Random partitioning of molecules at cell division

Dann Huh[a,b] and Johan Paulsson[a,1]

[a]Department of Systems Biology, Harvard University, Boston, MA 02115; and [b]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

Many RNAs, proteins, and organelles are present in such low numbers per cell that random segregation of individual copies causes large "partitioning errors" at cell division. Even symmetrically dividing cells can then by chance produce daughters with very different composition. The size of the errors depends on the segregation mechanism: Control systems can reduce low-abundance errors, but the segregation process can also be subject to upstream sources of randomness or spatial heterogeneities that create large errors despite high abundances. Here we mathematically demonstrate how partitioning errors arise for different types of segregation mechanisms and how errors can be greatly increased by upstream heterogeneity but remarkably hard to avoid through controlled partitioning. We also show that seemingly straightforward experiments cannot be straightforwardly interpreted because very different mechanisms produce identical fits and present an approach to deal with this problem by adding binomial counting noise and testing for convexity or concavity in the partitioning error as a function of the binomial thinning parameter. The results lay a conceptual groundwork for more effective studies of heterogeneity among growing and dividing cells, whether in microbes or in differentiating tissues.

At balanced growth the abundances of cellular components are on average doubled during each cell cycle and then halved at cell division. But individual cells can deviate greatly from the average. Stochastic chemical reactions create fluctuations during the cell cycle and stochastic partitioning of components creates further fluctuations at cell division (1–4). This process perturbs concentrations and indirectly shapes molecular mechanisms by placing evolutionary constraints on reaction rates and network topologies.

The variation coming from stochastic production has been closely studied in the last decade (5–13), emphasizing how it could be much greater than expected from Poisson statistics due to upstream sources of randomness (5–7), how it could be controlled (8, 9) or exploited (10, 11) depending on selective pressures, and how fluctuations can reveal properties of the underlying mechanisms (12, 13). Partitioning errors can contribute just as much to the overall heterogeneity and recent results suggest that much of the noise attributed to, e.g., gene expression may in fact originate in stochastic partitioning (14). But not even the basic guiding principles of partitioning have been identified: how partitioning errors depend on upstream spatial heterogeneity or self-control, how they affect the observable cell heterogeneity, or what they can teach us about mechanisms.

The goal of this study is to provide a mathematical understanding of these principles by quantitatively modeling the heterogeneity introduced by various types of segregation mechanisms. We compare simple *independent segregation* where each segregating unit has a constant and independent probability of ending up in either daughter cell, to *disordered segregation* where variation in the partitioning machinery or the intracellular milieu further randomizes levels between daughters, or to various types of *ordered segregation* where copies directly or indirectly interact with each other to create a more even distribution between daughters. We then consider how such partitioning errors can be distinguished experimentally and discuss future directions.

Results

For dividing cells with x copies of a randomly segregating component X, where L and R copies segregate to each daughter cell after division, we can define a statistical partitioning error in several equivalent ways (*SI Text*),

$$Q_x^2 = \underbrace{\frac{\left\langle (L-R)^2 \right\rangle}{\langle x \rangle^2}}_{\substack{\text{statistical difference} \\ \text{between daughters}}} = \underbrace{\frac{\left\langle \sigma_{L|x}^2 \right\rangle}{\langle L \rangle^2}}_{\substack{\text{variation in newborn cells} \\ \text{given x copies in mother}}} \quad [1]$$

$$= \underbrace{CV_L^2 - CV_x^2}_{\substack{\text{Population variation} \\ \text{added at division}}},$$

where the brackets $\langle \ldots \rangle$ denote averages over all dividing cells in the population, regardless of the distribution of x values, and σ and CV denote the standard deviation and the coefficient of variation, respectively. These three perspectives on the same statistical measure show how partitioning errors capture statistical differences between daughters, the perturbation in levels introduced at cell division, and the contributions to population heterogeneity introduced by cell division.

In the simplest example where molecules are statistically independent and each copy goes to either daughter with probability ½, the partitioning error over the whole population simply decreases with the square root of the number of molecules, $Q_x = 1/\sqrt{\langle x \rangle}$, regardless of fluctuations in x across cells. Because cells are not homogenous well-stirred reactors, for example due to active transport or molecular crowding (15–17), independent segregation may seem unlikely. However, spatial distributions can be uniform even in crowded environments, and even nonuniform spatial distributions can create independent partitioning statistics. Phenotypes consistent with independent segregation have indeed been reported for several cellular components despite complicated molecular mechanisms, e.g., endosomes and lysosomes (18) and symbiotic *Chlorella* cells (19). For macromolecules, one *Escherichia coli* study (20) suggested independent protein segregation, and another (6) showed that engineered mRNA transcripts follow independent or weakly disordered segregation, consistent with the observation that some transcripts clustered.

