wwPDB EM Validation Summary Report (i)

Mar 6, 2026 – 11:48 AM UTC

PDB ID	:	6X6J / pdb_00006x6j
EMDB ID	:	EMD-20021
Title	:	Cryo-EM Structure of CagX and CagY within the Helicobacter pylori PR
Authors	:	Sheedlo, M.J.; Chung, J.M.; Sawhney, N.; Durie, C.L.; Cover, T.L.; Ohi, M.D.; Lacy, D.B.
Deposited on	:	2020-05-28
Resolution	:	3.50 Å(reported)

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

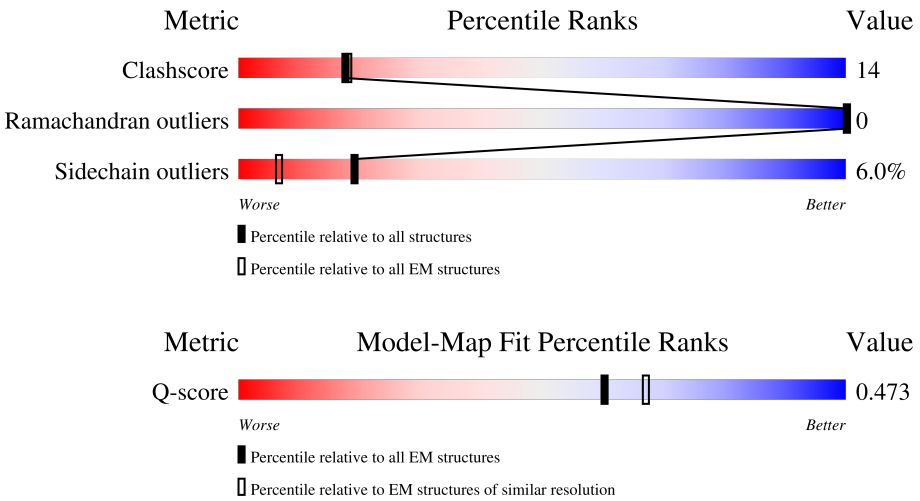
EMDB validation analysis	:	0.0.1.dev132
MolProbity	:	4-5-2 with Phenix2.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
EM percentile statistics	:	202505.v01 (Using data in the EMDB archive up until May 2025)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.50 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	EM structures (#Entries)	Similar EM resolution (#Entries, resolution range(Å))
Clashscore	229148	23984	-
Ramachandran outliers	224038	23583	-
Sidechain outliers	223484	23102	-
Q-score	-	25397	13950 (3.00 - 4.00)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	AX	521	20% 10% • 69%
1	BX	521	21% 9% • 69%
1	CX	521	20% 11% • 69%
1	DX	521	18% 12% • 69%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	EX	521	20% 11% 69%
1	FX	521	20% 10% 69%
1	GX	521	19% 11% 69%
1	HX	521	20% 11% 69%
1	IX	521	21% 9% 69%
1	JX	521	18% 12% 69%
1	KX	521	20% 9% 69%
1	LX	521	21% 9% 69%
1	MX	521	21% 9% 69%
1	NX	521	19% 12% 69%
1	OX	521	20% 10% 69%
1	PX	521	20% 11% 69%
1	QX	521	20% 11% 69%
2	AY	1927	5% 93%
2	BY	1927	93%
2	CY	1927	5% 93%
2	DY	1927	5% 93%
2	EY	1927	5% 93%
2	FY	1927	5% 93%
2	GY	1927	5% 93%
2	HY	1927	5% 93%
2	IY	1927	5% 93%
2	JY	1927	93%
2	KY	1927	5% 93%
2	LY	1927	5% 93%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	MY	1927	5% 93%
2	NY	1927	5% 93%
2	OY	1927	5% 93%
2	PY	1927	5% 93%
2	QY	1927	5% 93%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 40613 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Cag pathogenicity island protein (Cag8).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	AX	164	1339	858	227	252	2	0	0
1	BX	164	1339	858	227	252	2	0	0
1	CX	164	1339	858	227	252	2	0	0
1	DX	164	1339	858	227	252	2	0	0
1	EX	164	1339	858	227	252	2	0	0
1	FX	164	1339	858	227	252	2	0	0
1	GX	164	1339	858	227	252	2	0	0
1	HX	164	1339	858	227	252	2	0	0
1	IX	164	1339	858	227	252	2	0	0
1	JX	164	1339	858	227	252	2	0	0
1	KX	164	1339	858	227	252	2	0	0
1	LX	164	1339	858	227	252	2	0	0
1	MX	164	1339	858	227	252	2	0	0
1	NX	164	1339	858	227	252	2	0	0
1	OX	164	1339	858	227	252	2	0	0
1	PX	164	1339	858	227	252	2	0	0
1	QX	164	1339	858	227	252	2	0	0

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AX	?	-	GLU	deletion	UNP O25263
AX	516	GLU	LEU	conflict	UNP O25263
BX	?	-	GLU	deletion	UNP O25263
BX	516	GLU	LEU	conflict	UNP O25263
CX	?	-	GLU	deletion	UNP O25263
CX	516	GLU	LEU	conflict	UNP O25263
DX	?	-	GLU	deletion	UNP O25263
DX	516	GLU	LEU	conflict	UNP O25263
EX	?	-	GLU	deletion	UNP O25263
EX	516	GLU	LEU	conflict	UNP O25263
FX	?	-	GLU	deletion	UNP O25263
FX	516	GLU	LEU	conflict	UNP O25263
GX	?	-	GLU	deletion	UNP O25263
GX	516	GLU	LEU	conflict	UNP O25263
HX	?	-	GLU	deletion	UNP O25263
HX	516	GLU	LEU	conflict	UNP O25263
IX	?	-	GLU	deletion	UNP O25263
IX	516	GLU	LEU	conflict	UNP O25263
JX	?	-	GLU	deletion	UNP O25263
JX	516	GLU	LEU	conflict	UNP O25263
KX	?	-	GLU	deletion	UNP O25263
KX	516	GLU	LEU	conflict	UNP O25263
LX	?	-	GLU	deletion	UNP O25263
LX	516	GLU	LEU	conflict	UNP O25263
MX	?	-	GLU	deletion	UNP O25263
MX	516	GLU	LEU	conflict	UNP O25263
NX	?	-	GLU	deletion	UNP O25263
NX	516	GLU	LEU	conflict	UNP O25263
OX	?	-	GLU	deletion	UNP O25263
OX	516	GLU	LEU	conflict	UNP O25263
PX	?	-	GLU	deletion	UNP O25263
PX	516	GLU	LEU	conflict	UNP O25263
QX	?	-	GLU	deletion	UNP O25263
QX	516	GLU	LEU	conflict	UNP O25263

- Molecule 2 is a protein called Cag pathogenicity island protein (Cag7).

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	AY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	BY	135	1050	649	177	220	4	0	0

Continued on next page...

Continued from previous page...

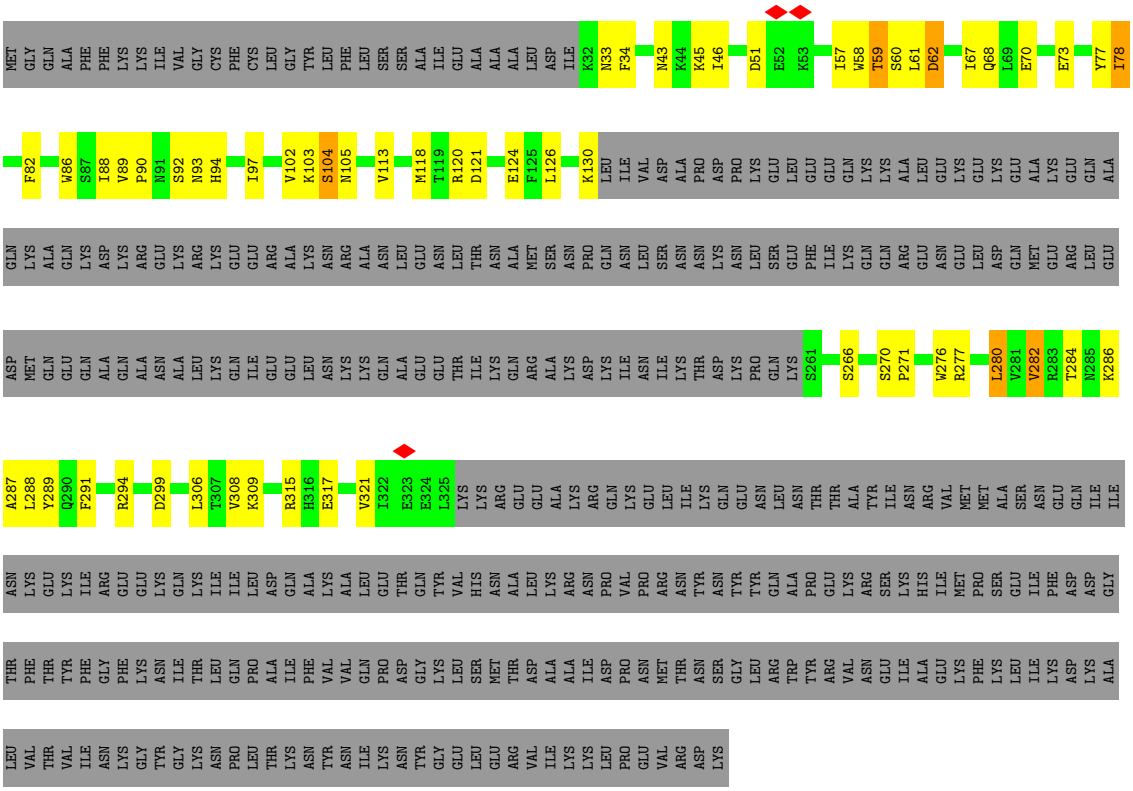
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	CY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	DY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	EY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	FY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	GY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	HY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	IY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	JY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	KY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	LY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	MY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	NY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	OY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	PY	135	1050	649	177	220	4	0	0
			Total	C	N	O	S		
2	QY	135	1050	649	177	220	4	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

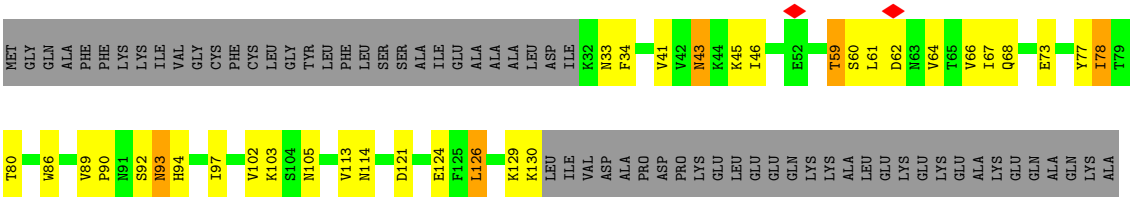
• Molecule 1: Cag pathogenicity island protein (Cag8)

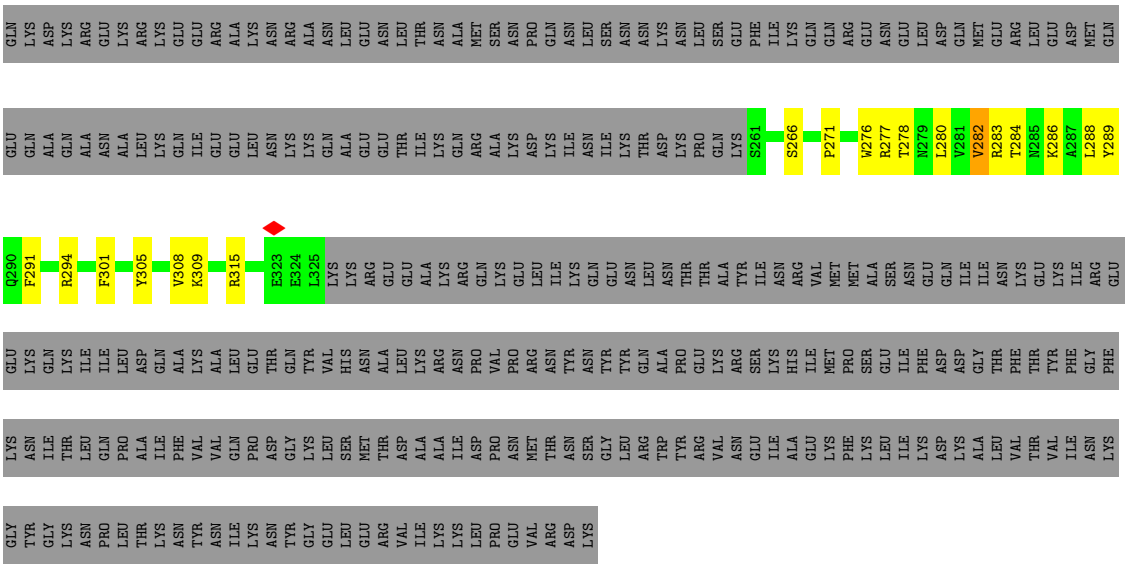
Chain AX:



• Molecule 1: Cag pathogenicity island protein (Cag8)

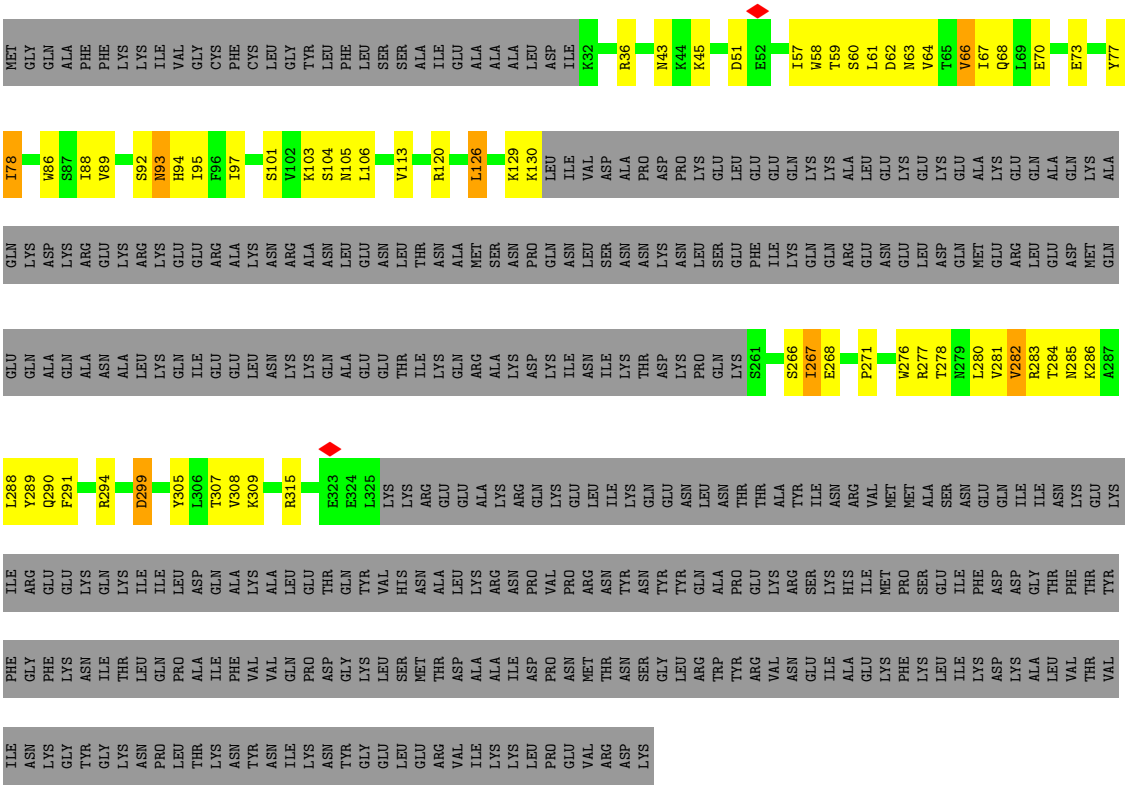
Chain BX:





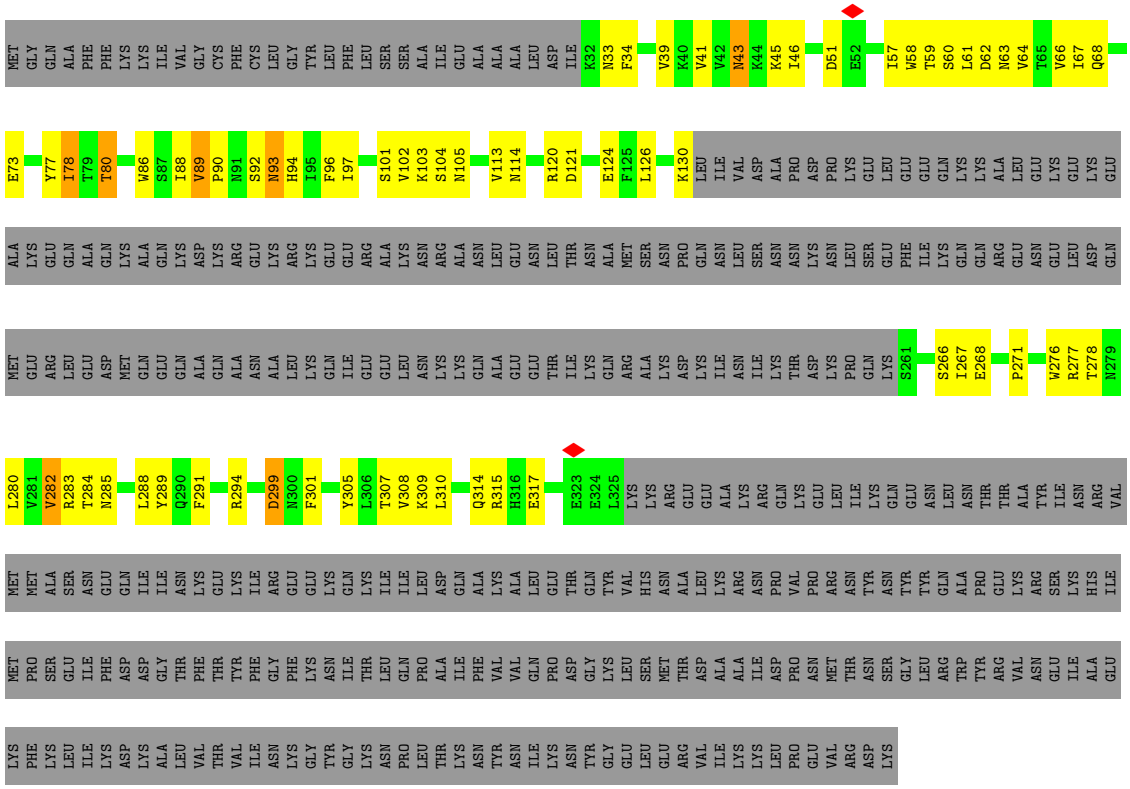
• Molecule 1: Cag pathogenicity island protein (Cag8)

Chain CX:



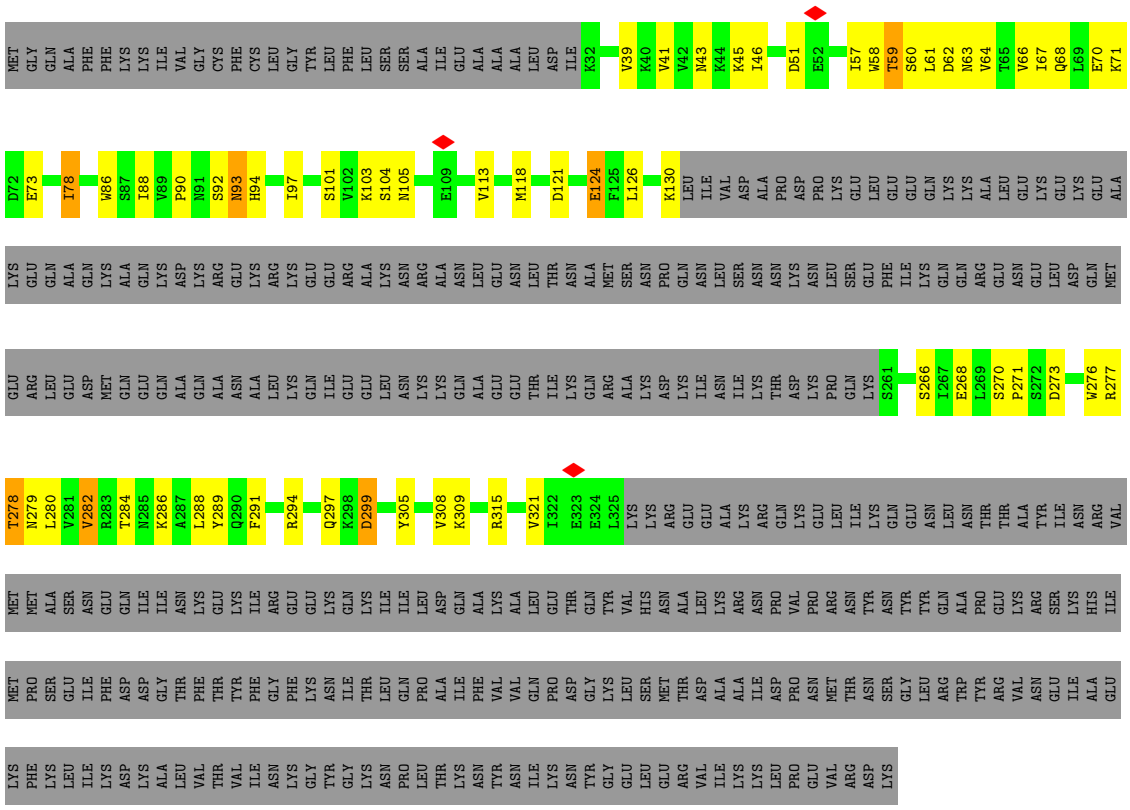
• Molecule 1: Cag pathogenicity island protein (Cag8)

Chain DX:



- Molecule 1: Cag pathogenicity island protein (Cag8)

Chain EX:



- Molecule 1: Cag pathogenicity island protein (Cag8)

Chain FX: 20% 10% • 69%

MET GLY GLN ALA PHE PHE LYS LYS ILE VAL GLY CYS PHE CYS LEU GLY TYR LEU PHE LEU SER SER ALA ILE GLU ALA ALA LEU ASP ILE K32 R36 V41 V42 N43 K44 K45 I46 E52 I57 W58 T59 S60 L61 D62 N63 V64 T65 V66 I67 Q68 E73 Y77

I78 T79 T80 W86 S87 I88 V89 P90 N91 S92 N93 H94 I95 F96 I97 S101 V102 K103 S104 N105 V113 D121 E124 F125 L126 K130 LEU ILE VAL ASP ALA PRO ASP PRO LYS GLU LEU GLU GLU GLN LYS LYS ALA LEU GLU LYS GLU LYS GLU ALA LYS GLU GLN ALA

GLN LYS ALA GLN LYS ASP LYS ARG GLU LYS ARG LYS GLU GLU ARG ALA LYS ASN ARG ALA ASN LEU GLU ASN LEU THR ASN ALA MET SER ASN PRO GLN ASN LEU SER ASN ASN LYS ASN LEU SER GLU PHE ILE LYS GLN GLN ARG GLU ASN GLU GLU ASP GLN MET GLU ARG LEU GLU

ASP MET GLN GLU GLN ALA GLN ALA ASN ALA ALA LEU LYS GLN ILE GLU GLU LEU ASN LYS LYS GLN ALA GLU GLU THR ILE ILE LYS GLN ARG ALA LYS ASP LYS ILE ASN ILE LYS THR ASP LYS PRO GLN LYS S261 D264 N265 S266 I267 P271 W276 R277 T278 N279 L280 V281 V282 R283

T284 N285 K286 Y289 R294 F301 Y305 V308 K309 L310 R315 L325 LYS LYS ARG GLU GLU ALA LYS ARG GLN LYS GLU LEU ILE LYS GLN GLU ASN LEU ASN THR THR ALA TYR ILE ASN ARG VAL MET MET ALA SER ASN GLU GLN ILE ILE ASN LYS GLU LYS ILE

ARG GLU GLU LYS GLN LYS ILE ILE LEU ASP GLN ALA LYS ALA LEU GLU THR GLN TYR VAL HIS ASN ALA LEU LYS ARG ASN PRO VAL PRO ARG ASN TYR ASN TYR TYR TYR GLN ALA PRO GLU LYS ARG SER LYS HIS ILE MET PRO SER GLU ILE PHE ASP ASP GLY THR PHE THR TYR PHE

GLY PHE LYS ASN ILE THR LEU GLN PRO ALA ILE PHE VAL VAL GLN PRO ASP GLY LYS LEU SER MET THR ASP ALA ALA ILE ASP PRO ASN MET THR ASN SER GLY LEU ARG TRP TYR ARG VAL ASN GLU ILE ALA GLU LYS PHE LYS LEU ILE LYS ASP LYS ALA LEU VAL THR VAL ILE

ASN LYS GLY TYR GLY LYS ASN PRO LEU THR LYS ASN TYR ASN ILE LYS ASN TYR GLY GLU LEU GLU ARG VAL ILE LYS LYS LEU PRO GLU VAL ARG ASP LYS

- Molecule 1: Cag pathogenicity island protein (Cag8)

