



**Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans**  
Katia Koelle *et al.*  
*Science* **314**, 1898 (2006);  
DOI: 10.1126/science.1132745

*This copy is for your personal, non-commercial use only.*

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of July 11, 2014 ):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/314/5807/1898.full.html>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/content/suppl/2006/12/19/314.5807.1898.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/314/5807/1898.full.html#related>

This article **cites 45 articles**, 20 of which can be accessed free:

<http://www.sciencemag.org/content/314/5807/1898.full.html#ref-list-1>

This article has been **cited by** 60 article(s) on the ISI Web of Science

This article has been **cited by** 71 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/314/5807/1898.full.html#related-urls>

This article appears in the following **subject collections**:

Epidemiology

<http://www.sciencemag.org/cgi/collection/epidemiology>

4. K. G. Miller, J. D. Wright, R. G. Fairbanks, *J. Geophys. Res. Sol. Earth Planets* **96**, 6829 (1991).
5. B. S. Wade, H. Pälike, *Paleoceanography* **19**, PA4019 (2004).
6. J. C. Zachos, N. J. Shackleton, J. S. Revenaugh, H. Pälike, B. P. Flower, *Science* **292**, 274 (2001).
7. Our detailed Oligocene Pacific data set is available in electronic form at a designated data repository ([www.pangaea.de](http://www.pangaea.de)). Accession codes are <http://doi.pangaea.de/10.1594/PANGAEA.547942>, <http://doi.pangaea.de/10.1594/PANGAEA.547798>, and <http://doi.pangaea.de/10.1594/PANGAEA.547800>.
8. Data were generated in five laboratories. A lower-resolution record across the entire interval (10) shows that there are no discernible interlaboratory offsets.
9. Methods are available as supporting material on Science Online.
10. C. H. Lear, Y. Rosenthal, H. K. Coxall, P. A. Wilson, *Paleoceanography* **19**, PA4015 (2004).
11. A. Tripati, H. Elderfield, L. Booth, J. Zachos, P. Ferretti, *Sci. Res., Proc. Ocean Drill. Prog.* (Ocean Drilling Program, College Station, TX, 2006), vol. 199.
12. M. Lyle *et al.*, *Init. Rep., Proc. Ocean Drill. Prog.* (Ocean Drilling Program, College Station, TX, 2002), vol. 199.
13. Measured  $\text{CaCO}_3$  content was used to calibrate calculated  $\text{CaCO}_3$  from physical property measurements (12).
14. H. Pälike *et al.*, *Sci. Res., Proc. Ocean Drill. Prog.* (Ocean Drilling Program, College Station, TX, 2005), vol. 199.
15. L. Lanci, J. M. Parés, J. E. T. Channell, D. V. Kent, *Earth Planet. Sci. Lett.* **237**, 617 (2005).
16. J. Laskar *et al.*, *Astron. Astrophys.* **428**, 261 (2004).
17. H. Pälike, J. Laskar, N. J. Shackleton, *Geology* **32**, 929 (2004).
18. J. Laskar, F. Joutel, F. Boudin, *Astron. Astrophys.* **270**, 522 (1993).
19. H. Pälike, N. J. Shackleton, *Earth Planet. Sci. Lett.* **182**, 1 (2000).
20. I. Raffi, *Phil. Trans. R. Soc. London Ser. A* **357**, 1975 (1999).
21. L. Lanci, J. M. Parés, J. E. T. Channell, D. V. Kent, *Earth Planet. Sci. Lett.* **226**, 207 (2004).
22. K. Billups, H. Pälike, J. E. T. Channell, J. C. Zachos, N. J. Shackleton, *Earth Planet. Sci. Lett.* **224**, 33 (2004).
23. N. J. Shackleton, M. A. Hall, I. Raffi, L. Tauxe, J. Zachos, *Geology* **28**, 447 (2000).
24. H. Pälike, J. Frazier, J. C. Zachos, *Quat. Sci. Rev.*, in press; available online at <http://dx.doi.org/10.1016/j.quascirev.2006.02.001>.
25. J. P. Kennett, N. J. Shackleton, *Nature* **260**, 513 (1976).
26. J. C. Zachos, T. M. Quinn, K. A. Salamy, *Paleoceanography* **11**, 251 (1996).
27. H. A. Paul, J. C. Zachos, B. P. Flower, A. Tripati, *Paleoceanography* **15**, 471 (2000).
28. The cycle naming scheme (5) incorporates the 405-ky cycle number, the geological epoch, and the nearest magnetic polarity chron.
29. S. Van Simaey, H. Brinkhuis, J. Pross, G. L. Williams, J. C. Zachos, *Geology* **33**, 709 (2005).
30. S. F. Pekar, N. Christie-Blick, M. A. Kominz, K. G. Miller, *Geology* **30**, 903 (2002).
31. W. S. Broecker, T.-H. Peng, *Tracers in the Sea* (LDGO Press, Palisades, NY, 1982).
32. T. H. van Andel, *Earth Planet. Sci. Lett.* **26**, 187 (1975).
33. R. M. DeConto, D. Pollard, *Nature* **421**, 245 (2003).
34. M. F. Loutre, D. Paillard, F. Vimeux, E. Cortijo, *Earth Planet. Sci. Lett.* **221**, 1 (2004).
35. J. C. Zachos, L. R. Kump, *Global Planet. Change* **47**, 51 (2005).
36. J. C. G. Walker, J. F. Kasting, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **97**, 151 (1992).
37. A description of the modified model equations can be found in (9).
38. B. N. Opdyke, J. C. G. Walker, *Geology* **20**, 733 (1992).
39. D. Paillard, L. Labeyrie, P. Yiou, *EOS Transactions AGU* **77**, 379 (1996).
40. J. Imbrie *et al.*, *Milankovitch and Climate, Part 1*, A. L. Berger, *et al.*, Eds. (Reidel, 1984), pp. 269–305.
41. The central bandpass frequency was  $2.5 \text{ My}^{-1}$ , with a bandwidth of  $\pm 0.5 \text{ My}^{-1}$ . Filtering was performed using (39).
42. N. J. Shackleton, N. D. Opdyke, *Quat. Res.* **3**, 39 (1973).
43. M. Ghil *et al.*, *Rev. Geophys.* **40**, 3.1 (2002).
44. This research used samples and data provided by the Ocean Drilling Program (ODP). ODP is sponsored by the U.S. National Science Foundation (NSF) and participating countries under the management of Joint Oceanographic Institutions (JOI), Inc. We thank the crew and shipboard scientists of ODP Leg 199, particularly I. Raffi and J. Backman for their biostratigraphy work. We thank L. Kump, J. Zachos, J. Walker, and J. Kasting for providing code and information for their models.

