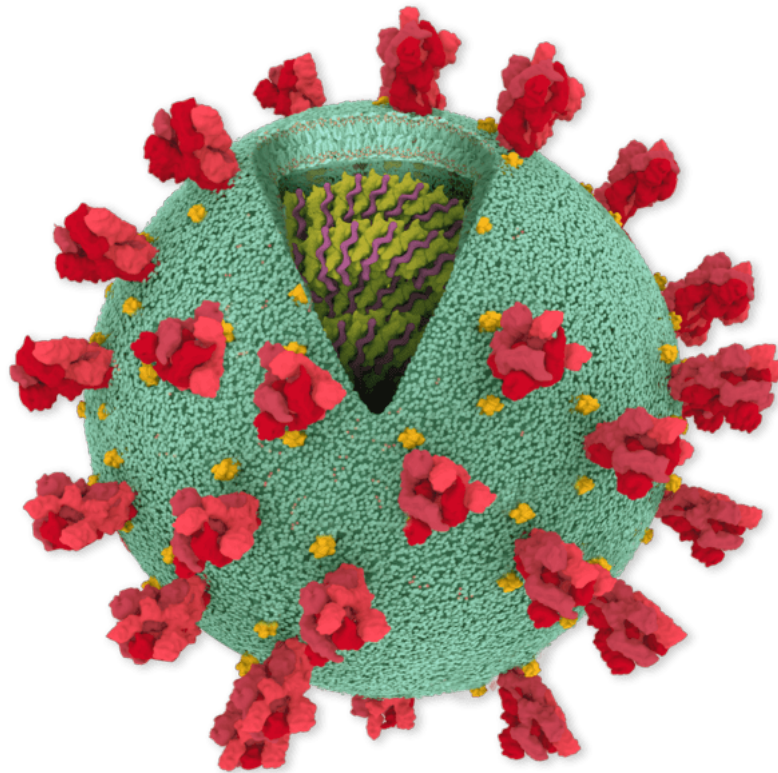


Bad News Wrapped in Protein: Inside the Coronavirus Genome

By Jonathan Corum and Carl Zimmer April 3, 2020



A virus is “simply a piece of bad news wrapped up in protein,” the biologists Jean and Peter Medawar wrote in 1977.

In January, scientists deciphered a piece of very bad news: the genome of SARS-CoV-2, the virus that causes Covid-19. The sample came from a 41-year-old man who worked at the seafood market in Wuhan where the first cluster of cases appeared.

Researchers are now racing to make sense of this viral recipe, which could inspire drugs, vaccines and other tools to fight the ongoing pandemic.

A String of RNA

Viruses must hijack living cells to replicate and spread. When the coronavirus finds a suitable cell, it injects a strand of RNA that contains the entire coronavirus genome.

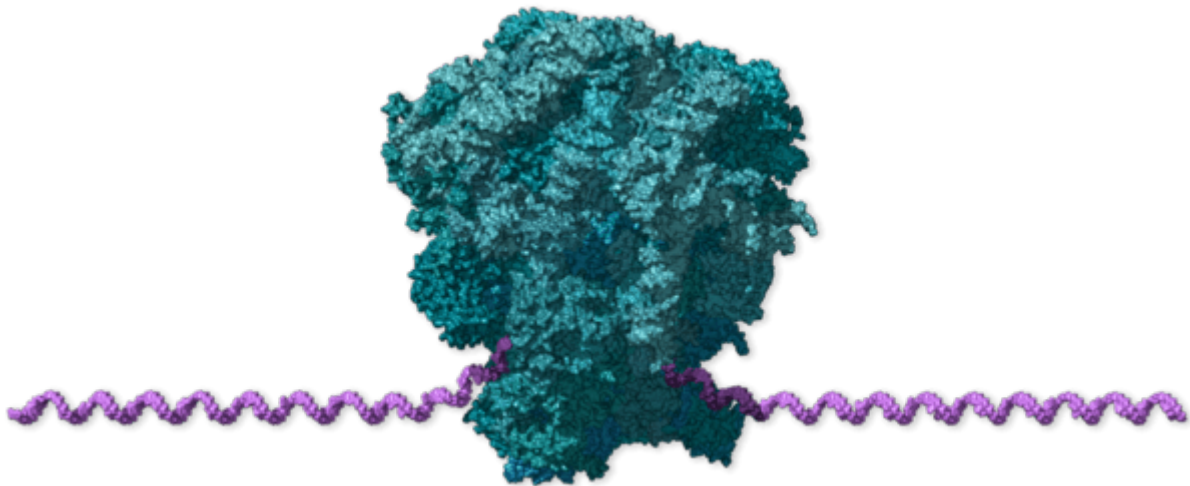


The genome of the new coronavirus is less than 30,000 “letters” long. (The human genome is over 3 billion.) Scientists have identified genes for as many as 29 proteins, which carry out a range of jobs from making copies of the coronavirus to suppressing the body’s immune responses.

The first sequence of RNA letters reads:

```
auuaaagguuuauaccuucccagguaacaaaccaaccaacuucgaucucuuguagaucuguucucuaa
acgaacuuuaaaauucuguguggcugucacucggcugcaugcuuagugcacucacgcagauaaauuaa
acuaauuacugucguugacaggacacgaguaacucgucuauucugcaggcugcuuacgguuucguccgug
uugcagccgaucaucagcacaucauagguuucguccgggugugaccgaaagguaag
```

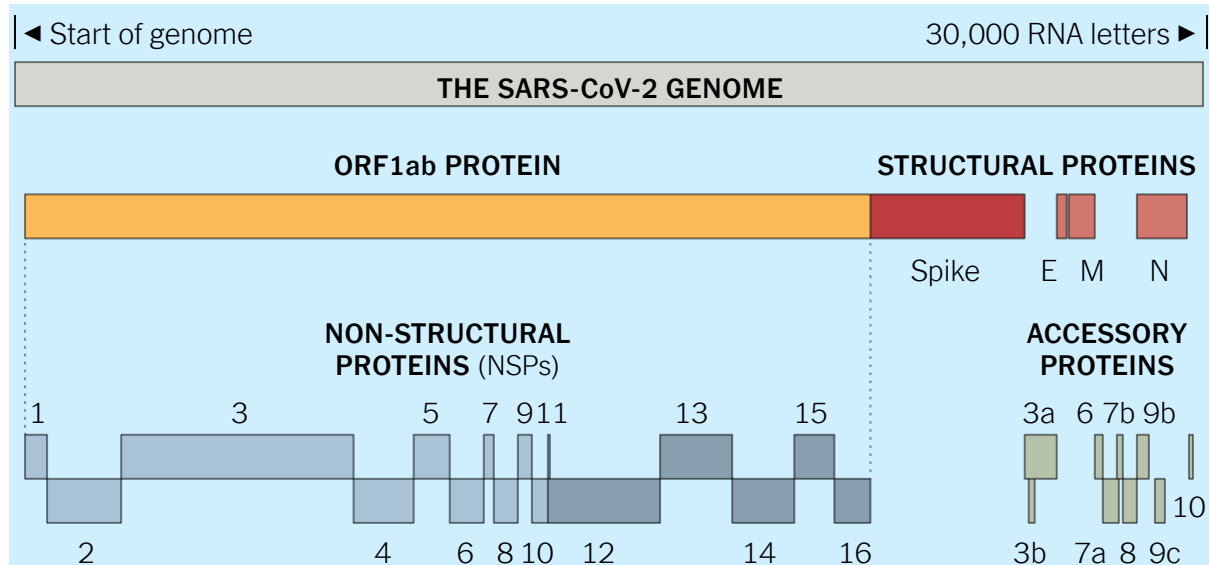
This sequence recruits machinery inside the infected cell to read the RNA letters — a, c, g and u — and translate them into coronavirus proteins.



The full coronavirus genome and the proteins it encodes are shown below.

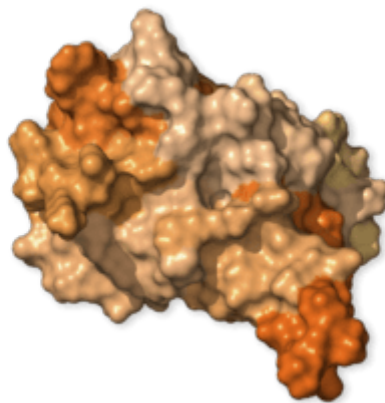
A Chain of Proteins · ORF1ab

The first viral protein created inside the infected cell is actually a chain of 16 proteins joined together. Two of these proteins act like scissors, snipping the links between the different proteins and freeing them to do their jobs.



