Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Cell, Vol. 105, 5–8, April 6, 2001, Copyright ©2001 by Cell Press

Virus Evolution: How Does an Enveloped Virus Make a Regular Structure?

Minireview

James H. Strauss* and Ellen G. Strauss*
Division of Biology
California Institute of Technology
Pasadena, California 91125

The evolution of viruses has been an exciting area of study, albeit an area that is fraught with difficulties because of the lack of a fossil record and because of the rapid sequence divergence exhibited by viruses. All viruses in collections available for study in the laboratory have been isolated within the last 70 years. Studies of the rate of sequence divergence in viruses over this period of time, all of which have focused on RNA viruses, have given estimates of 10^{-2} to 10^{-4} changes per nucleotide per year (Takeda et al., 1994; Weaver et al., 1997). Although these rates for the fixation of mutations of necessity assay changes in only the most variable positions in the viral genome, and there are clearly positions that change much more slowly, it is nonetheless clear that it is difficult to establish relationships between two viruses that last had a common ancestor, for example, a million years ago, based solely on sequence relationships. Furthermore, it has become increasingly clear in the last two decades that extensive recombination over the ages has complicated the evolutionary relationships among viruses belonging to different families (Strauss et al., 1996). To ascertain distant relationships among viruses, structural studies are of increasing importance, because the structure of a protein changes much less rapidly than does the amino acid sequence that forms the structure (Rossmann et al., 1974).

Two papers in this issue of *Cell* (Lescar et al., 2001; Pletnev et al., 2001) have now reported on the structure and position of the glycoproteins in the external surface of virions of alphaviruses (genus *Alphavirus*, family *Togaviridae*), using different methods. Lescar et al. (2001) used conventional crystallographic approaches to determine the structure of glycoprotein E1 of Semliki Forest virus. Pletnev et al. (2001) used cryoelectron microscopy to position glycoproteins E1 and E2 in intact virions of Sindbis virus. The two papers, taken together, support the surprising conclusion that E1 of alphaviruses is related to glycoprotein E of flaviviruses (genus *Flavivirus*, family *Flaviviridae*), whose structure had been previously determined (Rey et al., 1995), and that the structures of alphaviruses and flaviviruses are related, albeit distantly. Thus, these results support the hypothesis that alphavirus and flaviviruses are diverged from a common ancestor or, more precisely, that the structural components of these viruses have a common origin.

Ironically, early attempts by the International Committee for the Taxonomy of Viruses to classify viruses led to the inclusion of alphaviruses (originally called Group A arboviruses) and flaviviruses (first called Group B arboviruses) into the same family (*Togaviridae*). Classification was based on similarities in virus structure (enveloped, icosahedral viruses having plus strand RNA of 10–12 kb). However, the determination of the complete sequence of an alphavirus genome and of a flavivirus genome showed that the sequences were apparently unrelated and that the genome organizations were very different. Figure 1 compares the genomes of members of these two groups, illustrating that the gene order is different, the gene products are not congruent, and the translational and processing mechanisms used to produce the viral proteins are distinct. This led to the reclassification of the viruses into distinct families.

The Structures of Alphaviruses and Flaviviruses

Unlike many enveloped viruses, both alphaviruses (Strauss et al., 1995; Mancini et al., 2000) and flaviviruses (Ferlenghi et al., 2001) have a defined structure that makes them amenable to precise structural studies using the approach of cryoelectron microscopy (Baker et al., 1999). The alphavirus structure is illustrated in Figure 2. The nucleocapsid of the virus is composed of 240 molecules of capsid protein (shown in yellow-orange colors) arranged in a regular T=4 icosahedral lattice. The nucleocapsid is surrounded by a lipid envelope (green) derived from the host, in which are anchored two virus glycoproteins called E1 and E2 (blue). The glycoproteins penetrate the lipid bilayer and E2 makes contact with the capsid protein. E1 and a precursor to E2, variously called PE2 or p62, form a heterodimer that is essential for virus assembly. The E1-PE2 heterodimer is cleaved by furin to form an E1-E2 heterodimer during transport of the glycoproteins to the cell surface or during virus assembly. In the assembled virion, the E1-E2 heterodimers are arranged in a regular T=4 icosahedral lattice in which sets of three heterodimers form trimeric assemblies called spikes that project from the surface of the virion. Thus, alphaviruses are icosahedral structures that contain a lipid bilayer sandwiched between two protein shells, each of which possesses T=4 icosahedral symmetry.

