

# 1

## Delving Deep into the Structural Aspects of A Furin Cleavage Site Inserted into the Spike Protein of SARS-CoV-2: Supplementary materials

## Amino acid sequence alignment

QHD	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRS	45
6VSB	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRS	45
QHD	SVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGV	90
6VSB	SVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGV	90
QHD	YFASTEKSNIIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVFCEFOF	135
6VSB	YFASTEKSNIIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVFCEFOF	135
QHD	CNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE	180
6VSB	CNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE	180
QHD	GKQGNFKNLREFVFKNIDGYFKIYKHTPINLVRDLPQGFSALEP	225
6VSB	GKQGNFKNLREFVFKNIDGYFKIYKHTPINLVRDLPQGFSALEP	225
QHD	LVDLPIGINITRFQTLALHRSYLTTPGDSSSGWTAGAAAYVGYL	270
6VSB	LVDLPIGINITRFQTLALHRSYLTTPGDSSSGWTAGAAAYVGYL	270
QHD	QPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQT	315
6VSB	QPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQT	315
QHD	SNFRVQPTESIIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN	360
6VSB	SNFRVQPTESIIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN	360
QHD	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD	405
6VSB	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD	405
QHD	EVROIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYN	450
6VSB	EVROIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYN	450
QHD	YLYRLF RKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSY	495
6VSB	YLYRLF RKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSY	495
QHD	GFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN	540
6VSB	GFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN	540
QHD	FNFNGLTGTGVLTESNKKFLPFQOFGRDIADTTDAVRDPQTLLEIL	585
6VSB	FNFNGLTGTGVLTESNKKFLPFQOFGRDIADTTDAVRDPQTLLEIL	585
QHD	DITPCSEGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLT	630
6VSB	DITPCSEGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLT	630
QHD	PTWRVYSTGSNVFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQ	675
6VSB	PTWRVYSTGSNVFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQ	675

QHD	TQTNSP	RRAR	SVASQSI	IAYTMSLGAENSVAYSNN	SI	AIPTNFTI	720
6VSB	TQTNSP	GSAS	SVASQSI	IAYTMSLGAENSVAYSNN	SI	AIPTNFTI	720
QHD	SVTTEILP	VSMTKTSVDCTMYICGDST	ECNLLLO	YGSFCTQLNR			765
6VSB	SVTTEILP	VSMTKTSVDCTMYICGDST	ECNLLLO	YGSFCTQLNR			765
QHD	ALTGIAVE	QDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPS					810
6VSB	ALTGIAVE	QDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPS					810
QHD	KPSKR	SFIEDLLFNKVT	LADAGFIKQYGDCLG	DAARDL	ICAQKF		855
6VSB	KPSKR	SFIEDLLFNKVT	LADAGFIKQYGDCLG	DAARDL	ICAQKF		855
QHD	NGLTVLP	PPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM					900
6VSB	NGLTVLP	PPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM					900
QHD	QMAYRF	NGIGVTONVLYENQKLIANQFN	SAIGKIQDSL	SSTASAL			945
6VSB	QMAYRF	NGIGVTONVLYENQKLIANQFN	SAIGKIQDSL	SSTASAL			945
QHD	GKLQDV	VNQNQAALNTLVKQLSSNFGAISSVLNDILSRLD	KVEAE				990
6VSB	GKLQDV	VNQNQAALNTLVKQLSSNFGAISSVLNDILSRLD	PEAE				990
QHD	VOIDRL	LITGRLOSLOTYVTOQLIRAAEIRASANLAATKMSECVLG					1035
6VSB	VOIDRL	LITGRLOSLOTYVTOQLIRAAEIRASANLAATKMSECVLG					1035
QHD	QSKRVDF	CGKGYHLSFPOSAPHG	VVFLHV	TYVPAQEK	NFTTAPA		1080
6VSB	QSKRVDF	CGKGYHLSFPOSAPHG	VVFLHV	TYVPAQEK	NFTTAPA		1080
QHD	ICH	DGKAHFPREGVFVSN	GTHWFVTQRNFYEPQI	ITTDNTFVSGN			1125
6VSB	ICH	DGKAHFPREGVFVSN	GTHWFVTQRNFYEPQI	ITTDNTFVSGN			1125
QHD	CDVVIG	I	VNNTVYDPLQPELDSFKEELDKYFKNHTSPD	VLDGDIS			1170
6VSB	CDVVIG	I	VNNTVYDPLQPELDSFKEELDKYFKNHTSPD	VLDGDIS			1170
QHD	GINASV	VNIQKEIDRLNEVAKNLNESLIDLQELGKYE	QYIKWPWY				1215
6VSB	GINASV	VNIQKEIDRLNEVAKNLNESLIDLQELGKYE	QGSYIPE				1215
QHD	IWLGF	I	AGLIAIVMVTIMLCCMTSCCSC	L	KGCCSCGSCCKFDEDD		1260
6VSB	APRDGQ	A	YVRKDGEWVLLSTFLGRSLEV	L	FQGP	GHHHHHHHSAW	1260
QHD	SE	P	VLKGVKLHYT				1273
6VSB	SH	P	QFEKGGGSGG	GGSGGSAWSHPQFEK			1288