Research on other coronaviruses has given scientists a good understanding of what some of the SARS-CoV-2 proteins do. But other proteins are far more mysterious, and some might do nothing at all.

Cellular Saboteur · NSP1



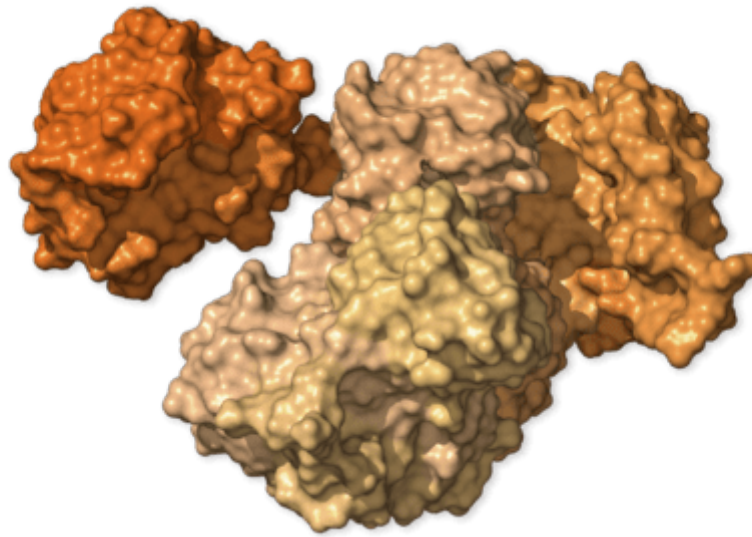
This protein slows down the infected cell's production of its own proteins. This sabotage forces the cell to make more virus proteins and prevents it from assembling antiviral proteins that could stop the virus.

```

auggagagccuuguccugguuuaacagagaaaacacaguccaacucaguugccuguuuuacagguucgc
gacgugcucguacguggcuuuggagacuccguggaggaggucuuauacagaggcacgucaacaucuuaaagau
ggcacuugugggcuuaguagaaguugaaaaaggcguuuugccucaacuugaacagcccuauguguucauca
aacguucggaugcucgaacugcaccucauggucauguuaugguugagcugguagcagaacucgaaggcauuc
aguacggucguaguggugagacacuugguguccuuguccucaugugggcgaaauaccaguggcuuaccgca
agguucuucucguaagaacgguaauaaaggagcugguggccauaguuaacggcgccgaucuaaaagucauu
ugacuuaggcgacgagcuuggcacugauccuuaugaagaauuucaagaaaacuggaacacuaaacauagcag
ugguguuacccgugaacucaugcgugagcuaaacggaggg

```

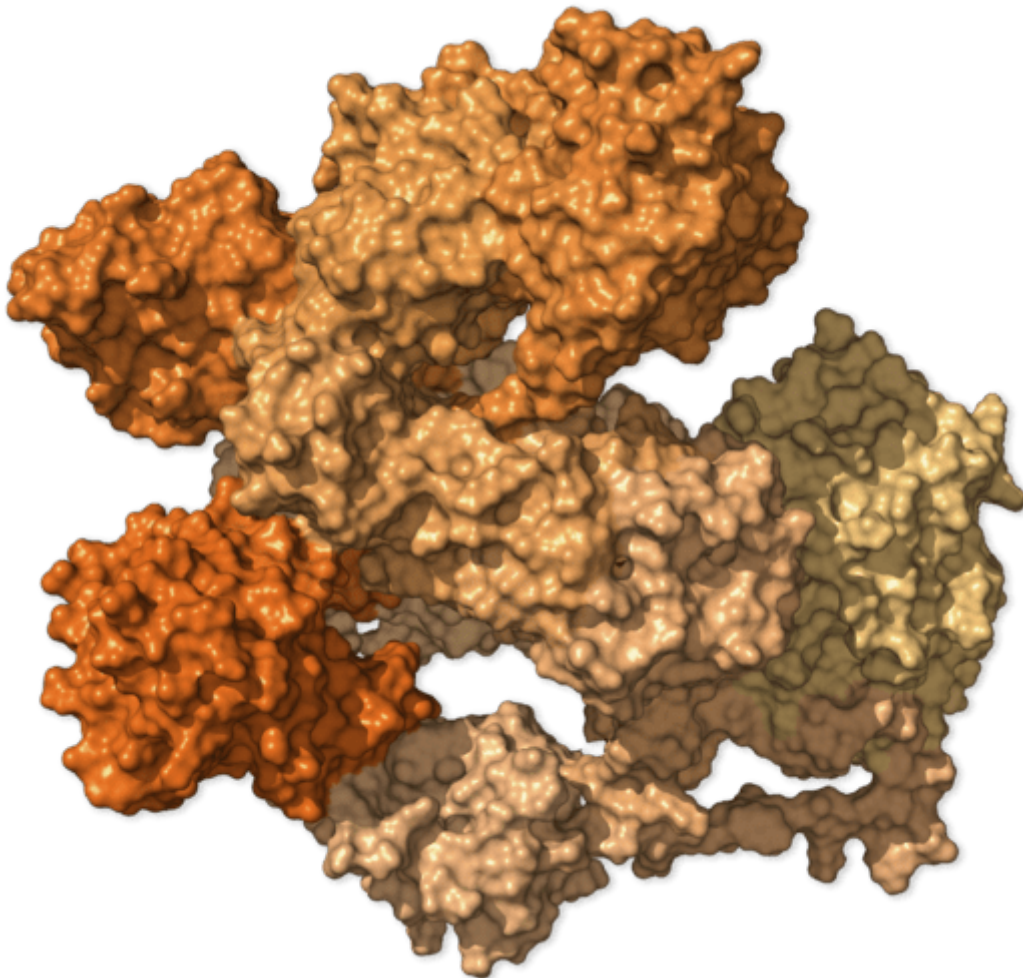
Mystery Protein · NSP2



Scientists aren't sure what NSP2 does. The other proteins it attaches to may offer some clues. Two of them help move molecule-filled bubbles called endosomes around the cell.

```
gcuaacacucgcuaugucgauaacaacuucuguggccugauggcuaccucucugagugcauuaaagaccuu
cuagcacgugcugguaaagcuucaugcacuuuguccgaacaacuggacuuuuuugacacuaagaggggugua
uacugcugccgugaacaugagcaugaaaauugcuugguacacggaacguucugaaaagagcaugaauugcag
acaccuuuugaaaauuuuuuuggcaaagaaaauuugacaccuucuuuuggggauguccaaaauuuugauuuu
ccuuuuuuuuccauaaucagaacuaaucaaccaagggguugaaaagaaaagcuugauggcuuuuugggguaga
auucgaucugucuauccaguugcgucaccaaauugaauucaaccaaaugugccuuuuaacucucaugaagugu
gaucuuugugguuagaaacuucauggcagacgggcgauuuuuguuuuuagccacuugcgaaauuuugggcacugag
aaauugacuaaaagaaggugccacuacuuugguuacuuaccccaaaugcuguuguuuuuuuuuuuuguc
cagcaugucacaaauucagaaguaggaccugagcauagucuuugccgaauaccuauaauagaauucuggcuuga
aaaccauucuuucguaagggugugcgcacuaauugccuuugggaggcugugugucucuuauuguugguugccaua
acaagugugccuauuugguuuccacgugcuagcgcuaacauagguuguaaccauacagguguuugggaga
agguuccgaaggucuuuauugacaaccuucuuugaaauacuccaaaaagagaaagucacaucaauuuugu
uggugacuuuuuuaacuuuauagaagagauugccauuuuuuuggaucuuuuuucugcuuccacaagugcuuu
uguggaaaacugugaaagguuuggaauuauaaagcauucacaaauuuuuguaauccugugguuuuuuuuu
aguuacaaaaggaagcuuuuuuagguugccuggaauuuuugguagaacagaaaucauacugaguccucuu
uauugcauuuugcaucagaggcugcuguguuuugacgaucuuuuuucuccgcacucuuuugaaacugcucaa
aaucucugugcuguuuuuacagaaggccgcuaauaacaauacuagauggaauuuucacaguauucacugagacuc
auugaugcuauagauguuacacucugauuuuggcuaucuaacaaucuaguuguaauugggccuacauuacagguggu
guuguucaguugacuucgcaguggcuaacuaacaucuuuuggcacuguuuauugaaaaacucaaaaccgguc
cuugauuuggcuuugaagagaaguuuuuaggaagguguagaguuuucuuagagacgguuugggaaaauuguuuuuuu
uauucuaaccugugcuugugaaauugucggugggacaaauugucaccugugcaaaggaauuuuagggagagugu
ucagacauucuuuuuagcuuguaaaauuuuuuuggcuuugugugcugacucuaucuuuuuugggaggcuu
aacuuuuuagccuugaauuuuagguuagaaacuuuugucacgcacuaaaagggauuguacagaaaguguguuu
aauccagagaagaaacuggccuacucaugccucuaaaagccccaaaagaaauuauucuuuagaggggaga
aacacuuccacagaaguguuuacagaggaaguugucuuugaaaacuggugauuuuacaaccauuagaaca
accuacuagugaagcuguuugaagcuccauugguugguacaccaguuuuuguaauuacgggcuuauuguugcucga
aaucuaagacacagaaaaguacugugcccuugcaccuauuauugaugguuacaaacaaauaccuucacacuca
aagcggu
```

