Optimizing ring assembly reveals the strength of weak interactions

Eric J. Deeds[a,1,2], John A. Bachman[b,1], and Walter Fontana[b,2]

[a]Center for Bioinformatics and Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047; and [b]Department of Systems Biology, Harvard Medical School, Boston, MA 02115

Edited by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved December 27, 2011 (received for review August 13, 2011)

Most cellular processes rely on large multiprotein complexes that must assemble into a well-defined quaternary structure in order to function. A number of prominent examples, including the 20S core particle of the proteasome and the AAA+ family of ATPases, contain ring-like structures. Developing an understanding of the complex assembly pathways employed by ring-like structures requires a characterization of the problems these pathways have had to overcome as they evolved. In this work, we use computational models to uncover one such problem: a deadlocked plateau in the assembly dynamics. When the molecular interactions between subunits are too strong, this plateau leads to significant delays in assembly and a reduction in steady-state yield. Conversely, if the interactions are too weak, assembly delays are caused by the instability of crucial intermediates. Intermediate affinities thus maximize the efficiency of assembly for homomeric ring-like structures. In the case of heteromeric rings, we find that rings including at least one weak interaction can assemble efficiently and robustly. Estimation of affinities from solved structures of ring-like complexes indicates that heteromeric rings tend to contain a weak interaction, confirming our prediction. In addition to providing an evolutionary rationale for structural features of rings, our work forms the basis for understanding the complex assembly pathways of stacked rings like the proteasome and suggests principles that would aid in the design of synthetic ring-like structures that self-assemble efficiently.

computational modeling | kinetic deadlock | ring complexes | self-assembly kinetics | glassy dynamics

The vast majority of cellular processes, from signal transduction to the synthesis and degradation of polypeptide chains, rely on the action of large macromolecular complexes (1). In order to carry out their functions, these complexes must adopt a well-defined quaternary structure (1–5). The efficient and effective assembly of these structures from a set of monomeric subunits is thus critically important to living systems. Although experimental work has revealed many details of complex assembly pathways (4–7), conceptual issues remain that are best understood through the analysis of models.

One such issue concerns the evolutionary pressures that have shaped assembly pathways. A similar question has arisen before in the theoretical study of protein folding (8, 9). In that case, it was helpful to consider a "null model," often called the "Levinthal paradox" (8), which immediately suggested a kinetic problem that protein sequences must overcome in order to fold quickly. Seen from that perspective, evolution has sculpted free energy landscapes that prevent the folding process from degenerating into a random search of conformational space (8, 9) or from producing overly stable intermediates (10–13). In the case of macromolecular assembly, the question revolves not around the free energy landscape that characterizes sequences that fold efficiently, but rather the evolution of the chemical potential landscape of a complex molecular interaction network that supports efficient assembly. If such "assembly landscapes" have been shaped by evolution, what problems have they evolved to overcome or avoid?

In this work, we begin to approach this question for a subset of macromolecular structures; namely, those consisting of rings. Rings represent a common "motif" in large macromolecular complexes (14), perhaps because of their general thermodynamic stability (15, 16) (see *SI Appendix*, Section 1) and their inherent symmetry. They are thus found in the context of signaling networks (e.g., the apoptosome; refs. 17 and 18), chaperones (e.g., GroEL; ref. 19), protein degradation [e.g., the proteasome (refs. 5 and 7), and ClpP (ref. 20) in bacteria], pore-forming endotoxins [e.g., the protective antigen (PA) of *Bacillus anthracis*; ref. 21], and many other biological processes. Previous studies employing assembly models of ring-like structures have focused on a few specific examples, such as ClpA (22) (an AAA+ family member) and the apoptosome (23). In this work, we focus on a simple but general model of ring assembly, a null model that allows us to illustrate a tension that arises between energetically local interactions and global topological constraints. The barriers induced by this tension can have a strong impact on assembly efficiency, and by understanding how such barriers can be overcome, we provide a basic insight into the evolutionary pressures that have shaped the assembly of a broad class of macromolecular structures.

Our principal finding is the existence of a "deadlocked plateau" in the assembly dynamics of rings and a simple strategy for avoiding it. Depending on the strengths of the molecular interactions between the subunits of the ring, this plateau can have a significant effect on the assembly efficiency of the structure. This is true both for cases in which assembly occurs from an initial condition in which all subunits are monomers, or when considering a steady-state scenario with constant synthesis of monomers and degradation/dilution of complexes. Assembly deadlocks are thus likely to exhibit significant evolutionary pressures on the interaction strengths in the ring. We have also found that, for heteromeric rings where the affinities between neighboring subunits can vary independently, inclusion of one or more "weak" interactions in the structure improves assembly efficiency dramatically. This computational observation leads us to predict that heteromeric ring-like structures will generally contain one interaction that is significantly weaker than the others. We tested this prediction by analyzing the solved structures of all heteromeric three-membered rings, and we found that the vast majority of them do in fact contain at least one weak interaction. Our work thus provides an evolutionary rationale for the structural features of ring-like complexes, in addition to suggesting simple principles that could prove useful in the design of self-assembling nanostructures (24).

Author contributions: E.J.D., J.A.B., and W.F. designed research; E.J.D. and J.A.B. performed research; E.J.D., J.A.B., and W.F. analyzed data; and E.J.D., J.A.B., and W.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]E.J.D. and J.A.B. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: deeds@ku.edu or walter@hms.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1113095109/-/DCSupplemental.

Results

Constructing a Model of Ring Assembly. The ring-like protein complexes we model in this work exhibit fairly rigid interaction geometries (such as the structure in Fig. 1*A*) and a well-defined number n of subunits. Assembly occurs due to binding reactions between intermediates ranging in size from 1 (monomers) to $n-1$. We account for the geometry of rings by distinguishing three cases for the association of two intermediates with lengths k and l: (*i*) If $k+l<n$, the binding reaction produces another intermediate species (Fig. 1*B*); these reactions occur with a uniform association rate α and a dissociation rate β that depends on the strength of the noncovalent bond being formed. (*ii*) If $k+l=n$, the interaction results in the formation of a ring structure; these reactions involve the (essentially) simultaneous formation of two interfaces (16) (Fig. 1*C*). Because rings are inherently very stable (15, 16) (see also *SI Appendix*, Section 1), the reverse rate for these reactions (γ) is generally very small. Ring formation is thus essentially irreversible for most of the parameters considered in this work, and the equilibrium yield of the ring approaches 100%. (*iii*) If $k+l>n$, the interaction would yield a protein complex with more than n components, resulting in a steric clash that prevents the interaction from occurring (Fig. 1*D*).

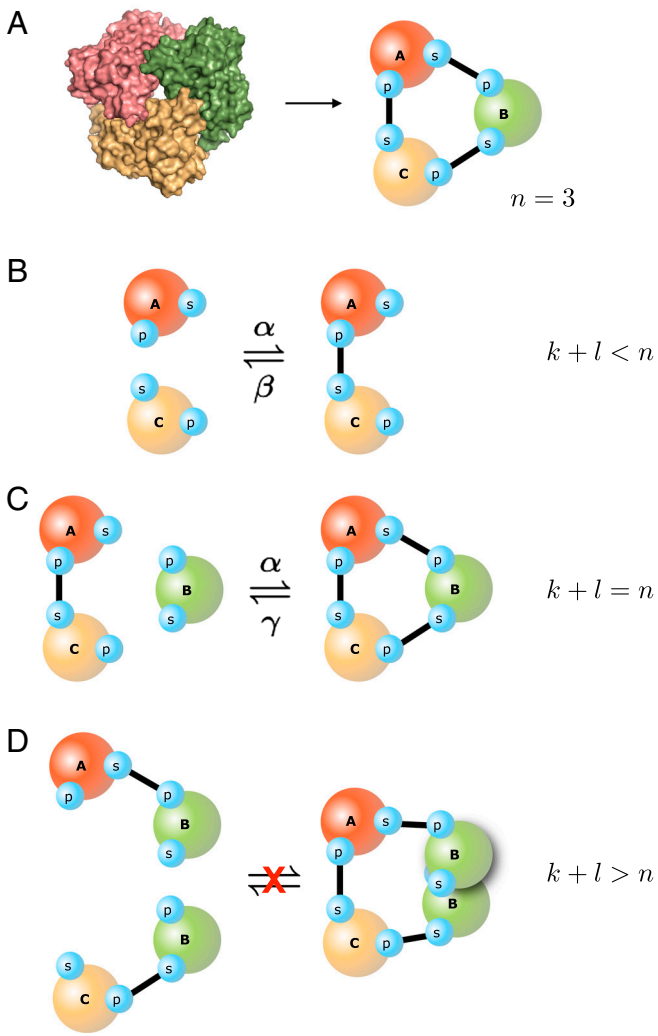

Fig. 1. Schematic of ring assembly. (*A*) A three-membered ring (X-ray structure from Protein Data Bank ID 2JB8) on the left is represented on the right as a graph involving three proteins, each with two binding interfaces (the small circles on the periphery of the nodes). (*B*) A pair of monomers bind to form a dimer, which represents the case where $k+l<n$. The forward (backward) rate constant of the interaction is denoted α (β). (*C*) A monomer binds to a dimer yielding the full three-membered ring, which represents the case where $k+l=n$. The forward rate of this reaction is taken to be α as in *B*, but the backward rate constant (γ) of the reaction is different. Because two interfaces are formed on the right-hand side of the reaction, $\gamma \ll \beta$ (see *SI Appendix*, Section 1). (*D*) Two dimers attempting to bind—i.e., $k+l>n$. These reactions do not occur because of steric hindrance.

The rings we consider also have a "sidedness"—that is, the monomers in the ring are not themselves internally symmetric; this characterizes most ring-like structures observed in nature (14). As a consequence, the subunits in our model have a distinct left and right side, and interactions can only occur between an interface on the right of one subunit and an interface on the left of another.

Our analysis largely focuses on the simplest case of ring assembly, where the parameters α, β, and γ in Fig. 1 depend only on the identity of the interface(s) involved. In this case, it is straightforward to derive from the reaction classes described above a system of ordinary differential equations (ODEs) describing the assembly dynamics of a ring of length n. The process for deriving these ODEs for both *homomeric* rings (where the individual subunits are indistinguishable) and *heteromeric* rings (where each subunit is distinct from every other subunit) is summarized in *Methods* and is described in detail in the *SI Appendix*, Section 2. The ODEs we obtain are integrated numerically using MATLAB (25) for a given set of parameters (i.e., monomer concentrations, association and dissociation rate constants; see *SI Appendix*, Section 2.6).

The Assembly Dynamics of Homomeric Rings. Fig. 2*A* depicts the assembly dynamics obtained from our model of a simple homomeric three-membered ring starting from an initial condition consisting of only monomers. When interaction affinities are very strong, the curves exhibit a characteristic "plateau." On very short timescales, monomers interact rapidly to form dimers; those dimers can subsequently interact with other monomers to form the full ring. After this initial phase, however, the monomers are depleted from the system but a significant concentration of dimers persists. Because these dimers cannot interact productively either with each other or with the full rings (Fig. 1*D*), the system is deadlocked until it reaches timescales on which dimers dissociate readily. At that point, monomers released by dissociation can interact with the remaining dimers, resulting in the formation of the full ring. For longer rings, the plateau occurs at lower concentrations of the full ring structure; thus, whereas approximately 65% of three-membered rings are formed in the plateau phase, only approximately 35% of seven-membered rings have formed at that point (see *SI Appendix*, Section 4.1.2 and Figs. S6 and S7).

The existence and duration of this deadlocked plateau strongly depends on the parameters of the system. To quantify that dependence, we considered the time T_X it takes a system with an initial condition of 100% monomers (all at equal concentration) to reach a state where X% of monomers are found in rings. We have plotted T_{99} as a function of the uniform interaction strength along the ring (represented by the dissociation constant for those interactions, $K_d \equiv \beta/\alpha$) for various initial monomer concentrations (Fig. 2*B*). Each concentration exhibits an affinity that minimizes the time to 99% yield (i.e., T_{99}). Stronger interactions result in considerably longer assembly times: In this "dissociation-challenged" regime, the duration of the deadlocked plateau increases with increasing affinity (see *SI Appendix*, Fig. S4). However, weaker interactions also result in longer assembly times: In this "association-challenged" regime, most dimers do not persist long enough to interact with monomers to create full rings. The value of the optimal affinity is proportional to the concentration of monomers in the system (see *SI Appendix*, Fig. S9).

Although the impact of deadlock on assembly can be quite dramatic, one may ask if this kinetic phenomenon is likely to be important for any given ring. In some cases, such as the apoptosome (23), a ring structure must be populated quickly as a part

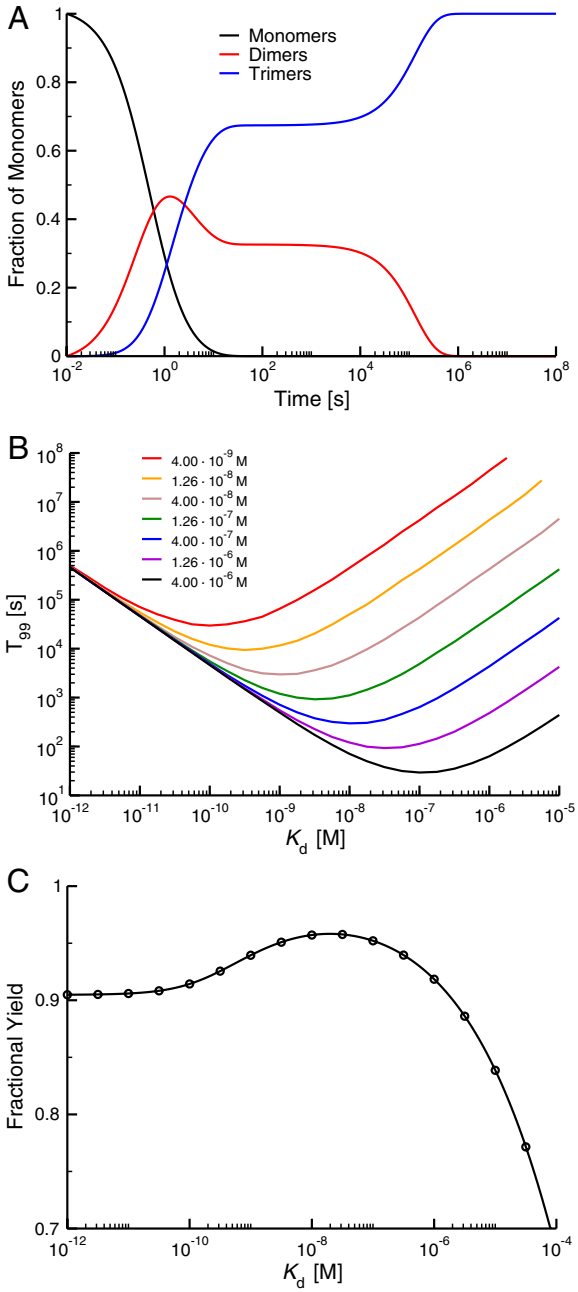

Fig. 2. Assembly of a homomeric three-membered ring. (*A*) In this graph, we consider the percentage of monomers in the various association states (monomer, dimer, and trimer) as a function of time. The affinities are uniformly very strong ($K_d = 10^{-12}$ M). The data are plotted on a logarithmic timescale because a linear scale obscures the existence of the plateau phase. The on-rate $\alpha = 2.53 \times 10^6$ M^{-1} s^{-1} and total subunit concentration $X_T = 400$ nM. (*B*) Variation in assembly time (measured by T_{99} as described in the text) with affinity (K_d) for various initial monomer concentrations X_T. All concentrations exhibit a distinct minimum in T_{99}; the K_d at which this minimum occurs is proportional to the total monomer concentration (see *SI Appendix*, Fig. S9). α as in *A*. (*C*) Steady-state yield (defined as the fraction of monomers in the full ring) as a function of affinity when subunit synthesis and degradation are taken into account according to model A (see *SI Appendix*, Sections 2.4 and 4.2). The synthesis and degradation parameters were chosen to yield the average concentration and half-life of proteins in *Saccharomyces cerevisiae* (28, 29), approximately 480 nM and 42 min, respectively. The solid curve represents an analytical solution of the steady-state yield and the circles represent steady-state results from the numerical integration of model A (see *SI Appendix*, Sections 2.4.1 and 3.2.1). The parameters in this case are α as in *A*, monomer synthesis rate $Q = 1.31 \times 10^{-10}$ M s^{-1}, degradation rate $\delta = 2.75 \times 10^{-4}$ s^{-1}, and $X_T = Q/\delta = 477$ nM.

of the propagation of a signal in a signaling network. In that scenario, a plateau could be detrimental because a large fraction of the monomers present in the system may not incorporate into the active molecule on the timescale of the response to signal (23). The interfaces in such rings may thus be under considerable evolutionary pressure to minimize the assembly time of the molecule. However, not all signaling molecules may be sensitive to short-timescale yield. If signaling is functional with the fraction of assembled structures at a level "below" the plateau, then there may be little evolutionary pressure on the affinities in the ring (see *SI Appendix*, Sections 4.1.1 and 4.1.3).

Not all rings in cells may need to assemble quickly; rings are often found as constitutively active and stable assemblies (such as the proteasome or GroEL) that are typically being lost from the cell by active protein degradation and/or dilution arising from cell growth and division. In this case, monomers must be constantly synthesized and assembled into the active structure in order to replace those that are lost. To explore the effect of the phenomena described in Fig. 2*B* on assembly when accounting for synthesis and degradation, we considered two models. In one case, every monomer in the system has the same probability of being degraded, regardless of the molecular context in which that subunit is found—we term this "model A." This case represents a likely scenario for active degradation by certain proteases (26, 27). In the second case ("model B"), all complexes have the same probability of being degraded, which corresponds to a situation in which all complexes are being diluted due to rapid cell growth as well as the activity of some proteases (26). In both models, monomers are synthesized at a constant rate; a full description of these models can be found in the *SI Appendix*, Section 2.4.

In this situation, steady-state assembly yield represents essentially the "return on investment" in the energy required for monomer synthesis because monomers that do not incorporate into the active structure are essentially wasted. In Fig. 2*C*, we plot the steady-state yield of the full complex vs. affinity for a homomeric three-membered ring under model A. The synthesis and degradation parameters in this case were chosen to represent the average concentration and half-life of proteins in *Saccharomyces cerevisiae* (28, 29) (approximately 480 nM and 42 min, respectively). Intermediate affinities maximize yield just as they minimize assembly time, although the magnitude of the effect depends on the parameters. In particular, if degradation rates become very low, the system approaches equilibrium and the greater thermodynamic stability observed for stronger interactions leads to higher yields for those structures (30). The results for model B are similar to those for model A, but with a smaller relative increase in yield (see *SI Appendix*, Section 4.2).

Heteromeric Ring Assembly and the Benefit of Weak Interactions. In heteromeric rings, every single subunit represents a distinct protein. In our models, all of the interactions between proteins along the ring are considered to be specific; that is, a subunit will only bind with its two neighboring proteins and not with any of the other subunits in the ring. When all of the subunit concentrations along the ring are equal, and all of the affinities between subunits are equivalent, one can show that the assembly dynamics of the heteromeric case is actually equivalent to the dynamics of homomeric rings described above (see *SI Appendix*, Section 2.3).

A major difference between homomeric and heteromeric rings, however, is that all of the interaction strengths along a heteromeric ring can be varied independently. We thus examined how changing the relative affinity along the ring influences assembly efficiency by considering a set of seven different affinities ($K_d = 10^{-12}, 10^{-11}, \ldots, 10^{-6}$ M) and constructing all of the unique configurations for a heteromeric ring of length n, where each of the affinities is chosen independently from that set (see *SI Appendix*, Section 4.3.1).

In Fig. 3*A*, we compare the 81 unique configurations for the heteromeric three-membered ring by ranking each configuration according to its T_{99} and its steady-state yield in models A and B. We find that rings containing one or two weak interactions tend to produce the highest yields and lowest assembly times in our models. As discussed above, for a three-membered ring, deadlock (and the corresponding reduction in assembly efficiency) occurs when the monomers are exhausted from the system before all of the dimers have been converted to the full ring. Inclusion of a single weak interaction, however, results in a single dimer that has a shorter half-life than the other two. When this dimer dissociates, the monomers that are produced can react with the other dimers to form the full ring. Inclusion of a weak interaction renders the system much more robust to changes in total subunit concentration (Fig. 3*B*). For heteromeric rings of length 4–7, we also find that the inclusion of one or more weak interactions is critical to optimizing assembly times and yield (see *SI Appendix*, Figs. S17–S19). As with homomeric rings, when degradation rates are very low, the system approaches equilibrium and a single weak interaction no longer produces maximal yields (30).

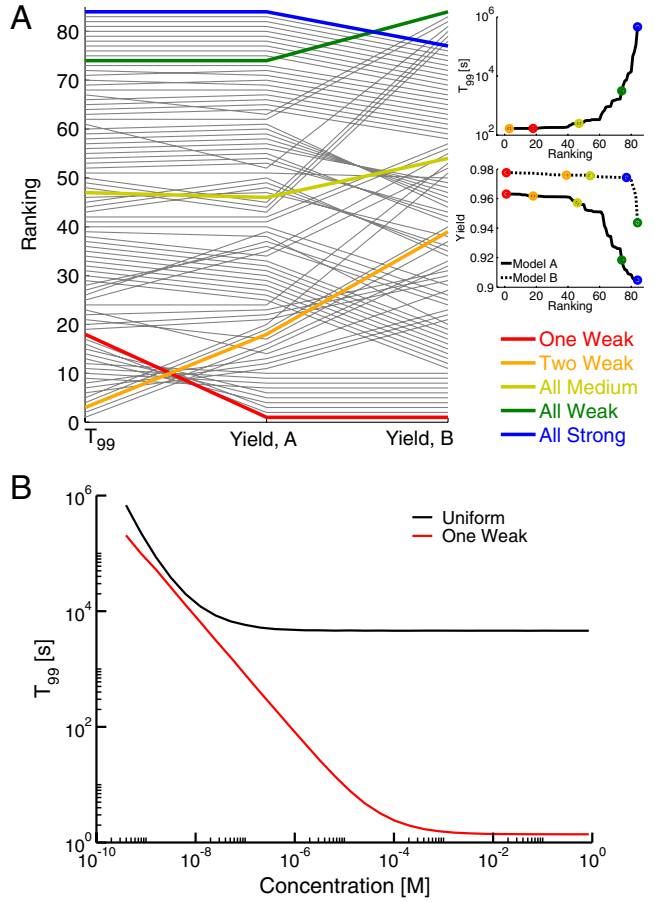

Fig. 3. Nonuniform affinities. (*A*) This plot shows the relative performance of different affinity configurations (gray lines) in the following categories: assembly time (measured by T_{99}), steady-state yield for model A (Yield, A), and steady-state yield for model B (Yield, B). Each affinity configuration is ranked from best (1), to worst (81). The insets display the differences in magnitude represented by the various ranks for T_{99} and yield. The configurations containing one or two weak interactions consistently have the best performance. Affinities vary; other parameters as for Fig. 2*C*. (*B*) Dependence of T_{99} on total subunit concentration for a configuration with three uniform interactions (K_d values of 10^{-10} M) and a configuration with one weak interaction (two interactions with $K_d = 10^{-12}$ M and one with $K_d = 10^{-6}$ M). The total thermodynamic stability of the ring is identical in the two cases. The assembly time is invariably faster with a single weak interaction than with uniform interactions. The value of parameter α is as for Fig. 2*A*.

The findings described in Fig. 3 suggest that rings may be under evolutionary pressure to exhibit at least one weak interaction, regardless of whether they need to assemble quickly in response to signals or assemble with high yield at steady-state (because most proteins are likely to be degraded at relatively high rates; ref. 28). To test this prediction, we considered the crystal structures of heteromeric three-membered rings. Using the database 3D Complex as a starting point (14), we constructed a dataset of 29 such rings (see *SI Appendix*, Section 5.1, and the *SI Table of Structures*) and computed the nonpolar surface area buried in subunit interactions as a proxy for affinity (31, 32). For each structure, we determined the weakest (W) and strongest (S) interaction using the software package Parameter Optimized Surfaces (POPS, ref. 33). The estimated probability density for the ratio between these two (i.e., W/S) in our dataset is shown in Fig. 4*A*. The distribution is approximately bimodal, with an overall average of 0.31; the majority of ring structures (24 out of 29) are found in the left peak of the distribution and have ratios consid-

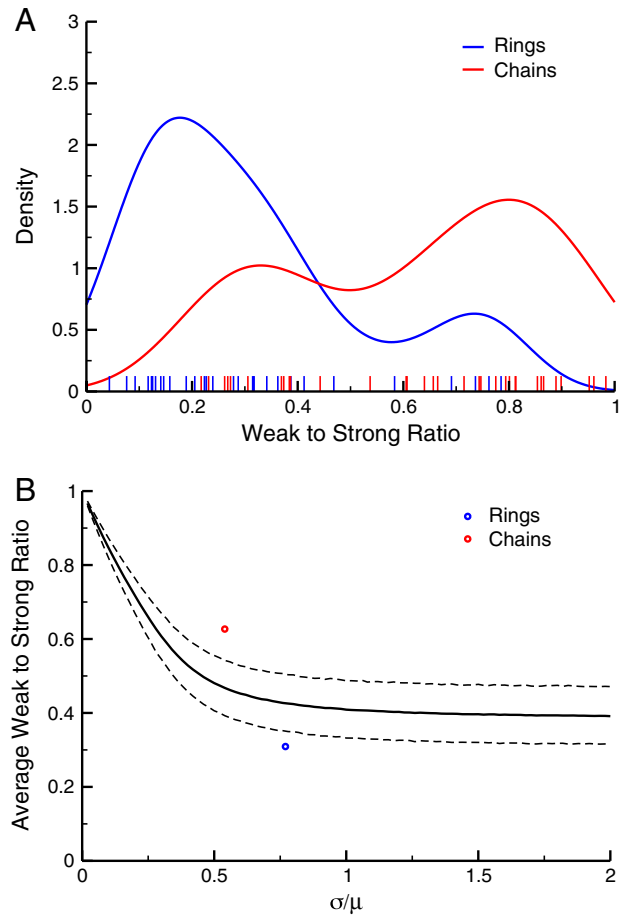

Fig. 4. Weak interactions in crystal structures. (*A*) Estimated probability densities of the weak-to-strong affinity ratio (W/S) for rings and chains. The "strength" of an interaction is measured by the change in solvent-accessible nonpolar surface area as described in the *SI Appendix*, Section 5.3. The density was estimated using the default kernel smoother in the R statistical computing language (37), and the original data points are indicated with short lines on the *x* axis. Ring structures exhibit a significantly smaller average value for this ratio than chains ($p = 10^{-5}$), although both distributions are bimodal. A minority of ring structures (5 of the 29 in this dataset) thus have strong weakest interactions and are more like chains, whereas a minority of chain structures (11 of the 33) have ratios more similar to those for rings. (*B*) Gaussian control for the average W/S ratios observed for chains and rings. The abscissa σ/μ indicates the ratio of the standard deviation to the mean for the underlying Gaussian distribution. The solid line represents the average W/S ratio in 104 random datasets, and the dashed lines represent the 95% confidence interval. The data for both rings and chains lie outside that interval.

erably smaller than 0.5 (see *SI Appendix*, Section 5.4). Other estimates of affinity (total buried surface area or Protein Interfaces, Surfaces, and Assemblies free energy; ref. 34) yielded similar results (see *SI Appendix*, Section 5.3). Interestingly, our buried surface area results would predict that the "strong" interactions in rings have an average K_d of approximately 10^{-12} M, whereas weak interactions have an average K_d of approximately 10^{-6} M (see *SI Appendix*, Section 5.6), which are precisely the values we employed in Figs. 2*A* and 3 *A* and *B*.

Because any heteromeric structure with three interactions will exhibit a weakest and a strongest, we performed two controls to evaluate the significance of the distribution we observed. In the first case, we examined the solved structures of chains. Four-membered chains have the same number of interactions as three-membered rings, but in contrast to rings their assembly efficiency is maximized when all of the interactions are uniformly strong (see *SI Appendix*, Sections 3.2.3 and 4.3.3). We constructed a dataset of 33 structures of heteromeric four-membered chains, whose bimodal W/S distribution is also shown in Fig. 4*A*. Chains have an average ratio (0.63) that is significantly higher than that for rings ($p = 10^{-5}$ based on a random permutation test; see *SI Appendix*, Section 5.4). This difference in average is mostly due to the fact that the weak interaction in chains is, on average, much stronger than the weak interaction in rings ($p = 6 \cdot 10^{-5}$), as we would expect from our findings on assembly efficiency *SI Appendix*, Fig. S25. The majority of chains (22 of the 33) are found in the right peak of the distribution, and in those cases the differences are even more pronounced.

