Hot new virus, deep connections

Roger W. Hendrix
Pittsburgh Bacteriophage Institute and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

Biologists of the 17th, 18th, and 19th centuries—call them natural historians—somehow got along without PCR, chip technology, mass spectrometry, or high-speed computers. From our vantage point in the 21st century we may wonder how our scientific ancestors, lacking our sophisticated tools, could have accomplished anything that we would recognize as forward progress in understanding how biology works. Yet they were in the enviable position that much of their biological world was completely unexplored; all they had to do to make a name for themselves was step out of their home habitat, usually Europe, and with a little luck they might find a new and wonderful organism, unlike anything previously known to science. Many examples of such discoveries can be cited (cycads, duck-billed platypuses, giant tortoises, Komodo dragons, and animalcules among them) and, in the most successful cases (think of Mr. Darwin), the new creature(s) dramatically enhanced our understanding of the structure and history of the biological world as a whole. Fortunately for the excitement quotient of modern-day natural historians, Mother Nature's reservoir of undiscovered bizarre and wonderful organisms is not yet empty, and a new one makes the transition from unknown to known with the report by Rice *et al.* (1) in this issue of PNAS.

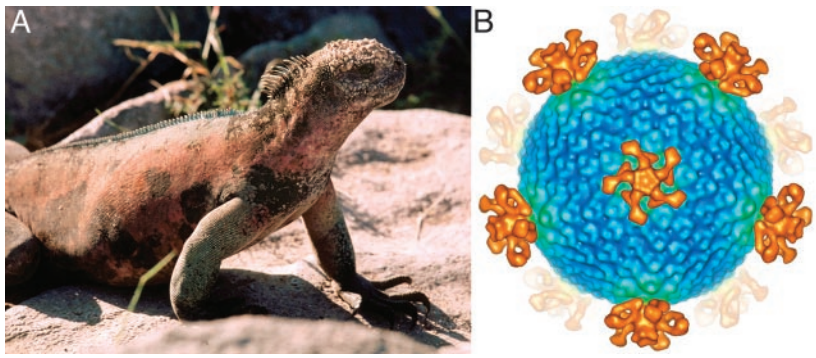

Fig. 1. Two organisms whose discovery has enriched our understanding of the biological world. (*A*) The Galapagos marine iguana, *Amblyrhynchus cristatus*. (*B*) Archaeal virus STIV, shown in the fivefold symmetrical view as it might appear to a cell about to be infected.

A New Virus

The new entry is a virus plucked from the near-boiling water of a thermal pool in Yellowstone National Park, and it is every bit as interesting to 21st century science as something like the Galapagos marine iguana (Fig. 1*A*) was to European science when it first came on the stage a few centuries ago. The new virus's host is the hyperthermophilic archaeon *Sulfolobus sulfataricus*, which grows happily at temperatures above 80°C and a pH of 2. Very few viruses of Archaea have been described to date [they amount to <1% of the viruses enumerated by the International Committee on the Taxonomy of Viruses (2)] but these early indications suggest that archaeal viruses likely are just as diverse as the more extensively characterized viruses of Bacteria and Eukarya. The viruses that infect the archaeal halophiles are so far confined to ones that have the same virion morphology and even occasional sequence similarity with the familiar tailed bacteriophages, but the viruses of the hyperthermophiles are a strange and diverse group with virion morphologies including filaments as well as shapes resembling food items such as lemons and corndogs. In this context, perhaps it is not surprising that the new virus would not look quite like anything described before. It is a spherical or, more properly, an icosahedrally symmetric virus (Fig. 1*B*), and, like most such viruses, the surface morphological features follow the rules enunciated by Caspar and Klug (3), although it has a previously undescribed triangulation number of 31. The most dramatic morphological feature of the virion is the protruding "turrets" that extend 13 nm above the capsid surface at the 12 fivefold symmetrical positions of the icosahedron. The function of the turrets is not known, but a plausible guess is that they have a role in attaching the virus to the cell and initiating infection. The morphological features are the basis for the authors' name for the virus: STIV, for *Sulfolobus* turreted icosahedral virus (1).

Early indications suggest that archaeal viruses likely are as diverse as viruses of Bacteria and Eukarya.

STIV has a moderately small, 17,663-bp circular double-stranded DNA genome, with 36 probable protein coding genes. Remarkably, only three of the predicted proteins match sequences in the public databases; all three are of unknown function, with two from other *Sulfolobus* viruses and the third from a *Sulfolobus* genome. The authors identify a fourth gene as encoding the 37-kDa major capsid subunit through analysis of the amino acid sequence. This preponderance of "pioneer" sequences is strongly reminiscent of the situation in the tailed bacteriophages; there, with the number of available genome sequences climbing past 200, it is still not uncommon for one-third or fewer of the predicted proteins of a newly sequenced phage genome to make database hits. In the bacteriophages, and also, one presumes, in the archaeviruses, this level of hits is due to a combination of a very large number of different kinds of genes in the viral population and an exceptionally high degree of sequence divergence in individual gene families (4). This finding emphasizes the richness of the undiscovered genetic diversity still out there.

Reminiscent of Known Viruses

By simply discovering and describing STIV, Rice *et al.* (1) have rendered a service to science. The properties of biological organisms cannot be predicted from first principles, because they are the products of evolution. As a consequence, we would never know what the marine iguana (Fig. 1*A*) has to tell us

See companion article on page 7716.