Untagging and Cutting · NSP3



NSP3 is a large protein that has two important jobs. One is cutting loose other viral proteins so they can do their own tasks. It also alters many of the infected cell's proteins.

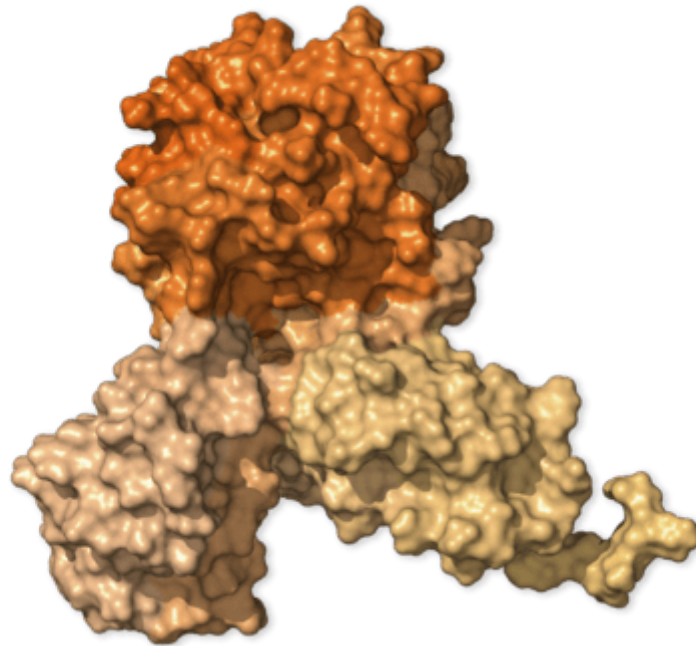
Normally, a healthy cell tags old proteins for destruction. But the coronavirus can remove those tags, changing the balance of proteins and possibly reducing the cell's ability to fight the virus.

```
gcaccaacaaagguuacuuiuuggugaugacacugugauagaagugcaagguuacaagagugugaauaucacu
uuugaacuugaugaaaggauugauaaaguacuuaaagagaagugcugccuauacaguugaacucgguaca
gaaguaaaugaguucgccuguguuugggcagaugcugucauaaaaacuugcaaccagauaucugaauuacu
uacaccacugggcauugauuuagaugaguggaguauggcuacauacuauuuuugaugagucuggugagu
uuaaaauuggcuucacauauguauuuguucuuucuaaccuccagaugaggauagaagaaggugauuguga
agaagaagaguuuugagccaucaacucaauaugaguaugguacugaagaugauuaccaagguaaaccuuugga
auuuggugccacuucugcugcucucaaccugaagaagagcaagaagaagauugguuagaugaugauaguca
acaaacuguuggucaacaagacggcagugaggacaaucaagacaacuacuauucaaacaauuguugaggu
ucaaccucaauuagagauggaacuuacaccaguuguucagacuauugaagugaauaguuuuagugguuuuu
uaaaacuucugacaauguauacauuaaaaaugcagacauuguggaagaagcuaaaaagguaaaaaccaacag
ugguuguuaaauugcagccaauguuuuaccuuuaaacauggaggagguguugcaggagccuuuuuagggcuacua
acaauugccaugcaaguugaauucugaugauuacauagcuacuaauaggaccacuuuuagugggugguaguugug
uuuuuagcggacacaaucuugcuuaaacacugucuucauguugucggcccaauguuaacaaaggugaagaca
uucaacuucuuuagagugcuuaugaaaauuuuuauucagcacgaaguucacuugcaccuuuuuauucagcug
```


<https://www.nytimes.com/interactive/2020/04/03/science/coronavirus-genome-bad-news-wrapped-in-protein.html?action=click&module=Well&pg...> 6/27

uggugcuuguauugacuguaugucgcgucauuuauugcgcagguagcaaaaagucacaacauugcuuugau
auggaacguuaaagauuucaugucauugucugaacaacuacgaaaacaaauacguagugcugcuaaaaaga
auaacuuaccuuuuuagguagacaugugcaacuacuagacaaguuguuaauguuguaacaacaaagauagcac
uuaggguggu

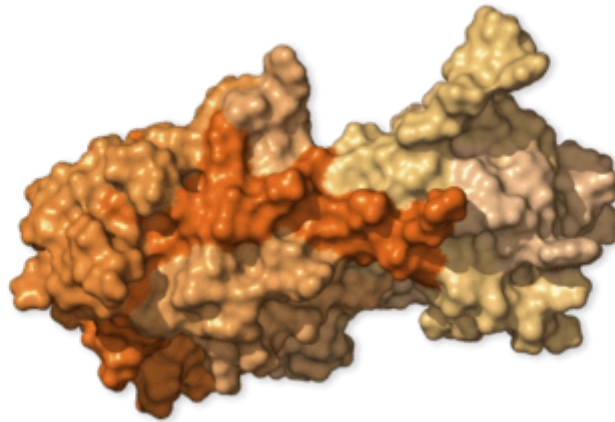
Bubble Maker · NSP4



Combining with other proteins, NSP4 helps build fluid-filled bubbles within infected cells. Inside these bubbles, parts for new copies of the virus are constructed.

[illegible]

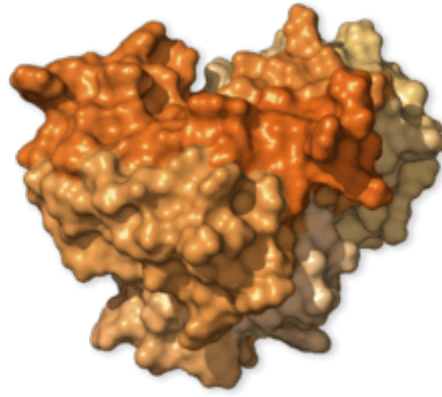
Protein Scissors · NSP5



This protein makes most of the cuts that free other NSP proteins to carry out their own jobs.

```
agugguuuuagaaaauggcauucccaucugguaaaguugagggguuguaugguacaaguaacuugugguaca
acuacacuuuacggucuuuggcuugaugacguaguuuacuguccaagacaugugaucugcaccucugaagac
augcuuaacccuaauuaugaagauuuacucauucguaagucuaaucauaauuucuuugguacaggcuggua
auguucacucaggguuauuggacauucuaugcaaaaauuguguacuuaagcuuaagguugauacagcca
auccuaagacaccuaagauuaaguuuguucgcauucacaccaggacagacuuuuucaguguuagcuuguuaca
augguucaccaucugguguuuaccaaugugcuauagaggcccauuuucacuaauaaggguucauuccuua
augguucaugugguaguguugguuuuacauagauuugacugugucuuuuuuguuacaugcaccuauug
gaauuaccaacuggaguucaugcuggcacagacuuagaagguaacuuuuuaggaccuuuuuguugacaggca
aacagcacaagcagcugguuacggacacacuaauuacaguuuauuguuuuagcuugguuguacgcugcugu
uauaaauggagacaggugguuucucaauucgauuuuaccacaacucuuuauugacuuuuuacuuugggcuaua
aguacaauuauagaaccucuaacacaagaccauguugacauacuaggaccuuuucugcucaaacuggaauug
ccguuuuagauaugugugcuucauuuuuagaauuacugcaaaaugguuagaauggacguaccauauugg
guagugcuuuuuagaagaugaauuuacaccuuuugauguuguuagacaauugcucagguguuacuucca
a
```

Bubble Factory · NSP6



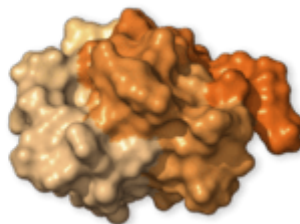
Works with NSP3 and NSP4 to make virus factory bubbles.

