Journal of Proteins and Proteomics
https://doi.org/10.1007/s42485-021-00060-3

ORIGINAL ARTICLE

The percentages of SARS-CoV-2 protein similarity and identity with SARS-CoV and BatCoV RaTG13 proteins can be used as indicators of virus origin

Mohammed Elimam Ahamed Mohammed[1,2]

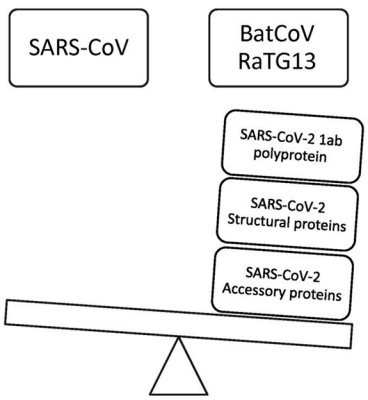
Received: 16 January 2021 / Revised: 8 March 2021 / Accepted: 22 March 2021

Abstract

There are three types of proteins in coronaviruses: nonstructural, structural, and accessory proteins. Coronavirus proteins are essential for viral replication and for the binding and invasion of hosts and the regulation of host cell metabolism and immunity. This study investigated the amino acid sequence similarity and identity percentages of 10 proteins in SARS-CoV-2, SARS-CoV and the *Rhinolophus affinis* bat coronavirus (BatCoV RaTG13). The investigated proteins were the 1ab polyprotein, spike protein, orf3a, the envelope protein, the membrane protein, orf6, orf7a, orf7b, orf8, and the nucleocapsid protein. The online sequence alignment service of The European Molecular Biology Open Software Suite (EMBOSS) was used to determine the percentages of protein similarity and identity in the three viruses. The results showed that the similarity and identity percentages of the SARS-CoV-2 and BatCoV RaTG13 proteins were both greater than 95%, while the identity and similarity percentages of SARS-CoV-2 and SARS-CoV were both greater than 38%. The proteins of SARS-CoV-2 and BatCoV RaTG13 have high identity and similarity compared to those of SARS-CoV-2 and SARS-CoV.

Graphic abstract

The proteins of the SARS-CoV-2 are most identical and similar to those of BatCoV RaTG13 than to the proteins of SARS-CoV



The proteins of the SARS-CoV-2 are most identical and similar to those of BatCoV RaTG13 than to the proteins of SARS-CoV

Keywords 1ab Polyprotein · Spike protein · Envelope protein · nsp3 · orf10

✉ Mohammed Elimam Ahamed Mohammed
mohammedelimam@yahoo.com

Extended author information available on the last page of the article

Published online: 09 April 2021

Introduction

Coronavirus disease 2019 (COVID-19) originated from a seafood market in Wuhan city (the capital of Hubei Province in southeastern China) and spread rapidly in more than 200 countries. By 2 Jul 2020, the total confirmed cases had reached more than 10.5 million, and 512,000 deaths had been reported. The symptoms of COVID-19 include cough, fever, headache, fatigue, sore throat, and malaise. The disease can lead to complications, such as pneumonia and severe acute respiratory syndrome (WHO 2020; Ahmad et al. 2020; Velavan and Meyer 2020). COVID-19 is transmitted through direct or indirect contact with respiratory droplets and biological samples such as urine, saliva, and stool (Shereen and Khan 2020). However, some studies proved the presence of the virus in air samples, and one study stated that the virus in air samples is viable for up to 3 h (Cheng et al. 2019; Ong et al. 2020; Liu et al. 2020; Doremalen et al. 2020).

Coronavirus 19 was named by the WHO and the International Committee on Taxonomy of Viruses (ICTV) as SARS-CoV-2, which is grouped in the same class of SARS-CoV (International Committee on Taxonomy of Viruses (ICTV) 2020). The two viruses belong to the family *Coronaviridae*, subfamily *Orthocoronavirinae,* genus *Betacoronavirus*; and subgenus *Sarbecovirus*, and the species is *severe acute respiratory syndrome-related coronavirus*. Bat coronavirus (BatCoV RaTG13) was isolated from animals of genus *Rhinolophus affinis.* Similar to SARS-CoV-2 and SARS-CoV, bat coronavirus BatCoV RaTG13 belongs to the *Betacoronavirus* family and has 96% genome sequence identity with the genome of SARS-CoV-2 (Zhou et al. 2020).

SARS-CoV-2, SARS-CoV, and bat coronavirus BatCoV RaTG13 have the same virion structure. They are RNA viruses with a nucleocapsid protein and an envelope. The viral envelope contains a bi-lipid membrane and three proteins: the spike protein, an envelope protein, and a membrane protein (Perlman and Netland 2009).

The three viruses contain two major genes: orf1ab and orf1a (comprising two-thirds of the total) and the structural and accessory protein genes (comprising one-third of the total). The Orf1ab and Orf1a genes are translated and hydrolysed to produce 16 nonstructural proteins (nsp1–nsp16), while the translation of the second gene produces the structural proteins spike (S), envelope (E), membrane (M), and nucleocapsid (N) and the accessory proteins orf3a, orf3b, orf6, orf7a, orf7b, orf8a, orf8b, orf9b and orf10. The number and type of accessory proteins differ according to the virus (Zhou et al. 2020; Yoshimoto 2020; Wang et al. 2020; Khailany et al. 2020; Wong et al. 2019; GenBank 2020).