# Supporting Online Material

[www.sciencemag.org/cgi/content/full/314/5807/1894/DC1](http://www.sciencemag.org/cgi/content/full/314/5807/1894/DC1)

Methods

SOM Text

Figs. S1 to S3

Tables S4 to S6

References

14 August 2006; accepted 6 November 2006

10.1126/science.1133822

# Epochal Evolution Shapes the Phylogenetics of Interpandemic Influenza A (H3N2) in Humans

Katia Koelle,<sup>1,2\*</sup> Sarah Cobey,<sup>1†</sup> Bryan Grenfell,<sup>2,3</sup> Mercedes Pascual<sup>1</sup>

Human influenza A (subtype H3N2) is characterized genetically by the limited standing diversity of its hemagglutinin and antigenically by clusters that emerge and replace each other within 2 to 8 years. By introducing an epidemiological model that allows for differences between the genetic and antigenic properties of the virus's hemagglutinin, we show that these patterns can arise from cluster-specific immunity alone. Central to the formulation is a genotype-to-phenotype mapping, based on neutral networks, with antigenic phenotypes, not genotypes, determining the degree of strain cross-immunity. The model parsimoniously explains well-known, as well as previously unremarked, features of interpandemic influenza dynamics and evolution. It captures the observed boom-and-bust pattern of viral evolution, with periods of antigenic stasis during which genetic diversity grows, and with episodic contraction of this diversity during cluster transitions.

**I**nterpandemic influenza causes substantial morbidity and mortality in humans. Annual winter epidemics yield cumulative attack

rates between 10 and 20% for influenza A (subtypes H3N2 and H1N1), and influenza B infections (1) and contribute heavily to deaths caused by respiratory infections worldwide. The virus is capable of evading immune recognition through continual antigenic drift of its surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA), complicating long-term control of the disease through vaccination (2). An understanding of the ecological and immunological processes driving influenza dynamics and evolution is therefore critical for anticipating and ultimately mitigating the effect of this infectious disease. Here we focus on the

phylogenetics (3) of H3N2, which has been present worldwide since its pandemic appearance in 1968.

One of the most striking characteristics of influenza A evolution is the limited standing diversity of the HA gene, despite the virus's high mutation rate. This limited diversity is evident in its phylogeny: The tree consists of a long trunk with short side branches that are indicative of high extinction rates of the lineages (4) (Fig. 1A). Epidemiological factors that contribute to this pattern include the short infectious period of the host and partial cross-immunity between similar strains (3). However, the most detailed model of interpandemic influenza to date suggests that these factors alone cannot account for limited diversity and that temporary strain-transcendent (generalized) immunity is necessary to restrict diversity (5).

More recently, differences between the genetic and antigenic evolution of the H3N2 virus's HA glycoprotein have been highlighted (6). A key unexplained pattern is that, although genetic change is gradual, antigenic change is punctuated. HA inhibition (HI) assays show that H3N2 sequences fall into groups, or clusters, with unique antigenic properties. Between 1968 and 2003, these clusters emerged and replaced each other within 2 to 8 years, exerting a major influence on vaccine strategy (2). Empirical evidence suggests that there is almost complete immunity between strains within a cluster (7). In contrast, cross-immunity is as low as 60 to 85% between clusters adjacent in time (7, 8) and is undetectable between temporally distant clus-

<sup>1</sup>Department of Ecology and Evolutionary Biology, 2019 Kraus Natural Science Building, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109-1048, USA.

<sup>2</sup>Center for Infectious Disease Dynamics (CIDD), Department of Biology, 208 Mueller Lab, Eberly College of Science, The Pennsylvania State University (PSU), University Park, PA 16802, USA. <sup>3</sup>Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA.

\*To whom correspondence should be addressed. E-mail: [kkoelle@psu.edu](mailto:kkoelle@psu.edu)

†These authors contributed equally to this work.

ters (9). A time series of influenza-related deaths highlights the importance of clusters and the timing of their transitions: Anomalously high mortality rates accompany cluster transitions (10, 11) (Fig. 1B).

