===== REVIEWS =====

Genetic Engineering Systems to Study Human Viral Pathogens from the Coronaviridae Family

S. O. Galkin[a, b], A. N. Anisenko[a, b, c], O. A. Shadrina[b, c], and M. B. Gottikh[b, c], *

[a] Bioengineering and Bioinformatics Department, Moscow State University, Moscow, 119991 Russia
[b] Chemistry Department, Moscow State University, Moscow, 119991 Russia
[c] Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow, 119991 Russia
*e-mail: gottikh@belozersky.msu.ru
Received May 12, 2021; revised June 9, 2021; accepted June 14, 2021

Abstract—The COVID-19 pandemic caused by the previously unknown SARS-CoV-2 *Betacoronavirus* made it extremely important to develop simple and safe cellular systems which allow manipulation of the viral genome and high-throughput screening of its potential inhibitors. In this review, we made an attempt at summarizing the currently existing data on genetic engineering systems used to study not only SARS-CoV-2, but also other viruses from the Coronaviridae family. In addition, the review covers the basic knowledge about the structure and the life cycle of coronaviruses.

INTRODUCTION

COVID-19 (Coronavirus infectious disease-2019), an infectious disease caused by a new strain of betacoronavirus, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), betacoronavirus, has already claimed millions of lives worldwide. Due to the high pathogenicity and contagiousness of the pathogen, the COVID-19 pandemic was declared by the World Health Organization. Research on SARS-CoV-2 is actively pursued in scientific centers all around the world, extensive work has been done to create vaccines, and significant efforts are aimed at finding drugs for the effective therapy and prevention of COVID-19.

Over the past 20 years, coronaviruses have been the etiological agents of three large-scale outbreaks of severe human respiratory infections: severe acute respiratory syndrome (SARS) in 2002, Middle East respiratory syndrome (MERS) in 2012, and the 2019 coronavirus infection (COVID-19), which affected almost every country and caused a worldwide health crisis [1, 2]. SARS-CoV-2 differs from SARS-CoV-1 and MERS-CoV in its long incubation period, high percentage of asymptomatic carriers, and ability to be transmitted from person to person in the absence of clinical symptoms in infected individuals [3]. This set of features of SARS-CoV-2 led to a local outbreak in Wuhan, China, leading to the pandemic. The lethality of COVID-19 strongly depends on the age and sex of the individual: the probability of lethality is highest in elderly people, especially males [4]. A significant correlation has been found between the severity of COVID-19 disease and concomitant chronic diseases or excess body weight [5].

Success in the study of highly pathogenic coronaviruses and, in particular, in the search for potential drugs for the therapy of the diseases they cause largely depends on understanding the peculiarities of the virus infectious cycle and, consequently, on the development of technologies used for research. These are mainly genetically engineered systems, which are numerous and varied, and each has its own advantages and disadvantages. The present review attempts to systematize the currently available systems for studying coronaviruses, and details the strengths and weaknesses of each of them.

CORONAVIRUS BIOLOGY

Members of the Coronaviridae family infect mammals and birds, causing respiratory diseases of varying severity. Until 2019, 6 coronaviruses circulated in the human population, of which *Betacoronavirus* 1

Abbreviations: ACE2, angiotensin-converting enzyme; ANPEP, aminopeptidase N; BAC, bacterial artificial chromosome; DI-RNA, defective interfering RNA; ERGIC, endoplasmic reticulum, Golgi intermediate compartment; HDV, hepatitis delta virus; NSP, non-structural proteins; ORF, open reading frame; RdRp, RNA-dependent RNA polymerase; TAR-cloning, transformation-associated recombination cloning; TGEV, transmissible gastroenteritis virus; TRS, transcription regulatory sequence; VLPs, virus-like particles; UTR, untranslated region.
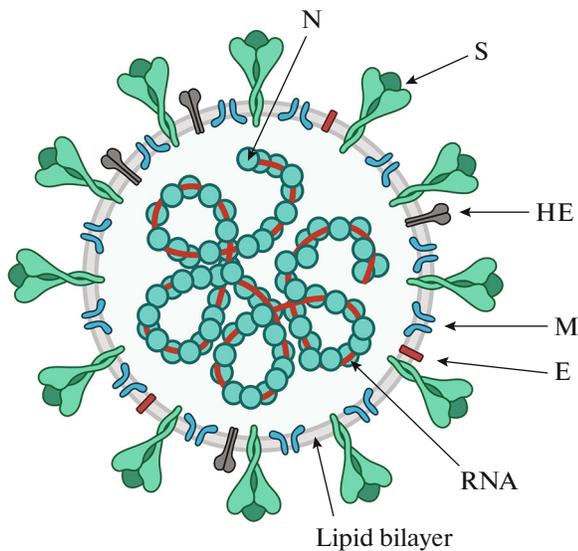
Fig. 1. Schematic representation of the Coronaviridae family virion structure. The helical symmetry nucleocapsid made up by the nucleocapsid (N) protein and one RNA genome molecule is surrounded by the supercapsid of cellular origin with embedded virus S, M, and E proteins. The hemagglutinin esterase (HE) protein is included in the virions of the human coronaviruses HCoV-OC43 and HCoV-HKU1, but is not found in HCoV-229E, HCoV-NL63, MERS-CoV, SARS-CoV-1, and SARS-CoV-2. The coronavirus virion contains about 300 S protein molecules, 2000 M protein molecules, 1000 N protein molecules, and 100 E protein molecules [11].

(HCoV-OC43), HCoV-229E, HCoV-NL63, and HCoV-HKU1 are considered seasonal respiratory viruses. The coronavirus associated with severe acute respiratory syndrome (SARS-CoV, recommended name from 2020: SARS-CoV-1; genus *Betacoronavirus*, subgenus *Sarbecovirus*) and the Middle East respiratory syndrome coronavirus (MERS-CoV; genus *Betacoronavirus*, subgenus *Merbecovirus*) that emerged in 2002 have proven highly pathogenic to humans. In 2019, humanity encountered another strain in the subgenus *Sarbecovirus* called SARS-CoV-2 [6]. This section describes the main details of the structure and life cycle of the members of the *Coronaviridae* family in general and the seven human pathogens in particular.

VIRION STRUCTURE

Coronaviruses are enveloped (+)RNA viruses belonging to group IV according to the Baltimore classification [7]. The average virion diameter is 80–120 nm. The virions are spherically shaped with characteristic spikes formed by the S protein, resembling a stellar corona. The schematic structure of a coronavirus virion is shown in Fig. 1. On the outside, the virion is covered by a supercapsid, a cell membrane derivative with integrated viral proteins. The following viral proteins can be found in the supercapsid of most human coronaviruses: the spike-forming glycoprotein (S) which interacts with the host cell surface and initiates the entry of virion components inside the cell; the membrane protein (M) which facilitates the interaction of virion components with each other and is crucial for successful assembly of viral particles [8], and the envelope protein (E), which forms pentameric transmembrane channels and acts as a viroporin [9]. The incorporation of the E-protein into the virion is not strictly necessary for the formation of infectious viral particles, but viruses which lack it are attenuated relative to the wild-type viruses [8]. HCoV-OC43 and HCoV-HKU1 viruses additionally contain a hemagglutinin-esterase (HE) protein in the envelope, which disrupts the S-protein receptor and thus promotes more efficient budding of new virions [10].

Inside the outer envelope there is a helical symmetry nucleocapsid consisting of N protein and genomic (+)-chain RNA. N protein is bound to the viral RNA and carries out structural functions including packaging and stabilization of the viral genome. In addition, N protein modulates the host cellular immune response to viral infection [8], and promotes more efficient synthesis of 3'-distal genes [12].

GENOME STRUCTURE

The genomic RNA of coronaviruses is unsegmented, capped at the 5'-end, and polyadenylated at the 3'-end. The distinctive feature of Coronaviridae genome is its high length of about 26–32000 bases [13], which is unique for RNA-containing viruses. On average a single-stranded unsegmented RNA in human viruses does not exceed 10–12000 bases [14]. It is supposed that the increased genome size of coronaviruses as compared to other viruses is due to the presence of editing activity in the RNA-dependent RNA polymerase (RdRp) complex, which reduces mutation rate [15].

Generally, the coronavirus genome can be subdivided into two principal parts. The first one contains the open reading frame (ORF) ORF1a/b, which encodes proteins mainly responsible for replication and transcription. The second part encodes structural and accessory proteins responsible for interaction with the cell surface and entry into the cell, as well as for adaptation to the host. The results of comparative analysis of the genome sequences of various coronaviruses confirm the functional significance of this division: in coronaviruses, the ORF1a/b conservation level is higher than that of the rest genome [16].

The number of ORFs in the genome differs in different coronavirus species and ranges from 6 to 11 [17]. The first ORF, ORF1a/b, makes up about 2/3 of the genome and encodes two polyproteins, pp1a and pp1ab. The pp1ab polyprotein is formed as a result of a ribosomal reading frame shift of one nucleotide backward (−1) just upstream of the ORF1a stop

Table 1. Coronavirus nonstructural protein functions

Protein name	Key functions
NSP1	Represses the translation of cellular mRNAs by binding the 18S rRNA of the 40S ribosome subunit [26]; promotes cellular mRNA cleavage in 5'UTR [27]
NSP2	Presumably, regulates cell survival signaling pathways; interacts with the prohibitin 1 and prohibitin 2 cell proteins [28]; not necessarily required for a successful infection [29]
NSP3	Processes pp1a and pp1ab, in particular NSP1−NSP4, owing to the presence of the papain-like protease-2 domain [30]; is a pore component in double-membrane-bound organelles which carry out viral genome and subgenomic RNA export into the cytoplasm [31]; possesses deubiquitination activity; cleaves mono-ADP-ribose from proteins thus suppressing cell interferon response [30]
NSP4	Participates in double-membrane-bound replicative organelle formation [32]
NSP5	3-Chymotrypsin-like protease, processes pp1a and pp1ab, in particular, NSP4−NSP11/NSP4−NSP16 [33]
NSP6	Prevents viral protein-containing vesicle fusion with lysosomes; participates in double-membrane-bound replicative organelle formation [34]
NSP7	Forms a hexadecameric complex with NSP8, which apparently works as a processivity factor [35] and primase [36, 37]
NSP8	Forms a hexadecameric complex with NSP7, which apparently works as a processivity factor [35] and primase [36, 37]; inhibits protein embedding into cell membrane via complex formation with signal recognition particle (SRP) 7SL RNA and 28S rRNA [26]
NSP9	Inhibits protein embedding into the cell membrane via complex formation with SRP 7SL RNA [26]; RNA polymerase complex component [19]
NSP10	Required for (−)-chain RNA synthesis during virus replication [38]; involved in pp1a/ab processing [39]; regulates the activity of the RdRp complex components NSP14 and NSP16 [40, 41]
NSP11	Short peptide formed in the course of pp1a processing. Function unknown [19]
NSP12	RdRp component with polymerase activity [19]
NSP13	Possesses helicase activity [42]
NSP14	Increases the accuracy of viral RNA synthesis due to the presence of proof-reading 3'→5' exonuclease activity [43]; participates in 5'-capping of viral RNAs, possesses N7-methyl transferase activity [44]
NSP15	Inhibits the activity of cellular double-stranded RNA (dsRNA) sensors; possesses functionally relevant endoribonuclease activity, with an unknown precise mechanism of action [45]
NSP16	Disturbs splicing through binding with U1/U2 small nuclear RNA [26]; participates in 5'-capping of viral RNAs, possesses 2'-O-methyl transferase activity [40]

codon. pp1a/pp1ab encodes non-structural proteins (NSPs) responsible for replication and transcription, suppression of cellular mRNA synthesis, and modulation of the cellular immune response [18].

