons. We explore one such reason in the form of bottlenecking the endocytosis process. Our method can be generalized to other networks and used to identify out-of-balance proteins. Our results provide insight into how network design, expression level regulation, and cell fitness are intertwined.


## Introduction

Protein copy numbers in yeast vary from a few to well over a million[1, 2]. Expression levels, along with a protein's binding partners and corresponding affinities, are critical determinants of a protein's function within the cell. In the context of multiprotein complexes–especially obligate complexes such as the ribosome–it is thought that protein concentrations are balanced according to the stoichiometry of the complex. This is referred to as the dosage balance hypothesis (DBH)[3–5]. Here, we expand this hypothesis to a network wide level, where proteins participate in multiple distinct complexes as well as transient interactions. In these more complex networks (Fig 1A), balance can be defined as having just enough copies of each protein to construct a target vector of complex abundances, with no proteins (or protein binding sites) in significant deficiency or excess. This generalized definition of balance reproduces the expected result for obligate complexes, where, for example, the ARP2/3 obligate complex (Fig 1B) would be balanced if all subunits had equal copy numbers.

For obligate complexes, dosage balance means that there are no leftover subunits, as these would be a waste of cell resources. However, even for proteins in non-obligate complexes a number of deleterious effects could be caused by imbalance. An overexpressed core or "bridge" subunit may sequester periphery subunits, paradoxically lowering the final number of complete complexes[5, 6]. Excess proteins may be prone to misinteractions, also called interaction promiscuity, with nonfunctional partners. Numerous studies have identified proteins with high intrinsic disorder as sensitive to overexpression[7–9], and these proteins have low, tightly regulated native expression levels[10, 11] indicating that misinteraction propensity and abundance are related. Underexpression carries its own dangers: a single underexpressed subunit will become a bottleneck for the whole complex. In addition, weakly expressed proteins are noisier[12] and thus less reliable for the cell. Male (XY) animal cells are known to employ "dosage compensation" mechanisms to increase the expression of X-chromosomal genes to be on par with female cells[13, 14], though for other genes it is the female cell that cuts expression levels in half[15], indicating that the cell preserves an optimized set of expression levels.

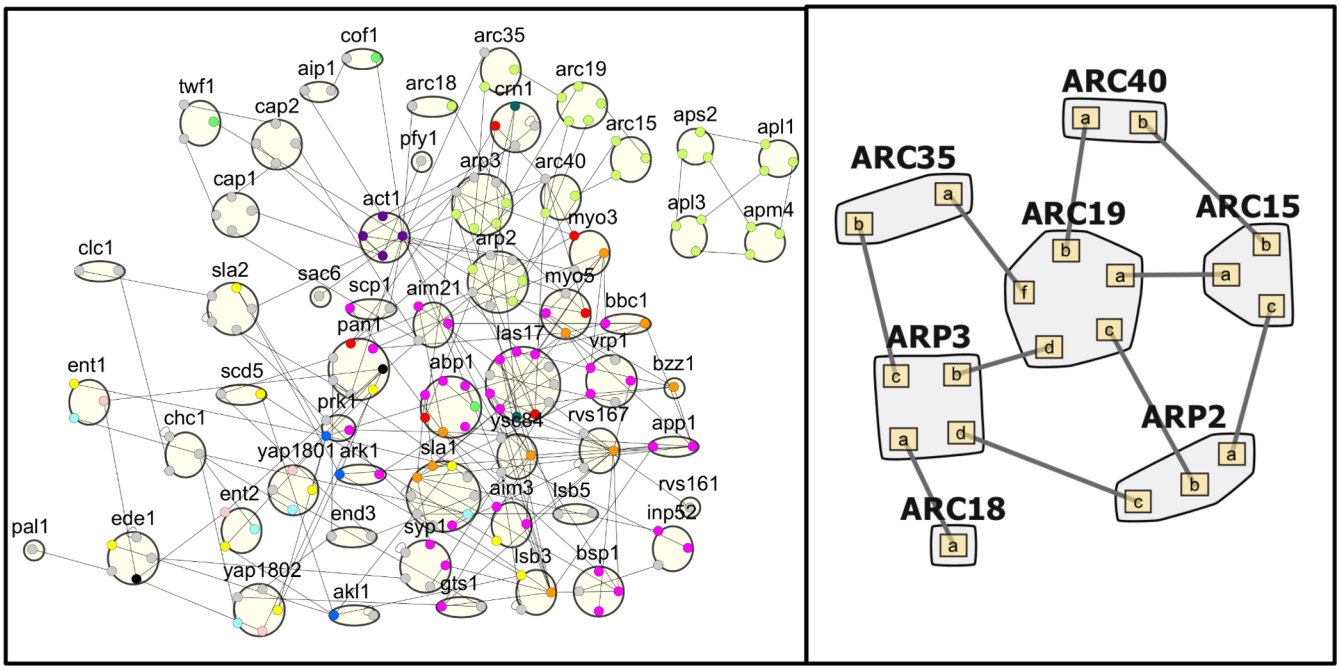



**Fig 1. Clathrin-mediated endocytosis network in yeast.** (Left) Site graph for the protein-protein interaction network (N = 56, E = 186), displaying interfaces used for binding interactions. Interfaces are color-coded according to domain type, the most common being SH3 domains (orange), Proline-rich regions (pink), phosphosites (yellow), acidic domains (red), and multi-protein complex subunit interfaces (light green). (Right) The ARP2/3 complex, a subset of the larger CME network.

https://doi.org/10.1371/journal.pcbi.1006022.g001

But optimized does not necessarily mean balanced. Imbalance may be necessary for functional reasons: signaling networks utilize underexpressed hubs to regulate which pathways are active as a given time[16]. Recent models show imbalance can be beneficial to complex assembly when affinity and kinetics are taken into account[17, 18]. A study of over 5,400 human proteins by Hein et al. found that strong interactions forming stable complexes are correlated with balance, but weak interactions are not, which may mean that the network as a whole is not balanced [19]. Finally, the concept of dosage balance being an optimal set of protein copy numbers generally relies on the assumption that proteins reach an equilibrium state of complex yield. Most processes in the cell do not occur at equilibrium and therefore deviations from balance could be beneficial in non-equilibrium models.

Here, we test the hypothesis that protein expression levels are significantly biased towards balance, even for complex PPINs that include weak and transient interactions. This first required us to develop a method to quantify stoichiometric balance in any arbitrary PPIN, given known binding interfaces and some observed copy numbers, which we call Stoichiometric Balance Optimization of Protein Networks (SBOPN). Copy number correlations thus are evaluated beyond direct binding partners to the more global network of interactors. We then can quantify the consequences of imbalance relative to perfect balance according to two criteria: 1) the deleterious consequences and cost of forming misinteractions, and 2) the potentially beneficial control of specific functional outcomes by modulating which complexes, given known binding affinities, actually assemble. Applied to the 56-protein, manually curated, interface-resolved CME PPIN [20], two of its sub-networks, as well as the ErbB PPIN[16], we find that stoichiometric balance in observed copy numbers is often significant, and observed imbalances, particularly of underexpressed proteins, could provide tuning knobs for functional outcomes.

The first consequence of imbalance we evaluate, misinteractions cost, has an indirect effect on function by allowing unbound proteins to bind to non-functional partners, sequestering components and thus affecting formation of specific complexes. They are believed to play a role in dosage sensitivity[7, 8, 21], and avoiding them has been shown to be an evolutionary force limiting protein diversity[22, 23], expression levels[24, 25], binding strengths[26], and protein network structure[23, 27]. Misinteractions, not being selected for by evolution, are weak and generally unstable, but there are far more ways for N proteins to misinteract (order $N^2$) than bind to their few functional partners (order N) [22, 23]. Cells have evolved a variety of mechanisms to increase specificity, such as allostery[28, 29], negative design[30, 31], compartmentalization[22], and temporal regulation of expression[32]. Copy number balance would be another such mechanism, as protein binding sites would saturate their stronger-binding functional partners.

The second and ultimately more direct consequence of imbalance we evaluate is that changes to copy numbers control which specific and functionally necessary complexes can form. When the central clathrin protein is knocked out in cells, for example, clathrin-mediated endocytosis (CME) is terminated, as clathrin is functionally irreplaceable[33]. The plasma membrane lipid $PI(4,5)P_2$ is also essential for CME, as it is required for recruiting the diverse cytosolic clathrin-coat proteins to the membrane to assemble vesicles[34]. Many clathrin-coat proteins, however, can be knocked out without fully terminating CME[35]. As the CME network illustrates (Fig 1), most of these proteins have multiple domains mediating interactions involving both competitive and non-competitive interactions. Adaptor proteins (proteins that bind to the membrane, to transmembrane cargo, and often to clathrin as well) exhibit redundancy in their binding partners that can partially explain how knock-outs to one protein can be rescued by the activity of related proteins. With simulation of simple kinetic models, we can then test these hypotheses, including for the non-equilibrium production of vesicles at the membrane. Although these models are far too simple to recapitulate the complexities of CME *in vivo*, they are nonetheless useful in highlighting potential bottlenecks in assembly due to copy numbers or binding affinities.

Quantifying balance in protein networks can thus lead to new insights, as unbalanced proteins may serve as assembly bottlenecks, or maintain alternate cellular functions outside of the network module being analyzed[18]. Dosage balance is also important for understanding dosage sensitivity[4, 21], a phenomenon where overexpression of a gene is detrimental or even lethal to cell growth. Studies estimate ~15% of genes in *S. cerevisiae* to be dosage sensitive[9, 36], but the negative effects of gene overexpression have been observed in several eukaryotic species including maize[4], flies[37], and humans[38–40]. Studying balance at a network-wide level is challenging because it requires resolved information about the interfaces proteins use to bind. A protein that binds noncompetitively with two partners requires equal abundance to its partners. But if the binding is competitive–i.e. the same interface is used to bind two different partners–the protein's abundance must equal the sum of that of its partners to have no leftovers (Fig 2). Classic protein-protein interactions networks (PPINs) lack this resolution, but recent studies have begun to add this information, creating what we refer to as interface-interaction networks (IINs)[16, 20, 41]. An IIN tracks not just protein partners but also the binding sites that proteins use to bind.

Our study of stoichiometric balance in larger, interface resolved PPINs is organized in the Results section in three parts. In the first part, we define a metric for quantifying stoichiometric balance and how noise in protein expression levels can be approximately accounted for. We apply our algorithm SBOPN to the CME PPIN [20, 41] and the ErbB PPIN [16], highlighting which proteins are over- and underexpressed relative to perfect balance. Although this analysis excludes temporal expression and binding affinity, it provides a starting point for the

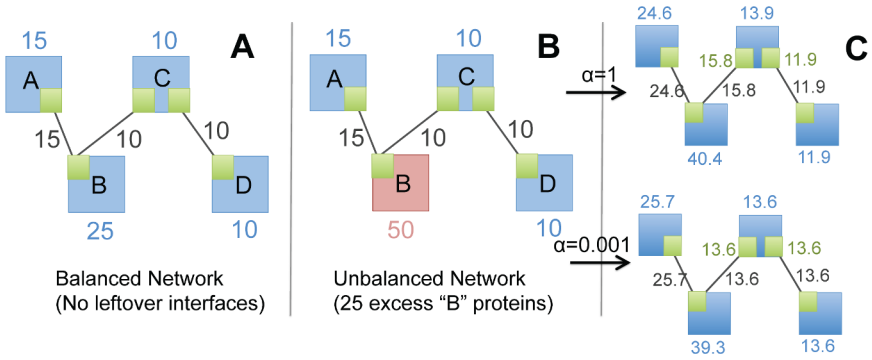



**Fig 2. Examples of balanced vs unbalanced copy numbers, and optimal solutions found by our algorithm.** A) A network with balanced copy numbers has just enough proteins (blue numbers) to form the desired number of complexes (black numbers). B) The copy numbers are unbalanced because an excess of "B" proteins is leftover after all possible complexes form. C) Starting from the network of (B) and using its copy numbers as $C_0$, our algorithm 'Stoichiometric Balance Optimization of Protein Networks' (SBOPN) solves for a balanced set of interface copy numbers (green text) that both 1) optimizes distance of the balanced interface copy numbers to $C_0$ and 2) constrains all interfaces on the same protein to the same copy number. The parameter α controls which of the two constraints is weighted more strongly. A low α (lower solution) forces all interfaces to the same copies on a protein. Higher α (upper solution) allows interfaces to vary to solve for copy numbers closer to $C_0$, as seen for protein "C". The protein copy number for "C" is calculated as the average over all its interface copy numbers.

https://doi.org/10.1371/journal.pcbi.1006022.g002

analysis of these features in the subsequent parts. In the second part, we switch to generalized interface-interaction network (IIN) topologies and network motifs to focus exclusively on how our first evaluation criteria, the cost of misinteractions under imbalance, is worse for strong binding proteins and for network topologies that resemble biological networks. In the third part, we return to the interface-resolved CME PPIN to evaluate the observed degree of stoichiometric balance in two smaller sub-networks of the CME network: the 7-subunit ARP2/3 complex and a simplified, nine protein, clathrin-coat forming module. In these sub-modules, we now can also evaluate our second criteria and assess how observed copy numbers influence proper multi-protein assembly given known binding affinities of interactions. Our simulations of (non-spatial) kinetic models demonstrate that stoichiometric balance does, in fact, improve multi-protein assembly relative to observed copy numbers, even for the nonequilibrium clathrin-coat assembly module. We speculate that the observed imbalances in clathrin adaptor proteins could offer a mechanism for making the vesicle formation process more tunable, since adaptor proteins are responsible for selecting cargo for endocytic uptake, which is the ultimate purpose of CME.

## Results

### Stoichiometric balance is measureable in large PPINs when interfaces are resolved

For a multi-subunit complex such as the ribosome or ARP2/3 complex (Fig 1B), all subunits bind together non-competitively to assemble a functional complex. Stoichiometric balance is simply having enough of each subunit to form complete complexes, with no subunit in excess. But quantifying balance in a general protein-protein interaction network is more challenging because some proteins will bind competitively, using the same interface for multiple interactions. Such proteins will need a higher concentration in order to saturate their functional partners (Fig 2). Thus, to establish stoichiometric balance in a PPIN the binding interfaces must be known. In previous work we analyzed several interface-resolved PPINs, including the

56-protein clathrin-mediated endocytosis (CME) network in yeast [20, 41] (Fig 1A), and the 127-protein ErbB signaling network in human cells[16].

To balance a network, a number of desired complexes may be assigned to each edge and then the number of required interface copies directly solved for. This is constrained with a starting set of copy numbers, $C_0$, otherwise the solution would be arbitrary. However, the inclusion of multiple interfaces per protein introduces a new constraint: interfaces on the same protein should have the same copy number. This constraint often makes nontrivial solutions (i.e. when none of the proteins are set to zero) impossible (see Methods). Therefore, we treat it as a soft constraint, using a parameter "α" to balance its influence. A high α allows more variation of interface copy numbers on the same protein (Fig 2C). We constructed and minimized an objective function using quadratic programming (Methods), which produces a new, optimally balanced set of copy numbers, $C_{balanced}$. For any given interface-resolved PPIN, there can be multiple locally optimized solutions of balanced copy numbers. In Fig 2C we illustrate solutions found by our algorithm SBOPN using the copy numbers of Fig 2B as $C_0$. If we apply our algorithm to Fig 2A, which is an already balanced network, it simply recovers the input copy numbers, such that $C_{balanced} = C_0$, regardless of α. Because our algorithm minimizes distance from $C_0$ to $C_{balanced}$, the optimal solutions produce both under and overexpressed proteins.