Regarding NS3, NS6, NS7a, NS7b and NS8 of BatCoV RaTG13, some published articles named them nonstructural proteins, and others named them accessory proteins (GenBank 2020; Fahmi et al. 2020; Tang et al. 2020; Li et al. 2020). These proteins are encoded by genes similar to those of structural and accessory proteins, and because they are comparable to the accessory proteins of SARS-CoV-2, they are considered accessory proteins.

This article investigated the protein sequence identity and similarity percentages of SARS-CoV-2 and compared them to the proteins of SARS-CoV and the BatCoV RaTG13.

Materials and methods

Study proteins

This 1ab polyprotein of SARS-CoV-2, SARS-CoV, and BatCoV RaTG13 was studied. Additionally, the structural and accessory proteins found in SARS-CoV-2 and BatCoV RaTG13 were studied, including the spike protein (S), orf3, envelope protein (E), membrane protein (M), orf6, orf7a, orf7b, orf8, and nucleocapsid protein (N) (Table 2). The amino acid sequences were obtained from the National Center for Biotechnology Information (NCBI) site (https://www.ncbi.nlm.nih.gov/protein) (Table 1).

Sequence alignment

The online sequence alignment service of The European Molecular Biology Open Software Suite (EMBOSS) was used to determine the percentages of protein similarity and identity of SARS-CoV-2, SARS-CoV, and RaTG13. The matrix of the sequence alignment was EBLOSUM62, and the gap extends penalties were 14 and 4. The sequence alignment service of the EMBOSS can be accessed at https://www.bioinformatics.nl/cgi-bin/emboss/matcher. As a confirmatory test, a coloured alignment display was generated for each protein using the service of multiple sequence alignment of the European Molecular Biology Laboratory—European Bioinformatics Institute (EMBL-EBI) available at https://www.ebi.ac.uk/Tools/msa/clustalo/.

Results and discussion

This study reports differences in the identity and similarity percentage of the proteins of SARS-CoV-2 versus SARS-CoV and of SARS-CoV-2 versus the bat coronavirus RaTG13. The differences suggest a bat origin over a SARS-CoV origin, and these differences were caused by different types of mutations including deletions, insertions and substitutions [Annex 1, Annex 2 in ESM, Figs. 1, 2, 3, 4, 5, 6, 7 and 8].

Table 1 The studied proteins of the three viruses

	Protein		SARS-CoV-2	SARS-CoV	RaTG13
1	1ab polyprotein	NCBI Code	YP_009724389.1	NP_828849.7	QHR63299.1
		Gene location	266..21555	265..21485	251..21537
		Amino acid number	7096	7073	7095
2	S protein	NCBI Code	YP_009724390.1	YP_009825051.1	QHR63300.2
		Gene location	21492..25259	21492..25259	21545..25354
		Amino acid number	1273	1255	1269
3	Orf3	NCBI Code	YP_009724391.1	YP_009825052.1	QHR63301.1
		Gene location	25393..26220	25268..26092	25363..26190
		Amino acid number	275	274	275
4	E protein	NCBI Code	YP_009724392.1	YP_009825054.1	QHR63302.1
		Gene location	26245..26472	26117..26347	26215..26442
		Amino acid number	75	76	75
5	M protein	NCBI Code	YP_009724393.1	YP_009825055.1	QHR63303.1
		Gene location	26523..27191	26398..27063	26493..27158
		Amino acid number	222	221	221
6	Orf6	NCBI Code	YP_009724394.1	YP_009825056.1	QHR63304.1
		Gene location	27202..27387	26913..27265	27169..27354
		Amino acid number	61	63	61
7	Orf7a	NCBI Code	YP_009724395.1	YP_009825057.1	QHR63305.1
		Gene location	27394..27759	27273..27641	27360..27725
		Amino acid number	121	122	121
8	Orf7b	NCBI Code	YP_009725318.1	YP_009825058.1	QHR63306.1
		Gene location	27756..27887	27638..27772	27722..27853
		Amino acid number	43	44	43
9	Orf8	NCBI Code	YP_009724396.1	YP_009825059.1 YP_009825060.1	QHR63307.1
		Gene location	27894..28259	27779..27898, 27864..28118	27860..28225
		Amino acid number	121	39, 84	121
10	N protein	NCBI Code	YP_009724397.2	YP_009825061.1	QHR63308.1
		Gene location	28274..29533	28120..29388	28240..29499
		Amino acid number	419	422	419

The number of amino acids and their sequences for the proteins were obtained from the National Center for Biotechnology Information (NCBI) site accessible at https://www.ncbi.nlm.nih.gov/protein

The 1ab polyprotein

The 1ab polyprotein of SARS-CoV-2, SARS-CoV and BatCoV RaTG13 is composed of 7096, 7073, and 7095 amino acids, respectively (Table 1). The amino acid sequence identity and similarity of the 1ab polyprotein of SARS-CoV-2 and BatCoV RaTG13 were 98.5% and 99.1%, respectively. The percentages of identity and similarity of the 1ab polyprotein of SARS-CoV-2 and SARS-CoV were 86.2% and 92.9%, respectively. The results show that SARS-CoV-2 most likely originates from the *Rhinolophus affinis* bat, not from a laboratory-modified SARS-CoV variant (Table 2). Large-scale mutations were reported for the 1ab protein of SARS-CoV-2, SARS-CoV, and the bat coronavirus RaTG13. However, more mutations in the 1ab polyprotein of SARS-CoV-2 and SARS-CoV were shared in common than those of SARS-CoV-2 and bat coronavirus RaTG13 [Annex 1].

