

RESEARCH

Open Access



Combination of Biodata Mining and Computational Modelling in Identification and Characterization of ORF1ab Polyprotein of SARS-CoV-2 Isolated from Oronasopharynx of an Iranian Patient

Reza Zolfaghari Emameh^{1*}, Hassan Nosrati² and Ramezan Ali Taheri³

Abstract

Background: Coronavirus disease 2019 (COVID-19) is an emerging zoonotic viral infection, which was started in Wuhan, China, in December 2019 and transmitted to other countries worldwide as a pandemic outbreak. Iran is one of the top ranked countries in the tables of COVID-19-infected and -mortality cases that make the Iranian patients as the potential targets for diversity of studies including epidemiology, biomedical, biodata, and viral proteins computational modelling studies.

Results: In this study, we applied bioinformatic biodata mining methods to detect CDS and protein sequences of ORF1ab polyprotein of SARS-CoV-2 isolated from oronasopharynx of an Iranian patient. Then through the computational modelling and antigenicity prediction approaches, the identified polyprotein sequence was analyzed. The results revealed that the identified ORF1ab polyprotein belongs to a part of nonstructural protein 1 (nsp1) with the high antigenicity residues in a glycine-proline or hydrophobic amino acid rich domain.

Conclusions: The results revealed that nsp1 as a virulence factor and crucial agent in spreading of the COVID-19 among the society can be a potential target for the future epidemiology, drug, and vaccine studies.

Keywords: SARS-CoV-2, COVID-19, ORF1ab, nsp1, Biodata mining, Protein Modelling

Introduction

Coronaviruses (CoVs) are positive strand RNA viruses belong to the order of Nidovirales and three families including *Arteriviridae*, *Coronaviridae*, and *Roniviridae* [1]. Based on the genetic studies, CoVs are classified to into four genera including alpha, beta, gamma, and delta CoVs. The diameter of CoVs is between 80 to 120 nm and their shape is spherical. The spike projections of

these virions give the appearance of solar corona to the CoVs. The main structural proteins of CoVs are envelope (E), membrane (M), nucleocapsid (N), and spike (S). The S proteins comprise N-linked signal peptide to be transferred to endoplasmic reticulum (ER) and consequently glycosylated in ER [2]. The homotrimeric structure of S glycoproteins on the surface of the CoVs mediate the attachment of virions to the cell receptors [3]. The size of positive-sense RNA genome of CoVs is between 26.2 and 31.7 kb. The RNA genome composes of six to ten open reading frames (ORFs). ORF1a as the longest part of the RNA encodes for the replicases and

* Correspondence: zolfaghari@nigeb.ac.ir

¹Department of Energy and Environmental Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), 14965/161, Tehran, Iran
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

ORF1b expresses for two large polyproteins including pp1a and pp1ab comprising about 4000 and 7000 amino acids. The expression of pp1ab polyprotein is essential for programmed ribosomal frame shifting signal by bridging between ORF1a and ORF1ab [4]. In the CoVs, the frameshifting signal is led to the expression of a RNA-dependent RNA polymerase (RdRP), which is required for the coronavirus replication [5]. The polyproteins of CoVs are cleaved by virus-encoded cysteine proteinases comprise papain- and chymotrypsin-like proteases into 16 nonstructural proteins (nsp) including the expression of nsp1 to nsp11 by ORF1a and encoding nsp12 to nsp16 by ORF1b [6]. The nsp3, nsp4, and nsp6 contain hydrophobic transmembrane domains, which are considered as the anchor sites of pp1a and pp1ab polyproteins to membranes during the first step of formation of replication-transcription complexes (RTC). Further study defined that two out of three hydrophobic domains in nsp3 and six out of seven hydrophobic domains in nsp6 span the membrane, while four hydrophobic domains in nsp4 span to lipid bilayer [7]. On the other hand, ORF1b-encoded nsps including nsp12 has the RdRP activity, nsp13 has the helicase activity, nsp14 has the 3' to 5' exonuclease and RNA cap N7-guanine methyltransferase and activities for proofreading in association with nsp7/nsp8/nsp12 complex, and nsp15 has the endoribonuclease activity. The nsp16 has the methyltransferase activity, which in combination with helicase/triphosphatase, nsp13, and 2'-O-MTase, a replication-transcription machinery is constituted to enable the CoVs in the RNA synthesis and processing steps [8].

CoVs cause zoonotic lethal human respiratory infections [9]. Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) was the causative agent of 2002–2003 outbreak that occurred in the Guangdong Province of China with mortality rate of 9% and 774 total deaths [10]. It is accepted that SARS-CoV was originated in Chinese bats that contain SARS-related CoVs with angiotensin converting enzyme 2 (ACE2) as the same host receptor, although the population working in the wet animal markets were the seropositive cases. In 2012, the CoVs were mutated to Middle East Respiratory Syndrome Coronavirus (MERS-CoV) or camel flu and obtained the human-to-human capability from the camel origin with mortality rate of 40% and 333 total deaths. The host cell receptor for MERS-CoV is Dipeptidyl peptidase 4 (DPP4), which is present in some other animal cells including bats, camels, horses, and rabbits [11, 12]. Up to 2019, the positive cases of MERS-CoV infection were 2374 and 823 total deaths from 27 countries [13]. Since the mouse model doesn't express the DPP4 cell receptor, the vaccine studies against the MERS-CoV infection were focused on other vaccine model animals including *Macaca mulatta* (Rhesus macaques) [14, 15], *Callithrix jacchus* (common marmoset) [15–17], *Camelus dromedarius*

(Dromedary camels) [18], hDPP4-transduced mice [19], transgenic mice expressing hDPP4 globally [20], hDPP4-humanized transgenic mice [21], CRISPR/Cas9-engineered mice [22], and hDPP4-knockin mice using CRISPR/Cas9 [23]. Since the big animals are not economic and easy handling, it is preferred that the smaller model animals with available testing vaccine efficiency methods to be applied in the MERS-CoV vaccine studies [24]. In addition, some potential vaccine candidates were produced against MERS-CoV infection using viral vectors including recombinant human adenovirus encoding for S protein [25–27], recombinant chimpanzee adenovirus encoding for S protein [28, 29], modified vaccinia virus Ankara encoding for S protein [29–31] and N protein [32], recombinant human adenovirus encoding for S protein with nanoparticle [33], DNA vaccine encoding for S protein [34–36], subunit vaccines for S protein, receptor binding domain of S protein, and recombinant N-terminal domain [37–48], virus-like particles encoding for S protein with nanoparticles [49–53], nanoparticles with ferritin displaying receptor-binding domain of S protein [54], inactivated whole- MERS-CoV [55–57], and live-attenuated MERS-CoV [58–62].

