Emergence of a synthetic recombinant Spike protein in the Wuhan SARS CoV-2 virus

Abstract
The origin and cause of the SARS CoV-2 Virus is still unknown despite the massive global impact through the pandemic.This study analyzed the Wuhan SARS CoV-2 spike protein gen sequence (genbank association no. MN908947.3) and compared manually with a synthetic recombinant BAT-SARS Spike protein SRBD gen sequence (genbank association no. FJ211860) that has been published in 2008 by Becker and Graham (1). Both genomes have been then compared with the SARS CoV(AY864806) genome of a SARS Beijing patient from 2006.

The results showed that the Wuhan SARS CoV-2 Spike protein shares 85% nucleotids with the synthetic recombinant BAT-SARS Spike domain which has been designed to simulate a cross-species transmission. While the synthetic BAT-SARS Spike protein matched the Wuhan SARS CoV-2 Spike protein, it showed no matches with the SARS CoV spike protein. Questions arise as the SARS CoV genome from 2006 shared over 60% nucleotids with the Wuhan SARS CoV-2 genome exempt for the spike protein domaine. Based on this research an engineered SARS CoV-2 with a designed synthetic recombinant BAT-SARS spike protein can be discussed and should be part of future researches.

Introduction

The severe respiratory disease COVID-19, previously named novel coronavirus or SARS CoV-2 was first noticed in late December 2019 in China (2), and declared officially on 11 March 2020 a pandemic by the World Health Organization (3)
Although genetically identified as a new type of the betacoronaviruses that causes SARS and the Middle East respiratory syndrome (MERS) in humans (4), frequent scientific researches indicate that the SARS CoV-2 is closely related to the bat derived SARS like-syndrom (5).

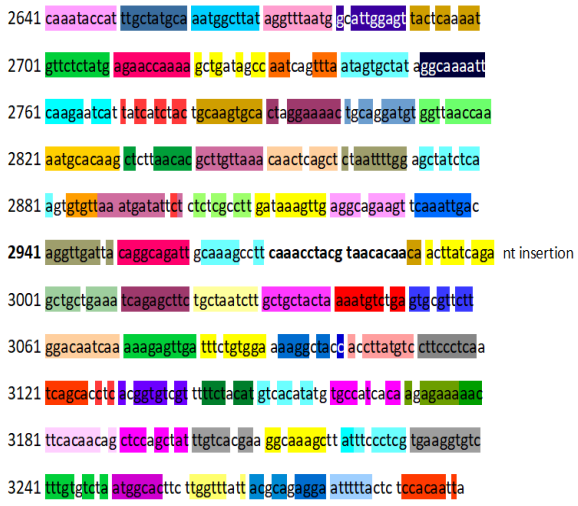
The SARS CoV-2 spike protein locates the human Angiotensin-Converting-Receptor 2 (ACE) cells and inserts his virus gen through a whole into the ACE2 cell (6)
Becker and Graham published in December 2008 a genome sequence as a result of a synthetic recombination of a novel chimeric virus which identified genetic cross-species transmission between a BAT SARS- like Coronavirus (BAT-SCoV) and the SARS Coronavirus.

The authors designed a consensus BAT-SCoV from 4 reported BAT-SCoV sequences (genbank association no. FJ211859). In a second step, they replaced the spike Receptor-Binding Domain (RBD) of the BAT-SCoV with the SARS CoV RBD simulating a recombination of mixed cross-species infection in vitro. Finally, the new synthetic recombinant BAT-SARS RBD (SRBD) (genbank association no. FJ211860) has been cultivated in cells and mice. The synthetic recombinant virus showed effective growth and infection in human and mice cells (1).

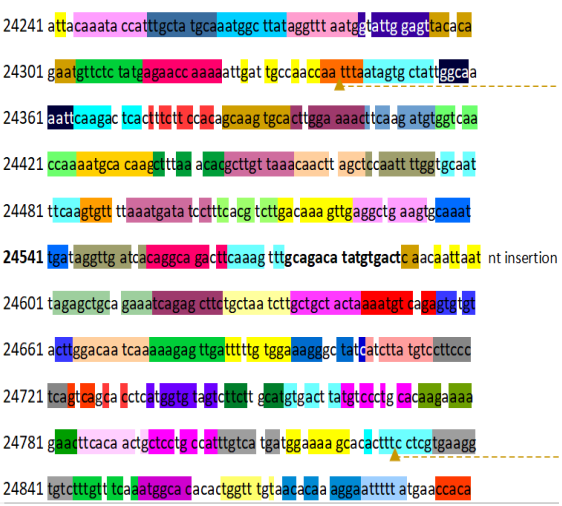
This study compares the SRBD sequence (genbank association no. FJ211860) with the Wuhan SARS CoV-2 virus genome (genbank association no. MN908947.3) to understand the structure and origin of the novel Corona virus. In a second step, this study compares both gen sequences with the SARS CoV genome (genbank association no. AY864806) from a stole sample of a Beijing SARS patient from 2006.

