wwPDB EM Validation Summary Report (i)

Feb 25, 2024 – 11:19 AM EST

PDB ID	:	6X6S
EMDB ID	:	EMD-22081
Title	:	Cryo-EM Structure of the Helicobacter pylori OMC
Authors	:	Sheedlo, M.J.; Chung, J.M.; Sawhney, N.; Durie, C.L.; Cover, T.L.; Ohi, M.D.; Lacy, D.B.
Deposited on	:	2020-05-29
Resolution	:	3.40 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:

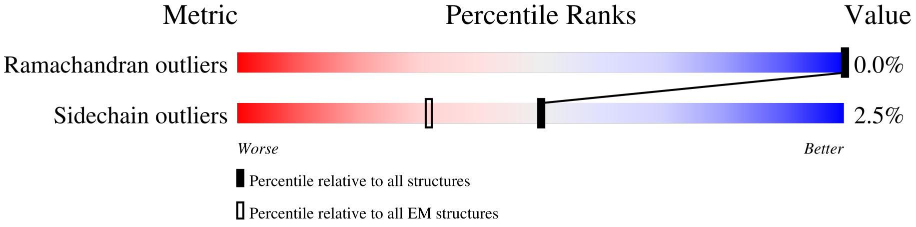
EMDB validation analysis	:	0.0.1.dev70
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

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

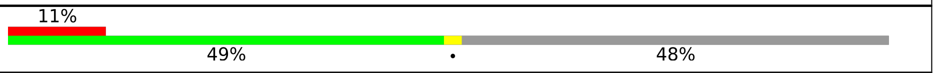
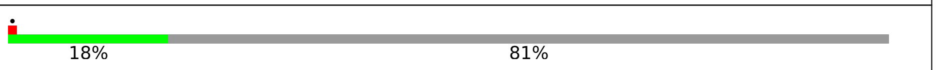
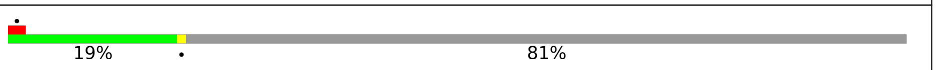
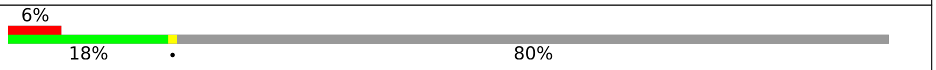
The reported resolution of this entry is 3.40 Å.

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)
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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	AA	481	
1	AB	481	
1	AC	481	
1	AD	481	
1	AE	481	
1	BA	481	
1	BB	481	
1	BC	481	
1	BD	481	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	BE	481	10% 21% • 78%
1	CA	481	11% 49% • 48%
1	CB	481	• 18% 81%
1	CC	481	• 19% 81%
1	CD	481	7% 19% • 80%
1	CE	481	10% 21% • 78%
1	DA	481	10% 49% • 48%
1	DB	481	• 18% • 81%
1	DC	481	• 18% • 81%
1	DD	481	7% 19% • 80%
1	DE	481	10% 20% • 78%
1	EA	481	11% 49% • 48%
1	EB	481	• 18% • 81%
1	EC	481	• 19% • 81%
1	ED	481	7% 18% • 80%
1	EE	481	11% 20% • 78%
1	FA	481	11% 49% • 48%
1	FB	481	• 18% 81%
1	FC	481	• 18% • 81%
1	FD	481	7% 19% • 80%
1	FE	481	10% 20% • 78%
1	GA	481	12% 49% • 48%
1	GB	481	• 19% 81%
1	GC	481	• 18% • 81%
1	GD	481	7% 18% • 80%

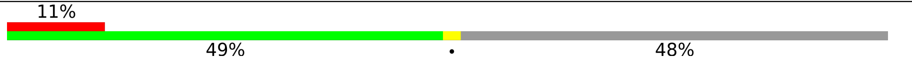
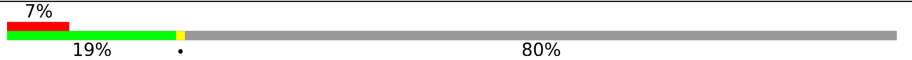
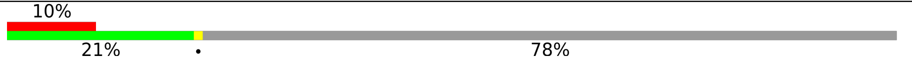
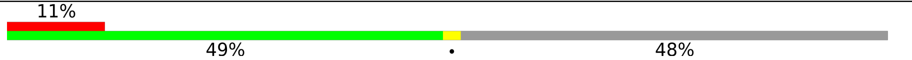
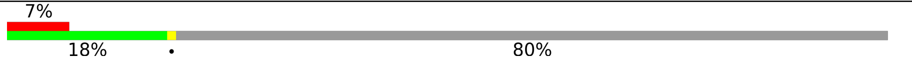
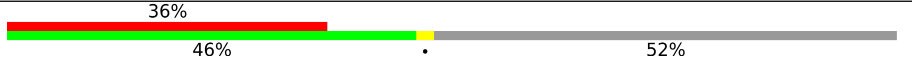
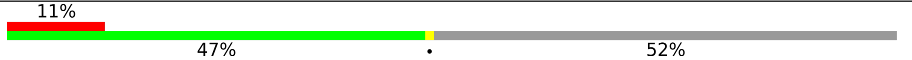
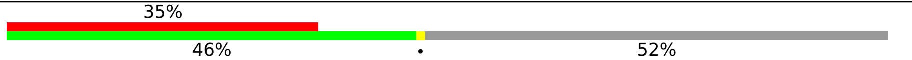
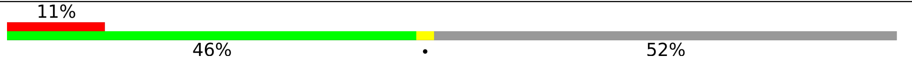
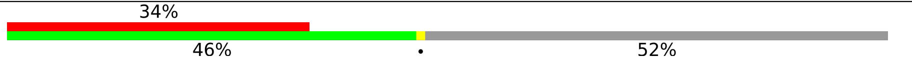
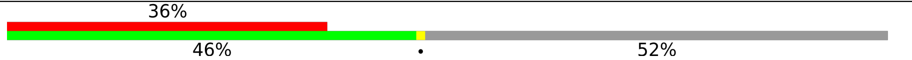
Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	GE	481	10% 20% • 78%
1	HA	481	11% 49% • 48%
1	HB	481	• 18% 81%
1	HC	481	• 19% • 81%
1	HD	481	7% 18% • 80%
1	HE	481	10% 20% • 78%
1	IA	481	11% 49% • 48%
1	IB	481	• 18% 81%
1	IC	481	• 19% 81%
1	ID	481	6% 18% • 80%
1	IE	481	10% 21% • 78%
1	JA	481	11% 49% • 48%
1	JB	481	• 18% 81%
1	JC	481	• 19% • 81%
1	JD	481	7% 18% • 80%
1	JE	481	10% 21% • 78%
1	KA	481	10% 49% • 48%
1	KB	481	• 19% 81%
1	KC	481	• 18% • 81%
1	KD	481	7% 19% • 80%
1	KE	481	10% 20% • 78%
1	LA	481	12% 49% • 48%
1	LB	481	• 19% 81%
1	LC	481	• 18% • 81%
1	LD	481	7% 19% • 80%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	LE	481	
1	MA	481	
1	MB	481	
1	MC	481	
1	MD	481	
1	ME	481	
1	NA	481	
1	NB	481	
1	NC	481	
1	ND	481	
1	NE	481	
2	AM	376	
2	Am	376	
2	BM	376	
2	Bm	376	
2	CM	376	
2	Cm	376	
2	DM	376	
2	Dm	376	
2	EM	376	
2	Em	376	
2	FM	376	
2	Fm	376	
2	GM	376	
2	Gm	376	

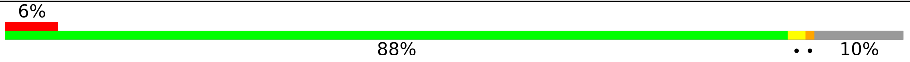
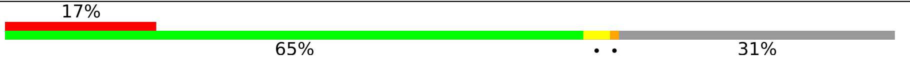
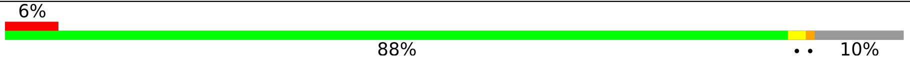
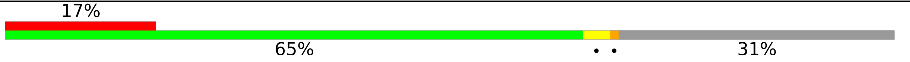
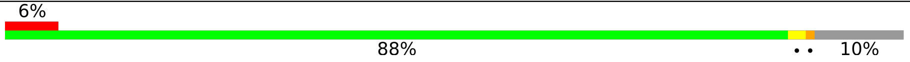
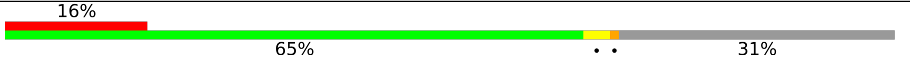
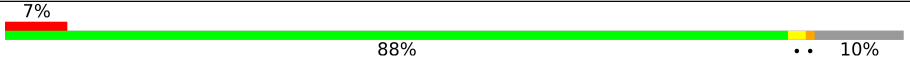
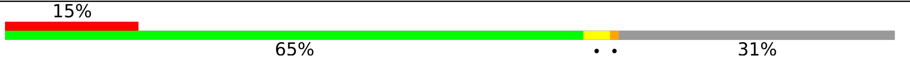
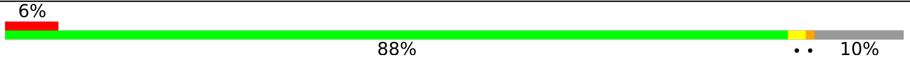
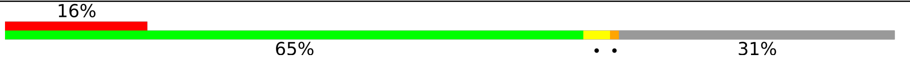
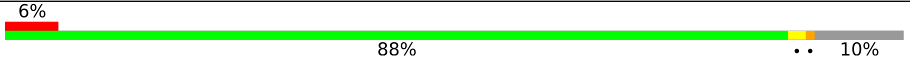
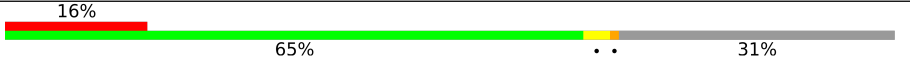
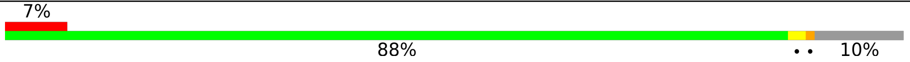
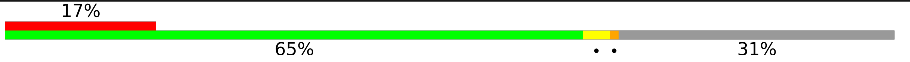
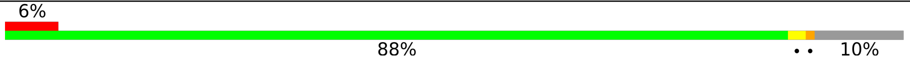
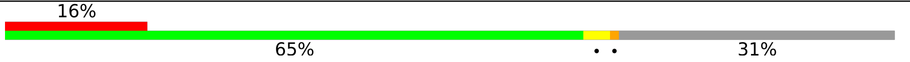
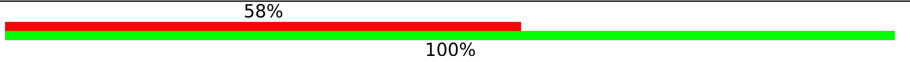
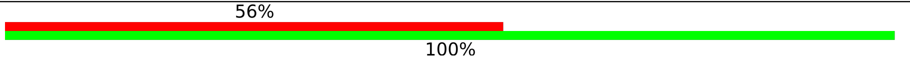
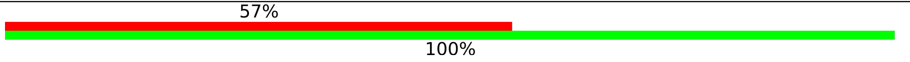
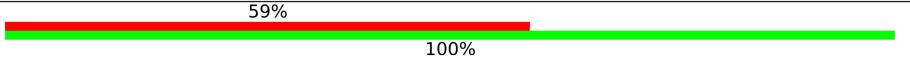
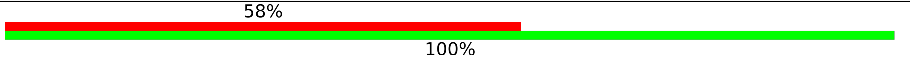
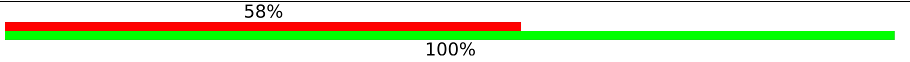
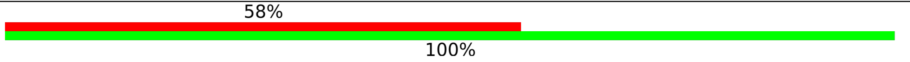
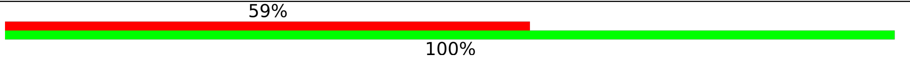
Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	HM	376	10% 47% • 52%
2	Hm	376	36% 45% • 52%
2	IM	376	10% 47% • 52%
2	Im	376	36% 46% • 52%
2	JM	376	11% 48% 52%
2	Jm	376	35% 46% • 52%
2	KM	376	11% 47% • 52%
2	Km	376	35% 47% • 52%
2	LM	376	10% 47% • 52%
2	Lm	376	35% 46% • 52%
2	MM	376	10% 47% • 52%
2	Mm	376	34% 47% • 52%
2	NM	376	10% 47% • 52%
2	Nm	376	35% 46% • 52%
3	AT	280	6% 88% •• 10%
3	At	280	17% 65% •• 31%
3	BT	280	6% 88% •• 10%
3	Bt	280	17% 65% •• 31%
3	CT	280	7% 88% •• 10%
3	Ct	280	15% 65% •• 31%
3	DT	280	7% 88% •• 10%
3	Dt	280	16% 65% •• 31%
3	ET	280	6% 88% •• 10%
3	Et	280	16% 65% •• 31%
3	FT	280	6% 88% •• 10%

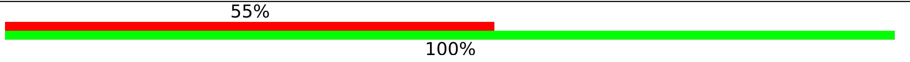
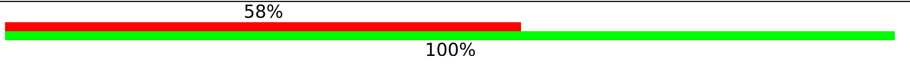
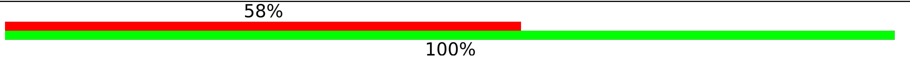
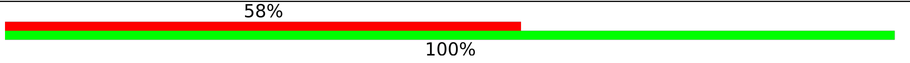
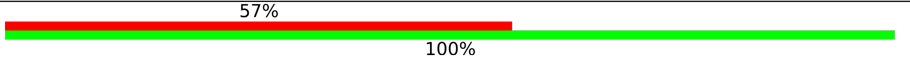
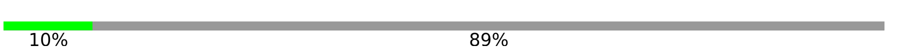
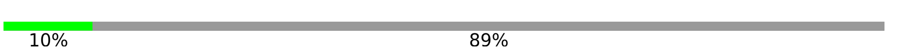
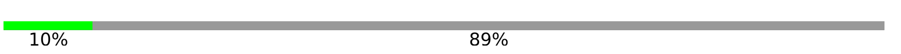
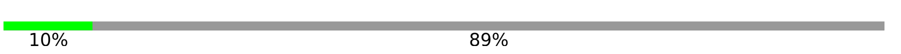
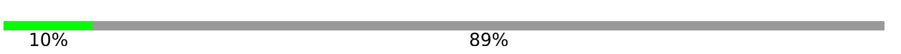
Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	Ft	280	
3	GT	280	
3	Gt	280	
3	HT	280	
3	Ht	280	
3	IT	280	
3	It	280	
3	JT	280	
3	Jt	280	
3	KT	280	
3	Kt	280	
3	LT	280	
3	Lt	280	
3	MT	280	
3	Mt	280	
3	NT	280	
3	Nt	280	
4	AU	160	
4	BU	160	
4	CU	160	
4	DU	160	
4	EU	160	
4	FU	160	
4	GU	160	
4	HU	160	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	IU	160	 56% 100%
4	JU	160	 55% 100%
4	KU	160	 58% 100%
4	LU	160	 58% 100%
4	MU	160	 58% 100%
4	NU	160	 57% 100%
5	AX	522	 • 30% •• 68%
5	BX	522	 • 29% •• 68%
5	CX	522	 • 30% •• 68%
5	DX	522	 • 30% •• 68%
5	EX	522	 • 30% •• 68%
5	FX	522	 • 30% •• 68%
5	GX	522	 • 30% •• 68%
5	HX	522	 • 30% •• 68%
5	IX	522	 • 30% •• 68%
5	JX	522	 • 30% •• 68%
5	KX	522	 • 30% •• 68%
5	LX	522	 • 29% •• 68%
5	MX	522	 • 30% •• 68%
5	NX	522	 • 30% •• 68%
6	AY	1927	 10% 89%
6	BY	1927	 10% 89%
6	CY	1927	 10% 89%
6	DY	1927	 10% 89%
6	EY	1927	 10% 89%

Continued on next page...

Continued from previous page...

Mol	**Chain**	**Length**	**Quality of chain**
6	FY	1927	10% 89%
6	GY	1927	10% 89%
6	HY	1927	10% 89%
6	IY	1927	10% 89%
6	JY	1927	10% 89%
6	KY	1927	10% 89%
6	LY	1927	10% 89%
6	MY	1927	10% 89%
6	NY	1927	10% 89%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 215978 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 Type IV secretion system apparatus protein Cag3.

Mol	Chain	Residues	Atoms	AltConf	Trace
1	AA	248	Total C N O S 1986 1254 329 395 8	0	0
1	AB	90	Total C N O S 728 468 122 135 3	0	0
1	AC	92	Total C N O S 741 476 125 137 3	0	0
1	AD	95	Total C N O S 778 508 124 143 3	0	0
1	AE	104	Total C N O S 847 549 136 159 3	0	0
1	BA	248	Total C N O S 1986 1254 329 395 8	0	0
1	BB	90	Total C N O S 728 468 122 135 3	0	0
1	BC	92	Total C N O S 741 476 125 137 3	0	0
1	BD	95	Total C N O S 778 508 124 143 3	0	0
1	BE	104	Total C N O S 847 549 136 159 3	0	0
1	CA	248	Total C N O S 1986 1254 329 395 8	0	0
1	CB	90	Total C N O S 728 468 122 135 3	0	0
1	CC	92	Total C N O S 741 476 125 137 3	0	0
1	CD	95	Total C N O S 778 508 124 143 3	0	0
1	CE	104	Total C N O S 847 549 136 159 3	0	0
1	DA	248	Total C N O S 1986 1254 329 395 8	0	0
1	DB	90	Total C N O S 728 468 122 135 3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	AltConf	Trace
1	DC	92	Total C N O S 741 476 125 137 3	0	0
1	DD	95	Total C N O S 778 508 124 143 3	0	0
1	DE	104	Total C N O S 847 549 136 159 3	0	0
1	EA	248	Total C N O S 1986 1254 329 395 8	0	0
1	EB	90	Total C N O S 728 468 122 135 3	0	0
1	EC	92	Total C N O S 741 476 125 137 3	0	0
1	ED	95	Total C N O S 778 508 124 143 3	0	0
1	EE	104	Total C N O S 847 549 136 159 3	0	0
1	FA	248	Total C N O S 1986 1254 329 395 8	0	0
1	FB	90	Total C N O S 728 468 122 135 3	0	0
1	FC	92	Total C N O S 741 476 125 137 3	0	0
1	FD	95	Total C N O S 778 508 124 143 3	0	0
1	FE	104	Total C N O S 847 549 136 159 3	0	0
1	GA	248	Total C N O S 1986 1254 329 395 8	0	0
1	GB	90	Total C N O S 728 468 122 135 3	0	0
1	GC	92	Total C N O S 741 476 125 137 3	0	0
1	GD	95	Total C N O S 778 508 124 143 3	0	0
1	GE	104	Total C N O S 847 549 136 159 3	0	0
1	HA	248	Total C N O S 1986 1254 329 395 8	0	0
1	HB	90	Total C N O S 728 468 122 135 3	0	0
1	HC	92	Total C N O S 741 476 125 137 3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					AltConf	Trace
1	HD	95	Total 778	C 508	N 124	O 143	S 3	0	0
1	HE	104	Total 847	C 549	N 136	O 159	S 3	0	0
1	IA	248	Total 1986	C 1254	N 329	O 395	S 8	0	0
1	IB	90	Total 728	C 468	N 122	O 135	S 3	0	0
1	IC	92	Total 741	C 476	N 125	O 137	S 3	0	0
1	ID	95	Total 778	C 508	N 124	O 143	S 3	0	0
1	IE	104	Total 847	C 549	N 136	O 159	S 3	0	0
1	JA	248	Total 1986	C 1254	N 329	O 395	S 8	0	0
1	JB	90	Total 728	C 468	N 122	O 135	S 3	0	0
1	JC	92	Total 741	C 476	N 125	O 137	S 3	0	0
1	JD	95	Total 778	C 508	N 124	O 143	S 3	0	0
1	JE	104	Total 847	C 549	N 136	O 159	S 3	0	0
1	KA	248	Total 1986	C 1254	N 329	O 395	S 8	0	0
1	KB	90	Total 728	C 468	N 122	O 135	S 3	0	0
1	KC	92	Total 741	C 476	N 125	O 137	S 3	0	0
1	KD	95	Total 778	C 508	N 124	O 143	S 3	0	0
1	KE	104	Total 847	C 549	N 136	O 159	S 3	0	0
1	LA	248	Total 1986	C 1254	N 329	O 395	S 8	0	0
1	LB	90	Total 728	C 468	N 122	O 135	S 3	0	0
1	LC	92	Total 741	C 476	N 125	O 137	S 3	0	0
1	LD	95	Total 778	C 508	N 124	O 143	S 3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					AltConf	Trace
1	LE	104	Total 847	C 549	N 136	O 159	S 3	0	0
1	MA	248	Total 1986	C 1254	N 329	O 395	S 8	0	0
1	MB	90	Total 728	C 468	N 122	O 135	S 3	0	0
1	MC	92	Total 741	C 476	N 125	O 137	S 3	0	0
1	MD	95	Total 778	C 508	N 124	O 143	S 3	0	0
1	ME	104	Total 847	C 549	N 136	O 159	S 3	0	0
1	NA	248	Total 1986	C 1254	N 329	O 395	S 8	0	0
1	NB	90	Total 728	C 468	N 122	O 135	S 3	0	0
1	NC	92	Total 741	C 476	N 125	O 137	S 3	0	0
1	ND	95	Total 778	C 508	N 124	O 143	S 3	0	0
1	NE	104	Total 847	C 549	N 136	O 159	S 3	0	0

There are 70 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
AB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
AC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
AD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
AE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
BA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
BB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
BC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
BD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
BE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
CA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
CB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
CC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
CD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
CE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
DA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
DB	275	ALA	GLN	conflict	UNP A0A2J9KJK3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
DC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
DD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
DE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
EA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
EB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
EC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
ED	275	ALA	GLN	conflict	UNP A0A2J9KJK3
EE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
FA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
FB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
FC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
FD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
FE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
GA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
GB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
GC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
GD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
GE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
HA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
HB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
HC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
HD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
HE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
IA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
IB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
IC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
ID	275	ALA	GLN	conflict	UNP A0A2J9KJK3
IE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
JA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
JB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
JC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
JD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
JE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
KA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
KB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
KC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
KD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
KE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
LA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
LB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
LC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
LD	275	ALA	GLN	conflict	UNP A0A2J9KJK3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
LE	275	ALA	GLN	conflict	UNP A0A2J9KJK3
MA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
MB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
MC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
MD	275	ALA	GLN	conflict	UNP A0A2J9KJK3
ME	275	ALA	GLN	conflict	UNP A0A2J9KJK3
NA	275	ALA	GLN	conflict	UNP A0A2J9KJK3
NB	275	ALA	GLN	conflict	UNP A0A2J9KJK3
NC	275	ALA	GLN	conflict	UNP A0A2J9KJK3
ND	275	ALA	GLN	conflict	UNP A0A2J9KJK3
NE	275	ALA	GLN	conflict	UNP A0A2J9KJK3

- Molecule 2 is a protein called Type IV secretion system apparatus protein CagM.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	AM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Am	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	BM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Bm	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	CM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Cm	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	DM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Dm	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	EM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Em	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	FM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Fm	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	GM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Gm	179	Total 1474	C 939	N 246	O 284	S 5	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					AltConf	Trace
2	HM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Hm	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	IM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Im	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	JM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Jm	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	KM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Km	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	LM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Lm	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	MM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Mm	179	Total 1474	C 939	N 246	O 284	S 5	0	0
2	NM	180	Total 1485	C 945	N 250	O 285	S 5	0	0
2	Nm	179	Total 1474	C 939	N 246	O 284	S 5	0	0

- Molecule 3 is a protein called Type IV secretion system apparatus protein CagT.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	AT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	At	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	BT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Bt	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	CT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					AltConf	Trace
3	Ct	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	DT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Dt	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	ET	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Et	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	FT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Ft	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	GT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Gt	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	HT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Ht	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	IT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	It	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	JT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Jt	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	KT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Kt	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	LT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Lt	192	Total 1615	C 1021	N 283	O 308	S 3	0	0
3	MT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Mt	192	Total 1615	C 1021	N 283	O 308	S 3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					AltConf	Trace
3	NT	253	Total 2083	C 1308	N 368	O 402	S 5	0	0
3	Nt	192	Total 1615	C 1021	N 283	O 308	S 3	0	0

- Molecule 4 is a protein called Unknown protein fragment.

