where $n=2$ and $n=4$ to the established number of $n=60$ boundary lipid sites per functional ATPase dimer. With the experimental data of Fig. 2, the best fit was generated when $a=2$, that is, only EL_{58}, EL_{59}, and EL_{60} appeared to possess ATPase activity[23]. The calculated regression curve is shown in Fig. 2. Additional activation curves for reconstituted Na^+K^+-ATPase and K^+-phosphatase activities were similarly generated by a computer treatment[23] of the experimental data in Ref. 14. The enzyme appeared to operate near a maximum of kinetic cooperativity. Lipid/protein equilibrium binding constants in the μM-range were obtained so that large free energy increments could be generated by changes in lipid/protein association. Such structural changes could therefore conceivably play a role in active transport.

Outlook

In the experiments of Fig. 2 and many similar studies, lipid activation was measured in the presence of excess non-activating detergent which served to disperse the lipid and the protein to a pseudo-homogeneous micellar solution. This situation does not exist in biological membranes. Here, two-dimensional concentrations have to be defined (e.g. as mole fraction units) and encounter probabilities have to be considered. In spite of the high lateral lipid mobility in most biological membranes, little free lipid may be present. Assuming, for example, a 1:1 weight ratio of protein to lipid and average mol. wt values of 32 000 and 800, respectively, one arrives at a number of 40 lipid molecules per protein molecule. About 20 boundary lipid sites are expected on a mol. wt 32 000 protein[4], so that about 50% of the lipid could be associated with proteins.

The above mechanism for kinetic cooperativity is so far supported by the kinetic studies on β-hydroxybutyrate dehydrogenase and Na^+K^+-ATPase. More generally, this mechanism could apply whenever cooperative effects are observed under conditions where ligand binding is non-cooperative. Spectroscopic lipid/protein binding studies[3–6] have given no indication for cooperativity, but direct lipid/protein binding methods will have to be developed in order to study lipid binding under the same experimental conditions as lipid activation. Such binding studies should also address the question whether drugs could act as analogues of boundary lipids. A possible example for this mode of action is presented by the phorboldiester tumor promoters which interact with protein kinase C as analogues of the endogeneous 1,2-diglyceride activator[24].

Acknowledgements

Studies from this laboratory have been supported by the Deutsche Forschungsgemeinschaft and in part by the Fonds der Chemischen Industrie.

