VIROLOGY 176, 306–307 (1990)

Recommendations of the Coronavirus Study Group for the Nomenclature of the Structural Proteins, mRNAs, and Genes of Coronaviruses

D. CAVANAGH,[1,*] D. A. BRIAN,[2] L. ENJUANES,[3] K. V. HOLMES,[4] M. M. C. LAI,[5] H. LAUDE,[6] S. G. SIDDELL,[7] W. SPAAN,[8] F. TAGUCHI,[9] AND P. J. TALBOT[10]

[1]*Division of Molecular Biology, AFRC Institute for Animal Health, Houghton Laboratory, Huntingdon, Cambridgeshire PE17 2DA, United Kingdom;* [2]*Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996-0845;* [3]*Centro de Biologia Molecular (CSIC-UAM), Universidad Autonoma, Canto Blanco, 28049 Madrid, Spain;* [4]*Department of Pathology, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, Maryland 20814-4799;* [5]*Department of Microbiology, USC School of Medicine, 2011 Zonal Avenue, Los Angeles, California 90033-1054;* [6]*INRA, Station de Recherches de Virologie et d'Immunologie Moléculaire, Centre de Recherches de Jouy-en-Josas, Domaine de Vilvert, 78350 Jouy-en-Josas, France;* [7]*Institut für Virologie und Immunbiologie der Universität Würzburg, Versbacher Strasse 7, 8700 Würzburg, West Germany;* [8]*Institute of Infectious Disease & Immunology, Virology Department, State University of Utrecht, Yalelaan 1, P.O. Box 80165, 3508 TD Utrecht, The Netherlands;* [9]*National Institute of Neuroscience, NCNP, 4-1-1 Ogawahigashi, Kodaira, Tokyo 187, Japan; and* [10]*Institut Armand-Frappier, Université du Québec, Virology Research Centre, Ville de Laval, Québec, Canada H7N 4Z3*

Received November 22, 1989; accepted November 28, 1989

We propose a nomenclature to replace the various systems currently in use to designate coronavirus structural proteins, mRNAs, and genes/open reading frames. The nonstructural proteins have not been addressed. © 1990 Academic Press, Inc.

Several names are currently used to refer to each of the three or four structural proteins of coronaviruses, with corresponding and different acronyms. Similarly, the mRNAs (reviewed in Refs. (*1*) and (*2*)) have been referred to by numbers or by letters. Finally, the genes/open reading frames (ORFs) have been designated by different authors in several different ways.

To overcome the confusion thus created, the Coronavirus Study Group of the Vertebrate Virus Subcommittee of the International Committee on Taxonomy of Viruses has reviewed the situation. At the Fourth International Symposium on Coronaviruses, held in July 1989 at King's College, Cambridge, England, the Group recommended a revised nomenclature to be used for all coronaviruses. The guidelines that have been formulated, if followed, will allow for newly identified mRNAs, genes, or ORFs to be named without creating confusion. The Study Group considered it inappropriate to make recommendations regarding proteins that are believed to be nonstructural because at the moment information is limited.

STRUCTURAL PROTEINS

The recommended acronyms for the structural proteins are shown in Table 1. The virion proteins have recently been reviewed (*2*).

* To whom requests for reprints should be addressed.

The large surface projection (spike or peplomer glycoprotein) has previously been referred to as S (*3*) or E2 (*4*). The acronym S may be used to denote the primary translation product and generally to refer to the spike glycoprotein. In some, though not all, coronaviruses, S is cleaved into two glycopolypeptides, the amino (N)-terminal S1 (E2B) and the carboxy (C)-terminal S2 (E2A) glycopolypeptides (the previous alternative acronyms are shown in parentheses (*5*)). The hemagglutinin–esterase (HE) glycoprotein has frequently been referred to by its approximate molecular weight, i.e., about 65,000 and 140,000 in its reduced and nonreduced forms, respectively, and more recently by the acronyms E3, H, and HA. Some of the coronaviruses do not possess a gene for HE, while some strains of others have an incomplete HE gene which is not expressed. Those coronaviruses which cause hemagglutination but do not have the HE protein (e.g., infectious bronchitis virus (IBV), transmissible gastroenteritis virus) have only poor hemagglutination activity. In contrast, those viruses with good hemagglutination activity do have the HE protein, although the presence of HE does not necessarily mean that the virus causes hemagglutination (e.g., some strains of murine hepatitis virus (MHV)). The HE of MHV exhibits amino acid homology with the HEF_1 subunit of the influenza C virus surface glycoprotein (*6*). It has been shown that the HE of bovine coro-

0042-6822/90 $3.00

navirus is a receptor-destroying enzyme with acetylesterase activity (*7*).