The analysis of a large body of data revealed that many NSPs simultaneously perform several functions in the viral replicative cycle, which allows the virus to cope with a wide range of tasks, from its own replication to counteracting the cellular immune response, without further enlarging the genome [19−22]. Table 1 provides up-to-date information on the most studied functions of NSPs. Several functional blocks can be distinguished among them: proteins responsible for the formation of the RdRp complex (NSP7−NSP16), proteins involved in the formation of double membrane-bound replicative organelles (NSP3−NSP6), proteins involved in the processing of pp1a and pp1ab polyproteins (NSP3, NSP5), and those involved in modulating cellular processes (NSP1, NSP2). The rest of the genome encodes the necessary structural proteins S, N, E, M, (HE) as well as a number of species-specific accessory proteins [23, 24]. It is believed that accessory proteins are not essential for viral replication, although they contribute to more efficient virion assembly both by suppressing intracellular immunity and by acting as scaffold proteins in virion assembly [25].

A schematic representation of the genome structure of human-infecting Coronaviridae is shown in Fig. 2. It should be noted that for all coronaviruses, the order of structural protein genes: ORF1a/b→S→E→M→N is conservative, but apparently does not carry any functional load, since the infectivity of the resulting viruses did not change when gene order was rearranged [46].
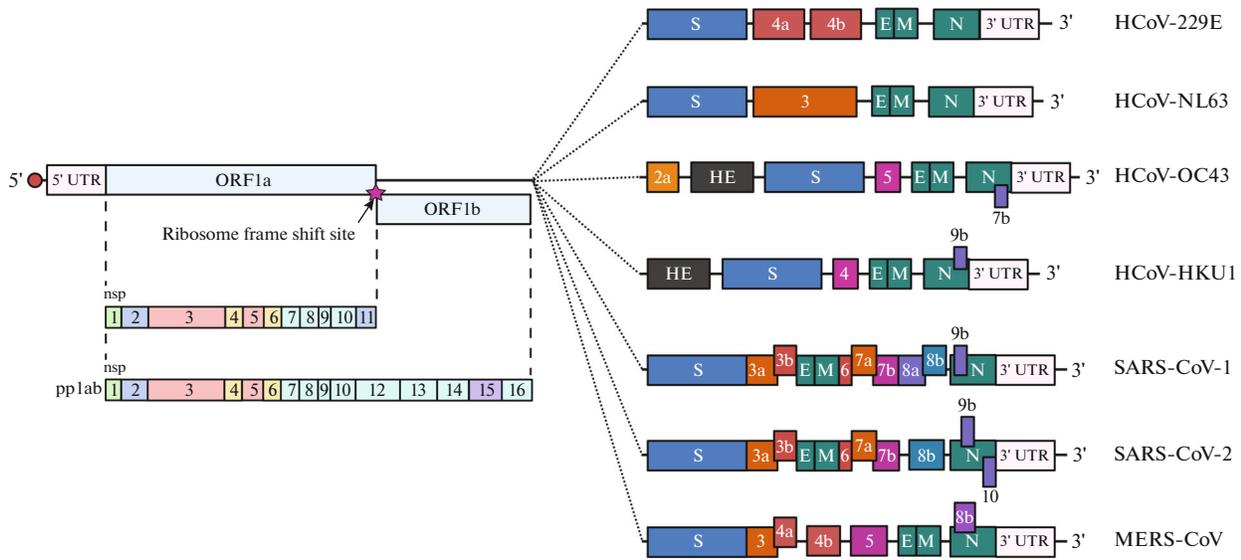
Fig. 2. Genome structure of coronaviruses circulating in human populations. In the figure, encoded virus proteins are indicated in capital letters to the right, virus-specific ORF nos. are indicated with numbers and small letters. UTR—untranslated region. Numbers to the left indicate encoded NSP nos.

LIFE CYCLE

Cell Entry

The life cycle of coronaviruses begins with virion binding to receptors on the host cell surface (Fig. 3). The viral surface glycoprotein S plays a key role in this process. It is this protein, the variability of which is higher than that of other viral proteins [16], which determines the species and tissue tropism of the pathogen. On the cell side, various surface components take part in the interaction with viral particles [47]. The HCoV-229E coronavirus uses aminopeptidase N (ANPEP) as a receptor. It is a metalloproteinase present on the surface of intestinal, lung, and kidney epithelial cells. In the intestine, this enzyme is involved in the hydrolysis of peptides produced by gastric and pancreatic proteases. The function of ANPEP in the lungs and kidneys is not fully established; most likely, it metabolizes various regulatory peptides [48]. For HCoV-NL63, the situation with cell receptors is still unclear. The S protein of this virus has been shown to interact with angiotensin-converting enzyme-2 (ACE2) [49], and the M protein binds to heparan sulfated proteins on the cell surface [50]. This M protein interaction may be a complementary factor for viral entry or increase the local concentration of virus on the cell surface. When S protein was removed from the studied virions, they retained the ability to infect cells even when ACE2 was blocked with antibodies, from which it could be concluded that there are alternative pathways for HCoV-NL63 entry into the cell, for example through the above-mentioned interaction of M protein with heparan sulfated proteins [50]. The HCoV-OC43 and HCoV-HKU1 viruses use 9-O-acetylated neuraminic acid-containing glycoproteins as a receptor [25]. These viruses, as mentioned earlier, have a hemagglutinin-esterase protein in their virion, which promotes more effective interaction with the cell surface and budding of new viral particles from the infected cell [51]. MERS-CoV uses dipeptidyl peptidase-4 (DPP4) as a receptor, a surface dimer protein that normally processes hormones and chemokines [52]. SARS-CoV-1 and SARS-CoV-2 use ACE2 as the main receptor [25], although SARS-CoV-2 also uses heparan sulfated proteins as a cofactor, interaction with which causes conformational changes in the S protein that contribute to its more efficient binding to ACE2 [53].

After binding to the receptor, the S protein is processed by the surface cellular proteases and undergoes conformational changes contributing to the fusion between the viral and cellular membranes. This mode of entry is called early entry. If surface proteases are for some reason unavailable, the late pathway of entry into the cell is activated. In this case, the virus first undergoes endocytosis. In the endosome, the cellular protease cathepsin L is activated under low pH conditions, which carries out S protein proteolysis, causing viral and endosomal membranes to fuse [55]. After this, the viral nucleocapsid consisting of genomic RNA and N protein appears in the cytoplasm, and the processes of viral transcription and replication begin.

Replication and Transcription

As mentioned above, coronavirus RNA contains a large number of ORFs (from 6 to 11) [24], which leads to certain difficulties in translating 3'-distal genes. Coronaviruses address this problem by forming subge-
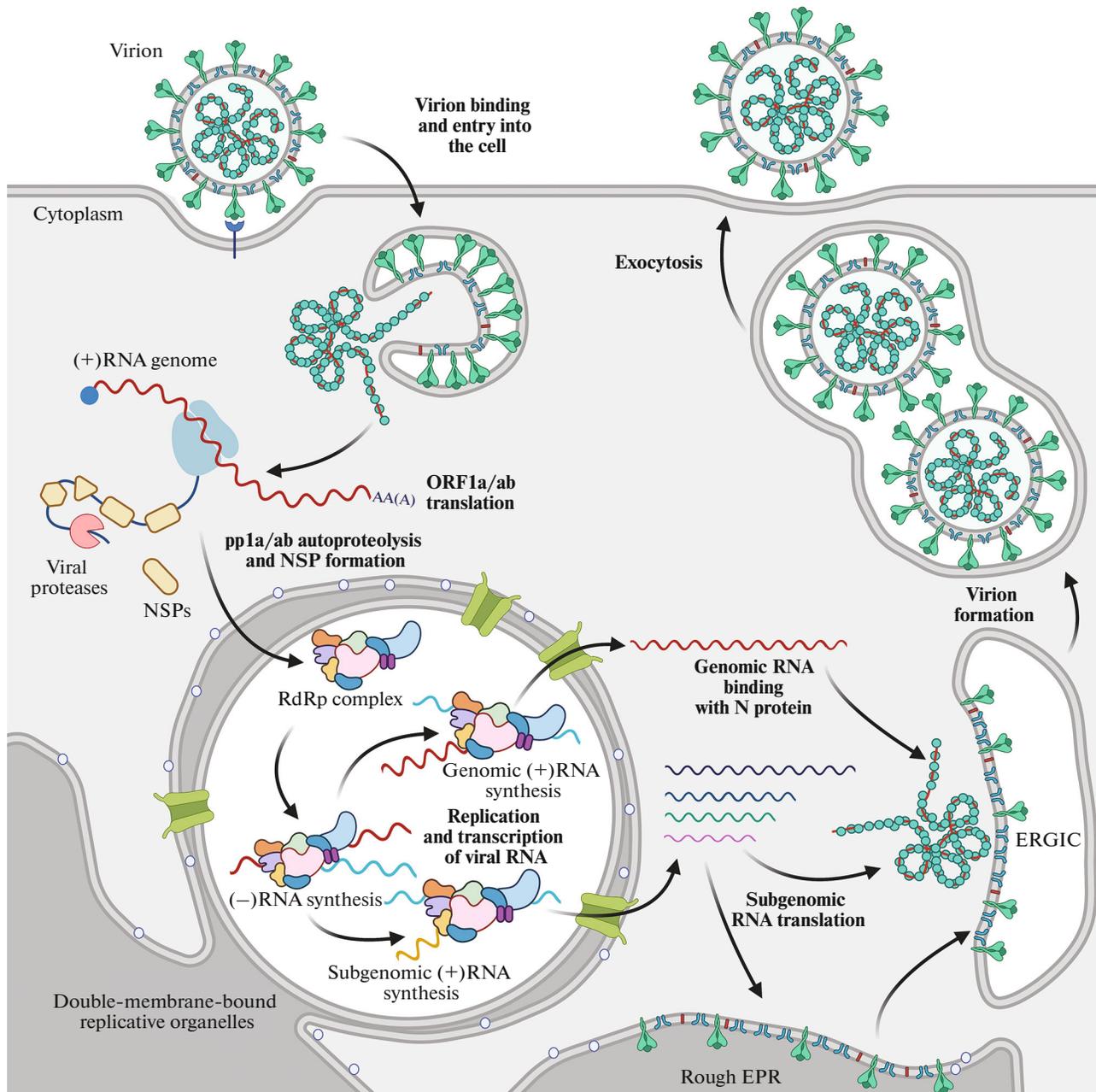
Fig. 3. Schematic representation of Coronaviridae family viruses. Infection starts with the virus particle binding with its receptor on the host cell surface which leads to virus and cell membrane fusion (early entry pathway), or virion encapsulation into endosomes and further virus and cell membrane fusion (late entry pathway). Viral RNA in complex with N protein enters the cytoplasm. ORF1a/b first open reading frame translation results in the pp1a and pp1ab polyproteins which further undergo autoproteolysis to obtain NSP2s. NSPs form the RdRp complex which carries out viral RNA replication and transcription. This process takes place in the double-membrane-bound organelles which are formed from the rough endoplasmic reticulum in the presence of pp1a and pp1ab components. Replication/transcription results in full-size and subgenomic viral RNAs. Structural and accessory proteins are translated from subgenomic RNAs. Structural proteins are assembled into the intermediate structure in the endoplasmic reticulum−Golgi intermediate compartment (ERGIC) which then form new virions together with the N protein/genomic viral RNA complex. Newly assembled virions are accumulated in the intracellular vesicles, and then leave the infected cell by exocytosis [modified from 54].