After the production of the 1ab polyprotein, some endopeptidases produce the 1a polyprotein and 16 nonstructural proteins (Snijder et al. 2016). The cleavage products of the 1ab polyprotein carry out a wide range of activities associated with the replication of the virus. The activities include binding and breakdown of ATP to produce ADP and phosphate, and the activities of different endopeptidases lead to the formation of nonstructural proteins (such as nonstructural proteins nsp3 and nsp5); furthermore, ribose-5-phosphate is produced through exonuclease activity, and new nucleotides are synthesized in association with methyltransferase, RNA polymerase and helicase functions for viral replication and prevention of supertwisting, and transcription is regulated through zinc finger proteins (Snijder et al. 2016).

```
LUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV:       MDLFMRFFTLRSITAQPVKIDNASPASTVHATATIPLQASLPFGWLVIGVAFLAVFQSAT   60
SARS-CoV-2:     MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFGWLIVGVALLAVFQSAS   60
Bat             MDLFMRIFTLGTVTLKQGEIKDATPSDSVRATATIPIQASLPFGWLIVGVAFLAVFQSAS   60
                ******:**: ::* :  :*.:*:*:. *:******:*********::***:*******:

SARS-CoV:       KIIALNKRWQLALYKGFQFICNLLLLFVTIYSHLLLVAAGMEAQFLYLYALIYFLQCINA   120
SARS-CoV-2:     KIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLEAPFLYLYALVYFLQSINF   120
Bat             KIITLKKRWQLALSKGIHFICNLLLLFVTVYSHLLLVAAGLEAPFLYLYALVYFLQSINF   120
                ***:*:******* **.:*:*********:**********:** *******:****.**

SARS-CoV:       CRIIMRCWLCWKCKSKNPLLYDANYFVCWHTHNYDYCIPYNSVTDTIVVTEGDGISTPKL   180
SARS-CoV-2:     VRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPYNSVTSSIVITSGDGTTSPIS   180
Bat             VRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPYNSVTSSIVITSGDGTTSPIS   180
                 ***** ******:***********:****: **********.:**:*.*** ::*

SARS-CoV:       KEDYQIGGYSEDRHSGVKDYVVVHGYFTEVYYQLESTQITTDTGIENATFFIFNKLVKDP   240
SARS-CoV-2:     EHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEP   240
Bat             EHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEP   240
                :.*******:*. .***** **:*.***. **** ***::****:*:.****:**:*.:*

SARS-CoV:       P-NVQIHTIDGSSGVANPAMDPIYDEPTTTTSVPL    274
SARS-CoV-2:     EEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL    275
Bat             EEHVQIHTIDGSSGVVNPAMEPIYDEPTTTTSVPL    275
                  :************.**.*:**************
```

Fig. 1 Sequence alignment of the orf3a accessory protein of SARS-CoV-2, SARS-CoV and the bat coronavirus RaTG13

```
CLUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV-2:     MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS   60
Bat             MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS   60
SARS-CoV:       MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPTVYVYS   60
                ************************************************:.****

SARS-CoV-2:     RVKNLNSSR-VPDLLV       75
Bat             RVKNLNSSR-VPDLLV       75
SARS-CoV:       RVKNLNSSEGVPDLLV       76
                ********. ******
```

Fig. 2 Sequence alignment of the envelope structural proteins of SARS-CoV-2, SARS-CoV and the bat coronavirus RatG13

The spike protein

The spike protein of SARS-CoV-2 contains 1273 amino acids, while the spike protein of SARS-CoV contains 1255 amino acids and that of BatCoV RaTG13 contains 1269 amino acids (Table 1). The spike protein of SARS-CoV-2 and that of SARS-CoV has an identity percentage of 76% and a similarity percentage of 86% (Table 2). The identity and similarity percentages of the spike protein of SARS-CoV-2 and the spike protein of RaTG13 are 97.4% and 98.4%, respectively (Table 2) [Annex 2 in ESM]. The identity and similarity percentages of the spike protein of SARS-CoV-2 and RaTG13 are higher than those of the spike protein of SARS-CoV-2 and SARS-CoV.

The spike protein of coronaviruses consists of three polypeptide chains with two domains: S1 and S2. The S1 and S2 domains are critical for binding host cell receptors (S1) and for fusing the virus with the membrane of the host cell. There is a hinge region between S1 and S2 that is a target for host cell proteases (Li 2016; Bosch et al. 2003). The spike

```
CLUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV:       -MADNGTITVEELKQLLEQWNLVIGFLFLAWIMLLQFAYSNRNRFLYIIKLVFLWLLWPV   59
SARS-CoV-2:     MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNRFLYIIKLIFLWLLWPV   60
Bat             -MADNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNRFLYIIKLIFLWLLWPV   59
                   .**********:**************:** ******:***********:********

SARS-CoV:       TLACFVLAAVYRINWVTGGIAIAMACIVGLMWLSYFVASFRLFARTRSMWSFNPETNILL   119
SARS-CoV-2:     TLACFVLAAVYRINWITGGIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILL   120
Bat             TLACFVLAAVYRINWITGGIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILL   119
                ***************:**********:*********:**********************

SARS-CoV:       NVPLRGTIVTRPLMESELVIGAVIIRGHLRMAGHSLGRCDIKDLPKEITVATSRTLSYYK   179
SARS-CoV-2:     NVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSYYK   180
Bat             NVPLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSYYK   179
                ****:***:****:**********:*****:*** ************************

SARS-CoV:       LGASQRVGTDSGFAAYNRYRIGNYKLNTDHAGSNDNIALLVQ     221
SARS-CoV-2:     LGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIALLVQ     222
Bat             LGASQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIALLVQ     221
                *******. *******.*************:.*.********
```

