Relating Three-Dimensional Structures to Protein Networks Provides Evolutionary Insights

Our findings reveal that TLR signaling is not required for robust antibody responses to antigen when given in four commonly used adjuvants, in particular FCA, which is widely thought to depend on TLR signaling for its adjuvant effect (*10*). However, ligands that signal through the MyD88 and TRIF pathways can costimulate these responses and affect the antibody class of the response, as has been known for LPS for many years (*3*), and is indicated by the modest boost seen in C57BL/6 mice immunized with Ribi adjuvant (Fig. 2). By exclusion, our data suggest the likelihood that non-TLR–mediated "innate" signals may be involved in the augmentation of adaptive antibody responses. It is also formally possible that MyD88/TRIF–independent modes of TLR signaling are yet to be discovered, possibly through TRAM and TIRAP, although so far there is no evidence for this.

Our results extend and clarify previous studies. A recent paper involving MyD88-deficient mice asserted that T cell–dependent antibody responses require activation of TLRs in B cells (*4*). However, that inference was based on the use of LPS as an adjuvant rather than more commonly used adjuvants such as FCA or alum. Indeed, an earlier study found that although MyD88-deficient mice failed to make an IgG2a response to ovalbumin given in FCA, they still made good IgG1 and IgE antibody responses, and antigen given in alum could promote an IgE response (*11*). However, in those studies, the remaining antibody responses in MyD88-deficient mice could conceivably have involved TLR signaling through TRIF-dependent pathways, a caveat that does not apply to our data. We clearly find that IgG2c (also known as IgG2a^b) responses to protein antigen given in alum or FCA are robust in $Myd88^{-/-}$;$Trif^{Lps2/Lps2}$ mice, in apparent contradiction to both studies of MyD88-deficient mice mentioned (*4, 11*) and other studies implicating a requirement of MyD88 or TLR signaling in the generation of IgG2a antibodies. It may be that TLR ligand–driven suppression of IgG2c/a production occurs in mice with intact TRIF signaling that lack MyD88; for example, through cytokine-driven polarization of the T helper response (*11, 12*). The IgG2c/a response is promoted by IFN-α/β or IFN-γ (*13, 14*). In viral infection, the IgG2c/a response is lost in mice lacking both IFN-α/βR and IFN-γR (*15*). Because we found good IgG2c responses upon immunization of $Myd88^{-/-}$; $Trif^{Lps2/Lps2}$ mice, our results may indicate that these immunizations can stimulate TLR-independent IFN production.

The antibody response to the T cell–independent type II antigen TNP-Ficoll was relatively normal in $Myd88^{-/-}$;$Trif^{Lps2/Lps2}$ mice, although it was slightly lower at later time points. Interestingly, the IgG3 component of the antigen-specific response of $Myd88^{-/-}$;$Trif^{Lps2/Lps2}$ mice was not significantly lower than normal despite their reduced preimmune total serum IgG3 levels. It has been argued that the class switch to IgG3 requires cytokines produced by accessory cells

(*16*). If this is so, it is clear from these results that production of the cytokines in question does not require TLR stimulation of B cells or accessory cells.

That B cells can respond to autologous DNA and RNA through TLR signaling (*17, 18*) raised the possibility that TLRs could affect preimmune development or maintenance of B cell subsets, particularly the marginal zone and B-1 compartments. However, our findings that $Myd88^{-/-}$;$Trif^{Lps2/Lps2}$ mice generate abundant B-1 and marginal zone B cells appear to rule out definitively a required role for TLR signaling in their development.

Our data do not contest the long-held understanding that TLR ligands can augment antibody responses. However, it is surprising, given the recent emphasis on the importance of TLRs in the initiation of the adaptive immune response, that we fail to find a deficit in the early antibody responses of $Myd88^{-/-}$;$Trif^{Lps2/Lps2}$ mice using conventional antigens and immunization regimens. Our data are more consistent with a model in which TLRs play roles in early microbial suppression, regulation of the antibody class, and sustaining antibody secretion at late times after immunization, rather than as an essential component of the self/nonself discrimination of the adaptive immune response. These data have implications for vaccine design because they indicate that robust antibody responses to moderate doses of antigens can be achieved when given in the total absence of TLR ligands. Because TLR-mediated signals can be toxic, our findings raise the possibility that unwanted

side effects of adjuvants may be avoided by excluding TLR ligands from adjuvants.