Mol	Chain	Residues	Atoms				AltConf	Trace
4	AU	160	Total 800	C 480	N 160	O 160	0	0
4	BU	160	Total 800	C 480	N 160	O 160	0	0
4	CU	160	Total 800	C 480	N 160	O 160	0	0
4	DU	160	Total 800	C 480	N 160	O 160	0	0
4	EU	160	Total 800	C 480	N 160	O 160	0	0
4	FU	160	Total 800	C 480	N 160	O 160	0	0
4	GU	160	Total 800	C 480	N 160	O 160	0	0
4	HU	160	Total 800	C 480	N 160	O 160	0	0
4	IU	160	Total 800	C 480	N 160	O 160	0	0
4	JU	160	Total 800	C 480	N 160	O 160	0	0
4	KU	160	Total 800	C 480	N 160	O 160	0	0
4	LU	160	Total 800	C 480	N 160	O 160	0	0
4	MU	160	Total 800	C 480	N 160	O 160	0	0
4	NU	160	Total 800	C 480	N 160	O 160	0	0

- Molecule 5 is a protein called Type IV secretion system apparatus protein CagX.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	AX	166	Total 1361	C 871	N 237	O 248	S 5	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					AltConf	Trace
5	BX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	CX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	DX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	EX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	FX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	GX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	HX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	IX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	JX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	KX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	LX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	MX	166	Total 1361	C 871	N 237	O 248	S 5	0	0
5	NX	166	Total 1361	C 871	N 237	O 248	S 5	0	0

- Molecule 6 is a protein called Cag pathogenicity island protein (Cag7).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	AY	203	Total 1529	C 982	N 246	O 292	S 9	0	0
6	BY	203	Total 1529	C 982	N 246	O 292	S 9	0	0
6	CY	203	Total 1529	C 982	N 246	O 292	S 9	0	0
6	DY	203	Total 1529	C 982	N 246	O 292	S 9	0	0
6	EY	203	Total 1529	C 982	N 246	O 292	S 9	0	0
6	FY	203	Total 1529	C 982	N 246	O 292	S 9	0	0

Continued on next page...

Continued from previous page...

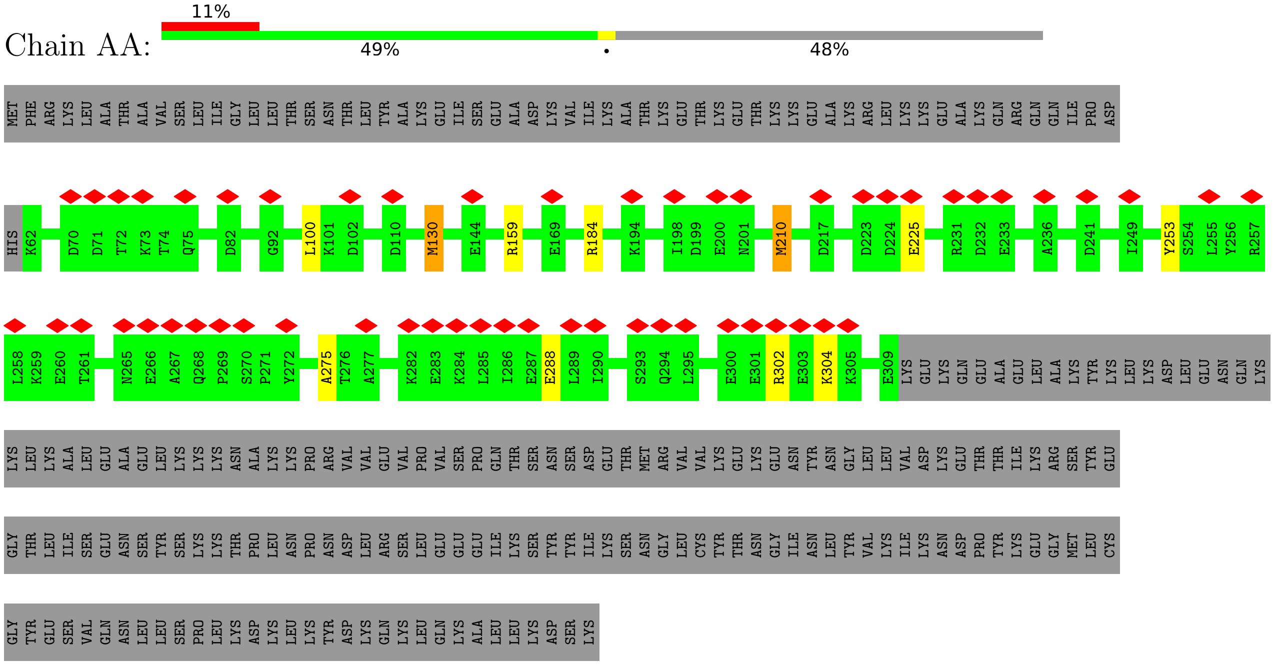
Mol	Chain	Residues	Atoms					AltConf	Trace
6	GY	203	Total	C	N	O	S	0	0
			1529	982	246	292	9		
6	HY	203	Total	C	N	O	S	0	0
			1529	982	246	292	9		
6	IY	203	Total	C	N	O	S	0	0
			1529	982	246	292	9		
6	JY	203	Total	C	N	O	S	0	0
			1529	982	246	292	9		
6	KY	203	Total	C	N	O	S	0	0
			1529	982	246	292	9		
6	LY	203	Total	C	N	O	S	0	0
			1529	982	246	292	9		
6	MY	203	Total	C	N	O	S	0	0
			1529	982	246	292	9		
6	NY	203	Total	C	N	O	S	0	0
			1529	982	246	292	9		

3 Residue-property plots ⓘ

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.

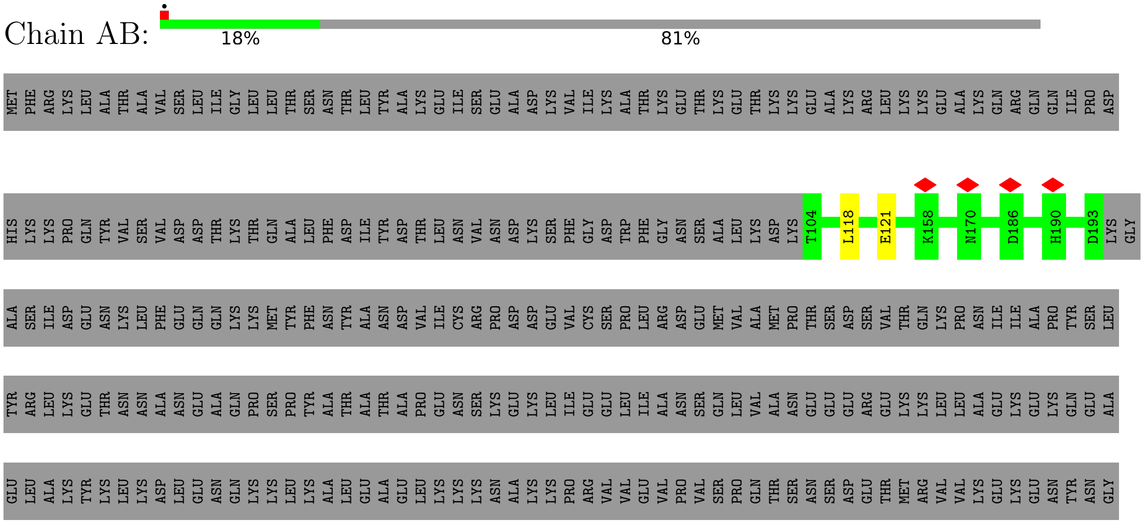
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain AA:



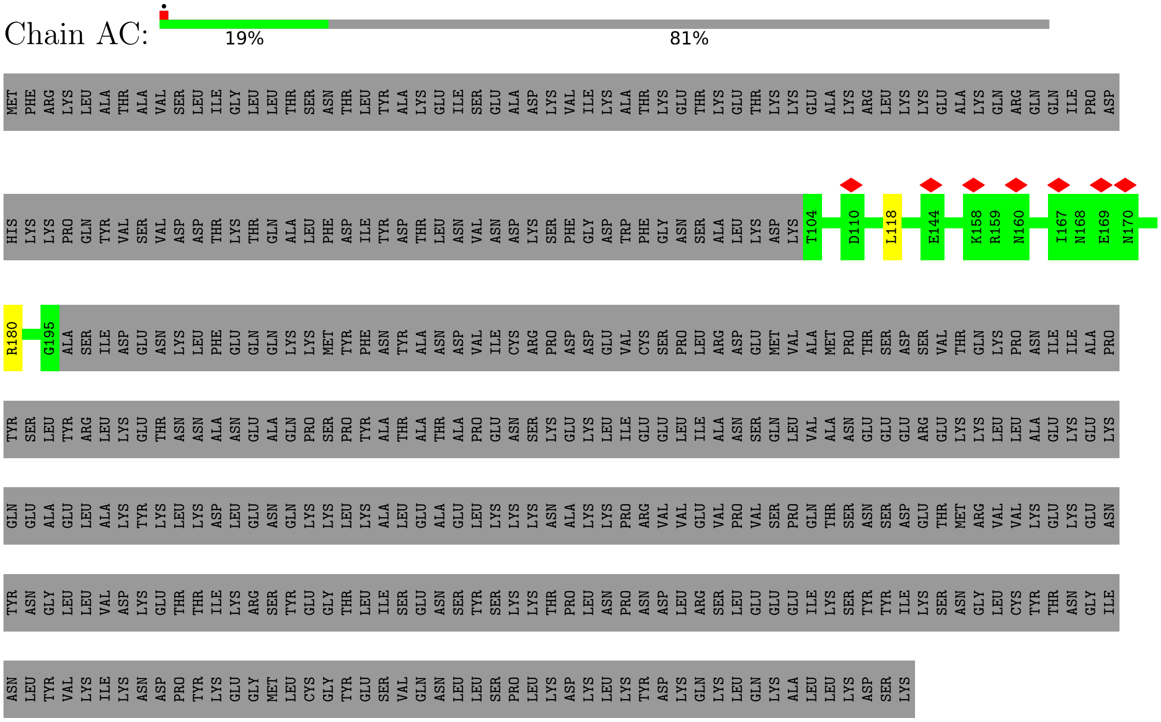
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain AB:



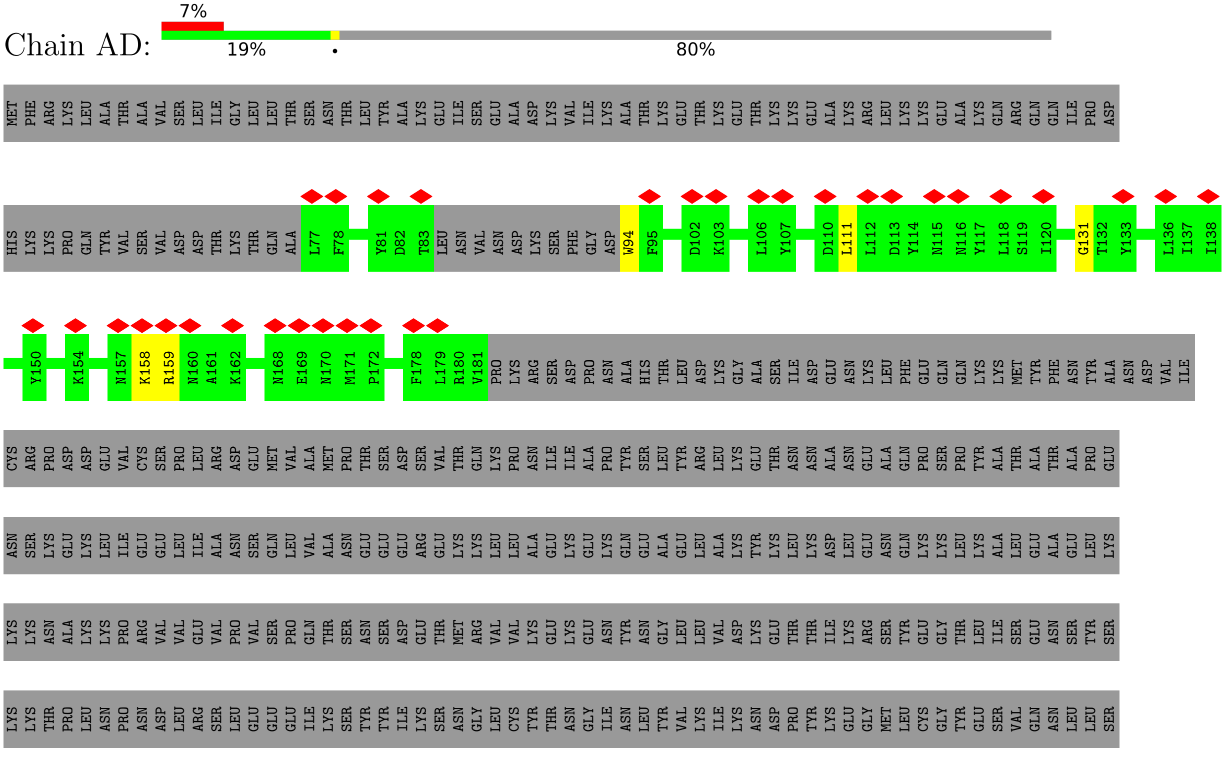
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain AC:

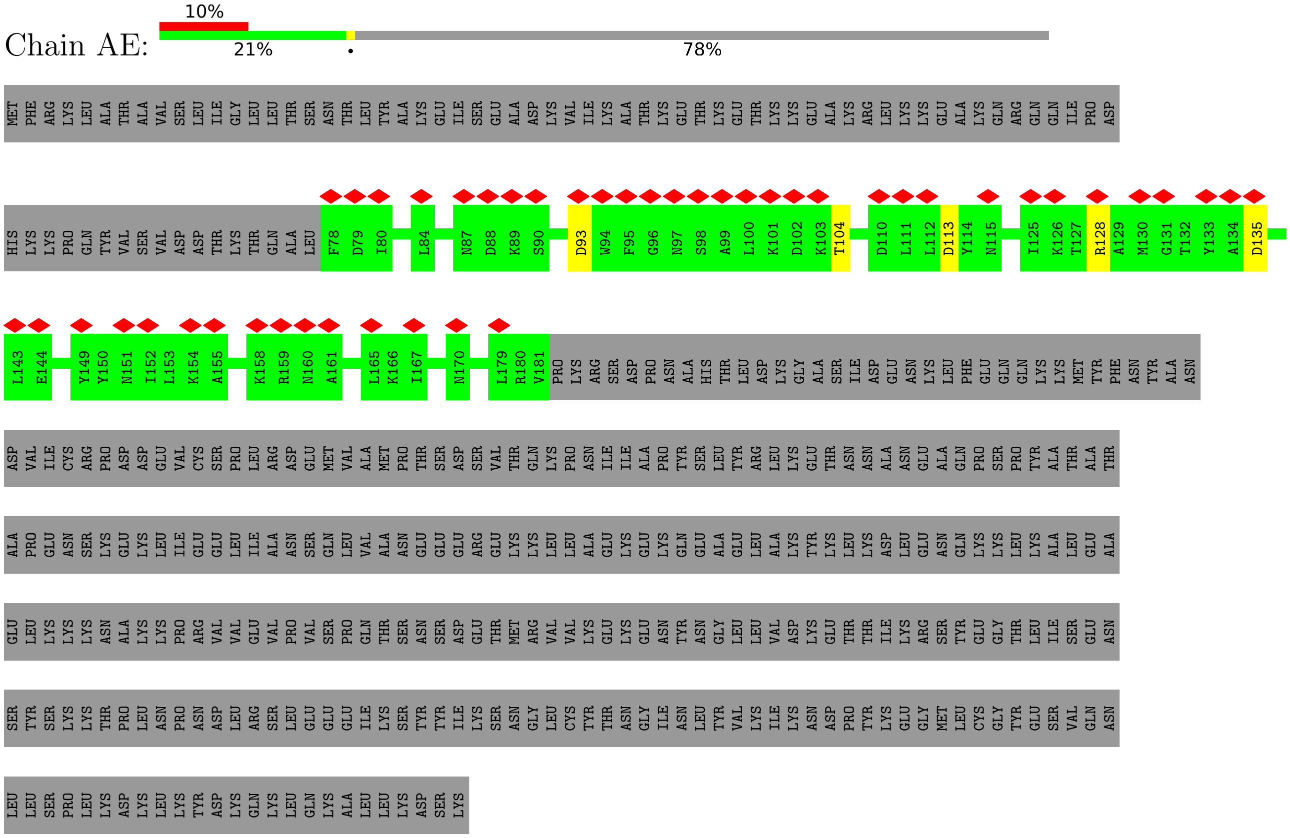


- Molecule 1: Type IV secretion system apparatus protein Cag3

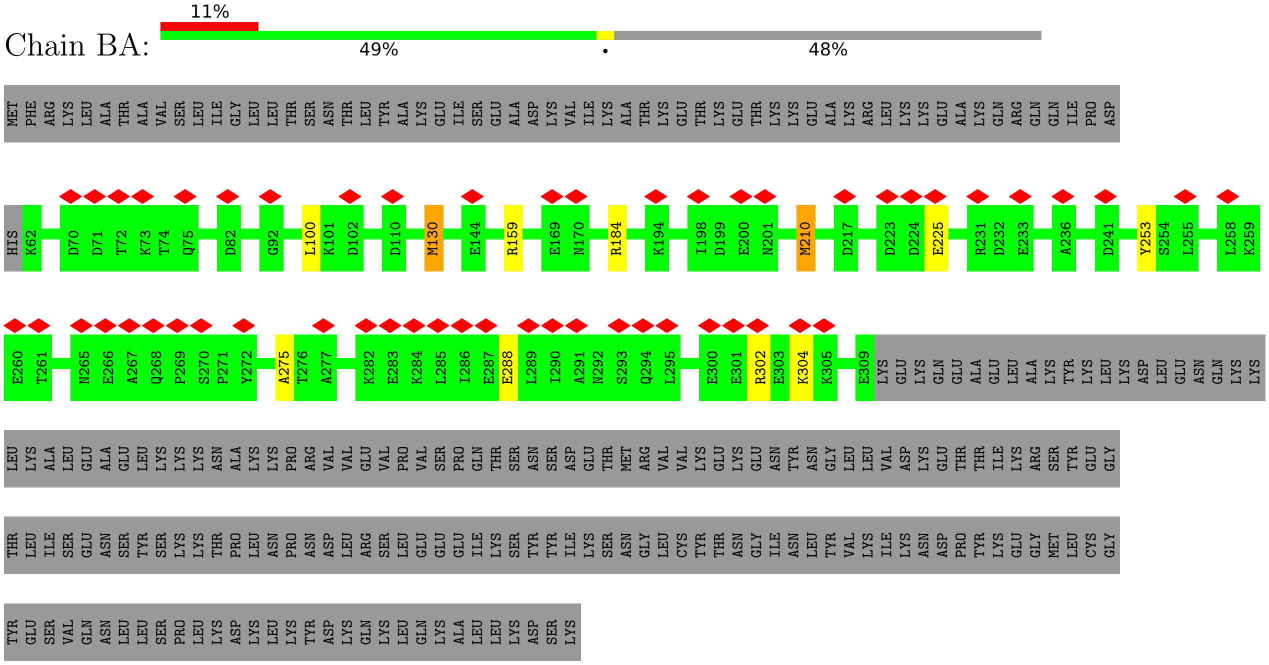
Chain AD:



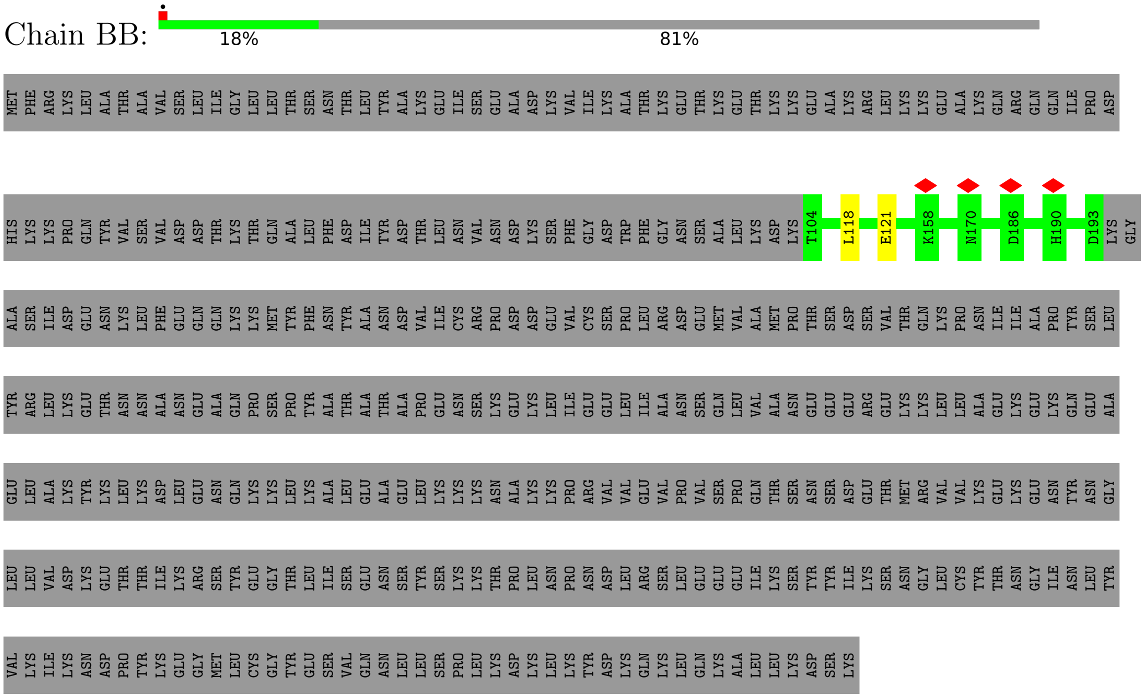
- Molecule 1: Type IV secretion system apparatus protein Cag3



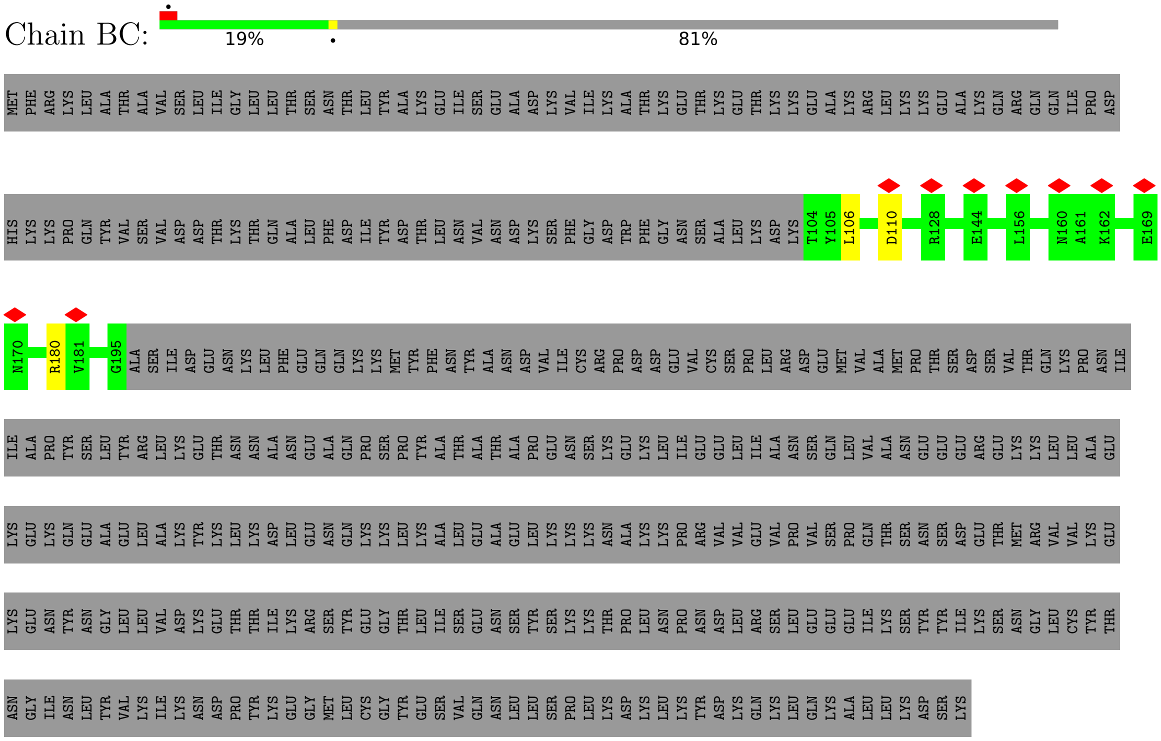
- Molecule 1: Type IV secretion system apparatus protein Cag3



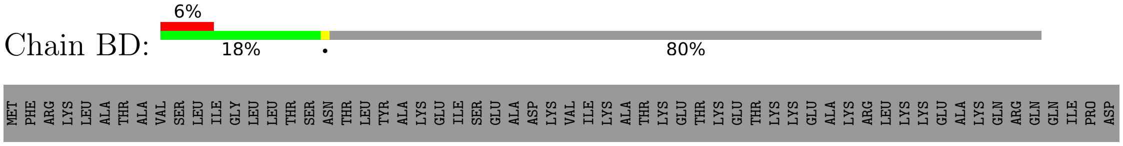
- Molecule 1: Type IV secretion system apparatus protein Cag3