Fig. 3 The membrane proteins of SARS-CoV-2, SARS-CoV and bat coronavirus RaTG13 coloured according to the sequence alignment

```
CLUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV-2:     MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSLTENKYSQLDEEQPMEI   60
Bat             MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSLTENKYSQLDEEQPMEI   60
SARS-CoV:       MFHLVDFQVTIAEILIIIMRTFRIAIWNLDVIISSIVRQLFKPLTKKNYSELDDEEPMEL   60
                ***************:******:::***** **. *:::* * **:::**:**:*:***:

SARS-CoV-2:     D--    61
Bat             D--    61
SARS-CoV:       DYP    63
                *
```

Fig. 4 Sequence alignment of the orf6 accessory protein of SARS-CoV-2, SARS-CoV and the bat coronavirus RaTG13

protein of SARS-CoV-2 has a furin cleavage site in the hinge region. The furin cleavage site is composed of four amino acids (681–684). The presence of the furin cleavage site may be critical for the high transmission rate of SARS-CoV-2 compared to other coronaviruses (Walls et al. 2020).

Orf3a

The accessory protein orf3a of SARS-CoV-2 contains 275 amino acids, and its gene (25393.0.26220) is located between the spike and E protein genes. The orf3a protein of SARS-CoV contains 274 amino acids, while the NS3 of BatCoV RaTG13 is composed of 275 amino acids (Table 1). The amino sequence alignment of orf3a of SARS-CoV-2 and SARS-CoV showed that the sequence identity was 72.4% and that the sequence similarity was 85.1%. The similarity percentage of orf3a in SARS-CoV-2 and SARS-CoV was 90.2, not 85.1% as reported by Yashimito (2020), which may be due to the different software programs used in the two studies (Yoshimoto 2020). Orf3a (SARS-CoV-2) and NS3 (BatCoV RaTG13) were characterized by 97.8% identity and 98.9% similarity (Table 2, Fig. 1).

Orf3a plays different roles in the virus including 1) viral envelope assembly and 2) host cell binding and infusion by interacting with the structural proteins (M, S, and E) and the accessory protein (7a) of SARS-CoV (Brunn et al. 2007). In host organisms, the highest immunogenicity of the N-terminus of orf3a is known to have a strong

```
CLUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV:       MKIILFLTLIVFTSCELYHYQECVRGTTVLLKEPCPSGTYEGNSPFHPLADNKFALTCTS   60
SARS-CoV-2:     MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNSPFHPLADNKFALTCFS   60
Bat             MKIILFLVLVTLATCELYHYQECVRGTTVLLKEPCSSGTYEGNSPFHPLADNKFALTCFS   60
                *******.*:.:::********************** ********************** *

SARS-CoV:       THFAFACADGTRHTYQLRARSVSPKLFIRQEEVQQELYSPLFLIVAALVFLILCFTIKRK   120
SARS-CoV-2:     TQFAFACPDGVKHVYQLRARSVSPKLFIRQEEV-QELYSPIFLIVAAIVFITLCFTLKRK   119
Bat             TQFAFACPDGVKHVYQLRARSVSPKLFIRQEEV-QELYSPIFLIIAAIVFITLCFTLKRK   119
                *:***** **.:*.****************** ******:***:**:**: ****:***

SARS-CoV:       TE     122
SARS-CoV-2:     TE     121
Bat             TE     121
                **
```

Fig. 5 The orf7a accessory protein of SARS-CoV-2, SARS-CoV and bat coronavirus RaTG13 and its alignment

Fig. 6 Sequence alignment of the orf7b accessory protein of SARS-CoV-2, SARS-CoV and the bat coronavirus RaTG13

```
CLUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV-2:     MIELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNETCHA-   43
Bat             MSELSLIDFYLCFLAFLLFLVLIMLIIFWFSLELQDHNETCHA-   43
SARS-CoV:       MNELTLIDFYLCFLAFLLFLVLIMLIIFWFSLEIQDLEEPCTKV   44
                * **:*****************************:** :* *
```

```
CLUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV-2:     MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIEL   60
Bat             MKLLVFLGILTTVTAFHQECSLQSCAQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIEL   60
SARS-CoV:       MKLLIVLTCISLCSC--ICTVVQRCASNKPHVLEDPCKVQHSARS-CoVYP-o---rfbM   53
                **:*:.*  ::  :.      :* *:.::*:*::*** ::. ::    *        : :

SARS-CoV-2:     CVD------EAGSKS----------PIQYIDIGNYTVSCL-PFTINCQEPKLGSLVVRCS   103
Bat             CVD------EVGSKS----------PIQYIDIGNYTVSCS-PFTINCQEPKLGSLVVRCS   103
SARS-CoV:       CLKILVRYNTRGNTYSTAWLCALGKVLPFHRWHTMVQTCTPNVTINCQDPAGGALIARCW   113
                *:.        *..            : :    . . :*   .*****:*  *:*:.**

SARS-CoV-2:     FYEDFLEYHDVRVVLDFI-----        121
Bat             FYEDFLEYHDVRVVLDFI-----        121
SARS-CoV:       YLHEGHQTAAFRDVLVVLNKRTN        136
                : .:  :   .* ** .:
```