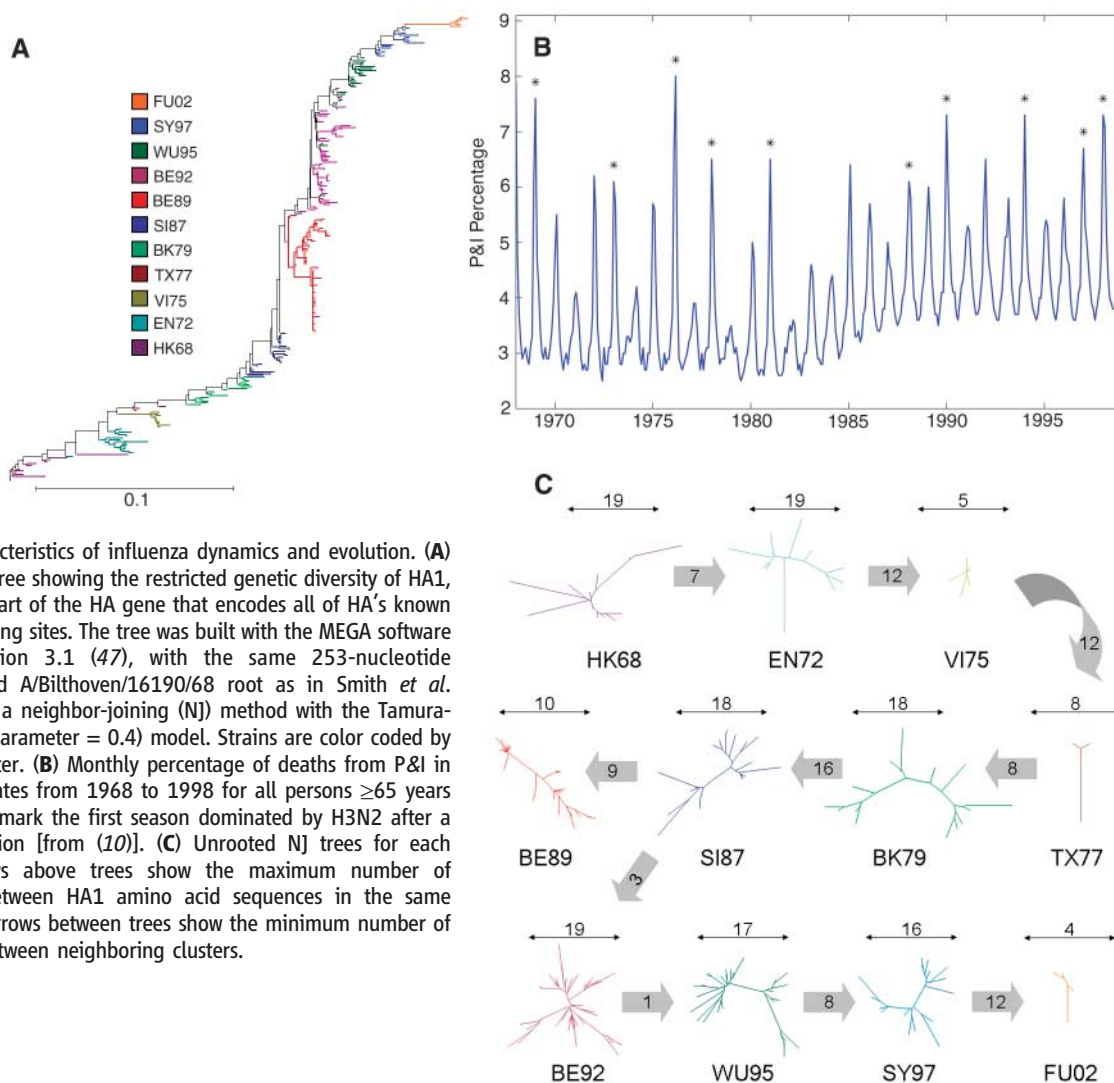
Here we present a model that offers an explanation for both limited HA diversity and punctuated antigenic changes. It is a phylogenetic model, in that it simulates the interaction of epidemiological dynamics and pathogen evolution using a single underlying framework (3). Unlike previous models that focus on individual strains [e.g. (5, 12)], our model emphasizes the concept of clusters. We first looked at the genetic sequences analyzed by Smith and coauthors (6) by cluster designation (Fig. 1C). Within a cluster, the HA sequences can differ by a substantial number of amino acids while retaining their antigenic similarity. For example, two sequences in the HK68 cluster differ by 19 amino acids. Furthermore, transitions between clusters, and therefore large changes in antigenic properties, can result from as few as one amino acid change (as in the case of the SI87 to BE89 and the BE92 to WU95 cluster transitions).

Traditional multistrain disease models (including those for influenza) directly relate the degree of cross-immunity between strains to the distances between their sequences. Distance metrics often employ the Hamming distance between two strain sequences, and they are used in both bit-string models (with alleles 0 or 1 at each locus) (13, 14) and in more realistic models of amino acid evolution (5). By design, distance metrics set the degree of cross-immunity high between strains with high sequence similarity and low between strains with low sequence similarity. However, these distance metrics are incongruent with observations of cross-immunity between influenza clusters (Fig. 1C). They cannot provide a framework in which a single amino acid change can markedly release a strain from population-level host immunity and in which 19 changes can have little antigenic effect.

Although it may be argued that amino acid changes that precipitate cluster jumps occur at key influential sites, a closer investigation of the substitutions associated with cluster transitions suggests that few sites fit this model. Of the 43 sites associated with cluster transitions, 35 also

exhibited some degree of neutral polymorphism (6). (We categorize a site as carrying some degree of neutral polymorphism if the site exhibits any variation in amino acid use within at least one cluster.) In the remaining eight sites, substitutions were always accompanied by cluster transitions (table S1); however, these sites could not account for every transition. Four transitions were associated only with sites capable of showing neutral polymorphism. Site-directed mutagenesis of strains belonging to different clusters also showed that identical substitutions can have different effects on hemadsorption, depending on differences in local structure (15). Thus, a model of invariably influential and neutral sites is unlikely to generate a realistic topology.

**Neutral networks map genetic-to-antigenic change.** We therefore take an alternative approach for modeling cross-immunity—one in which phenotype is determined by a context-dependent interaction of amino acids. Every amino acid site in an epitope is potentially important. Whether an amino acid replacement in a site precipitates a cluster jump is determined



**Fig. 1.** Characteristics of influenza dynamics and evolution. **(A)** Phylogenetic tree showing the restricted genetic diversity of HA1, which is the part of the HA gene that encodes all of HA's known antibody binding sites. The tree was built with the MEGA software package version 3.1 (47), with the same 253-nucleotide sequences and A/Bilthoven/16190/68 root as in Smith *et al.* (2004), using a neighbor-joining (NJ) method with the Tamura-Nei (gamma parameter = 0.4) model. Strains are color coded by antigenic cluster. **(B)** Monthly percentage of deaths from P&I in the United States from 1968 to 1998 for all persons  $\geq 65$  years old. Asterisks mark the first season dominated by H3N2 after a cluster transition [from (10)]. **(C)** Unrooted NJ trees for each cluster. Arrows above trees show the maximum number of differences between HA1 amino acid sequences in the same cluster, and arrows between trees show the minimum number of differences between neighboring clusters.

not only by the change at that site, but also by the genetic background in which the change is made. Epitope structure is thus mediated through high-level interactions between the amino acids of a HA sequence. Extensive theoretical and empirical studies have focused on understanding protein genotype-to-phenotype mapping in light of these interactions. These studies have shown that sequence space is inhabited by “neutral networks” (16–20). Neutral networks for proteins are defined as sets of amino acid sequences (of length  $L$ ) that are connected by one-mutation neighbors and map into the same conformation. The number of sequences folding into the same conformation varies, with the size distribution of the conformational sets consisting of many rare shapes and only a few highly designable, or frequent, shapes (21). The presence of an underlying neutral network topology affects evolutionary dynamics on the phenotype landscape: In simulations where an optimal shape (phenotype) is defined exogenously, an initial phenotype will evolve toward this target shape in punctuated steps (22, 23). During the periods of phenotypic stasis separating the punctuated shape changes, sequences diffuse through genotype space along neutral or almost neutral networks (24, 25). This diffusion is critical for gaining access to adjacent networks that enable more dramatic phenotypic change (22, 26). The stepwise emergence of these phenotypic innovations, guided by the process of neutral diffusion, has been termed epochal evolution (27, 28).