The benefit of this method is that the distance between $C_0$ and $C_{balanced}$ gives you a relative estimate of how "balanced" $C_0$ already is, and thus a metric from which to evaluate the significance of balance in the observed copy numbers. Using real copy numbers taken from Kulak et al.[2], $C_{real}$, as $C_0$, we calculated both chi-square distance (CSD) and Jensen-Shannon distance (JSD) between $C_{real}$ and $C_{balanced}$ (Methods). The former metric looks at differences between absolute values and penalizes high deviations more strongly than low deviations, whereas the latter converts both vectors to distributions and measures the similarity between them. We do not expect any networks to have $C_{real}$ that is already perfectly optimized, such that $C_{real} = C_{balanced}$. To establish the significance of both distance metrics, we generated 5,000 sets of random $C_0$ vectors, sampled from a yeast concentration distribution. We then measured the CSD and JSD from $C_0$ to $C_{balanced}$ for each of these random copy number vectors. If $C_{real}$ is balanced, its distance metrics should have a significant p-value relative to yeast copy numbers selected randomly from the yeast distribution. The C++ code for our SBOPN algorithm and example input and output files may be downloaded at https://github.com/mjohn218/StoichiometricBalance.

**Accounting for noise in observed copy number measurements.** Even constitutively expressed genes do not have a constant abundance; they vary due to both extrinsic and intrinsic noise [42]. Taniguchi et al. found that the abundance of a single protein in *E. coli* follows a gamma distribution [12]. Therefore, one reason copy number balance should not be expected to be perfectly matched is due to inherent fluctuations in protein copy numbers. Our algorithm, however, ultimately assigns a single copy number to each interface in the network to optimize perfect balance, when realistically a range of values would be more appropriate.

Our method does provide one mechanism to allow a range of copy number values for a single protein, and that is through allowing each interface on a single protein to have distinct values. This range can be tuned through our parameter α, which biases solutions towards equivalent interface copies per protein when set to zero. As the α parameter increases, more variation is observed (Fig 2C). For example, one interface may be assigned 200 copies and another on the same protein 300 copies. If the protein is usually expressed within the 150–350 copy range, this solution is more realistic than enforcing both copy numbers to be exactly 250.

We therefore systematically characterized how variations in α changed the "noise", or variability in interface copy numbers on each protein. Taniguchi et al. found that yeast proteins

with high abundance (~1,000 or more copies) had a noise ($\sigma^2/\mu^2$) upper limit of about 0.5 with ungated data and 0.1 with gated data[12]. For $\alpha \leq 0.03$, we found that proteins with mean interface copy numbers above 1,000 had less than 0.1 noise, indicating that such a solution is possible. (S1 Fig). Low abundance proteins exhibit higher noise in terms of expression level[12, 43], and this feature is also observed in our model. We therefore used values of $\alpha$ in the 0.01 to 2 range based on this analysis (S1 Fig).

**Protein copy numbers in yeast clathrin-mediated endocytosis are balanced.** As Fig 3A and 3B shows, at $\alpha = 1$ the p-value for JSD was found to be statistically significant ($p = 0.0054$) but the p-value for chi-square distance was not ($p = 0.157$). We analyzed the real copy numbers (S2 Table) before and after balancing and found that the protein cofilin was highly overexpressed (Fig 3C) meaning that it had to be greatly lowered to achieve balance. This resulted in a skewed CSD for $C_{real}$, which the change in cofilin dominated. We therefore re-tested the degree of balance when cofilin was removed from the network. At $\alpha = 1$, both JSD ($p = 0.0012$) and CSD ($p = 0.022$) were statistically significant (Fig 3D), indicating that these 55 proteins are balanced compared to random copy numbers. These results were robust to changes in $\alpha$, but the p-values tended to be lowest when $\alpha$ was in the 0.01 to 2 range. The absolute distance from $C_{real}$ to $C_{balanced}$ lowered as $\alpha$ was raised, plateauing when $\alpha \geq 10$.

Because protein complexes that strongly bind are thought to be more balanced than weak interactions, we repeated the analysis on the full 56-protein network after removing one of two modules from the network: the four protein subunits of the AP complex, and the seven proteins in the ARP2/3 complex. Without the former, the p-value increased to 0.0088 for JSD and 0.197 for CSD, indicating less overall balance. Removing only the ARP2/3 complex similarly raised the p-values to 0.023 and 0.24. This trend held when cofilin was also removed.

The four AP subunits that form the obligate AP-2 complex are fairly close in abundance, as are the clathrin heavy chain and clathrin light chain proteins, which is consistent with the pressure for strong binding proteins to be more tightly balanced.

**Stoichiometric balance is not measured without proper interface binding interactions.** To test whether balance depended mostly on protein network structure rather than the child interface interaction network (IIN) structure, we ran this analysis again using random IINs for the same parent protein network, again excluding cofilin. In other words, we randomized whether proteins bind competitively or noncompetitively, using a rewiring method from Holland et al.[41]. For 20 random IINs, we found that the real copy numbers were significantly less balanced. For $\alpha = 1$, the same analysis obtained p-values of $0.44 \pm 0.12$ for CSD and $0.24 \pm 0.13$ for JSD. Thus, the protein copy numbers are balanced according to the underlying interface network.

**Observed protein imbalances can highlight functional relationships.** Finally, by looking at the relative change between $C_{real}$ and $C_{balanced}$, we could examine which proteins are underexpressed in the network. We note that because our method minimizes the distance from $C_{real}$ to $C_{balanced}$, the optimal solution has comparable number of over and underexpressed proteins. As Fig 3E shows, the five most underexpressed proteins are PRK1 (by a factor of nearly 40), ARK1, AIP1, APP1, and YSC84. PRK1 and ARK1 are both kinases; they form transient interactions with their partners for the purpose of phosphorylation. Since a single kinase can phosphorylate many proteins relatively quickly, rather than form stable complexes with each target, there is a sensible functional explanation for why these proteins can be underexpressed relative to their partners by such a large margin. Similarly, APP1 is a phosphatase. The protein AIP1 is an actin binding protein that targets a binding surface of actin without any competition from other actin binders, and also binds the highly expressed cofilin. Its low abundance relative to actin and cofilin could indicate it acts as a bottleneck in regulating cofilin-actin interactions, or perhaps more simply, that functionally it is not needed at a 1:1 stoichiometry with the

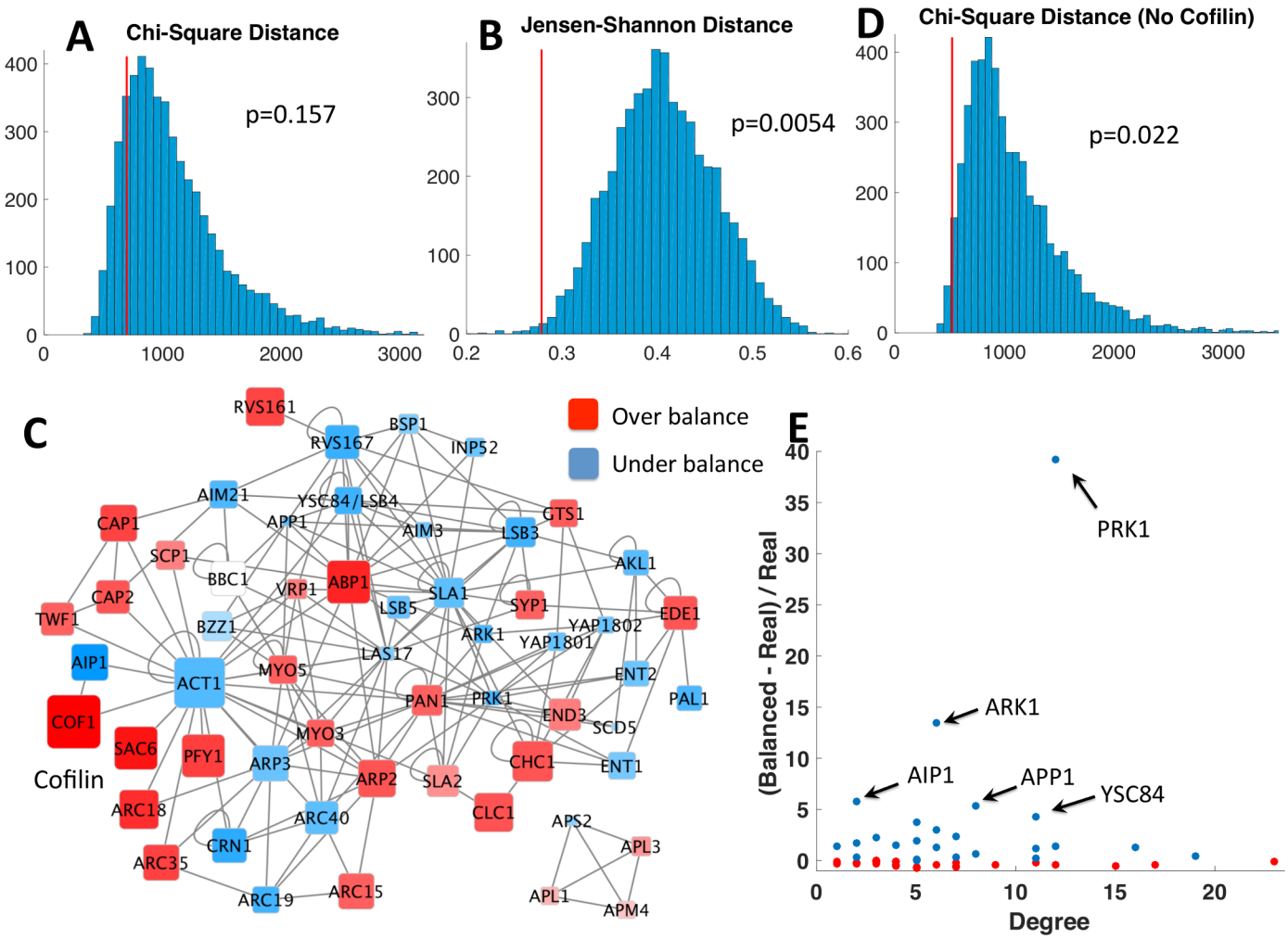



**Fig 3. Clathrin-mediated endocytosis proteins are balanced. (A,B)** Histograms for chi-square distance and Jensen-Shannon distance between the real protein copy numbers and their copy numbers after balancing. Compared to 5,000 sets of random sampled copy numbers, the real copy numbers had a statistically significant Jensen-Shannon distance, but not chi-square distance. **(C)** Graph of CME network, showing which proteins were overexpressed (red) or underexpressed (blue) compared to the balanced copy numbers. Cofilin was highly overexpressed, which led to a high chi-square distance. **(D)** Histogram for chi-square distance when cofilin was removed from the network. It is now statistically significant, indicating that the other 55 proteins are balanced compared to random copy numbers. **(E)** The five most underexpressed proteins were two kinases (PRK1 and ARK1), one phosphatase (APP1), and two partners of Actin (AIP1 and YSC84). The former three bind transiently to their partners, so there is no functional need for them to be balanced. The latter two are discussed in the text.

https://doi.org/10.1371/journal.pcbi.1006022.g003

ubiquitous actin protein. YSC84 has 13 binding partners, and 10 of these partners all bind the YSC84 SH3 domain, including the relatively highly expressed ABP1. Although many of these binding partners (all proline rich domains-PRDs) also have additional partners of their own, ABP1's PRD is specific to YSC84's SH3 domain[41]. As we return to in the discussion, under-expression could indicate a functional regulatory role for this protein, or indicate transient interactions with partners. Identifying underexpressed proteins and which of their interface binding partners apply pressure to increase copy numbers is a useful first step in hypothesizing about the temporal dynamics of such proteins within the cell.

Actin is overexpressed compared to its partners, excluding cofilin, which can be likely attributed to its primary role as a central component of the cell cytoskeleton. Clathrin, another protein that polymerizes, is also overexpressed, the reasons for which are investigated in part 3, "Beyond misinteractions: Multi-protein functional assemblies are sensitive to stoichiometric

balance". Cofilin's high expression is the most imbalanced, and Kulak et al. also found it highly expressed in HeLa cells and *S. pombe* [2]. The protein acts to sever actin filaments, without which the cytoskeleton cannot reorganize[44] and cells cannot migrate[45]. Highly expressed proteins will enable faster complex formation, so one possible advantage of its high abundance is making rapid reorganization of the cytoskeleton possible.

**Ras and MAP3K proteins in the ErbB signaling network are underexpressed.** We applied our algorithm SBOPN to another IIN from the literature: that of the 127 protein human ErbB signaling network, characterized by Kiel et al.[16]. Our algorithm optimizes copy numbers to the full network structure even if not all individual target copy numbers are available. Thus, we measured the distance between the real ($C_{real}$) and optimized ($C_{balanced}$) copy numbers for the 115 of the 127 proteins for which we could assign expression levels from HeLa cells (Methods, S2 Table). We compared results to copy numbers randomly sampled from a HeLa protein concentration distribution.

Because this is a signaling network where the majority of interactions are phosphorylation, we expected these transient interactions to bias the copy numbers against significant balance. However, while JSD was not found to be significant ($p = 0.274$), CSD was ($p = 0.022$). This result held when copy numbers were shuffled rather than randomly sampled (JSD: $p = 0.120$, CSD: $p = 0.019$). As stated above, CSD is dominated by large deviations. Thus, while the network as a whole is not balanced, there appears to be no dramatic overexpression.

The three Ras proteins (HRAS, NRAS, and KRAS) were found to be underexpressed (S2 Fig), confirming the findings of Kiel et al. using simpler comparisons of Ras copy numbers to all binding partners [16]. Also found to be underexpressed were all five MAP3K proteins (RAF1, MAP3K1, MAP3K11, MAP3K2, and MAP3K4) in the network. MAP3K proteins are the top layer in MAPK cascades, a signaling motif consisting of three proteins (a MAP3K, MAP2K, and MAPK) occasionally bound together via a scaffold protein[46]. The membrane bound receptors ErbB2 and ErbB3 were similarly underexpressed. These results suggest strategic underexpression of certain upstream proteins, potentially to control specific outputs from diverse inputs[16], and to amplify a signal as it travels "downstream" in a signaling network. Underexpression of upstream proteins is not a universal rule, however, and may depend on the type of interaction and the dynamics of the signaling network.

## Imbalance increases misinteractions dependent on the network topology and binding affinities of proteins

In this second part, we investigate how the cost of imbalance, measured solely in terms of misinteractions, depends on general properties of proteins, including binding affinity and number of binary partners. In a stoichiometrically balanced network, proteins will be driven to saturate their stronger-binding functional partners. Any "leftover" proteins, however, may misinteract, or form non-functional complexes that, while weak, are combinatorically numerous.

**Misinteractions are minimized under balanced copy numbers and are largely independent of network motif structure.** Complex formation and misinteractions must be evaluated at the level of individual protein binding interfaces, and we thus study small network motifs that have been previously characterized in real biological interface interaction networks (IINs) to control binding specificity [41]. Of these five motifs (Fig 4A), the hub and square motif are the most common in biological IINs relative to random networks [41]. The chain, triangle, and flag motif are selected against due to the challenges in optimizing such binding interfaces for strong selective binding and against misinteractions.[23, 27, 41] The motif defines the functional or "specific" interactions, which we allow at equal binding strengths. However, all other possible protein-protein interactions were allowed as misinteractions, which occur at weaker

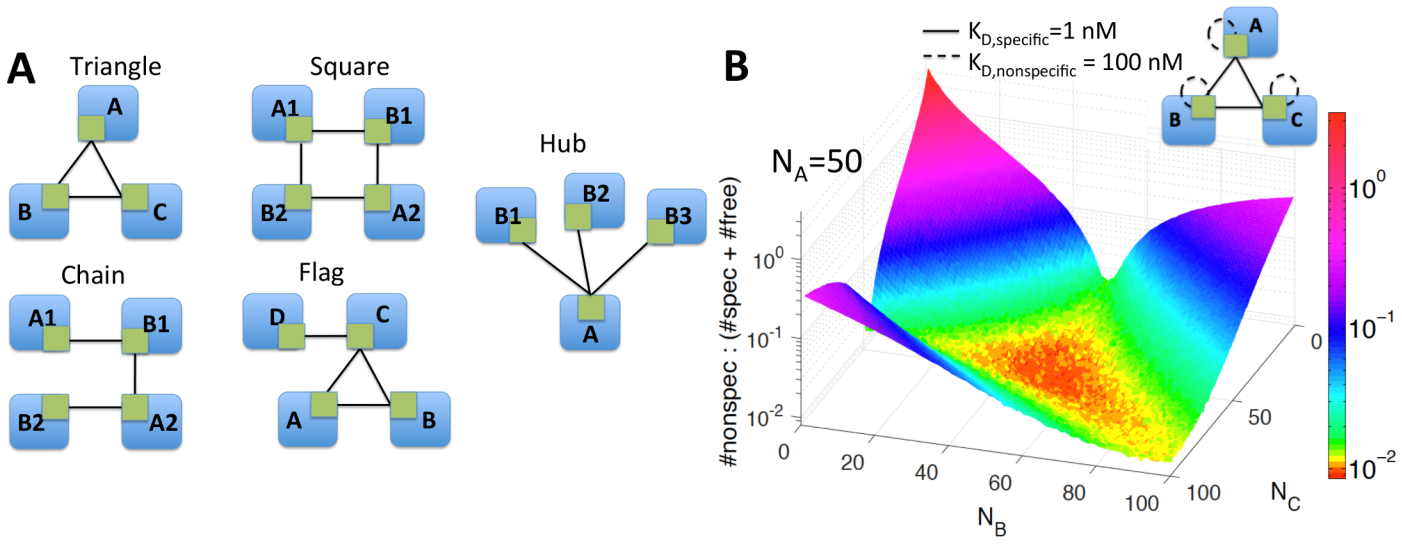



**Fig 4. Misinteractions in network motifs from biological IINs. (A)** Five network motifs that have been shown to impact specificity of binding in biological IINs were tested for the effects of imbalance on misinteractions. **(B)** Surface plot obtained for the triangle network. The z-axis is the frequency of misinteractions at steady-state (Eq 1) averaged across 1000 runs. The x and y axes are the number of B and C proteins; the number of A proteins is fixed at 50. As one protein becomes overexpressed, misinteractions increase exponentially.

https://doi.org/10.1371/journal.pcbi.1006022.g004

strength than the specific interactions. Because each node represents an interface (each on its own protein in this case), all binding was competitive.