References and Notes
1. K. Takeda, S. Akira, *Int. Immunol.* **17**, 1 (2005).
2. B. Beutler *et al.*, *Annu. Rev. Immunol.* **24**, 353 (2006).
3. J. Andersson, O. Sjoberg, G. Moller, *Transplant. Rev.* **11**, 131 (1972).
4. C. Pasare, R. Medzhitov, *Nature* **438**, 364 (2005).
5. D. Nemazee, A. Gavin, K. Hoebe, B. Beutler, *Nature* **441**, E4 (2006).
6. K. Hoebe *et al.*, *Nature* **424**, 743 (2003).
7. M. Yamamoto *et al.*, *Science* **301**, 640 (2003).
8. E. Janssen *et al.*, *Immunity* **24**, 787 (2006).
9. Materials and methods are available as supporting material on *Science* Online.
10. C. Pasare, R. Medzhitov, *Curr. Opin. Immunol.* **15**, 677 (2003).
11. M. Schnare *et al.*, *Nat. Immunol.* **2**, 947 (2001).
12. T. Kaisho *et al.*, *Int. Immunol.* **14**, 695 (2002).
13. C. M. Snapper, W. E. Paul, *Science* **236**, 944 (1987).
14. F. D. Finkelman *et al.*, *J. Exp. Med.* **174**, 1179 (1991).
15. M. F. van den Broek, U. Muller, S. Huang, M. Aguet, R. M. Zinkernagel, *J. Virol.* **69**, 4792 (1995).
16. C. M. Snapper *et al.*, *J. Exp. Med.* **175**, 1367 (1992).
17. E. A. Leadbetter *et al.*, *Nature* **416**, 603 (2002).
18. C. M. Lau *et al.*, *J. Exp. Med.* **202**, 1171 (2005).
19. We thank D. Kono and A. Theofilopoulos for comments on the manuscript, A. Rolink for FGK4.5 antibody, and G. Nemerow and N. Sarvetnick for use of instruments. Supported by NIH grants RO1GM44809 to D.N., AI050241 to B.B., and T32AI07606 to B.D.

Relating Three-Dimensional Structures to Protein Networks Provides Evolutionary Insights

Philip M. Kim,[1] Long J. Lu,[1] Yu Xia,[4,5] Mark B. Gerstein[1,2,3]*

Most studies of protein networks operate on a high level of abstraction, neglecting structural and chemical aspects of each interaction. Here, we characterize interactions by using atomic-resolution information from three-dimensional protein structures. We find that some previously recognized relationships between network topology and genomic features (e.g., hubs tending to be essential proteins) are actually more reflective of a structural quantity, the number of distinct binding interfaces. Subdividing hubs with respect to this quantity provides insight into their evolutionary rate and indicates that additional mechanisms of network growth are active in evolution (beyond effective preferential attachment through gene duplication).

Protein interaction networks are principal components of a systems-level description of the cell (*1–4*). Many previous studies have explored global aspects of network topology, clearly linking it to protein function, expression dynamics, and other genomic features (*5–9*). In particular, a protein's degree (number of interaction partners) is an important factor, and proteins with high degree (hubs) have been found

to be essential (*3, 7*). However, most network studies have not considered the structural and chemical aspects of interactions; only recently have there been proposals to use structural information for systems biology (*10*). One specific problem with the current treatment is that protein interaction networks do not differentiate between many types of relationships—e.g., high-affinity and direct versus loose and transient. Sometimes,