As an additional control, we considered a case in which all of the interactions in the structure were drawn from the same underlying Gaussian distribution (see *SI Appendix*, Section 5.5). Fig. 4*B* shows that both rings and chains exhibit average ratios outside the 95% confidence intervals for this model, indicating that it is unlikely to describe either case. Although we cannot rule out a situation in which affinities are drawn from some other underlying distribution, Fig. 4*B* suggests that the parameters of the distribution could well be under selective pressure to produce rings that meet the affinity requirements for efficient assembly.

Discussion

A number of physical and biological systems, such as glasses and proteins, consist of many concurrently and locally interacting parts. It has long been appreciated that the functional behavior and evolutionary dynamics of these systems are governed by free energy landscapes with many local optima arising from conflicting interactions that are impossible to satisfy simultaneously (e.g., "frustration"; refs. 9 and 30). As was shown for the folding of proteins in the ß-trefoil family (10–13), such situations generate a trade-off in which the desirable stability of native contacts (i.e., interactions present in the final configuration) may conflict with the need to undo them should they be generated in the "wrong" temporal order, preventing further native contacts from forming. Prematurely formed native contacts that are too strong have the potential to slow down the required backtracking and significantly delay the overall folding process. Contacts that are too weak, however, destabilize the entire folding process.

In this work, we expand this idea into the realm of assembly, specifically the assembly of rings, where concurrent exploration of all possible assembly pathways leads to an analogous phenomenon, but in the context of a (partially) bimolecular reaction network. Glassy dynamics arises when earlier reactions use up components needed in subsequent reactions (35), thus slowing down the overall kinetics of the final product. Excessive affinity between subunits causes their rapid sequestration into stable intermediates, choking subsequent bimolecular reactions in which these subunits are needed and causing them to be dominated by the dissociation of stable intermediates (corresponding to the "backtracking" in the ß-trefoil case). The inclusion of a single weak interaction in a heteromeric, three-membered ring optimally solves this conundrum by destabilizing only a single intermediate, whose rapid dissociation regenerates monomers ready to react with the other, stable dimers to form the full ring. These results suggest that the chemical potential landscape governing assembly kinetics must evolve features that avoid reaction deadlock, much as free energy landscapes in protein folding must evolve to destabilize certain intermediates in topologically frustrated folds (10–13). Our data analysis of available structures indicates that the "single weak interaction" strategy is likely employed by the majority of evolved heteromeric three-membered rings (Fig. 4*A*). This strategy might serve as a useful guide in the design of synthetic ring-like structures that quickly assemble with high yield (24).

Because assembly arises from a network of bimolecular association and unimolecular dissociation reactions, assembly systems can exhibit features that are not readily observed in the unimolecular isomerization process of protein folding (8, 9). For instance, overexpressing just one subunit of a three-member heteromeric ring severely exacerbates deadlock (see *SI Appendix*, Section 4.5), reinforcing the fact that the operant concern in assembly is a landscape of chemical potential. In addition, assembly systems may employ unique strategies such as subcellular localization of subunits or extensive allosteric interactions among subunits (36) to overcome deadlock. Although our preliminary findings indicate that allostery offers little benefit over the single weak interaction strategy for single rings (see *SI Appendix*, Section 4.4), such approaches may be employed extensively in more complex structures like the proteasome or ribosome (4–7). Our work indicates that the problems of intramolecular folding and intermolecular assembly may share a level of abstraction that enables lessons from landscape theory (9–13), developed in the context of protein folding, to assist in rationalizing the complex assembly mechanisms observed for macromolecular machines.

Methods

Mathematical Model. The mathematical framework we use for modeling the dynamics of ring assembly is explained in detail in the *SI Appendix*, Section 2. We provide a brief description of our approach here. For any homomeric ring of length n, there are n different molecular species that could be generated, ranging from monomers (size 1) to the full ring (size n). The concentration of any species of size j is denoted X_j. For any species of size $j < n$, there are six distinct physical processes that will influence its concentration: (*i*) an increase in X_j resulting from the dissociation of any larger intermediate that contains it as a subcomplex; (*ii*) an increase in X_j resulting from a binding interaction between two smaller intermediates; (*iii*) a decrease in X_j resulting from an interaction with some other intermediate to form a larger complex, but not the full ring; (*iv*) a decrease in X_j when it dissociates to form smaller intermediates; (*v*) a decrease in X_j resulting from an interaction with its complementary intermediate to form the full ring; and (*vi*) an increase in X_j resulting from the dissociation of the full ring.

For the full ring, there are only two processes that affect its concentration: (*i*) an increase in X_n resulting from a binding reaction between two intermediates, and (*ii*) a decrease in X_n due to the dissociation of the full ring.

From the processes listed above we can derive a system of ODEs describing the time evolution of the concentration of any intermediate X_j and the full ring X_n (see *SI Appendix*, Section 2.1). Heteromeric rings are modeled in much the same way; the main difference is that there are n distinct molecular species for each size class j (depending on the identities of the subunits in the complex), but only one molecular species for the full ring. Given that only "neighboring" heteromeric intermediates can interact with one another, it is straightforward to derive the ODEs for the heteromeric case (see *SI Appendix*, Section 2.2). We add synthesis and degradation to the model (based either on model A or model B) by including a constant synthesis term (denoted by the variable Q) to the kinetic equation for monomers and the appropriate first-order degradation terms (with a constant degradation rate δ; see *SI Appendix*, Section 2.4). Our model for homomeric chains is described in the *SI Appendix*, Section 2.5.

All systems of ODEs were numerically integrated using the "ode15s" function in MATLAB (25).

Structural Data. As mentioned in the text, we used the database 3D Complex as a basis for obtaining both the heteromeric three-membered ring and four-membered chain structures (14). In both cases, we chose the "QS-90" level of the 3D Complex hierarchy in order to avoid counting very closely related structures (which are often simply mutants of a single protein) in the dataset. Of the 82 rings in this set, many represent situations quite distinct from that considered in our model. For instance, antibody–antigen complexes often form three-membered rings (involving the heavy and light chains of the antibody, which bind each other and the antigen), but such structures have not evolved to assemble with all three chains present. Rather, the antibody chains are assembled first in cells, and only when secreted (or expressed on a cell surface) do they interact with the antigen. Similarly, a number of "three-membered" rings in 3D Complex involve proteases bound to a protein inhibitor. In those cases, the two chains of the protease are actually synthesized as a long polypeptide chain that is cleaved during maturation of the zymogen. The interaction between these chains thus does not arise as a result of a bimolecular reaction, but rather a unimolecular folding reaction, and as such the assembly of these structures is not considered in our model. In total, 53 of the 82 three-membered rings were deemed to not conform to the assumptions of our model, leaving 29 structures for the analysis in Fig. 4. Similarly, of the 104 four-membered heteromeric chains we obtained from 3D Complex, 60 were disregarded for reasons similar to the ones cited for rings, and 11 were actually found to be rings on further analysis. A full description of the datasets and their construction can be found in *SI Appendix*, Sections 5.1 and 5.2. A detailed list of all structures included in the datasets, or excluded for one of the reasons cited, is also provided in the *SI Table of Structures*.

Statistical Methods. To test if the affinity distributions we observed exhibited significantly different averages, we performed a simple permutation test using the "twot.permutation" function provided by the Data Analysis and Graphics (DAAG) package in R (37) with 10^5 replicates. The p value reported represents the fraction of these permuted datasets with a difference of means greater than the difference we observed. The Gaussian control in Fig. 4*B* was obtained by sampling three affinities from an underlying Gaussian distribution with an average and standard deviation similar to that observed for both our dataset of three-membered rings and our dataset of four-membered chains. A more detailed description of the affinity distributions can be found in the *SI Appendix*, Sections 5.4 and 5.5.

ACKNOWLEDGMENTS. We would like to thank Tom Kolokotrones for his suggestions and his comments on the manuscript. E.J.D. was supported by a postdoctoral fellowship from the National Institutes of Health.

1. Alberts B (1998) The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell* 92:291–294.
2. Korostelev A, Noller HF (2007) The ribosome in focus: New structures bring new insights. *Trends Biochem Sci* 32:434–441.
3. Bashan A, Yonath A (2008) Correlating ribosome function with high-resolution structures. *Trends Microbiol* 16:326–335.
4. Staley JP, Woolford JL (2009) Assembly of ribosomes and spliceosomes: Complex ribonucleoprotein machines. *Curr Opin Cell Biol* 21:109–118.
5. Murata S, Yashiroda H, Tanaka K (2009) Molecular mechanisms of proteasome assembly. *Nat Rev Mol Cell Biol* 10:104–115.
6. Kressler D, Hurt E, Bassler J (2010) Driving ribosome assembly. *Biochim Biophys Acta* 1803:673–683.
7. Marques AJ, Palanimurugan R, Matias AC, Ramos PC, Dohmen RJ (2009) Catalytic mechanism and assembly of the proteasome. *Chem Rev* 109:1509–1536.
8. Shakhnovich E (2006) Protein folding thermodynamics and dynamics: Where physics, chemistry, and biology meet. *Chem Rev* 106:1559–1588.
9. Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14:70–75.
10. Capraro DT, Roy M, Onuchic JN, Jennings PA (2008) Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? *Proc Natl Acad Sci USA* 105:14844–14848.
11. Chavez LL, Gosavi S, Jennings PA, Onuchic JN (2006) Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. *Proc Natl Acad Sci USA* 103:10254–10258.
12. Gosavi S, Chavez LL, Jennings PA, Onuchic JN (2006) Topological frustration and the folding of interleukin-1 beta. *J Mol Biol* 357:986–996.
13. Gosavi S, Whitford PC, Jennings PA, Onuchic JN (2008) Extracting function from a beta-trefoil folding motif. *Proc Natl Acad Sci USA* 105:10384–10389.
14. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA (2006) 3D complex: A structural classification of protein complexes. *PLoS Comput Biol* 2:e155.
15. Saiz L, Vilar JMG (2006) Stochastic dynamics of macromolecular-assembly networks. *Mol Syst Biol* 2 2006.0024.
16. Bray D, Lay S (1997) Computer-based analysis of the binding steps in protein complex formation. *Proc Natl Acad Sci USA* 94:13493–13498.
17. Qi S, et al. (2010) Crystal structure of the Caenorhabditis elegans apoptosome reveals an octameric assembly of CED-4. *Cell* 141:446–457.
18. Shi Y (2006) Mechanical aspects of apoptosome assembly. *Curr Opin Cell Biol* 18:677–684.
19. Braig K, et al. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature* 371:578–586.
20. Wang J, Hartling JA, Flanagan JM (1997) The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* 91:447–456.
21. Petosa C, Collier RJ, Klimpel KR, Leppla SH, Liddington RC (1997) Crystal structure of the Anthrax toxin protective antigen. *Nature* 385:833–838.
22. Kress W, Mutschler H, Weber-Ban E (2007) Assembly pathway of an AAA+ protein: Tracking ClpA and ClpAP complex formation in real time. *Biochemistry* 46:6183–6193.
23. Nakabayashi J, Sasaki A (2006) A mathematical model for apoptosome assembly: The optimal cytochrome *c*/Apaf-1 ratio. *J Theor Biol* 242:280–287.
24. Yin P, Choi HMT, Calvert CR, Pierce NA (2008) Programming biomolecular self-assembly pathways. *Nature* 451(7176):318–322.
25. MATLAB (2010) MATLAB Ver 7.11. (MathWorks, Natick, MA).
26. Sharma S, Hoskins JR, Wickner S (2005) Binding and degradation of heterodimeric substrates by ClpAP and ClpXP. *J Biol Chem* 280:5449–5455.
27. Moore SD, Baker TA, Sauer RT (2008) Forced extraction of targeted components from complex macromolecular assemblies. *Proc Natl Acad Sci USA* 105:11685–11690.
28. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci USA* 103:13004–13009.
29. Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
30. Wolynes PG (1996) Symmetry and the energy landscapes of biomolecules. *Proc Natl Acad Sci USA* 93:14249–14255.
31. Horton N, Lewis M (1992) Calculation of the free energy of association for protein complexes. *Protein Sci* 1:169–181.
32. Bougouffa S, Warwicker J (2008) Volume-based solvation models out-perform area-based models in combined studies of wild-type and mutated protein-protein interfaces. *BMC Bioinformatics* 9:448.
33. Cavallo L, Kleinjung J, Fraternali F (2003) POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res* 31:3364–3366.
34. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774–797.
35. Awazu A, Kaneko K (2009) Ubiquitous "glassy" relaxation in catalytic reaction networks. *Phys Rev E* 80:041931.
36. Williamson JR (2008) Cooperativity in macromolecular assembly. *Nat Chem Biol* 4:458–465.
37. R Development Core Team (2010) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).

Optimizing ring assembly reveals the strength of weak interactions

Supporting Information Appendix

Eric J. Deeds[1*], John A. Bachman[2*] and Walter Fontana[2]

[1]Center for Bioinformatics and Department of Molecular Biosciences, The University of Kansas, 2030 Becker Dr., Lawrence, KS 66047

[2]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston MA 02115, USA

Email: Eric Deeds - deeds@ku.edu; John Bachman - bachman@fas.harvard.edu; Walter Fontana - walter@hms.harvard.edu;

[*]These authors contributed equally

Contents

List of Figures

Section summaries

1—Thermodynamics of rings

In this section we provide a brief explanation for the fact that rings are typically orders of magnitude more stable than acyclic intermediates.

2—Mathematical model of ring assembly

In sections 2.1 and 2.2 we provide a detailed derivation of the kinetic equations for the assembly mechanism that defines our model of homo- and heteromeric ring formation. Some care is required to properly account for combinatorial factors when formulating a model in terms of monomers with explicit interfaces with elementary interactions and associated rate constants.

In section 2.3 we prove that the equations for homomeric ring assembly are a rescaled version of those for heteromeric ring assembly when all affinities are equal and all intermediates in a size class have equal initial concentration.

In section 2.4 we extend our model of ring assembly to include synthesis and degradation. We consider two variants, A and B, that differ in the degradation mechanism. In model A, each *monomer* is removed at a rate proportional to its concentration, implying that the degradation machinery can pry individual monomers out of a complex (including the full ring). In model B, each *complex* is removed at a rate proportional to its concentration.

Since they lack a global geometric constraint, the assembly of chains provides an insightful contrast to the assembly of rings, not the least because a weak interaction makes the ring case more similar to the chain case, while retaining the stability intrinsic to rings. In section 2.5 we define the kinetic equations for assembly and synthesis/degradation of type A and B for heteromeric chains of four monomers. (We discuss chain optimization in section 4.3.3.)

3—Equilibrium and steady-state solutions for the homomeric 3-membered ring

In section 3.1 we provide an analytical solution for the concentration of the three-membered ring at equilibrium based on the thermodynamic considerations of section 1. Even for comparatively weak interactions, small amounts of monomers suffice to push the equilibrium almost entirely to the side of rings, explaining why the yields we observe are very close to 100% (see also Figure 2A of the main text) in the parameter regimes we consider.

In section 3.2.1 we derive the equation for the steady-state yield of the three-member ring under type-A degradation shown in Figure 2C of the main text. In section 3.2.2 we derive the equation for the steady-state curve of 3-ring assembly under type-B degradation. We compare type A and type B models in section 4.2.2 (see Figure 12).

In section 3.2.3 we show that, in contrast to the ring case, strengthening the interactions of chains under a type A degradation scenario always increases yield (see sections 3.2.1 and 3.2.2; see also

Figure 2C of the main text).

4—Additional results

In section 4.1 we examine the formation of the plateau phase for 3-rings as a function of affinity and concentration. The plateau begins to appear at a critical interaction strength (depending on total monomer concentration) and increases in duration with increasing affinity, while its height remains unaffected. The plateau height, however, decreases with increasing ring size. We also examine the chevron plots of Figure 2B in the main text as a function of ring size and target yields.

In section 4.2 we compare in more detail the assembly kinetics of the degradation models A and B described in section 2.4. We examine the effect of interaction strength and synthesis rate on yield for models A and B using the equations derived in sections 3.2.1 and 3.2.2. Interactions that are weaker or stronger than an optimal value reduce steady-state assembly yield.

In section 4.3 we assess the effect of affinity configurations on ring assembly with respect to efficiency (time to 99% equilibrium yield or steady-state yield for the two versions with degradation). We first detail how to properly enumerate distinct affinity configurations and then sample configurations for heteromeric 4-, 5-, and 6-rings, extending the results for 3-rings reported in the main text. The main observation is that rings with two weak interactions perform well, presumably because they can assemble as two "quasi-chains" and then snap together. We compare these results with the case of heteromeric chain assembly, showing that universally stronger interactions are always better.

While heterogeneity of affinities is one solution to assembly bottlenecks, a strict sequential ("hierarchical") assembly—based, for example, on allosteric interactions—might be another, as touched upon in the main text. In section 4.4 we compare three assembly scenarios (uniform interactions, one weak interaction, hierarchical), suggesting that the weak interaction scenario yields the fastest assembly.

In section 4.5 we consider cases in which the concentration of subunits varies in a three-membered heteromeric ring.

5—Analysis of structural data

In this section we describe the structures in the dataset and the statistical procedures underlying their analysis, summarized as Figure 4 in the main text.

1 Thermodynamics of rings

In order to construct a model of ring assembly, we must first have a general characterization of the chemical reactions that are involved in their formation. For simplicity, we begin by considering the case of a single, homomeric ring of length n. As discussed in the main text, if two intermediates, consisting of k and l subunits respectively, react with one another, there are essentially three classes of reaction:

1. $k + l < n$
   The intermediates react and produce another intermediate of length $k + l$ and a single non-covalent interaction is formed.

2. $k + l = n$
   The reaction produces a fully formed ring by creating two non-covalent interactions.

3. $k + l > n$
   In our models, we imagine that the molecules in question are fairly rigid with well-defined interaction angles. Thus, if $k + l > n$, formation of the product of the reaction would result in two subunits occupying the same space (see Fig. 1D in the main text). We thus ignore reactions of type 3.

Type 1 reactions involve the formation of a single non-covalent interaction in a reversible manner:

$$X_k + X_l \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} X_{k+l}$$

where "X_k" denotes the chemical species corresponding to an intermediate with k subunits. We denote the forward rate of this reaction (or the "on" rate) as α, and the backward (or "off") rate as β. In our work, we do not consider small variations in the on rate that might arise for intermediates of differing size, and we take α to be a constant for all the type 1 reactions in the system. We can write the standard free energy of formation of the interaction in type 1 reactions as:

$$\Delta G_b^0(1) = \Delta G_i^0 + \Delta G_p^0, \tag{1}$$

where $\Delta G_b^0(1)$ is the overall standard free energy of formation of the interaction. This free energy is related in the usual way to the dissociation constant of the reaction by $K_D(1) = \beta/\alpha = c_0 e^{\Delta G_b^0(1)/RT}$, where c_0 is some standard reference concentration (here taken to be 1 molar). In equation 1, we have decomposed this binding energy into two components, adopting the notation of Saiz and Vilar [1]. ΔG_i^0 represents the free energy of formation of the interface itself and includes contributions from both the desolvation of the two protein interfaces and the molecular contacts (e.g. electrostatic and Van der Waals interactions) formed upon binding. ΔG_p^0 represents the positional entropy loss entailed when taking two proteins that can freely diffuse around a particular molar volume and confining them to a given binary complex [1].

Type 2 reactions can be written in a very similar manner to type 1 reactions above. We have:

$$X_k + X_l \underset{\gamma}{\overset{\alpha}{\rightleftharpoons}} X_n$$

In this case, we assume that the forward rate of reaction is again equal to α and we introduce a third parameter, γ, to represent the dissociation rate for a fully formed ring. Since type 2 reactions constrain the motion of exactly the same number of molecules as type 1 reactions, the change in positional entropy for both reactions is equivalent [1]. A type 2 reaction, however, involves the formation of 2 interfaces, rather than just one (see Fig. 1C in the main text). If we consider a case where all of the interactions along the ring have identical thermodynamic properties, we have:

$$\begin{aligned} \Delta G_b^0(2) &= 2\Delta G_i^0 + \Delta G_p^0 \\ &= 2\Delta G_b^0(1) - \Delta G_p^0 \end{aligned} \tag{2}$$

where the second equation relates the energetics of type 2 reactions to the energetics of type 1 reactions. Since the association rate α is taken to be equivalent for both cases here, equation 2 indicates that the reverse rate of a type 2 reaction (γ) will differ significantly from β. Defining $K_\gamma \equiv \gamma/\alpha$ as the dissociation constant for reactions of type 2, we have:

$$\begin{aligned} \gamma &= \alpha \cdot c_0 \cdot e^{\left(2\Delta G_b^0(1) - \Delta G_p^0\right)/RT} \\ &= \left(\frac{\alpha}{c_0}\right) K_D(1)^2 \cdot e^{-\Delta G_p^0/RT}. \end{aligned} \tag{3}$$

If all of the individual reactions of type 1 are favorable for a given ring (i.e. $K_D < 1$ M for all reactions of type 1)—and since $\Delta G_p^0 > 0$ by definition—we will have $\gamma \ll \beta$. Thus, for the types of rings discussed in this work, the rate of dissociation for the full ring is generally many orders of magnitude smaller than the rate of dissociation for any of the acyclic intermediates.

We can define similar equations for cases in which the interactions in a ring do not have the same strengths (e.g. the heteromeric cases described in section 2.2 below). In our notation, each interaction in such a context has a unique label (see section 2.2.1); we define the type 1 affinity of any given interaction "i" as $K_D(i) \equiv \frac{\beta_i}{\alpha}$, where β_i denotes the off rate for interaction i. Hence in any type 2 reaction in which interactions "i" and "j" are formed, we have:

$$\gamma_{i,j} = \left(\frac{\alpha}{c_0}\right) K_D(i) \cdot K_D(j) \cdot e^{-\Delta G_p^0/RT}. \tag{4}$$

In our analysis of ring assembly we treat the affinities of the interactions along the ring as our principal variable and examine how changes in affinities influence assembly efficiency. As indicated above, we treat α as a constant—that is, variations in the dissociation constant (either between configurations or across the interactions in a given ring) are taken to represent changes in β and γ (see Figure 1 for an example of this thermodynamic picture for two different dimers forming along a ring). Changing the value of α represents an alternative method of modulating

the affinity. Interestingly, changes in α can be mirrored by changes in concentration (assuming that all reactions still have the same on rate). This is perhaps easiest to see by considering the stochastic version of the deterministic association rate. If we define this stochastic rate as "a", we have that $\alpha = a \cdot N_A \cdot V$ where N_A is Avogadro's number and V is the volume of the system [2]. In this construction, we note that any change to α can be conceptualized as either a change in the fundamental frequency of collisions (a) while keeping concentration constant, or as a change in the volume of the system (while keeping total particle numbers constant). The latter scenario corresponds to a change in concentration. Thus, changing K_D by varying α is equivalent to changing concentration scales while keeping α constant. The effect of concentration on assembly dynamics can be seen in section 4.1.1, Figure 5.

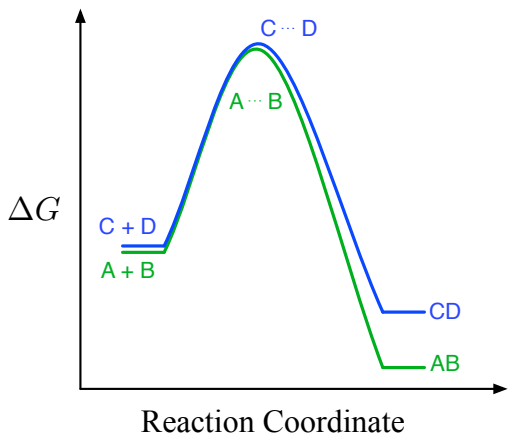

Figure 1: Schematic free energy landscape for a case in which differences in affinities are entirely represented by differences in off rates. Here we have two different binding reactions: A binds B and C binds D. "A + B" and "C + D" represent the unbound states on the far left of the schematic reaction coordinate; the unbound states in this case have roughly the same free energy. The transition states (represented by "A $\cdots$ B" and "C $\cdots$ D") also have approximately the same free energy; the change in free energy from the unbound state to the transition state is identical in both cases (giving identical values of α). However, the bound states ("AB" and "CD") exhibit very different free energies, and the difference in free energy change between the transition state and the bound state results in a much higher value of β for the C-D binding reaction compared to the A-B binding reaction.

In this work we take a "ballpark" value for the association rate of proteins as $\alpha \approx 10^6$ M^{-1} s^{-1} [3].

Since we take α to be constant, for any given value of K_D, if we know ΔG_p^0, we can calculate both β and γ for any particular temperature T. Here we focus on $T = 300$ K, so $RT \approx 0.6$ kcal mol^{-1}. Estimates of ΔG_p^0 vary in the literature [1], but here we set $\Delta G_p^0 = 9$ kcal mol^{-1} [4], which is taken to be a constant for all reactions. Changing the value of ΔG_p^0 will essentially modulate the stability of rings compared to acyclic intermediates, although even considerably smaller values (say, 6 kcal mol^{-1}) still result in $\gamma \ll \beta$ for the parameter values we consider here (see equation 3). We leave investigation of the effects of modulating the positional entropy term to future work.

2 Mathematical model of ring assembly

2.1 Assembly of homomeric ring complexes

2.1.1 Notation

We first consider a homomeric ring structure consisting of n identical subunits, any of which can bind to any other. The subunits are treated as identical but have a "sidedness," that is, each subunit has two distinct interfaces, a left and a right; the left side of one can bind to the right side of another, while two left sides and two right sides cannot bind each other.

As in section 1, we adopt the notation that X_j denotes the sub-complex containing j subunits, with $1 \leq j \leq n$. In the discussion that follows, we also use X_j to refer to the *concentration* of the complex with j subunits—the meaning of "X_j" in any given case will be clear from the context. In the differential equations that follow, X_1 thus denotes the concentration of monomers, X_2 denotes the concentration of dimers, and so on, with X_n denoting the concentration of the full ring structure.

As discussed in section 1, we denote the on rate as α, the off rate as β, and the ring breakage rate (the off rate for type 2 reactions) as γ.

2.1.2 Structure, symmetries, and rate constants

A model of assembly must necessarily take into account essential structural aspects of the molecular species that are generated in the process. This has consequences for how we write our rate equations.

The kinetic description of a reaction splits into a time-dependent monomial of concentrations representing mass-action and a time-independent term, the rate constant, representing a reaction *mechanism*. Models of assembly, like ours, aim at studying the consequences of mechanisms and are therefore defined in terms of thermodynamically motivated rate constants that pertain to elementary interactions (see section 1). This requires that we account explicitly for combinatorial factors intrinsic to a reaction mechanism rather than absorbing them into an overall rate constant.

To illustrate the issue, consider the dimerization of a monomer. The formation of an asymmetric dimer will occur at twice the rate than the formation of a symmetric dimer. Yet, this distinction cannot be expressed in a notation that is too terse in structural detail, such as $A + A \longrightarrow A_2$. The situation clarifies instantly when making a minimum of structure explicit. For example, consider identical monomers, each with two distinct binding sites, black and white. The asymmetric case might be represented as ○□● + ○□● ⟶ ○□●○□● and the symmetric case as ○□● + ○□● ⟶ ○□●●□○ . The difference in the respective reaction rates is accounted for by noting that the mechanism producing the asymmetric dimer allows for two "reaction paths", each involving a distinct black/white combination, while the mechanism producing the symmetric dimer admits only one reaction path (since the "white" sites in the symmetric case cannot bind one another). If we wish to ascribe the same fundamental reaction rate constant to any given "site–site" interaction for both cases, the "apparent" rate constant for the asymmetric reaction will be twice that of the symmetric reaction. By representing molecular structure that is relevant to the problem at hand, we can thus reason about combinatorial factors intrinsic to a mechanism.

Since combinatorial factors apply per reaction event, we must also account for factors that arise from the distinguishability of reactants. Say we have a system of reactions: $A + A \longrightarrow A_2$ and $A + A_2 \longrightarrow A_3$. In the first case, the reactants are *indistinguishable* from one another, while in the latter case the reactants are distinguishable. If we wish to have the same per-site rate constant for these reactions (as in section 1) then we must consider the number of distinct instances of each reaction. For A binding to A_2, we will have $N(A) \cdot N(A_2)$ such instances, where $N(A)$ is the number of A molecules in the system; for A binding to itself, we will have $N(A)(N(A) - 1)/2$ distinct reaction instances. In the second case, passing to the limit of continuous concentrations $[A]$, we can neglect the linear term, but must preserve the factor $1/2$. Taking explicitly into account the combinatorial factors arising from *both* mechanism and reactant indistinguishability permits the formulation of rate equations in terms of rate constants that are taken to be the same across reactions. In our example, the asymmetric reaction has rate $v = \alpha \cdot 2 \cdot [A]^2/2 = \alpha \cdot [A]^2$, while the symmetric case has half that rate: $v = \alpha \cdot [A]^2/2$; in both cases $d[A]/dt = -2v$ and $d[A_2]/dt = v$ due to the fact that the reaction "consumes" two molecules of A but produces only one molecule of A_2.

An analogous situation arises with the assembly of homomeric rings when two intermediates of equal length k react with one another. The asymmetric case occurs when the combined length $2k$ of the product is less than the length n of the ring (Figure 2A). The symmetric case occurs when $k = n/2$ and is due to the simultaneous formation of two interactions upon ring closure (Figure 2B), as argued in section 1.