nomic RNAs [56, 57]. Among (+)RNA viruses, two main mechanisms for their synthesis are found: through the presence of RdRp internal binding sequences in the genome (Caliciviridae, Bromoviri-dae, Virgaviridae, Togaviridae families) or of premature transcription termination sites (Nodaviridae, Togaviridae, Roniviridae, Tombusviridae families) (Figs. 4a and 4b) [58]. However, in case of coronavi-

ruses, the process of subgenomic RNA synthesis is discontinuous [40, 41, 43]: during the synthesis of the (−)RNA chain, the RdRp-complex dissociates from the (+)RNA matrix at certain sites, binds in the 5'UTR region, and continues RNA synthesis (Fig. 4b) [59]. The positions of dissociation and subsequent binding of RdRp are determined by transcription regulatory sequences (TRS). In some cases RdRp skipping is possible during (+)RNA-chain synthesis: in this case RdRp dissociates from TRS located in 5' UTR and re-binds with TRS located upstream of the start of the protein-coding sequence (Fig. 4b) [60]. The shortened RNAs obtained in the process of discontinuous transcription serve as a template for subsequent translation of viral proteins. Interestingly, it is the process of interrupted transcription that allows the virus to use specific mechanisms of cellular translation repression without affecting the translation of viral mRNAs. It is believed that the presence of the leader sequence within all viral mRNAs protects them from degradation by NSP1-protein, while cellular mRNAs are cleaved by it near their 5'-end [56]. If the RdRp-complex does not dissociate during RNA chain elongation, full-length RNA complementary to the genomic RNA is synthesized, and genomic (+)RNAs are then synthesized using it as a template, which are further used for the assembly of new viral particles and the synthesis of pp1a and pp1ab polyproteins. It has been shown that active synthesis of viral proteins starts as early as five hours after infection [61]. At the same time, the content of viral RNA in the cell significantly increases, which is caused not only by replication and discontinuous transcription, but also by the suppression of cellular mRNA synthesis by the viral NSP1 protein [62].

Virion Assembly and Budding

A distinctive feature of the replication cycle of coronaviruses is the new virions' assembly site. Most enveloped viruses are assembled in the cell plasma membrane, whereas coronavirus virions are assembled in the intermediate compartment between the endoplasmic reticulum and the Golgi, which is referred to as the endoplasmic reticulum−Golgi intermediate compartment (ERGIC). It was shown that after translation the structural S, M, E, and N proteins accumulate in ERGIC [63]. Genomic (+)RNA binds to the N protein to form the nucleocapsid. M protein participates in the recognition of a special genome packaging signal (package sequence, PS) in the viral RNA [64]. The ability of N protein to recognize PS depends on the virus species, for example, the packaging of mouse hepatitis virus (MHV) genomic RNA occurs via the N-independent mechanism [65], whereas that of the SARS-CoV-1 genome is N-dependent [66]. The packaging signal is located within ORF1a/b, which ensures that only full-length RNAs are incorporated into newly assembled virions [67, 68].

The assembly center of new viral particles is the M protein, which interacts with the S, N, and E proteins and genomic (+)RNA [69]. After M, S, E proteins and the N-protein−RNA complex appears in ERGIC, mature viral particles are assembled and budded. The assembled virions accumulate in intracellular vesicles, which, in turn, are released into the extracellular space as a result of exocytosis or lysis of the infected cell [70].

EXPERIMENTAL SYSTEMS TO STUDY CORONAVIRUS REPLICATION

Success in the study of highly pathogenic viruses, which include some coronaviruses, and in the search for potential drugs for their therapy depends on the development of technologies that allow manipulation of viruses and, in particular, of their genome. Traditionally, full-length viral systems, i.e., infectious virus particles, are used to study viral replication. The advantages of such systems include maximum correspondence to the real infection conditions. During infection, viral proteins come into contact with each other and with a multitude of cellular factors, forming complex networks of protein interactions that are recreated in these systems. However, the use of infectious viruses carries a high risk of infecting the operator, so only certified laboratories with appropriate equipment are allowed to work with them. Moreover, in the case of RNA-containing viruses, when working with full-size virus systems, it is quite difficult to obtain virus variants mutant for the genes of interest. In the case of coronaviruses, manipulations with their genomes are also extremely difficult due to the large size of their genomic RNA, which, as mentioned above, ranges from 26 to 32 000 bases [13]. In this context, much attention is paid both to the modification of full-length viral systems and to the development of simpler and safer systems that allow the study of individual stages of the coronavirus life cycle and searching for their inhibitors. We will consider in detail the currently available variants of modified full-length viral systems that can be used to obtain viral RNA containing the necessary mutations, as well as the safer, although often less informative, pseudoviral systems and replicons.

MODIFIED FULL-SIZE VIRUS SYSTEMS

Two approaches are used to obtain variously modified viral coronavirus genomes: direct modification of genomic RNA (defective interfering RNA, DI-RNA) or obtaining cDNA containing the necessary modifications, which is then used for synthesis of genomic RNA and subsequent assembly of viral particles.

DI-RNA-Systems

The study of coronavirus replication mechanisms began with the use of viral DI-RNAs containing only parts of viral genome: viral regulatory sequences
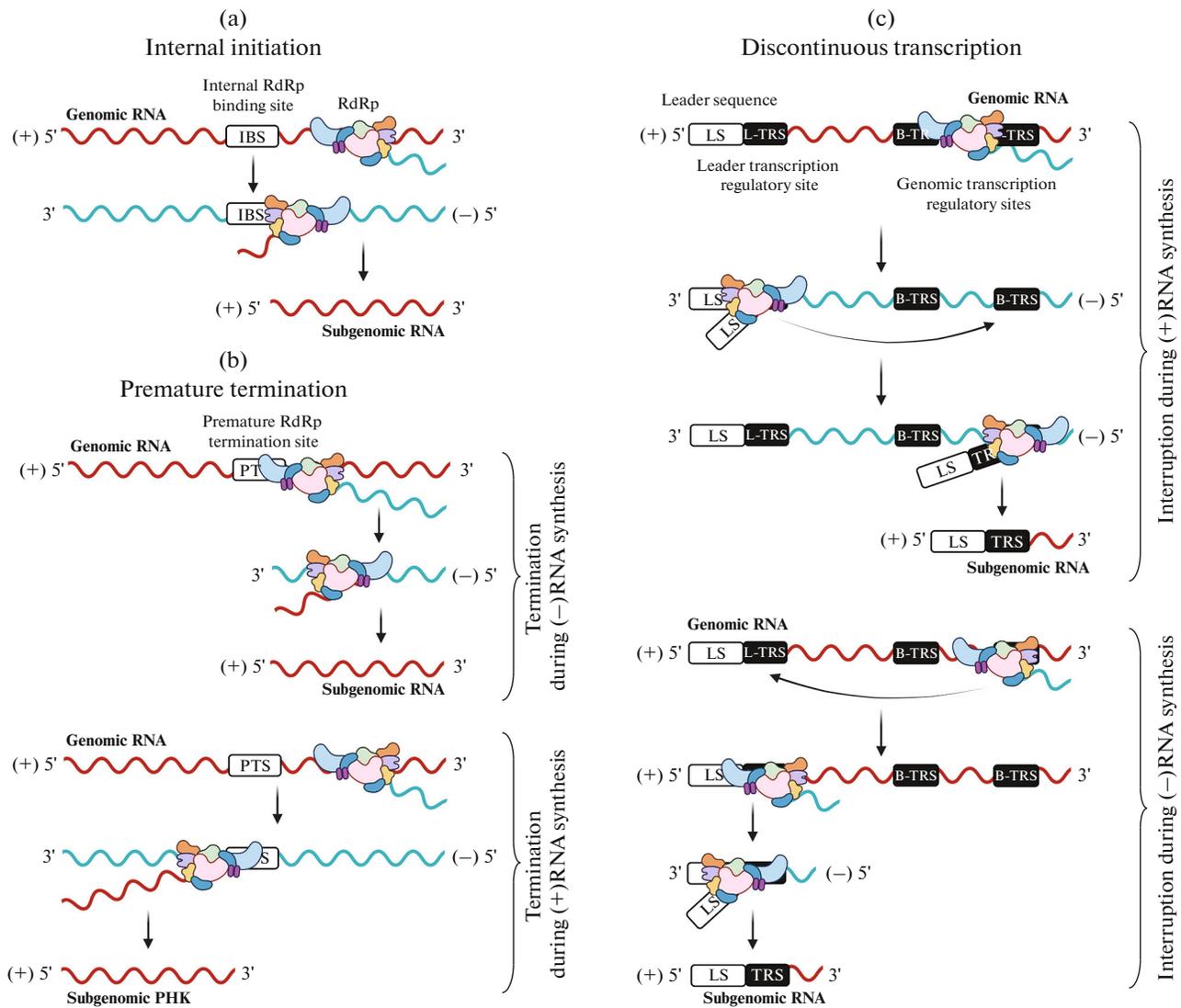
Fig. 4. Schematic representation of coronavirus subgenomic RNA synthesis strategies. (a) Internal initiation model. IBS—internal RdRp binding site. (b) Premature transcription termination model. Termination may occur both during (+)-chain and (−)-chain RNA synthesis. PTS—premature termination site for RdRp. (c) Discontinuous transcription model. Transcription may be interrupted both during (+)-chain and (−)-chain RNA synthesis [58]. LS—leader sequence; TRS—transcription regulatory sequence; L-TRS—leader TRS, and B-TRS—TRS in the coding region.