We can now interpret influenza clusters in terms of neutral networks (Fig. 1C). Within each cluster, strains have similar conformations of their HA epitopes; host antibodies are thus more likely to recognize these strains as antigenically equivalent, and strain cross-immunity is close to complete. These strains can therefore be classified as belonging to a set of neutral networks with similar antigenic phenotypes. Sharply reduced cross-immunity between influenza clusters arises from antigenic escape by the HA protein and is precipitated by amino acid substitutions that substantially change the structure of one or more epitopes.

**Dynamical consequences of the genetic-antigenic map.** To determine the effect of this genotype topology on the dynamics and evolution of influenza, we modeled the disease by coupling an epidemiological transmission model to a genotype-phenotype (GP) model that implements neutral networks. The GP model is used to map strains, or genotypes, into antigenic phenotypes. To this end, we extended a simple model that allows for a tunable degree of neutrality (29, 30). It is a generalization of the NK fitness-landscape model, which allows epistatic interactions between loci to affect fitness (31).

We modeled influenza's HA as five distinct epitopes, with each epitope represented by a separate GP map (32). At a given epitope, amino acids interact with a small number of their neighbors [ $K = 1$ , where  $K$  is the degree of

epistasis or context dependency] to determine the epitope's overall shape. This level of context dependency captures the size distribution of the neutral networks seen in lattice models of protein folding (21) (Fig. 2A). Our genetic-antigenic map for HA assumes statistical properties of cross-immunity between strains that are one amino acid apart (Fig. 2B). Specifically, point mutations resulting in an amino acid replacement in an epitope belong to one of three types. Mutations are neutral (100% cross-immunity) if the epitope in which the replacement occurs does not change its conformation. This occurs when the two sequences belong to the same neutral network at the mutated epitope. When the mutated sequence belongs to another neutral network, two different cases are considered: Mutations are almost neutral (93% cross-immunity) if the epitope changes its conformation only slightly, whereas they are classified as escape mutations if the epitope changes its structure significantly (80% cross-immunity). These escape mutations precipitate cluster transitions and occur relatively infrequently (32) (Fig. 2).

Given this underlying genotype space topology, we simulated the dynamics of influenza and the concurrent evolution of its HA in a temperate-latitude population. We based our influenza transmission model on a multistrain model that

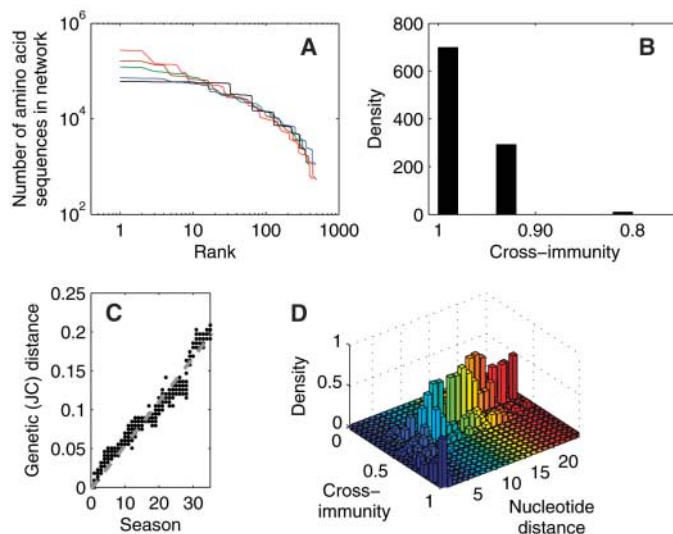
uses a status-based approach for tracking classes of hosts (12). However, instead of modeling the number of individuals susceptible, infected, and recovered to each individual strain, we modeled the number of individuals susceptible, infected, and recovered to each neutral ensemble. We define a neutral ensemble  $i$  as the set of strains that have identically shaped epitopes; i.e., that reside on the same neutral network at each epitope. This compartmental formulation eliminates the need to track the infection histories of individual hosts and facilitates future mathematical analyses of the dynamics. Without mutation, the dynamics of a system with  $n$  neutral ensembles are given by Eqs. 1a and 1b:

$$\frac{dI_i}{dt} = [\beta(t)I_i + mp_i] \frac{S_i}{N} - (v + \mu_D)I_i \quad (1a)$$

$$\frac{dS_i}{dt} = \mu_B N - \sum_{j=1}^n [\beta(t)I_j + mp_j] \sigma_{ij} \frac{S_i}{N} - \mu_D S_i \quad (1b)$$

where the subscript  $i$  denotes the neutral ensemble, and  $\sigma_{ij}$  denotes the degree of cross-immunity between neutral ensembles  $i$  and  $j$ . As in standard epidemiological models,  $\beta(t)$  is the seasonally varying transmission rate,  $v$  is the recovery rate,

**Fig. 2.** Statistical properties of the model's neutral networks and patterns of genetic and antigenic change under the neutral-network framework. (A) Distribution of neutral-network sizes for each of the five modeled HA epitopes. Epitopes one to five each contain approximately  $1.5^L = 438$  neutral networks, or distinct conformations. For each epitope, the networks were generated with a neutralized NK fitness-landscape model (30), with  $L = 15$ ,  $K = 1$ , alphabet size ( $A$ ) = 2, and neutrality parameter ( $F$ ) = 2 (see supporting online material text). (B) Cross-immunity  $\sigma_{ij}$  between strains  $i$  and  $j$  that differ by a single amino acid. We let a neutral mutation of an epitope  $\epsilon$  result in  $x_{ij}(\epsilon) = 100\%$  recognition, an almost-neutral mutation result in  $x_{ij}(\epsilon) = 65\%$  recognition, and an escape mutation result in  $x_{ij}(\epsilon) = 0\%$  recognition. The degree of cross-immunity between two strains is assumed to depend additively on recognition at each epitope, such that  $\sigma_{ij} = \frac{1}{5} \sum_{\epsilon=1}^5 x_{ij}(\epsilon)$ . In the case of a non-neutral mutation, we let there be a low chance [ $p = 1/40$  (1 out of 40)] that the adjacent neutral networks have conformations that are antigenically very different, yielding an escape mutant that may precipitate a cluster transition. The remainder of the time, a non-neutral mutation is considered almost neutral, yielding an antigenically similar conformation. With cross-immunity additive across epitopes, two sequences that are one mutation apart have either full cross-immunity, 93% cross-immunity, or 80% cross-immunity. The distribution was generated by computing the degree of cross-immunity between two strains of Hamming distance that are one amino acid apart for 1000 randomly sampled strain pairs. (C) Genetic [Jukes-Cantor (JC)] distance from the founder strain to the annual strain samples over the time course of the simulation. (D) Frequency distribution of the degree of cross-immunity between the simulation-generated HA strains shown in Fig. 3D.