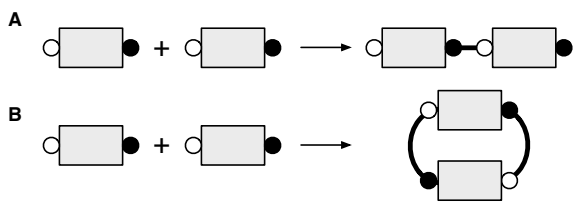

Figure 2: Structured reactions provide a rationale for explicit combinatorial factors needed for rate equations employing thermodynamic rate constants. The figure illustrates the asymmetric and symmetric dimerization reactions that arise in homomeric ring assembly. The grey boxes represent identical chains of equal length $k > 1$. Asymmetric dimer formation (A) arises when chains of equal length combine to form a chain twice as long but shorter than the full ring. Symmetric dimerization (B) arises only for rings of even length and corresponds to identical chains, each half the ring size, simultaneously forming two interactions and completing the ring, as detailed in section 1. As discussed in the text, the reaction velocity of the asymmetric case is twice the velocity of the symmetric case.

For the sake of less cluttered equations, we will not name molecular species in a manner that fully describes their relevant structure, but use a "shorthand" notation instead as described in section 2.1.1. The drawback of such a notation is that one can forget the implications that structural details can have for determining reaction rates. It is therefore useful to enumerate the various reaction classes and their distinct kinetic contribution to homomeric ring assembly. We essentially have bimolecular association reactions in which intermediates form other intermediates or the full ring, and unimolecular dissociation reactions in which molecules break apart to form smaller intermediates. We denote the kinetic term that a reaction class contributes to species X_k with D_k.

1. **The bimolecular association reaction $X_k + X_l \to X_j$, with $j < n$.** Since every chain X_k has two "sides," there are two paths for this reaction to proceed: X_l binds on the right of X_k and X_l binds on the left of X_k. Both paths occur with rate α, yielding a reaction rate $v = 2 \cdot \alpha \cdot X_k \cdot X_l$ with $D_k = D_l = -v$ and $D_j = v$.

2. **The bimolecular association reaction $2\ X_k \to X_j$, with $j < n$.** This represents two copies of a molecule of length k binding to form an intermediate of length $j = 2 \cdot k$. Since one of the X_k's can bind on either side of the other (Figure 2A), the rate of reaction is $v = 2 \cdot \alpha \cdot X_k^2/2 = \alpha \cdot X_k^2$ with $D_k = -2v$ and $D_j = v$.

3. **The bimolecular association reaction $X_k + X_l \to X_n$.** In this case, there is only one reaction path for X_l to bind to X_k: if the right side of X_k binds to the left side of X_l, then the left side of X_k must simultaneously bind the right side of X_l in order to form the full ring (see Figure 1C in the main text). The rate for this reaction is $v = \alpha \cdot X_k \cdot X_l$, with $D_k = D_l = -v$ and $D_n = v$.

4. **The bimolecular association reaction $2\ X_k \to X_n$.** This represents two copies of a molecule of length $k = n/2$ binding to one another to form the full ring (Figure 2B); this reaction only applies to rings of even length. As discussed above, the rate of this reaction is $v = \alpha \cdot X_k^2/2$ with $D_k = -2v$ and $D_n = v$.

5. **The dissociation reaction $X_j \to X_k + X_l$ with $j < n$.** This represents the dissociation of a chain X_j into two distinct smaller chains X_k and X_l. The dissociation rate of any single interaction is denoted β, but in this case there are always *two* interactions in X_j that could break in order to give X_k and X_l; in other words, X_k can break off from either the right or left end of X_j. The rate of the reaction is thus $v = 2 \cdot \beta \cdot X_j$ with $D_j = -v$ and $D_k = D_l = v$.

6. **The dissociation reaction $X_j \to 2\ X_k$ with $j < n$.** This represents the dissociation of a chain X_j into two copies of X_k. Of course, j must be even in this case. There is always only one interaction in X_j that can break in order to form 2 X_k's (namely, the interaction in the "middle" of X_j) so the rate for this reaction is just $v = \beta \cdot X_j$ with $D_j = -v$ and $D_k = 2v$.

7. **The dissociation reaction $X_n \to X_k + X_l$.** This represents the dissociation of the full ring into two different intermediates of length k and l. As discussed in section 1, any two interactions in X_n will break with rate γ. Since there are always n ways to choose two interactions in X_n that will form X_k and X_l when both are broken, the rate of this reaction is $v = n \cdot \gamma \cdot X_n$ with $D_n = -v$ and $D_k = D_l = v$.

8. **The dissociation reaction $X_n \to 2\ X_k$.** This represents the dissociation of the full ring into two copies of the intermediate X_k; this reaction obviously only applies to rings of even length. In this case, there are only $n/2$ distinct ways of choosing two interactions to break in order to form two copies of X_k, so the rate of the reaction is $v = n/2 \cdot \gamma \cdot X_n$ with $D_n = -v$ and $D_k = 2v$.

In the following section, we aggregate over these types of reactions in order to construct a full system of ODEs describing the deterministic time evolution of the concentrations of all molecular species.

2.1.3 Ordinary differential equations

Assembly Intermediates. Assuming mass-action kinetics as described in section 2.1.2 above, the ordinary differential equations describing the change in concentration of any ring-assembly intermediate X_j, $1 \leq j < n$, can be written as follows:

$$\frac{dX_j}{dt} = M_j^{(1)} + M_j^{(2)} + M_j^{(3)} + M_j^{(4)} + M_j^{(5)} + M_j^{(6)} \qquad 1 \leq j < n, \tag{5}$$

where the $M_j^{(i)}$ represent groups of terms describing the exchange of mass between intermediates resulting from six specific binding and unbinding processes. The processes are:

1. An increase in X_j resulting from the dissociation of higher-order intermediates containing X_j as a sub-complex (positive terms):

   $$M_j^{(1)} = 2 \cdot \beta \sum_{l=j+1}^{n-1} X_l.$$

   Note that this term covers both reactions in which only one molecule of X_j is produced (class 5 reactions in section 2.1.2) and reactions where 2 molecules of X_j are produced (class 6 reactions in section 2.1.2).

2. An increase resulting from smaller intermediates binding together to form X_j (positive terms):

   $$M_j^{(2)} = \alpha \sum_{l=1}^{j-1} X_l \cdot X_{j-l}.$$

   Note that the sum in the above term will count asymmetric pairs of intermediates twice, but will count symmetric pairs only once. To illustrate this, suppose we have $j = 6$ (with of course $n > 6$). The sum in $M_j^{(2)}$ will contain the following terms: $X_1 \cdot X_5$, $X_2 \cdot X_4$, $X_3 \cdot X_3$, $X_4 \cdot X_2$ and $X_5 \cdot X_1$. Every type of interaction occurs twice except for the $j/2$ case. $M_j^{(2)}$ thus correctly represents the difference between class 1 reactions (where the two intermediates are of different length) and class 2 reactions (where the two intermediates have the same length), as described in section 2.1.2.

3. A decrease resulting from the binding of X_j to complementary intermediates to form higher-order complexes, but not the full ring (negative terms):

   $$M_j^{(3)} = -2 \cdot \alpha \cdot X_j \sum_{l=1}^{n-j-1} X_l. \tag{6}$$

   Note that, unlike $M_j^{(2)}$, symmetric and asymmetric cases are considered equally here. This is due to the fact that both class 1 and class 2 association reactions have the same net effect on the concentration of the reactants (see section 2.1.2).

4. A decrease resulting from the breakage of one of the $j-1$ interactions within the complex X_j (negative term):

$$M_j^{(4)} = -\beta \cdot X_j \cdot (j-1).$$

This term covers both class 5 and class 6 dissociation reactions (see section 2.1.2).

5. A decrease resulting from the binding of X_j to its complementary intermediate to form the full ring (negative term):

$$M_j^{(5)} = -\alpha \cdot X_j \cdot X_{n-j}.$$

Note that when X_j interacts with its complement X_{n-j} to form the full ring X_n, *two* new interactions are formed simultaneously (see section 1 and Fig. 1C in the main text). This term covers both class 3 and class 4 association reactions in section 2.1.2, since both have the same net effect on the reactants.

6. An increase resulting from the breakage of the full ring to yield an intermediate of length j (positive term). The rate γ represents the breakage of *two* interactions in the full ring to yield the smaller intermediate (see section 1):

$$M_j^{(6)} = n \cdot \gamma \cdot X_n.$$

This term covers both class 7 and class 8 dissociation reactions, since both cases have the same net effect on the products (see section 2.1.2).

Full Ring. The dynamics of the concentration of the full ring (X_n) can be described in terms of two processes: the binding of complementary intermediates to form the ring, and the dissociation of the ring due to the simultaneous breaking of two interactions. These terms are denoted $M_n^{(1)}$ and $M_n^{(2)}$, respectively:

$$\frac{dX_n}{dt} = M_n^{(1)} + M_n^{(2)}. \tag{7}$$

1. The first term, $M_n^{(1)}$, models the process of the binding of complementary intermediates to form the ring, and hence consists of positive terms:

$$M_n^{(1)} = \alpha \sum_{l=1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{1}{1+\delta_{j,n/2}} \right) X_l \cdot X_{n-l},$$

where the floor function $\lfloor x \rfloor$ returns the largest integer smaller than x and arises to ensure correct counting for both odd and even rings. The $\delta_{j,n/2}$ term is the standard Kronecker delta and is included to represent the difference between type 3 and type 4 association reactions (see section 2.1.2): when $j \neq n/2$, the reaction occurs at rate $\alpha \cdot X_l \cdot X_{n-l}$, but when $j = n/2$ the reaction occurs with rate $(\alpha/2) \cdot X_{n/2}^2$.

2. The second term, $M_n^{(2)}$, models the process of the breakage of the ring by dissociation of two interactions. There are $\binom{n}{2}$ ways of choosing any two of the n interactions in the full ring to dissociate with rate γ, yielding the following expression:

$$M_n^{(2)} = -\binom{n}{2} \cdot \gamma \cdot X_n.$$

Note that this covers the net effect of all class 7 and class 8 dissociation reactions described in section 2.1.2.

2.1.4 Example: homomeric three-ring

The full set of differential equations for the homomeric three-ring can therefore be given as follows:

$$\begin{aligned}
\frac{dX_1}{dt} &= 2\beta X_2 - 2\alpha X_1^2 - \alpha X_1 X_2 + 3\gamma X_3 \\
\frac{dX_2}{dt} &= \alpha X_1^2 - \beta X_2 - \alpha X_2 X_1 + 3\gamma X_3 \\
\frac{dX_3}{dt} &= \alpha X_1 X_2 - 3\gamma X_3. \qquad (8)
\end{aligned}$$

2.2 Assembly of heteromeric rings

2.2.1 Notation

We now consider a ring structure with n distinct subunits, which we label $x_0, x_1, \ldots, x_{n-1}$—in this case, the subscripts refer to the *identity* of the subunit, not the size of a sub-complex as in the heteromeric case. Because the subunits are distinct, and each subunit can bind only to its neighbors, the ring is *heteromeric* (see Fig. 1 in the main text). In our notation, the indices for the subunits represent the equivalence classes $[0], [1], \ldots, [n-1]$ of the integer modular arithmetic group $\mathbb{Z}_n$, where n is the ring length. That is, counting, addition, and subtraction on these indices in the discussion below should be understood to occur modulo n. In classifying the various sub-complexes of the full ring we adopt the convention that we enumerate the subunits in increasing order, identifying the complex by the index of the subunit from which we begin counting; so for example, in the case of a ring of length five, the trimer containing the subunits x_4, x_0 and x_1 can be identified as a molecule containing three subunits starting from subunit x_4, the "first" subunit of the complex. We refer to the subunits that come first in counting order as occupying the "left" end of the complex, while those that come last in counting order are on the "right" end: in the trimer described above, subunit x_4 is on the left end, while subunit x_1 is on the right end.

Let $x_{i,j}$ denote the ring assembly intermediate starting from subunit i and containing j total subunits. For example, in the case of a five-membered ring, $x_{2,1}$ denotes the monomer x_2, while $x_{3,4}$ denotes the tetramer starting at subunit x_3 and including subunits x_4, x_0, and x_1. For any given sub-complex of size $j < n$ there are n distinct sub-complexes of that size. An issue arises in

our notation due to the fact that any of the variables $x_{i,n}$ could be used to denote the concentration of the full ring, since they all refer to the same species. For simplicity, we use the variable "x_n" to denote the full ring, with $x_n \equiv x_{i,n}\ \forall\ i$. Since the individual species in the ring x_i have indices ranging from 0 to $n-1$, the variable x_n refers unambiguously to the full ring.

As in the homomeric case, we use the variable "$x_{i,j}$" to represent both the concentration of a given species and as a label for the species itself; the meaning of any given $x_{i,j}$ will be clear from the context.

Since the binding interfaces between subunits are distinct and may vary in strength, we can no longer apply a single off rate β to all dissociation reactions as we did in the homomeric case. We denote the off rate between two subunits x_i and x_{i+1} as β_i; that is, off rates are labeled from the left side. Similarly, when describing the rate of ring breakage, we can no longer apply a single rate γ. Since the ring breakage rate is determined by the strength of the interactions at the two points of breakage, in the heteromeric case we must denote the two interactions that break. As in equation 4, we use the notation $\gamma_{i,j}$ to denote the rate of ring breakage occurring at two junctions: between subunits i and $i+1$, and j and $j+1$. This rate $\gamma_{i,j}$ is determined by the strengths of the interactions between these subunits (dissociation constants $K_D(i)$ and $K_D(j)$). See equation 4 in section 1 for a description of how to calculate $\gamma_{i,j}$ from these affinities.

The set of chemical reactions possible with heteromeric rings is similar to those described for homomeric rings in section 2.1.2. The difference here is that no reaction can ever consume or produce two copies of the same molecule since no intermediate in this system can ever react with another copy of itself in a productive way. We can thus ignore class 2, 4, 6 and 8 reactions as defined in section 2.1.2.

2.2.2 Ordinary differential equations

Assembly Intermediates. In analogy to equation 5 for the homomeric case, the differential equations describing the change in concentration of an assembly intermediate $x_{i,j}$ have the following form:

$$\frac{dx_{i,j}}{dt} = T^{(1)}_{i,j} + T^{(2)}_{i,j} + T^{(3)}_{i,j} + T^{(4)}_{i,j} + T^{(5)}_{i,j} + T^{(6)}_{i,j} \qquad 1 \leq j < n,\ 0 \leq i < n, \tag{9}$$

where the $T^{(k)}_{i,j}$ represent groups of terms describing the exchange of mass between intermediates resulting from the same six binding and unbinding processes described above for the homomeric case. In the heteromeric case, these terms have the the following structure:

1. An increase in $x_{i,j}$ resulting from the dissociation of $x_{i,j}$ from either end of higher-order intermediates, yielding a term consisting of two sums:

$$T^{(1)}_{i,j} = \beta_{i+j-1} \sum_{l=j+1}^{n-1} x_{i,l} + \beta_{i-1} \sum_{l=1}^{n-j-1} x_{i-l,j+l}.$$

2. An increase resulting from smaller intermediates binding together to form $x_{i,j}$ (positive

terms):

$$T_{i,j}^{(2)} = \alpha \sum_{l=1}^{j-1} x_{i,l} \cdot x_{i+l,j-l}.$$

3. A decrease resulting from the binding of $x_{i,j}$ to complementary intermediates to form higher-order complexes (negative terms), but not the full ring. Since binding can occur at either of the two ends of $x_{i,j}$, this term consists of two sums:

$$T_{i,j}^{(3)} = -\alpha \cdot x_{i,j} \left(\sum_{l=1}^{n-j-1} x_{i+j,l} + \sum_{l=1}^{n-j-1} x_{i-l,l} \right).$$

4. A decrease resulting from the breakage of one of the $j-1$ interactions within the complex $x_{i,j}$ (negative term):

$$T_{i,j}^{(4)} = -x_{i,j} \sum_{k=i}^{i+j-2} \beta_k,$$

where the sum is limited by $k = i+j-2$ since the "final" interaction that can be broken in the complex occurs between subunits $i+j-2$ and $i+j-1$ and as such is indexed as β_{i+j-2} in our notation.

5. A decrease resulting from the binding of $x_{i,j}$ to its complementary intermediate to form the full ring (negative term):

$$T_{i,j}^{(5)} = -\alpha \cdot x_{i,j} \cdot x_{i+j,n-j}.$$

6. An increase resulting from the breakage of the full ring to yield the intermediate $x_{i,j}$ (positive term). There is only one pair of interactions that can break to produce the intermediate $x_{i,j}$: the interaction joining subunit x_{i-1} to subunit x_i (at the "left" end of $x_{i,j}$) and the interaction joining subunit x_{i+j-1} to subunit x_{i+j} (at the "right" end). This compound rate is denoted $\gamma_{i-1,i+j-1}$, yielding the following expression:

$$T_j^{(6)} = \gamma_{i-1,i+j-1} \cdot x_n.$$

Full Ring. As in the homomeric case, there is a single equation for the full ring, and it incorporates two processes: the binding of complementary intermediates to form the ring, and the breakage of the ring due to the dissociation of two interactions. These terms are denoted $T_n^{(1)}$ and $T_n^{(2)}$, respectively:

$$\frac{dx_n}{dt} = T_n^{(1)} + T_n^{(2)}. \tag{10}$$

1. The first term, $T_n^{(1)}$, models the binding of complementary intermediates to form the ring,

and hence consists of positive terms:

$$T_n^{(1)} = \alpha \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \left(\sum_{i=0}^{n-1} \left(\frac{1}{1+\delta_{j,n/2}} \right) x_{i,j} \cdot x_{i+j,n-j} \right).$$

The Kronecker delta ($\delta_{j,n/2}$) arises to compensate for the fact that the inner sum above *double-counts* the distinct reactions that can occur between intermediates of length $n/2$.

2. The second term, $T_n^{(2)}$, models the process of the breakage of the ring by dissociation of two interactions. Since dissociation can occur at any two interactions, we must sum over all of the unique $\gamma_{i,j}$ to get the total rate of ring breakage. Note that because the breakage of the two interactions is understood to occur simultaneously the order of interaction breakage does not matter: the rate $\gamma_{i,j}$ is therefore equivalent to $\gamma_{j,i}$ and only one of these permutations is counted when taking the sum.

$$T_n^{(2)} = -x_n \sum_{i=0}^{n-2} \left(\sum_{j=i+1}^{n-1} \gamma_{i,j} \right).$$

2.2.3 Example: heteromeric three-ring

As a specific example, the full set of equations for the change in concentration of $x_{i,j}$ for a three-membered ring is therefore:

$$\begin{aligned}
\frac{dx_{0,1}}{dt} &= \beta_0 x_{0,2} + \beta_2 x_{2,2} - \alpha x_{0,1}(x_{1,1} + x_{2,1}) - \alpha x_{0,1} x_{1,2} + \gamma_{2,0} x_3 \\
\frac{dx_{1,1}}{dt} &= \beta_1 x_{1,2} + \beta_0 x_{0,2} - \alpha x_{1,1}(x_{2,1} + x_{0,1}) - \alpha x_{1,1} x_{2,2} + \gamma_{0,1} x_3 \\
\frac{dx_{2,1}}{dt} &= \beta_2 x_{2,2} + \beta_1 x_{1,2} - \alpha x_{2,1}(x_{0,1} + x_{1,1}) - \alpha x_{2,1} x_{0,2} + \gamma_{1,2} x_3 \\
\frac{dx_{0,2}}{dt} &= \alpha x_{0,1} x_{1,1} - \beta_0 x_{0,2} - \alpha x_{0,2} x_{2,1} + \gamma_{2,1} x_3 \\
\frac{dx_{1,2}}{dt} &= \alpha x_{1,1} x_{2,1} - \beta_1 x_{1,2} - \alpha x_{1,2} x_{0,1} + \gamma_{0,2} x_3 \\
\frac{dx_{2,2}}{dt} &= \alpha x_{2,1} x_{0,1} - \beta_2 x_{2,2} - \alpha x_{2,2} x_{1,1} + \gamma_{1,0} x_3 \\
\frac{dx_3}{dt} &= \alpha(x_{0,1} x_{1,2} + x_{1,1} x_{2,2} + x_{2,1} x_{0,2}) + x_3(\gamma_{0,1} + \gamma_{0,2} + \gamma_{1,2}).
\end{aligned}$$

2.3 Homomeric ring assembly arises as a special case of heteromeric ring assembly

Using the equations shown in Sections 2.1.3 and 2.2.2, we can show that the assembly of homomeric rings arises as a special case of heteromeric ring assembly where the affinities and initial subunit concentrations are all equal.

Assembly Intermediates. We start by considering heteromeric ring assembly. The concentration of all heteromeric intermediates of length $j < n$, denoted as X'_j, is simply the sum of the concentrations of the various $x_{i,j}$ intermediates of that length:

$$X'_j \equiv \sum_{i=0}^{n-1} x_{i,j} \qquad 1 \leq j < n.$$

Consider the case where the initial concentrations of each subunit are equal, and where affinities between subunits are uniform—that is, $\beta_i = \beta \ \forall i$ and thus $\gamma_{i,j} = \gamma \ \forall (i,j)$. In this situation, at time $t = 0$ terms 1-6 in equation 9 will be equal for any two intermediates of length $j < n$; that is, we will have $T^x_{i,j} = T^x_{k,j}$ for any term x in equation 9 and any pair of species $x_{i,j}$ and $x_{k,j}$. At $t = 0$ we thus have that the differential equations for any individual intermediate of length j will be equal to the differential equations for all the other intermediates of length j. Since there is no source of symmetry breaking in this case, the differential equations will remain identical throughout the time evolution of the system, and so we have $x_{i,j} = x_{k,j}$ at all times. This yields:

$$X'_j = \sum_{k=0}^{n-1} x_{k,j} = n \cdot x_{i,j} \qquad 1 \leq j < n$$

$$x_{i,j} = \frac{X'_j}{n}, \tag{11}$$

where we have arbitrarily chosen an $x_{i,j}$ to represent all intermediates of length j since they are all equal. By the chain rule, the differential equation for the change in concentration of X'_j is the sum of the equations for all $x_{i,j}$:

$$\frac{dX'_j}{dt} = \sum_{k=0}^{n-1} \frac{dx_{k,j}}{dt} = n \cdot \frac{dx_{i,j}}{dt} \qquad 1 \leq j < n. \tag{12}$$

As in the homomeric case (equation 5) and heteromeric case (equation 9), the differential equation $\frac{dX'_j}{dt}$ consists of terms corresponding to the six processes of binding and unbinding:

$$\frac{dX'_j}{dt} = S^{(1)}_j + S^{(2)}_j + S^{(3)}_j + S^{(4)}_j + S^{(5)}_j + S^{(6)}_j.$$

We start from the heteromeric equations defining the terms of $\frac{dx_{i,j}}{dt}$, denoting the heteromeric on rate as α_t to distinguish it from the homomeric on rate, denoted α_m. Using the relations given by equations 11 and 12, we can write the equations for $\frac{dX'_j}{dt}$ as follows:

1. We can write $S^{(1)}_j$ by starting with the heteromeric term $T^{(1)}_{i,j}$, substituting $\frac{X'_j}{n}$ for any $x_{i,j}$ (by equation 11), substituting β for any β_i, and multiplying by the factor n (by equation

12), to get :

$$\begin{aligned} S_j^{(1)} &= n \cdot \beta \left(\sum_{l=j+1}^{n-1} \frac{X'_l}{n} + \sum_{l=1}^{n-j-1} \frac{X'_{j+l}}{n} \right) \\ &= \beta \left(\sum_{l=j+1}^{n-1} X'_l + \sum_{l=1}^{n-j-1} X'_{j+l} \right) \\ &= \beta \left(\left(X'_{j+1} + X'_{j+2} + \ldots + X'_{n-1} \right) + \left(X'_{j+1} + X'_{j+2} + \ldots + X'_{n-1} \right) \right). \end{aligned}$$

As the final line above demonstrates, the two summations in the expression sum over the same set of terms and hence can be replaced by a single sum multiplied by a factor of two. We thus find that $S_j^{(1)}$ is equivalent to the term $M_j^{(1)}$:

$$S_j^{(1)} = 2 \cdot \beta \sum_{l=j+1}^{n-1} X'_l = M_j^{(1)}.$$

2. We repeat the above procedure, starting with the terms in $T_{i,j}^{(2)}$ for the heteromeric case and substituting $\frac{X'_j}{n}$ for any $x_{i,j}$ and multiplying by the factor n:

$$\begin{aligned} S_j^{(2)} &= n \cdot \alpha_t \sum_{l=1}^{j-1} \frac{X'_l}{n} \cdot \frac{X'_{j-l}}{n} \\ &= \frac{\alpha_t}{n} \sum_{l=1}^{j-1} X'_l \cdot X'_{j-l}. \end{aligned}$$

Note that, if we introduce the relationship $\alpha_m = \frac{\alpha_t}{n}$, this yields:

$$S_j^{(2)} = \alpha_m \sum_{l=1}^{j-1} X'_l \cdot X'_{j-l} = M_j^{(2)}.$$

We thus have that the heteromeric term $S_j^{(2)} = M_j^{(2)}$ when we set the homomeric association rate α_m to equal the heteromeric association rate α_t, divided by the ring length n.

3. We calculate $S_j^{(3)}$ from $T_{i,j}^{(3)}$:

$$\begin{aligned} S_j^{(3)} &= -n \cdot \alpha_t \cdot \frac{X'_j}{n} \left(\sum_{l=1}^{n-j-1} \frac{X'_l}{n} + \sum_{l=1}^{n-j-1} \frac{X'_l}{n} \right) \\ &= -2 \cdot \frac{\alpha_t}{n} \cdot X'_j \sum_{l=1}^{n-j-1} X'_l \\ &= -2 \cdot \alpha_m \cdot X'_j \sum_{l=1}^{n-j-1} X'_l = M_j^{(3)}, \end{aligned}$$

where we have again used the substitution $\alpha_m = \frac{\alpha_t}{n}$.

