

例題 3.7

Covid-19 の原因ウイルス (SARS-CoV-2) のスパイクタンパク質について検索を実行する。スパイクタンパク質は、抗体の主なターゲットとなる部分で、ウイルス感染のための研究対象として重要である。SARS-CoV-2 のスパイクタンパク質の構造については、PDB の今月の分子 (翻訳版) を参照して欲しい。SARS-CoV-2 のスパイクタンパク質の Swiss-Prot のデータ識別名は SPIKE_SARS2、アクセッション番号は P0DTC2 である。この配列に対して BLAST 検索を実行する。検索対象の DB は Swiss-Prot を選択する。検索の結果、トップに現れるのは、SARS-CoV-2 のスパイクタンパク質の配列そのものだが、その次にヒットするのは、2003 年に報告された重症急性呼吸器症候群 (SARS) の原因ウイルスのスパイクタンパク質である。両者のアラインメントをとり、以下の 2 カ所の部位について、UniProtKB の内容を参考に調べよ。

1. ACE2 (アンギオテンシン変換酵素 2) の認識部位。ACE2 は肺、心臓、腎臓、腸の細胞で見られ、これらの細胞がウイルスによる感染の対象となる。
2. SARS-CoV-2 では、フリリンによる切断部位の周辺に 4 残基 (12 塩基) の挿入が見られ、その部分の多塩基性により、多様なプロテアーゼで切断される可能性が高まり、それだけ多くの組織で増殖することになる。

解答

Swiss-Prot の配列 SPIKE_SARS2 をクエリとし、Swiss-Prot データベースに対して BLAST 検索を行う。

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange

Or, upload file ファイルが選択されていません。

Job Title

Choose Search Set

Databases Standard databases (nr etc.) Experimental databases

Compare Select to compare standard and experimental database

Standard Database

Program Selection

Algorithm blastp (protein-protein BLAST)

BLAST Search Database swissprot using Blastp (protein-protein BLAST)

検索結果は以下の通り。

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

select all 50 sequences selected

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Severe acute res...	Severe acute res...	2553	2553	99%	0.0	100.00%	1273	P0DTC2.1
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Severe acute res...	Severe acute res...	1973	1973	98%	0.0	77.04%	1255	P59594.1
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Bat SARS coron...	Bat SARS coron...	1935	1935	98%	0.0	76.57%	1242	Q3LZX1.1
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Bat SARS CoV R...	Bat SARS CoV R...	1924	1924	98%	0.0	75.93%	1241	Q3I5J5.1
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Bat CoV 279/2005	Bat CoV 279/2005	1910	1910	98%	0.0	75.38%	1241	Q00475.1
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Piroistrellus bat c...	Piroistrellus bat c...	582	582	77%	0.0	37.44%	1352	A3EXD0.1
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Roussetus bat c...	Roussetus bat c...	572	572	94%	0.0	32.08%	1274	A3EXG6.1
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Bat coronavirus (...)	Bat coronavirus (...)	561	561	78%	3e-176	36.61%	1350	Q004F2.1
RecName: Full=Spike glycoprotein; Short=S.glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Betacoronavirus (...)	Betacoronavirus (...)	534	534	75%	4e-166	35.32%	1353	K9N5Q8.1