$\mu_B$  is the birth rate, and  $\mu_D$  is the death rate. Because clusters consist of neutral ensembles with high levels of cross-immunity, the model formulation yields dynamics that are essentially susceptible-infected-recovered (SIR) within single clusters. The model without immigration results in the stochastic extinction of influenza during the summer troughs; we therefore introduced the immigration rate parameter  $m$  to allow an influx of strains from areas outside the region. We let  $p_i$  represent the proportion of strains in the outside region that belong to neutral ensemble  $i$ . For simplicity, rather than modeling the outside region dynamically, we assumed that the strains in this region are derived from the strains present in the temperate region during the previous year (32). A more complex model that incorporates the region dynamically produces similar results as long as there is sufficient asynchrony between the geographic patches and provided that no genetic bottlenecks are otherwise produced.

The resulting disease dynamics reproduce influenza's well-known annual outbreaks in temperate regions, with seasonal attack rates of 1 to 13% (Fig. 3, A and B). The model also captures the sequential replacement of influenza clusters that is seen empirically, with cluster-transition years having higher incidence levels (Fig. 3A) and correspondingly higher annual attack rates (Fig. 3B), echoing the pneumonia and influenza (P&I) mortality rates that are representative of influenza's time series (10)

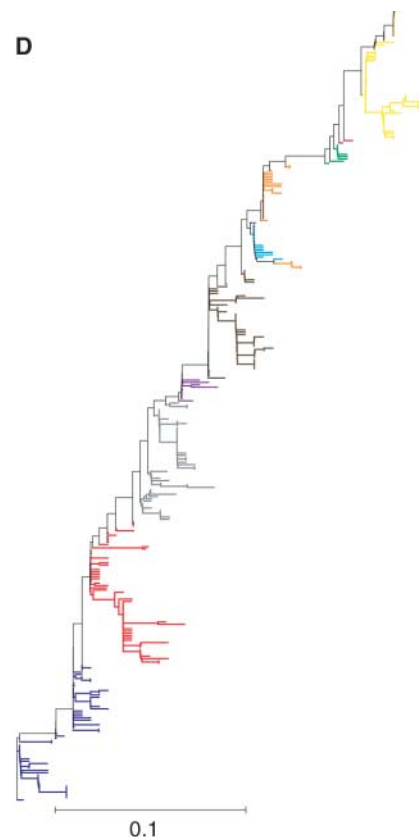
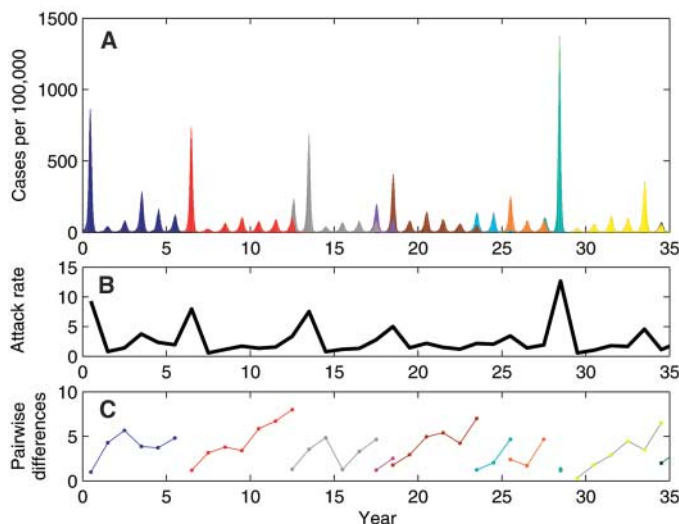
(Fig. 1B). The temporal patterns of genetic and antigenic change can now also be visualized (Fig. 2, C and D). Figure 2C shows that genetic distances increase gradually from the founder strain, consistent with the gradual genetic change of H3N2's HA1 observed by Smith and co-authors (6). However, antigenic change, as measured by cross-immunity, evolves in punctuated steps between clusters. In Fig. 2D, we plotted cross-immunity between genetically distinct strains that were sampled over the time course of the simulation. Under influenza's epochal evolution, a roughly linear relationship between antigenic differences and genetic differences emerges. This relationship is in agreement with the previously observed relationship between genetic and antigenic distances (6). However, the roughly linear relationship between these distances over the long run does not imply that genetic and antigenic differences are linearly related at shorter time scales. Figure 1C and the previous findings of punctuated antigenic change by Smith and colleagues (6) clearly show that antigenic and genetic distances can be uncorrelated at this temporal scale. The neutral network topology thus accounts for both the short and the long time-scale patterns of antigenic and genetic distance relationships.

The model also predicts an emergent dynamic property of influenza that arises from the interaction between cluster jumps and herd immunity. Specifically, the model frequently generates a refractory year of low attack rates after

cluster-transition years. During cluster transitions, many hosts become infected and recover, and hence there are few susceptible hosts to infect in the following year. Two years after a cluster-transition year, enough susceptible hosts have been replenished to cause another large annual outbreak. The replenishment of susceptible hosts comes from a combination of new births and the slight antigenic evolution of the strains that compose the cluster. The existence of this refractory period is empirically suggested—a pattern that, to our knowledge, has gone unnoticed. After a year dominated by a new cluster (and not immediately followed by another cluster), the incidence of H3N2 often drops below that of the following “normal” year—a year that is neither dominated by an invading cluster nor possibly refractory (10, 33, 34). This pattern occurs in six out of seven cases after the HK68 cluster. Interestingly, also in six out of seven potentially refractory years, subtypes H1N1 and/or B predominate.