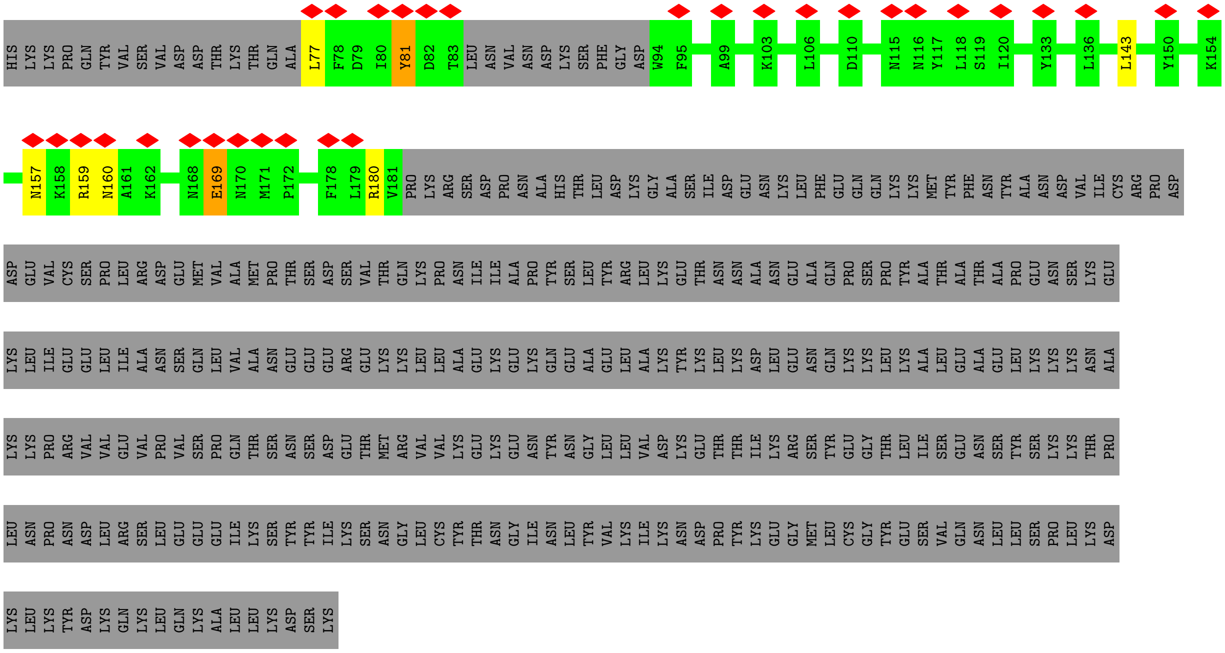


- Molecule 1: Type IV secretion system apparatus protein Cag3



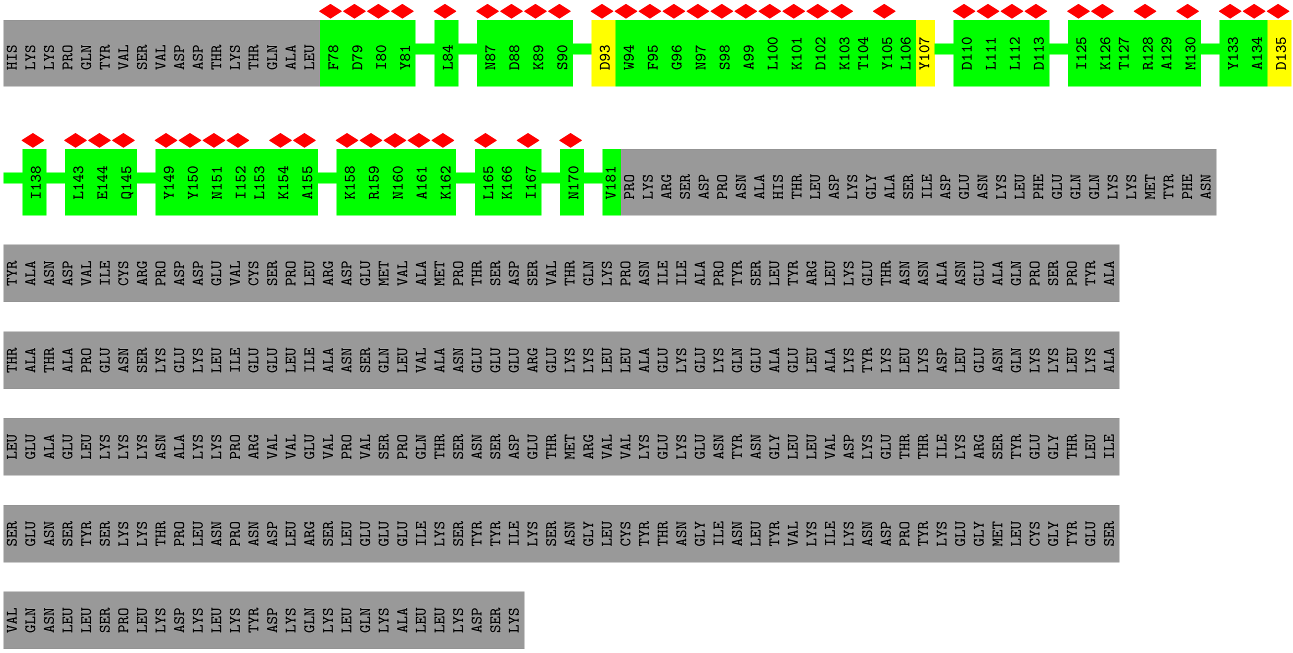
- Molecule 1: Type IV secretion system apparatus protein Cag3





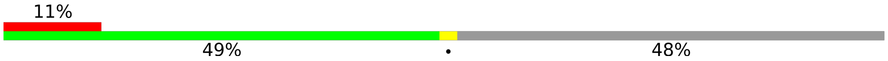
- Molecule 1: Type IV secretion system apparatus protein Cag3

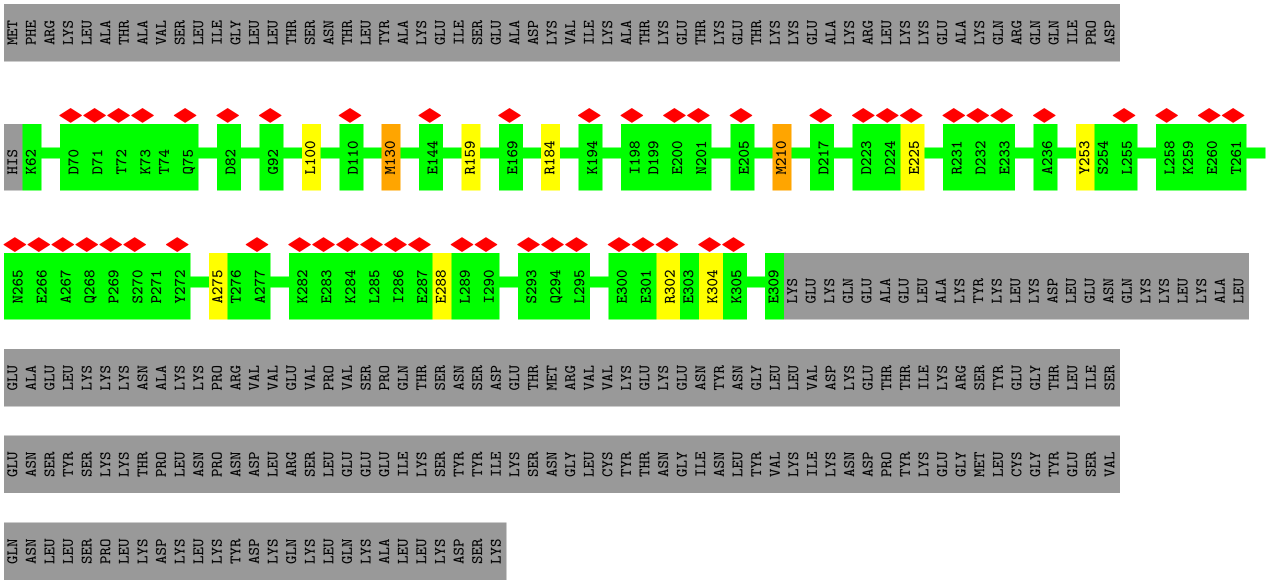
Chain BE:



- Molecule 1: Type IV secretion system apparatus protein Cag3

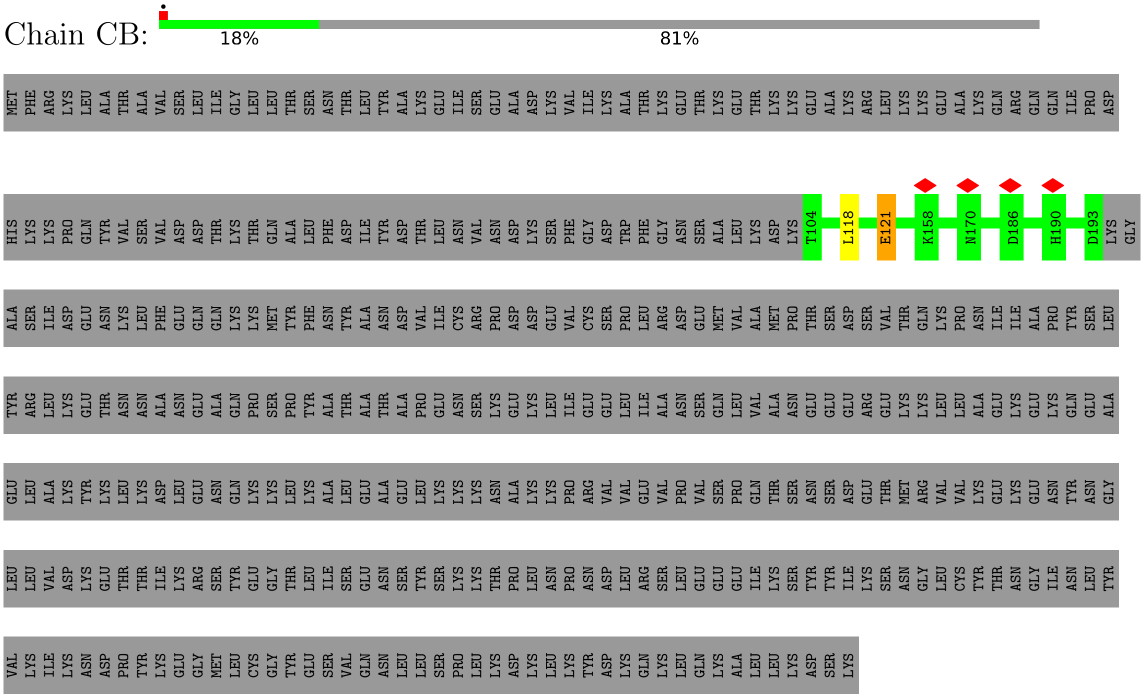
Chain CA:





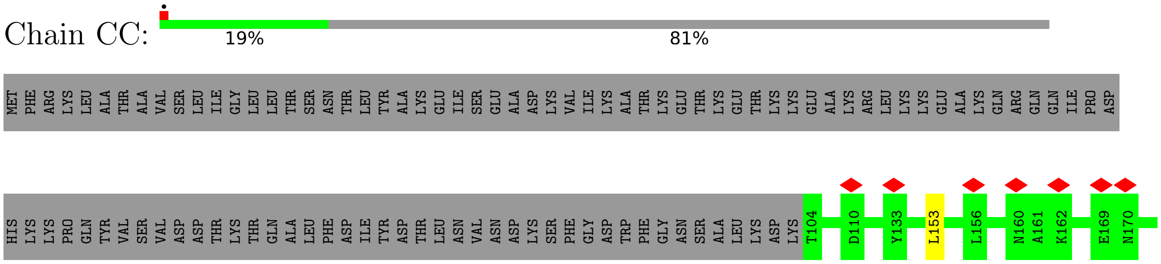
- Molecule 1: Type IV secretion system apparatus protein Cag3

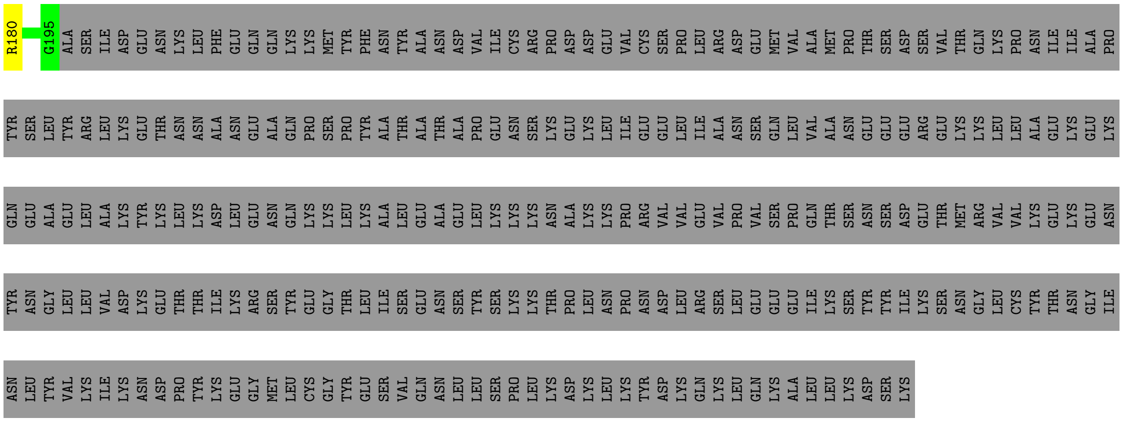
Chain CB:



- Molecule 1: Type IV secretion system apparatus protein Cag3

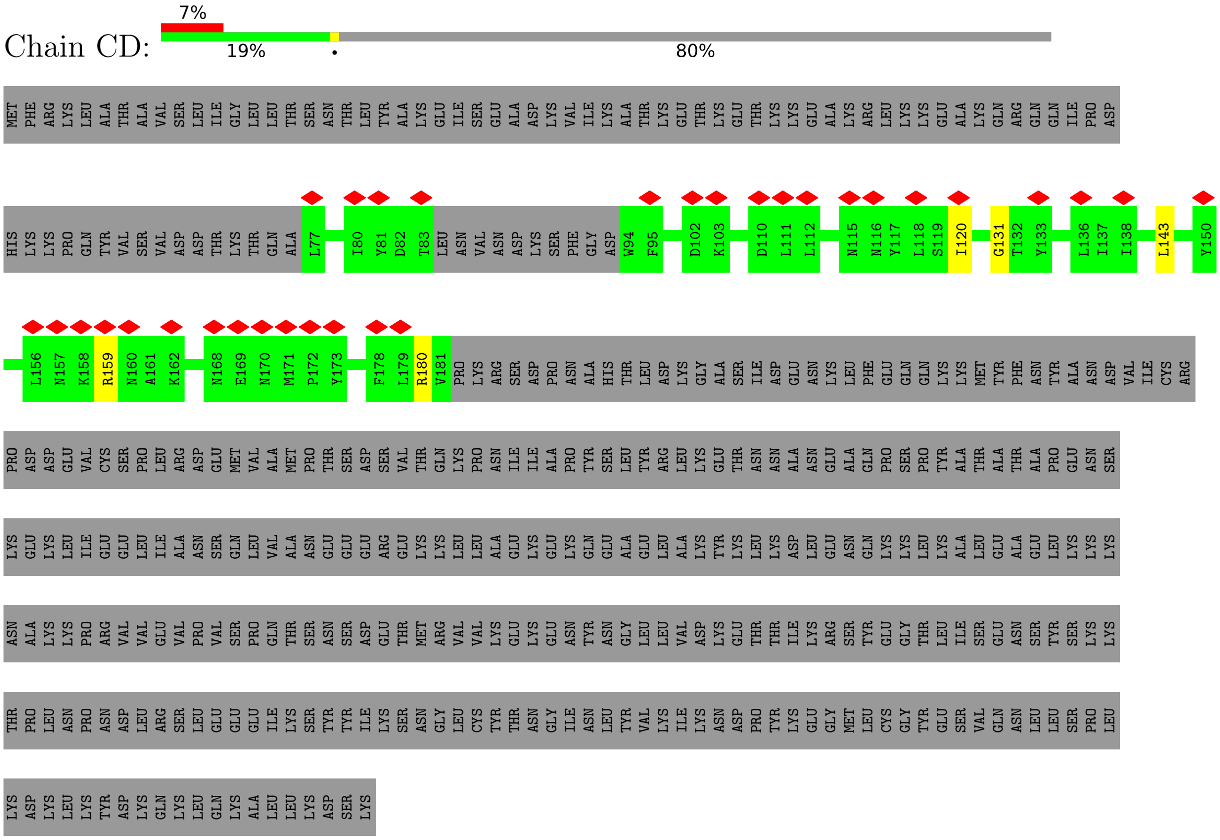
Chain CC:





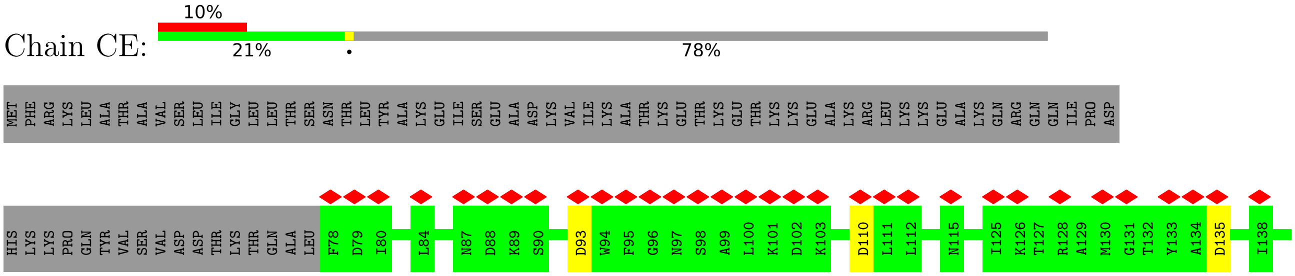
- Molecule 1: Type IV secretion system apparatus protein Cag3

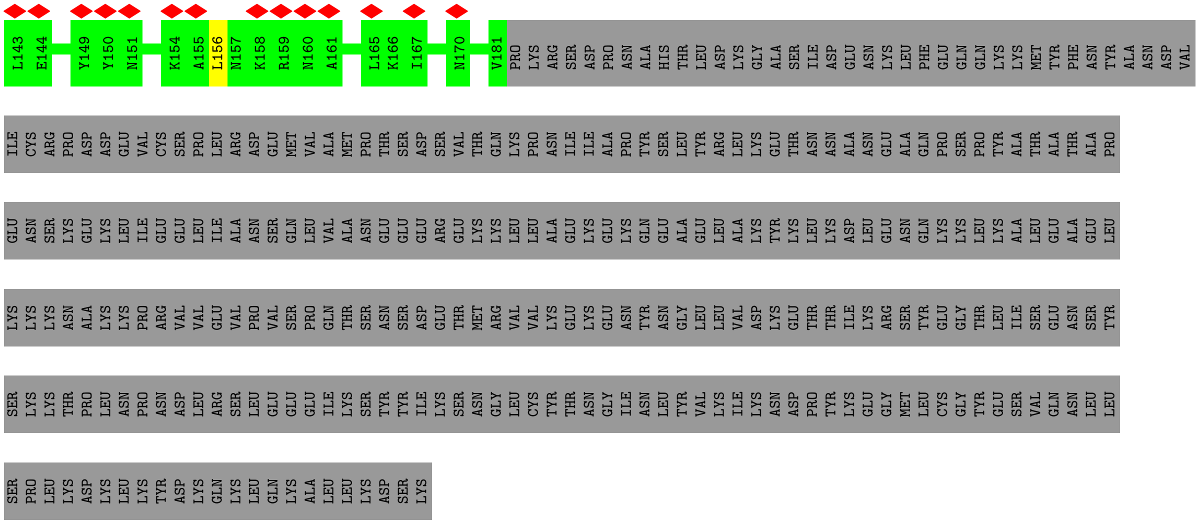
Chain CD:



- Molecule 1: Type IV secretion system apparatus protein Cag3

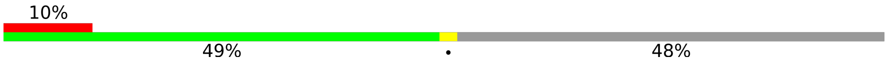
Chain CE:

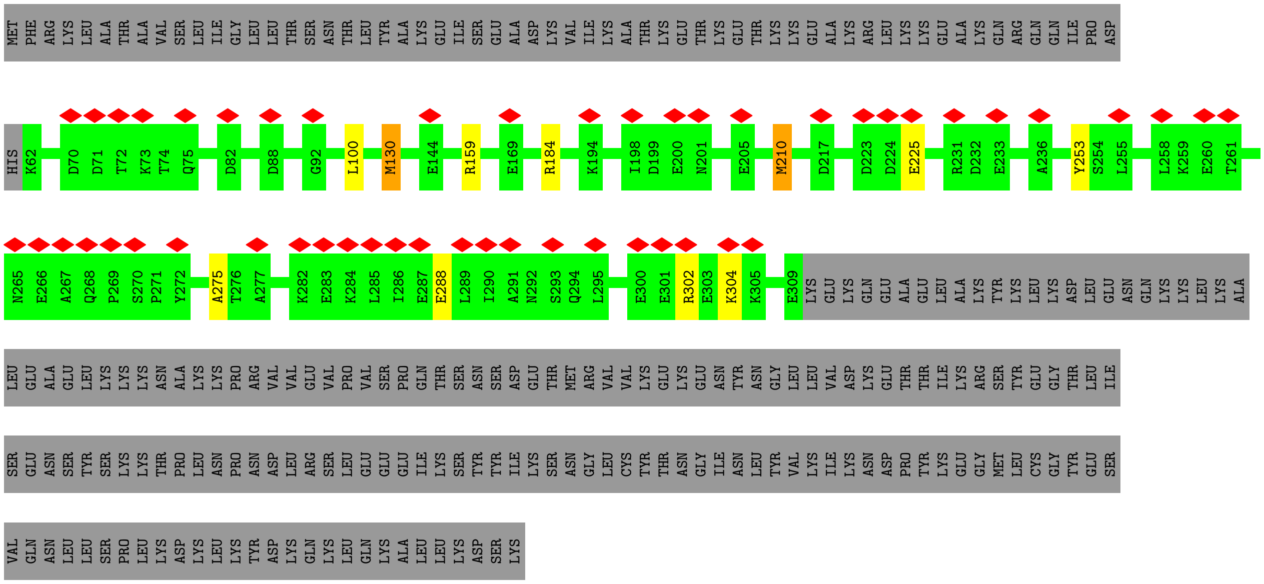




- Molecule 1: Type IV secretion system apparatus protein Cag3

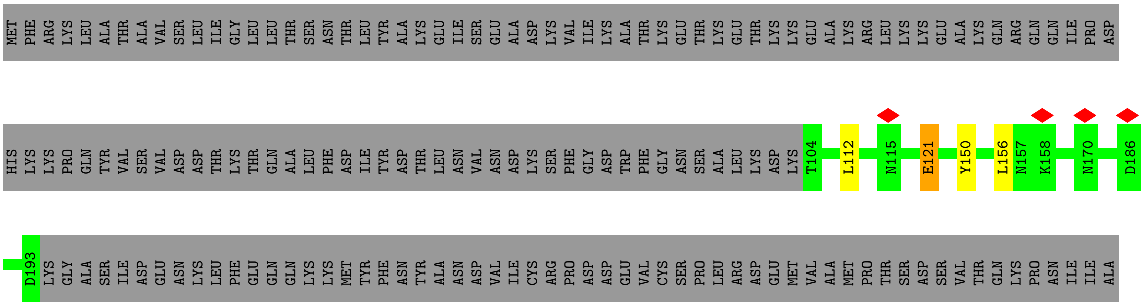
Chain DA:

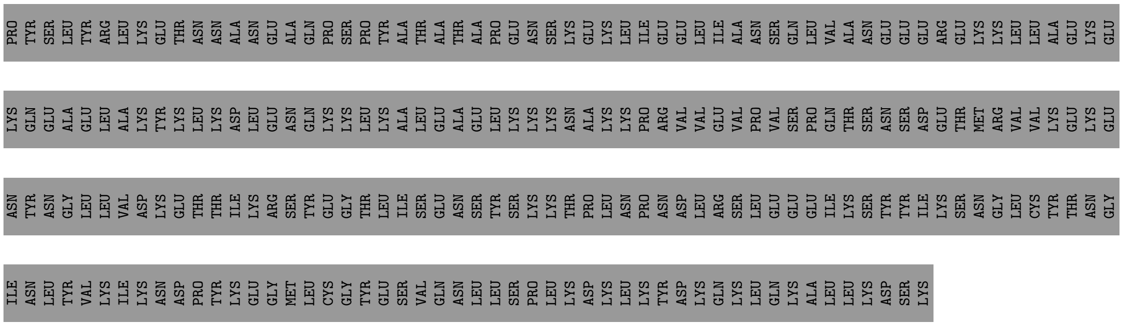




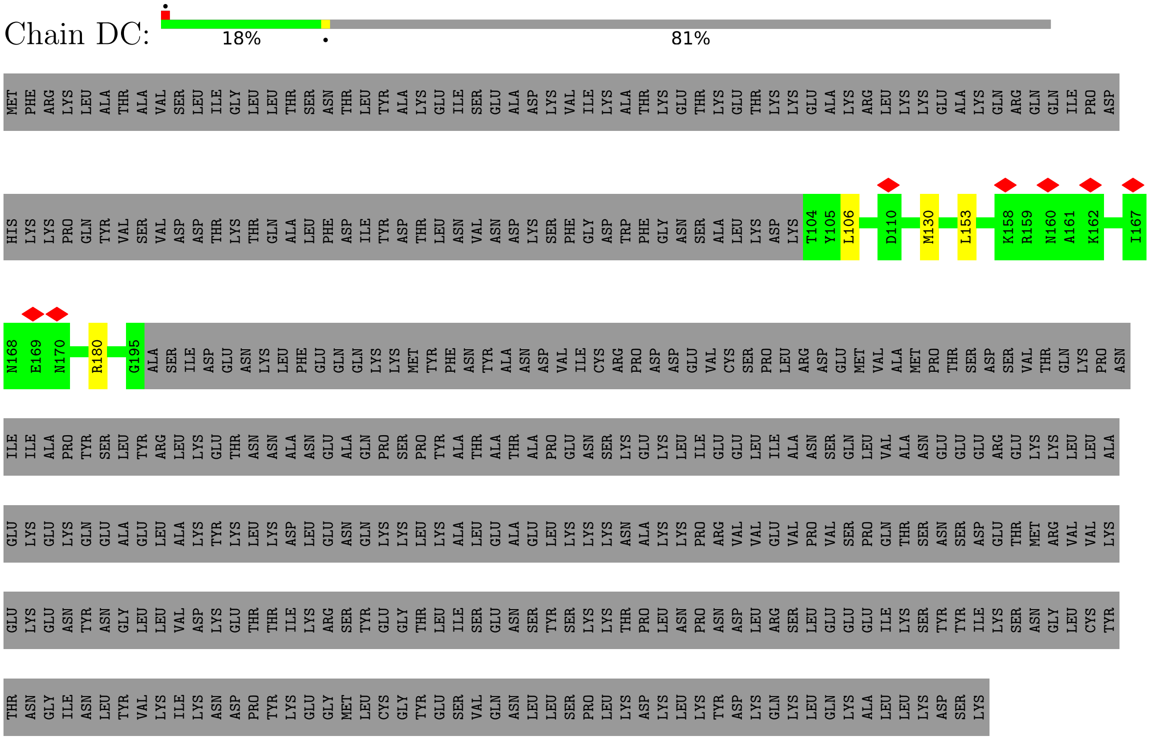
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain DB:

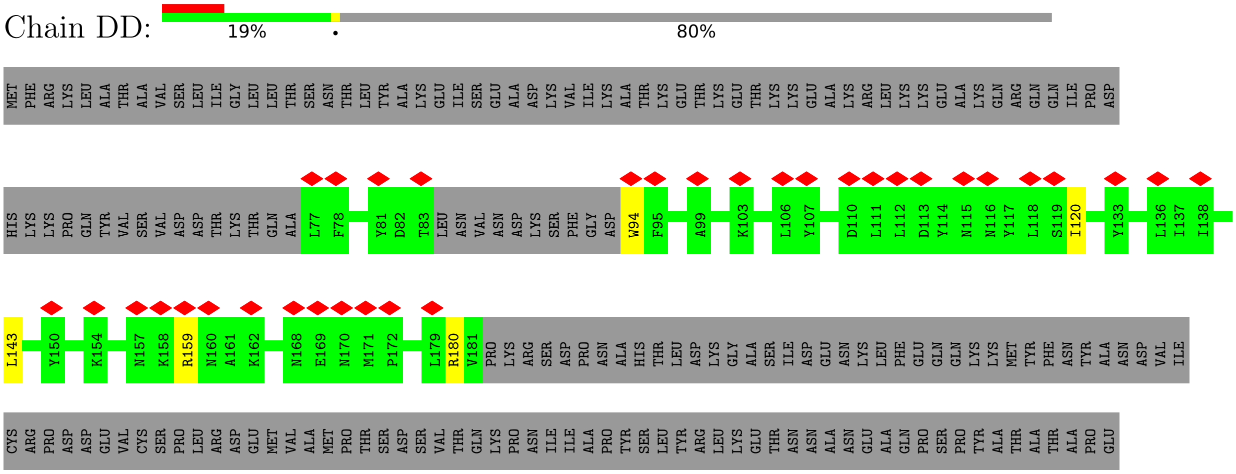


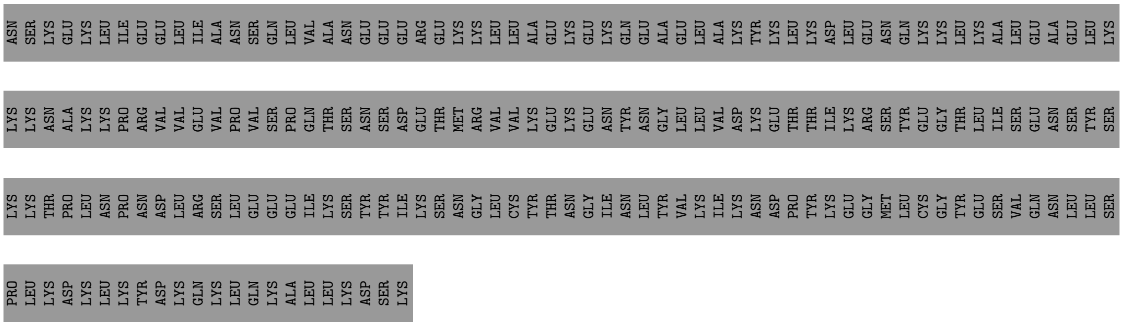


- Molecule 1: Type IV secretion system apparatus protein Cag3

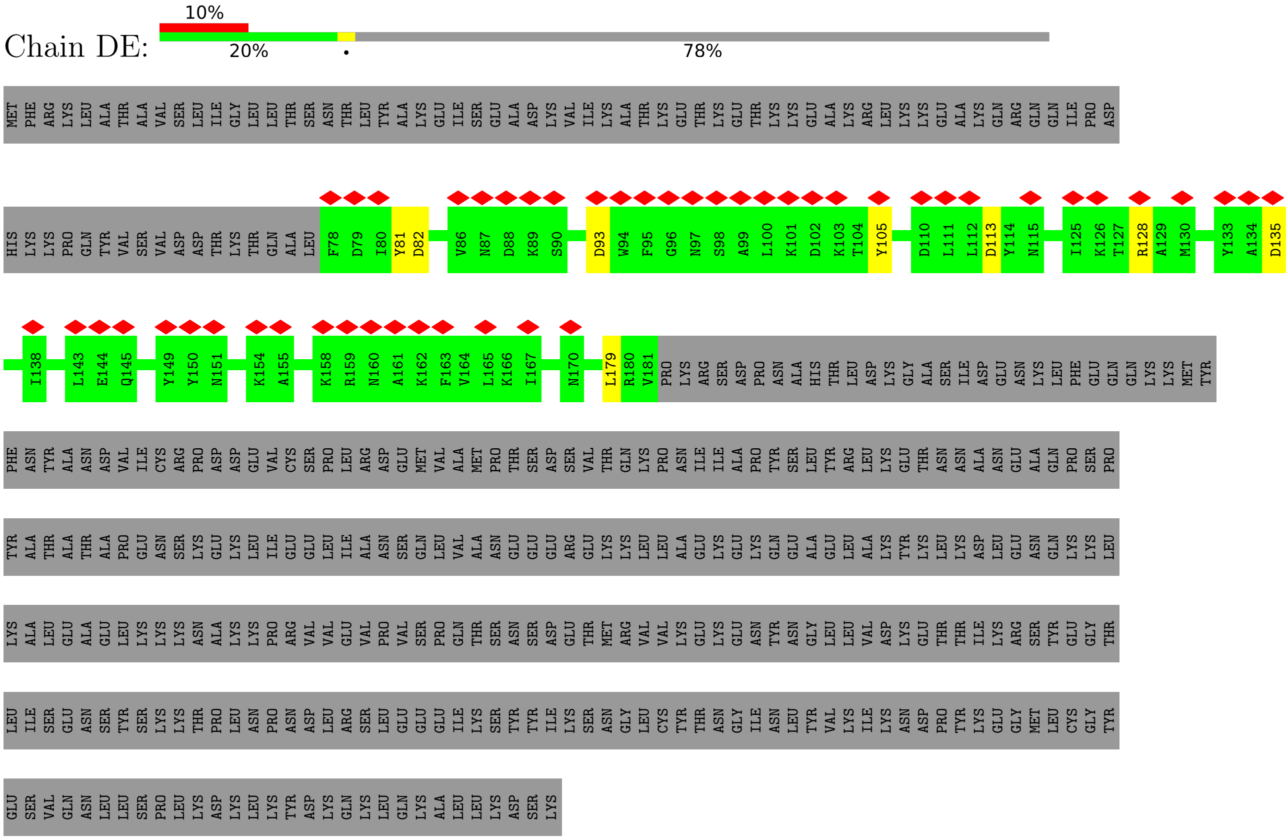


- Molecule 1: Type IV secretion system apparatus protein Cag3

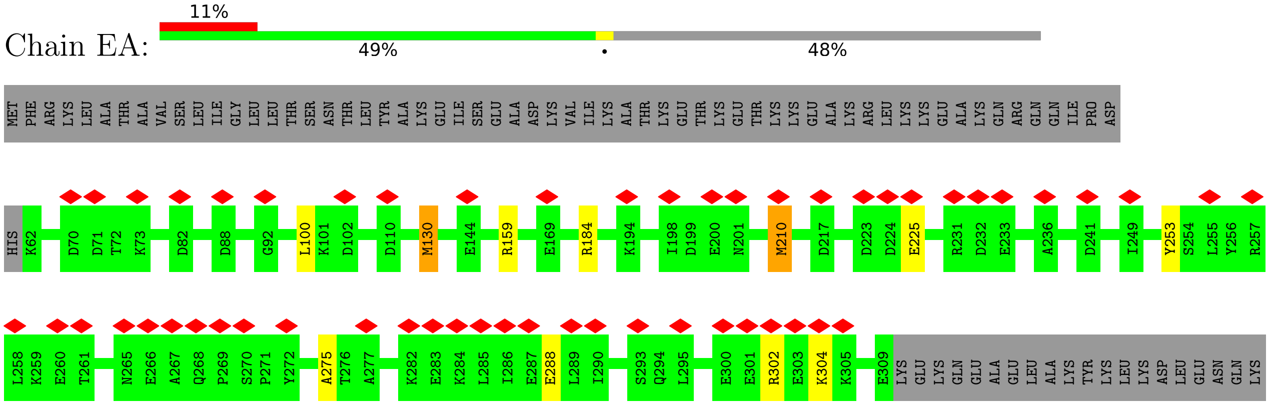


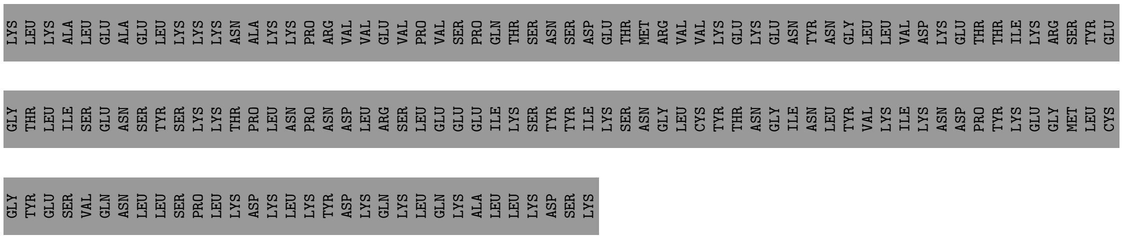


- Molecule 1: Type IV secretion system apparatus protein Cag3

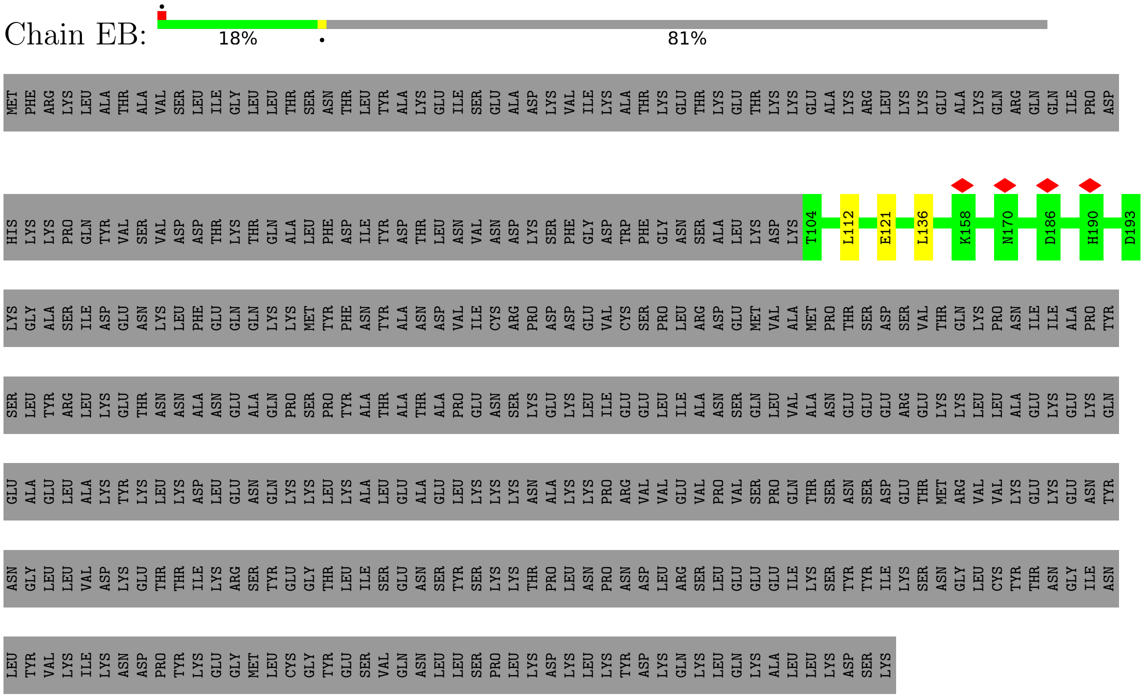


- Molecule 1: Type IV secretion system apparatus protein Cag3

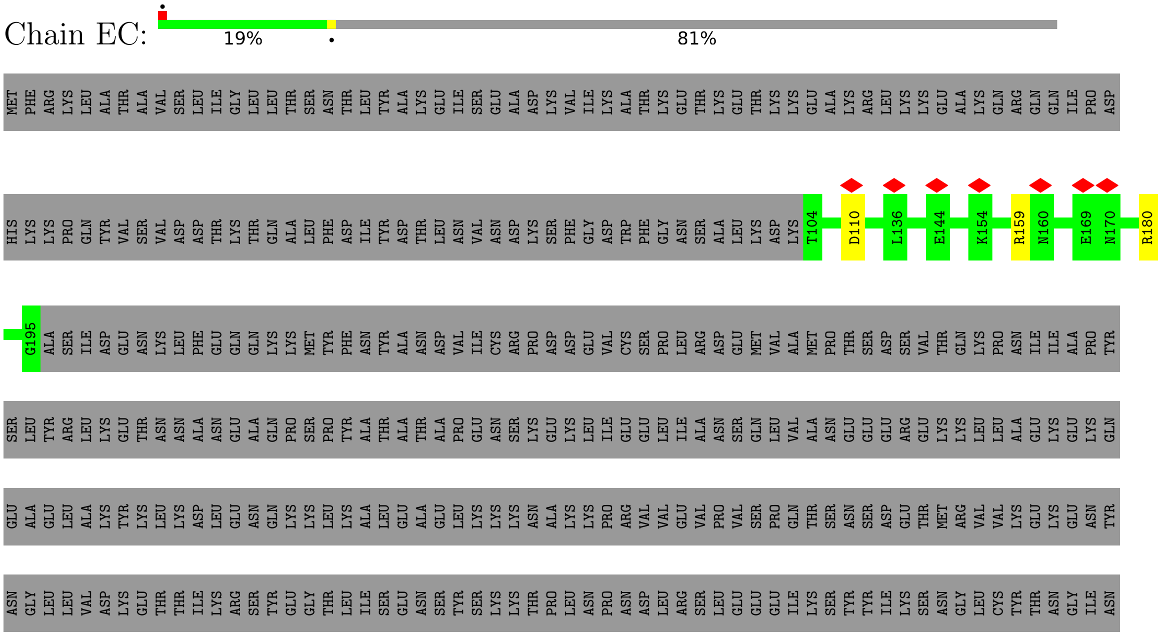




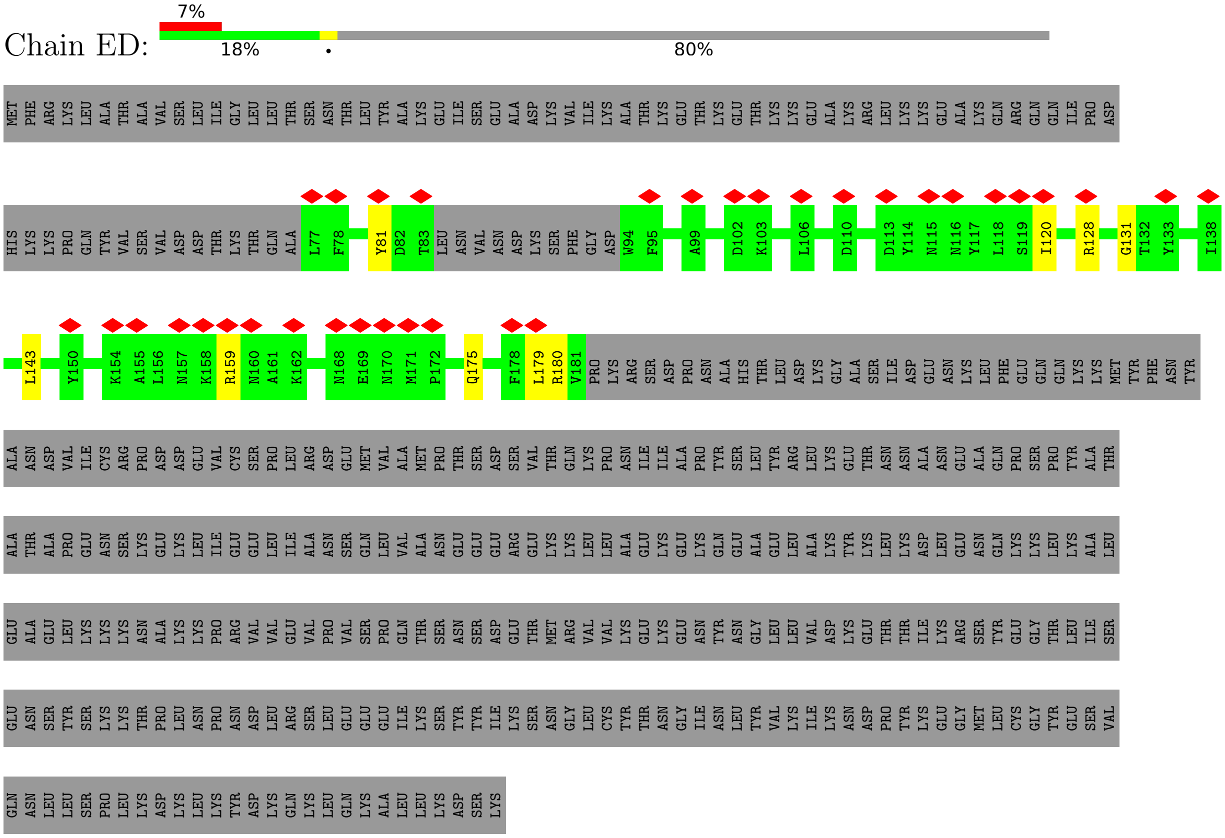
- Molecule 1: Type IV secretion system apparatus protein Cag3



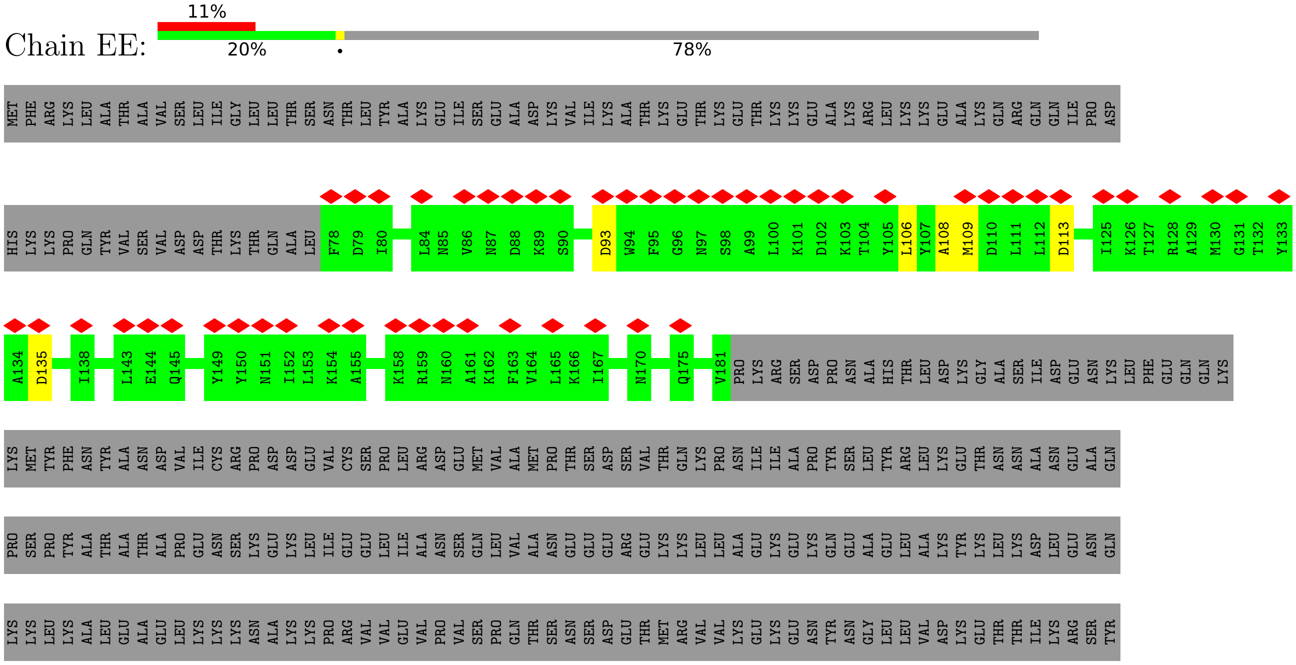
- Molecule 1: Type IV secretion system apparatus protein Cag3

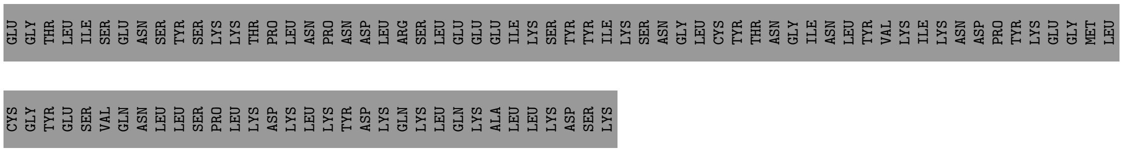


- Molecule 1: Type IV secretion system apparatus protein Cag3

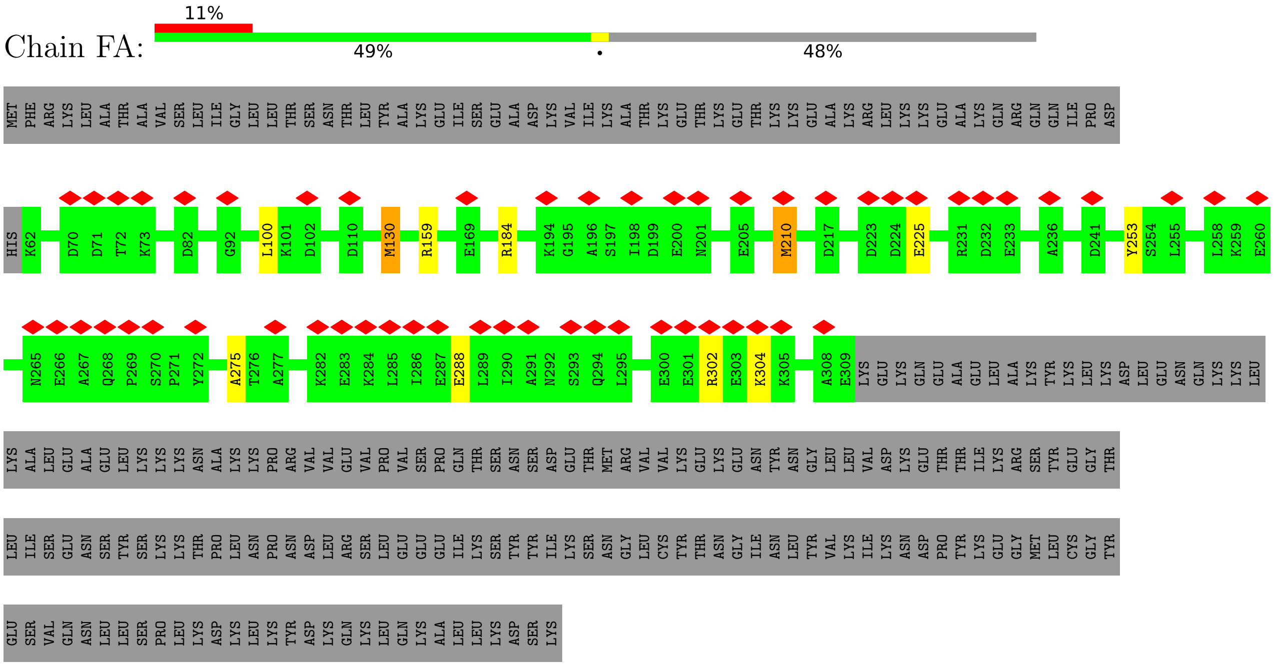


- Molecule 1: Type IV secretion system apparatus protein Cag3

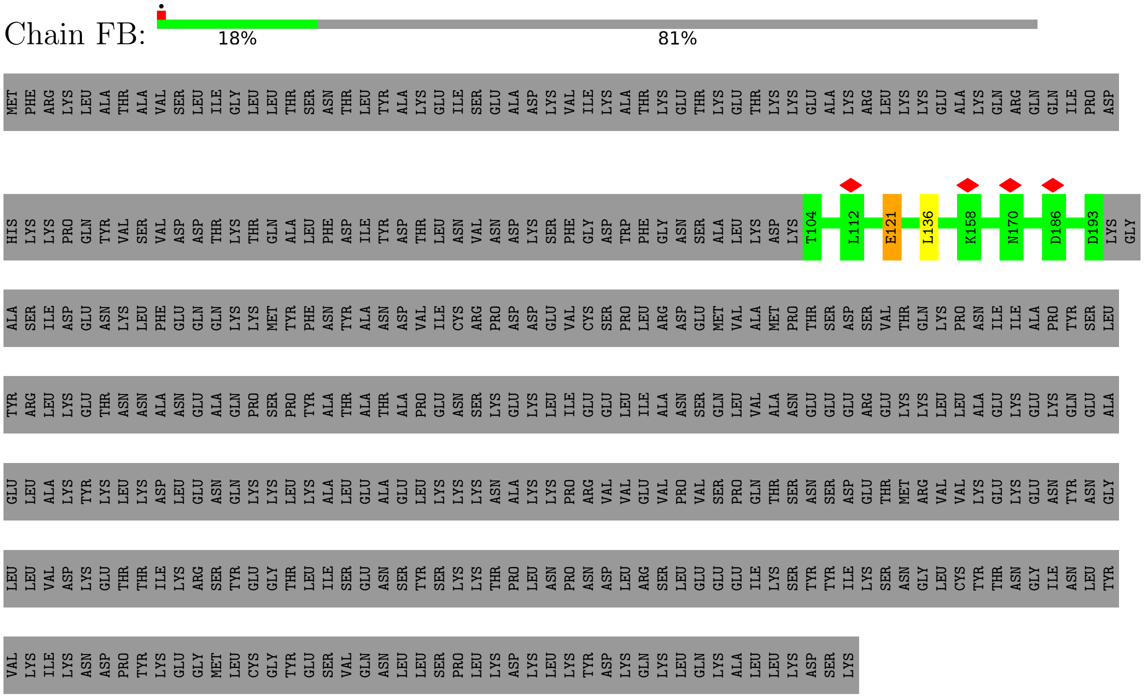




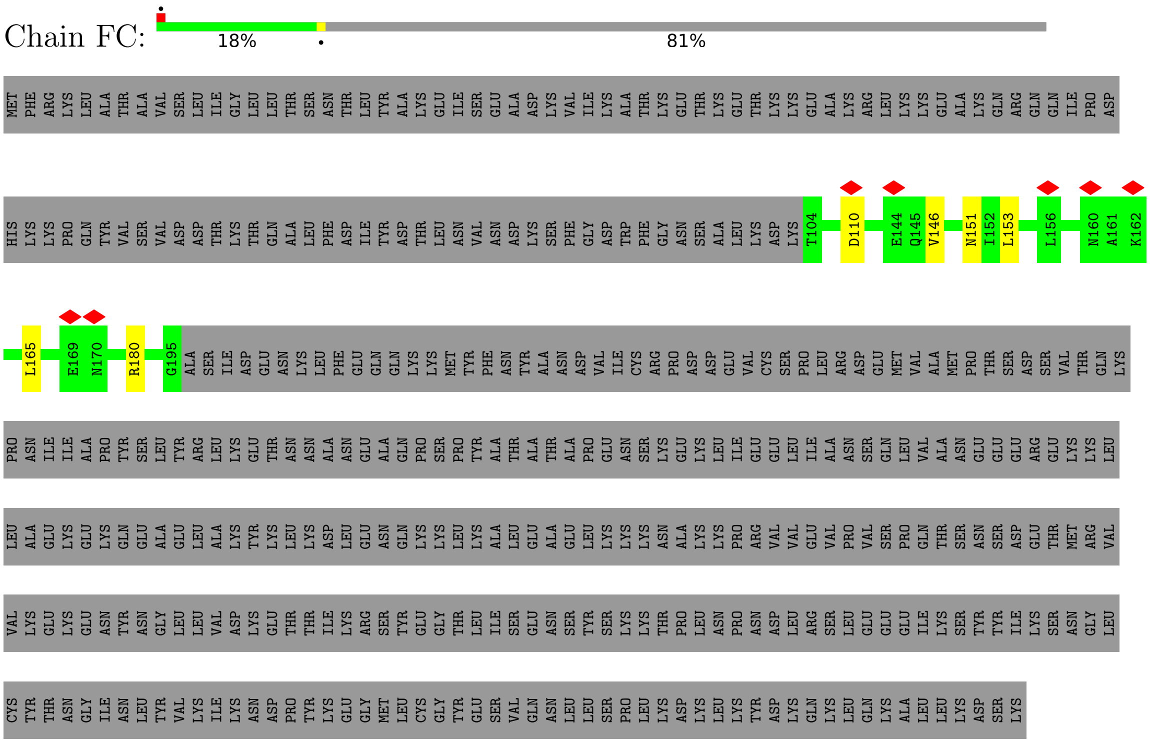
- Molecule 1: Type IV secretion system apparatus protein Cag3