☒ non conserved

☒ ≥50% conserved

**Figure 1.1** Alignment of amino acid sequences of QHD43416.1(QHD, GenBank access code for S protein of SARS-CoV-2) and 6VSB (PDB access code for a Cryo-EM structure of S protein of SARS-CoV-2). Here, 6VSB represents the structural template for homology structural modeling with the SwissModel server (<https://swissmodel.expasy.org/interactive>).

CHAPTER 1. DELVING DEEP INTO THE STRUCTURAL ASPECTS OF A FURIN CLEAVAGE SITE  
 INSERTED INTO THE SPIKE PROTEIN OF SARS-COV-2: SUPPLEMENTARY MATERIALS

4

6VSB	MFVFLVLLPLVSSQCVNLTTRTQLPP	AYTNSFTRGVVYYPDKVFRS	45			
6VSBA	.....	AYTNSFTRGVVYYPDKVFRS	19			
6VSBB	.....	AYTNSFTRGVVYYPDKVFRS	19			
6VSBC	.....	AYTNSFTRGVVYYPDKVFRS	19			
6VSB	SVLHSTQDLFLPFFSNVTWFH	AIHVSSGTNGTKRFDNP	VLPFNDGV	90		
6VSBA	SVLHSTQDLFLPFFSNVTWFH	FDNP	.....	VLPFNDGV	52	
6VSBB	SVLHSTQDLFLPFFSNVTWFH	NP	.....	VLPFNDGV	50	
6VSBC	SVLHSTQDLFLPFFSNVTWFH	DN	.....	VLPFNDGV	51	
6VSB	YFAST	EKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVFCEFOF	135			
6VSBA	YFAST	N...IIRGWIFGTTLDSKTQSLLIVNNATNVVIKVFCEFOF	94			
6VSBB	YFAST	EKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVFCEFOF	95			
6VSBC	YFAST	N...IIRGWIFGTTLDSKTQSLLIVNNATNVVIKVFCEFOF	93			
6VSB	CNDPFLG	VYYHKNNKSWMSE	FRVYSSANNCTFEYVSQPF	MD	LE	180
6VSBA	CNDPFLG	E	.....	FRVYSSANNCTFEYVSQPF	LKNLR	126
6VSBB	CNDPFL	SE	.....	FRVYSSANNCTFEYVSQPF	LKNLR	127
6VSBC	CNDPF	F	.....	RVYSSANNCTFEYVSQPF	LKNLR	122
6VSB	GKQGNFKNLR	EFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEP	225			
6VSBA	.....	EFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEP	161			
6VSBB	.....	EFVFKNIDGYFKIYSKHTP	P.....QGFSALEP	155		
6VSBC	.....	EFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEP	157			
6VSB	LVDLPIGINITRFQTLALHR	SYLTPGDSSSGWTAG	AAAYYVGYL	270		
6VSBA	LVDLPIGINITRFQTLALHR	G	.....	AAAYYVGYL	192	
6VSBB	LVDLPIGINITRFQTLA	AYY	.....	.....VGYL	180	
6VSBC	LVDLPIGINITRFQTLALH	G	.....	AAAYYVGYL	187	
6VSB	QPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQT	315				
6VSBA	QPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQT	237				
6VSBB	QPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQT	225				
6VSBC	QPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQT	232				
6VSB	SNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN	360				
6VSBA	SNFRVQPTESIVR	.....LCPFGEVFNATRFASVYAWNRKRISN	276			
6VSBB	SNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN	270				
6VSBC	SNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN	277				
6VSB	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD	405				
6VSBA	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD	321				
6VSBB	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD	315				
6VSBC	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD	322				
6VSB	EVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDS	KVGGNYN	450			
6VSBA	EVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDS	YN	.....	361		
6VSBB	EVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDS	YN	.....	355		
6VSBC	EVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDS	YN	.....	362		
6VSB	YLYRFLFRKS	NLKPFFERDISTEIIY	QAGSTPCNGVEGFNCYFPLQSY	495		
6VSBA	.....	YLYRPLQSYG	.....	371		
6VSBB	.....	YLYRNLKPFFERDISTEIIY	NC	.....	YFPLQ	380
6VSBC	.....	YLYRNLKPFFERDISTENC	.....	.....	YFP	383