Balanced copy numbers are relatively easy to design for these simple network motifs, and the optimization of the first part,"Stoichiometric balance is measureable in large PPINs when interfaces are resolved", is not necessary. We study imbalanced copy numbers by simply varying the copy numbers of two proteins in each network over a wide range while keeping the remaining proteins constant. For each set of copy numbers, we ran the system to equilibrium using the Gillespie algorithm[47]. We could then measure the total number of specific and non-specific complexes formed ($N_{specific}$, $N_{nonspecific}$), as well as unbound proteins ($N_{free}$), and use this to evaluate the cost of being out-of-balance in terms of misinteraction frequency:

$$Cost(C_0) = \frac{N_{nonspecific}(C_0)}{N_{specific}(C_0) + N_{free}(C_0)} \quad (1)$$

averaged across 1,000 runs, where $C_0$ is the vector of initial copy numbers.

The frequency of misinteractions is lowest when the protein copy numbers are balanced. Fig 4B shows the results for the triangle network. For example, when all three proteins have equal abundance of 50 copies, about 25 of each specific complex are formed, and minimal proteins are leftover. Cost also remains low when two proteins are equally overexpressed, as these excess proteins can bind to each other. The instances where misinteractions are the most frequent are when one protein is overexpressed, as this protein has no specific partners left and thus will self-bind: a misinteraction for this motif. Similar surface plots were obtained for all five network motifs (S3 Fig).

Notably, with balanced copy numbers, the frequency of misinteractions is almost entirely dependent on the relative strength, or energy gap, between specific and nonspecific binding (Fig 5A) and there was little difference among the five networks. The slope varies slightly from one motif to another, and we confirmed that this can be calculated relatively accurately based on the ratio of specific versus non-specific interactions possible for that motif. Furthermore,

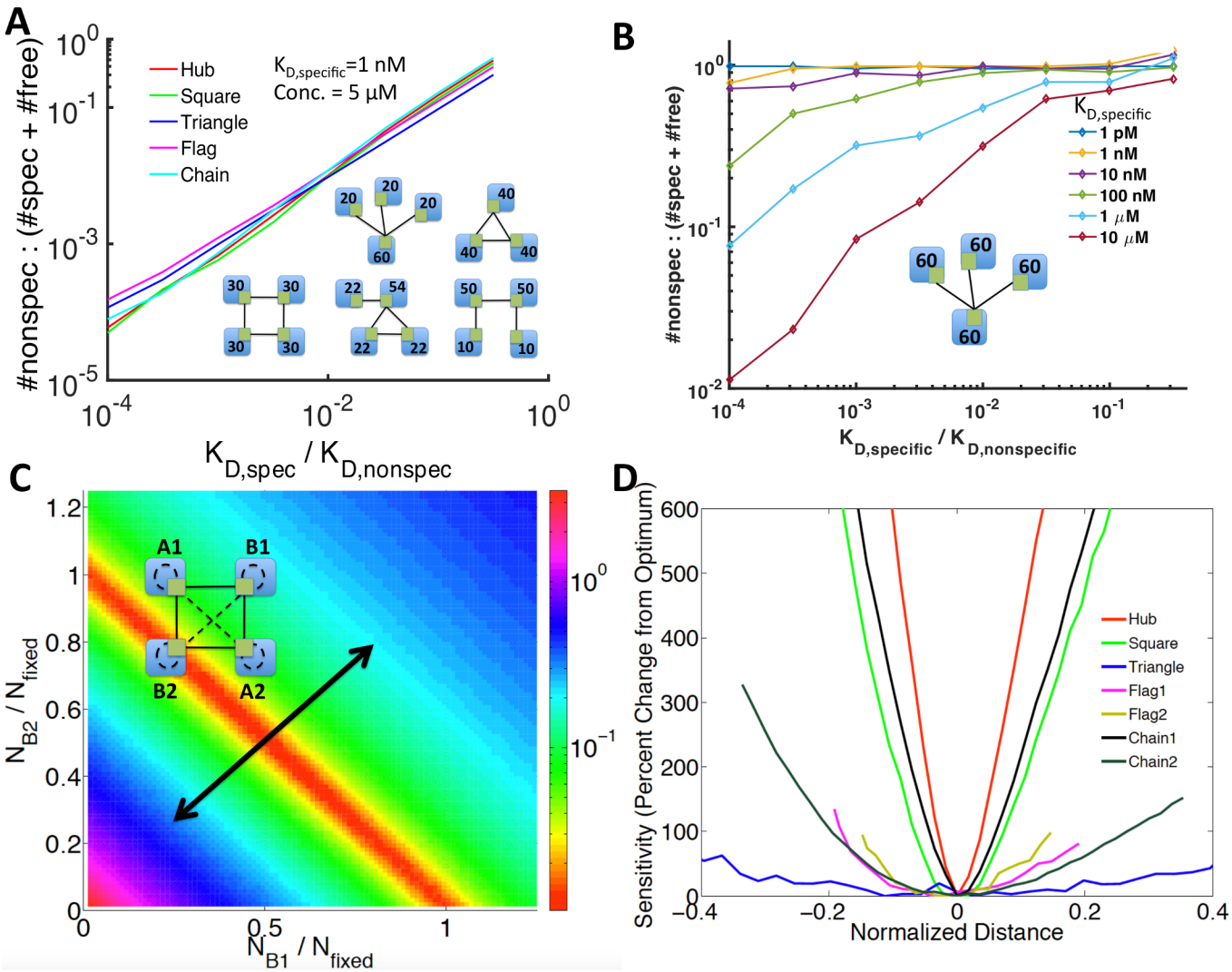



**Fig 5. Misinteractions are motif dependent only when concentrations are imbalanced. (A)** At balanced concentrations, misinteraction frequency increased linearly with the ratio of $K_{D,specific}$ to $K_{D,\ nonspecific}$. It was also roughly equal for all five network motifs. **(B)** At unbalanced concentrations, misinteractions can occur even at a large energy gap (low K_D ratio), unless the overall binding is weak (i.e. red curve). **(C)** Surface plot for the square network, measuring the ratio of (#nonspecific complexes: #specific complexes + free proteins) when A1 and A2 are fixed while B1 and B2 are varied. The principal component (black line) is shown across the region of lowest misinteraction frequency. **(D)** Cost sensitivity to concentration imbalance varies significantly between motifs. The "distance" is measured along the principal component of the surface plots as you move away from the optimal region. Two different pairs of fixed proteins were analyzed for the chain and flag networks. The hub and square networks were the most sensitive to imbalance, while the flag and triangle were the least.

https://doi.org/10.1371/journal.pcbi.1006022.g005

the results were similar when we varied the absolute strength of specific binding from 1nM; under balanced conditions it affects the number of free proteins ($N_{free}$) relative to total complexes formed. Thus, under balanced copy numbers, the cost of misinteractions is not strongly dependent on specific binding affinities.

**Misinteractions for imbalanced copy-numbers are worse for biologically common motifs and strong binding proteins.** Unlike the similar cost of misinteractions under balanced copy numbers, the five networks noticeably differ in sensitivity to imbalanced copy numbers. In general, as copy numbers become more imbalanced, the misinteraction cost grows. To quantify this rate for each network motif, we measured the percent change in cost as one travels along the principal components away from the balanced copy numbers (Fig 5C; S3

Fig). The hub and square motifs were found to be the most sensitive, showing a rapid increase in cost of misinteractions as imbalance grows, whereas the flag and triangle motifs were found to be the least. (Fig 5D). The triangle motif has the least sensitivity and it also has the fewest misinteractions possible; it can form 3 specific complexes and only 3 misinteracting complexes. The robustness of this module also then extends to the flag motif, which contains a triangle.

The motifs most sensitive to imbalance, the hub and square motif, are also the motifs most common in biological networks [27, 41]. In previous work, we demonstrated that these motifs are evolutionarily selected for in biological networks because binding interfaces that interact through these specific motifs are much easier to simultaneously design for high specificity (strong $K_{D,specific}$) and for weak nonfunctional interactions (weak $K_{D,nonspecific}$) [27, 41]. Although these motifs thus produce more selective binding interfaces, our results show that there is more pressure to maintain copy number balance in these biologically common motifs to prevent misinteractions.

Importantly, unlike the results for balanced copy numbers, strong binding proteins are highly prone to misinteractions under imbalanced conditions (Fig 5B). Weak-binding proteins form minimal complexes overall, and thus imbalances in copy numbers do not strongly influence their binding patterns. Strong binding proteins, on the other hand, are driven to bind to any unbound interface, even when the gap separating specific and non-specific binding is large. This is because although the nonspecific binding affinities are orders of magnitude weaker than the specific binding affinity, for a strong binder ($K_D$ = 1nM), the nonspecific interactions will be strong enough ($K_D$~10μM) to form stable complexes (Fig 5). The number of possible misinteracting partners is also approximately given by the total number of interfaces in the cell. Thus, leftover copies of these proteins frequently misinteract. This supports the observations that strong binding proteins should be tightly regulated to maintain stoichiometric balance[19], and therefore avoid misinteractions. For weak binding proteins, on the other hand, misinteraction cost is not a significant pressure favoring copy number balance.

**Larger networks with biological topologies produce more misinteractions under copy number imbalance.** Our analysis of network motifs above demonstrated that topologies common in biological IINs are actually more prone to misinteractions when copy numbers are imbalanced. We find here that the same trend applies to much larger networks that again exhibit biological topologies (Fig 6). To show this, we analyzed 500 IINs that differed in three properties: motif frequencies; degree distribution; and density, which was determined by the size of the network (90–200 proteins for 150 edges). The biological-like IINs have motif frequencies biased to hub and square motifs, they have a degree distribution that is power-law like or "scale-free", meaning, broadly speaking, that a few "hub" proteins have many connections while the majority are specialized for a few interactions, and they tend to be sparse; interfaces in the CME IIN have an average degree of only 2.06[41]. For simplicity, here we will assume each interface is on its own protein, such that the PPIN is the same as the IIN. Balanced copy numbers are assigned to each network using our optimization method described above based on network structure (also see Methods), and imbalanced copy numbers are defined by randomly sampling copy numbers from the yeast distribution. Specific and non-specific $K_d$ values for each possible binding interaction were initially taken from a previous study[27], where the gap between specific and non-specific binding was optimized based on selecting amino-acid sequences for each interface [27].

As expected, when copy numbers are balanced rather than imbalanced via random assignments, all networks produced fewer misinteractions. The networks that, under balanced copy numbers, produced the fewest misinteractions were the networks most like biological IINs: they were sparse networks and they had optimized topologies favoring square and hub motifs

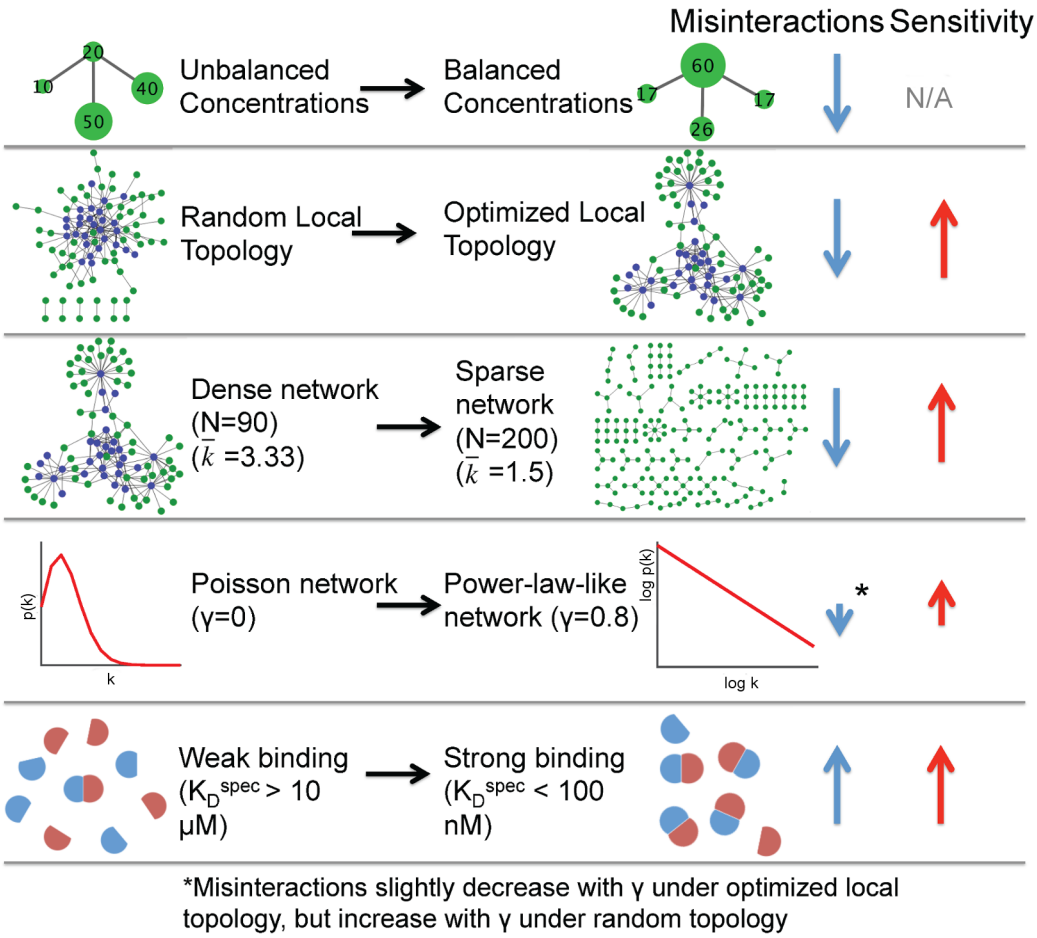



**Fig 6. Biological IIN topologies have more misinteractions under imbalance.** Shown are trends in misinteraction frequency under balanced concentrations (blue arrows) and sensitivity to imbalance (red arrows). Several features that make networks perform better under balanced concentrations make them perform worse under unbalanced concentrations: sparseness, a topology that matches with real interface networks, and a power-law degree distribution. Strong average binding caused both increased misinteractions and increased sensitivity.

https://doi.org/10.1371/journal.pcbi.1006022.g006

(Fig 6). Because these IINs also had larger energy gaps separating $K_{D,Specific}$ and $K_{D,Nonspecific}$ [27], we verified that when all networks were assigned the same $K_{D,Specific}$ and $K_{D,Nonspecific}$ (1000-fold different), the biological IINs indeed produced fewer misinteractions under balanced copy numbers (S4 Fig), although the difference was relatively small. Hence, overall, the results are similar to the findings with motifs, that for balanced copy numbers, misinteractions are not strongly influenced by network structure.