in fact, interactions are reported that connect two proteins that never touch each other physically but are only linked through a third protein (*11*, *12*).
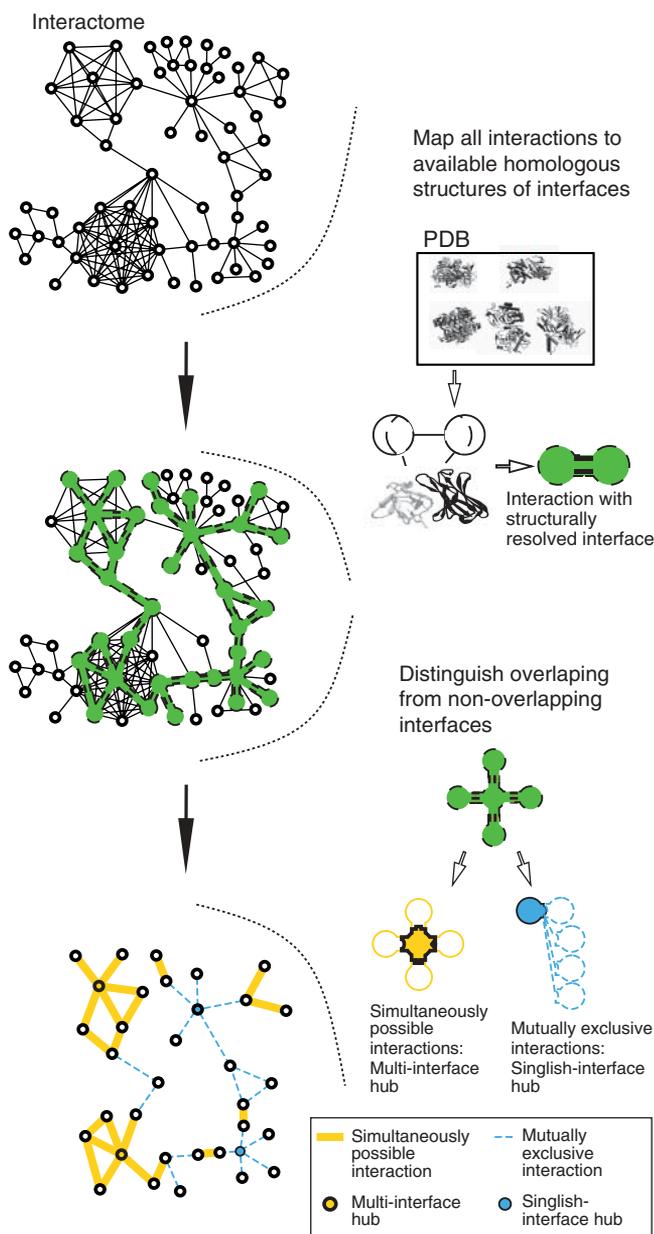
Here we address this problem by combining structural modeling with network analysis. In particular, we compiled a consensus yeast interaction network from various sources (*13*), filtering out low-confidence interactions by using statistical methodologies (*4*). We then annotated

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. [2]Department of Computer Science, Yale University, New Haven, CT 06520, USA. [3]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA. [4]Bioinformatics Program, Boston University, Boston, MA 02215, USA. [5]Department of Chemistry, Boston University, Boston, MA 02215, USA.

*To whom correspondence should be addressed. E-mail: mark.gerstein@yale.edu

many of the edges in this network structurally on the basis of sequence similarity to known complexes (Fig. 1). We used simple three-dimensional (3D)–structural exclusion to distinguish the interfaces of each interaction. Consider two or more proteins interacting with a common partner protein. If they use the same interface on the partner (as known from the structures), the interactions are classified as mutually exclusive. Conversely, if they use different interfaces, the interactions are simultaneously possible (Fig. 1). The network resulting from this analysis (the structural interaction network, SIN) contains 873 nodes (proteins) and 1269 edges (interactions), 438 of which are mutually exclusive (fig. S1). It contains parts of 147 complexes, suggesting that it covers a representative range of interactions. [For the SIN data set and further discussion, see SOM Text, tables S1 and S3, and (*13*)]

After building the SIN, we examined its two different kinds of interactions with respect to the properties of the linked proteins. As shown in Table 1, proteins connected by simultaneously possible interactions are more likely to share the same function than are those connected by mutually exclusive ones [in terms of Gene Ontology (GO) cellular component, molecular function, and biological process designations]; also, they are more likely to be expressed at the same time. Consequently, we expect most of the mutually exclusive interactions to be temporary or transient, because they cannot occur at the same time. Likewise, the simultaneously possible interactions are enriched in permanent associations, connecting members of the same complex (table S4).

Turning to global statistics of the SIN, we find that it has a degree distribution with a notably shorter tail than either the complete yeast interactome or a core, filtered subset of this (fig. S2). In particular, hubs in the SIN have a maximum of 14 interaction partners. This is similar to the number of close-packed neighbors in crystal lattices (12 in hexagonal packing) and reflects the direct, physical constraints on interactions in the SIN. In contrast, in early yeast interactomes some hubs had >200 interaction partners, and even in newer data sets, >30 partners are noted for some proteins (*7*, *14*).