Many mathematical and computational models of stochastic processes in cells have included independent segregation of molecules (21–26), for example the pioneering gene expression

Fig. 1. Disordered segregation can increase partitioning errors greatly. (*A*) *Upper*: Variation in division site or random segregation of other large components causes fluctuations in daughter size or in the volume accessible to the segregating component. *Lower*: In simulations (symbols) x is picked from Poisson distributions and volume variation is generated by assuming either (truncated) Gaussian variation in the size (circles) or random segregation of another large component (triangle), where Q_{Vol} is determined numerically. The exact analytical curve is dashed. The shaded area corresponds to experimental results for many microbes where $3\% < Q_{Vol} < 7\%$ (27–29). (*B*) *Upper*: Segregating molecules (dots) are randomly grouped into vesicles (gray circles) and the vesicles are independently partitioned into daughter cells. *Lower*: Simulation results (symbols) and analytical expressions that are approximate (solid lines) and exact (dashed line) for cases of Eqs. **4** and **5**, respectively. For the case of Eq. **4**, x and v are sampled from Poisson distributions. For the case of Eq. **5**, the value of v is first sampled from a Poisson distribution and x is then sampled from a Poisson distribution with average sv for each given v, so that $q = 1$; hence the second term of Eq. **5** is zero in this example. Simulations using different average $\langle x \rangle$ produce indistinguishable plots for virtually all values of $\langle x \rangle$ (*SI Text*).
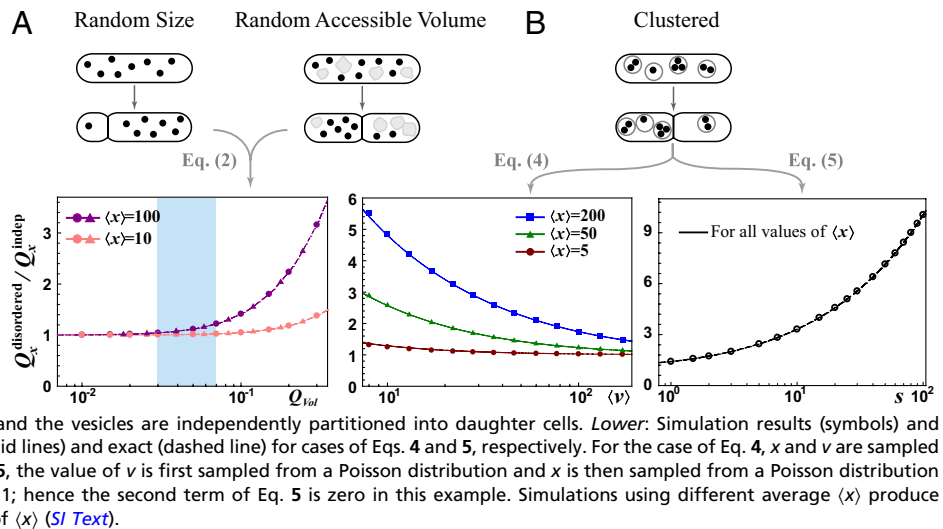
analyses by Berg (22) and Rigney (23), but because independent segregation directly implies binomial partitioning errors, such models are not used to study partitioning itself but rather to provide realistic contexts for studying other mechanisms. Here we analyze stochastic processes that account for nontrivial

Box 1.

To increase the applicability of the models, we specify the probabilistic rules of partitioning as broadly as possible. For example, in the trivial case of independent partitioning we assume only that each molecule has an independent probability of being in either cell half. Such independence can be physically realized in many different ways, for example by having a well-mixed cytoplasm, by making immobile molecules with independent probabilities in either cell half, or by active segregation mechanisms that randomly pick molecules and move them to either cell half. However, the derivations of partitioning errors are often more tractable by picking one specific dynamic model that exactly instantiates the general assumptions. The detailed Markov processes (*SI Text*) used here are thus merely "mock processes" that provably generate the correct partitioning errors corresponding to the assumptions in the main text. In a two-component model where one daughter has L copies of the component of interest (total of x) and n copies of some upstream factor (total of v), one such mock process may include the jumps

$$L \xrightarrow{f_{\text{jump}}(n,L)} L+1 \quad \text{and} \quad n \xrightarrow{g_{\text{jump}}(n)} n+1$$
$$L \xrightarrow{f_{\text{jump}}(v-n,x-L)} L-1 \qquad n \xrightarrow{g_{\text{jump}}(v-n)} n-1, \qquad \textbf{[B1]}$$

and the corresponding Markov process is then solved for the distribution or variance across cells. For independent partitioning we could use $f_{\text{jump}} = x - L$, which gives the expected $Q_x^2 = 1/\langle x \rangle$.

To further account for arbitrary heterogeneity in the total numbers x and v between dividing cells, we apply the law of total variance to separate the error given x and v from the error in x and v themselves, i.e., $\sigma^2(L) = \sigma^2(\langle L \mid x, v \rangle) + \langle \sigma^2(L \mid x, v) \rangle = \sigma^2(x)/4 + \langle \sigma^2(L \mid x, v) \rangle$, where normalizing by $\langle L \rangle^2 = \langle x \rangle^2/4$ gives Eq. **1** (*SI Text*).

mechanisms as well as arbitrary intrinsic and extrinsic fluctuations in x (Box 1). Many results are analytically exact (denoted '='), and most approximations (denoted '≈') are highly accurate as shown by comparing with exact simulations.