Once copy numbers were imbalanced, however, the biological-like IINs produced a sharper increase in misinteractions (higher sensitivity-Fig 6). This is consistent with the trends from the previous subsection, where the biological motifs of hub and square motifs were also more sensitive to imbalance. Sparse networks are more sensitive to imbalance because they have more interfaces (N) that can possibly misinteract (order $N^2$). The only network feature that did not have a significant trend in controlling misinteractions either for balanced or unbalanced copy numbers was the degree-distribution. For power-law network topologies compared to Poisson networks, misinteractions could be higher or lower depending on the local motifs or the network sparseness (Fig 6; S4 Fig). Thus, local topology and density was more important than the overall degree distribution.

Finally, because highly abundant proteins are thought to have low average affinity to avoid misinteractions, we increased the absolute strength of $K_{D,Specific,}$ while keeping the gap between $K_{D,Specific}$ and $K_{D,Nonspecific}$ the same. Stronger affinity did indeed lead to both more nonspecific complexes and higher sensitivity to copy number imbalance. This result is consistent with the previous subsection and confirms that strong binding affinities can be paradoxically deleterious to specific complex formation.

## Beyond misinteractions: Multi-protein functional assemblies are sensitive to stoichiometric balance

In the second part, "Imbalance increases misinteractions dependent on the network topology and binding affinities of proteins", we only studied binary, competitive interactions. But proteins often bind noncompetitively into higher complexes, and they may interact weakly and thus form few complexes, in which case imbalance may have functional benefits [17, 18]. Furthermore, the above models looked at equilibrium results, whereas many biological systems exhibit non-equilibrium dynamics. We created kinetic models of two modules from the CME network with observed imbalances: the ARP2/3 complex and a simplified vesicle forming protein subset. Simulating higher complex formation is challenging because of the exponentially large number of possible species, so we used NFSim[48], a stochastic solver of chemical kinetics that is rule-based, enabling an efficient tracking of higher-order complexes as they appear in time.

**The ARP2/3 complex has higher yield under stoichiometric balance.** One unexpected imbalance we found was that of the isolated, 7-component ARP2/3 complex. The complex has one highly underexpressed subunit, ARC19. ARC19 is a core subunit, binding to five other subunits (Fig 1B). Because of this, it is more likely to form misinteractions (due to its five interfaces) and be a part of incorrect complexes (e.g. complexes of the form ARC19 –ARC40 –ARP2 –ARC19 are incorrect because they contain two ARP19 proteins). Therefore, we tested whether the observed copy numbers might improve formation of complete ARP2/3 complexes.

Ultimately, we found that balanced copy numbers always improved formation of complete ARP2/3 complexes relative to the observed copy numbers, whether or not misinteractions were modeled (S5 Fig). We simulated simplified complex assembly using arbitrary rate constants and two sets of copy numbers: those observed from Kulak et al. and stoichiometrically balanced (in this case equal) copy numbers for each subunit. We measured "yield" as the number of proteins in full complexes divided by the number of proteins in all complexes, including misassembled or incomplete. Some cooperatively was allowed in that if three proteins in a trimer were held together by two binding events, the third binding event could occur at a faster rate (due to all three subunits being localized together). Binding to the core subunit ARC19 was also set to be 10-fold stronger than peripheral bindings, as this increased yield. But no matter what parameter ranges we used, we could not increase the yield of the Kulak copy numbers (max ~13%) versus the balanced copy numbers (max ~50%). Because ARC19 has ~5-fold underexpression compared to the other 6 subunits, incomplete complexes dominate. The results held when we also allowed ARC19 to form misinteractions.

Imbalances in copy numbers have been shown to actually improve the yield for self-assembly, but the optimal copy numbers must take on specific ratios of components to optimize yield [17]. Here, we see that the ARP2/3 subunits do not exhibit optimal expression for yield in our model. One possible explanation is that the ARC19 subunit has distinct thermodynamics or kinetics that are critical for controlling assembly. This would suggest that this subunit has conserved expression across all organisms. However, this is not the case. We compared the

expression levels of the seven subunits with data from three other studies. Two also found ARC19 to be underexpressed[49, 50], whereas one[1] found it to be overexpressed. However, Chong et al. also found ARP2 to be underexpressed, whereas Kulak et al. found it to be overexpressed. We also compared the abundance of human homologs from five studies[2, 19, 51–53] and found similar issues with noise, though only one found ARC19's homolog to be underexpressed. (S5 Fig) Thus, no conservation of subunit expression levels is observed. Without a more structurally and biochemically accurate model for the ARP2/3 components, it is difficult to assess whether the low expression of ARC19 does provide some benefit in assembly yield. As we return to in the discussion, several other factors may explain the imbalance, such as noise in expression levels or in measurements of expression levels, or additional roles in the cell for some ARP2/3 subunits.

**A simplified clathrin-coated vesicle forming module enables a kinetic study of imbalance effects on non-equilibrium assembly.** For our final analysis, we test the effects of copy number balance on a more complex, non-equilibrium model of clathrin-coat assembly for vesicle formation. Our minimal model for vesicle formation includes nine cytoplasmic proteins plus the plasma membrane lipid recruiter $PI(4,5)P_2$, with the biochemical parameters taken from the literature for all known binding interface interactions (Fig 7; Table 1). In clathrin-mediated endocytosis, clathrin triskelia consisting of three heavy chains (CHC1) and three lights chains (CLC1) are recruited to the membrane via adaptor proteins that bind lipids (ENT1 & 2, SYP1, SLA2, YAP1801) and in some cases also transmembrane cargo (ENT1 & 2, YAP1801). Clathrin polymerize to form a hexagonal clathrin cage of ~100 triskelia [54] that helps deform the plasma membrane into spherical membrane vesicles of ~100 nm in diameter. Additional non-membrane-binding scaffold proteins help stabilize the assembly (EDE1, YAP1802). Importantly, the assemblies do not have to exhibit a perfect stoichiometry of components, unlike the ARP2/3 complex, in order to function, with variable compositions shown to produce clathrin-coated structures *in vitro*[35, 55, 56]. To measure vesicle formation in our model, we therefore make the assumption that completed vesicles contain 100 triskelia [54] in a complex on the membrane. Once a completed model vesicle is formed, all components that are a part of this complex are recycled, unbound, back to the cytoplasm, keeping total protein concentrations fixed.

We emphasize that this minimal model is based on the known concentrations and binding properties of the component proteins, and thus we are not attempting to optimize the model to best describe *in vivo* observations. Furthermore, this kinetic model does not account for biomechanics of the membrane budding or coupling to the cytoskeleton, or molecular structure, which are important features of CME. As we see in our simulations, our vesicles form ~10 times faster than vesicle formation *in vivo*. However, clathrin-coated vesicles (pre-scission) are observed to assemble *in vitro* with minimal components, without the cytoskeleton or any energy sources[35, 56]. We thus included in our model all proteins from the larger CME network (Fig 1) that directly connect clathrin coat assembly to the membrane surface, linking the assembly process with the ultimate endocytic goal of transmembrane receptor and cargo uptake. Our model thus represents a useful qualitative framework to assess how stoichiometric balance in clathrin-coat components can impact vesicle formation and thus cargo uptake.

An important feature that our model does capture is the reduction in dimensionality (3D to 2D) which accompanies binding to the membrane surface [59]. Once localized to the membrane via either lipid binding or recruitment by other proteins, proteins are concentrated in units of $Area^{-1}$, with binding constants of $K_d^{2D} = K_d^{3D}/(2\sigma)$, where $\sigma$ is a lengthscale in the nanometer range[73], as discussed in Ref. [59]. Transitioning to the membrane can drive dramatic increases in complex formation due to higher effective concentrations of components [59]. In our simulations here, we find that this is a critical factor controlling vesicle formation.

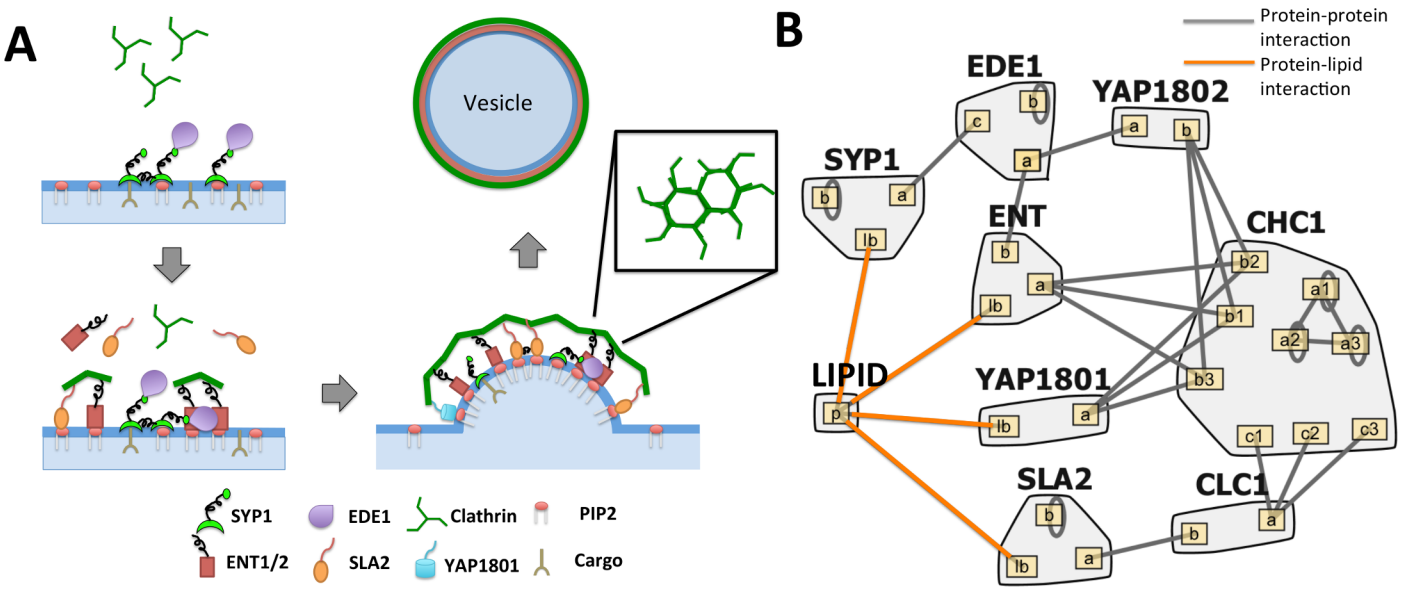



**Fig 7. Clathrin membrane recruitment model. (A)** In clathrin-mediated endocytosis, adaptor proteins bind to the lipid membrane and recruit clathrin triskelia to the surface. These triskelia assemble a hexagonal cage around the plasma membrane vesicle. **(B)** Binding model of the clathrin module. Included are seven adaptor or accessory proteins (SYP1, EDE1, YAP1801/2, ENT1/2, and SLA2), clathrin heavy chains already assumed to be in trimer form, and clathrin light chains. Five of the adaptor/accessory proteins can bind directly to the lipid membrane. Picture generated with Rulebender.

https://doi.org/10.1371/journal.pcbi.1006022.g007

Besides this division between the cytoplasm and the membrane surface, there is no other spatial resolution. A full list of model assumptions can be found in the S1 Text.

**Adaptor proteins are underexpressed and can tune vesicle formation.** We first evaluated whether this nine-protein module (Fig 7) was significantly balanced using SBOPN. The clathrin heavy chains and light chains are close in expression, as expected since these two have a strong binding affinity (~ 1nM)[66]. But clathrin was overexpressed compared to its adaptor proteins by over 3-fold. Functionally, a full triskelia has up to six binding sites for adaptor proteins, but only one needs to be bound to localize it to the membrane. Hence, it is not strictly necessary for the adaptor proteins to be balanced. However, we found that when balanced copy numbers were used instead of observed copy numbers, vesicles formed faster and with fewer components (Fig 8A) Thus the biological copy numbers do not appear optimized for maximum vesicle formation, though they are sufficient to drive vesicle formation.

Our model assumes these proteins are well-mixed throughout the cytosol, but cells can spatially regulate proteins, altering the local concentration. We simulate this by altering the expression of the adaptor proteins in our model. Knocking out either SLA2 or ENT1/2 pushes the copy numbers even further out-of-balance, and nearly halts vesicle formation (Fig 8C and 8D). Increasing their expression increases vesicle formation because they are below saturation. Decreasing the other adaptor or scaffold proteins also increases imbalance and has a negative effect on the speed of vesicles, although it is less severe. Clathrin-coat assembly is quite sensitive to these membrane-binding protein concentrations because they not only recruit clathrin to the membrane, but they stabilize the triskelion in 2D, where they can then exploit reduced dimensionality to drive binding [59]. If clathrin polymerized effectively in solution, far fewer adaptor proteins would be needed to link large clathrin-cages to the membrane surface. We speculate that this sensitivity to the membrane-binding adaptor proteins and their observed underexpression could allow the cell to better tune productive vesicle formation to occur only when enough cargo is localized [74]. The adaptor proteins ultimately localize the cargo bound

**Table 1. Parameters for clathrin membrane recruitment model.** See S1 Text for further notes.

| Parameter | Description | Value | Source(s) |
|---|---|---|---|
| Vol_CP | Cytosol volume | 37.2 $\mu m^3$ | Jorgensen *Science* 2002[57]; Alberts *Molecular Biology of the Cell 6th Ed.* 2015[58] |
| SA_PM | Plasma membrane surface area | 75.7 $\mu m^2$ | Jorgensen *Science* 2002[57] |
| σ | $K_{D,3D}$ to $K_{D,2D}$ conversion | 1 nm | Yogurtcu *PLoS Comp Biol* 2018 [59] |
| Kd_CHC_CHC | Clathrin heavy chain polymerization | 100 μM | Wakeham *EMBO J* 2003[60] |
| Kd_CHC_ENT | Clathrin heavy chain binding to ENT1/2 | 22 μM | Miele *Nat Struc Mol Biol* 2004[61] |
| Kd_CHC_YAP | Clathrin heavy chain binding to YAP1801/2 | 160 μM | Zhuo *J Mol Biol* 2010[62] |
| Kd_EDE_ENT | EDE1 to ENT1/2 binding | 12 μM | de Beer *Nat Struc Biol* 2000[63] |
| Kd_EDE_YAP | EDE1 to YAP1802 binding | 0.6 μM | Morgan *J Biol Chem* 2003[64] |
| Kd_EDE_EDE | EDE1 dimerization | 0.127 μM | Boeke *Mol Syst Biol* 2014[65] |
| Kd_CHC_CLC | Clathrin heavy chain to light chain binding | 0.1 nM | Winkler & Stanley *EMBO J* 1983[66] |
| Kd_CLC_SLA | Clathrin light chain to SLA2 binding | 22 μM | Engqvist-Goldstein *JCB* 2001[67]; Miele *Nat Struc Mol Biol* 2004[61] |
| Kd_SLA_SLA | SLA2 dimerization | 1 nM | Wilbur *J Biol Chem* 2008[68] |
| Kd_SYP_SYP | SYP1 dimerization | 2.5 μM | Henne *Structure* 2007[69] |
| Kd_SYP_EDE | SYP1 to EDE1 binding | 0.227 μM | Boeke *Mol Sys Biol* 2014[65] |
| Kd_L_ENT | ENT1/2 binding to lipid | 0.02 μM | Stahelin *J Biol Chem* 2003[70] |
| Kd_L_YAP | YAP1801 binding to lipid | 0.3 μM | Stahelin *J Biol Chem* 2003[70] |
| Kd_L_SLA | SLA2 binding to lipid | 0.2 μM | Stahelin *J Biol Chem* 2003[70] |
| Kd_L_SYP | SYP1 binding to lipid | 53 μM | Moravcevic *Structure* 2015[71] |
| k_off | | 1 $s^{-1}$ | |
| L_0 | Density of PtdIns(3,4)$P_2$ lipids | 25,292 lipids/$\mu m^2$ | Yoon *Nat Chem* 2011[72] |
| CHC1_0 | Total clathrin heavy chain trimers | 6426 | Kulak *Nat Methods* 2014[2] |
| CLC1_0 | Total clathrin light chains | 14538 | Kulak *Nat Methods* 2014[2] |
| EDE1_0 | EDE1 total proteins | 5964 | Kulak *Nat Methods* 2014[2] |
| ENT_0 | ENT1/2 total proteins | 3075 | Kulak *Nat Methods* 2014[2] |
| YAP1801_0 | YAP1801 total proteins | 357 | Kulak *Nat Methods* 2014[2] |
| YAP1802_0 | YAP1802 total proteins | 264 | Kulak *Nat Methods* 2014[2] |
| SLA2_0 | SLA2 total proteins | 3904 | Kulak *Nat Methods* 2014[2] |
| SYP1_0 | SYP1 total proteins | 2467 | Kulak *Nat Methods* 2014[2] |
| T_vesicle | Triskelia in a vesicle | 100 | McMahon & Boucrot *Nat Rev Mol Cell Biol* 2011[54] |
| k_dump | Rate of deletion for a complex of > = 100 triskelia | 1000 $s^{-1}$ | Arbitrarily high rate |
| k_recyc | Rate of protein recycling to the cytoplasm | 1000 $s^{-1}$ | Arbitrarily high rate |

https://doi.org/10.1371/journal.pcbi.1006022.t001

membrane receptors to clathrin-coated sites, a process called cargo loading[75, 76]. By increasing or decreasing the local concentration of adaptors, clathrin recruitment can be halted or sped up. With balanced copy numbers, the process is more stable to perturbations in copy numbers, and therefore less efficiently tuned.