Fig. 7 The orf8 accessory protein of SARS-CoV-2, SARS-CoV and bat coronavirus RaTG13 and its alignment

protective effect on humoural immunity (Zhong et al. 2006). Orf3a has a cysteine-rich domain that possesses potassium ion channel activity by interacting with the S and E proteins (Brunn et al. 2007; Zeng et al. 2004). The C-terminus of orf3a arrests the host cell cycle by depleting cyclin D3 and facilitates apoptosis of host cells by interacting with the M protein (Yuan et al. 2007; Marra et al. 2003; Law et al. 2005).

```
CLUSTAL O(1.2.4) multiple sequence alignment


SARS-CoV-2:      MSDNGPQ-NQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQH   59
Bat              MSDNGPQ-NQRNAPRITFGGPSDSTGSNQNGERSGARPKQRRPQGLPNNTASWFTALTQH   59
SARS-CoV:        MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQH   60
                 ******* ***.*********:***..**** *.*** **********************

SARS-CoV-2:      GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA   119
Bat              GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA   119
SARS-CoV:        GKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA   120
                 ***:*:*************.**************:********:*****************

SARS-CoV-2:      GLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGG   179
Bat              GLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGG   179
SARS-CoV:        SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLPKGFYAEGSRGG   180
                 .*******:**:********************* **** **********************

SARS-CoV-2:      SQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQ   239
Bat              SQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGSDAALALLLLDRLNQLESKMSGKGQ   239
SARS-CoV:        SQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSGKGQ   240
                 ***********..**********.******..*.::*****************:*****

SARS-CoV-2:      QQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
Bat              QQQSQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK   299
SARS-CoV:        QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYK   300
                 ***.********************** **********************:*********

SARS-CoV-2:      HWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDA   359
Bat              HWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDA   359
SARS-CoV:        HWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA   360
                 ********************************** **********:***:**********

SARS-CoV-2:      YKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA   419
Bat              YKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA   419
SARS-CoV:        YKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADST   420
                 ****************:**:* ******** *********:****:***:***.*.: .:

SARS-CoV-2:      --     419
Bat              --     419
SARS-CoV:        QA     422
```

Fig. 8 The nucleocapsid structural protein of SARS-CoV-2, SARS-CoV and bat coronavirus RaTG13 and the coloured sequence alignment

Envelope protein (E protein)

The E protein of SARS-CoV-2, SARS-CoV, and BatCoV RaTG13 consists of 75, 76, and 75 amino acids, respectively (Table 1). The percentages of the identity and similarity of the E protein inSARS-CoV-2 and SARS-CoV are 94.7 and 96.1, respectively; these percentages were 94.7 and 97.4 in Yoshimoto (2020) (Table 2, Fig. 2). The E protein of SARS-CoV-2 and the E protein of BatCoV RaTG13 are 100% identical and similar. The results strongly favour a bat origin of SARS-CoV-2 over a SARS-CoV origin. The E protein contains three domains, the C-terminus, N-terminus, and transmembrane, with different functions in the virus and in host cells (Schoeman and Fielding 2019).

The E protein plays different functions in viral replication and the interaction of the virus with host organisms and cells, such as assembly of the virion envelope, suppression of host cell stress responses, facilitation of viral replication and vitality, and as an ion channel to induce the release of virions from host cells (Nieto-Torres et al. 2011; Álvarez et al. 2010; Corse and Machamer 2003; Yuan et al. 2006; Ruch and Machamer 2012).

Table 2 The percentages of identity and similarity of the SARS-CoV-2 proteins compared to those of SARS-CoV and RaTG13 (bat coronavirus)

	Protein		Identity %	Similarity %
1	1 ab polyprotein	SARS-CoV-2 and SARS-CoV	86.2	92.9
		SARS-CoV-2 and RaTG13	98.5	99.1
2	Spike protein	SARS-CoV-2 and SARS-CoV	76	86
		SARS-CoV-2 and RaTG13	97.4	98.4
3	Orf3a (NS3)	SARS-CoV-2 and SARS-CoV	72.4	85.1
		SARS-CoV-2 and RaTG13	97.8	98.9
4	E protein	SARS-CoV-2 and SARS-CoV	94.7	96.1
		SARS-CoV-2 and RaTG13	100	100
5	M protein	SARS-CoV-2 and SARS-CoV	90.5	96.4
		SARS-CoV-2 and RaTG13	99.5	99.5
6	Orf6 (NS6)	SARS-CoV-2 and SARS-CoV	68.9	88.5
		SARS-CoV-2 and RaTG13	100	100
7	Orf7a (NS7a)	SARS-CoV-2 and SARS-CoV	85.2	90.2
		SARS-CoV-2 and RaTG13	97.5	99.2
8	Orf7b (NS7b)	SARS-CoV-2 and SARS-CoV	85.4	90.2
		SARS-CoV-2 and RaTG13	97.7	97.7
9	Orf8 (NS8)	SARS-CoV-2 and SARS-CoV	38.9 44.4	77.8 66.7
		SARS-CoV-2 and RaTG13	95	95.9
10	N protein	SARS-CoV-2 and SARS-CoV	90.5	94.3

Membrane protein

The membrane protein (M) of SARS-CoV-2, SARS-CoV, and BatCoV RaTG13 is composed of 222, 221, and 221 amino acids, respectively (Table 1). The percentages of identity and similarity of the amino acid sequences of the M protein of SARS-CoV-2 and SARS-CoV are 90.5 and 96.4, respectively, while those of the M protein of SARS-CoV-2 and BatCoV RaTG13 are both 99.5% (Table 2, Fig. 3). The M protein has three domains, the N-terminus, C-terminus, and transmembrane, with different functions (Neuman et al. 2010).

The M protein is important for the assembly, transport, and release of the virus from host cell organelles (Ma et al. 2008; Siu et al. 2008). The M protein of SARS-CoV inhibits the transcription of interferon-1, which leads to the inhibition of the innate immunity of host organisms (Siu et al. 2009).

