

# Emergence of symmetry in homooligomeric biological assemblies

Ingemar André<sup>\*†</sup>, Charlie E. M. Strauss<sup>†‡</sup>, David B. Kaplan<sup>§</sup>, Philip Bradley<sup>¶</sup>, and David Baker<sup>\*||\*\*</sup>

<sup>\*</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195; <sup>†</sup>Los Alamos National Laboratory, Los Alamos, NM 87545; <sup>§</sup>Institute for Nuclear Theory, Box 351550, Seattle, WA 98195-1550; <sup>¶</sup>Fred Hutchinson Cancer Research Center, Seattle, WA 98109; and <sup>||</sup>Howard Hughes Medical Institute, Seattle, WA 98195

Contributed by David Baker, August 7, 2008 (sent for review May 19, 2008)

**Naturally occurring homooligomeric protein complexes exhibit striking internal symmetry. The evolutionary origins of this symmetry have been the subject of considerable speculation; proposals for the advantages associated with symmetry include greater folding efficiency, reduced aggregation, amenability to allosteric regulation, and greater adaptability. An alternative possibility stems from the idea that to contribute to fitness, and hence be subject to evolutionary optimization, a complex must be significantly populated, which implies that the interaction energy between monomers in the ancestors of modern-day complexes must have been sufficient to at least partially overcome the entropic cost of association. Here, we investigate the effects of this bias toward very-low-energy complexes on the distribution of symmetry in primordial homooligomers modeled as randomly interacting pairs of monomers. We demonstrate quantitatively that a bias toward very-low-energy complexes can result in the emergence of symmetry from random ensembles in which the overall frequency of symmetric complexes is vanishingly small. This result is corroborated by using explicit protein–protein docking calculations to generate ensembles of randomly docked complexes: the fraction of these that are symmetric increases from 0.02% in the overall population to >50% in very low energy subpopulations.**

evolution | protein structure | Monte Carlo Simulation | modeling

The origin of the striking symmetry of homooligomeric protein assemblies has been the subject of much speculation. The potential benefits of symmetric assemblies to an organism include increased coding economy (1), greater stability, reduced aggregation, robustness to errors in synthesis (2, 3), and amenability to allosteric regulation (3). It has also been proposed that structural symmetry arises because it makes proteins more designable and foldable, or is a result of domain swapping (4–7). Monod and colleagues argued that symmetrical arrangements enhance the sensitivity of selection because “effects of single mutations occurring in a symmetrical oligomer . . . should be greatly amplified as compared to the effects of similar mutations in a monomer or non-symmetrical oligomer,” since each mutation at the interface effectively counts twice (3). In analogy to atomic clusters it has been suggested that symmetric oligomeric arrangements may, in general, be low in energy (8). Shakhnovich and coworkers argued that homodimers are favored over heterodimers because the mean energies of homodimeric interfaces are likely to be lower than those of heterodimeric interfaces (9, 10). Many of these arguments are discussed in greater detail in an excellent review by Goodsell and Olsen (2).

As with many evolutionary questions, it is difficult to assess which of the factors described above contributed to the symmetry of modern-day assemblies and to what extent. However, with the advent of molecular modeling, it is now possible to assess the sufficiency of quantitative hypotheses in this area. In this article we develop a quantitative model for the emergence of symmetry in biological assemblies that is based on explicit docking calculations on realistic protein models.

The key premise underlying our approach is that selection is only likely to operate on primordial complexes with sufficient initial interaction energy to at least partially overcome the entropic costs of association of the monomers; evolution can only optimize a complex that is populated sufficiently to confer a benefit to the organism. For example, contrast the effects of a mutation that stabilizes an interface by 1 kcal/mol on the population of binding modes with interaction energies of 0 and –8 kcal/mol, respectively. Assuming that the free-energy cost arising from the entropy loss in association is 10 kcal/mol, the mutation increases the population of the first binding mode from  $5 \times 10^{-8}$  to  $3 \times 10^{-7}$ , and the second, from 0.035 to 0.18. Because the population of the first binding mode is vanishingly small even after mutation, it is unlikely to contribute to fitness, and hence, the second, much lower energy binding mode is a more likely starting point for evolutionary optimization by mutation and selection.