4. To calculate $S_j^{(4)}$, we begin with $T_{i,j}^{(4)}$, choosing $i = 0$ for simplicity in the following sums:

$$\begin{aligned} S_j^{(4)} &= -n \cdot \frac{X'_j}{n} \sum_{k=0}^{j-2} \beta_k \\ &= -X'_j \cdot (\beta_0 + \beta_1 + \ldots + \beta_{j-2}). \end{aligned}$$

Since all $\beta_i = \beta$, and since there are $j - 1$ terms in the sum, we can substitute to yield an expression identical to the homomeric equation:

$$S_j^{(4)} = -\beta \cdot X'_j \cdot (j - 1) = M_j^{(4)}.$$

5. We repeat the procedure starting with $T_{i,j}^{(5)}$, again noting the transformation $\alpha_m = \frac{\alpha_t}{n}$:

$$\begin{aligned} S_j^{(5)} &= -n \cdot \alpha_t \cdot \frac{X'_j}{n} \cdot \frac{X'_{n-j}}{n} \\ &= -\alpha_m \cdot X'_j \cdot X'_{n-j} = M_j^{(5)}. \end{aligned}$$

6. We repeat the procedure for $T_{i,j}^{(6)}$, which describes the process of ring breakage. We denote the concentration of the full heteromeric ring in terms of the combined heteromeric intermediates X'_j as X'_n. Since in the heteromeric case there is only instance of the full ring x_n, $X'_n = x_n$. Due to the assumed uniformity of affinities, $\gamma_{i,j} = \gamma$:

$$S_j^{(6)} = n \cdot \gamma \cdot x_n = M_j^{(6)}.$$

Full Ring. To determine the equation for the full heteromeric ring X'_n in terms of the combined heteromeric intermediates X'_j, we start with the equation for the full heteromeric ring as given in equation 10 and make the appropriate substitutions. As in the heteromeric and homomeric cases, the equation consists of two terms describing the processes of ring formation and ring breakage:

$$\frac{dX'_n}{dt} = S_n^{(1)} + S_n^{(2)}.$$

1. We start with the first term, $S_n^{(1)}$, describing the process of ring formation. As in the derivation for the assembly intermediates, we substitute $\frac{X'_j}{n}$ for any $x_{i,j}$; in this case, however, we do not multiply the entire term by n, since $X'_n = x_n$ in contrast to the assembly intermediates, where $X'_{i,j} = n \cdot x_{i,j}$.

$$S_n^{(1)} = \alpha_t \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \left(\sum_{i=0}^{n-1} \left(\frac{1}{1+\delta_{j,n/2}} \right) \frac{X'_j}{n} \cdot \frac{X'_{n-j}}{n} \right).$$

We note that the terms in the inner sum do not depend on i, and thus this sum represents multiplication by a factor of n. Remembering $\alpha_m = \frac{\alpha_t}{n}$, we have:

$$\begin{aligned} S_n^{(1)} &= n \cdot \alpha_t \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{1}{1+\delta_{j,n/2}} \right) \frac{X'_j}{n} \cdot \frac{X'_{n-j}}{n} \\ &= \alpha_m \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{1}{1+\delta_{j,n/2}} \right) X'_j \cdot X'_{n-j} = M_n^{(1)}. \end{aligned}$$

2. We repeat the procedure for the second term, $S_n^{(2)}$:

$$\begin{aligned} S_n^{(2)} &= -X'_n \sum_{i=0}^{n-2} \left(\sum_{j=i+1}^{n-1} \gamma_{i,j} \right) \\ &= -X'_n \sum_{i=0}^{n-2} \left(\gamma_{i,i+1} + \gamma_{i,i+2} + \ldots + \gamma_{i,n-1} \right). \end{aligned}$$

There are $n-1-i$ terms in the inner sum, and since $\gamma_{i,j} = \gamma$ due to interaction uniformity, this yields:

$$\begin{aligned} S_n^{(2)} &= -X'_n \sum_{i=0}^{n-2} \gamma(n-1-i) \\ &= -\gamma \cdot X'_n \left((n-1) + (n-2) + \ldots + (n-1-(n-3)) + (n-1-(n-2)) \right) \\ &= -\gamma \cdot X'_n \left((n-1) + (n-2) + \ldots + 2 + 1) \right) \\ &= -\gamma \cdot X'_n \sum_{k=1}^{n-1} k. \end{aligned}$$

We can rewrite this summation, $\sum_{k=1}^{n-1} k$, in terms of the well-known sum of the first n

natural numbers: $\sum_{k=1}^{n-1} k = (\sum_{k=1}^{n} k) - n = n(n+1)/2 - n$. This yields:

$$\begin{aligned} S_n^{(2)} &= -\left(\frac{n^2+n-2n}{2}\right) \cdot \gamma \cdot X'_n \\ &= -\frac{n(n-1)}{2} \cdot \gamma \cdot X'_n \\ &= -\binom{n}{2} \cdot \gamma \cdot X'_n = M_n^{(2)}. \end{aligned}$$

Summary. As described above, we can show that every $S_j^{(x)}$ is equivalent to the corresponding $M_j^{(x)}$ term. This indicates that we can describe the time evolution of the size-class variables X'_j in the heteromeric case with the same exact equations governing the dynamics of the homomeric X_j concentrations. The assembly dynamics of any given heteromeric ring in which the following conditions hold:

1. The off rates β_i and thus the ring breakage rates $\gamma_{i,j}$ are equal for all subunits in the heteromeric case
2. For any given length j, the initial concentrations of the complexes $x_{i,j}$ are equal for all i

can thus be represented by a homomeric system with on rate $\alpha_m = \frac{\alpha_t}{n}$. The rescaling of the association rate arises naturally in this case: given some $k + l \leq n$, a given homomeric X_k molecule can react with any X_l molecule. Any heteromeric $x_{i,k}$ molecule, however, can only react with one of the n members of X'_l. Homomeric molecules thus have n more binding options than heteromeric ones, and dividing the association rate by n compensates for this fact.

2.4 Modeling synthesis and degradation of rings

The models described in sections 2.1.3 and 2.2.2 are meant to capture the dynamics of ring assembly when a closed pool of monomers is allowed to spontaneously assemble. This situation mimics the conditions of many *in vitro* assembly experiments (e.g., [5]), or cases in which assembly is characterized by rapid activation of monomers (e.g. the apoptosome [6]). As discussed in the main text, however, most cells do not synthesize a large number of monomers and only then allow them to interact—assembly in the majority of cases is likely to be characterized by a steady-state process in which the cell constantly produces new monomers to compensate for continuous decreases in the concentration of both assembly intermediates and the full ring due to cell growth or active protein degradation.

We must thus understand whether the dynamic deadlock we have identified (i.e. the plateau in Fig. 2A in the main text) has any bearing on the assembly of rings in the presence of synthesis and degradation. These two processes could each have a role in facilitating assembly: the degradation of persisting intermediates alters or removes the incompatible complexes responsible for deadlock, while the synthesis of new monomers allows any persisting intermediates to find compatible binding partners and complete the process of assembly.

To explore how subunit synthesis and degradation of complexes would affect ring assembly dynamics, we designed additional models incorporating these two processes. Existing experimental evidence suggests that many distinct "degradation" scenarios are possible, depending on the particular cell type in question. We consider two specific cases.

In the first model, which we denote "subunit deletion" or simply "model A," we assume that each individual *subunit* has a certain likelihood of being degraded, irrespective of its membership in a protein complex. When a subunit is deleted from a complex, the other subunits are left intact. In the context of a system in which proteins are post-translationally tagged and targeted for degradation by proteases, this implies that (1) the probability of a protein being tagged is insensitive to binding context, and (2) the degradation machinery is capable of removing individual subunits from complexes. Experimental work on the bacterial protease ClpXP suggests that it can extract and degrade a tagged subunit from a complex in precisely this fashion, leaving other bound (but untagged) subunits intact [7]; this capability of ClpXP has also been exploited for use experimentally as "molecular tweezers" [8].

In the second model, denoted "whole-complex degradation" or "model B," we instead assume that each *complex* has a certain likelihood of being degraded—this includes both monomers and fully assembled rings. The simplest physical interpretation of this model is one of continuous dilution: for example, if the cell grows continuously in volume, the concentrations of all complexes, monomers and rings alike, will be reduced. In the context of active protein degradation processes this model assumes that (1) the propensity for degradation is a property of complexes rather than subunits, and hence that all complexes have an equal likelihood of being tagged regardless of size, and (2) that the degradation machinery consumes the entire complex in one interaction. In contrast to ClpXP, the bacterial protease ClpAP appears to use this alternative mechanism, degrading whole heterodimeric complexes even in cases where only subunit has been tagged [7].

As described above, there is some experimental evidence for the degradation processes described in model A and model B. These two models, however, should not be considered comprehensive, nor mutually exclusive—the dynamics of degradation processes in cells are likely to depend on complex size, specific subunit affinity for targeting enzymes, the number and type of degradative post-translational modifications, and many other factors in a highly complex manner. In this work we consider these two cases separately for the sake of simplicity; we leave exploration of the effects of more complicated degradation processes to future work.

In the following sections we describe the ordinary differential equations used to model homomeric and heteromeric ring assembly in the presence of model A- or model B-type degradation. In both cases, production of monomers occurs at a rate Q that is considered to be constant in time, and all first-order degradation processes occur with a rate δ.

2.4.1 Homomeric model A: subunit deletion

Assembly Intermediates with Lengths between 2 and n - 1. As shown in equation 5, the ordinary differential equations for the change in concentration of any homomeric ring-assembly intermediate $X_j, 1 \leq j < n$ can be written as the sum of six groups of terms, $M_j^{(1)} + \ldots + M_j^{(6)}$, describing six specific binding and unbinding processes. To incorporate subunit deletion into our

equations describing the concentrations of assembly intermediates we add additional terms to each $\frac{dX_j}{dt}$. For intermediates of length $2 \leq j < n-1$ (we deal with monomers in a separate section below), these terms take the form:

$$\frac{dX_j}{dt} = M_j^{(1)} + \ldots + M_j^{(6)} + M_j^{(A1)} + M_j^{(A2)}. \qquad 2 \leq j < n-1 \tag{13}$$

The physical processes modeled by these two additional terms are as follows:

1. A decrease in X_j resulting from deletion of a monomer from the complex (negative term). In model A, each monomer has the same independent chance of being degraded; since there are j monomers in X_j, the deletion rate δ is multiplied by j:

$$M_j^{(A1)} = -j \cdot \delta \cdot X_j. \tag{14}$$

2. An increase in X_j resulting from deletion of a monomer from a larger intermediate X_k of size $j < k < n$, yielding X_j (positive terms). Because there are exactly two monomers that can be deleted from X_k to yield a smaller intermediate of length j, the degradation rate δ is multiplied by two:

$$M_j^{(A2)} = 2 \cdot \delta \cdot \sum_{k=j+1}^{n-1} X_k.$$

Monomers. The equation for monomers X_1 is identical to equation 13 for intermediates smaller than length $n-1$, with the addition of a single additional term, Q, denoting the rate of synthesis of new monomers:

$$\frac{dX_1}{dt} = M_1^{(1)} + \ldots + M_1^{(6)} + M_1^{(A1)} + M_1^{(A2)} + Q.$$

Assembly Intermediates of Length n - 1. The equation for assembly intermediates X_{n-1} is similar to the equations for smaller intermediates; however, since the only larger intermediate (from which a monomer can be deleted to yield X_{n-1}) is the full ring X_n, the equation has the following structure:

$$\frac{dX_{n-1}}{dt} = M_{n-1}^{(1)} + \ldots + M_{n-1}^{(6)} + M_{n-1}^{(A1)} + M_{n-1}^{(A3)}.$$

In this expression $M_{n-1}^{(A1)}$ is as in equation 14 with $j = n-1$, and $M_{n-1}^{(A3)}$ is as follows:

$$M_{n-1}^{(A3)} = n \cdot \delta \cdot X_n.$$

The factor n is applied to the rate δ due to the fact that there are n subunits that can be deleted from the full ring X_n to yield the intermediate X_{n-1}.

Full Ring. The formation of the full ring X_n with the addition of degradation includes the two processes described for equation 7, with the addition of a term to describe the degradation of the

full ring:

$$\frac{dX_n}{dt} = M_n^{(1)} + M_n^{(2)} + M_n^{(A4)}.$$

Because there are n subunits that can be deleted from the full ring, each with rate δ, the term $M_n^{(A4)}$ has the following form:

$$M_n^{(A4)} = -n \cdot \delta \cdot X_n.$$

Example: Homomeric Three-Ring. The full set of differential equations for the homomeric three-ring with subunit deletion is as follows:

$$\begin{aligned}
\frac{dX_1}{dt} &= 2\beta X_2 - 2\alpha X_1^2 - \alpha X_1 X_2 + 3\gamma X_3 - \delta X_1 + 2\delta X_2 + Q \\
\frac{dX_2}{dt} &= \alpha X_1^2 - \beta X_2 - \alpha X_2 X_1 + 3\gamma X_3 - 2\delta X_2 + 3\delta X_3 \\
\frac{dX_3}{dt} &= \alpha X_1 X_2 - 3\gamma X_3 - 3\delta X_3.
\end{aligned} \tag{15}$$

2.4.2 Heteromeric model A: subunit deletion

We now consider the case of heteromeric ring assembly with model A degradation, where individual subunits can be deleted individually from assembled complexes. Each subunit x_i may have a different rate of deletion, denoted δ_i; the synthesis rate of monomer x_i is denoted Q_i. Though we describe here the form of the equations allowing for differences between the different δ_i and Q_i, in our simulations we considered only the case where the synthesis and degradation rates of all subunits are equal, that is $\delta_i = \delta$ and $Q_i = \mathrm{Q}$.

Assembly Intermediates with Lengths between 2 and n - 1. As shown in equation 9, the ordinary differential equations for the change in concentration of any heteromeric ring-assembly intermediate $x_{i,j}, 1 \leq j < n,\ 0 \leq i < n$ can be written as the sum of six groups of terms, $T_{i,j}^{(1)} + \ldots + T_{i,j}^{(6)}$, describing six specific binding and unbinding processes. As in the case of homomeric rings, to incorporate subunit deletion into our equations describing the concentrations of assembly intermediates we add additional terms to each $\frac{dx_{i,j}}{dt}$. For intermediates of length $2 \leq j < n-1$ (as for the homomeric case we deal with monomers in a separate section below), these terms take the form:

$$\frac{dx_{i,j}}{dt} = T_{i,j}^{(1)} + \ldots + T_{i,j}^{(6)} + T_{i,j}^{(A1)} + T_{i,j}^{(A2)} \qquad 2 \leq j < n-1,\ 0 \leq i < n. \tag{16}$$

The physical processes modeled by these two additional terms are as follows:

1. A decrease in $x_{i,j}$ resulting from deletion of a monomer from the complex (negative terms). Since any of the subunits within $x_{i,j}$ can be deleted, we sum over the deletion rates for all j

members of $x_{i,j}$:

$$T_{i,j}^{(A1)} = -x_{i,j} \sum_{k=0}^{j-1} \delta_{i+k}. \tag{17}$$

2. An increase in $x_{i,j}$ resulting from deletion of a monomer from a larger intermediate of size $j < k < n$, yielding $x_{i,j}$ (positive terms). For deletion of a single subunit to result in the production of $x_{i,j}$, $x_{i,j}$ must be on one of the two "ends" of the larger complex. In the complexes where $x_{i,j}$ is on the "left end" (using the notation described in section 2.2.1), $x_{i,j}$ will be produced when the subunit that links $x_{i,j}$ to the complex is deleted; this is subunit x_{i+j}. In complexes where $x_{i.j}$ is on the "right end", subunit x_{i-1} must be deleted in order to obtain $x_{i,j}$. We thus obtain:

$$T_{i,j}^{(A2)} = \delta_{i+j} \sum_{k=j+1}^{n-1} x_{i,k} + \delta_{i-1} \sum_{k=j+1}^{n-1} x_{i+j-k,k},$$

where the first sum considers all the left cases, and the second all the right cases.

Monomers. The equation for monomers $x_{i,1}$ is identical to equation 16 for intermediates smaller than length $n-1$, with the addition of a single additional term, Q_i, denoting the rate of synthesis of monomer i:

$$\frac{dx_{i,1}}{dt} = T_{i,1}^{(1)} + \ldots + T_{i,1}^{(6)} + T_{i,1}^{(A1)} + T_{i,1}^{(A2)} + Q_i.$$

Assembly Intermediates of Length n - 1. The equation for assembly intermediates $x_{i,n-1}$ is similar to the equations for smaller intermediates; however, since the only larger intermediate from which a monomer can be deleted to yield $x_{i,n-1}$ is the full ring x_n, we have:

$$\frac{dx_{i,n-1}}{dt} = T_{i,n-1}^{(1)} + \ldots + T_{i,n-1}^{(6)} + T_{i,n-1}^{(A1)} + T_{i,n-1}^{(A3)}.$$

In this expression $T_{i,n-1}^{(A1)}$ is as in equation 17 with $j = n-1$. $T_{i,n-1}^{(A3)}$ describes the rate of formation of $x_{i,n-1}$ by deletion of a subunit from the full ring; since the only monomer that can be deleted from the full ring to result in $x_{i,n-1}$ is the monomer x_{i-1}, $T_{i,n-1}^{(A3)}$ is written:

$$T_{i,n-1}^{(A3)} = \delta_{i-1} \cdot x_n.$$

Full Ring. The equation for formation of the full ring x_n includes the two assembly processes described in equation 10, with the addition of a term to describe the degradation of the full ring:

$$\frac{dx_n}{dt} = T_n^{(1)} + T_n^{(2)} + T_n^{(A4)}.$$

Since any of the subunits x_i can be deleted from the full ring, each with its own rate δ_i, the term has the following form:

$$T_n^{(A4)} = -x_n \sum_{i=0}^{n-1} \delta_i.$$

Example: Heteromeric Three-Ring. The full set of differential equations for the heteromeric three-ring with subunit deletion is as follows:

$$\begin{aligned}
\frac{dx_{0,1}}{dt} &= \beta_0 x_{0,2} + \beta_2 x_{2,2} - \alpha x_{0,1}(x_{1,1} + x_{2,1}) - \alpha x_{0,1} x_{1,2} + \gamma_{2,0} x_3 \\
&\quad -\delta_0 x_{0,1} + \delta_1 x_{0,2} + \delta_2 x_{2,2} + Q_0 \\
\frac{dx_{1,1}}{dt} &= \beta_1 x_{1,2} + \beta_0 x_{0,2} - \alpha x_{1,1}(x_{2,1} + x_{0,1}) - \alpha x_{1,1} x_{2,2} + \gamma_{0,1} x_3 \\
&\quad -\delta_1 x_{1,1} + \delta_2 x_{1,2} + \delta_0 x_{0,2} + Q_1 \\
\frac{dx_{2,1}}{dt} &= \beta_2 x_{2,2} + \beta_1 x_{1,2} - \alpha x_{2,1}(x_{0,1} + x_{1,1}) - \alpha x_{2,1} x_{0,2} + \gamma_{1,2} x_3 \\
&\quad -\delta_2 x_{2,1} + \delta_0 x_{2,2} + \delta_1 x_{1,2} + Q_2 \\
\frac{dx_{0,2}}{dt} &= \alpha x_{0,1} x_{1,1} - \beta_0 x_{0,2} - \alpha x_{0,2} x_{2,1} + \gamma_{2,1} x_3 \\
&\quad -x_{0,2}(\delta_0 + \delta_1) + \delta_2 x_3 \\
\frac{dx_{1,2}}{dt} &= \alpha x_{1,1} x_{2,1} - \beta_1 x_{1,2} - \alpha x_{1,2} x_{0,1} + \gamma_{0,2} x_3 \\
&\quad -x_{1,2}(\delta_1 + \delta_2) + \delta_0 x_3 \\
\frac{dx_{2,2}}{dt} &= \alpha x_{2,1} x_{0,1} - \beta_2 x_{2,2} - \alpha x_{2,2} x_{1,1} + \gamma_{1,0} x_3 \\
&\quad -x_{2,2}(\delta_2 + \delta_0) + \delta_1 x_3 \\
\frac{dx_3}{dt} &= \alpha(x_{0,1} x_{1,2} + x_{1,1} x_{2,2} + x_{2,1} x_{0,2}) + x_3(\gamma_{0,1} + \gamma_{0,2} + \gamma_{1,2}) \\
&\quad -x_3(\delta_0 + \delta_1 + \delta_2).
\end{aligned}$$

2.4.3 Homomeric model B: whole-complex degradation

Assembly Intermediates. As shown in equation 5, the ordinary differential equations for the change in concentration of any ring-assembly intermediate $X_j, 1 \leq j < n$ can be written as the sum of six groups of terms, $M_j^{(1)} + \ldots + M_j^{(6)}$, describing six specific binding and unbinding processes. To incorporate whole-complex degradation into our equations describing the concentrations of assembly intermediates we add an additional term, $M_j^{(B1)}$, to each $\frac{dX_j}{dt}$. The resulting equation for intermediates of length $2 \leq j < n$ (monomers are dealt with in a separate section below) is as follows:

$$\frac{dX_j}{dt} = M_j^{(1)} + \ldots + M_j^{(6)} + M_j^{(B1)} \qquad 2 \leq j < n.$$

Since in this model of degradation all complexes have an equal rate δ of being degraded, the degradation term $M_j^{(B1)}$ is simply:

$$M_j^{(B1)} = -\delta \cdot X_j \qquad 1 \leq j \leq n. \tag{18}$$

Monomers. In our equation for monomers X_1, the degradation term $M_1^{(B1)}$ is as described above in equation 18, but we also account for the synthesis of new monomers by incorporating the synthesis rate Q:

$$\frac{dX_1}{dt} = M_1^{(1)} + \ldots + M_1^{(6)} + M_1^{(B1)} + Q.$$

Full Ring. The equation for the formation of the full ring X_n consists of the same two terms as described for equation 7, with the addition of the degradation term $M_n^{(B1)}$:

$$\frac{dX_n}{dt} = M_n^{(1)} + M_n^{(2)} + M_n^{(B1)}. \tag{19}$$

$M_n^{(B1)}$ takes the same form as equation 18:

$$M_n^{(B1)} = -\delta \cdot X_n.$$

Example: Homomeric Three-Ring. The full set of differential equations for the homomeric three-ring with whole-complex degradation is as follows:

$$\begin{aligned}
\frac{dX_1}{dt} &= 2\beta X_2 - 2\alpha X_1^2 - \alpha X_1 X_2 + 3\gamma X_3 - \delta X_1 + Q \\
\frac{dX_2}{dt} &= \alpha X_1^2 - \beta X_2 - \alpha X_2 X_1 + 3\gamma X_3 - \delta X_2 \\
\frac{dX_3}{dt} &= \alpha X_1 X_2 - 3\gamma X_3 - \delta X_3.
\end{aligned} \tag{20}$$

2.4.4 Heteromeric model B: whole-complex degradation

Because the whole-complex degradation model applies the same degradation rate to all complexes, the differential equations for heteromeric ring assembly incorporating this degradation model require nearly identical modifications as those for homomeric ring assembly, described in section 2.4.3. Note that in this case while there is a single degradation rate δ that applies to all complexes, but we allow for different rates of monomer synthesis—the synthesis rate for monomer x_i is denoted Q_i. As in the case of the subunit deletion model for heteromeric rings (section 2.4.2). in our simulations we only consider the case where these rates are all equal (i.e. $Q_i = Q\ \forall i$).

Assembly intermediates. As shown in equation 9, the ordinary differential equations for the change in concentration of any heteromeric ring-assembly intermediate $x_{i,j}, 1 \leq j < n,\ 0 \leq i < n$ can be written as the sum of six groups of terms, $T_{i,j}^{(1)} + \ldots + T_{i,j}^{(6)}$, describing six specific binding and

unbinding processes. As in the case of homomeric rings, to incorporate whole-complex degradation into our equations describing the concentrations of assembly intermediates we add a single additional term $T_{i,j}^{(B1)}$ to each $\frac{dx_{i,j}}{dt}$. The resulting equation for intermediates of length $2 \leq j < n$ (we deal with monomers in a separate section below), is as follows:

$$\frac{dx_{i,j}}{dt} = T_{i,j}^{(1)} + \ldots + T_{i,j}^{(6)} + T_{i,j}^{(B1)} \qquad 2 \leq j < n,\ 0 \leq i < n. \tag{21}$$

Since in this model of degradation all complexes have an equal rate δ of being degraded, the degradation term $T_{i,j}^{(B1)}$ is simply

$$T_{i,j}^{(B1)} = -\delta \cdot x_{i,j} \qquad 1 \leq j \leq n. \tag{22}$$

Monomers. In our equation for monomers $x_{i,1}$, the degradation term $T_{i,1}^{(B1)}$ is as described above in equation 22, but we also account for the synthesis of new monomers by incorporating the synthesis rate Q_i:

$$\frac{dx_{i,j}}{dt} = T_{i,1}^{(1)} + \ldots + T_{i,1}^{(6)} + T_{i,1}^{(B1)} + Q_i.$$

Full Ring. The formation of the full ring x_n with the addition of whole-complex degradation includes the two processes described for equation 10, with the addition of the term $T_n^{(B1)}$:

$$\frac{dx_n}{dt} = T_n^{(1)} + T_n^{(2)} + T_n^{(B1)}.$$

Here $T_n^{(B1)}$ is as described in equation 22:

$$T_n^{(B1)} = -\delta \cdot x_n.$$

Example: Heteromeric Three-Ring. The full set of differential equations for the heteromeric three-ring with whole-complex degradation is as follows:

$$
\begin{aligned}
\frac{dx_{0,1}}{dt} &= \beta_0 x_{0,2} + \beta_2 x_{2,2} - \alpha x_{0,1}(x_{1,1} + x_{2,1}) - \alpha x_{0,1} x_{1,2} + \gamma_{2,0} x_3 \\
&\quad -\delta x_{0,1} + Q_0 \\
\frac{dx_{1,1}}{dt} &= \beta_1 x_{1,2} + \beta_0 x_{0,2} - \alpha x_{1,1}(x_{2,1} + x_{0,1}) - \alpha x_{1,1} x_{2,2} + \gamma_{0,1} x_3 \\
&\quad -\delta x_{1,1} + Q_1 \\
\frac{dx_{2,1}}{dt} &= \beta_2 x_{2,2} + \beta_1 x_{1,2} - \alpha x_{2,1}(x_{0,1} + x_{1,1}) - \alpha x_{2,1} x_{0,2} + \gamma_{1,2} x_3 \\
&\quad -\delta x_{2,1} + Q_2 \\
\frac{dx_{0,2}}{dt} &= \alpha x_{0,1} x_{1,1} - \beta_0 x_{0,2} - \alpha x_{0,2} x_{2,1} + \gamma_{2,1} x_3 \\
&\quad -\delta x_{0,2} \\
\frac{dx_{1,2}}{dt} &= \alpha x_{1,1} x_{2,1} - \beta_1 x_{1,2} - \alpha x_{1,2} x_{0,1} + \gamma_{0,2} x_3 \\
&\quad -\delta x_{1,2} \\
\frac{dx_{2,2}}{dt} &= \alpha x_{2,1} x_{0,1} - \beta_2 x_{2,2} - \alpha x_{2,2} x_{1,1} + \gamma_{1,0} x_3 \\
&\quad -\delta x_{2,2} \\
\frac{dx_3}{dt} &= \alpha x_{0,1} x_{1,2} + \alpha x_{1,1} x_{2,2} + \alpha x_{2,1} x_{0,2} + x_3(\gamma_{0,1} + \gamma_{0,2} + \gamma_{1,2}) \\
&\quad -\delta x_3.
\end{aligned}
$$

2.5 Assembly of Heteromeric Chains

In order to provide a contrast to the affinity configurations that optimize the assembly efficiency of rings (section 4.3.2), we considered a model of the assembly of heteromeric chains (whose optimization we consider in section 4.3.3). We derived equations for the assembly of chains in a manner exactly analogous to that for rings as described in section 2.2. Given the similarities between the ring and chain case, we will not describe ODEs for chains of arbitrary length. We will instead restrict our discussion to chains of length four; these chains have exactly the same number of interactions as three-membered rings, allowing for direct comparison between the two cases (e.g. Fig. 4A in the main text).

In the equations that follow, we denote the monomers that make up the four-chain c_1, c_2, c_3, and c_4. By analogy to the notation for rings, $c_{i,j}$ is used to denote the concentration of the species that starts at subunit c_i and is of length j. β_i denotes the off rate for the interaction connecting subunit c_i to c_{i+1}. As with rings, α denotes the on rate, δ the degradation rate, and Q the monomer synthesis rate.