The phylogenetic studies and sequence analyses of SARS-CoV-2 and some SARS-related CoVs revealed that all use ACE2 as the host cell receptor [63]. Evolutionary, human SARS-CoVs and bat SARS-CoVs such as LYRa11, Rs3367, Rf1, Cp and Rp3 share a common ancestor, while SARS-CoV and MERS-CoV are distantly related to each other [64, 65]. Receptor-binding domain (RBD) of S protein, which is responsible for binding to ACE2 of cell host receptor, is considered as the major part evolving in the beta CoVs so 29 unique RBDs were phylogenetically identified in three distinct clades [66].

Based on the outbreak of SARS-CoV-2 as a novel member of CoVs in December 2019 in Wuhan, China, the causative agent of coronavirus disease 2019 (COVID-19), severity of symptoms, high human-to-human transmission rate, pandemic epidemiological situation, and high mortality rate (>2,000,000 infected cases and >120,000 deaths worldwide till mid-April 2020) [67], it is an urge to study SARS-CoV-2 in all aspects to discover potential pharmaceutical and vaccine candidates against COVID-19.

In this study, we evaluated the partial DNA sequence and the encoded ORF1ab polyprotein isolated from the oronasopharynx of an Iranian patient through combination of biodata mining and computational modelling methods to identify whether a potential domain is available for the stimulation of human immune system to consider it as a potential target of drug and vaccine studies.

Results

Sequence Analysis

The multiple sequence alignment (MSA) analysis of the coding sequence (CDS) from the Iranian patient

(Accession Number: MT152900) and the Chinese CDS query (Accession Number: NC_045512.2) revealed that the Iranian CDS sequence was located between nucleotides 237 and 558 (total length: 322 bases) of the Chinese CDS query with 100% sequence identity (Fig. 1a). In addition, the MSA analysis of the partial ORF1ab polyprotein from the Iranian patient (Accession Number: QIH55230) and the query protein sequence from Wuhan, China (Accession Number: YP_009724389.1) revealed that the Iranian protein sequence was a part of ORF1ab polyprotein from Wuhan, China (Accession Number: YP_009724389.1) with 100% sequence identity (Fig. 1b) (The complete information related to the Fig. 1a and b has been presented in the supplementary material 1).

Protein Modelling

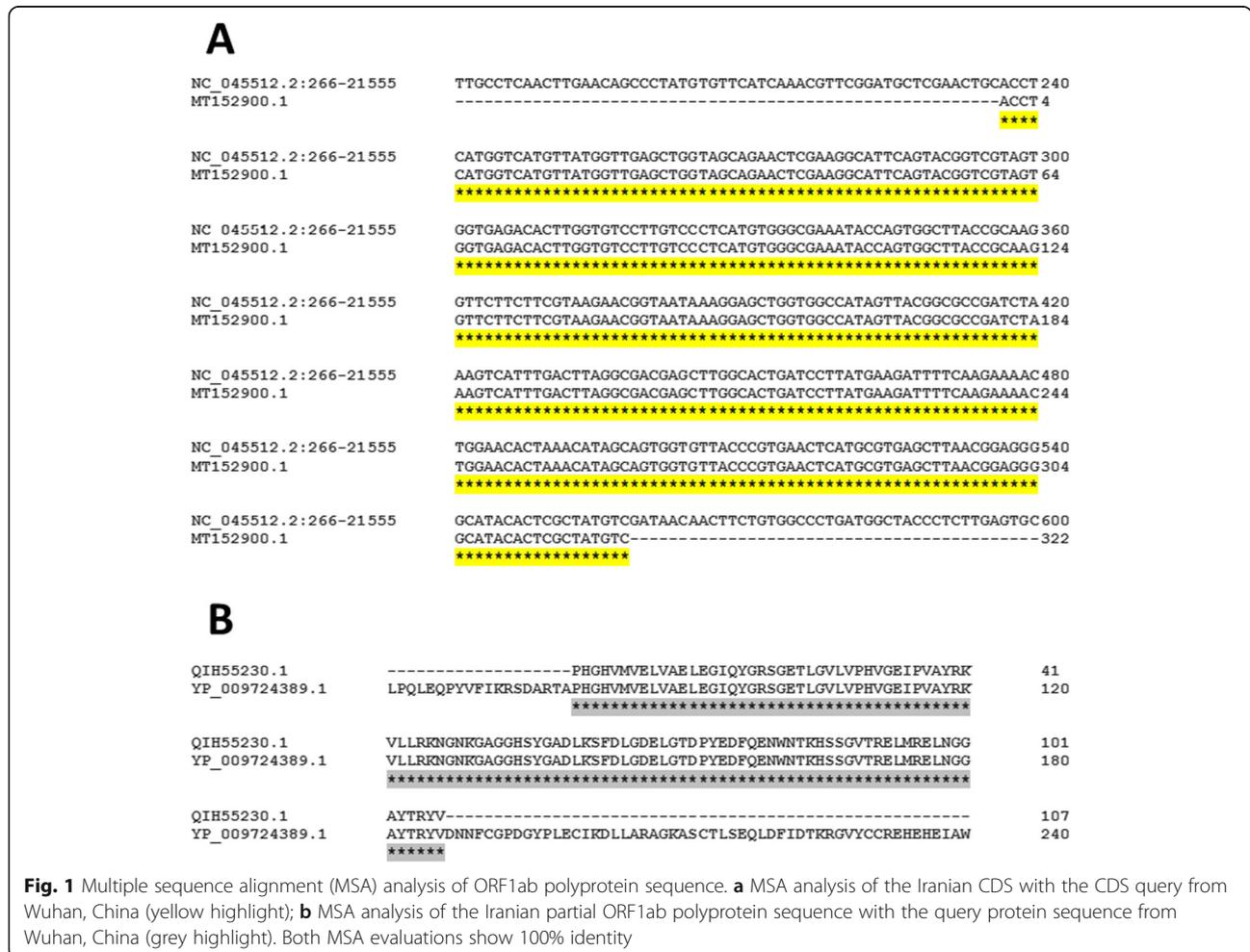
The protein modelling of the partial ORF1ab polyprotein sequence from the Iranian patient (Accession Number: QIH55230) in the RCSB PDB Protein Data Bank revealed that a 47-amino acid sequence of NMR entry ID: 2GDT belonged to nsp1 from the SARS-CoV with *E*-

