

<https://www.ncbi.nlm.nih.gov/Structure/SARS-CoV-2.html>

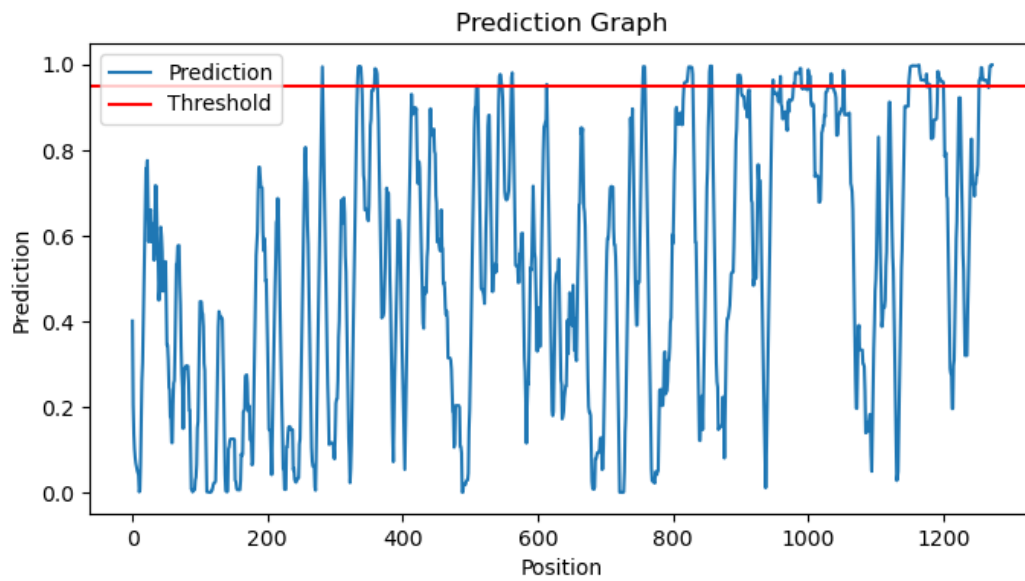
surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]
NCBI Reference Sequence: YP_009724390.1

https://www.ncbi.nlm.nih.gov/protein/YP_009724390.1?report=fasta

MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVT
WFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIRGWIFGTTLDSKTQSLIVNNATNVV
IKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKN
LREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTGDS
SSGWTAGAAAYYVGYLQPRTFLLKYNENGITDAVDCALDPLSETKCTLKSFTVEKGIYQTSN
FRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCY
GVSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDS
KVGGNYNLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGY
QPYRVVVLSEFLLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQGFGR
DIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPT
WRVYSTGSNVFQTRAGCLIGAEHVNNSECDIPIGAGICASYQTQTNSPRRARSVASQSIIAY
TMSLGAENSVAYSNNISAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSF
CTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLF
NKVTLADAGFIKQYGDCLGDIAARDLCAQKFNGLTVLPPLLDEMIQYTSALLAGTITSGWT
FGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSSTASALGKLQD
VVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDLKVEAEVQIDRLITGRLQSLQTYVTQQLIR
AAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF
TTAPAICHGKAHFPREGVFVSNNGTHWFTQNRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVY
DPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVNIQKEIDRLNEVAKNLNESLIDLQEL
GKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSEPVLK
GVKLHYT

N
NLCPFGE
SNCVA
VV
NGLT
FQ
D
YGSF
IEDLLFNKVTAD
FNGL
PFAMQ
Q
N
ILSRDLKVEAE
LQSL
GQ
PQ
EELDKYFKNHTSPDVDLGDISGINASV
NIQK
LNESLIDLQ
CKFDEDDSEPVKLG