Assembly of a Heteromeric Chain of Length Four

$$
\begin{aligned}
\frac{dc_{1,1}}{dt} &= \beta_1(c_{1,2}+c_{1,3}+c_{1,4}) - \alpha c_{1,1}(c_{2,1}+c_{2,2}+c_{2,3}) \\
\frac{dc_{2,1}}{dt} &= \beta_2(c_{2,2}+c_{2,3}) + \beta_1 c_{1,2} - \alpha c_{2,1}(c_{3,1}+c_{3,2}+c_{1,1}) \\
\frac{dc_{3,1}}{dt} &= \beta_3 c_{3,2} + \beta_2(c_{1,3}+c_{2,2}) - \alpha c_{3,1}(c_{4,1}+c_{2,1}+c_{1,2}) \\
\frac{dc_{4,1}}{dt} &= \beta_3(c_{1,4}+c_{2,3}+c_{3,2}) - \alpha c_{4,1}(c_{3,1}+c_{2,2}+c_{1,3}) \\
\frac{dc_{1,2}}{dt} &= \beta_2(c_{1,3}+c_{1,4}) + \alpha c_{1,1}c_{2,1} - \alpha c_{1,2}(c_{3,1}+c_{3,2}) - \beta_1 c_{1,2} \\
\frac{dc_{2,2}}{dt} &= \beta_3 c_{2,3} + \beta_1 c_{1,3} + \alpha c_{2,1}c_{3,1} - \alpha c_{2,2}(c_{4,1}+c_{1,1}) - \beta_2 c_{2,2} \\
\frac{dc_{3,2}}{dt} &= \beta_2(c_{1,4}+c_{2,3}) + \alpha c_{3,1}c_{4,1} - \alpha c_{3,2}(c_{2,1}+c_{1,2}) - \beta_3 c_{3,2} \\
\frac{dc_{1,3}}{dt} &= \beta_3 c_{1,4} + \alpha(c_{1,1}c_{2,2}+c_{1,2}c_{3,1}) - \alpha c_{1,3}c_{4,1} - c_{1,3}(\beta_1+\beta_2) \\
\frac{dc_{2,3}}{dt} &= \beta_1 c_{1,4} + \alpha(c_{2,1}c_{3,2}+c_{2,2}c_{4,1}) - \alpha c_{2,3}c_{1,1} - c_{2,3}(\beta_2+\beta_3) \\
\frac{dc_{1,4}}{dt} &= \alpha(c_{1,1}c_{2,3}+c_{1,2}c_{3,2}+c_{1,3}c_{4,1}) - c_{1,4}(\beta_1+\beta_2+\beta_3)
\end{aligned}
$$

Assembly of a Heteromeric Chain with Model A Degradation

$$
\begin{aligned}
\frac{dc_{1,1}}{dt} &= \beta_1(c_{1,2}+c_{1,3}+c_{1,4}) - \alpha c_{1,1}(c_{2,1}+c_{2,2}+c_{2,3}) - \delta c_{1,1} + \delta(c_{1,2}+c_{1,3}+c_{1,4}) + Q \\
\frac{dc_{2,1}}{dt} &= \beta_2(c_{2,2}+c_{2,3}) + \beta_1 c_{1,2} - \alpha c_{2,1}(c_{3,1}+c_{3,2}+c_{1,1}) - \delta c_{2,1} + \delta(c_{2,2}+c_{2,3}+c_{1,2}) + Q \\
\frac{dc_{3,1}}{dt} &= \beta_3 c_{3,2} + \beta_2(c_{1,3}+c_{2,2}) - \alpha c_{3,1}(c_{4,1}+c_{2,1}+c_{1,2}) - \delta c_{3,1} + \delta(c_{3,2}+c_{1,3}+c_{2,2}) + Q \\
\frac{dc_{4,1}}{dt} &= \beta_3(c_{1,4}+c_{2,3}+c_{3,2}) - \alpha c_{4,1}(c_{3,1}+c_{2,2}+c_{1,3}) - \delta c_{4,1} + \delta(c_{1,4}+c_{2,3}+c_{3,2}) + Q \\
\frac{dc_{1,2}}{dt} &= \beta_2(c_{1,3}+c_{1,4}) + \alpha c_{1,1}c_{2,1} - \alpha c_{1,2}(c_{3,1}+c_{3,2}) - \beta_1 c_{1,2} - 2\delta c_{1,2} + \delta(c_{1,3}+c_{1,4}) \\
\frac{dc_{2,2}}{dt} &= \beta_3 c_{2,3} + \beta_1 c_{1,3} + \alpha c_{2,1}c_{3,1} - \alpha c_{2,2}(c_{4,1}+c_{1,1}) - \beta_2 c_{2,2} - 2\delta c_{2,2} + \delta(c_{2,3}+c_{1,3}) \\
\frac{dc_{3,2}}{dt} &= \beta_2(c_{1,4}+c_{2,3}) + \alpha c_{3,1}c_{4,1} - \alpha c_{3,2}(c_{2,1}+c_{1,2}) - \beta_3 c_{3,2} - 2\delta c_{3,2} + \delta(c_{1,4}+c_{2,3}) \\
\frac{dc_{1,3}}{dt} &= \beta_3 c_{1,4} + \alpha(c_{1,1}c_{2,2}+c_{1,2}c_{3,1}) - \alpha c_{1,3}c_{4,1} - c_{1,3}(\beta_1+\beta_2) - 3\delta c_{1,3} + \delta c_{1,4} \\
\frac{dc_{2,3}}{dt} &= \beta_1 c_{1,4} + \alpha(c_{2,1}c_{3,2}+c_{2,2}c_{4,1}) - \alpha c_{2,3}c_{1,1} - c_{2,3}(\beta_2+\beta_3) - 3\delta c_{2,3} + \delta c_{1,4} \\
\frac{dc_{1,4}}{dt} &= \alpha(c_{1,1}c_{2,3}+c_{1,2}c_{3,2}+c_{1,3}c_{4,1}) - c_{1,4}(\beta_1+\beta_2+\beta_3) - 4\delta c_{1,4}
\end{aligned}
$$

Assembly of a Heteromeric Chain with Model B Degradation

$$
\begin{aligned}
\frac{dc_{1,1}}{dt} &= \beta_1(c_{1,2}+c_{1,3}+c_{1,4}) - \alpha c_{1,1}(c_{2,1}+c_{2,2}+c_{2,3}) - \delta c_{1,1} + Q \\
\frac{dc_{2,1}}{dt} &= \beta_2(c_{2,2}+c_{2,3}) + \beta_1 c_{1,2} - \alpha c_{2,1}(c_{3,1}+c_{3,2}+c_{1,1}) - \delta c_{2,1} + Q \\
\frac{dc_{3,1}}{dt} &= \beta_3 c_{3,2} + \beta_2(c_{1,3}+c_{2,2}) - \alpha c_{3,1}(c_{4,1}+c_{2,1}+c_{1,2}) - \delta c_{3,1} + Q \\
\frac{dc_{4,1}}{dt} &= \beta_3(c_{1,4}+c_{2,3}+c_{3,2}) - \alpha c_{4,1}(c_{3,1}+c_{2,2}+c_{1,3}) - \delta c_{4,1} + Q \\
\frac{dc_{1,2}}{dt} &= \beta_2(c_{1,3}+c_{1,4}) + \alpha c_{1,1}c_{2,1} - \alpha c_{1,2}(c_{3,1}+c_{3,2}) - \beta_1 c_{1,2} - \delta c_{1,2} \\
\frac{dc_{2,2}}{dt} &= \beta_3 c_{2,3} + \beta_1 c_{1,3} + \alpha c_{2,1}c_{3,1} - \alpha c_{2,2}(c_{4,1}+c_{1,1}) - \beta_2 c_{2,2} - \delta c_{2,2} \\
\frac{dc_{3,2}}{dt} &= \beta_2(c_{1,4}+c_{2,3}) + \alpha c_{3,1}c_{4,1} - \alpha c_{3,2}(c_{2,1}+c_{1,2}) - \beta_3 c_{3,2} - \delta c_{3,2} \\
\frac{dc_{1,3}}{dt} &= \beta_3 c_{1,4} + \alpha(c_{1,1}c_{2,2}+c_{1,2}c_{3,1}) - \alpha c_{1,3}c_{4,1} - c_{1,3}(\beta_1+\beta_2) - \delta c_{1,3} \\
\frac{dc_{2,3}}{dt} &= \beta_1 c_{1,4} + \alpha(c_{2,1}c_{3,2}+c_{2,2}c_{4,1}) - \alpha c_{2,3}c_{1,1} - c_{2,3}(\beta_2+\beta_3) - \delta c_{2,3} \\
\frac{dc_{1,4}}{dt} &= \alpha(c_{1,1}c_{2,3}+c_{1,2}c_{3,2}+c_{1,3}c_{4,1}) - c_{1,4}(\beta_1+\beta_2+\beta_3) - \delta c_{1,4}
\end{aligned}
$$

2.6 A note on numerical methods

Numerical integration of the ODEs described in the above sections was carried out in MATLAB 7 [9], using the "ode15s" function with parameters AbsTol $= 1 \cdot 10^{-15}$ and RelTol $= 2.22045 \cdot 10^{-14}$. All analytical work was performed using Mathematica 7 [10]. The statistical analyses described in sections 4.1.2 and 5 were carried out using the R statistical computing package [11].

3 Equilibrium and steady-state solutions for the homomeric 3-membered ring

3.1 Equilibrium results without synthesis or degradation

In section 2.1.4, we provided an example of the system of ordinary differential equations representing the dynamics of a simple, homomeric three-membered ring (equation 8). It is straightforward to solve for the concentrations of each of the intermediates at equilibrium. To review, the concentration of the monomer is represented by X_1, the dimer by X_2 and the trimer by X_3. We also introduce the variable $X_T \equiv \sum_{i=1}^{n} i \cdot X_i$ to represent the total concentration of

protein in the system. Setting the three differential equations in equation 8 to 0, we have:

$$\begin{aligned} X_3 &= \frac{X_1 \cdot X_2}{3K_\gamma} \\ X_2 &= \frac{X_1^2}{K_D}, \end{aligned}$$

where $K_\gamma = \gamma/\alpha = \frac{1}{c_0} \cdot K_D^2 \cdot e^{-\Delta G_p^0/RT}$ is used for notational convenience (see section 1). Combining the above relationships with the conservation of mass results in the cubic equation:

$$X_1^3 \left(\frac{1}{K_\gamma \cdot K_D} \right) + X_1^2 \left(\frac{2}{K_D} \right) + X_1 - X_T = 0. \tag{23}$$

Equation 23 can be solved for X_1 trivially using the cubic formula (using Mathematica [10]), with the positive, real solution representing the equilibrium value of X_1. Given the length of the solution, we do not reproduce the full formula here.

To examine the distribution of mass among the various intermediates, we introduce $Y_i \equiv i \cdot X_i/X_T$ which represents the fraction of total protein in the system found in the intermediate of length i. A plot of the equilibrium values of Y_1, Y_2 and Y_3 as a function of X_T is shown in Figure 3. There is a transition between monomers and trimers as X_T is decreased; note that the stability of rings is such that this transition occurs at small values of X_T even when the individual affinities are fairly weak ($K_D \sim 10^{-5}$ M). Dimers are always rare at equilibrium ($Y_2 < 10^{-5}$), regardless of X_T.

3.2 Steady-state results including synthesis and degradation

3.2.1 Model A: subunit deletion

In section 2.4.1, we provided an example of model A-type synthesis and degradation for the homomeric three-membered ring (equation 15). To calculate the change in total monomer concentration, X_T, over time, we recall that $X_T = X_1 + 2X_2 + 3X_3$ for a homomeric three-membered ring and find:

$$\begin{aligned} \frac{dX_T}{dt} &= \frac{dX_1}{dt} + 2\frac{dX_2}{dt} + 3\frac{dX_3}{dt} \\ &= Q - \delta X_1 + 2\delta X_2 - 4\delta X_2 + 6\delta X_3 - 9\delta X_3 \\ &= Q - \delta(X_1 + 2X_2 + 3X_3) \\ &= Q - \delta X_T. \end{aligned}$$

This gives the obvious result at steady state:

$$X_T = \frac{Q}{\delta}. \tag{24}$$

To calculate the steady-state concentration of the various species in this case, we begin by

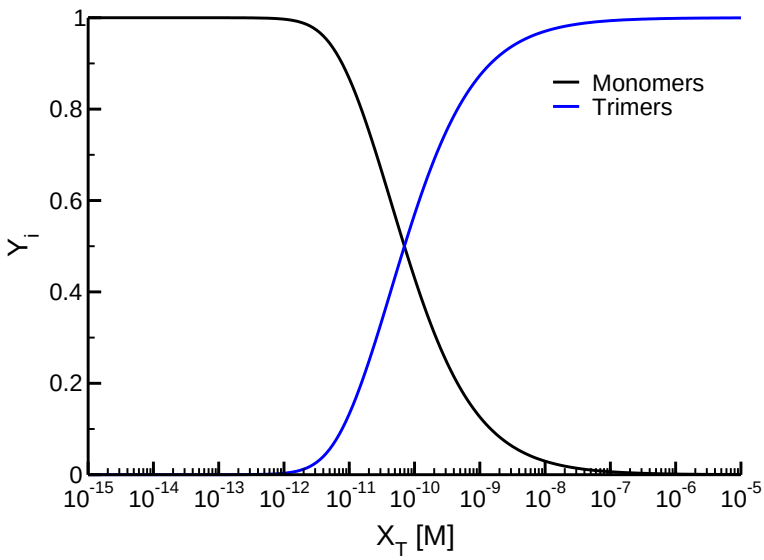

Figure 3: The equilibrium fraction of monomers and trimers for a homomeric three-membered ring as a function of total subunit concentration. "Y_i" represents Y_1 or Y_3 as defined in the text. Y_2 is so small at all values of X_T that it cannot be seen here. In this case, we have $K_D = 10^{-5}$, which gives $K_\gamma \sim 10^{-17}$. The transition between monomers and trimers is centered around an X_T of 10^{-10} in this case.

considering X_3:

$$\begin{aligned} 0 &= \alpha X_1 \cdot X_2 - 3\gamma X_3 - 3\delta X_3 \\ X_3 &= \frac{X_1 \cdot X_2}{3\left(K_\gamma + K_\delta\right)} \,, \end{aligned} \tag{25}$$

where $K_\gamma \equiv \gamma/\alpha$ was defined in section 1 and a new constant $K_\delta \equiv \delta/\alpha$ is introduced. Turning to X_2, we have:

$$\begin{aligned} 0 &= \alpha X_1^2 - \beta X_2 - \alpha X_1 \cdot X_2 + 3\gamma X_3 - 2\delta X_2 + 3\delta X_3 \\ &= \alpha X_1^2 - \beta X_2 - 2\delta X_2 \\ X_2 &= \frac{X_1^2}{K_D + 2K_\delta} \,, \end{aligned} \tag{26}$$

where the reduction between the first and second lines is due to the identity in equation 25 above. Combining equation 26 with equation 25 gives:

$$X_3 = \frac{X_1^3}{3\left(K_\gamma + K_\delta\right)\left(K_D + 2K_\delta\right)} \,. \tag{27}$$

From the differential equation for X_1 we obtain:

$$0 = 2\beta X_2 - 2\alpha X_1^2 - \alpha X_1 \cdot X_2 + 3\gamma X_3 - \delta X_1 + 2\delta X_2 + Q.$$

Dividing through by δ and recalling equation 24, we obtain:

$$0 = 2\left(\frac{K_D}{K_\delta}\right) X_2 - 2\frac{X_1^2}{K_\delta} - \frac{X_1 \cdot X_2}{K_\delta} + 3\left(\frac{K_\gamma}{K_\delta}\right) X_3 - X_1 + 2X_2 + X_T.$$

Substituting using equations 27 and 26, we obtain the cubic equation:

$$\begin{aligned} 0 &= X_1^3 \left(\frac{K_\gamma}{K_\delta\left(K_\gamma + K_\delta\right)\left(K_D + 2K_\delta\right)} - \frac{1}{K_\delta\left(K_D + 2K_\delta\right)}\right) + \\ &\quad + X_1^2 \left(\frac{2K_D}{K_\delta\left(K_D + 2K_\delta\right)} + \frac{2}{K_D + 2K_\delta} - \frac{2}{K_\delta}\right) - X_1 + X_T. \end{aligned} \tag{28}$$

As in section 3.1, it is straightforward to solve the above equation, but due to the length of the formula we do not reproduce it here.

Although equation 28 is true for any set of parameters, recall from section 1 that $\gamma \ll \beta$. For the parameters that we consider here in this work (K_D's generally lower/stronger than 10^{-5} M and protein half-lives less than 10^8 s), we have $\gamma \ll \delta$. We can obtain a slightly simplified version of equation 28 for the parameters we consider by noting that $K_D + K_\gamma \approx K_D$ and $K_\gamma / K_D \approx 0$:

$$0 = \frac{-X_1^3}{K_\delta\left(K_D + 2K_\delta\right)} + X_1^2 \left(\frac{2K_D}{K_\delta\left(K_D + 2K_\delta\right)} + \frac{2}{K_D + 2K_\delta} - \frac{2}{K_\delta}\right) - X_1 + X_T. \tag{29}$$

The results displayed in Fig. 2C of the main text, where we plot Y_3 vs. K_D, are calculated using the positive real solution of equation 29.

Given the solutions to either equation 28 or 29, it is natural to ask whether one can obtain a closed-form solution for the value of K_D that maximizes Y_3 by solving the equation $dY_3/dK_D = 0$. Unfortunately, the derivatives in question are extremely complex and as such we have not yet found an analytical solution for the K_D that provides maximum yield.

3.2.2 Model B: whole-complex degradation

A mathematical description of model B-type degradation can be found in section 2.4.3. Here we consider the case of a homomeric three-membered ring, equation 20. As with model A above, we can easily show:

$$\begin{aligned} \frac{dX_T}{dt} &= Q - \delta\left(X_1 + 2X_2 + 3X_3\right) \\ &= Q - \delta X_T, \end{aligned}$$

which at steady state yields $X_T = Q/\delta$ as with model A. We first solve for X_3 at steady state:

$$X_3 = \frac{X_1 \cdot X_2}{3K_\gamma + K_\delta} \ . \tag{30}$$

In the case of model B, degradation of the full ring does not produce a dimer, and so the $\alpha X_1 \cdot X_2$ term in the equation for dX_2/dt does not cancel as was the case for model A. This yields a considerably more complicated equation for X_2:

$$X_2 = \frac{X_1^2}{X_1\left(1 - \frac{3K_\gamma}{3K_\gamma + K_\delta}\right) + K_D + K_\delta} \ . \tag{31}$$

The equation for X_1 at steady state can be written:

$$0 = 2\left(\frac{K_D}{K_\delta}\right)X_2 - 2\frac{X_1^2}{K_\delta} - \frac{X_1 \cdot X_2}{K_\delta} + 3\left(\frac{K_\gamma}{K_\delta}\right)X_3 - X_1 + X_T. \tag{32}$$

Using equations 30 and 31 we can rewrite equation 32, which after simplifying and collecting terms of the same order yields a cubic equation:

$$\begin{aligned} 0 \ = \ & -3K_\delta X_1^3 - 3K_\delta X_1^2(K_\delta + 2K_\gamma) - K_\delta X_1(K_D(K_\delta + 3K_\gamma) + K_\delta(K_\delta + 3K_\gamma - X_T)) + \\ & + K_\delta X_T(K_D + K_\delta)(K_\delta + 3K_\gamma). \end{aligned} \tag{33}$$

As discussed at the end of section 3.2.1 above, we generally have $K_\gamma \ll K_\delta$. Using this fact, we can obtain a somewhat simpler version of equation 33:

$$0 = -3K_\delta X_1^3 - 3K_\delta^2 X_1^2 - K_\delta^2 X_1(K_D + K_\delta - X_T) + K_\delta^2 X_T(K_D + K_\delta). \tag{34}$$

Analytical results for model B are calculated using the positive real solution of equation 34. As with model A, we have not obtained a closed-form analytical solution for the K_D that maximizes Y_3 using either equation 33 or 34.

Despite their similarities, the dependence of Y_3 on the parameters of the system differs considerably between model A and model B. A detailed analysis of these differences (obtained using the analytical results presented here in addition to results from numerical integration) can be found in section 4.2.2.

3.2.3 Model A degradation for a homomeric chain

As discussed above, for homomeric rings we have found that intermediate affinities can maximize the yield of the full structure at steady state for both of our models of degradation. Here we examine this phenomenon in the case of chain-like structures by considering a simple homodimer under model A degradation. In this case, we have only two types of intermediates (with variables X_1 and X_2 for the concentration of the monomer and dimer, respectively). For this homodimer,

we have the following simple system of ODEs:

$$\begin{aligned}\frac{dX_1}{dt} &= -2\alpha X_1^2 + 2\beta X_2 + Q - \delta X_1 + 2\delta X_2 \\ \frac{dX_2}{dt} &= \alpha X_1^2 - \beta X_2 - 2\delta X_2. \end{aligned} \tag{35}$$

It is simple to verify that, for $X_T \equiv X_1 + 2X_2$, we have $X_T = Q/\delta$ at steady state. From equation 35, we have:

$$X_2 = \frac{X_1^2}{K_D + 2K_\delta} \tag{36}$$

at steady state. Combining this result with the first equation in 35, we get the quadratic equation:

$$0 = X_1^2 \left(\frac{2K_D}{K_\delta(K_D + 2K_\delta)} + \frac{2}{K_D + 2K_\delta} - \frac{2}{K_\delta} \right) - X_1 + X_T$$

for X_1 at steady state. Solving this yields:

$$X_1 = \frac{1}{4}\left(-K_D - 2K_\delta + \sqrt{K_D + 2K_\delta} \cdot \sqrt{8X_T + K_D + 2K_\delta}\right). \tag{37}$$

Unlike the ring models discussed above, it is simple to differentiate Y_2 with respect to K_D to find the K_D at which yield of the homodimer is maximized. We get:

$$\frac{dY_2}{dK_D} = \frac{-4X_T - K_D - K_\delta + \sqrt{K_D + 2K_\delta} \cdot \sqrt{8X_T + K_D + 2K_\delta}}{4X_T\sqrt{K_D + 2K_\delta} \cdot \sqrt{8X_T + K_D + 2K_\delta}}, \tag{38}$$

where we have $dY_2/dK_D < 0$ for $X_T \neq 0$. To show this, simply note that $dY_2/dK_D \geq 0$ implies:

$$\begin{aligned} -(4X_T + K_D + 2K_\delta) + \sqrt{K_D + 2K_\delta} \cdot \sqrt{8X_T + K_D + 2K_\delta} &\geq 0 \\ \frac{4X_T}{\sqrt{K_D + 2K_\delta}} + \sqrt{K_D + 2K_\delta} &\leq \sqrt{8X_T + K_D + 2K_\delta} \\ \left(\frac{4X_T}{\sqrt{K_D + K_\delta}}\right)^2 + 8X_T + K_D + 2K_\delta &\leq 8X_T + K_D + 2K_\delta \\ X_T^2 &\leq 0. \end{aligned} \tag{39}$$

Since X_T is real, equation 39 can only be satisfied when $X_T = 0$, where dY_2/dK_D is trivially equal to 0 for all values of K_D. Strengthening the interactions (i.e. decreasing K_D) thus *always* increases the yield of the chain. Although this result is very intuitive, it nonetheless provides a counterpoint to the behavior of ring-like structures, where increasing affinity does not necessarily increase the yield of the complex (e.g. Fig. 2C in the main text). Our numerical results on heteromeric chains of length four (section 4.3.3) are consistent with the above results.

4 Additional results

4.1 Ring assembly dynamics and deadlock

4.1.1 Effects of affinity and concentration on ring assembly

As discussed in the main text, homomeric rings with uniformly strong interactions between subunits can exhibit a long-lasting assembly deadlock, or "plateau," during which formation of the fully assembled ring is blocked due to the persistence of incompatible intermediates (see Fig. 2A in the main text). In this section we further illustrate the effects of affinity and subunit concentration on assembly dynamics.

As shown in Figure 4A, weakening affinities progressively delays the process of ring assembly due to the lack of stability of assembly intermediates. With a monomer concentration of 400 nM, at very weak affinities ($K_D > 10^{-4}$ M) the balance of protein complexes at equilibrium begins to shift away from trimers towards monomers (see also Figure 3 in section 3.1). By contrast, strong interactions induce the plateau, shown in Figure 4B. While increasing affinities progressively increases the duration of the plateau phase, the plateau always halts assembly at the same concentration of fully assembled complex—that is, the plateau "height" is invariant for a particular ring size. These results demonstrate that for ring assembly efficiency, there exists an *optimal* affinity for any given ring size and subunit concentration.

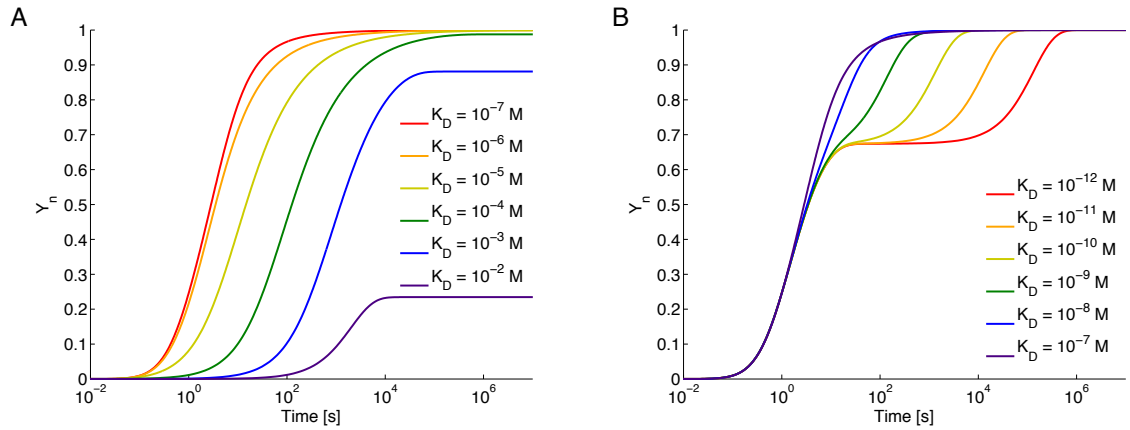

Figure 4: Effect of affinity on ring assembly dynamics. Shown are the results of numerical simulation of the ordinary differential equations for the three-membered homomeric ring (equation 8). $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}, initial monomer concentration of 400 nM. (A) Weaker-than-optimal interactions. (B) Stronger-than-optimal interactions.

The initial concentration of monomers also affects the ring assembly dynamics, as shown in Figure 5. Reducing the monomer concentration delays assembly, due to the lower frequency of interaction between subunits (e.g., the red curve). Increasing the monomer concentration induces a plateau phase, but, importantly, higher concentrations induce the plateau at progressively *earlier* timescales. While this allows for shorter waiting times to reach percentage yields smaller than the plateau percentage (purple curve), it ultimately does not affect the time taken to achieve yields greater than the plateau (overlap of the purple, blue and green curves). By contrast,

increasing affinities while keeping monomer concentration constant induces a plateau phase that progressively *delays* assembly for these higher yields, while leaving the time taken to reach the sub-plateau yields completely unaffected (Figure 4B).

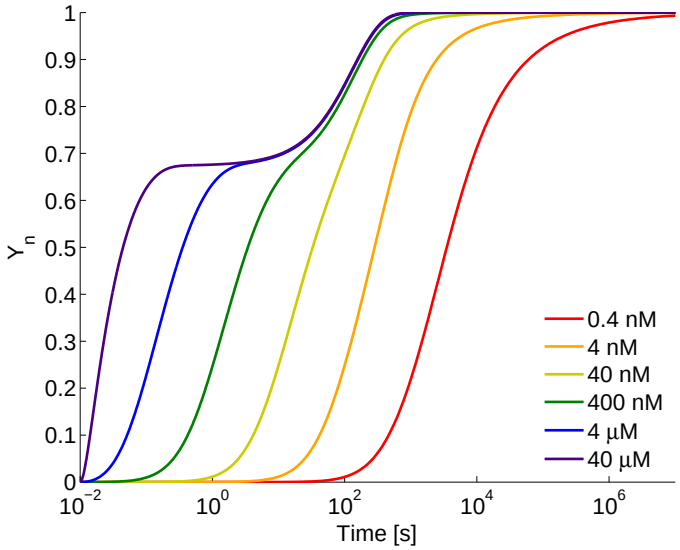

Figure 5: Effect of initial monomer concentration on ring assembly dynamics. Numerical simulations of the ordinary differential equations for the three-membered homomeric ring (equation 8). $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}, $\beta = 2.53 \cdot 10^{-3}$ s^{-1}, $K_D = \beta/\alpha = 10^{-9}$ M.

The different role played by affinity vs. concentration in ring assembly dynamics can be explained as follows. Since the plateau phase results from the persistence of incompatible intermediates, the time of *initiation* of the plateau phase is determined by the rate of complex formation, which is in turn determined by the on rate α and the concentration of monomers. However, the *resolution* of the plateau phase is determined by the timescale of dissociation of the interactions: the more stable the incompatible intermediates are, the longer the average waiting time until they dissociate and return subunits to the pool.

These observations also allow us to make the prediction that in experimental assays of the assembly of homomeric rings (e.g., of the type described by Kress et al. [12]), one should always be able to induce the plateau by increasing the monomer concentration. Actually observing the plateau, however, requires that the time resolution of the experiment be sufficiently small, which may prove difficult in certain cases. It is also important to note that the plateau effect will be difficult to observe unless the data from the assay is plotted on a logarithmic time scale.

4.1.2 Scaling of plateau height with ring length

As mentioned in the main text, we find that the length of the ring in question has a strong influence on the concentration of fully formed rings observed during the "plateau" phase of the dynamics. We find that the value of $Y_n = n \cdot X_n/X_T$ observed at the plateau is an invariant function of concentration for fixed ring length (Figures 4 and 5) and that increasing ring length

decreases Y_n in the plateau phase (Figure 6).

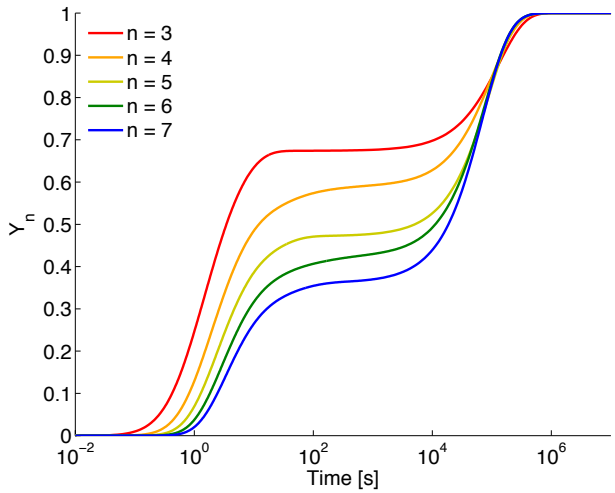

Figure 6: Effect of ring length on ring assembly dynamics. Numerical simulations of the ordinary differential equations for homomeric ring assembly described in section 2.1.3. $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}, $K_D = 10^{-12}$ M, initial monomer concentration of 400 nM. The plateau occurs at progressively lower yields as ring length increases.