required to maintain RNA replication and sequences encoding parts of viral proteins. Such RNAs are capable of replication, but exclusively in cells infected with the corresponding coronavirus, and in some cases they are also packaged into non-infectious virus-like particles (VLP) [67, 71, 72] (Fig. 5). In this way, new variants of viral particles carrying the necessary modifications can be obtained. DI-RNAs can also contain a reporter gene and if genome replication and transcription occur the efficiency of viral enzymes can be assessed using such systems. This is one of the first and most primitive systems, and with it a large body of data on the replication and packaging mechanisms of coronavirus genomes has been obtained. For example, the packaging signals of MHV [73−75], transmissible gastroenteritis virus (TGEV) [67], and SARS-CoV-2 [68] were determined using DI-RNA. M protein interaction with PS RNA has been demonstrated for MHV using this system [64]. Using SARS-CoV-2 as an example, it was shown that when the virus-infected cells are transfected with defective RNA, it begins to actively replicate and become incorporated into virions resulting in a reduction in infectious viral particle production by approximately half 24 h after DI-RNA transfection. Based on these data, it was suggested that DI-RNA could be used as a therapeutic agent to reduce the viral load [68].
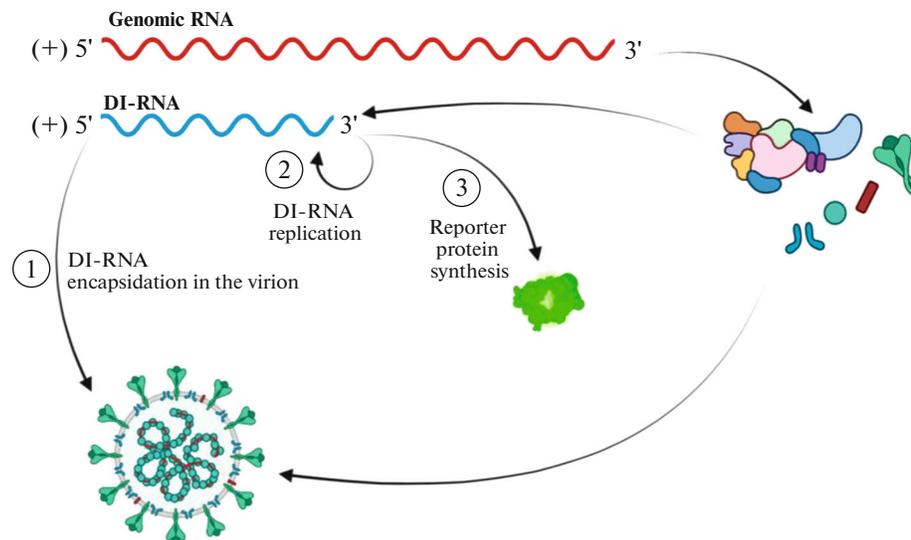
Fig. 5. DI-RNA system functioning principle. In coronavirus-infected cells, DI-RNAs may be included in the virion (1), and/or replicated (2), and/or express reporter proteins (3) in the presence of viral proteins translated from genomic coronavirus RNA.

Infectious cDNA Inserted in Artificial Bacterial Chromosomes

The development of molecular cloning methods made it possible to assemble full-length viral cDNAs part-wise. For the first time, full-length cDNA of the coronavirus genome was obtained for TGEV by cloning into an artificial bacterial chromosome (BAC) [76]. As part of the BAC, the viral cDNA was put under the control of the cytomegalovirus (CMV) promoter, and at the 3'-end it was bordered by the poly(A) sequence, the hepatitis delta virus (HDV) ribozyme sequence, and the BGH transcription terminator (Fig. 6a). HDV ribozyme catalyzes RNA chain breakage between the last nucleotide of the coronavirus genome and the first nucleotide of the ribozyme, which makes it possible to obtain RNA containing a 3'-end sequence identical to the viral 3'-end sequence [77]. This precise matching is necessary due to the fact that the 3'UTR sequence of coronaviruses is crucial for successful replication and transcription of viral RNA [78]. After BAC transfection, full-genome viral RNA and proteins are synthesized in cells and assembled into infectious viral particles. Using BAC to work with the cDNA of coronavirus genomes has several advantages. First, it becomes possible to efficiently generate unlimited amounts of the required cDNA in *Escherichia coli*. Second, it is possible to transfect mammalian cells with this bacterial chromosome with high efficiency, which allows intracellular expression of viral RNA without resorting to more complicated techniques for its production, such as in vitro transcription. Third, the BAC sequence is relatively easy to modify using Red-recombinase and restriction endonuclease I-SceI, which makes it possible to modify the virus genes under study [79].

Subsequently, the approach of cloning viral cDNA in BAC, in addition to TGEV, was applied to create infectious clones of the following human pathogens: HCoV-OC43 [80], MERS-CoV [81], SARS-CoV-1 [82], and SARS-CoV-2 [83] (Table 2).

Infectious cDNA Obtained Using TAR-Cloning

Transformation-associated recombination (TAR) is a cloning method that exploits the specific features of *Saccharomyces cerevisiae* cells, in which homologous recombination of overlapping DNA fragments occurs with high frequency (Fig. 6b) [91]. The use of yeast cells to obtain and generate full-length coronavirus cDNA has several advantages over the bacterial system. First, long DNA sequences are generally less stable in bacteria than in yeast. Second, it has been shown that sequence fragments of the gene encoding ORF1a/b can be toxic to bacteria [76]. Third, the high efficiency of TAR cloning makes it possible to quickly and without great difficulty introduce mutations into the viral genome: all that is required is to make changes in one or more transformed fragment. This method has been used to assemble and modify the long genomes of the DNA viruses, CMV (genome size ~236 kbp) [92, 93], and herpes simplex virus (genome size ~152 kbp) [94, 95]. As for the human pathogens in the *Coronaviridae* family, cDNAs of MERS-CoV, HCoV-229E, HCoV-HKU1, and SARS-CoV-2 genomes were obtained by TAR-cloning (Table 2) [84]. Note that the SARS-CoV-2 cDNA was obtained by this method in just two months [84].
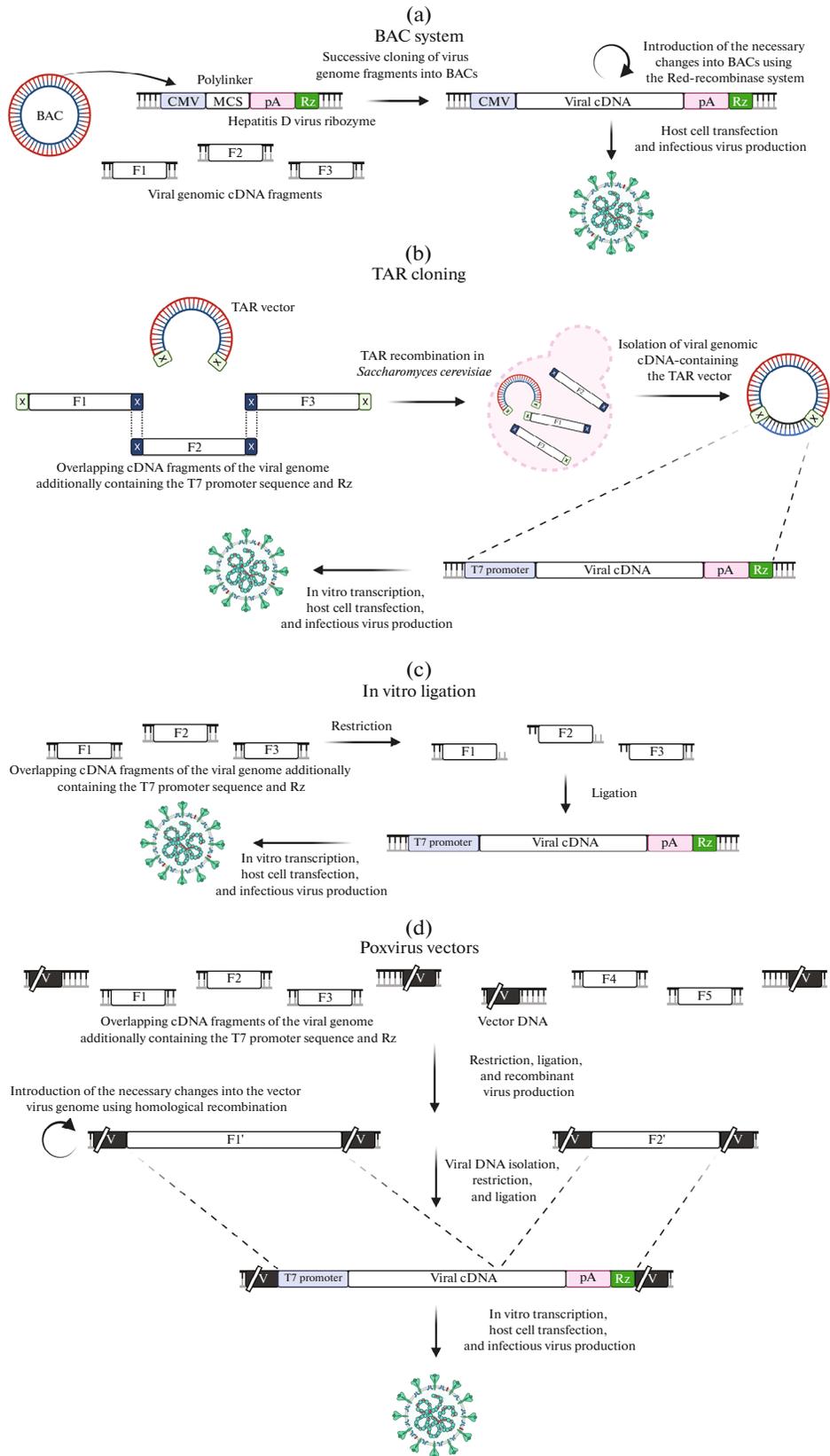
Fig. 6. Schematic representation of full-size virus cDNA production using artificial bacterial chromosomes (a), TAR cloning technique (b), in vitro ligation technique (c), or recombinant poxviruses (d). CMV—cytomegalovirus promoter; pA—poly(A) tail; Rz—HDV ribozyme, and MCS—polylinker.

Table 2. Approaches used to obtain full-size coronavirus cDNAs capable of human infection

Virus	Approach used to obtain cDNA			
	BAC	TAR-cloning	In vitro ligation	Poxvirus vectors
HCoV-OC43	+ [80][a]	−	−	−
HCoV-229E	−	+ [84]	−	+ [85]
HCoV-NL63	−	−	+ [86]	−
HCoV-HKU1	−	+ [84]	−	−
MERS-CoV	+ [81]	+ [84]	+ [87]	−
SARS-CoV-1	+ [82]	−	+ [88]	+ [89]
SARS-CoV-2	+ [83]	+ [84]	+ [90]	−

[a] The reference is indicated in brackets.

Infectious cDNA Obtained by In Vitro Ligation

Another approach to obtain coronavirus genomes is to assemble a full-length viral cDNA from smaller fragments flanked by unique restriction sites. Sequences recognized by type I or type III restriction endonucleases can be used as such sites. In the latter case restriction sites are lost after ligation, so it becomes impossible to further modify the cDNA, because the recognition site and the break introduction site are spatially separated in the case of type III restriction endonucleases. In vitro ligation can use both restriction sites that already exist in the genome under study and those artificially created. In this case restriction sites are introduced using synonymous substitutions that do not change the amino acid sequence of viral proteins. Also, during cDNA preparation, the T7 promoter sequence required for RNA production in a cell-free system using T7 RNA polymerase is introduced into the start of the genome thus assembled [88, 96] (Fig. 6c). The resulting genomic RNA is delivered to the target cells by transfection or electroporation, followed by viral transcription and replication processes, as well as the assembly of infectious viral particles. The disadvantages of this method include the need to modify the native nucleotide sequence of the viral genome in some cases to introduce additional sites for restriction endonucleases or to remove random T7 terminator sequences found in the coronavirus genome. However, the use of type III restriction endonucleases, which recognize asymmetric DNA sequences and make cuts at a certain distance from the biding site minimizes the need to introduce synonymous substitutions. An undoubted advantage of this approach is the ability to fragment the viral genome in such a way as to break the genes of proteins that exhibit toxicity when produced in bacteria into separate fragments. Full-length cDNAs of human HCoV-NL63 [86], SARS-CoV-1 [88], MERS-CoV [87], and SARS-CoV-2 [90] coronaviruses have been obtained using in vitro ligation (Table 2).