References

1 Sandermann, H. (1978) *Biochim. Biophys. Acta* 515, 209–237
2 Bohnenberger, E. and Sandermann, H. (1983) *Eur. J. Biochem.* 132, 645–650
3 Griffith, O. H., Brotherus, J. R. and Jost, P. C. (1982) in *Lipid–Protein Interactions*, Vol. 2 (Jost, P. C. and Griffith, O. H., eds), pp. 225–237, J. Wiley & Sons
4 Marsh, D. (1983) *Trends Biochem. Sci.* 8, 330–333
5 Caffey, M. and Feigenson, G. W. (1981) *Biochemistry* 20, 1949–1961
6 Simmonds, A. C., East, J. M., Jones, O. T., Rooney, E. K., McWhirter, J. and Lee, A. G. (1982) *Biochim. Biophys. Acta* 693, 398–406
7 Seelig, A. and Seelig, J. (1978) *Hoppe-Seyler's Z. Physiol. Chem.* 359, 1747–1756
8 Hoffmann, W., Sarzala, M. G., Gomez-Fernandez, J. C., Goni, F. M., Restall, C. J., Chapman, D., Heppeler, G. and Kreutz, W. (1980) *J. Mol. Biol.* 141, 119–132
9 Chapman, D., Gomez-Fernandez, J. C. and Goni, F. M. (1982) *Trends Biochem. Sci.* 7, 67–70
10 Jost, P. C., Griffith, O H., Capaldi, R. A. and Vanderkooi, G. (1973) *Proc. Natl Acad. Sci. USA* 70, 480–484
11 Warren, G. B., Birdsall, N. J. M., Lee, A. G. and Metcalfe, J. C. (1974) in *Membrane Proteins in Transport and Phosphorylation* (Azzone, G. F., Klingenberg, M. E., Quagliariello, E. and Siliprandi, N., eds), pp. 1–12. North-Holland Publishing Company
12 Rupley, J. A., Gratton, E. and Careri, G. (1983) *Trends Biochem. Sci.* 8, 18–22
13 Koenig, S. H. (1980) in *Water in Polymers* (Rowland, S. P., ed.), *ACS Symposium Series No. 127*, pp. 157–176, American Chemical Society
14 Ottolenghi, P. (1979) *Eur. J. Biochem.* 99, 113–131
15 Sandermann, H. (1982) *Eur. J. Biochem.* 127, 123–128
16 Koshland, D. E. Jr., Nemethy, G. and Filmer, D. (1966) *Biochemistry* 5, 365–385
17 Monod, J., Wyman, J. and Changeux, J.-P. (1965) *J. Mol. Biol.* 12, 88–118
18 Tyuma, I., Imai, K. and Shimizu, K. (1973) *Biochemistry* 12, 1491–1498
19 Blake, R. II., Hager, L. P. and Gennis, R. B. (1978) *J. Biol. Chem.* 253, 1963–1971
20 Cortese, J. D., Vidal, J. C., Churchill, P., McIntyre, O. and Fleischer, S. (1982) *Biochemistry* 21, 3899–3908
21 Cortese, J. D. and Vidal, J. C. (1983) *Arch. Biochem. Biophys.* 224, 351–357
22 Hess, G. P., Cash, D. J. and Aoshima, H. (1983) *Annu. Rev. Biophys. Bioeng.* 12, 443–473
23 Sandermann, H. and Gottwald, B. A. (1983) *Biochim. Biophys. Acta* 732, 332–335
24 Nishizuka, Y. (1983) *Trends Biochem. Sci.* 8, 13–16
25 Møller, J. V., Andersen, J. P. and leMaire, M. (1982) *Mol. Cell. Biochem.* 42, 83–107
26 Jørgensen, P. L. (1982) *Biochim. Biophys. Acta* 694, 27–68
27 Dean, W. L. and Gray, R. D. (1982) *J. Biol. Chem.* 257, 14679–14685
28 Grover, A. K., Slotboom, A. J., de Haas, G. H. and Hammes, G. G. (1975) *J. Biol. Chem.* 250, 31–38
29 Isaacson, Y. A., Deroo, P. W., Rosenthal, A. F., Bittman, R., McIntyre, J. O., Bock, H.-G., Gazzotti, P. and Fleischer, S. (1979) *J. Biol. Chem.* 254, 117–126
30 Vik, S. B., Georgevich, G. and Capaldi, R. A. (1981) *Proc. Natl Acad. Sci. USA* 78, 1456–1460

Slipped-mispairing and the evolution of introns

Gordon P. Moore

Introns are regions of non-coding nucleotides which interrupt most eukaryotic genes. Comparison of homologous genes in related species shows that the number and position of introns are usually highly conserved during evolution, but that intron length can change relatively quickly. Detailed comparison of the DNA sequence of introns suggests that a common mechanism for varying length is insertion or deletion caused by 'slipped-mispairing' of short, direct repeats during DNA replication.

Introns, or intervening sequences, were first discovered in genes of animal viruses. They were subsequently found in genes coding for ribosomal RNA, transfer RNA and many proteins in all eukaryotes which have been examined. Although introns do not occur in prokaryotic genes, almost all genes in higher eukaryotes are interrupted, and some contain as many as 50 introns, comprising more than 90% of the gene[1]. Introns are transcribed along with coding DNA (exons), but are spliced out of RNA prior to transport from the nucleus for translation into protein.