Disordered Segregation Can Greatly Increase Partitioning Errors.
Partitioning errors can be substantial even for components present in high numbers due to upstream sources of randomness. For example, the cytoplasmic volume available to a component can vary between the two daughters either because of errors in the position of the septum or because large cytoplasm-excluding structures such as vacuoles segregate randomly (Fig. 1*A*). The individual copies then become statistically dependent even if they never interact, increasing the partitioning error. Here we make the single assumption that each molecule independently occupies a cell half with a probability that is proportional to the available volume in that half and that the molecules take up negligible space. The exact partitioning error then follows:

$$Q_x^2 = \frac{1 - Q_{\text{Vol}}^2}{\langle x \rangle} + Q_{\text{Vol}}^2 (\text{CV}_x^2 + 1), \qquad \textbf{[2]}$$

where Q_{Vol} is the partitioning error of relative available volume. The generality of this result allows us to evaluate phenotypic data where mechanisms are unknown. Accounting for the unequal cell division measured for several microbes (27–29) shows that this volume variation should have a minor impact on partitioning errors (Fig. 1*A*).

Many molecules also form complexes and clusters or segregate in vesicles or organelles (Fig. 1*B*). The partitioning error then depends both on the distribution of copies among groups and on the segregation of the groups themselves. We assume that (*i*) each molecule (x copies before division) segregates independently into vesicles (v copies before division) and (*ii*) each vesicle segregates independently to either daughter (*SI Text*). The partitioning error then exactly follows:

$$Q_x^2 = \frac{1}{\langle x \rangle} - \frac{1}{\langle x \rangle^2} \left\langle \frac{x}{v} \right\rangle + \frac{1}{\langle x \rangle^2} \left\langle \frac{x^2}{v} \right\rangle. \qquad \textbf{[3]}$$

Inspecting Eq. **3** shows that the partitioning error further depends on the correlations between x and v across mother cells. We consider two cases to illustrate the effect. First, if x and v are statistically independent, the partitioning error follows

$$Q_x^2 \approx \frac{1}{\langle x \rangle} + \frac{(1 + \mathrm{CV}_x^2)(1 + \mathrm{CV}_v^2)}{\langle v \rangle}, \qquad [4]$$

which is accurate for a wide range of parameters (Fig. 1B). Second, if the average number of molecules for a given number of vesicles is proportional to the number of vesicles, $\langle x \mid v \rangle = sv$, so that s corresponds to the average number of molecules per vesicle, the partitioning error exactly follows

$$Q_x^2 = \frac{1+s}{\langle x \rangle} + \frac{s}{\langle x \rangle^2}(q-1), \qquad [5]$$

where $q = \langle \sigma^2(x|v)/\langle x|v \rangle \rangle$, and the second term typically is negligible. In both cases, clustered segregation could increase segregation errors greatly for small v or large s (Fig. 1B). The seemingly structured nature of many cell types—with numerous complexes, clusters, and organelles of varying sizes—can thus have an enormous randomizing effect at cell division. In the extreme where every molecule segregates to the same daughter, $Q_x^2 = 1 + \mathrm{CV}_x^2$ (*SI Text*).

There are virtually no quantitative studies of disordered segregation (the conceptually similar "extrinsic noise" was only recently (7, 30) quantified in genetic networks) but these results suggest a substantial impact on heterogeneity. However, some partitioning mechanisms may also be less disordered than they appear. For example, the Min proteins that help control cell division in *E. coli* oscillate spatially between cell poles (31), suggesting that the majority of proteins would end up in the same cell half after septation. But the system is also presumably under selection to avoid large fluctuations that could interfere with control, and recent experiments (32, 33) indeed show surprisingly small partitioning errors.

Ordered Segregation Requires Extreme Parameters to Greatly Suppress Partitioning Errors. For components present in low numbers cells must either tolerate large partitioning errors or use control mechanisms to reduce them. Because little is known quantitatively about the mechanisms we consider three principles: size exclusion, binding to spindles or other cellular sites, and pair formation.

Cells could reduce partitioning errors passively if the segregating components occupy sufficiently large volume, which would create more available space in the cell half with fewer copies (Fig. 2A) (34). For very large components the error reduction would depend on exact shapes, but because even organelles are fairly small compared to the total cell volume, we assume (*i*) each individual component occupies a cell half with a probability that is proportional to the available volume, (*ii*) each molecule takes up a fraction K of the total cell volume, and (*iii*) the two cell halves have equal available volumes except for the location of component X studied. Such mechanisms produce a partitioning error

$$Q_x^2 \approx \frac{1}{\langle x \rangle}\big(1 - K\langle x \rangle (\mathrm{CV}_x^2 + 1)\big), \qquad [6]$$

where $K\langle x \rangle$ is the average total fraction of cell volume occupied by all segregating X units (*SI Text*). However, the effect is significant only if the organelles take up a large total fraction (Fig. 2A) and, except for chromosomes in some organisms, Eq. **6** suggests that most cell components occupy too small total volumes for self volume exclusion to substantially affect partitioning errors (34). Eq. **6** also shows that partitioning errors decrease with the variation in organelle numbers, because a larger fraction of cells then have so many organelles that volume exclusion is significant.