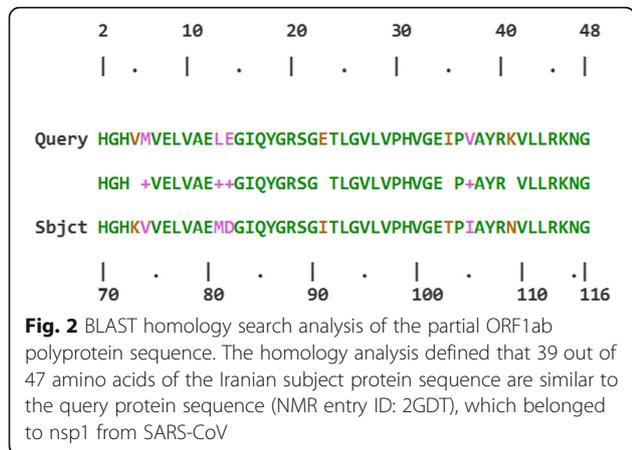
value: 4.20992E-16 and 83% identity, was the most similar model for the partial ORF1ab polyprotein sequence from the Iranian patient (Accession Number: QIH55230) (Fig. 2).

The visualization of the partial ORF1ab polyprotein sequence from the Iranian patient (Accession Number: QIH55230) by the NGL (WebGL) viewer revealed that the 3D model of the subject protein sequence has a considerable overlap with the query sequence. This 3D model overlap demonstrated that the partial ORF1ab polyprotein sequence from the Iranian patient (Accession Number: QIH55230) has a protein structure with close similarity to the nsp1 from SARS-CoV (Fig. 3).

Antigenicity Prediction

The antigenicity prediction of partial ORF1ab polyprotein sequence from the Iranian patient (Accession Number: QIH55230) defined three antigenic domains including 22-, 13-, and 7-amino acids sequences. The most antigenic domain was the 22-amino acids domain located between Thr92 and Arg113 as the most





antigenic domain. In the first ranked antigenic domain, 13 out of 22 amino acids were hydrophobic (Table 1).

Further evaluation of the constituent amino acids of first ranked antigenic domain defined that the hydrophobic and hydrophilic amino acids were located at the surface of the subject protein and consequently access of human immune system, while other amino acids were buried and out of human immune system (Fig. 4).

Discussion

SARS-CoV-2 infected many people from several cities of Iran since February 2020, which some studies are performing on different aspects of COVID-19.

The MSA analysis of the CDS and partial ORF1ab polyprotein sequence from the Iranian patient revealed that

both CDS and partial ORF1ab polyprotein sequences of the Iranian sample were 100% identical to the query CDS sequence from Wuhan, China (Accession Number: NC_045512.2) and the query protein sequence from Wuhan, China (Accession Number: YP_009724389.1), respectively. These identities demonstrated that COVID-19 in Iran had the Wuhan origin of China, which was transmitted by human-to-human epidemiological pattern following a pandemic outbreak in other Asian Southeast countries such as Hong Kong, Japan and South Korea [68]. The protein modelling of the partial ORF1ab polyprotein sequence from the Iranian patient and detecting the NMR structure with PDB entry ID: 2GDT approved that the subject protein sequence from the Iranian patient is a part of nsp1 from SARS-CoV-2, which was 83% identical to the nsp1 from SARS-CoV. As it was identified, nsp1 is encoded by ORF1a and is highly conserved, crucial to the virus replication, survival in the society and spread among susceptible populations, and can be a potential virulence factor in COVID-19 through accelerating the cellular RNA degradation and consequently blocking the human immune response [69]. Since the BLASTP *E*-value scores for nsp1 from various isolates of SARS-CoV showed high percentage of identity, it was highly possible that the analyzed protein sequence was nsp1 and the Iranian patient had been affected by the virulence effect of identified nsp1 of SARS-CoV-2. The antigenicity prediction of the partial ORF1ab polyprotein or nsp1 sequence from the Iranian patient defined that firstly hydrophobic and secondly hydrophilic amino acids of the first ranked antigenic domain of partial nsp1 of the patient displayed higher