Chain GX: 19% 11% • 69%

MET GLY GLN ALA PHE PHE LYS LYS ILE VAL GLY CYS PHE CYS LEU GLY TYR LEU PHE LEU SER SER ALA ILE GLU ALA ALA LEU ASP ILE K32 R36 N43 K44 K45 I46 A47 Y48 E52 I57 W58 T59 S60 L61 D62 N63 V66 I67 Q68 L69 E70 K71 D72 E73

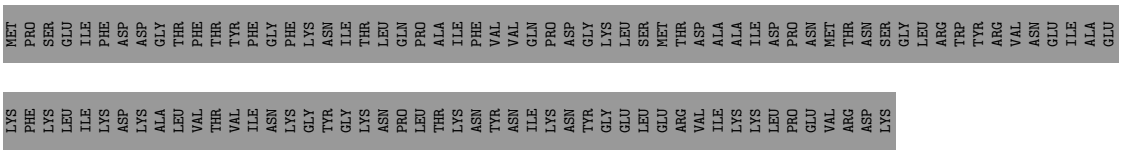
Y77 I78 W86 V89 P90 N91 S92 N93 H94 I95 F96 I97 K103 S104 N105 V113 D121 E124 F125 L126 K130 LEU ILE VAL ASP ALA PRO ASP PRO LYS GLU LEU GLU GLU GLN LYS LYS ALA LEU GLU LYS GLU LYS GLU ALA LYS GLU GLN ALA GLN LYS ALA

GLN LYS ASP LYS ARG GLU LYS ARG LYS GLU GLU ARG ALA LYS ASN ARG ALA ASN LEU GLU ASN LEU THR ASN ALA MET SER ASN PRO GLN ASN LEU SER ASN ASN LYS ASN LEU SER GLU PHE ILE LYS GLN GLN ARG GLU ASN GLU LEU ASP GLN MET GLU ARG LEU GLU ASP MET GLN

GLU GLN ALA GLN ALA ASN ALA LEU LYS GLN ILE GLU GLU LEU ASN LYS LYS GLN ALA GLU GLU THR ILE LYS GLN ARG ALA LYS ASP LYS ILE ASN ILE LYS THR ASP LYS PRO GLN LYS S261 D264 N265 S266 I267 E268 P271 S272 D273 W276 R277 T278 N279 L280 V281 V282 R283

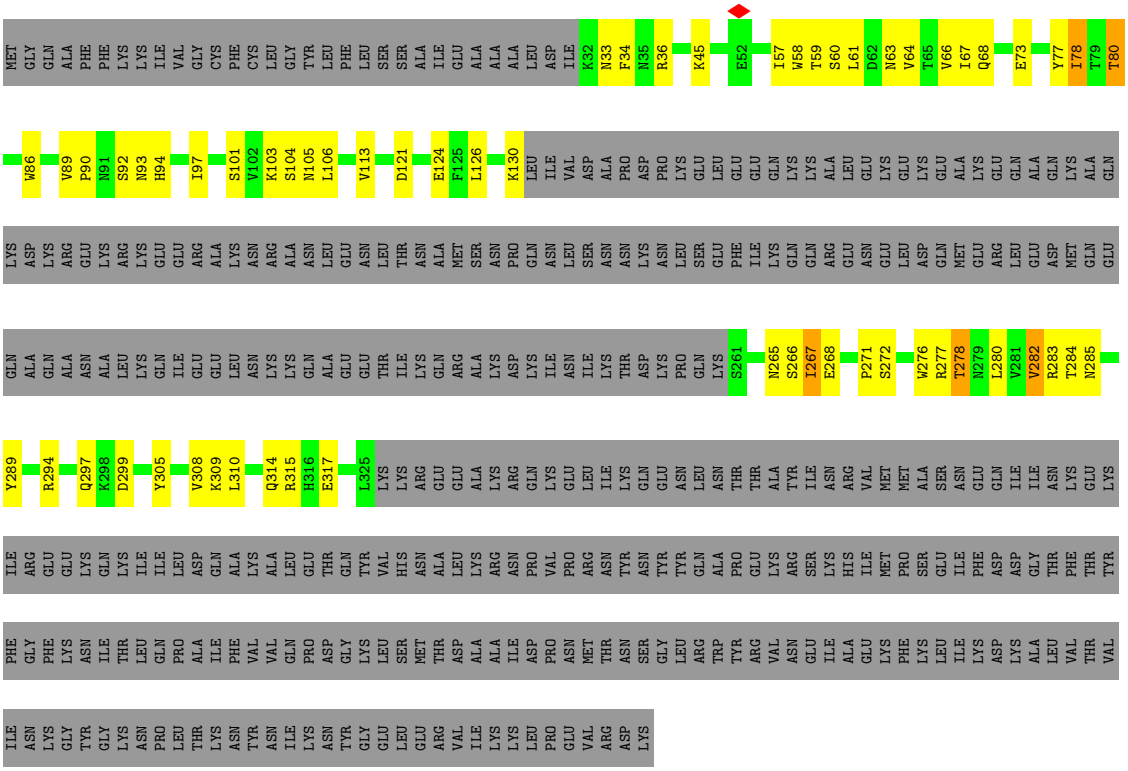
T284 N285 K286 Y289 Q290 F291 R294 Q297 K298 D299 N300 F301 Y305 L306 T307 V308 K309 L310 E311 Y312 P313 Q314 R315 H316 E317 E323 E324 L325 LYS LYS ARG GLU GLU ALA LYS ARG GLN LYS GLU LEU ILE LYS GLN GLU ASN LEU ASN THR THR ALA TYR ILE ASN ARG VAL

MET MET ALA SER ASN GLU GLN ILE ILE ASN LYS GLU LYS ILE ARG GLU GLU LYS GLN LYS ILE ILE LEU ASP GLN ALA LYS ALA LYS GLU GLU THR GLN TYR VAL HIS ASN ALA LEU LYS ARG ASN PRO VAL PRO ARG ASN TYR ASN GLN TYR TYR GLN ALA PRO GLU LYS ARG SER LYS HIS ILE



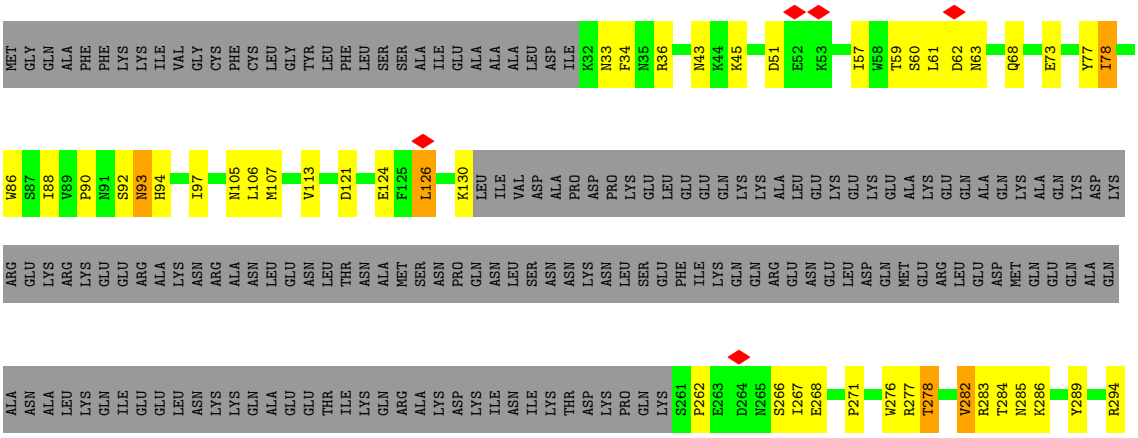
• Molecule 1: Cag pathogenicity island protein (Cag8)

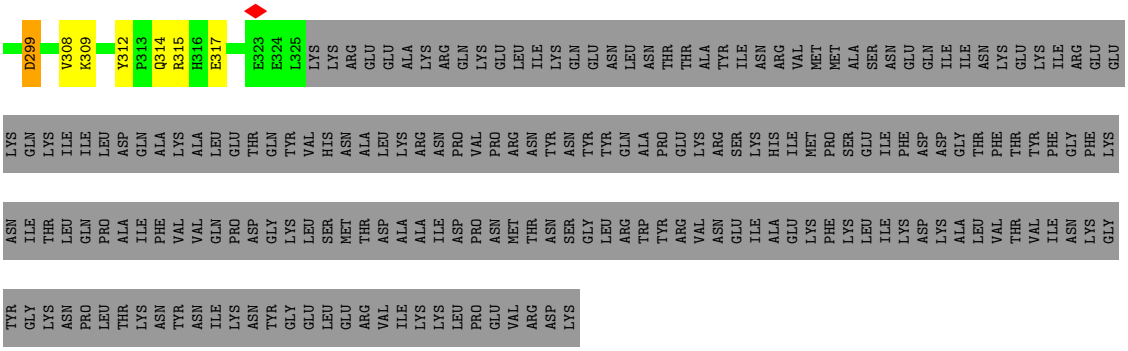
Chain HX:



• Molecule 1: Cag pathogenicity island protein (Cag8)

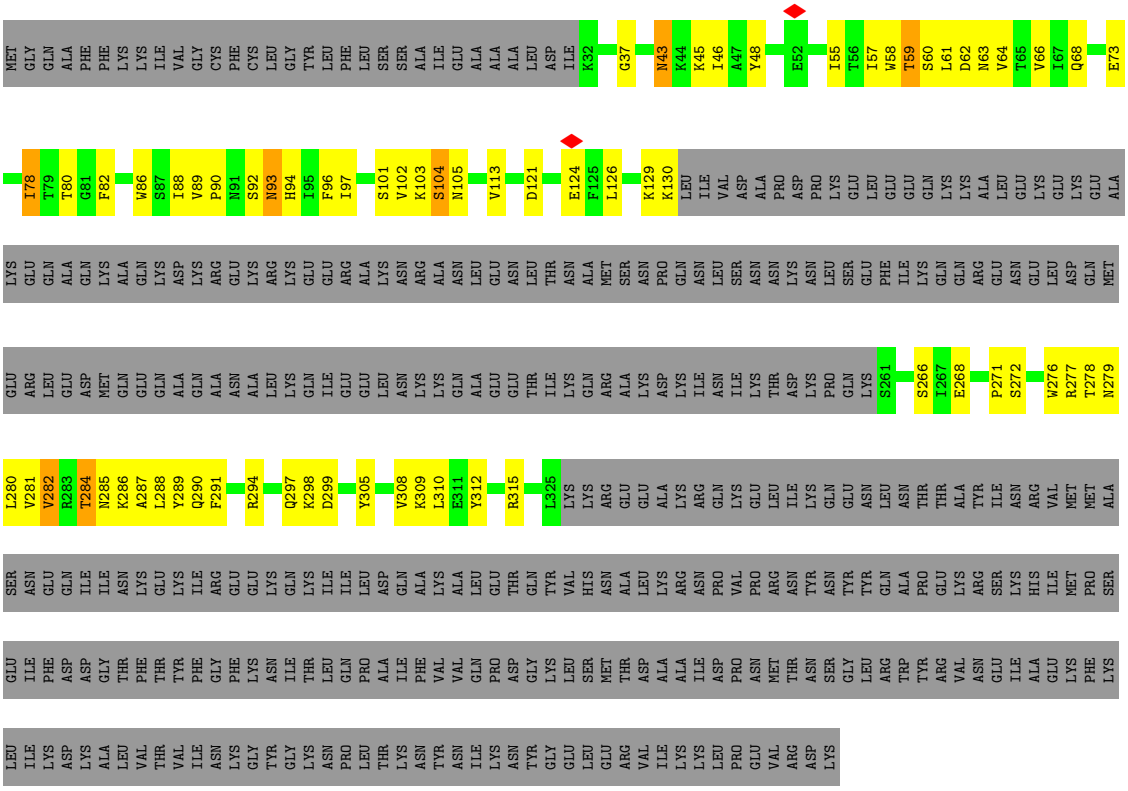
Chain IX:





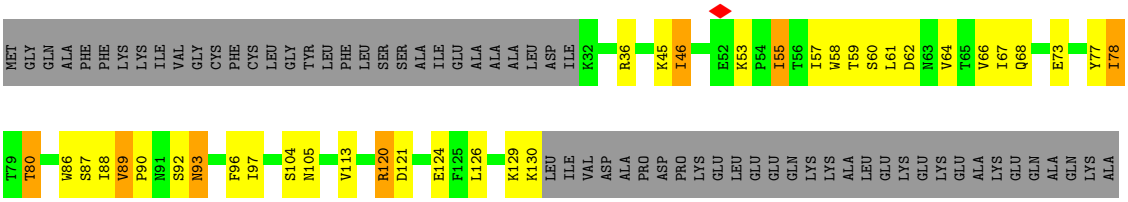
- Molecule 1: Cag pathogenicity island protein (Cag8)

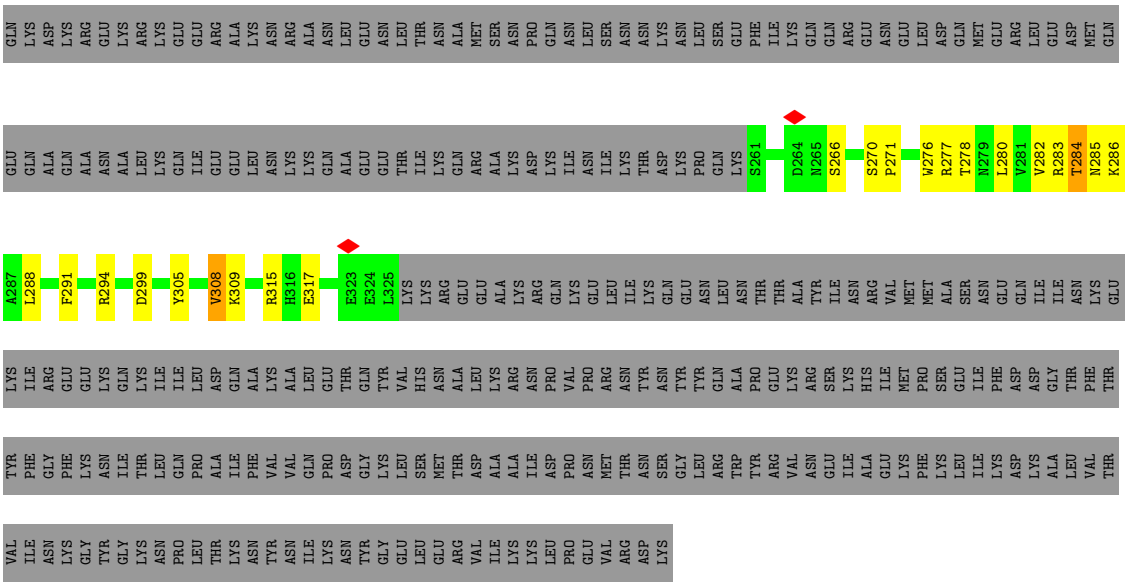
Chain JX: 18% 12% • 69%



- Molecule 1: Cag pathogenicity island protein (Cag8)

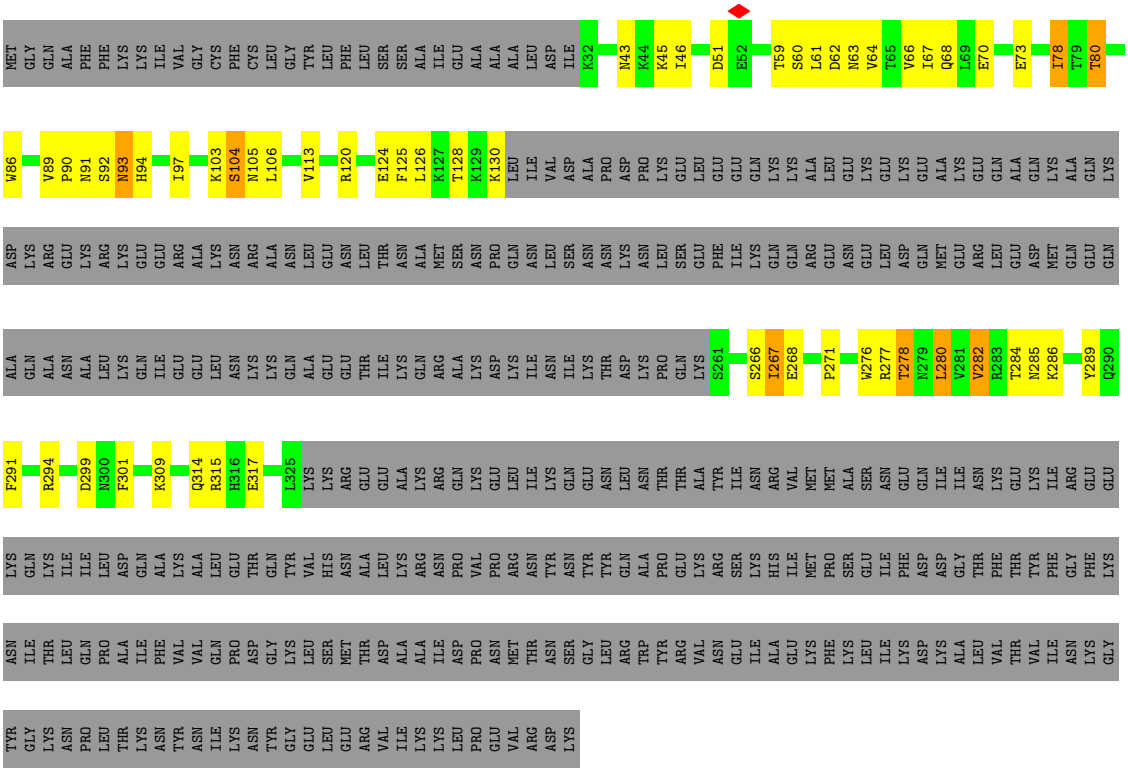
Chain KX: 20% 9% • 69%





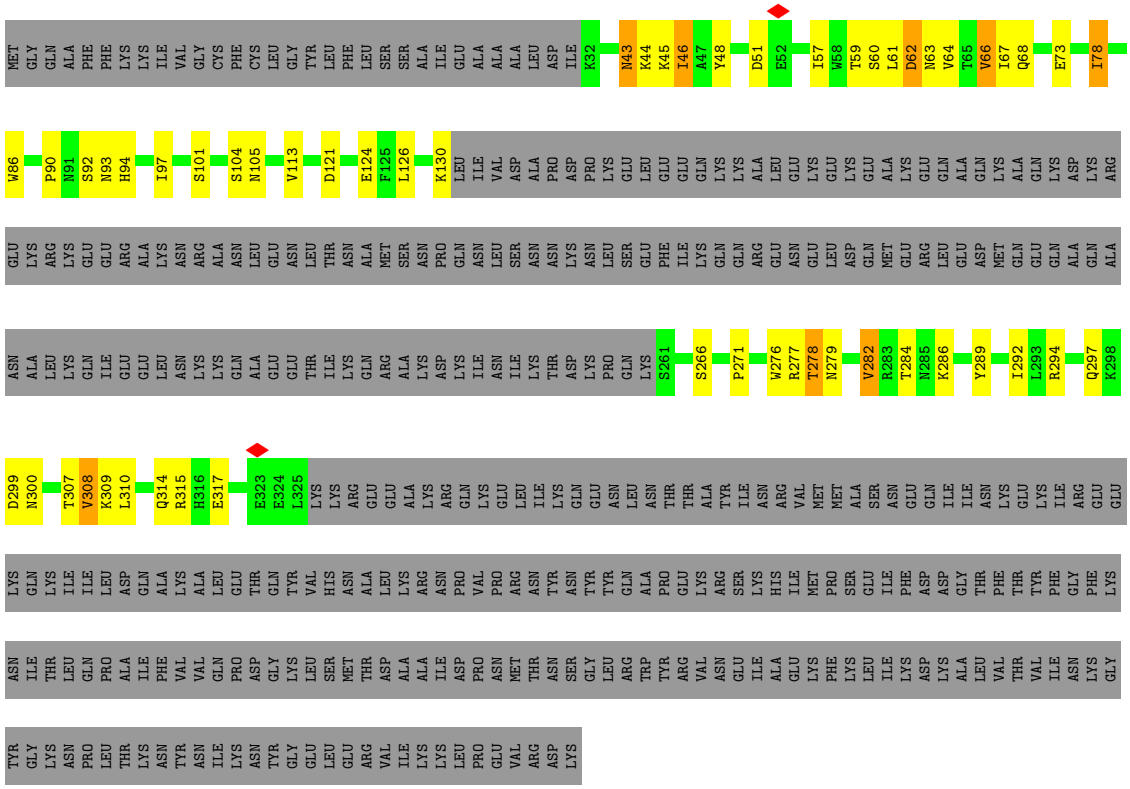
- Molecule 1: Cag pathogenicity island protein (Cag8)

Chain LX:



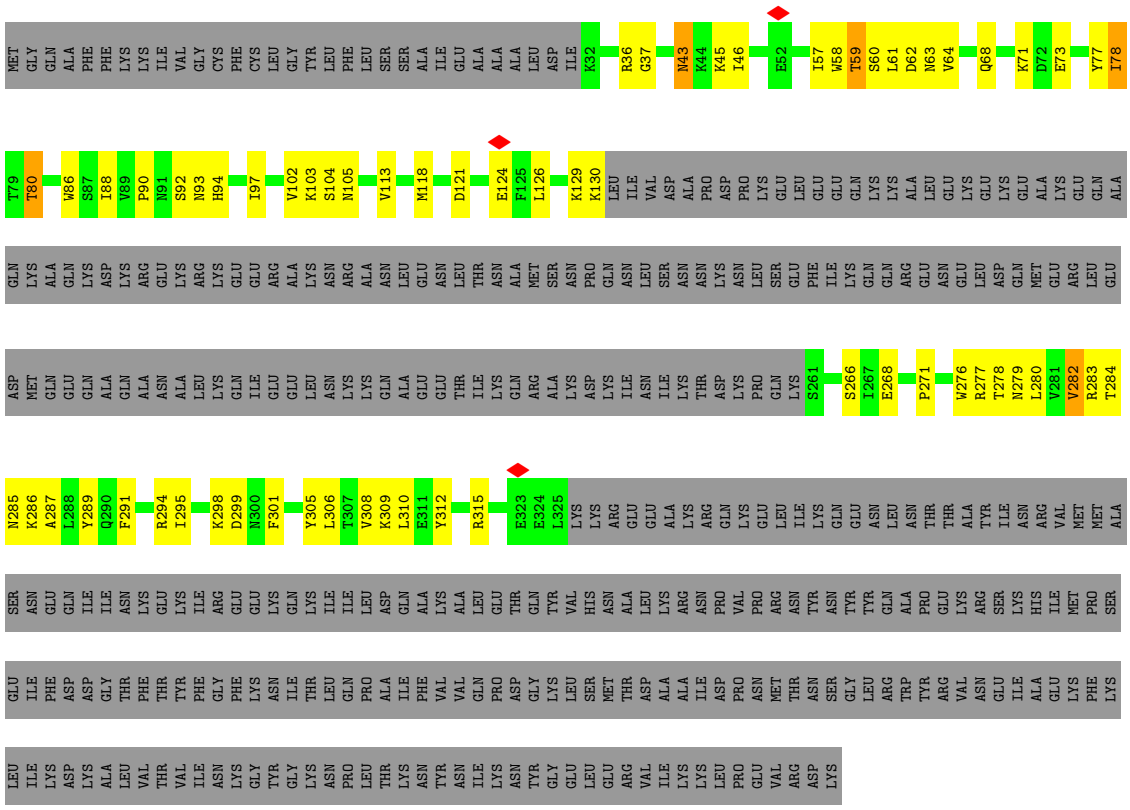
- Molecule 1: Cag pathogenicity island protein (Cag8)

Chain MX:



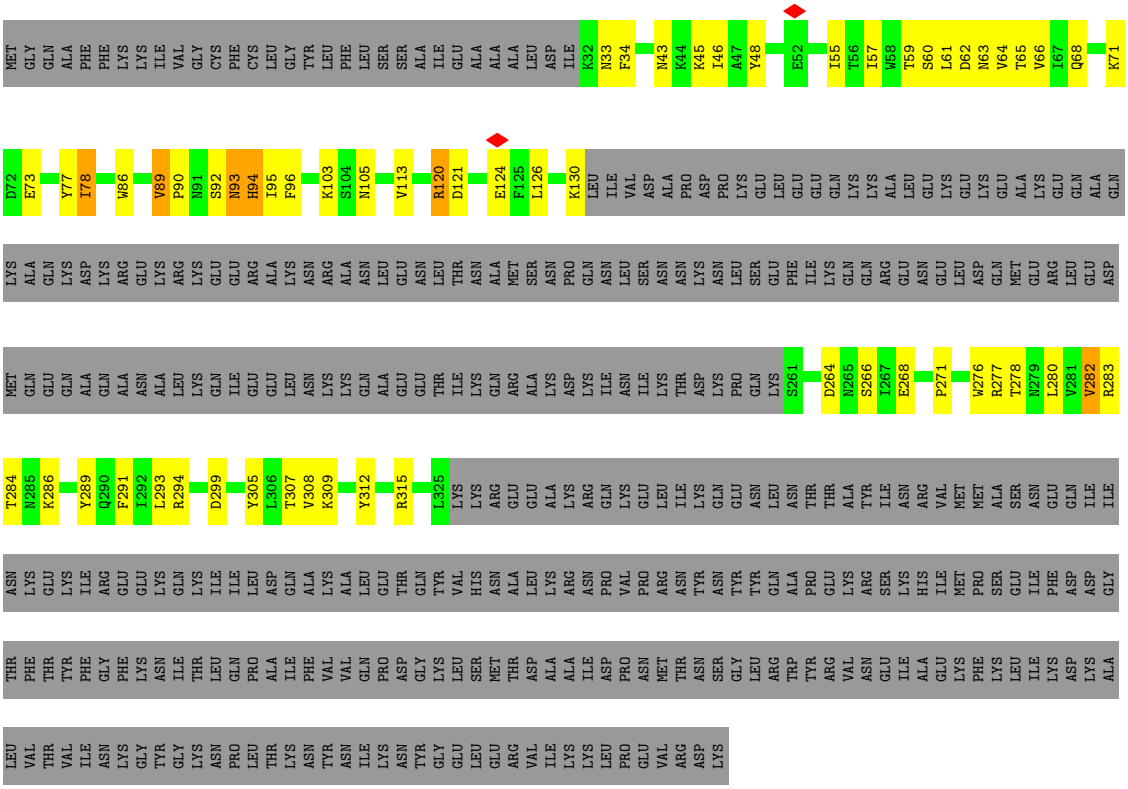
- Molecule 1: Cag pathogenicity island protein (Cag8)