Even if organelles hitchhiked with chromosomal segregation to reduce partitioning errors (Fig. 2B), for example by binding to mitotic spindles or spindle poles (18, 35), accurate partitioning is still hard to achieve. In cells that by chance have fewer binding sites v than organelles x, the unbound organelles will segregate randomly, and in cells with fewer organelles than binding sites, more organelles could bind to sites on one side than the other (1). To demonstrate the inaccuracy even in an idealized case we assume that (*i*) free organelles segregate independently and (*ii*) organelles always bind available sites, with no cells simultaneously having free organelles and free sites. The partitioning error then follows

$$Q_x^2 \approx \frac{1}{\langle x \rangle}\Bigg(1 - \frac{1}{\langle x \rangle} \sum_{v=0}^{\infty}\Bigg[\bigg(\frac{1}{v} - Q_v^2\bigg)\sum_{x=0}^{v} x^2 P(x,v) \\ + (v - v^2 Q_v^2)\sum_{x=v+1}^{\infty} P(x,v)\Bigg]\Bigg), \qquad [7]$$

where $P(x, v)$ is the probability of x organelles and v binding sites across dividing cells, and Q_v is the partitioning error of the binding sites for a fixed v. Analyzing this equation further (Fig. 2B) shows that errors can be reduced even when the numbers are not perfectly matched—more organelles bound to sites in one cell half means relatively more free sites in the other half—but also that the effect is small. For example, even if the averages are perfectly matched, $\langle v \rangle = \langle x \rangle = 100$, each daughter cell receives exactly $v/2$ binding sites, and the only sources of noise in x and v are independent low-abundance Poisson fluctuations, the error Q_x would be reduced less than threefold compared to uncontrolled, independent partitioning (Fig. 2B). Control works much worse yet when the averages $\langle x \rangle$ and $\langle v \rangle$ are lower or unequal, the organelles bind weakly, or the independent fluctuations are larger—with the slight twist that if the averages are unequal, even uncorrelated fluctuations could slightly improve control by making it more likely that the numbers are matched in at least some fraction of the individual cells, i.e., ensuring that sometimes $x \approx v$ even if $\langle x \rangle \neq \langle v \rangle$. With v instead as specific DNA binding sites, so that v may be approximated as fixed and without partitioning error, Eq. **7** simplifies greatly. If we further assume that distribution $P(x)$ is approximately symmetric, the partitioning error can be approximated (*SI Text*) by

$$Q_x^2 \approx \frac{1}{\langle x \rangle}\bigg(\frac{\langle |x - \langle x \rangle| \rangle}{\langle x \rangle} - \frac{\mathrm{CV}_x^2}{2}\bigg),$$

where for all symmetric distributions $\mathrm{CV}_x \leq 1$, since $x \geq 0$. For example, if both the CV and the normalized absolute deviation are ~10%, the binding sites can reduce errors only threefold and the effect is smaller yet with increasing fluctuations in x.

The accuracy of chromosomal partitioning instead comes from the fact that the individual molecules form pairs with each other, using spindles to split each pair into separate daughter cells. The underlying molecular mechanisms are highly complex for chromosomes, but they are more tractable for, e.g., plasmid R1, where actin-like ParM proteins are attached to a pair of R1 plasmids via the ParR protein and push each plasmid to the opposite poles of *E. coli* (36, 37). Here we model the accuracy of such systems in terms of the phenotypic pairing parameters (Fig. 2C). We assume that (*i*) molecules form pairs by some arbitrary mechanism, with k as the average fraction of paired molecules that may depend on x; (*ii*) the two units of each pair are separated into different daughters with probability p and into the same daughter with probability $1 - p$; and (*iii*) unpaired molecules (fraction $1 - k$) segregate independently. For all such mechanisms—regardless of pairing mechanism—the partitioning error exactly follows

Fig. 2. Ordered segregation requires extreme parameters to substantially reduce partitioning errors. (*A*) *Upper*: Segregating units with non-negligible size (gray circles) exclude each other and thereby promote more even segregation. *Lower*: Simulation results (symbols) and analytical approximation (solid lines) plotted as a function of the average total volume fraction occupied by the segregating units, with *x* sampled from Poisson distributions for each $\langle x \rangle$ (circles and triangles) or assuming that synthesis occurs in a geometric burst with average burst size of 10 (diamonds). All results are truncated to ensure $Kx < 1$. The shaded area corresponds to the physiological regime for most cellular