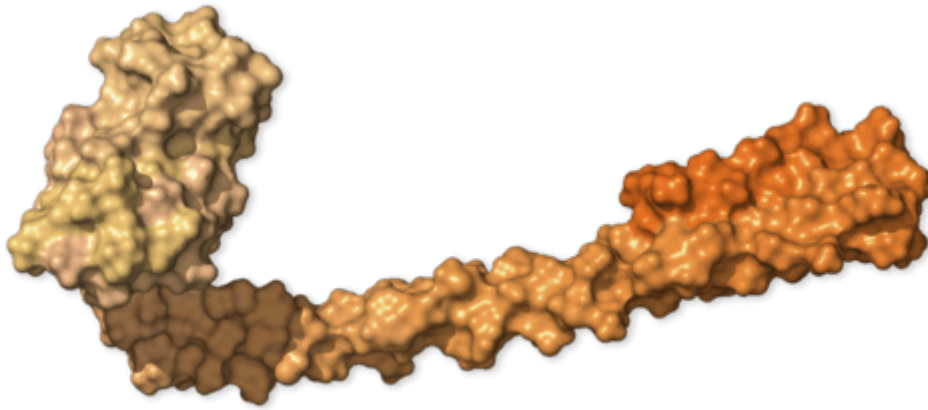
```

agugcagugaaaagaacaaucaagggguacacaccacugguuguuacucacaauuuugacuucacuuuuagu
uuuaguccagaguacucaauaggucuuuguucuuuuuuuuuguaugaaaugccuuuuuaccuuuugcuau
gguaauuauugcuauugucugcuuuugcaaugauguuugucaaacauaagcaugcauuucucuguuuguuu
uuguuaccuucucugccacugugcuuauuuuuauauggucuaauagccugcuaguugggugaugcguaau
augacaugguuggauaugguugauacuaguuugucugguuuuuuagcuaaaagacuguguuuauugauugcauca
gcuguaguguuacuaauccuuauagacagcaagaacuguguaugaugaugggugcuaggagaguguggacacu
uaugaauugucuugacacucguuuuauaaaguuuuauaugguaaugcuuuagaucaagccauuuccaugugg
gcucuuaauaauucucuguuacuucuaacuacucagguguaguuaacaacugucauguuuuuggccagagguuu
guuuuuauuguguguugaguauugcccuauuuuucuucauaacugguaauacacuucaguguauaaugcuagu
uuauuguuuucuuaggcuauuuuuuguacuuguuacuuuggccucuuuuuguuuacucaaccgcuacuuuagacu
gacucuugguguuuuaugauuacuuaguuuucacacaggaguuuagauauauggaauucacagggacuacuc
ccaccaagaauagcauagaugccuucacacucaacauuuuuuuguuggguguugggucacacuuuguauuc
aaaguagccacuguacag

```

Copy Assistants · NSP7 and NSP8



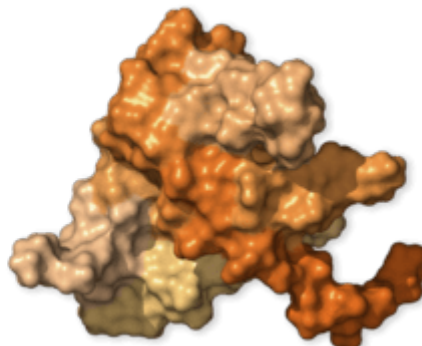


These two proteins help NSP12 make new copies of the RNA genome, which can ultimately end up inside new viruses.

```
ucuaaaaugucagauguaaagugcacaucauguagucuuacucucaguuuugcaacaacucagaguagaauca
ucaucuaaaauugugggcuaauguguccaguuacacaugacauucucuuagcuaaagauacuacugaagc
cuuugaaaaaaugguuucacuacuuucuguuuuugcuuuccaugcaggugcuguagacauaaacaagcuu
ugugaagaaaugcuggacaacagggaaccuuacaa
```

```
gcuauagccucagaguuuaguucccuuccaucauauagcagcuuuugcuacugcucaagaagcuuauagagcag
gcuguugcuaauggugauucugaaguuguucuuaaaaaguugaagaagucuuugaauguggcuaaaucuga
auuugaccgugaugcagccaugcaacguaaguuggaaaagauggcugaucuaagcuauagacccaaauguaua
aacaggcuagaucugaggacaagagggcaaaaguuaacuagugcuauagcagacaauugcuuuucacuauugcuua
gaaaguuggauaaugaugcacucaacaacauuaucacaauagcaagagaugguuguguucccuugaacaua
auaccucuuaacaacagcagccaaacuaaugguugucuaaccagacuaauaacacauaauaaaaauacgugugau
gguacaacaauuuacuuauugcaucagcauugugggaaauccaacagguuguagaugcagauaguaaaaauugu
ucaacuuaugugaaaauuaguauggacaauucaccuaauuuagcauggccucuuaauuguaacagcuuuuagg
gccaaucugcugucaaaauacag
```

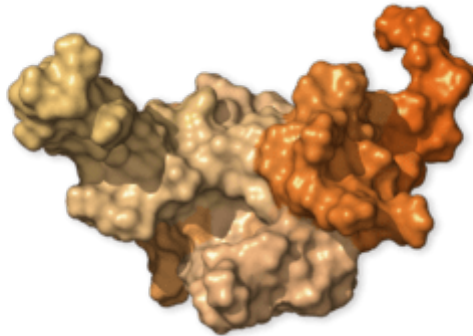
At the Heart of the Cell · NSP9



This protein infiltrates tiny channels in the infected cell's nucleus, which holds our own genome. It may be able to influence the movement of molecules in and out of the nucleus — but for what purpose, no one knows.

aaauaugagcuuaguccugugcacuacgacagaugucuugugcugccgguacuacacaaacugcuugcacu
 gaugacaauugcguaagcuuacuacaacacacaaaggaggguagguuuuguacuugcacuguaucggaau
 uacaggauuugaaaugggcuagauucccuaagagugauggaacugguacuauacagaacuggaaccac
 cuuguagguuuuguuacagacacaccuaaagguccuaaagugaaguauuuauacuuuauuaaaggauuaa
 acaaccuaaaauagagguagguacuugguaguuuagcugccacaguacgucuacaa

Genetic Camouflage · NSP10



Human cells have antiviral proteins that find viral RNA and shred it. This protein works with NSP16 to camouflage the virus's genes so that they don't get attacked.

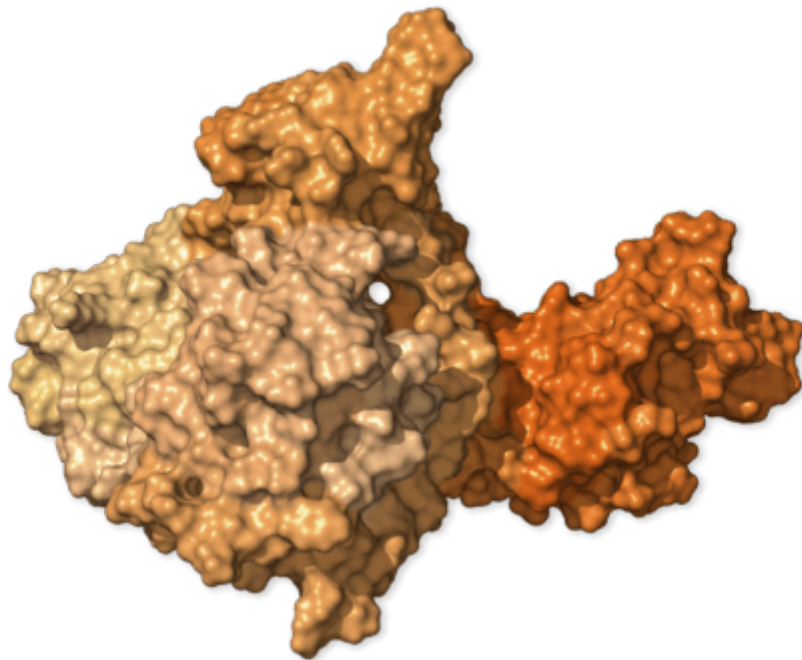
gcugguaaugcaacagaagugccugccaaaucaacuguaauaauuuucugugcuuuugcuguagauugcugcu
 aaagcuuacaaagauuaucuaagcuaguggggggacaaccaaucacuaauuguguuaagauguuguguacacac
 acugguacuggucaggcaauaacaguuacaccggaagccaauauggaucaagaauccuuugguggugcaucg
 uguugucuguacugccguugccacauagaucauccaaauccuaaaggauuuugugacuuaaaagguaaguau
 guacaaauaccuacaacuugugcuaaugaccugugggguuuuacacuuaaaaacacagucuguaccgucugc
 gguauguggaaagguuauggcuguaguugugaucaacuccgcgaacccaugcuucag