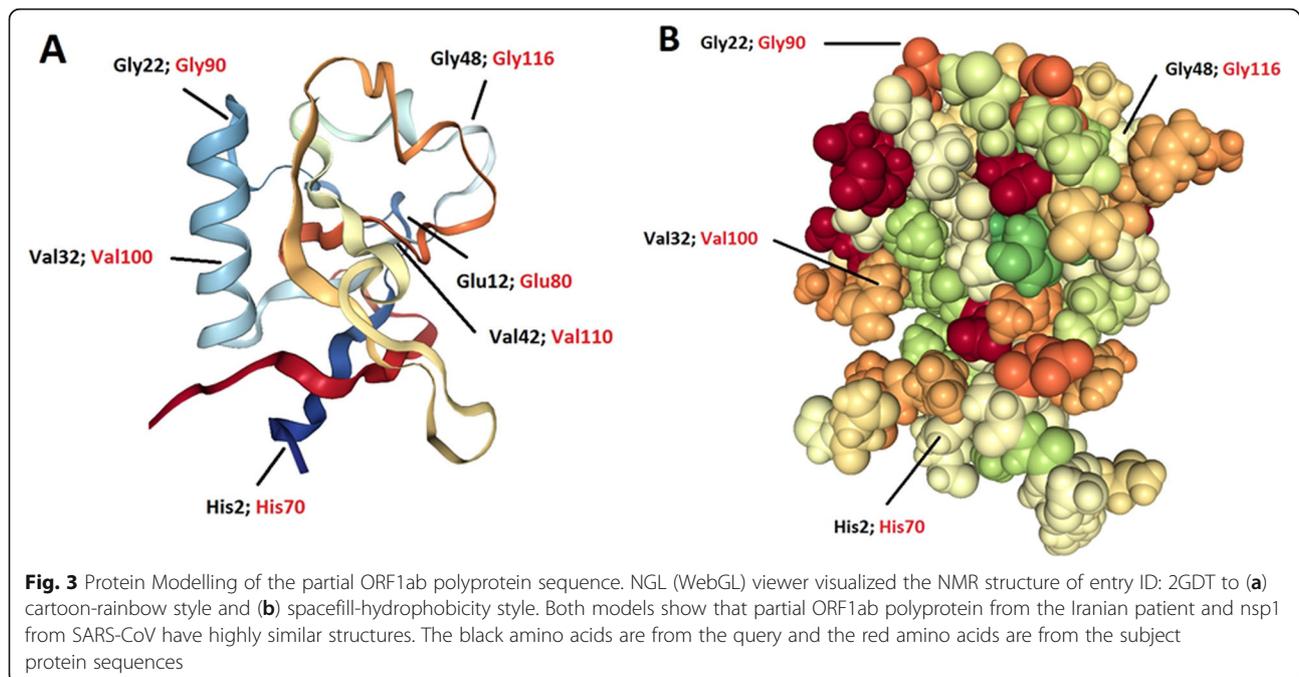


Table 1 Antigenicity prediction of the partial ORF1ab polyprotein sequence

Rank	Score	Length	Residues*
1	1.206	22	24 – TLGVLVPHVGEIPVAYRKVLLR – 45
2	1.169	13	4 – HVMVELVAELEGI – 16
3	1.047	7	58 – GADLKSF – 64

*: The grey highlights are hydrophobic, yellow highlights are hydrophilic, cyan highlights are positive charged, green highlights are negative charged, purple highlights are aromatic amino acids.

antigenic properties with accessibility to the human immune system such as Gly94, Pro98, and Gly101. The previous studies showed that diversity of pathogenic and venomous living organisms produce glycine-proline rich antigens in the secretions or venoms [70–73]. Furthermore, glycine, proline, as well as hydrophobic amino acids were exposed on the surface of partial nsP1 of SARS-CoV-2 to play as a part of a virulence factor (nsP1) and stimulate the human immune system [74, 75].

Conclusions

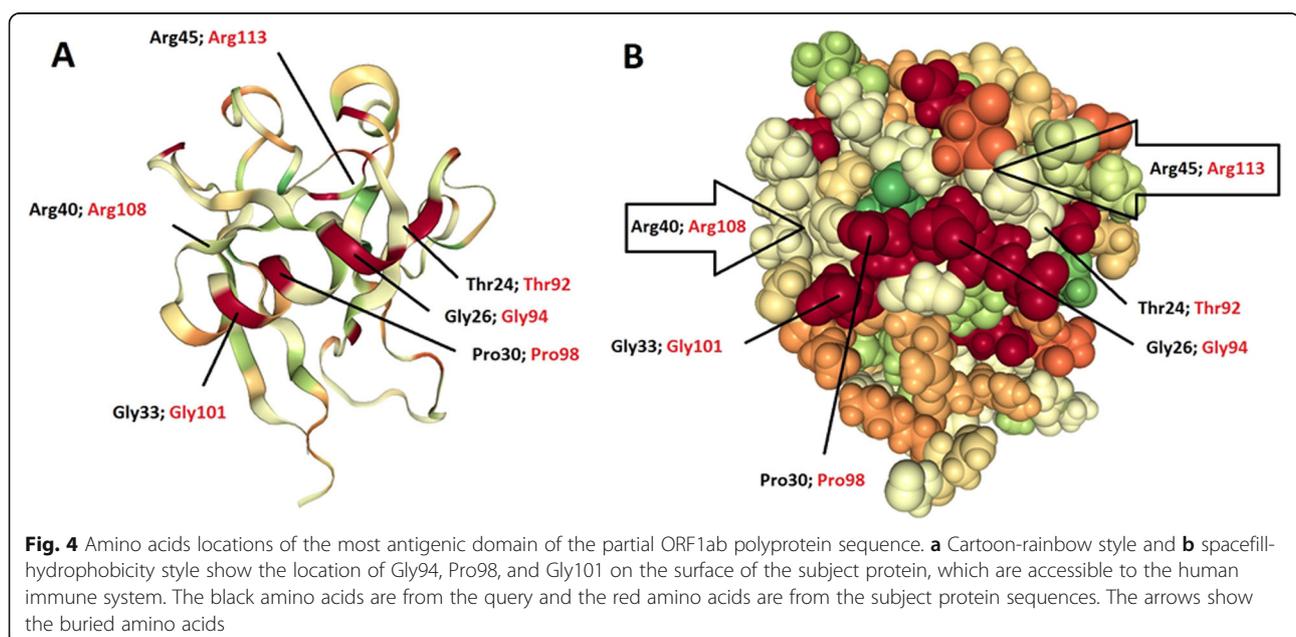
Although the identified protein sequence from an Iranian patient was a part of nsP1 from SARS-CoV-2 and could be a virulence and survival factor in the spreading of the COVID-19 among the population, there are some other potentials in nsP1 to make it attractive for future therapeutic and preventive strategies in pharmaceutical and vaccine manufacturers.