The structure of flaviviruses is less well understood. However, it is known that the assembly process resembles that of alphaviruses. A glycoprotein heterodimer consisting of virus glycoproteins called prM (for precursor to M) and E is first formed. Both glycoproteins penetrate the bilayer, and cleavage of prM to M by furin occurs during assembly. Flaviviruses appear to utilize T=3 symmetry, however, and there are no projecting spikes on the surface of the virion. Instead the glycoproteins, or at least glycoprotein E, lies flat on the surface of the virion to form a fairly smooth layer that coats the surface of the lipid bilayer (Kuhn and Rossmann, 1995; Rey et al., 1995).

In alphaviruses, it was widely assumed that E1 and E2, which form an easily demonstrable heterodimer in the virion, both contribute to the projecting spikes. Thus, the absence of spikes in flaviviruses, together with the difference in triangulation numbers, suggested that the structures of alphaviruses and flaviviruses were unrelated. However, the paper by Lescar et al. (2001) shows that the fold of alphavirus E1 is related to that for flavi-

*E-mail: straussj@caltech.edu

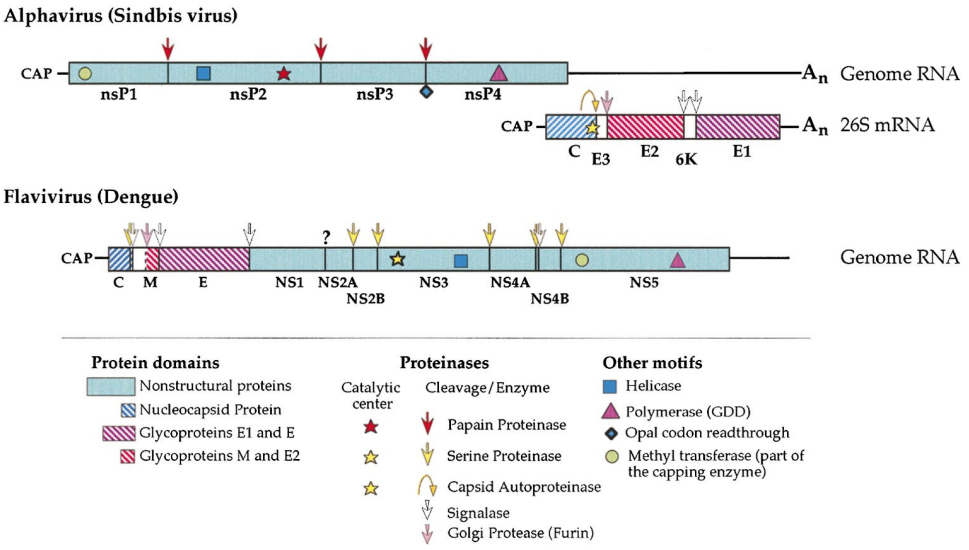

Figure 1. Comparison of the Genome Organizations of Alphaviruses and Flaviviruses

virus E, suggesting that the two proteins have a common ancestral source. They propose a fit of E1 into the cryo-EM density of Semliki Forest virus determined to 9 Å resolution (Mancini et al., 2000). The paper by Pletnev et al. (2001) shows that the bulk of E1 does not contribute to the outer portions of the spike but, instead, forms a layer closely apposed to the lipid bilayer, analogous to the position of E in the flavivirion. They also show that E2 projects upward to the full-length of the spike. Thus, E1 forms what has been called the skirt that surrounds the lipid bilayer and part of the lower domains of the spikes, whereas E2 forms the projecting part of the spike. The absence of spikes in flaviviruses could then be due to a difference between the cleaved E2 and M.

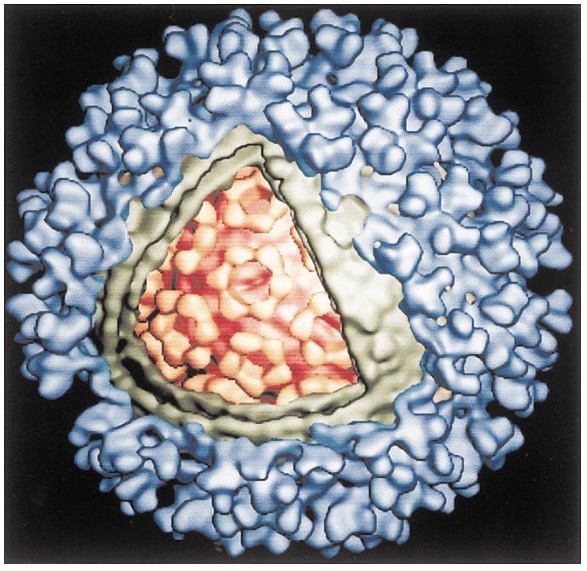

Figure 2. Cryoelectron Microscopic Reconstruction of an Alphavirus Virion (Ross River Virus)

E2 is ~420 amino acids in size, of which about 360 residues form the ectodomain, whereas M is only about 75 residues long, of which about 38 residues are present in the ectodomain. Thus, one can imagine that an immature flavivirion containing prM (about 170 residues) rather than M might more resemble the alphavirus structure, with short projecting spikes, but cleavage to M removes the spikes.