Copy Machine · NSP12

caugccuaaacaugcuuagaauuauggccucacuuguucucgcaaacauacaacguguuguagcuuguc
 acaccguuucuaauagauuagcuaaagagugugcuaagauuagagugaaauggucauguguggcgguucacu
 auauguuaaaccagguggaaccucaucaggagaugccacaacugcuuaugcuaauaguguuuuaacauu
 ugucaagcugucacggccaauuuuauugcacuuuuauacuagauugguaacaaaauugccgauaaguauugc
 cgcauuuacaacacagacuuuauagugucucuauagaaauagagauguugacacagacuuugugaauag
 uuuuacgcauuuugcguaaacuuucuaaugaugauacucucugacgaugcuguuguguguuucuuuagc
 acuuuugcaucuaaggucuaugugcuaagcauaaagaacuuuaagucaguucuuuauuaucaaaacaauu
 uuuuauugcugaagcaaaauuggacugagacugaccuuacuaaaggaccucaugaauuuugcucuca
 acauacaauugcuaguuuaacaggguagaauuauuguguaaccuuccuuaaccagauccaucaagaauccuagg
 ggccggcuguuuuugaugaauaucguaaaaacagaugguacacuuuauugaacggguucgugucuuuagc
 uauagaugcuuaccacuuacuaaacaucuaaagcaggaugcugaugucuuuauuugacuuaaca
 auacauaagaaagcuacaugaugaguuuacaggacacauuuagacauugaauucuguuauugcuuacuauga
 uaacacuuaagguauugggaaccugaguuuuauaggcuauguacacaccgcauacagucuuacag

Another sequence, NSP11, overlaps part of the same stretch of RNA. But it's not clear if the tiny protein encoded by this gene has any function at all.

Unwinding RNA · NSP13

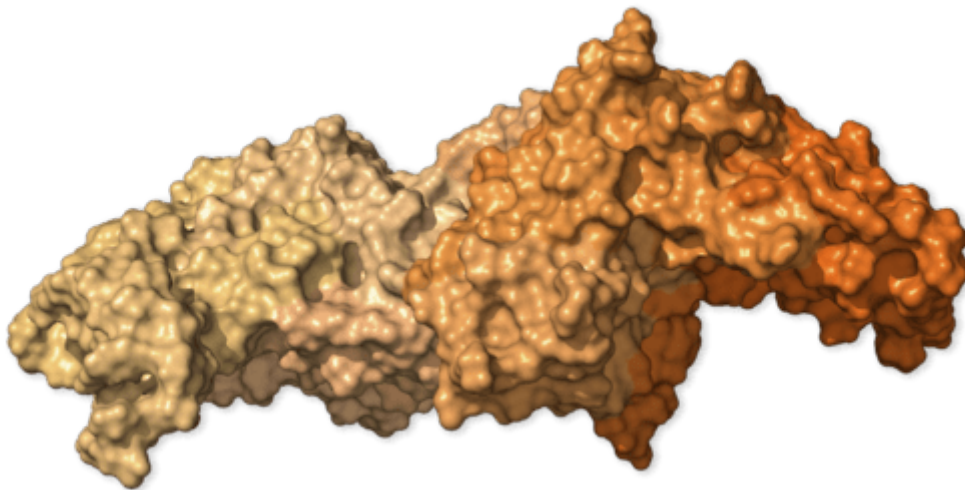


Normally, virus RNA is wound into intricate twists and turns. Scientists suspect that NSP13 unwinds it so that other proteins can read its sequence and make new copies.

gcuguuggggcuuguguucuuugcaauucacagacuucuuuagauguggugcuugcuaucguagaccuuc
 uuauuguuaaaugcuguuacgaccaugucauaucacacauaaaauagucuuugcuguuaauccguau
 guuugcaaugcuccagguugugaugucacagaugugacuaacuuuacuuaggagguagagcuauuauugu
 aaucacauaaaccaccauuaguuuuuccauugugucuaauggacaaguuuuugguuuauauaaaaauaca
 uguguugguagcgauaauuguuacugacuuaaauugcaauugcaacauugacuggacaaaugcuggugauuac
 auuuuagcuaacaccuguacugaaagacuaagcuuuuugcagcagaaacgcuaaagcuacugaggagaca
 uuuaaacugucuuuagguauugcuacuguaagugcugucugacagagaauuacauucuuuauagg
 gaaguugguaaaccuagaccaccacuuuacccgaaauuauugcuuuuacugguuauugcuuaacuaaaaaacagu

aaaguacaaauaggagaguacaccuuugaaaaaggugacuauaggugaugcuguuguuuaccgagguaca
 acaacuuaacaaauaaauguuggugauuuuuugugcugacaucauacaguaauggccauuaagugcaccu
 acacuagugccacaaagagcacuauguuagaauuacuggcuuauacccaacacucaauaucucagaugagu
 uuucuaagcaauguugcaaaauuacaaaagguugguauugcaaaaguuuacacuccaggggaccaccug
 guacugguaagagucuuuuugcuauuggccuagcucucuaacccuucugcucgcauaguguaucagcuu
 gcucucaugccgcuguugaugcacuaugugagaaggcauuuuuuuuugccuauagauaaauguaguaga
 auuuauaccugcacgugcucguguagaguguuuugauaaaaucaaaagugaauucaacauuagaacaguaugc
 uuuuguacuguaaaugcauugccugagacgacagcagauauaguugcuuugaugaaauucaauggccaca
 aauuaugauuuagaguguugucuaugccagauuacgugcuaagcacuauguguaacauuggcgaccugcuca
 auuaccugcaccacgcacauugcuaacuaagggcacacuaagaaccagaauuuucaaauucaguguguagacu
 uauaaaaacuauagguccagacauguuccucggaacuugucggcguuguccugcugaaauuguugacacugu
 gagugcuuugguuuuauagaauaaagcuuaaagcacauaaagacaaauacagcucaauugcuuuuuuuuuuu
 uuauaaggguuuuauacacgcaugauguuuuacucugcauuuacaggccacaaauaggcgugguaagaga
 auuccuuacacguaaccucugcuuggagaaaagcugucuuuuuuuacccuuauuuuacagaaugcuguagc
 cucaaagaauuuugggacuaccaacucaaaacuguuugauuacacagggcucagaauaugacuauugcauuu
 cacucaaaaccacugaaacagcucacucuuuguaaanguaaacagaauuuuauuguugcuauuaccagagcaaa
 aguaggcauacuuugcauauaugucugauagagaccuuuauagacaaguugcauuuacaagucuuugaaauuc
 cacguaggaaugugggcaacuuuacaa

Viral Proofreader · NSP14

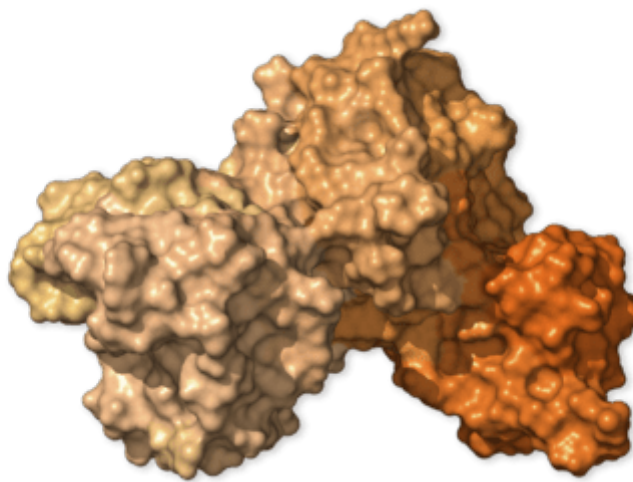


As NSP12 duplicates the coronavirus genome, it sometimes adds a wrong letter to the new copy. NSP14 cuts out these errors, so that the correct letter can be added instead.