6VSB	GFQPTNG	VGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN	540		
6VSBA	...	FQPTVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN	413		
6VSBB	SYG	FQPTVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN	425		
6VSBC	LQSYGFQ	VGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN	428		
6VSB		FNFNGLTGTGVLTESNKKFLPFQGFGRDIADTTDAVRDPOTLEIL	585		
6VSBA		FNFNGLTGTGVLTESNKKFLPFQGFGRDIADTTDAVRDPOTLEIL	458		
6VSBB		FNFNGLTGTGVLTESNKKFLPFQGFGRDIADTTDAVRDPOTLEIL	470		
6VSBC		FNFNGLTGTGVLTESNKKFLPFQGFGRDIADTTDAVRDPOTLEIL	473		
6VSB		DITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEV	PVAIHADQLT	630	
6VSBA		DITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEV	SN.....	495	
6VSBB		DITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEV	TGSN.....	509	
6VSBC		DITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEV	N.....	509	
6VSB	PTRVYSTGSN	VFQTRAGCLIGAEHVNNSYECDIPIGAGICA	SYQ	675	
6VSBA	.....	VFQTRAGCLIGAEHVNNSYECDIPIGAGICA	.....	526	
6VSBB	.....	VFQTRAGCLIGAEHVNNSYECDIPIGAGICA	.....	540	
6VSBC	.....	VFQTRAGCLIGAEHVNNSYECDIPIGAGICA	.....	540	
6VSB	TQTNSPGSASS	VASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTI		720	
6VSBA	.....	VASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTI		560	
6VSBB	.....	VASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTI		574	
6VSBC	.....	VASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTI		574	
6VSB		SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNR		765	
6VSBA		SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNR		605	
6VSBB		SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNR		619	
6VSBC		SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNR		619	
6VSB		ALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPS		810	
6VSBA		ALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPS		650	
6VSBB		ALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPS		664	
6VSBC		ALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPS		664	
6VSB	KPSK	RSFIEDLLFN	KVTLADAGFIKQYGDCLGDIAARD	LICAKKF	855
6VSBA	K...	RSFIEDLLFN	.....	KVTLQKF	668
6VSBB	K...	RSFIEDLLFN	.....	KVTLQKF	682
6VSBC	K...	RSFIEDLLFN	.....	KVTLQKF	682
6VSB		NGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM		900	
6VSBA		NGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM		713	
6VSBB		NGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM		727	
6VSBC		NGLTVLPPLLTDemiaQYTSALLAGTITSGWTFGAGAALQIPFAM		727	
6VSB		QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDLSLSTASAL		945	
6VSBA		QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDLSLSTASAL		758	
6VSBB		QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDLSLSTASAL		772	
6VSBC		QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDLSLSTASAL		772	
6VSB		GKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDPPEAE		990	
6VSBA		GKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDPPEAE		803	
6VSBB		GKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDPPEAE		817	
6VSBC		GKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDPPEAE		817	