Infectious cDNA Obtained Using Poxvirus-Based Vectors

An alternative strategy for obtaining and modifying infectious cDNAs could be the use of vectors based on the smallpox virus (Poxviridae family). A full-length coronavirus cDNA with a size of about 30 000 bp can easily be produced as part of the smallpox virus genome, which has an extremely large size (about 200 000 bp). As part of this approach, the first step involves in vitro ligation of overlapping coronavirus cDNA fragments to obtain full-length cDNA, which is incorporated into the poxvirus vector after a purification step (Fig. 6d). However, in some cases, it is possible to clone several large cDNA fragments into separate vectors [85]. For example, in the case of obtaining SARS-CoV-1 cDNA from eight initial fragments, two poxviral vectors were obtained that contained fragments 1–20288 and 20272–29727 of the SARS-CoV-1 genome; the first vector also included the T7 promoter sequence. This was followed by another in vitro DNA ligation of the obtained recombinant viruses followed by in vitro transcription and transfection of eukaryotic cells with the generated RNA to obtain full-fledged infectious SARS-CoV-1 viral particles [89]. The introduction of an additional step using poxviral vectors, although it complicates the described in vitro ligation approach, still has several advantages. For example, the genomes of recombinant poxviruses are stable and efficiently replicate in cell culture, which makes it possible to overcome the problem of toxicity of some coronavirus sequences to bacteria [97]. It is also important that there are approaches that make it possible to efficiently introduce mutations into the genome of recombinant poxviruses using homologous recombination, and this leads to great opportunities for modifying the viral genes under study. Coronavirus HCoV-229E [85] and SARS-CoV-1 [89] cDNAs have been obtained using recombinant poxviruses (Table 2). However, this approach is rather complicated, and perhaps this can explain the absence of articles that use poxvirus-based vectors to obtain full-length SARS-CoV-2 cDNA.

PSEUDOVIRUS SYSTEM

As an alternative to full-length viruses, so-called pseudoviruses are often used in research. Pseudoviruses, also referred to as VLPs, are replication incompetent versions of original viruses. In general terms, a pseudoviral system consists of at least two DNA fragments: one encodes a viral genome from which one or more genes encoding important structural viral proteins have been deleted, and a reporter gene has been inserted in their place, allowing the effects on viral replication to be evaluated in cells; the second DNA fragment encodes the missing segment of the genome. Joint expression of the two fragments in special packaging cells leads to the assembly of viral particles capable of infecting the target cells and reproducing certain stages of the coronavirus life cycle, but unable to produce new viral particles in transduced cells because they contain a shortened genome. The undoubted and most important advantage of pseudoviruses is their safety: they do not require any special approvals and the appropriate levels of protection to work with them are lower.

Pseudoviruses are used in both basic and applied research. VLP-based antiviral vaccines have already been developed (see review [98]). They are also used as vectors for targeted drug delivery and gene therapy [99], as a positive control in various test-systems for diagnosis of viral infections [100], and as models to study various stages of the virus life cycle under near-physiological conditions.

Many different pseudoviral systems have been developed for pathogens in the family *Coronaviridae*. Most of the studies using VLPs are devoted to various aspects of the structure of coronaviruses, as well as the mechanisms of virion entry into the cell and detachment from it. Pioneering studies have been performed on MHV [101]. It was shown that the expression of only two viral proteins, M and E, is required for the assembly and efficient budding of pseudoviral particles. The S protein, although included in the VLP, was not necessary for the production of viral particles. It was surprising that N-protein not only did not affect the assembly of pseudoviral particles, but also was not included in their composition when co-expressed with M and E proteins [65]. It is worth noting that such unusual results are consistent with the data that the MHV genomic RNA packaging signal interacts with M rather than N protein [64]. VLPs were also obtained for HCoV-NL63. The minimum set of structural proteins for pseudovirus formation in this case included only M and E proteins [102]. In the case of SARS-CoV-1, the data on the minimal set of proteins for the formation of pseudovirus particles are quite contradictory. According to one study, the expression of two proteins, M and N, is required for the formation of pseudovirus particles in the cell. However, effective pseudovirus budding occurred only when three proteins, S, M, and N, were co-expressed. E protein,

despite its inclusion in VLP during co-expression with M and N proteins, was not necessary for pseudovirus formation [103]. According to another study, SARS-CoV-1-based pseudoviruses require expression of M, E, and N proteins for efficient assembly and budding [104]. This difference is probably due to differences in the packaging cells used: HEK-293 and Vero E6, respectively. In the case of SARS-CoV-2, proteins M, S, and E are required for VLP assembly in HEK-293T cells [105]. Interestingly, differences in VLP assembly in HEK-293T and Vero E6 cells were also observed for SARS-CoV-2, when S, E, M, and N proteins were co-expressed in Vero E6 cells, the S protein was more efficiently incorporated into the VLP than in HEK-293T cells [106].

A smaller number of studies are devoted to the production of VLPs for investigating the amplification and transcription of coronavirus RNA. For example, a special cellular system was developed for SARS-CoV-2, which allows the study of many fundamental aspects of the biology of the virus [107]. As part of this work, the N protein gene in the SARS-CoV-2 genome was replaced with a green fluorescent protein (GFP) gene. To assemble VLPs, the N protein was expressed ectopically in the packaging cells. As a result of this approach, VLPs cannot be generated in cells that do not express N protein. Culturing N protein-expressing packer cells for one month did not result in recombination acts contributing to the formation of the full-length SARS-CoV-2 genome and infectious virus production, confirming the safety of such systems. Nevertheless, further measures were proposed to improve the safety of the created pseudoviral system. For this purpose, the N protein gene was split into two parts and inserted into different vectors. The complete N-protein was formed as a result of intein-mediated protein splicing [107]. The created pseudoviral particles can be used for a wide range of tasks: fundamental study of SARS-CoV-2 biology, screening of potential inhibitors of virus penetration, transcription and replication, creation of vaccines and detection of neutralizing antibodies in patients. Through the use of this system much has already been learned about SARS-CoV-2. For example, it was found that N-protein contacts the components of cellular stress granules G3BP1 and G3BP2, and potential inhibitors of virus replication were also found [107].

The use of pseudoviruses based on the *Coronaviridae* family has an important practical application. A large number of studies are devoted to the development of VLP-based vaccines [102, 106−108]. A VLP vector capable of transducing human dendritic cells was created based on the neurotropic coronavirus HCoV-229E [109]. The RNA in the obtained pseudovirus encodes the complete ORF1a/b gene as well as three reporter genes, which can be subsequently replaced with target genes. Through discontinuous transcription, all three genes can be expressed in transduced cells. The uniqueness of the created vector lies
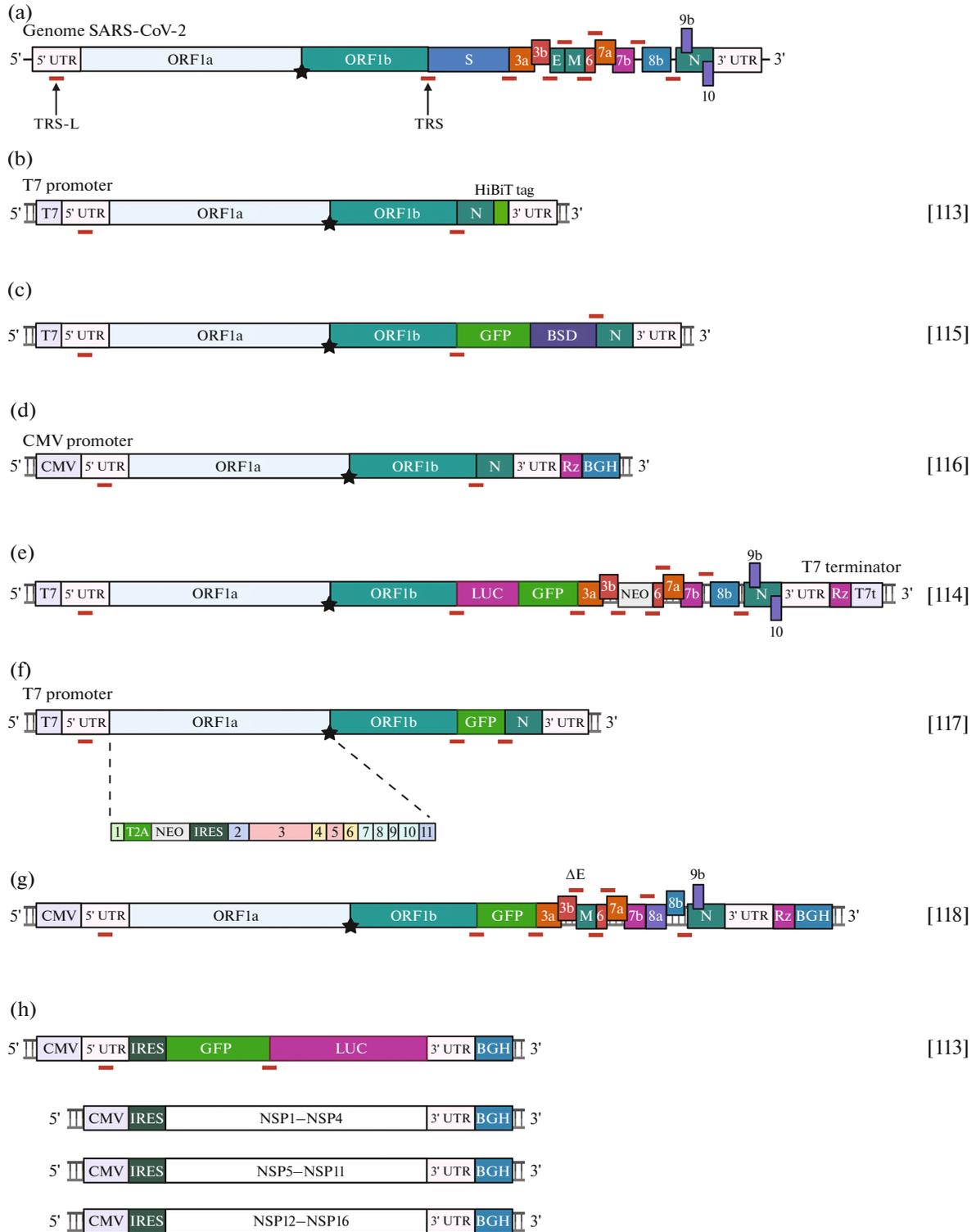
Fig. 7. Key structural elements of coronavirus replicons. (a) SARS-CoV-2 schematic genome structure (provided as a reference structure). (b–h) Structure of several different replicons [112–118]; regulatory elements (5'UTR, 3'UTR, TRS) are always encoded in a replicon, pp1a/ab components are absent only in minimal replicons (h) [118]. Replicon coding part (body) may be put under the control of either T7, or a cellular promoter. In the first case, it is possible to both carry out in vitro transcription using the obtained RNA for cell transfection and transfect cells with the DNA construct with the simultaneous T7 RNA polymerase expression [112–118]. Encoded proteins are indicated with letters across the diagram. TRS are indicated with red lines. GFP—green fluorescent protein; LUC—luciferase; BSD—blasticidin-resistance protein; NEO—neomycin-resistance protein, and IRES—internal ribosome entry site.

in the fact that several reporter genes with a total length of about 6 thousand bases can fit into one pseudovirus at once. However, this system has a significant drawback: in addition to the target genes, the entire polyprotein pp1a/ab is expressed, some fragments of which exhibit cytotoxicity [110].