Within hubs in the SIN, we compared those with many physical interfaces (as detected by our approach) to those with a few—to uncouple degree from interface number (which are correlated). We defined hubs by setting an arbitrary cutoff of ≥5 interaction partners; variations in this cutoff did not affect our results (table S5). We detected differences in the properties between multi- and single-interface hubs. However, more statistically significant differences were evident if we distinguished between hubs with one or two interfaces (singlish-interface) and those with more

Fig. 1. The creation of the structural interaction network (SIN) data set. All interactions from the filtered protein interaction data set are mapped to Pfam domains (*30*). The Pfam domains are mapped to known structures of protein interactions by means of iPfam (*31*). Only those interactions in which both interaction partners (or a homologous domain of either) can be found in a 3D structure of a protein complex are kept. All interactions are then classified into mutually exclusive and simultaneously possible by 3D structural exclusion. When a protein has more than one simultaneously possible interaction, the number of interaction interfaces is counted.

Table 1. Differences of simultaneously possible versus mutually exclusive interactions with respect to GO annotations (shared by the interacting proteins) and coexpression correlation coefficients. GO biological process, molecular function, and cellular component are taken from SGD lite (*26*) and coexpression correlation from the compendium expression data set (*27*). All differences are significant, with $P \ll 0.01$.

	Simultaneously possible interactions	Mutually exclusive interactions
Fraction with same biological process	14%	24%
Fraction with same molecular function	18%	33%
Fraction with same cellular component	12%	27%
Coexpression correlation	0.17	0.23

than two interfaces (multi-interface). First, we examined the essentiality of both kinds of hubs. Although hubs in general are more likely than other proteins to be essential for cellular viability (3), as shown in Table 2, multi-interface hubs are twice as likely to be essential as singlish-interface ones, which, in turn, are no more likely to be essential than the average protein in the SIN. This result suggests the notion of hubs having a higher essentiality due to their network centrality is somewhat incomplete: It is the number of interaction interfaces that leads to higher essentiality.

Furthermore, Table 2 shows that multi-interface hubs are more likely to be coexpressed with their neighbors than are singlish-interface ones. This provides a straightforward structural explanation for the existence of two types of expression dynamics for hubs (10), date and party hubs (7). In particular, singlish-interface hubs seem to correspond to date hubs (which are expressed at

different times than their interaction partners; table S7), and multi-interface hubs correspond to party hubs (which are expressed at the same times as their interaction partners). It is quite reasonable that the interaction partners of singlish-interface hubs are not coexpressed, because they would compete for the same binding interface. On the other hand, for the partners of multi-interface hubs, it makes sense to be expressed simultaneously, because they bind to different interfaces. Multi-interface hubs, in fact, correspond to central members of protein complexes, as is evident from cross-referencing them with known complexes (table S8). A representative multi-interface hub, for example, is Arp2p, a member of the Arp2/3 complex. Conversely, a good example for a singlish-interface hub is Snf1p, a central protein kinase (see SOM Text and table S6).

There has been some controversy over whether hubs are slower-evolving than other proteins (15–18). A commonly used measure of evolutionary rate is the dN/dS ratio (the ratio of non-synonymous to synonymous substitutions, also referred to as K_a/K_s ratio). Table 2 shows that it is significantly lower for multi-interface hubs than for the average SIN protein, but not so for singlish-interface hubs. Although the dependence of evolutionary rate on protein degree has been attributed to an underlying effect of expression level (18), we find that the relationship of evolutionary rate to the number of interfaces is independent of expression level (whereas that to the degree is not) (fig. S4). The aforementioned controversy may have arisen because previous studies did not differentiate between singlish and multi-interface hubs. A larger number of interfaces may give rise to a lower evolutionary rate because a larger fraction of residues participate in interactions. Indeed, Fig. 2 shows that the variation in a protein's evolutionary rate can be accounted for better by changes in the fraction of its accessible surface area involved in interactions than by degree. This result can be explained simply from a structural point of view: The average mutational rate for exposed surface residues is more than twice as high as for those at an interface, which, in turn, is slightly higher than the one for buried residues (table S9) (19). Thus, as suggested previously (20) and shown in our analysis, the proportion of a protein's available surface area involved in interactions should correlate inversely with evolutionary rate.