We investigate the extent to which such a bias toward very-low-energy primordial complexes could have contributed to the emergence of the highly symmetric homooligomeric complexes ubiquitous in nature today. We show that there are two competing factors that determine the fraction of complexes with a given degree of symmetry in a subpopulation: for complexes of increasing symmetry, the number of possible configurations decreases while the variance in their interaction energy grows. We show by using both simple analytic models and explicit protein–protein docking simulations that the combination of these two counteracting factors leads to a predominance of symmetric structures in very low energy binding modes. In contrast with most previous suggestions, in our model, symmetrical complexes are not favored because they provide specific functional advantages, but because the initial pool of protein complexes available for competitive selection is heavily biased toward symmetrical binding modes.

## Results and Discussion

To determine whether a bias toward low energy alone can give rise to highly symmetric complexes, we first quantify the extent of symmetry in randomly docked complexes, and show that a majority of naturally occurring complexes are more symmetric than 99.98% of random complexes. Next, we show that the variance in the interaction energy distribution for randomly docked complexes increases with increasing symmetry. We then combine these results and find that increasingly stringent bias toward low-energy structures yields increasingly symmetric pop-

Author contributions: I.A., C.E.M.S., D.B.K., P.B., and D.B. designed research; I.A., C.E.M.S., D.B.K., and D.B. performed research; I.A., C.E.M.S., D.B.K., and D.B. analyzed data; and I.A., C.E.M.S., and D.B. wrote the paper.

The authors declare no conflict of interest.

<sup>†</sup>I.A. and C.E.M.S. contributed equally to this work.

<sup>\*\*</sup>To whom correspondence should be addressed. E-mail: dabaker@u.washington.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0807576105/DCSupplemental](http://www.pnas.org/cgi/content/full/0807576105/DCSupplemental).

© 2008 by The National Academy of Sciences of the USA

**Table 1. Increase in symmetry produced by refinement and selection**

Protein monomer	PDB ID code	Length	$P(S_{\text{dev}} < 0.2 \text{ \AA})$ in %		
			Random ensemble	Refined ensemble (low/high resolution)	Low-energy ensemble (lowest 2.5%)
Protein L	1hz6	61	0.02	0.35/0.31	2.9
Ubiquitin	1ogw	76	0.03	0.18/0.19	1.9
Ribosome binding factor	1tif	59	0.03	0.29/0.35	2.9
cheY	2chy	128	0.02	0.10/0.19	3.2
Fatty acid binding protein	1lfo	127	0.01	0.10/0.15	2.6

The fraction of symmetrical structures ( $S_{\text{dev}} < 0.2 \text{ \AA}$ ) in a ensemble of dimer binding modes after random docking, after energy refinement by using low/high-resolution descriptions of the proteins, and for the lowest 2.5% energy structures.

ulations. At each step, the results from simple model calculations are corroborated by using explicit protein–protein docking simulations; in particular, we find that the fraction of symmetric complexes generated in protein docking simulations (11) increases from 0.02% in the overall population to over 50% in very-low-energy subpopulations.

**Extent of Symmetry in Randomly Generated Complexes.** We begin by comparing the symmetry of naturally occurring homodimeric proteins to that of randomly generated homodimeric complexes. To quantify the notion of *nearly* symmetric requires a metric for deviations from perfect symmetry in homodimer complexes. Let A and B be any two atoms in one protein monomer, and A' and B' be the corresponding atoms in the other monomer. If the dimer is perfectly symmetric, then

$$\text{distance}(A, B') = \text{distance}(A', B) \text{ for all pairs of atoms.}$$

We define the symmetry measure  $S_{\text{dev}}$  to be the average of  $|\text{dist}(A, B') - \text{dist}(A', B)|$  over all pairs of  $\alpha$ -carbon atoms. This atomic-distance-based measure is useful, among other reasons, because the system energy is largely a function of interatomic distances. Perfect symmetry corresponds to  $S_{\text{dev}} = 0$ .