6VSB	<span style="background-color: #0000FF; color: #0000FF;">VOIDRLITG</span> RLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLG	1035
6VSBA	<span style="background-color: #0000FF; color: #0000FF;">VOIDRLITG</span> .....	812
6VSBB	<span style="background-color: #0000FF; color: #0000FF;">VOIDRLITG</span> .....	826
6VSBC	<span style="background-color: #0000FF; color: #0000FF;">VOIDRLITG</span> .....	826
6VSB	QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTVPAQEKNFTTAPA	1080
6VSBA	.....	812
6VSBB	.....	826
6VSBC	.....	826
6VSB	ICHDGKAHFPPREGVVFVSNNGTHWFVTQRNFYEPQIITTDNTFVSGN	1125
6VSBA	.....	812
6VSBB	.....	826
6VSBC	.....	826
6VSB	CDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDIS	1170
6VSBA	.....	812
6VSBB	.....	826
6VSBC	.....	826
6VSB	GINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQGSYIPE	1215
6VSBA	.....	812
6VSBB	.....	826
6VSBC	.....	826
6VSB	APRDGQAYVRKDGWVLLSTFLGRSLEVLVLFQGPGRHHHHHHHSAW	1260
6VSBA	.....	812
6VSBB	.....	826
6VSBC	.....	826
6VSB	SHPQFEKGGGSGGGGSGGSAWSHPQFEK	1288
6VSBA	.....	812
6VSBB	.....	826
6VSBC	.....	826

  non conserved  
   $\geq 50\%$  conserved

**Figure 1.2** Alignment of amino acid sequences of 6VSB (experimental) and its three chains (A, B and C) in the structural model of 6VSB.

QHD	MVFVLVLLPLVSSQCVNLTTTRTQLPP	AYTNSFTRGVYYPD	40
modelA	.....	AYTNSFTRGVYYPD	14
modelB	.....	AYTNSFTRGVYYPD	14
modelC	.....	AYTNSFTRGVYYPD	14
QHD	KVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD		80
modelA	KVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD		54
modelB	KVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD		54
modelC	KVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD		54
QHD	NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV		120
modelA	NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV		94
modelB	NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV		94
modelC	NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV		94
QHD	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY		160
modelA	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY		134
modelB	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY		134
modelC	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY		134
QHD	SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGY		200
modelA	SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGY		174
modelB	SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGY		174
modelC	SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGY		174
QHD	FKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFOT		240
modelA	FKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFOT		214
modelB	FKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFOT		214
modelC	FKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFOT		214
QHD	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYN		280
modelA	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYN		254
modelB	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYN		254
modelC	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYN		254
QHD	ENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV		320
modelA	ENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV		294
modelB	ENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV		294
modelC	ENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV		294
QHD	QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN		360
modelA	QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN		334
modelB	QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN		334
modelC	QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN		334
QHD	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF		400
modelA	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF		374
modelB	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF		374
modelC	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF		374
QHD	VIRGDEVROIAPGQTGKIADYNYKLPDDFTGCVIAWNSNN		440
modelA	VIRGDEVROIAPGQTGKIADYNYKLPDDFTGCVIAWNSNN		414
modelB	VIRGDEVROIAPGQTGKIADYNYKLPDDFTGCVIAWNSNN		414
modelC	VIRGDEVROIAPGQTGKIADYNYKLPDDFTGCVIAWNSNN		414