Chain NX:



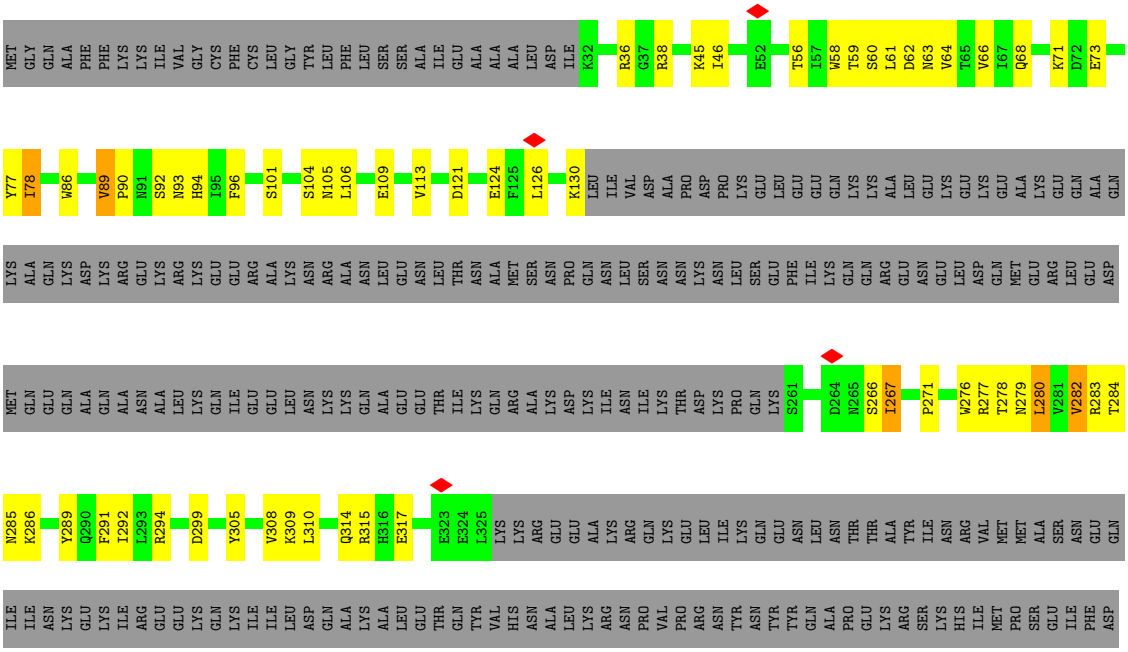
- Molecule 1: Cag pathogenicity island protein (Cag8)

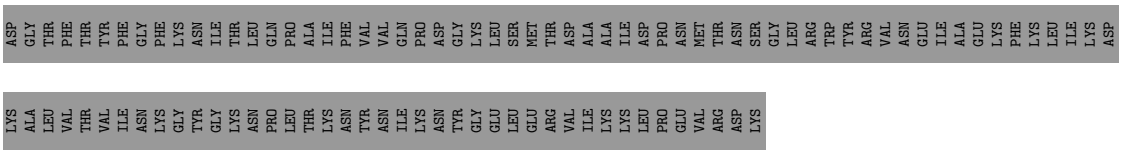
Chain OX:



- Molecule 1: Cag pathogenicity island protein (Cag8)

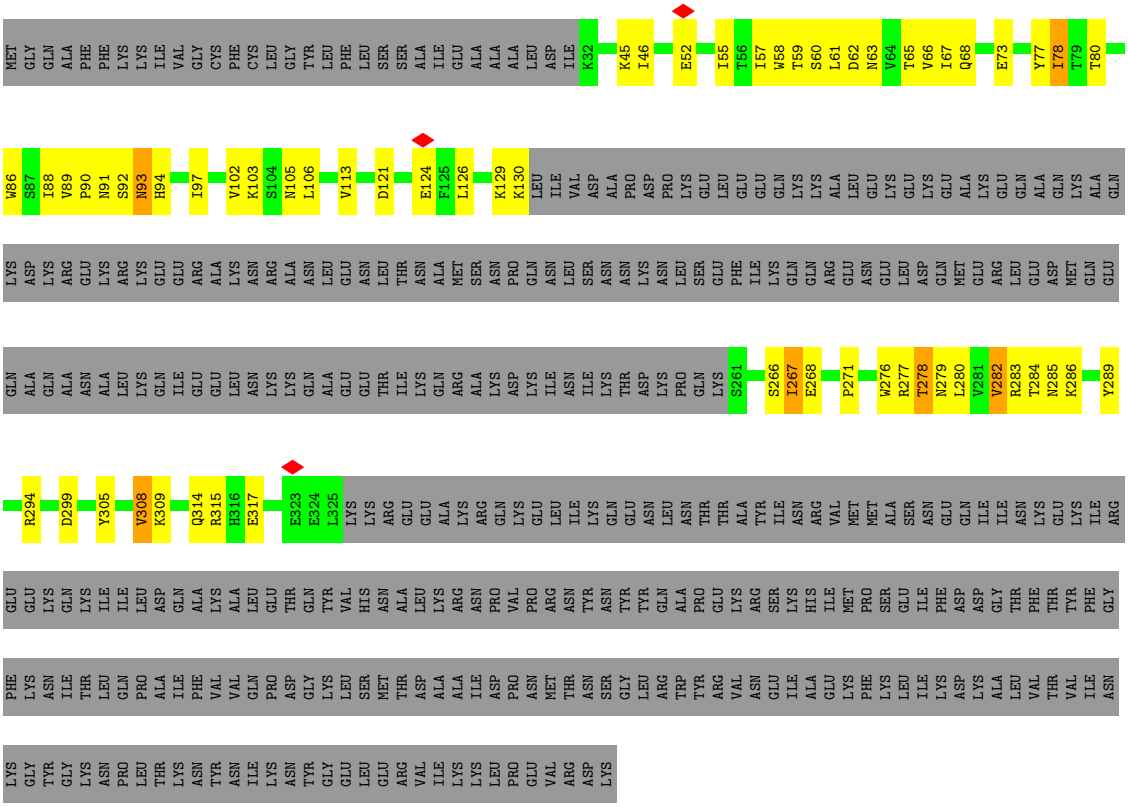
Chain PX:





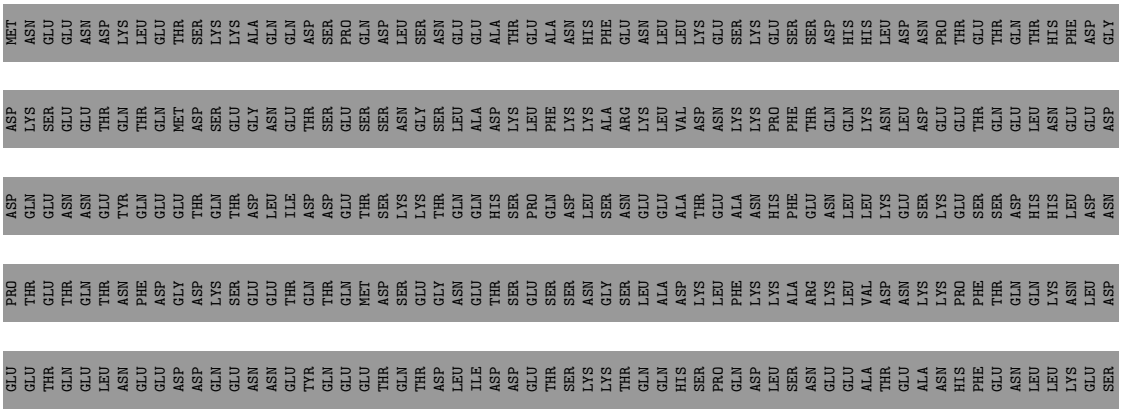
• Molecule 1: Cag pathogenicity island protein (Cag8)

Chain QX: 20% 11% 69%

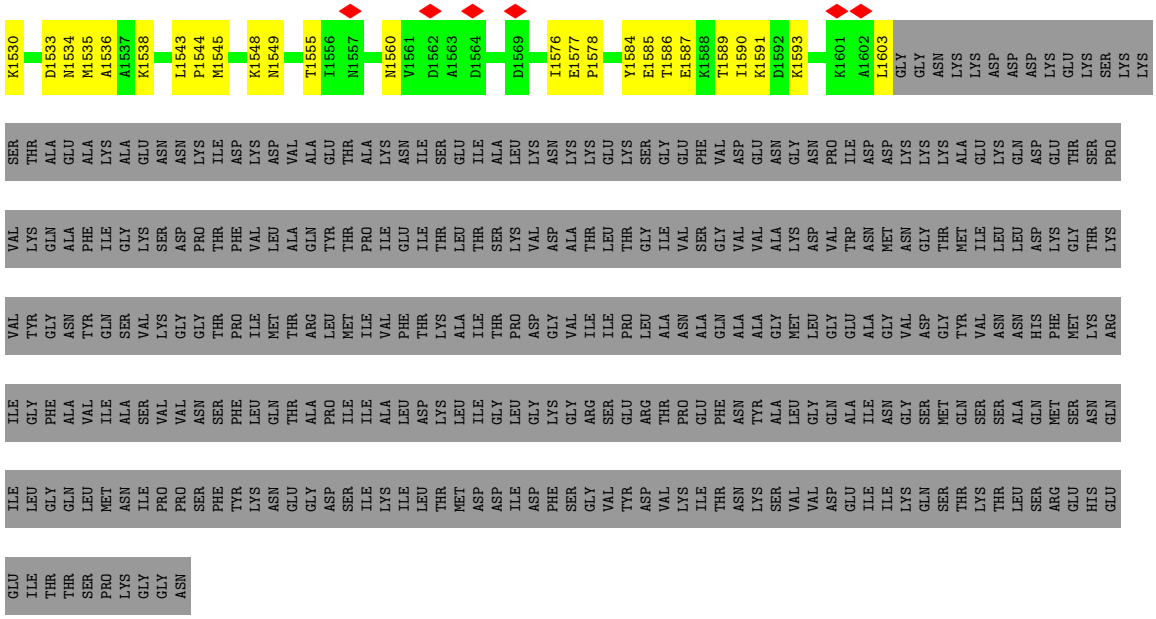


• Molecule 2: Cag pathogenicity island protein (Cag7)

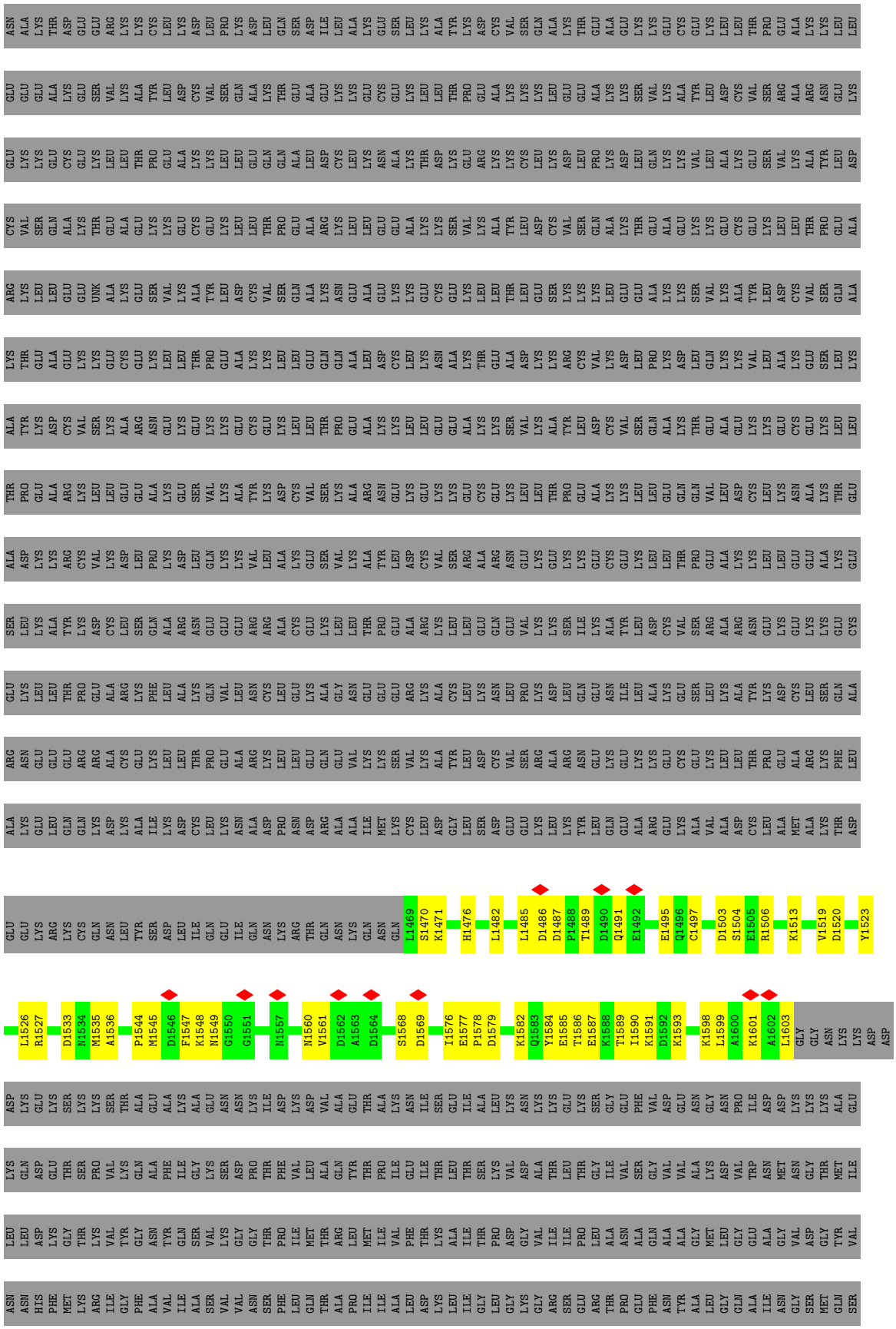
Chain AY: 5% 93%

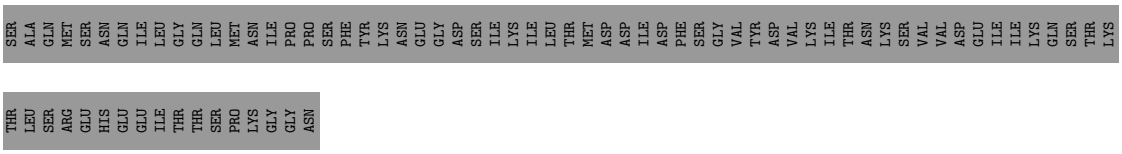


L1469
S1470
K1471
D1486
D1487
P1488
T1489
D1490
Q1491
E1492
C1497
S1502
D1503
S1504
E1505
R1506
I1509
R1514
E1518
V1519
D1520
Y1523
L1526



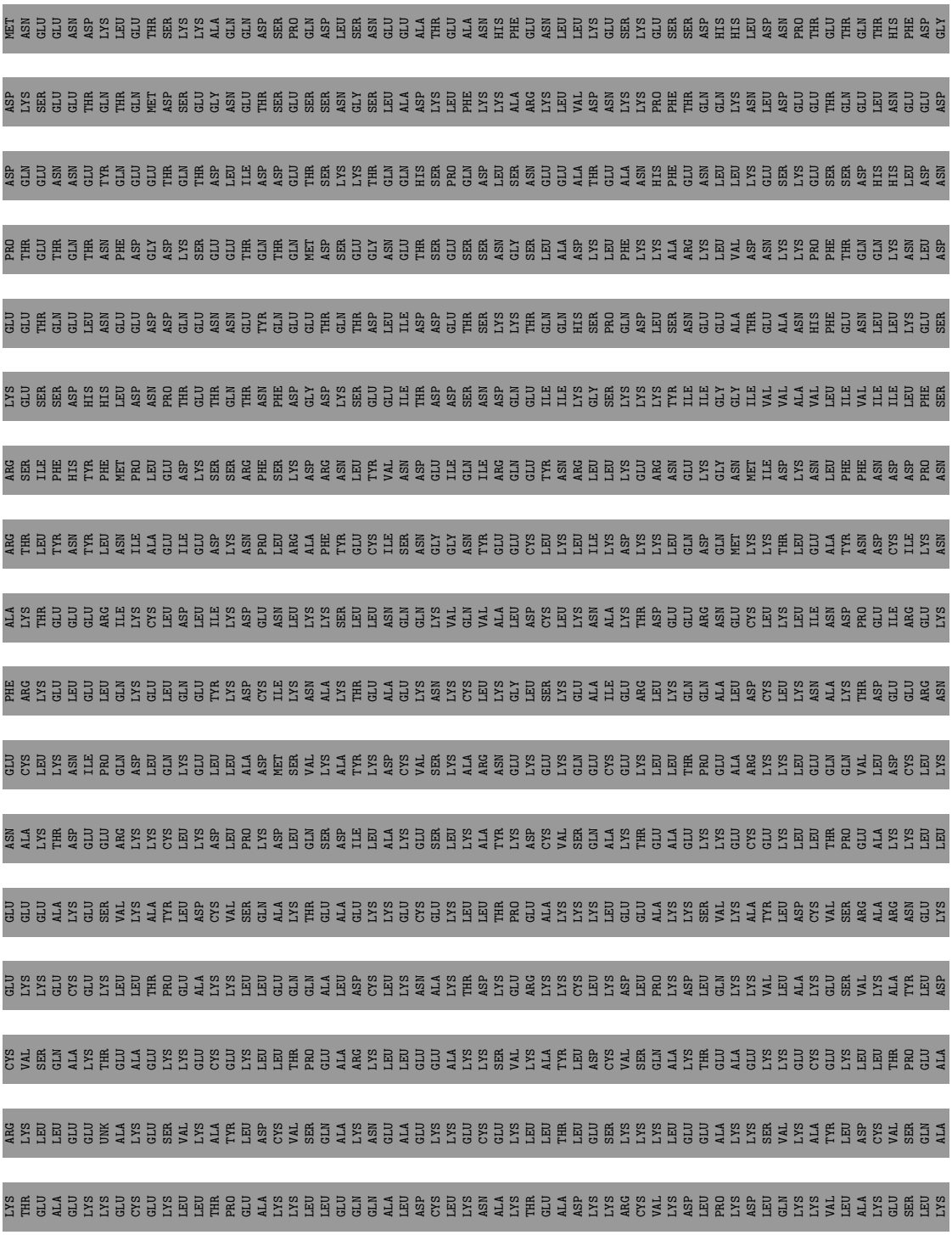
• Molecule 2: Cag pathogenicity island protein (Cag7)

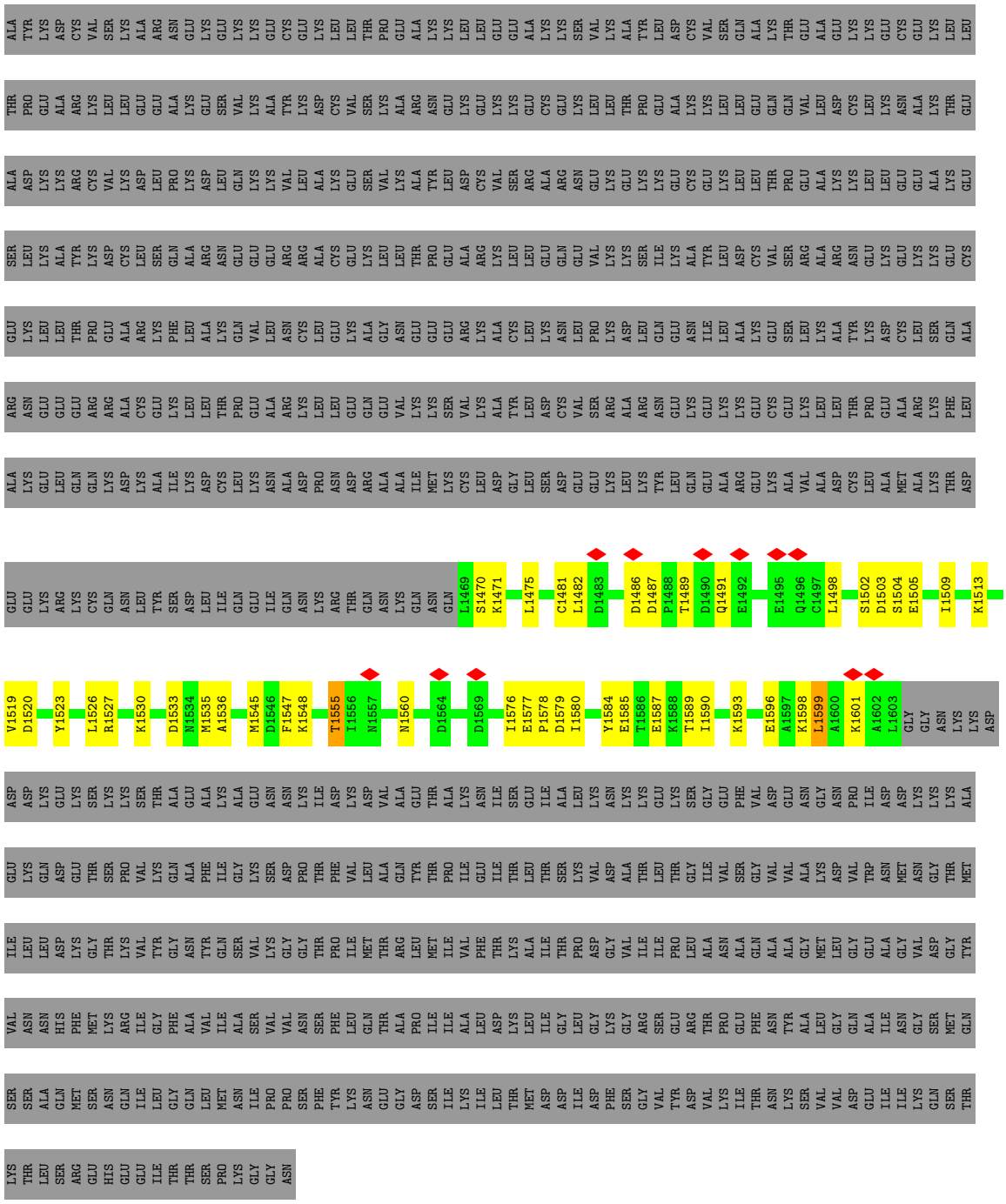

L1526 R1527 D1533 N1534 M1535 A1536 P1544 M1545 D1546 F1547 K1548 N1549 G1550 G1551 N1557 N1560 V1561 D1562 A1563 D1564 S1568 D1569 I1576 E1577 P1578 D1579 K1582 Q1583 Y1584 E1585 T1586 E1587 K1588 T1589 I1590 K1591 D1592 K1593 K1598 L1599 A1600 K1601 A1602 L1603 GLY GLY ASN LYS LYS ASP ASP
L1469 S1470 K1471 H1476 L1482 L1485 D1486 D1487 P1488 T1489 D1490 Q1491 E1492 E1495 Q1496 C1497 D1503 S1504 E1505 R1506 K1513 V1519 D1520 Y1523



- Molecule 2: Cag pathogenicity island protein (Cag7)

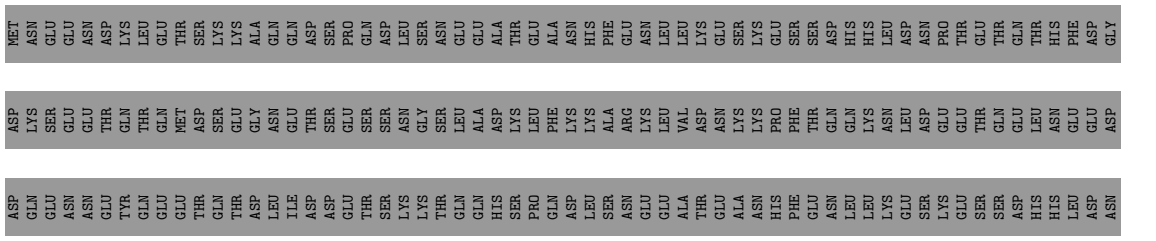
Chain CY: 5% 93%





- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain DY: 5% 93%



PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS LYS TYR ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE VAL ILE ILE LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU GLU CYS LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN

ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ILE LYS ASP GLU ASN LEU LYS LYS SER LEU LEU ASN GLN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LEU ILE ASN ASP PRO GLU ILE ARG GLU LYS

