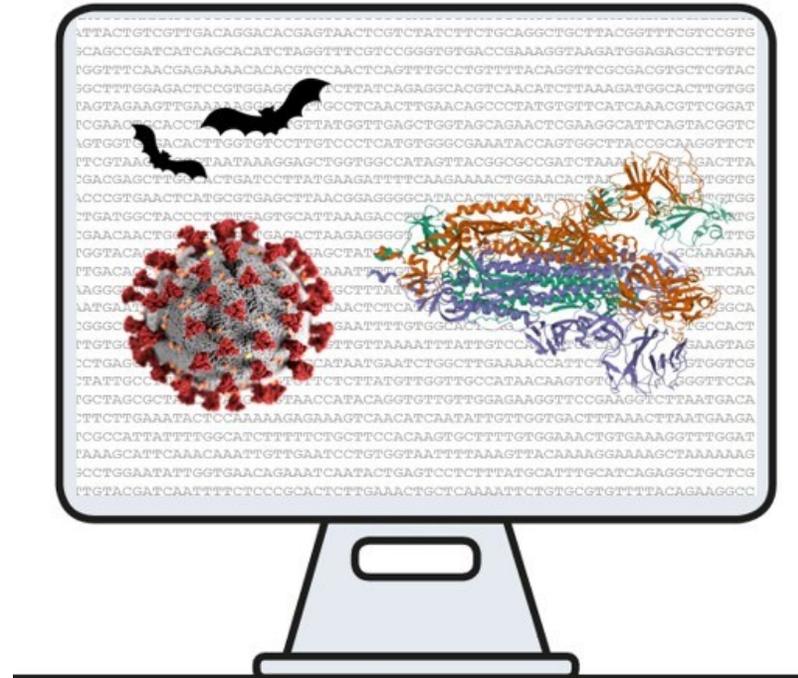


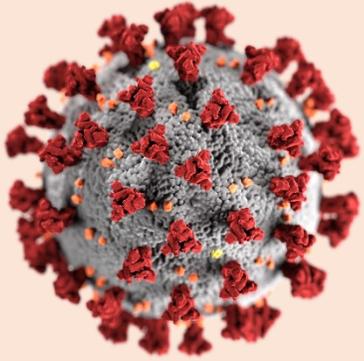
A la découverte du nouveau coronavirus et de la [bioinformatique](#)...

*Cet atelier permet de
... découvrir SARS-CoV-2 et des
banques de données & outils
bioinformatiques utilisés par les
chercheurs du monde entier!*



*...de comprendre l'importance
d'avoir librement accès à toutes
les données (**open data**)*

*...de mieux comprendre les démarches
scientifiques et les défis...*



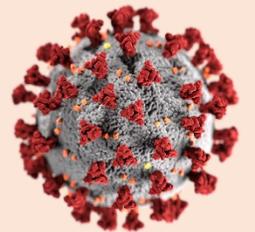
1 - Introduction

Biologie

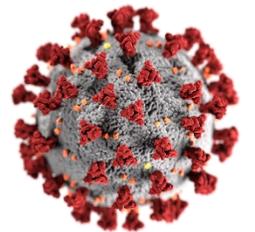
Des virus ?

Des virus qui infectent l'homme ?

SARS-CoV-2 ?



C'est quoi un virus ?



Un virus est une particule microscopique qui contient du matériel génétique (ADN ou ARN). Un virus ne peut se répliquer qu'en pénétrant dans une cellule et en utilisant la machinerie cellulaire.

Il existe des virus qui infectent des animaux et d'autres qui infectent des végétaux. Des virus peuvent aussi infecter les bactéries: ce sont les bactériophages.

S'ils provoquent des maladies, les virus sont considérés comme pathogènes.

Des virus, des chiffres et nous...

Un article de 2011 publié dans Nature Microbiology estimait qu'il y avait plus d'un quintillion - un suivi de 30 zéros- de virus sur Terre.

Dans une étude de 2018, Suttle a découvert que plus de 800 millions de virus se déposaient sur chaque mètre carré de Terre chaque jour.

Nous touchons des centaines de millions de virus chaque jour.

8 % du génome humain est d'origine virale. Une protéine d'origine virale, la syncytine a joué un rôle clé dans l'évolution des mammifères, permettant l'apparition du placenta.

[Autres protéines humaines d'origine virale](#) (UniProtKB)

Un virus peut produire plus de 1000 copies de lui-même par jour

"Ce sont les organismes les plus diversifiés de notre planète (...) et nous ne savons encore rien d'eux".

Les virus qui infectent les bactéries (bactériophages) jouent un rôle important pour en régulant les populations de bactéries nocives pour l'homme. Ils pourraient être une alternative aux antibiotiques dans certains cas (phagothérapie).

Le monde vivant ne pourrait pas exister sans les virus !

"Nous avalons plus d'un milliard de virus chaque fois que nous allons nager"

Les virus jouent un rôle clé dans la régulation du carbone dans l'océan.

[Vidéo 1](#) (2:29)

C'est quoi un virus ?

Quel(s) rôle(s) jouent les virus dans l'environnement ?



The image shows a YouTube video player interface. At the top, there is a search bar with the text "Search" and a magnifying glass icon. Below the search bar, the video player is displayed. The video content features a white background with a large red and grey virus particle illustration in the top left and bottom right corners. In the center, the text "C'est quoi un virus ?" is displayed with a play button icon. Below the title, the presenter's name "Phillippe Le Mercier" is listed, followed by his credentials: "Biologiste, Doctorat en virologie - Institut Pasteur", "ViralZone, Groupe Swiss-Prot, Genève", and "SIB Swiss Institute of Bioinformatics". Logos for "SIB Swiss Institute of Bioinformatics" and "Unil Université de Lausanne Service Culture et Médiation scientifique" are visible in the top right corner. The video player controls at the bottom show a play button, a progress bar at 0:00 / 2:29, and icons for closed captions, settings, and full screen.

Des virus qui infectent l'homme ?

850'000 virus susceptibles de passer du règne animal à l'homme...

6 grandes pandémies depuis le début du 20ème siècle...

Importance de la protection de l'environnement...(RTS, 30 nov 2020)

Plusieurs milliers de virus différents 'infectent' l'homme

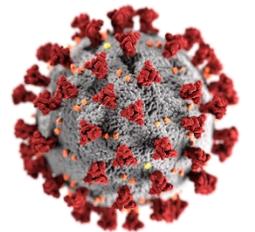
(on ne les a pas encore tous découverts...),

mais seulement ~ 160 sont responsables d'une pathologie et ont donc été bien étudiés

(grippe, SIDA, herpès, hépatite, verrue, ...)

[Le virome humain \(Scientific American\)](#)

“Biologists estimate that 380 trillion viruses are living on and inside your body right now—10 times the number of bacteria.”



Il existe environ 160 virus 'pathogènes' connus pour infecter l'homme.
 Les modes d'infection (Transmission) sont variés de même que les pathologies associées (Disease)

<https://viralzone.expasy.org/678>



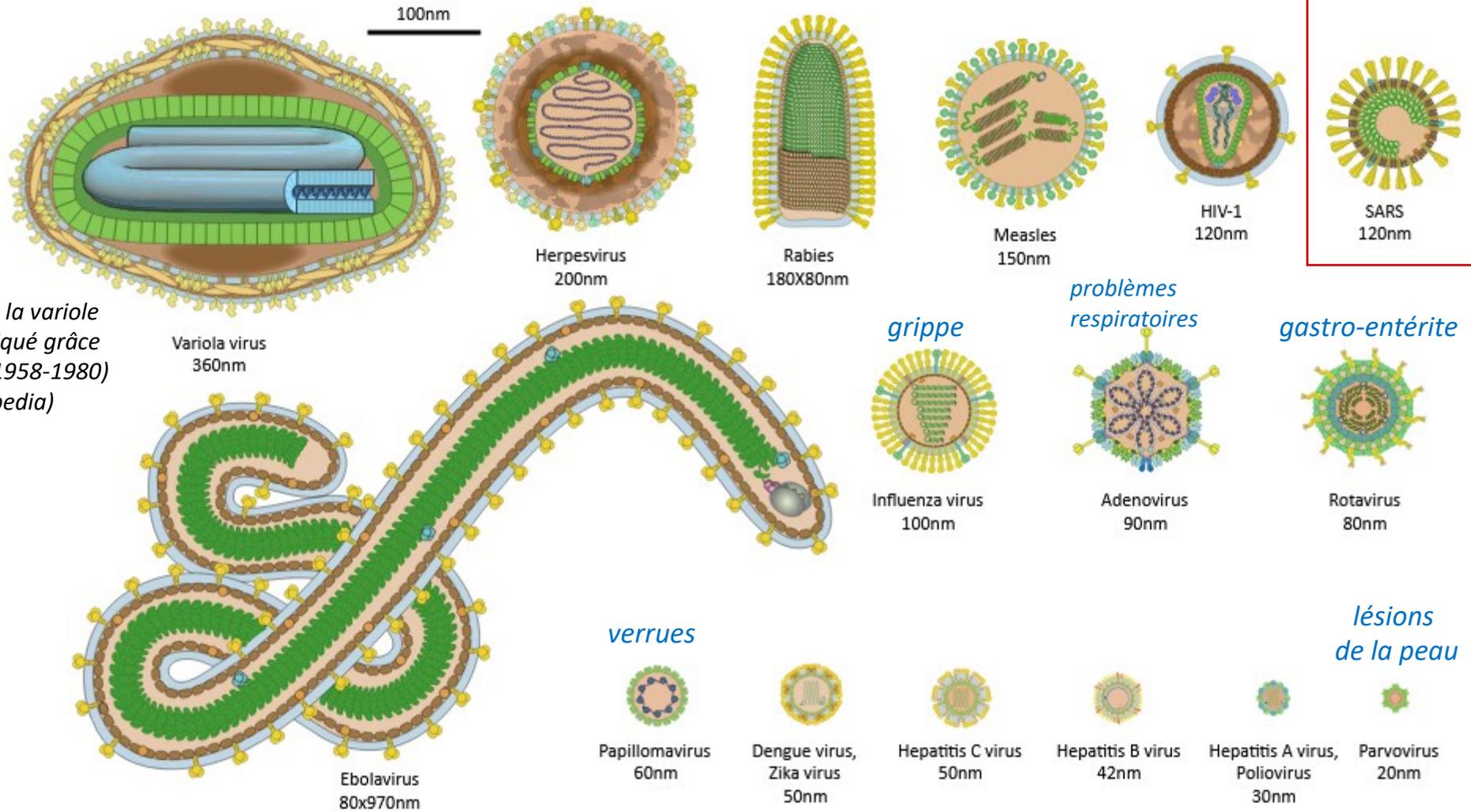
Human viruses and associated pathologies

The table below displays the list of human viral pathogens, with transmission and general facts about associated pathologies.

(See [human viruses by Baltimore classification](#))

Virus	Genus, Family	Host	Transmission	Disease
Adeno-associated virus	Dependovirus , Parvoviridae	Human, vertebrates	Respiratory	None
Aichi virus	Kobuvirus , Picornaviridae	Human	Fecal-oral	Gastroenteritis
Australian bat lyssavirus	Lyssavirus , Rhabdoviridae	Human, bats	Zoonosis, animal bite	Fatal encephalitis
BK polyomavirus	Polyomavirus , Polyomaviridae	Human	Respiratory fluids or urine	None
Banna virus	Seadornavirus , Reoviridae	Human, cattle, pig, mosquitoes	Zoonosis, arthropod bite	Encephalitis
Barmah forest virus	Alphavirus , Togaviridae	Human, marsupials, mosquitoes	Zoonosis, arthropod bite	Fever, joint pain
Bunyamwera virus	Orthobunyavirus , Bunyaviridae	Human, mosquitoes	Zoonosis, arthropod bite	Encephalitis
Bunyavirus La Crosse	Orthobunyavirus , Bunyaviridae	Human, deer, mosquitoes, tamias	Zoonosis, arthropod bite	Encephalitis
Bunyavirus snowshoe hare	Orthobunyavirus , Bunyaviridae	Human, rodents, mosquitoes	Zoonosis, arthropod bite	Encephalitis
Cercopithecine herpesvirus	Lymphocryptovirus , Herpesviridae	Human, monkeys	Zoonosis, animal bite	Encephalitis
Chandipura virus	Vesiculovirus , Rhabdoviridae	Human, sandflies	Zoonosis, athropod bite	Encephalitis
Chikungunya virus	Alphavirus , Togaviridae	Human, monkeys, mosquitoes	Zoonosis, arthropod bite	Fever, joint pain

Des exemples de virus 'pathogènes' qui infectent l'homme, leur taille respective et les pathologies associées

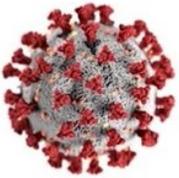


Le virus de la variole a été éradiqué grâce au vaccin (1958-1980) (Wikipedia)

[Vidéo 2](#) (2:33)

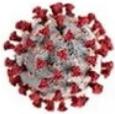
Pourquoi certains virus rendent malades ?
D'où viennent les virus qui infectent l'homme ?
Quel est le rôle des chauve-souris ?

☰ YouTube^{CH} 🔍



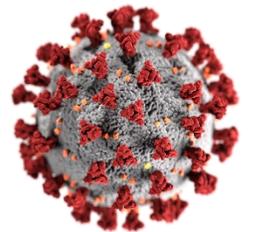
Pourquoi certains virus 'rendent malades' ?

Phillippe Le Mercier
Biologiste, Doctorat en virologie - Institut Pasteur
ViralZone, Groupe Swiss-Prot, Genève
SIB Swiss Institute of Bioinformatics

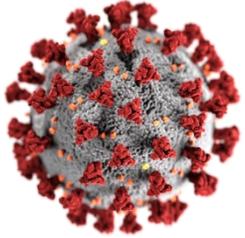


▶ ⏪ 🔊 0:00 / 2:33     

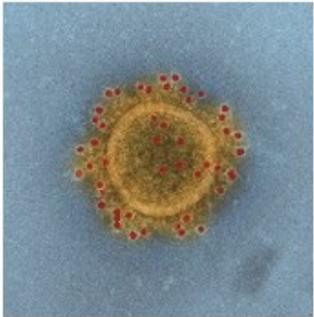
Les coronavirus et SARS-CoV-2



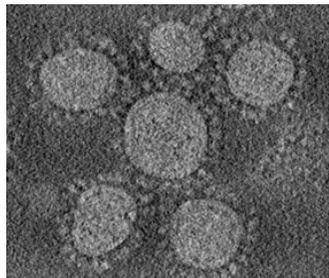
Les coronavirus sont une grande famille de virus, dont la plupart sont inoffensifs pour l'homme. Quatre types de coronavirus (OC43, 229E, NL63 et HKU1) sont endémiques et connus pour provoquer des rhumes; trois types de coronavirus peuvent causer de graves infections pulmonaires:



- Severe Acute Respiratory Syndrome-related Coronavirus 2
- **SARS-CoV-2**
- responsable de la pandémie de Coronavirus disease-19 (COVID-19)
- réservoir: chauve-souris (?)
- [Génome](#) (référence)



- Middle-East Respiratory Syndrome Coronavirus
- **MERS-CoV**
- responsable d'épidémies récurrentes depuis 2012
- réservoir: chameau, chauve-souris (?)
- [Génome](#) (référence)

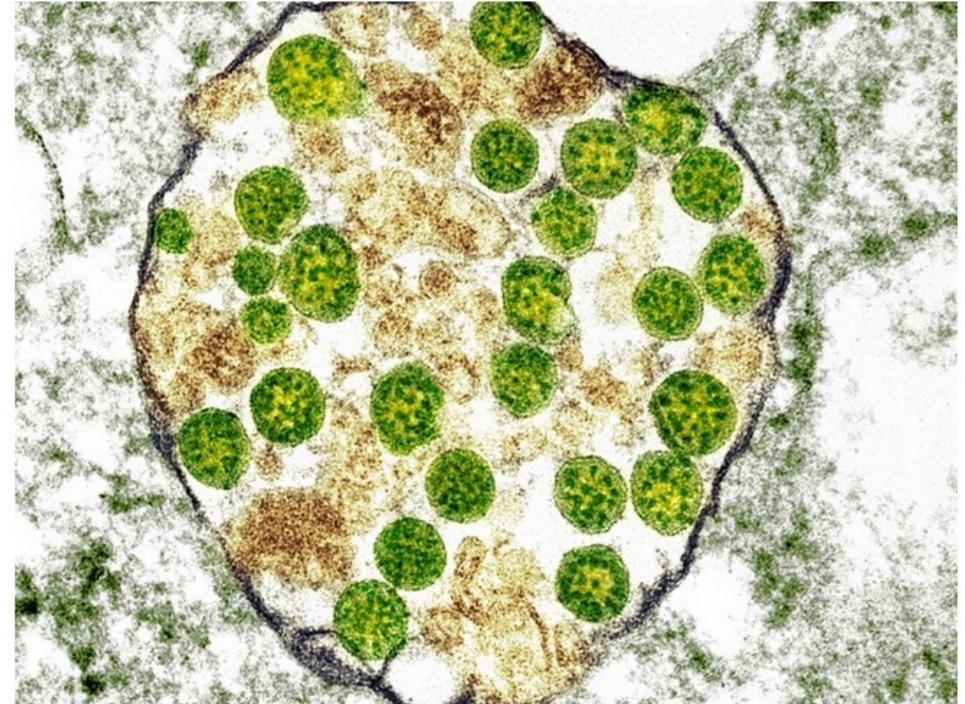
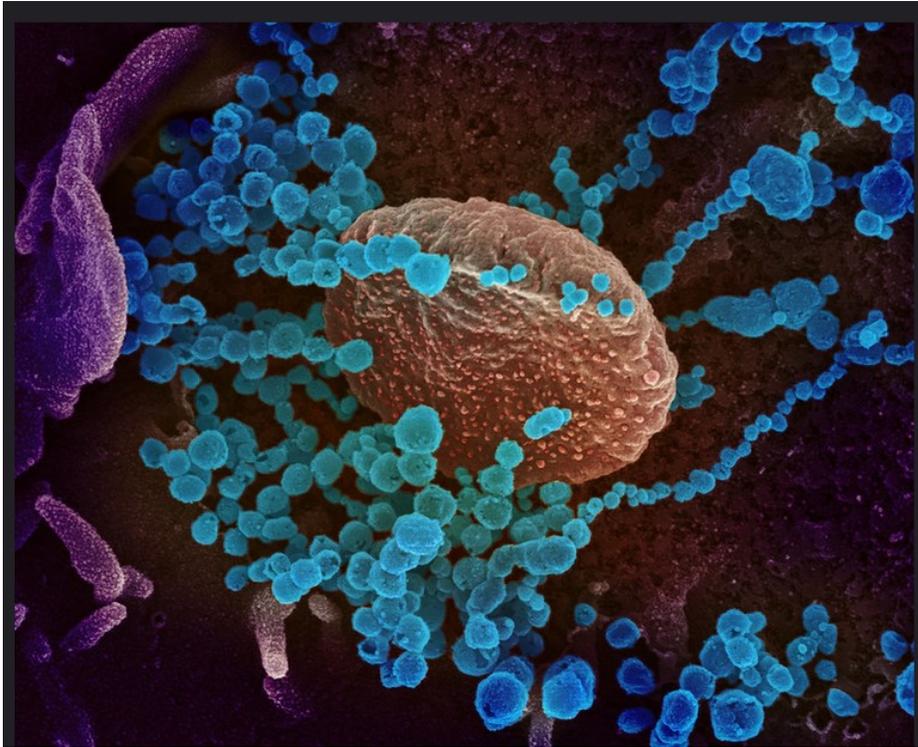


- Severe Acute Respiratory Syndrome-related Coronavirus
- **SARS-CoV**
- responsable d'une épidémie en 2003 qui a touché plus de 30 pays
- réservoir: chauve-souris; animal intermédiaire: civette
- [Génome](#) (référence)

Les estimations concernant l'apparition des premiers coronavirus varient considérablement, de 10 000 ans à 300 millions d'années.

[Plus d'info sur SARS-CoV-2](#)

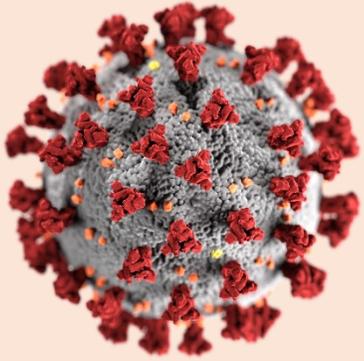
Des images de SARS-CoV-2 à l'extérieur et à l'intérieur d'une cellule....



A cell infected with particles of SARS-CoV-2. Credit: Cynthia S. Goldsmith and Azaibi Tamin/CDC/SPL

<https://www.flickr.com/photos/niaid/49557550751/>

<https://www.nature.com/articles/d41586-020-00502-w>



2 - Le génome de SARS-CoV-2

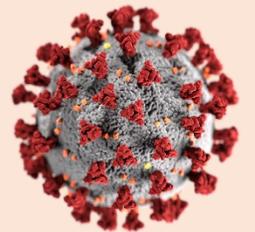
Bioinformatique

Banque de données de séquences ADN & ARN: [GenBank](#)

Outil de recherche de séquences dans les génomes: [BLAT@UCSC](#)

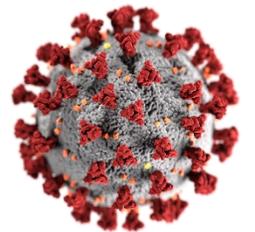
Biologie

[C'est quoi un génome?](#)



La séquence du génome de SARS-CoV-2

```
ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCCTCTAA  
CGAACTTTAAAAATCTGIGTGGCTGTCACCTCGGCTGCAIGCTTAGTGCACCTCACGCAGTATAATTAATAA  
TAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTTCGTCCGT  
TTGCAGCCGATCATCAGCACATCTAGGTTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGT  
CCTGGTTTTCAACGAGAAAACACACGTCCTCAACTCAGTTTGCCTGTTTTTACAGGTTTCGCGACGTGCTCGTA  
GTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTG  
CTTAGTAGAAGTTGAAAAAGGGCGTTTTTGCCTCAACTTGAACAGCCCTATGTGTTTCATCAAACGTTCCGGA
```



Le 10 janvier 2020, une équipe chinoise a soumis une première séquence du génome de SARS-CoV-2 à la banque de données GenBank.



Zhang Yongzhen a choisi de partager la séquence du génome de SARS-CoV2: cette séquence est devenue 'La séquence de référence'.