The discovery that genes occur in coding blocks, rather than as a continuous unit, has

Gordon P. Moore is at the Division of Biological Sciences, The University of Michigan, Ann Arbor, MI 48109, USA.

had a great impact on genetics and cell biology. In spite of a great deal of effort, we do not yet understand the function of introns, how they evolved, or the mechanisms by which they are cleaved out of RNA. Nevertheless, some important generalizations about introns can now be made and some interesting exceptions to these rules have been found. As a generalization, it seems that introns are spliced out in a defined order, and it is probable that their removal is somehow linked to transport out of the nucleus and subsequent stability of the mRNA. Further, enzymes which catalyse RNA splicing seem to recognize specific sequences at the intron/exon border[2] since mutations in this region result in abnormal splicing and usually inactivate the gene. In addition to splicing enzymes, a class of small, nuclear RNA molecules appear to be involved in the splicing process[3]. Although splicing enzymes are thought to be required for removal of most introns, a recent finding is that some introns are spliced out of RNA in the complete absence of enzyme catalysis ('self-splicing' RNA)[4]. Most introns, of course, do not code for protein. Particularly interesting exceptions are introns of some mitochondrial genes which code for proteins which apparently have a role in the process of intron splicing itself[5]. With this brief summary as an introduction, we turn now to the subject of intron length and position.

Constancy of intron position – variation of intron length

A striking fact has emerged from examination of individual genes in related species. The positions of introns tend to remain constant during evolution. As an example, β-like globins from such diverse sources as human, rabbit, chicken, mouse and frog have introns in exactly the same positions[6]; even the distantly related plant protein leghemoglobin has introns in the same two positions[7]. Although there are exceptions to this rule [such as the rat preproinsulin gene 1 (Ref. 8)] these probably represent cases where introns have been lost, rather than gained or translocated.

Introns are thought to occur between 'domains' of the corresponding protein[9]. Although it is rather vaguely defined, the word 'domain' denotes a structural and/or functional unit. Occurrence of introns between domains suggests that introns arose through translocation of independent genes to positions near each other and fusion to form a new, intron-containing gene. Examples of genes which have introns between domains, suggesting that such 'exon-shuffling'[10] has taken place, include the immunoglobulins[11], α-fetoprotein[12] and ovomucoid[13].

In contrast to evolutionary conservation of intron position, both the base sequence and the length of introns change much more rapidly than coding DNA. The sequence change results from single nucleotide substitutions which are probably not selected against because most introns do not have a coding function. Nucleotide replacements, however, cannot explain evolutionary change in the length of introns. Continuing the example of the β-like globin genes, the length of one of the two introns varies from 116 nucleotides in mouse to 130 in human; the other intron varies from 573 nucleotides in rabbit to 906 in human[6]. The remainder of this review will focus on the molecular mechanism of length changes during the evolution of introns.

Slipped-mispairing as a mechanism of mutation

It has been known since Seymour Benzer's classic mutational studies in the rII cistrons of bacteriophage T4, that all genetic loci are not equally mutable[14]. Rather, there are 'hot-spots' which mutate as much as 100 times more rapidly than other sites. In prokaryotes, where such hot-spots have been cloned and analysed[15], they are found to be associated with the presence of short, directly repeated sequences of nucleotides. These repeats can be amplified (insertion) or lost (deletion) at a high frequency. Non-repetitive sequences between the repeat units can also be amplified or lost during the process.

Why should the presence of repeated sequences result in a hot-spot for mutation? A model has been presented by Streisinger *et al.*[16], which postulates that adjacent, direct repeats can mispair during DNA replication causing a loop. If new DNA synthesis occurs using the mispaired DNA as template, an insertion is generated; if DNA of the loop is cleaved, deletion is the result. It can be difficult to distinguish which process (insertion or deletion) has occurred when presented with only two sequences for comparison. Since only the deletion aspect of the model has been rigorously tested by mutational analysis, this discussion will focus on the kind of slipped-misparing deletion shown diagrammatically in Fig. 1, with the understanding that what appears to be a deletion in one sequence may, in fact, be insertion in its homologue.