gcugaaaauguaacaggacucuuuuuagauuguaguaagguaauacacuggguuacauccuacacaggcaccu
 acacaccucaguguugacacuaaaaucaaaacugaagguuuuauuguguuagacauaccuggcauaccuaaggac
 augaccuauagaagacucaucucuaugauggguuuuuuuuauagaauuaucaaguuuauugguuaccuuaacau
 uuuaucacccgcgaagaagcuauaagacauguacgugcauggauuggcuucgaugucgaggggugucaugcu
 acuagagaagcuguuugguaccauuuacuuuacagcuagguuuuuuacagguuguuaaccuaguugcugua
 ccuacagguuauuguuagauacaccuaauaauacagauuuuuuccagaguuaugugcuuaaaccaccgcccuggagau
 cauuuuuuaacaccucauaccacuuauguacaaaggacuuccuuggaauuguagugcguauaaagauuuguaca
 aauguuaagugacacacuuuuuuuucucucugacagagucguauuuugucuuauugggcacauuggcuuugagu
 ugacaucuauagaaguauuuuugugaaaaauaggaccugagcgcaccuguuugucuaugugauagacgugccacau
 gcuuuuuccacugcuucagacacuuauugccuguuuggcauauucuaauuggauuugauuacgucuaauaauuc
 cguuuaugauguuaucaacaauuggguuuuacagguaaccuacaaagcaaccaugaucuguaauuguca
 aguccaugguaaagcacauguagcuaguugugaugcaaucaugacuaggugucuaugcuguccacgagugcuu
 uguuaagcguguugacuggacuaauugaauuauccuauaauuggugaugaacugaagaauuauugcggcuuguag

aaagguucaacacaugguuguuaaagcugcauuauuagcagacaaaauucccaguucuuucacgacauuggua
accuuaaagcuauuaaguguguaccucaagcugauguagaauuggaaguucuaugaugcacagccuuguagug
acaaagcuuaauaaaauagaagaauuaauucuaauucuuaugccacacauucugacaaauucacagauugguguau
gccuauuuuuggaaauugcaaugucgauagauauccugcuauuuccauuguuuguagauuugacacuagagugc
uaucuuaaccuuaacuugccugguugugaugguggcaguuuuguanguaaaauaaacaugcauuccacacaccag
uuuuugauaaaagugcuuuuuguuaauuuuaaaacaauuaccuuuuuucuuuacucugacaguccaugugagu
cucaugggaaaacaaguagugucagauauagauuauguaccacuaaagucugcuacguguauaacacguugca
auuuagguggugcugucuguagacaucaugcuauaugaguacagauuguauucugaugcuuauaacaugauga
ucucagcuggcuuuagcuuguggguuuacaaacaauuugauacuuuaaaccucuggaacacuuuuacaagac
uucag

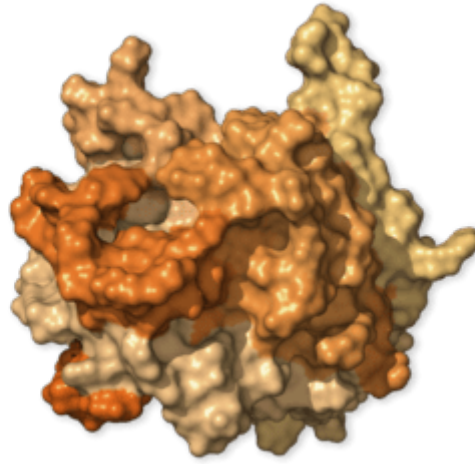
Cleaning Up · NSP15



Researchers suspect that this protein chops up leftover virus RNA as a way to hide from the infected cell's antiviral defenses.

agaguuuagaaaauguggcuuuuuauguuguaaaauaagggacacuuugauggacaacaggggugaaguaccag
uuucuaucuuuaauaacacuguuuacacaaaaguuugaugguguguugauguaagaauuguuugaaaaaauaaaca
acauuaccuguuauuguagcauuugagcuuugggcuaaagcgcaacuuuaaaccaguaccagaggugaaa
auacucaauaauuuggguguggacauugcugcuauuacugugaucugggacuacaaaagagaugcuccagca
cauaauaucuacuauuggguguuuguuucuaugacugacauagccaagaaaaccaacugaaacgauuugugcacca
cucacugucuuuuuugaugguagaguugauggucaaguagacuuuuuagaaaugcccguaauggguguucuu
auuacagaagguaguguuaaaagguuuacaaccaucuguaagguccaaacaagcuagucuuuauuggagucaca
uuauuugggagaagccguaaaaacacaguucaauuuuuuaaagaaaguugaugguguuguccaacaauuaccu
gaaacuuacuuuacucagaguagaaauuuacaagaauuuuaaaccaggagucaaauggaaaauugauuucuuu
gaauuagcuauuggaugaauucauugaacgguaauaaauagaaggcuaugccuucgaacauaucguuuuug
gagauuuuagucauagucaguuaagguguuuacauucacugauuuggacuagcuaaacguuuuaaggaauac
cuuuugaauuagaagauuuuuauccuauuggacaguacaguuaaaaaacuauuucuaacagaugcgcaaacag
guucaucuaaguguguguguuucuguuauugauuuuuuacuuugaugauuuuuguugaaauaauaaaauccca
agauuuuauucuguauguuucuaagguugucaaagugacuuuugacuauacagaaauuucuuuauugcuuuggug
uaaagaugqccauquaqaacauuuuacccaaaauuacaa

More Camouflage · NSP16



NSP16 works with NSP10 to help the virus's genes hide from proteins that chop up viral RNA.

```

ucuagucaagcguggcaaccggguguugcuaugccuaaauuuacaaaugcaagaugcuauuagaa
aagugugaccuucaaaaauuauggugauagugcaacauuaccuaaaggcauaauggauggucgcaaaaau
acucaacugugucaauauuuuaacacauuaacauuagcuguaccuauaauaugagaguuauacauuuuggu
gcugguucugauaaaggaguugcaccagguacagcuguuuuaagacagugguugccuacggguacgcugcuu
gucgauucagaucuuuaugacuuugucucugaugcagauucaacuuugauuggugauugugcaacuguaau
acagcuauuaaaugggaucucuuuuuagugauauguacgaccuaagacuaaaaauguuacaaaagaa
aaugacucuaaagaggguuuuuucacuuacauuuguggguuuuauacaacaaaagcuagcucuuaggaggu
uccguggcuauaaagauaacagaacauucuuuggaugcugaucuuuuauaagcucaugggacacuucgcaug
guggacagccuuuguuacuaaugugaaugcgucaucaucugaagcauuuuuaauuggauguaauuaucuuug
caaaccacgcgaacaaauagaugguuauugcaugcaugcaaaauacauuuuuggaggaaucacaauc
aauucaguugucuuuccuauucuuuuuuugacaugaguaauuuucccuuaauuaagggguacugcuguuug
ucuuuaaaagaaggucuuuacauuagauuuuuuauucucuucuuuaguaaagguaagacuuauuuuagaga
aaacaacagaguuguuuuuuucuaugugauguuucuuuuaacaacuaaacgaaca

```

Spike Protein · S

The spike protein is one of four structural proteins — S, E, M and N — that form the outer layer of the coronavirus and protect the RNA inside. Structural proteins also help assemble and release new copies of the virus.



The S proteins form prominent spikes on the surface of the virus by arranging themselves in groups of three. These crownlike spikes give coronaviruses their name.



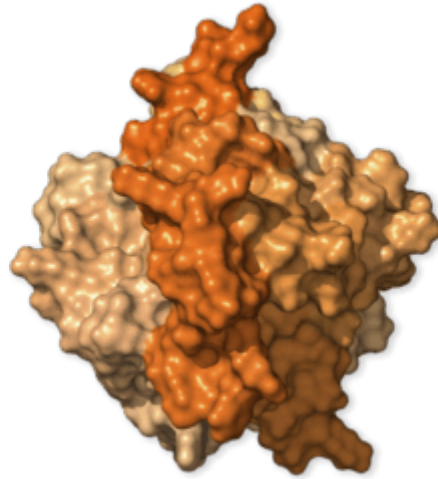
Part of the spike can extend and attach to a protein called ACE2 (in yellow below), which appears on particular cells in the human airway. The virus can then invade the cell.