*E-mail: rhx@pitt.edu.

about how the biological world works if we did not first know that it exists. Similarly, the hole in our view of the biological world that the new virus (Fig. 1*B*) fills is a hole that we did not even know was there before this bit of biological exploration. However, the really interesting story here comes from going beyond the simple existence of STIV to a careful examination of its capsid structure. It is apparent from the cryoelectron microscopy structure that the hexons, i.e., the groups of subunits sitting in the positions where one usually finds a hexamer, are actually threefold symmetric rather than sixfold symmetric. This situation was first seen with adenovirus, where the hexons are trimers of the subunit. The x-ray structure of the adenovirus hexon (5) showed that each subunit is made of two structurally similar β-barrel ("jelly roll") domains; these are arranged in the trimeric hexon to produce a quasi-sixfold array of β-barrels, allowing the hexon to fit neatly into the quasi-sixfold symmetry of the surrounding protein lattice. It was a surprise a few years later when an x-ray structure of the hexon protein from *Escherichia coli* bacteriophage PRD1 (6) showed that it has essentially the same double β-barrel fold and is arranged similarly in the capsid lattice, despite the lack of any recognizable sequence similarity. More recently, it was shown that the same is true for the hexon of the virion of PBCV-1 (7), a virus that infects the eukaryotic alga *Chlorella*. In the case of STIV, Rice *et al.* (1) show that the high-resolution structures of the hexon proteins of adenovirus and phage PRD1 can be fit quite convincingly into the density envelope of the STIV cryoelectron microscopy structure, positioned the same way with respect to the capsid lattice as they are in their home capsid. Direct confirmation that STIV has the same double β-barrel fold as the other trimeric hexon proteins will have to await a high-resolution structure, but for now the smart money is heavily on the hypothesis that the STIV protein is the fourth member of this protein structure family and the first for a virus of Archaea.

Evolutionary Connections

Attempts to understand the evolutionary history of viruses have been frustrated by the facts that viruses do not leave fossils and that making conclusions about phylogenetic distance based on morphological similarity of virions is often misleading (8). The situation has gotten better with the availability of numerous genome sequences, which have made it possible to deduce some of the mechanisms of viral evolution. However, the sequences have revealed two complicating issues. First, viruses are the champions of horizontal swapping of genes, which means that their genomes have been through a sort of genetic Cuisinart; second, and more problematic for understanding deep evolutionary connections, viral protein sequences that we believe on other grounds to be homologous (that is, to share common ancestry) have often diverged to the point that no similarity in their sequences is detectable. This is where detailed structural similarities like the ones described in the previous paragraph come to the rescue. The assumption is that the structural similarities in the capsid proteins of adenovirus, phage PRD1, algal virus PBCV-1, and now archaeal virus STIV imply a common ancestry for those viruses (or strictly speaking, for the genes responsible for capsid structure), despite the absence of any surviving sequence similarity. In addition, there are two other groups of large viruses for which similar ancestral connections can be inferred across domains of life. First, the tailed phages of Bacteria and Archaea share enough similarities of virion structure and assembly mechanisms with the herpesviruses to make a compelling case for shared ancestry (9), and, second, the reoviruses of eukaryotic hosts and bacteriophages of the ϕ6 family share a characteristic T = 2:13 double capsid structure, segmented double-stranded RNA genome, and replication mechanism, leading to the same conclusion (10).

The simplest interpretation of these observations, and my own personal favorite, is that there were already viruses resembling modern adenoviruses, herpesviruses, and reoviruses active before the divergence of cellular life into the contemporary domains of Bacteria, Archaea, and Eukarya, $\approx$3 billion years ago. In this view, different lines of each of these virus types diverged in parallel with the cellular forms, with each viral line coevolving with one of the three cellular domains down to the present. The main alternative views are that the similar structures and assembly mechanisms arose independently and therefore do not imply common ancestry, or that each virus type arose more recently in one of the three domains and spread horizontally to the others. Which of these views (or which combination of them) is right can only become clearer as we isolate and characterize more new viruses and learn more about the viruses already in hand.

1. Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J. E., Douglas, T. & Young, M. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 7716–7720.
2. Ackermann, H. W. (2001) *Arch. Virol.* **146,** 843–857.
3. Caspar, D. L. D. & Klug, A. (1962) *Cold Spring Harbor Symp. Quant. Biol.* **27,** 1–24.
4. Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., *et al.* (2003) *Cell* **113,** 171–182.
5. Roberts, M. M., White, J. L., Grütter, M. G. & Burnett, R. M. (1986) *Science* **232,** 1148–1151.
6. Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. (1999) *Cell* **98,** 825–833.
7. Nandhagopal, N., Simpson, A. A., Gurnon, J. R., Yan, X., Baker, T. S., Graves, M. V., Van Etten, J. L. & Rossmann, M. G. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 14758–14763.
8. Lawrence, J. G., Hatfull, G. F. & Hendrix, R. W. (2002) *J. Bacteriol.* **184,** 4891–4905.
9. Newcomb, W. W., Juhas, R. M., Thomsen, D. R., Homa, F. L., Burch, A. D., Weller, S. K. & Brown, J. C. (2001) *J. Virol.* **75,** 10923–10932.
10. Butcher, S. J., Dokland, T., Ojala, P. M., Bamford, D. H. & Fuller, S. D. (1997) *EMBO J.* **16,** 4477–4487.