Despite the underexpression of adaptor proteins, we observed a very high adaptor to triskelia ratio in completed vesicles (~19). A single triskelion can bind three SLA2 and three ENT1/2 proteins, which can bind three EDE1 and SYP1 proteins, leading to a seeming saturation of 12 adaptors per triskelion. However, most of these proteins can also dimerize with a strong affinity, allowing them to bind to other complexes of adaptor proteins. Our model lacks steric hindrance that would otherwise prevent this high level of aggregation, but nonetheless there is a clear gap in strength between adaptor protein interactions and clathrin interactions (Table 1). These weak clathrin interactions, particularly polymerization (~100 μM)[60], prevent spontaneous cage formation in the cytosol. It is the aggregation of adaptor proteins and localization

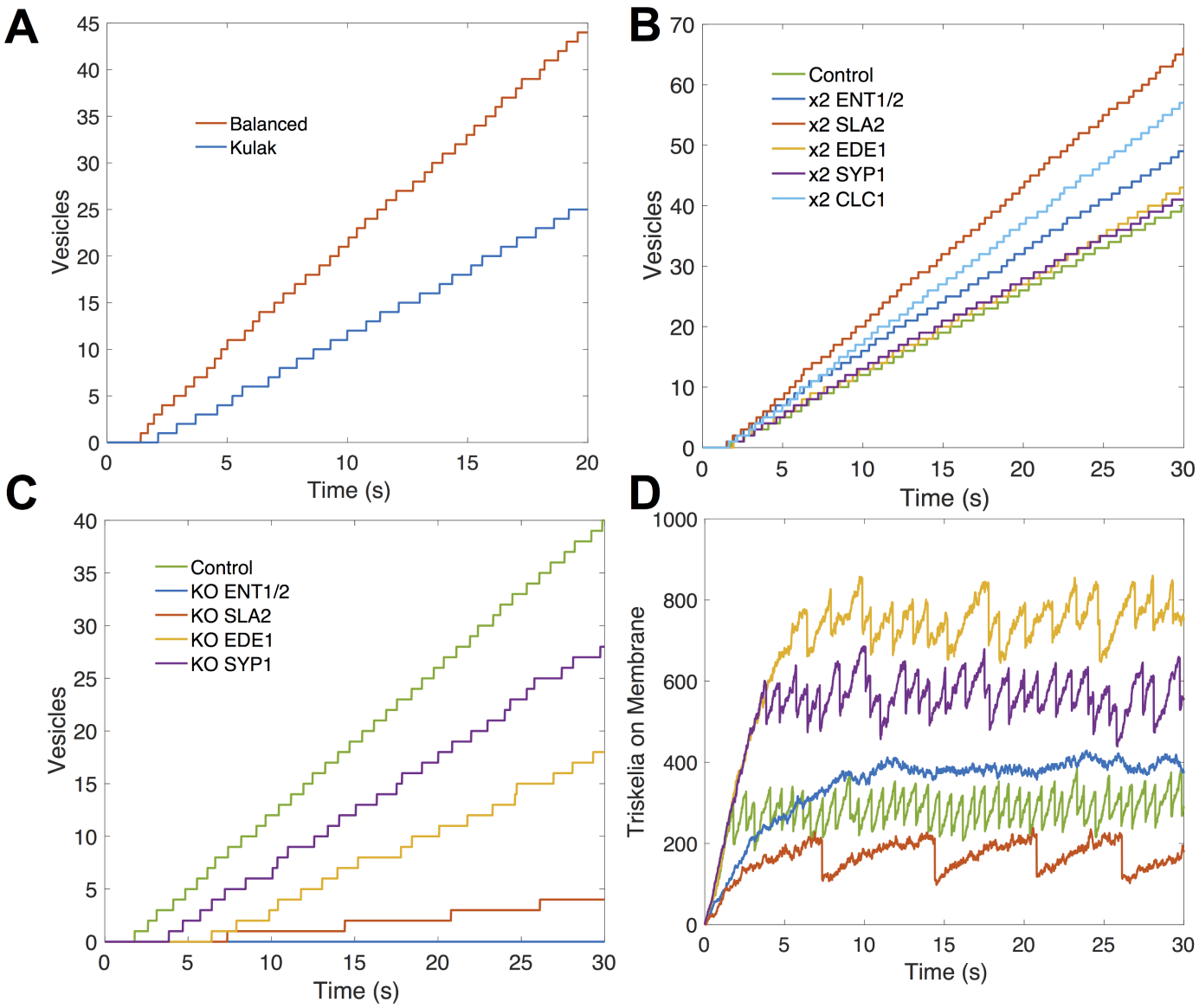
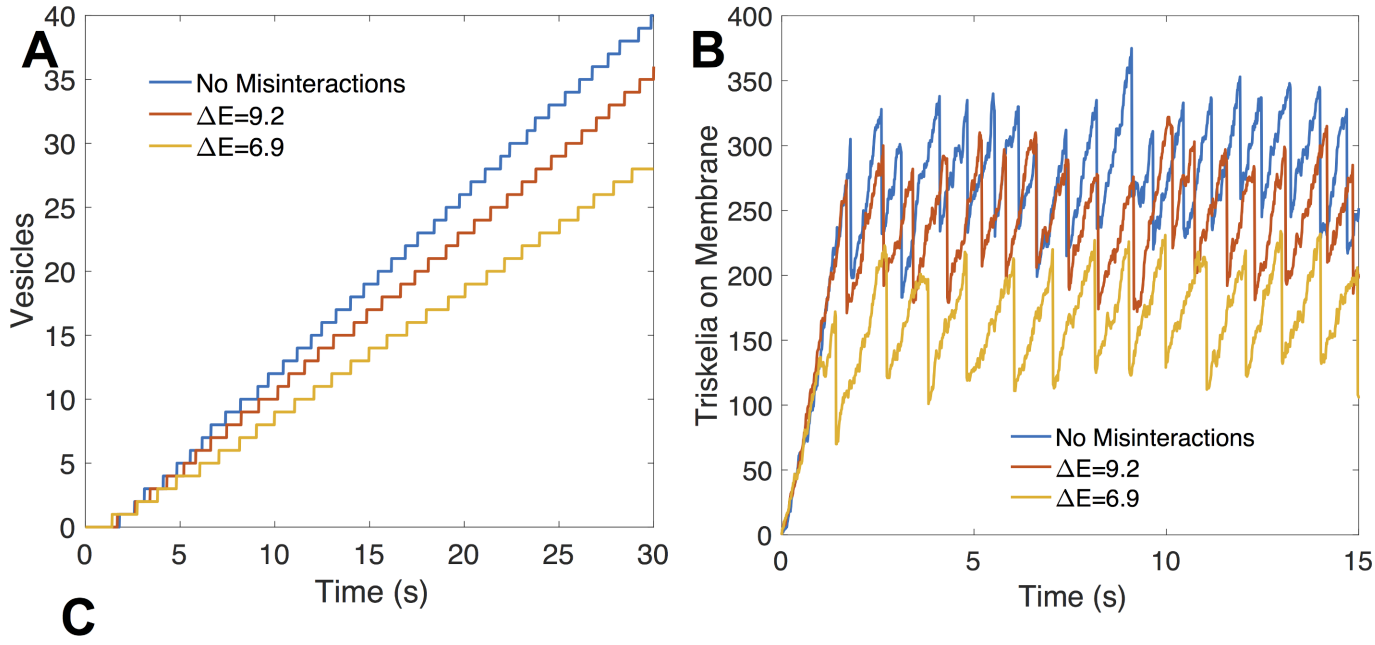



**Fig 8. Vesicle formation is tunable with adaptor proteins. (A)** Vesicles were formed faster with balanced copy numbers, indicating that the biological copy numbers are not optimized for maximum vesicle formation. **(B)** Adaptor proteins in the network were underexpressed. Vesicle frequency could be increased by doubling their concentrations. **(C,D)** The system is sensitive to adaptor protein knockouts. Knocking out either SYP1 or ENT1/2 nearly halts vesicle formation. SYP1 and EDE1 appear to have an aggregating effect, allowing vesicles to form with less triskelia on the membrane.

https://doi.org/10.1371/journal.pcbi.1006022.g008

to the 2D cell membrane that allows cage formation to occur; at least 81% of triskelia were brought to the membrane by adaptor proteins. This suggests another possible reason for over-expression of clathrin: to compensate for lower binding affinity by saturating adaptor proteins.

**Misinteractions have a significant impact for the strong-binding interactions.** To determine the overall influence of misinteractions on vesicle formation, and its dependence on protein binding affinity, we added misinteractions at two different strengths (Methods), with an average ratio of $K_{D,nonspecific}$ to $K_{D,specific}$ of 10,000 and 1,000. Despite the weakness of the misinteractions, they decreased the frequency of vesicle formation (Fig 9A and 9B), though this effect was overall less significant than that of copy number alteration (Fig 8).




C

| Misinteractions | Adaptors / Triskelia | Excess CHC1[1] | Excess CLC1[1] | ENT1/2 | SLA2 | SYP1 | EDE1 | YAP1801/2 |
|---|---|---|---|---|---|---|---|---|
| None | 18.7 | 29 | 36 | 553 | 366 | 479 | 470 | 3 |
| ΔE=9.2 | 22.5 | 71 | 88 | 674 | 438 | 576 | 560 | 2 |
| ΔE=6.9 | 33.2 | 208 | 262 | 988 | 668 | 857 | 802 | 6 |

[1]Excess means not part of a complete triskelia complex. The molecules may be in a partial complex

**Fig 9. Misinteractions interfere with clathrin recruitment. (A)** Adding misinteractions to the network decreased vesicle formation and **(B)** interferes with recruitment of triskelia to the membrane. This was caused by aggregates containing too many adaptor proteins, draining them from the cytoplasmic pool. **(C)** Average adaptor proteins in each vesicle. With strong misinteractions, vesicle aggregates contained many adaptors and incomplete triskelia.

https://doi.org/10.1371/journal.pcbi.1006022.g009

In part 2, "Imbalance increases misinteractions dependent on the network topology and binding affinities of proteins", we found that strong-binding proteins are more sensitive to stoichiometric balance because they are prone to misinteractions. The strongest binders in the network are the Clathrin heavy-chain to light chain interaction (Table 1), and they are both more highly expressed relative to the adaptor partners. Misinteractions dramatically increased the number of both heavy and light chains that were not properly assembled into triskelion (~10 fold), because they became trapped in misinteractions (Fig 9C). For the weaker binding adaptor proteins, the misinteractions increased non-functional aggregation but to a much lower extent, resulting in about 2-fold increase of adaptor proteins in vesicle complexes. Although this 2-fold increase may seem high given the weakness of the misinteractions, it is driven by the localization of these adaptor proteins on the membrane, which concentrates the proteins and promotes binding between any pair of available binding interfaces [59].

Ultimately, misinteractions reduced the frequency of vesicle formation because each vesicle contained a very large aggregate of proteins that drained the cytoplasmic pool of adaptors needed to form new vesicles. The adaptor protein composition is shown in Fig 9C. Without misinteractions, vesicles had an average of 18.7 adaptor proteins per full triskelia, whereas strong misintearctions increased the ratio to 33.2. An interesting consequence of misinteractions is that it initially sped up the formation of the first vesicle, due to the large aggregates assembling on the membrane. However, subsequent vesicles were slower to accumulate than without misinteractions. In contrast, without misinteractions, the speed of initial vesicle formation always correlates with the speed of subsequent vesicles formed.

## Discussion

### Measuring stoichiometric balance in protein-protein networks determines unexpected correlations in protein expression levels

The metrics and SBOPN algorithm we have developed objectively determine whether a protein is under or overexpressed relative to not only its direct binding partners, but to a larger network including partners of partners. This global evaluation is thus sensitive to the size of the network, but directly captures how the multiple binding interfaces of a protein can control its competition for binding partners. In the interface-resolved CME network, we have shown evidence of imperfect, but statistically significant stoichiometric balance. However, the original 56-protein network was overall unbalanced due to the high overexpression of the actin binding protein cofilin. The size of the network clearly matters, in the small modules, we are statistically out-of-balance, but on a larger scale, still in balance. Outliers are emphasized in smaller networks. At the same time, leaving out additional partners can provide some explanation for the observed imbalance. Imbalance may also indicate possible missing interactions in the network. Despite the simplicity of our metric, our method was still able to highlight both correlated concentrations and proteins that violate balance for functional reasons, such as the kinase PRK1. Furthermore, the observed balances can suggest possible mechanisms of assembly, for example, that can then be studied using kinetic modeling, as we did here. What our results emphasize is that correlations are highly important: functionality can be obliterated with significant imbalance, and misinteractions can also be overwhelming with significant imbalance.

Although we only applied our stoichiometric balance analysis to the 56 protein CME network, two smaller modules of this network, and the 127-protein ErbB network, these networks are significantly larger than the obligate complexes previous studied for copy number balance [5, 6]. Our networks also contain a much larger variety of binding interaction strengths and competitive and non-competitive interactions. As we showed above, balance depended on the protein network's underlying IIN. While it would be beneficial to repeat this analysis on a larger network, there is a paucity of manually curated IINs in the literature. There are various larger automatically constructed IINs, constructed with homology modeling[77, 78], but our previous work found these automatic IINs suffer from various inaccuracies and differ significantly from manually curated IINs in topology[41].

### Limitations of measuring stoichiometric balance for larger PPINs

The SBOPN method only accounts for the binding interface network structure and observed copy numbers. A missing feature of our stoichiometric balance metric is that proteins within a network can be expressed with both spatial and temporal variation. For a small binding network this is not a major concern, since proteins in the same complex tend to be co-expressed

[79] and co-localized so they may bind. But as network size is scaled up, the probability of all proteins being equally present reduces. Such temporal and spatial variations could be taken into account in the construction of the network, leaving out proteins that are not functional at the same time.

A natural extension to our measure of stoichiometric balance would be to also account for binding affinities of interactions in addition to the binding interface network structure and observed copy numbers. Our results here and previous studies[19] indicate that balance should be more tightly constrained for strong binding proteins. However, one benefit to leaving affinities out of the measurement is that biochemical data is in even more limited availability than binding interface data. Our existing metric can thus be much more easily applied to a variety of networks. Furthermore, by picking out highly correlated expression levels, our method can then indicate which interactions might be quite strong, or vice-versa, which may be transient or weak.

