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*).

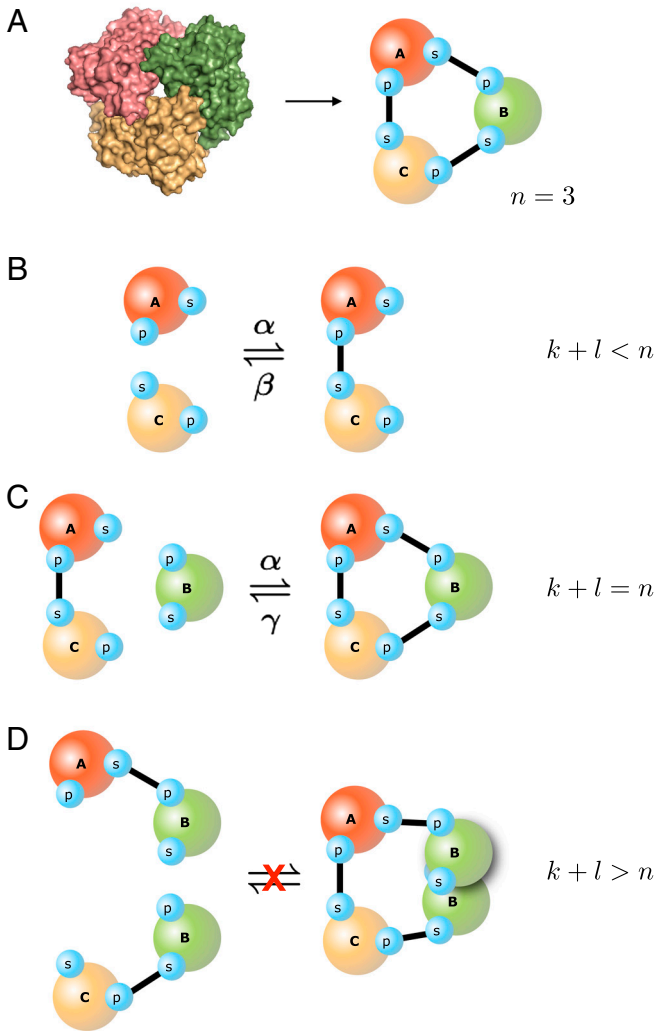

Fig. 1. Schematic of ring assembly. (*A*) A three-membered ring (X-ray structure from Protein Data Bank ID 2JB8) on the left is represented on the right as a graph involving three proteins, each with two binding interfaces (the small circles on the periphery of the nodes). (*B*) A pair of monomers bind 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

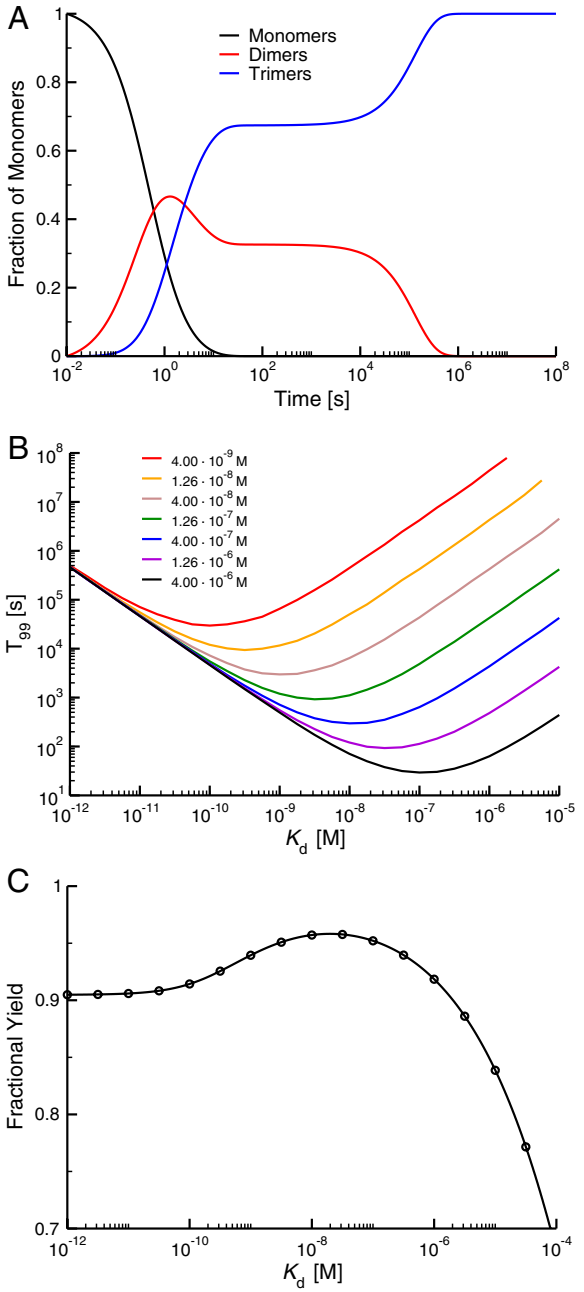

Fig. 2. Assembly of a homomeric three-membered ring. (*A*) In this graph, we consider the percentage of monomers in the various association states (monomer, dimer, and trimer) as a function of time. The affinities are uniformly very strong ($K_d = 10^{-12}$ M). The data are plotted on a logarithmic timescale 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).