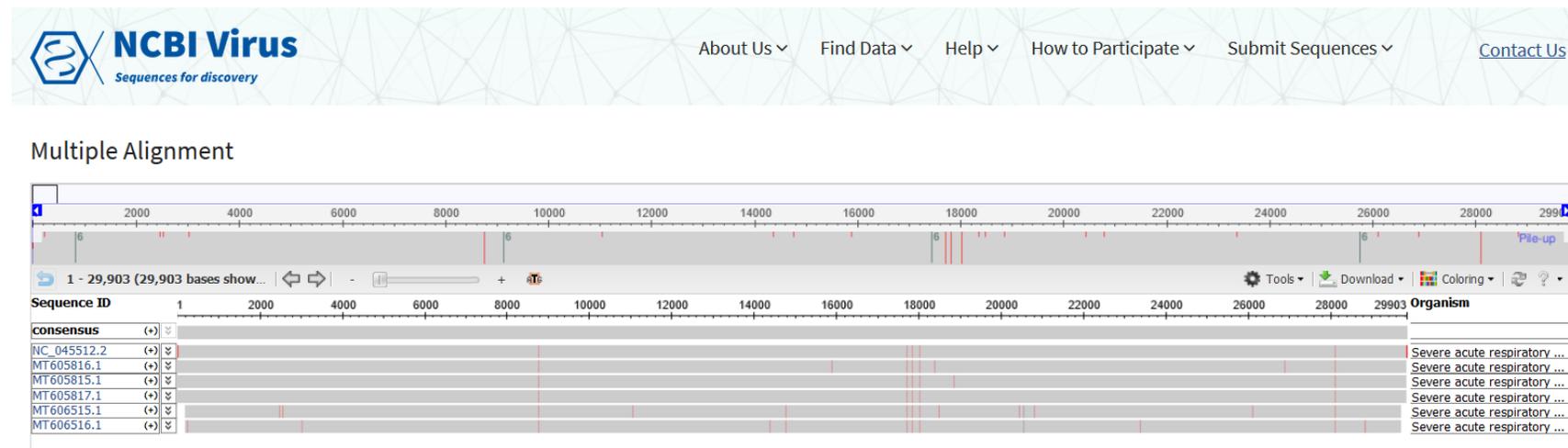
La séquence du génome de SARS-CoV-2 est composée de [29'903 nucléotides*](#).

** le génome de SARS-CoV-2 est un ARN simple brin. Dans les banques de données comme GenBank, on représente ce type de génome sous forme de séquence ADN (T à la place de U).*

Depuis ce premier génome, **plusieurs dizaines de milliers de génomes de SARS-CoV-2 ont été séquencés dans différents pays**. Plusieurs centres de recherche donnent librement accès à ces données. Et c'est essentiel !

Exemple (NCBI): [Severe acute respiratory syndrome coronavirus 2 data hub](#)

- Ce site permet de comparer des génomes: les différences par rapport à la séquence du génome de référence (consensus) apparaissent sous forme de barres verticales
- Exemple: le génome du virus 'chinois' et plusieurs génomes de virus 'américains'



Expert - Pour avoir accès aux séquences : copier/coller le **numéro d'accèsion/Sequence ID** dans l'URL:
<https://www.ncbi.nlm.nih.gov/nuccore/MT911538.1?report=fasta>

Voici les génomes de virus (provenant de pays différents, à plusieurs mois d'intervalle):

[Séquence de SARS-CoV-2 Chine](#) (Wuhan, 10 janvier 2020) ; séquence de 'référence'

[Séquence de SARS-CoV-2 USA](#) (Minnesota, 21 août 2020)

[Séquence de SARS-CoV-2 Egypt](#) (7 octobre 2020)

[Séquence de SARS-CoV-2 B 1.1.7 UK](#) (8 Janvier 2021)

Comparer ces séquences (copier/coller 2 séquences dans l'outil [align@UniProt](#))

Cela peut prendre plusieurs minutes....

Trouver les différences (= les mutations) et en particulier celle située en position 23'403:

ceci est une mutation



```
NC_045512.2  
MT911538.1
```

```
23401 GGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTG 23460  
23347 GGGTGTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTG 23406  
**_*.....*
```

Résultat alignement génome de référence / génome B.1.1.7 (UK): [fichier txt](#)



NC 045512.2	23281	TGATGCTGCCGTGATCCACAGACACTTGAGATTCTTGACATTACACCATGTTCTTTGG	23340
MT911538.1	23227	TGATGCTGCCGTGATCCACAGACACTTGAGATTCTTGACATTACACCATGTTCTTTGG	23286

NC 045512.2	23341	TGGTGTCCAGTGTATAACACCAGGAACAATACTTCTAACCAGGTTGCTGTTCTTTATCA	23400
MT911538.1	23287	TGGTGTCCAGTGTATAACACCAGGAACAATACTTCTAACCAGGTTGCTGTTCTTTATCA	23346

NC 045512.2	23401	GGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTG	23460
MT911538.1	23347	GGGTGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTG	23406
		** . *****	

ceci est une mutation



NC 045512.2
MT911538.1

23401	GGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTG	23460
23347	GGGTGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTG	23406
	** . *****	

NC 045512.2	23761	GACATCAGTAGATTGTACAATGTACATTTGTGGTGATTCAACTGAATGCAGCAATCTTTT	23820
MT911538.1	23707	GACATCAGTAGATTGTACAATGTACATTTGTGGTGATTCAACTGAATGCAGCAATCTTTT	23766

NC 045512.2	23821	GTTGCAATATGGCAGTTTTTGTACACAATTAACCCTGCTTTAACTGGAATAGCTGTTGA	23880
MT911538.1	23767	GTTGCAATATGGCAGTTTTTGTACACAATTAACCCTGCTTTAACTGGAATAGCTGTTGA	23826

NC 045512.2	23881	ACAAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAATTTACAAAAACACCACC	23940
MT911538.1	23827	ACAAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAATTTACAAAAACACCACC	23886

NC 045512.2	23941	AATTAAGATTTTGGTGGTTTTAATTTTTTCAAAAATATTACCAGATCCATCAAAACCAAG	24000
MT911538.1	23887	AATTAAGATTTTGGTGGTTTTAATTTTTTCAAAAATATTACCAGATCCATCAAAACCAAG	23946

NC 045512.2	24001	CAAGAGGTCATTTATTGAAGATCTACTTTTTCAACAAGTGCACCTGCAGATGCTGGCTT	24060
MT911538.1	23947	CAAGAGGTCATTTATTGAAGATCTACTTTTTCAACAAGTGCACCTGCAGATGCTGGCTT	24006

NC 045512.2	24061	CATCAACAATATGGTGATTGCCTTGGTGATATTGCTGCTAGAGACCTCATTGTGCACA	24120
MT911538.1	24007	CATCAACAATATGGTGATTGCCTTGGTGATATTGCTGCTAGAGACCTCATTGTGCACA	24066

NC 045512.2	24121	AAAGTTTAAACGGCCTTACTGTTTTGCCACCTTTGCTCACAGATGAAATGATTGCTCAATA	24180
MT911538.1	24067	AAAGTTTAAACGGCCTTACTGTTTTGCCACCTTTGCTCACAGATGAAATGATTGCTCAATA	24126

N = le nucléotide n'a pas pu être identifié

Le coronavirus mute relativement peu. En tout cas moins qu'une grippe, une gastro-entérite ou une hépatite. "Mais il mute suffisamment pour qu'on arrive à le reconnaître et à identifier ses 'ancêtres'", note Didier Trono.

La génomique à la rescousse pour étudier l'évolution du coronavirus

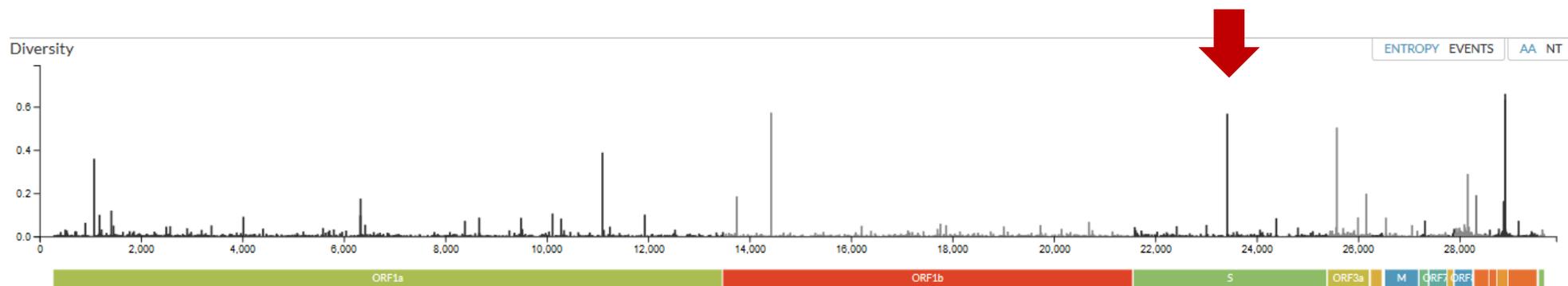
<https://www.rts.ch/info/sciences-tech/11559187-la-genomique-a-la-rescousse-pour-etudier-levolution-du-coronavirus.html>

*Les coronavirus sont parmi les rares virus à ARN dotés d'un mécanisme de 'correction génomique' - qui empêche le virus d'accumuler des mutations. Les médicaments qui 'tuent' les virus (comme l'hépatite C) en induisant des mutations, ne marchent donc pas contre SARS-CoV-2!
Nature **581**, 22-26 (2020)*

Les barres verticales représentent les mutations les plus fréquemment observées parmi les quelque 29'000 nucléotides (comparaison de milliers de génomes)

Été 2020

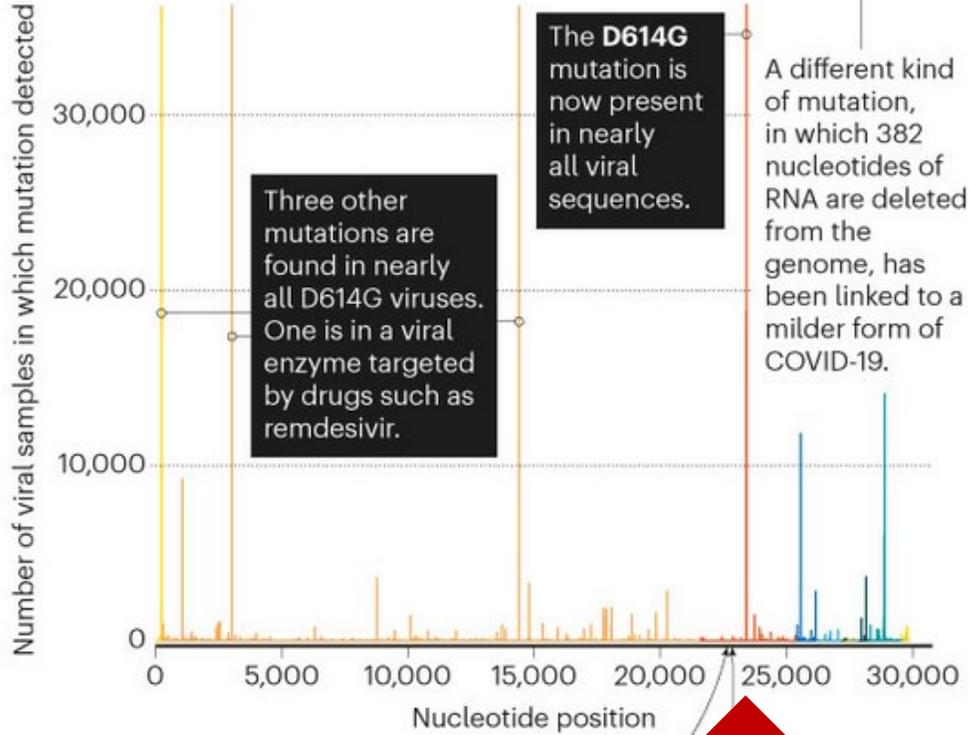
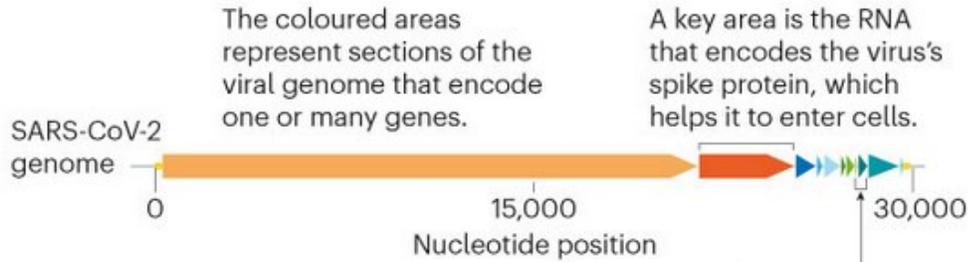
"Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear"



<https://nextstrain.org/ncov/global> -5139 genomes sampled between Dec 2019 and Sep 2020)

A CATALOGUE OF CORONAVIRUS MUTATIONS

Various mutations have been detected in SARS-CoV-2 genomes, including the most prevalent one, D614G. The virus's genetic code has just under 30,000 nucleotides of RNA, or letters, that spell out at least 29 genes. The most common mutations are single-nucleotide changes.

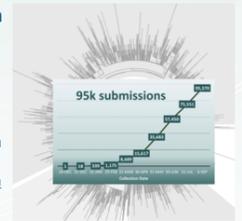


In Focus

95,000 viral genomic sequences of hCoV-19 shared with unprecedented speed via GISAID

Since the start of the COVID-19 outbreak and the identification of the pandemic virus, laboratories around the world are generating viral genome sequence data with unprecedented speed, enabling real-time progress in the understanding of the new disease and in the research and development of candidate medical countermeasures. Sequence data are essential to design and evaluate diagnostic tests, to track and trace the ongoing outbreak, and to identify potential intervention options. [Listen to PRI's Elana Gordon.](#)

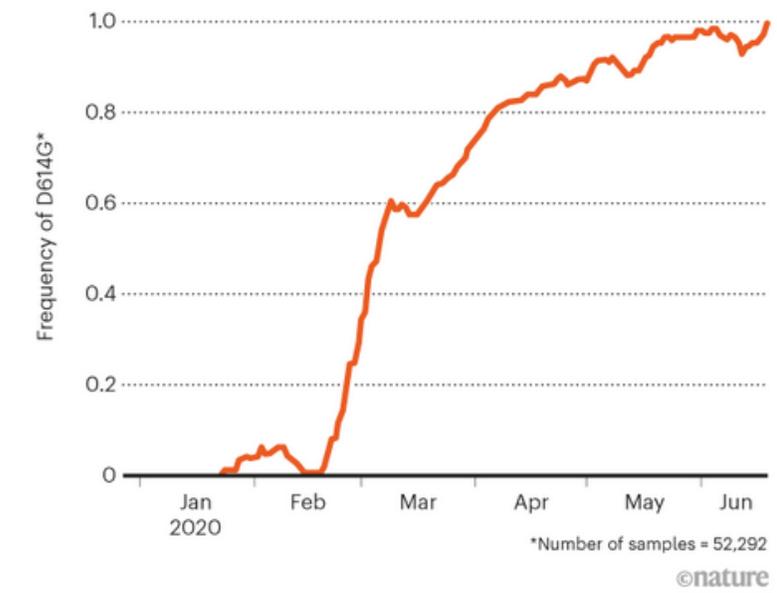
GISAID data Submitters and Curators ensure real-time data sharing of hCoV-19 remains reliable, to enable rapid progress in the understanding of the new COVID-19 disease and in the research and development of candidate medical countermeasures.



<https://www.gisaid.org/>

GLOBAL SPREAD

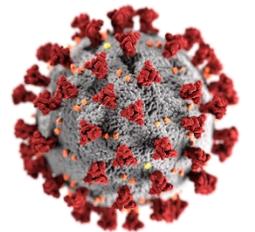
By the end of June, the D614G mutation was found in almost all SARS-CoV-2 samples worldwide.



The coronavirus is mutating: does it matter ?

La spécificité du test de détection PCR

Spécificité des amorces (primers)



Séquencer le génome de SARS-CoV-2 a permis de rapidement mettre en place un **test PCR** pour détecter la présence du virus dans des **frottis** naso-pharyngés (nez) ou oro-pharyngés (gorge).

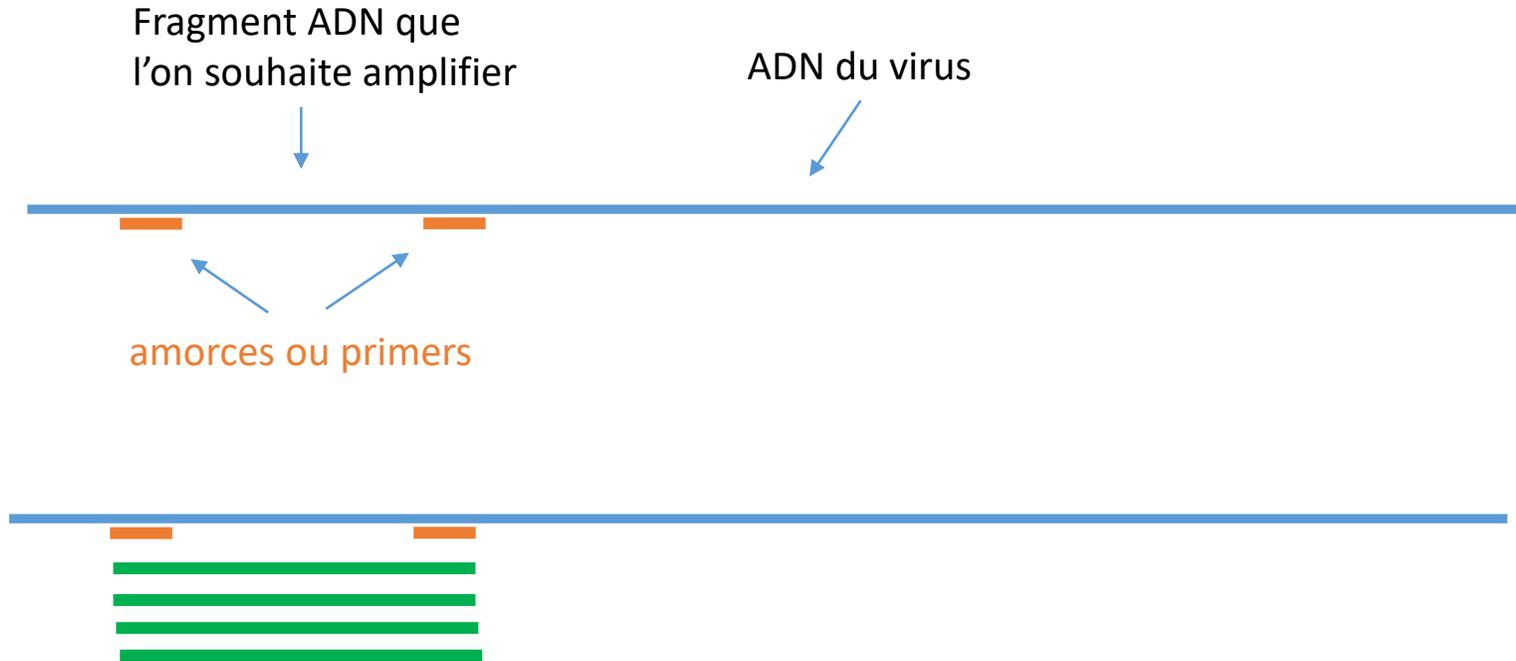
Le test peut être 'positif' à partir d'une centaine de virus présents dans le frottis!



Photo d'illustration. • © JOE RAEDLE / GETTY IMAGES NORTH AMERICA / Getty Images via AFP

La PCR est une technique de laboratoire qui permet d'amplifier de façon sélective un fragment d'ADN, c'est-à-dire d'en produire des milliers de copies.

Deux petites séquences d'ADN d'environ 15 nucléotides appelées **amorces ou primers**, sont choisies avec soin. Elles sont complémentaires (100% identique) au **brin d'ADN** et 'encadrent' le fragment à amplifier



Amplification du **fragment** grâce à plusieurs cycles de PCR

Si le virus est présent dans l'échantillon testé, les fragments d'ADN amplifiés seront visibles sur un gel par exemple.

Virus +	Virus -
	

Afin de valider le test, et selon les tests, plusieurs fragments (différentes régions dans le génome) sont amplifiés.

[Un peu de math...](#)
[PCR statistiques: VPN, VPP](#)

Primers (amorces) utilisés pour le test PCR

Voici des exemples de primers utilisés pour les tests PCR:

- Primer no 1 CTCGAACTGCACCTCATGG
- Primer no 2 GGCATACACTCGCTATGTC

! Petit casse-tête (géré par les [programmes bioinformatiques](#) qui permettent de sélectionner les bons primers ...)

Biologie (brins complémentaires)

```
CTCGAACTGCACCTCATGG (primer)
|||||
GAGCTTGACGTGGAGTACC (ADN/ARN)
```

Bioinformatique (alignement)

```
CTCGAACTGCACCTCATGG (primer)
|||||
CTCGAACTGCACCTCATGG (ADN/ARN)
```

Cette activité est basée sur le **point de vue bioinformatique...**

- 1a.** Combien de fois retrouvez-vous la séquence d'un primer dans le génome du virus (manuellement)?
- Chercher la séquence d'un primer dans le [génom du virus](#) (Utiliser Ctrl F ou commandF sur mac)

Région que l'on souhaite amplifier

ADN du virus

amorces

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[GenBank](#) [Graphics](#)

>NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

```
ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAA
CGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACCTCACGCAGTATAATTAATAAC
TAATTACTGTCGTTGACAGGACACAGAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCGGTG
TTGCAGCCGATCATCAGCACATCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTGTC
CCTGGTTTCAACGAGAAAACACACGTTCCAACTCAGTTTGCCCTGTTTACAGGTTCCGCGACGTGCTCGTAC
GTGGCTTTGGAGACTCCGTGGAGGAGGCTTATCAGAGGCACGTCAACATCTTAAAGATGGCAGCTTGTGG
CTTAGTAGAAGTTGAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTTTCATCAAACGTTTCGGAT
CTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAAGCTCGAAGGCATTCAGTACGGTC
GTAGTGGTGAGACACTTGGTGTCTTGTCCCTCATGTGGCGAAATACCAGTGGCTTACCGCAAGGTTCT
TCCTCGTAAGAACCGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTGACTTA
GGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAAGCTGGAACACTAAACATAGCAGTGGTG
TTACCCGTGAACCTCATGCGTGAAGCTTAAAGAGGGGATACACTCGCTATGTCGATAACAACCTTCTGTGG
CCCTGATGGCTACCTCTTGTAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTTCATGCACITTTG
TCCGAACAACCTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTTG
CTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATGACAGACCTTTTGAATTAATTTGGCAAGAA
ATTTGACACCTTCAATGGGAAATGTCAAAATTTTGTATTTCCCTTAAATTCATAATCAAGACTATTCAA
CCAAGGGTTGAAAAGAAAAGCTTGTATGGCTTTATGGGTAGAATTCGATCTGCTATCCAGTTGCGTCAC
CAAATGAATGCAACCAATGTGCCCTTCAACTCTCATGAAGTGTGATCATTGTGGTGAACCTTCATGGCA
```

➤ Primer no 1 CTCGAACTGCACCTCATGG

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[GenBank](#) [Graphics](#)

>NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

```
ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAA
CGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACCTCACGCAGTATAATTAATAAC
TAATTACTGTCGTTGACAGGACACAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCGGTG
TTGCAGCCGATCATCAGCACATCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTGTC
CCTGGTTTCAACGAGAAAACACACGTTCCAACTCAGTTTGCCCTGTTTACAGGTTCCGCGACGTGCTCGTAC
GTGGCTTTGGAGACTCCGTGGAGGAGGCTTATCAGAGGCACGTCAACATCTTAAAGATGGCAGCTTGTGG
CTTAGTAGAAGTTGAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTTTCATCAAACGTTTCGGAT
GCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAAGCTCGAAGGCATTCAGTACGGTC
GTAGTGGTGAGACACTTGGTGTCTTGTCCCTCATGTGGCGAAATACCAGTGGCTTACCGCAAGGTTCT
TCCTCGTAAGAACCGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTGACTTA
GGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAAGCTGGAACACTAAACATAGCAGTGGTG
TTACCCGTGAACCTCATGCGTGAAGCTTAAAGAGGGGATACACTCGCTATGTCGATAACAACCTTCTGTGG
CCCTGATGGCTACCTCTTGTAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTTCATGCACITTTG
CCTGATGGCTACCTCTTGTAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTTCATGCACITTTG
TCCGAACAACCTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTTG
CTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATGACAGACCTTTTGAATTAATTTGGCAAGAA
ATTTGACACCTTCAATGGGAAATGTCAAAATTTTGTATTTCCCTTAAATTCATAATCAAGACTATTCAA
CCAAGGGTTGAAAAGAAAAGCTTGTATGGCTTTATGGGTAGAATTCGATCTGCTATCCAGTTGCGTCAC
CAAATGAATGCAACCAATGTGCCCTTCAACTCTCATGAAGTGTGATCATTGTGGTGAACCTTCATGGCA
```

➤ Primer no 2 GGCATACTCGCTATGTC

Primers (amorces) utilisés pour le test PCR

Voici des exemples de primers utilisés pour les tests PCR:

- Primer no 1: CTCGAACTGCACCTCATGG
- Primer no 2: GGCATACACTCGCTATGTC

1b. Combien de fois retrouvez-vous la séquence d'un primer dans le génome du virus (avec un outil bioinformatique) ?