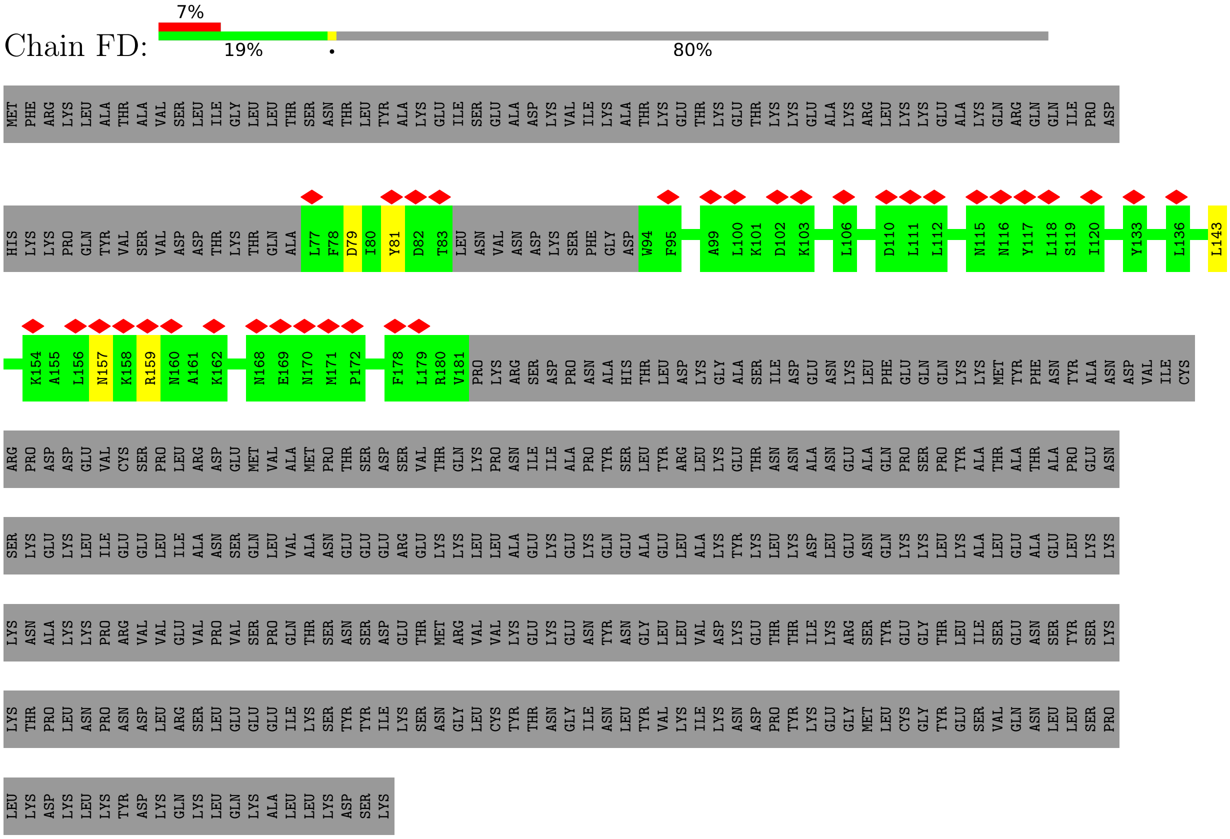
- Molecule 1: Type IV secretion system apparatus protein Cag3



- Molecule 1: Type IV secretion system apparatus protein Cag3

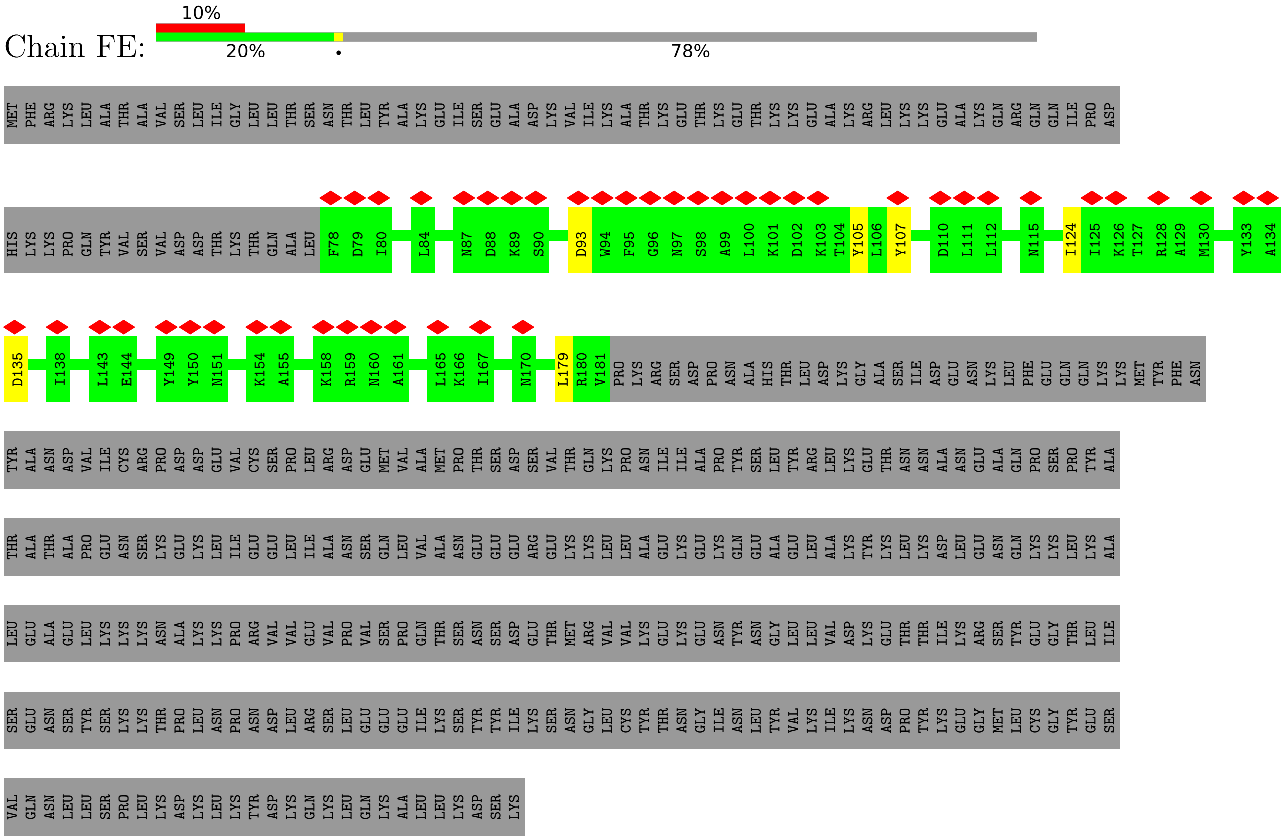


- Molecule 1: Type IV secretion system apparatus protein Cag3



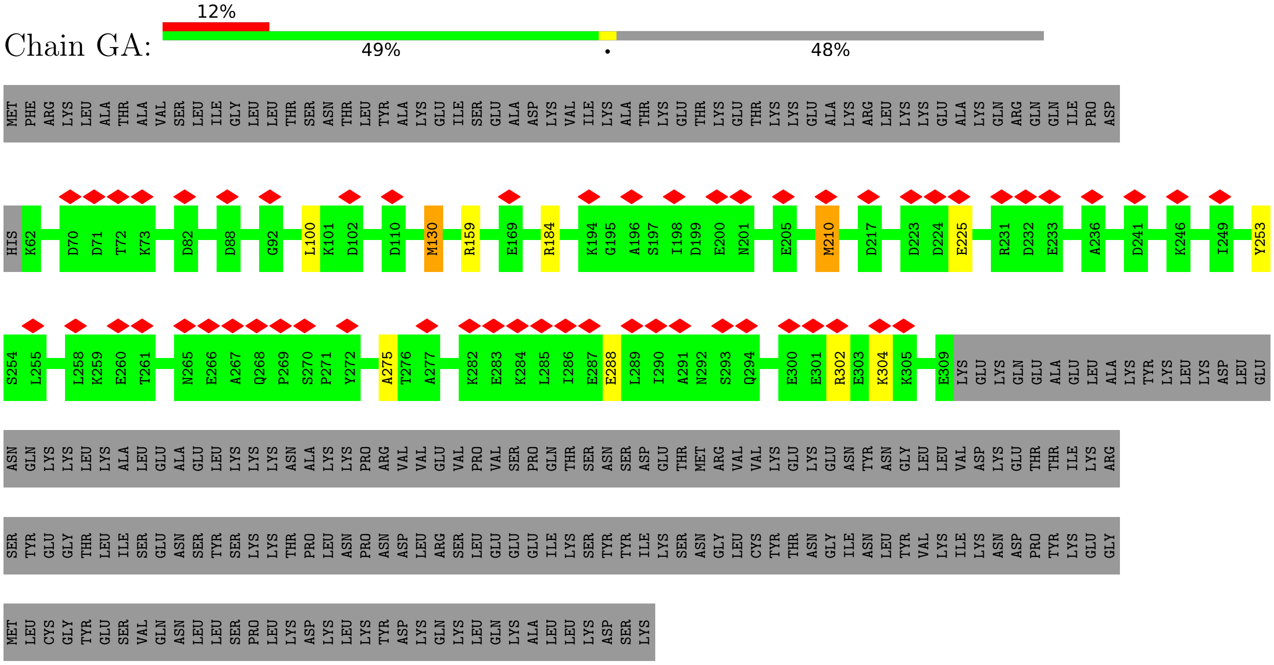
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain FE:



- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain GA:

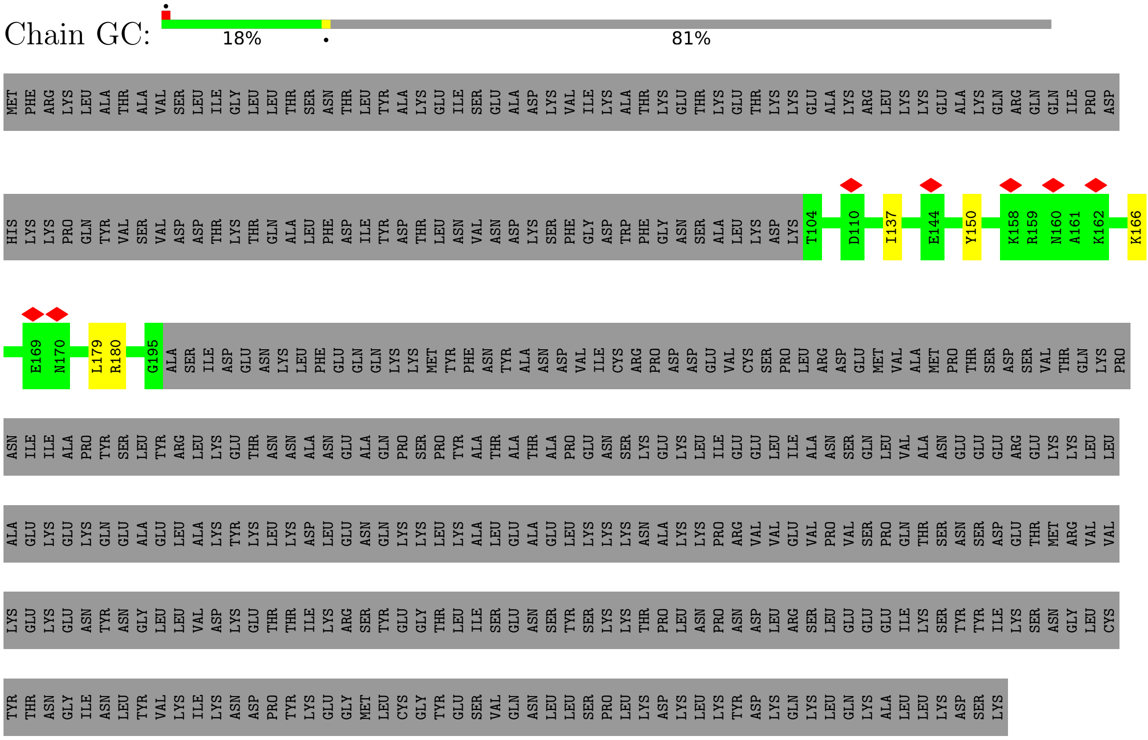


- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain GB:

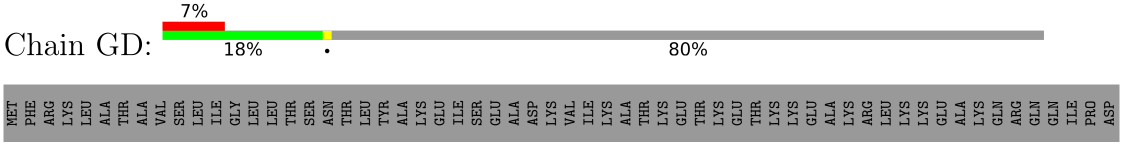
- Molecule 1: Type IV secretion system apparatus protein Cag3

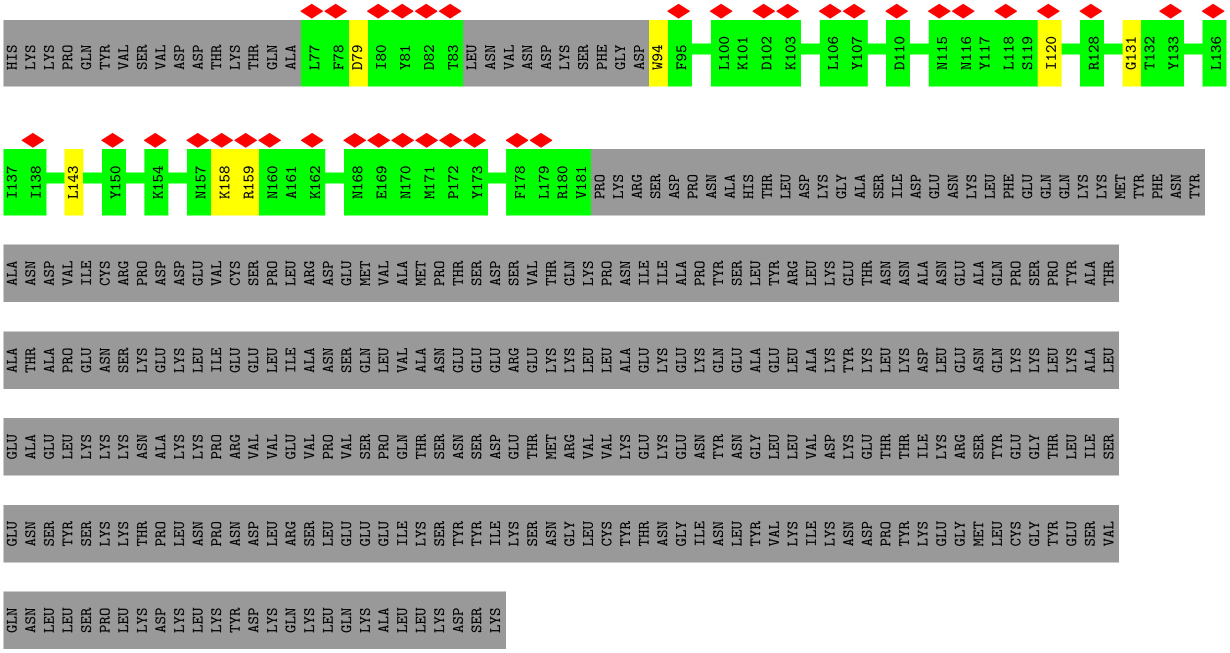
Chain GC:



- Molecule 1: Type IV secretion system apparatus protein Cag3

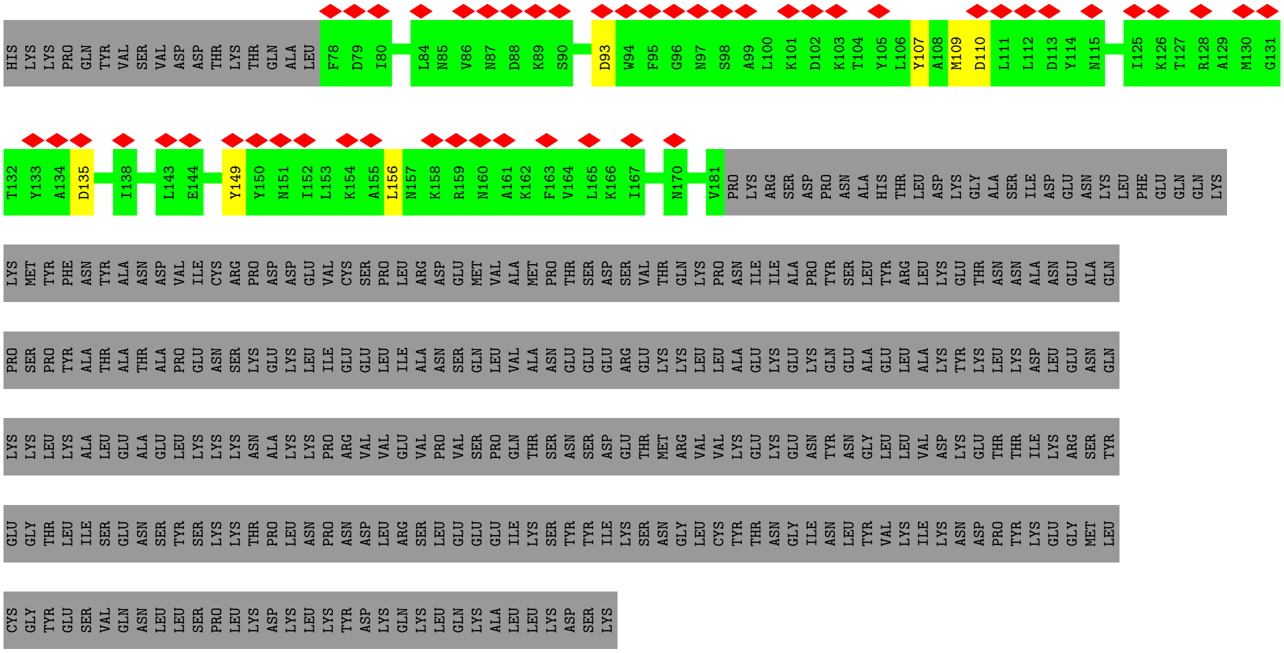
Chain GD:





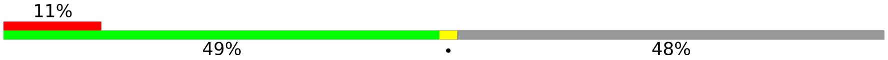
- Molecule 1: Type IV secretion system apparatus protein Cag3

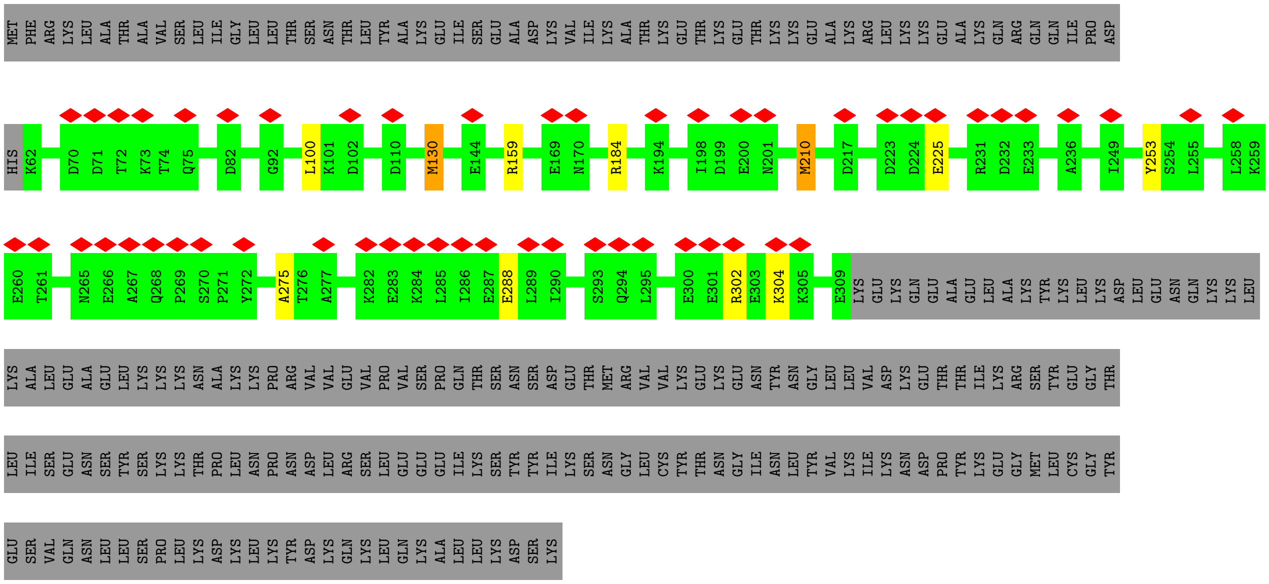
Chain GE:



- Molecule 1: Type IV secretion system apparatus protein Cag3

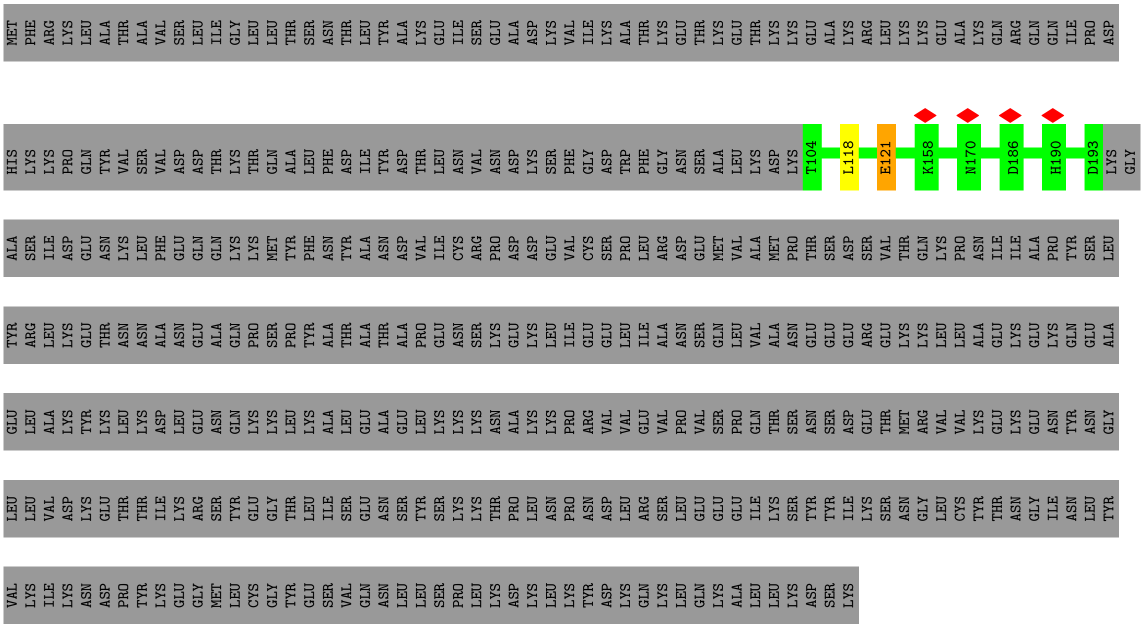
Chain HA:





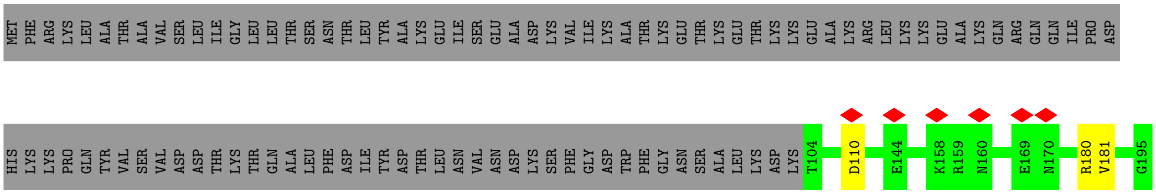
- Molecule 1: Type IV secretion system apparatus protein Cag3

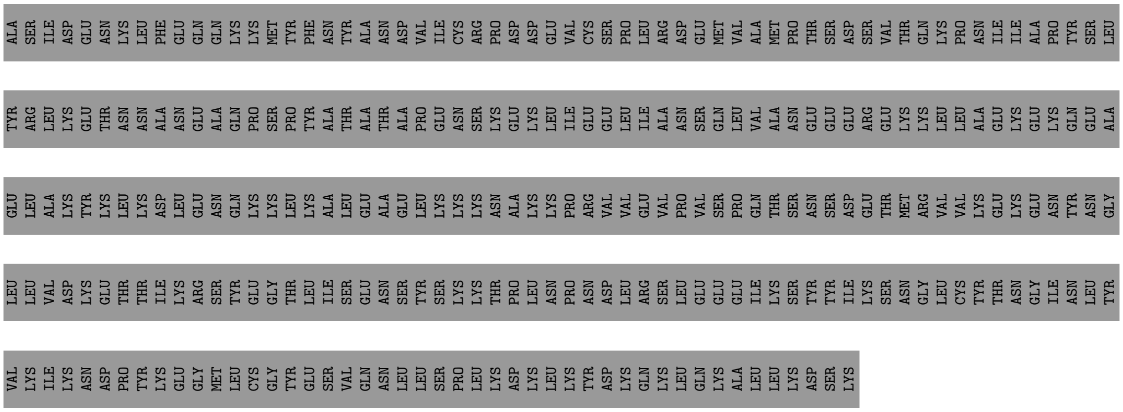
Chain HB:



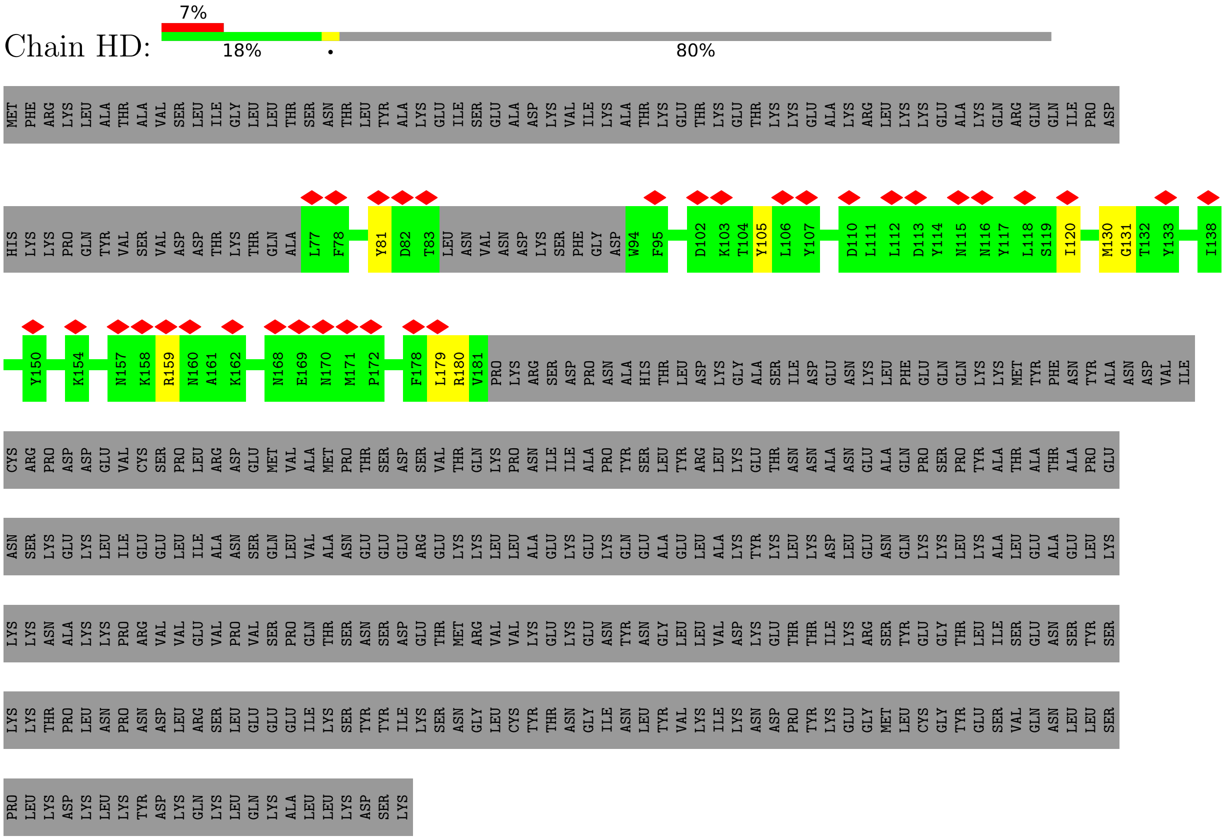
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain HC:

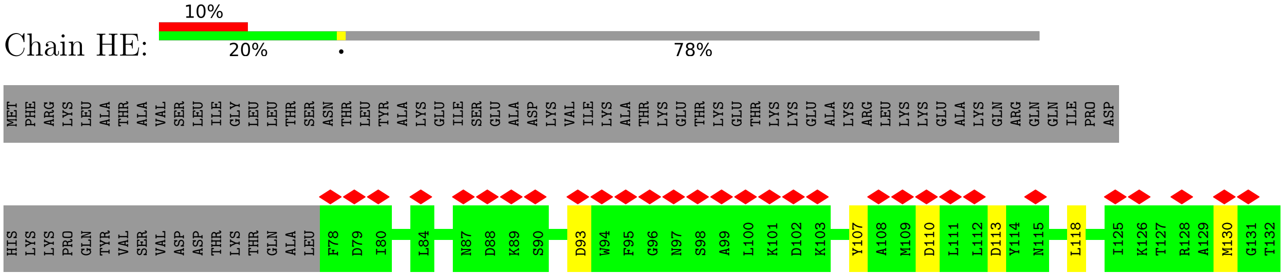


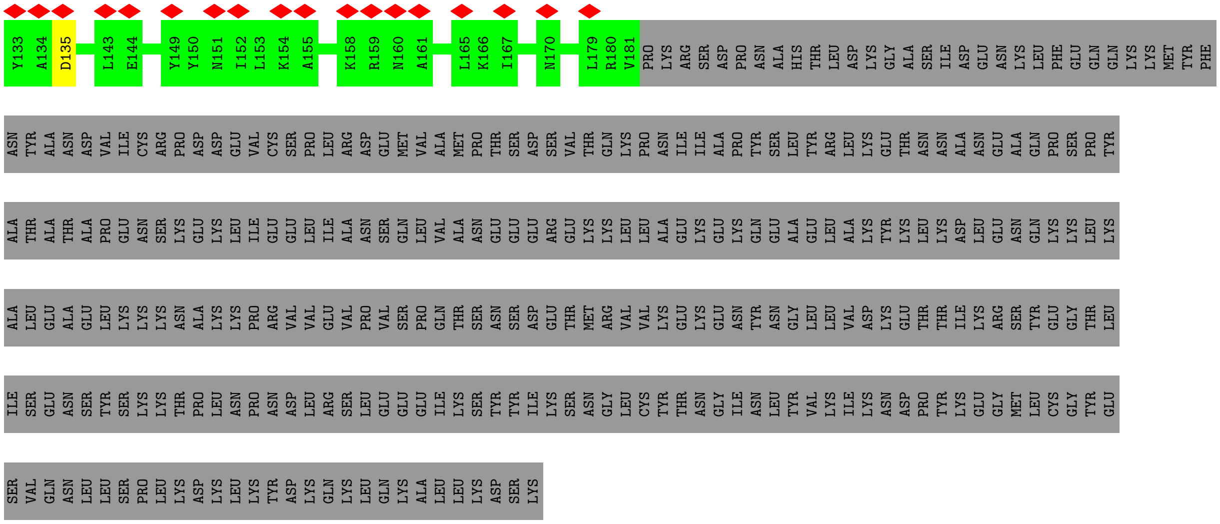


- Molecule 1: Type IV secretion system apparatus protein Cag3

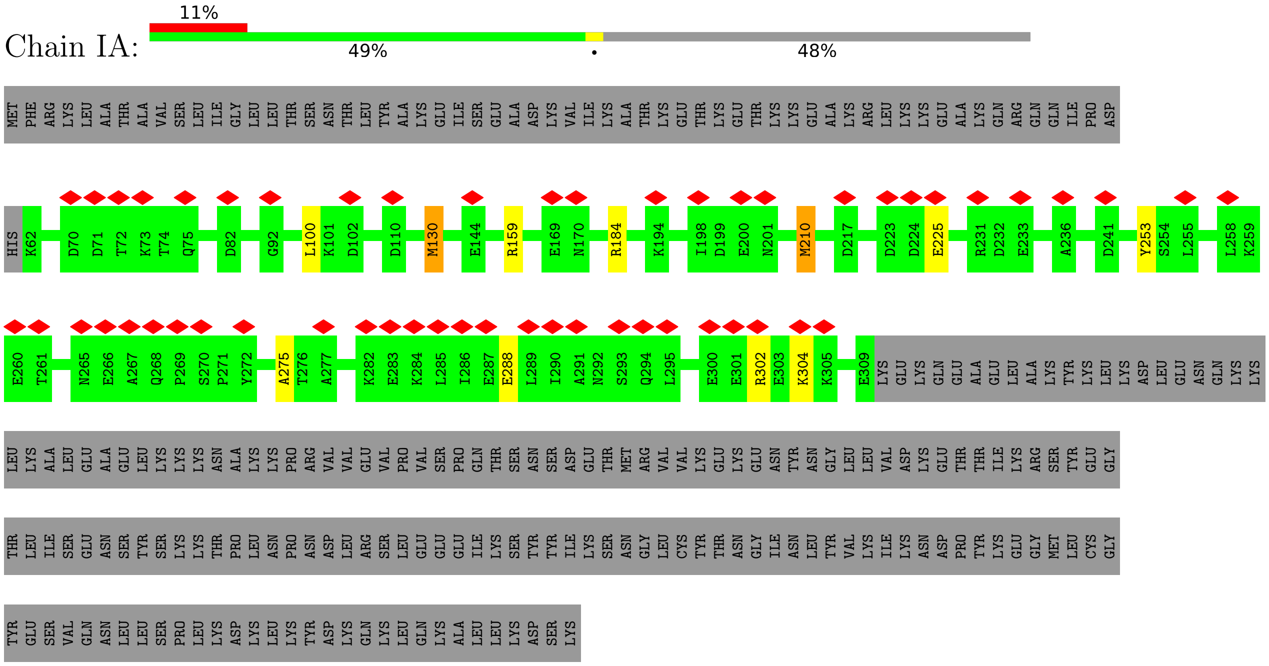


- Molecule 1: Type IV secretion system apparatus protein Cag3

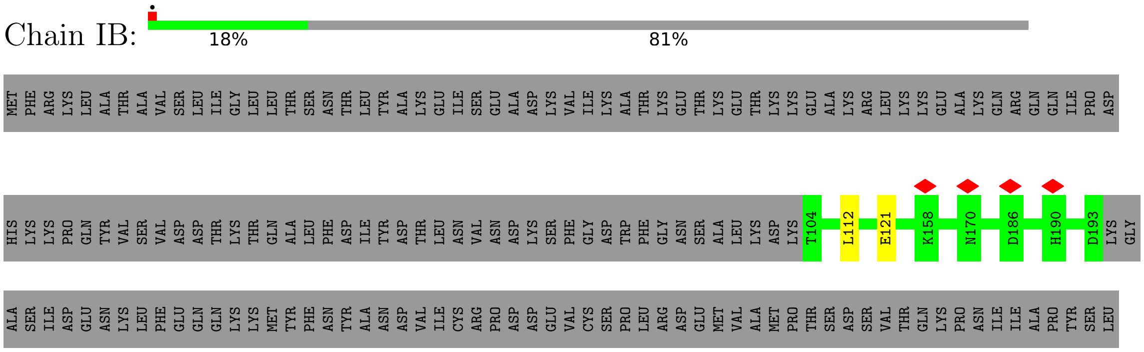


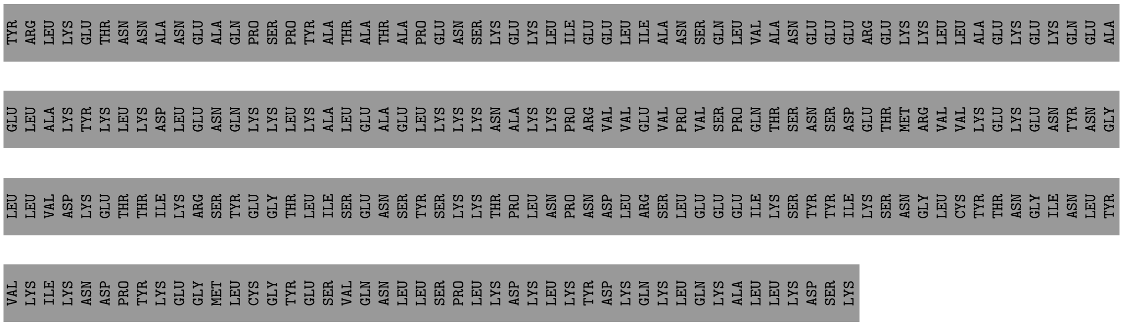


- Molecule 1: Type IV secretion system apparatus protein Cag3



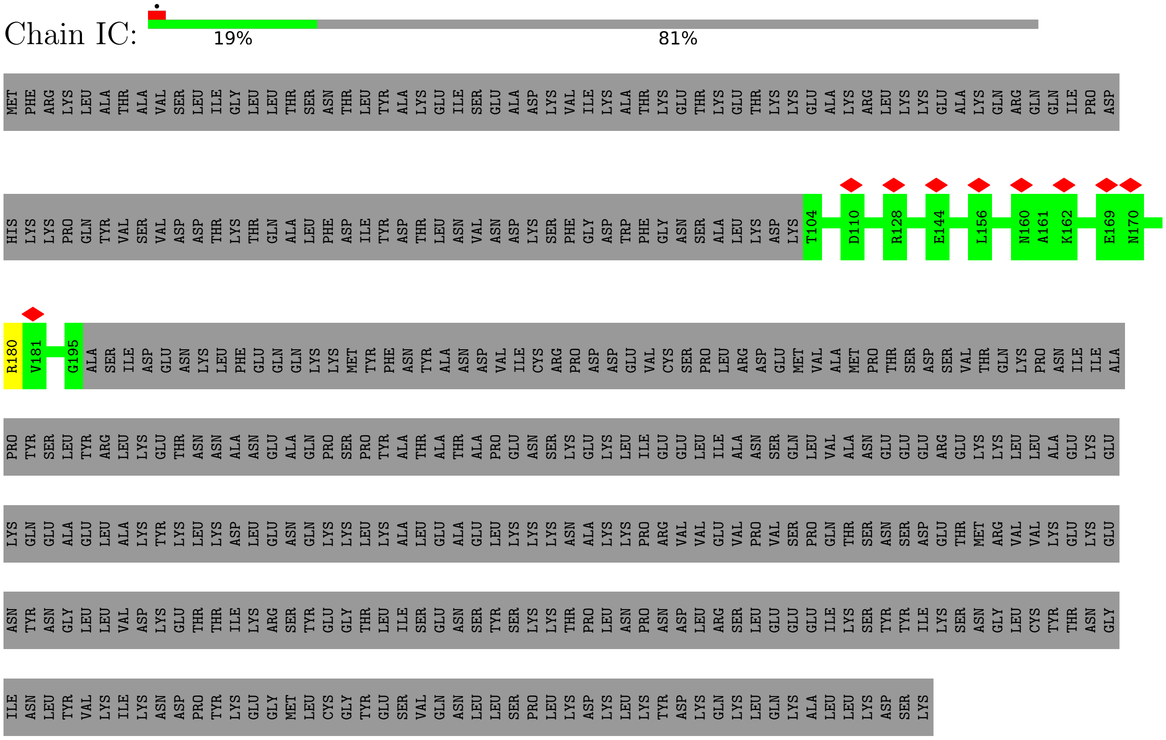
- Molecule 1: Type IV secretion system apparatus protein Cag3





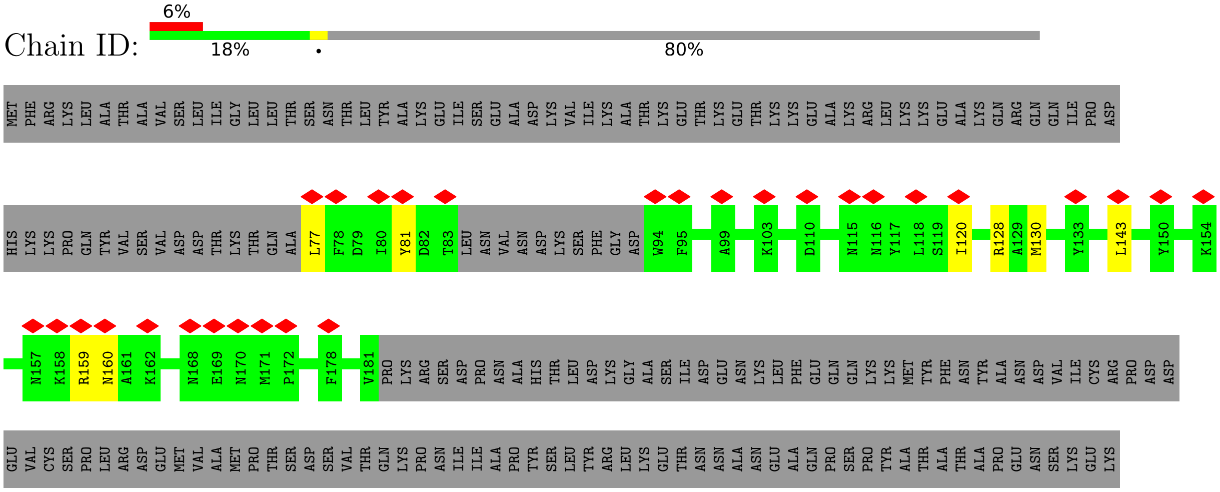
- Molecule 1: Type IV secretion system apparatus protein Cag3

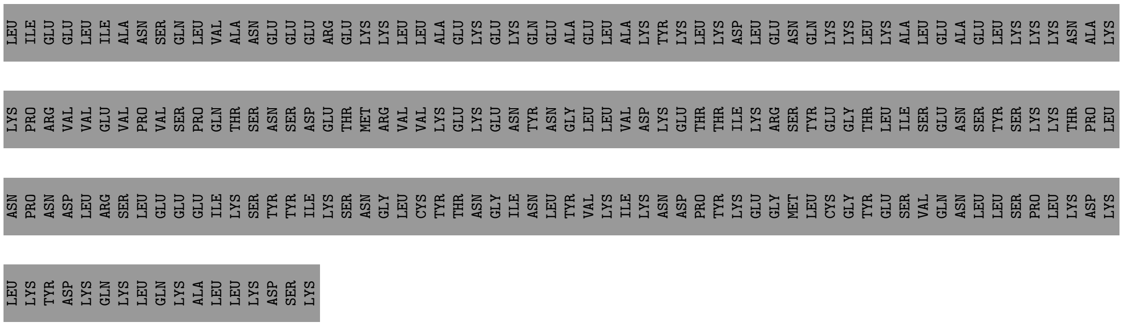
Chain IC:



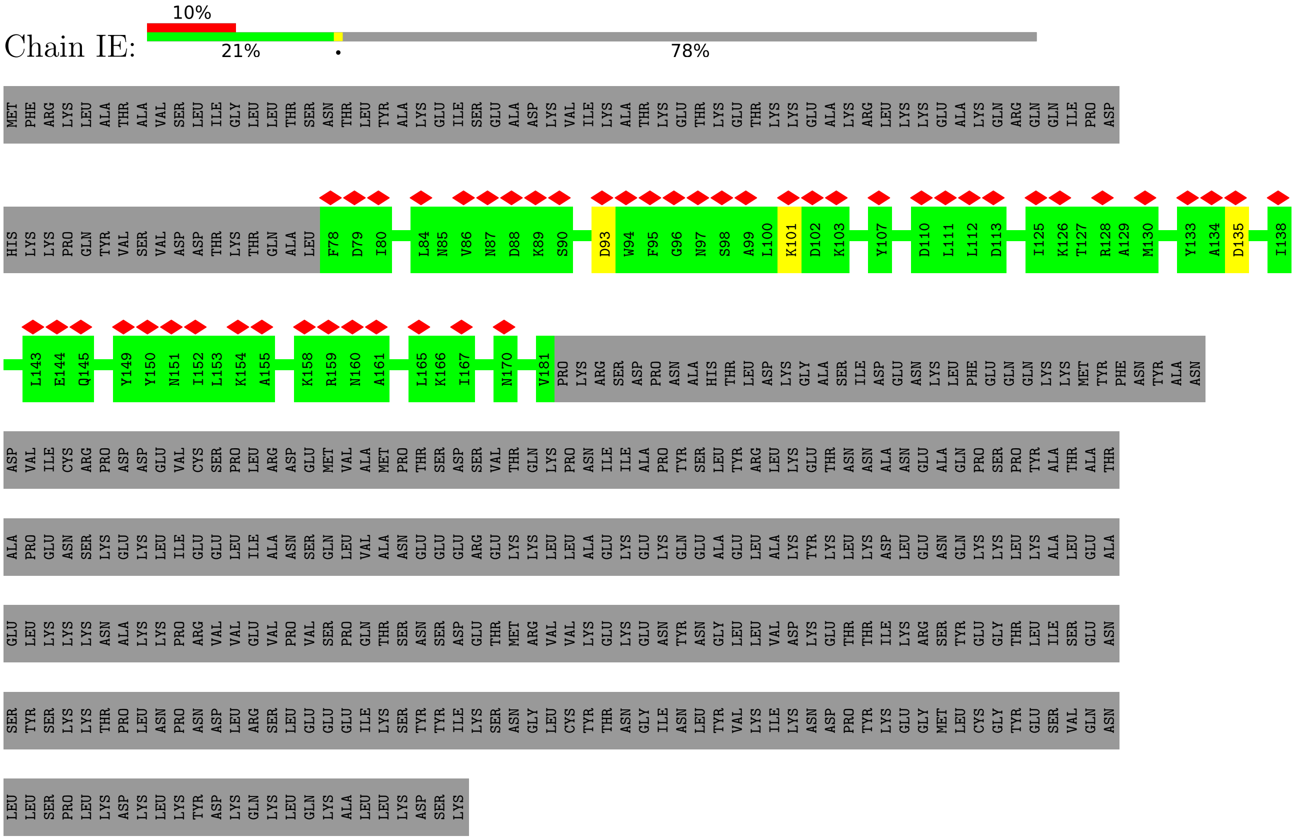
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain ID:

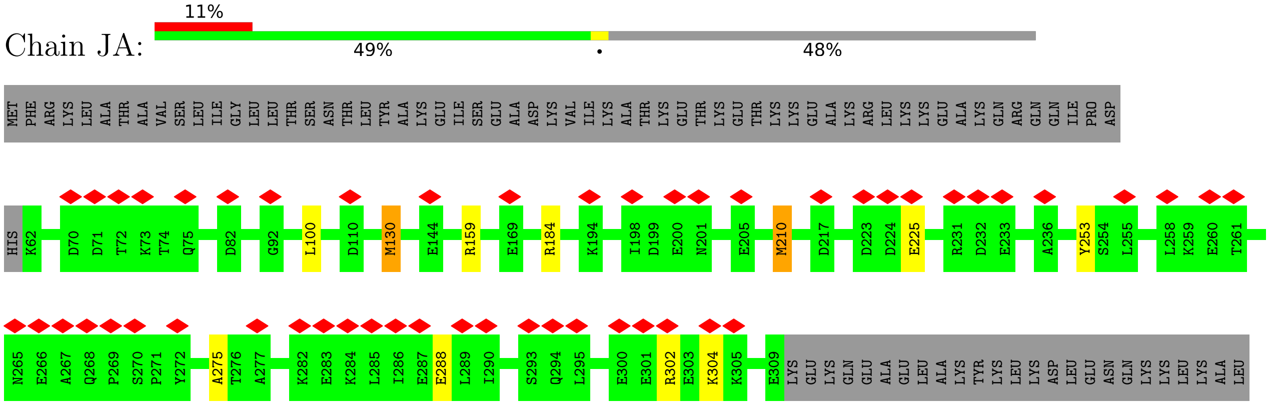


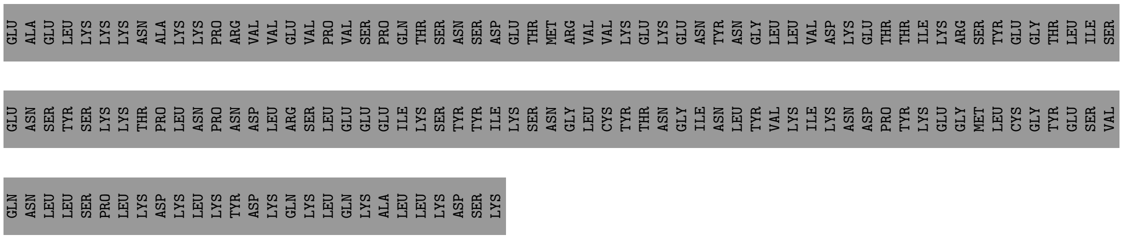


- Molecule 1: Type IV secretion system apparatus protein Cag3



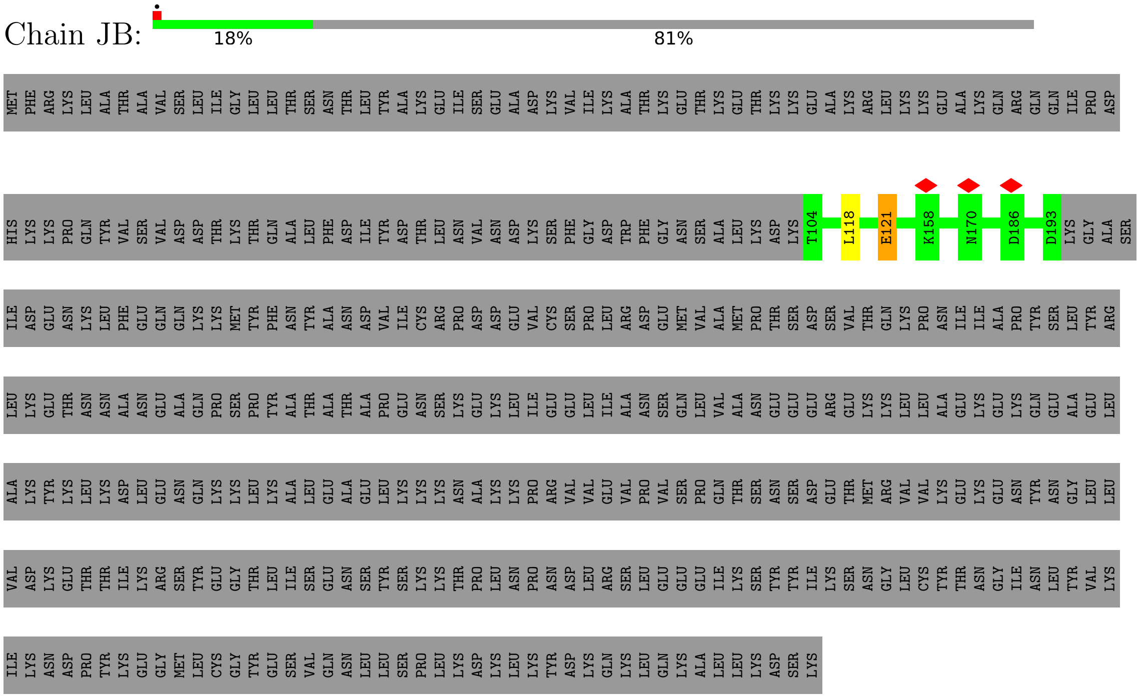
- Molecule 1: Type IV secretion system apparatus protein Cag3





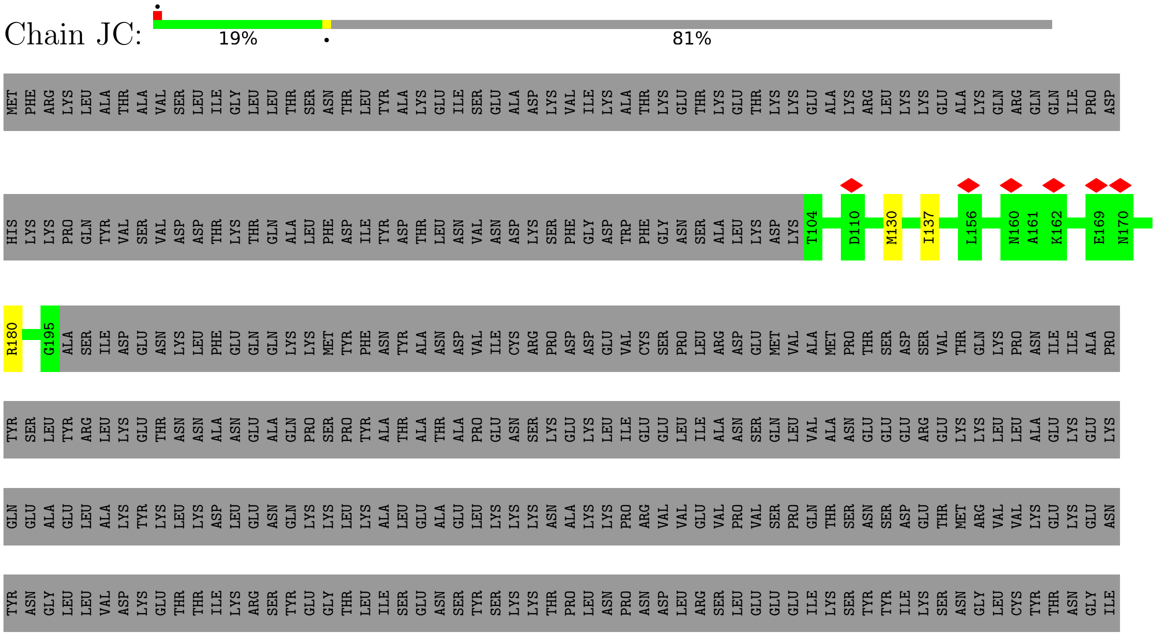
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain JB:

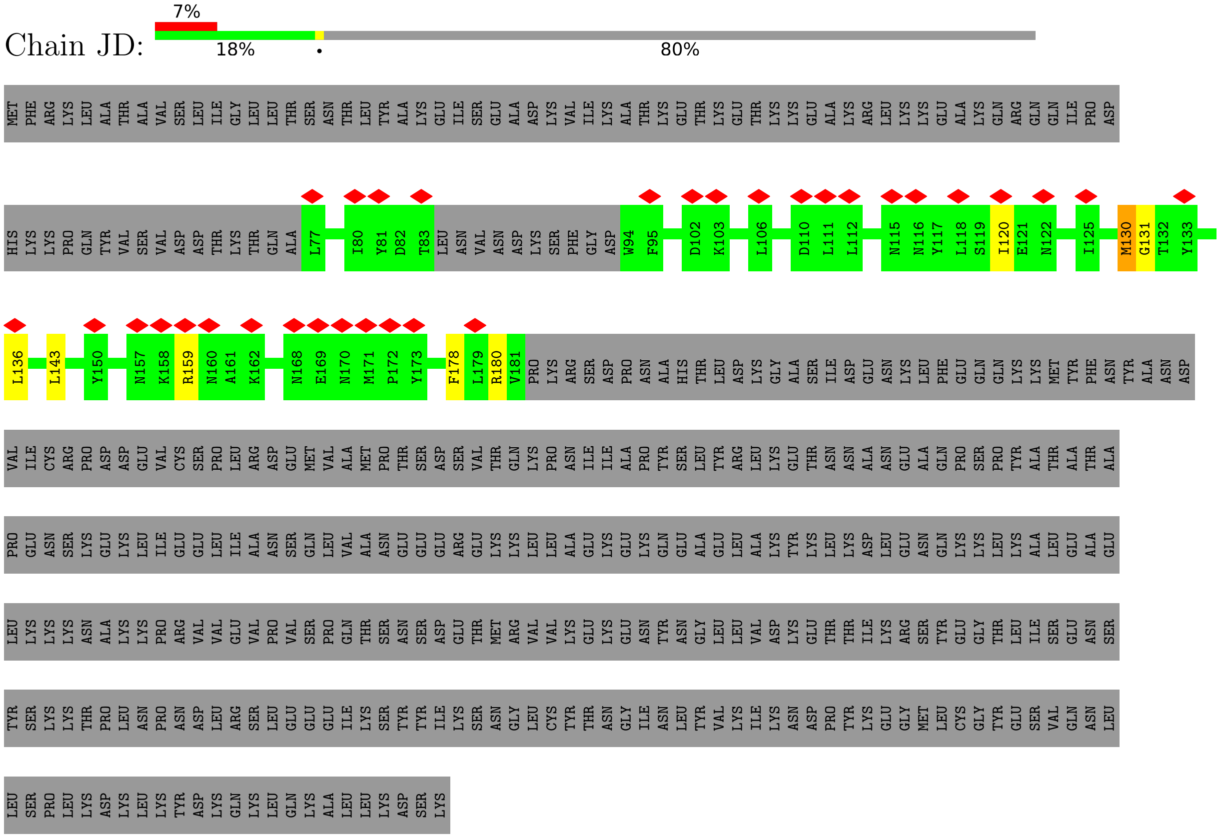


- Molecule 1: Type IV secretion system apparatus protein Cag3

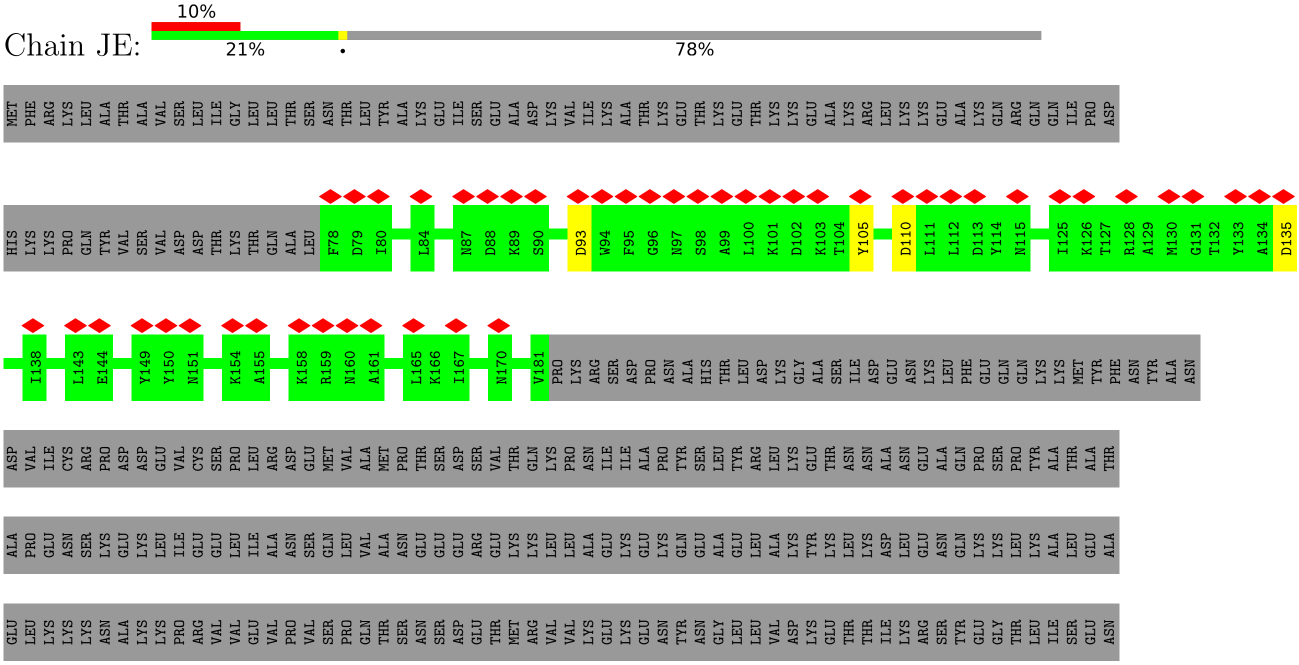
Chain JC:



- Molecule 1: Type IV secretion system apparatus protein Cag3

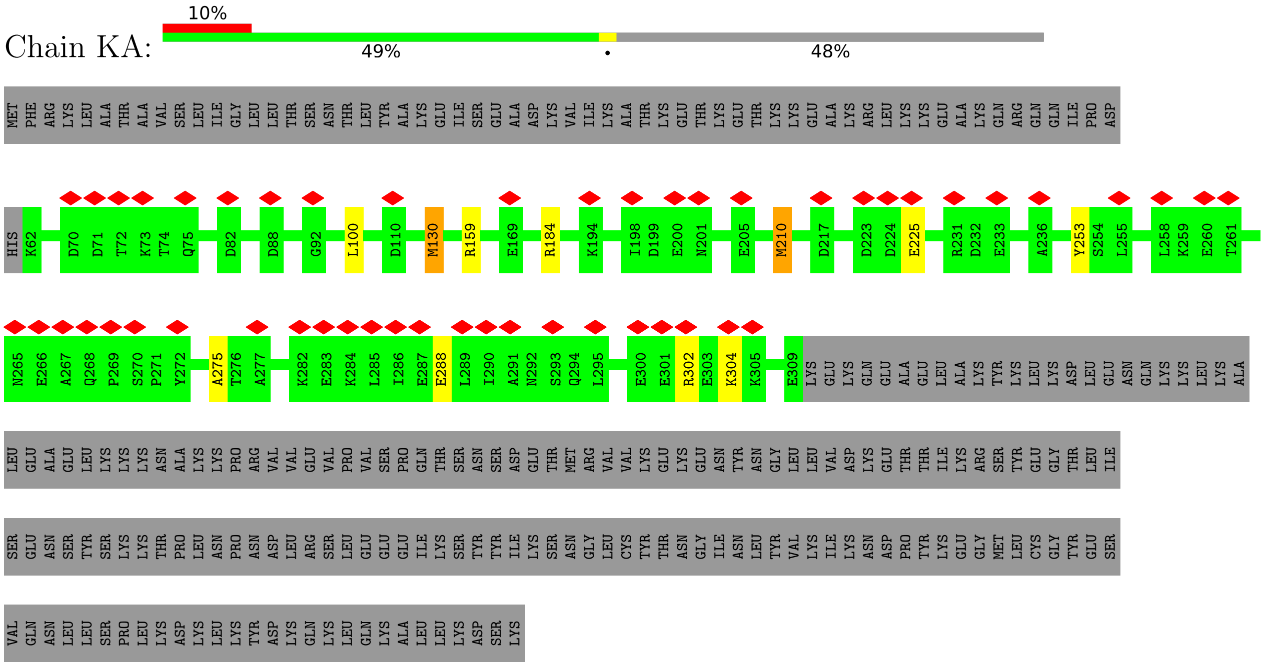


- Molecule 1: Type IV secretion system apparatus protein Cag3



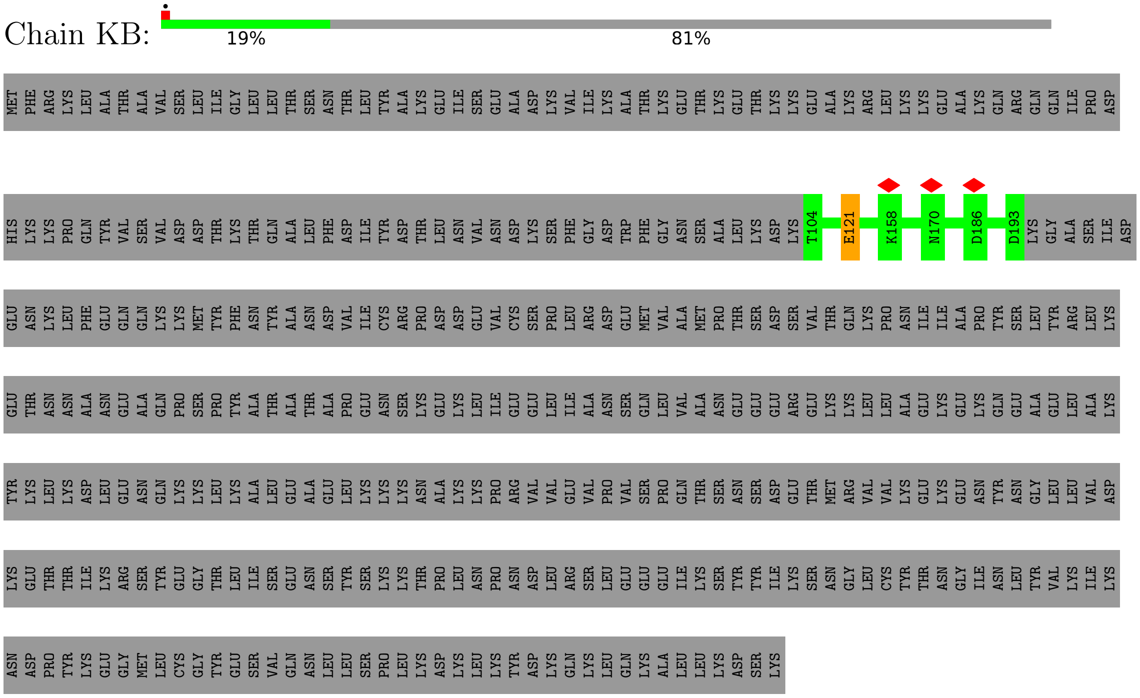
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain KA:

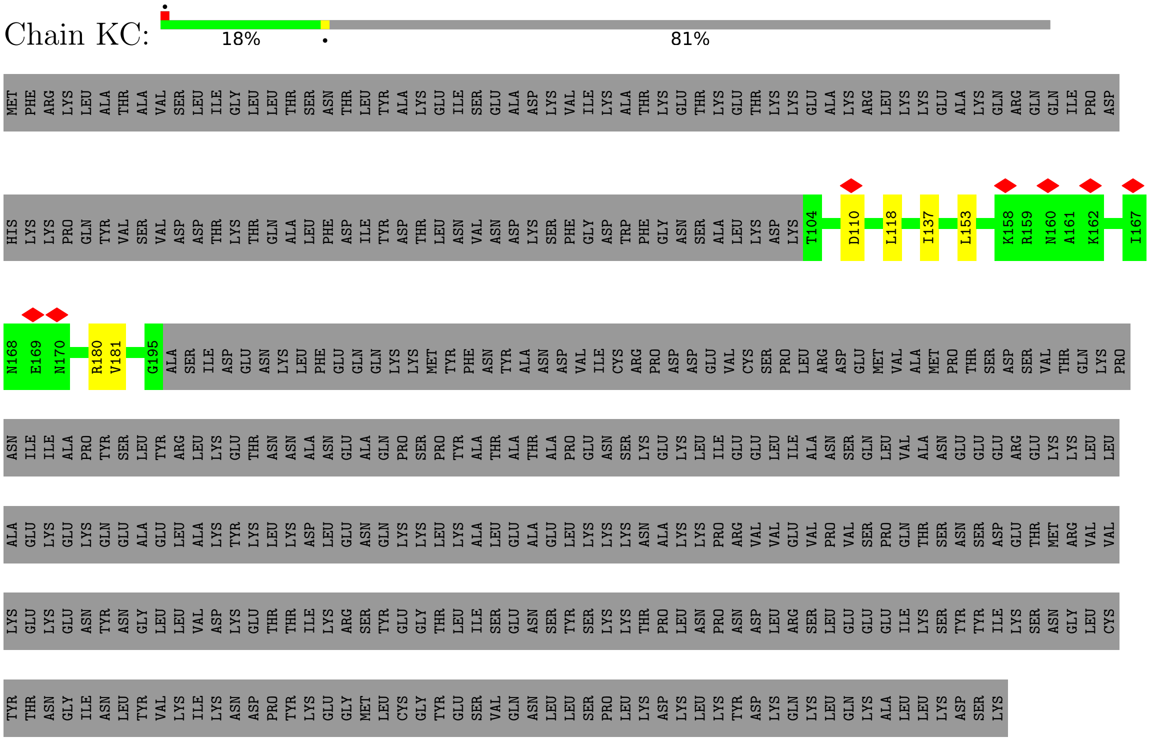


- Molecule 1: Type IV secretion system apparatus protein Cag3

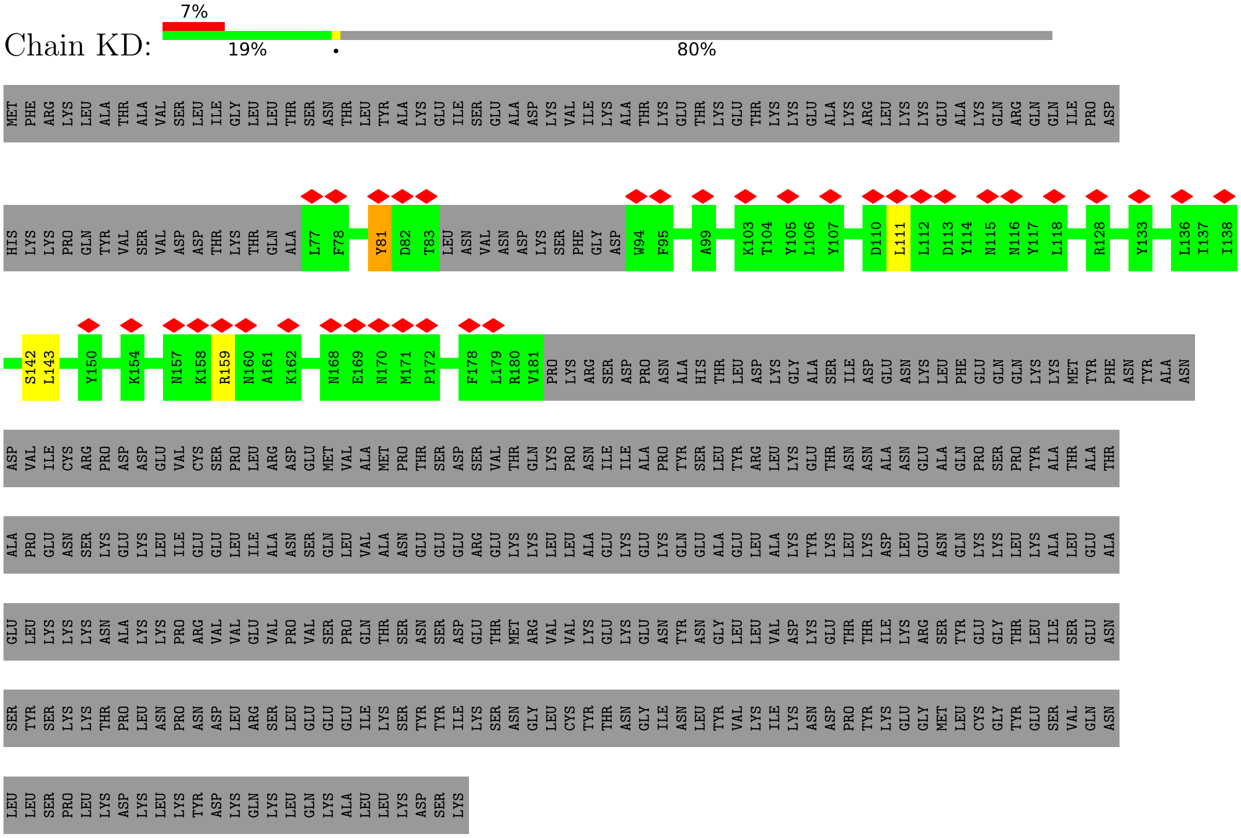
Chain KB:



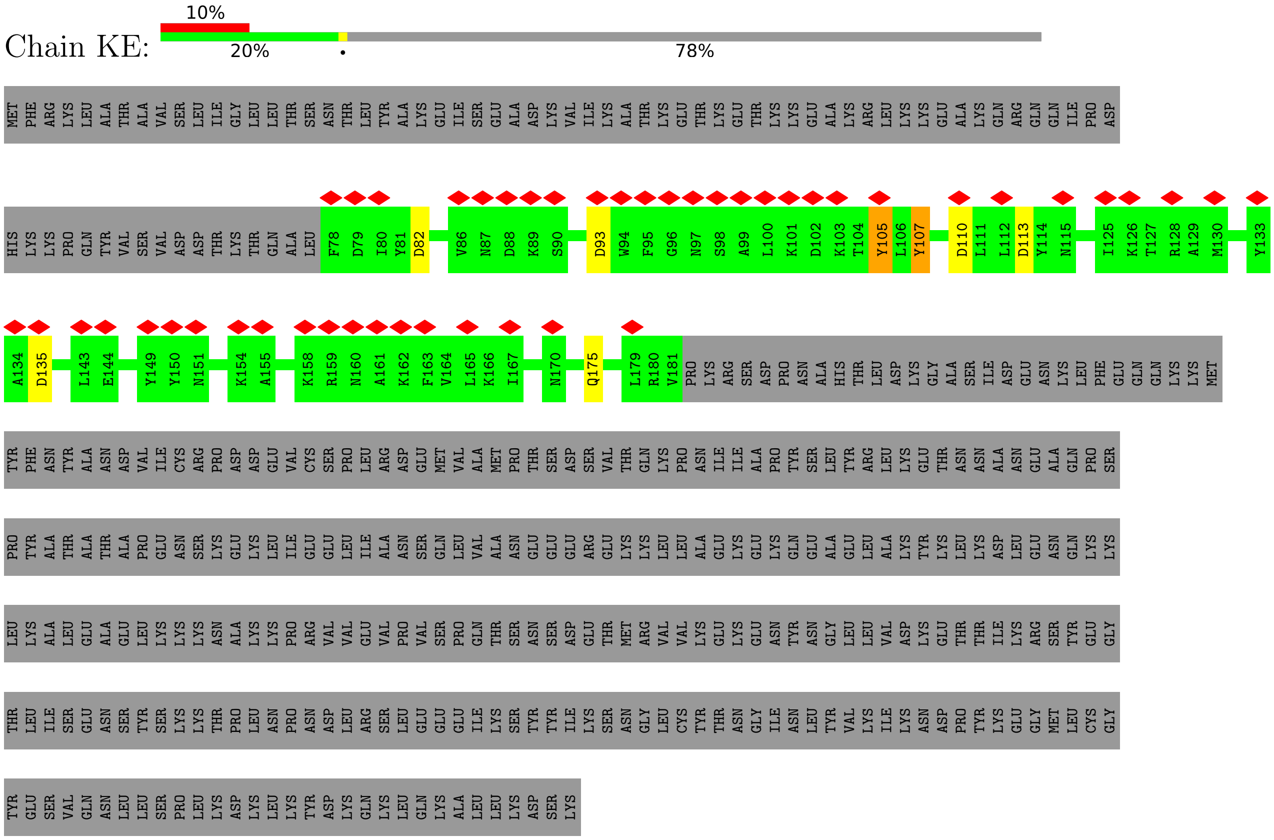
- Molecule 1: Type IV secretion system apparatus protein Cag3