Examples of slipped-mispairing deletion in prokaryotes

The slipped-mispairing mechanism was proposed to explain mutational hot-spots in T4 phage. Subsequent analysis in the *lac*I gene of *E. coli* showed that deletions of 9–123 base pairs arise spontaneously and rapidly between repeats as short as 5–8 base pairs[15]. Deletions associated with direct repeats have also been reported in phage T7 (Ref. 17) and at other loci of *E. coli*[18]. Recently, Albertini *et al.*[19] used an *E. coli* mutant with fused *lac*I and *Z* genes to detect and analyse a large collection of spontaneous deletions. Almost all of the deletions occurred between short repeats as predicted by the slipped-mispairing model. Moreover, the longer the repeat homology, the more frequently the deletion occurred. Although perfect homology of the repeat units is not necessary in order for slipped-mispairing to occur, Albertini *et al.* were able to make the demonstration complete by showing that introduction of sufficient mismatch into the repeat units dramatically reduced the frequency of deletion.

Surprisingly, homologies as short as 14 base pairs can mediate deletions as large as 700–1000 base pairs. The frequency of such deletions is strongly correlated with the length of the homology and is probably affected by the distance between homologies, the surrounding sequences and, perhaps, the sequence of the repeat itself. It is immensely satisfying for geneticists to find a molecular explanation for Benzer's 23-year-old observation of mutational hot-spots and to find that, just as he predicted, the nucleotide sequence of particular loci can render them more or less susceptible to mutation.

The role of slipped-mispairing in evolution of introns

To what extent can the slipped-mispairing mechanism, discovered in prokaryotes, be extrapolated to eukaryotes and, in particular, to the evolution of intron structure? The question is not easy to answer because generation and selection of mutants is difficult in higher eukaryotes. Also, the non-coding nature of most introns renders them refractile to mutant analysis with the exception (as noted above) of the regions directly adjacent to the exon which are involved in splicing.

A different approach, comparison of the DNA sequence of homologous introns in related species, has revealed the influence of slipped-mispairing on intron structure. When the intron sequences are aligned, events of insertion or deletion are evident; the presence of direct repeats near the borders of these regions indicates that slipped-mispairing has been the mechanism of the mutation. Unfortunately, such interspecies comparison can be equivocal because, during the course of evolution, secondary events of mutation can obscure the process. For example, the direct repeats responsible for slipped-mispairing may have mutated beyond recognition or the

introns may be difficult to align unambiguously.

In my laboratory we have encountered an interesting case in which the effect of slipped-mispairing on introns is particularly clear (Ref. 20 and Foran, D., Johnson, P. and Moore, G., submitted for publication). We have sequenced two actin genes from the sea urchin *Strongylocentrotus franciscanus* which are physically linked and separated by about 5 kilobases. The two genes are extremely similar (they differ by only 1.7% of nucleotides in the coding region) and probably derive from a recent event of gene duplication. This situation is ideal because we can observe what has happened to introns in the two genes during the relatively short time since they diverged from one another. In molecular terms, this is equivalent to finding a well-preserved fossil whose details have not been obscured by the effects of weathering.

In both actin genes there is an intron at the codon which specifies amino acid 204 of the protein. One intron is 172 nucleotides in length and the other is 191. When the sequences are aligned, there are only five bases which do not match (which again shows that these genes have only recently diverged). However, there are ten deletions; these are obviously responsible for the difference in overall length. Five of the deletions are 3–20 nucleotides in length, while 5 are a single base. Every one of the large deletions is associated with a short, direct repeat suggesting that it was generated by slipped-mispairing. All of the single base deletions occur in runs of A (adenine) or T (thymine) and were probably generated by the same mechanism.