Orf6

The orf6 protein of SARS-CoV-2 and BatCoV RaTG13 contains 61 amino acids, while it contains 63 amino acids in SARS-CoV (Table 1). The orf6 protein of SARS-CoV-2 and SARS-CoV is characterized by an identity percentage of 68.9% and a similarity percentage of 88.5%. The percentages of orf6 protein identity and similarity in SARS-CoV-2 and BatCoV RaTG13 are both 100% (Table 2, Fig. 4).

The functions of orf6 include (1) participation in the formation of replication/transcription to facilitate viral replication, (2) induction of an increase in the number of virions during infection, (3) contribution to virus evasion of the host immune system and (4) involvement in the formation of double-membrane vesicles (DMVs) in host cells to ensure virus assembly (Kumar et al. 2007; Narayanan et al. 2008; Gunalan et al. 2011).

Orf7a

Orf7a of SARS-CoV-2 and BatCoV RaTG13 contains 121 amino acids, while orf7a of SARS-CoV contains 122 amino acids (Table 1). The percentages of orf7a identity and similarity in SARS-CoV-2 and SARS-CoV are 85.2 and 90.2, respectively. The orf7a protein of SARS-CoV-2 and the orf7a protein of the BatCoV RaTG13 share an identity percentage of 97.5% and similarity percentage of 99.2% (Table 2, Fig. 5).

Orf 7a of SARS-CoV is a transmembrane protein divided into four regions from the N-terminus: (1) the first 15 amino acids are broken down by the infected host cells; (2) amino acids 16–96 form the intracellular domain; (3) amino acids 97–117 with a collective hydrophobic nature form the transmembrane domain; and 4) the C-terminus consists of the last five amino acids (Liu et al. 2014).

Orf7a plays a role in virus binding to and invasion of host cells by interacting with the S, M, E, and orf3a proteins (Narayanan et al. 2008; Tan et al. 2006). Orf7a does not

contribute to the replication of the virus (Liu et al. 2014; Tan et al. 2006; Yount et al. 2005; Schaecher et al. 2007). Orf7a plays some roles in host cells, such as triggering apoptosis, downregulating protein synthesis, arresting the cell cycle at the G0/G1 phase, and activating cytokine production (Narayanan et al. 2008; Liu et al. 2014; Tan et al. 2006; Schaecher et al. 2007).

Orf7b

The orf7b protein in both SARS-CoV-2 and BatCoV RaTG13 contains 43 amino acids, while orf7b in SARS-CoV contains 44 amino acids (Table 1). The orf7b protein of SARS-CoV-2 and the orf7b protein of SARS-CoV are characterized by an identity percentage of 85.4 and a similarity percentage of 90.2. On the other hand, the identity and similarity percentages of the orf7b protein of SARS-CoV-2 and that of BatCoV RaTG13 are 97.7% each (Table 2, Fig. 6).

The orf7b protein contains three domains: an N-terminal domain (external), a C-terminal domain (in the cytoplasm), and a transmembrane hydrophobic domain (Liu et al. 2014).

It has been reported that orf7b is not involved in virus replication (Liu et al. 2014; Tan et al. 2006; Yount et al. 2005; Schaecher et al. 2007). The anti-orf7b antibody concentration is increased in SARS-CoV patients, which shows that orf7b is highly immunogenic and can be used in vaccination trials (Schaecher et al. 2007; Guo et al. 2004).

Orf8

The identity and similarity of orf8 in the SARS-CoV-2 and orf8a in SARS-CoV are 38.9 and 77.8, respectively. The orf8 protein of SARS-CoV-2 is 44.4% identical and 66.7% similar to orf8b of SARS-CoV (Fig. 10). The identity and similarity of SARS-CoV-2 and BatCoV RaTG13 orf8 are 95% and 95.9%, respectively (Table 2, Fig. 7). However, the orf8 protein of SARS-CoV-2 has 121 amino acids compared to 39 and 84 amino acids in the orf8a protein variants of SARS-CoV, and 121 amino acids for orf8 of BatCoV RaTG13 (Table 1).

Orf8a and orf8b of SARS-CoV are not needed for viral replication. In host cells, they are localized in vesicle-like structures in mitochondria, the endoplasmic reticulum, cytosol and nucleus of host cells. orf8a and orf8b of SARS-CoV stimulate cellular DNA synthesis and caspase-dependent apoptosis (Keng and Tan 2009).

Nucleocapsid protein (N protein)

The N protein of SARS-CoV-2 and BatCoV RatG13 consists of 419 amino acids, while that of SARS-CoV consists of 422 amino acids (Table 1). The N protein of SARS-CoV-2 and the N protein of SARS-CoV are 90.5% identical and 94.3% similar, while those of SARS-CoV-2 and BatCoV RaTG13 are 99% identical and similar (Table 2, Fig. 8). The N-protein is an RNA-binding protein with three domains: an N-terminal domain that binds RNA, a C-terminal domain critical for dimerization, and a disordered central region rich in serine and arginine (SR) (Kang et al. 2020).

The N protein is essential for the formation of helical viral RNA, induction of the replication and transcription of the virus, and control of host cell metabolism to ensure the viral replication process and to regulate host cell apoptosis and the cell cycle (Kang et al. 2020; Cong et al. 2020; Surjit et al. 2006). Moreover, the N protein is very immunogenic and induces the host immune system to respond against SARS-CoV (Lin et al. 2003).

Conclusion

The SARS-CoV-2 proteins and the BatCoV RaTG13 share high identity and similarity compared to the SARS-CoV-2 and SARS-CoV proteins. The findings of this study proved the usefulness of determining the percentages of protein identity and similarity in determining the origin of viruses.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42485-021-00060-3.