The gene for the spike protein in SARS-CoV-2 has an insertion of 12 genetic letters: `ccucggcgggca`. This mutation may help the spikes bind tightly to human cells — a crucial step in its evolution from a virus that infected bats and other species.

A number of scientific teams are now designing vaccines that could prevent the spikes from attaching to human cells.

Escape Artist · ORF3a



The SARS-CoV-2 genome also encodes a group of so-called “accessory proteins.” They help change the environment inside the infected cell to make it easier for the virus to replicate.

The ORF3a protein pokes a hole in the membrane of an infected cell, making it easier for new viruses to escape. It also triggers inflammation, one of the most dangerous symptoms of Covid-19.

```
auggauuuguuuuagagaaucuuacacaaauuggaacuguaacuuguagcaaggugaaaaucaaggauugcuacu
ccuucagauuuuguucgcgcuacugcaacgauaccgauacaagccucacuccuuucggauggcuuauugu
uggcguugcacuucuuugcuguuuuuucagagcgcuuccaaaaucuaaaccucuaaaaagagauggcaacuagc
acucuccaaggguugucacuuuguuuugcaacuugcuguuguuguuuguaacaguuuacucacaccuuuugcu
cguugcugcuggccuugaagccccuuuucucuaucuuuauugcuuuagucuacuucuuugcagaguauaaacu
uuguaagaauaauaaugaggcuuugcuguugcuggaaaugccguuccaaaaacccauuacuuuauaugc
caacuauuuuucuuugcuggcauacuaauuguuacgacuauuguauaccuuacaauaguguaacuucuuca
auugucauuacuucaggugauggcacaacaaguccuauuucugaacaugacuaccagauuggguguuauacu
gaaaaaugggaaucuggaguaaaagacuguguuguauuacacaguuaucacuuucagacuauuaccagcug
uacucaacucaauugaguacagacacugguguugaacauguuaccuucuaucuaaauaaaauuguugau
gagccugaagaacauguccaaaucacacaauccgacgguucauccggaguuguuaauccaguaauggaacca
auuuauaugaugaaccgacgacgacuacuagcgugccuuuguuagcacaagcugaugaguacgaacuu
```

ORF3b overlaps the same RNA, but scientists aren’t sure if SARS-CoV-2 uses this gene to make proteins.

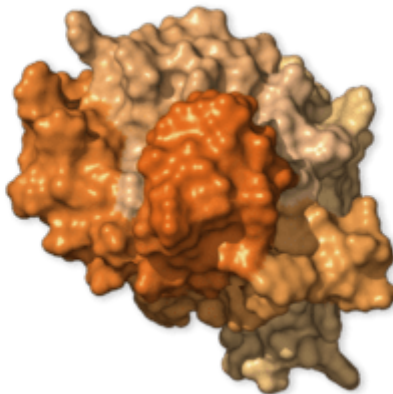
Envelope Protein · E



The envelope protein is a structural protein that helps form the oily bubble of the virus. It may also have jobs to do once the virus is inside the cell. Researchers have found that it latches onto proteins that help turn our own genes on and off. It's possible that pattern changes when the E protein interferes.

```
auguacucauucguuucggaagagacagguacguuaauaguuaauagcguacuucuuuuuucugcuuucgug
guauucugcuaguuaacacuagccauccuuacugcgcuucgauugugugcguacugcugcaauauuguua
acgugagucuuguuaaaacuuuuuuacguuuacucucguguuaaaaaucugaaucuucuaagaguuc
cugaucuucuggucuaaacgaacuaaaauuuauuuaguuuuuucuguuuggaacuuuaauuuuagcc
```

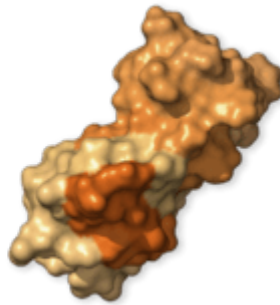
Membrane Protein · M



Another structural protein that forms part of the outer coat of the virus.

```
auggcagauuccaacgguacuauuaccguugaagagcuuaaaaagcuccuugaacaauggaaccuagua
auagguuuccuauuccuuacauggauuugucuucuacaauuugccuaugccaacaggaauagguuuuuguau
auaauuaaguuaauuuuccucuggcuguuauggccaguaacuuuagcuuguuuuugugcuugcugcuguu
uacagaauaaaauuggaucaccgguggaauugcuauucgcaauggcuugucuuguaggcuugauguggcucagc
uacuucuuugcuucuuucagacuguuugcgcguaucgcguuccauguggucauucaauccagaaacuaacauu
cuucucaacgugccacuccauggcacuaauucugaccagaccgcuucuaagaaagugaacucguaaucggagcu
gugaucuuucguggacaucuuucguauugcuggacaccaucuaggacgcugugacaucaaggaccugccua
aagaaucacuguugcuacauacgaacgcuucuuuuuacaaaauugggagcuucgcagcguguagcaggug
acucagguuuugcugcauacagucgcuacaggauuggcaacuauaaaauaaacacagaccauuccaguagca
gugacaauauugcuuugcuuguacaguaagugacaacag
```

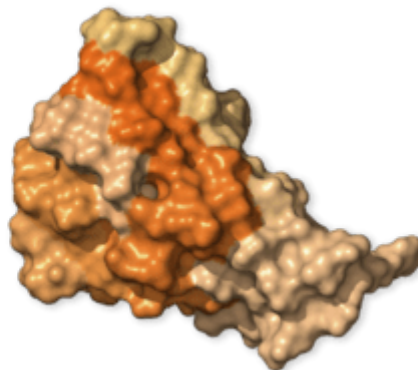
Signal Blocker · ORF6



This accessory protein blocks signals that the infected cell would send out to the immune system. It also blocks some of the cell's own virus-fighting proteins, the same ones targeted by other viruses such as polio and influenza.

```
auguuucaucucguugacuuucagguuacuauagcagagauuuacuaauuuuuuagaggacuuuuuuaagu
uuccauuuggaaucuugauuacaucaaaaaccucauaauuuuuuuuuuaucaagucacuaacugagaa
aaauuucuaauuagaugaagagcaaccauggagauugauuuuuuacgaac
```

Virus Liberator · ORF7a



When new viruses try to escape a cell, the cell can snare them with proteins called tetherin. Some research suggests that ORF7a cuts down an infected cell's supply of tetherin, allowing more of the viruses to escape. Researchers have also found that the protein can trigger infected cells to commit suicide — which contributes to the damage Covid-19 causes to the lungs.

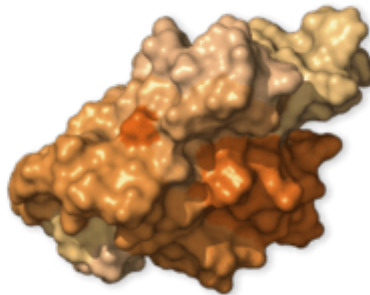
```

augaaaaauuucuuuuucuuuggcacugauaacacucgcuacuugugagcuuuaucauaccaagagugugu
uagagguacaacaguacuuaaagaaccuugcucuucuggaacauacgagggcaauucaccauuucauc
cucuagcugauaacaauuugcacugacuugcuuuagcacucaaauugcuuuugcuuguccugacggcgua
aaacacgucuaucaguuaacgugccagaucauuuacacuaaacuguucaucagacaagaggaaguucaaga
acuuuacucuccaauuuuucuuauuguugcggaauaguguuuauaacacuugcuucacacucaaaaga
aagacagaauugaauuacuuuauuugacuucuuuugugcuuuuuagccuuucugcuauuccuug
uuuuauuuaugcuuauuauuuuugguucucacuugaacugcaagaucuaauggaaacuugucacgccu
aaacgaac

```

ORF7b overlaps this same stretch of RNA, but it's not clear what, if anything, the gene does.