components (34) and shows that the effect should be minor. (*B*) *Upper*: Organelles (dots) compete for available binding sites (ends of astral, gray) and unbound organelles are partitioned independently. *Lower*: Simulation results (symbols) and analytical approximations (solid lines) when binding sites are distributed evenly between the two daughters. The circles and triangles correspond to *v* and *x* sampled from Poisson distributions, using $\langle v \rangle = 100$, and diamonds show *x* and *v* sampled from processes where synthesis occurs in a geometrically distributed burst, with average burst sizes of 20 and 10 for *v* and *x*, respectively. The results show how modest the effects are when *x* and *v* are not exactly matched. (*C*) *Upper*: Among $(x - \delta_{x,odd})/2$ possible pairs in each cell, a fraction *r* of molecules (dots) binomially forms pairs, where $\delta_{x,odd} = 1$ if *x* is odd and zero otherwise. Paired molecules segregate separately with probability *p* whereas unpaired molecules segregate independently. Under these assumptions, $k = r(1 - P_o/\langle x \rangle)$, where P_o is the probability that *x* is an odd number. *Lower*: Simulations (symbols), exact analytical results (Eq. 8, dashed line), and analytical approximations ($k = r$ in Eq. 8, solid line) when *x* is sampled from a Poisson distribution of average 10. The results show that both *r* and *p* must approach 100% for efficient control. Derivations are given in *SI Text*.
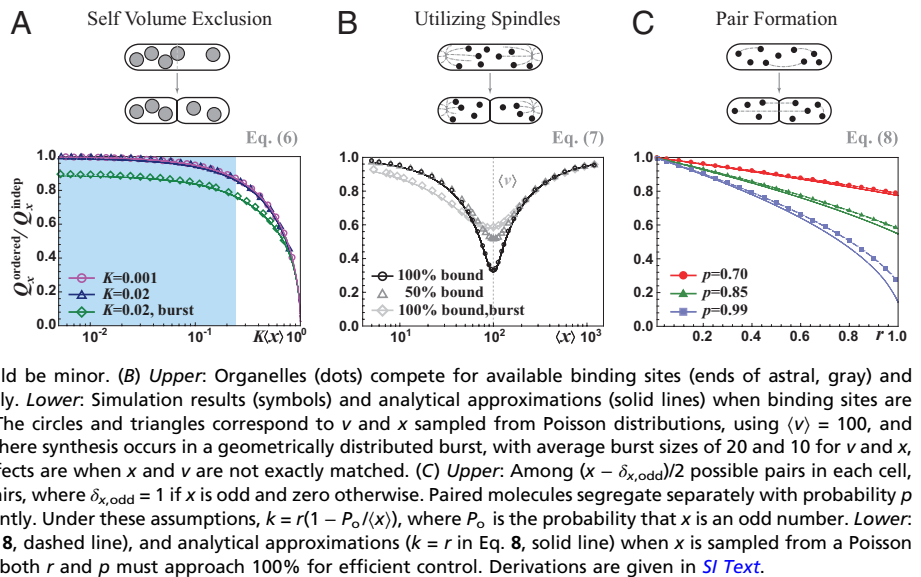
$$Q_x^2 = \frac{1 - (2p - 1)k}{\langle x \rangle}. \qquad [8]$$

As seen from Fig. 2*C*, substantially reducing the errors is again highly nontrivial. For example, a fivefold reduction of the error requires both pairing and splitting probabilities of at least 90–95%. Chromosomes achieve extremely high *k* and *p*, using elaborate control mechanisms, but for most other cellular components such measures are perhaps not cost-effective.

Despite the challenges above, some studies suggest that segregation indeed is moderately ordered for several organelles. For example, during mitosis the chloroplasts in mesophyll protoplasts of *Nicotiana tabacum* reportedly use actin filaments to segregate more accurately than expected from independent segregation (38), and similar effects have been observed for mitochondrial inheritance in scorpion spermatogenesis (39–41). Studies of Golgi particles in mitotic HeLa cells in turn showed twofold smaller Q_x compared with independent segregation for $\langle x \rangle \approx 130$ (42), whereas chloroplasts in unicellular algae (43) showed a threefold reduction of Q_x given the measured $\langle x \rangle \approx 20$. We know of no quantitative measurement showing ordered partitioning of RNAs or proteins. Some RNAs bind to spindles (44), but also seem to form clusters of random size, with unknown net accuracy. The most precise non-DNA segregation mechanism reported to our knowledge is for carboxysomes in cyanobacteria (45), which achieve a fivefold reduction in Q_x compared with independent segregation, for $x = 6$.

To Suppress Partitioning Errors, Each Source of Randomness Needs Separate Control. For disordered segregation the task of reducing partitioning errors gets harder yet, not just because of the larger errors but also because fluctuations are introduced at multiple levels. For gene expression, multiple sources of noise can be suppressed by a single feedback loop (46) because deviations in abundances can trigger homeostatic responses regardless of what caused the deviations. Disordered segregation, however, as when proteins are sorted into clusters that then are sorted into the two cell halves, is fundamentally different. Even if the first sorting process is perfectly controlled by a homeostatic feedback system, the randomness of cluster inheritance can still produce large protein partitioning errors. The same is true if the second sorting is perfect but the first is not.

For a quantitative example we assume cooperative control over the sorting process of proteins into clusters and over the

sorting of vesicles into each daughter cell. Specifically we assume that each molecule in a vesicle containing x_1 copies migrates randomly to any other vesicle with a probability proportional to $x_1^{h_1}$ and that each vesicle migrates to the other cell half with a probability proportional to $x_2^{h_2}$. Further assuming $\langle x \mid v \rangle = sv$ as in Fig. 1*B*, the partitioning error approximately follows

$$Q_x^2 \approx \left(\frac{1}{h_1} + \frac{s}{h_2} \right) \times \frac{1}{\langle x \rangle}, \qquad [9]$$

where *s* is the average number of proteins per cluster. Because the two noise terms are inversely proportional to their respective sorting efficiencies, and the partitioning error Q_x is proportional to the square root of their sum, both processes must thus be separately and tightly controlled for accurate segregation. The same is true for most combinations of disordered and ordered segregation above: As opposed to negative feedback control of production rates, control of segregation corrects only errors produced by the process being controlled.

Does Segregation Control Serve to Reduce the Risk of Complete Loss? The results above raise the question of why ordered segregation would evolve if its effects on the heterogeneity are marginal anyway. For the many stable components that serve as templates either for their own production (e.g., mitochondria, chloroplasts, and plasmids) or where *de novo* production is slow [e.g., peroxisomes (47), centrioles (48), or bacterial flagella (49)], the greatest risk may be that all copies segregate to the same daughter. In fact, though ordered partitioning seems to have a small effect on standard deviations—forcing cells to tolerate some fluctuations—such mechanisms can still greatly reduce the risk of complete loss (Fig. 3). Reducing losses at cell division may then be the main selective driving force for ordered partitioning.