## Noise and variability in experimental copy number measurements can limit observed balance

In this study we used yeast copy numbers from Kulak et al. because it was the most comprehensive. The other three studies we used for comparison did not cover all 56 proteins in our network. However, for the proteins we could compare, we found significant discrepancies between relative abundances. Light chains are weakly expressed in other studies, for example[1, 49, 50]. A few possible reasons for this exist. The first is that fluorescence data is inherently noisy. Experimentalists must deal with background noise, interference with protein localization due to the large fluorescent tags, and cross interactions with other proteins [80]. The second is that cell lines can accrue mutations over time that decrease or increase gene expression, a phenomenon observed with HeLa cells[81]. Finally, cells may alter gene expression for regulatory reasons, so the environment in which cells are grown may alter gene expression.

## Perfect balance is not observed, even if it would improve both misinteractions or equilibrium complex yield

We do not expect the cell to perfectly optimize the yield of all of its many assemblies. Each network we have evaluated here is ultimately part of a larger, global cellular network. Perfectly optimizing isolated, local modules does not appear to be a significant pressure for the cell, particularly when a sufficient balance, such as we observe for the vesicle-forming module, maintains functionality. Additionally, these processes, such as in the vesicle forming model discussed below, typically do not occur at equilibrium. Therefore, the concept of minimizing 'leftover' proteins based on expected equilibrium complexes formed is a simplification. Correlations in copy numbers are nonetheless often significant relative to randomly assigned copy numbers.

We found that copy number imbalance can lead to misinteractions and the features of biological IINs (power-law-like degree distribution, square and hub motifs, sparseness) typically have less misinteractions under balance copy numbers but more misinteractions under imbalance. These networks thus should require more tightly controlled balance to avoid misinteractions. But misinteractions are of course not the only pressure on copy numbers. For multiprotein assembly in an obligate complex (ARP2/3) and in a minimal model of vesicle formation for CME, we found that the functional cost of imbalance was dominated more by its impact on determining specific functional complexes than avoiding misinteractions. Nonetheless, the fact that misinteractions can decrease vesicle formation, by sequestering away adaptor

proteins into large aggregates, shows that misinteractions are worse than simply having an excess of free proteins. If this result can be generalized, it may have important implications for mechanistic modeling of biological systems, as misinteractions or system error is rarely taken into account.

## Observed imbalances in the non-equilibrium vesicle forming module could provide benefits to assembling cargo-selective vesicles

Although the functional effects of copy number balance are usually discussed in the context of number of complete complexes at equilibrium, we have shown that non-equilibrium dynamics can be affected as well. While the clathrin heavy chains and light chains were balanced with each other, they were overexpressed compared to their adaptor proteins, and this limited the frequency of vesicle formation. Although we found that perfectly balanced copy numbers therefore improved vesicle formation frequency compared to observed copy numbers, we speculate that specific imbalances could still be selected for evolutionarily. There are various possible reasons for this imbalance: the function of endocytosis is cargo uptake, and there is a cargo loading process before endocytosis occurs.[75, 76] Hence to maximize function, controlled endocytosis around high-cargo areas of the membrane may be preferably to frequent, spontaneous endocytosis, and the adaptor proteins can serve as an intentional bottleneck in the process. Clathrin, which cannot directly bind to the membrane, may be kept at a high expression in the cytosol so that there are enough triskelia to quickly form a vesicle no matter where the endocytic site occurs. However, the observed underexpression could also be because there are other adaptor proteins not included in our model, or because clathrin interactions have weaker affinities than interactions between adaptor proteins and must saturate them.

Finally, the predictions of our minimal vesicle-forming model are ultimately limited by the approximations we made to simulate the clathrin coat assembly and vesicle formation. Our model vesicles formed about 10 times faster than is observed *in vivo*. To fully capture the dynamics of this complex process, an ideal model would include all the proteins in our CME network (Fig 1), and include both the known biochemistry of binding interactions and the physics and biomechanics of membrane bending and scission. In yeast, the cytoskeleton is needed to help induce membrane budding, after which energy-consuming proteins such as dynamin scission off the vesicle from the plasma membrane for transport into the cell [76, 82]. However, such a modeling approach does not exist, due to the computational limitations of simulating such large complexes and membrane remodeling, and the lack of biochemical data.

Based on the model we did construct, however, there are some more specific limitations. The first is that while rule-based modeling is a convenient way to model complex formation, some theoretical aggregates may be impossible due to steric hindrance. Our model predicted that a vesicle of 100 triskelia could contain ~1900 additional proteins. Assuming each vesicle is a sphere with 100nm diameter, the allowable surface area per adaptor/scaffold protein would only be ~17nm$^2$, which is too small to accommodate the excluded volume of the large, disordered regions of proteins such as ENT1 and 2[83]. Second, we did not include cooperatively in our model. Molecules localized in the same aggregate do not interact at a faster rate in conventional rule-based modeling. Clathrin triskelia weakly polymerize, as noted above, but the aggregation effect of the adaptor proteins–especially the SYP1/EDE1 complex–localizes triskelia close together, allowing them to bind strongly. In future work we will consider effects of cooperativity on assembly, as well as construct more detailed spatial and structural models of the vesicle forming process.

## Methods

### Stoichiometric Balance Optimization for Protein Networks (SBOPN) algorithm

A stoichiometrically balanced network has the copy numbers of each interface matched to the copy numbers of all pairwise complexes it participates in (Fig 2). Balanced copy numbers are obtained by assigning a number of desired complexes to each edge in the interface binding network. The balanced copy numbers of each interface can then be calculated from the equation:

$$Ax = C \tag{2}$$

Where "A" is a binary matrix with $N_{int}$ rows (one for each interface) and $M_{edge}$ columns (one for each pairwise interaction). $A_{i,j} = 1$ if the interface *i* is used in the interaction *j*, or 2 if a self-interaction, and 0 otherwise. "x" is the vector of desired pairwise complexes ($M_{edge}$ x 1), and "C" is the number of interface copy numbers ($N_{int}$ x 1). In Fig 10 we illustrate this procedure for a small toy network.

If desired pairwise complexes, x, is specified, interface copy numbers, C, can directly be solved for using Eq 1, but if interface copy numbers, C, are specified, x will not, in general, have an exact or nontrivial solution unless C is balanced. This is because all entries of x must be >0 or some other minimum value, as negative copies cannot exist. This produces a hard constraint on x. Given a vector C, an optimal solution to x must be solved for using quadratic programming rather than linear least-squares.

Our goal is to select for an optimal x given an input set of copy numbers "$C_0$". This is a soft constraint on the optimal x, because the input $C_0$ may not be balanced. Once an optimal x is found, forward solving Eq 1 will in general not perfectly recover $C_0$. $C_0$ can constrain all interfaces or a subset of them. To constrain a protein is to constrain all interfaces on it. We introduce a third constraint on the optimal x: the copy numbers of interfaces on the same protein should be equal. This often makes nontrivial solutions impossible (Fig 10), so it is also a soft constraint. Combining all of these constraints, the optimal desired number of complexes "x"

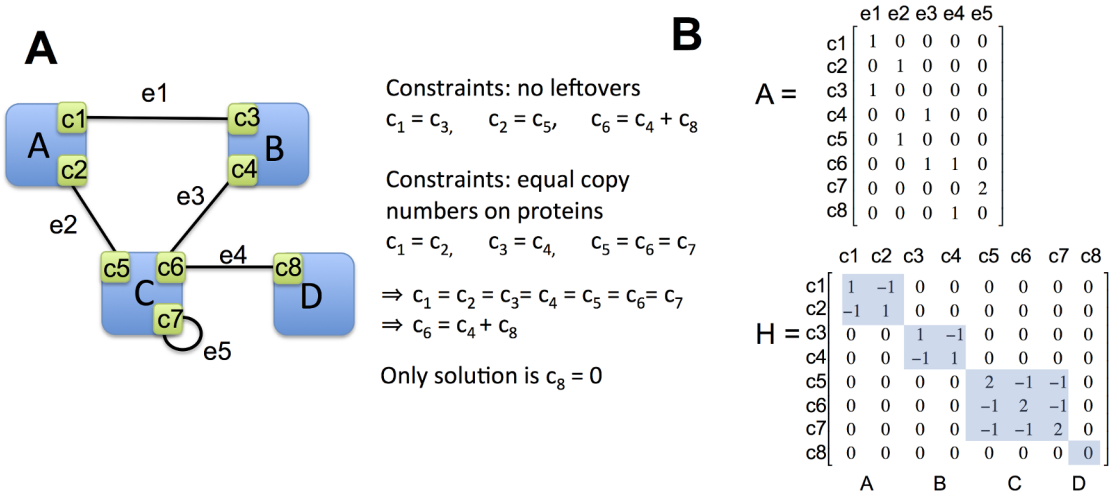



**Fig 10. Example network for constructing inputs to the SBOPN algorithm. (A)** This example PPIN with interfaces resolved has no nontrivial balanced solution when all constraints are applied. **(B)** The "A" and "H" matrices that are used as inputs for the SPOBN method are shown for the left network.

https://doi.org/10.1371/journal.pcbi.1006022.g010

can be found by minimizing the equation:

$$\min_x[\alpha(Ax - C_0)^T Z(Ax - C_0) + (Ax)^T H(Ax)],\ x \geq 0 \quad (3)$$

Where each variable is defined as follows:

A: $N_{int}$ x $M_{edge}$ matrix defining which interfaces are used in which interaction, i.e. pairwise complex.

x: $M_{edge}$ x 1 vector of desired pairwise complex copy numbers

$C_0$: $N_{int}$ x 1 vector of constrained copy numbers.

Z: $N_{int}$ x $N_{int}$ diagonal matrix that selects which interfaces are constrained. Entries = 1 if the interface is constrained and = 0 otherwise. If all interfaces are constrained, Z equals the identity matrix.

H: $N_{int}$ x $N_{int}$ permutated block diagonal matrix with positive and negative entries such that $H^*C = 0$ if interfaces on the same protein have equal copy numbers. Each block corresponds to a protein (Fig 10).

α: 1x1 scaling parameter which determines the relative weight of the $C_0$ soft constraint vs the equal interfaces soft constraint.

For any vector x, Eq 2 produces a positive scalar value. The equation was minimized using the OOQP (object-oriented quadratic programming) 0.99.26 package for C++[84]. Quadratic programming is necessary due to the constraint of $x \geq 0$. Eq 2 can be converted into a quadratic equation of the form

$$\frac{1}{2}x^T Qx + d^T x + r \quad (4)$$

Using

$Q = 2\alpha A^T Z A + 2A^T H A$

$d^T = -2\alpha C_0{}^T Z^T A$

$r = \alpha C_0{}^T Z C_0$

"r" can be ignored by the solver when minimizing the equation since it is a constant term.

Once $x_{min}$ is found via Eq 3, the optimized interface copy numbers can obtained by forward solving $A^* x_{min} = C_{balanced,int}$. Interfaces on the same protein will not necessarily have equal copy numbers due to the competing constraints of Eq 2 (Fig 2C). We can assign a single copy number to each protein by averaging over all interface copy numbers on that protein to give $C_{balanced}$, a vector of protein copy numbers. These values were used when calculating which proteins were over or underexpressed in the networks. Distance from $C_0$ to $C_{balanced}$ was used as a metric to determine relative balance (see below).

## Biological protein copy numbers

For the yeast CME network, $C_0$ was used to constrain all 56 proteins (Z = Identity matrix) because copy numbers from Kulak et al. were available[2]. For the ErbB signaling network, only 115 out of 127 proteins with available expression level data were constrained. 100 of these proteins were constrained with HeLa copy number estimations from Kulak et al. [2], while estimated copy numbers for 15 additional proteins were added from four additional studies [19, 51–53], leaving 12 proteins with unknown expression data. See S2 Table for all values.

## Measuring the degree of stoichiometric balance in observed concentrations

Using the optimized copy numbers, $C_{balanced}$, we can then ask, how close are the original, biologically observed copy numbers to these optimally balanced values? If the original copy numbers are already perfectly balanced, then they will match the optimal copy numbers. If they are

imperfect, then the two distributions will differ. We use two metrics to quantify the distance between the observed and optimized concentrations: chi-square distance (CSD)

$$\sqrt{\sum_i \frac{(X_i - Y_i)^2}{(X_i + Y_i)}} \tag{5}$$

and Jensen-Shannon Distance (JSD) after converting both vectors (X and Y) to distributions (x and y)

$$\sqrt{\frac{1}{2}(D_{KL}(x\|z) + D_{KL}(y\|z))} \tag{6}$$

Where z = (x+y)/2 and $D_{KL}$ is the Kullback-Leibler divergence

$$D_{KL}(x\|y) = \sum_i x_i log \frac{x_i}{y_i} \tag{7}$$

For cases where Z≠I (i.e. not all interfaces were constrained) only distance between constrained interfaces was measured.

## Small network motifs

Binding for the five 3- or 4-node network motifs; triangle, chain, square, 4-node hub, and flag; was simulated using the Gillespie algorithm[47]. Besides the specific binary interactions, nonspecific interactions were allowed at a strength determined by an "energy gap" between binding energies, though in practice we defined the ratio nonspecific $K_D$ to specific $K_D$ by factors of 10. This corresponded to a linear difference in free energies via the equations:

$$K_{D,specific} = c_0 e^{-\Delta E_1 / K_B T}$$

$$K_{D,nonspecific} = c_0 e^{-\Delta E_2 / K_B T}$$

$$\frac{K_{D,specific}}{K_{D,nonspecific}} = e^{-(\Delta E_1 - \Delta E_2) / K_B T}$$

The networks were simulated under various initial concentrations. The steady-state ratio of Eq 1 was recorded, where $N_{nonspecific}$ is the number of nonspecific binary complexes, $N_{specific}$ is the number of specific binary complexes, and $N_{free}$ is the number of free proteins. Ratios were averaged across 5,000 runs.

To generate surface plots, two proteins were chosen to be variable while the remaining proteins were given fixed copy numbers. Because the flag motif produced asymmetric plots, two different choices of variable proteins were used. (S3 Fig) Surface plots were generated using Matlab.

We calculated sensitivity by determining the principal component of the surface plot data (i.e. the vector of greatest variance) and measuring the percent change in ratio from the optimum along this vector. For better comparison, we normalized distance along the surface plots via dividing the abundance of the variable proteins by the abundance of the fixed proteins.

Motifs with purely noncompetitive interactions were not considered, because the interface network would then consist entirely of pairs, such as the IIN for Fig 1B. The balance is simple for pairs: all interfaces have the same copy numbers. We limited our analysis of Results part 2, "Imbalance increases misinteractions dependent on the network topology and binding

affinities of proteins", to small competitive motifs where we could enumerate all possible complexes and study effects of concentration variation systematically.

## Analysis of complex IIN topologies

For the large network analysis we used the 500 networks from Johnson *et al*, *J Phys Chem B* 2013[27]. 25 sets of 10 networks each were randomly generated using two parameters: number of nodes (90, 110, 125, 150, 200), keeping the number of edges fixed at 150; and the preferential attachment exponent "γ" from Goh, 2001[85]. γ = 0 corresponds to a binomial, Erdos-Renyi network, whereas γ = 1 corresponds to a power-law or "scale-free" network. Values of 0, 0.2, 0.4, 0.6, and 0.8 were used. Finally, a local topology optimization algorithm that decreased the frequency of chain and triangle motifs and increased hub motifs was applied to each network, for 500 networks in total. All networks assume competitive (binary) binding.

Rather than assign an arbitrary specific and nonspecific $K_D$ for the networks, we used the relative binding energies determined for each network in the source paper. This was determined by a physics-based Monte Carlo optimization scheme of amino acid residues, as described in Johnson, 2011[23]. The minimum energy gap between specific and nonspecific interactions could be measured as a relative metric of the network's propensity for misinteractions. Because the binding strengths were relative, we could alter the average binding strength to determine the effects on misinteractions. This was varied between 7 values of 1 nM to 1 mM, using factors of 10. Finally, to obtain results more comparable to the simple networks, we also ran simulations where each specific interaction had $K_D$ = 100 nM and each nonspecific interaction had $K_D$ = 100 μM.

