Cold Spring Harbor Laboratory
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

HOME | ABOUT | SUBMIT | ALERTS / RSS | CHANNELS

Search
Advanced Search

bioRxiv is receiving many new papers on coronavirus 2019-nCoV. A reminder: these are preliminary reports that have not been peer-reviewed. They should not be regarded as conclusive, guide clinical practice/health-related behavior, or be reported in news media as established information.

New Results

This article has been withdrawn. Click here for details

Previous Next

Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag

119 comments

Prashant Pradhan, Ashutosh Kumar Pandey, Akhilesh Mishra, Parul Gupta, Praveen Kumar Tripathi, Manoj Balakrishnan Menon, James Gomes, Perumal Vivekanandan, Bishwajit Kundu

doi: https://doi.org/10.1101/2020.01.30.927871

Posted January 31, 2020.

This article is a preprint and has not been certified by peer review [what does this mean?].

Download PDF
Supplementary Material
Email
Share
Citation Tools

Abstract | Full Text | Info/History | Metrics | Preview PDF

Abstract

We are currently witnessing a major epidemic caused by the 2019 novel coronavirus (2019-nCoV). The evolution of 2019-nCoV remains elusive. We found 4 insertions in the spike glycoprotein (S) which are unique to the 2019-nCoV and are not present in other coronaviruses. Importantly, amino acid residues in all the 4 inserts have identity or similarity to those in the HIV-1 gp120 or HIV-1 Gag. Interestingly, despite the inserts being discontinuous on the primary amino acid sequence, 3D-modelling of the 2019-nCoV suggests that they converge to constitute the receptor binding site. The finding of 4 unique inserts in the 2019-nCoV, all of which have identity /similarity to amino acid residues in key structural proteins of HIV-1 is unlikely to be fortuitous in nature. This work provides yet unknown insights on 2019-nCoV and sheds light on the evolution and pathogenicity of this virus with important implications for diagnosis of this virus.

Subject Area
Evolutionary Biology

Subject Areas
All Articles
Animal Behavior and Cognition
Biochemistry
Bioengineering
Bioinformatics
Biophysics
Cancer Biology
Cell Biology
Clinical Trials*
Developmental Biology
Epidemiology
Evolutionary Biology
Genetics
Genomics
Immunology
Microbiology
Molecular Biology
Neuroscience
Paleontology
Pathology
Pharmacology and Toxicology
Physiology
Plant Biology
Scientific Communication and Education
Synthetic Biology
Systems Biology
Zoology

* The Clinical Trials and Epidemiology subject categories are now closed to new submissions following the completion of bioRxiv's clinical research pilot project and launch of the dedicated health sciences server medRxiv (submit.medrxiv.org). New papers that report results of Clinical Trials must now be submitted to medRxiv. Most new Epidemiology papers also should be submitted to medRxiv, but if a paper contains no health-related information, authors may choose to submit it to another bioRxiv subject category (e.g., Genetics or Microbiology).

We use cookies on this site to enhance your user experience. By clicking any link on this page you are giving your consent for us to set cookies. Continue Find out more

Coronaviruses (CoV) are single-stranded positive-sense RNA viruses that infect animals and humans. These are classified into 4 genera based on their host specificity: *Alphacoronavirus, Betacoronavirus, Deltacoronavirus and Gammacoronavirus* (Snijder et al., 2006). There are seven known types of CoVs that includes 229E and NL63 (Genus Alphacoronavirus), OC43, HKU1, MERS and SARS (Genus Betacoronavirus). While 229E, NL63, OC43, and HKU1 commonly infect humans, the SARS and MERS outbreak in 2002 and 2012 respectively occurred when the virus crossed-over from animals to humans causing significant mortality (J. Chan et al., n.d.; J. F. W. Chan et al., 2015). In December 2019, another outbreak of coronavirus was reported from Wuhan, China that also transmitted from animals to humans. This new virus has been temporarily termed as 2019-novel Coronavirus (2019-nCoV) by the World Health Organization (WHO) (J. F.-W. Chan et al., 2020; Zhu et al., 2020). While there are several hypotheses about the origin of 2019-nCoV, the source of this ongoing outbreak remains elusive.

The transmission patterns of 2019-nCoV is similar to patterns of transmission documented in the previous outbreaks including by bodily or aerosol contact with persons infected with the virus. Cases of mild to severe illness, and death from the infection have been reported from Wuhan. This outbreak has spread rapidly distant nations including France, Australia and USA among others. The number of cases within and outside China are increasing steeply. Our current understanding is limited to the virus genome sequences and modest epidemiological and clinical data. Comprehensive analysis of the available 2019-nCoV sequences may provide important clues that may help advance our current understanding to manage the ongoing outbreak.

The spike glycoprotein (S) of cornonavirus is cleaved into two subunits (S1 and S2). The S1 subunit helps in receptor binding and the S2 subunit facilitates membrane fusion (Bosch et al., 2003; Li, 2016). The spike glycoproteins of coronoviruses are important determinants of tissue tropism and host range. In addition the spike glycoproteins are critical targets for vaccine development (Du et al., 2013). For this reason, the spike proteins represent the most extensively studied among coronaviruses. We therefore sought to investigate the spike glycoprotein of the 2019-nCoV to understand its evolution, novel features sequence and structural features using computational tools.

Methodology

Retrieval and alignment of nucleic acid and protein sequences

We retrieved all the available coronavirus sequences (n=55) from NCBI viral genome database (https://www.ncbi.nlm.nih.gov/) and we used the GISAID (Elbe & Buckland-Merrett, 2017) [https://www.gisaid.org/] to retrieve all available full-length sequences (n=28) of 2019-nCoV as on 27 Jan 2020. Multiple sequence alignment of all coronavirus genomes was performed by using MUSCLE software (Edgar, 2004) based on neighbour joining method. Out of 55 coronavirus genome 32 representative genomes of all category were used for phylogenetic tree development using MEGAX software (Kumar et al., 2018). The closest relative was found to be SARS CoV. The glycoprotein region of SARS CoV and 2019-nCoV were aligned and visualized using Multalin software (Corpet, 1988). The identified amino acid and nucleotide sequence were aligned with whole viral genome database using BLASTp and BLASTn. The conservation of the nucleotide and amino acid motifs in 28 clinical variants of 2019-nCoV genome were presented by performing multiple sequence alignment using MEGAX software. The three dimensional structure of 2019-nCoV glycoprotein was generated by using SWISS-MODEL online server (Biasini et al., 2014) and the structure was marked and visualized by using PyMol (DeLano, 2002).