Estimates of the  $S_{\text{dev}}$  probability distribution,  $P(S_{\text{dev}})$ , for random homodimers were obtained by randomly orienting pairs of monomeric protein structures (Table 1) followed by a translation to bring them into atomic contact with no energy minimization. The  $S_{\text{dev}}$  probability distributions for different proteins (Fig. 1A, yellow, green, and brown lines) are very similar, with an approximately linear increase in probability density with increasing  $S_{\text{dev}}$  starting from zero at perfect symmetry. To investigate the generality of this result, we determine  $P(S_{\text{dev}})$  for a simple model system composed of a pair of identical spheres of radius  $r_g$  with atoms uniformly distributed on their surfaces. For finite distance between the two spheres, this calculation (supporting information (SI) Appendix) involves an integral that must be evaluated numerically. In the limit of infinite distance between the spheres, the integral can be evaluated analytically, and the result is simple:

$$P(S_{\text{dev}}^{\infty}) = \frac{9}{8r_g^2} S_{\text{dev}}^{\infty}; \quad 0 < S_{\text{dev}}^{\infty} < \frac{4}{3} r_g. \quad [1]$$

As shown in Fig. 1A, the interacting sphere model (blue lines) qualitatively recapitulates the results for the random protein complexes (yellow, green, and brown lines). The infinite-distance solution (dashed blue line) is a reasonable approximation to the exact solution even at small intermonomer distances (solid blue line), in particular, when the deviation,  $S_{\text{dev}}$ , from perfect symmetry is small.

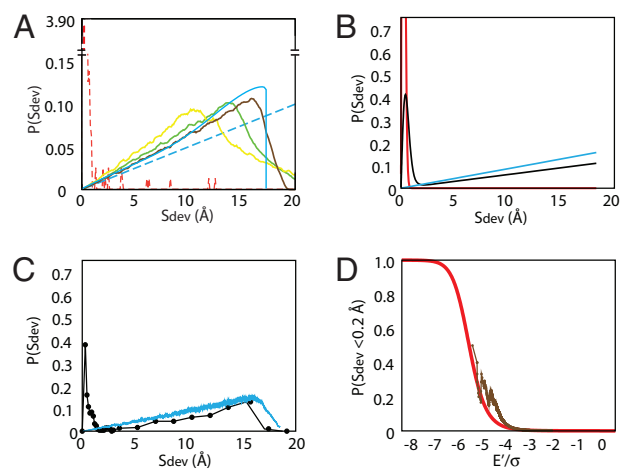
The empirical  $S_{\text{dev}}$  distribution for a set of 796 native protein homodimers contrasts strikingly with that of randomly generated dimeric complexes as illustrated in Fig. 1A (red line). The

majority of native structures are more symmetric than >99.98% of randomly docked complexes; the probability of observing a random complex with  $S_{\text{dev}} < 0.2$ , as is typical for native homodimers, is  $\approx 1/5,000$ .

The contrast between the high symmetry of native complexes and low symmetry of random homodimers frames the central objective of this article: to establish a quantitative explanation for the high symmetry of oligomeric assemblies observed in nature.

### Interaction Energy Distribution for Randomly Generated Complexes.

Consider a set of randomly generated homodimeric interfaces, each of which contains  $N$  independent interactions across the interface. Supposing each of these interactions is drawn inde-



**Fig. 1.** Emergence of symmetry from ensembles of randomly docked homodimeric complexes. (A) Comparison of  $S_{\text{dev}}$  distributions for interacting spheres, complexes formed by randomly docking monomeric proteins, and native protein homodimers. Solid blue, numerical solution of supplementary Eq. 1 for contacting spheres; dashed blue, analytic solution (Eq. 1) for spheres at infinite distance; Yellow (1hz6), green (1hz6) and brown (2chy), numerical results for random protein homodimer complexes generated by explicit protein docking. Red,  $S_{\text{dev}}$  distribution of 796 naturally occurring protein homodimer structures. The y axis is broken to accommodate the sharp peak near  $S_{\text{dev}} = 0$  for the naturally occurring complexes. (B) Increase in symmetry in random homodimeric complexes with increasingly stringent energy-based selection obtained by numerically integrating Eq. 3. Blue, no energy threshold; black, energy  $E < -4.6 \sigma$ ; red,  $E < -10.0 \sigma$  (A more negative threshold implies a tighter binding for function.) (C) Numerical results from Rosetta all-atom protein docking calculations with 2chy. Blue,  $S_{\text{dev}}$  distribution for randomly generated docked complexes (no energy threshold); black,  $S_{\text{dev}}$  distribution for very-low-energy docked complexes with  $E < -4.6 \sigma$  below the mean. (D) Increase in the fraction of symmetric structures ( $S_{\text{dev}} < 0.2 \text{ \AA}$ ) in population after energy-based selection; brown, result from 2chy docking simulations; red, simple model result (SI Appendix, Eq. S8). As discussed in the SI Appendix, we estimate  $\sigma$  to be typically  $\approx 1 \text{ kcal/mol}$  (Table S1).