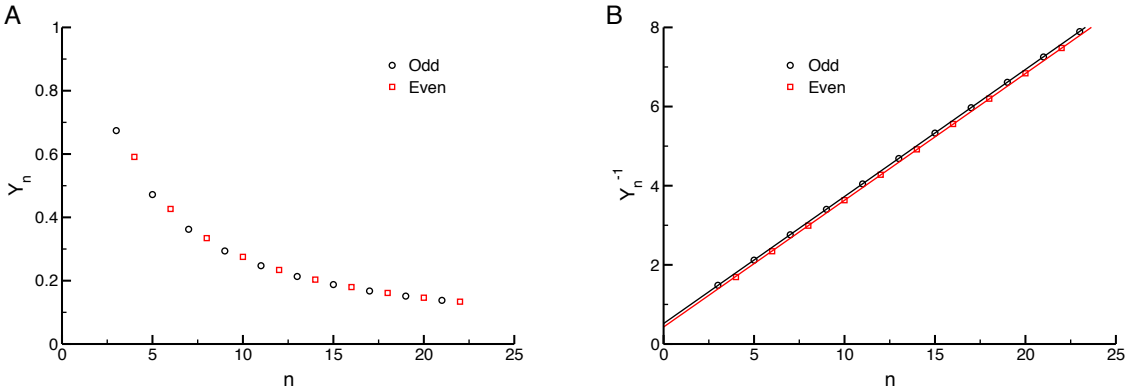

Figure 7: Effect of ring length on plateau height. Both panels represent data from numerical simulations of the ordinary differential equations for homomeric ring assembly described in section 2.1.3. $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}. Initial monomer concentrations set to $n \cdot 40$nM. Plateaus calculated numerically by sampling the fraction of full complex at $t = 10^8$ seconds in the presence of very strong interactions (K$_D = 10^{-24}$ M). (A) Here we plot Y_n directly for rings of length 3 to 22. Note the difference between rings with odd and even lengths. (B) Here we plot the same data as in (A), but using the inverse of the plateau height (Y_n^{-1}) for the y coordinate. The solid lines represent the best linear fits to the data, with $Y_n^{-1} = 0.320n + 0.521$ for odd rings and $Y_n^{-1} = 0.320n + 0.430$ for even rings.

The regular relationship between "plateau height" and ring length is shown in Figure 7.

Interestingly, we observe that odd and even rings have slightly different scaling (Figure 7A). In both cases, the relationship between Y_n^{-1} and n is approximately linear at the plateau (Figure 7B). For odd rings, we have $Y_n^{-1} = 0.320n + 0.521$ with $p < 2 \cdot 10^{-16}$ for both coefficients. For even rings, we have $Y_n^{-1} = 0.320n + 0.430$, again with $p < 2 \cdot 10^{-16}$ for both coefficients. Although the difference in intercepts is slight (Figure 7B), we find that the intercept terms are likely distinct, in that their 95% confidence intervals do not overlap. The slope terms, however, are statistically indistinguishable.

At present, we have not determined an exact analytical relationship that can explain the scaling of plateau height with n that we observe.

4.1.3 Assembly time as a function of affinity, ring length, and concentration

The dependence of ring assembly dynamics on affinity, concentration, and ring length, described in sections 4.1.1 and 4.1.2, can be summarized by plotting assembly time as a function of these three variables, as shown in Figure 8. As described in the main text we describe assembly efficiency using T_X. This is the amount of time it takes a system that starts out with 100% monomers to reach X% of the maximum concentration of the full ring. Each panel within Figure 8 consists of a series of curves showing the assembly time of the ring as a function of affinity; each curve within the panel represents a starting monomer concentration, as denoted in the figure legend. The first row of panels represent the results for a ring of length three; the second row, a ring of length seven. The columns of the plot show the assembly-time curves for target yields of 99%, 75% and 50% (i.e., T_{99}, T_{75}, and T_{50}).

The plots in Figure 8 highlight a number of the dynamic effects described in depth in sections 4.1.1 and 4.1.2 above:

- In each plot, the curves have distinct minima at which assembly time is optimized for that ring length and concentration. Assembly time increases when affinities are either stronger or weaker than this optimal level. The notable exception to this pattern is the T_{50} curve for the three-ring, in which each curve is nearly flat when interactions are strong. This is due to the fact that for the three-ring, 50% yield is reached *before* the plateau phase begins, so affinity has minimal effect on the T_{50} (see Figure 4B). By contrast, for the ring of length seven, T_{50} increases when interactions are strong: this is due to the fact that the plateau phase has already occurred (at approximately 36% yield—see Figures 6 and 7A), and hence its duration becomes rate-limiting for assembly measured by T_{50}.
- Generally, increasing concentration decreases assembly time by increasing the rate of association between subunits (violet curves are always lower than red curves). However, when interactions are very strong, increasing concentration has minimal effect on assembly time because the rate of dissociation of the incompatible intermediates during the plateau phase becomes the rate-limiting process (see Figure 5).
- For larger rings, having weaker-than-optimal interactions incurs a greater "penalty" in terms of assembly time than for shorter rings. This can be seen in the much greater upward slope of the families of curves at weaker-than-optimal interactions for rings of length seven.

This phenomenon is due to the fact that, when weak affinities are the limiting factor in assembly, the larger intermediates that must form during the assembly of larger rings are more likely to break apart before finding a suitable binding partner to form the full ring.

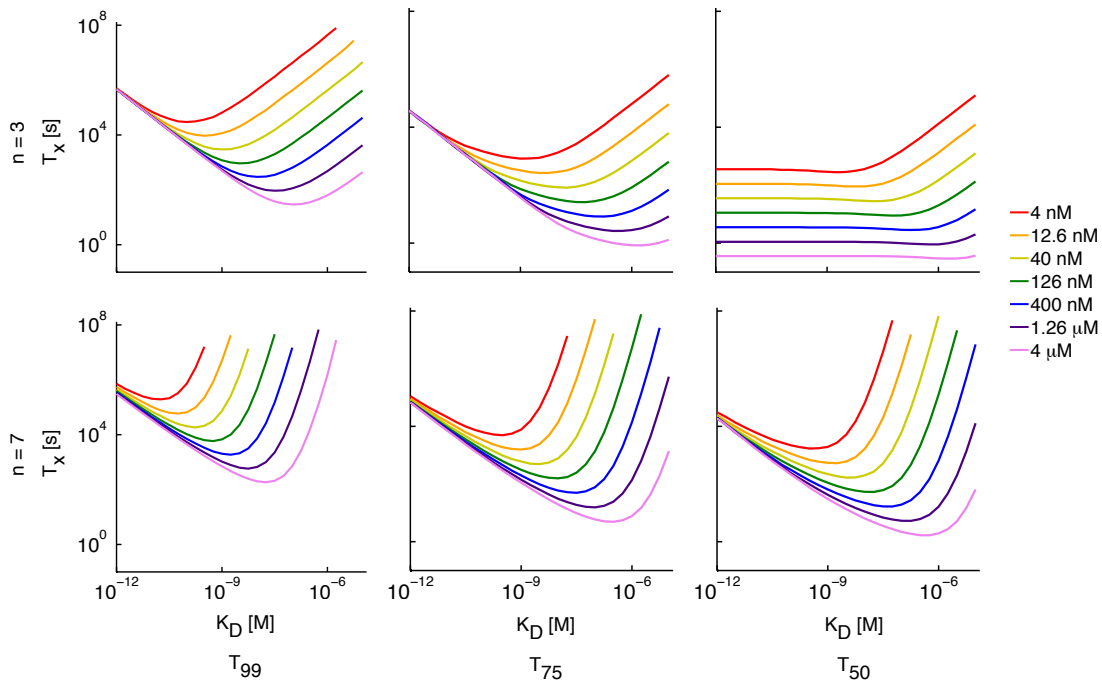

Figure 8: Assembly time as a function of affinity, concentration, and ring length. Times to reach 99%, 75% or 50% yield (T_{99}, T_{75}, and T_{50}, respectively) were calculated by numerical simulation of the ordinary differential equations for homomeric ring assembly described in section 2.1.3. $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1} for all simulations.

An additional aspect of the dynamics that can be seen in Figure 8 is that the affinity yielding optimal concentration (the minimum of each curve) changes as a function of the concentration. This is shown clearly for the homomeric three- and seven-rings in Figure 9. The relationship between the K_D yielding the fastest assembly (lowest T_{99}) and the initial monomer concentration is linear with a numerically calculated relationship of $y = \frac{1}{40}x$ for the three-ring, and $y = 4.457 \cdot 10^{-3}x - 4.521 \cdot 10^{-14}$ for the seven-ring.

4.2 Synthesis and degradation

4.2.1 Ring assembly dynamics with synthesis and degradation

As discussed in sections 2.4 and 3.2, we also explored the effect of synthesis and degradation on the dynamics of ring assembly. In this section we include additional numerical and analytical results from these models.

Figure 10 shows the assembly dynamics of the homomeric three-ring in the case of model A

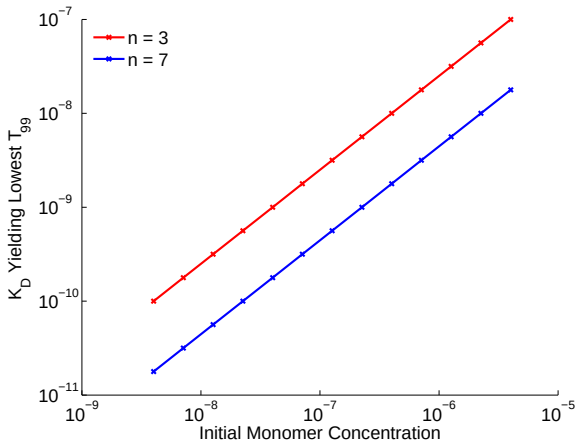

Figure 9: Optimal affinity as a function of monomer concentration. For each concentration shown, results were calculated by varying the K_D in increments of $10^{\frac{1}{4}}$, performing numerical simulation of the differential equations given in section 2.1.3, and identifying the K_D leading to the lowest T_{99}. $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}. The plot is scaled logarithmically to show the entire range of numerical data. A fit yields a slope of 1, that is, a linear relationship between optimal K_D and monomer concentration, as detailed in the text.

(Figure 10A) and model B (Figure 10B) degradation. In the presence of synthesis and degradation, assembly reaches a steady state where the yield of the full ring is less than 100%, and the concentrations of intermediate species are non-zero (compare to Fig. 2A in the main text). Furthermore, the steady-state yield of the ring is higher in the presence of model B degradation than for model A degradation, using nominally identical rates for synthesis and degradation.

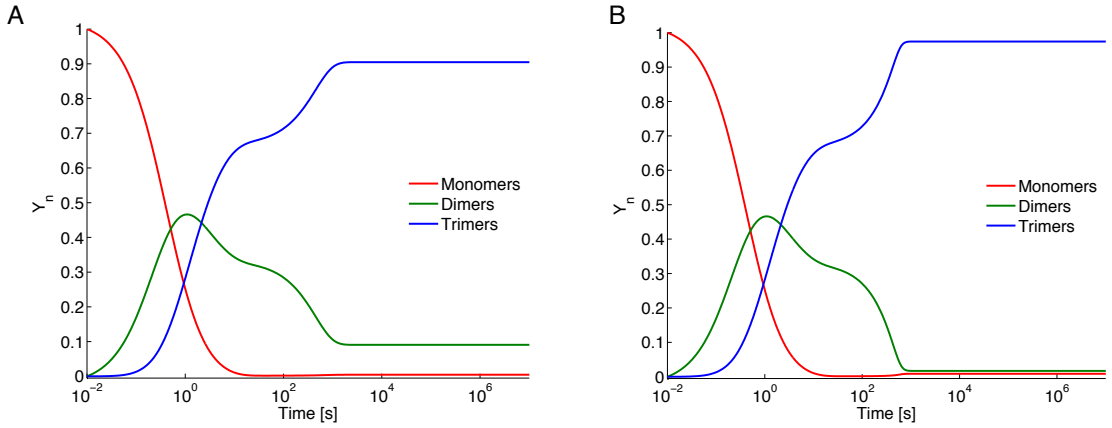

Figure 10: Ring assembly in the presence of synthesis and degradation. Results based on numerical simulation of the ordinary differential equations for the homomeric three-ring described in sections 2.4.1 and 2.4.3. The parameter values for both the simulation in both plots are identical. As described in the main text, the parameter values $X_T = Q/\delta = 477$ nM and $\delta = 2.75 \cdot 10^{-4}$ s^{-1} are taken from averages in available experimental data on yeast proteins [13, 14]. From these numbers, we can calculate $Q = X_T \cdot \delta = 1.31 \cdot 10^{-10}$ M s^{-1}. For these simulations, we also have $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}, $K_D = 10^{-12}$ M. (A) Degradation model A. (B) Degradation model B.

4.2.2 *Effect of K_D and synthesis rate on assembly yield*

As with assembly time (see Figure 4), the steady-state yield of the full complex in the presence of synthesis and degradation is sensitive to the affinities between subunits. Figure 11 shows the effect of varying the affinity on the dynamics of formation of the full ring, in the presence of model A-type degradation. Going from very strong interactions to intermediate interactions, the steady-state yield improves (red vs. green curve). However, further weakening of the interactions rapidly diminishes yield (green vs. blue curve). This result indicates that in the presence of synthesis and degradation, interactions that are weaker or stronger than an optimal value will diminish the yield of the ring.

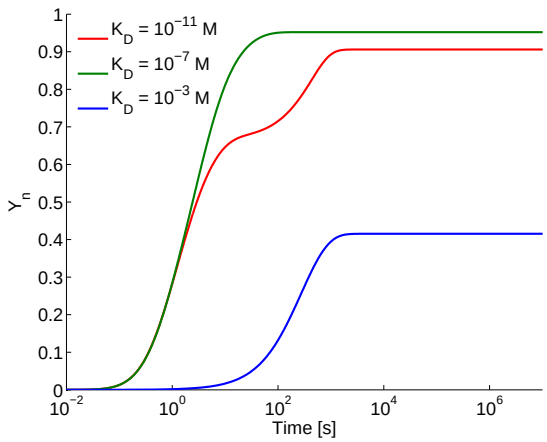

Figure 11: Assembly dynamics of full ring vs. affinity in the presence of synthesis and degradation. Degradation model A, other parameters as for Figure 10.

In the case of the homomeric three-ring, we can explore this effect both numerically and analytically (see section 3.2 and Fig. 2C in the main text). The results of this analysis for both model A and model B are shown in Figure 12. As also demonstrated in Figure 11, very strong interactions can suppress the assembly of the full complex, though the extent of this effect is dependent on the model for degradation that is assumed (see also Figure 14).

These differences between the degradation models can be understood by referring to the timecourses of assembly in the presence of strong interactions, shown in Figure 10. In model A degradation (Figure 10A), the likelihood of a complex experiencing a deletion event increases in proportion with its length, since the deletion rate δ applies to subunits rather than complexes. The fully assembled rings are thus subject to a higher rate of degradation than lower-order intermediates. In addition, deletion of a subunit from a trimer produces a dimer; since dimers cannot interact to form trimers and will tend to persist in the presence of strong interactions, model A degradation leads to a steady state in which dimers occupy a relatively larger fraction of total mass, at the expense of trimers. In model B degradation (Figure 10B), monomers, dimers, and trimers are all equally likely to experience degradation, and degradation of a trimer removes it completely from the system rather than giving rise to an assembly-incompatible dimer. Therefore when the model B system has reached steady state it produces a relatively larger yield of trimers than model A, and a lower balance of dimers. Since these simulations were performed

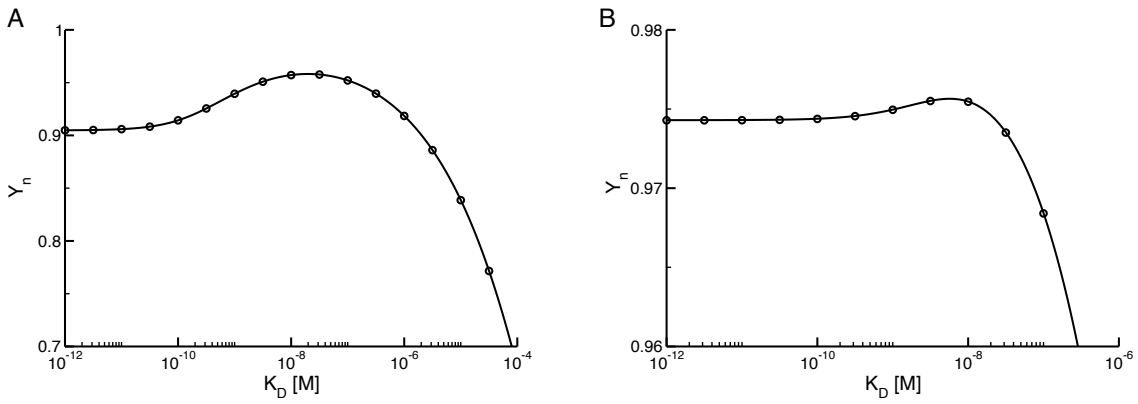

Figure 12: Effect of binding strength on steady state yield of fully assembled complex in the presence of synthesis and degradation. As in Figure 10, $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}, $Q = 1.31 \cdot 10^{-10}$ M s^{-1}, $\delta = 2.75 \cdot 10^{-4}$ s^{-1}, $X_T = Q/\delta = 477$ nM. In both cases, we can consider these curves to represent the effect of varying affinity on rings formed from "average" yeast proteins [13,14]. (A) Model A degradation. Solid line indicates analytical solution of equation 29; circles represent results from numerical integration of equation 15. (B) Model B degradation. Solid line indicates analytical solution of equation 34; circles represent results from numerical integration of equation 20. Note difference in scale between (A) and (B).

with strong interactions, the steady state concentration of monomers is fairly low in both cases.

The steady state yield of the full ring depends not only on the affinity, but also on the rates of synthesis and degradation as well as the total concentration of protein in the system. In an analogous fashion to Figure 12, Figure 13 shows plots of yield vs. K_D, but here each curve represents a different value for the total amount of mass in the system X_T, produced by changing the synthesis rate Q while holding the degradation rate δ constant (changing Q and not δ allows us to look specifically at the effect of changes in X_T while leaving the parameter K_δ, the ratio between the degradation rate δ and the association rate α, unchanged; see section 3.2). For each curve, the yield Y_n is calculated relative to the total amount of protein at steady state, given by Q/δ (see section 3.2).

The curves in Figure 13 become higher for increasing X_T, indicating that for any given affinity, increasing the total amount of protein also increases the proportion of mass that assembles into the full ring. In addition, the effect of affinity optimality also changes as a function of X_T. At very *low* X_T, degradation, rather than deadlock, is the limiting factor in assembly, and hence alleviating deadlock via affinity optimization has minimal effect in boosting steady state yield. At very *high* X_T, the larger influx of new monomers quickly alleviates deadlock, which again mitigates the effect of overly strong interactions (purple curve). At *intermediate* synthesis rates, strong interactions lead to a significant fraction of deadlocked intermediates at steady state, which is alleviated by weakening interactions until an optimal yield is reached (green curve). However, beyond the point of optimality, weakening affinities leads to reduced yield, regardless of the total amount of protein in the system.

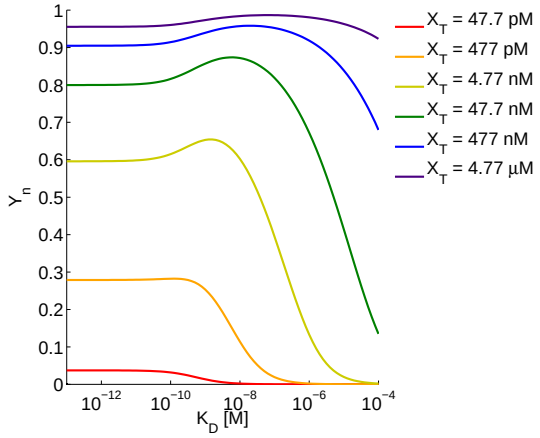

Figure 13: Steady state yield as a function of K_D and total protein concentration, model A degradation. Results generated by numerical simulation of the ordinary differential equations described in section 2.4.1; different values of X_T were obtained by changing the value of the synthesis rate Q while holding δ at a constant value of $2.75 \cdot 10^{-4}$ s^{-1}. $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}.

The magnitude of the affinity optimality effect for both degradation models is summarized in Figure 14. This figure tracks the increase in yield that can be achieved by using optimal interactions rather than very strong interactions, for a range of values of X_T. As discussed above for Figure 13, the improvement in yield that can be gained by using optimal interactions is dependent on the total amount of protein X_T, and reaches a maximum of approximately 7.5% for model A at $X_T = 24$ nM with $\delta = 2.75 \cdot 10^{-4}$. In addition, as discussed for Figure 12, model B degradation exhibits a much less pronounced affinity optimality effect for all values for X_T, with a maximal improvement in yield of only 0.51% at $X_T = 6.0$ nM.

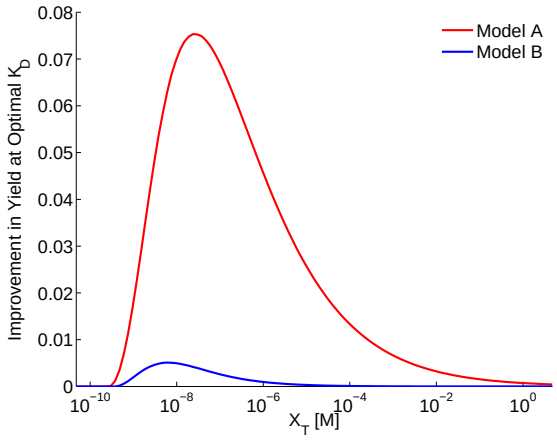

Figure 14: Improvement in assembly yield at optimal K_D vs. very strong interactions ($K_D = 10^{-13}$) at different values of X_T. Results generated by numerical simulation of the ordinary differential equations described in sections 2.4.1 and 2.4.3; different values of X_T were obtained by changing the value of the synthesis rate Q while holding δ at a constant value of $2.75 \cdot 10^{-4}$ s^{-1}. $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}.

Though the results shown in Figures 13 and 14 help to elucidate the underlying dynamics, our analytical solutions for the homomeric three-ring for both model A (equation 29) and model B (equation 34) reveal that the steady state yield is a complex function not only of K_D and X_T, but also of the ratio of the degradation rate to the association rate, K_δ. We leave a thorough characterization of the relative contributions of these different parameters to future work. However, taken together, these results indicate that ring assembly efficiency can be inhibited by uniformly strong interactions in the presence of synthesis and degradation, though the magnitude of this effect depends on the particular model of degradation that is assumed and the choice of synthesis and degradation parameters.

4.3 Effect of affinity configurations on assembly

To determine the effect of affinity configuration on the assembly efficiency of heteromeric rings, we systematically sampled the space of possible configurations and performed numerical simulations of assembly for each one. In this section we describe our methods for enumerating configurations and present additional results for the assembly of 4-, 5-, and 6-membered heteromeric rings.

4.3.1 Enumerating distinct affinity configurations for heteromeric rings

To consider a large space of possible affinity configurations, we allowed the affinity at each subunit interface to vary over seven orders of magnitude, with a K_D of either 10^{-12}, 10^{-11}, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, or 10^{-6} M. If each subunit interface were considered to be distinct, this would imply $7^3 = 343$ unique configurations for the homomeric three-ring, $7^4 = 2401$ for the four-ring, and so on. However, as shown in Figure 15, such an approach would include many redundant configurations differing only in the (arbitrary) labeling of the subunits. For example, having a *strong* interaction between x_0 and x_1, a *medium* strength interaction between x_1 and x_2, and a *weak* interaction between x_2 and x_0 (a configuration we will denote $S - M - W$, with the abbreviated affinities enumerated in subunit counting order—see section 2.2.1), is equivalent in its assembly properties to the configuration having a *medium* strength interaction between x_0 and x_1, a *weak* interaction between x_1 and x_2, and a *strong* interaction between x_2 and x_0 (a configuration we denote $M - W - S$, using the same convention). The equivalence arises from the fact that the $S - M - W$ arrangement of interactions can be converted into the $M - W - S$ arrangement simply by shifting the labels of the subunits around the ring one position while preserving their sequential ordering. By thus considering the *rotational* symmetry of the interactions around the ring we can ignore these dynamically identical configurations.

In addition, since the affinity configurations under consideration are two-dimensional and do not distinguish a "top" or "bottom" for the ring, to identify unique configurations we must also note that the subunit counting direction—clockwise vs. counter-clockwise—also does not affect the dynamics. Changing the subunit counting direction is equivalent to *reflecting* the subunit labels about an axis in the plane, which can also be imagined as "flipping" the ring structure over. As shown in Figure 15, if the subunit labels for the ring with the affinity configuration $S - M - W$ are "flipped over" or reflected, it results in the nominally distinct configuration $W - M - S$; while the counting order of the interactions has changed from clockwise to counter-clockwise, the

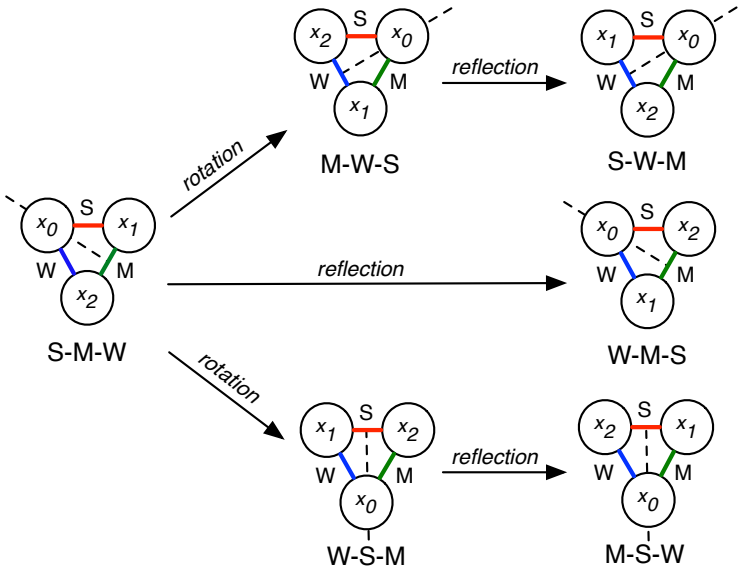

Figure 15: Ring symmetries. A unique affinity configuration can be given several different names depending on how subunits are labeled or counted. To identify all unique affinity configurations, all possible notationally distinct configurations are considered and then redundant configurations are eliminated. Redundant configurations are identified by their equivalence to an existing configuration via rotational symmetry, corresponding to a change in the placement of the first subunit label x_0 along the ring, or by symmetry of reflection, corresponding to a change in the direction in which subunits are enumerated. By considering these two types of symmetry, the six notationally distinct affinity configurations shown can be seen to be equivalent, and represented by a single member, $S - M - W$.

assembly dynamics of the configuration have not changed. Thus, nominally distinct configurations that are identical by the symmetries of rotation and reflection can be ignored, and only one representative of the family of equivalent configurations need be considered for analysis.

Proceeding in this fashion, we enumerated the unique affinity configurations for the 3-, 4-, 5-, and 6-membered rings, allowing the K_D at each interface to vary over seven orders of magnitude as described above. With the exclusion of configurations that are redundant by symmetry, we obtain a significantly reduced number of possible configurations (see Table 1).

Ring Length	**Unique Configurations**
3	81
4	406
5	1,855
6	10,528
7	60,028

Table 1: Numbers of distinct affinity configurations for rings of different lengths after accounting for symmetries (see Figure 15).

4.3.2 Numerical simulation results for the heteromeric 3-, 4-, 5-, and 6-rings

For each of the affinity configurations enumerated as above we ran numerical simulations of heteromeric ring assembly using the systems of ordinary differential equations described in sections 2.2.2, 2.4.2 and 2.4.4.

For the three-ring (Fig. 3A in the main text, reproduced here as Figure 16), the results show that configurations involving either one or two weak interactions assemble more efficiently than configurations involving uniform interactions. As shown in the inset plots, the differences in efficiency between configurations with one or two weak interactions are very small in absolute terms. It is also worth noting that the relative T_{99} rankings of one- versus two-weak interaction configurations is dependent in part on monomer concentration, while the model A and model B rankings are not.

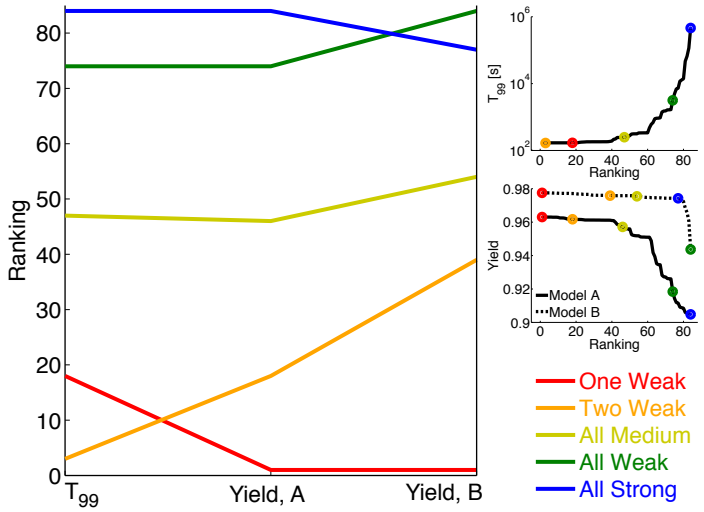

Figure 16: Ranking the assembly efficiency and yield of affinity configurations for the heteromeric three-ring. This figure corresponds to Fig. 3A in the main text and is included here for completeness. For calculation of T_{99}, simulations were performed with initial monomer concentrations of 477 nM for each subunit. For simulations of model A and model B degradation, $Q = 1.31 \cdot 10^{-10}$ M s^{-1} and $\delta = 2.75 \cdot 10^{-4}$ s^{-1}. $X_T = Q/\delta = 477$ nM. For all simulations, $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}. "One Weak" denotes a configuration with binding strengths (i.e., K_Ds) of $10^{-12}, 10^{-12}$, and 10^{-6} M. "Two Weak": $10^{-12}, 10^{-6}$, and 10^{-6} M; "All Medium": $10^{-8}, 10^{-8}$, and 10^{-8} M; "All Weak": $10^{-6}, 10^{-6}$, and 10^{-6} M; "All Strong": $10^{-12}, 10^{-12}$, and 10^{-12} M.