The Study Group has recommended that the integral membrane glycoprotein, M (previously also designated E1 (*4*)), should not be referred to as a matrix protein. The structure of this protein is substantially different from that of the matrix proteins of para- and orthomyxoviruses. The nucleocapsid (N) protein is sometimes referred to as the nucleoprotein, with the same acronym.

mRNAs

mRNAs are to be referred to by numbers 1, 2, 3 . . . in order of decreasing size. The genome-sized mRNA is therefore mRNA 1. Consequently, the mRNAs of IBV, previously denoted as F, E, D, C, B, and A (*8*), should henceforth be referred to as 1, 2, 3, 4, 5, and 6, respectively.

Coronaviruses differ in the number of their subgenomic mRNAs. It is recognized that mRNAs of different coronaviruses encoding the same protein may have different numbers; e.g., the S glycoprotein of IBV and MHV is encoded by mRNAs 2 and 3, respectively. However, this helps to highlight a major difference.

When a protein has a name and an acronym, e.g., spike S, the corresponding mRNA can be referred to as mRNA 2 (S). It might be useful to label figures in this way. On other occasions, one might refer to "mRNA 2" or the S mRNA, as appropriate, depending on the context in which the mRNA is being discussed.

When "new" mRNAs are identified, the use of numbers should be continued. For example, in all strains of murine hepatitis virus examined mRNA 2 encodes, in the 5′ ORF, a 30,000 molecular weight protein. Some strains, e.g., MHV-JHM, have more recently been shown to have an additional mRNA of a size intermediate between that of mRNA 2 and mRNA 3 (S) (*9*). This intermediate mRNA, which encodes HE, should be referred to as "mRNA 2-1." A dash (-), not a point, should be used.

GENES/ORFs

Coronavirus mRNAs form a 3′ coterminal nested set and the sequences that are absent from the next smallest mRNA are often called the "unique regions." Therefore, with the exception of the smallest mRNA, all the mRNAs are structurally polycistronic. However, it is believed (see (*2*)) that only the unique regions are translationally active. These regions may contain one or more ORFs.

TABLE 1

ACRONYMS FOR CORONAVIRUS STRUCTURAL PROTEINS

Name	Acronym
Spike glycoprotein	S
N-Terminal cleavage product	S1
C-Terminal cleavage product	S2
Hemagglutinin–esterase glycoprotein	HE
Integral membrane glycoprotein	M
Nucleocapsid protein	N

Genes/ORFs are to be referred to by letters. When the corresponding protein has a name, the acronym (uppercase) should be used, e.g., S, HE. Otherwise, the gene/ORF should be referred to by the number of the mRNA, plus a letter (lowercase) when there is more than one ORF. For example, the unique region of mRNA3 of IBV has three ORFs, 3a, 3b, 3c. MHV-A59, in contrast to MHV-JHM, lacks a complete ORF encoding the HE protein and does not generate a mRNA for HE; consequently HE is not expressed. Thus, MHV-JHM has ORF 2a and gene HE (ORF 2b), and the corresponding ORFs of MHV-A59 are referred to as ORFs 2a and 2b.

If a newly investigated coronavirus is shown to have a genome organization very similar to that of an extensively studied coronavirus, the designations of the mRNAs and genes should be based as closely as possible on those established for the previously studied virus. In this way the nomenclature will help to reflect similarities.

REFERENCES

1. SIDDELL, S. G., WEGE, H., and TER MEULEN, V., *J. Gen. Virol.* **64,** 761–776 (1983).
2. SPAAN, W., CAVANAGH, D., and HORZINEK, M. C., *J. Gen. Virol.* **69,** 2939–2952 (1988).
3. CAVANAGH, D., *J. Gen. Virol.* **53,** 93–103 (1981).
4. STURMAN, L. S., and HOLMES, K. V., *Virology* **77,** 650–660 (1977).
5. STURMAN, L. S., RICARD, C. S., and HOLMES, K. V., *Virology* **56,** 904–911 (1985).
6. LUYTJES, W., BREDENBEEK, P. J., NOTEN, A. F. H., HORZINEK, M. C., and SPAAN, W., *Virology* **166,** 415–422 (1988).
7. VLASAK, R., LUYTJES, W., LEIDER, J., SPAAN, W., and PALESE, P., *J. Virol.* **62,** 4686–4690 (1988).
8. STERN, D. F., and SEFTON, B. M., *J. Virol.* **50,** 22–29 (1984).
9. SHIEH, C.-K., LEE, H.-J., YOKOMORI, K., LA MONICA, N., MAKINO, S., and LAI, M. M. C., *J. Virol.* **63,** 3729–3736 (1989).