KLHYT

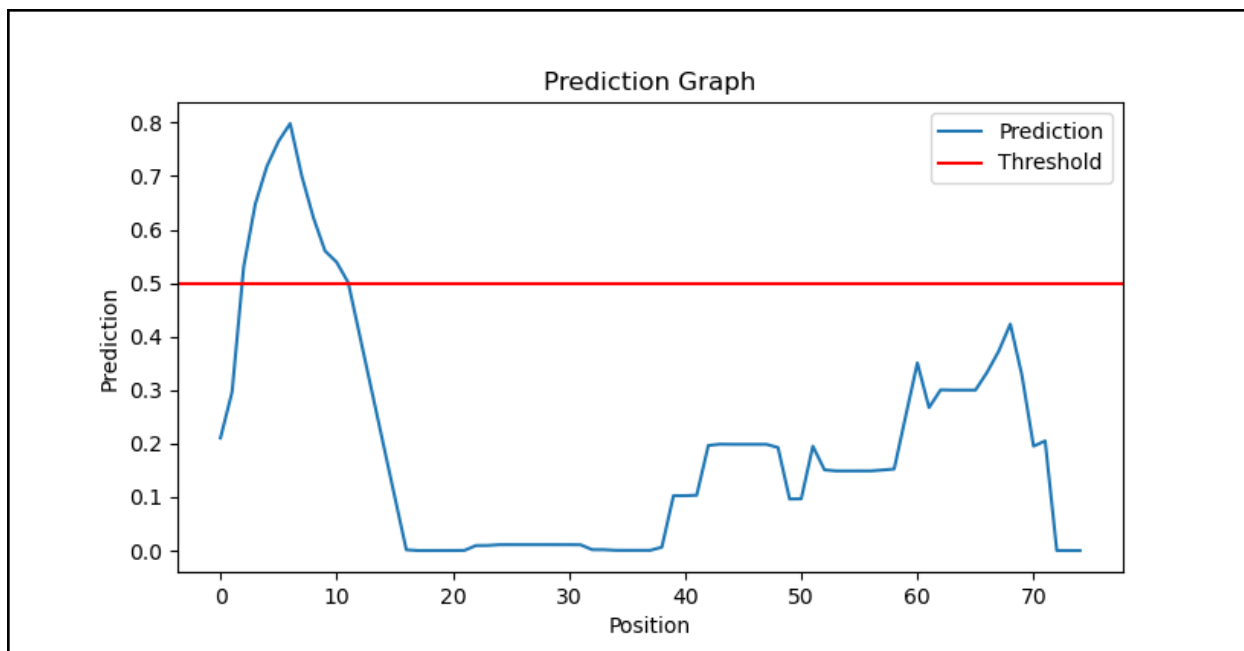


envelope protein [Severe acute respiratory syndrome coronavirus 2]
NCBI Reference Sequence: YP_009724392.1