Based on the highly conserved sequence of nsP1 among the isolates of SARS-CoVs, it can be an appropriate candidate in the molecular epidemiology of COVID-19 in the pandemic outbreaks.

Methods

Sequence Analysis

To obtain the data for DNA sequences of SARS-CoV-2 from Iran, we used NCBI Virus database (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>) [76]. The accession number of MT152900 was selected that was related to the nucleotide sequence with 322 b in length isolated from the oronasopharynx of an Iranian patient annotated on the NCBI Virus database on 2020-02-26. The accession number MT152900 was annotated on the NCBI Virus database with the following details: Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/MHKN-1/human/2020/IRN ORF1ab polyprotein (orf1ab) gene, partial cds (Karbalaie Niya, M.H., et al.).



The accession number of the relevant protein sequence is QIH55230, which was nominated as the partial ORF1ab polyprotein with 107 amino acids in length.

The CDS sequence of the Iranian patient (Accession Number: MT152900) and the query CDS sequence from Wuhan, China (Accession Number: NC_045512.2) were compared using MSA and Clustal Omega algorithm (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) [77]. In addition, the partial ORF1ab polyprotein from the Iranian patient (Accession Number: QIH55230) and the query protein sequence from Wuhan, China (Accession Number: YP_009724389.1) were compared using MSA and Clustal Omega algorithm as well.

Protein Modelling

The partial ORF1ab polyprotein sequence from the Iranian patient (Accession Number: QIH55230) was searched in the RCSB PDB Protein Data Bank (<https://www.rcsb.org/>) [78]. A Basic Local Alignment Search Tool (BLAST) was employed by the Data Bank to identify the most identical crystalized protein structure to the subject protein sequence. Then, the NMR structure for the most identical crystalized protein structure would be visualized by NGL (WebGL) viewer [79].

Antigenicity Prediction

The antigenicity prediction of the partial ORF1ab polyprotein sequence from the Iranian patient (Accession Number: QIH55230) was performed using EMBOSS antigenic explorer (<https://www.bioinformatics.nl/cgi-bin/emboss/antigenic>) [80]. This web tool predicts the potentially regions of the subject protein sequence through application of Kolaskar and Tongaonkar method on the hydrophobic residues of a protein domain. In addition, the amino acids location of the most antigenic domain of the subject protein was evaluated using NGL (WebGL) viewer within RCSB PDB Protein Data Bank.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12575-020-00121-9>.

Additional file 1.

Abbreviations

ACE2: Angiotensin converting enzyme 2; BLAST: Basic Local Alignment Search Tool; CDS: Coding sequence; COVID-19: Coronavirus disease 2019; DPP4: Dipeptidyl peptidase 4; E: Envelope; ER: Endoplasmic reticulum; M: Membrane; MERS-CoV: Middle East Respiratory Syndrome Coronavirus; MSA: Multiple sequence alignment; N: Nucleocapsid; NCBI: National Center for Biotechnology Information; NMR: Nuclear magnetic resonance; nsp: Nonstructural protein; ORF: Open reading frame; PDB: Protein data bank; RBD: Receptor-binding domain; RdRP: RNA-dependent RNA polymerase; RNA: Ribonucleic acid; RTC: Replication-transcription complexes; S: Spike; SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

Acknowledgements

We thank the Deputy of Research and COVID-19 National Committee of the National Institute of Genetic Engineering and Biotechnology (NIGEB) from the Islamic Republic of Iran for preparing the condition to perform this study. No funding organizations had any role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

Authors' Contributions

All authors participated in the design of the study. RZE designed and carried out the search and biodata mining related to the SARS-CoVs, COVID-19, and gene/protein sequences from NCBI database. HN and RAT helped in the bio-data mining and preparing the protein modelling files. RZE performed the MSA, protein modelling, and antigenicity prediction analyses. RZE drafted the first version of the manuscript. All authors participated in writing further versions and read and approved the final manuscript.

Funding

To perform this study, RZE received a research grant support from the National Institute of Genetic Engineering and Biotechnology (NIGEB) of the Islamic Republic of Iran.

Availability of Data and Materials

All data analyzed in this study were prepared from databases including NCBI Virus, RCSB PDB Protein Data Bank, and EMBOSS antigenic explorer and were included in this article.

Ethics Approval and Consent to Participate

The data of the ORF1ab polyprotein of SARS-CoV-2 isolated from oronasopharynx of an Iranian patient was available online in the NCBI Virus database with the following accession numbers: MT152900 and QIH55230.

Consent for Publication

All authors have read and approved the final version of the manuscript.

Competing Interests

The authors declare that they have no conflict of interests.

Author details

¹Department of Energy and Environmental Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), 14965/161, Tehran, Iran.

²Department of Materials Engineering, Tarbiat Modares University, Tehran, Iran.

³Nanobiotechnology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran.

Received: 16 March 2020 Accepted: 8 April 2020

Published online: 21 April 2020

References

1. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol*. 2009;7(6):439–50.
2. Zheng J, Yamada Y, Fung TS, Huang M, Chia R, Liu DX. Identification of N-linked glycosylation sites in the spike protein and their functional impact on the replication and infectivity of coronavirus infectious bronchitis virus in cell culture. *Virology*. 2018;513:65–74.
3. Belouzard S, Millet JK, Licitra BN, Whittaker GR. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses*. 2012;4(6):1011–33.
4. Plant EP, Sims AC, Baric RS, Dinman JD, Taylor DR. Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production. *Viruses*. 2013;5(1):279–94.
5. Plant EP, Dinman JD. The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front Biosci*. 2008;13:4873–81.
6. Lindner HA, Fotouhi-Ardakani N, Lytyn V, Lachance P, Sulea T, Menard R. The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J Virol*. 2005;79(24):15199–208.
7. Oostra M, Hagemeijer MC, van Gent M, Bekker CP, te Lintelo EG, Rottier PJ, et al. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J Virol*. 2008;82(24):12392–405.
8. Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, et al. One severe acute respiratory syndrome coronavirus protein

- complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci U S A*. 2014;111(37):E3900–9.
9. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*. 2015;1282:1–23.
 10. Zhong NS, Zheng BJ, Li YM, Poon XZH, Chan KH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet*. 2003;362(9393):1353–8.
 11. Raj VS, Mou H, Smits SL, Dekkers DH, Muller MA, Dijkman R, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*. 2013;495(7440):251–4.
 12. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 2012;367(19):1814–20.
 13. Ahmadzadeh J, Mobaraki K. Epidemiological status of the Middle East respiratory syndrome coronavirus in 2019: an update from January 1 to march 31, 2019. *Int J Gen Med*. 2019;12:305–11.
 14. Prescott J, Falzarano D, de Wit E, Hardcastle K, Feldmann F, Haddock E, et al. Pathogenicity and viral shedding of MERS-CoV in immunocompromised rhesus macaques. *Front Immunol*. 2018;9:205.
 15. Yu P, Xu Y, Deng W, Bao L, Huang L, Xu Y, et al. Comparative pathology of rhesus macaque and common marmoset animal models with Middle East respiratory syndrome coronavirus. *PLoS One*. 2017;12(2):e0172093.
 16. Falzarano D, de Wit E, Feldmann F, Rasmussen AL, Okumura A, Peng X, et al. Infection with MERS-CoV causes lethal pneumonia in the common marmoset. *PLoS Pathog*. 2014;10(8):e1004250.
 17. Yeung ML, Yao Y, Jia L, Chan JF, Chan KH, Cheung KF, et al. MERS coronavirus induces apoptosis in kidney and lung by upregulating Smad7 and FGF2. *Nat Microbiol*. 2016;1:16004.
 18. Adney DR, van Doremalen N, Brown VR, Bushmaker T, Scott D, de Wit E, et al. Replication and shedding of MERS-CoV in upper respiratory tract of inoculated dromedary camels. *Emerg Infect Dis*. 2014;20(12):1999–2005.
 19. Zhao J, Li K, Wohlford-Lenane C, Agnihothram SS, Fett C, Zhao J, et al. Rapid generation of a mouse model for Middle East respiratory syndrome. *Proc Natl Acad Sci U S A*. 2014;111(13):4970–5.
 20. Agrawal AS, Garron T, Tao X, Peng BH, Wakamiya M, Chan TS, et al. Generation of a transgenic mouse model of Middle East respiratory syndrome coronavirus infection and disease. *J Virol*. 2015;89(7):3659–70.
 21. Pascal KE, Coleman CM, Mujica AO, Kamat V, Badithe A, Fairhurst J, et al. Pre- and postexposure efficacy of fully human antibodies against spike protein in a novel humanized mouse model of MERS-CoV infection. *Proc Natl Acad Sci U S A*. 2015;112(28):8738–43.
 22. Cockrell AS, Yount BL, Scobey T, Jensen K, Douglas M, Beall A, et al. A mouse model for MERS coronavirus-induced acute respiratory distress syndrome. *Nat Microbiol*. 2016;2:16226.
 23. Fan C, Wu X, Liu Q, Li Q, Liu S, Lu J, et al. A Human DPP4-Knockin Mouse's Susceptibility to Infection by Authentic and Pseudotyped MERS-CoV. *Viruses*. 2018;10(9):448.
 24. Yong CY, Ong HK, Yeap SK, Ho KL, Tan WS. Recent advances in the vaccine development against Middle East respiratory syndrome-coronavirus. *Front Microbiol*. 2019;10:1781.
 25. Hashem AM, Algaissi A, Agrawal AS, Al-Amri SS, Alhabbab RY, Sohrab SS, et al. A highly immunogenic, protective, and safe adenovirus-based vaccine expressing Middle East respiratory syndrome coronavirus S1-CD40L fusion protein in a transgenic human Dipeptidyl peptidase 4 mouse model. *J Infect Dis*. 2019;220(10):1558–67.
 26. Guo X, Deng Y, Chen H, Lan J, Wang W, Zou X, et al. Systemic and mucosal immunity in mice elicited by a single immunization with human adenovirus type 5 or 41 vector-based vaccines carrying the spike protein of Middle East respiratory syndrome coronavirus. *Immunology*. 2015;145(4):476–84.
 27. Kim E, Okada K, Kenniston T, Raj VS, AlHajri MM, Farag EA, et al. Immunogenicity of an adenoviral-based Middle East respiratory syndrome coronavirus vaccine in BALB/c mice. *Vaccine*. 2014;32(45):5975–82.
 28. Munster VJ, Wells D, Lambe T, Wright D, Fischer RJ, Bushmaker T, et al. Protective efficacy of a novel simian adenovirus vaccine against lethal MERS-CoV challenge in a transgenic human DPP4 mouse model. *NPJ Vaccines*. 2017;2:28.
 29. Alharbi NK, Padron-Regalado E, Thompson CP, Kupke A, Wells D, Sloan MA, et al. ChAdOx1 and MVA based vaccine candidates against MERS-CoV elicit neutralising antibodies and cellular immune responses in mice. *Vaccine*. 2017;35(30):3780–8.
 30. Volz A, Kupke A, Song F, Jany S, Fux R, Shams-Eldin H, et al. Protective efficacy of recombinant modified Vaccinia virus Ankara delivering Middle East respiratory syndrome coronavirus spike glycoprotein. *J Virol*. 2015;89(16):8651–6.