QHD	LDSKVGGNYYLYRLFRKSNLKPFFERDISTEIQAGSTPC	480
modelA	LDSKVGGNYYLYRLFRKSNLKPFFERDISTEIQAGSTPC	454
modelB	LDSKVGGNYYLYRLFRKSNLKPFFERDISTEIQAGSTPC	454
modelC	LDSKVGGNYYLYRLFRKSNLKPFFERDISTEIQAGSTPC	454
QHD	NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHA	520
modelA	NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHA	494
modelB	NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHA	494
modelC	NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHA	494
QHD	PATVCGPKKSTNLVKNKCVNFFNGLTGTGVLTESNKKFL	560
modelA	PATVCGPKKSTNLVKNKCVNFFNGLTGTGVLTESNKKFL	534
modelB	PATVCGPKKSTNLVKNKCVNFFNGLTGTGVLTESNKKFL	534
modelC	PATVCGPKKSTNLVKNKCVNFFNGLTGTGVLTESNKKFL	534
QHD	PFQQFGRDIADTTDAVRDPQMLEILDITPCSFGGVSVITP	600
modelA	PFQQFGRDIADTTDAVRDPQMLEILDITPCSFGGVSVITP	574
modelB	PFQQFGRDIADTTDAVRDPQMLEILDITPCSFGGVSVITP	574
modelC	PFQQFGRDIADTTDAVRDPQMLEILDITPCSFGGVSVITP	574
QHD	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS	640
modelA	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS	614
modelB	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS	614
modelC	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS	614
QHD	NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQNS	680
modelA	NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQNS	654
modelB	NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQNS	654
modelC	NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQNS	654
QHD	PRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI	720
modelA	PRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI	694
modelB	PRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI	694
modelC	PRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI	694
QHD	SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFC	760
modelA	SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFC	734
modelB	SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFC	734
modelC	SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFC	734
QHD	TQLNRALTGIAVEQDKNTOEVFAQVKQIYKTPPIKDFGGF	800
modelA	TQLNRALTGIAVEQDKNTOEVFAQVKQIYKTPPIKDFGGF	774
modelB	TQLNRALTGIAVEQDKNTOEVFAQVKQIYKTPPIKDFGGF	774
modelC	TQLNRALTGIAVEQDKNTOEVFAQVKQIYKTPPIKDFGGF	774
QHD	NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC	840
modelA	NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC	814
modelB	NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC	814
modelC	NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC	814
QHD	LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG	880
modelA	LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG	854
modelB	LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG	854
modelC	LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG	854



QHD	TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQ	920
modelA	TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQ	894
modelB	TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQ	894
modelC	TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQ	894
QHD	KLIANQFNNSAIGKIQDSLSTASALGKLDVNVNQAQALN	960
modelA	KLIANQFNNSAIGKIQDSLSTASALGKLDVNVNQAQALN	934
modelB	KLIANQFNNSAIGKIQDSLSTASALGKLDVNVNQAQALN	934
modelC	KLIANQFNNSAIGKIQDSLSTASALGKLDVNVNQAQALN	934
QHD	TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGR	1000
modelA	TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITG.	973
modelB	TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITG.	973
modelC	TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITG.	973
QHD	LQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV	1040
modelA	.....	973
modelB	.....	973
modelC	.....	973
QHD	DFCGKGYHLMSEFPQSAPHGVVFLHVTVYVPAQEKNFTTAPA	1080
modelA	.....	973
modelB	.....	973
modelC	.....	973
QHD	ICHGDKAHFPREGVVFVSNGTHWFVTQRNFYEPQIITTDNT	1120
modelA	.....	973
modelB	.....	973
modelC	.....	973
QHD	FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHT	1160
modelA	.....	973
modelB	.....	973
modelC	.....	973
QHD	SPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL	1200
modelA	.....	973
modelB	.....	973
modelC	.....	973
QHD	QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSC	1240
modelA	.....	973
modelB	.....	973
modelC	.....	973
QHD	CSCCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT	1273
modelA	.....	973
modelB	.....	973
modelC	.....	973

non conserved

$\geq 50\%$  conserved

**Figure 1.3** Alignment of amino acid sequences of QHD43416.1 (QHD) and three chains (A, B and C) of the homology structural model (supplementary file **model.pdb**) of S protein of SARS-CoV-2.