https://www.ncbi.nlm.nih.gov/protein/YP_009724392.1?report=fasta

MYSFVSEETGTLIVNSVLLFLAFVVFLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLN
SSRVPDLLV

SFVSEETGTL

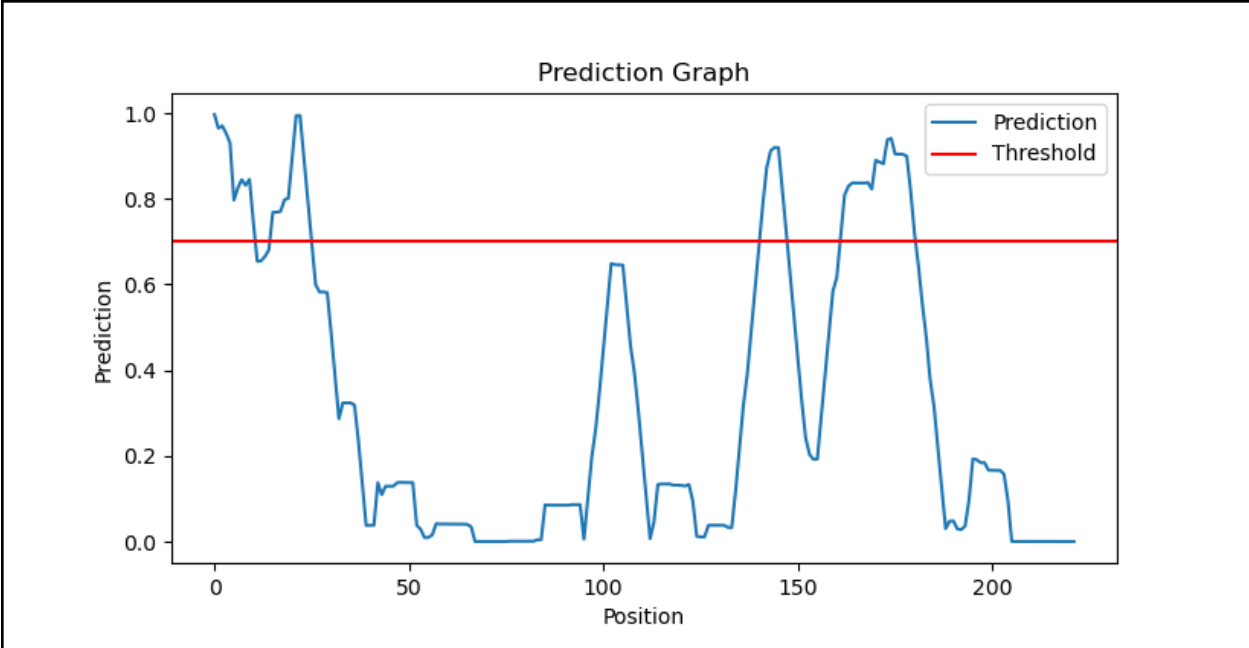


membrane glycoprotein [Severe acute respiratory syndrome coronavirus 2]
 NCBI Reference Sequence: YP_009724393.1

https://www.ncbi.nlm.nih.gov/protein/YP_009724393.1?report=fasta

MADSNGTITVEELKKLLEQWNLVIGFLFTWICLLQFAYANRNRFLYIIKLIFLWLLWPVTLACFV
 LAAYRINWITGGIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILLNVPLHGTILTR
 PLLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSTYYKLGASQRVAGDSGFAAY
 SRYRIGNYKLNTDHSSSSDNIALLVQ