The Evolution of Enveloped Viruses

The parallels between the assembly of alphaviruses and flaviviruses and the similarities in structure revealed by the present studies suggest that an enveloped virus with an icosahedral structure arose long ago and has diverged into these two familes. Many other enveloped viruses whose structures are more or less known use quite different assembly mechanisms (van Regenmortel et al., 2000). Influenza virus (family *Myxoviridae*) or respiratory syncytial virus (family *Paramyxoviridae*), for example, have helical nucleocapsids. Although each virus has two glycoproteins in the envelope, one of which is cleaved by furin during transport or assembly, these two glycoproteins do not interact during assembly. Instead, each glycoprotein forms spikes that consist of homomultimers. The assembled virions are not uniform in composition or structure. Rather than forming regular structures in which the envelope proteins are present in quasi-equivalent positions, virions are pleomorphic and do not have a fixed composition. Retroviruses, such as HIV, use yet another assembly mechanism. The immature nucleocapsids are assembled using radial symmetry and are not icosahedral in structure (Wilk et al., 2001). The glycoproteins do not form a regular icosahedral lattice and the final composition of the virion is not rigorously constrained. In fact, nucleocapsids will bud in the absence of glycoproteins. Thus, it is possible that the ability to form a regular icosahedral, enveloped virus may have arisen only once. It is interesting to note that hepatitis C virus (genus *Hepacivirus*, family *Flaviviridae*), which is only distantly related to the members of the

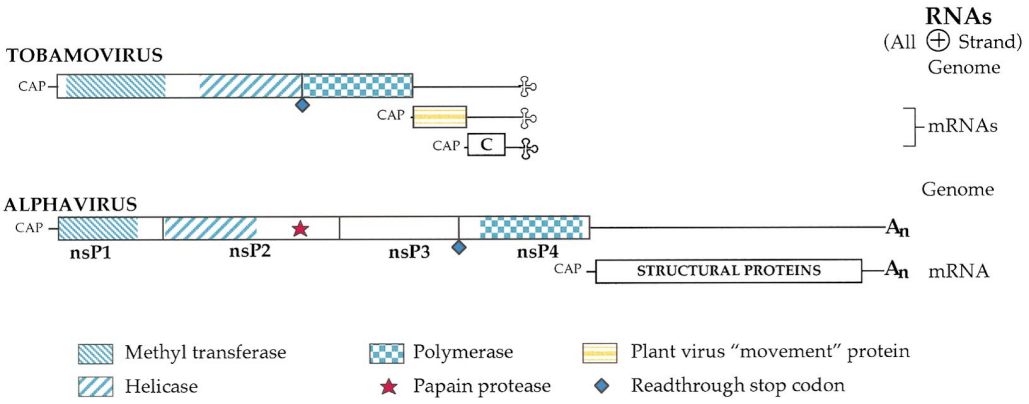

Figure 3. Comparison of the Genome Organizations of a Tobamovirus (Tobacco Mosaic Virus) and an Alphavirus (Sindbis Virus)

genus *Flavivirus*, is also thought to possess a glycoprotein with the same fold as the flavivirus E protein (Yagnik et al., 2000). Although the structure of hepaciviruses is unknown, it may be similar to that of flaviviruses. Further afield, it would be of interest to determine the structure of arteriviruses (family *Arteriviridae*), which are enveloped viruses whose genome organization and mechanism of RNA synthesis are related to those of coronaviruses (family *Coronaviridae*), but whose nucleocapsid is icosahedral rather than helical like coronaviruses (Snijder and Meulenberg, 1998).

Are Alphaviruses and Flaviviruses Related by Linear Descent from a Common Ancestor?

The results of Lescar et al. (2001) and of Pletnev et al. (2001) make it likely that the glycoproteins that form the surface of the virions of alphaviruses and flaviviruses derived from a common ancestor that arose long ago. What of the rest of the genomes? Did the entire genome result by linear descent from a common ancestor, or was recombination involved? The genome organizations of the two families of viruses, illustrated in Figure 1, make clear that recombination was at least partially responsible for the current genomes. It seems likely that during evolution the module that encodes the structural genes became associated with different modules encoding genes for RNA replication. This hypothesis is supported by comparisons of alphaviruses with a number of plant viruses. Plant viruses belonging to at least four different families have been found to possess RNA replication genes that are homologous to genes found in alphaviruses. However, the viruses have clearly undergone multiple recombination events in which new proteins have been introduced from other viruses or from the host during the evolution of this diverse group of viruses (Strauss et al., 1996). A comparison of tobacco mosaic virus (TMV) and Sindbis virus is shown in Figure 3. Although several of the RNA replication proteins of Sindbis virus and TMV are homologous (shaded areas in the nonstructural proteins), and the mechanisms of RNA synthesis are related, the structural genes of these two viruses are unrelated and the structures of the resulting virions are entirely different. TMV is not enveloped and the virion, which contains a single species of capsid protein, is a helical structure with the appearance of a rigid rod (Figure 4). Thus, it is not uncommon for structural modules to evolve independently of RNA replication modules, with mixing to form new combinations effected by recombination.