PHE ARG LYS GLU LEU GLU LEU GLN LYS GLU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA LYS THR GLU ALA GLU LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU ALA ILE GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN

GLU CYS LEU LYS ASN ILE PRO GLN ASP LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS GLN GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LYS LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN SER ASP ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU

GLU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS

GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP LYS GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP

CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA

ARG LYS LEU LEU GLU GLU UNK ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS ASN GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR LEU GLU SER LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA

LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR GLU ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER LEU LYS

ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU

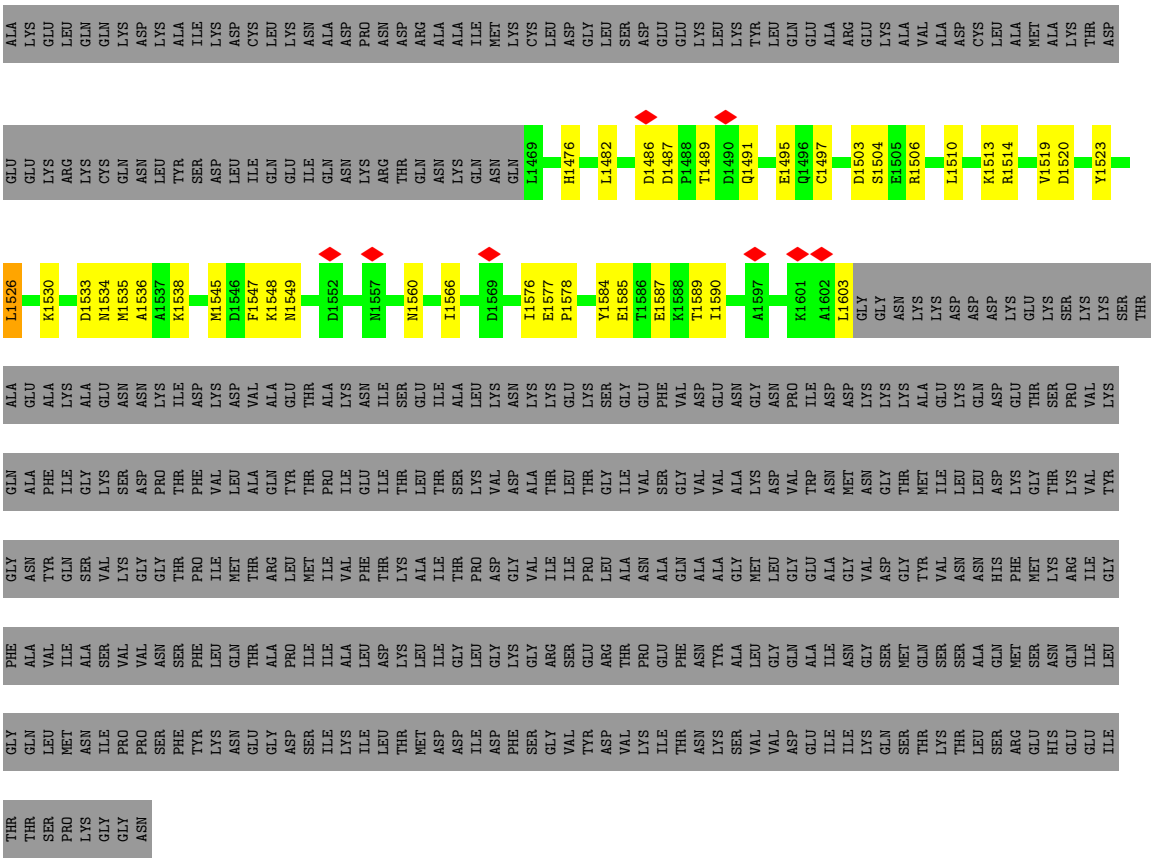
THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS ASN ALA LYS THR GLU

ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS GLU

SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER ILE LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS

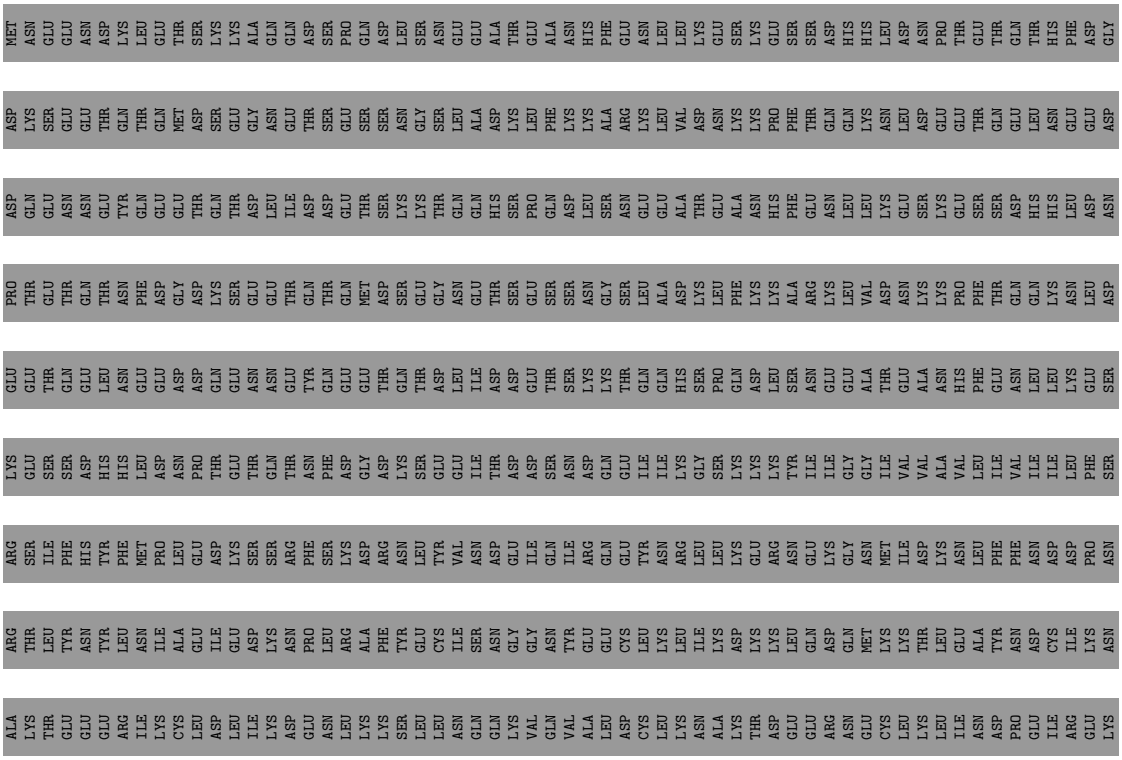
GLU LYS LEU LEU THR PRO GLU ALA ARG LYS PHE LEU ALA LYS GLN VAL LEU ASN CYS LEU GLU LYS ALA GLY ASN GLU GLU ARG LYS ALA CYS LEU LYS ASN LEU PRO LYS ASP LEU GLN GLU ASN ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA

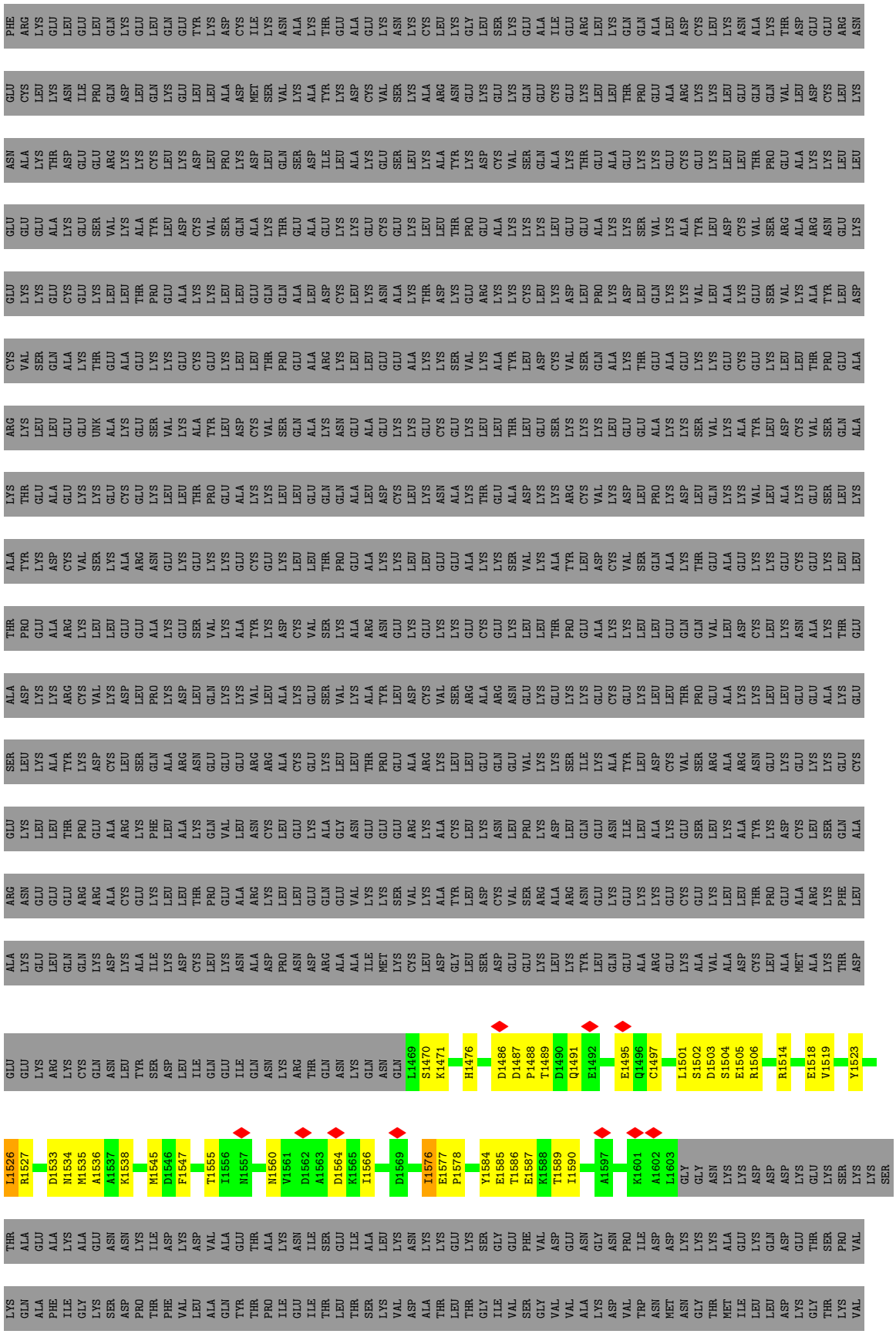
ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS PHE LEU



- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain EY: 5% 93%




L1526 R1527 D1533 N1534 M1535 A1536 A1537 K1538 M1545 D1546 F1547 T1555 I1556 N1557 N1560 V1561 D1562 A1563 D1564 K1565 I1566 D1569 I1576 E1577 P1578 Y1584 E1585 T1586 E1587 K1588 T1589 I1590 A1597 K1601 A1602 L1603
L1469 S1470 K1471 H1476 D1486 D1487 P1488 T1489 D1490 Q1491 E1492 E1495 Q1496 C1497 L1501 S1502 D1503 S1504 E1505 R1506 R1514 E1518 V1519 Y1523

TYR GLY ASN TYR GLN SER VAL LYS GLY GLY THR PRO ILE MET THR ARG LEU MET ILE VAL PHE THR LYS ALA ILE THR PRO ASP GLY VAL ILE ILE PRO LEU ALA ASN ALA GLN ALA ALA GLY MET LEU GLY GLU ALA GLY VAL ASP GLY TYR VAL ASN ASN HIS PHE MET LYS ARG ILE

GLY PHE ALA VAL ILE ALA SER VAL VAL ASN SER PHE LEU GLN THR ALA PRO ILE ILE ALA LEU ASP LYS LEU ILE GLY LEU GLY LYS GLY ARG SER GLU ARG THR PRO GLU PHE ASN TYR ALA LEU GLY GLN ALA ILE ASN GLY SER MET GLN SER SER ALA GLN MET SER ASN GLN ILE

LEU GLY GLN LEU MET ASN ILE PRO PRO SER PHE TYR LYS ASN GLU GLY ASP SER ILE LYS ILE LEU THR MET ASP ASP ILE ASP PHE SER GLY VAL TYR ASP VAL LYS ILE THR ASN LYS SER VAL VAL ASP GLU ILE ILE LYS GLN SER THR LYS THR LEU SER ARG GLU HIS GLU GLU

ILE THR THR SER PRO LYS GLY GLY ASN

- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain FY: 5% • 93%

MET ASN GLU GLU ASN ASP LYS LEU GLU THR SER LYS LYS ALA GLN GLN ASP SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR HIS PHE ASP GLY

ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP GLU GLU THR GLN GLU LEU ASN GLU ASP

ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN

PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS LYS TYR ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE VAL ILE ILE LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU GLU CYS LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN

ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ILE LYS ASP GLU ASN LEU LYS LYS SER LEU LEU ASN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LEU ILE ASN ASP PRO GLU ILE ARG GLU LYS

PHE ARG LYS GLU LEU GLU LEU GLN LYS GLU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA LYS THR GLU ALA GLU LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU ALA ILE GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU ARG ASN

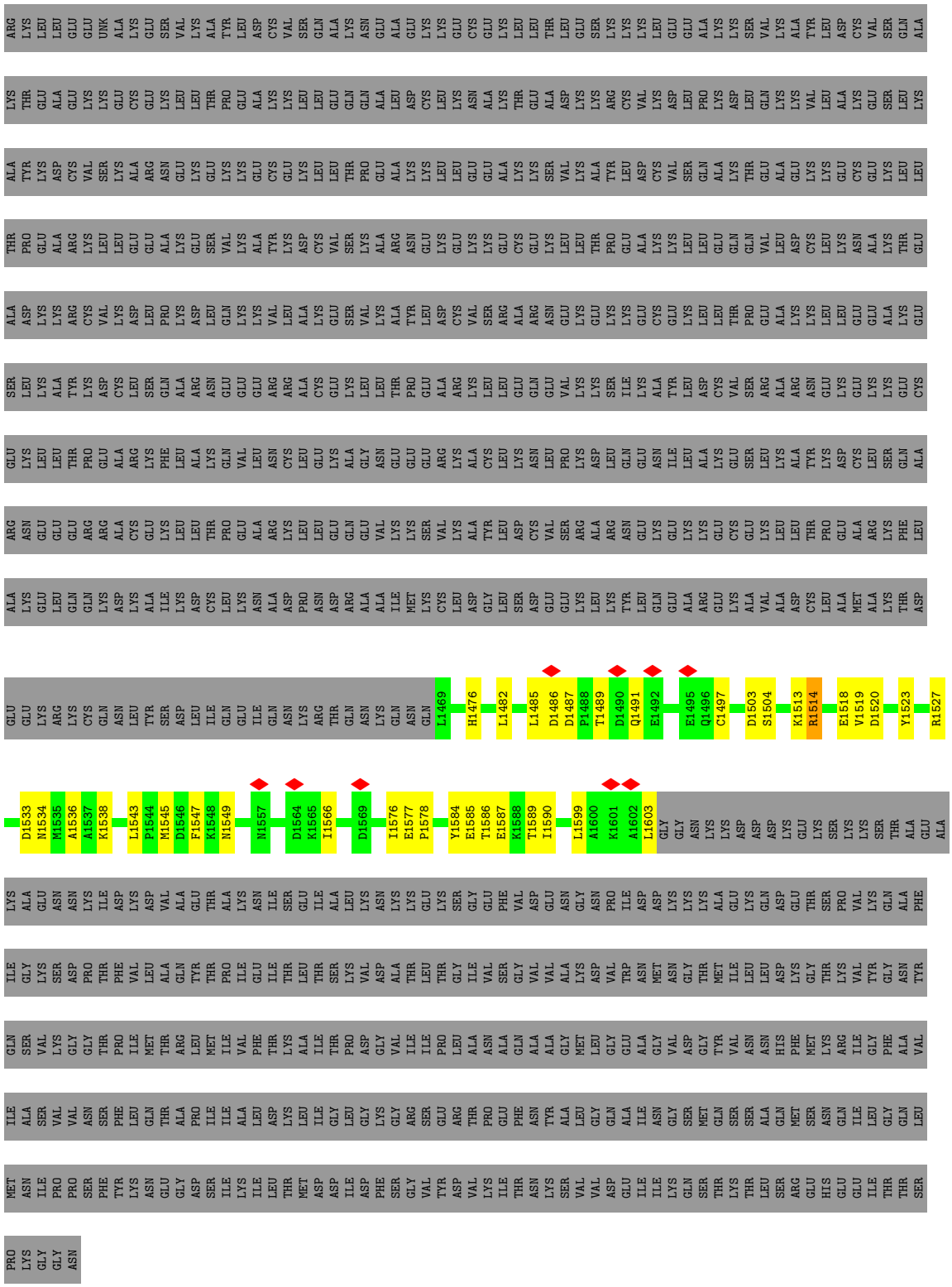
GLU CYS LEU LYS ASN ILE PRO GLN ASP LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS GLN GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LYS LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN SER ASP ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU

GLU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS

GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP LYS GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP

CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA



- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain GY: 5% • 93%

ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP GLU GLU THR GLN GLU LEU ASN GLU GLU ASP

ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN

PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS LYS TYR ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE VAL ILE ILE LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU GLU CYS LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN

ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ILE LYS ASP GLU ASN LEU LYS LYS SER LEU LEU ASN GLN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LEU ILE ASN ASP PRO GLU ILE ARG GLU LYS

PHE ARG LYS GLU LEU GLU LEU GLN LYS GLU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA LYS THR GLU ALA GLU LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU ALA ILE GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN

GLU CYS LEU LYS ASN ILE PRO GLN ASP LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS GLN GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LYS LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN SER ASP ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU

GLU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS

GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP LYS GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP

CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA

ARG LYS LEU LEU GLU GLU UNK ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS ASN GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR LEU GLU SER LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA

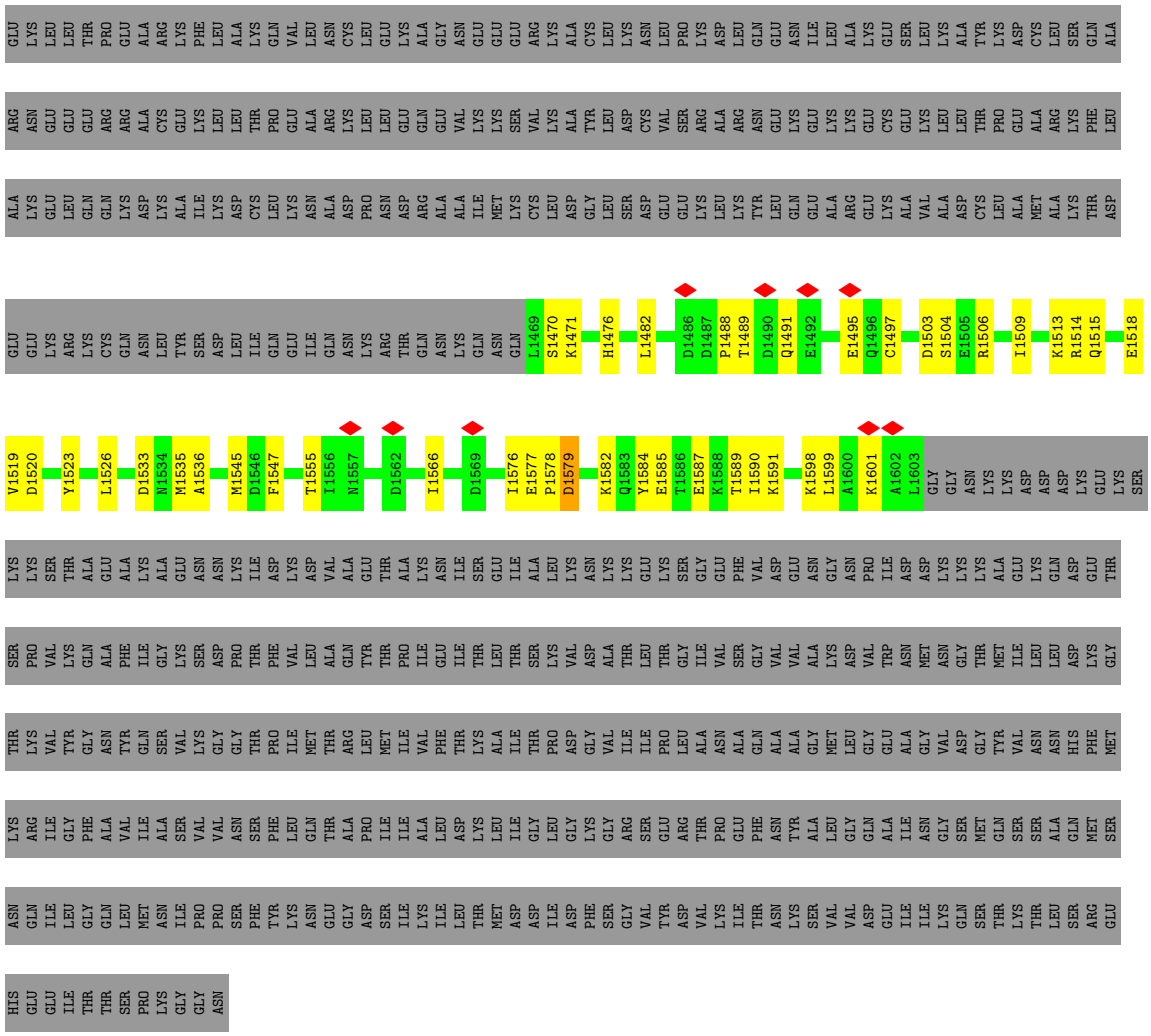
LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR GLU ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER LEU LYS

ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU

THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS ASN ALA LYS THR GLU

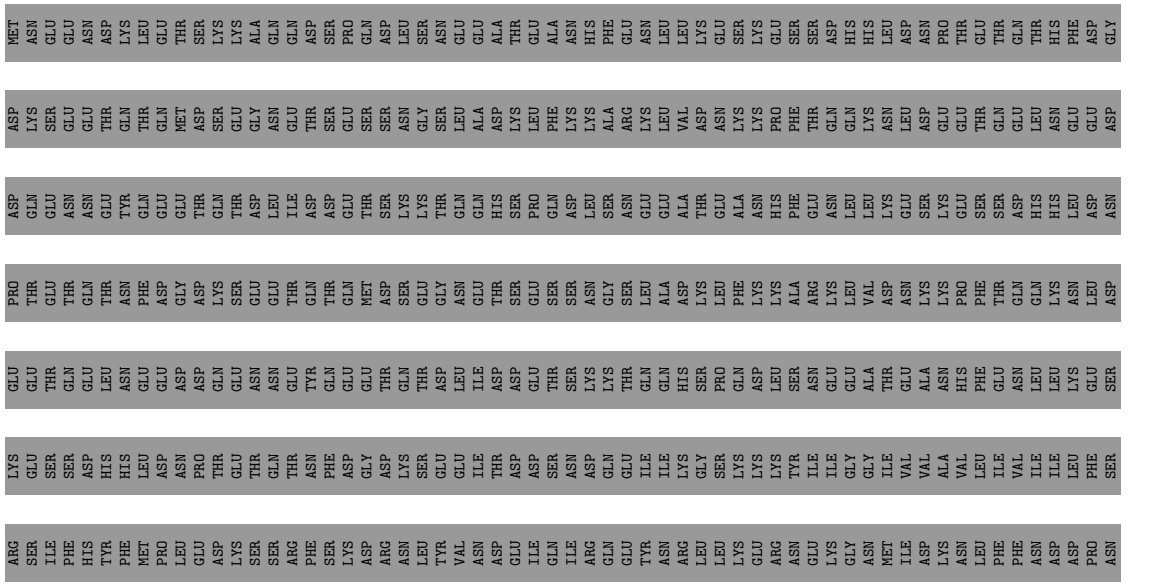
ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS GLU