For the four-ring (Figure 17), the configuration containing alternating strong and weak interactions outranked other configurations based on T_{99} and model A yield; for model B, however, the configuration with a single weak interaction resulted in the highest yield. Configurations with either one or two weak interactions (with the rest strong) outperformed configurations with uniform interactions or more than two weak interactions.

The results for the five-ring (Figure 18) were similar to those for the four-ring: again, configurations with two weak interactions assembled with the lowest values for T_{99}, slightly ahead of the one-weak interaction configuration. However, the one-weak interaction configuration

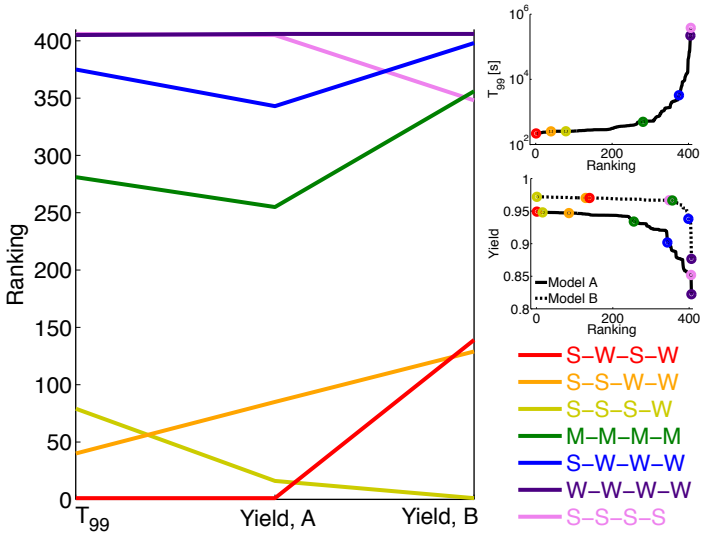

Figure 17: Ranking the assembly efficiency and yield of affinity configurations for the heteromeric four-ring. Parameters as for Figure 16. "S", "M", and "W" denote strong ($K_D = 10^{-12}$ M), medium ($K_D = 10^{-8}$ M), and weak ($K_D = 10^{-6}$ M) interactions, respectively.

produced the highest model B yield. Configurations with uniform affinities assembled both with low efficiency and low yield. Interestingly, changing the strength of one interaction in the configuration $S - S - S - W - W$ from strong to medium (resulting in the configuration $S - S - M - W - W$) dramatically reduces its performance relative to the other configurations.

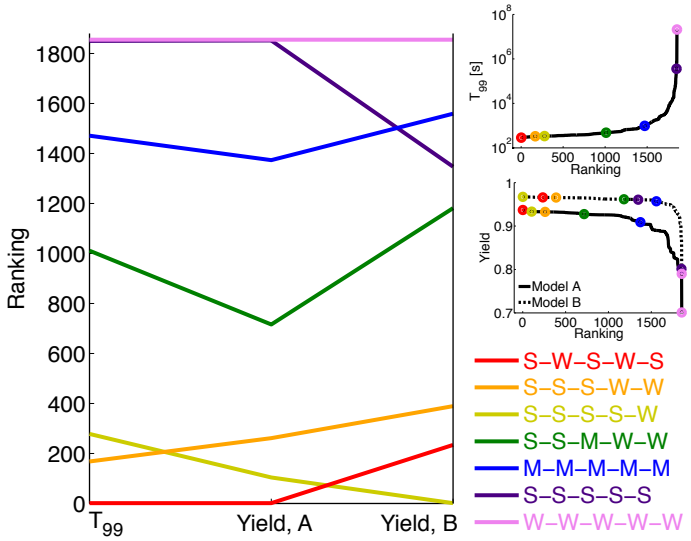

Figure 18: Ranking the assembly efficiency and yield of affinity configurations for the heteromeric five-ring. Parameters as for Figure 16; "S", "M", and "W" denote strong ($K_D = 10^{-12}$ M), medium ($K_D = 10^{-8}$ M), and weak ($K_D = 10^{-6}$ M) interactions, respectively.

The results for the six-ring (Figure 19) also indicate the assembly efficiency of rings with either one or two weak interactions. Both configurations with two weak interactions slightly outperform the one-weak interaction configuration for T_{99} and model A yield, while the one-weak interaction produces a higher model B yield. All three outperform uniform configurations or configurations with more than two weak interactions. Interestingly, the configuration with alternating strong and weak interactions does not perform as well by any metric as the configurations with either one or two weak interactions. This result, along with the results for the four-ring, in which alternating strong-weak interactions performed very well, suggests that the key to assembly efficiency is not alternating strong and weak interactions in and of itself, but the presence of *two* weak interactions. This may be due to the fact that a ring with two weak interactions may be able to assemble in two halves that then "snap" together to form a stable ring.

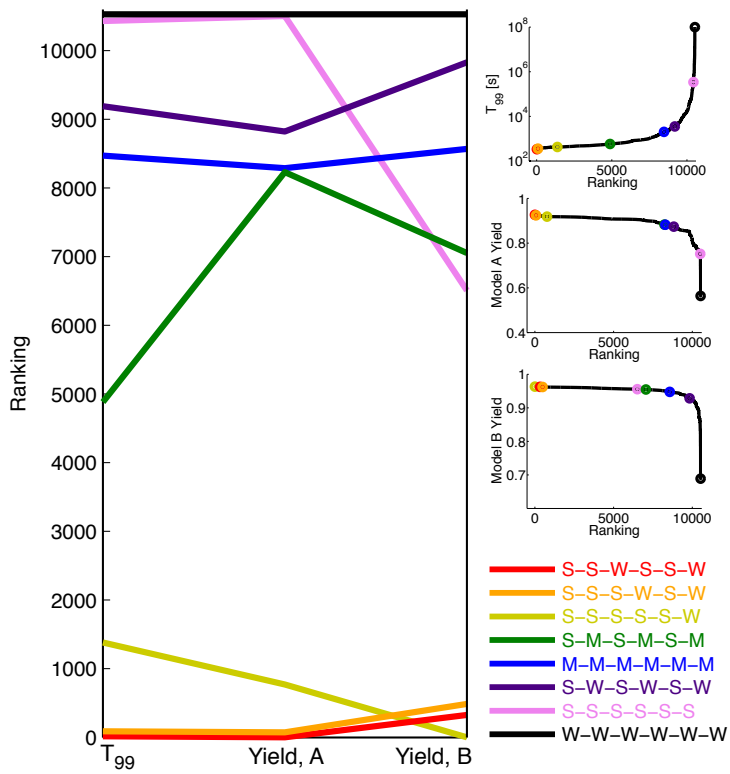

Figure 19: Ranking the assembly efficiency and yield of affinity configurations for the heteromeric six-ring. Parameters as for Figure 16. "S", "M", and "W" denote strong ($K_D = 10^{-12}$ M), medium ($K_D = 10^{-8}$ M), and weak ($K_D = 10^{-6}$ M) interactions, respectively.

4.3.3 Optimizing the assembly of chains

To provide a contrast to our results for heteromeric three-membered rings (Figure 16), we considered how affinity configuration influences the assembly efficiency of four-membered heteromeric chains. Although these two structures share the same number of interactions, chains lack the rotational symmetry of rings, and as such there are over twice as many distinct

configurations in this case (196 for the chains vs. 81 for the rings). Given some affinity configuration (and its attendant dissociation rates β_1 to β_3), we performed numerical simulations of simple heteromeric chain assembly, assembly with model A degradation, and assembly with model B assembly (the equations used for these calculations are listed in section 2.5).

The relative performance of these affinity configurations is compared in Figure 20. Since chains are generally much less stable than rings (see section 1), most of the configurations sampled here did not assemble to a level of 99%, making it impossible to define T_{99} as for rings. In simulations that did not consider synthesis and degradation, we thus calculated equilibrium yield instead ("Eq. Yield" in Figure 20). Note that, in stark contrast to the behavior for rings, for chains stronger is always better, regardless of the efficiency metric in question.

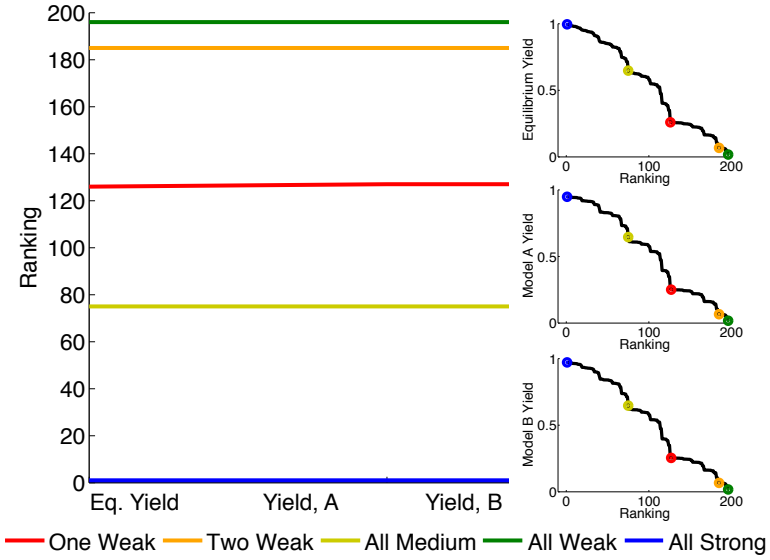

Figure 20: Ranking the assembly efficiency and yield of affinity configurations for the heteromeric 4-membered chain. Given that most affinity configurations do not assemble to 99% at the concentration considered here, in this case we use equilibrium yield (the 'Eq. Yield' column to the far left of the ranking plot) rather than T_{99} to characterize assembly efficiency in the absence of synthesis and degradation. The equilibrium yields were calculated at a total concentration of 477 nM for each monomer type. For simulations of model A and model B degradation, $Q = 1.31 \cdot 10^{-10}$ M s^{-1} and $\delta = 2.75 \cdot 10^{-4}$ s^{-1}. $X_T = Q/\delta = 477$ nM. For all simulations, $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}. "One Weak" denotes a configuration with binding strengths (i.e., K_Ds) of $10^{-12}, 10^{-12}$, and 10^{-6} M. "Two Weak": $10^{-12}, 10^{-6}$, and 10^{-6} M; "All Medium": $10^{-8}, 10^{-8}$, and 10^{-8} M; "All Weak": $10^{-6}, 10^{-6}$, and 10^{-6} M; "All Strong": $10^{-12}, 10^{-12}$, and 10^{-12} M.

4.4 Hierarchical assembly pathways

As discussed in section 4.3 above, introducing weak interactions into heteromeric ring structures can dramatically improve their assembly efficiency according to a wide variety of measures. In this section we compare an alternative mechanism by which kinetic assembly bottlenecks can be addressed: namely, the sequential, stepwise assembly of subunits to form the ring. In the case of the heteromeric three-ring, this type of assembly would imply that, for example, x_0 can bind x_1,

but the binding of x_1 to x_2 is contingent on x_1's being bound to x_0, and the binding of x_2 to x_0 is similarly contingent on x_2's being bound to x_1. This model represents a case where allosteric interactions—or other sources of hierarchical structure—force assembly to proceed sequentially, thus preventing the accumulation of assembly intermediates that are incompatible. Yin and coworkers have deployed precisely this approach to optimize the assembly process of cyclic DNA nanostructures [15]. To compare the assembly efficiency of this allosteric approach with the biophysical strategies described in section 4.3, we created and analyzed a model of stepwise assembly as described below.

4.4.1 Mathematical model of stepwise assembly

Before proceeding, we describe here the set of ordinary differential equations used to model the assembly of a heteromeric three-ring via sequential, stepwise assembly. We use the notational convention for heteromeric rings described in section 2.2.1. In this highly simplified model, there is no binding between x_1 and x_2, or between x_2 and x_0, unless x_0 has first bound to x_1. As a result, the concentrations of the dimers $x_{1,2}$ and $x_{2,2}$ are zero for all time. In addition, because the ring breakage rate is much smaller than the dissociation rate (i.e., $\gamma << \beta$) for the parameter values we consider below, for simplicity we ignore the process of ring breakage in this case (i.e. we set $\gamma = 0$).

The ODEs are as follows:

$$\begin{aligned}
\frac{dx_{0,1}}{dt} &= \beta x_{0,2} - \alpha x_{0,1} x_{1,1} \\
\frac{dx_{1,1}}{dt} &= \beta x_{0,2} - \alpha x_{0,1} x_{1,1} \\
\frac{dx_{2,1}}{dt} &= -\alpha x_{2,1} x_{0,2} \\
\frac{dx_{0,2}}{dt} &= \alpha x_{0,1} x_{1,1} - \beta x_{0,2} - \alpha x_{0,2} x_{2,1} \\
\frac{dx_3}{dt} &= \alpha x_{2,1} x_{0,2}.
\end{aligned} \tag{40}$$

4.4.2 Comparing sequential assembly with weak interactions

To compare between sequential assembly and the biophysical strategies discussed in section 4.3, we numerically integrated the ODEs from equation 40. In Figure 21 we compare the assembly dynamics of this sequential model, a ring containing a single very weak interaction, and a ring with uniformly strong interactions. In this case we have chosen affinity configurations such that the uniform case and the single weak interaction case exhibit identical thermodynamic stabilities. We find that the single weak interaction configuration always assembles faster than the sequential case, although the magnitude of this difference varies with total monomer concentration.

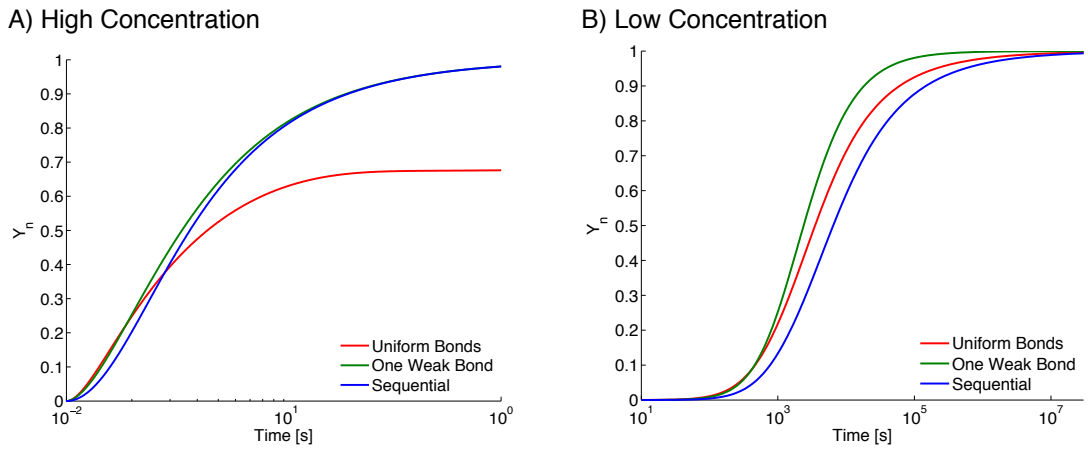

Figure 21: Assembly timecourses for three-membered heteromeric rings with uniform interactions (K_Ds of 10^{-9}), one weak interaction (K_Ds of $10^{-12}, 10^{-12}$, and 10^{-3}), and sequential assembly (uniform K_Ds of 10^{-9}). $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}. For simplicity, the ring breakage rate γ is set to zero in all cases. (A) Monomer concentration of 40 μM. (B) Monomer concentration of 0.4 nM. Although the three cases here have equivalent thermodynamic stability, they display different assembly kinetics. Note that the "one weak interaction" configuration consistently assembles faster than the sequential case, though the magnitude of this difference depends on concentration.

The differences observed in Figure 21 can be best understood in terms of a schematic view of the assembly process, Figure 22. In this illustration, we consider a pool of monomers A, B, and C that interact with each other to form a heteromeric ring. For simplicity, we do not explicitly consider trimer formation in this schematic, focusing entirely on the process of monomers binding to form dimers.

For the non-sequential assembly models (Figure 22, panels A and B), there are three "paths" that the system can take to assemble full rings from a starting pool of monomers: 1) bind A and B, then C; 2) bind A and C, then B; 3) bind B and C, then A. Configurations with uniformly strong interactions attempt to take all three paths simultaneously (represented by the fact that all possible dimers are present in Figure 22A). When concentrations are high, the system consumes all possible monomers too quickly, and since the interactions are strong, a plateau is induced (as discussed above and in the main text). During the plateau phase, assembly via any given path can only proceed when the system "backtracks" from one of the other paths through dissociation of a dimer. Uniformly strong interactions thus lead to unavoidable deadlocks at high concentrations (Figure 21A).

When one of the interactions along the ring is weak, although the system can in theory take all three possible assembly paths, only two of those paths will actually be taken by the majority of proteins in the system. If we make the $A - C$ interaction weak, any monomers attempting to take path number "2" by first forming an $A - C$ interaction will ultimately be unsuccessful, since this interaction will tend to break soon after formation. This fact is represented schematically in Figure 22B by the lack of $A - C$ dimers. As the schematic demonstrates, any monomers that take a given path (say, by forming the stable $A - B$ dimer) are guaranteed to have access to the

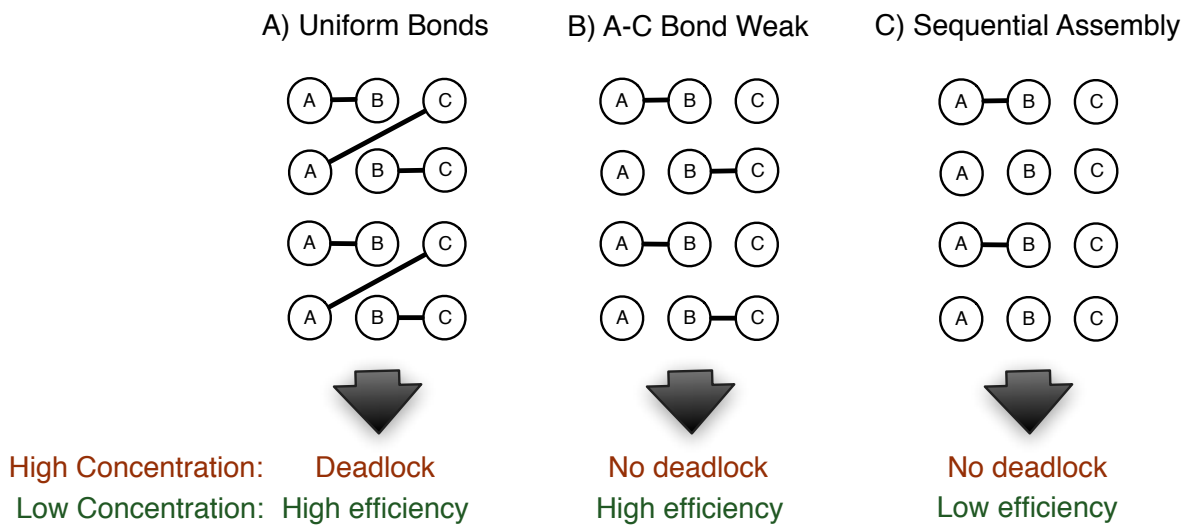

Figure 22: Schematic diagram of ring assembly. Shown are the interactions that would form after an arbitrary amount of time given a pool of monomers at equal concentrations and equivalent association rates. For simplicity, only the formation of dimers is considered here. (A) Uniform interactions: all possible dimers can form. (B) One weak interaction: $A - B$ and $B - C$ dimers can form, but encounters between A and C do not result in a stable interaction. (C) Sequential assembly: only encounters between A and B result in a stable interaction. Since the association rates are the same as in (A) and (B), the number of $A - B$ dimers formed is the same (two of the four possible).

cognate monomer needed to complete assembly (C in this example). Configurations with a single weak interaction thus avoid the problem of deadlock at high concentrations and achieve efficient assembly across a wide variety of conditions (see Fig. 3 in the main text).

In the sequential case, assembly can only proceed down a single path—in our example, this is path "1" (Figure 22C). This strategy avoids the potential for deadlock, and as such we do not see a plateau here (Figure 21A). The fact that sequential assembly occurs more slowly than "single weak interaction" assembly arises from the fact that the latter can take two paths concurrently, while the former is forced to take only one. Since the association rate is assumed to be identical for both cases, weak interaction assembly initially exhibits twice the number of productive reactions (a fact schematized by the existence of both $A - B$ and $B - C$ dimers in Figure 22B vs. only $A - B$ dimers in Figure 22C).

We thus find that inclusion of a single weak interaction in a "concurrently" assembling trimer provides the best of both worlds—fast assembly at low concentrations while avoiding deadlock at high concentrations. It is important to note that our analysis here is not exhaustive, and there may be conditions in which sequential assembly will be favored over the concurrent case. We leave further exploration of these alternative approaches to future work.

4.5 Non-uniform concentrations

Since each protein in a heteromeric ring will be transcribed and translated from a separate gene, such rings have the capacity to not only demonstrate variation in affinities (as discussed above) but also the total concentration of each subunit. Such differences could easily arise from the inherent stochastic nature of gene expression [16], or from differences in gene regulation between subunits in the cell. To test the potential effects of such variation, we considered the assembly dynamics of a three-membered heteromeric ring in a situation where one subunit has a higher total concentration than the other two.

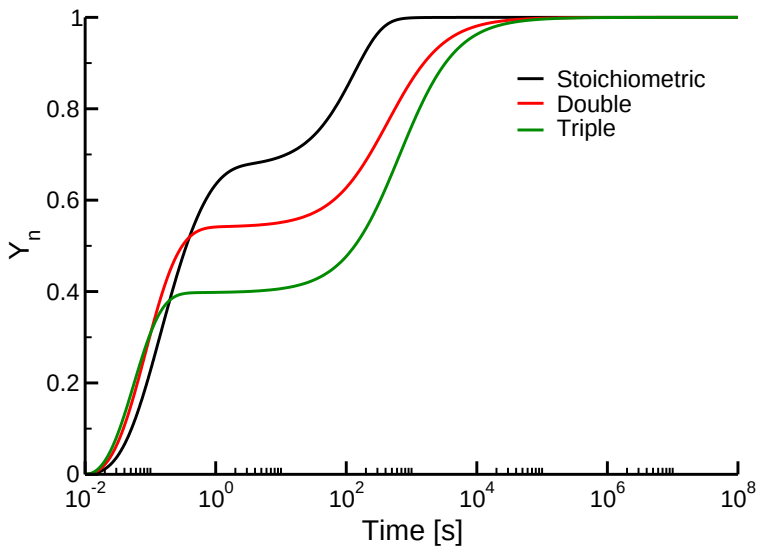

Figure 23: Assembly dynamics of a three-member heteromeric ring with non-uniform subunit concentrations. The black curve represents a case where all subunits are in stoichiometric concentrations (i.e. the situation considered extensively above and in the main text). The red curve represents a case where a single subunit is at double the concentration of the other two, and the green curve represents a case where one subunit is triple the concentration of the other two. As one can see, as the difference in concentrations increases, the plateau becomes "lower" and persists longer (i.e. T_{99} increases dramatically). In all plots, two of the subunits are present at 4 μM concentration, while the other varies from 4 to 12 μM. The affinities in this case are uniform with K_Ds of 10^{-9} M, and the parameter $\alpha = 2.53 \cdot 10^6$ M^{-1} s^{-1}.

As one can see from Figure 23, increasing the concentration of a single subunit exacerbates deadlock, resulting in a deadlocked plateau that occurs at a lower assembly yield and that persists longer. This occurs because the subunit that is at higher concentration (say, the "A" subunit of a "ABC" heteromeric ring) rapidly binds to the other two subunits, forming a comparatively large number of AB and AC dimers that must dissociate in order for assembly to proceed to completion. These findings highlight the fact that the dynamics of assembly in this case depend not solely on K_Ds (i.e. the free energy of binding) but also on subunit concentrations (which influence the *chemical potential* of the bimolecular reactions in question). It is currently unclear if either hierarchical assembly or simple affinity configurations can overcome the increased deadlock

resulting from non-uniform concentrations. We leave complete consideration of how chemical potential landscapes might evolve to manage differences in subunit concentration to future work.

5 Analysis of structural data

5.1 Structures for heteromeric three-membered rings

Our work strongly implies that structures that include at least one weak interaction will enjoy an evolutionary advantage, either in terms of assembly time or steady-state yield (Fig. 3A of the main text). We explored the solved structures of ring-like complexes in order to assess if there was any evidence for the existence of such weak interactions in the ring-like complexes found in living systems.

We began by assembling a set of three-membered heteromeric rings of known structure. We used the database 3D Complex [17] as a starting point for collecting the structures, and based our analysis on all heteromeric three-membered complexes in that database with a ring-like topology. In order to reduce the redundancy of the data set (i.e. to avoid considering two very closely related or identical rings as different examples of evolutionarily optimized structures), we utilized the "QS-90" level of the 3D Complex hierarchy [17]. At QS-90, complexes with greater than 90% sequence identity are grouped together into a single class, from which a single representative structure is taken. Using the QS-90 level of the hierarchy allows us to ignore cases where multiple mutant forms of the same complex, or very closely related complexes, have solved structures in the PDB.

We curated the resulting 82 heteromeric three-rings in the database in order to remove structures in which the biology of assembly did not match the case considered by our model. Specifically, we removed structures in the following four classes:

1. **Antibody-Antigen Complexes** Antibodies consist of two polypeptide chains (Heavy and Light) that interact extensively with each other. In many cases, both chains interact with an antigen, thus forming a ring-like topology. Biologically, however, antibodies are synthesized and secreted in the absence of antigen, and only then bind to the antigen in question. Our model does not cover this case, and so we do not consider this type of complex in our data set. It is important to note, however, that the interactions to the antigen generally involve much smaller surface area than the Heavy-Light interaction, but even though these structures support the conclusions of our model we cannot be certain that the evolutionary pressures on this system are equivalent to those implied by our model. Of the 82 ring structures in the initial dataset, 28 belonged to this class.

2. **Integral Membrane Complexes** The vast majority of these cases involve the extracellular domains of dimeric membrane-bound receptors binding to monomeric cytokines. This situation, in which two members of the complex are constrained to a membrane surface and one can diffuse in three-dimensional space, presents a very different set of assembly challenges compared to the model considered in this work. In addition, some cytokine-receptor binding events induce conformational changes in the receptor that influence receptor dimerization, an effect which is also neglected in our model. As with the

antibody case, most of these structures contain at least one "weak" interaction, but we nonetheless ignore them due to the fact that they do not conform to the assumptions our model. Of the 82 ring structures in the initial dataset, 15 belonged to this class.

3. **Complexes Produced through Proteolysis** This class consists of cases where a subset of the individual chains in the structure are produced when a single chain is cleaved in two. The majority of these cases involve proteases (e.g. trypsin) complexed with protease inhibitors. Proteases such as trypsin are synthesized and fold as a single polypeptide chain (the zymogen). Activation of the zymogen involves a proteolytic cleavage event in which this single chain is cut in two. This produces two polypeptide chains that are intricately folded with one another. In these structures, the interaction between the two chains of the protease is not formed through bimolecular association, but rather through folding as a single chain and subsequent cleavage. This situation is clearly distinct from the assembly dynamics considered in our models, although it is again the case that the protease-inhibitor interactions are considerably weaker than the interaction between the two proteases (a fact which would support our model if the data were included). Of the 82 structures in the initial dataset, 8 belonged to this class.

4. **Miscellaneous** This class consisted of one structure in which binding between two of the subunits was induced by a small molecule (FK506), and one case in which the complex assembles around DNA. Neither case conforms to the assumptions of our model, and so these two structures are also removed from the set.

After curating the 82 structures of heteromeric three-membered rings from 3D Complex, we obtained 29 structures for which the biological system represented by the structure seemed to represent a case similar to that considered in our model. Of these 29 structures, many are enzymes (e.g. glutamine amidotransferase) and many serve regulatory functions (e.g. the complex of the transcription factor NF-κB with its regulator IκBα). A list of all structures can be found in a table provided as additional supplementary material.

5.2 Structures for heteromeric four-membered chains

To serve as a contrast to the case of the three-membered rings discussed above, we also considered heteromeric four-membered chains. A heteromeric four-membered chain contains exactly the same number of interactions as a three-membered ring, making it possible to perform a direct comparison between the two types of structures. This comparison is particularly informative due to the fact that optimizing assembly in the case of chains will tend to favor uniformly strong interactions (as demonstrated by the analyses in sections 3.2.3 and 4.3.3).