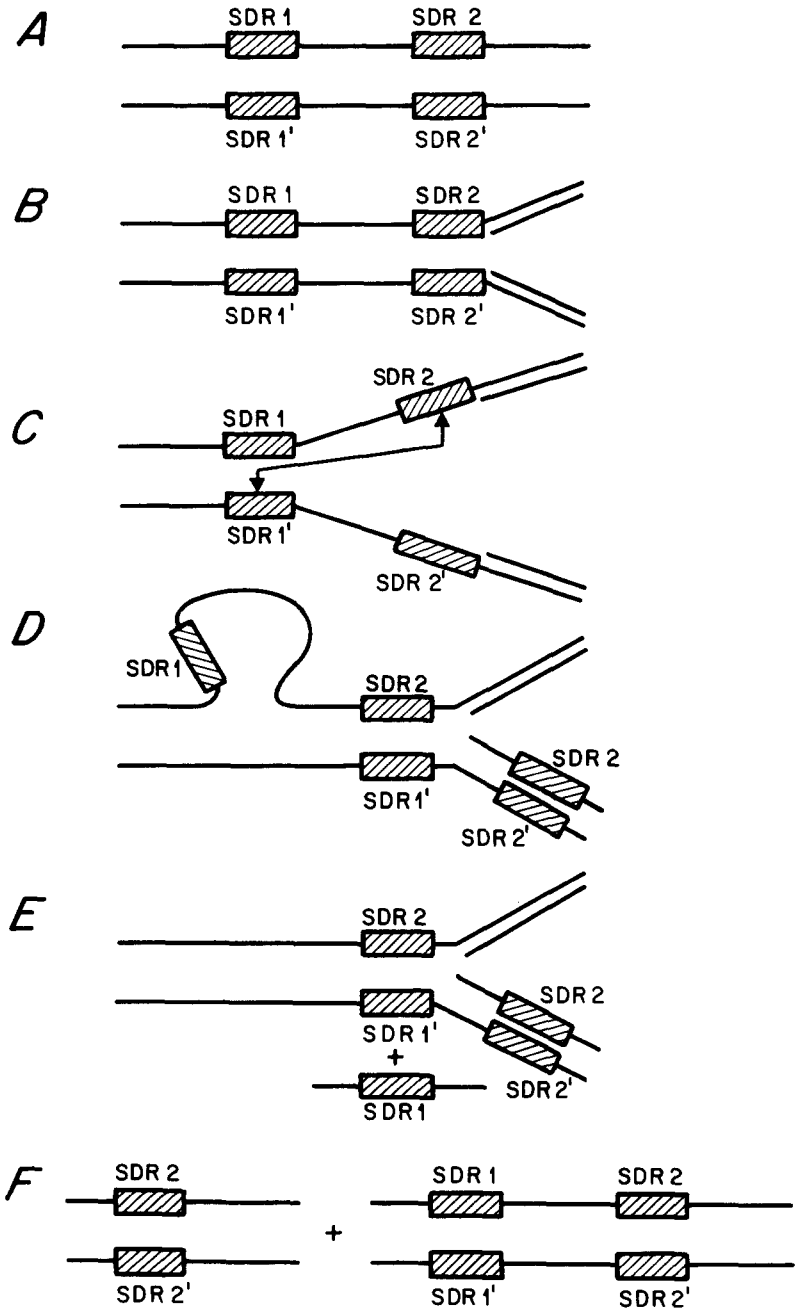

Fig. 1. The generation of deletions by 'slipped-mispairing' of direct repeats during replication of DNA. Parallel solid lines represent complimentary strands of DNA, while shaded boxes labeled SDR represent short, direct repeats and their compliments on the opposite strand. (A) Double strand of DNA containing two short, direct repeats; (B) replication fork; (C) the strands separate; (D) slipped-mispairing of the repeats results in a loop whose size is dependent upon the distance between the repeats; (E) cleavage of the loop; (F) one daughter has undergone a deletion of SDR 1 and flanking DNA.

Why is slipped-mispairing so common in introns?