We compared our model's results with the characteristic phylogenetic patterns of H3N2 HA1. A phylogenetic reconstruction of the sequences obtained from yearly sampling of the model's infected individuals generated a tree with a long trunk and short side branches that is characteristic of influenza (4) (Fig. 3D). To better understand the factors that determine the shape of this phylogenetic tree, we tracked the diversity of strains present in the population over time, as measured by average pairwise nucleotide

**Fig. 3.** Influenza dynamics and evolution from model simulations. **(A)** Time series of infected cases, color coded by cluster designation. Clusters remained dominant for less than a decade, in congruence with the empirically suggested two- to eight-season dominance of H3N2's clusters (6). **(B)** Annual attack rates, showing increases during cluster-transition years and a refractory period during each subsequent year. **(C)** Average pairwise nucleotide differences over time, showing a boom-and-bust pattern of genetic diversity that is associated with cluster-transition years. Nucleotide differences were calculated annually from 20 randomly sampled strains along cluster designations. Simulations were started at the endemic equilibrium of a single random strain. **(D)** Phylogenetic tree showing the restricted HA diversity generated by the model. The neighbor-joining (JC distance) tree was built with 10 randomly sampled strains from every year. Lineages are color coded by antigenic cluster. Simulations were run with an initial population size of two million. We let the duration of infection  $\nu^{-1} = 8$  days (48), the birth rate  $\mu_B = 14/1000$  per year, the death rate  $\mu_D = 8/1000$  per year, the mutation rate  $\delta = 2 \times 10^{-5}$  per base per day, and the probability of an adjacent network being antigenically discontinuous  $\rho = 1/40$ . We used an average basic reproductive rate  $R_0$  of 5 [consistent with estimates from (49)], with seasonal sinusoidal forcing  $c = 0.35$ . The strength of the interaction between the temperate region and the outside region is determined by the transmissibility-reduction factor  $r^{-1} = 1/150$  and the immigration rate of  $m = 50$  per day (32).



tide differences between strains (Fig. 3C). As a cluster emerges and spreads, strains diffuse through genotype space with weak positive selection from the slight antigenic changes accumulating at the epitopes, increasing strain diversity. The appearance of a novel phenotype severely curtails diversity in the old cluster. This contraction results from a selective sweep, with the mutant strain having higher fitness (because of the higher levels of susceptible hosts available to it), thereby allowing it to competitively exclude all the strains in the previous cluster. The invasion of the new cluster is facilitated by the accumulated diversity of the old cluster: The novel antigenic cluster can have up to 80% cross-immunity with strain variants of the old cluster that share four out of five of its epitopes; however, it has only 52% cross-immunity with strain variants of the old cluster that differ slightly in four out of five of the epitopes. The emergence of new clusters, their rapid expansion, and their decline create a boom-and-bust cycle of explosive diversity and contraction.

To validate this model, we computed the average pairwise nucleotide differences in the H3N2 sequences given by Smith and coauthors (6). Despite the limited number of antigenically typed samples, a boom-and-bust pattern is evident. There is a significant increase in nucleotide differences as the clusters age ( $P < 0.02$ ) (Fig. 4A), suggesting diffusion in genotype space. A plot of diversity over time (1968 to 2003) shows that this overall increase occurs in seven out of nine clusters and is greatest while the cluster is still dominant (Fig. 4B). This pattern in

diversity also holds when analysis is restricted to residues located in the five epitopes, with pairwise nucleotide differences growing from 0 to 3 to approximately 5 to 13 in 3 to 4 years. A further match between the model and H3N2 HA1 is observed when tree-balance metrics are employed to determine the degree of skewness within clusters (32). These types of quantitative measures should be used in the future to compare different phylodynamic models.

**A minimal theory for influenza.** By incorporating a GP mapping based on neutral networks into a simple multistrain transmission model, we have shown that major features of influenza's interepidemic ecology and evolution can be explained. These features include sequential cluster transitions and limited standing HA diversity in the viral population. Another model, which assumes that genotypic differences approximate phenotypic differences, has been successful in reproducing clustered strain appearances (12). However, this model is deliberately restricted to a much lower dimensional space [one-dimensional (1D) or 2D], which constrains the direction of evolution. Models that incorporate higher dimensions of mutational possibilities have not shown evidence for self-organized clusters; thus, explosive diversity results instead (5, 14). To limit viral diversity, these models have had to invoke temporary strain-transcending immunity (5, 14). Both of these models also predict fluctuations in strain diversity; however, these decreases in diversity are not associated with increases in the attack rate. It is also unclear whether or not these models generate antigenic clusters. The existence

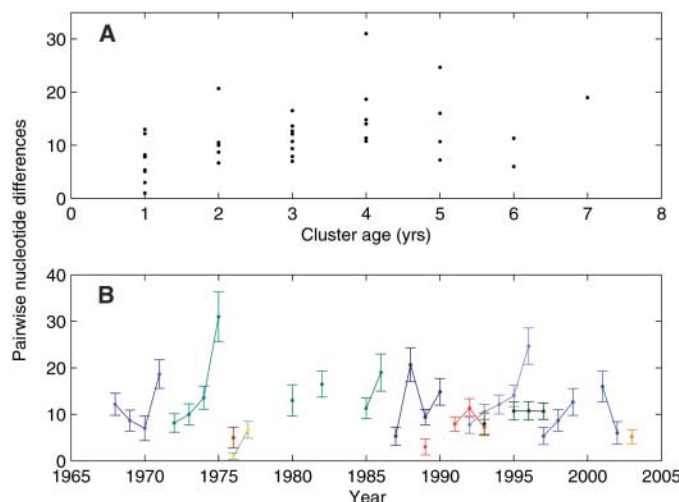
of generalized immunity is suggested by cellular immune responses in mice [(35, 36), but see (37)], although the amount of protection required in models of human populations appears to be comparatively greater. Here, we have shown that generalized immunity is unnecessary to restrict the interepidemic diversity of HA. Our model indicates that weak within-cluster selection and the selective sweeps that accompany cluster transitions are sufficient. Empirical patterns corroborate this model (Fig. 1B and Fig. 4).