- Utiliser [BLAT@UCSC](#)
- Copier/coller la séquence d'un primer
- Choisir 'genome' SARS-CoV-2
- Cliquer 'Submit'

2. Combien de fois retrouvez-vous la séquence d'un primer dans le génome humain? Le génome humain est composé de quelque 3 milliards de nucléotides...

- Utiliser [BLAT@UCSC](#)
- Choisir 'genome' human
- Copier/coller la séquence d'un primer
- Cliquer 'Submit'

1b. Combien de fois retrouvez-vous la séquence d'un primer dans le génome du virus (avec un outil bioinformatique [BLAT@UCSC](#)) ?

BLAT Search Genome

Genome: Search all
SARS-CoV-2

Assembly: Dec. 2013 (GRCh38/hg38)

Query type: BLAT's guess

Sort output: query,score

Output type: hyperlink

CTCGAACTGCACCTCATGG

All Results (no minimum matches)

Submit I'm feeling lucky Clear

BLAT Search Results

Go back to [NC_045512v2:492-510](#) on the Genome Browser.

Custom track name: blat YourSeq

Custom track description: blat on YourSeq

Build a custom track with these results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	YourSeq	19	1	19	19	100.0%	NC_045512v2	+	492	510	19

Le primer est retrouvé une seule fois dans le génome du virus (100 % identique)

2. Combien de fois retrouvez-vous la séquence d'un primer dans le génome humain ([BLAT@UCSC](#)) ? Le génome humain est composé de quelque 3 milliards de nucléotides...

BLAT Search Genome

Genome: Search all

Human

Assembly:

Dec. 2013 (GRCh38/hg38)

Query type:

BLAT's guess

Sort output:

query,score

Output type:

hyperlink

CTCGAACTGCACCTCATGG

All Results (no minimum matches)

Submit

I'm feeling lucky

Clear



Genomes

Genome Browser

Tools

Mirrors

Downloads

My Data

Projects

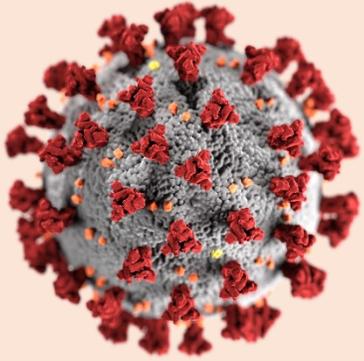
Help

About Us

Human (hg38) BLAT Results

Sorry, no matches found (with score at least 20)

Le primer n'est pas retrouvé dans le génome humain (3 milliards de lettres !)



3 - Les protéines de SARS-CoV-2

Bioinformatique

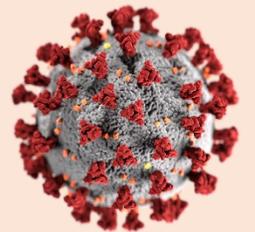
Banque de données de protéines: [UniProtKB/Swiss-Prot](#)

Outil de traduction ADN -> protéine: [translate @ ExpASY](#)

Banque de données de structures 3D: [PDB](#)

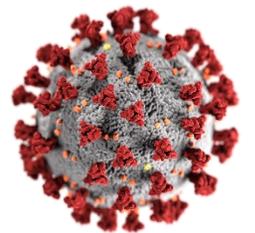
Biologie

[C'est quoi une protéine?](#)



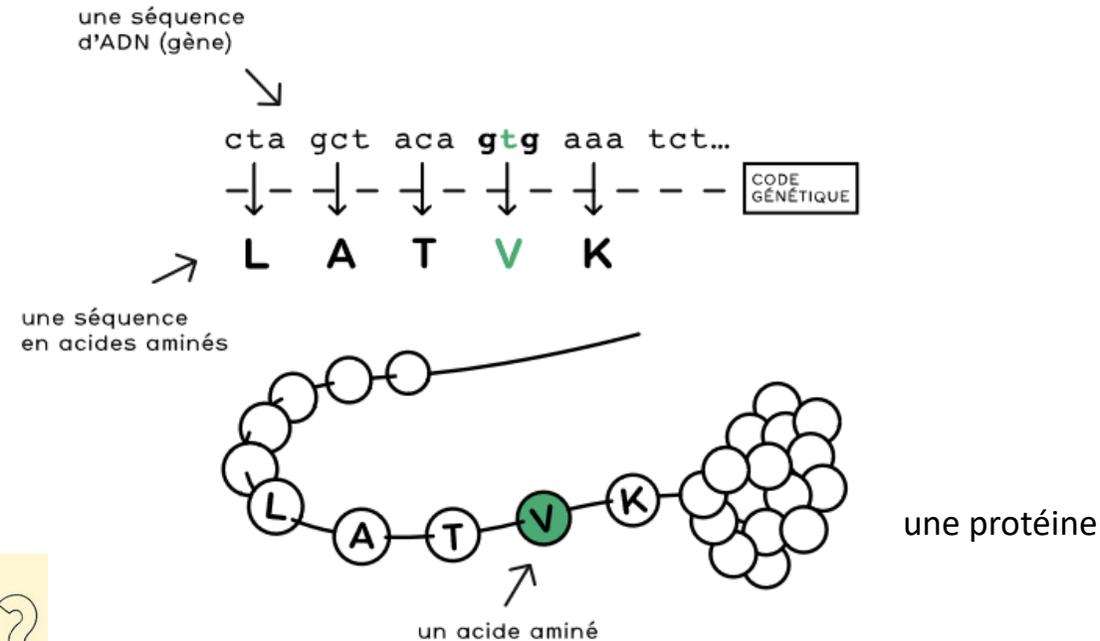
Du génome aux protéines

[Bad News Wrapped in Protein: Inside the Coronavirus Genome](#) (New-York Times)



L'analyse de la séquence du génome du virus a permis de trouver les séquences en acides aminés des protéines du virus. Environ 29 gènes codant pour des protéines ont été identifiés. **Nous allons nous intéresser au gène codant pour la protéine Spike.**

DES NUCLÉOTIDES AUX ACIDES AMINÉS...



UNE PROTÉINE?
présentation de quelques protagonistes importants de cette b.d.

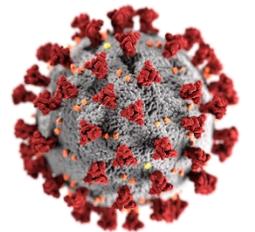
[Une protéine?](#)

[C'est quoi une protéine?](http://www.precisionmed.ch)
www.precisionmed.ch

La protéine Spike de SARS-CoV-2

*Spike ou spicule en français

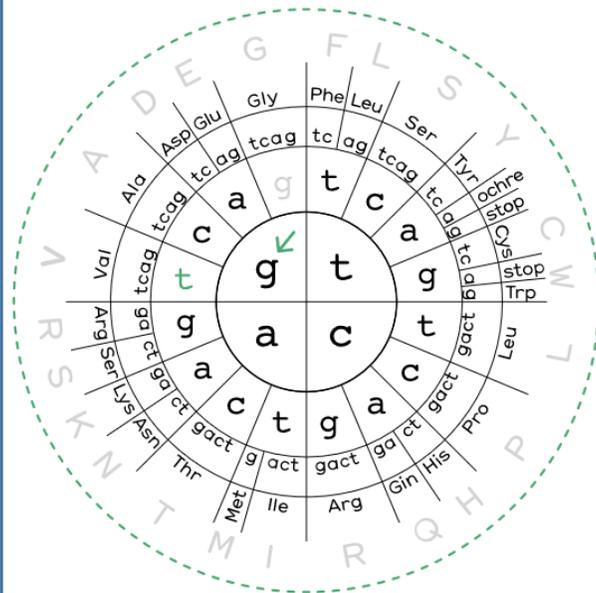
«Une seule protéine située à la surface du nouveau coronavirus est responsable de sa transmission chez l'humain et de la pandémie qui en a découlé: la protéine Spike. Poussés par les développements technologiques majeurs de ces dernières années, les scientifiques ont rapidement déterminé sa composition et sa structure 3D, ce qui a grandement aidé le développement des vaccins.» [Radio Canada](#)



Du génome aux protéines

Voici le début de la séquence du gène codant pour la protéine Spike:

```
atgtttgt ttttcttggt ttattgccac tagtctctag tcagtggtt
aatcttaca ccagaactca attaccccct gcatacacta attctttcac
acgtgggtgt tattaccctg acaaagtttt cagatcctca gttttacatt
caactcagga cttggttctta cttttctttt ccaatggtac ttggttccat
gctatacatg tctctgggac caatgggtact aagagggttg ataaccctgt
cctaccattt aatgatgggtg tttattttgc ttccactgag aagtctaaca
taataagagg ctggatTTTT ggtactactt tagattcgaa gaccaggtcc
ctacttattg ttaataacgc tactaatggtt ...
```



Quels sont les premiers acides aminés de la protéine Spike?

atg ttt gtt ttt ctt gtt tta ttg ...
M F V

```
>Gene_Spike
atggttfgttttttctgtttttattgcoactagtctctagtcagtggttaattttacaacc
agaactcaattaccocotgcatacactaatctttccacgtggtgtttattaccotgac
aaagttttcagatccctcagttttacattcaactcaggacttggtttaccctttctttcc
aatgttacttgggtccatgctatacatgtctctgggccaatgggactaagaggtttgat
aacctgtctccacttaataatgaggtgtttttttgctccactgagaagctaaacata
ataagaggtggatttttggtaactcttagatctgaagaccagctccctacttatgtt
aataacgctactaatgtgttttaaaagctgtgaaattcaatttgaatgatccattt
ttgggtgtttattacacaaaaaacaaaaagttgggtgaaagtgagttcagagtttat
tctagtgcgaataatggcaacttttgaatattgtctctcagcctttcttatggacctgaa
gaaaaacagggtaatttcaaaaacttagggaatttgggtttaaagaatttgaggttat
tttaaaatatttcaagcacacgcttataatttagtgctgatctccctcagggtttt
tcggcttttagaacacttggtagattggcaatagggtataacatcaactaggttcaaac
tacttgctttacatagaagttatttgactcctgggtgattctctctcaggttggacgct
ggtgctgcagcttatttagtgggttatcttcaacctaggactttctataaaataaat
gaaaaatggaaocattacagatgctgtagactgtgcaactgcaacctctcagaaacaaag
tgtacgttgaatctctcactgtgaaaaaggaatctatacaacttctaacttagagct
caaccaacagaactattgttagatttcttaattacaacttggcccttttgggtgaa
gttttaacgccacagatttgactctgtttatgcttggaaacaggaagagaatcagcaac
tgtgttgcgtatttctgtcctataataatccgcacttcttccacttttaagtggtat
ggagtgctcctactaaataaataatgactctgttactaatgtctatgagatctcttt
gtaattagaggtgatgaagtcagcaaaatcgctccagggaacactggaagatgtgat
tataattataaattaccagatgatttacaggctgcttatagcttggaaattcaaacat
cttgattcagaagttgggtgtaattataat taccgtgataagattttaggaagctaat
ctcaaacctttgagagagatatttcaactgaaatctatcaggccggtagcaacctgtt
aatgggttggaggttttaattgttacttcttcaactcaatggtttccaacccaact
aatgggttgggttacaacacacagagtagtagtacttctttgaaactctcaatgca
cagcaactgttgggactaaaaagctactaatttgggttaaaacaaaatgtgctcaat
ttcaactcaatgggtttaaaggcagcaggtgttcttactgagcttaacaaaaagtttctg
cctttccaacaaatttggcagagacattgctgacactactgatgctgctcagctcacaag
caacttgagattcttgacatcacacatgttcttttgggtgctcaggtttataacacca
ggaacaaaacttcaaccaggttctgttctttatcagatgttaactgcacagaaagct
cctgttctcattcagatcaacttactcctacttggcgtgtttattctacaggttct
aatgttttcaaacagctgcaaggtgttttaattaggggtgaaactgtcaacaaactcat
gagtgtagacatacccaatttggcaggtatattgctgtagttatcagaactcagcaactct
cctcggcgggcaactgtagctagctcaactcaactcactgctcaactgctcactgtgt
gcagaaaatcagttgcttactctaaactctattgccaatacccaaaaatttactatt
agtgttaccacagaaattctaccaggtcttatgccaagacatcagtagattgtacaatg
taacatttgggtgatttcaactgaaatgcagcaactcttttggcaatggcaggtttttgt
acacaaataaaccgtgcttcaactggaatagctgttgaacaagcaaaaaaccccagaa
gtttttgcacaagtcacaacaaatttcaaaaaacccaacttaaaagatttgggtgtttt
aatttttcacaatattaccagatccatcaaaacaaagcaagaggtcattttagaagt
ctactttcaacaaagtgcacttgcagatgctggcttcaacaaatattggtgattgc
ctgggtgataattgctgctagacactcaatttggcaaaaaagtttaacggccttactggt
ttgcaaccttgcctcagatgaaatgattgctcaatacactctgcaactgttagcgggt
acaactcctctgttgggacttgggcaaggtgctgcaatacaaaactattgctatg
caaatggcttataggtttaaaggtatggagttacaacagaaatgtctctatgagaaccaa
aaatgattgccaacaaatttaaatggtctattggcaaaatcaagactcacttctcc
acagcaagtgcaacttggaaaactcaagatgtggtcaacaaaaatgcacaagctttaaac
acgcttgttaaaccaacttagctccaatttgggtgcaatttcaagttttaaagataatc
tttcaagcttggcaaaaagttgaggtgaaagtgcaaatgtaggttgcacagggcaga
cttcaaaagttgcagacatattgactcaacaaatatttagaggtgcagaaatcagagct
ctgctaatcttgcctactaaaatgtcagaggtggtacttggcaaatcaaaaagagtt
gatttttgggaaagggctacatcttatgtcctccctcagtcagcaactcaggtgta
gtctcttgcactgactattgtcctcgcacaaagaaagaaacttcaacactgctcctgccc
atttgcactgagggaaaagcaacttctcctggaaggtgtctttgttcaaatggcaaca
caactggtttgtaacacaaaggaatttttgaaccacaaatcattactacagacaacaca
tttgtgctggtaactgtgatttgaataggaattgtcaacaacacagtttatgatcct
ttgcaacctgaaattagactcaatcaaggaggagttagataaatattttaaagaatcaca
tcacagatggtgatttaggtgacatctctggcatttaattgctcaggttgaacatcoaa
aaagaaattgaccgctcaatgaggttgcacaaagaaatttaaagaaatctctcagatctc
caagaaactggaaagttatgagcagtaataaaatggccatggtacattggctaggttt
atagctggctgattgccaatgtaattgggtgcaaatattgcttgcctgtagcagcaggtg
ttaggtgtctcaaggcgtgtgttcttgggtgaaactcctgcaaaattgtagaagcagca
ctgagccaggtgctcaaggagtcacaaattacattacacata
```

(1) Voici la [séquence complète du gène](#) codant pour la protéine Spike de SARS-CoV-2.

Utiliser l'outil [Translate@Expasy](#) pour traduire cette séquence de nucléotides en séquence d'acides aminés

La séquence en acides aminés de la protéine Spike se trouve dans le 'Frame 1'



Cette séquence, sous forme d'ARNm, est présente dans les vaccins mRNA (Pfizer et Moderna)

(2) Expert: 'Voir' le gène codant pour la protéine Spike dans le génome de SARS-CoV-2:

- Depuis: https://www.ncbi.nlm.nih.gov/nucleotide/NC_045512.2
- Chercher 'Spike' dans la page (Utiliser Ctrl F ou commandF sur mac)
- Cliquer sur 'gene'



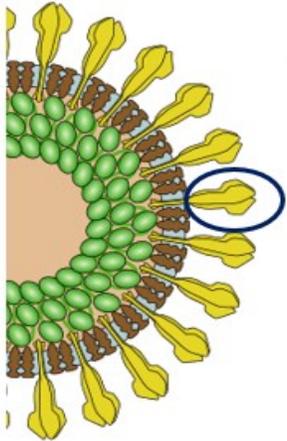
gene

```
21563..25384
/gene="S"
/locus_tag="GU280_gp02"
/gene_synonym="spike glycoprotein"
/db_xref="GeneID:43740568"
```

```
21361 aatc0aaatt agttgcttct ctattcttta ttggacatga gtaaaatttc ctttaaata
21421 aggggtactg ctgttatgct tttaaaagaa ggtcaaatca atgatatgat tttatctctt
21481 cttagtaaag gttagacttat aattagagaa acaacacagag ttgttatctc tagttaggtt
21541 ctgtttaa caactaaacgaa caatggtttgt tttcttggtt ttattggcaac tagtcttag
21601 tcagtggttt aatcttcaaa ccagaactca attaccctct gcatacacta attctttcac
21661 acgtggtgtt tattaccctg acaaaagttt cagatcctca gttttacatt caactcagga
21721 ctgttcttta cctttctttt ccaatggtac ttggttccat gctatcacatg tctctgggac
21781 caatggta ct aagaggtttg ataaccctgt cctaccattt aatgatgggt tttatttgg
21841 ttccactgag aagctcaaca taataagagg ctggattttt ggtactactt tagattcga
21901 gaccagctcc ctactattgt ttaataacgc tactaagttt gttattaaag tctgtgaatt
21961 tcaattttgt aatgatccat ttttgggtgt ttattaccac aaaaacaaca aaagttgga
```



La protéine Spike



Découvrez la [séquence en acides aminés](#) de la protéine Spike de SARS-CoV-2 dans la banque de données UniProtKB/Swiss-Prot.

Quelle est la longueur (nombre d'acides aminés) de la protéine Spike?

Comparer la séquence de 'référence' avec la séquence de Spike B 1.1.7 UK ([séquences \(pdf\)](#))

Voici une représentation de la **structure 3D de la protéine Spike** et de sa séquence en acides aminés (bande de données PDB)

Depuis:

<https://www.rcsb.org/3d-view/6vxx>

Cliquer sur un acide aminé pour voir où il est localisé dans la structure 3D.

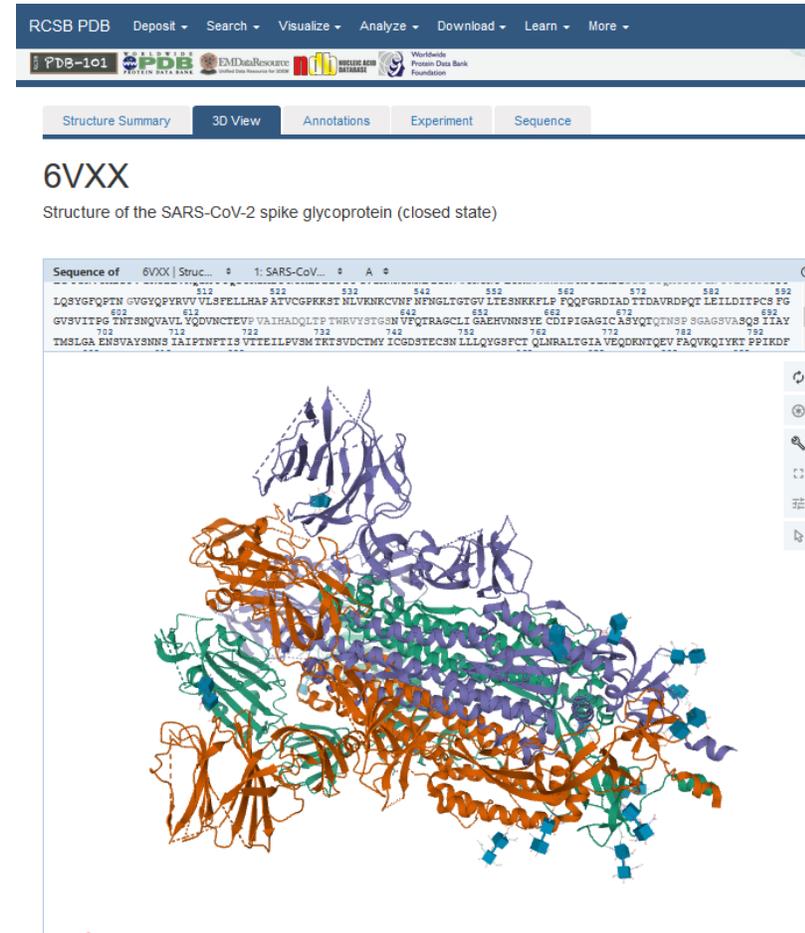
➔ Localiser en particulier l'acide aminé en position 614, un variant très présent pendant l'été 2020

"Making sense of mutation: what **D614G** means for the COVID-19 pandemic remains unclear"

Expert: comparer les 2 formes de la protéine Spike

- 6vxx: Spike forme fermée
- 6vsb: Spike forme ouverte

➔ Localiser un variant 'anglais': N501Y
➔ Localiser les variants 'sud africains': K417N, E484K and N501Y



RCSB PDB Deposit Search Visualize Analyze Download Learn More

PDB-101 PDB EMDataResource

Structure Summary 3D View Annotations Experiment Sequence

6VXX
Structure of the SARS-CoV-2 spike glycoprotein (closed state)