There are other examples of introns which are affected by deletion or insertion caused by slipped-mispairing. These include introns of the human β-like globin genes[21], chorion genes of silkmoth[22] and actin genes of other sea urchin species[23]. The linked *S. franciscanus* actin genes described above allow us to estimate the rate of slipped-mispairing deletion relative to simple base changes. Those accustomed to analysis of mutations in structural genes would expect base changes to far outnumber deletions. In the introns, however, deletions are twice as frequent as base changes. Thus, the dominant event which leads to divergence of these introns during evolution does not appear to be nucleotide replacement, but rather mutation caused by slipped-mispairing.

Slipped-mispairing is not the only mechanism which can result in rapid change of intron length during evolution, indeed we discuss other mechanisms elsewhere (Foran, D., Johnson, P. and Moore, G., submitted for publication). Nevertheless, we are forced to ask why slipped-mispairing is so prevalent in introns. Although the question cannot be answered with certainty, introns have a property in common with other non-coding sequences which probably renders them particularly susceptible to slipped-mispairing: a surprisingly high proportion of A and T in their sequences. The introns of the sea urchin actin genes described above are 72% A or T, and this is fairly typical of introns in general. It is intuitively evident asymmetry with respect to base composition increases the frequency of random repeated sequences.

Several workers, including myself, have described equations which quantitate the effect of base composition on frequency of repeats[24,25]. Given that repeats which result in slipped-mispairing can be quite short, and that perfect homology is not required, the effect of high A + T content could be to increase dramatically the number of short, direct repeats relative to what would be expected if base ratios were more equal. This could lead to a high rate of slipped-mispairing and consequent mutation. It should be stressed that it is not high A + T content *per se* which promotes slipped-mispairing, but rather asymmetry of base composition leading to an increase in the number of short repeats. A high fraction of G + C would have the same effect, although this is not commonly the case in introns. We are left then, with the secondary question of why introns are composed predominantly of A and T. Although there has been some speculation[26], the answer to this question, like many others regarding the evolution and function of introns, remains unknown.