This model made several simplifying assumptions that should be investigated. We assumed that only HA determines the antigenic phenotype. This is certainly not the whole story: Antibody responses to NA and internal proteins, cellular immune responses, and non-immune-mediated traits contribute to viral fitness and hence might affect influenza's dynamics and diversity patterns (38, 39). However, the fact that HA diversity and associated variations in herd immunity allow our model to explain emergent evolutionary and epidemic properties underlines the fact that HA is the major (phylo)dynamical component to build on. Teasing out whole-genome influences on influenza cluster transitions is an important area for future work. We have also assumed that there is no interaction between H3N2, H1N1, or influenza B, despite suggestive instances of subtype replacement (e.g., of H1N1 by H2N2 and H2N2 by H3N2). Although the model presented here predicts refractory periods after cluster-transition years during which H1N1/B outbreaks often occur, it does not include a mechanism to explain why H1N1/B outbreaks do not seem to occur in seasons with high H3N2 incidence. Some degree of generalized immunity (5) or ecological interference (40) may be needed to account for these interannual subtype patterns and pandemic subtype replacements. These processes do not, however, seem necessary to restrict the intrasubtype diversity of HA, nor can they parsimoniously explain aspects of H3N2's interannual variability.

Our model also opens up general questions about the time-series analyses of influenza dynamics at the population level. In the absence of strain data, influenza has been modeled as a simple SIRS system (41–43), with viral evolution causing transition from the recovered class to the susceptible class. In light of the punctuated cluster transitions identified by Smith and colleagues (6) and the implications of the transitions expanded upon in this paper, it is becoming evident that influenza dynamics can be approximated as a serial SIR system, with abrupt changes in the level of susceptible hosts when a new cluster appears. The development of such mathematical and statistical approaches may allow the short-term dynamics of influenza to be predicted without recourse to complex evolutionary models. Improved subtype-specific flu surveillance is the key to full utilization of such methods (44).

Further investigation of the biological basis of neutrality in GP mappings could yield new

**Fig. 4.** Patterns of genetic diversity in HA1. (A) Average pairwise nucleotide differences between the HA1 regions of circulating strains as a function of cluster age. Cluster age was calculated from the first year in which a strain from the cluster appeared in the data set. Each strain was assumed to circulate only during the year it was isolated. Average pairwise differences were calculated as the means within groups of isolates, distinguished by cluster and year, using the MEGA software package version 3.1 (47). The slope is significantly positive ( $P < 0.02$ ), as was determined by least-trimmed-squares regression (LTSR) and 1000 bootstrap samples. (B) Average pairwise nucleotide differences in HA1 by cluster and year. The average number of nucleotide differences increased in all nine clusters for which between-year comparisons are possible using least-squares regression, and in seven of nine clusters (EN72, VI75, BK79, SI87, BE89, BE92, and SY97) using LTSR. Vertical bars indicate 95% confidence levels, inferred from 500 bootstrap samples. More sequences were available for EN72, SI87, BE89, BE92, WU95, and SY97 than for other clusters (fig. S1). The patterns in diversity shown in (A) and (B) were similar when considering only the subset of codons belonging to epitopes A to E, as identified in (50) and (6). For some clusters, the growth in pairwise nucleotide differences was interrupted by sampling discontinuities (i.e., years for which one or no sequences of that cluster were sampled; see fig. S1).



insights into the phylodynamics of disease systems in general. Critically, neutrality is mediated not only by the accessibility of different structures to the pathogen but also by the hosts' abilities to detect these differences. The phenotype of the pathogen thus depends on the specificity of the host response. The availability of different phenotypes under this framework may establish major differences between the phylogenetic characteristics of different diseases [e.g., influenza A and measles (3)] and the same disease in different hosts experiencing different ecologies (e.g., influenza A in swine, avian, and equine hosts). There is also evidence that the specificity of the host antibody response to influenza varies within a host population (45). This heterogeneity might have important consequences for the diffusion of strains along antigenically similar networks in genotype space.

This study highlights the necessity of coupling molecular evolution with population-level models to understand the basic aspects of a biological system. In the case of influenza, a more detailed understanding of the structural basis of antigenicity and the dynamics of immune recognition of viral epitopes is of utmost importance. The roles that neutrality and context dependency play in enabling phenotypic change (46) need to be addressed in virological studies. These efforts could help to identify the GP map of influenza's HA and are critical for vaccine development, because antigenic distances between epidemic strains and vaccine strains determine vaccine efficacy (2). Further inquiries into population-level processes that affect influenza dynamics and evolution, such as the extent of epidemiological mixing and the global circulation of the virus in humans and other host species, are also necessary parts of this research. Integrating findings from these fields will be critical to understanding and managing influenza.