Sequence of 6VXX | Struc... 1: SARS-CoV... A

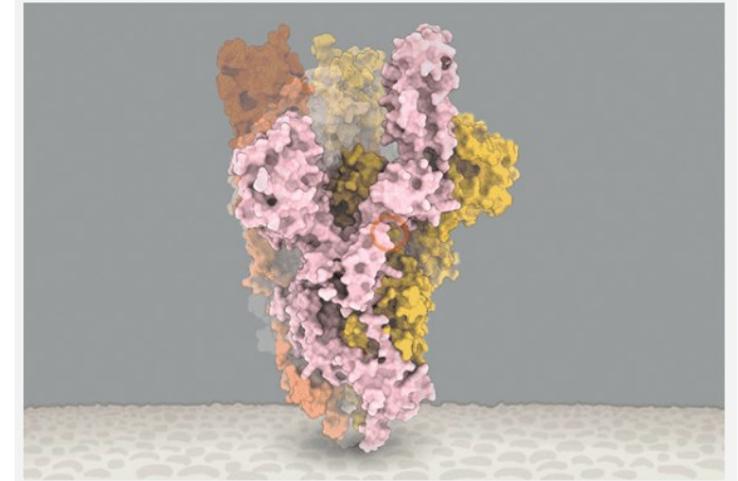
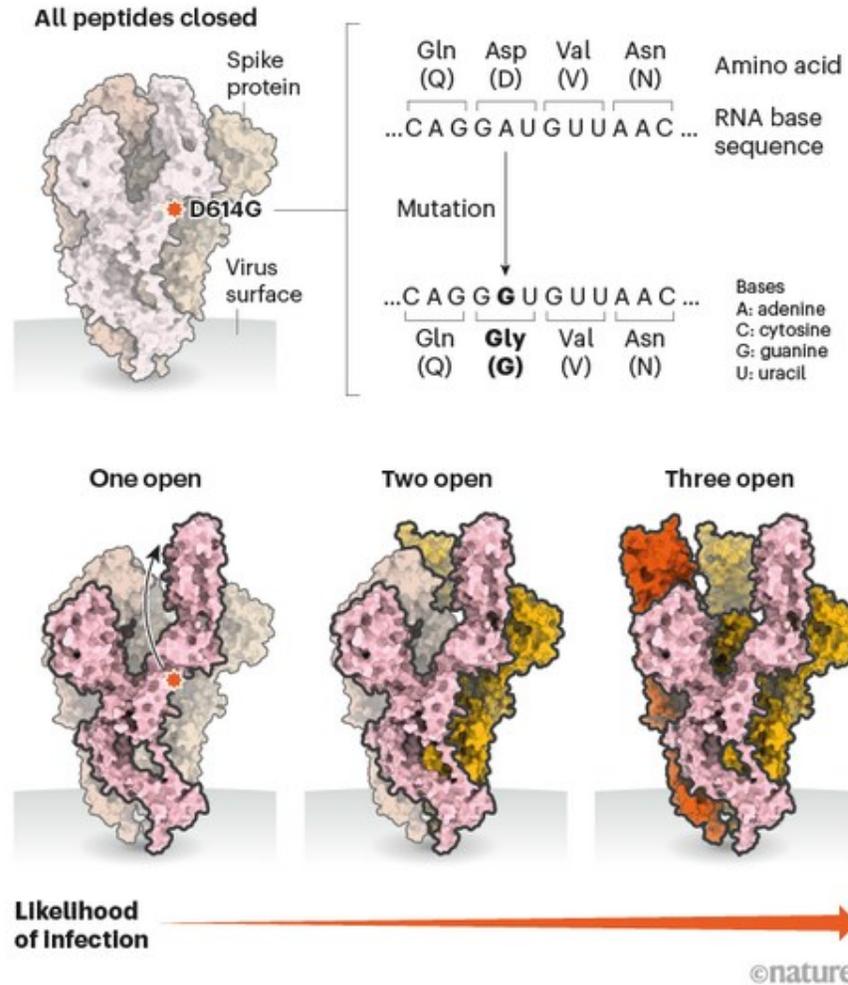
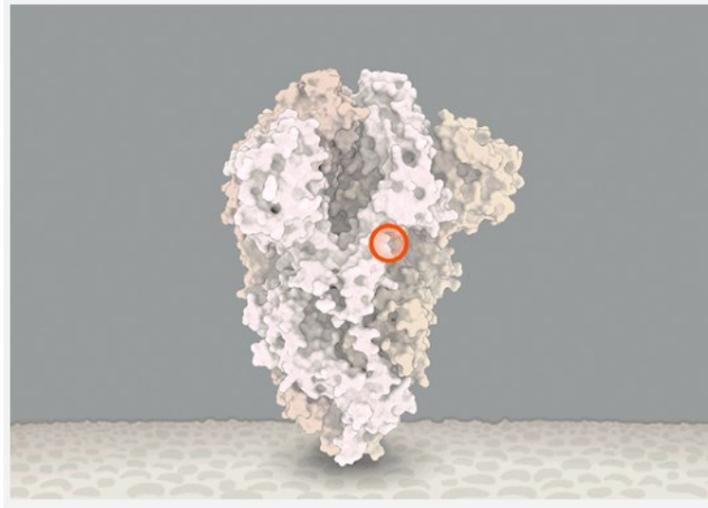
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

Note: les cubes bleus représentent des sucres

[Nature, août 2020](#)

THE MUTATION THAT LOOSENS THE SPIKE PROTEIN

Spike proteins on SARS-CoV-2 bind to receptors on human cells, helping the virus to enter. A spike protein is made up of three smaller peptides in 'open' or 'closed' orientations; when more are open, it's easier for the protein to bind. The D614G mutation — the result of a single-letter change to the viral RNA code — seems to relax connections between peptides. This makes open conformations more likely and might increase the chance of infection.

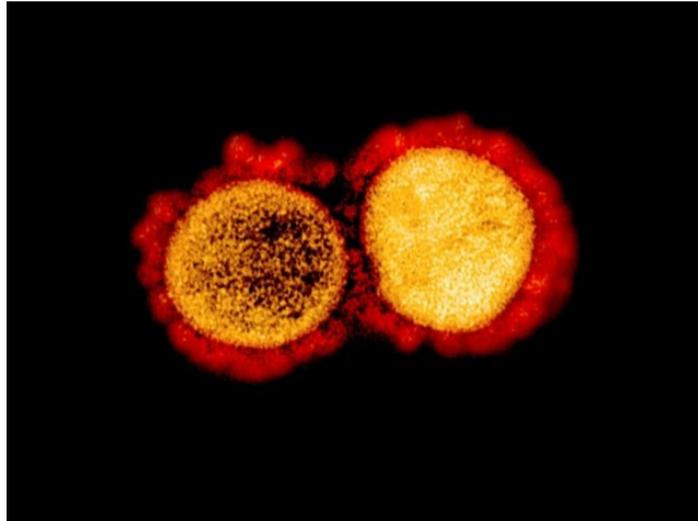


'Closed' and 'open' conformations of the spike protein on SARS-CoV-2, which binds to receptors on human cells. A common mutation (circled) seems to make the protein favour open conformations, which might mean the virus can enter cells more easily. (Source: Structural data from K. Shen & J. Luban)

[The coronavirus is mutating: does it matter ?](#)

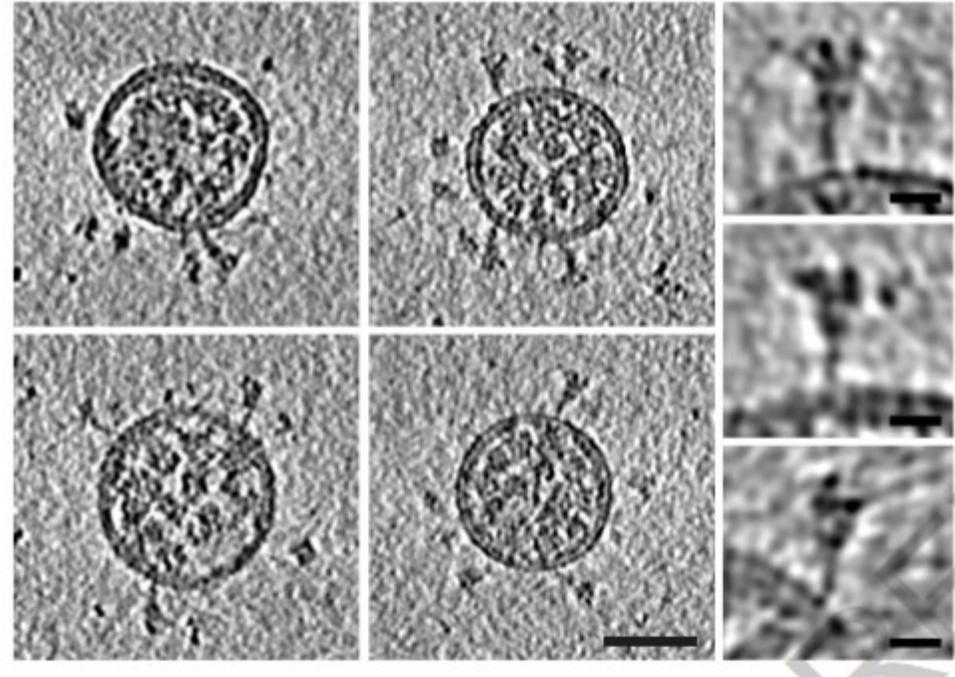
doi: 10.1038/d41586-020-02544-6

Images cryo-EM de la protéine Spike à la surface du virion



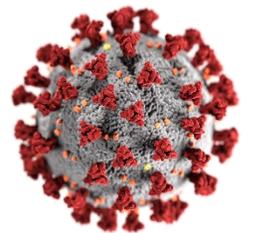
Une image du virus SARS-CoV-2 transmise par les Instituts nationaux de santé (NIH/NIAD) le 11 août 2020.

NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES/AFP/ARCHIVES - HANDOUT

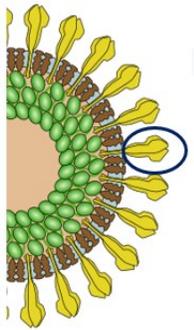


https://www.nature.com/articles/s41586-020-2665-2_reference.pdf

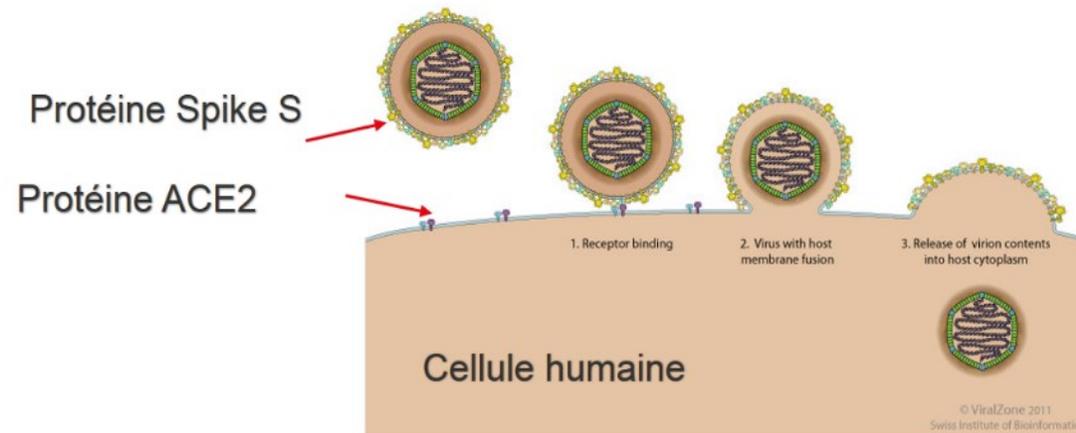
Comment le virus pénètre dans les cellules humaines ?



Les protéines Spike et ACE2



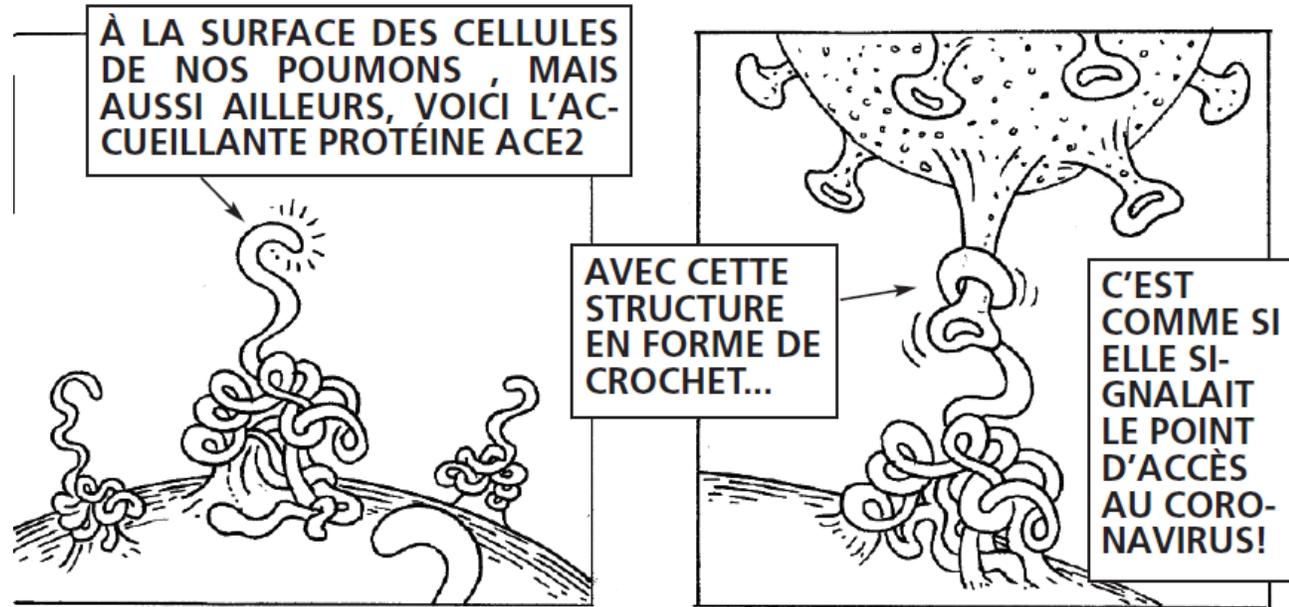
La protéine **Spike** permet au virus de pénétrer dans nos cellules, en interagissant entre autre avec une protéine humaine appelée **ACE2**, présente à la surface des cellules humaines.



<https://viralzone.expasy.org/9096>

Note: d'autres protéines humaines sont impliquées dans ce processus: elles coupent Spike et induisent ainsi la production d'un 'peptide de fusion' (TMPRSS2, Furin)

angiotensin-converting enzyme 2



La toux d'un voisin qui envoie 10 particules virales pourrait suffire à déclencher une simple infection de la gorge. Mais si le voisin est plus proche et qu'il envoie 100 particules virales, le virus pourrait descendre jusqu'aux poumons et être plus difficile à éliminer... Le virus peut se 'reproduire' dans la gorge (et être ainsi présent dans la salive) avant même que les symptômes ne se manifestent...
Nature **581**, 22-26 (2020)

[Protein Spotlight comics ACE2](#): Entrez donc !

[Vidéo 3](#) (4:54)

Comment SARS-CoV-2 entre dans les cellules humaines ?

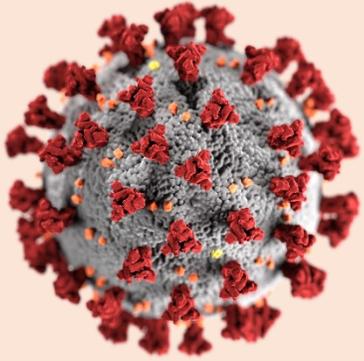
Pourquoi des symptômes aussi différents, comme par exemple la perte d'odorat ?

Pourquoi une infection avec SARS-CoV ou SARS-CoV-2 ne conduit pas aux mêmes symptômes ?

Pourquoi SARS-CoV n'a pas été à l'origine d'une pandémie en 2003 ?



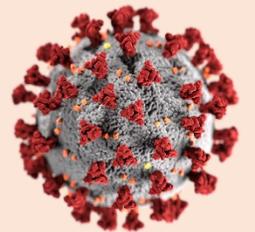
The image shows a YouTube video player interface. At the top, there is a search bar with the text "Search" and a magnifying glass icon. Below the search bar, the video player displays a slide with a white background and a black border. The slide features a 3D model of a coronavirus particle in the top left and bottom right corners. The main text on the slide reads "Comment le coronavirus SARS-CoV-2 entre dans les cellules humaines ?" with a play button icon over the word "cellules". In the bottom left corner of the slide, the presenter's name "Phillippe Le Mercier" is listed, along with his credentials: "Biologiste, Doctorat en virologie - Institut Pasteur", "ViralZone, Groupe Swiss-Prot, Genève", and "SIB Swiss Institute of Bioinformatics". In the top right corner of the slide, there are logos for "SIB Swiss Institute of Bioinformatics" and "Unil Université de Lausanne Service Culture et Médiation scientifique". At the bottom of the video player, there is a progress bar showing "0:00 / 4:54" and various control icons like play, volume, and full screen.



4 - A la recherche de l'origine de SARS-CoV-2

Bioinformatique

Banque de données de taxonomie: [Taxonomy@NCBI](https://www.ncbi.nlm.nih.gov/taxonomy)



Les coronavirus infectent de nombreuses espèces de mammifères



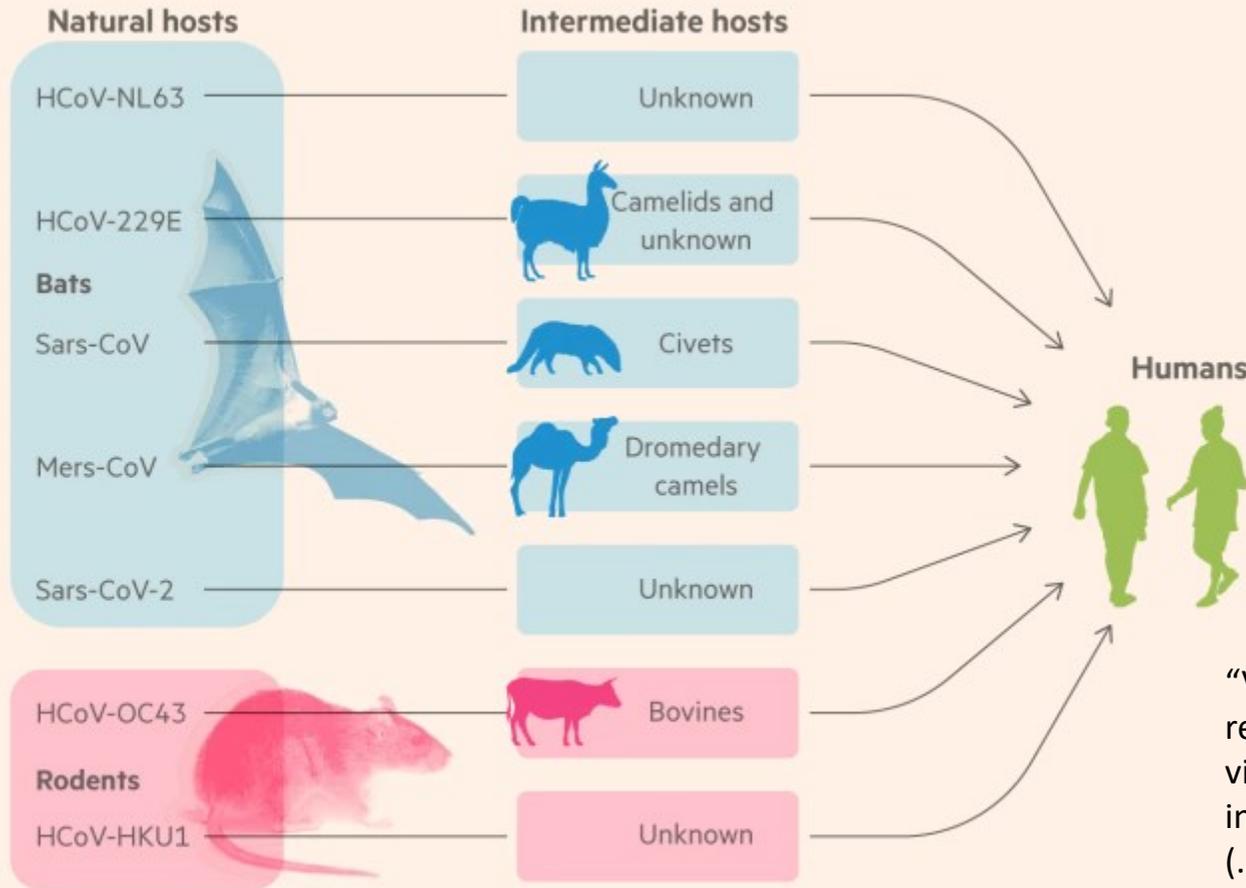
- SARS coronavirus (*homme*)
- Civet coronavirus (*civette*)
- Hedgehog coronavirus (*hérisson*)
- Bat coronavirus (*chauve-souris*)
- Rabbit coronavirus (*lapin*)
- Camel coronavirus (*chameau*)
- Dog coronavirus (*chien*)
- Rat coronavirus (*rat*)
- Bovine coronavirus (*vache*)
- Equine coronavirus (*cheval*)
- Yak coronavirus (*yak*)
- Pangolin coronavirus (*pangolin*)
- Porcine coronavirus (*cochon*)



Certains coronavirus provoquent également des bronchites chez les poulets et une maladie intestinale mortelle chez les porcs

*La classification des virus est particulièrement complexe...
Il n'est pas toujours facile de s'y retrouver...*

Animal hosts of human coronaviruses



Source: International Journal of Biological Sciences
© FT

What bats can teach us about developing immunity to Covid-19 | Free to read

Efforts to develop effective drugs or vaccines depend on understanding how the virus outwits the immune system

[Financial Time](#)

“Viruses love bats. (...)The big difference is that bats’ remarkable immune system tames and tolerates many viruses that cause havoc when they spread to humans, including the coronavirus responsible for Covid-19. (...) Coronaviruses have been evolving in bats for thousands or millions of years. (...)Viruses are much more virulent when they spread to humans from bats than from other mammals,” says Prof Crespi. “Yet they seem to do little harm to the bats themselves.” ”

Question à se poser pour étudier l'origine de SARS-CoV-2:

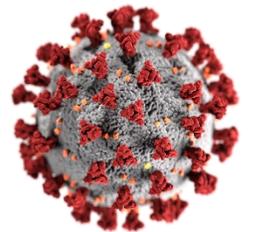
Quel coronavirus est le plus similaire à SARS-CoV-2 ?

Bioinformatique

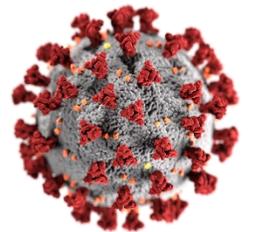
Outil de recherche de similarité: [BLAST](#)

Banque de données de protéines: [UniProtKB/Swiss-Prot](#)

Note: Il existe plusieurs milliers de coronavirus de chauve-souris....



Comparer la protéine Spike de différents coronavirus (chauve-souris, civette, pangolin, ...)



La protéine Spike chez différents coronavirus (1)

Voici des séquences partielles de la protéine Spike de différents coronavirus infectant différentes espèces, différentes années.

Quelle séquence est la plus similaire à human SARS-CoV-2 ?

Quelle(s) hypothèse(s) pourriez-vous faire ?