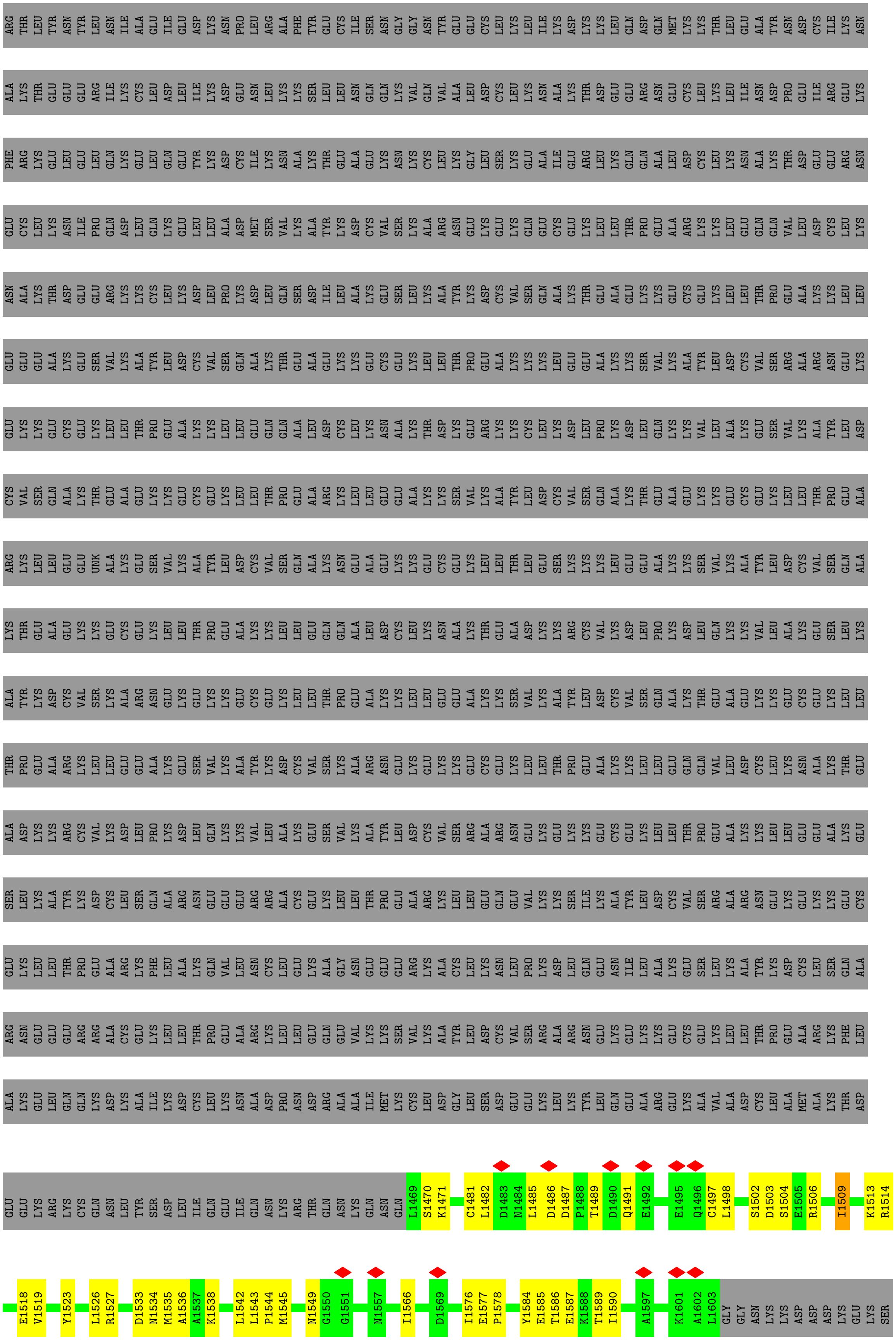
SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER ILE LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS



• Molecule 2: Cag pathogenicity island protein (Cag7)

Chain HY: 5% 93%




GLU GLU LYS ARG LYS CYS GLN ASN LEU TYR SER ASP LEU ILE GLN GLU ILE GLN ASN LYS ARG THR GLN ASN LYS GLN ASN GLN L1469 S1470 K1471 C1481 L1482 D1483 N1484 L1485 D1486 D1487 P1488 T1489 D1490 Q1491 E1492 E1495 Q1496 C1497 L1498 S1502 D1503 S1504 E1505 R1506 I1509 K1513 R1514
E1518 V1519 Y1523 L1526 R1527 D1533 N1534 M1535 A1536 A1537 K1538 L1542 L1543 P1544 M1545 N1549 G1550 G1551 N1557 I1566 D1569 I1576 E1577 P1578 Y1584 E1585 T1586 E1587 K1588 T1589 I1590 A1597 K1601 A1602 L1603 GLY GLY ASN LYS LYS ASP ASP ASP LYS GLU LYS SER

LYS LYS SER THR ALA GLU ALA LYS ALA GLU ASN ASN LYS ILE ASP LYS ASP VAL ALA GLU THR ALA LYS ASN ILE SER GLU ILE ALA LEU LYS ASN LYS LYS LYS GLU LYS SER GLY GLU PHE VAL ASP GLU ASN GLY ASN PRO ILE ASP ASP LYS LYS LYS ALA GLU LYS GLN ASP LYS GLU THR

SER PRO VAL LYS GLN ALA PHE ILE GLY LYS SER ASP PRO THR PHE VAL LEU ALA GLN TYR THR PRO ILE GLU ILE THR LEU THR SER LYS VAL ASP ALA THR LEU THR GLY ILE VAL SER GLY VAL VAL ALA LYS ASP VAL TRP ASN MET ASN GLY THR MET ILE LEU LEU ASP LYS GLY

THR LYS VAL TYR GLY ASN TYR GLN SER VAL LYS GLY GLY THR PRO ILE MET THR ARG LEU MET ILE VAL PHE THR LYS ALA ILE THR PRO ASP GLY VAL ILE ILE PRO LEU ALA ASN ALA GLN ALA ALA GLY MET LEU GLY GLU ALA GLY VAL ASP GLY TYR VAL ASN ASN HIS PHE MET

LYS ARG ILE GLY PHE ALA VAL ILE ALA SER VAL VAL ASN SER PHE LEU GLN THR ALA PRO ILE ILE ALA LEU ASP LYS LEU ILE GLY LEU GLY LYS ARG SER GLU ARG THR PRO GLU PHE ASN TYR ALA LEU GLY GLN ALA ILE ASN GLY SER MET GLN SER SER ALA GLN MET SER

ASN GLN ILE LEU GLY GLN LEU MET ASN ILE PRO PRO SER PHE TYR LYS ASN GLU GLY ASP SER ILE ILE LYS ILE LEU THR MET ASP ASP ILE ASP PHE SER GLY VAL TYR ASP VAL LYS ILE THR ASN LYS SER VAL VAL ASP GLU ILE ILE ILE LYS GLN SER THR LYS THR LEU SER ARG GLU

HIS GLU GLU ILE THR THR SER PRO LYS GLY GLY ASN

- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain IY: 5% 93%

MET ASN GLU GLU ASN ASP LYS LEU GLU THR SER LYS LYS ALA GLN GLN ASP SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR HIS PHE ASP GLY

ASP LYS SER GLU GLU THR GLN THR MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP GLU GLU THR GLN GLU LEU ASN GLU GLU ASP

ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN

PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

GLU GLU THR GLN GLU LEU ASN ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU GLU ALA THR GLU GLU ASN ASP HIS HIS LEU ASP ASN

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS TYR ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE ILE LEU LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU GLU ALA ASP PRO GLU ILE ASN

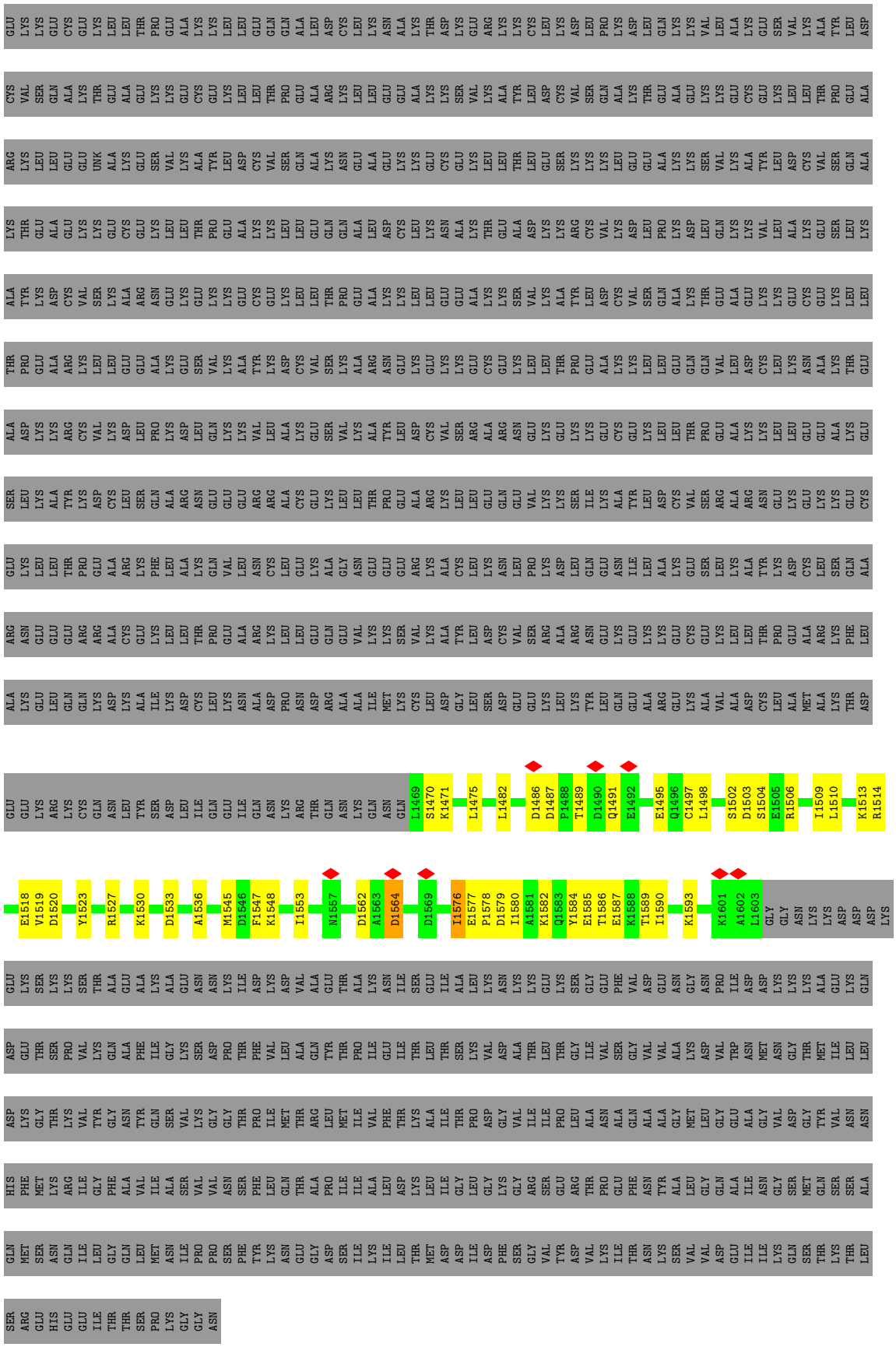
ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ASP ASN CYS LEU LEU ASN GLN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LEU ILE ASN ASP PRO GLU ILE ARG GLU LYS

PHE ARG LYS GLU LEU GLU LEU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA THR GLU ALA LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU ALA ILE GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU ARG ASN

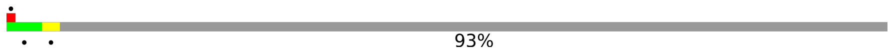
GLU CYS LEU LYS ASN ILE PRO GLN LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS SER LYS GLN GLU CYS GLU LYS LYS LEU LEU THR PRO GLU ALA ARG LYS LYS LYS LEU LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LEU ASP ASP LYS LYS LEU ALA ILE LEU LEU LYS GLU SER LEU LYS ALA TYR LYS ASP CYS VAL SER GLN ALA LYS THR GLU ALA LYS LYS LEU LEU LYS LYS LEU LEU

GLU GLU GLU ALA LYS GLU SER VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS



- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain JY: 

MET ASN GLU GLU ASN ASP LYS LEU GLU THR SER LYS LYS ALA GLN GLN ASP SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR HIS PHE ASP GLY

ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP GLU GLU THR GLN GLU LEU ASN GLU ASP

ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN

PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS LYS TYR ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE VAL ILE ILE LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU GLU CYS LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN

ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ILE LYS ASP GLU ASN LEU LYS LYS SER LEU LEU ASN GLN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LEU ILE ASN ASP PRO GLU ILE ARG GLU LYS

PHE ARG LYS GLU LEU GLU LEU GLN LYS GLU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA LYS THR GLU ALA GLU LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU ALA ILE GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU ARG ASN

GLU CYS LEU LYS ASN ILE PRO GLN ASP LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS GLN GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LYS LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN SER ASP ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU

GLU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS

GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP LYS GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP

CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA

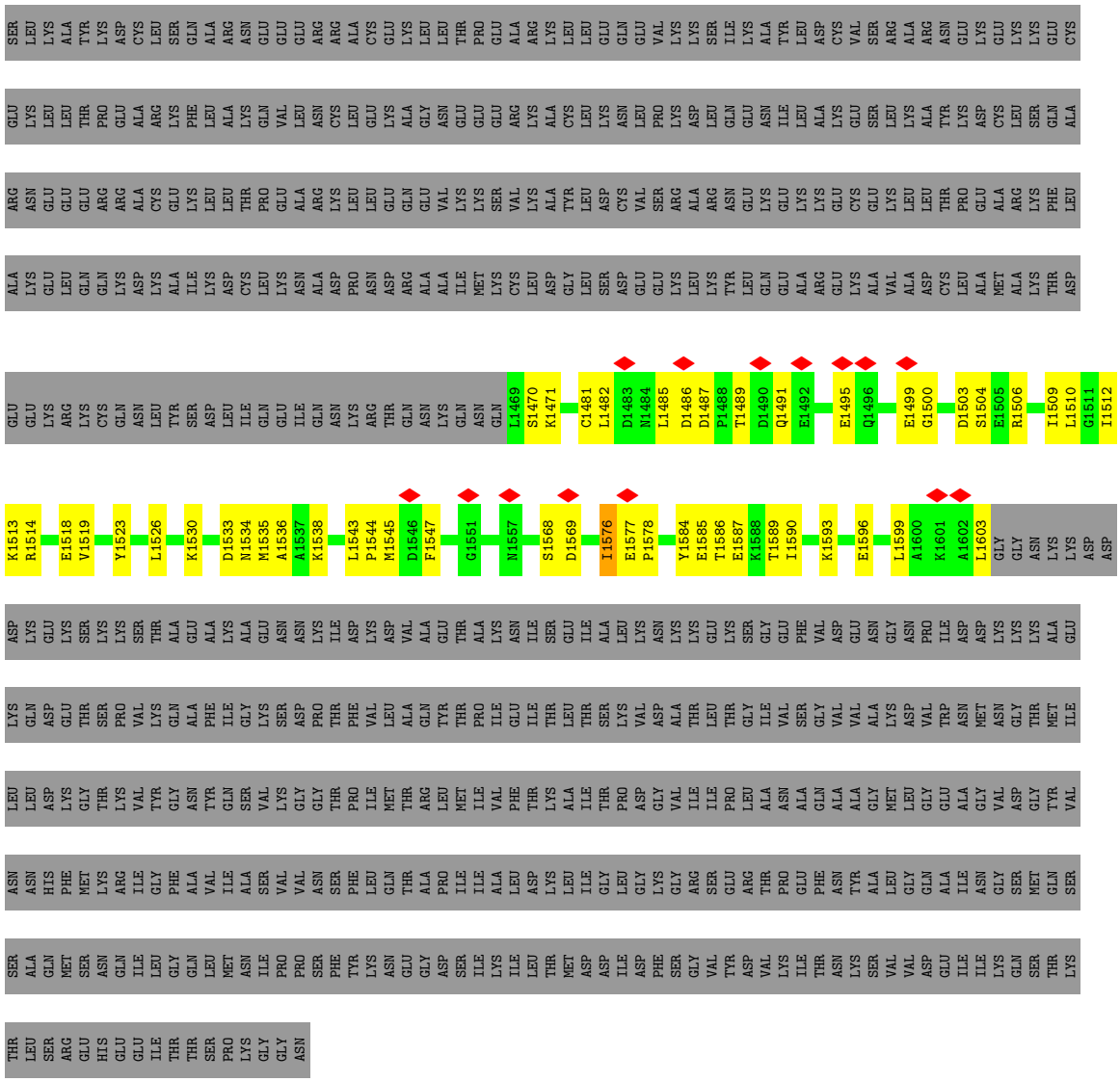
ARG LYS LEU LEU GLU GLU UNK ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS ASN GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR LEU GLU SER LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA

LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR GLU ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER LEU LYS

ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU

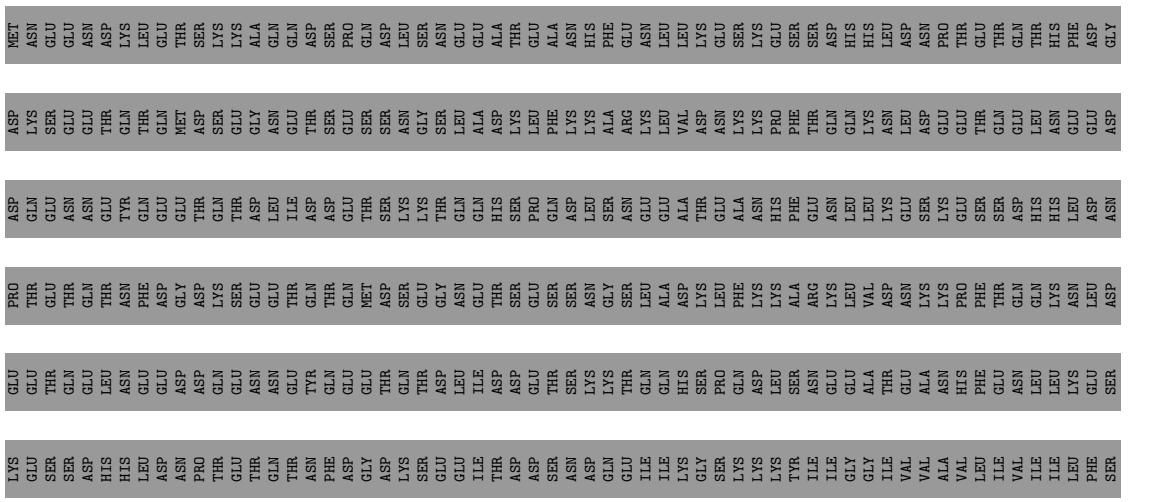
THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS ASN ALA LYS THR GLU

ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS GLU

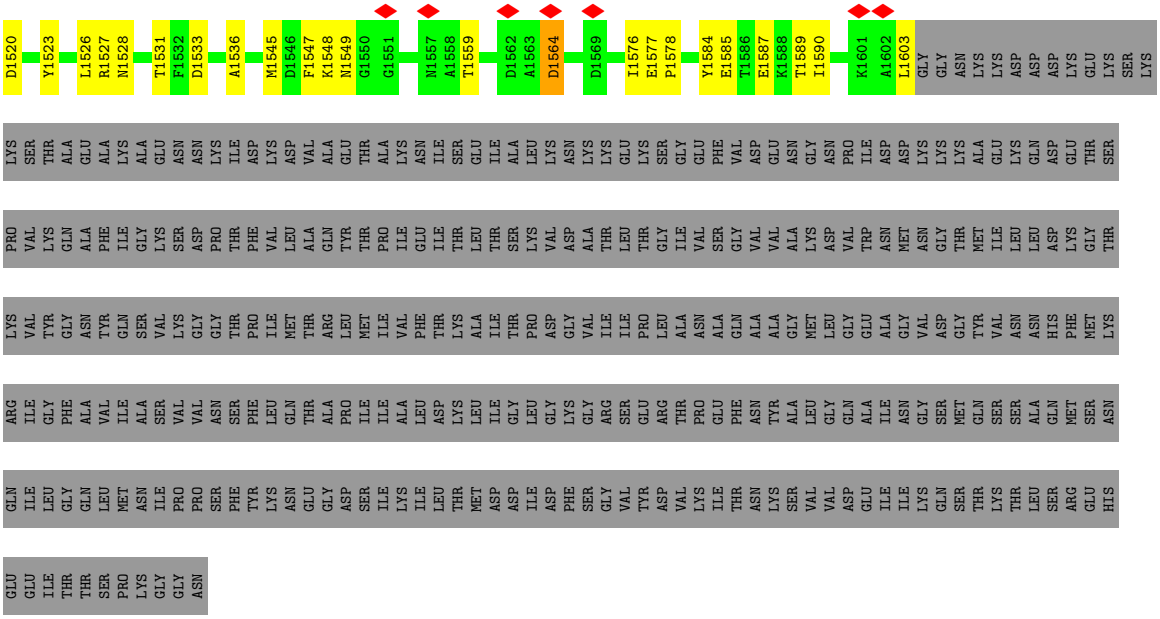


• Molecule 2: Cag pathogenicity island protein (Cag7)

Chain KY: 5% 93%

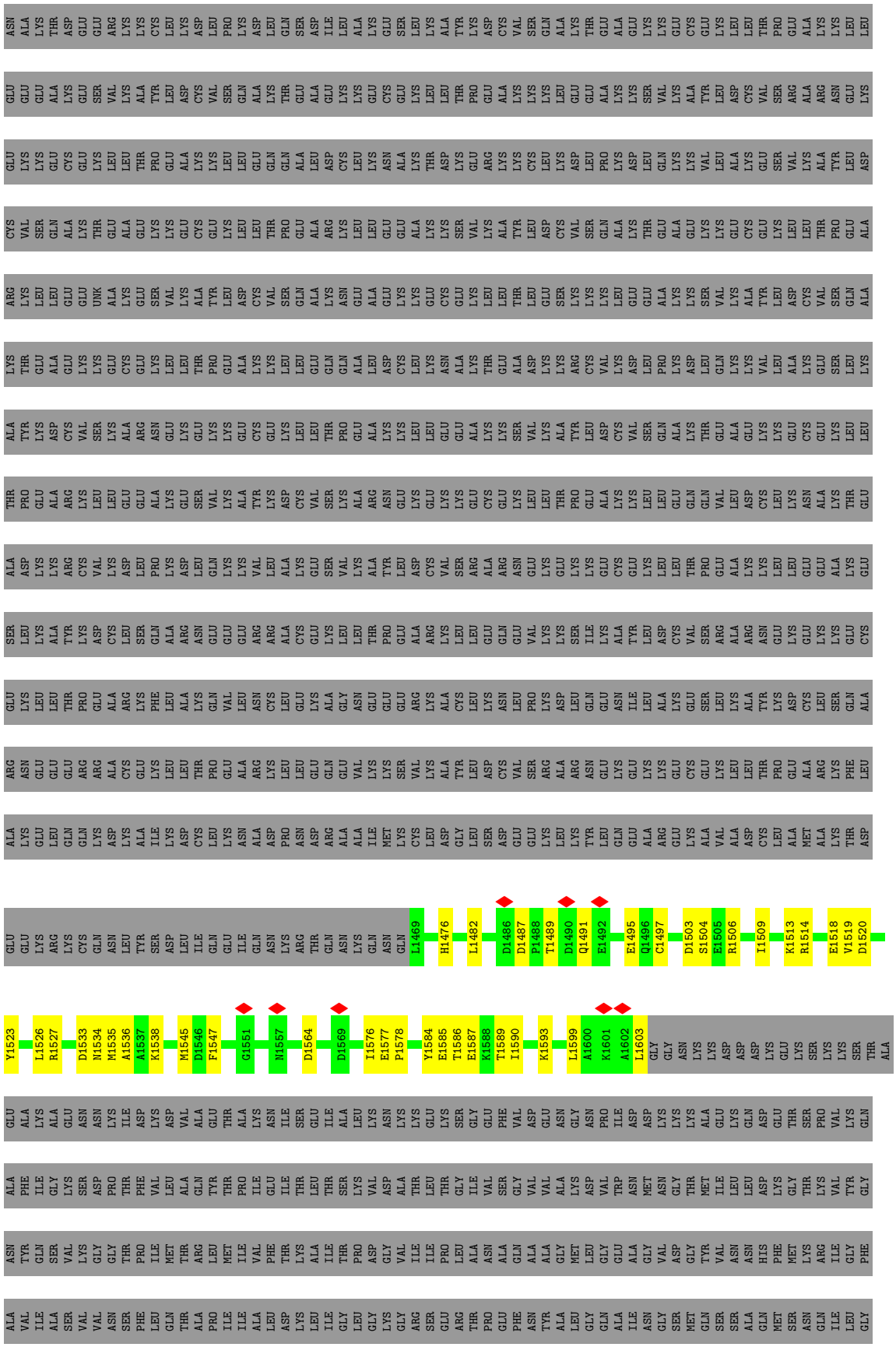


L1469
S1470
K1471
S1479
L1482
D1486
D1487
P1488
T1489
D1490
Q1491
E1492
E1495
Q1496
C1497
D1503
S1504
E1505
R1506
I1509
K1513
R1514
E1518
V1519



- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain LY: 5% • 93%


Y1523
L1526
R1527
D1533
N1534
M1535
A1536
A1537
K1538
M1545
D1546
F1547
G1551
N1557
D1564
D1569
I1576
E1577
P1578
Y1584
E1585
T1586
E1587
K1588
T1589
I1590
K1593
L1599
A1600
K1601
A1602
L1603
L1469
H1476
L1482
D1486
D1487
P1488
T1489
D1490
Q1491
E1492
E1495
Q1496
C1497
D1503
S1504
E1505
R1506
I1509
K1513
R1514
E1518
V1519
D1520

GLN LEU MET ASN ILE PRO PRO SER PHE TYR LYS ASN GLU GLY ASP SER ILE LYS ILE LEU THR MET ASP ASP ILE ASP PHE SER GLY VAL TYR ASP VAL LYS ILE THR ASN LYS SER VAL VAL ASP GLU ILE ILE LYS GLN SER THR LYS THR LEU SER ARG GLU HIS GLU GLU ILE THR

THR SER PRO LYS GLY GLY ASN

- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain MY: 5% 93%

MET ASN GLU GLU ASN ASP LYS LEU GLU THR SER LYS LYS ALA GLN GLN ASP SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR HIS PHE ASP GLY

ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS HIS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP GLU GLU THR GLN GLU LEU GLU ASP

ASP GLN GLU ASN ASN GLU TYR GLN GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER LYS GLU GLU GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU THR SER SER ASP HIS LEU ASP ASN

PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ILE LYS ALA ARG LYS LEU LEU VAL ASP ASN LYS LYS PRO THR PHE THR GLN GLN LYS ASN HIS HIS LEU ASP

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN GLN ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS LYS TYR ILE ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE VAL ILE ILE LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU GLU CYS LEU LYS LEU LEU LYS ASP LYS LYS LEU GLN GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN

ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ILE LYS ASP GLU ASN LEU LYS LYS SER LEU LEU ASN GLN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LYS LEU LYS GLU GLU

PHE ARG LYS GLU LEU GLU LEU GLN LYS GLU GLU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA LYS THR GLU ALA ALA LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN

GLU CYS LEU LYS ASN ILE PRO GLN ASP LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA ALA TYR LYS ASP CYS VAL SER LYS ALA LYS LYS GLN GLU LYS GLN GLN CYS GLU LYS LEU LEU THR PRO GLU ALA ALA LYS GLN GLN VAL LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LYS ASP LEU LYS ASP LEU ILE LEU ALA LYS LYS GLU SER LEU LYS ALA ALA LYS GLU LYS LYS LYS GLU SER GLN ALA LYS THR GLU ALA LYS LYS LYS LYS LEU LEU

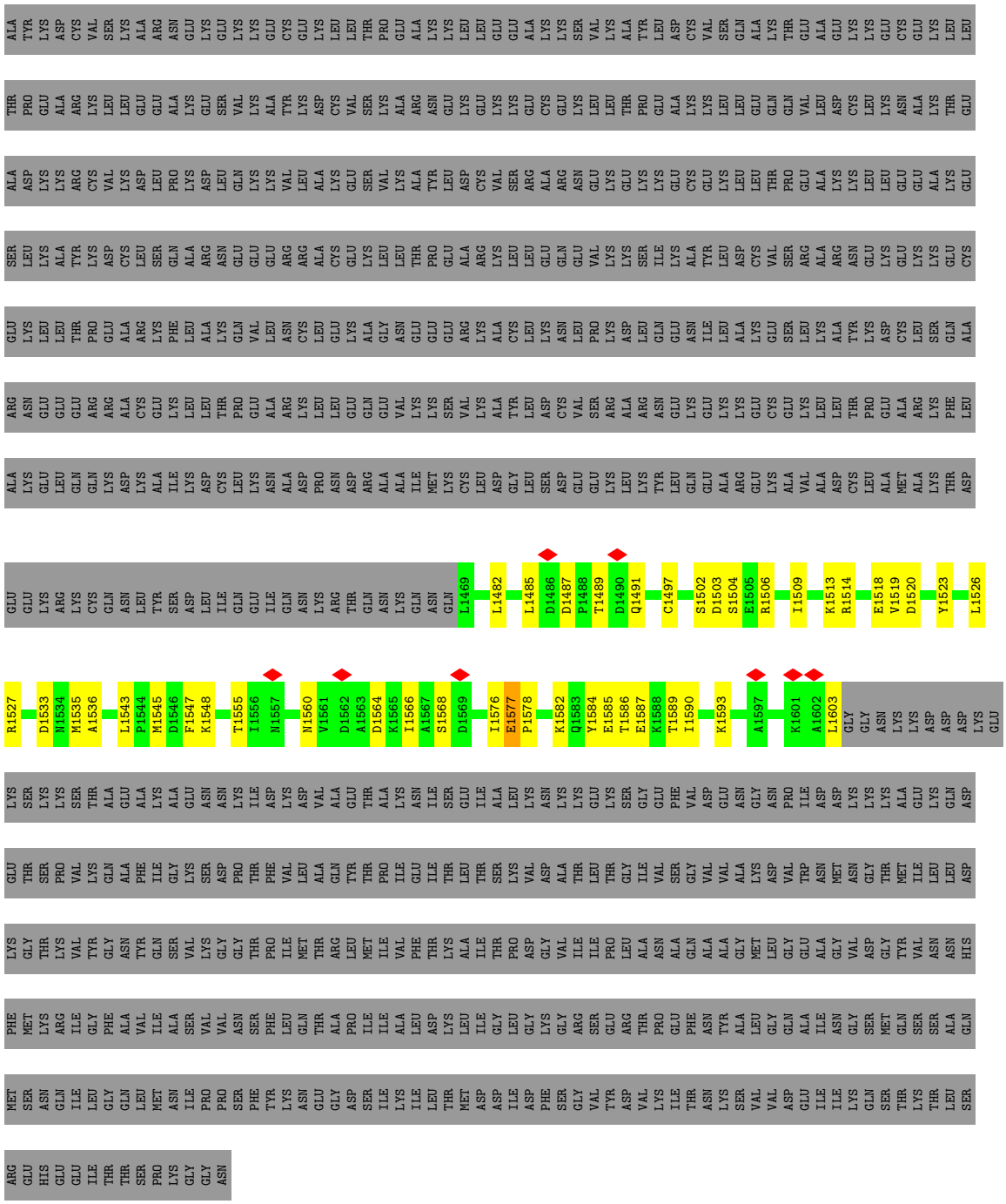
GLU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER ALA LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS

GLU LYS LYS GLU CYS GLU LYS LYS LEU ALA ALA THR PRO GLU ALA LYS LYS LEU LEU LEU LEU LYS LYS LYS LYS LEU LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP

CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU ALA GLN LEU LEU THR PRO GLU ALA ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA LYS LYS GLU LYS LYS LYS LEU GLU GLU LYS LYS LYS GLU GLU LYS LEU LEU ALA

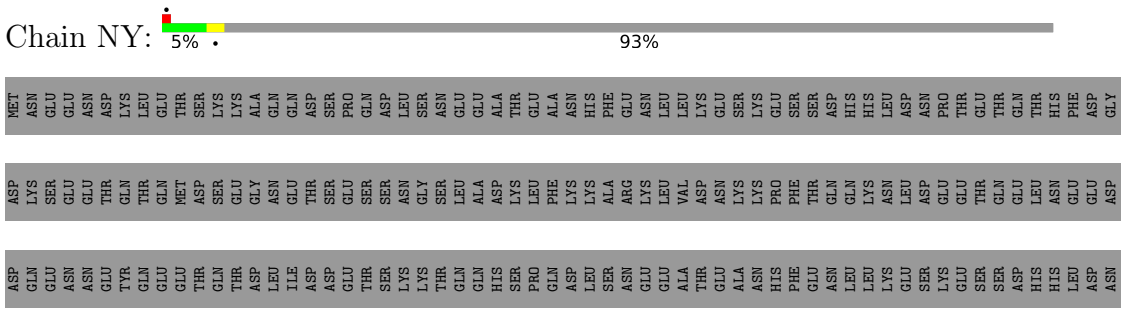
ARG LYS LEU LEU GLU UNK ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS LYS ASN GLU ALA LYS LYS GLU CYS GLU LYS LYS THR LEU LEU LEU GLU GLU SER LYS LYS LYS LEU GLU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA

LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU LEU THR PRO GLU ALA LYS LYS LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR GLU ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU LYS LYS GLU SER LEU LYS



- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain NY:



PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS LYS TYR ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE VAL ILE ILE LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU GLU CYS LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN

ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ILE LYS ASP GLU ASN LEU LYS LYS SER LEU LEU ASN GLN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LEU ILE ASN ASP PRO GLU ILE ARG GLU LYS

PHE ARG LYS GLU LEU GLU LEU GLN LYS GLU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA LYS THR GLU ALA GLU LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU ALA ILE GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN

GLU CYS LEU LYS ASN ILE PRO GLN ASP LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS GLN GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LYS LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN SER ASP ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU

GLU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS

GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP LYS GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP

CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA

ARG LYS LEU LEU GLU GLU UNK ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS ASN GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR LEU GLU SER LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA

LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR GLU ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER LEU LYS

ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU

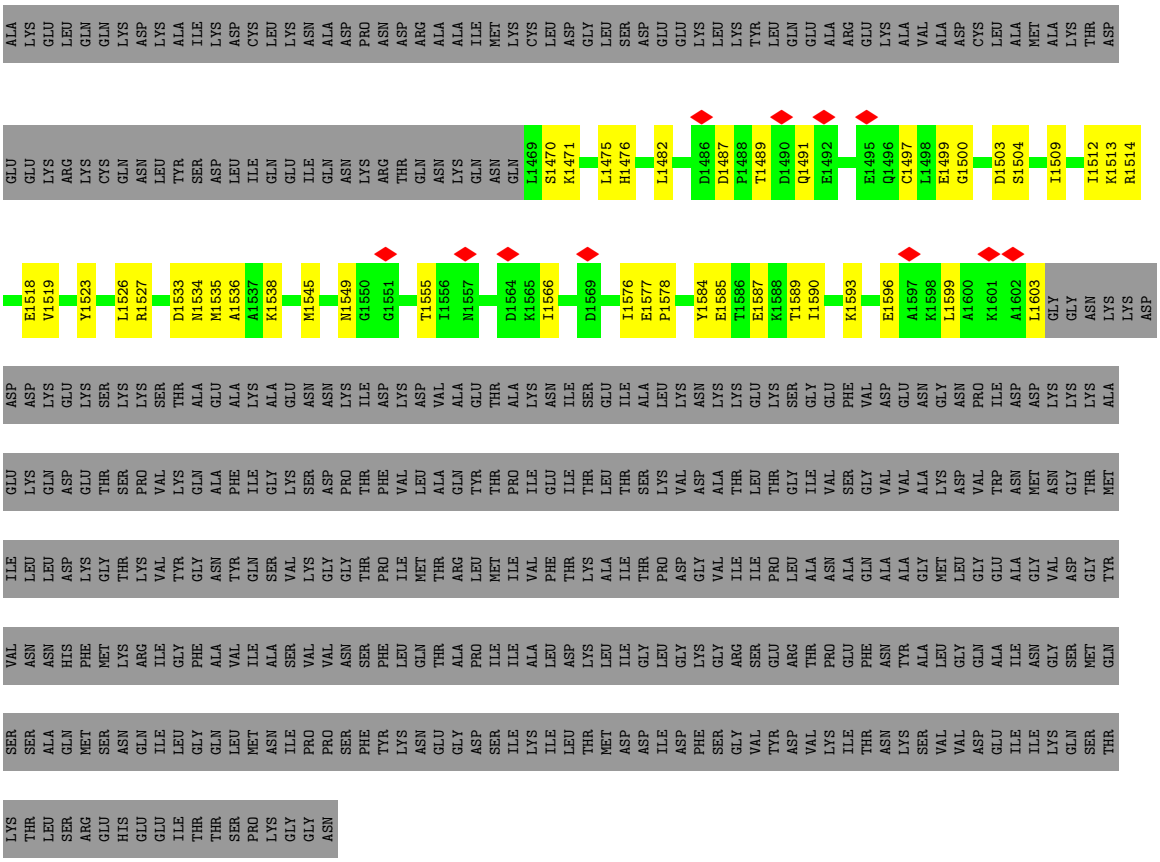
THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS ASN ALA LYS THR GLU

ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS GLU

SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER ILE LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS

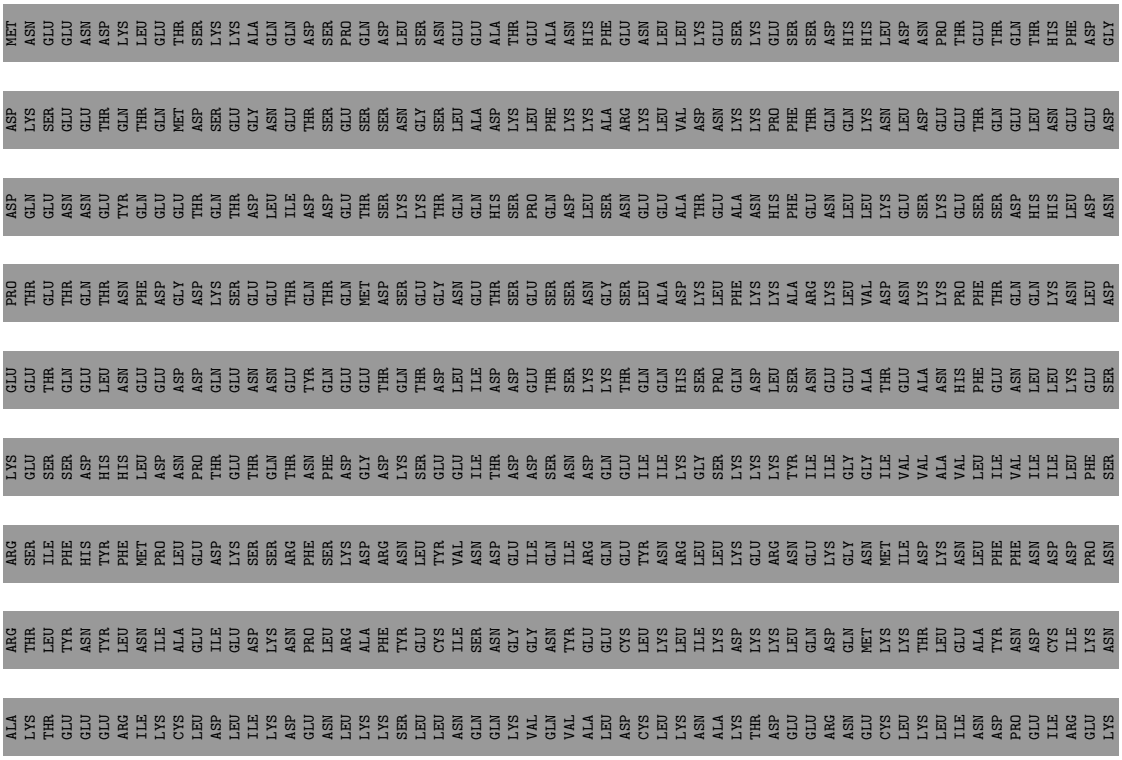
GLU LYS LEU LEU THR PRO GLU ALA ARG LYS PHE LEU ALA LYS GLN VAL LEU ASN CYS LEU GLU LYS ALA GLY ASN GLU GLU ARG LYS ALA CYS LEU LYS ASN LEU PRO LYS ASP LEU GLN GLU ASN ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA

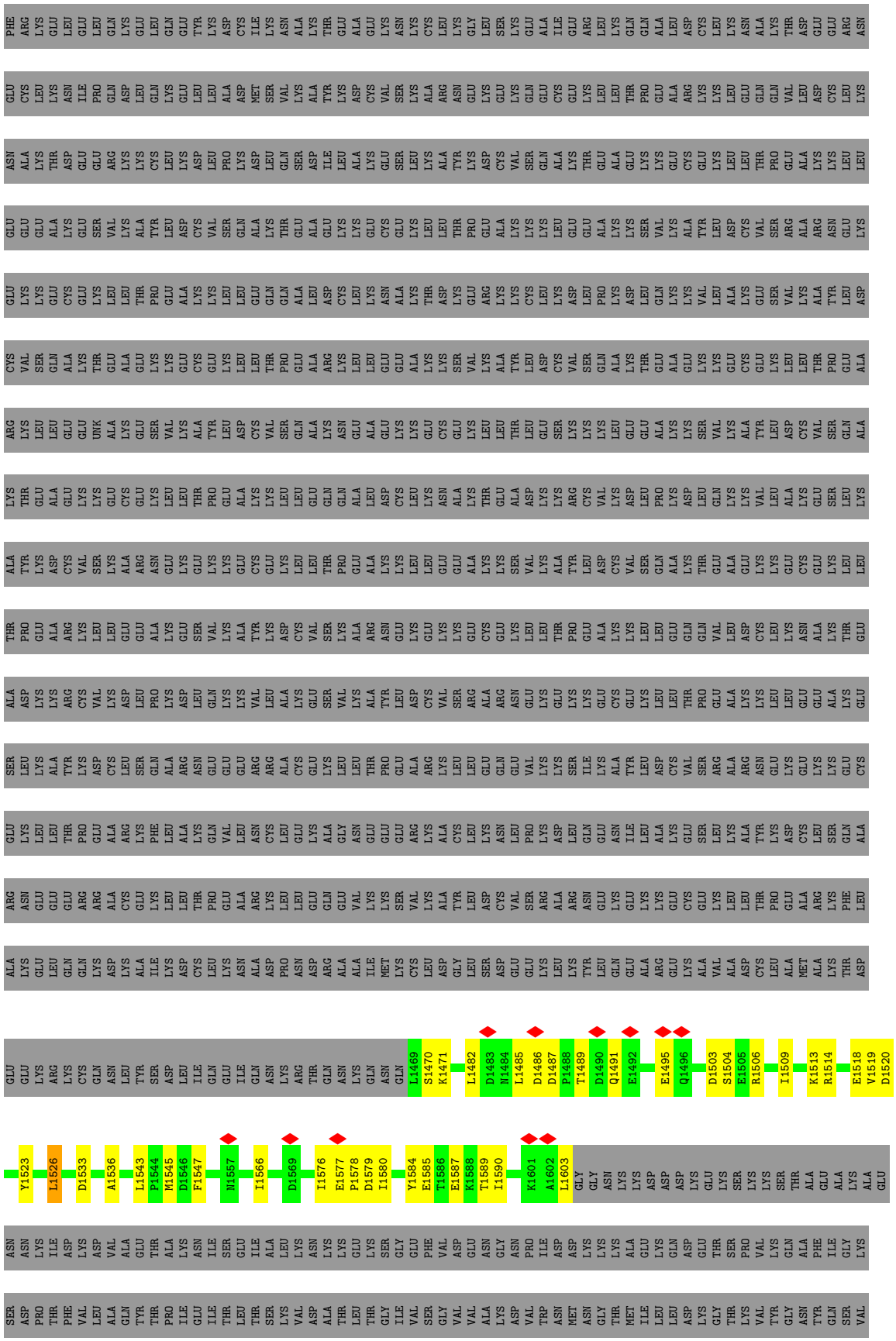
ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS PHE LEU



• Molecule 2: Cag pathogenicity island protein (Cag7)

Chain OY: 5% • 93%




Y1523
L1526
D1533
A1536
L1543
P1544
M1545
D1546
F1547
N1557
I1566
D1569
I1576
E1577
P1578
D1579
I1580
Y1584
E1585
T1586
E1587
K1588
T1589
I1590
K1601
A1602
L1603
L1469
S1470
K1471
L1482
D1483
N1484
L1485
D1486
D1487
P1488
T1489
D1490
Q1491
E1492
E1495
Q1496
D1503
S1504
E1505
R1506
I1509
K1513
R1514
E1518
V1519
D1520

LYS GLY GLY THR PRO ILE MET THR ARG LEU MET ILE VAL PHE THR LYS ALA ILE THR PRO ASP GLY VAL ILE ILE PRO LEU ALA ASN ALA GLN ALA ALA GLY MET LEU GLY GLU ALA GLY VAL ASP GLY TYR VAL ASN ASN HIS PHE MET LYS ARG ILE GLY PHE ALA VAL ILE ALA SER

VAL VAL ASN SER PHE LEU GLN THR ALA PRO ILE ILE ALA LEU ASP LYS LEU ILE GLY LEU GLY LYS GLY ARG SER GLU ARG THR PRO GLU PHE ASN TYR ALA LEU GLY GLN ALA ILE ASN GLY SER MET GLN SER SER ALA GLN MET SER ASN GLN ILE LEU GLY GLN LEU MET ASN ILE

PRO PRO SER PHE TYR LYS ASN GLU GLY ASP SER ILE LYS ILE LEU THR MET ASP ASP ILE ASP PHE SER GLY VAL TYR ASP VAL LYS ILE THR ASN LYS SER VAL VAL ASP GLU ILE ILE LYS GLN SER THR LYS THR LEU SER ARG GLU HIS GLU GLU ILE THR THR SER PRO LYS GLY

GLY ASN

- Molecule 2: Cag pathogenicity island protein (Cag7)

Chain PY: 5% • 93%

MET ASN GLU GLU ASN ASP LYS LEU GLU THR SER LYS LYS ALA GLN GLN ASP SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR HIS PHE ASP GLY

ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP GLU GLU THR GLN GLU LEU ASN GLU GLU ASP

ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN

PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS LYS TYR ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE VAL ILE ILE LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU GLU CYS LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN

ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ILE LYS ASP GLU ASN LEU LYS LYS SER LEU LEU ASN GLN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LEU ILE ASN ASP PRO GLU ILE ARG GLU LYS

PHE ARG LYS GLU LEU GLU LEU GLN LYS GLU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA LYS THR GLU ALA GLU LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU ALA ILE GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU ARG ASN

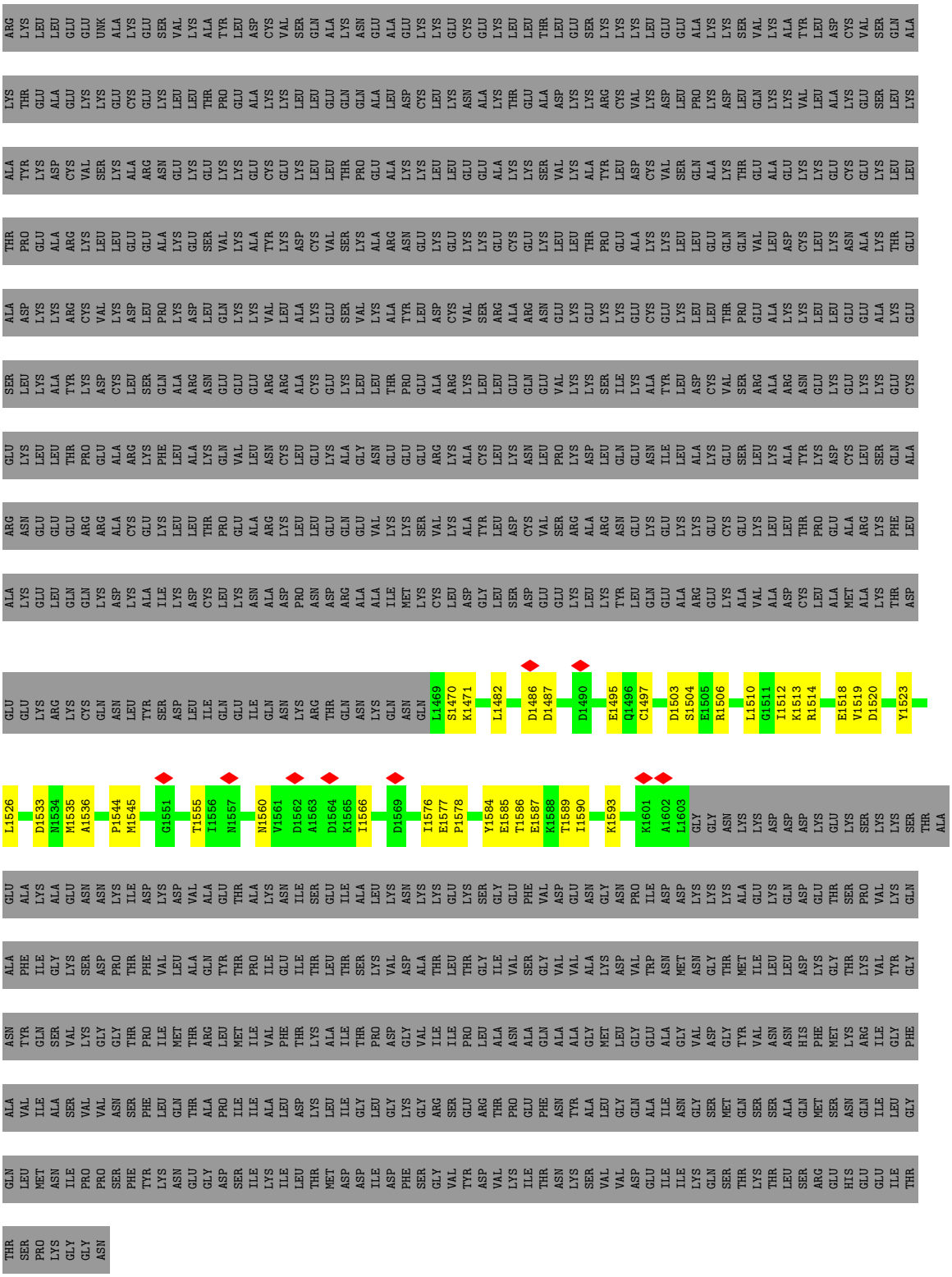
GLU CYS LEU LYS ASN ILE PRO GLN ASP LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS GLN GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LYS LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN SER ASP ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU

GLU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS

GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP LYS GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP

CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA



• Molecule 2: Cag pathogenicity island protein (Cag7)

Chain QY: 5% • 93%

ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP GLU GLU THR GLN GLU LEU ASN GLU GLU ASP

ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN

PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU ILE THR ASP ASP SER ASN ASP GLN GLU ILE ILE LYS GLY SER LYS LYS LYS TYR ILE ILE GLY GLY ILE VAL VAL ALA VAL LEU ILE VAL ILE ILE LEU PHE SER

ARG SER ILE PHE HIS TYR PHE MET PRO LEU GLU ASP LYS SER SER ARG PHE SER LYS ASP ARG ASN LEU TYR VAL ASN ASP GLU ILE GLN ILE ARG GLN GLU TYR ASN ARG LEU LEU LYS GLU ARG ASN GLU LYS GLY ASN MET ILE ASP LYS ASN LEU PHE PHE ASN ASP ASP PRO ASN

ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU GLU CYS LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN

ALA LYS THR GLU GLU GLU ARG ILE LYS CYS LEU ASP LEU ILE LYS ASP GLU ASN LEU LYS LYS SER LEU LEU ASN GLN GLN LYS VAL GLN VAL ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN GLU CYS LEU LYS LEU ILE ASN ASP PRO GLU ILE ARG GLU LYS

PHE ARG LYS GLU LEU GLU LEU GLN LYS GLU LEU GLN GLU TYR LYS ASP CYS ILE LYS ASN ALA LYS THR GLU ALA GLU LYS ASN LYS CYS LEU LYS GLY LEU SER LYS GLU ALA ILE GLU ARG LEU LYS GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP GLU GLU ARG ASN

GLU CYS LEU LYS ASN ILE PRO GLN ASP LEU GLN LYS GLU LEU LEU ALA ASP MET SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS GLN GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LYS LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS

ASN ALA LYS THR ASP GLU GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN SER ASP ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU

GLU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS

GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR ASP LYS GLU ARG LYS LYS CYS LEU LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP

CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA

ARG LYS LEU LEU GLU GLU UNK ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS ASN GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR LEU GLU SER LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA

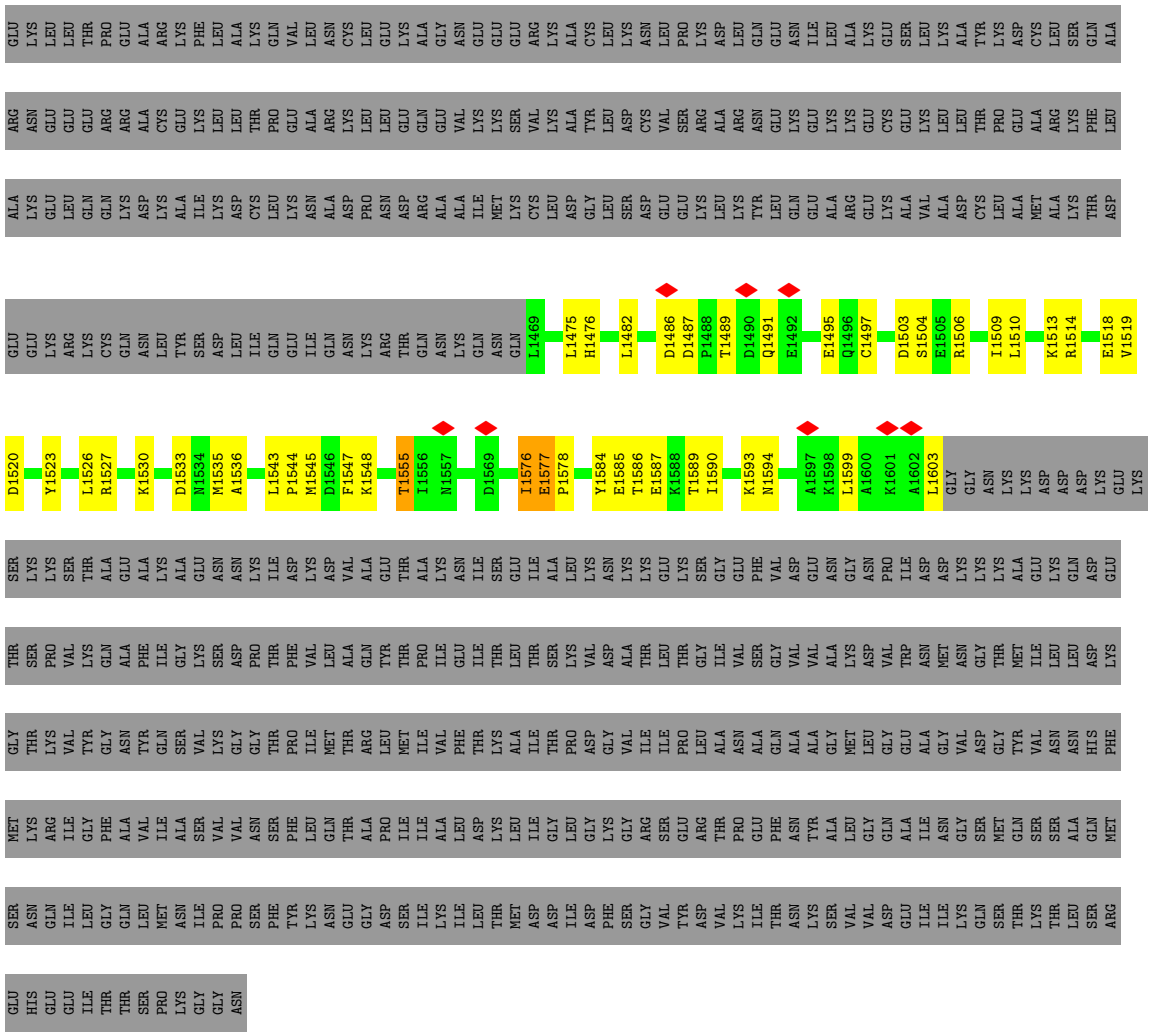
LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN ALA LEU ASP CYS LEU LYS ASN ALA LYS THR GLU ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER LEU LYS

ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU

THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS ASN ALA LYS THR GLU

ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS GLU

SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER ILE LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS


GLU LYS LEU LEU THR PRO GLU ALA ARG LYS PHE LEU ALA LYS GLN VAL LEU ASN CYS LEU GLU LYS ALA GLY ASN GLU GLU GLU ARG LYS ALA CYS LEU LYS ASN LEU PRO LYS ASP LEU GLN GLU ASN ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA
ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS PHE LEU
ALA LYS GLU LEU GLN GLN LYS ASP LYS ALA ILE LYS ASP CYS LEU LYS ASN ALA ASP PRO ASN ASP ASP ARG ALA ALA ILE MET LYS CYS LEU ASP GLY LEU SER ASP GLU GLU LYS LEU LYS TYR LEU GLN GLU ALA ARG GLU LYS ALA VAL ALA LYS ASP CYS LEU ALA ALA MET ALA LYS THR ASP
GLU GLU LYS ARG LYS CYS GLN ASN LEU TYR SER ASP LEU ILE GLN GLU ILE GLN ASN LYS GLN ASN GLN L1469 L1475 H1476 L1482 D1486 D1487 P1488 T1489 D1490 Q1491 E1492 E1495 Q1496 C1497 D1503 S1504 E1505 R1506 I1509 L1510 K1513 R1514 E1518 V1519
D1520 Y1523 L1526 R1527 K1530 D1533 N1534 M1535 A1536 L1543 P1544 M1545 D1546 F1547 K1548 T1555 I1556 N1557 D1569 I1576 E1577 P1578 Y1584 E1585 T1586 E1587 K1588 T1589 I1590 K1593 N1594 A1597 K1598 L1599 A1600 K1601 A1602 L1603 GLY GLY ASN LYS LYS ASP ASP ASP LYS GLU LYS
SER LYS LYS SER THR ALA GLU ALA LYS ALA GLU GLU ASN ASN LYS ASP VAL ALA GLN TYR ILE ILE GLN LYS ILE ASP PRO VAL LYS ASN ILE SER GLU ILE ALA LEU LYS ASN LYS GLU GLU LYS SER GLY GLU PHE VAL ASP GLU ASN GLY ASN PRO ILE ASP ASP LYS LYS LYS ALA GLU LYS GLN ASP GLU
THR SER PRO VAL LYS GLN ALA PHE ILE GLY LYS SER ASP PRO THR PHE VAL LEU ALA GLN TYR THR PRO ILE GLU ILE THR LEU THR SER LYS VAL ASP ALA THR LEU THR GLY ILE VAL SER GLY VAL VAL ALA LYS ASP VAL TRP ASN MET ASN GLY THR MET ILE LEU LEU ASP LYS
GLY THR LYS VAL TYR GLY ASN TYR GLN SER VAL LYS GLY GLY THR PRO ILE MET THR ARG LEU MET ILE VAL PHE THR LYS ALA ILE THR PRO ASP GLY VAL ILE ILE PRO LEU ALA ASN ALA GLN ALA ALA GLY MET LEU GLY GLU ALA GLY VAL ASP GLY TYR VAL ASN ASN HIS PHE
MET LYS ARG ILE GLY PHE ALA VAL ILE ALA SER VAL VAL ASN SER PHE LEU GLN THR ALA PRO ILE ILE ALA LEU ASP LYS LEU ILE GLY LEU GLY LYS GLY ARG SER GLU ARG THR PRO GLU PHE ASN TYR ALA LEU GLY GLN ALA ILE ASN GLY SER MET GLN SER SER ALA GLN MET
SER ASN GLN ILE LEU GLY GLN LEU MET ASN ILE PRO PRO SER PHE TYR LYS ASN GLU GLY ASP SER ILE LYS ILE LEU THR MET ASP ASP ILE ASP PHE SER GLY VAL TYR ASP VAL LYS ILE THR ASN LYS SER VAL VAL ASP GLU ILE ILE LYS GLN SER THR LYS THR LEU SER ARG
GLU HIS GLU GLU ILE THR THR SER PRO LYS GLY GLY ASN

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	20929	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	59.2	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.148	Depositor
Minimum map value	-0.069	Depositor
Average map value	-0.000	Depositor
Map value standard deviation	0.002	Depositor
Recommended contour level	0.03	Depositor
Map size (Å)	836.39996, 836.39996, 836.39996	wwPDB
Map dimensions	510, 510, 510	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.64, 1.64, 1.64	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| >5$	RMSZ	$\#\|Z\| >5$
1	AX	0.32	0/1365	0.39	0/1845
1	BX	0.32	0/1365	0.40	0/1845
1	CX	0.32	0/1365	0.40	0/1845
1	DX	0.33	0/1365	0.40	0/1845
1	EX	0.33	0/1365	0.41	0/1845
1	FX	0.32	0/1365	0.41	0/1845
1	GX	0.32	0/1365	0.40	0/1845
1	HX	0.33	0/1365	0.41	0/1845
1	IX	0.32	0/1365	0.40	0/1845
1	JX	0.33	0/1365	0.41	0/1845
1	KX	0.32	0/1365	0.39	0/1845
1	LX	0.32	0/1365	0.40	0/1845
1	MX	0.32	0/1365	0.41	0/1845
1	NX	0.33	0/1365	0.40	0/1845
1	OX	0.32	0/1365	0.41	0/1845
1	PX	0.32	0/1365	0.39	0/1845
1	QX	0.33	0/1365	0.42	0/1845
2	AY	0.24	0/1061	0.38	0/1432
2	BY	0.24	0/1061	0.38	0/1432
2	CY	0.24	0/1061	0.40	0/1432
2	DY	0.24	0/1061	0.38	0/1432
2	EY	0.24	0/1061	0.38	0/1432
2	FY	0.25	0/1061	0.37	0/1432
2	GY	0.24	0/1061	0.38	0/1432
2	HY	0.25	0/1061	0.41	0/1432
2	IY	0.24	0/1061	0.37	0/1432
2	JY	0.25	0/1061	0.41	0/1432
2	KY	0.25	0/1061	0.38	0/1432
2	LY	0.24	0/1061	0.38	0/1432
2	MY	0.24	0/1061	0.38	0/1432
2	NY	0.24	0/1061	0.38	0/1432
2	OY	0.24	0/1061	0.40	0/1432
2	PY	0.24	0/1061	0.37	0/1432
2	QY	0.24	0/1061	0.37	0/1432

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#|Z| >5	RMSZ	#|Z| >5
All	All	0.29	0/41242	0.39	0/55709

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	AX	1339	0	1355	47	0
1	BX	1339	0	1355	37	0
1	CX	1339	0	1355	53	0
1	DX	1339	0	1355	52	0
1	EX	1339	0	1355	44	0
1	FX	1339	0	1355	43	0
1	GX	1339	0	1355	48	0
1	HX	1339	0	1355	43	0
1	IX	1339	0	1355	42	0
1	JX	1339	0	1355	53	0
1	KX	1339	0	1355	41	0
1	LX	1339	0	1355	40	0
1	MX	1339	0	1355	43	0
1	NX	1339	0	1355	50	0
1	OX	1339	0	1355	43	0
1	PX	1339	0	1355	40	0
1	QX	1339	0	1355	45	0
2	AY	1050	0	1038	36	0
2	BY	1050	0	1038	40	0
2	CY	1050	0	1038	35	0
2	DY	1050	0	1038	36	0
2	EY	1050	0	1038	38	0
2	FY	1050	0	1038	31	0
2	GY	1050	0	1038	33	0
2	HY	1050	0	1038	34	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	IY	1050	0	1038	38	0
2	JY	1050	0	1038	37	0
2	KY	1050	0	1038	32	0
2	LY	1050	0	1038	31	0
2	MY	1050	0	1038	39	0
2	NY	1050	0	1038	37	0
2	OY	1050	0	1038	27	0
2	PY	1050	0	1038	29	0
2	QY	1050	0	1038	37	0
All	All	40613	0	40681	1098	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 14.

The worst 5 of 1098 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:LX:126:LEU:O	1:LX:130:LYS:HB2	1.65	0.96
1:MX:126:LEU:O	1:MX:130:LYS:HB2	1.68	0.93
1:QX:126:LEU:O	1:QX:130:LYS:HB2	1.68	0.93
1:NX:126:LEU:O	1:NX:130:LYS:HB2	1.71	0.91
1:JX:126:LEU:O	1:JX:130:LYS:HB2	1.71	0.91

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	AX	160/521 (31%)	154 (96%)	6 (4%)	0	100 100
1	BX	160/521 (31%)	151 (94%)	9 (6%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	CX	160/521 (31%)	154 (96%)	6 (4%)	0	100	100
1	DX	160/521 (31%)	156 (98%)	4 (2%)	0	100	100
1	EX	160/521 (31%)	153 (96%)	7 (4%)	0	100	100
1	FX	160/521 (31%)	152 (95%)	8 (5%)	0	100	100
1	GX	160/521 (31%)	153 (96%)	7 (4%)	0	100	100
1	HX	160/521 (31%)	155 (97%)	5 (3%)	0	100	100
1	IX	160/521 (31%)	154 (96%)	6 (4%)	0	100	100
1	JX	160/521 (31%)	151 (94%)	9 (6%)	0	100	100
1	KX	160/521 (31%)	157 (98%)	3 (2%)	0	100	100
1	LX	160/521 (31%)	154 (96%)	6 (4%)	0	100	100
1	MX	160/521 (31%)	155 (97%)	5 (3%)	0	100	100
1	NX	160/521 (31%)	156 (98%)	4 (2%)	0	100	100
1	OX	160/521 (31%)	152 (95%)	8 (5%)	0	100	100
1	PX	160/521 (31%)	154 (96%)	6 (4%)	0	100	100
1	QX	160/521 (31%)	154 (96%)	6 (4%)	0	100	100
2	AY	133/1927 (7%)	127 (96%)	6 (4%)	0	100	100
2	BY	133/1927 (7%)	128 (96%)	5 (4%)	0	100	100
2	CY	133/1927 (7%)	129 (97%)	4 (3%)	0	100	100
2	DY	133/1927 (7%)	126 (95%)	7 (5%)	0	100	100
2	EY	133/1927 (7%)	127 (96%)	6 (4%)	0	100	100
2	FY	133/1927 (7%)	123 (92%)	10 (8%)	0	100	100
2	GY	133/1927 (7%)	129 (97%)	4 (3%)	0	100	100
2	HY	133/1927 (7%)	126 (95%)	7 (5%)	0	100	100
2	IY	133/1927 (7%)	128 (96%)	5 (4%)	0	100	100
2	JY	133/1927 (7%)	127 (96%)	6 (4%)	0	100	100
2	KY	133/1927 (7%)	126 (95%)	7 (5%)	0	100	100
2	LY	133/1927 (7%)	127 (96%)	6 (4%)	0	100	100
2	MY	133/1927 (7%)	126 (95%)	7 (5%)	0	100	100
2	NY	133/1927 (7%)	127 (96%)	6 (4%)	0	100	100
2	OY	133/1927 (7%)	129 (97%)	4 (3%)	0	100	100
2	PY	133/1927 (7%)	127 (96%)	6 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	QY	133/1927 (7%)	128 (96%)	5 (4%)	0	100	100
All	All	4981/41616 (12%)	4775 (96%)	206 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	AX	152/469 (32%)	141 (93%)	11 (7%)	13	38
1	BX	152/469 (32%)	139 (91%)	13 (9%)	10	33
1	CX	152/469 (32%)	140 (92%)	12 (8%)	11	36
1	DX	152/469 (32%)	135 (89%)	17 (11%)	6	25
1	EX	152/469 (32%)	137 (90%)	15 (10%)	7	29
1	FX	152/469 (32%)	140 (92%)	12 (8%)	11	36
1	GX	152/469 (32%)	136 (90%)	16 (10%)	6	27
1	HX	152/469 (32%)	142 (93%)	10 (7%)	15	41
1	IX	152/469 (32%)	143 (94%)	9 (6%)	18	44
1	JX	152/469 (32%)	139 (91%)	13 (9%)	10	33
1	KX	152/469 (32%)	136 (90%)	16 (10%)	6	27
1	LX	152/469 (32%)	136 (90%)	16 (10%)	6	27
1	MX	152/469 (32%)	141 (93%)	11 (7%)	13	38
1	NX	152/469 (32%)	144 (95%)	8 (5%)	20	47
1	OX	152/469 (32%)	140 (92%)	12 (8%)	11	36
1	PX	152/469 (32%)	142 (93%)	10 (7%)	15	41
1	QX	152/469 (32%)	140 (92%)	12 (8%)	11	36
2	AY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	BY	115/1724 (7%)	113 (98%)	2 (2%)	53	70
2	CY	115/1724 (7%)	109 (95%)	6 (5%)	21	47

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	DY	115/1724 (7%)	111 (96%)	4 (4%)	32	57
2	EY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	FY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	GY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	HY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	IY	115/1724 (7%)	110 (96%)	5 (4%)	26	51
2	JY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	KY	115/1724 (7%)	111 (96%)	4 (4%)	32	57
2	LY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	MY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	NY	115/1724 (7%)	114 (99%)	1 (1%)	70	76
2	OY	115/1724 (7%)	112 (97%)	3 (3%)	40	62
2	PY	115/1724 (7%)	113 (98%)	2 (2%)	53	70
2	QY	115/1724 (7%)	108 (94%)	7 (6%)	17	43
All	All	4539/37281 (12%)	4268 (94%)	271 (6%)	19	44

5 of 271 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	FY	1497	CYS
2	HY	1497	CYS
2	OY	1520	ASP
1	HX	64	VAL
1	GX	299	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 113 such sidechains are listed below:

Mol	Chain	Res	Type
2	EY	1476	HIS
2	QY	1557	ASN
2	HY	1583	GLN
2	QY	1549	ASN
2	OY	1476	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

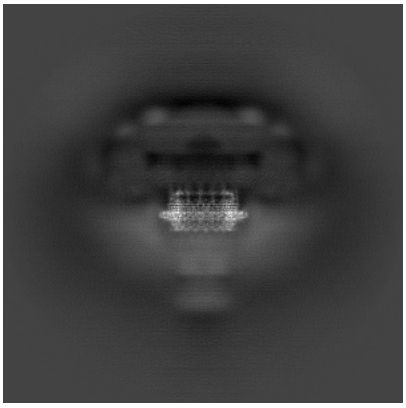
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-20021. These allow visual inspection of the internal detail of the map and identification of artifacts.

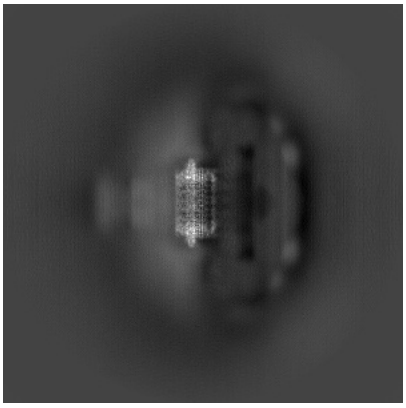
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

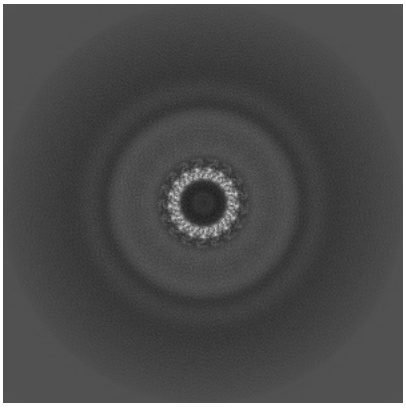
6.1.1 Primary map



X



Y

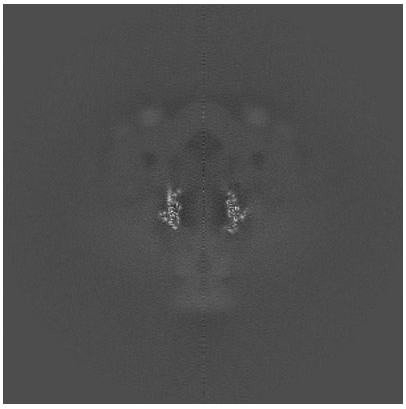


Z

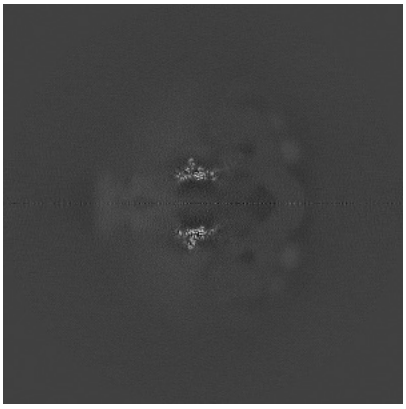
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

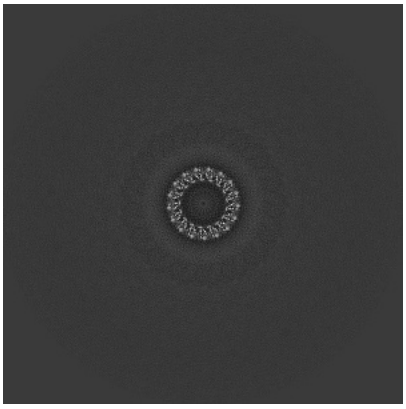
6.2.1 Primary map



X Index: 255



Y Index: 255

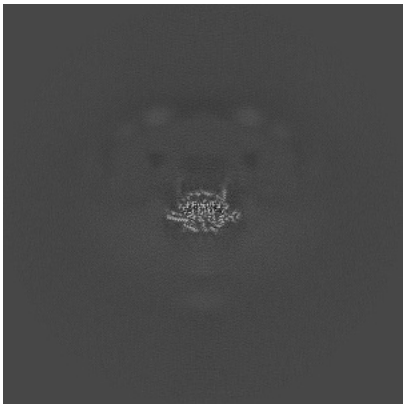


Z Index: 255

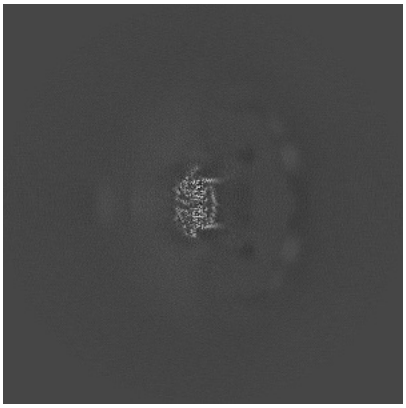
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

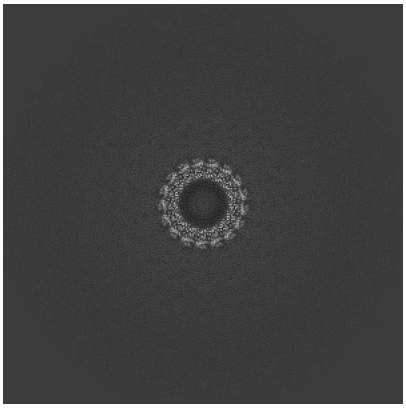
6.3.1 Primary map



X Index: 288



Y Index: 223

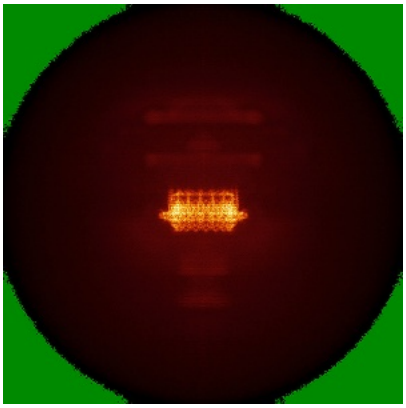


Z Index: 243

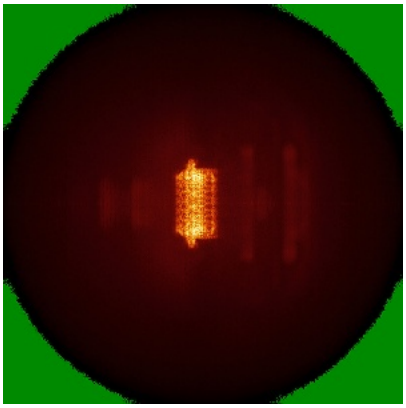
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) (i)