Networks were simulated to steady state using the Gillespie algorithm[47] under five differing sets of copy numbers (CNs) for free proteins: equal CNs for each protein, random CNs sampled from a yeast protein concentration distribution (performed 20 times) and three forms of balanced CNs using the network architecture. Any set of CNs without leftovers–i.e. having exactly enough proteins to create a certain number of specific complexes–is considered "balanced", and thus there are infinite solutions. The first balanced set assumed an equal number of each type of specific complex, which results in protein CNs proportional to the protein's number of partners. The remaining balanced CNs were determined by finding "x" to minimize a simplified form of Eq 2:

$$\min_x (A^* x - C_0)^T (A^* x - C_0) \tag{8}$$

Here there is only one interface on each protein, and all the proteins are constrained, so there is no need for a Z matrix, the α scaling parameter, or the second term. $C_0$ is either equal copy numbers or randomly sampled copy numbers. After $x_{min}$ is found via quadratic programming (see above), the balanced CNs are obtained by forward solving $C_{balanced} = A^* x_{min}$.

To measure nonspecific complex formation, a modified ratio was used:

$$Cost(C_0) = \frac{2N_{nonspecific}(C_0)}{2N_{specific}(C_0) + N_{free}(C_0)} \tag{9}$$

to compare total individual proteins in each bound or unbound state, rather than number of unbound or bound states. To measure sensitivity, the ratio under unbalanced CNs ($C_0$) divided by the ratio under balanced CNs ($C_{balanced}$) was calculated. A higher ratio indicates higher sensitivity to CN balancing.

## ARP2/3 complex

The kinetic model was simulated using the stochastic simulation method (the Gillespie algorithm). Binding interactions were encoded via the rule-based language BioNetGen and simulated via the Network Free Simulation (NFSim) software [48]. Trimer cooperativity was modeled by increasing the rate of the third reaction if three members of a correct trimer were held together by two reactions. For example, if A is bound to B is bound to C, and a binding between A and C is possible, that reaction rate was set to be arbitrarily high. Reaction rates were arbitrary, but interactions with the core subunit ARC19 were set to be ~10 fold stronger than interactions between periphery subunits, as this increased yield. Yield was measured via the equation

$$Yield = \frac{N_{desired}}{N_{desired} + N_{undesired}} \tag{10}$$

Where $N_{desired}$ is the number of *proteins* in complete complexes (equal to seven times the number of complex complexes) and $N_{undesired}$ is the number of proteins in incomplete or misbound complexes. Completely free proteins were ignored.

## Simulating clathrin recruitment to the membrane

A subnetwork of nine proteins–clathrin heavy chain (CHC1), clathrin light chain (CLC1), SLA2, ENT1/2, EDE1, SYP1, and YAP1801/2 –was defined based on known binding interactions (Table 1). Because the existence of multiple interfaces, allowing noncompetitive binding, results in a large number of possible species we simulated our model using the Network Free Simulator (NFSim)[48]. Binding dissociation constants were obtained from the literature, including for protein-lipid binding. For simplicity, the heavy chains were already assumed to be in trimer form, and ENT1/2 was combined into a single protein as the binding partners were the same. Binding constants were pulled from the literature. (Table 1)

The cell membrane and the cell cytoplasm function as different compartments with different volumes, but NFSim is not integrated with BioNetGen's compartment language. We bypassed this problem by doubling the number of rules: besides the main rule for each reaction, an additional rule stated that if both proteins are on the cell membrane then the $k_{on}$ rate should be increased according to the membrane volume. Cell membrane 'volume' was determined by multiplying the membrane surface area by a factor $2\sigma = 2$ nm to capture the change in binding affinities between 3D and 2D (see S1 Text).

Since our primary goal was to measure clathrin recruitment to the membrane, any complex on the membrane with at least 100 triskelia (a complex of three CHC1 and three CLC1) was considered a "vesicle" and deleted at a high rate $k_{dump}$. Proteins in the vesicle were then added back to the cytoplasmic pool at a rate $k_{recyc}$, which was set to be equal to $k_{dump}$ to indicate fast recycling. However, we clarify that even fast recycling is not instantaneous, and that proteins are added back one at a time rather than all at once. Fast vesicle formation thus could still drain the pool of adaptor proteins.

Misinteraction strengths were determined by calculating the geometric mean of the dissociation constants of each interface, as this provided a $K_D$ based on the arithmetic mean of the binding energies.

$$\begin{aligned} K_{D,mean} &= \sqrt[n]{K_{D,1}K_{D,2}\ldots K_{D,n}} \\ &= \sqrt[n]{e^{-\Delta E_1/K_BT}\cdot e^{-\Delta E_1/K_BT}\cdot\ldots e^{-\Delta E_1/K_BT}} \\ &= \sqrt[n]{e^{-(\Delta E_1+\Delta E_2+\ldots\Delta E_n)/K_BT}} \\ &= e^{-(\Delta E_1+\Delta E_2+\ldots\Delta E_n)/nK_BT} \end{aligned}$$

The $K_D$ of a misinteraction between two interfaces was set to be:

$$f\sqrt{K_{D,mean,1}K_{D,mean,2}} \tag{11}$$

where f = 10,000 (weak misinteractions, corresponding to an energy gap of ~9.21) or 1,000 (stronger misinteractions, energy gap of ~6.91)

Network maps were generated using Cytoscape[86] and RuleBender[87]. Plots were generated in MATLAB. C++ code for the network balancing algorithm SBOPN is available at https://github.com/mjohn218/StoichiometricBalance, and may be applied to any interface-resolved network. The CME and ErbB networks are provided as example inputs.

## Supporting information

**S1 Text. Notes on the vesicle forming module.**
(PDF)

**S1 Table. Model parameters with notes.**
(PDF)

**S2 Table. Protein Copy Numbers for the CME and ErbB networks.**
(XLSX)

**S1 Fig. Effects of the α parameter on interface copy number noise. (A)** Noise is calculated as the variance of the copy numbers assigned to interfaces on the same protein divided by the square of their average copy number. It does not refer to expression level noise. A high "α" parameter allowed greater variance, but even a low α could not remove noise entirely because there are no balanced solutions where all proteins can have interfaces of equal copy number. Noise had a sigmoidal relationship with log(α). **(B)** Example of interface noise on a protein. **(C,D)** Scatter plot of protein interface copy number noise vs a protein's balanced "abundance", the average of their interface copy numbers. The black line is where noise is inverse of abundance. The red line is noise = 0.1, which is expected to be the upper limit of noise when abundance exceeds ~1000 copy numbers [12]. For a low α, proteins varied widely in the amount of noise they have, though high-abundance proteins tended to have less noise, and were below the 0.1 threshold. As α was raised, proteins approached the same level of noise.
(TIFF)

**S2 Fig. Ras and MAP3K proteins in the ErbB network are underexpressed.** The ErbB network, which consists mainly of phosphorylation interactions, was not found to be statistically balanced based on the Jensen-Shannon divergence. However, certain proteins of note were found to be underexpressed, such as the three Ras proteins (HRAS, KRAS, and NRAS), and the MAP3K layer (RAF1, BRAF, ARAF, MAP3K1, MAP3K2, MAP3K4, and MAP3K11). Also underexpressed were the ErbB receptors and the hub SRC. These suggest a strategic imbalance of upstream proteins (in the case of MAPK cascades) or network bottlenecks (Ras proteins or SRC). Highlighted are the Ras proteins (blue), MAP3Ks (orange), MAP2Ks (green), and MAPKs (red).
(TIFF)

**S3 Fig. Misinteraction frequency in the small network motifs. (A)** Small networks used to construct the surface plots. For all simulations, two proteins had variable concentrations (blue) while the others had fixed concentrations (pink). **(B)** Surface plots of misinteraction frequency (color bar-Eq 1 main text). Misinteraction frequency is measured as Nnonspecific / (Nspecific + Nfree); that is, number of nonspecific complexes divided by all other species; at steady-state as described in the main text. Each plot corresponds to each respective network in **A**. The X

and Y-axes are the concentrations of the variable proteins divided by the total concentrations of the fixed proteins. The black line is the principal component, which was used as an axis to measure the sensitivity of misinteractions as one moved away from a local minimum. For the chain we used two arbitrary local minima because the absolute minimum was when B2 = 0, a trivial solution. For the flag network we used two different sets of fixed and variable proteins because the surface plots were asymmetric. **(C)** The sensitivity of each network to misinteraction frequency as the protein concentrations moved away from an optimum (local minimum). Sensitivity is measured as percent change from the optimal (lowest) misinteraction frequency.
(TIFF)

**S4 Fig. Effects of optimized local topology on misinteractions. (A)** Misinteraction frequency of networks under randomly sampled (left) and balanced copy numbers (right) when fixed energy gaps were used (KD, specific = 100nM, KD,nonspecific = 100μM). Networks with optimized topology and a power-law-like distribution ($\gamma = 0.8$) performed best under balanced copy numbers but worse under imbalance. **(B)** Heat map of misinteraction frequency under balanced copy numbers vs degree distribution and network density. Denser networks always had more misinteractions, but the effects of degree distribution depended on whether the local topology was optimized or not.
(TIFF)

**S5 Fig. ARP2/3 complex has higher yield under balanced copy numbers. (A)** Contact map of the seven subunits of the complex, generated with RuleBender{Smith, 2012 #488}**(B)** Under varying misinteraction strengths, the yield for the balanced copy numbers was always higher than for the observed copy numbers from Kulak et al.{Kulak, 2014 #276} Yield was measured as $N_{desired} / (N_{desired} + N_{undesired})$, which refer to the number of proteins in either desired (complete) complexes or undesired (incomplete or misassembled) complexes. **(C)** The observed copy number distribution was not found to be conserved between studies in either yeast or humans. Bar plots are from five studies of the ARP2/3 subunits in human cells. The red bar is for the addition of the "subunit 5-like" protein. Only one study (Hein et al.) found ARC19's equivalent, subunit 4, to be underexpressed{Hein, 2015 #277}.
(TIFF)

## Acknowledgments

We gratefully acknowledge computational resources from Maryland Advanced Computing Cluster (MARCC) and the Hopkins High Performance Cluster (HHPC). We thank Daisy Duan for helping to collect rate parameters and the Johnson lab members for helpful feedback on the manuscript.

## Author Contributions

**Conceptualization:** David O. Holland, Margaret E. Johnson.

**Data curation:** David O. Holland, Margaret E. Johnson.

**Formal analysis:** David O. Holland, Margaret E. Johnson.

**Funding acquisition:** Margaret E. Johnson.

**Investigation:** David O. Holland, Margaret E. Johnson.

**Methodology:** David O. Holland, Margaret E. Johnson.

**Project administration:** Margaret E. Johnson.

**Resources:** Margaret E. Johnson.

**Software:** David O. Holland, Margaret E. Johnson.

**Supervision:** Margaret E. Johnson.

**Validation:** David O. Holland, Margaret E. Johnson.

**Visualization:** David O. Holland, Margaret E. Johnson.

**Writing – original draft:** David O. Holland, Margaret E. Johnson.

**Writing – review & editing:** David O. Holland, Margaret E. Johnson.

## References


1. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. Nature. 2003; 425(6959):737–41. https://doi.org/10.1038/nature02046 PMID: 14562106.
2. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods. 2014; 11(3):319–24. https://doi.org/10.1038/nmeth.2834 PMID: 24487582.
3. Veitia RA, Bottani S, Birchler JA. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. Trends in genetics: TIG. 2008; 24:390–7. https://doi.org/10.1016/j.tig.2008.05.005 PMID: 18585818.
4. Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci U S A. 2012; 109(37):14746–53. https://doi.org/10.1073/pnas.1207726109 PMID: 22908297; PubMed Central PMCID: PMCPMC3443177.
5. Veitia RA, Potier MC. Gene dosage imbalances: action, reaction, and models. Trends Biochem Sci. 2015; 40(6):309–17. https://doi.org/10.1016/j.tibs.2015.03.011 PMID: 25937627.
6. Oberdorf R, Kortemme T. Complex topology rather than complex membership is a determinant of protein dosage sensitivity. Molecular systems biology. 2009; 5:253. https://doi.org/10.1038/msb.2009.9 PMID: 19293832.
7. Tomala K, Korona R. Evaluating the fitness cost of protein expression in Saccharomyces cerevisiae. Genome Biol Evol. 2013; 5(11):2051–60. https://doi.org/10.1093/gbe/evt154 PMID: 24128940; PubMed Central PMCID: PMCPMC3845635.
8. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. Cell. 2009; 138(1):198–208. https://doi.org/10.1016/j.cell.2009.04.029 PMID: 19596244.
9. Makanae K, Kintaka R, Makino T, Kitano H, Moriya H. Identification of dosage-sensitive genes in Saccharomyces cerevisiae using the genetic tug-of-war method. Genome Res. 2013; 23(2):300–11. https://doi.org/10.1101/gr.146662.112 PMID: 23275495; PubMed Central PMCID: PMCPMC3561871.
10. Gsponer J, Babu MM. Cellular strategies for regulating functional and nonfunctional protein aggregation. Cell Rep. 2012; 2(5):1425–37. https://doi.org/10.1016/j.celrep.2012.09.036 PMID: 23168257; PubMed Central PMCID: PMCPMC3607227.
11. Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science. 2008; 322(5906):1365–8. https://doi.org/10.1126/science.1163581 PMID: 19039133; PubMed Central PMCID: PMCPMC2803065.
12. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science. 2010; 329(5991):533–8. https://doi.org/10.1126/science.1188308 PMID: 20671182; PubMed Central PMCID: PMCPMC2922915.
13. Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, et al. Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. Nat Genet. 2011; 43(12):1179–85. https://doi.org/10.1038/ng.948 PMID: 22019781; PubMed Central PMCID: PMCPMC3576853.
14. Georgiev P, Chlamydas S, Akhtar A. Drosophila dosage compensation: males are from Mars, females are from Venus. Fly (Austin). 2011; 5(2):147–54. https://doi.org/10.4161/fly.5.2.14934 PMID: 21339706; PubMed Central PMCID: PMCPMC3127063.
15. Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GA. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome.

Proc Natl Acad Sci U S A. 2012; 109(14):5346–51. https://doi.org/10.1073/pnas.1116763109 PMID: 22392987; PubMed Central PMCID: PMCPMC3325647.