pendently from a distribution with standard deviation  $E$ , the distribution for the total interaction energy has a standard deviation  $\sigma_{\text{asym}} = \sqrt{NE}$ . For the special case of symmetric homodimer interfaces, there are  $N/2$  distinct interactions, each of which occurs twice, and the total interaction energy has standard deviation  $\sigma_{\text{sym}} = \sqrt{N/2}(2E) = \sqrt{2}\sigma_{\text{asym}}$ . For higher-order symmetric oligomers, the same argument yields  $j$ -fold greater variance in the interaction energy distribution for  $j$ -fold symmetric oligomers over their asymmetric counterparts. This is a generalization of the result obtained by Shakhnovich and coworkers for flat disks containing a single type of interaction center in their study of the evolutionary advantages of homodimers over heterodimers (9, 10).

We tested the prediction of a broader energy distribution for symmetrical protein arrangements by using docking simulations. As shown in Fig. S1, the distribution of interaction energies in randomly generated  $C_2$  symmetric complexes is indeed broader than that of the randomly generated asymmetric complexes described above; the variance is  $\approx 1.8$  times greater in the symmetric case, reasonably close to the theoretical estimate of 2. The distributions are Gaussian, as might be anticipated from the Central Limit Theorem applied to a sum of random interactions (see SI Appendix for discussion).

The analysis is readily extended to complexes that deviate from perfect symmetry. The interaction energy is the sum of pairs of (A, B') and (A', B) interactions; in the  $S_{\text{dev}} = 0 \text{ \AA}$  limit these are perfectly correlated and for large values of  $S_{\text{dev}}$  they are completely uncorrelated. In the intervening partially symmetric regime, the interaction energies are partially correlated. The difference in the AB' distance and the A'B distance is on average  $S_{\text{dev}}$ , and hence the correlation  $\rho(S_{\text{dev}})$  between the AB' and A'B interaction energies decays as  $S_{\text{dev}}$  approaches the characteristic length of the interaction potential. Assuming for concreteness a Gaussian decay in the interaction energy correlation over length  $L$  ( $\approx 1 \text{ \AA}$ ), we obtain (see SI Appendix and Fig. S2)

$$\sigma(S_{\text{dev}}) = \sigma_{\text{asym}} \sqrt{1 + \rho(S_{\text{dev}})} = \sigma_{\text{asym}} \sqrt{1 + \exp(-S_{\text{dev}}^2/2L^2)}. \quad [2]$$

The Gaussian energy distribution for complexes with symmetry  $S_{\text{dev}}$  is then

$$P(E|S_{\text{dev}}) = \frac{\text{const}}{\sigma(S_{\text{dev}})} \exp\left(-\frac{1}{2}\left(\frac{E}{\sigma(S_{\text{dev}})}\right)^2\right). \quad [3]$$

**Emergence of Symmetry.** Combining Eqs. 1 and 3, we find that the frequency of complexes with interaction energy  $E$  and symmetry  $S_{\text{dev}}$  is

$$\begin{aligned} P(S_{\text{dev}}, E) &= P(S_{\text{dev}})P(E|S_{\text{dev}}) \\ &= \text{const} \times \frac{S_{\text{dev}}}{\sigma(S_{\text{dev}})} \exp\left(-\frac{1}{2}\left(\frac{E}{\sigma(S_{\text{dev}})}\right)^2\right). \end{aligned} \quad [4]$$

Evolutionary selection can operate only on complexes sufficiently populated to contribute to function; thus, we are interested in the distribution of  $S_{\text{dev}}$  for complexes with binding energy strong enough to at least partially overcome the 10–15 kcal/mol entropic cost of association (12). We obtain the  $S_{\text{dev}}$  distribution as a function of the binding energy threshold  $E'$  required for a contribution to fitness by numerically integrating  $P(S_{\text{dev}}, E)$  from  $-\infty$  to  $E'$ . As shown in Fig. 1B, with increasingly stringent thresholds (i.e., more negative  $E'$ ), the probability distribution shifts from predominantly asymmetric (i.e., high- $S_{\text{dev}}$ ) complexes in the random population (blue) to increasingly symmetric (black); when the energy threshold drops below  $-10\sigma$