To assemble this dataset we began with the 104 heteromeric four-membered chains found at the QS-90 level of the 3D Complex hierarchy. We removed structures in the following classes from the data set:

1. **Integral membrane complexes** As with the heteromeric rings discussed above, a number of the chain structures in this data set included one or more proteins that exist as integral membrane proteins in the cell. In general, these structures involved the extracellular

domains of receptors (e.g. the T-cell receptor) complexed with ligands and/or other receptors. As with the three-membered rings, we did not consider the particular evolutionary pressures that might arise in the case of interactions involving integral membrane proteins. Of the 104 chain structures in the initial dataset, 23 belong to this class.

2. **Biological assembly mis-annotated** The 3D Complex database is constructed on the basis of the biological assemblies included in default PDB files [17]. As has been noted elsewhere [18, 19], these complexes are sometimes distinct from the assembly considered biologically relevant in the very manuscript in which the crystal structure at issue is reported. In this case, if the paper reporting a particular structure, or other relevant literature, indicated that the biologically active form of the complex was not a four-membered chain, we removed that structure from our data set. In some cases, we found that authors of the paper reporting the structure cited direct size-exclusion chromatography results indicating that the biologically relevant form of the molecule was a dimer, despite the fact that the "biological assembly" in the PDB was reported as a tetramer. Of the 104 chain structures in the initial dataset, we identified 18 for which there was strong evidence that the biologically relevant form of the complex was not in fact a four-membered chain.

3. **Antibody-antigen complexes** A number of structures in the data set consisted of antibodies binding to various antigens. As with the rings above, we removed these structures from our data set as they represent a case of assembly quite different from that considered here. Of the 104 chain structures in the initial dataset, 8 consisted of antibody-antigen complexes.

4. **Complexes produced through proteolysis** This class consisted of protease and lectin molecules in which at least two of the chains in the final structure are synthesized as a single polypeptide sequence which is later cleaved to give the final, active structure. Since two or more of the chains in these structures do not interact with one another through a bimolecular association event, we did not include these cases in our analysis. Of the 104 chain structures in the initial dataset, 7 involved complexes in which this type of proteolytic cleavage was involved.

5. **DNA-binding complexes** These structures consisted of complexes that assemble around specific DNA sequences. Since assembly on DNA is not considered in our model, we did not include these cases in our analysis. Of the 104 chain structures in the initial dataset, 4 involved complexes assembling on DNA.

After curating the data set into the above classes, we obtained a dataset of 44 heteromeric four-membered chains; as discussed in section 5.3 below, 11 of these structures actually represented four-membered rings upon further analysis. Of the remaining 33 structures, most represent either enzymes or enzyme-inhibitor complexes. A list of these structures is provided as an additional supplementary table.

5.3 Calculating changes in non-polar surface area

Examining affinity distributions in rings and chains using solved protein structures involves estimating the binary binding affinities between components of the structure. Here we focus on measuring the change in Solvent-Accessible Non-Polar Surface Area ($\Delta\text{SASA}_{\text{NP}}$), which has been shown to correlate with binding affinities in some studies [20, 21]. We used the software package POPS [22] to perform this calculation. We proceeded by creating three separate PDB files: file (1) contains only the atoms (ATOM records) that belong to the residues of the first chain (say, chain "`A`"), file (2) contains only the atoms that belong to the residues of the second chain ("`B`"), and file (3) contains the atoms from both chains ("`A + B`"). Next, we calculated the non-polar solvent-accessible surface area for each file separately using POPS. This area is marked as "hydrophobic" in the POPS output. We then calculated $\Delta\text{SASA}_{\text{NP}}$ as the difference between the sum of these areas for each domain separately and the area for the domains combined:

$$\Delta\text{SASA}_{\text{NP}}(\texttt{A},\texttt{B}) = \text{SASA}_{\text{NP}}(\texttt{A}) + \text{SASA}_{\text{NP}}(\texttt{B}) - \text{SASA}_{\text{NP}}(\texttt{A}+\texttt{B}). \tag{41}$$

In other words, $\Delta\text{SASA}_{\text{NP}}$ is calculated as the SASA_{NP} of file (1) plus the SASA_{NP} of file (2) minus the SASA_{NP} of file (3). We used the definition in equation 41 to calculate $\Delta\text{SASA}_{\text{NP}}$ for every pair of chains in the structure of interest.

For the curated three-membered heteromeric ring structures, all of the cases yielded the expected ring-like topology when subjected to this analysis. The case with the four-membered chains was more complex. Many of the structures in this case actually contained more than 3 interactions (defined as $\Delta\text{SASA}_{\text{NP}}$ greater than some cutoff). Since the only way to include more than 3 unique interactions in a graph of four nodes involves creating a cycle, this analysis indicated that a number of the structures considered as "chains" in 3D Complex actually contained rings. We used the minimum "affinity" observed for rings (129.67 Å^2) as a cutoff and found that 11 of the 44 curated chains actually exhibited significant ring-like structure, a fact that we confirmed through simple visual analysis of the structure itself in each case. This analysis left 33 "true" four-membered heteromeric chains in our data set.

It is important to note that equation 41 represents a very rough and imperfect measure of affinity [20, 21, 23]. As such, we also explored using total interface area (as opposed to non-polar area), fractional surface area (defined as the area of the interface divided by the entire SASA of the binary complex), and the ΔG values annotated in PISA [24]. We found that these alternative definitions of affinity were strongly correlated with $\Delta\text{SASA}_{\text{NP}}$ ($R^2 > 0.8$ in each case), thus yielding nearly identical results to those discussed for $\Delta\text{SASA}_{\text{NP}}$ below. Of course, all of these definitions are imperfect [23], but in the absence of empirical data regarding affinities in this case, the structural analysis presented here represents the only available test of our predictions.

5.4 Comparing rings and chains

As discussed above, every structure in our ring or chain data sets contains exactly three interactions with $\Delta\text{SASA}_{\text{NP}} > 129.67$ Å^2. For any structure, we can thus define the largest interface (or "Strong" interaction, denoted S), the smallest interface (or "Weak" interaction, denoted W), and the interface in between the two (i.e. the "Medium" interaction, denoted M).

Fig. 4A in the main text represents a summary comparison of the S and W interactions for rings and chains, which we expand upon here.

Figure 24 shows a histogram, the kernel-smoothed density as well as a "rug" plot of the raw data for the S interactions of both rings and chains, while Figure 25 represents the same plot for the W interactions. In the case of S interactions, the mean $\Delta\mathrm{SASA}_{\mathrm{NP}}$ for rings is $\overline{\mathrm{S_R}} = 2099$ Å^2, while for chains we have $\overline{\mathrm{S_C}} = 1631$ Å^2. To test if this difference in means is significant, we performed a simple random permutation test with 10^5 replicates in the statistical package R [11]. We found that the difference is not significant after correcting for multiple comparisons (the uncorrected $p = 0.0441$). For weak interactions, we have $\overline{\mathrm{W_R}} = 531$ Å^2 for rings and $\overline{\mathrm{W_C}} = 914$ Å^2 for chains. The difference in means for weak interactions was considerably more significant ($p = 6 \cdot 10^{-5}$). From this analysis we can conclude that the strong interactions in rings are, on average, stronger than those for chains (with the caveat of weak statistical significance), while the weak interactions in rings are considerably weaker than the weak interactions in chains.

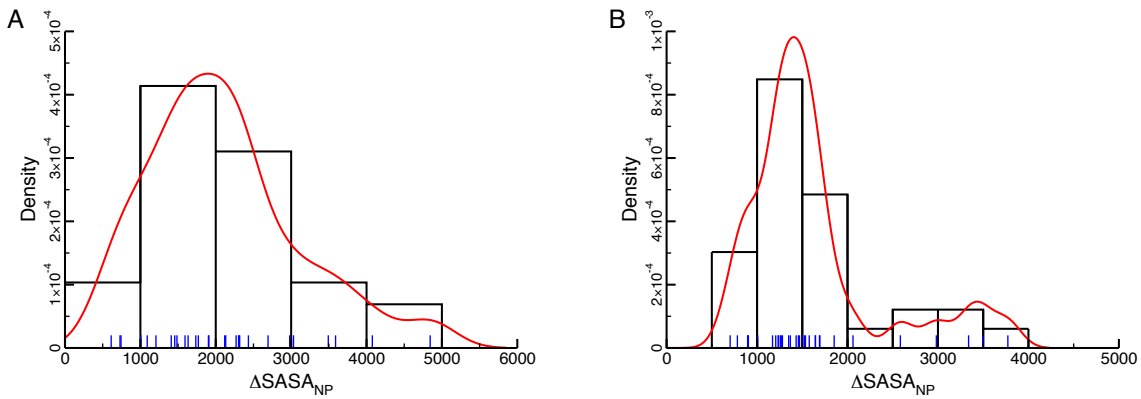

Figure 24: Strong interactions in both rings and chains. (A) Here we plot the distribution of "S" affinities for rings. The black boxes represent a histogram of the data, while the red line corresponds to a kernel-smoothed density. The blue lines on the x-axis are a "rug" plot of the data, where each line represents the affinity for a single ring in the data set. The rug plot is provided to give a sense for the data underlying both the histogram and the kernel-smoothed density. (B) A plot as in panel A, but for the S affinities in chains.

It is important to note that the assembly properties of a ring or chain with a particular set of affinities will vary strongly with total monomer concentration (see section 4.1.1). We thus also considered the weak-to-strong interaction ratios (W/S); a plot of the ratio densities for rings and chains can be found in Fig. 4B of the main text. Figure 26 shows this density, as well as histograms and rug plots, for rings and chains separately. Again using a permutation test, we found that the mean ratios for rings, $\overline{(\mathrm{W/S})_{\mathrm{R}}} = 0.309$, is significantly smaller than that for chains $\overline{(\mathrm{W/S})_{\mathrm{C}}} = 0.627$ ($p = 10^{-5}$). Comparison of both the absolute affinity and relative affinity distributions reveals that the weakest interaction in rings is significantly weaker than the weakest interaction in chains, as our assembly models would predict.

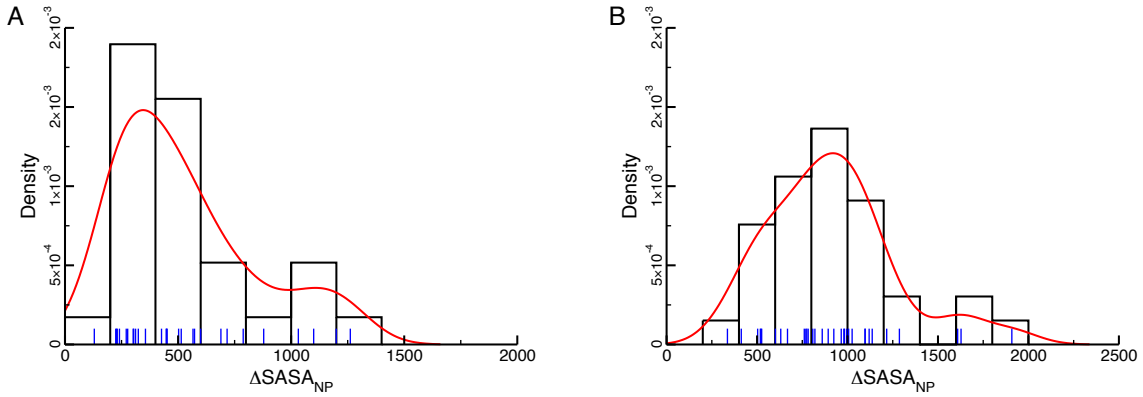

Figure 25: Weak interactions in both rings and chains. (A) Here we plot the distribution of "W" affinities for rings. The black boxes represent a histogram of the data, while the red line corresponds to a kernel-smoothed density. The blue lines on the x-axis are a "rug" plot of the data, where each line represents the affinity for a single ring in the data set. The rug plot is provided to give a sense for the data underlying both the histogram and the kernel-smoothed density. (B) A plot as in panel A, but for the W affinities in chains.

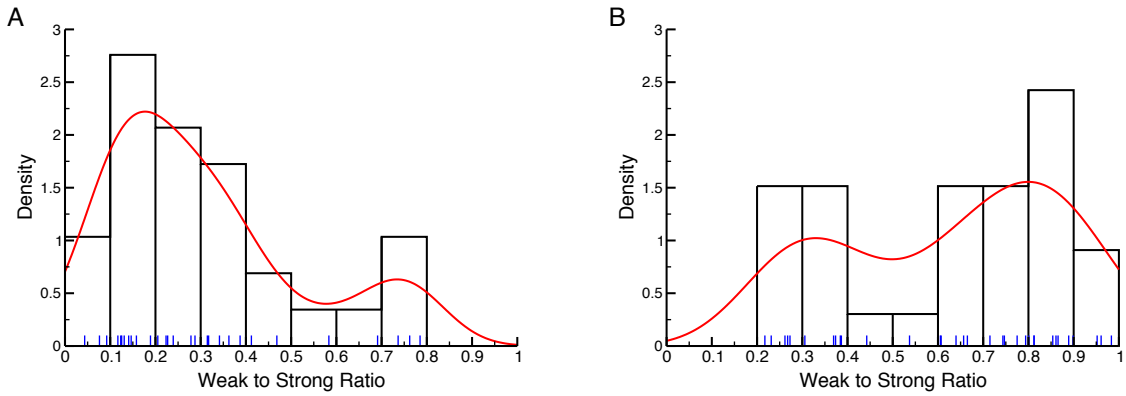

Figure 26: The weak-to-strong ratio for both rings and chains. (A) Here we plot the distribution of the weak to strong ratio (W/S) for rings. The black boxes represent a histogram of the data, while the red line corresponds to a kernel-smoothed density. The blue lines on the x-axis are a "rug" plot of the data, where each line represents the affinity for a single ring in the data set. The rug plot is provided to give a sense for the data underlying both the histogram and the kernel-smoothed density. (B) A plot as in panel A, but for the W/S ratios found in chains.

Interestingly, the kernel-smoothed density estimates and the histograms for both rings and chains demonstrate considerable bimodality (see Figure 26). For the rings, we can divide the data into

the first peak (with $(\mathrm{W/S})_\mathrm{R} < 0.5$) and the second peak ($(\mathrm{W/S})_\mathrm{R} > 0.5$). As can be seen from Figure 26, the majority of rings (24 of 29) belong to the first peak, and taking this peak alone we have $\overline{(\mathrm{W/S})_\mathrm{R}} = 0.225$, which is significantly smaller than the average for the entire sample. The 5 points belonging to the second peak have a much higher average, with $\overline{(\mathrm{W/S})_\mathrm{R}} = 0.712$. Similarly, we can use a ratio of 0.5 to divide the chains into two peaks, and for 11 structures in the smaller ratio peak we have $\overline{(\mathrm{W/S})_\mathrm{C}} = 0.319$ while for the 23 points in the second peak we have $\overline{(\mathrm{W/S})_\mathrm{C}} = 0.781$. As Figure 27 demonstrates, normal quantile-quantile plots for the major peaks in both cases (the smaller-ratio peak for the rings and the larger-ratio peak for the chains) reveal that both can be well-approximated by Gaussian distributions (the smaller peaks in both cases contain too few data points to support conclusions regarding normality). Although we do not have enough data to make this point conclusively, we can speculate that both ratio distributions are the result of two combined Gaussian distributions, one with a smaller average ratio, and one with a larger.

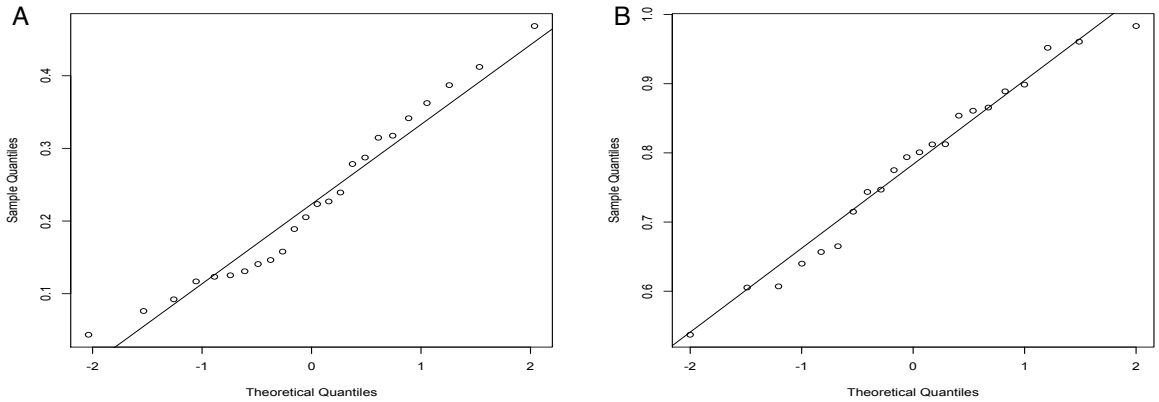

Figure 27: Quantile-Quantile plots comparing the ratio distributions of rings and chains with normal distributions. (A) Here we compare the quantiles of the small-ratio W/S distribution for rings (e.g. those rings with W/S < 0.5) to the quantiles of the normal distribution. The "Sample Quantiles" on the y-axis are taken from our data on rings, while the "Theoretical Quantiles" for the normal distribution are computed in R. The solid black line represents a linear fit to the Q-Q data. The linear fit in this case is fairly good, with strong statistical significance for both the slope and intercept terms ($p < 2 \cdot 10^{-16}$). Although there are slight systematic deviations from the straight line throughout the range of quantiles, this result indicates that the small-ratio distribution is approximately Gaussian in character. (B) In this case we compare the quantiles of the large-ratio W/S distribution for chains (e.g. those chains with W/S > 0.5) to the quantiles of the normal distribution. As in panel A, the "Sample Quantiles" on the y-axis are taken from our data on chains, while the "Theoretical Quantiles" for the normal distribution are computed in R. The solid line is again a linear fit to the Q-Q data; the fit in this case is excellent, with $p < 2 \cdot 10^{-16}$ for both the slope and intercept terms. There is less systematic deviation from the fit in this case, indicating that the large-W/S peak for chains can be fairly well approximated Gaussian distribution.

As can be seen from Fig. 4A in the main text, the smaller-ratio peak for the chains overlaps with that peak for the rings, and vice versa. Using a permutation test, we found that smaller-ratio chains had a significantly higher average than smaller-ratio rings (although the significance is

fairly weak after correcting for multiple comparisons, with an uncorrected $p = 0.0199$), while the averages for the larger-ratio chains and rings were statistically indistinguishable ($p = 0.254$).

We thus find that 24 of the 29 rings considered in our data set conform to our prediction that rings will contain at least one "weak" interaction. The 5 remaining structures do not seem to contain a weak interaction (i.e. they seem to belong to the chain distribution rather than the ring one). These cases may represent situations in which the cell can accommodate sub-optimal assembly efficiency, or cases in which mechanisms not considered in our model (e.g. chaperone-mediated assembly) are involved.

5.5 Gaussian control

Although the above analysis indicates that rings and chains, on average, conform to the expectations of our model, it is unclear to what extent our observations represent evolutionary pressures on affinities. For instance, say we have some positive continuous random variable X that follows an underlying probability density p. Sampling three instances of X from this distribution will naturally result in a "largest" and "smallest" value for the sample, and the ratio of these two numbers will always be less than 1.

We thus conducted a control to test whether we would observe W/S ratio distributions similar to the results for rings and chains when the interactions themselves are sampled from a single underlying distribution. We considered a simple model in which a "structure" is constructed by sampling 3 $\Delta\mathrm{SASA_{NP}}$ from a Gaussian distribution. We set the mean μ of the Gaussian to be the mean of all the interactions ($A = S \cup M \cup W$) in a given data set; for rings we have $\overline{\mathrm{A_R}} = 1255$ Å^2 and for chains we have $\overline{\mathrm{A_C}} = 1296$ Å^2. The standard deviation σ of the Gaussian was set to the sample standard deviation in each case: $s(\mathrm{A_R}) = 966$ Å^2 and $s(\mathrm{A_C}) = 700$ Å^2.

Of course, $\Delta\mathrm{SASA_{NP}}$ values cannot be negative, and indeed in section 5.3 we defined a cutoff for considering only interactions with $\Delta\mathrm{SASA_{NP}} > 129.67$ Å^2 as valid. In order to mimic these constraints, we must thus introduce a minimum affinity (129.67 Å^2) and reject affinities below that cutoff. If we only implement a lower bound, however, the set of sampled interactions from the above procedure will exhibit a mean significantly different from the underlying Gaussian used to construct the data. To prevent this from happening, we implement an upper bound such that the z-score of this upper bound is equal to the absolute value of the z-score for the chosen minimum (that is, $z_{max} = -z_{min}$). This allows us to construct a distribution of random affinities from the underlying Gaussian with a minimum possible affinity that is nonetheless symmetric and exhibits the defined average.

Each random structure sampled from the distribution as defined above has a S, M and W interaction, and for each structure we calculate the W/S ratio. A "model" data set is constructed from N such structures, where $N = 29$ for rings and 33 for chains to mimic the distributions we observe in the real data. We constructed 10^4 such data sets (for a total of $2.9 \cdot 10^5$ structures in the case of the rings) and asked what fraction of these random datasets exhibited $\overline{\mathrm{W/S}}$ ratios smaller than or equal to that observed for the rings and larger than or equal to that observed for chains.

In every case, we find that this Gaussian control is unlikely to explain the data: $p = 9 \cdot 10^{-4}$ for

rings and $p < 10^{-4}$ for chains.

Fig. 4C in the main text is meant to summarize the results of this control graphically. In that case, we have a single Gaussian distribution with an average taken to be approximately that observed for rings and chains ($\mu = 1250$ Å^2). We vary the standard deviation from $\sigma = 25$ Å^2 to 2500 Å^2 and maintain a cutoff of 129.67 Å^2. In this case, we take $N = 30$ for each data set and we construct 10^4 data sets for each value of σ. In Figure 4C we plot the average W/S and 95% confidence intervals for random data sets constructed this way as a function of σ/μ.

It is important to note that we have controlled here only for one type of underlying distribution; namely a Gaussian with a particular minimum affinity cutoff. Although this control is clearly unlikely to produce the data, one could potentially find some other single underlying distribution of affinities that could. As Fig. 4B in the main text indicates, however, even if this is the case, one could argue that evolution has selected parameters for this underlying distribution (e.g. μ and σ) such that configurations with optimal assembly characteristics are likely.

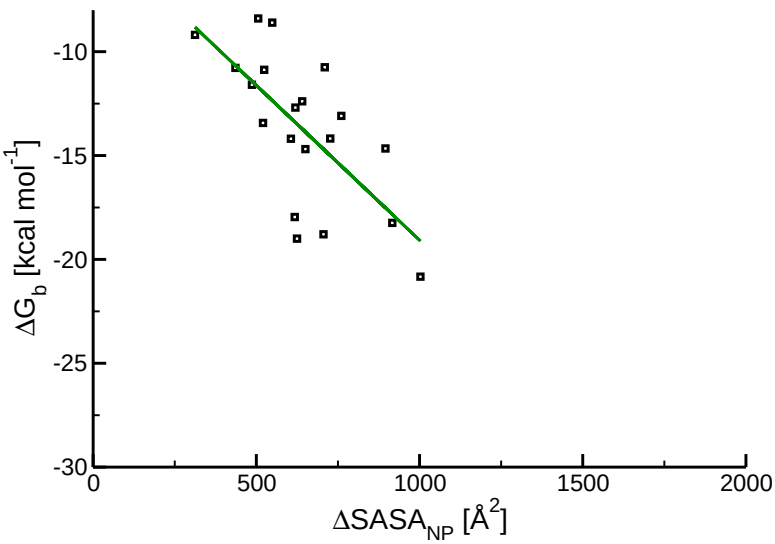

Figure 28: Relationship between the change in solvent-accessible non-polar surface area and binding free energy. The squares in the plot represent data taken from Table 1 in reference [21]; in this case the ΔG_b values are obtained from experimental measurements, and we determined the $\Delta\text{SASA}_{\text{NP}}$ directly from the corresponding crystal structures using POPS [22] as described in section 5.3 above. The green line is a linear fit to the data, which yields an $R^2 = 0.47$.

5.6 Affinities for the interactions in the crystal structures of rings

One can use available crystal structures of interacting proteins for which affinities are known to investigate the quantitative relationship between $\Delta\text{SASA}_{\text{NP}}$ and ΔG_b^0. Using a recently-published

data set of 20 such structures [21], we found a roughly linear relationship between the two, with an R^2 of 0.47 (see Figure 28). Although the correlation is imperfect, the linear fit allows us to map $\Delta\text{SASA}_{\text{NP}}$ values into ΔG_b^0 values and thus K_Ds for the interactions in our data set. The equation we obtain in this case is:

$$\Delta G_b^0(\mathtt{A},\mathtt{B}) = -0.015 \cdot \Delta\text{SASA}_{\text{NP}}(\mathtt{A},\mathtt{B}) - 4.17$$

with $\Delta\text{SASA}_{\text{NP}}(\mathtt{A},\mathtt{B})$ given in Å^2 and $\Delta G_b^0(\mathtt{A},\mathtt{B})$ in kcal mol^{-1}. Assuming approximately room temperatures (i.e. $RT \approx 0.6$ kcal mol^{-1}), the average K_D for strong bonds in our rings data set is $8.0 \cdot 10^{-12}$, and the average K_D for weak bonds is $1.8 \cdot 10^{-6}$. Interestingly, these are very close to the values used for Figures 2A and 3B in the main text, as well as the optimum values obtained in our analysis of heteromeric rings (Fig. 3A of the main text). As mentioned above (and as is clear from Figure 28), $\Delta\text{SASA}_{\text{NP}}$ is only a very rough measure of actual binding affinity [20, 21, 23]; these results simply indicate that the K_D values we use for "strong" and "weak" bonds in the text (e.g. 10^{-12} and 10^{-6} M, respectively) are at least broadly consistent with the range of affinities one would expect given the buried surface areas in the crystal structures of homomeric rings.

References

1. Saiz L, Vilar JM (2006) Stochastic dynamics of macromolecular-assembly networks. *Mol Syst Biol* 2:2006 0024.

2. Henriksen NE, Hansen FY (2008) *Theories of Molecular Reaction Dynamics: The Microscopic Foundation of Chemical Kinetics* (Oxford University Press, Oxford).

3. Camacho CJ, Kimura SR, DeLisi C, Vajda S (2000) Kinetics of desolvation-mediated protein-protein binding. *Biophys J* 78:1094–105.

4. Minh DD, et al. (2005) The entropic cost of protein-protein association: a case study on acetylcholinesterase binding to fasciculin-2. *Biophys J* 89:L25–7.

5. Grziwa A, et al. (1994) Dissociation and reconstitution of the thermoplasma proteasome. *Eur J Biochem* 223:1061–7.

6. Nakabayashi J, Sasaki A (2006) A mathematical model for apoptosome assembly: the optimal cytochrome c/apaf-1 ratio. *Journal of Theoretical Biology* 242:280–7.

7. Sharma S, Hoskins JR, Wickner S (2005) Binding and degradation of heterodimeric substrates by clpap and clpxp. *J Biol Chem* 280:5449–55.

8. Moore SD, Baker TA, Sauer RT (2008) Forced extraction of targeted components from complex macromolecular assemblies. *Proc Natl Acad Sci USA* 105:11685–90.

9. The MathWorks, Inc. (2010) *MATLAB Version 7.11* (The MathWorks, Inc., Natick, Massachusetts).

10. Wolfram Research (2008) *Mathematica Version 7* (Wolfram Research, Champaign, Illinois).

11. R Development Core Team (2010) *R: A Language and Environment for Statistical Computing* (The R Foundation for Statistical Computing, Vienna, Austria).

12. Kress W, Mutschler H, Weber-Ban E (2007) Assembly pathway of an aaa+ protein: Tracking clpa and clpap complex formation in real time. *Biochemistry* 46:6183–6193.

13. Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.

14. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci USA* 103:13004–9.

15. Yin P, Choi HMT, Calvert CR, Pierce NA (2008) Programming biomolecular self-assembly pathways. *Nature* 451:318–22.

16. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–6.

17. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA (2006) 3d complex: a structural classification of protein complexes. *PLoS Comput Biol* 2:e155.

18. Xu Q, et al. (2008) Statistical analysis of interface similarity in crystals of homologous proteins. *Journal of molecular biology* 381:487–507.

19. Xu Q, Dunbrack RL (2011) The protein common interface database (protcid)–a comprehensive database of interactions of homologous proteins in multiple crystal forms. *Nucleic Acids Research* 39:D761–70.

20. Horton N, Lewis M (1992) Calculation of the free energy of association for protein complexes. *Protein Sci* 1:169–81.

21. Bougouffa S, Warwicker J (2008) Volume-based solvation models out-perform area-based models in combined studies of wild-type and mutated protein-protein interfaces. *BMC bioinformatics* 9:448.

22. Cavallo L, Kleinjung J, Fraternali F (2003) Pops: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Research* 31:3364–6.

23. Kastritis PL, et al. (2011) A structure-based benchmark for protein-protein binding affinity. *Protein Sci* 20:482–491.

24. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* 372:774–97.