Segregation Mechanisms Are Hard to Infer Experimentally. Measurement errors could make precise mechanisms appear imprecise, whereas ignoring legitimate outliers has the opposite effect. In most organelle studies, where it is possible to count individual components, another potential problem is the combination of crowding and low-resolution microscopy, producing systematic tendencies to undercount components in cells that by chance have a high abundance and thereby making mechanisms appear more precise than they actually are. One solution to this problem is to use quantitative microscopy to determine total abundances
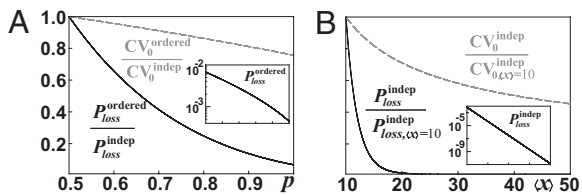
SYSTEMS BIOLOGY

Fig. 3. Ordered segregation and increased averages can dramatically reduce the rate of complete loss. (A) Relative reduction of the probability P_{loss} that all units segregate to the same daughter (solid line), as well as CV_0 (dashed line), comparing ordered vs. independent segregation. We assume that 90% of the units form pairs ($r = 0.9$ model in Fig. 2C), and x is sampled from a Poisson distribution with mean 10. (B) Relative reduction of the loss rate and CV_0 with the average number of molecules, again calculated assuming a Poisson distribution for x. Insets: Absolute loss rates P_{loss} as functions of p and $\langle x \rangle$, respectively. Derivations are given in SI Text.

and then test how the partitioning errors scale with the abundance in the mother cell. However, the error Q_x is proportional to the inverse square root of the average for several (but not all) of the mechanisms above, where it is the proportionality constant that represents the accuracy of segregation,

$$Q_x^2 \approx \frac{A}{\langle x \rangle}, \text{ where } \begin{cases} A < 1, \text{ ordered partitioning} \\ A = 1, \text{ independent partitioning} \\ A > 1, \text{ disordered partitioning} \end{cases} \quad \textbf{[10]}$$

as in Figs. 2 and 3. The problem is that if levels are reported in arbitrary units of fluorescence rather than in integer numbers of molecules, as in most fluorescent protein studies and several measurements of macronuclear DNA in ciliates (50, 51), then A and $\langle x \rangle$ are not known separately. Very different partitioning mechanisms—from the very ordered to the very disordered—can then display exactly the same scaling behavior and fit the same experiments (Fig. 4A). For example, the only study to our knowledge that quantified protein segregation statistics (20) showed that the partitioning error for a CI-YFP fusion protein in the λ-cascade of E. coli was inversely proportional to the abundance in the mother cell and assumed independent segregation ($A = 1$) to infer the average number of proteins. Assuming independent partitioning seems plausible: Most cytoplasmic proteins in E. coli probably follow independent segregation, and even if many CI-YFP molecules would be unspecifically bound to the chromosomes, these copies may also effectively segregate independently. However, it is important to note that such experiments *assume* rather than *demonstrate* independent segregation. If proteins formed clusters or were subject to ordered segregation, the fits could be equally good (Eq. **10**) and the inferred averages would be incorrect. Both counting and quantification strategies may thus generate qualitatively misleading results, and segregation measurements demand the same attention as, e.g., gene expression noise.

When the integer numbers of molecules cannot be directly counted, the type of segregation mechanism can be alternatively inferred by introducing an additional binomial source of noise (Fig. 4B). For example, a protein could be translationally fused to a photoconvertible fluorescent protein (PC-FP). If each PC-FP has an independent probability $0 < u < 1$ of photoconversion for a certain amount of light (which should be demonstrated), the partitioning error of the photoconverted proteins (l and r copies in each daughter cell) follows (SI Text)

$$\langle (l-r)^2 \rangle = \left[\langle x \rangle + \left(\langle (L-R)^2 \rangle - \langle x \rangle \right) u \right] u. \quad \textbf{[11]}$$

The partitioning error of the fluorescence from the photoconverted PC-FP, which is proportional to $\langle (l-r)^2 \rangle$, is thus
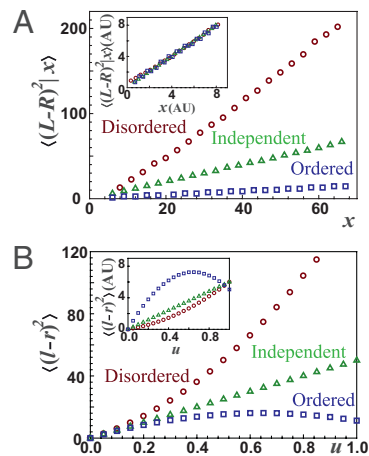