Results

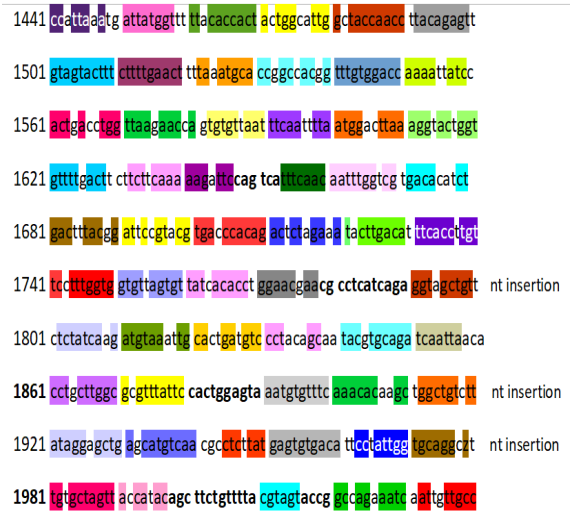
1. The Wuhan SARS CoV-2 gene fragments (nt 21807 – 21933) and (nt 22321 – 25384) matched with the ones from the synthetic recombinant SARS BAT-SRBD (nt 244 – nt 380) and (nt 701 - 3721). The pictures 1 and 3 show a screenshot of the SRBD gen sequence that matches with the SARS CoV-2 spike protein sequence in picture 2 and 4. Noticable are multiple point mutations and frequent insertion of short nucleotid sequences in both genomes, here in bold.



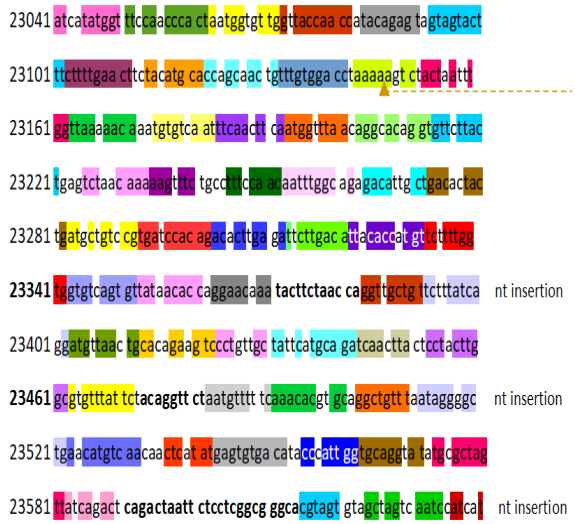
Pic 1 SRBD gen



Pic 2 Wuhan SARS CoV-2



Pic 3 SRBD gen



Pic 4 Wuhan SARS CoV-2

However, the SRBD gene domaine had two noticable gen sequences which show no matches to the Wuhan SARS Spike protein domain. The first gap was at Wuhan SARS CoV-2 (nt 21601- 21796), likewise at SRBD (nt 1 - 243) and a second gap at Wuhan SARS CoV-2 (nt 21934 – 22278) compared to SRBD (nt 382 - 691). Alltogether, the results show that the recombinant SRBD Spike protein gen matches in 85% with the Wuhan SARS CoV-2 Spike protein gene.

2. The Wuhan SARS CoV-2 genome (nt 1- nt 29870) matched in 19,551 nucleotids or 60% of the genome with the SARS CoV genome (nt 29730). Noticable are multiple point mutations in both genomes. Table 1 lists the nucleotid matches based on virus protein structure.

SARS CoV-2 (no. MN908947.3)	SARS CoV (AY864806)	Virus protein
nt 6 - 1181	nt 6 - 1183	Leader protein
nt 4861 - 21598	nt 4861 - 4800	Proteinase, TM-2, 3C Proteinase, Transmembran, Binding protein, growth factor, RN Polymerase, Helicase, Exonuclease, Endonuclease
nt 28231 - 29870	Nt 28084 - 29727	N, 10

3. The recombinant SRBD (nt 3271) spike protein gene showed no matches with the SARS CoV spike protein gene.

Discussion

The genome of a species is unique like a fingerprint. The emergence of a synthetic recombinant BAT-SARS spike protein in the Wuhan SARS genome has different explanations. L.F Wang et al. explained that different horseshoe bat species from the genus Rhinolophus has been identified as a reservoir for number of viruses with close genetic relationship to the SARS CoV virus which can be responsible for the past SARS outbreakes in 2002/2003 and 2003/2004. The discovery of viruses from the SARS-like CoV group in bats support the transmission between different virus groups (7).

However, several research groups like Wang or Marra et al. seperately identified the SARS CoV-2 virus as a new virus without genetic relationship to any human Coronaviruses (8).
The discussion leads to the question, if the BAT SARS-like CoV evolved naturally through transmission why the Wuhan SARS CoV-2 Spike domain bears no resemblance with the SARS CoV Spike domain,although both genomes share half of the genome, exempt for the most important virus part, the spike protein?