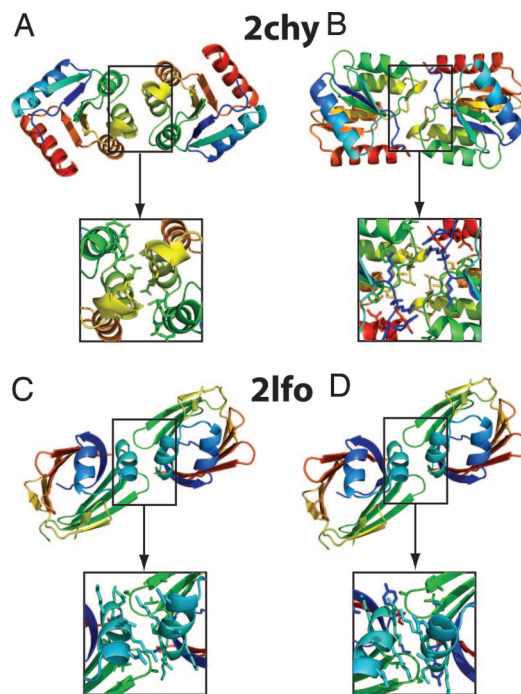


Fig. 2. Very-low-energy complexes generated by random docking of monomeric proteins are predominantly highly symmetrical. (Left) The most symmetrical binding mode of the three lowest-energy structures after high-resolution docking of the normally monomeric proteins 2chy (A) and 2lfo (C). (Right) Low-energy symmetric binding modes after sequence evolution (see SI Appendix): 2chy (B) and 2lfo (D). Sequence optimization lowers the interaction energies and can change the relative energy of different binding modes, but does not fundamentally change their symmetry.

(red), the distribution is similar to that of the highly symmetric natural homodimers (Fig. 1A, red). In the intermediate case (black), the distribution is bimodal with dominantly asymmetric and dominantly symmetric subpopulations; we quantify the abruptness of this transition between the red and blue modes further below.

We investigated whether this striking emergence of symmetry is manifested in explicit protein docking calculations. The randomly docked complexes described above were refined by using the Rosetta Monte Carlo protein docking protocol (11) to produce well packed protein-like interfaces. For each of the five protein test cases, this energy-based refinement alone considerably increased the extent of symmetry in the population (Table 1; see the SI Appendix and Fig. S3 for discussion). We then selected the subset of refined structures with energies below different cutoff thresholds and determined the  $S_{\text{dev}}$  distribution in each low-energy subpopulation. As shown in Fig. 1C for a typical protein, this energy-based enrichment (black) shifts the original random distribution (blue) dramatically toward low values of  $S_{\text{dev}}$ , and closely parallels the analytic model results (Fig. 1B).

The increase in the fraction of nearly symmetric structures ( $S_{\text{dev}} < 0.2 \text{ \AA}$ ) with more stringent energy-based selection threshold is illustrated in Fig. 1D. The simple model predicts a relatively abrupt phase-transition behavior as the energy cutoff for selection drops below  $-5\sigma$  (Fig. 1D, red). This increase in symmetry after energy-based selection is recapitulated in the populations of refined docked protein complexes: increasingly stringent bias toward low-energy structures produces populations with increasing fractions of nearly symmetric ( $S_{\text{dev}} < 0.2 \text{ \AA}$ ) structures (Fig. 1D, brown line, and Table 1). Examples of

very-low-energy docked complexes with high symmetry are shown on the left side of Fig. 2.

The preceding analysis suggests that the primordial complexes with populations sufficient to contribute to fitness and hence be subject to natural selection were very likely to be symmetrical. In our model, these states will form the starting point for subsequent evolutionary optimization by mutation and selection. As discussed earlier, Monod *et al.* (3) suggested that symmetrical interfaces have a further advantage at this evolutionary optimization stage because the effect of a mutation is effectively amplified by symmetry. Indeed, the argument presented in the paragraphs preceding Eq. 2 carries over immediately to the mutation case: the standard deviation of the interaction energy associated with single mutations at interfaces is  $\sqrt{2}$ -fold greater for symmetric complexes than for asymmetric complexes because each mutation effectively occurs twice. However, this effect is counteracted to some extent by the lower frequency of mutations at symmetric interfaces because asymmetric interfaces can contain up to twofold more independently mutable residues. We investigated the trade-off between these two effects by carrying out explicit simulations of sequence evolution on the randomly generated protein complexes described above. As described in the *SI Appendix* and Fig. S4, in some but not all cases, cycles of sequence evolution and selection increase the extent of symmetry in the population. Examples of low-energy symmetric structures resulting from cycles of mutation and selection are shown on the right side of Fig. 2.