## References and Notes

- N. J. Cox, K. Subbarao, *Annu. Rev. Med.* **51**, 407 (2000).
- D. J. Smith, S. Forrest, D. H. Ackley, A. S. Perelson, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14001 (1999).
- B. T. Grenfell *et al.*, *Science* **303**, 327 (2004).
- W. M. Fitch, R. M. Bush, C. A. Bender, N. J. Cox, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7712 (1997).
- N. M. Ferguson, A. P. Galvani, R. M. Bush, *Nature* **422**, 428 (2003).
- D. J. Smith *et al.*, *Science* **305**, 371 (2004).
- P. W. Gill, A. M. Murphy, *Med. J. Aust.* **2**, 761 (1977).
- G. Meiklejohn, T. C. Eickhoff, P. Graves, I. Josephine, *J. Infect. Dis.* **138**, 618 (1978).
- S. Nakajima, E. Nobusawa, K. Nakajima, *Virology* **274**, 220 (2000).
- S. K. Greene, E. L. Ionides, M. L. Wilson, *Am. J. Epidemiol.* **163**, 316 (2006).
- D. C. Wiley, I. A. Wilson, J. J. Skehel, *Nature* **289**, 373 (1981).
- J. R. Gog, B. T. Grenfell, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17209 (2002).
- M. Girvan, D. S. Callaway, M. E. J. Newman, S. H. Strogatz, *Phys. Rev. E Stat. Nonlinear. Soft Matter Phys.* **65**, 031915 (2002).
- F. Tria, M. Lässig, L. Peliti, S. Franz, *J. Stat. Mech.* **1**, P07008 (2005).
- K. Nakajima, E. Nobusawa, A. Nagy, S. Nakajima, *J. Virol.* **79**, 6472 (2005).
- K. F. Lau, K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 638 (1990).
- D. J. Lipman, W. J. Wilbur, *Proc. R. Soc. London Ser. B* **245**, 7 (1991).
- A. Babajide *et al.*, *J. Theor. Biol.* **212**, 35 (2001).
- E. Bornberg-Bauer, *Biophys. J.* **73**, 2393 (1997).
- U. Bastolla, H. E. Roman, M. Vendruscolo, *J. Theor. Biol.* **200**, 49 (1999).
- H. Li, R. Helling, C. Tang, N. Wingreen, *Science* **273**, 666 (1996).
- P. Schuster, W. Fontana, *Phys. D* **133**, 427 (1999).
- W. Fontana, P. Schuster, *J. Theor. Biol.* **194**, 491 (1998).
- M. A. Huynen, *J. Mol. Evol.* **43**, 165 (1996).
- M. A. Huynen, P. F. Stadler, W. Fontana, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 397 (1996).
- E. van Nimwegen, J. P. Crutchfield, *Bull. Math. Biol.* **62**, 799 (2000).
- E. van Nimwegen, J. P. Crutchfield, M. Mitchell, *Phys. Lett. A* **229**, 144 (1997).
- E. van Nimwegen, J. P. Crutchfield, M. Mitchell, *Theor. Comput. Sci.* **229**, 41 (1999).
- L. Barnett, in *Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life* (MIT Press, Cambridge, MA, 1998), pp. 18–27.
- M. E. J. Newman, R. Engelhardt, *Proc. R. Soc. London Ser. B* **265**, 1333 (1998).
- S. A. Kauffman, S. Johnsen, *J. Theor. Biol.* **149**, 467 (1991).
- Materials and methods are available as supporting online material on Science Online.
- T. A. Reichert *et al.*, *Am. J. Epidemiol.* **160**, 492 (2004).
- W. W. Thompson *et al.*, *JAMA* **289**, 179 (2003).
- P. G. Thomas, R. Keating, D. J. Hulse-Post, P. C. Doherty, *Emerg. Infect. Dis.* **12**, 48 (2006).
- J. W. Yewdell, J. R. Bennink, G. L. Smith, B. Moss, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1785 (1985).
- S. R. Crowe, S. C. Miller, D. L. Woodland, *Vaccine* **24**, 452 (2006).
- J. R. Gog, G. F. Rimmelzwaan, A. D. M. Osterhaus, B. T. Grenfell, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11143 (2003).
- E. C. Holmes *et al.*, *PLoS Biol.* **3**, e300 (2005).
- P. Rohani, C. J. Green, N. B. Mantilla-Beniers, B. T. Grenfell, *Nature* **422**, 885 (2003).
- J. Dushoff, J. B. Plotkin, S. A. Levin, D. J. D. Earn, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16915 (2004).
- Y. Xia, J. R. Gog, B. T. Grenfell, *Appl. Stat.* **54**, 659 (2005).
- S. K. Greene, J. S. Koopman, M. L. Wilson, in *Proceedings of the International Conference on Options for the Control of Influenza V*, Y. Kawaoka, Ed. (Elsevier, San Diego, CA, 2004), pp. 795–798.
- D. J. Smith, *Science* **312**, 392 (2006).
- K. Sato *et al.*, *Epidemiol. Infect.* **132**, 399 (2004).
- A. Wagner, in *Robustness and Evolvability in Living Systems*, S. A. Levin, S. Strogatz, Eds. (Princeton Studies in Complexity (Princeton Univ. Press, Princeton, NJ, 2005), pp. 223–227.
- S. Kumar, K. Tamura, M. Nei, *Brief. Bioinform.* **5**, 150 (2004).
- R. Douglas, in *The Influenza Viruses and Influenza*, E. Kilbourne, Ed. (Academic Press, New York, 1975), pp. 395–418.
- C. C. Spicer, C. J. Lawrence, *J. Hyg. Camb.* **93**, 105 (1984).
- R. M. Bush, W. M. Fitch, C. A. Bender, N. J. Cox, *Mol. Biol. Evol.* **16**, 1457 (1999).
- We thank R. Fouchier, D. Smith, S. Greene, M. Boni, E. Holmes, D. Alonso, G. Sawicki, A. King, S. Allesina, I. Volkov, and J. Lloyd-Smith for useful discussions and comments; L. Fernando-Chaves and D. Ruiz-Moreno for feedback on statistical analyses; three anonymous reviewers for helpful comments; and the Center for the Study of Complex Systems at the University of Michigan and the High Performance Computing group at PSU for computing resources. This work was supported by the James S. McDonnell Foundation (through a Centennial Fellowship to M.P.), a CIDD postdoctoral fellowship to K.K., and an NSF Graduate Research Fellowship to S.C.

## Supporting Online Material

[www.sciencemag.org/cgi/content/full/314/5807/1898/DC1](http://www.sciencemag.org/cgi/content/full/314/5807/1898/DC1)

Materials and Methods

SOM Text

Figs. S1 to S6

Table S1

References

19 July 2006; accepted 30 October 2006  
10.1126/science.1132745

# Structure of Dual Function Iron Regulatory Protein 1 Complexed with Ferritin IRE-RNA

William E. Walden,<sup>1</sup> Anna I. Selezneva,<sup>1</sup> Jérôme Dupuy,<sup>2</sup> Anne Volbeda,<sup>2</sup> Juan C. Fontecilla-Camps,<sup>2</sup> Elizabeth C. Theil,<sup>3</sup> Karl Volz<sup>1\*</sup>

Iron regulatory protein 1 (IRP1) binds iron-responsive elements (IREs) in messenger RNAs (mRNAs), to repress translation or degradation, or binds an iron-sulfur cluster, to become a cytosolic aconitase enzyme. The 2.8 angstrom resolution crystal structure of the IRP1:ferritin H IRE complex shows an open protein conformation compared with that of cytosolic aconitase. The extended, L-shaped IRP1 molecule embraces the IRE stem-loop through interactions at two sites separated by ~30 angstroms, each involving about a dozen protein:RNA bonds. Extensive conformational changes related to binding the IRE or an iron-sulfur cluster explain the alternate functions of IRP1 as an mRNA regulator or enzyme.

**I**ron regulatory protein 1 (IRP1) is a cytosolic, RNA binding protein that regulates the translation or stability of mRNAs encoding proteins for iron transport, storage, and use. IRP1 has an alternate function as cytosolic (c-) aconitase when a [4Fe-4S] cluster is bound (1–5).

The distribution of IRP1 between these mutually exclusive activities requires no new protein synthesis; iron excess or starvation promotes c-aconitase or RNA binding activity, respectively (6). Assembly and disassembly of the [4Fe-4S] cluster appears to be an effective mechanism for