Finally, we examined network evolution from a structural perspective. The existing scale-free network topology (the dominance of hubs) may have evolved through preferential attachment (21). Gene duplication is one possible cause of such an evolutionary process (22, 23), but other factors, such as preferential rewiring, could contribute as well (23, 24). As depicted in Fig. 3, if a hub evolves by duplication, its interaction partners are expected to be enriched in paralogs (products of homologous genes originating from within-genome duplications). As expected (25), we found that two proteins are significantly more likely to be
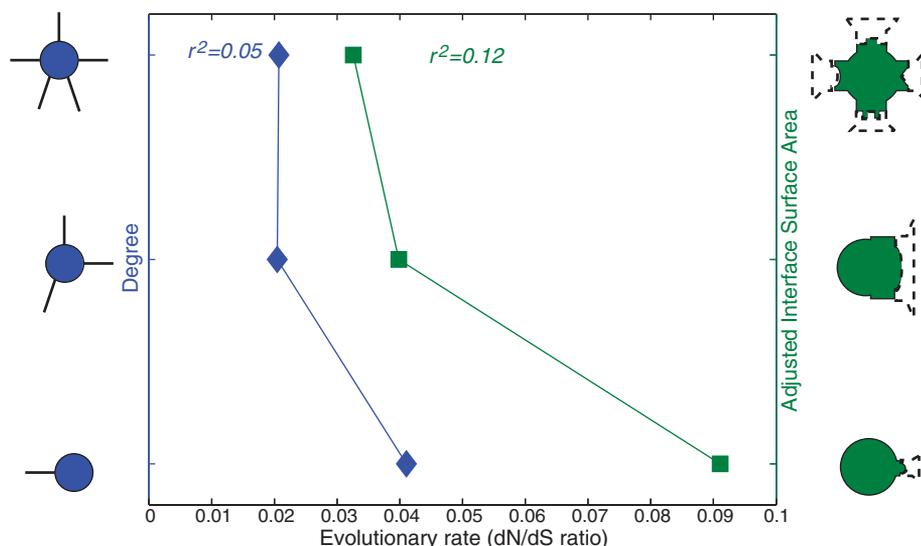

Fig. 2. Dependence of the average evolutionary rate (dN/dS ratio) of a protein with the degree and the interacting accessible surface area (adjusted by protein size, as estimated from molecular weight). For the degree correlation coefficient, we get $r^2 = 0.05$, and for the adjusted interface surface area, $r^2 = 0.12$, suggesting that more than twice as much of the variation in dN/dS is accounted for by adjusted interface surface area (12%) than by the degree (5%).

Table 2. Correlation of genomic features with singlish and multi-interface hubs. The fraction of proteins that are products of essential genes (28), the average expression correlation with their neighbors (27), and the evolutionary rate [dN/dS ratio, from (29)] was calculated for the entire proteome, the entire SIN, singlish-interface protein hubs, and multi-interface protein hubs. The P-values of the differences between the whole data set and the singlish interface hubs (all-singlish) and the singlish and multi-interface hubs (singlish-multi) were calculated with the Wilcoxon rank-sum test (see Methods in the SOM).

	Entire proteome	All in data set	P-value (all-singlish)	Singlish-interface hubs only	P-value (singlish-multi)	Multi-interface hubs only
Protein essentiality	18.6%	32.3%	0.9	31.8%	<0.01	64.9%
Expression correlation		0.20	0.3	0.17	<0.05	0.25
Evolutionary rate	0.077	0.047	0.5	0.051	<0.01	0.029

Table 3. Fraction of protein pairs that are paralogs of each other. Random pair: randomly chosen protein pair from our data set (average); Same partner: fraction of pairs with the same interaction partner that are paralogs; Same partner, same interface: Fraction of pairs that bind to the same interface that are paralogs; Same partner, different interface: fraction of pairs with the same interaction partner, but different interacting interface that are paralogs (calculated from the platinum standard set only; see Methods in the SOM).