Results

Uncanny similarity of novel inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag

Our phylogentic tree of full-length coronaviruses suggests that 2019-nCoV is closely related to SARS CoV [Fig1]. In addition, other recent studies have linked the 2019-nCoV to SARS CoV. We therefore compared the spike glycoprotein sequences of the 2019-nCoV to that of the SARS CoV (NCBI Accession number: AY390556.1). On careful examination of the sequence alignment we found that the 2019-nCoV spike glycoprotein contains 4 insertions [Fig.2]. To further investigate if these inserts are present in any other corona virus, we performed a multiple sequence alignment of the spike glycoprotein amino acid sequences of all available

Proudly supported by Chan Zuckerberg Initiative coronaviruses (n=55) [refer Table S.File1] in NCBI refseq (ncbi.nlm.nih.gov) this includes one sequence of 2019-nCoV[Fig.S1]. We found that these 4 insertions [inserts 1, 2, 3 and 4] are unique to 2019-nCoV and are not present in other coronaviruses analyzed. Another group from China had documented three insertions comparing fewer spike glycoprotein sequences of coronaviruses. Another group from China had documented three insertions comparing fewer spike glycoprotein sequences of coronaviruses (Zhou et al., 2020).

Fig.S1 Download figure | Open in new tab

Multiple sequence alignment of glycoprotein of *coronaviridae* family, representing all the four inserts.

Figure 1: Download figure | Open in new tab

Maximum likelihood genealogy show the evolution of 2019-nCoV:
The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. The tree with the highest log likelihood (12458.88) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. This analysis involved 5 amino acid sequences. There were a total of 1387 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

Figure 2: Download figure | Open in new tab

Multiple sequence alignment between spike proteins of 2019-nCoV and SARS.
The sequences of spike proteins of 2019-nCoV (Wuhan-HU-1, Accession NC_045512) and of SARS CoV (GZ02, Accession AY390556) were aligned using MultiAlin software. The sites of difference are highlighted in boxes.

Figure 3. Download figure | Open in new tab

Modelled homo-trimer spike glycoprotein of 2019-nCoV virus.
The inserts from HIV envelop protein are shown with colored beads, present at the binding site of the protein.

We then analyzed all available full-length sequences (n=28) of 2019-nCoV in GISAID (Elbe & Buckland-Merrett, 2017) as on January 27, 2020 for the presence of these inserts. As most of these sequences are not annotated, we compared the nucleotide sequences of the spike glycoprotein of all available 2019-nCoV sequences using BLASTp. Interestingly, all the 4 insertions were absolutely (100%) conserved in all the available 2019-nCoV sequences analyzed [Fig.S2, Fig.S3].

Fig.S2: Download figure | Open in new tab

All four inserts are present in the aligned 28 Wuhan 2019-nCoV virus genomes obtained from GISAID. The gap in the Bat-SARS Like CoV in the last row shows that insert 1 and 4 is very unique to Wuhan 2019-nCoV.

Fig.S3 Download figure | Open in new tab

Phylogenetic tree of 28 clinical isolates genome of 2019-nCoV including one from bat as a host.

We then translated the aligned genome and found that these inserts are present in all Wuhan 2019-nCoV viruses except the 2019-nCoV virus of Bat as a host [Fig.S4]. Intrigued by the 4 highly conserved inserts unique to 2019-nCoV we wanted to understand their origin. For this purpose, we used the 2019-nCoV local alignment with each insert as query against all virus genomes and considered hits with 100% sequence coverage. Surprisingly, each of the four inserts aligned with short segments of the Human immunodeficiency Virus-1 (HIV-1) proteins. The amino acid positions of the inserts in 2019-nCoV and the corresponding residues in HIV-1 gp120 and HIV-1 Gag are shown in Table 1. The first 3 inserts (insert 1,2 and 3) aligned to short segments of amino acid residues in HIV-1 gp120. The insert 4 aligned to HIV-1 Gag. The insert 1 (6 amino acid residues) and insert 2 (6 amino acid residues) in the spike glycoprotein of 2019-nCoV are 100% identical to the residues mapped to HIV-1 gp120. The insert 3 (12 amino acid residues) in 2019-nCoV maps to HIV-1 gp120 with gaps [see Table 1]. The insert 4 (8 amino acid residues) maps to HIV-1 Gag with gaps.

Supplementary Fig 4. Download figure | Open in new tab

Genome alingment of Coronaviridae family. Highlighted black sequences are the inserts represented here.

Table 1: View inline | View popup

Aligned sequences of 2019-nCoV and gp120 protein of HIV-1 with their positions in primary sequence of protein. All the inserts have a high density of positively charged residues. The deleted fragments in insert 3 and 4 increase the positive charge to surface area ratio. *please see Supp. Table 1 for accession numbers

Although, the 4 inserts represent discontiguous short stretches of amino acids in spike glycoprotein of 2019-nCoV, the fact that all three of them share amino acid identity or similarity with HIV-1 gp120 and HIV-1 Gag (among all annotated virus proteins) suggests that this is not a random fortuitous finding. In other words, one may sporadically expect a fortuitous match for a stretch of 6-12 contiguous amino acid residues in an unrelated protein. However, it is unlikely that all 4 inserts in the 2019-nCoV spike glycoprotein fortuitously match with 2 key structural proteins of an unrelated virus (HIV-1).

The amino acid residues of inserts 1, 2 and 3 of 2019-nCoV spike glycoprotein that mapped to HIV-1 were a part of the V4, V5 and V1 domains respectively in gp120 [Table 1]. Since the 2019-nCoV inserts mapped to variable regions of HIV-1, they were not ubiquitous in HIV-1 gp120, but were limited to selected sequences of HIV-1 [refer S.File1] primarily from Asia and Africa.