REPLICONS

Replicons are self-sustaining RNAs containing all the regulatory elements of the viral genome. In the presence of viral nonstructural proteins that ensure viral replication and transcription, new RNA molecules are synthesized from the replicons [111]. The required nonstructural proteins can be encoded in the replicon itself or expressed in the cell from additional vectors. Replicons do not contain structural proteins and do not produce infectious viral particles, so they represent a safe alternative to full-length viral systems. Reporter genes are often encoded within the replicons, making it relatively easy to assess the efficiency of RNA synthesis under appropriate conditions. Pseudoviruses are also well suited for studying individual stages of the coronavirus life cycle, including genome replication, transcription, and translation, but nevertheless, it is replicons that are widely used in this kind of research because they are much easier to handle and safer than VLP, however they are farther from the conditions of real infection.

In general, all coronavirus replicons are arranged similarly: the 5'- and 3'-terminal sequences of the replicons coincide completely with the 5'- and 3'UTR of coronaviruses, respectively. In the study of discontinuous transcription, the sequence corresponding to the intergenic region of the coronavirus under study and regulating the process of RdRp complex skipping is placed inside the replicon [112–118]. As can be seen from Fig. 7, in the vast majority of cases the replicons contain a large ORF1a and 1b reading frame, which encodes nonstructural proteins involved in viral RNA replication, as well as the gene sequence of N protein required for efficient viral RNA synthesis [119, 120]. Replicons often contain reporter protein genes in the place of genes for various structural proteins of the virus. Such reporters are under the control of viral regulatory TRS sequences taken from the genes of the corresponding structural proteins TRS-S (from protein S), TRS-N (from protein N), and TRS-M (from protein M). In the case of minimal replicons (Fig. 7h) that do not encode pp1a/b and N protein, the viral proteins that provide replication/transcription must be expressed in the cells under study from additional vectors.

Replicons are frequently used for high-throughput screening of viral replication and transcription inhibitors [81, 112, 113, 115, 117, 118, 121–125]. SARS-CoV-1-based replicons have found application in testing the inhibitors of viral 3CL^{pro} and PL^{pro} proteases and viral helicase, as well as inhibitors of reading frame shift during translation [126–128]. In some studies, stable replicon-containing cell lines were obtained for which, in addition to the reporter genes, an antibiotic resistance gene, most commonly blasticidin and zeocin resistance, was incorporated into the replicon [115, 123]. It is convenient to use replicons to assess the effect of the expression of various proteins on the processes of viral genome replication and transcription. The use of replicons makes it possible to reliably show that one or another effect on the viral life cycle is related to the effect on the replication/transcription processes rather than virion assembly and budding. For example, using replicons, Wang et al. [117] showed that NSP16 but not NSP1 and NSP2 expression is required for effective SARS-CoV-1 RNA replication. And recently, Luo et al. [118] found that the replication efficiency of SARS-CoV-2 RNA was significantly increased in the presence of the ORF6 expression product.

CONCLUSIONS

In the present review, we attempted to summarize the data on the genetically engineered systems developed to date for studying human pathogens in the *Coronaviridae* family. For each approach, examples of possible uses were provided and the advantages and disadvantages of each system were outlined. We do not claim to cover all the systems used to study coronaviruses and the results obtained with their help, since coronavirus research, and primarily SARS-CoV-2 research, is actively carried out worldwide and new data on the mechanisms of pathogenesis of new coronavirus infection, SARS-CoV-2 replication inhibitors, and drugs and vaccines under development appear every day. We hope that the data presented in this review will be of interest to those involved in coronavirus research and will guide them in choosing the right system for their specific scientific tasks.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest. The authors declare no conflict of interest.

Statement on the welfare of humans or animals. This article does not contain any studies involving humans or animals performed by any of the authors.

ADDITIONAL INFORMATION

All images were obtained using the BioRender software.

REFERENCES

1. Yin Y., Wunderink R.G. 2018. MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology*. **23**, 130−137.

2. Wu F., Zhao S., Yu B., Chen Y.-M., Wang W., Song Z.-G., Hu Y., Tao Z.-W., Tian J.-H., Pei Y.-Y., Yuan M.-L., Zhang Y.-L., Dai F.-H., Liu Y., Wang Q.-M., et al. 2020. A new coronavirus associated with human respiratory disease in China. *Nature*. **579**, 265−269.

3. Rothe C., Schunk M., Sothmann P., Bretzel G., Froeschl G., Wallrauch C., Zimmer T., Thiel V., Janke C., Guggemos W., Seilmaier M., Drosten C., Vollmar P., Zwirglmaier K., Zange S., Wölfel R., Hoelscher M. 2020. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N. Engl. J. Med*. **382**, 970−971.

4. Perez-Saez J., Lauer S.A., Kaiser L., Regard S., Delaporte E., Guessous I., Stringhini S., Azman A.S., Serocov-POP Study Group. 2021. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. *Lancet Infect. Dis*. **21**, e69−e70.

5. Rebello C. J., Kirwan J.P., Greenway F.L. 2020. Obesity, the most common comorbidity in SARS-CoV-2: is leptin the link? *Int. J. Obes. (Lond.)*. **44**, 1810−1817.

6. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. 2020. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol*. **5**, 536−544.

7. Baltimore D. 1971. Expression of animal virus genomes. *Bacteriol. Rev*. **35**, 235−241.

8. Schoeman D., Fielding B.C. 2019. Coronavirus envelope protein: current knowledge. *Virol. J*. **16**, 69.

9. Sarkar M., Saha S. 2020. Structural insight into the role of novel SARS-CoV-2 E protein: a potential target for vaccine development and other therapeutic strategies. *PLoS One*. **15,** e0237300.

10. Lang Y., Li W., Li Z., Koerhuis D., van den Burg A.C.S., Rozemuller E., Bosch B.-J., van Kuppeveld F.J.M., Boons G.-J., Huizinga E.G., van der Schaar H.M., de Groot R.J. 2020. Coronavirus hemagglutinin-esterase and spike proteins coevolve for functional balance and optimal virion avidity. *Proc. Natl. Acad. Sci. U. S. A*. **117**, 25759−25770.

11. Bar-On Y.M., Flamholz A., Phillips R., Milo R. 2020. SARS-CoV-2 (COVID-19) by the numbers. *eLife*. **9**, e57309.

12. Almazán F., Galán C., Enjuanes L. 2004. The nucleoprotein is required for efficient coronavirus genome replication. *J. Virol*. **78**, 12683−12688.

13. Masters P.S. 2006. The molecular biology of coronaviruses. *Adv. Virus Res*. **66**, 193−292.

14. Belshaw R., Pybus O.G., Rambaut A. 2007. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res*. **17**, 1496−1504.

15. Sanjuán R., Nebot M.R., Chirico N., Mansky L.M., Belshaw R. 2010. Viral mutation rates. *J. Virol*. **84**, 9733−9748.

16. Chen Y., Liu Q., Guo D. 2020. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol*. **92**, 418−423.

17. Song Z., Xu Y., Bao L., Zhang L., Yu P., Qu Y., Zhu H., Zhao W., Han Y., Qin C. 2019. From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses*. **11**, 59.

18. Sa Ribero M., Jouvenet N., Dreux M., Nisole S. 2020. Interplay between SARS-CoV-2 and the type I interferon response. *PLoS Pathog*. **16** (7), e1008737.

19. Snijder E.J., Decroly E., Ziebuhr J. 2016. The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv. Virus Res*. **96**, 59−126.

20. Raj R. 2021. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochem. Biophys. Rep*. **25**, 100847.

21. Khan M.T., Irfan M., Ahsan H., Ahmed A., Kaushik A.C., Khan A.S., Chinnasamy S., Ali A., Wei D.-Q. 2021. Structures of SARS-CoV-2 RNA-binding proteins and therapeutic targets. *Intervirology*. **64**, 55−68.

22. Yoshimoto F.K. 2020. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J*. **39**, 198−216.

23. Guo Y.-R., Cao Q.-D., Hong Z.-S., Tan Y.-Y., Chen S.-D., Jin H.-J., Tan K.-S., Wang D.-Y., Yan Y. 2020. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak − An update on the status. *Mil. Med. Res*. **7**, 11.

24. Cui J., Li F., Shi Z.-L. 2019. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol*. **17**, 181−192.

25. Artika I.M., Dewantari A.K., Wiyatno A. 2020. Molecular biology of coronaviruses: Current knowledge. *Heliyon*. **6**, e04743.

26. Banerjee A.K., Blanco M.R., Bruce E.A., Honson D.D., Chen L.M., Chow A., Bhat P., Ollikainen N., Quinodoz S.A., Loney C., Thai J., Miller Z.D., Lin A.E., Schmidt M.M., Stewart D.G., et al. 2020. SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. *Cell*. **183**, 1325−1339.e21.

27. Gaglia M M., Covarrubias S., Wong W., Glaunsinger B.A. 2012. A common strategy for host RNA degradation by divergent viruses. *J. Virol*. **86**, 9527−9530.

28. Cornillez-Ty C.T., Liao L., Yates J.R., Kuhn P., Buchmeier M.J. 2009. Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *J. Virol*. **83**, 10314−10318.

29. Graham R.L., Sims A.C., Brockway S.M., Bari, R.S., Denison M.R. 2005. The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication. *J. Virol*. **79**, 13399−13411.

30. Lei J., Kusov Y., Hilgenfeld R. 2018. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res*. **149**, 58−74.

31. Wolff G., Limpens R.W.A.L., Zevenhoven-Dobbe J.C., Laugks U., Zheng S., de Jong A.W.M., Koning R.I., Agard D.A., Grünewald K., Koster A.J., Snijder E.J., Bárcena M. 2020. A molecular pore spans the double

membrane of the coronavirus replication organelle. *Science.* **369**, 1395−1398.

32. Oostra M., te Lintelo E.G., Deijs M., Verheije M.H., Rottier P.J.M., de Haan C.A.M. 2007. Localization and membrane topology of coronavirus nonstructural protein 4: Involvement of the early secretory pathway in replication. *J. Virol.* **81**, 12323−12336.

33. Tomar S., Johnston M.L., St John S.E., Osswald H.L., Nyalapatla P.R., Paul L.N., Ghosh A.K., Denison M.R., Mesecar A.D. 2015. Ligand-induced dimerization of Middle East respiratory syndrome (MERS) coronavirus nsp5 protease (3CLpro): Implications for *nsp5* regulation and the development of antivirals. *J. Biol. Chem.* **290**, 19403−19422.