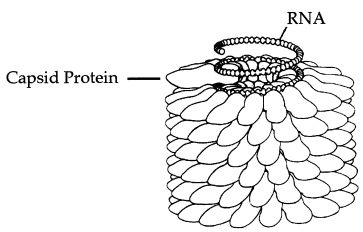

Figure 4. Diagrammatic View of the Helical Tobacco Mosaic Virus Virion

The Ecology of Alphaviruses and Flaviviruses

There are currently recognized about 75 flaviviruses and 25 alphaviruses. Both families have a worldwide distribution. Many of these viruses, but especially many of the flaviviruses, cause enormous numbers of cases of human illness each year. To cite a few flavivirus examples, the dengue viruses, of which there are four serotypes, cause an estimated 100 million cases of dengue fever each year; yellow fever virus is still widespread in Africa and Latin America; Japanese encephalitis virus is widespread in Asia; and West Nile virus has recently appeared in the Americas (the New York City area) for the first time. A hallmark of both alphaviruses and flaviviruses is that most of the viruses are transmitted by blood-sucking arthropods, that is, the viruses alternate between infection of an arthropod host and a vertebrate host. Although a number of viruses belonging to other families of RNA viruses are known to be arboviruses (so called because they are *ar*thropod-*bo*rne), only certain genera in the family *Bunyaviridae* rival the alphaviruses and flaviviruses in the number of and geographic range occupied by arboviruses (van Regenmortel et al., 2000). Perhaps the (related) structures of alphaviruses and flaviviruses have been important in allowing them to occupy and largely dominate the ecological niche represented by arboviruses.

Selected Reading

Baker, T.S., Olson, N.H., and Fuller, S.D. (1999). Microbiol. Mol. Biol. Rev. *63*, 862–922.

Ferlenghi, I., Clarke, M., Thomas, D., Rutten, T., Allison, S.L., Schalich, J., Heinz, F.X., Harrison, S.C., Rey, F.A., and Fuller, S.D. (2001). Mol. Cell *7*, 603–614.

Kuhn, R.J., and Rossmann, M.G. (1995). Nature *375*, 275–276.

Lescar, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., Wengler, G., and Rey, F.A. (2001). Cell *105*, this issue, 137–148.

Mancini, E.J., Clarke, E., Gowen, B.E., Rutten, T., and Fuller, S.D. (2000). Mol. Cell *5*, 255–266.

Pletnev, S.V., Zhang, W., Mukhopadhyay, S., Fisher, B.R., Hernandez, R., Brown, D.T., Baker, T.S., Rossmann, M.G., and Kuhn, R.J. (2001). Cell *105*, this issue, 127–136.

Rey, F., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). Nature *375*, 291–298.

Rossmann, M.G., Moras, D., and Olson, K.W. (1974). Nature *250*, 194–199.

Snijder, E.J., and Meulenberg, J.J.M. (1998). J. Gen. Virol. *79*, 961–979.

Strauss, J.H., Strauss, E.G., and Kuhn, R.J. (1995). Trends Microbiol. *3*, 346–350.

Strauss, E.G., Strauss, J.H., and Levine, A.J. (1996). Virus evolution. In Fields Virology, 3rd Edition, B.N. Fields, D.M. Knipe, P.M. Howley, et al., eds. (Philadelphia: Lippincott-Raven Publishers), pp. 153–171.

Takeda, N., Tanimura, M., and Miyamura, K. (1994). J. Virol. *68*, 854–862.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., and Wickner, R.B. (2000). Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses (San Diego, CA: Academic Press), 1162 pp.

Weaver, S.C., Kang, W.L., Shirako, Y., Rümenapf, T., Strauss, E.G., and Strauss, J.H. (1997). J. Virol. *71*, 613–623.

Wilk, T., Gross, I., Gowen, B.E., Rutten, T., de Haas, F., Welker, R., Kräusslich, H.-G., Boulanger, P., and Fuller, S.D. (2001). J. Virol. *75*, 759–771.

Yagnik, A.T., Lahm, A., Meola, A., Roccasecca, R.M., Ercole, B.B., Nicosia, A., and Tramontano, A. (2000). Proteins Struct. Function Genet. *40*, 355–366.