The HIV-1 Gag protein enables interaction of virus with negatively charged host surface (Murakami, 2008) and a high positive charge on the Gag protein is a key feature for the host-virus interaction. On analyzing the pI values for each of the 4 inserts in 2019-nCoV and the corresponding stretches of amino acid residues from HIV-1 proteins we found that a) the pI values were very similar for each pair analyzed b) most of these pI values were 10±2 [Refer Table 1]. Of note, despite the gaps in inserts 3 and 4 the pI values were comparable. This uniformity in the pI values for all the 4 inserts merits further investigation.

As none of these 4 inserts are present in any other coronavirus, the genomic region encoding these inserts represent ideal candidates for designing primers that can distinguish 2019-nCoV from other coronaviruses.

The novel inserts are part of the receptor binding site of 2019-nCoV

To get structural insights and to understand the role of these insertions in 2019-nCoV glycoprotein, we modelled its structure based on available structure of SARS spike glycoprotein (PDB: 6ACD.1.A). The comparison of the modelled structure reveals that although inserts 1,2 and 3 are at non-contiguous locations in the protein primary sequence, they fold to constitute the part of glycoprotein binding site that recognizes the host receptor (**Kirchdoerfer et al., 2016**) (Figure 4). The insert 1 corresponds to the NTD (N-terminal domain) and the inserts 2 and 3 correspond to the CTD (C-terminal domain) of the S1 subunit in the 2019-nCoV spike glycoprotein. The insert 4 is at the junction of the SD1 (sub domain 1) and SD2 (sub domain 2) of the S1 subunit (**Ou et al., 2017**). We speculate, that these insertions provide additional flexibility to the glycoprotein binding site by forming a hydrophilic loop in the protein structure that may facilitate or enhance virus-host interactions.

Evolutionary Analysis of 2019-nCoV

It has been speculated that 2019-nCoV is a variant of Coronavirus derived from an animal source which got transmitted to humans. Considering the change of specificity for host, we decided to study the sequences of spike glycoprotein (S protein) of the virus. S proteins are surface proteins that help the virus in host recognition and attachment. Thus, a change in these proteins can be reflected as a change of host specificity of the virus. To know the alterations in S protein gene of 2019-nCoV and its consequences in structural re-arrangements we performed *in-sillico* analysis of 2019-nCoV with respect to all other viruses. A multiple sequence alignment between the S protein amino acid sequences of 2019-nCoV, Bat-SARS-Like, SARS-GZ02 and MERS revealed that S protein has evolved with closest significant diversity from the SARS-GZ02 (**Figure 1**).

Insertions in Spike protein region of 2019-nCoV

Since the S protein of 2019-nCoV shares closest ancestry with SARS GZ02, the sequence coding for spike proteins of these two viruses were compared using MultiAlin software. We found four new insertions in the protein of 2019-nCoV- “GTNGTKR” (IS1), “HKNNKS” (IS2), “GDSSSG” (IS3) and “QTNSPRRA” (IS4) (**Figure 2**). To our surprise, these sequence insertions were not only absent in S protein of SARS but were also not observed in any other member of the *Coronaviridae* family (Supplementary figure). This is startling as it is quite unlikely for a virus to have acquired such unique insertions naturally in a short duration of time.

Insertions share similarity to HIV

The insertions were observed to be present in all the genomic sequences of 2019-nCoV virus available from the recent clinical isolates (**Supplementary Figure 1**). To know the source of these insertions in 2019-nCoV a local alignment was done with BLASTp using these insertions as query with all virus genome. Unexpectedly, all the insertions got aligned with Human immunodeficiency Virus-1 (HIV-1). Further analysis revealed that aligned sequences of HIV-1 with 2019-nCoV were derived from surface glycoprotein gp120 (amino acid sequence positions: 404-409, 462-467, 136-150) and from Gag protein (366-384 amino acid) (**Table 1**). Gag protein of HIV is involved in host membrane binding, packaging of the virus and for the formation of virus-like particles. Gp120 plays crucial role in recognizing the host cell by binding to the primary receptor CD4.This binding induces structural rearrangements in GP120, creating a high affinity binding site for a chemokine co-receptor like CXCR4 and/or CCR5.

Discussion

The current outbreak of 2019-nCoV warrants a thorough investigation and understanding of its ability to infect human beings. Keeping in mind that there has been a clear change in the preference of host from previous coronaviruses to this virus, we studied the change in spike protein between 2019-nCoV and other viruses. We found four new insertions in the S protein of 2019-nCoV when compared to its nearest relative, SARS CoV. The genome sequence from the recent 28 clinical isolates showed that the sequence coding for these insertions are conserved amongst all these isolates. This indicates that these insertions have been preferably acquired by the 2019-nCoV, providing it with additional survival and infectivity advantage. Delving deeper we found that these insertions were similar to HIV-1. Our results highlight an astonishing relation between the gp120 and Gag protein of HIV, with 2019-nCoV spike glycoprotein. These proteins are critical for the viruses to identify and latch on to their host cells and for viral assembly (**Beniac et al., 2006**). Since surface proteins are responsible for host tropism, changes in these proteins imply a change in host specificity of the virus. According to reports from China, there has been a gain of host specificity in case 2019-nCoV as the virus was originally known to infect animals and not humans but after the mutations, it has gained tropism to humans as well.

Moving ahead, 3D modelling of the protein structure displayed that these insertions are present at the binding site of 2019-nCoV. Due to the presence of gp120 motifs in 2019-nCoV spike glycoprotein at its binding domain, we propose that these motif insertions could have provided an enhanced affinity towards host cell receptors. Further, this structural change might have also increased the range of host cells that 2019-nCoV can infect. To the best of our knowledge, the function of these motifs is still not clear in HIV and need to be explored. The exchange of genetic material among the viruses is well known and such critical exchange highlights the risk and the need to investigate the relations between seemingly unrelated virus families.