 31. Song F, Fux R, Provacia LB, Volz A, Eickmann M, Becker S, et al. Middle East respiratory syndrome coronavirus spike protein delivered by modified vaccinia virus Ankara efficiently induces virus-neutralizing antibodies. *J Virol*. 2013;87(21):11950–4.
 32. Veit S, Jany S, Fux R, Sutter G, Volz A. CD8+ T Cells Responding to the Middle East Respiratory Syndrome Coronavirus Nucleocapsid Protein Delivered by Vaccinia Virus MVA in Mice. *Viruses*. 2018;10(12):718.
 33. Jung SY, Kang KW, Lee EY, Seo DW, Kim HL, Kim H, et al. Heterologous prime-boost vaccination with adenoviral vector and protein nanoparticles induces both Th1 and Th2 responses against Middle East respiratory syndrome coronavirus. *Vaccine*. 2018;36(24):3468–76.
 34. Muthumani K, Falzarano D, Reuschel EL, Tingey C, Flingai S, Villarreal DO, et al. A synthetic consensus anti-spike protein DNA vaccine induces protective immunity against Middle East respiratory syndrome coronavirus in nonhuman primates. *Sci Transl Med*. 2015;7(301):301ra132.
 35. Al-Amri SS, Abbas AT, Siddiq LA, Alghamdi A, Sanki MA, Al-Muhanna MK, et al. Immunogenicity of candidate MERS-CoV DNA vaccines based on the spike protein. *Sci Rep*. 2017;7:44875.
 36. Chi H, Zheng X, Wang X, Wang C, Wang H, Gai W, et al. DNA vaccine encoding Middle East respiratory syndrome coronavirus S1 protein induces protective immune responses in mice. *Vaccine*. 2017;35(16):2069–75.
 37. Wang Y, Tai W, Yang J, Zhao G, Sun S, Tseng CK, et al. Receptor-binding domain of MERS-CoV with optimal immunogen dosage and immunization interval protects human transgenic mice from MERS-CoV infection. *Hum Vaccin Immunother*. 2017;13(7):1615–24.
 38. Adney DR, Wang L, van Doremalen N, Shi W, Zhang Y, Kong WP, et al. Efficacy of an Adjuvanted Middle East Respiratory Syndrome Coronavirus Spike Protein Vaccine in Dromedary Camels and Alpacas. *Viruses*. 2019;11(3):212.
 39. Pallesen J, Wang N, Corbett KS, Wrapp D, Kirchdoerfer RN, Turner HL, et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc Natl Acad Sci U S A*. 2017;114(35):E7348–E57.
 40. Tai W, Zhao G, Sun S, Guo Y, Wang Y, Tao X, et al. A recombinant receptor-binding domain of MERS-CoV in trimeric form protects human dipeptidyl peptidase 4 (hDPP4) transgenic mice from MERS-CoV infection. *Virology*. 2016;499:375–82.
 41. Ma C, Wang L, Tao X, Zhang N, Yang Y, Tseng CK, et al. Searching for an ideal vaccine candidate among different MERS coronavirus receptor-binding fragments—the importance of immunofocusing in subunit vaccine design. *Vaccine*. 2014;32(46):6170–6.
 42. Du L, Kou Z, Ma C, Tao X, Wang L, Zhao G, et al. A truncated receptor-binding domain of MERS-CoV spike protein potently inhibits MERS-CoV infection and induces strong neutralizing antibody responses: implication for developing therapeutics and vaccines. *PLoS One*. 2013;8(12):e81587.
 43. Ma C, Li Y, Wang L, Zhao G, Tao X, Tseng CT, et al. Intranasal vaccination with recombinant receptor-binding domain of MERS-CoV spike protein induces much stronger local mucosal immune responses than subcutaneous immunization: implication for designing novel mucosal MERS vaccines. *Vaccine*. 2014;32(18):2100–8.
 44. Nyon MP, Du L, Tseng CK, Seid CA, Pollet J, Naceanceno KS, et al. Engineering a stable CHO cell line for the expression of a MERS-coronavirus vaccine antigen. *Vaccine*. 2018;36(14):1853–62.
 45. Zhang N, Channappanavar R, Ma C, Wang L, Tang J, Garron T, et al. Identification of an ideal adjuvant for receptor-binding domain-based subunit vaccines against Middle East respiratory syndrome coronavirus. *Cell Mol Immunol*. 2016;13(2):180–90.
 46. Lan J, Deng Y, Chen H, Lu G, Wang W, Guo X, et al. Tailoring subunit vaccine immunity with adjuvant combinations and delivery routes using the Middle East respiratory coronavirus (MERS-CoV) receptor-binding domain as an antigen. *PLoS One*. 2014;9(11):e112602.
 47. Lan J, Yao Y, Deng Y, Chen H, Lu G, Wang W, et al. Recombinant receptor binding domain protein induces partial protective immunity in rhesus macaques against Middle East respiratory syndrome coronavirus challenge. *EBioMedicine*. 2015;2(10):1438–46.
 48. Jiaming L, Yanfeng Y, Yao D, Yawei H, Linlin B, Baoying H, et al. The recombinant N-terminal domain of spike proteins is a potential vaccine against Middle East respiratory syndrome coronavirus (MERS-CoV) infection. *Vaccine*. 2017;35(1):10–8.