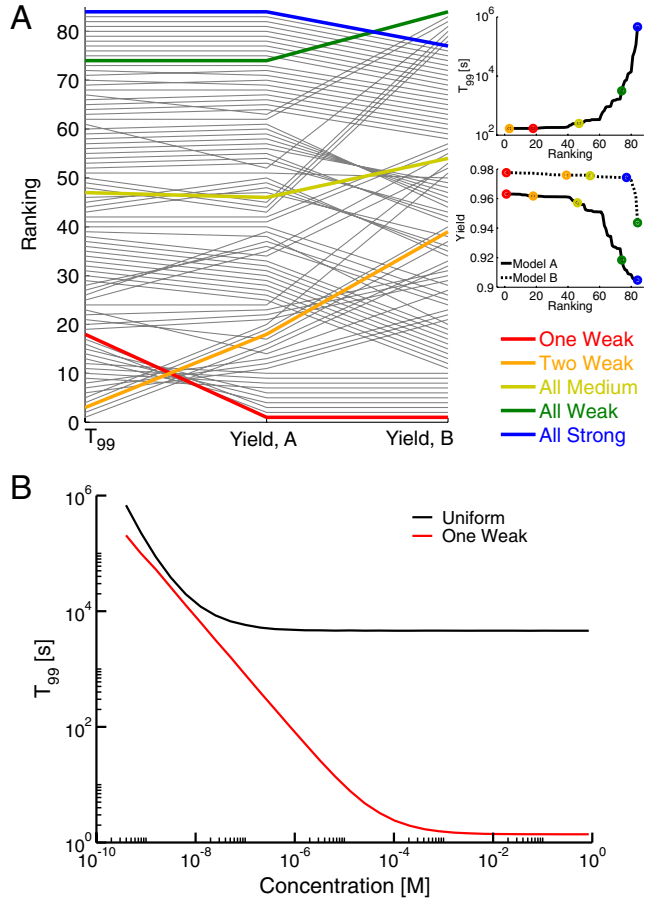

Fig. 3. Nonuniform affinities. (*A*) This plot shows the relative performance of different affinity configurations (gray lines) in the following categories: assembly time (measured by T_{99}), steady-state yield for model A (Yield, A), and steady-state yield for model B (Yield, B). Each affinity configuration 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-

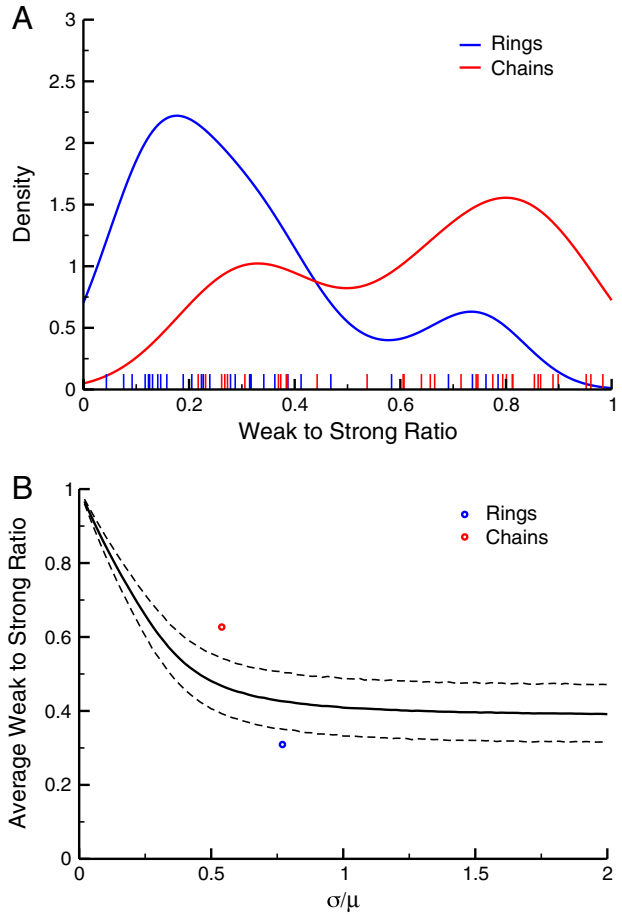

Fig. 4. Weak interactions in crystal structures. (*A*) Estimated probability densities of the weak-to-strong affinity ratio (W/S) for rings and chains. The "strength" of an interaction is measured by the change in solvent-accessible nonpolar surface area as described in the *SI Appendix*, Section 5.3. 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).