16. Kiel C, Verschueren E, Yang J-S, Serrano L. Integration of protein abundance and structure data reveals competition in the ErbB signaling network. Sci Signal. 2013; 6:ra109. https://doi.org/10.1126/scisignal.2004560 PMID: 24345680.
17. Murugan A, Zou J, Brenner MP. Undesired usage and the robust self-assembly of heterogeneous structures. Nat Commun. 2015; 6:6203. https://doi.org/10.1038/ncomms7203 PMID: 25669898.
18. Matalon O, Horovitz A, Levy ED. Different subunits belonging to the same protein complex often exhibit discordant expression levels and evolutionary properties. Curr Opin Struct Biol. 2014; 26:113–20. https://doi.org/10.1016/j.sbi.2014.06.001 PMID: 24997301.
19. Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell. 2015; 163(3):712–23. https://doi.org/10.1016/j.cell.2015.09.053 PMID: 26496610.
20. Johnson ME, Hummer G. Interface-resolved network of protein-protein interactions. Plos Comput Biol. 2013; 9:e1003065. https://doi.org/10.1371/journal.pcbi.1003065 PMID: 23696724.
21. Moriya H. Quantitative nature of overexpression experiments. Mol Biol Cell. 2015; 26(22):3932–9. https://doi.org/10.1091/mbc.E15-07-0512 PMID: 26543202; PubMed Central PMCID: PMCPMC4710226.
22. Zhang J, Maslov S, Shakhnovich EI. Constraints imposed by non-functional protein-protein interactions on gene expression and proteome size. Mol Syst Biol. 2008; 4:210. Epub 2008/08/07. https://doi.org/10.1038/msb.2008.48 PMID: 18682700; PubMed Central PMCID: PMC2538908.
23. Johnson ME, Hummer G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. P Natl Acad Sci USA. 2011; 108(2):603–8. https://doi.org/10.1073/Pnas.1010954108 ISI:000286097700035. PMID: 21187424
24. Heo M, Maslov S, Shakhnovich E. Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. Proc Natl Acad Sci U S A. 2011; 108(10):4258–63. https://doi.org/10.1073/pnas.1009392108 PMID: 21368118; PubMed Central PMCID: PMCPMC3054035.
25. Yang J-R, Liao B-Y, Zhuang S-M, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. P Natl Acad Sci USA. 2012; 109:E831–40. https://doi.org/10.1073/pnas.1117408109 PMID: 22416125.
26. Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. P Natl Acad Sci USA. 2012; 109:20461–6. https://doi.org/10.1073/pnas.1209312109 PMID: 23184996.
27. Johnson ME, Hummer G. Evolutionary pressure on the topology of protein interface interaction networks. The journal of physical chemistry B. 2013; 117:13098–106. https://doi.org/10.1021/jp402944e PMID: 23701316.
28. Tzeng SR, Kalodimos CG. Protein dynamics and allostery: an NMR view. Curr Opin Struct Biol. 2011; 21(1):62–7. https://doi.org/10.1016/j.sbi.2010.10.007 PMID: 21109422.
29. Reichard P. Ribonucleotide reductases: substrate specificity by allostery. Biochem Biophys Res Commun. 2010; 396(1):19–23. https://doi.org/10.1016/j.bbrc.2010.02.108 PMID: 20494104.
30. Schreiber G, Keating AE. Protein binding specificity versus promiscuity. Curr Opin Struct Biol. 2011; 21(1):50–61. https://doi.org/10.1016/j.sbi.2010.10.002 PMID: 21071205; PubMed Central PMCID: PMCPMC3053118.
31. Zarrinpar A, Park S-H, Lim WA. Optimization of specificity in a cellular protein interaction network by negative selection. Nature. 2003; 426:676–80. https://doi.org/10.1038/nature02178 PMID: 14668868.
32. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet. 2002; 31(1):64–8. https://doi.org/10.1038/ng881 PMID: 11967538.
33. Motley A, Bright NA, Seaman MN, Robinson MS. Clathrin-mediated endocytosis in AP-2-depleted cells. The Journal of cell biology. 2003; 162(5):909–18. https://doi.org/10.1083/jcb.200305145 PMID: 12952941; PubMed Central PMCID: PMCPMC2172830.
34. Jost M, Simpson F, Kavran JM, Lemmon MA, Schmid SL. Phosphatidylinositol-4,5-bisphosphate is required for endocytic coated vesicle formation. Curr Biol. 1998; 8(25):1399–402. PMID: 9889104.
35. Dannhauser PN, Ungewickell EJ. Reconstitution of clathrin-coated bud and vesicle formation with minimal components. Nat Cell Biol. 2012; 14(6):634–9. https://doi.org/10.1038/ncb2478 PMID: 22522172.
36. Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, et al. Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell. 2006; 21(3):319–30. https://doi.org/10.1016/j.molcel.2005.12.011 PMID: 16455487.

37. Zhou J, Lemos B, Dopman EB, Hartl DL. Copy-number variation: the balance between gene dosage and expression in Drosophila melanogaster. Genome Biol Evol. 2011; 3:1014–24. https://doi.org/10.1093/gbe/evr023 PMID: 21979154; PubMed Central PMCID: PMCPMC3227403.

38. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet. 2009; 10:451–81. https://doi.org/10.1146/annurev.genom.9.081307.164217 PMID: 19715442; PubMed Central PMCID: PMCPMC4472309.

39. McElroy JP, Krupp LB, Johnson BA, McCauley JL, Qi Z, Caillier SJ, et al. Copy number variation in pediatric multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England). 2013; 19:1014–21. https://doi.org/10.1177/1352458512469696 PMID: 23239789.

40. Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen TH, et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res. 2004; 64 (9):3060–71. PMID: 15126342.

41. Holland DO, Shapiro BH, Xue P, Johnson ME. Protein-protein binding selectivity and network topology constrains global and local properties of interface binding networks. Sci Rep. 2017; 7:5631. https://doi.org/10.1038/s41598-017-05686-2 PMID: 28717235

42. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002; 297:1183–6. https://doi.org/10.1126/science.1070919 PMID: 12183631

43. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, et al. Noise in protein expression scales with natural protein abundance. Nat Genet. 2006; 38(6):636–43. https://doi.org/10.1038/ng1807 PMID: 16715097.

44. Chen Q, Pollard TD. Actin filament severing by cofilin dismantles actin patches and produces mother filaments for new patches. Curr Biol. 2013; 23(13):1154–62. https://doi.org/10.1016/j.cub.2013.05.005 PMID: 23727096; PubMed Central PMCID: PMCPMC4131202.

45. Bravo-Cordero JJ, Magalhaes MA, Eddy RJ, Hodgson L, Condeelis J. Functions of cofilin in cell locomotion and invasion. Nat Rev Mol Cell Biol. 2013; 14(7):405–15. https://doi.org/10.1038/nrm3609 PMID: 23778968; PubMed Central PMCID: PMCPMC3878614.

46. Levchenko A, Bruck J, Sternberg PW. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. Proc Natl Acad Sci U S A. 2000; 97(11):5818–23. PMID: 10823939; PubMed Central PMCID: PMCPMC18517.

47. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977; 81:2340–61.

48. Sneddon MW, Faeder JR, Emonet T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. Nature methods. 2011; 8(2):177–U12. https://doi.org/10.1038/nmeth.1546 WOS:000286654600017. PMID: 21186362

49. Chong YT, Koh JL, Friesen H, Duffy SK, Cox MJ, Moses A, et al. Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. Cell. 2015; 161(6):1413–24. https://doi.org/10.1016/j.cell.2015.04.051 PMID: 26046442.

50. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006; 441(7095):840–6. https://doi.org/10.1038/nature04785 PMID: 16699522.

51. Wisniewski JR, Hein MY, Cox J, Mann M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. Mol Cell Proteomics. 2014; 13(12):3497–506. https://doi.org/10.1074/mcp.M113.037309 PMID: 25225357; PubMed Central PMCID: PMCPMC4256500.

52. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol. 2011; 7:548. https://doi.org/10.1038/msb.2011.81 PMID: 22068331; PubMed Central PMCID: PMCPMC3261714.

53. Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, et al. The quantitative proteome of a human cell line. Mol Syst Biol. 2011; 7:549. https://doi.org/10.1038/msb.2011.82 PMID: 22068332; PubMed Central PMCID: PMCPMC3261713.

54. McMahon HT, Boucrot E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. Nat Rev Mol Cell Biol. 2011; 12(8):517–33. https://doi.org/10.1038/nrm3151 PMID: 21779028.

55. Mishra SK, Keyel PA, Hawryluk MJ, Agostinelli NR, Watkins SC, Traub LM. Disabled-2 exhibits the properties of a cargo-selective endocytic clathrin adaptor. The EMBO journal. 2002; 21(18):4915–26. https://doi.org/10.1093/emboj/cdf487 PMID: 12234931; PubMed Central PMCID: PMCPMC126284.

56. Kelly BT, Graham SC, Liska N, Dannhauser PN, Honing S, Ungewickell EJ, et al. Clathrin adaptors. AP2 controls clathrin polymerization with a membrane-activated switch. Science. 2014; 345 (6195):459–63. https://doi.org/10.1126/science.1254836 PMID: 25061211; PubMed Central PMCID: PMCPMC4333214.

57. Jorgensen P, Nishikawa JL, Breitkreutz BJ, Tyers M. Systematic identification of pathways that couple cell growth and division in yeast. Science. 2002; 297(5580):395–400. https://doi.org/10.1126/science.1070850 PMID: 12089449.

58. Alberts B. Molecular biology of the cell. Sixth edition. ed. New York, NY: Garland Science, Taylor and Francis Group; 2015. 1 volume (various pagings) p.

59. Yogurtcu ON, Johnson ME. Cytoplasmic proteins can exploit membrane localization to trigger functional assembly. PLoS Comp Biol. 2018;Accepted.

60. Wakeham DE, Chen CY, Greene B, Hwang PK, Brodsky FM. Clathrin self-assembly involves coordinated weak interactions favorable for cellular regulation. The EMBO journal. 2003; 22(19):4980–90. https://doi.org/10.1093/emboj/cdg511 PMID: 14517237; PubMed Central PMCID: PMCPMC204494.

61. Miele AE, Watson PJ, Evans PR, Traub LM, Owen DJ. Two distinct interaction motifs in amphiphysin bind two independent sites on the clathrin terminal domain beta-propeller. Nat Struct Mol Biol. 2004; 11 (3):242–8. https://doi.org/10.1038/nsmb736 PMID: 14981508.

62. Zhuo Y, Ilangovan U, Schirf V, Demeler B, Sousa R, Hinck AP, et al. Dynamic interactions between clathrin and locally structured elements in a disordered protein mediate clathrin lattice assembly. J Mol Biol. 2010; 404(2):274–90. https://doi.org/10.1016/j.jmb.2010.09.044 PMID: 20875424; PubMed Central PMCID: PMCPMC2981644.

63. de Beer T, Hoofnagle AN, Enmon JL, Bowers RC, Yamabhai M, Kay BK, et al. Molecular mechanism of NPF recognition by EH domains. Nat Struct Biol. 2000; 7(11):1018–22. https://doi.org/10.1038/80924 PMID: 11062555.

64. Morgan JR, Prasad K, Jin S, Augustine GJ, Lafer EM. Eps15 homology domain-NPF motif interactions regulate clathrin coat assembly during synaptic vesicle recycling. J Biol Chem. 2003; 278(35):33583–92. https://doi.org/10.1074/jbc.M304346200 PMID: 12807910.

65. Boeke D, Trautmann S, Meurer M, Wachsmuth M, Godlee C, Knop M, et al. Quantification of cytosolic interactions identifies Ede1 oligomers as key organizers of endocytosis. Mol Syst Biol. 2014; 10:756. https://doi.org/10.15252/msb.20145422 PMID: 25366307; PubMed Central PMCID: PMCPMC4299599.

66. Winkler FK, Stanley KK. Clathrin heavy chain, light chain interactions. EMBO J. 1983; 2(8):1393–400. PMID: 10872336; PubMed Central PMCID: PMCPMC555288.

67. Engqvist-Goldstein AE, Warren RA, Kessels MM, Keen JH, Heuser J, Drubin DG. The actin-binding protein Hip1R associates with clathrin during early stages of endocytosis and promotes clathrin assembly in vitro. J Cell Biol. 2001; 154(6):1209–23. https://doi.org/10.1083/jcb.200106089 PMID: 11564758; PubMed Central PMCID: PMCPMC2150824.

68. Wilbur JD, Chen CY, Manalo V, Hwang PK, Fletterick RJ, Brodsky FM. Actin binding by Hip1 (huntingtin-interacting protein 1) and Hip1R (Hip1-related protein) is regulated by clathrin light chain. J Biol Chem. 2008; 283(47):32870–9. https://doi.org/10.1074/jbc.M802863200 PMID: 18790740; PubMed Central PMCID: PMCPMC2583295.

69. Henne WM, Kent HM, Ford MG, Hegde BG, Daumke O, Butler PJ, et al. Structure and analysis of FCHo2 F-BAR domain: a dimerizing and membrane recruitment module that effects membrane curvature. Structure. 2007; 15(7):839–52. https://doi.org/10.1016/j.str.2007.05.002 PMID: 17540576.

70. Stahelin RV, Long F, Peter BJ, Murray D, De Camilli P, McMahon HT, et al. Contrasting membrane interaction mechanisms of AP180 N-terminal homology (ANTH) and epsin N-terminal homology (ENTH) domains. The Journal of biological chemistry. 2003; 278(31):28993–9. https://doi.org/10.1074/jbc.M302865200 PMID: 12740367.

71. Moravcevic K, Alvarado D, Schmitz KR, Kenniston JA, Mendrola JM, Ferguson KM, et al. Comparison of Saccharomyces cerevisiae F-BAR domain structures reveals a conserved inositol phosphate binding site. Structure. 2015; 23(2):352–63. https://doi.org/10.1016/j.str.2014.12.009 PMID: 25620000; PubMed Central PMCID: PMCPMC4319572.

72. Yoon Y, Lee PJ, Kurilova S, Cho W. In situ quantitative imaging of cellular lipids using molecular sensors. Nat Chem. 2011; 3(11):868–74. https://doi.org/10.1038/nchem.1163 PMID: 22024883; PubMed Central PMCID: PMCPMC3205457.

73. Wu Y, Vendome J, Shapiro L, Ben-Shaul A, Honig B. Transforming binding affinities from three dimensions to two with application to cadherin clustering. Nature. 2011; 475(7357):510–3. https://doi.org/10.1038/nature10183 PMID: 21796210; PubMed Central PMCID: PMCPMC3167384.

74. Loerke D, Mettlen M, Yarar D, Jaqaman K, Jaqaman H, Danuser G, et al. Cargo and dynamin regulate clathrin-coated pit maturation. Plos Biol. 2009; 7(3):e57. https://doi.org/10.1371/journal.pbio.1000057 PMID: 19296720; PubMed Central PMCID: PMCPMC2656549.

75. Weinberg J, Drubin DG. Clathrin-mediated endocytosis in budding yeast. Trends Cell Biol. 2012; 22 (1):1–13. https://doi.org/10.1016/j.tcb.2011.09.001 ISI:000299450400001. PMID: 22018597

76. Boettner DR, Chi RJ, Lemmon SK. Lessons from yeast for clathrin-mediated endocytosis. Nat Cell Biol. 2011; 14(1):2–10. https://doi.org/10.1038/ncb2403 PMID: 22193158.

77. Mosca R, Ceol A, Aloy P. Interactome3D: adding structural details to protein networks. Nat Methods. 2013; 10(1):47–53. https://doi.org/10.1038/nmeth.2289 PMID: 23399932.

78. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nature biotechnology. 2012; 30(2):159–64. Epub 2012/01/19. https://doi.org/10.1038/nbt.2106 PMID: 22252508.

79. Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. Nature. 2003; 424:194–7. https://doi.org/10.1038/nature01771 PMID: 12853957.

80. Crivat G, Taraska JW. Imaging proteins inside cells with fluorescent tags. Trends Biotechnol. 2012; 30 (1):8–16. https://doi.org/10.1016/j.tibtech.2011.08.002 PMID: 21924508; PubMed Central PMCID: PMCPMC3246539.

81. Landry JJ, Pyl PT, Rausch T, Zichner T, Tekkedil MM, Stutz AM, et al. The genomic and transcriptomic landscape of a HeLa cell line. G3 (Bethesda). 2013; 3(8):1213–24. https://doi.org/10.1534/g3.113.005777 PMID: 23550136; PubMed Central PMCID: PMCPMC3737162.

82. Lu R, Drubin DG, Sun Y. Clathrin-mediated endocytosis in budding yeast at a glance. J Cell Sci. 2016; 129(8):1531–6. https://doi.org/10.1242/jcs.182303 PMID: 27084361; PubMed Central PMCID: PMCPMC4852772.

83. Busch DJ, Houser JR, Hayden CC, Sherman MB, Lafer EM, Stachowiak JC. Intrinsically disordered proteins drive membrane curvature. Nat Commun. 2015; 6:7875. https://doi.org/10.1038/ncomms8875 PMID: 26204806; PubMed Central PMCID: PMCPMC4515776.

84. Gertz EM, Wright SJ. Object-oriented software for quadratic programming. ACM Transactions on Mathematical Software. 2003; 29:58–81.

85. Goh K-I, Kahng B, Kim D. Universal behavior of load distribution in scale-free networks. Phys Rev Lett. 2001; 87:278207.

86. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13(11):2498–504. https://doi.org/10.1101/gr.1239303 PMID: 14597658; PubMed Central PMCID: PMCPMC403769.

87. Smith AM, Xu W, Sun Y, Faeder JR, Marai GE. RuleBender: integrated modeling, simulation and visualization for rule-based intracellular biochemistry. Bmc Bioinformatics. 2012; 13 Suppl 8:S3. https://doi.org/10.1186/1471-2105-13-S8-S3 PMID: 22607382; PubMed Central PMCID: PMCPMC3355338.