Conclusions

Our analysis of the spike glycoprotein of 2019-nCoV revealed several interesting findings: First, we identified 4 unique inserts in the 2019-nCoV spike glycoprotein that are not present in any other coronavirus reported till date. To our surprise, all the 4 inserts in the 2019-nCoV mapped to short segments of amino acids in the HIV-1 gp120 and Gag among all annotated virus proteins in the NCBI database. This uncanny similarity of novel inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag is unlikely to be fortuitous. Further, 3D modelling suggests that atleast 3 of the unique inserts which are non-contiguous in the primary protein sequence of the 2019-nCoV spike glycoprotein converge to constitute the key components of the receptor binding site. Of note, all the 4 inserts have pI values of around 10 that may facilitate virus-host interactions. Taken together, our findings suggest unconventional evolution of 2019-nCoV that warrants further investigation. Our work highlights novel evolutionary aspects of the 2019-nCoV and has implications on the pathogenesis and diagnosis of this virus.

Footnotes

- ⸸$ Equal contribution

References

Beniac, D. R., Andonov, A., Grudeski, E., & Booth, T. F. (2006). Architecture of the SARS coronavirus prefusion spike. *Nature Structural and Molecular Biology*, **13**(8), 751–752. https://doi.org/10.1038/nsmb1123 Google Scholar

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., & Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gku340 Google Scholar

Bosch, B. J., van der Zee, R., de Haan, C. A. M., & Rottier, P. J. M. (2003). The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. *Journal of Virology*, **77**(16), 8801–8811. https://doi.org/10.1128/jvi.77.16.8801-8811.2003 Abstract/FREE Full Text Google Scholar

Chan, J. F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K. K.-W., Yuan, S., & Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections*, **9**(1), 221–236. https://doi.org/10.1080/22221751.2020.1719902 Google Scholar

Chan, J. F. W., Lau, S. K. P., To, K. K. W., Cheng, V. C. C., Woo, P. C. Y., & Yuen, K.-Y. (2015). *Middle East Respiratory Syndrome Coronavirus: Another Zoonotic Betacoronavirus Causing SARS-Like Disease*. https://doi.org/10.1128/CMR.00102-14 Google Scholar

Chan, J., To, K., Tse, H., Jin, D., microbiology, K. Y.-T. in, & 2013, undefined. (n.d.). Interspecies transmission and emergence of novel viruses: lessons from bats and birds. Elsevier. Google Scholar

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*. https://doi.org/10.1093/nar/16.22.10881 Google Scholar

DeLano, W. L. (2002). The PyMOL Molecular Graphics System, Version 1.1. Schr{ö}dinger LLC. https://doi.org/10.1038/hr.2014.17 Google Scholar

Du, L., Zhao, G., Kou, Z., Ma, C., Sun, S., Poon, V. K. M., Lu, L., Wang, L., Debnath, A. K., Zheng, B.-J., Zhou, Y., & Jiang, S. (2013). Identification of a Receptor-Binding Domain in the S Protein of the Novel Human Coronavirus Middle East Respiratory Syndrome Coronavirus as an Essential Target for Vaccine Development. *Journal of Virology*, **87**(17), 9939–9942. https://doi.org/10.1128/jvi.01048-13 Abstract/FREE Full Text Google Scholar

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkh340 Google Scholar

Elbe, S., & Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*. https://doi.org/10.1002/gch2.1018 Google Scholar

Kirchdoerfer, R. N., Cottrell, C. A., Wang, N., Pallesen, J., Yassine, H. M., Turner, H. L., Corbett, K. S., Graham, B. S., McLellan, J. S., & Ward, A. B. (2016). Pre-fusion structure of a human coronavirus spike protein. *Nature*. https://doi.org/10.1038/nature17200 Google Scholar

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msy096 Google Scholar

Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology*, **3**(1), 237–261. https://doi.org/10.1146/annurev-virology-110615-042301 Google Scholar

Murakami, T. (2008). Roles of the interactions between Env and Gag proteins in the HIV-1 replication cycle. *Microbiology and Immunology*, **52**(5), 287–295. https://doi.org/10.1111/j.1348-0421.2008.00008.x CrossRef PubMed Google Scholar

Ou, X., Guan, H., Qin, B., Mu, Z., Wojdyla, J. A., Wang, M., Dominguez, S. R., Qian, Z., & Cui, S. (2017). Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nature Communications*. https://doi.org/10.1038/ncomms15216 Google Scholar

Snijder, E. J., van der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J. J. M., van der Meulen, J., Koerten, H. K., & Mommaas, A. M. (2006). Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. *Journal of Virology*, **80**(12), 5927–5940. https://doi.org/10.1128/JVI.02501-05 Abstract/FREE Full Text Google Scholar

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., … Shi, Z.-L. (2020). Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *BioRxiv*. https://doi.org/10.1101/2020.01.22.914952 Google Scholar

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*, NEJMoa2001017. https://doi.org/10.1056/NEJMoa2001017 Google Scholar

View Abstract

Blog posts linking to this article:

Booster Shots - latimes.com, 03 Mar 2020

«Life.ru» — информационный портал, 06 Feb 2020

nic prostszego, 04 Feb 2020

カレントアウェアネス・ポータル - カレントアウェアネス-R, 04 Feb 2020

Medical Microbiology, 04 Feb 2020

Invest Records, 03 Feb 2020

Simons FoundationSimons Foundation, 03 Feb 2020

Climate Science, 02 Feb 2020

The Reference Frame, 02 Feb 2020

Tweets referencing this article:

한옥대문
@9j4zWzlueneZgXi

RT @Harok_QAC: 인도 과학자는 우한 바이러스에서 그동안 코로나 바이러스에서 없었던 에이즈 바이러스 삽입 발견. 이것은 우연이 아니다. 중국 우한 바이러스 설계 가능성 암시. 만약 사실리라면 정말 무서운 일이다.

01 Feb 2020

Holier Than Thou
@BhariDhaatu

RT @Sharanyashettyy: So the conspiracy stories will be proven right? Coronavirus got accidentally leaked from a lab in Wuhan, China?

01 Feb 2020

119 Comments BioRxiv Disqus' Privacy Policy Login

Recommend 311 Tweet Share Sort by Newest

Join the discussion…

LOG IN WITH OR SIGN UP WITH DISQUS ?

Name

Huihuang • 9 days ago

This is a false claim: read the following from American Association for the Advancement of Science. https://www.bioworld.com/ar...