Wang et al. suggested that the genome sequence of all bat SARS-like Co was almost identical with the SARS CoV genome particularly for the Spike protein domain. This study did not confirm the observations. And showed no matches between the spike gene sequence of the SARS CoV and the BAT SARS- like synthetic recombinant.

Also discussed must be the possibility that the BAT-SARS synthetic recombinant Spike protein had not evolved by natural transmission such as sustained transmission or virus adaptation. Wang et al. reported that a direct natural transmission from animal to humans is unlikely because of the unability of the BAT SARS- like virus to use the ACE2 receptor for human infection.
The possibility of a designed cross-species transmission by replacing a recombinant receptor binding protein can overcome the limiting factor for animal to human infection and could be an argument for a engineered SARS CoV-2 virus. Wang et al. already predicted in 2006 that experimental infection of the SARS CoV with different bat species through substantial genetic changes by recombination and adaption like point mutations can improve the ability of direct human infection. As Becker and Graham reported, no studies to the date of their publication has

used a synthetic approach to assess potential mechanism of zoonotic emergence of a noncultivable virus.

However, the possibility of synthetic constructed RNA virus strains with gene fragments from different virus groups was already successfully presented by B.Wei et al. in 2008. This study confirmed that an amored-RNA containing long chimeric RNA with 5 different virus fragments such as three SARS CoV fragments, one HCV fragment, and one H5N1 Influenza virus fragment was feasable for assay controls in clinical laboratories (9).

Another feature of the sytnthetic recombinant SRBD gene sequence is worth discussing. If theoretically the SARS CoV-2 was engineered with synthetic recombinant SRBD, then specific gene mechanism for RNA-sequencing techniques such as editing, processing, and stabilizing would be necessary. Studies by Alice Barkan showed that designed pentacotripeptide protein repeat proteins (PRP) bind proteins to specific RNA sequences, activate, and stabilize specific RNAs in vivo (10,11). This study observed frequent short nucleotid sequences at same positions in the SARS CoV-2 spike protein domain and the SRBD spike domain. The results suggested that these nucleotid sequences might be PRP proteins which are required for engineering novel SARS viruses with specific recombinant RNA spike proteins. However, the emergence of PRPs within the SARS CoV-2 genome was not part of this study and should be part of future researches.

To Summarize, the observation of the SARS CoV-2, the SARS CoV and the chimeric recombinant virus strain gave plausible conclusions for an engineered SARS CoV-2 virus rather than a virus that has been evolved through natural transmission and natural selection process.

This study likes to emphasize one aspect of the Becker and Graham research. While the study suggested already in 2008 that such synthetic reconstructed cross-species transmissions allow for testing of vaccines against future zoonotic virus strains, interestingly, it predicted at the same time that future SARS CoV vaccines would provide significant protection against other zoonotic natural or also deliberately designed SARS CoV threads. The questions arises if future vaccine production and distribution were required if we don't make laboratory experiments with engineered zoonotic cross-species virus threads.

Literature

1. Michelle M. Becker, Rachel L. Graham, Eric F. Donaldson, Barry Rockx, Amy C. Sims, Timothy Sheahan, Raymond J. Pickles, Davide Corti , Robert E. Johnston , Ralph S. Baric, and Mark R. Denison (2008) Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. PNAS December 16, vol. 105, no. 50: 19944 –19949

2. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. Lancet 395, 565–574 (2020)

3. W. H. Organization, Novel Coronavirus (COVID-19) Situation (2020); https://experience.arcgis.com/experience/685d0ace521648f8a5beeeee1b9125cd.

4. N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan (2020) A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733.

5. Chih-Cheng Lai, Tzu-Ping Shih,Wen-Chien Ko, Hung-Jen Tang, and Po-Ren Hsuehe Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges (2020) Ijantimicag 55,3:105924.

6. Lane B. Scheiber II (MD), Lane B. Scheiber (Scd) (2020) Cure and Vaccine, Striking Back at Covid-19: Volume 6

7. Lin-Fa Wang, Zhengli Shi, Shuyi Zhang, Hume Field,Peter Daszak, and Bryan T. Eaton (2006) Review of Bats and SARS Emerging Infectious Diseases. Vol. 12, No. 12

8. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, et al.(2003) The genome sequence of the SARS-associated coronavirus. Science.300:1399–404.

9. Baojun Wei, Yuxiang Wei, Kuo Zhang,Changmei Yang, Jing Wang,Ruihuan Xu,Sien Zhan,Guigao Lin,Wei Wang,Min Liu,Lunan Wang, Rui Zhang, Jinming Li (2008) Construction of Armored RNA Containing Long-Size Chimeric RNA by Increasing the Number and Affinity of the Pac Site in Exogenous RNA and Sequence Coding Coat Protein of the MS2 Bacteriophage. Intervirology51:144–150

10. Jana Prikryl, Margarita Rojas, Gadi Schuster, Alice Barkan, (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. PNAS,vol. 108 : no. 1, 415–420.

11. Rafael G. Miranda, James J. McDermott, Alice Barkan (2018) RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR–RNA interactions Nucleic Acids. Research,Vol. 46, No. 5 :2613–2623