Human SARS-CoV-2	(2020)	IRGDEVQRQIAPGQTGKIAD
Pangolin coronavirus	(2020)	VRGDEVQRQIAPGQTGRIAD
Human SARS-CoV	(2003)	VKGDDVQRQIAPGQTGVVIAD
Bat coronavirus	(2020)	ITGDEVQRQIAPGQTGKIAD



wikipedia



wikipedia

La protéine Spike chez différents coronavirus (2)

Voici des séquences partielles de la protéine Spike de différents coronavirus infectant différentes espèces, différentes années.

```
>Human_SARS2020
FSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD
>Civet_2003_coronavirus
FSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIAD
>Pangolin_2020_coronavirus
FSTFKCYGVSPTKLNDLCFTNVYADSFVVRGDEVQRQIAPGQTGRIAD
>Human_SARS2003
FSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIAD
>Bat_2020_coronavirus
FSTFKCYGVSPTKLNDLCFTNVYADSFVITGDEVQRQIAPGQTGKIAD
>Human_MERS2012
VNDFTCsqISPAAIASNCYSSLILDYFSYPLSMKSDLSVSSAGPISQ
>Bat_2007_coronavirus
VDEFSCNGISPDSIARGCYSTLTVDYFAYPLSMKSYIRPGSAGNIPL
```

La protéine Spike chez différents coronavirus (2)

Comparer les séquences partielles de la protéine Spike de différents coronavirus, en faisant des alignements 2 à 2 avec la protéine Spike de SARS-CoV-2 (2020) à l'aide de [Align@UniProt](#).

Remplir le tableau suivant:

% identité	Human_SARS2020
Human_SARS2020	100
Human_SARS2003	
Civet_2003_coronavirus	
Pangolin_2020_coronavirus	
Bat_2020_coronavirus	
Human_MERS2012	
Bat_2007_coronavirus	

% identité	Human_SARS2020
Human_SARS2020	100
Human_SARS2003	82.2
Civet_2003_coronavirus	82.2
Pangolin_2020_coronavirus	86.7
Bat_2020_coronavirus	85.9
Human_MERS2012	18.1
Bat_2007_coronavirus	22.7

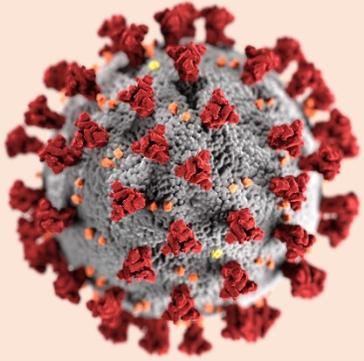
Réponse:

Quelle(s) chaîne(s) de transmission de SARS-CoV-2 serait envisageable selon ces résultats ?

homme -> homme	1
cochon -> homme	2
civette -> homme	3
chauve-souris -> homme	4
pangolin -> homme	5
chauve-souris -> pangolin -> chauve-souris -> homme	6

Attention: il s'agit de découvrir la démarche scientifique et de formuler des hypothèses et en aucun cas des conclusions !
Des conclusions sur les chaînes de transmission sont impossibles à faire avec si peu de données!

L'échantillonnage (présence ou absence d'une séquence (coronavirus du chameau 2012)) et/ou des erreurs de séquençage peuvent influencer de manière significative l'interprétation



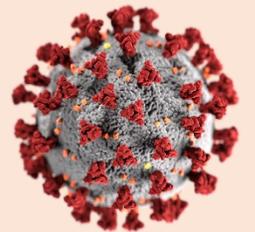
5 – A la recherche d'un traitement ...

Biologie:

[C'est quoi un médicament?](#)

Bioinformatique

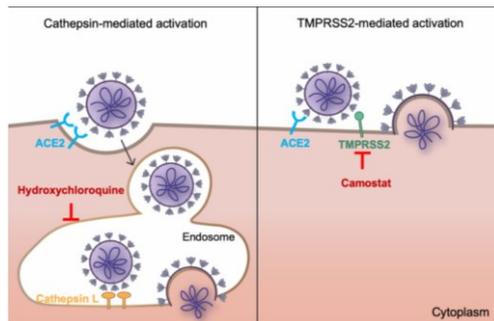
Banque de données de structure 3D: [PDB](#)





<https://viralzone.expasy.org/9078>

Un schéma du cycle d'infection du virus dans une cellule humaine et des **principaux traitements** en voie d'investigation: **vaccin** (neutralizing antibody) et **molécules médicament** ciblant différentes voies biologiques et protéines du coronavirus.



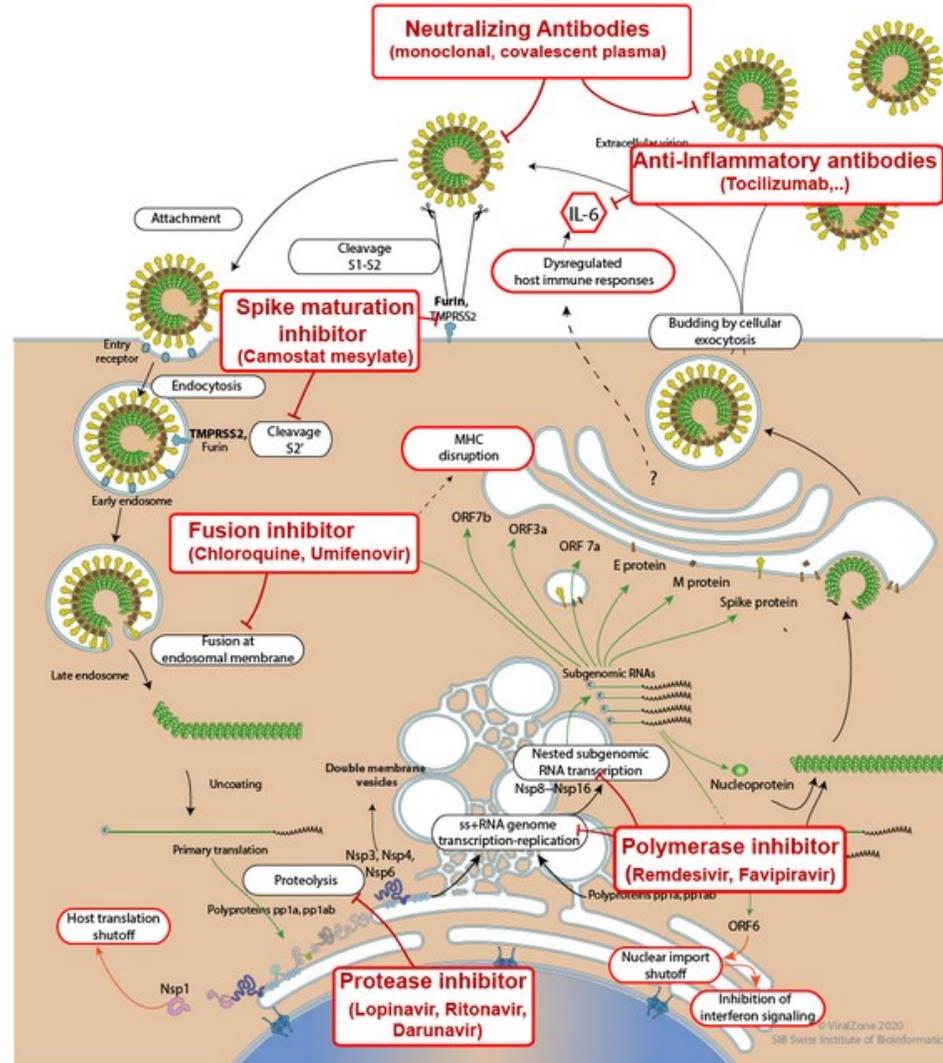
Le virus pourrait se passer de la 'voie endosome': ceci pourrait expliquer l'échec de la chloroquine...

<https://pubmed.ncbi.nlm.nih.gov/33465165/>

Antiviral drugs

There is no antiviral drug globally accepted for treatment, but many drugs are under investigations. About 21 trials of antivirals are under development (3th May 2020). Antiviral drugs are difficult to find, because of the cell-parasitic nature of viruses. It has been long to get efficient drugs against HIV or Hepatitis C. COVID-19 vaccine may arrive before we get efficient antivirals.

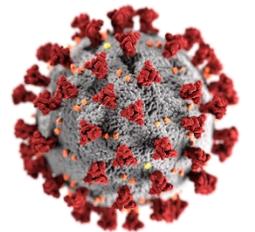
External resource

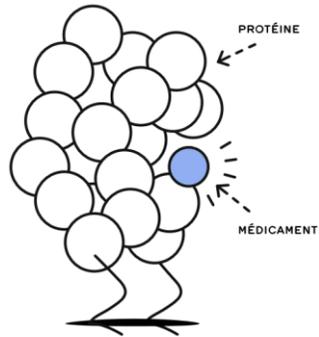


ANTIVIRAL DRUGS This table displays the main drugs under investigation

Médicaments ?

[C'est quoi un médicament?](#)





L'interaction protéine-médicament est un peu à l'image de l'interaction d'une clé avec une serrure: elle dépend pour beaucoup de la forme du médicament et de la protéine.

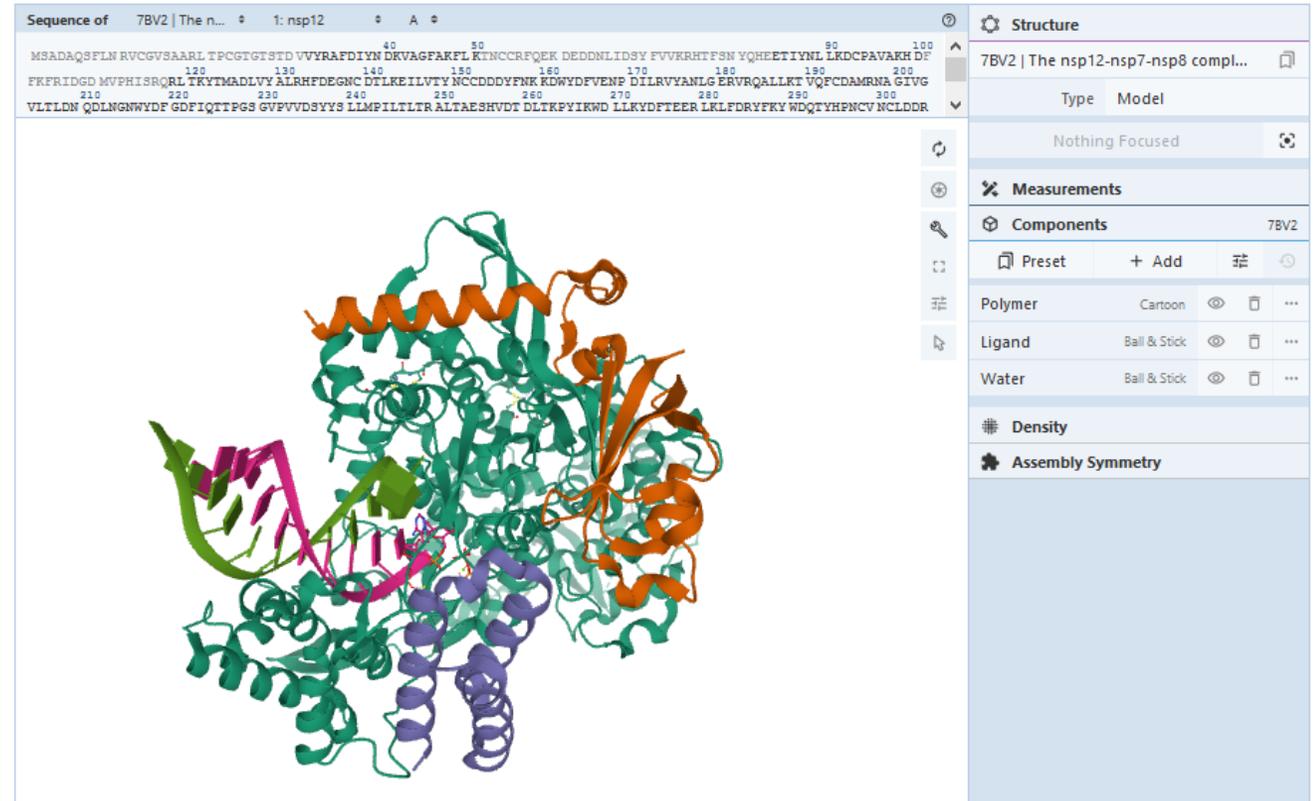
Structure 3D de la polymérase du coronavirus en présence du médicament Remdesivir (Publication)

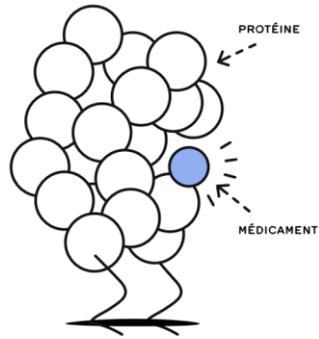
- Visualiser l'ARN en train d'être dupliqué (double brin) et le médicament

- Comment les chercheurs conçoivent-ils les médicaments de demain? [Atelier Drug Design](#)

7BV2

The nsp12-nsp7-nsp8 complex bound to the template-primer RNA and triphosphate form of Remdesivir(RTP)

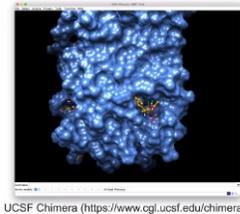




L'interaction protéine-médicament est un peu à l'image de l'interaction d'une clé avec une serrure: elle dépend pour beaucoup de la forme du médicament et de la protéine.

[Structure 3D de la protéase](#) du coronavirus en présence d'un inhibiteur potentiel ([Publication](#))

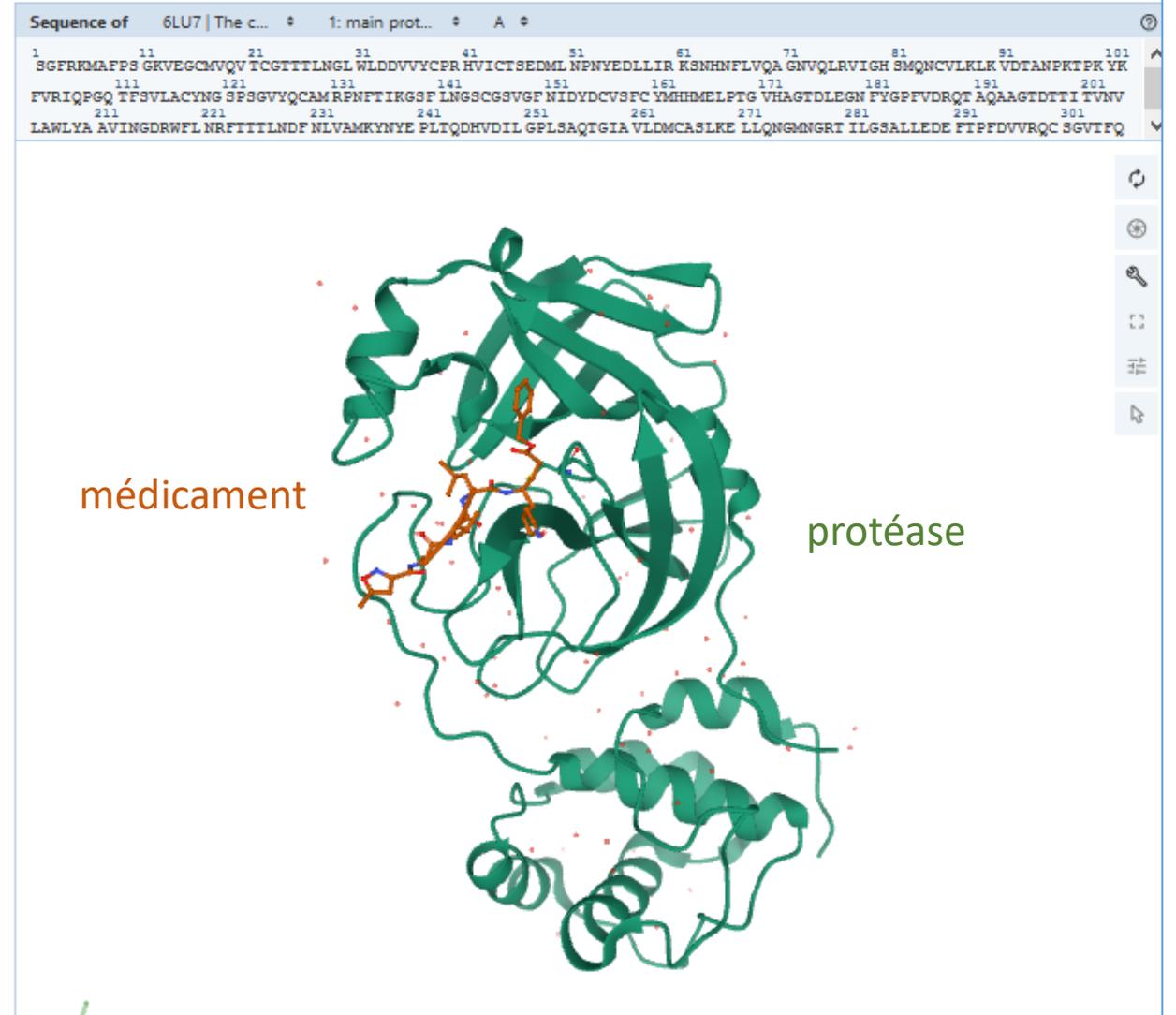
La même protéase avec un autre médicament (en cours de validation): <https://www.rcsb.org/3d-view/5RG1/>



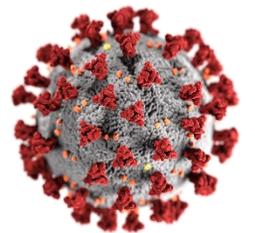
- Comment les chercheurs conçoivent-ils les médicaments de demain? [Atelier Drug Design](#)
- [La protéase dans UniProtKB/Swiss-Prot](#) (3C-like proteinase)

6LU7

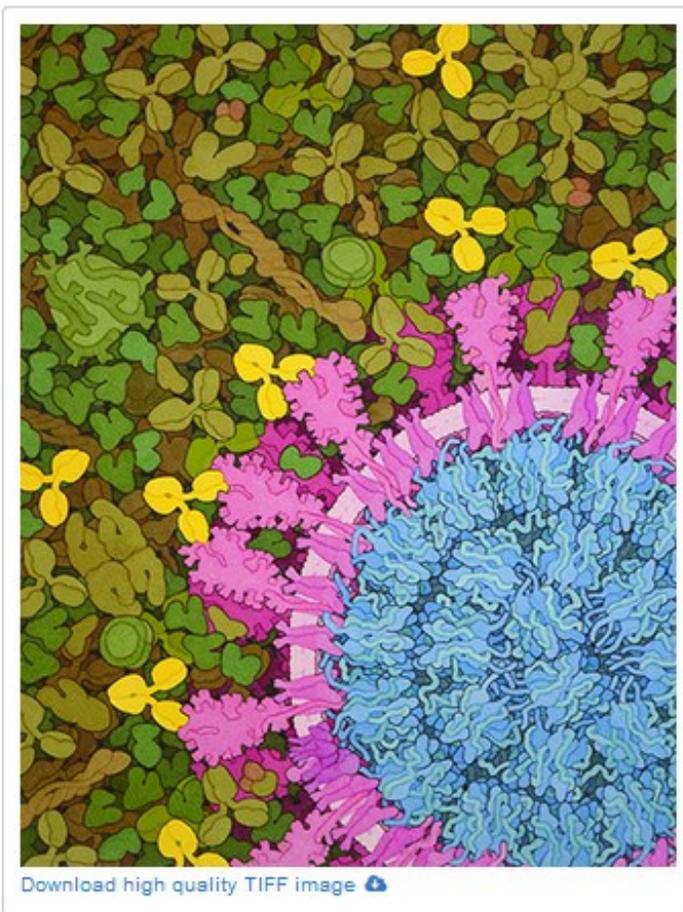
The crystal structure of COVID-19 main protease in complex with an inhibitor N3



Vaccin ?



SARS-CoV-2 and Neutralizing Antibodies, 2020



<http://pdb101.rcsb.org/sci-art/goodsell-gallery/sars-cov-2-and-neutralizing-antibodies>

Acknowledgement: David S. Goodsell, RCSB Protein Data Bank and Springer Nature; doi: 10.2210/rcsb_pdb/goodsell-gallery-025

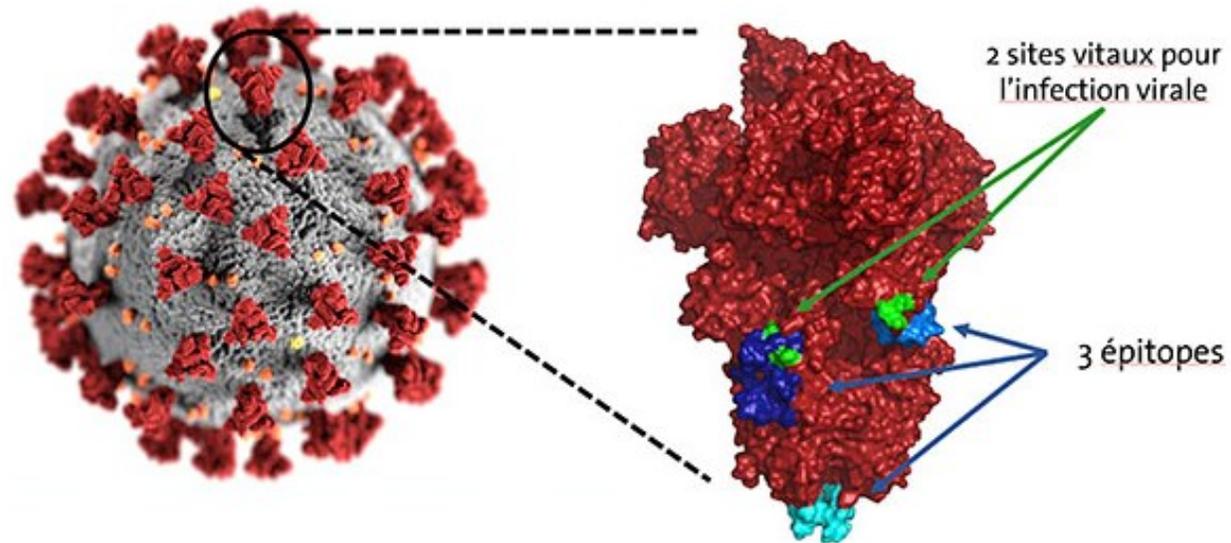
This painting shows a cross section through SARS-CoV-2 surrounded by blood plasma, with neutralizing antibodies in bright yellow. The painting was commissioned for the cover of a special COVID-19 issue of Nature, presented 20 August 2020.

It incorporates information from two cryoelectron microscopy studies that explore the shape and distribution of spikes and the nucleoprotein:

Yao H et al. (2020) Molecular architecture of the SARS-CoV-2 virus. bioRxiv preprint DOI: [10.1101/2020.07.08.192104](https://doi.org/10.1101/2020.07.08.192104)

Ke Z et al. (2020) Structures, conformations and distributions of SARS-CoV-2 spike protein trimers on intact virions. bioRxiv preprint DOI: [10.1101/2020.06.27.174979](https://doi.org/10.1101/2020.06.27.174979)

Des scientifiques de l'UNIGE et des HUG ont identifié les parties de l'enveloppe du SARS-Cov-2 qui sont le plus fréquemment visées par les anticorps. Elles représentent une cible potentielle pour le développement d'un vaccin.