Fig. 4. Segregation mechanisms are difficult to infer. (A) Examples of ordered, disordered, and independent segregation mechanisms were simulated, and the conditional mean-squared error, given x copies in the mother cell, was plotted as a function of x. This measure is closely related to the partitioning error and was used to interpret experiments for protein segregation in E. coli (20). To illustrate disordered segregation we used the second case of Fig. 1B (Eq. 5) where the number v of vesicles is sampled from a Poisson distribution with average of 25, and x is sampled from a Poisson distribution with average $2v$ for each given v; hence $\langle x \mid v \rangle = 2v$. To illustrate ordered segregation we used the pairing mechanism of Fig. 2C, with $r = 0.8$ and $p = 0.99$. All three curves are nearly perfectly approximated by $\langle (L - R)^2 \mid x \rangle = ax$. Inset: When levels are measured in arbitrary units, and the three mechanisms are assayed in separate experiments (different units), the curves are indistinguishable and can be laid on top of each other. Each mechanism thus fits the other datasets equally well. (B) Segregation mechanisms can be distinguished without counting the units by utilizing photoconvertible proteins. As the fraction of conversion (u) increases, the partitioning errors of the photoconverted molecules show convex, concave, and straight curves for disordered, ordered, and independent partitioning, respectively. All parameters are the same as in A with $\langle x \rangle = 50$. Inset: The same scaling as in the Inset of A is applied to make the fluorescent level arbitrary, and the shapes of the curves allow us to distinguish the partitioning mechanism.

a parabolic function of u, controllable by light. For disordered protein segregation the curve will be convex ($\langle (L - R)^2 \rangle > \langle x \rangle$), for ordered segregation it will be concave ($\langle (L - R)^2 \rangle < \langle x \rangle$), and for independent segregation it will be a straight line ($\langle (L - R)^2 \rangle = \langle x \rangle$). It is not necessary to measure the conversion fraction u: The shape of the curve is sufficient, as long as u is not too small. If u is also measured, which is possible with PC-FP but difficult with photoactivatable proteins, we can further fit the curve to a parabola and calculate $\langle (L - R)^2 \rangle / \langle x \rangle$ by $u_{\min,\max} = (2 - 2\langle (L - R)^2 \rangle / \langle x \rangle)^{-1}$, which gives the value of A for the cases in Eq. **10**. The intuition behind this effect is that for ordered partitioning, binomial sampling introduces significant noise relative to the original fluctuations when u is small so that $\langle (l - r)^2 \rangle$ overestimates the real partitioning error, but as u approaches 1 this additional noise is eliminated, and hence Eq. **11** becomes concave. For disordered partitioning, the dominant binomial sampling errors at low u instead hides a large natural error and hence $\langle (l - r)^2 \rangle$ underestimates the actual partitioning error, an effect that again disappears as u approaches 1. The curve of the partitioning error as a function of u then becomes convex. In practice, the highest value of u that can be achieved with many current photoconvertible proteins may be in the range where high-quality data are required to practically use this approach. However, more such proteins are being engineered and higher u can be achieved by photobleaching the FPs.

Outlook

Appropriate segregation of components at cell division is key to cell proliferation. However, although the heterogeneity caused by random births and deaths of molecules during the cell cycle has received enormous attention in recent years, the heterogeneity caused by partitioning errors has been virtually ignored in comparison. Many theoretical studies consider independent segregation and binomial errors. But just as gene expression is poorly approximated by a homogenous Poisson process due to feedback control or upstream sources of heterogeneity, cells are not bags of enzymes where all individual molecules segregate independently, but rather contain complexes, vesicles, cytoskeletons, and organelles. By considering these features we show how surprisingly hard it is for cells to eliminate partitioning errors through controlled segregation, how easily such errors are amplified by upstream factors, and how difficult it is to interpret seemingly direct partitioning measurements.

In addition to identifying these effects, we systematically investigate what patterns to expect from partitioning and how to more effectively design experiments. So far, measurements have been possible for only a handful of systems—often in obscure and poorly characterized processes and often with too low accuracy to discriminate among different types of segregation mechanisms—whereas the partitioning errors of crucial proteins and RNAs in model organisms have barely been measured at all. Although these studies are very interesting in their own right, not all components that have been counted thus necessarily count so much toward our general understanding of segregation, whereas the components that count cannot necessarily be counted (to paraphrase an alleged Einstein quote). This problem is likely to change in the near future. Several of the major experimental challenges have recently been overcome (6, 12, 52) and much of the necessary mathematical theory is now in place to address the central questions—How random is segregation and how much does it contribute to cellular heterogeneity?—quantifying one of the most central aspects of life at the level of individual cells.