The analysis in this article has focused primarily on oligomeric forms in which there is a single type of interaction patch between monomers. Routes to more complex oligomers have been discussed by Monod *et al.* (3), Hansen (13), Teichmann *et al.* (14, 15), and Bennett *et al.* (4, 5). Teichmann and coworkers have shown that higher-order oligomers are often built up of smaller symmetric dimers and trimers that associate by using additional interaction patches (15). Our findings of increased likelihood of symmetry among lowest-energy complexes are relevant to both the early stage and late stage in acquisition of oligomer structure in their model; for example, in a tetramer, the initial forming of dimers from monomers, and subsequently the formation of tetramer from interacting dimers.

As noted in the introduction, Shakhnovich and coworkers have studied a question related to the problem investigated here; the energetic advantage of homodimeric versus heterodimeric complexes (9, 10). In the two-dimensional interacting disk model used in those studies, homodimeric complexes are necessarily symmetric, so the problem of the origin of symmetry of homodimeric complexes did not arise. For this model the authors found a twofold greater variance in the homodimer energy distribution than in the heterodimer energy distribution, which parallels our finding for symmetric versus asymmetric homodimeric complexes in this article. Although provocative, in the absence of a model for the decrease in the number of states with increasing symmetry, the earlier work does not indicate whether selection for low energy alone is sufficient to result in the emergence of symmetric structures. (See *SI Appendix* for further discussion.)

The importance of estimating the prior probability of symmetric complexes is illustrated by the apparent inconsistency between the prevalence in nature of homodimers (and dimers of dimers) (15) and the  $j$ -fold increase in the variance in energy in cyclic oligomers with  $J$  subunits as noted in the above analysis of the interaction energy distribution for random complexes, which should favor the higher-order symmetric structures in the low-energy subpopulation. As emphasized earlier, the fraction of symmetric structures in the low-energy population depends not only on the energy variance, but also on the prior probability of forming a symmetric complex. The probability of randomly forming a symmetric complex is much smaller for higher-order

cyclic oligomers<sup>††</sup>, and the balance between the two factors favors symmetric dimers over higher-order symmetric oligomers.

The key result of this article is that symmetric assemblies can arise solely from a bias toward complexes with sufficiently favorable interaction energy to be significantly populated despite the entropic cost of association. Particularly notable is the almost complete shift toward the  $S_{\text{dev}}$  distribution of native complexes (Fig. 1A, red) produced by selecting the lowest-energy complexes (Fig. 1C, black) from large populations of randomly docked complexes (Fig. 1C, blue) generated in Rosetta all-atom docking simulations. This enrichment occurs because the increase in the variance of the interaction energy with increasing symmetry more than compensates for the decrease in the density of states with increasing symmetry. As shown in the *SI Appendix*, if the variance in the interaction energy becomes large, as perhaps is the case of partially folded monomers that can reconfigure to optimize interactions with other monomers, symmetry can arise at equilibrium in the absence of selection, which may have relevance to amyloid formation. Similar considerations may also have relevance to the nucleation of symmetry from intrinsically nonsymmetric monomers in nonbiological assemblies such as crystals and clusters. Finally, although we have shown that the pool of randomly arising candidates for eventual competitive selection are *a priori* extremely biased toward symmetry, it must be emphasized that, as in any argument about history, it is nearly impossible to evaluate the relative contributions of different potential causes; along with the initial bias described here, the factors noted by Monod *et al.* (3, 16) and others summarized in the introduction likely contributed to the high symmetry of modern homooligomers.

## Methods

The protein-modeling software Rosetta was used to perform the protein-protein docking and protein design calculations (11). Rosetta uses real-space Monte Carlo minimization to find the lowest-energy conformation of binding partners. The first step in the docking protocol consists of a low-resolution search where only the backbone atoms and a single "centroid" atom representing the side chains are modeled. Random starting points for docking are generated by randomizing the rotational degrees of freedom of the binding partners, and then translating them into contact. Fig. 1A and column 4 of Table 1 was generated at this stage. In the subsequent high-resolution docking stage, a detailed physically realistic all-atom energy function (17) is optimized by using Monte Carlo minimization protocol. Each step in this protocol consists of (i) a small random perturbation to the rigid body degrees of freedom, (ii) discrete optimization of the side-chain rotamer conformations, and (iii) quasi-Newton optimization of the side chain and rigid body degrees of freedom (11).