1 • Reply • Share ›

Timothy Takemoto • 17 days ago

I wonder if the authors have missed a correlation between nCov and HIV because they ignore the first protein in the first sequence "GTNGTKR", matching it with TNGTKR in HIV when the HIV gene sequence continues with the omitted G making TNGTKRG which, since proteins are often in circular, or spiral, "heptad repeat sequences", may be functionally or morphologically the same as GTNGTKR.

If this were the case it would make the chance of overlap twenty times less likely since there are 20 amino acids coded by gene sequences.

I am a psychologist not a biologist.

3 • Reply • Share ›

ResearchGuy • 23 days ago

I see several commenters have asked about the COMPOUND probability of ALL FOUR sequences occurring naturally in what would seem to be a section of nCoV that affects what kind of cells that can infect. I have seen no answers to that question. Someone, preferably several someone's, please answer it. If I had the skills I would do so myself, but I don't think any of the posts have even stated the exact individual probabilities so I can't multiply them together for myself.

4 • Reply • Share ›

jean-claude perez • 25 days ago

Please read https://www.preprints.org/m...

1 • Reply • Share ›

Ironman • a month ago

In the 2011 research paper by Y. Kawaoka and two colleagues at his animal virology lab at the University of Wisconsin-Madison, titled "HIV reverse-binding protein is essential for influenza A virus replication and promotes genome-trafficking in late-stage infection". Published in the Journal of Virology, September 2011. Is this current published study a reflection of the scientific reality of binding HIV to flu viruses as suggested by the Indian team? If so what purpose would that serve in Wuhan?

• Reply • Share ›

Hannah Davis • a month ago

Out of curiosity, I just did a quick check of how many HIV-1 protein sequences there are in the NCBI database, because I suspected that the virus might be over-represented.

Via the NCBI Taxonomy Browser, out of 6 012 978 total viral protein sequences, 1 169 134 are from HIV-1 alone.

For comparison, there are only 57 759 protein sequences in the database from ALL Coronaviridae combined.

This over-representation of HIV-1 in the database, combined with its famously high mutation rate, makes it VERY likely that any given short NT or AA sequence will show up in one or more HIV-1 sequence via blastp. Even if you ignore the effects of selection on viruses that may need to fold their proteins/interact with membranes/etc. in the same way.

13 • Reply • Share ›

David Murphy • a month ago • edited

Where's the link to the specific HIV-1 protiens?

Why does it not cite the relevant nuccore entries?

Looking at an example HIV-1 GP120 protein sequence (which I have to since I don't see a specific one cited)

https://www.ncbi.nlm.nih.go...

I see none of these sequences:

GTNGTKR

HKNNKS

GDSSSG

QTNSPRRA

Running a blastp alignment between these sequences and the linked HIV-1 GP120 the only match I find is *2 proteins long.*

https://blast.ncbi.nlm.nih....

Also, just for fun, lets run a BLASTN between Wuhan-Hu-1

GenBank: MN908947.3

https://www.ncbi.nlm.nih.go...

And HIV-1 complete genome

GenBank: AF033819.3

https://www.ncbi.nlm.nih.go...

With the *most permissive settings possible* we do get some hits.

4 hits with an "expect"(number of times we expect to see matches of this type by chance in the given search) score of 2.8
8 hits with an "expect" score of 9.8

AKA: chance

https://blast.ncbi.nlm.nih....

None of them are any of these 4 "inserts"

This is entirely bunk and bunk with bad citations and sourcing at that.

5 • 2 • Reply • Share ›

John Detwiler → David Murphy • 24 days ago

they give the links in their supplemental file. I easily accessed them by replacing the accession number in any blue link.

1 • Reply • Share ›

Timothy Takemoto → David Murphy • a month ago • edited

I know nothing about this field but the authors do give the references for the proteins in which they found the "inserts" in their supplementary file linked above.
https://www.biorxiv.org/con...
The contents of the supplementary file is pasted below
country Accesion ID Subtype Insert
Thailand AFU28737.1 CRF01_AE Insert 1
Thailand AFU28711.1 CRF01_AE Insert 1
Thailand AFU28717.1 CRF01_AE Insert 1
Thailand AFU28733.1 CRF01_AE Insert 1
Thailand AFU28693.1 CRF01_AE Insert 1
Thailand AFU28721.1 CRF01_AE Insert 1
Thailand AFU28699.1 CRF01_AE Insert 1
Thailand AFU28729.1 CRF01_AE Insert 1
Thailand AFU28705.1 CRF01_AE Insert 1
Thailand AFU28725.1 CRF01_AE Insert 1
Kenya ALB06757.1 G Insert 2
India ACL98861.1 C Insert 3
India ACL98864.1 C Insert 3
India ACL98860.1 C Insert 3
India ACL98859.1 C Insert 3
India AKR75206.1 C Insert 4

Thus no two inserts occur in the same HIV protein.
Further, while the whole "insert" is indeed gtngtkr in their Figure 1, in their Table 1 they drop the first protein to leave.

For the first of the above Accesions
https://www.ncbi.nlm.nih.go...
1 mrvketqrnw pnlwkwgtli lglviicsas dnlwvtvyyg vpvwrdadtt lfcasdakah
61 dtemhniwat hacvptdpnp qeirlenvte nfnmwknnmv eqmqedvisl wdqglkpcvk
121 ltplcvtlnc tnanltnvsl ttnntfpsfn itgniteevr ncsfnmttel rdkqqkvyal
181 fykldivpid rdknnssnse yrlincntsv ikqacpkvsf dpipihyctp agyailkcnd
241 kkfngtgpck nvssvqcthg ikpvvstqll lngslaeeei iirsenltnn aktiivhlne
301 svkincirps nntrtsipig pgqvfyrtgd iigdirkayc kvngtewykv ltgvtgklke
361 hfpeknisfq ppsggdpeit mhhfncr*g*ef fycnttnlfn nsc*tngtkr*g cnetiilpcr
421 ikqiinmwqg vgqamyappi rglinctsni tgilltrdgg ndtgnetfrp gggnikdnwr
481 nelykykvvq ieplgiaptr akrrvvqrek ravglgamif gflgaagstm gaasitltvq
541 arqllsgivq qqsnllraie aqqhllqltv wgikqlqarv laverylkdq kflglwgcsg
601 kiicptavpw natwsnrsye eiwnnmtwie wereisnytn qiyeiltqsq dqqdrnerdl
661 leldkwaslw nwfditnwlw yirifimivg gliglriifa vlsivnrvrq gysplsfqtp
721 thqqrepdrp erieegggeq grdrsvrlvn gflalawddl rslclfsyhr lrdfiliaar
781 tvellgrssl qglrrgwegl kylgnllayw iqelktsais llnavaiava ewtdrvieva
841 qrawrailhi prrirqgler alv