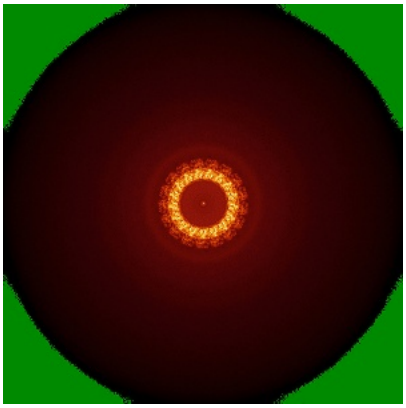
6.4.1 Primary map



X



Y



Z

The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views (i)

6.5.1 Primary map

X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.03. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

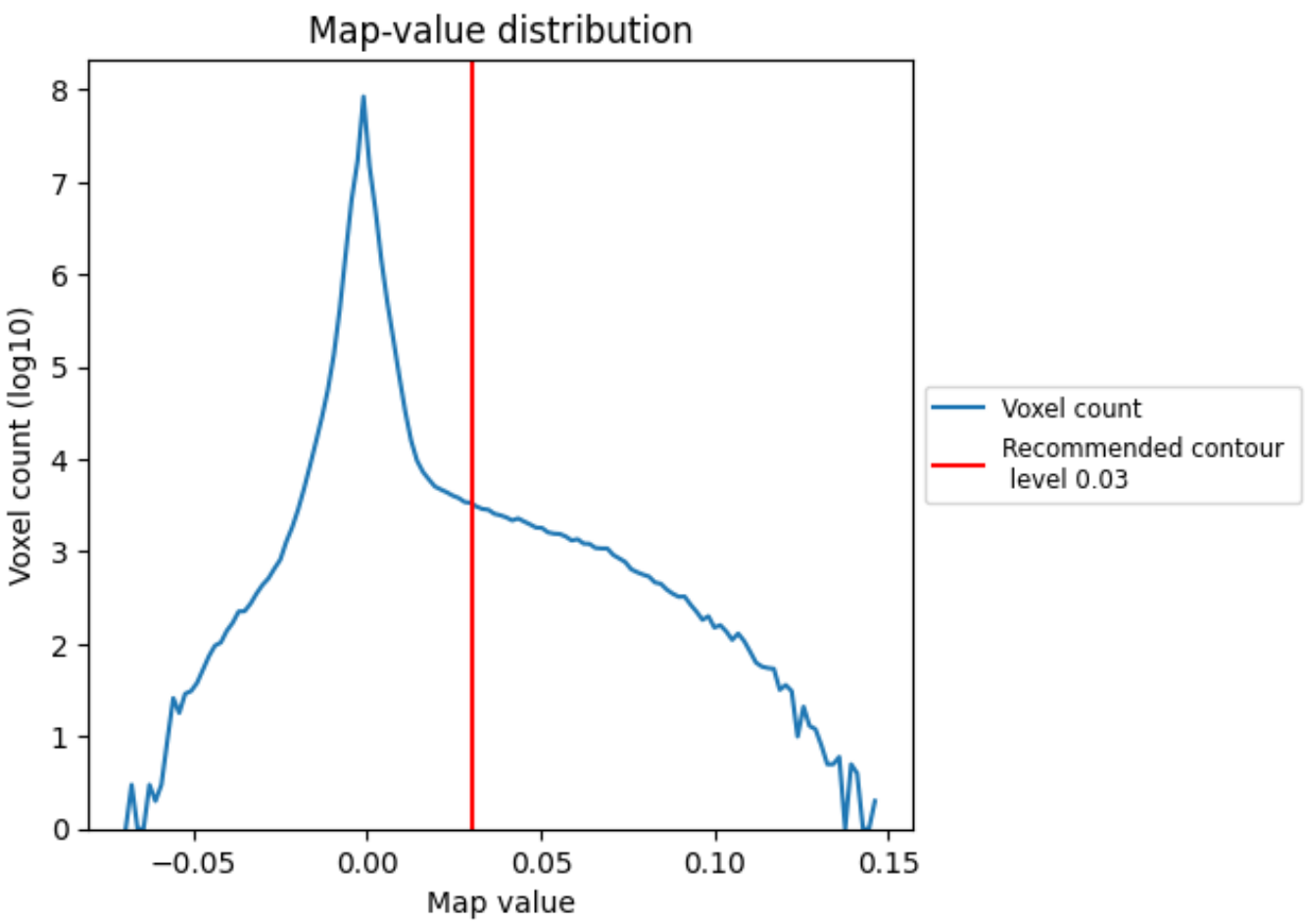
6.6 Mask visualisation (i)

This section was not generated. No masks/segmentation were deposited.

7 Map analysis (i)

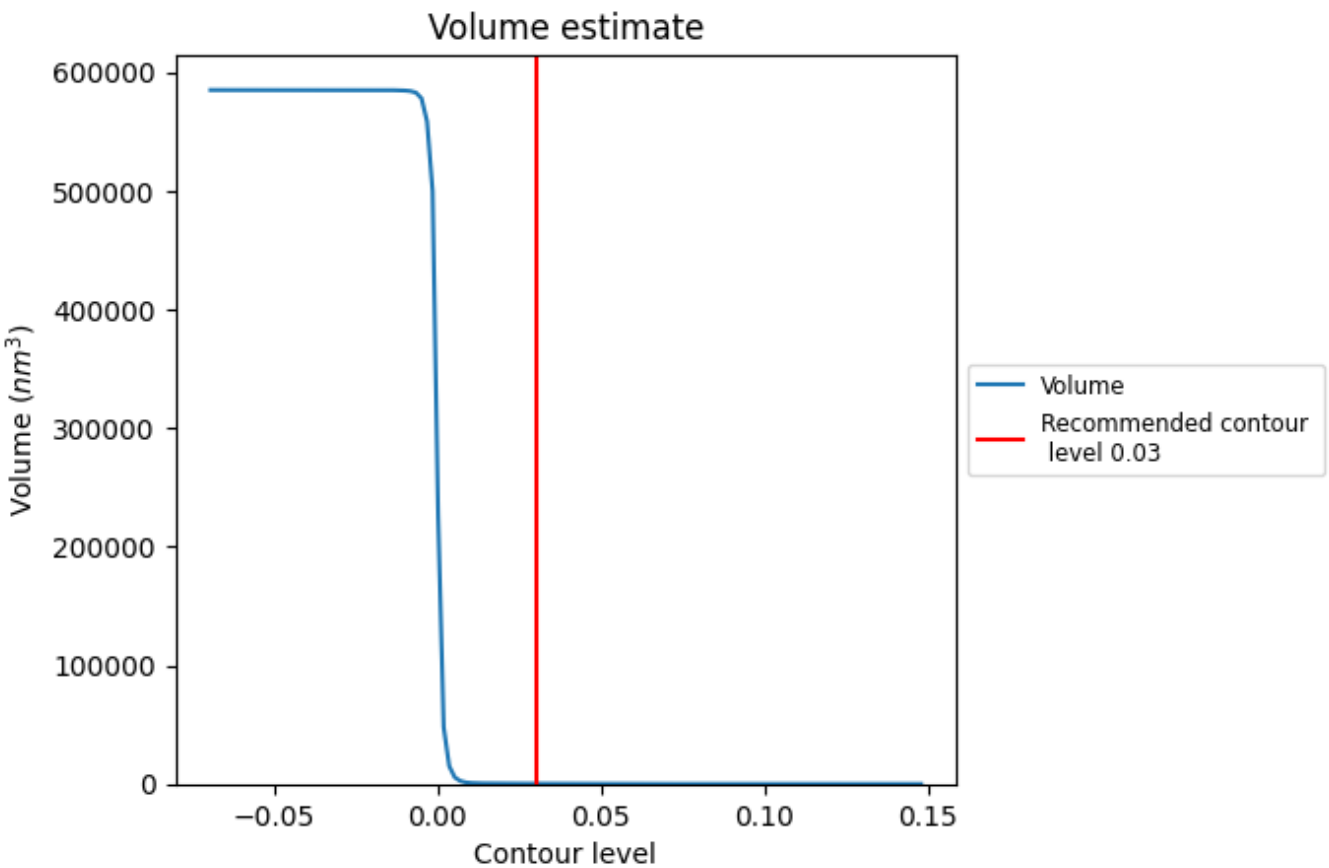
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution (i)



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

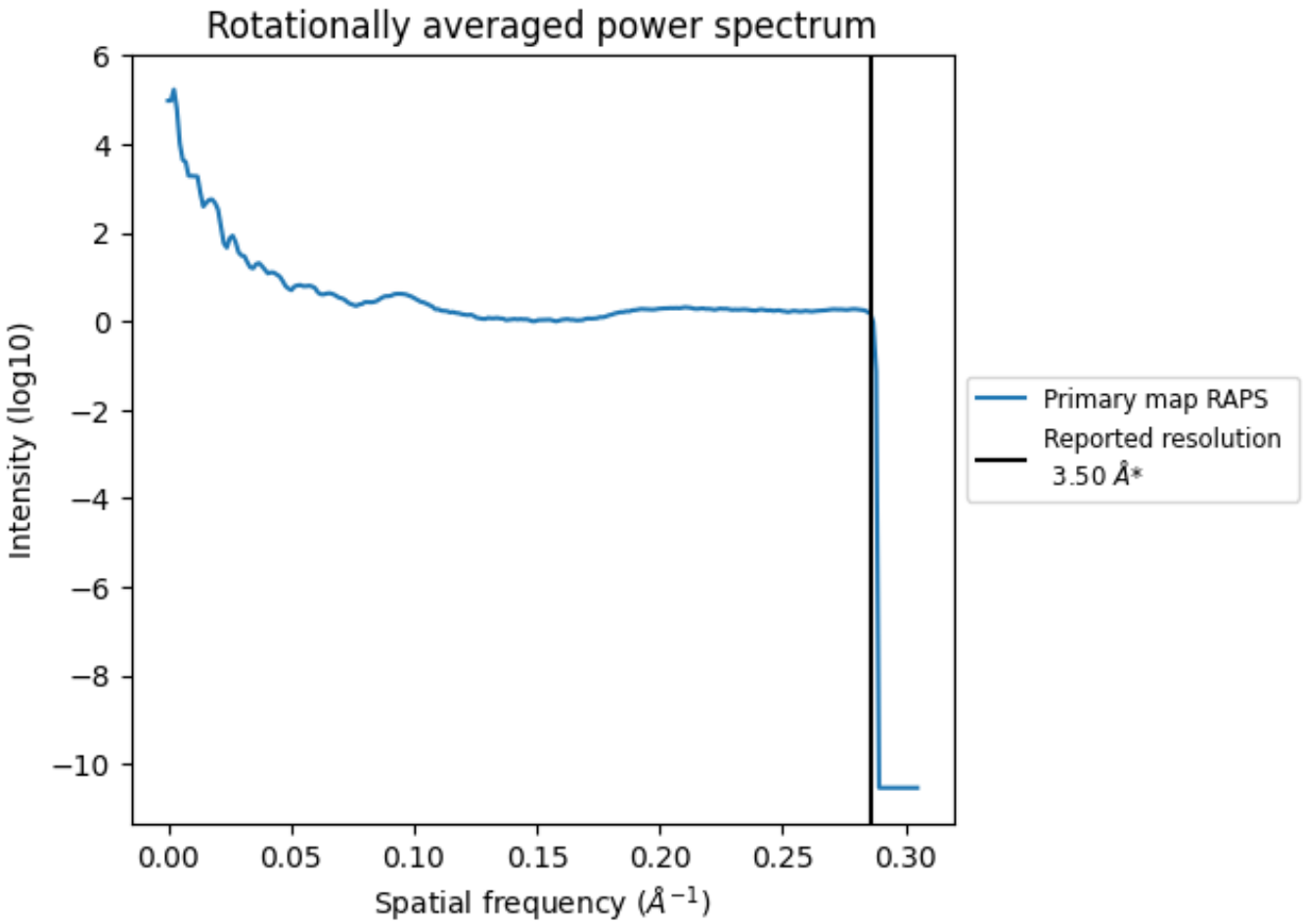
7.2 Volume estimate (i)



The volume at the recommended contour level is 244 nm^3; this corresponds to an approximate mass of 220 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum (i)

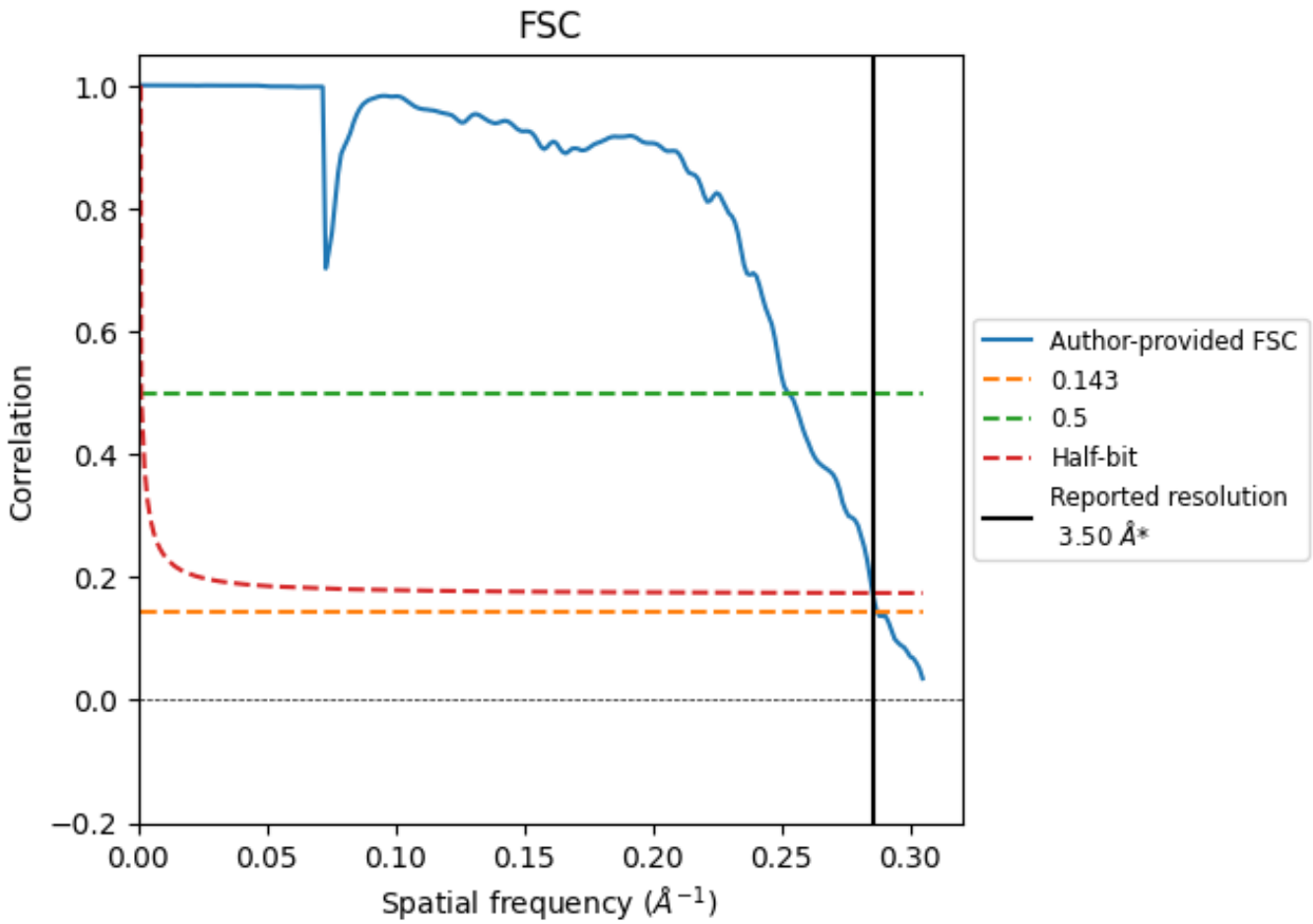


*Reported resolution corresponds to spatial frequency of 0.286 Å^{-1}

8 Fourier-Shell correlation (i)

Fourier-Shell Correlation (FSC) is the most commonly used method to estimate the resolution of single-particle and subtomogram-averaged maps. The shape of the curve depends on the imposed symmetry, mask and whether or not the two 3D reconstructions used were processed from a common reference. The reported resolution is shown as a black line. A curve is displayed for the half-bit criterion in addition to lines showing the 0.143 gold standard cut-off and 0.5 cut-off.

8.1 FSC (i)



*Reported resolution corresponds to spatial frequency of 0.286 Å^{-1}

8.2 Resolution estimates (i)

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.50	-	-
Author-provided FSC curve	3.48	3.96	3.50
Unmasked-calculated*	-	-	-

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

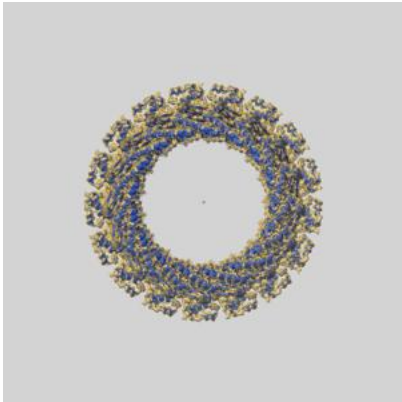
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-20021 and PDB model 6X6J. Per-residue inclusion information can be found in section 3 on page 8.

9.1 Map-model overlay (i)

X

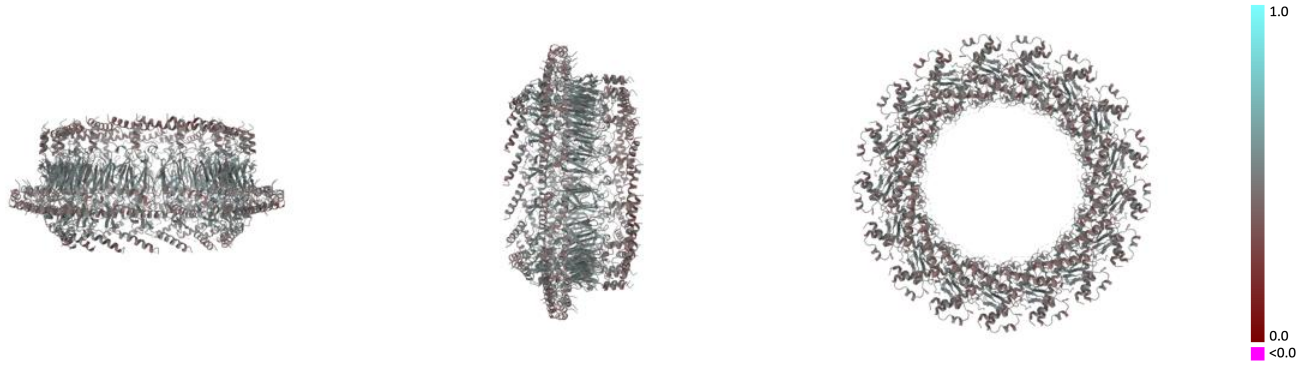
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.03 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

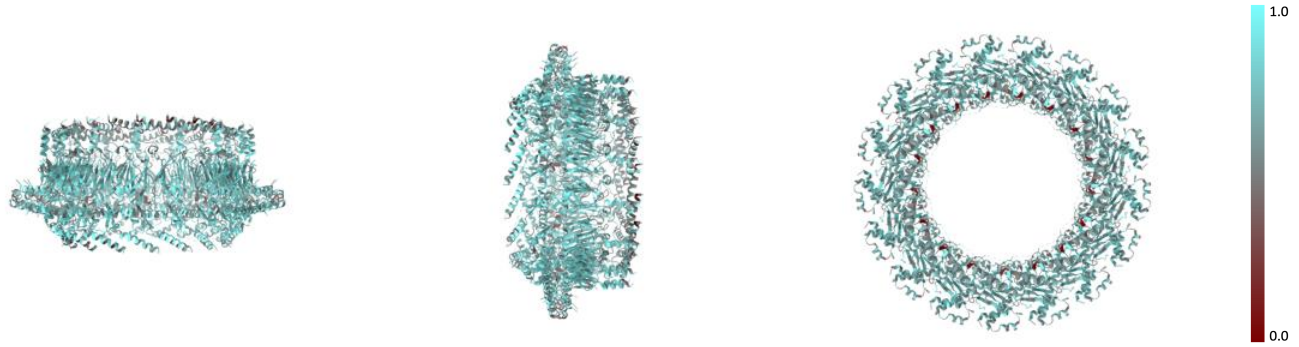
9.2 Q-score mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

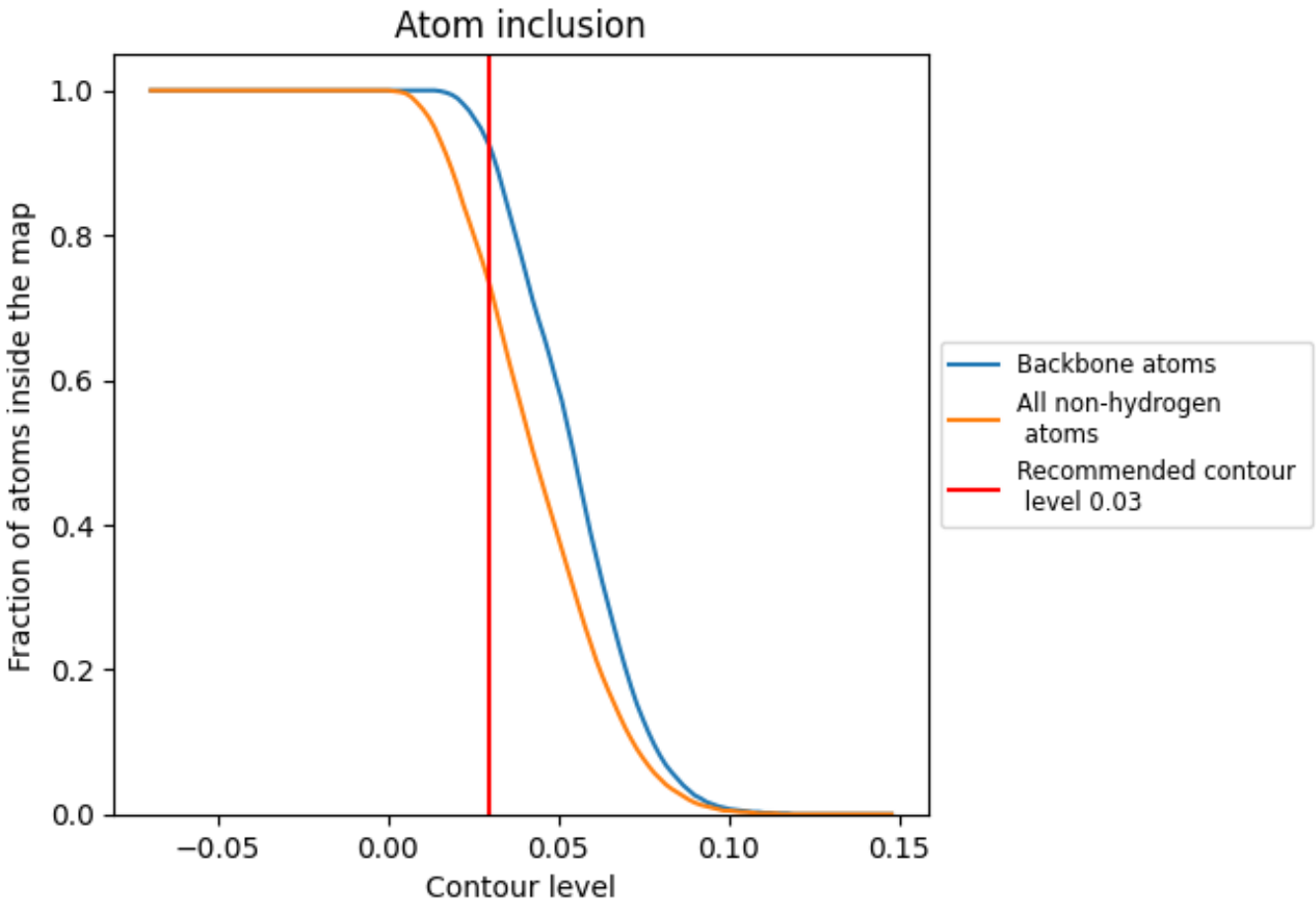
9.3 Atom inclusion mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.03).

9.4 Atom inclusion (i)



At the recommended contour level, 92% of all backbone atoms, 73% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (0.03) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7300	0.4730
AX	0.7670	0.4910
AY	0.6670	0.4470
BX	0.7720	0.4910
BY	0.6810	0.4480
CX	0.7720	0.4910
CY	0.6720	0.4460
DX	0.7740	0.4930
DY	0.6730	0.4530
EX	0.7800	0.4950
EY	0.6740	0.4500
FX	0.7790	0.4930
FY	0.6740	0.4470
GX	0.7800	0.4950
GY	0.6780	0.4480
HX	0.7740	0.4920
HY	0.6760	0.4460
IX	0.7720	0.4950
IY	0.6790	0.4450
JX	0.7700	0.4880
JY	0.6600	0.4430
KX	0.7800	0.4920
KY	0.6700	0.4480
LX	0.7740	0.4920
LY	0.6830	0.4470
MX	0.7690	0.4930
MY	0.6730	0.4470
NX	0.7670	0.4920
NY	0.6760	0.4470
OX	0.7760	0.4930
OY	0.6680	0.4470
PX	0.7740	0.4910
PY	0.6790	0.4490
QX	0.7770	0.4960
QY	0.6820	0.4500