Localisation des trois sites ciblés le plus fréquemment par les anticorps humains sur les clous (ou "spikes") du coronavirus responsable du COVID-19 @UNIGE

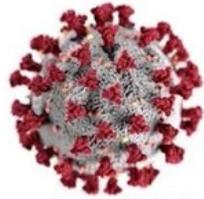
<https://www.unige.ch/communication/communiqués/2020/deux-faibles-potentiels-dans-la-cuirasse-du-coronavirus/>

[Vidéo 4](#) (2:26)

Vaccins ou médicaments anti-viraux : qu'est-ce qui est le plus facile à développer ?

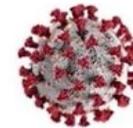


Search



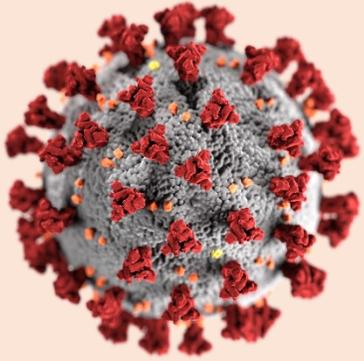
Vaccins ou médicaments anti-viraux:
qu'est-ce qui est le plus 'facile' à développer ?

Phillippe Le Mercier
Biologiste, Doctorat en virologie - Institut Pasteur
ViralZone, Groupe Swiss-Prot, Genève
SIB Swiss Institute of Bioinformatics



0:00 / 2:25



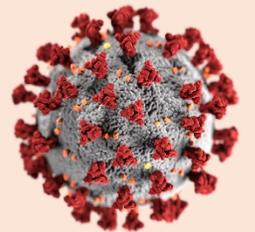


Un exemple de recherche menée au SIB

Référence:

Christian Sigrist, Alan Bridge, Philippe Le Mercier

DOI:<https://doi.org/10.1016/j.antiviral.2020.104759> (pdf)



Le virus SARS-CoV-2 est un cousin très proche du virus SARS-CoV,
responsable de l'épidémie de 2003,
qui avait infecté plus de 8'000 personnes dans 30 pays différents.



Swiss Institute of
Bioinformatics

Des chercheurs du SIB ont eu l'idée de **comparer** la séquence en acide aminé de la protéine Spike de différents coronavirus.

Ils ont comparé en particulier la protéine Spike du coronavirus SARS-CoV-2 avec la protéine Spike du coronavirus SARS-CoV, le virus qui avait infecté l'homme en 2003.

Comparer ces 2 séquences en acide aminé:

>Spike_SARS-CoV-2

RISNCSVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTG

>Spike_SARS-CoV

KISNCSVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVQRQIAPGQTG

Vous pouvez le faire manuellement ou à l'aide de l'outil bioinformatique [Align@UniProt](#)

Combien de différences ?

Retrouvez-vous 3 acides aminés consécutifs 'RGD' dans une des séquences ? Laquelle ?

Comparer les séquences en acide aminé complètes des protéines Spike de SARS-CoV-2 et de SARS-CoV

Spike_SARS-CoV-2: [lien](#)

Spike_SARS-CoV: [lien](#)

Copier et coller les séquences dans [l'outil Align@Uniprot](#)

Chercher le motif RGD dans l'alignement (Ctrl F)

A quelle position dans la séquence de SARS-CoV-2 le retrouvez-vous ?

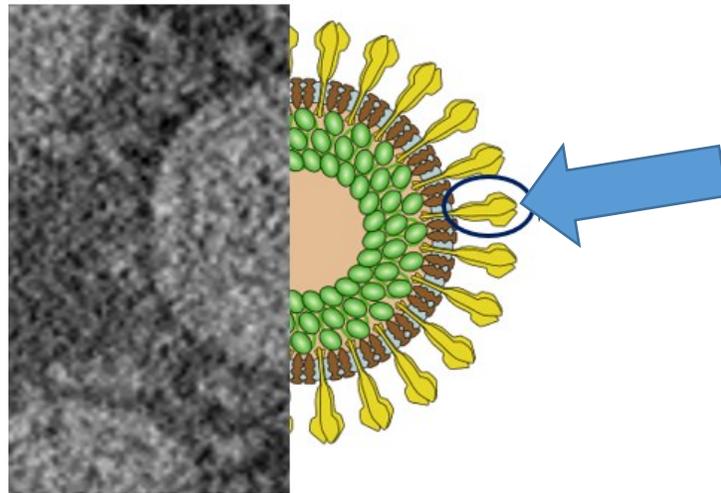


Swiss Institute of
Bioinformatics

Ces chercheurs ont ainsi découvert la présence d'un **motif RGD** dans la séquence de la protéine de SARS-CoV-2, motif qui est absent dans la séquence de la protéine Spike de SARS-CoV.

Ce motif est très célèbre en virologie, car il permet aux virus d'entrer dans les cellules humaines en utilisant une autre porte d'entrée que ACE2, une autre protéine appelée **intégrine**.

Pour que le motif RGD puisse jouer un rôle dans l'interaction avec les cellules humaines, il doit se trouver à la surface de la protéine Spike



Est-ce que le motif RGD est localisé à la surface de la protéine Spike ?

Pour répondre à cette question, il faut étudier la structure 3D de la protéine Spike.

Voici une représentation de la **structure 3D de la protéine Spike** et de sa séquence en acides aminés (bande de données PDB):

RCSB PDB Deposit Search Visualize Analyze Download Learn More

PDB-101 PDB EMDatabank Nucleic Acid Database Worldwide Protein Data Bank Foundation

Structure Summary **3D View** Annotations Experiment Sequence

6VXX

Structure of the SARS-CoV-2 spike glycoprotein (closed state)

Sequence of 6VXX | Struc... 1: SARS-CoV... A

```
LQSYGFQPIN GVGYPYRVV VLSPELLHAP ATVCGPKRST NLVNRKCVNF NFNGLTGTGV LTESNKKFLP FQQFGRDIAD TIDAVRDPQT LEILDITPCS PG
GVSVIITPG TINTSNQAVL YQDVNCTEVP VAIHADQLIP TRWVYSTGSH VFQTRAGCLI GAELVNSNVE CDIPIGAGIC ASVYQTINSP SGAGSVASQS IIAV
TMSLGA ENSVAYSNNIS IAIPTNFTIS VTTEILPVSMTKTSVDCITMY ICGDSTECNS LLLQVGSFCT QLNRLALGIA VEQDRNTQEV FAQVKQIYKT PFIKDF
```

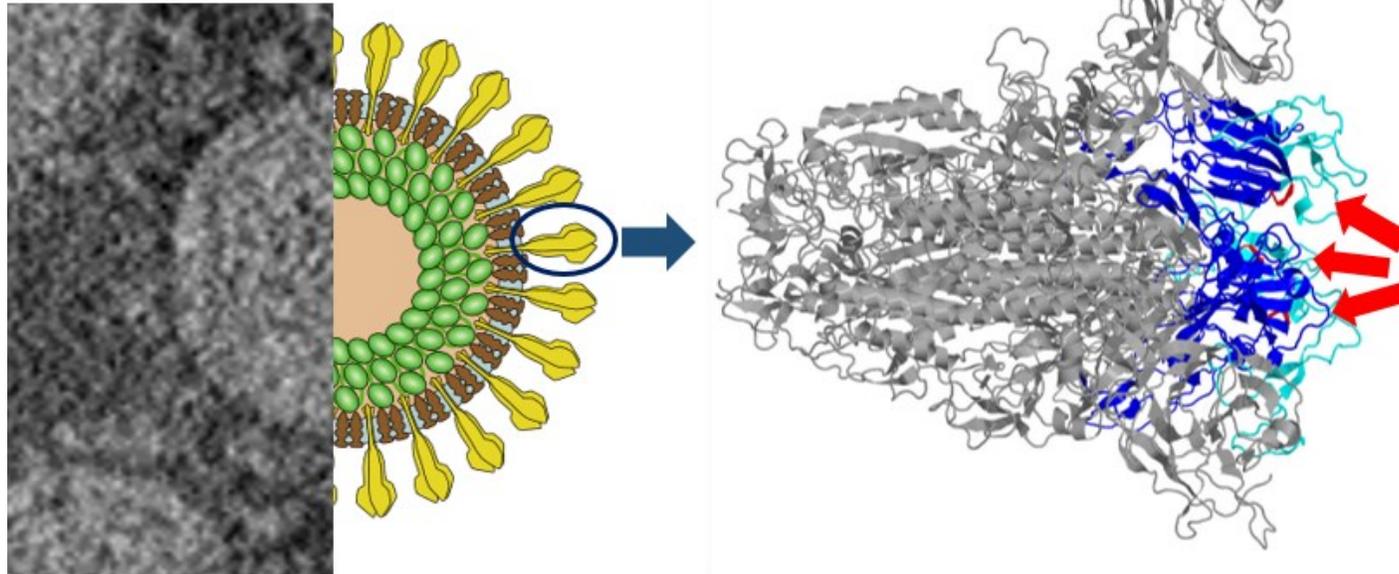
<https://www.rcsb.org/3d-view/6vxx>

Localiser le motif RGD dans la structure 3D de la protéine Spike

Truc: le motif RGD est en position 403-405 dans la séquence

[Nature, août 2020](#)

En utilisant les données connues de la structure 3D de la protéine Spike du SARS-CoV-2, les chercheurs du SIB ont montré que le motif RGD (en rouge) est localisé à la surface de la protéine Spike, proche de la région qui est impliquée dans l'interaction avec les protéines 'récepteur' humaines (en bleu).



En résumé: Spike, le motif RGD et les intégrines

Le motif RGD, retrouvé à la surface de la protéine Spike du virus SARS CoV-2, est connu pour favoriser l'interaction avec d'autres protéines humaines, appelées intégrines.

Ce motif a été retrouvé chez tous les virus SARS-CoV-2 qui ont été séquencés jusqu'à maintenant. Il est possible que le virus ait acquis ce motif au cours de son évolution et ait ainsi acquis la capacité de lier les intégrines pour favoriser son entrée dans les cellules de son hôte, mais cela reste à prouver.

Il existe actuellement peu de molécules antivirales efficaces contre le SARS-CoV-2. Les agents qui bloquent la liaison aux intégrines pourraient constituer une piste prometteuse. Les bloqueurs connus de la liaison aux intégrines comprennent entre autre l'anticorps natalizumab utilisé pour le traitement de la sclérose en plaques et la maladie de Crohn.



Antiviral Research

Volume 177, May 2020, 104759



A potential role for integrins in host cell entry by SARS-CoV-2

Christian JA Sigrist , Alan Bridge , Philippe Le Mercier  

Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, Switzerland

Received 20 February 2020, Revised 26 February 2020, Accepted 28 February 2020, Available online 1 March 2020.



Swiss Institute of
Bioinformatics

<https://doi.org/10.1016/j.antiviral.2020.104759>

Grâce à cette information, des médecins espagnols ont continué à traiter un patient Covid-19 avec un médicament ciblant les intégrines



ELSEVIER

Multiple Sclerosis and Related Disorders
Volume 44, September 2020, 102250



Correspondence

Covid-19 in a patient with multiple sclerosis treated with natalizumab: May the blockade of integrins have a protective role?

Clara Aguirre ^a  , Virginia Meca-Lallana ^a, Ana Barrios-Blandino ^b, Beatriz del Río ^a, Jose Vivancos ^c

Show more 

<https://doi.org/10.1016/j.msard.2020.102250> [Get rights and content](#)

<https://doi.org/10.1016/j.msard.2020.102250>

Grâce entre autre à cette information, des chercheurs sont à la recherche d'un médicament qui bloque l'interaction entre Spike et les intégrines.

The Integrin Binding Peptide, ATN-161, as a Novel Therapy for SARS-CoV-2 Infection

Brandon J Beddingfield ^{1 2}, Naoki Iwanaga ³, Prem P Chapagain ^{4 5}, Wenshu Zheng ⁶, Chad J Roy ^{1 2}, Tony Y Hu ⁶, Jay K Kolls ³, Gregory J Bix ^{7 8 9 10 11}

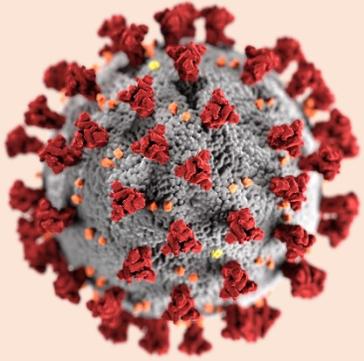
Affiliations + expand

PMID: 33102950 PMID: [PMC7566794](#) DOI: [10.1016/j.jacbts.2020.10.003](#)

[Free PMC article](#)

Abstract

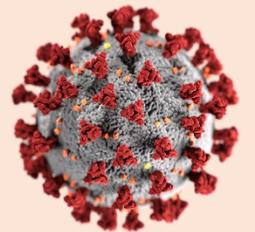
Many efforts to design and screen therapeutics for the current severe acute respiratory syndrome coronavirus (SARS-CoV-2) pandemic have focused on inhibiting viral host cell entry by disrupting ACE2 binding with the SARS-CoV-2 spike protein. This work focuses on the potential to inhibit SARS-CoV-2 entry through a hypothesized $\alpha 5\beta 1$ integrin-based mechanism, and indicates that inhibiting the spike protein interaction with $\alpha 5\beta 1$ integrin (+/- ACE2), and the interaction between $\alpha 5\beta 1$ integrin and ACE2 using a novel molecule ATN-161 represents a promising approach to treat COVID-19.



6 - SARS-CoV-2 et HIV

Bioinformatique

Recherche de similarité et notion de 'hasard'



SARS-CoV-2: un virus provenant d'une manipulation humaine ?


FAKE NEWS **FACTS**


- SARS-Cov-2 est très similaire à de nombreuses souches de coronavirus circulant dans la nature en Asie, avant et après la pandémie.
- Ces virus sont connus pour sauter d'une espèce à l'autre sans problème, donc rien de surprenant.
- L'analyse génétique de SARS-CoV-2 présente une organisation génomique et des protéines en tout point similaires aux autres virus sauvages.

Les thèses selon lesquelles une partie d'HIV a été insérée par l'homme dans le virus sont de mauvaises interprétation des résultats de recherche de similarité (BLAST)

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

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>

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

Abstract

Info/History

Metrics

 Preview PDF

Abstract

This paper has been withdrawn by its authors. They intend to revise it in response to comments received from the research community on their technical approach and their interpretation of the results. If you have any questions, please contact the corresponding author.

**FAKE NEWS
FACTS**

[Ce papier](#) a été démenti et retiré. Les auteurs ont utilisé des fragments de 6 à 10 acides aminés: il existe près de 1 millions de séquences d'HIV hypervariables, c'est inévitable qu'il y ait des similarités juste par hasard. Ce papier, même si il n'a jamais été publié, a fait le buzz...

Pour les experts (1):

[Emerg Microbes Infect.](#) 2020; 9(1): 378–381.

PMCID: PMC7033698

Published online 2020 Feb 14. doi: [10.1080/22221751.2020.1727299](https://doi.org/10.1080/22221751.2020.1727299)

PMID: [32056509](https://pubmed.ncbi.nlm.nih.gov/32056509/)

HIV-1 did not contribute to the 2019-nCoV genome

[Chuan Xiao](#),^{a,CONTACT} [Xiaojun Li](#),^b [Shuying Liu](#),^c [Yongming Sang](#),^d [Shou-Jiang Gao](#),^e and [Feng Gao](#)^{b,f}

▶ [Author information](#) ▶ [Article notes](#) ▶ [Copyright and License information](#) [Disclaimer](#)

Voici des exemples de séquences d'acides aminés retrouvées dans les protéines de SARS-CoV-2 utilisées pour 'démontrer' que SARS-CoV-2 contient des morceaux de génome de HIV:

TNGTKR
HKNNKS
RSYLTPGDSSSG
QTNSPRRA

La 'E value' est une probabilité de trouver le même résultat par hasard. Plus cette valeur est petite (< 0) plus le match est 'validé'.

Faire un Blast 'protein' au NCBI: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Sélectionner Organism: "Human immunodeficiency virus (taxid:12721)"

Cliquer sur 

Créer une séquence aléatoire avec les mêmes 'lettres' ([edit sequence]) et refaire un BLAST.

Que pouvez-vous conclure?

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7033698/>

Pour les experts (2):

Voici un morceau du gène codant pour la protéine Spike:

AATGGTACTAAGAGGTTTGATAACCCTG

Faire un Blast 'nucléotide' au NCBI: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.

Sélectionner Organism: "Human immunodeficiency virus (taxid:12721)"

Cliquer sur **BLAST**

La 'E value' est une probabilité de trouver le même résultat par hasard. Plus cette valeur est petite (< 0) plus le match est 'validé'.

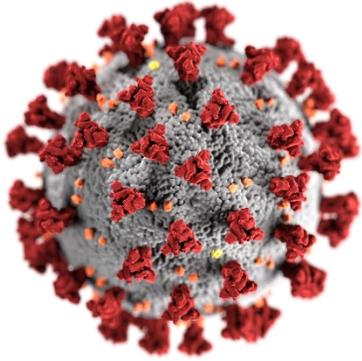
Créer une séquence aléatoire avec les mêmes 'lettres' ([edit sequence]) et refaire un BLAST.

Que pouvez-vous conclure?

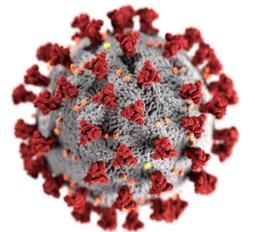
A HIV-1 isolate 19828.PPH11 from Netherlands envelope glycoprotein (env) gene, partial cds				
Sequence ID:	HQ644953.1	Length: 1143	Number of Matches: 1	Range 1: 967 to 994
Score	Expect	Identities	Gaps	Strand
38.3 bits(41)	7.5	25/28(89%)	0/28(0%)	Plus/Plus
Query	86	AATGGTACTAAGAGGTTTGATAACCCTG	113	
Sbjct	967	AATGGTACTAAAAGGTTAGATAACACTG	994	
B HIV-1 isolate patient B clone 16.3 from Netherlands envelope glycoprotein (env) gene, complete cds				
Sequence ID:	HQ386166.1	Length: 2580	Number of Matches: 1	Range 1: 2493 to 2523
Score	Expect	Identities	Gaps	Strand
39.2 bits(42)	2.1	27/31(87%)	0/31(0%)	Plus/Minus
Query	351	CCTAAAAGTTCCTTGTGAATAACTGTATTATT	381	
Sbjct	2523	CCTAAAAGTTCCTTGTGAATAATTCTATAATT	2493	

la séquence aléatoire. On peut en conclure que la similarité entre la séquence codante de la protéine S et le génome du VIH n'est pas significative. Les alignements ont été réalisés sur le site BLAST du NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Figure 6. Recherche de similarités entre les séquences codant pour la protéine spike de CoV2 et le génome de HIV. A. Alignement le plus significatif entre la séquence codant pour la protéine S de SARS-CoV-2 (query) et le génome du VIH (subject). B. Contrôle négatif : alignement le plus significatif entre une séquence aléatoire, obtenue en mélangeant les nucléotides de la séquence précédente, et le génome du VIH. Noter la valeur du score *expect*, qui indique le nombre de faux-positifs attendus au hasard. Ce score présente pour les deux alignements des valeurs supérieures à 1, et est même plus élevé pour l'alignement de la séquence de CoV que pour



Cet atelier est issu d'une collaboration entre
L'éprouvette, laboratoire public de l'Université de Lausanne
(Service Culture et Médiation Scientifique)
et
le groupe Swiss-Prot - [SIB Swiss Institute of Bioinformatics](https://www.sib.ac.ch/)



1

Dernières nouvelles... Le vaccin à ARNm

Cette séquence est la **séquence du gène codant pour la protéine Spike de SARS-CoV-2.**

Elle est présente dans les vaccins contre SARS-CoV-2 dits ‘vaccins à ARN messenger (ou ARNm)’. Ces vaccins contiennent une séquence ARNm composée au final de 4284 nucléotides, qui a été modifiée afin de permettre une production optimale de la protéine Spike dans les cellules.

20 années de recherche concentrées dans un vaccin...

“A 30 microgram dose turns out to actually contain 30 micrograms of RNA. (...) This consists of around 13,000 billion repetitions of the same 4284 characters. In addition, there is a clever lipid (fatty) packaging system that gets the mRNA into our cells.(...)”