34. Cottam E.M., Whelband M.C., Wileman T. 2014. Coronavirus NSP6 restricts autophagosome expansion. *Autophagy.* **10**, 1426−1441.

35. Zhai Y., Sun F., Li X., Pang H., Xu X., Bartlam M., Rao Z. 2005. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat. Struct. Mol. Biol.* **12**, 980−986.

36. Imbert I., Guillemot J.-C., Bourhis J.-M., Bussetta C., Coutard B., Egloff M.-P., Ferron F., Gorbalenya A.E., Canard B. 2006. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J.* **25**, 4933−4942.

37. te Velthuis A.J.W., van den Worm S.H.E., Snijder E.J. 2012. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both *de novo* initiation and primer extension. *Nucleic Acids. Res.* **40**, 1737−1747.

38. Sawicki S.G., Sawicki D.L., Younker D., Meyer Y., Thiel V., Stokes H., Siddell S.G. 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. *PLoS Pathog.* **1**, e39.

39. Donaldson E.F., Sims A.C., Graham R.L., Denison M.R., Baric R.S. 2007. Murine hepatitis virus replicase protein nsp10 is a critical regulator of viral RNA synthesis. *J. Virol.* **81**, 6356−6368.

40. Bouvet M., Debarnot C., Imbert I., Selisko B., Snijder E.J., Canard B., Decroly E. 2010. *In vitro* reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog.* **6**, e1000863.

41. Bouvet M., Imbert I., Subissi L., Gluais L., Canard B., Decroly E. 2012. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9372−9377.

42. Shu T., Huang M., Wu D., Ren Y., Zhang X., Han Y., Mu J., Wang R., Qiu Y., Zhang D.-Y., Zhou X. 2020. SARS-coronavirus-2 Nsp13 possesses NTPase and RNA helicase activities that can be inhibited by bismuth salts. *Virol. Sin.* **35**, 321−329.

43. Minskaia E., Hertzig T., Gorbalenya A.E., Campanacci V., Cambillau C., Canard B., Ziebuhr J. 2006. Discovery of an RNA virus 3'→5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5108−5113.

44. Chen Y., Cai H., Pan J., Xiang N., Tien P., Ahola T., Guo D. 2009. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7

methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3484−3489.

45. Deng X., Hackbart M., Mettelman R.C., O'Brien A., Mielech A.M., Yi G., Kao C.C., Baker S.C. 2017. Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci. A.* **114**, E4251−E4260.

46. Forni D., Cagliani R., Clerici M., Sironi M. 2017. Molecular evolution of human coronavirus genomes. *Trends Microbiol.* **25**, 35−48.

47. Li F. 2016. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* **3**, 237−261.

48. Yeager C.L., Ashmun R.A., Williams R.K., Cardellichio C.B., Shapiro L.H., Look A.T., Holmes K.V. 1992. Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature.* **357**, 420−422.

49. Wu K., Li W., Peng G., Li F. 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19970−19974.

50. Naskalska A., Dabrowska A., Szczepanski A., Milewska A., Jasik K.P., Pyrc K. 2019. Membrane protein of human coronavirus NL63 is responsible for interaction with the adhesion receptor. *J. Virol.* **93**, e00355-19.

51. Desforges M., Desjardins J., Zhang C., Talbot P.J. 2013. The acetyl-esterase activity of the hemagglutinin-esterase protein of human coronavirus OC43 strongly enhances the production of infectious virus. *J. Virol.* **87**, 3097−3107.

52. Raj V.S., Mou H., Smits S.L., Dekkers D.H.W., Müller M.A., Dijkman R., Muth D., Demmers J.A.A., Zaki A., Fouchier R.A.M., Thiel V., Drosten C., Rottier P.J.M., Osterhaus A.D.M.E., Bosch B.J., Haagmans B.L. 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature.* **495**, 251−254.

53. Kalra R.S., Kandimalla R. 2021. Engaging the spikes: heparan sulfate facilitates SARS-CoV-2 spike protein binding to ACE2 and potentiates viral infection. *Sig. Transduct. Target. Ther.* **6** (1), 39.

54. Hartenian E., Nandakumar D., Lari A., Ly M., Tucker J.M., Glaunsinger B.A. 2020. The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910−12934.

55. Tang T., Bidon M., Jaimes J.A., Whittaker G.R., Daniel S. 2020. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral. Res.* **178**, 104792.

56. Sola I., Almazán F., Zúñiga S., Enjuanes L. 2015. Continuous and discontinuous RNA synthesis in coronaviruses. *Annu. Rev. Virol.* **2**, 265−288.

57. Sawicki S.G., Sawicki D.L., Siddell S.G. 2007. A contemporary view of coronavirus transcription. *J. Virol.* **81**, 20−29.

58. Sztuba-Solińska J., Stollar V., Bujarski J.J. 2011. Subgenomic messenger RNAs: Mastering regulation of (+)-strand RNA virus life cycle. *Virology.* **412**, 245−255.

59. Zúñiga S., Sola I., Alonso S., Enjuanes L. 2004. Sequence motifs involved in the regulation of discontin-

uous coronavirus subgenomic RNA synthesis. *J. Virol.* **78**, 980–994.

60. Wu H.-Y., Brian D.A. 2007. 5'-proximal hot spot for an inducible positive-to-negative-strand template switch by coronavirus RNA-dependent RNA polymerase. *J. Virol.* **81**, 3206–3215.

61. Bojkova D., Klann K., Koch B., Widera M., Krause D., Ciesek S., Cinatl J., Münch C. 2020. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature.* **583**, 469–472.

62. Finkel Y., Gluck A., Nachshon A., Winkler R., Fisher T., Rozman B., Mizrahi O., Lubelsky Y., Zuckerman B., Slobodin B., Yahalom-Ronen Y., Tamir H., Ulitsky I., Israely T., Paran N., et al. 2021. SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis. *Nature.* **594**, 240–245.

63. Woo J., Lee E.Y., Lee M., Kim T., Cho Y.-E. 2019. An *in vivo* cell-based assay for investigating the specific interaction between the SARS-CoV N-protein and its viral RNA packaging sequence. *Biochem. Biophys. Res. Commun.* **520**, 499–506.

64. Narayanan K., Makino S. 2001. Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *J. Virol.* **75**, 9059–9067.

65. Vennema H., Godeke G.J., Rossen J.W., Voorhout W.F., Horzinek M.C., Opstelten D.J., Rottier P.J. 1996. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* **15**, 2020–2028.

66. Hsieh P.-K., Chang S.C., Huang C.-C., Lee T.-T., Hsiao C.-W., Kou Y.-H., Chen I.-Y., Chang C.-K., Huang T.-H., Chang M.-F. 2005. Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *J. Virol.* **79**, 13848–13855.

67. Masters P.S. 2019. Coronavirus genomic RNA packaging. *Virology.* **537**, 198–207.

68. Yao S., Narayanan A., Majowicz S., Jose J., Archetti M. 2021. A synthetic defective interfering SARS-CoV-2. *Peer J.* **9**, e11686.

69. Kuo L., Hurst-Hess K.R., Koetzner C.A., Masters P.S. 2016. Analyses of coronavirus assembly interactions with interspecies membrane and nucleocapsid protein chimeras. *J. Virol.* **90**, 4357–4368.

70. Orenstein J.M., Banach B., Baker S. 2008. Morphogenesis of coronavirus HCoV-NL63 in cell culture: a transmission electron microscopic study. *Open Infect. Dis. J.* **2**, 52–58.

71. Makino S., Lai M.M. 1990. Studies of coronavirus DI RNA replication using in vitro constructed DI cDNA clones. *Adv. Exp. Med. Biol.* **276**, 341–347.

72. Kim K.H., Narayanan K., Makino S. 1998. Characterization of coronavirus DI RNA packaging. *Adv. Exp. Med. Biol.* **440**, 347–353.

73. Bos E.C., Dobbe J.C., Luytjes W., Spaan W.J. 1997. A subgenomic mRNA transcript of the coronavirus mouse hepatitis virus strain A59 defective interfering (DI) RNA is packaged when it contains the DI packaging signal. *J. Virol.* **71**, 5684–5687.

74. Makino S., Keck J.G., Stohlman S.A., Lai M.M. 1986. High-frequency RNA recombination of murine coronaviruses. *J. Virol.* **57**, 729–737.

75. Williams G.D., Chang R.-Y., Brian D.A. 1999. A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication. *J. Virol.* **73**, 8349–8355.

76. Almazán F., González J.M., Pénzes Z., Izeta A., Calvo E., Plana-Durán J., Enjuanes L. 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5516–5521.

77. Strauss-Soukup J.K., Strobel S.A. 2001. Ribozyme enzymology. In: *RNA.* Eds Söll D., Nishimura S., Moore P. Oxford: Pergamon, pp. 187–206.

78. Yang D., Leibowitz J.L. 2015. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* **206**, 120–133.

79. Yang J., Sun B., Huang H., Jiang Y., Diao L., Chen B., Xu C., Wang X., Liu J., Jiang W., Yang S. 2014. High-efficiency scarless genetic modification in *Escherichia coli* by using lambda red recombination and I-SceI cleavage. *Appl. Environ. Microbiol.* **80**, 3826–3834.

80. St-Jean J.R., Desforges M., Almazán F., Jacomy H., Enjuanes L., Talbot P.J. 2006. Recovery of a neurovirulent human coronavirus OC43 from an infectious cDNA clone. *J. Virol.* **80**, 3670–3674.

81. Almazán F., DeDiego M.L., Sola I., Zuñiga S., Nieto-Torres J.L., Marquez-Jurado S., Andrés G., Enjuanes L. 2013. Engineering a replication-competent, propagation-defective middle east respiratory syndrome coronavirus as a vaccine candidate. *mBio,* **4,** e00650-13.

82. Pfefferle S., Krähling V., Ditt V., Grywna K., Mühlberger E., Drosten C. 2009. Reverse genetic characterization of the natural genomic deletion in SARS-coronavirus strain Frankfurt-1 open reading frame 7b reveals an attenuating function of the 7b protein in vitro and in vivo. *Virol. J.* **6**, 131.

83. Ye C., Chiem K., Park J.-G., Oladunni F., Platt R. N., Anderson T., Almazan F., de la Torre J.C., Martinez-Sobrido L. 2020. Rescue of SARS-CoV-2 from a single bacterial artificial chromosome. *mBio.* **11** (5), e02168-20. https://doi.org/10.1128/mBio.02168-20

84. Thi Nhu Thao T., Labroussaa F., Ebert N., V'kovski P., Stalder H., Portmann J., Kelly J., Steiner S., Holwerda M., Kratzel A., Gultom M., Schmied K., Laloli L., Hüsser L., Wider M., et al. 2020. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature.* **582**, 561–565.

85. Eriksson K.K., Makia D., Thiel V. 2008. Generation of recombinant coronaviruses using vaccinia virus as the cloning vector and stable cell lines containing coronaviral replicon RNAs. *Methods Mol. Biol.* **454**, 237–254.

86. Donaldson E.F., Yount B., Sims A.C., Burkett S., Pickles R.J., Baric R.S. 2008. Systematic assembly of a full-length infectious clone of human coronavirus NL63. *J. Virol.* **82**, 11948–11957.