1. Warren G (1993) Membrane partitioning during cell division. *Annu Rev Biochem* 62: 323–348.
2. Marshall WF (2007) Stability and robustness of an organelle number control system: Modeling and measuring homeostatic regulation of centriole abundance. *Biophys J* 93:1818–1833.
3. Birky CW, Jr. (1983) The partitioning of cytoplasmic organelles at cell division. *Int Rev Cytol Suppl* 15:49–89.
4. Nordström K, Austin SJ (1989) Mechanisms that contribute to the stable segregation of plasmids. *Annu Rev Genet* 23:37–69.
5. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31:69–73.
6. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* 123:1025–1036.
7. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
8. Raj A, Rifkin SA, Andersen E, van Oudenaarden A (2010) Variability in gene expression underlies incomplete penetrance. *Nature* 463:913–918.
9. Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. *Nature* 405:590–593.
10. Süel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB (2006) An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440:545–550.
11. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459:428–432.
12. Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. *Science* 311:1600–1603.
13. Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440:358–362.
14. Huh D, Paulsson J (2011) Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet* 43:95–100.
15. Weiss M, Elsner M, Kartberg F, Nilsson T (2004) Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophys J* 87:3518–3524.
16. Kulkarni RP, Castelino K, Majumdar A, Fraser SE (2006) Intracellular transport dynamics of endosomes containing DNA polyplexes along the microtubule network. *Biophys J* 90:L42–L44.
17. Golding I, Cox EC (2006) Physical nature of bacterial cytoplasm. *Phys Rev Lett* 96: 098102.
18. Bergeland T, Widerberg J, Bakke O, Nordeng TW (2001) Mitotic partitioning of endosomes and lysosomes. *Curr Biol* 11:644–651.
19. McAuley PJ (1990) Partitioning of symbiotic Chlorella at host cell telophase in the green hydra symbiosis. *Philos Trans R Soc Lond B Biol Sci* 329:47–53.
20. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.
21. Brenner N, Farkash K, Braun E (2006) Dynamics of protein distributions in cell populations. *Phys Biol* 3:172–182.
22. Berg OG (1978) A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol* 71:587–603.
23. Rigney DR (1979) Stochastic model of constitutive protein levels in growing and dividing bacterial cells. *J Theor Biol* 76:453–480.
24. Paulsson J, Ehrenberg M (2001) Noise in a minimal regulatory network: Plasmid copy number control. *Q Rev Biophys* 34:1–59.
25. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99:12795–12800.
26. Rausenberger J, Kollmann M (2008) Quantifying origins of cell-to-cell variations in gene expression. *Biophys J* 95:4523–4528.
27. Tyson JJ (1989) Effects of asymmetric division on a stochastic model of the cell division cycle. *Math Biosci* 96:165–184.
28. Nanninga N, Koppes LJH, de Vries-Tijssen FC (1979) The cell cycle of Bacillus subtilis as studied by electron microscopy. *Arch Microbiol* 123:173–181.
29. Trueba FJ (1982) On the precision and accuracy achieved by Escherichia coli cells at fission about their middle. *Arch Microbiol* 131:55–59.
30. Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304:1811–1814.
31. Raskin DM, de Boer PA (1999) Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*. *Proc Natl Acad Sci USA* 96: 4971–4976.
32. Juarez JR, Margolin W (2010) Changes in the Min oscillation pattern before and after cell birth. *J Bacteriol* 192:4134–4142.
33. Di Ventura B, Sourjik V (2011) Self-organized partitioning of dynamically localized proteins in bacterial cell division. *Mol Syst Biol* 7:457.
34. Birky CW, Skavaril RV (1984) Random partitioning of cytoplasmic organelles at cell division: The effect of organelle and cell volume. *J Theor Biol* 106:441–447.
35. Shima DT, Cabrera-Poch N, Pepperkok R, Warren G (1998) An ordered inheritance strategy for the Golgi apparatus: Visualization of mitotic disassembly reveals a role for the mitotic spindle. *J Cell Biol* 141:955–966.
36. Møller-Jensen J, et al. (2003) Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Mol Cell* 12:1477–1487.
37. Garner EC, Campbell CS, Weibel DB, Mullins RD (2007) Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. *Science* 315:1270–1274.
38. Sheahan MB, Rose RJ, McCurdy DW (2004) Organelle inheritance in plant cell division: The actin cytoskeleton is required for unbiased inheritance of chloroplasts, mitochondria and endoplasmic reticulum in dividing protoplasts. *Plant J* 37:379–390.
39. Wilson EB (1916) The distribution of the chondriosomes to the spermatozoa in scorpions. *Proc Natl Acad Sci USA* 2:321–324.
40. Wilson EB (1931) The distribution of sperm-forming materials in scorpions. *J Morphol Physiol* 52:429–483.
41. Hood RD, Watson OF, Deason TR, Benton CLB, Jr. (1972) Ultrastructure of scorpion spermatozoa with atypical axonemes. *Cytobios* 5:167–177.
42. Shima DT, Haldar K, Pepperkok R, Watson R, Warren G (1997) Partitioning of the Golgi apparatus during mitosis in living HeLa cells. *J Cell Biol* 137:1211–1228.
43. Hennis AS, Birky CW, Jr. (1984) Stochastic partitioning of chloroplasts at cell division in the alga Olisthodiscus, and compensating control of chloroplast replication. *J Cell Sci* 70:1–15.
44. Blower MD, Feric E, Weis K, Heald R (2007) Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules. *J Cell Biol* 179: 1365–1373.
45. Savage DF, Afonso B, Chen AH, Silver PA (2010) Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* 327:1258–1261.
46. Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–418.
47. Motley AM, Hettema EH (2007) Yeast peroxisomes multiply by growth and division. *J Cell Biol* 178:399–410.
48. Marshall WF, Vucica Y, Rosenbaum JL (2001) Kinetics and regulation of de novo centriole assembly. Implications for the mechanism of centriole duplication. *Curr Biol* 11:308–317.
49. Neidhardt FC (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology* (American Society for Microbiology, Washington, DC), p 123.
50. Doerder FP (1979) Regulation of macronuclear DNA content in Tetrahymena thermophila. *J Protozool* 26:28–35.
51. Berger JD (1979) Regulation of macronuclear DNA content in Paramecium tetraurelia. *J Protozool* 26:18–28.
52. Huang B, et al. (2007) Counting low-copy number proteins in a single cell. *Science* 315: 81–84.

SYSTEMS BIOLOGY