- Un article (expert): «...[the source code of the BioNTech/Pfizer SARS-CoV-2 vaccine](#)»
- Un article (SimplyScience): [Contre le coronavirus, voici le vaccin à ARNm !](#)

«The coronavirus is here to stay — here’s what that means

A vaccine that is 90% effective at blocking transmission will need to reach at least 55% of the population to achieve temporary herd immunity as long as some social distancing measures — such as face masks and many people working from home — remain in place to keep transmission in check, according to a model developed by Alexandra Hogan at Imperial College London and her colleagues. (A vaccine would need to reach almost 67% of people to provide herd immunity if all social distancing measures were lifted.) But if the rate of transmission increases because of a new variant, or if a vaccine is less effective than 90% at blocking transmission, vaccine coverage will need to be greater to blunt circulation.»

https://www.nature.com/articles/d41586-021-00396-2?utm_source=Nature+Briefing&utm_campaign=6aea13811d-briefing-dy-20210216&utm_medium=email&utm_term=0_c9dfd39373-6aea13811d-45521262

```
>Gene_Spike
atgtttgtttttctgtttttatgccactagctctagtcagtggttaattctacaacc
agaactcaatccccccgcatcacataatcttccacacgtggtgttattaccctgac
aaagtttcagatcctcagttttacattcaactcaggactgttcttaccttctttcc
aatgttacttggtccatgatacattgctctgggccaatggtactaagaggtttgat
aacctgtccaccatttaatgatggttttatttgcctccactgagaagttaacata
ataagaggtggattttggtagctttagattcgaagaccagctccctacttattgtt
aataacgctactaattgtgtattaaagctgtgaattccaatttggtaagccattt
ttgggtgtttataccacaaaaacaagaattggatggaaagtgtgtcagagtttat
ctagtcgcaataattgcacttttgaatgtctctcagcctttcttattggacctgaa
ggaaaaacaggttaattcaaaaacttaggaaattgtgtttaaagaattgtggttat
ttaaaaatattctaaagcacacgctataattttagtgogtgcctccctcaggtttt
tcggctttagaacctgtgttagtttgcacatagttataacactcactagtttcaaac
ttactgtcttaacatagaagttatttgaactcctggtgattcttccagttggacagc
ggtgtgcagcttattatgggttatcttcaactcaggttttctataaaataataat
gaaaaaggaaactacagatgctgtagctgtgcactgacccctctcagaacaaaag
tgtactgtgaaactcctcactgtagaaaaagaaactcaaaccttcaacttagagtc
caaccaacagaactattgttagattctcaataattcaaaccttggccctttggtag
gtttttaaaccaccagattgcatctgtttatgtgtggaacaggaagaatcagcaac
tgtgtgtgctgattctgtctataataatccgcatctttccactttaagtttat
ggatgtctcctactaaatataatgatctgtcttactaattgtctatcagatctcatt
gtaattagaggtgtagaagtcagacaaactcctccaggcacaactggaagattgctgt
tataatataaaattaccagatgatttaccagctgctgtatagattggaattcaacaat
ctgtattcaaggttgggtgtaattataataccctgtatagatttttaggaagctaat
ctcaaacctttgagagagatttcaactgaaactcactcagccggtgacacacactgt
aatgtgtgaaagttttaaattgtaacttcccttaacatcaatggtttccaaccaact
aatgtgtgttaccacacatcagagtagtagtactttcttggaaactctacatgca
ccagcaactgtttgtggacctaaaagctcactcaatttgggttaaaaaacaaatgtcact
ttcaactcaatggtttaaaccagcagactgttcttactgagctcaaaaaaagtttctg
ccttccaacaattggcagagacattgctgacactactgctgctcctgtatccacag
acactgagattctgacattacacactgttcttttgggtggtgtaggttataacacca
ggaaacaaactctcaaccaggttgcgttctttatcaggatgttaactgcacagaagtc
cctgtgctattcagatcaacttaactcactggtgcttttattctcaaggttct
aatgttttcaaacacgtgcagcgtgtttaaagggtgaaactgcacaactcatat
gaggtgacataccactggtgcaagttatgcgctagttatcagactcagactaattct
cctcggcggcagctagtgtagctcaactccatctgctcactatgtcacttgggt
gcagaaaattcagttgcttactctaaactctattgccataccacaaaatttactatt
aggttaccacagaattctaccagttctatgaccaagacatcagtagattgacaatg
tacatttgggtgattcaactgaaatgcaactcttttggcaataggcagtttttgg
acacaattaaacgtgcttttaactggaatagctgttgaacaaagacaaaaacccaagaa
gttttgcacaagtcaaaaatttacaacacacccaactaaagatttgggtggttt
aattttcaacaatttaccagatctcaactcaaaacaaagcaagaggtcattttagaat
ctactttcaacaagtgcactcagatgctgctcactcaacaacaaatggtgattgct
ctgtgtgattgctgctgagagacactcatttggcacaagaattttaaaggccttactgtt
ttgccactttgctcagagatgaaatgattgctcaactcctcagctgttaagcgggt
acaactctctgggtggacctttgggtcaggtgctgcatcaacaacactttgctatg
caaatggcttattagttttaaagtttggagttacacagaaatgtctctatgagaacca
aaattgttgcacaacaaattaaatgtgctatggcaaaattcaagactcacttctctcc
acagcaagtgcactggaaaaactcaagatggtgctcaacaaatgcaacaagctttaa
acgctgtttaaacaacttagctccaaatttgggtgcaatttcaaggtttttaaagatac
ctttcagctctgacaagaattgagcgtgaaatgcaaatgtaggtgtgacacagcaga
cttcaaaagttgacacataatgtgactcaacaataatagagctgcagaacacagagct
ctctgtaactctgctgctactaaaatgtagaggtgtgactggcaacacaaaagagtt
gattttgtgaaaaggctatcactatgctctccctcagctcagcactcaggtgta
gtctcttgcagtgactatgtcctgcacaaagaaagaaactcacaactcctcctgccc
atgttgcagtaggaaagcaacttccctcgtgaaaggttcttggttcaaatggcaaca
cactggtttgtaacacaaagaaatttttgaacacaaactcaactacagacaacaca
ttgtgtcgtgtaactgtgagttgtaataagaaattgcaacaacacagtttatgatcct
ttgcaactgaaatagactcactcaaggaggaattagataaaatattttaaagaatcataca
tcaccagatgttgaattaggtgacactctggcactaaatgctcagttgtaaacctcaa
aaagaaattgacgcctcaatgaggttggcaagaattttaaataactctcactcagctc
caagaactggaaagattgagcagtatataaaatggccatggtacattggcaggtttt
atagctggctgtgtgcatagtaattggtgacaattatgcttgcgtgtatgaccaggtgc
tgtagttgctcaaggcgtgtgttctgtggatcctgctgcaaaattgtaggaagcagac
ctgagccagctgcaaaagagtcacaaattacattacacata
```

2

Dernières nouvelles... Le variant anglais

La 'lignée britannique' du SARS-CoV-2 (B.1.1.7) a été découverte le 8 décembre 2020:

- 17 'nouvelles' mutations sont présentes dans le génome du virus B.1.1.7

Comparaison des génomes de SARS-CoV-2:

- [génomme de référence \(NC_045512.2\)](#)
- [génomme du variant anglais B.1.1.7 \(LR991698\)](#)

(! seule une toute petite partie de l'alignement est visible ici)

Les différences entre le génome de SARS-CoV-2 (génomme de référence) et le génome du virus 'anglais B.1.1.7' apparaissent en clair dans cet alignement.

Ces différences correspondent aux 'nouvelles' mutations présentes dans le virus anglais.

ceci est une mutation



LR991698.2	23023	ATCATATGGT	23082
NC_045512.2	23041	ATCATATGGT	23100
LR991698.2	23083	TTCTTTTGA	23142
NC_045512.2	23101	TTCTTTTGA	23160
LR991698.2	23143	GGTAAAAA	23202
NC_045512.2	23161	GGTAAAAA	23220
LR991698.2	23203	TGAGTCTA	23262
NC_045512.2	23221	TGAGTCTA	23280
LR991698.2	23263	TGATGCTG	23322
NC_045512.2	23281	TGATGCTG	23340
LR991698.2	23323	GTCAGTGT	23382
NC_045512.2	23341	GTCAGTGT	23400
LR991698.2	23383	GGTGTTA	23442
NC_045512.2	23401	GGATGTT	23460
LR991698.2	23443	GCGTGT	23502
NC_045512.2	23461	GCGTGT	23520
LR991698.2	23503	TGAACAT	23562
NC_045512.2	23521	TGAACAT	23580
LR991698.2	23563	TTATCAG	23622
NC_045512.2	23581	TTATCAG	23640
LR991698.2	23623	TGCCTAC	23682
NC_045512.2	23641	TGCCTAC	23700

2

Dernières nouvelles...
Le variant anglais

Comparaison des séquences en acide aminé de la [protéine Spike de référence](#) et de la protéine Spike du virus anglais (B.1.1.7) ([Accès aux séquences](#))

Les différences (changements d'acide aminé / mutations / variants) apparaissent en clair dans cet alignement.

P0DTC2 SPIKE_SARS2	1	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVVYDKVFRSSVLHSTQDLFLPFFS	60	601	GTNTSNQVAVLYQDVNCTEVPVAIHADOLTPTWRVYSTGNSVFPOTRAGCLIGAEHVNNYSY	660
Spike_UK_B.1.1.7	1	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVVYDKVFRSSVLHSTQDLFLPFFS	60	601	GTNTSNQVAVLYQDVNCTEVPVAIHADOLTPTWRVYSTGNSVFPOTRAGCLIGAEHVNNYSY	660
P0DTC2 SPIKE_SARS2	61	NVTWFHAIHVSCTNGTKRFDPVLPFNDGVYFASTEKSNIRGWIFGTTLDSKTQSLLIIV	120	661	ECDIPIGAGICASYQOTNSPRRARSVASOSIIAYTMSLGAENSVAYSNNNSIAIPIINFTI	720
Spike_UK_B.1.1.7	61	NVTWFHAIHVSCTNGTKRFDPVLPFNDGVYFASTEKSNIRGWIFGTTLDSKTQSLLIIV	120	661	ECDIPIGAGICASYQOTNSPRRARSVASOSIIAYTMSLGAENSVAYSNNNSIAIPIINFTI	720
P0DTC2 SPIKE_SARS2	121	NNATNVVIRKVFQPCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFIMDL	180	721	SVTTEILPVSMTKRTSVDCTMYICGDSPECSNLLLOYGSFCTQLNRALTGIAVEQDKNTQE	780
Spike_UK_B.1.1.7	121	NNATNVVIRKVFQPCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFIMDL	180	721	SVTTEILPVSMTKRTSVDCTMYICGDSPECSNLLLOYGSFCTQLNRALTGIAVEQDKNTQE	780
P0DTC2 SPIKE_SARS2	181	GKQGNFKNLRREFVFKNIDGYFKIYSKHTPINLVRDLPOGFSALEPLVDLPIGINITRFQT	240	781	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKSRSPFIEDLLFNKVTLADAGFIKQYGD	840
Spike_UK_B.1.1.7	181	GKQGNFKNLRREFVFKNIDGYFKIYSKHTPINLVRDLPOGFSALEPLVDLPIGINITRFQT	240	781	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKSRSPFIEDLLFNKVTLADAGFIKQYGD	840
P0DTC2 SPIKE_SARS2	241	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTEFLKYNENGTITDAVDCALDPLSETK	300	841	LGDI AARDL ICAQKFNGLTVLPLLTDEMI AQYTSALLAGTITSGWTFGAGAALQIPFAM	900
Spike_UK_B.1.1.7	241	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTEFLKYNENGTITDAVDCALDPLSETK	300	841	LGDI AARDL ICAQKFNGLTVLPLLTDEMI AQYTSALLAGTITSGWTFGAGAALQIPFAM	900
P0DTC2 SPIKE_SARS2	301	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYANNRRRISN	360	901	QMAYRFNGIGVTONVLYENOKLIANQFNSAIGKIQDLSSTASALGKLQDVVNQNAQALN	960
Spike_UK_B.1.1.7	301	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYANNRRRISN	360	901	QMAYRFNGIGVTONVLYENOKLIANQFNSAIGKIQDLSSTASALGKLQDVVNQNAQALN	960
P0DTC2 SPIKE_SARS2	361	CVADYSVLNYSASFSTFKCYGVSPKLNLDLCTNVYADSFVIRGDEVROIAPGQTGKIAD	420	961	TLVKQLSSNFGAISSVLDNIIARLDKVEAEVQIDRLITGRQLQSLQTYVTQQLIRAAEIRA	1020
Spike_UK_B.1.1.7	361	CVADYSVLNYSASFSTFKCYGVSPKLNLDLCTNVYADSFVIRGDEVROIAPGQTGKIAD	420	961	TLVKQLSSNFGAISSVLDNIIARLDKVEAEVQIDRLITGRQLQSLQTYVTQQLIRAAEIRA	1020
P0DTC2 SPIKE_SARS2	421	YNYKLPDDFTGCVIAWNSNNLDSKVGNNYLYRFRKSNLKPFFERDITSTEIYOAGSTPC	480	1021	SANLAATKMSSECVLQSKRVDFCGKGYHLMSPPOSAPHGVVFLHVTVVPAQEKNFITAPA	1080
Spike_UK_B.1.1.7	421	YNYKLPDDFTGCVIAWNSNNLDSKVGNNYLYRFRKSNLKPFFERDITSTEIYOAGSTPC	480	1021	SANLAATKMSSECVLQSKRVDFCGKGYHLMSPPOSAPHGVVFLHVTVVPAQEKNFITAPA	1080
P0DTC2 SPIKE_SARS2	481	NGVEGFNCYFPLOSYGQPTNGVYQPYRVVVLVSLFELLHAPATVCCPKKSTNLVKNKCVN	540	1081	ICHDKGAHFPRGEGVFSVNGTHWFVTQRNFYEQIITHTNTFVSGNCDVVIIGVNNTVYDE	1140
Spike_UK_B.1.1.7	481	NGVEGFNCYFPLOSYGQPTNGVYQPYRVVVLVSLFELLHAPATVCCPKKSTNLVKNKCVN	540	1081	ICHDKGAHFPRGEGVFSVNGTHWFVTQRNFYEQIITHTNTFVSGNCDVVIIGVNNTVYDE	1140
P0DTC2 SPIKE_SARS2	541	FNFNGLTGTGVLTESNKKFLPFQOQGRDIADTTDAVRDPQTEILLDITPCSFGGVSVITE	600	1141	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL	1200
Spike_UK_B.1.1.7	541	FNFNGLTGTGVLTESNKKFLPFQOQGRDIADTTDAVRDPQTEILLDITPCSFGGVSVITE	600	1141	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL	1200
				1201	QELGKYEQYIKWPWYIWLGFITAGLIAIVMVTIMLCMTSCCSCLGKCCSCGSCCKFDEDD	1260
				1201	QELGKYEQYIKWPWYIWLGFITAGLIAIVMVTIMLCMTSCCSCLGKCCSCGSCCKFDEDD	1260
				1261	SEPVLGKVKLHYT	1273
				1261	SEPVLGKVKLHYT	1273

[align@UniProt](#)

+ 501 **N>Y** In strain: B.1.1.7, 501YV2; May enhance affinity to human ACE2 receptor. (UniProt)

Liste des 'variants Spike UK' selon [UniProtKB](#): **N501Y**, P681H, T716I, S982A, D1118H

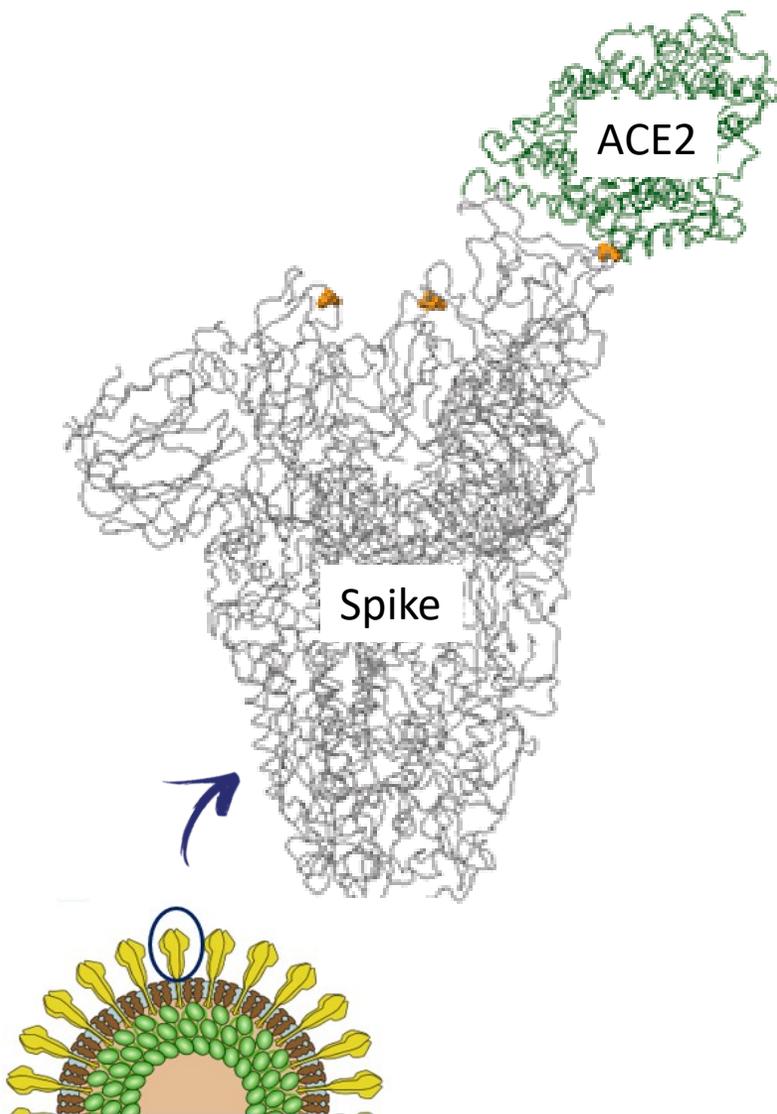
2

Dernières nouvelles... Le variant anglais

La guerre des clones....

La 'variante britannique' du SARS-CoV-2 (appelée B.1.1.7), est un virus parmi d'autres qui s'est développé particulièrement rapidement en Angleterre.

- 17 mutations sont présentes dans le génome du virus B.1.1.7: elles entraînent toutes des changements d'acides aminés dans les protéines du virus, dont 9 mutations dans la protéine Spike.
 - La mutation **N501Y** de la protéine Spike (en jaune sur cette image) pourrait modifier les interactions de la protéine avec les récepteurs présents sur les cellules humaines (ACE2).
- Le PDG de BioNTech, Uğur Şahin, a souligné que la protéine Spike codée par l'ARN messager du vaccin était composée de 1270 acides aminés: 'seulement' 9 d'entre eux sont différents dans la variante britannique du virus.
- La propagation du variant B.1.1.7 a eu lieu pendant une période propice à la transmission des virus respiratoires (janvier, février). Il faut souligner que plus le virus SARS-CoV-2 circulera, plus il aura la possibilité de 'muter'.



<https://science.sciencemag.org/content/371/6524/9>

3

Dernières nouvelles...
Le variant Afrique du Sud (501Y.V2)

Comparaison des séquences en acide aminé de la [protéine Spike de référence](#) et de la protéine Spike du virus d'Afrique du Sud (501Y.V2) ([Accès aux séquences](#))

Les différences (changements d'acide aminé / mutations / variants) apparaissent en clair dans cet alignement.

P0DTC2 SPIKE_SARS2	1	MFVFLVLLPLVSSQCVNLTTRTQLPYPATNSFTRGVYVYDVKVFRSSVLHSTQDLFLPFFS	60	P0DTC2 SPIKE_SARS2	601	GNNTSNQVAVLYQDVNCTEVPVAIHADQLTPTNRVYSTGNSVFEQTRAGCLIGAEHVNSY	660
Spike_South_Africa_50..	1	MFVFLVLLPLVSSQCVNLTTRTQLPYPATNSFTRGVYVYDVKVFRSSVLHSTQDLFLPFFS	60	Spike_South_Africa_50..	601	GNNTSNQVAVLYQDVNCTEVPVAIHADQLTPTNRVYSTGNSVFEQTRAGCLIGAEHVNSY	660
P0DTC2 SPIKE_SARS2	61	NVTWFHAIHVSGTNGTKREFNPVLPFNQDGVYFASTEKSNIRGWIPIGTTLDKSTQSLIIV	120	P0DTC2 SPIKE_SARS2	661	ECDFIPGAGICASYQIQNSPRRARSVASQSIILAVTMSLIGENSVAKSNNSIALIPNFTI	720
Spike_South_Africa_50..	61	NVTWFHAIHVSGTNGTKREFNPVLPFNQDGVYFASTEKSNIRGWIPIGTTLDKSTQSLIIV	120	Spike_South_Africa_50..	661	ECDFIPGAGICASYQIQNSPRRARSVASQSIILAVTMSLIGENSVAKSNNSIALIPNFTI	720
P0DTC2 SPIKE_SARS2	121	NNAINVVIKVCPEPQCNDPFLGVYVYHKNKSKWMESEFRVYSSANNCTFEVYSQPFLMDLE	180	P0DTC2 SPIKE_SARS2	721	SVTTEILPVSMTKTSVDCCTMYICGDSSTECNLLQYGSFCTQLNRAITGIAVEQDRNTQE	780
Spike_South_Africa_50..	121	NNAINVVIKVCPEPQCNDPFLGVYVYHKNKSKWMESEFRVYSSANNCTFEVYSQPFLMDLE	180	Spike_South_Africa_50..	721	SVTTEILPVSMTKTSVDCCTMYICGDSSTECNLLQYGSFCTQLNRAITGIAVEQDRNTQE	780
P0DTC2 SPIKE_SARS2	181	GKQGNFKNLRREFVFNIDGYFKIYKHTPINLVRDLQPGFSALEPLVDLPIGINITRFQT	240	P0DTC2 SPIKE_SARS2	781	YFAQYKQIYKTPPIKDFGGFNFSQILPDPKSKRSFIEDLLENKVLADAGFIKQYGGC	840
Spike_South_Africa_50..	181	GKQGNFKNLRREFVFNIDGYFKIYKHTPINLVRDLQPGFSALEPLVDLPIGINITRFQT	240	Spike_South_Africa_50..	781	YFAQYKQIYKTPPIKDFGGFNFSQILPDPKSKRSFIEDLLENKVLADAGFIKQYGGC	840
P0DTC2 SPIKE_SARS2	241	LLALHRSYLTGPDSSSGWTAGAAAYVGYLQPTFLFKYENGTITDAVDCALDPLSETR	300	P0DTC2 SPIKE_SARS2	841	LGDTAARDLICAKFNGLTLPPLLTDEMIQYTSALLAGTITSGWTFGAGALQIPFAM	900
Spike_South_Africa_50..	241	LLALHRSYLTGPDSSSGWTAGAAAYVGYLQPTFLFKYENGTITDAVDCALDPLSETR	300	Spike_South_Africa_50..	841	LGDTAARDLICAKFNGLTLPPLLTDEMIQYTSALLAGTITSGWTFGAGALQIPFAM	900
P0DTC2 SPIKE_SARS2	301	CTLRSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPGGEVFNATRFASVYAMNKRKISN	360	P0DTC2 SPIKE_SARS2	961	RLVKQLSSNFGAISVVMNDLISRLDRVEAEVQIDRLITGRISLQSLQTVVQQLIRAAEIRA	1020
Spike_South_Africa_50..	301	CTLRSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPGGEVFNATRFASVYAMNKRKISN	360	Spike_South_Africa_50..	961	RLVKQLSSNFGAISVVMNDLISRLDRVEAEVQIDRLITGRISLQSLQTVVQQLIRAAEIRA	1020
P0DTC2 SPIKE_SARS2	361	CVADYSVLYNSASFSTFKCYGVSPTKLNLDLCTNIVYADSFVIRGDEVQIAGQTSNIAD	420	P0DTC2 SPIKE_SARS2	1021	SANLAATRMSECVLQSRVDFPCGKGYHLSFPQSAPHGVVFLHVTYVPAQEKNTTAPA	1080
Spike_South_Africa_50..	361	CVADYSVLYNSASFSTFKCYGVSPTKLNLDLCTNIVYADSFVIRGDEVQIAGQTSNIAD	420	Spike_South_Africa_50..	1021	SANLAATRMSECVLQSRVDFPCGKGYHLSFPQSAPHGVVFLHVTYVPAQEKNTTAPA	1080
P0DTC2 SPIKE_SARS2	421	NYNKLDDPFTCCVIAWNSNLDKSVGGNYYLYRLFRKSNLKPFRDITSTEIYQAGSTPC	480	P0DTC2 SPIKE_SARS2	1081	ICHDKRAHFPRGQVFSNGTHWFVTRQNFYEPQIITDNTFVSGNCDVVIGIVNNIVYDE	1140
Spike_South_Africa_50..	421	NYNKLDDPFTCCVIAWNSNLDKSVGGNYYLYRLFRKSNLKPFRDITSTEIYQAGSTPC	480	Spike_South_Africa_50..	1081	ICHDKRAHFPRGQVFSNGTHWFVTRQNFYEPQIITDNTFVSGNCDVVIGIVNNIVYDE	1140
P0DTC2 SPIKE_SARS2	481	NGVLSGFRVLPFLQSYVQPTVYVYKQPYRVVVLSEFLLHAPAVCGPKKSTNLVKNKCVN	540	P0DTC2 SPIKE_SARS2	1141	LQPELDSFKEELDKVFNHSTSPDVLGDISGINASVNNIQEIDRLNEVARNLINESLIDE	1200
Spike_South_Africa_50..	481	NGVLSGFRVLPFLQSYVQPTVYVYKQPYRVVVLSEFLLHAPAVCGPKKSTNLVKNKCVN	540	Spike_South_Africa_50..	1141	LQPELDSFKEELDKVFNHSTSPDVLGDISGINASVNNIQEIDRLNEVARNLINESLIDE	1200
P0DTC2 SPIKE_SARS2	541	FNENGLTGTGVLTESNKRFLFPQQFGRDIDDTDAVRDPQTLIELDITPCSGGGVSVITP	600	P0DTC2 SPIKE_SARS2	1261	QELGKYEYQIKWFWYIWLGFAGLIAIVMVTIMLCMTSCCCLGKCCSCGSCCKFDEDD	1260
Spike_South_Africa_50..	541	FNENGLTGTGVLTESNKRFLFPQQFGRDIDDTDAVRDPQTLIELDITPCSGGGVSVITP	600	Spike_South_Africa_50..	1261	QELGKYEYQIKWFWYIWLGFAGLIAIVMVTIMLCMTSCCCLGKCCSCGSCCKFDEDD	1260
P0DTC2 SPIKE_SARS2				P0DTC2 SPIKE_SARS2	1261	SEFVLRGKVLKHT	1273
Spike_South_Africa_50..				Spike_South_Africa_50..	1261	SEFVLRGKVLKHT	1273

[align@UniProt](https://www.uniprot.org/align)

Liste des 'variants Spike du virus d'Afrique du Sud' selon [UniProtKB](#): L18F, D80A, D215G, R246I, K417N, **E484K**, **N501Y**, A570D, A701V

Publication : <https://www.medrxiv.org/content/10.1101/2020.12.21.20248640v1>

"Eight lineage-defining mutations in the spike protein, including three at important residues in the receptor-binding domain (K417N, E484K and N501Y) that may have functional significance."