MADSNGTITVE
LLEQWNLVIG
AVILRGH
KDLPKEITVATSRTLSTYYKL

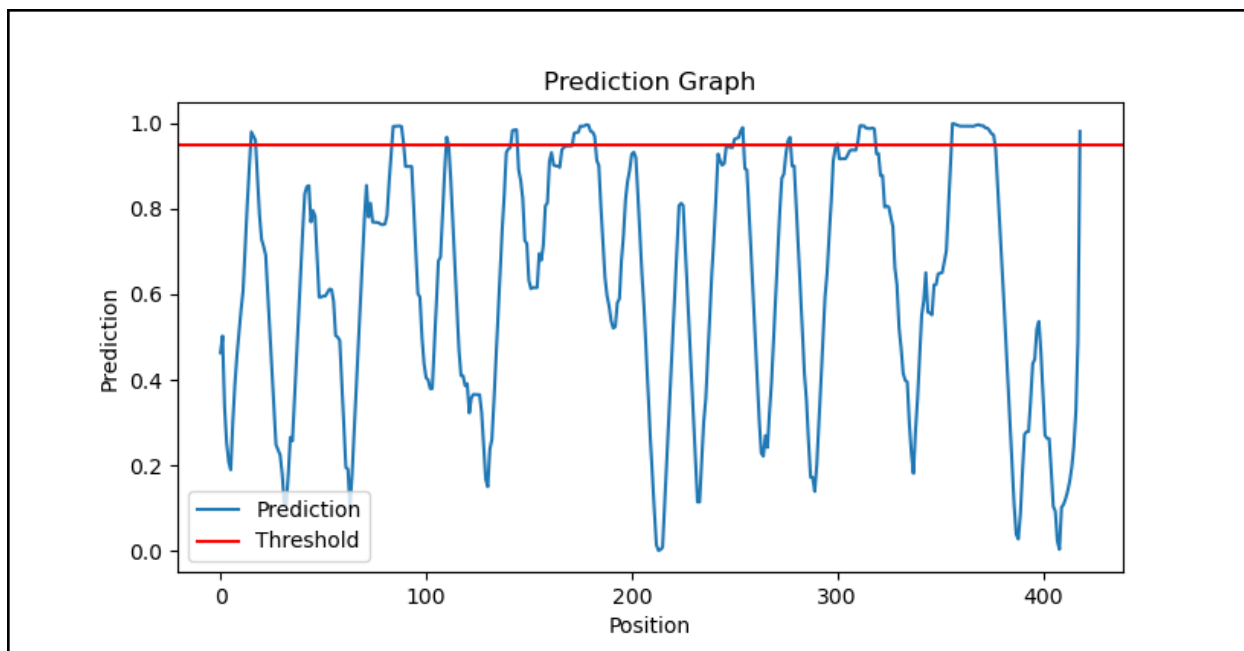


nucleocapsid phosphoprotein [Severe acute respiratory syndrome coronavirus 2]
NCBI Reference Sequence: YP_009724397.2

https://www.ncbi.nlm.nih.gov/protein/YP_009724397.2?report=fasta

MSDNGPQNQRNAPRITFGGSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQH
GKEDLKFPARGQGVPIINTSSPDDQIGYYRRATRRIRGGDGKMKDLSRWYFYLLGTGPEAG
LPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIIVLQLPQGTTLPKGFYAEGSRGGSQAS
SRSSSRSRNSSRNSTPGSSRGTSARMAGNGGDAALALLLDRLNQLESKMSGKGQQQQG
QTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQI
AQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDKDPNFKDQVILLNKHIDAYKTFPPT
PKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA

TFG
GYRRA
Y
KDH
AEGSRGGSQAS
AAEAS
RG
W
SAFFGMSR
IDAYKTFPPT
PKKDKKKKAD
A



Bolded sequences are considered for cross-referencing. Only sequences with length ≥ 5 amino acids are selected.

Selected stretches (**confident consensus achieved**)

NLCPFGE
SNCVA
IEDLLFNKVTLAD
PFAMQ
ILSRDKVEAE
EELDKYFKNHTSPDVLGDISGINASV
LNESLIDLQ
CKFDEDDSEPVLKG
KLHYT

SFVSEETGTL

MADSNGTITVE
LLEQWNLVIG
AVILRGH
KDLPKEITVATSRTLSTYYKL

GYRRA
AEGSRGGSQAS
AAEAS
ASAFFGMSR
IDAYKTFPPTEPKDKKKKKAD

Multi-epitope molecule sequences

Spike

NLCPFGESNCVAIEDLLFNKVTADILSRDKVEAEEELDKYFKNHTSPDVDLGDISGINASV
LNESLIDLQCKFDEDDSEPVKLGKLHYT

Envelope

SFVSEETGTL

Membrane

MADSNGTITVELLEQWNLVIGKDLPKEITVATSRTLSTYYKL

Nucleocapsid

GYRRRAEGSRGGSQASAAEASIDAYKTFPPTEPKDKKKKAD

All (rigid linker between different antigen sources, flexible linkers between the same source)

EAAAKEAAAKEAAAK,NLCPFGE,SNCVA,GGGGSGGGGSGGGGS,IEDLLFNKVTAD,GG
GGSGGGGSGGGGS,ILSRDKVEAE,GGGGSGGGGSGGGGS,EELDKYFKNHTSPDVDLG
DISGINASV,GGGGSGGGGSGGGGS,LNESLIDLQ,GGGGSGGGGSGGGGS,CKFDEDDSE
PVLKG,GGGGSGGGGSGGGGS,KLHYT,EAAAKEAAAKEAAAK,SFVSEETGTL,EAAAKEA
AAKEAAAK,MADSNGTITVE,LLEQWNLVIG,GGGGSGGGGSGGGGS,KDLPKEITVATSRTL
STYYKL,EAAAKEAAAKEAAAK,GYYRRA,GGGGSGGGGSGGGGS,AEGSRGGSQAS,GGG
GSGGGGSGGGGS,AAEAS,EAAAKEAAAKEAAAK,IDAYKTFPPTEPKDKKKKAD,EAAA
KEAAAKEAAAK