	Random pair	Same partner	Same partner, same interface	Same partner, different interface
Fraction paralogs	0.23%	4.10%	8.10%	0.00%
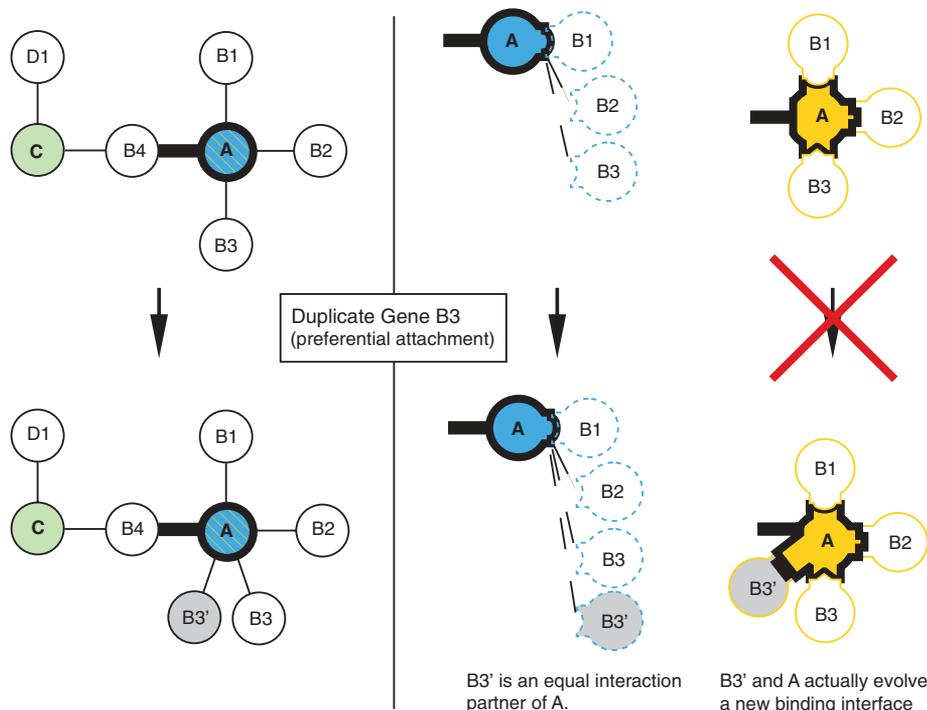
Fig. 3. The concept of network evolution by gene duplication. A given protein may acquire a new interaction by duplication of an existing one. Given equal likelihood of any gene to be duplicated, a protein with many partners is more likely to get a new partner than one with few—hence, there is effective preferential attachment. For singlish-interface hubs, this mechanism is straightforward. However, for multi-interface hubs, it would then require coevolution of the hub and the duplicated gene to form a new interface.

paralogs if they share a common partner (Table 3). However, this is true only if they also share an interaction interface. We did not find enrichment for paralogs among interaction partners binding to different interfaces (Table 3). That is, our analysis is consistent with the evolution of singlish-interface hubs through duplication-mutation, whereas it does not support such an evolution of multi-interface hubs (Fig. 3). Because multi-interface hubs are often parts of larger protein complexes, it appears that protein-complex evolution could follow a different mechanism.

From our 3D structural analysis of protein interaction networks, we find that we can distinguish two fundamentally different types of network edges. On the one hand, we find a group of interactions that are simultaneously possible and a set of multi-interface hubs associated with these. Multi-interface hubs correspond, in many respects, to our "classic" notion of network hubs. They are more likely to be essential and more conserved. They are most likely members of large and stable complexes. However, they do not follow canonical models of network evolution, growing through gene duplication. On the other hand, we find a second class of interactions, mutually exclusive ones, which have a transient character and occur in singlish-interface hubs. Singlish-interface hubs are distinctly "nonclassical": They are neither likely to be essential, nor conserved. However, in respect to network growth they do follow the canonical preferential gene duplication model.