Génome: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA694014>

3

Dernières nouvelles...
Le variant Afrique du Sud (501Y.V2)

Comparaison des séquences en acide aminé de la [protéine Spike de référence](#) et de la protéine Spike du virus d'Afrique du Sud (501Y.V2) ([Accès aux séquences](#))

Les différences (changements d'acide aminé / mutations / variants) apparaissent en clair dans cet alignement.

P0DTC2 SPIKE_SARS2	1	MFVFLVLLPLVSSQCVNLTTRTQLPYPAYTNSFTRGVYVYDVKVFRSSVLHSTQDLFLPFFS	60	P0DTC2 SPIKE_SARS2	601	GNLNSQVAVLYQDVNCTEVPVAIHADQLTPTNRVYSTGSMVFQTRAGCLIGAEHVNSY	660
Spike_South_Africa_50..	1	MFVFLVLLPLVSSQCVNLTTRTQLPYPAYTNSFTRGVYVYDVKVFRSSVLHSTQDLFLPFFS	60	Spike_South_Africa_50..	601	GNLNSQVAVLYQDVNCTEVPVAIHADQLTPTNRVYSTGSMVFQTRAGCLIGAEHVNSY	660
P0DTC2 SPIKE_SARS2	61	NVTWFHAIHVSGTNGTKREFNPVLPFNQDGVYFASTEKSNIRGWIPIGTTLDKSTQSLIIV	120	P0DTC2 SPIKE_SARS2	661	ECDFIPGAGICASYQQTNSPRRARSVASQSIILAVTMSLIGENSVAKSNNSIALPNTPTI	720
Spike_South_Africa_50..	61	NVTWFHAIHVSGTNGTKREFNPVLPFNQDGVYFASTEKSNIRGWIPIGTTLDKSTQSLIIV	120	Spike_South_Africa_50..	661	ECDFIPGAGICASYQQTNSPRRARSVASQSIILAVTMSLIGENSVAKSNNSIALPNTPTI	720
P0DTC2 SPIKE_SARS2	121	NNAINVVIKVCPEPQCNDPFLGVYVYHKNKSKWMESEFRVYSSANNCTFEVYSQPFLMDLE	180	P0DTC2 SPIKE_SARS2	721	SVTTEILPVSMTKTSVDCCTMYICGDSSTECNLLQYGSFCTQLNRAITGIAVEQDRNTOE	780
Spike_South_Africa_50..	121	NNAINVVIKVCPEPQCNDPFLGVYVYHKNKSKWMESEFRVYSSANNCTFEVYSQPFLMDLE	180	Spike_South_Africa_50..	721	SVTTEILPVSMTKTSVDCCTMYICGDSSTECNLLQYGSFCTQLNRAITGIAVEQDRNTOE	780
P0DTC2 SPIKE_SARS2	181	GKQGNFKNLRREFVFNIDGYFKIYKHTPINLVRDLQPGFSALEPLVDLPIGINITRFQT	240	P0DTC2 SPIKE_SARS2	781	VFAQVYKQIYKTPPIKDFGGNFESQILPDPKSKRSFIEDLLENKVLADAGFIKQYGGC	840
Spike_South_Africa_50..	181	GKQGNFKNLRREFVFNIDGYFKIYKHTPINLVRDLQPGFSALEPLVDLPIGINITRFQT	240	Spike_South_Africa_50..	781	VFAQVYKQIYKTPPIKDFGGNFESQILPDPKSKRSFIEDLLENKVLADAGFIKQYGGC	840
P0DTC2 SPIKE_SARS2	241	LLALHRSYLTFGDSSSGWTAGAAAYVGYLQPTFLFKYNENGTITDAVDCALDPLSETR	300	P0DTC2 SPIKE_SARS2	841	LGDIAARDLCAQKFNGLTLPPLLTDEMIQYTSALLAGTITSGWTFGAGALQIPFAM	900
Spike_South_Africa_50..	241	LLALHRSYLTFGDSSSGWTAGAAAYVGYLQPTFLFKYNENGTITDAVDCALDPLSETR	300	Spike_South_Africa_50..	841	LGDIAARDLCAQKFNGLTLPPLLTDEMIQYTSALLAGTITSGWTFGAGALQIPFAM	900
P0DTC2 SPIKE_SARS2	301	CTLRSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPGGEVFNATRFASVYAMNKRKISN	360	P0DTC2 SPIKE_SARS2	901	QMAVRFNGIGVTONVLYENKLIANQFNSAIGRIQDLSSTASALGKLDVNVNQAQALN	960
Spike_South_Africa_50..	301	CTLRSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPGGEVFNATRFASVYAMNKRKISN	360	Spike_South_Africa_50..	901	QMAVRFNGIGVTONVLYENKLIANQFNSAIGRIQDLSSTASALGKLDVNVNQAQALN	960
P0DTC2 SPIKE_SARS2	361	CVADYSVLYNSASFSTFKCYGVSPTKLNLDLCTNIVYADSFVIRGDEVVQIAGQTSIAD	420	P0DTC2 SPIKE_SARS2	961	ILVKQLSSNFGAISSVIMDLISRLDRVEAEVQIDRLITGRISLQTVVQQLIRAAEIRA	1020
Spike_South_Africa_50..	361	CVADYSVLYNSASFSTFKCYGVSPTKLNLDLCTNIVYADSFVIRGDEVVQIAGQTSIAD	420	Spike_South_Africa_50..	961	ILVKQLSSNFGAISSVIMDLISRLDRVEAEVQIDRLITGRISLQTVVQQLIRAAEIRA	1020
P0DTC2 SPIKE_SARS2	421	NYNKLDDPDTCCVIAWNSNLDKSVGGNYYLYRLFRKSNLKPFRDISEIYQAGSTPC	480	P0DTC2 SPIKE_SARS2	1021	SANLAATRMSECVLQSRVDFPCGRGYHLSMFPQSAPHGVVFLHVTYVPAQEKNTTAPA	1080
Spike_South_Africa_50..	421	NYNKLDDPDTCCVIAWNSNLDKSVGGNYYLYRLFRKSNLKPFRDISEIYQAGSTPC	480	Spike_South_Africa_50..	1021	SANLAATRMSECVLQSRVDFPCGRGYHLSMFPQSAPHGVVFLHVTYVPAQEKNTTAPA	1080
P0DTC2 SPIKE_SARS2	481	NGVLSFRNLPFLQSYVQPTINQVYRIVVVLSEFLLHAPAVCGPKKSTNLVKNKCVN	540	P0DTC2 SPIKE_SARS2	1081	ICHDKRAHFPRGQVFSNGTHWFVTRQNFYEPQIITDNTFVSG	1140
Spike_South_Africa_50..	481	NGVLSFRNLPFLQSYVQPTINQVYRIVVVLSEFLLHAPAVCGPKKSTNLVKNKCVN	540	Spike_South_Africa_50..	1081	ICHDKRAHFPRGQVFSNGTHWFVTRQNFYEPQIITDNTFVSG	1140
P0DTC2 SPIKE_SARS2	541	FNENGLTGTGVLTESNKRFLFPQQGRDIDDTTDAVRDPQTLLEILDITPCSGGGVSVITP	600	P0DTC2 SPIKE_SARS2	1201	QELGKYEQYIKWPIYIWLGFAGLIAIVMVTIMLCMTSCCSCG	1260
Spike_South_Africa_50..	541	FNENGLTGTGVLTESNKRFLFPQQGRDIDDTTDAVRDPQTLLEILDITPCSGGGVSVITP	600	Spike_South_Africa_50..	1201	QELGKYEQYIKWPIYIWLGFAGLIAIVMVTIMLCMTSCCSCG	1260
P0DTC2 SPIKE_SARS2	541	FNENGLTGTGVLTESNKRFLFPQQGRDIDDTTDAVRDPQTLLEILDITPCSGGGVSVITP	600	P0DTC2 SPIKE_SARS2	1261	SEFVLRGKVLHHT	1320
Spike_South_Africa_50..	541	FNENGLTGTGVLTESNKRFLFPQQGRDIDDTTDAVRDPQTLLEILDITPCSGGGVSVITP	600	Spike_South_Africa_50..	1261	SEFVLRGKVLHHT	1320

align@UniProt

Cette mutation semble spécifique au variant sud africain

Liste des 'variants Spike du virus d'Afrique du Sud' selon [UniProtKB](#): L18F, D80A, D215G, R246I, K417N, **E484K**, **N501Y**, A570D, A701V

Publication : <https://www.medrxiv.org/content/10.1101/2020.12.21.20248640v1>
"Eight lineage-defining mutations in the spike protein, including three at important residues in the receptor-binding domain (K417N) have functional significance."

Cette mutation est aussi retrouvée dans le variant anglais

Génome: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA694014>

4

Dernières nouvelles...

Le variant Brésilien (P1, P.1 ou P.1.1.28.1)

Les mutations Spike du variant brésilien P1:

L18F, T20N, P26S, D138Y, R190S, **K417(T ou N)**, **E484K**, **N501Y**, D614G, H655Y, T1027I, V1176F

(Philippe Le Mercier, UniProtKB)

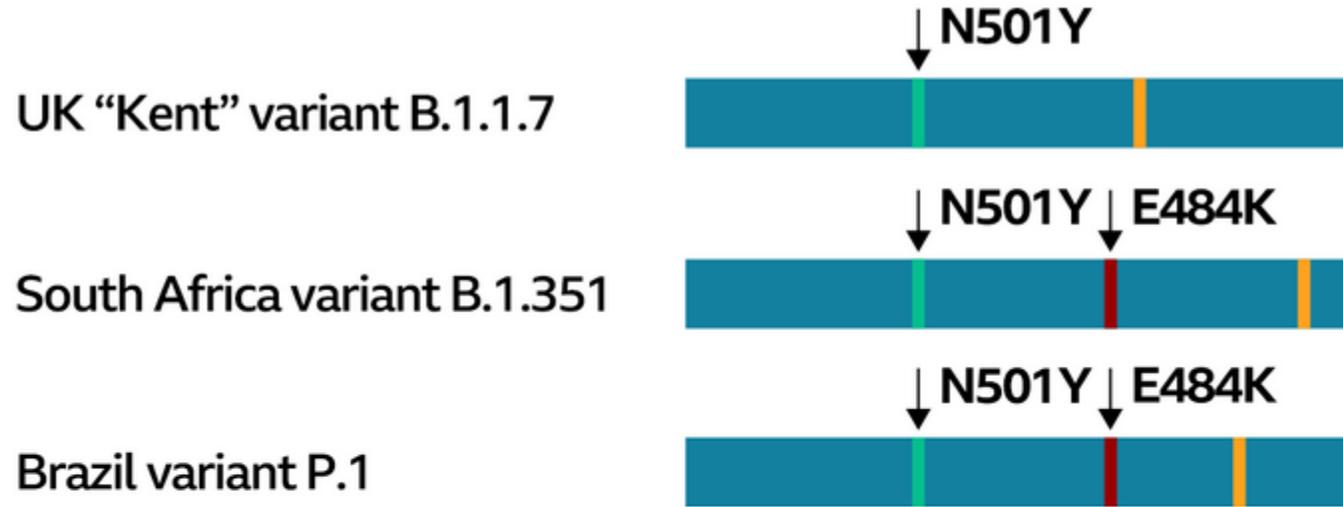
12 nonsynonymous mutations including K417T, E484K, and N501Y in the receptor-binding domain of the spike protein (variant P.1 for Pangolin variant 1).

Abstract: Cases of SARS-CoV-2 infection in Manaus, Brazil, resurged in late 2020, despite high levels of previous infection there. Through genome sequencing of viruses sampled in Manaus between November 2020 and January 2021, we identified the emergence and circulation of a novel SARS-CoV-2 variant of concern, lineage P.1, that acquired 17 mutations, including a trio in the spike protein (**K417T, E484K and N501Y**) associated with increased binding to the human ACE2 receptor. Molecular clock analysis shows that P.1 emergence occurred around early November 2020 and was preceded by a period of faster

https://wwwnc.cdc.gov/eid/article/27/4/21-0138_article

4

Dernières nouvelles...
Le variant Brésilien (P1, P.1 ou P.1.1.28.1)



K417T/N, E484K, and N501Y

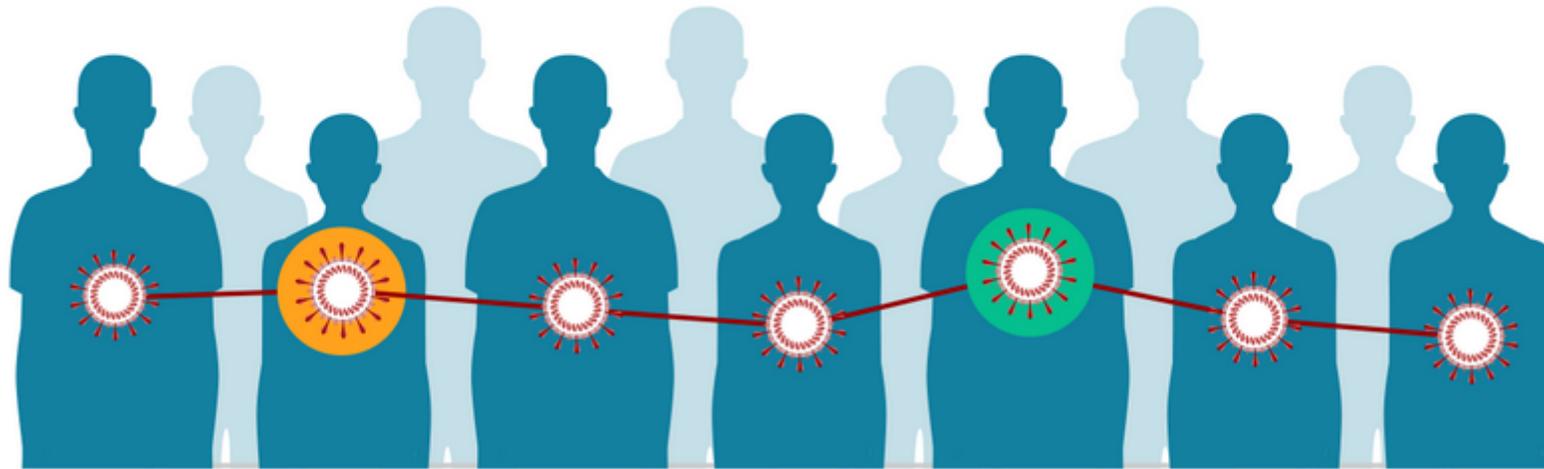
N501Y mutation seen in UK, South Africa and Brazil variants may help the virus spread more easily.

E484K mutation seen in South Africa, Brazil and some UK variants may affect the antibody response.

<https://www.bbc.com/news/health-55659820>

1 High numbers of cases increase risk of mutations

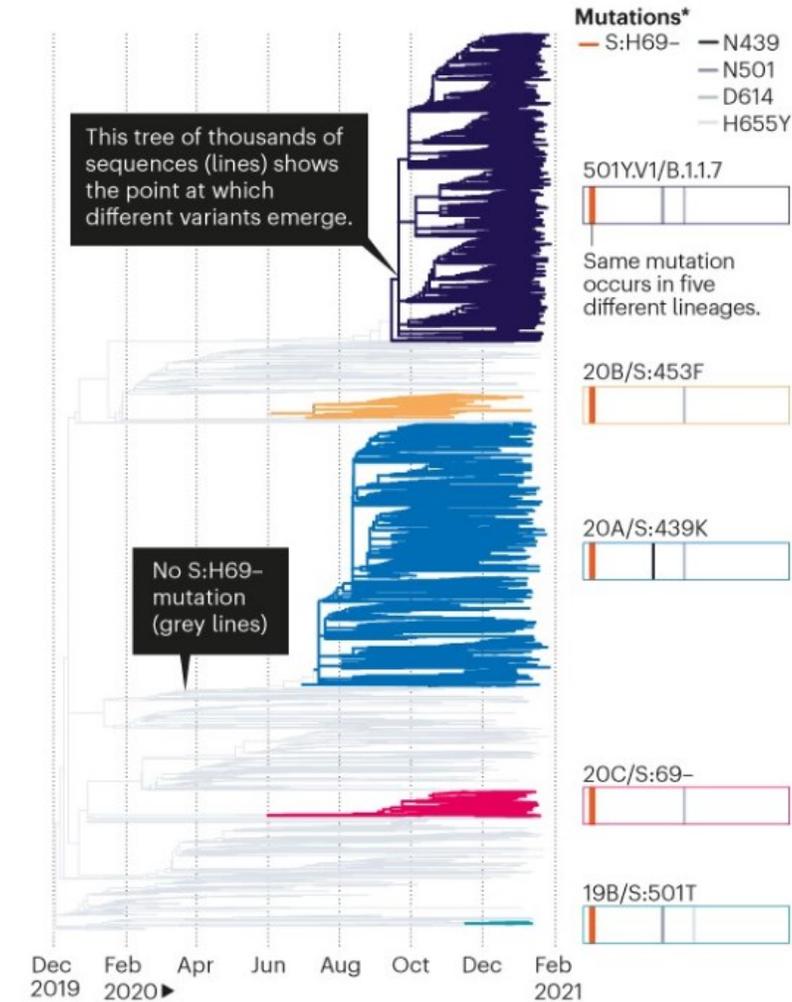
The more a virus spreads, the more chance it has to mutate. Thousands of small changes have been seen in coronavirus so far - most with little impact.



<https://www.bbc.com/news/health-55659820>

MUTATION WATCH

The same mutations are arising in different variants of SARS-CoV-2 and could affect virus spread, lethality or vaccine efficacy. The S:H69- mutation in the spike protein has occurred in different lineages of the tree (coloured lines), suggesting it might give the virus an evolutionary advantage — a key fact to consider when designing health policies and vaccines.



*Only a few mutations are shown for simplicity; this does not represent the full list.

©nature

As this article went to press, laboratories around the world had sequenced **more than 610,000 SARS-CoV-2 samples**; that number could well exceed one million by the end of the pandemic.

“Even in the United Kingdom, certain regions are better represented in sequencing data than are others. The tools we have to infer geographical spread from samples do not typically account for these biases. Without correction, the United Kingdom would be labelled a common source of variants just because of its high rate of sequencing, not because of how variants are spreading.”

“One of the three main variants of concern, 501Y.V1 (or B.1.1.7), has a deletion in the spike protein. This causes the diagnostic PCR test to fail to detect the spike gene — a problem called S drop-out. But 501Y.V1 is not the only variant that causes such drop-outs, so confirmation can come only from sequencing.”

Nature 591, 30-33 (2021)

doi: <https://doi.org/10.1038/d41586-021-00525-x>

Mutations et vaccins

L'idée est de vérifier que les mutations présentes dans la protéine Spike des virus 'variants' ne sont pas localisées dans des régions connues ou prédites pour être reconnues par le système immunitaire (épitopes)

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7571903/pdf/main.pdf>

Table 2
Predicted epitopes for spike protein of SARS-CoV-2 (aa319-541).

Allele	start	end	length	Sequence	score	rank
H2-K ^d	350	358	9	VYAWNRRKI	0.745883	0.04
H2-K ^d	504	512	9	GYQPYRVVV	0.740992	0.04
H2-D ^d	525	533	9	CGPKKSTNL	0.537362	0.02
H2-D ^d	382	390	9	VSPKLNLDL	0.499642	0.02
H2-K ^d	379	387	9	CYGVSPTKL	0.498969	0.14
H2-K ^d	494	503	10	SYGFQPTNGV	0.457492	0.17
H2-D ^d	524	533	10	VCGPKKSTNL	0.432674	0.03
H2-K ^d	327	335	9	VRFNITNL	0.3807	0.22
H2-K ^d	395	402	8	VYADSFVI	0.372275	0.22
H2-D ^d	505	513	9	YQPYRVVVV	0.35273	0.06
H2-K ^d	504	513	10	GYQPYRVVV	0.339094	0.26
H2-K ^d	328	335	8	RFPNITNL	0.319452	0.28
H2-K ^d	399	407	9	SFVRGDEV	0.319166	0.28
H2-D ^d	503	511	9	VGYQPYRVV	0.313391	0.1
H2-D ^d	503	513	11	VGYQPYRVVV	0.303071	0.11
H2-K ^d	378	387	10	KCYGVSPTKL	0.265013	0.36
H2-D ^d	369	377	9	YNSASFSTF	0.245941	0.17
H2-D ^d	410	418	9	IAPGQTGKI	0.242132	0.18
H2-D ^d	445	452	8	VCGNYYNL	0.233027	0.19
H2-D ^d	523	533	11	TVCCPKKSTNL	0.213929	0.24
H2-K ^d	472	480	9	IYQAGSTPC	0.204669	0.48
H2-D ^d	327	335	9	VRFNITNL	0.20042	0.28
H2-D ^d	522	533	12	ATVCGPKKSTNL	0.188822	0.32
H2-D ^d	328	335	8	RFPNITNL	0.156347	0.45
H2-D ^d	520	533	14	APATVCGPKKSTNL	0.155495	0.46
H2-D ^d	444	452	9	KVGGNYYNL	0.153646	0.47
H2-D ^d	370	377	8	NSASFSTF	0.152266	0.48
H2-D ^d	503	510	8	VGYQPYRV	0.151477	0.49

<http://covid19.datamonkey.org>

 CTL EPITOPES OVERLAPPING THIS SITE

YGFQPTNGV (262969) 

HLA-C*12:32, HLA-C*03:76, HLA-C*03:55, HLA-C*03:57, HLA-C*03:79

FQPTNGVGY (262981) 

HLA-B*15:103, HLA-B*15:127, HLA-B*15:132, HLA-B*15:156

- Un site pour découvrir les virus: <https://viropourtous.ch/>