The *SI Appendix* contains a more detailed description of the methodology, notably a more detailed description of the docking and design calculations, derivations of the analytical results, detailed descriptions of the numerical simulations used to generate Fig. 1, and simulations investigating the effect of protein shape on the distribution of  $S_{\text{dev}}$  (Fig. S5).

<sup>††</sup>The number of degrees of freedom required to specify an oligomer structure is 6 [(number of subunits) – 1] for asymmetric complexes and 4 for symmetric complexes, so the fraction of randomly formed complexes that are symmetric is exponentially decreasing with exponent [6 (number of subunits) – 10]. Dimers of dimers have 6 degrees of freedom versus the 4 degrees of freedom for cyclic tetramers, which may also contribute to their greater prevalence. The loss of entropy on binding also favors dimers over higher-order oligomers because the entropic cost of forming a higher-order oligomer increases linearly with the number of subunits.

**ACKNOWLEDGMENTS.** We thank Ora Schueler-Furman, John Moult, Rhiju Das, and David Eisenberg for stimulating conversations about the origin of symmetry. This work was supported by a Knut and Alice Wallenberg Foundation postdoctoral fellowship (to I.A.), by Defense Threat Reduction Agency Contract MIPR7K08970172 (to C.E.M.S.), in part by Spanish MEC grant SAB2006-0089 and project FPA2006-05423 (to D.B.K.), by the regional Comunidad de Madrid HEPHACOS project (D.B.K.), and by Department of Energy Grant DE-FG02-00ER41132 (to D.B.K.), and by the National Institutes of Health and Howard Hughes Medical Institute.

1. Crick FH, Watson JD (1956) Structure of small viruses. *Nature* 177(4506):473–475.
2. Goodsell DS, Olson AJ (2000) Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29:105–153.
3. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: A plausible model. *J Mol Biol* 12:88–118.
4. Bennett MJ, Choe S, Eisenberg D (1994) Domain swapping: Entangling alliances between proteins. *Proc Natl Acad Sci USA* 91:3127–3131.
5. Bennett MJ, Schlunegger MP, Eisenberg D (1995) 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci* 4(12):2455–2468.
6. Levy Y, Cho SS, Shen T, Onuchic JN, Wolynes PG (2005) Symmetry and frustration in protein energy landscapes: a near degeneracy resolves the Rop dimer-folding mystery. *Proc Natl Acad Sci USA* 102:2373–2378.
7. Li H, Helling R, Tang C, Wingreen N (1996) Emergence of preferred structures in a simple model of protein folding. *Science* 273(5275):666–669.
8. Wolynes PG (1996) Symmetry and the energy landscapes of biomolecules. *Proc Natl Acad Sci USA* 93:14249–14255.
9. Lukatsky DB, Shakhnovich BE, Mintseris J, Shakhnovich EI (2007) Structural similarity enhances interaction propensity of proteins. *J Mol Biol* 365(5):1596–1606.
10. Lukatsky DB, Zeldovich KB, Shakhnovich EI (2006) Statistically enhanced self-attraction of random patterns. *Phys Rev Lett* 97(17):178101.
11. Gray JJ, et al. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 331(1):281–299.
12. Brady GP, Sharp KA (1997) Entropy in protein folding and in protein-protein interactions. *Curr Opin Struct Biol* 7(2):215–221.
13. Hanson KR (1966) Symmetry of protein oligomers formed by isologous association. *J Mol Biol* 22(2):405–409.
14. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA (2008) Assembly reflects evolution of protein complexes. *Nature* 453(7199):1262–1265.
15. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA (2006) 3D complex: A structural classification of protein complexes. *PLoS Comput Biol* 2(11):e155.
16. Monod J (1969) *Symmetry and Function of Biological Systems at the Macromolecular Level*, eds Engstrom A, Strandberg B. (Wiley, New York), pp 15–29.
17. Das R, Baker D (2008) Macromolecular modeling with rosetta. *Annu Rev Biochem* 77:363–382.