References

1 de Crombrugghe, B. and Pastan, I. (1982) *Trends Biochem. Sci.* 7, 11–13

2 Mount, S. (1982) *Nucl. Acids. Res.* 10, 459–570

3 Mount, S., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J. (1983) *Cell* 33, 509–518

4 Kruger, K., Grabowski, P. Zaug, A., Sands, J., Gottshling, D. and Cech, T. (1982) *Cell* 31, 147–157

5 Anziano, P., Hanson, D., Mahler, H. and Perlman, P. (1982) *Cell* 30, 925–932

6 Jeffreys, A. (1982) in *Genome Evolution* (Dover, G. A. and Flavell, R. B., eds), pp. 157–177. Academic Press

7 Jenson, E., Palvdan, K., Hyldig-Nielsen, J., Jørgensen, P. and Marker, K. (1981) *Nature* (London) 291, 677

8 Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) *Cell* 21, 555–566

9 Blake, C. (1979) *Nature* (London) 277, 598

10 Gilbert, W. (1978) *Nature* (London) 271, 501

11 Sanko, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979) *Nature* (London) 280, 288–294.

12 Eiferman, F., Young, P., Scott, R. and Tilghman, M. (1981) *Nature* (London) 294, 713

13 Croose, G., Simonsen, C., McEwan, R. and Schimke, R. (1982) *J. Biol. Chem.* 257, 7887–7892

14 Benzer, S. (1961) *Proc. Natl Acad. Sci. USA* 46, 1585–1594

15 Farabaugh, P. and Miller, J. (1978) *J. Mol. Biol.* 126, 847–857

16 Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* 31, 77–84

17 Studier, F., Rosenberg, A., Simon, M. and Dunn, J. (1977) *J. Mol. Biol.* 135, 917–929

18 Wu, A., Chapman, A., Platt, T., Guarente, L. and Beckwith, J. (1980) *Cell* 19, 829–836

19 Albertini, A., Hofer, M., Calos, M. and Miller, J. (1982) *Cell* 29, 319–328

20 Johnson, P., Foran, D. and Moore, G. *Mol. Cell. Biol.* (in press)

21 Efstratiadis, A. *et al.* (1980) *Cell* 21, 653–668

22 Jones, C. and Kafatos, F. (1980) *Cell* 22, 855–867

23 Schular, M., McOsker, P. and Keller, E. (1983) *Mol. Cell. Biol.* 3, 448–456

24 Day, G. and Blake, R. (1982) *Nucl. Acids. Res.* 10, 8323–8332

25 Moore, G., Moore, A. and Grosman, L. I. *J. Theor. Biol.* (in press)

26 Gojobori, T., Li, W. and Graur, D. (1982) *J. Mol. Evol.* 18, 360–369

Molecular variation of influenza surface antigens

L. Van Rompuy, W. Min Jou, M. Verhoeyen, D. Huylebroeck and W. Fiers

Although vaccination can prevent viral diseases such as poliomyelitis and measles in mankind, and has eradicated smallpox, influenza remains a very common cause of illness. The main reason for this lies in the frequent antigenic changes of the virus surface antigens so that the protective effects of a previous immune response are circumvented.

Influenza viruses (Fig. 1) bear two surface antigens, a hemagglutinin (HA) and a neuraminidase (NA). Although both antigens change independently, the hemagglutinin is considered the more important because only antibodies directed against this antigen can prevent infection. The changes may be drastic (antigenic shift) or relatively minor (antigenic drift). While there may be many practical reasons why scientists have been interested in this virus, they have always been especially attracted by the mechanism of variation itself. Are the virus surface antigens presented to the immune system in such a way that it has to select a predetermined and finite series of variant strains (drift) which, in a final step, leads to shift, or does the viral replicase produce variants in a special systematic way?

In recent years, several lines of investigation have converged to yield a rather detailed molecular picture of influenza surface antigens.

Classification of strains

Influenza virus strains are divided into types A, B and C, according to the serological cross-reactivity of their major internal protein, the nucleoprotein. Type A is the most extensively studied since it is the most pathogenic type for humans. Influenza B and C are less pathogenic to man and, unlike type A viruses, are not known to occur in animals. Within the type A viruses, a further classification is based upon the immunological cross-reactivity of the HA and NA surface antigens. The appearance of completely new subtype surface antigens in the human population is called antigenic shift. Among type A viruses, 13 HA subtypes (designated H1 to H13) and 9 NA subtypes (called N1 to N9) have been distinguished so far. Of these only three influenza A combinations, with the antigenic composition H1N1 (1918–1957, 1977–present), H2N2 (1957–1968) and H3N2 (1968–present) have been observed in humans this century.

The nomenclature system can be understood as follows: A/Aichi/68 (H3N2) refers to an influenza A virus isolated from man in the Aichi area (of Japan) in 1968 with an HA of subtype H3 and an NA of the

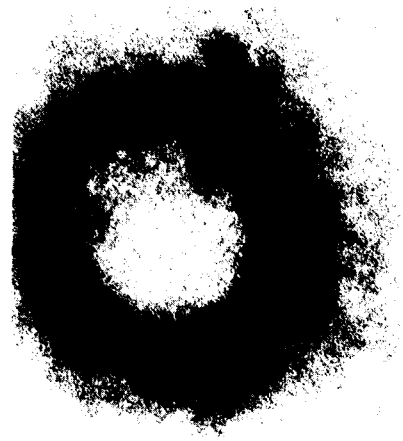

Fig. 1. Electron micrograph of an influenza A type virion (×200 000). From *The Influenza Viruses and Influenza* (Kilbourne, E. D., ed.), Academic Press; reproduced with permission from P. W. Choppin and R. W. Compans (1975). *The surface antigens are visible as spikes.*

The authors are at the Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B9000 Ghent, Belgium.