Declarations

Conflict of interest The author declares no conflict of interest.

References

Ahmad S, Hafeez A, Siddqui SA, Ahmad M, Mishra S (2020) A review of COVID-19 (Coronavirus Disease-2019) diagnosis treatments and prevention. EJMO 4(2):116–125

Álvarez E, DeDiego ML, Nieto-Torres JL, Jiménez-Guardeño JM, Marcos-Villar L, Enjuanes L (2010) The envelope protein of severe acute respiratory syndrome coronavirus interacts with the non-structural protein 3 and is ubiquitinated. Virology 402(2):281–291

Bosch BJ, van der Zee R, de Haan CA, Rottier PJ (2003) The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J Virol. 77(16):8801–8811

Cheng VCC, Wong SC, Chen JHK, Yip CCY, Chuang VWM, Tsang OTY, Sridhar S, Chan JFW, Ho P, Yuen K (2020) Escalating infection control response to the rapidly evolving epidemiology of the coronavirus disease 2019 (COVID-19) due to SARS-CoV-2 in Hong Kong. Infect Control Hosp Epidemiol 41(5):493–498

Cong YY, Ulasli M, Schepers H, Mauthe M, V'kovski P, Kriegenburg F, Thiel V, de Haan CAM, Reggiori F (2020) Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in corona viral life cycle. J Virol 94(4):01925–02019

Corse E, Machamer CE (2003) The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. Virology 312(1):25–34

Fahmi M, Kubota Y, Ito M (2020) Nonstructural proteins NS7b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV. Infect Genet Evol. 81:104272

GenBank: MN996532.1. Bat coronavirus RaTG13, complete genome. https://www.ncbi.nlm.nih.gov/nuccore/MN996532. Last updated on 24 March 2020. Accessed on 4 Jul 2020

Gunalan V, Mirazimi A, Tan Y (2011) A putative diacidic motif in the SARS-CoV ORF6 protein influences its subcellular localization and suppression of expression of co-transfected expression constructs. BMC Res Notes 4:446. https://doi.org/10.1186/1756-0500-4-446

Guo JP, Petric M, Campbell W, McGeer PL (2004) SARS corona virus peptides recognized by antibodies in the sera of convalescent cases. Virology 324(2):251–256

International Committee on Taxonomy of Viruses (ICTV) (2020) ICTV 2019 Master Species List (MSL35). https://talk.ictvonline.org/files/master-species-lists/m/msl/9601

Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X, He S, Zhou Z, Zhou Z, Chen Q, Yan Y, Zhang C, Shan H, Chen S (2020) Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharmaceutica Sinica B. https://doi.org/10.1016/j.apsb.2020.04.009

Keng CT, Tan YJ (2009) Molecular and biochemical characterization of the SARS-CoV accessory proteins ORF8a, ORF8b and ORF8ab. Mol Biol SARS-Coronavirus 177–191

Khailany RA, Safdar M, Ozaslan M (2020) Genomic characterization of a novel SARS-CoV-2. Gene Rep 19:100682

Kumar P, Gunalan V, Liu B, Chow VTK, Druce J, Birch C, Catton M, Fielding BC, Tan YJ, Lal SK (2007) The nonstructural protein 8 (nsp8) of the SARS coronavirus interacts with its ORF6 accessory protein. Virology 366(2):293–303

Law PTW, Wong CH, Au TCC, Chuck C, Kong S, Chan PKS, To K, Lo AWI, Chan JYW, Suen Y, Chan HYE, Fung K, Waye MMY, Sung JJY, Lo YMD, Tsui SKW (2005) The 3a protein of severe acute respiratory syndrome-associated coronavirus induces apoptosis in Vero E6 cells. J Gen Virol. 86(Pt 7):1921–1930

Li F (2016) Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virol. 3(1):237–261

Li Y, Yang X, Wang N, Wang H, Yin B, Yang X, Jiang W (2020) The divergence between SARS-CoV-2 and RaTG13 might be overestimated due to the extensive RNA modification. Future Virol. https://doi.org/10.2217/fvl-2020-0066

Lin Y, Shen X, Yang RF, Li YX, Ji YY, He YY, Shi MD, Lu W, Shi TL, Wang J, Wang HX, Jiang HL, Shen JH, Xie YH, Wang Y, Pei G, Shen BF, Wu JR, Sun B (2003) Identification of an epitope of SARS-coronavirus nucleocapsid protein. Cell Res 13:141–145

Liu DX, Fung TS, Chong KK, Shukla A, Hilgenfeld R (2014) Accessory proteins of SARS-CoV and other coronaviruses. Antiviral Res 109:97–109

Liu Y, Zhi Z, Chen Y, Guo M, Liu Y, Gali NK, Sun L, Duan Y, Cai J, Westerdahl D, Liu X, Xu K, Ho K, Kan H, Fu Q, Lan K (2020) Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. Nature 582:557–560

Ma H, Fang C, Hsieh Y, Chen S, Li H, Lo S (2008) Expression and membrane integration of SARS-CoV M protein. J Biomed Sci 15:301–310

Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, Khattra J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, Vogrig R, Ward D, Watson B, Brunham RC, Krajden M, Petric M, Skowronski DM, Upton C, Roper RL (2003) The Genome sequence of the SARS-associated coronavirus. Science 300(5624):1399–1404

Narayanan K, Huang C, Makino S (2008) SARS coronavirus accessory proteins. Virus Res. 133:113–121

Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, Droese B, Klaus JP, Shinji Makino S, Sawicki SG, Siddell SG, Stamou DG, Wilson IA, Kuhn P, Buchmeier MJ (2010) A structural analysis of M protein in coronavirus assembly and morphology. J Struct Biol 174:11–22

Nieto-Torres JL, DeDiego ML, Álvarez E, Jiménez-Guardeño JM, Regla-Nava JA, Llorente M, Kremer L, Shuo S, Enjuanes L (2011) Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. Virology 415(2):69–82

Ong SWX, Tan YK, Chia PY, Lee TH, Ng OT, Wong MSY, Marimuthu K (2020) Air, Surface environmental, and personal protective equipment contamination by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from a symptomatic patient. JAMA 323(16):1610–1612

Perlman S, Netland J (2009) Coronaviruses post-SARS: update on replication and pathogenesis. Nat Rev Microbiol. 7:439–450

Ruch TR, Machamer CE (2012) The coronavirus E protein: assembly and beyond. Viruses 4(3):363–382

Schaecher SR, Touchette E, Schriewer J, Buller RM, Pekosz A (2007) Severe acute respiratory syndrome coronavirus gene 7 products contribute to virus-induced apoptosis. J Virol 81(20):11054–11068

Schoeman D, Fielding BC (2019) Coronavirus envelope protein: current knowledge. Virol J 16:69

Shereen MA, Khan S (2020) COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. J Adv Res 24:91–98

Siu YL, Teoh KT, Lo J, Chan CM, Kien F, Escriou N, Tsao SW, Nicholls JM, Altmeyer R, Peiris JSM, 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(22):11318–11330

Siu K, Kok K, Ng MJ, Poon VKM, Yuen K, Zheng B, Jin D (2009) Severe acute respiratory syndrome coronavirus M protein inhibits type I interferon production by impeding the formation of TRAF3•TANK•TBK1/IKKϵ complex. J Biol Chem 284:16202–16209

Snijder EJ, Decroly E, Ziebuhr J (2016) The nonstructural proteins directing coronavirus RNA synthesis and processing. Adv Virus Res. 96:59–126

Surjit M, Liu B, Chow VT, Lal SK (2006) The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells. J Biol Chem. 281:10669–10681

Tan YJ, Lim SG, Hong W (2006) Understanding the accessory viral proteins unique to the severe acute respiratory syndrome (SARS) coronavirus. Antiviral Res 72(2):78–88

Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J, Lu J (2020) On the origin and continuing evolution of SARS-CoV-2. NATL Sci Rev 7(6):1012–1023

van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, Tamin A, Harcourt JL, Thornburg NJ, Gerber SI, Lloyd-Smith JO, de Wit E, Munster VJ (2020) Aerosol and surface stability of SARS-CoV-2 as Compared with SARS-CoV-1. N Engl J Med. 382:1564–1567

Velavan TP, Meyer CG (2020) The COVID-19 epidemic. Trop Med Int Health 25(3):278–280

von Brunn A, Teepe C, Simpson JC, Pepperkok R, Friedel CC, Zimmer R, Roberts R, Baric R, Haas J (2007) Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. PLoS ONE 2(5):e459

Walls AC, Park Y, Tortorici MA, Wall A, McGuire AT, Veesler D (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 181(2):281–292

Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, Zhang Z (2020) The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol 92:667–674

WHO (2020) WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/. Accessed on 3 Jul 2020

Wong G, Bi YH, Wang QH, Chen XW, Zhang ZG, Yao YG (2020) Zoonotic origins of human coronavirus 2019 (HCoV-19/SARS-CoV-2): why is this work important? Zool Res 41(3):213–219

Yoshimoto FK (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

Yount B, Roberts RS, Sims AC, Deming D, Frieman MB, Sparks J, Denison MR, Davis N, Baric RS (2005) Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice. J Virol 79(23):14909–14922

Yuan Q, Liao Y, Torres J, Tam JP, Liu DX (2006) Biochemical evidence for the presence of mixed membrane topologies of the severe acute respiratory syndrome coronavirus envelope protein expressed in mammalian cells. FEBS Lett 580(13):3192–3200

Yuan X, Yao Z, Wu J, Zhou Y, Shan Y, Dong B, Zhao Z, Hua P, Chen J, Cong Y (2007) G1 phase cell cycle arrest induced by SARS-CoV 3a protein via the cyclin D3/pRb pathway. Am J Respir Cell Mol Biol. 37(1):9–19

Zeng R, Yang RF, Shi MD, Jiang M, Xie Y, Ruan H, Jiang X, Shi L, Zhou H, Zhang L, Wu X, Lin Y, Ji Y, Xiong L, Jin Y, Dai E, Wang X, Si B, Wang J, Wang H, Wang C, Gan Y, Li Y, Cao J, Zuo J, Shan S, Xie E, Chen S, Jiang Z, Zhang X, Wang Y, Pei G, Sun B, Wu J (2004) Characterization of the 3a protein of SARS-associated coronavirus in infected vero E6 cells and SARS patients. J Mol Biol. 341(1):271–279

Zhong X, Guo Z, Yang H, Peng L, Xie Y, Wong T, Lai S, Guo Z (2006) Amino terminus of the SARS coronavirus protein 3a elicits strong, potentially protective humoral responses in infected patients. J Gen Virol 87:369–373

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273

Authors and Affiliations

Mohammed Elimam Ahamed Mohammed[1,2]

[1] Department of Chemistry, Faculty of Science, King Khalid University, Abha, Saudi Arabia

[2] Unit of Bee Research and Honey Production, Faculty of Science, King Khalid University, Abha, Saudi Arabia