The second is
https://www.ncbi.nlm.nih.go...
1 mrvketqrnw pnlwkwgtli lglviicsas dnlwvtvyyg vpvwrdaett lfcasdakah
61 dtemhniwat hacvptdpnp qeihlenvte nfnmwknnmv eqmqedvisl wdqglkpcvk
121 ltplcvtlnc tnanltnarl ttnntfpsfn itgniteevr ncsfnmttel rdkqqkvyal
181 fykldivpid rdknnssnse yrlincntsv ikqacpkvsf dpipihyctp agyailkcnd
241 kkfngtgpck nvssvqcthg ikpvvstqll lngslaeeei iirsenltnn aktiivhlne
301 svkincirps nntrtsipig pgqvfyrtgd iigdirkayc kvngtewykv ltgvtgklke
361 hfpeknisfq ppsggdpeit mhhfncr*g*ef fycnttnlfn nsc*tngtkrg* cnetiilpcr
421 ikqiinmwqg vgqamyappi rglinctsni tgilltrdgg ndtgnetfrp gggnikdnwr
481 nelykykvvq ieplgiaptr akrrvvqrek ravglgamif gflgaagstm gaasitltvq
541 arqllsgivq qqsnllraie aqqhllqltv wgikqlqarv laverylkdq kflglwgcsg
601 kiicptavpw natwsnrsye eiwnnmtwie wereisnytn qiyeiltqsq dqqdrnerdl
661 leldkwaslw nwfditnwlw yirifimivg gliglriifa vlsivnrvrq gysplsfqtp
721 thqqrepdrp erieegggeq grdrsvrlvn gflalawddl rslclfsyhr lrdfiliaar
781 tvellgrssl qglrrgwegl kylgnllayw iqelktsais llnavaiava ewtdrvieva
841 qrawrailhi prrirqgler alv
which has the same "tngtkr"

The authors claim that
"In other words, one may sporadically expect a fortuitous match for a stretch of 6-12 contiguous amino acid residues in an unrelated protein. However, it is unlikely that all 4 inserts in the 2019-nCoV spike glycoprotein fortuitously match with 2 key structural proteins of an unrelated virus (HIV-1). "

But this *may* not be not as unlikely as the say bearing in mind the very large absolute number of HIV virus proteins in the database, and the fact that there are an especially large relative number of HIV viruses in the database compared to the number of other virus proteins.

Searching the database for "virus" the hits are

Human immunodeficiency virus 1 (1,165,203)
Influenza A virus (929142)
Hepacivirus C (198829)
Hepatitis B virus (161444)
Influenza B virus (138359)
All other taxa (2,496,781)
Even assuming that "all other taxa" are all viruses this implies that in the database, 1 in 3.7 viruses are variants of HIV 1.

However, if as they say these "inserts" (the differences between the new virus and SARS), might equally well come from any of the proteins in the database, there are "162,117,248 items" that contain the word "protein" so only 1 in 150 "protein"s are HIV 1.

• Reply • Share ›

Russell Hughes → David Murphy • a month ago

You are correct total misinformation top to bottom

• Reply • Share ›

Dottor Leopardo • a month ago

identity /similarity to amino acid residues in key structural proteins of HIV-1 is QUITE COMMON in nature (starting from the human Red Blood Cells surface).

4 • Reply • Share ›

Jason Weir • a month ago • edited

I give credit to the authors who have now withdrawn this manuscript and have stated that they are reworking it to incorporate feedback from the research community.

3 • 1 • Reply • Share ›

Death Syndrome • a month ago

If it express gp120 so why it doesn't infect immune system?

• Reply • Share ›

jay wang • a month ago

I BlastPed the region that spans the first 2 “insertions”, and found no so called insertions at all in the alignments with other bat-cov viruses. On the contrary the alignments showed very natural diversities around the two ”insertion” regions. The two “insertions” are obviously products of evolution， not engineered artifacts！ Furthermore， even human has huge number of proteins that are homologous with those of vegetables， so why it is a surprise that there are short homologous regions between 2 viruses？

• Reply • Share ›

jay wang • a month ago

I BlastPed the region that spans the first 2 “insertions”, and found no so called insertions at all in the alignments with other bat-cov viruses. On the contrary the alignments showed very natural diversities around the two ”insertion” regions. The two “insertions” are obviously products of evolution， not engineered artifacts！ Furthermore， even human has huge number of proteins that are homologous with those of vegetables， so why it is a surprise that there are short homologous regions between 2 viruses？

4 | 19 • Reply • Share ›

Chris Avellonea • a month ago

Can someone please explain why these "inserts" (wherever they're from, HIV or other species) form a conformational cluster when the protein is folded? What's more, does this conformational cluster have a physiological function, i.e., a binding site? Please, someone with expertise help explain this.

9 | 2 • Reply • Share ›

Connor • a month ago

I'm a logician so I'm fond of synopsis
I'm not a geneticist

Reading through the board though

There was bat virus = A
There's now ncov19 = B

The difference between A and B rna is the insertion/addition of 4 sequences
These 4 sequences are present in HIV
They are also present in some related bat stuff and also not that uncommon generally

There is also an artefact of the complexity of this stuff that a possible and previously seen (1919 flu) change in a very small common sequence can precipitate the rest
We don't know the correlations there

What I'd like an opinion on is

How likely is it that
The change from A to B happened in such a time frame commensurate with the emergence of B purely as an organic process

Can I get a breakdown of probabilities (what u guys call 'E values'

Thanks

15 | 1 • Reply • Share ›

Emilian Snarski → Connor • 20 days ago

http://virological.org/t/ta...

| • Reply • Share ›

Brian Hanley → Connor • a month ago

100% probability. If you read the comments, these guys cherry picked matches to HIV that aren't even in the top 100 matches. It's a nonsense paper that is utterly irresponsible.

| • Reply • Share ›

AronF • a month ago

We really ought to be given the 'E value' of the BLAST hits to the HIV sequences in order to determine the actual "unlikely[hood]" of the 2019-nCov matching the HIV-1 inserts. That is the very reason the E-value exists. The fact of 100% coverage means little if there is an abundance of hits with 95% coverage for example - in this case the E value would be high.