References and Notes

1. A. L. Barabasi, Z. N. Oltvai, *Nat. Rev. Genet.* **5**, 101 (2004).
2. A. C. Gavin *et al.*, *Nature* **440**, 631 (2006).
3. H. Jeong, S. P. Mason, A. L. Barabasi, Z. N. Oltvai, *Nature* **411**, 41 (2001).
4. R. Jansen *et al.*, *Science* **302**, 449 (2003).
5. U. de Lichtenberg, L. J. Jensen, S. Brunak, P. Bork, *Science* **307**, 724 (2005).
6. R. Kelley, T. Ideker, *Nat. Biotechnol.* **23**, 561 (2005).
7. J. D. Han *et al.*, *Nature* **430**, 88 (2004).
8. I. Lee, S. V. Date, A. T. Adai, E. M. Marcotte, *Science* **306**, 1555 (2004).
9. P. M. Kim, thesis (MIT, Cambridge, MA, 2003), pp. 1–125.
10. P. Aloy, R. B. Russell, *Nat. Rev. Mol. Cell Biol.* **7**, 188 (2006).
11. P. Aloy, R. B. Russell, *Nat. Biotechnol.* **22**, 1317 (2004).
12. G. D. Bader, C. W. Hogue, *Nat. Biotechnol.* **20**, 991 (2002).
13. See supporting information on *Science* Online for details. For updates to the SIN and interactive viewing using tYNA, see http://sin.gersteinlab.org.
14. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
15. H. B. Fraser, A. E. Hirsh, L. M. Steinmetz, C. Scharfe, M. W. Feldman, *Science* **296**, 750 (2002).
16. I. K. Jordan, Y. I. Wolf, E. V. Koonin, *BMC Evol. Biol.* **3**, 1 (2003).
17. S. Wuchty, *Genome Res.* **14**, 1310 (2004).
18. J. D. Bloom, C. Adami, *BMC Evol. Biol.* **3**, 21 (2003).
19. W. S. Valdar, J. M. Thornton, *Proteins* **42**, 108 (2001).
20. H. B. Fraser, A. E. Hirsh, *BMC Evol. Biol.* **4**, 13 (2004).
21. R. Albert, A. L. Barabasi, *Rev. Mod. Phys.* **74**, 47 (2002).
22. A. Vazquez, *Phys. Rev. E* **67**, 056103 (2003).
23. A. Wagner, *Proc. Biol. Sci.* **270**, 457 (2003).
24. E. Eisenberg, E. Y. Levanon, *Phys. Rev. Lett.* **91**, 138701 (2003).
25. J. Espadaler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7151 (2005).
26. L. Issel-Tarver *et al.*, *Methods Enzymol.* **350**, 329 (2002).
27. T. R. Hughes *et al.*, *Cell* **102**, 109 (2000).
28. G. Giaever *et al.*, *Nature* **418**, 387 (2002).
29. D. P. Wall *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5483 (2005).
30. A. Bateman *et al.*, *Nucleic Acids Res.* **30**, 276 (2002).
31. R. D. Finn, M. Marshall, A. Bateman, *Bioinformatics* **21**, 410 (2005).
32. We thank J. Korbel, T. Gianoulis, A. Paccanaro, and H. Yu for helpful comments on the manuscript. We also thank the anonymous referees for valuable suggestions. This work was supported by the NIH. Y.X. was supported by a fellowship from the Jane Coffin Childs Memorial Fund for Medical Research.

Characterizing a Mammalian Circannual Pacemaker

Gerald A. Lincoln,[1]* Iain J. Clarke,[2] Roelof A. Hut,[3] David G. Hazlerigg[4]

Many species express endogenous cycles in physiology and behavior that allow anticipation of the seasons. The anatomical and cellular bases of these circannual rhythms have not been defined. Here, we provide strong evidence using an in vivo Soay sheep model that the circannual regulation of prolactin secretion, and its associated biology, derive from a pituitary-based timing mechanism. Circannual rhythm generation is seen as the product of the interaction between melatonin-regulated timer cells and adjacent prolactin-secreting cells, which together function as an intrapituitary "pacemaker-slave" timer system. These new insights open the way for a molecular analysis of long-term timing mechanisms.

Endogenous circannual rhythms drive many long-term cycles in physiology and behavior in long-lived vertebrates (*1*, *2*) including reproduction (*3*), hibernation (*4*, *5*), migration (*6*), and pelage growth (*7*), but the anatomical and cellular bases of such rhythm generation remain a mystery. We investigated whether a circannual rhythm may be generated through a pituitary mechanism, itself dependent on the circadian system. We focused on the anterior pituitary control of prolactin secretion, but similar cell-cell interactions in the brain may