Mystery Protein · ORF8



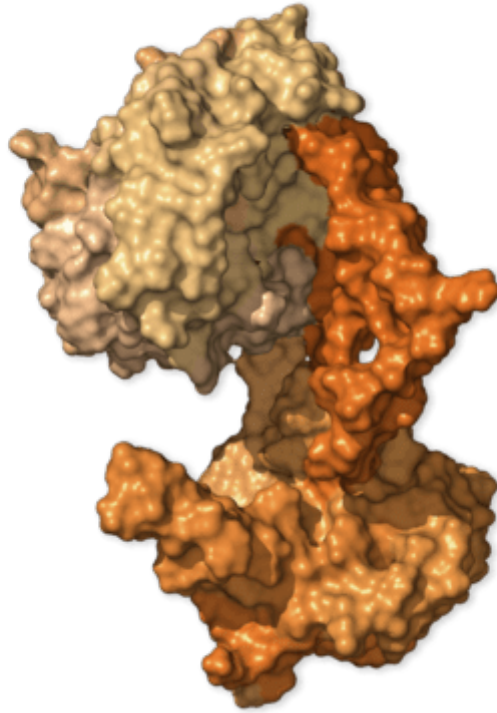
The gene for this accessory protein is dramatically different in SARS-CoV-2 than in other coronaviruses. Researchers are debating what it does.

```

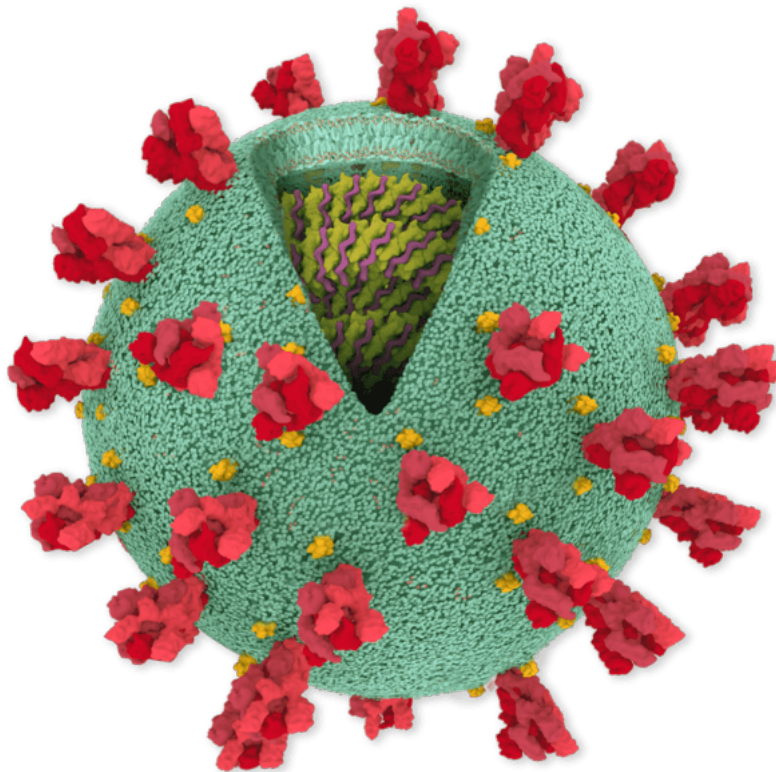
augaaauuucuuuuucuuaggaaucaucacaacugugagcugcauuuaccaagauguaguuuacaguca
uguacuacaacaucaaccauauuguaguugagacccguguccuauucacuucuaauucuaaaugguauuuaga
guaggagcuagaaaucagcaccuuuaauugaauugugcguggaugaggcugguucuaaaucacccaauucag
uacaucgauaucgguaauuauacaguuuuccuguuuaccuuuuacaauuaauugccaggaaccuaaaauugg
guagucuuguagugcguguuucguuucuaugaagacuuuuuagaguaucaugacguucguguuguuuuagauu
ucaucuaaacgaacaaacuaaa

```

Nucleocapsid Protein · N



The N protein protects the virus RNA, keeping it stable inside the virus. Many N proteins link together in a long spiral, wrapping and coiling the RNA:

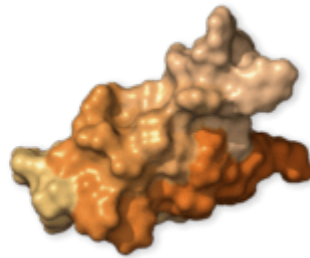


```
augucugauaauggaccccaaaucagcgaaaugcaccccgcauuacguuugguggacccucagauuca
acuggcaguaaccagaauggagaacgcaguggggcgcgaucaaaacaacgucggccccaagguuuaccca
aauaacugcgucuugguucaccgcucucacucaacauggcaaggaagaccuuuuuucccucgaggaca
aggcguuccaauuaacaccauagcaguccagaugaccaaaauuggcuacuaccgaagagcuaccagacga
auucgugguggugacgguaaaaugaagaucucaguccaagaugguauuucuacuaccuaggaacugggc
```

cagaagcuggacuucccuauggugcuaacaaagacggcaucauauggguugcaacugagggagccuuga
 auacacaaaagaucacauuggcacccgcaauccugcuaacaaugcugcaaucgugcuacaacuuccuca
 aggaacaacauugccaaaaggcuucacgcagaagggagcagagggcggcagucagcagcucucguuc
 cucaucacguagucgcaacaguucaagaaaaucaacuccaggcagcagaggggaacuucuccugcuaga
 auggcuggcaauggcggugaugcugcucuuugcugcugcuugacagauugaaccagcuugagagca
 aaugucugguaaaaggccaacaacaaggccaacugucacuaagaaaucugcugcugagggcuucua
 agaagccucggcaaaaacguacugccacuaaaagcauacaauguaacacaagcuuucggcagacgugguc
 cagaacaaacccaaggaaaauuuggggaccaggaacuaaucagacaaggaacugauuacaaacauuggc
 cgaaaauugcacaauuugcccccagcgcucacgcuucucggaugucgcgcauuggcauggaagucacac
 cuucgggaacgugguugaccuacacaggugccaucaaaauuggaugacaaagaucacaaauucaaagauca
 agucauuuugcugaauaagcauauugacgcauacaaaacauuccaccaaagagccuaaaaaggacaa
 aaagaagaaggcugaugaaacucaagccuuaccgcagagacagaagaaacagcaaacugugacucucucuc
 cugcugcagauuuggaugauuucuccaaauugcaacaauccaugagcagugcugacucaacucaggc
 cuaaacucaugcagaccacacaaggcag

The accessory proteins ORF9b and ORF9c overlap this same stretch of RNA. ORF9b blocks interferon, a key molecule in the defense against viruses, but it's not clear if ORF9c is used at all.

Mystery Protein · ORF10



Close relatives of the SARS-CoV-2 virus don't have the gene for this tiny accessory protein, so it's hard to know what it's for yet — or even if the virus makes proteins from it.

augggcuaauaaaacguuuucgcuuuuccguuuacgauauauagucucucuuugugcagaauugaauucucgu
 aacuacauagcacaaguagauguuuaacuuuaauncucacauag

End of the Line



The coronavirus genome ends with a snippet of RNA that stops the cell's protein-making machinery. It then trails away as a repeating sequence of aaaaaaaaaaaaaa

```
caaucuuuaucaguguguaacauuagggaggacuugaaagagccaccacauuuucaccgaggccacgcgga  
guacgaucgaguguaacagugaacaauugcuagggagagcugccuauauggaagagcccuaauguguaaaaua  
auuuuaguagugcuaucucccaugugauuuuaauagcuucuuaggagaugacaaaaaaaaaaaaaaaaa  
aaaaaaaaaaaaaaaaa
```

Note: The four letters of DNA are A, C, G and T. In RNA molecules like the coronavirus genome, the T (thymine) is replaced with U (uracil).

Sources: Fan Wu et al., Nature; National Center for Biotechnology Information; Dr. David Gordon, University of California, San Francisco; Dr. Matthew B. Frieman and Dr. Stuart Weston, University of Maryland School of Medicine; Dr. Pleuni Pennings, San Francisco State University; David Haussler and Jason Fernandes, U.C. Santa Cruz Genomics Institute; Journal of Virology; Annual Review of Virology.

Model sources: Coronavirus by Maria Voigt, RCSB Protein Data Bank headquartered at Rutgers University–New Brunswick; Ribosome from Heena Khatter et al., Nature; Proteins from Yang Zhang's Research Group, University of Michigan.