2 番目にヒットした配列 P59594.1 (SPIKE_SARS) は SARS の原因ウイルスのスパイクタンパク質である。両者のアラインメントは以下の通り。

Sequence ID: [P59594.1](#) Length: 1255 Number of Matches: 1

Range 1: 28 to 1255 [GenPept](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1973 bits(5112)	0.0	Compositional matrix adjust.	963/1250(77%)	1097/1250(87%)	24/1250(1%)
Query 26		PAYTN--SFTRGVVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSNGTKRFDNPV			83
Sbjct 28		P Y T S RGVVYYPD++FRS L+ TQDLFLPF+SNVT FH I+ F NPV			80
Query 84		LPFNDGVYFASTEKSNIRGWIFGTTLDLSKTQSLIVNNATNVVIKVFCEQFCNDPFLGV			143
Sbjct 81		+PF DG+YFA+TEKSN++RGW+FG+T+++K+QS++I+NN+TNVVI+ C F+ C++PF V			140
Query 144		YHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEKQGNFKNLREFVFKNIDGYFKI			203
Sbjct 141		+ ++ ++ +A NCTFEY+S F +D+ K GNFK+LREFVFKN DG+ +			196
Query 204		YSKHTPINLVRDLPPQGFSALEPLVDLPIGINITRFQTLALHRSYLTPGDSSSGWTAGAA			263
Sbjct 197		Y + PI++VRDLP GF+ L+P+ LP+GINIT F+ +L + +P W AA			250
Query 264		AYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCTLKSFVEKGIYQTSNFRVQPT			323
Sbjct 251		AY+VGYL+P TF+LKY+ENGTITDAVDC+ +PL+E KC++KSF ++KGIYQTSNFRV P+			310
Query 324		ESIVRFPNITNLCPPFGEVFNATRFASVYAWNKRKISNCVADYSVLYNSASFSTFKCYGVS			383
Sbjct 311		+VRFPNITNLCPPFGEVFNAT+F SVYAW RK+ISNCVADYSVLYNS FSTFKCYGVS			370
Query 384		PTKLNLDLFCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNLDL			443
Sbjct 371		TKLNLDLFC+NVYADSFV++GD+VRQIAPGQTG IADYNYKLPDDF GCV+AWN+ N+D+			430
Query 444		KVGGNYNYLRLFRKSNLKPFFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGV			503
Sbjct 431		GNVNY YR R L+PFFERDIS + PC NCY+PL YGF T G+			489
Query 504		GYQPYRVVVLSEFELLHAPATVCGPKKSTNLVKNKCVNENFNGLTGTGVLTESNKKFLPFQ			563
Sbjct 490		GYQPYRVVVLSEFELL+APATVCGPK ST+L+KN+CVNENFNGLTGTGVL T S+K+F PFQ			549
Query 564		QFGRDIADTTDAVRDPQTEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVA			623
Sbjct 550		QFGRD++D TD+VRDP+T EILDI+PCSFGGVSVITPGTN S++VAVLYQDVNCT+V A			609
Query 624		IHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNSYECDIPIGAGICASYQTQNSPRR			683
Sbjct 610		IHADQLTP WR+YSTG+NVFQT+AGCLIGAEHV+ SYECDIPIGAGICASY T +			665
Query 684		ARSVASQSI IAYTMSLGAENSVAYSNNIAIPTNFTISVTEILPVSMTKTSVDCTMYIC			743
Sbjct 666		RS + +SI+AYTMSLGA++S+AYSNN+IAIPTNF+IS+TTE++PVSM KTSVDC MYIC			725

UniProtKB の P0DTC2 (SPIKE_SARS2) と P59594 (SPIKE_SARS) のアノテーション情報を比較する。

- (1) ACE2 結合モチーフとして、SPIKE_SARS2 では 437-508、SPIKE_SARS では 424-494 が登録されている。上記のアラインメントで一致部分が少ないことがわかる。
- (2) フリン切断部位として、SPIKE_SARS2 では 685-686、SPIKE_SARS では 667-668 が登録されている。その N 末側に SPIKE_SARS では 4 残基の挿入が見られる。

Query	744	GDSTECNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFS	803
Sbjct	726	GDSTEC+NLLLQYGSFCTQLNRAL+GIA EQD+NT+EVFAQVKQ+YKTP +K FGGFNFS	785
Query	804	QILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPP	863
Sbjct	786	QILPDP KP+KRSFIEDLLFNKVTLADAGF+KQYG+CLGDI ARDLICAQKFNGLTVLPP	845
Query	864	LLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLI	923
Sbjct	846	LLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKI	905
Query	924	ANQFNSAIGKIQDLSSTASALGKLDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSR	983
Sbjct	906	ANQFN AI +IQ+SL++T++ALGKLDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSR	965
Query	984	LDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFC	1043
Sbjct	966	LDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFC	1025
Query	1044	GKGYHLMSFPQSAPHGVVFLHVTVVPAQEKNFTTAPAICHGKAHFPREGVFSNGTHWF	1103
Sbjct	1026	GKGYHLMSFPQ+APHGVVFLHVTVVPAQE+NETTAPAICH+GKA+FPREGVFW NGT WF	1085
Query	1104	VTQRNFYEPQIIITDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD	1163
Sbjct	1086	+TQRNF+ PQIIITDNTFVSGNCDVVIGI+NNTVYDPLQPELDSFKEELDKYFKNHTSPD	1145
Query	1164	VLDGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAG	1223
Sbjct	1146	VLDGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWY+WLGFIAG	1205
Query	1224	LIAIVMVTImlcmtscsclkgccscgscckFDEDDSEPVLKGVKLHYT	1273
Sbjct	1206	LIAIVMVTI+LCCMTSCCSCLKG CSCGSCCKFDEDDSEPVLKGVKLHYT	1255