49. Wang C, Zheng X, Gai W, Zhao Y, Wang H, Wang H, et al. MERS-CoV virus-like particles produced in insect cells induce specific humoral and cellular immunity in rhesus macaques. *Oncotarget*. 2017;8(8):12686–94.
50. Coleman CM, Liu YV, Mu H, Taylor JK, Massare M, Flyer DC, et al. Purified coronavirus spike protein nanoparticles induce coronavirus neutralizing antibodies in mice. *Vaccine*. 2014;32(26):3169–74.
51. Coleman CM, Venkataraman T, Liu YV, Glenn GM, Smith GE, Flyer DC, et al. MERS-CoV spike nanoparticles protect mice from MERS-CoV infection. *Vaccine*. 2017;35(12):1586–9.
52. Wang C, Zheng X, Gai W, Wong G, Wang H, Jin H, et al. Novel chimeric virus-like particles vaccine displaying MERS-CoV receptor-binding domain induce specific humoral and cellular immune response in mice. *Antivir Res*. 2017;140:55–61.
53. Lan J, Deng Y, Song J, Huang B, Wang W, Tan W. Significant spike-specific IgG and neutralizing antibodies in mice induced by a novel chimeric virus-like particle vaccine candidate for Middle East respiratory syndrome coronavirus. *Virology*. 2018;33(5):453–5.
54. Kim YS, Son A, Kim J, Kwon SB, Kim MH, Kim P, et al. Chaperone-mediated assembly of ferritin-based Middle East respiratory syndrome-coronavirus nanoparticles. *Front Immunol*. 2018;9:1093.
55. Deng Y, Lan J, Bao L, Huang B, Ye F, Chen Y, et al. Enhanced protection in mice induced by immunization with inactivated whole viruses compare to spike protein of middle east respiratory syndrome coronavirus. *Emerg Microbes Infect*. 2018;7(1):60.
56. Agrawal AS, Tao X, Algaissi A, Garron T, Narayanan K, Peng BH, et al. Immunization with inactivated Middle East respiratory syndrome coronavirus vaccine leads to lung immunopathology on challenge with live virus. *Hum Vaccin Immunother*. 2016;12(9):2351–6.
57. Wirblich C, Coleman CM, Kurup D, Abraham TS, Bernbaum JG, Jahrling PB, et al. One-Health: a Safe, Efficient, Dual-Use Vaccine for Humans and Animals against Middle East Respiratory Syndrome Coronavirus and Rabies Virus. *J Virol*. 2017;91(2):e02040–16.
58. Menachery VD, Gralinski LE, Mitchell HD, Dinnon KH 3rd, Leist SR, Yount BL Jr, et al. Middle East Respiratory Syndrome Coronavirus Nonstructural Protein 16 Is Necessary for Interferon Resistance and Viral Pathogenesis. *mSphere*. 2017;2(6):00346–17.
59. Almazan F, DeDiego ML, Sola I, Zuniga S, Nieto-Torres JL, Marquez-Jurado S, et al. Engineering a replication-competent, propagation-defective Middle East respiratory syndrome coronavirus as a vaccine candidate. *mBio*. 2013;4(5):e00650–13.
60. Malczyk AH, Kupke A, Prufer S, Scheuplein VA, Hutzler S, Kreuz D, et al. A highly immunogenic and protective Middle East respiratory syndrome coronavirus vaccine based on a recombinant measles virus vaccine platform. *J Virol*. 2015;89(22):11654–67.
61. Bodmer BS, Fiedler AH, Hanauer JRH, Prufer S, Muhlebach MD. Live-attenuated bivalent measles virus-derived vaccines targeting Middle East respiratory syndrome coronavirus induce robust and multifunctional T cell responses against both viruses in an appropriate mouse model. *Virology*. 2018;521:99–107.
62. Liu R, Wang J, Shao Y, Wang X, Zhang H, Shuai L, et al. A recombinant VSV-vectored MERS-CoV vaccine induces neutralizing antibody and T cell responses in rhesus monkeys after single dose immunization. *Antivir Res*. 2018;150:30–8.
63. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 2020. <https://doi.org/10.1016/j.cell.2020.02.052>.
64. Sun Z, Thilakavathy K, Kumar SS, He G, Liu SV. Potential Factors Influencing Repeated SARS Outbreaks in China. *Int J Environ Res Public Health*. 2020;17(5):1633.
65. Wang L, Fu S, Cao Y, Zhang H, Feng Y, Yang W, et al. Discovery and genetic analysis of novel coronaviruses in least horseshoe bats in southwestern China. *Emerg Microbes Infect*. 2017;6(3):e14.
66. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol*. 2020;5(4):562–9.
67. Zhang JJ, Dong X, Cao YY, Yuan YD, Yang YB, Yan YQ, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy*. 2020. <https://doi.org/10.1111/all.14238>.
68. Thompson RN. Novel Coronavirus Outbreak in Wuhan, China, 2020: Intense Surveillance Is Vital for Preventing Sustained Transmission in New Locations. *J Clin Med*. 2020;9(2).
69. Connor RF, Roper RL. Unique SARS-CoV protein nsp1: bioinformatics, biochemistry and potential effects on virulence. *Trends Microbiol*. 2007;15(2):51–3.
70. Yarawsky AE, English LR, Whitten ST, Herr AB. The Proline/Glycine-rich region of the biofilm adhesion protein Aap forms an extended stalk that resists compaction. *J Mol Biol*. 2017;429(2):261–79.
71. Tchernychev B, Cabilly S, Wilchek M. The epitopes for natural polyreactive antibodies are rich in proline. *Proc Natl Acad Sci U S A*. 1997;94(12):6335–9.
72. Kim TY, Kang SY, Ahn IY, Cho SY, Hong SJ. Molecular cloning and characterization of an antigenic protein with a repeating region from *Clonorchis sinensis*. *Korean J Parasitol*. 2001;39(1):57–66.
73. Dalla Valle L, Nardi A, Alibardi L. Isolation of a new class of cysteine-glycine-proline-rich beta-proteins (beta-keratins) and their expression in snake epidermis. *J Anat*. 2010;216(3):356–67.
74. Zolfaghari Emameh R, Barker H, Hytonen VP, Tolvanen ME, Parkkila S. Beta carbonic anhydrases: novel targets for pesticides and anti-parasitic agents in agriculture and livestock husbandry. *Parasit Vectors*. 2014;7:403.
75. Zolfaghari Emameh R, Barker HR, Tolvanen ME, Parkkila S, Hytonen VP. Horizontal transfer of beta-carbonic anhydrase genes from prokaryotes to protozoans, insects, and nematodes. *Parasit Vectors*. 2016;9:152.
76. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res*. 2015;43(Database issue):D571–7.
77. Sievers F, Higgins DG. Clustal omega for making accurate alignments of many protein sequences. *Protein Sci*. 2018;27(1):135–45.
78. Goodsell DS, Dutta S, Zardecki C, Voigt M, Berman HM, Burley SK. The RCSB PDB "molecule of the month": inspiring a molecular view of biology. *PLoS Biol*. 2015;13(5):e1002140.
79. Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlic A, Rose PW. NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics*. 2018;34(21):3755–8.
80. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*. 1990;276(1–2):172–4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