2 | • Reply • Share ›

Zhenguo Zhang • a month ago

The study shows something interesting, but the title is just exaggeration. Furthermore, if you look at Table 1, the inserts are not so similar to HIV proteins except insert 1 and 2. Also these inserts have been found in the bat virus (Fig. S2). But these findings has potential to explain the contagiousness of this new virus.

8 | 1 • Reply • Share ›

Milwaukee • a month ago

The 1/30/2020 Lancet study of 99 patients from Wuhan found that T-cell count was significantly suppressed. If there are GP120 amino acids present in 2019-nCOV spike protein that could allow it to bind to CD4, it could explain the low T-cell count.

| • Reply • Share ›

JoshP • a month ago

Interesting observations and discussion below, requires further inquiry. Things to keep in mind: (1) Blasting each insert individually unsurprisingly identifies multiple hits. But I am unable to find any virus besides HIV-1 that contains ALL FOUR inserts in their totality together (in the gp120 and gag proteins). This can be confirmed at both the nucleotide and amino acid levels. This, to my mind, changes the probability calculus and requires further study to understand the basis of these findings. (2) A great deal of weight is being placed on this single uploaded sequence (RaTG13, QHR63300.1), stated to be from 2013 Yunnan bat feces, deposited by the Wuhan Institute of Virology on 1/29/20. Clearly it will improve confidence in the scientific community if the provenance of this one critical linking sequence can be independently verified as an actual, naturally-occurring sequence isolated as stated in 2013 from bat feces. (3) Interestingly, the nucleotide sequences of inserts 2 and 3 in RaTG13 are identical to those in the 2019-nCoV virus, not even a single synonymous substitution since 2013. This strikes me as intriguing and indeed worthy of further study. It will be interesting to see how stable the nucleotide sequences of inserts 2 and 3 remain in this human outbreak in thinking about how likely it is that they would be unaltered since 2013 in the natural bat (or other reservoir or intermediate) population.

13 | • Reply • Share ›

Hannah Davis → JoshP • a month ago • edited

One issue with putting so much emphasis on the "inserts" (which, I should note, are actually NOT present in their entirety in HIV; if you compare the inserts in fig. 2 to the inserts actually matched to various HIV-1 seqs in Table 1, the latter sequences are actually shorter, and only the first two match 100% even after being shortened to 6 nucleotides) is that 1) HIV-1 mutates a lot, and 2) there are a ton of HIV-1 sequences in the database.

In other words, there's a fairly good chance that virtually ANY sufficiently short sequence of amino acids that is in any way compatible with being a functional RNA virus will be found in some HIV-1 sequence somewhere. This is an expected artefact of the amount of time/money that has gone into studying HIV-1 via sequencing.

The authors didn't even find exact matches in the same HIV-1 genome, but in separate ones. You can check this yourself - look at the seqs that they say match "Insert 1" and look for "Insert 2" in them. There are some vaguely similar sequences, I guess, but being vaguely similar to a mere 6 amino acid sequence isn't very impressive - especially given that there are numerous better matches in non-viruses and viruses alike!

And Insert 2, *as shown in Fig 2*, doesn't actually match ANY HIV-1 sequence: the full seq is "YYHKNNKS" (note: this seems arbitrary, because it includes both a short gap and then four amino acids to the right of the gap, "NNKS", that differ in SARS and 2019-nCoV, but none of the subsequent differing amino acids). The shortened sequence that they used to find a 100% match to a single HIV-1 database entry is "HKNNKS" - and they would not have had a 100% match without shortening it, as the previous two amino acids in the HIV-1 seq are "GT", not "YY".

2 | • Reply • Share ›

Kairui Sun → JoshP • a month ago

Thank you Josh for your insights. I looked up the most recent genome sequenced for the 2019-nCoV (GenBank: MT007544.1) and found a mutation in the nucleotide sequence for insert 3, resulting in a Ser to Arg mutation. Since all other nucleotides for the Spike protein align perfectly among 5 2019-nCoV sequences I aligned, it's very unlikely that it is a sequencing error. How would you interpret this result given your point about the stability of the insert?

Reply • Share ›

Dr. Andy Crispr • a month ago

I think their findings are interesting and merit further scientific investigation. Specifically, because of its novelty, we still have quite a low number of nCoV samples sequenced and, given the virus' high mutation rate and given the intrinsic variability of genome editing techniques, it would be interesting to see if more sequence data would come up that potentially further underscores this proposed connection to HIV. The fact that this connection, if weak, has been found so early after the first discovery/sequencing of the virus is astounding.

10 | 3 • Reply • Share ›

Laurie McGee • a month ago

SARS Coronavirus and HIV minute similarity are mentioned in this research back in 2003. https://link.springer.com/a...

11 | 1 • Reply • Share ›

Aaron Holmgren → Laurie McGee • a month ago

From the abstract..."which are unique to the 2019-nCoV and are not present in other coronaviruses"

1 | • Reply • Share ›

Max Combest → Laurie McGee • a month ago • edited

The link you include says explicitly that there is no sequence homology between the two. They are using HIV as an example to compare modes of entry into host cells. They work similarly, but use completely different proteins. So they propose using a treatment working from the CoV sequence analogous to an HIV treatment that uses its corresponding sequence.