87. Scobey T., Yount B.L., Sims A.C., Donaldson E.F., Agnihothram S.S., Menachery V.D., Graham R.L., Swanstrom J., Bove P.F., Kim J.D., Grego S., Randell S.H., Baric R.S. 2013. Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16157–16162.

88. Yount B., Curtis K.M., Fritz E.A., Hensley L.E., Jahrling P.B., Prentice E., Denison M.R., Geisbert T.W., Baric R.S. 2003. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12995−13000.

89. van den Worm S.H.E., Eriksson K.K., Zevenhoven J.C., Weber F., Züst R., Kuri T., Dijkman R., Chang G., Siddell S.G., Snijder E. J., Thiel V., Davidson A.D. 2012. Reverse genetics of SARS-related coronavirus using vaccinia virus-based recombination. *PLoS One.* **7**, e32857.

90. Xie X., Muruato A., Lokugamage K.G., Narayanan K., Zhang X., Zou J., Liu J., Schindewolf, C., Bopp N.E., Aguilar P.V., Plante K.S., Weaver S.C., Makino S., LeDuc J.W., Menachery V.D., Shi P.Y. 2020. An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe.* **27**, 841−848.e3.

91. Kouprina N., Larionov V. 2016. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma.* **125**, 621.

92. Cunningham C., Gatherer D., Hilfrich B., Baluchova K., Dargan D.J., Thomson M., Griffiths P.D., Wilkinson G.W.G., Schulz T.F., Davison A.J. 2010. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J. Gen. Virol.* **91**, 605.

93. Vashee S., Stockwell T.B., Alperovich N., Denisova E.A., Gibson D.G., Cady K.C., Miller K., Kannan K., Malouli D., Crawford L.B., Voorhies A.A., Bruening E., Caposio P., Früh K. 2017. Cloning, assembly, and modification of the primary human cytomegalovirus isolate Toledo by yeast-based transformation-associated recombination. *mSphere.* **2**, e00331-17.

94. Oldfield L.M., Grzesik P., Voorhies A.A., Alperovich N., MacMath D., Najera C.D., Chandra D.S., Prasad S., Noskov V.N., Montague M. G., Friedman R.M., Desai P.J., Vashee S. 2017. Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proc. Natl. Acad. Sci. USA.* **114**, E8885−E8894.

95. Smith S., Reuven N., Mohni K.N., Schumacher A.J., Weller S.K. 2014. Structure of the herpes simplex virus 1 genome: manipulation of nicks and gaps can abrogate infectivity and alter the cellular DNA damage response. *J. Virol.* **88**, 10146−10156.

96. Yount B., Curtis K.M., Baric R.S. 2000. Strategy for systematic assembly of large RNA and DNA genomes: Transmissible gastroenteritis virus model. *J. Virol.* **74**, 10600−10611.

97. Merchlinsky M., Moss B. 1992. Introduction of foreign DNA into the vaccinia virus genome by *in vitro* ligation: recombination-independent selectable cloning vectors. *Virology.* **190**, 522−526.

98. Mohsen M. O., Zha L., Cabral-Miranda G., Bachmann M. F. 2017. Major findings and recent advances in virus-like particle (VLP)-based vaccines. *Semin. Immunol.* **34**, 123−132.

99. Nooraei S., Bahrulolum H., Hoseini Z.S., Katalani C., Hajizade A., Easton A.J., Ahmadian G. 2021. Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. *J. Nanobiotechnology.* **19**, 59.

100. Chan S.K., Du P., Ignacio C., Mehta S., Newton I.G., Steinmetz N.F. 2021. Biomimetic virus-like particles as severe acute respiratory syndrome coronavirus 2 diagnostic tools. *ACS Nano.* **15**, 1259−1272.

101. Bos E.C., Luytjes W., Spaan W.J. 1997. The function of the spike protein of mouse hepatitis virus strain A59 can be studied on virus-like particles: cleavage is not required for infectivity. *J. Virol.* **71**, 9427−9433.

102. Naskalska A., Dabrowska A., Nowak P., Szczepanski A., Jasik K., Milewska A., Ochman M., Zeglen S., Rajfur Z., Pyrc K. 2018. Novel coronavirus-like particles targeting cells lining the respiratory tract. *PLoS One.* **13**, e0203489.

103. Huang Y., Yang Z., Kong W., Nabel G.J. 2004. Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: Implications for assembly and vaccine production. *J. Virol.* **78**, 12557−12565.

104. Siu Y.L., Teoh K.T., Lo J., Chan C.M., Kien F., Escriou N., Tsao S.W., Nicholls J.M., Altmeyer R., Peiris J.S., Bruzzone R., Nal B. 2008. The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J. Virol.* **82**, 11318−11330.

105. Swann H., Sharma A., Preece B., Peterson A., Eldridge C., Belnap D.M., Vershinin M., Saffarian S. 2020. Minimal system for assembly of SARS-CoV-2 virus like particles. *Sci. Rep.* **10**, 21877.

106. Xu R., Shi M., Li J., Song P., Li N. 2020. Construction of SARS-CoV-2 virus-like particles by mammalian expression system. *Front. Bioeng. Biotechnol.* **8,** 862.

107. Ju X., Zhu Y., Wang Y., Li J., Zhang J., Gong M., Ren W., Li S., Zhong J., Zhang L., Zhang Q.C., Zhang R., Ding Q. 2021. A novel cell culture system modeling the SARS-CoV-2 life cycle. *PLoS Pathog.* **17**, e1009439.

108. Wang C., Zheng X., Gai W., Zhao Y., Wang H., Wang H., Feng N., Chi H., Qiu B., Li N., Wang T., Gao Y., Yang S., Xia X. 2016. MERS-CoV virus-like particles produced in insect cells induce specific humoural and cellular imminity in rhesus macaques. *Oncotarget.* **8**, 12686−12694.

109. Talbot P.J., Ekandé S., Cashman N.R., Mounir S., Stewart J.N. 1993. Neurotropism of human coronavirus 229E. *Adv. Exp. Med. Biol.* **342**, 339−346.

110. Thiel V., Karl N., Schelle B., Disterer P., Klagge I., Siddell S.G. 2003. Multigene RNA vector based on coronavirus transcription. *J. Virol.* **77**, 9790−9798.

111. Khromykh A.A. 2000. Replicon-based vectors of positive strand RNA viruses. *Curr. Opin. Mol. Ther.* **2**, 555−569.

112. Kotaki T., Xie X., Sh, P.-Y., Kameoka M. 2021. A PCR amplicon-based SARS-CoV-2 replicon for antiviral evaluation. *Sci. Rep.* **11**, 2229.

113. He X., Quan S., Xu M., Rodriguez S., Goh S.L., Wei J., Fridman A., Koeplinger K.A., Carroll S.S., Grobler J.A., Espeseth A.S., Olsen D.B., Hazuda D.J., Wang D. 2021. Generation of SARS-CoV-2 reporter replicon for high-throughput antiviral screening and testing. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2025866118.

114. Ge F., Luo Y., Liew P.X., Hung E. 2007. Derivation of a novel SARS−coronavirus replicon cell line and its

application for anti-SARS drug screening. *Virology*. **360**, 150−158.

115. Almazán F., DeDiego M.L., Galán C., Escors D., Álvarez E., Ortego J., Sola I., Zuñiga S., Alonso S., Moreno J.L., Nogales A., Capiscol C., Enjuanes L. 2006. Construction of a severe acute respiratory syndrome coronavirus infectious cDNA clone and a replicon to study coronavirus RNA synthesis. *J. Virol.* **80**, 10900−10906.

116. Hertzig T., Scandella E., Schelle B., Ziebuhr J., Siddell S.G., Ludewig B., Thiel V. 2004. Rapid identification of coronavirus replicase inhibitors using a selectable replicon RNA. *J. Gen. Virol.* **85**, 1717−1725.

117. Wang J.-M., Wang L.-F., Shi Z.-L. 2008. Construction of a non-infectious SARS coronavirus replicon for application in drug screening and analysis of viral protein function. *Biochem. Biophys. Res. Commun.* **374**, 138−142.

118. Luo Y., Yu F., Zhou M., Liu Y., Xia B., Zhang X., Liu J., Zhang J., Du Y., Li R., Wu L., Zhang X., Pan T., Guo D., Peng T., Zhang H. 2021. Engineering a reliable and convenient SARS-COV-2 replicon system for analysis of viral RNA synthesis and screening of antiviral inhibitors. *mBio*. **12,** e02754-20. https://doi.org/10.1128/mBio.02754-20

119. McBride R., van Zyl M., Fielding B.C. 2014. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*. **6**, 2991−3018.

120. Dutta N.K., Mazumdar K., GordyJ.T. 2020. The nucleocapsid protein of SARS-COV-2: a target for vaccine development. *J. Virol.* **94**, e00647-20.

121. Ge F., Xiong S., Lin F.S., Zhang Z.P., Zhang X.E. 2008. High-throughput assay using a GFP-expressing replicon for SARS-CoV drug discovery. *Antiviral Res.* **80**, 107−113.

122. Pfefferle S., Schöpf J., Kögl M., Friedel C.C., Müller M.A., Carbajo-Lozoya J., Stellberger T., von Dall'Armi E., Herzog P., Kallies S., Niemeyer D., Ditt V., Kuri T., Züst R., Pumpor K., et al. 2011. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS Pathog.* **7**, e1002331.

123. Jin Y.-Y., Lin H., Cao L., Wu W.-C., Ji Y., Du L., Jiang Y., Xie Y., Tong K., Xing F., Zheng F., Shi M., Pan J.A., Peng X., Guo D. 2021. A convenient and biosafe replicon with accessory genes of SARS-CoV-2 and its potential application in antiviral drug discovery. *Virol. Sin.* **36** (5), 913−923. https://doi.org/10.1007/s12250-021-00385-9

124. Wang B., Zhang C., Lei X., Ren L., Zhao Z., Wang J., Huang H. 2021. Construction of non-infectious SARS-CoV-2 replicons and their application in drug evaluation. *Virol. Sin*. **36**, 890−900. https://doi.org/10.1007/s12250-021-00369-9

125. Zhang Q.-Y., Deng C.-L., Liu J., Li J.-Q., Zhang H.-Q., Li N., Zhang Y.-N., Li X.-D., Zhang B., Xu Y., Ye H.-Q. 2021. SARS-CoV-2 replicon for high-throughput antiviral screening. *J. Gen. Virol.* **102**, 001583.

126. Ahn D.-G., Lee W., Choi J.-K., Kim S.-J., Plant E.P., Almazán F., Taylor D.R., Enjuanes L., Oh J.-W. 2011. Interference of ribosomal frameshifting by antisense peptide nucleic acids suppresses SARS coronavirus replication. *Antiviral Res.* **91**, 1−10.

127. Adedeji A.O., Marchand B., te Velthuis A.J.W., Snijder E.J., Weiss S., Eoff R.L., Singh K., Sarafianos S.G. 2012. Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS One*. **7**, e36521.

128. Ivanov K.A., Hertzig T., Rozanov M., Bayer S., Thiel V., Gorbalenya A.E., Ziebuhr J. 2004. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12694−12699.

Translated by E. Martynova