4 | • Reply • Share ›

Prashant Pradhan • a month ago

This is a preliminary study. Considering the grave situation, it was shared in BioRxiv as soon as possible to have creative discussion on the fast evolution of SARS-like corona viruses. It was not our intention to feed into the conspiracy theories and no such claims are made here. While we appreciate the criticisms and comments provided by scientific colleagues at BioRxiv forum and elsewhere, the story has been differently interpreted and shared by social media and news platforms. We have positively received all criticisms and comments. To avoid further misinterpretation and confusions world-over, we have decided to withdraw the current version of the preprint and will get back with a revised version after reanalysis, addressing the comments and concerns. Thank you to all who contributed in this open-review process.
: Authors of the Manuscript

44 | 10 • Reply • Share ›

dave → Prashant Pradhan • a month ago

What are the similarities between 2019-nCov and Ebola?

| • Reply • Share ›

Luis Yáñez → Prashant Pradhan • a month ago

The main concern is the fact that finding 6 aminoacids conserved between two cherry-picked sequences is enough for you to call it an insertion, rather than considering homoplasy as the most likely explanation first.

| • Reply • Share ›

Luis Yáñez → Prashant Pradhan • a month ago

It's very sad that they have used your data for creating these ridiculous conspiracy theories, this is not your fault. However, you should be ashamed for considering the conservation of 6 amino acids enough to jump to conclusions about 'insertions' rather than considering homoplasy first, as the most likely answer

| • Reply • Share ›

Dzogchen • a month ago • edited

This report highlights the dangers of assuming significance to a highly improbable yet random occurrence. If one calculates the probability of finding all four peptides within the HIV-1 genome it is improbable but that does not infer non-randomness. Lots of highly improbable events happen in nature that are in fact random. Even if you constrain to just the viral sequences in the database which are nearly 6 million residues (protein) the probability is quite low that all 4 peptides match HIV-1 but the authors fell into the trap of assigning significance to randomness.

9 | 1 • Reply • Share ›

torque • a month ago

Jason, I took a look at the blast results. The Wuhan seafood market virus does seem to match the bat coronavirus. However, if you click on the Accession (QHR63250.1 and QHR63300.1) you can see that both were submitted on the same day, 27-JAN-2020 by CAS Key Laboratory of Special Pathogens, Wuhan Institute of Virology. There are some subtle differences in "ORIGIN". It may be instructive to see what those differences are.

8 | • Reply • Share ›

reuns → torque • a month ago • edited

QHR63300.1 is the decoded spike protein from the genome RatG13 (now indexed as MN996532.1) the closest known bat virus to nCov, and QHR63250.1 is nCov decoded spike protein.

| • Reply • Share ›

torque → reuns • a month ago • edited

@reuns et al. How do I locate other bat coronavirus' that have one or more of these "inserts"? (For anyone else that was wondering, the notes indicate QHR63300.1 was collected 24-Jul-2013 - fecal swab from bats. But still it was submitted into the system on 27-Jan-2020.)

1 | • Reply • Share ›

戸谷剛 • a month ago

.. Is this suggesting that the primary Coronavirus infects to HIV patients in Wuhan and mutated with transferring from hiv gene ?? Very aweful scenarios but possible.

3 | • Reply • Share ›

Matthew → 戸谷剛 • a month ago

It would be possible for nCov to mutate inside a person with HIV and pick up traits. Based off of what the HIV virus is currently doing to a person's RNA by the time the nCov takes its taste of the proteins that have already been altered. This nCov mutates.

| • Reply • Share ›

Mingchiu Fung → 戸谷剛 • a month ago

Can a CoV and HIV undergo recombination? I doubt it.

1 | • Reply • Share ›

Professor2 → Mingchiu Fung • 22 days ago

They certainly can if they infect the same cell. However, they don't so far as I know.

| • Reply • Share ›

Ironman → Mingchiu Fung • a month ago

You may want to examine the paper: "HIV reverse-binding protein is essential for influenza A virus replication and promotes genome-trafficking in late-stage infection". Published in the Journal of Virology, September 2011.

Reply • Share ›

Jing Hou → 戸谷剛 • a month ago

That's unlikely as the coding sequences are actually different, abeit having the same peptide stretches randomly

Reply • Share ›

illumined • a month ago

A layman has a question if I may. The arguments against HIV having been artificially spliced into this coronavirus make sense, but leaves me to wonder if there isn't really a relationship been this novel coronavirus and HIV then why does the coronavirus respond to HIV protease inhibitors? Is it just something about those types of drugs that makes it work in diseases beyond HIV? Thank you.

3 Reply • Share ›

Thomas Peskie → illumined • a month ago • edited

HIV uses a lot of the same proteins that corona viruses in general use. These drugs were also tested on SARS with some success.

Source: https://www.sciencemag.org/...

5 Reply • Share ›

ewyler → illumined • a month ago

Hi, regarding the protease inhibitor: both HIV and Coronaviruses (and other viruses) depend on proteases (proteins which cleave other proteins) for proper function. In human cells this is only rarely observed, but many viruses make the proteins in huge chunks (polyproteins), which are subsequently cleaved into individual proteins that each carry out different functions. This is discussed to be generally a favorable property for viruses (https://www.ncbi.nlm.nih.go.... Whether this is a principle based on an ancient common ancestor of all viruses, or a product of "convergent evolution" (i.e. that the same principle has been acquired over time by different viruses) remains elusive. Since these HIV protease inhibitors are known to be non-toxic to humans, it therefore makes sense to test them whether they work also against the Coronavirus protease. I would however say it would be big luck if they do.

4 Reply • Share ›

Badguy → illumined • a month ago

Most probably, it responds to the
HIV protease inhibitors since it share those 4 inserts with HIV.

2 Reply • Share ›

Niklas E. → Badguy • a month ago

I think that's unlikely to be the reason since the inserts are in the spike proteins (the surface of the virus) and not in the proteases. So even if there are some sequences from HIV there are not in the part that is targeted by the HIV drugs.

Reply • Share ›

N_Anthem • a month ago

How about the "phenotypic mixing"? If there is an early case that has been infected by both CoV & HIV,wouldn't there be a chance bringing out the nCoV with that 4 inserts ?

Reply • Share ›

PedroB → N_Anthem • a month ago

How would that explain their positions within the sequence chain and their direct effect on contributing to the structure of the receptor binding site? That question is really bothering me so if anyone has a solid explanation for it I'd really appreciate it. Thanks.

1 Reply • Share ›

Load more comments

Subscribe Add Disqus to your siteAdd DisqusAdd Do Not Sell My Data

Back to top