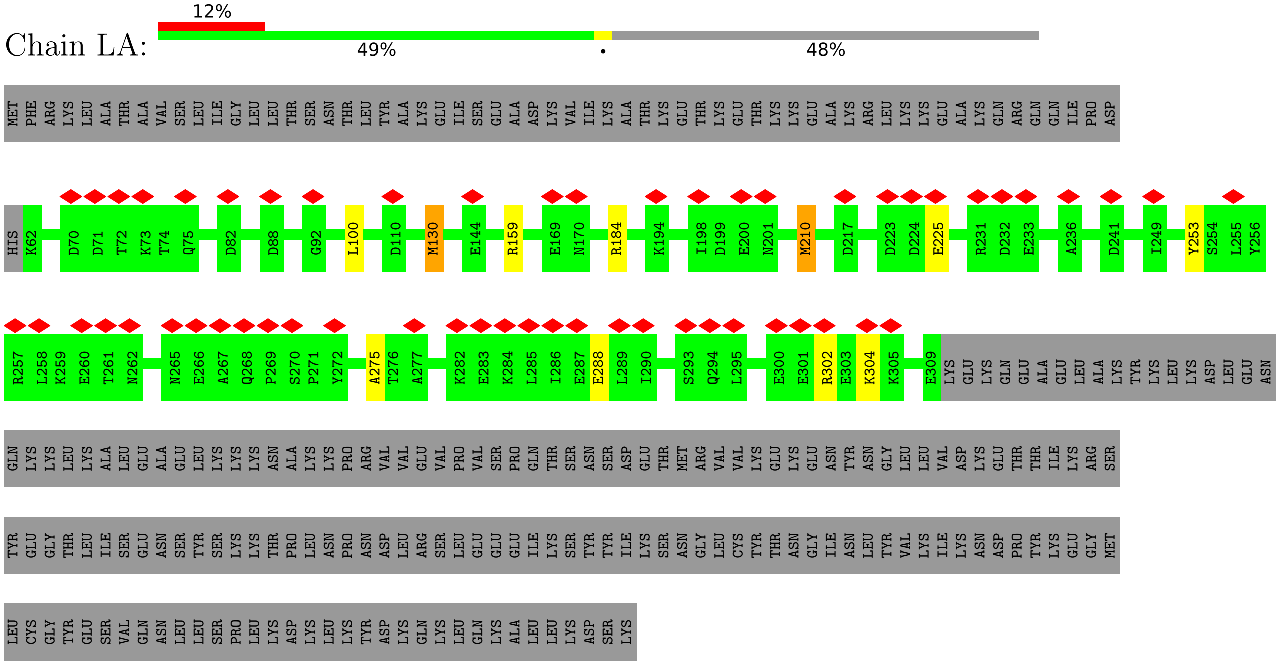
- Molecule 1: Type IV secretion system apparatus protein Cag3



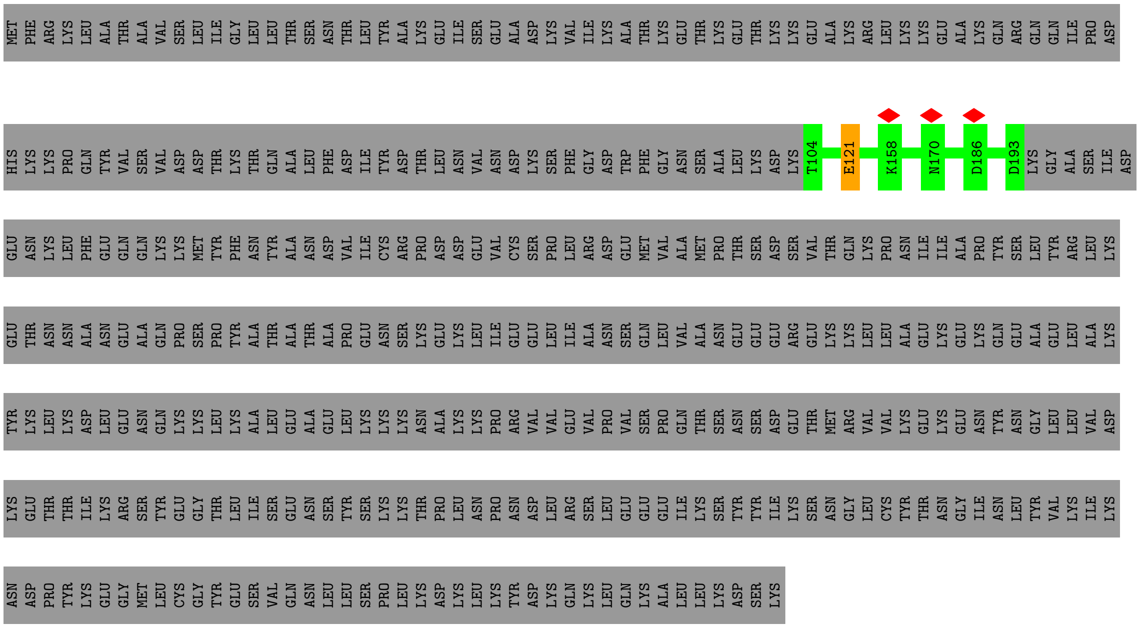
- Molecule 1: Type IV secretion system apparatus protein Cag3



- Molecule 1: Type IV secretion system apparatus protein Cag3

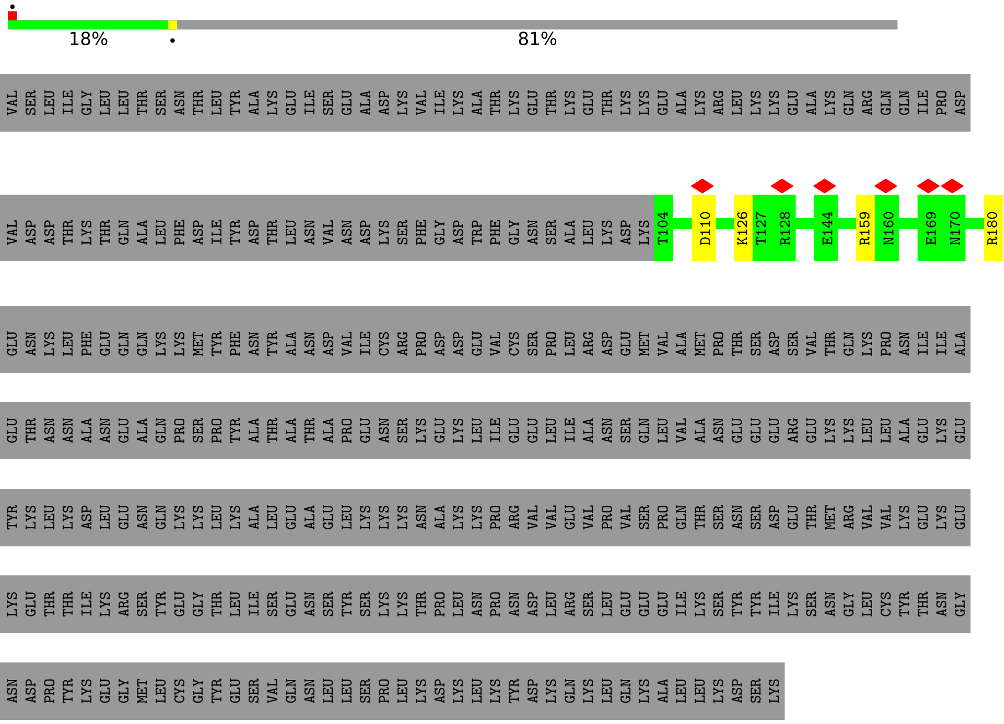


- Molecule 1: Type IV secretion system apparatus protein Cag3



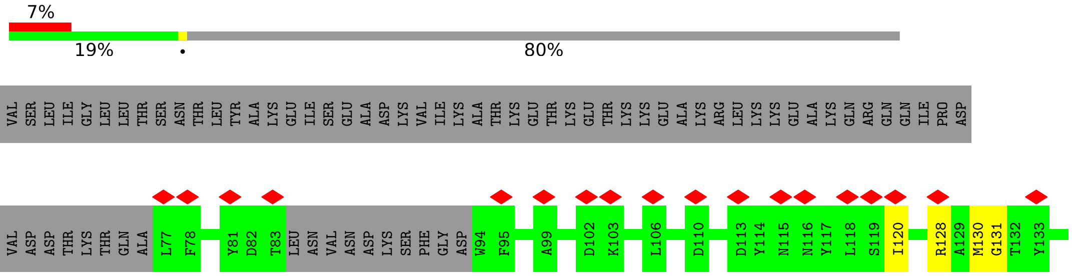
- Molecule 1: Type IV secretion system apparatus protein Cag3

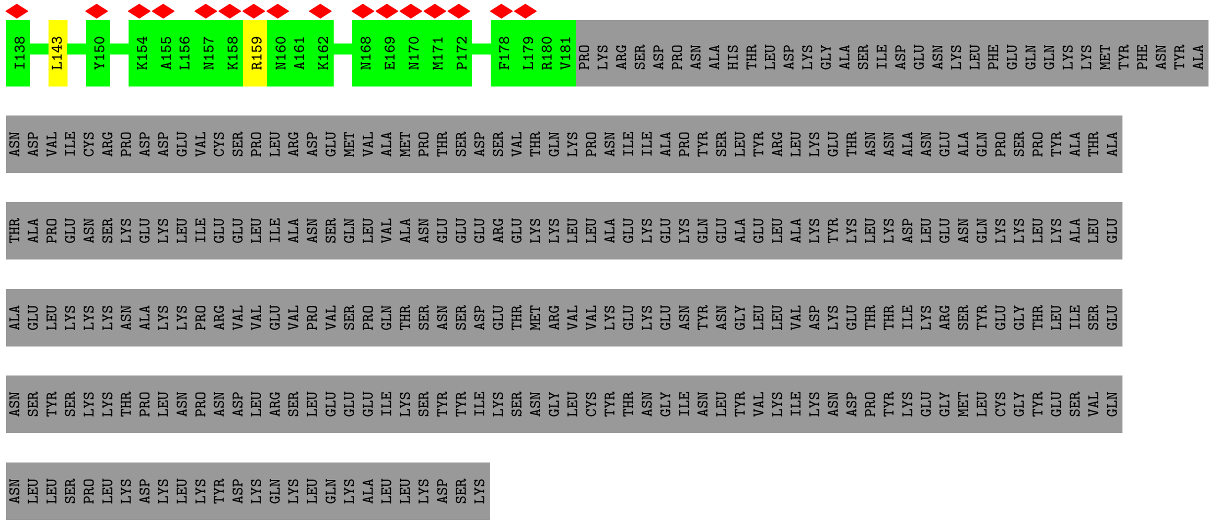
Chain LC:



- Molecule 1: Type IV secretion system apparatus protein Cag3

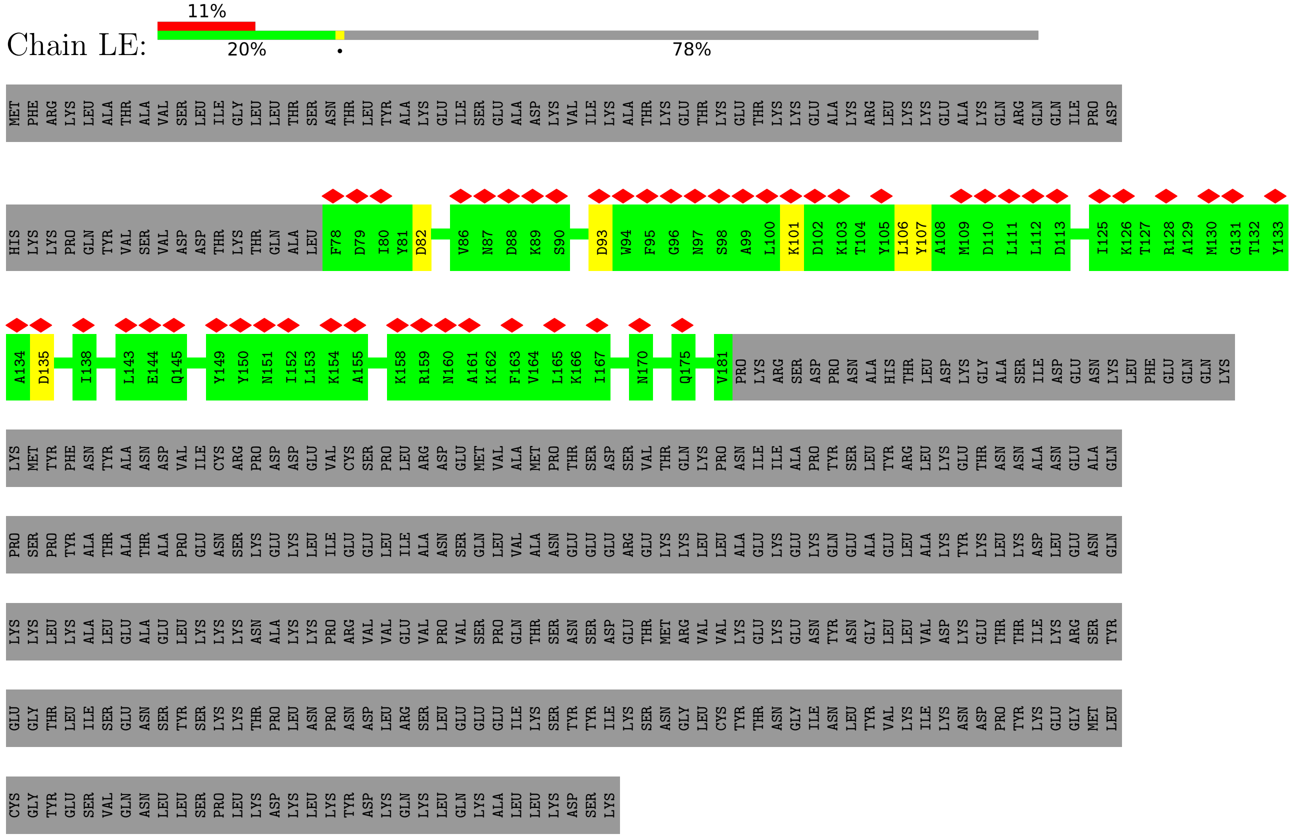
Chain LD:





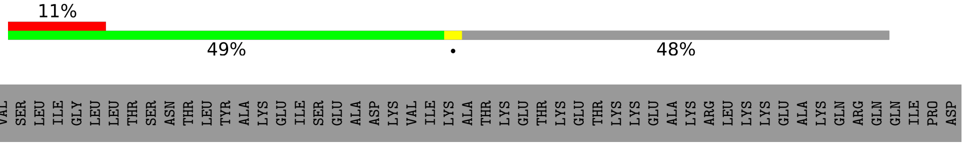
- Molecule 1: Type IV secretion system apparatus protein Cag3

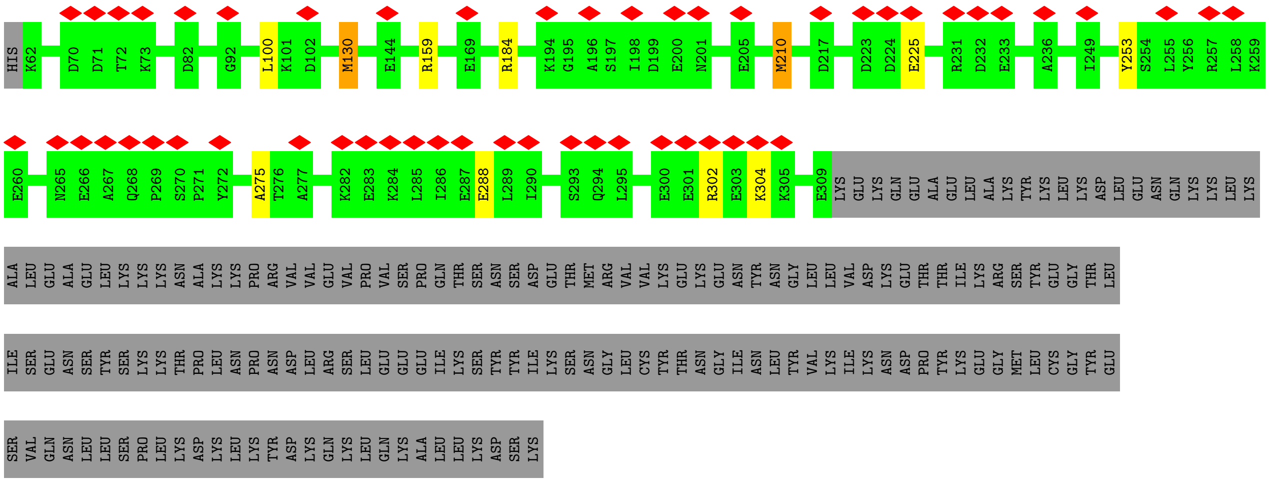
Chain LE:



- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain MA:



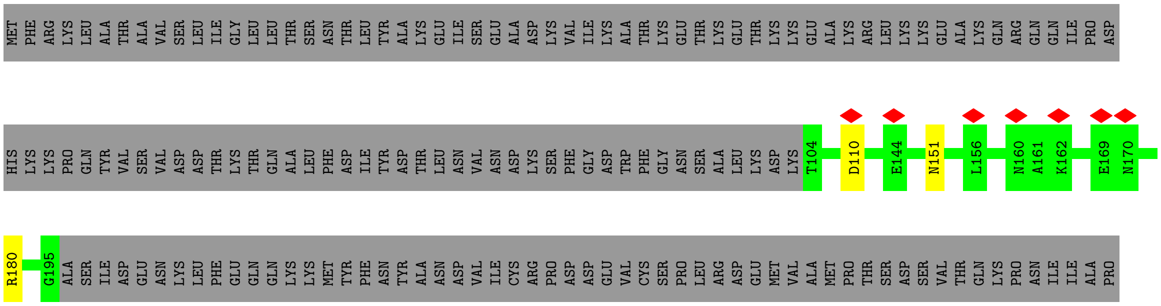


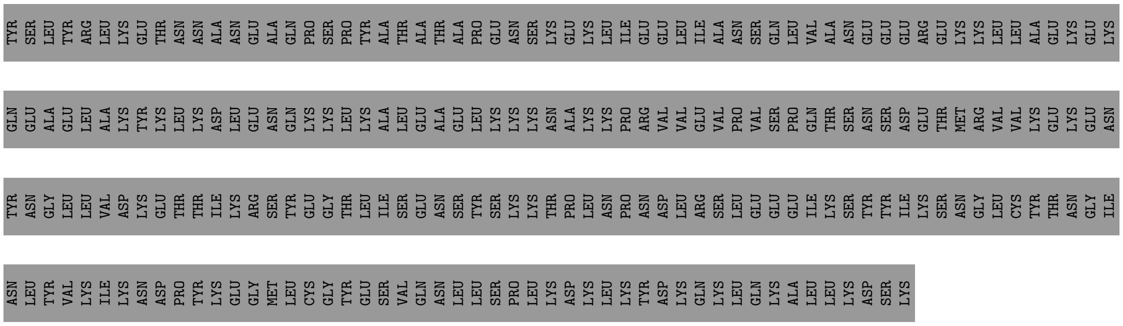
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain MB:

- Molecule 1: Type IV secretion system apparatus protein Cag3

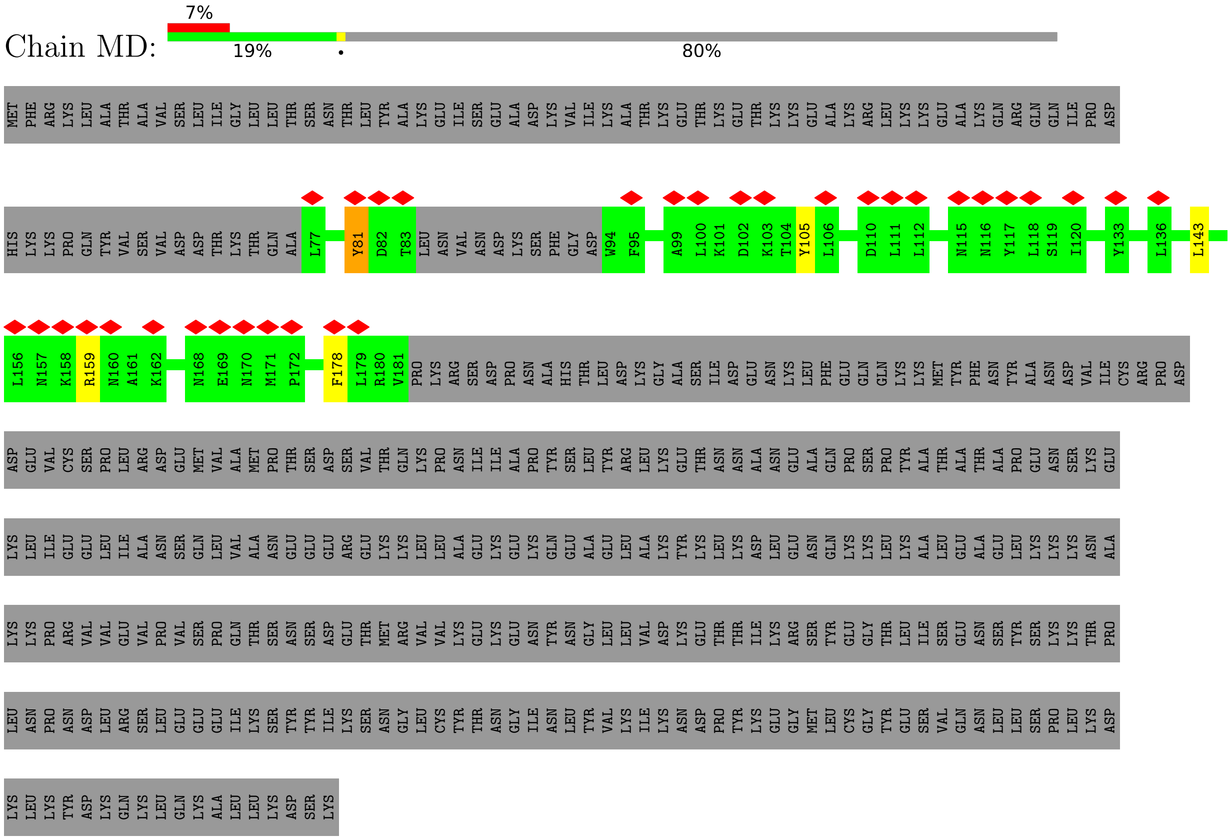
Chain MC:





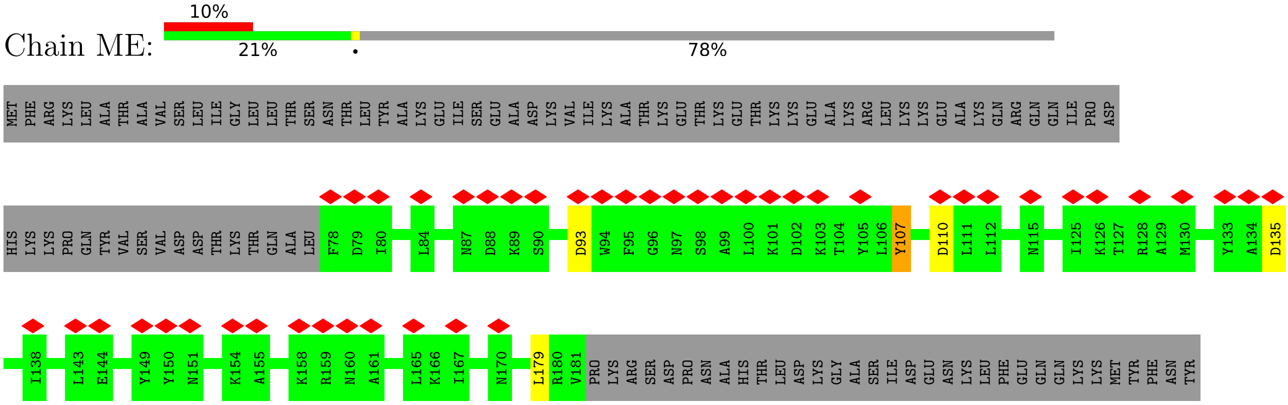
- Molecule 1: Type IV secretion system apparatus protein Cag3

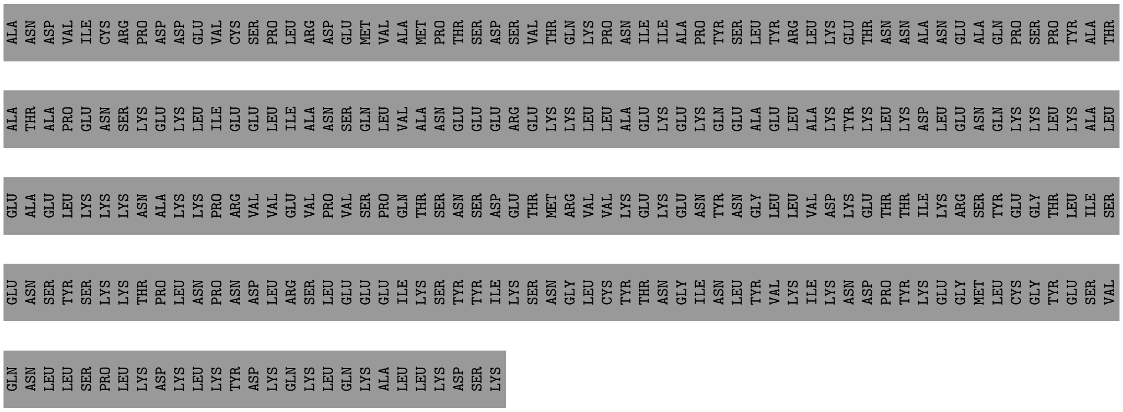
Chain MD:



- Molecule 1: Type IV secretion system apparatus protein Cag3

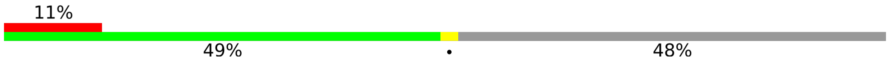
Chain ME:

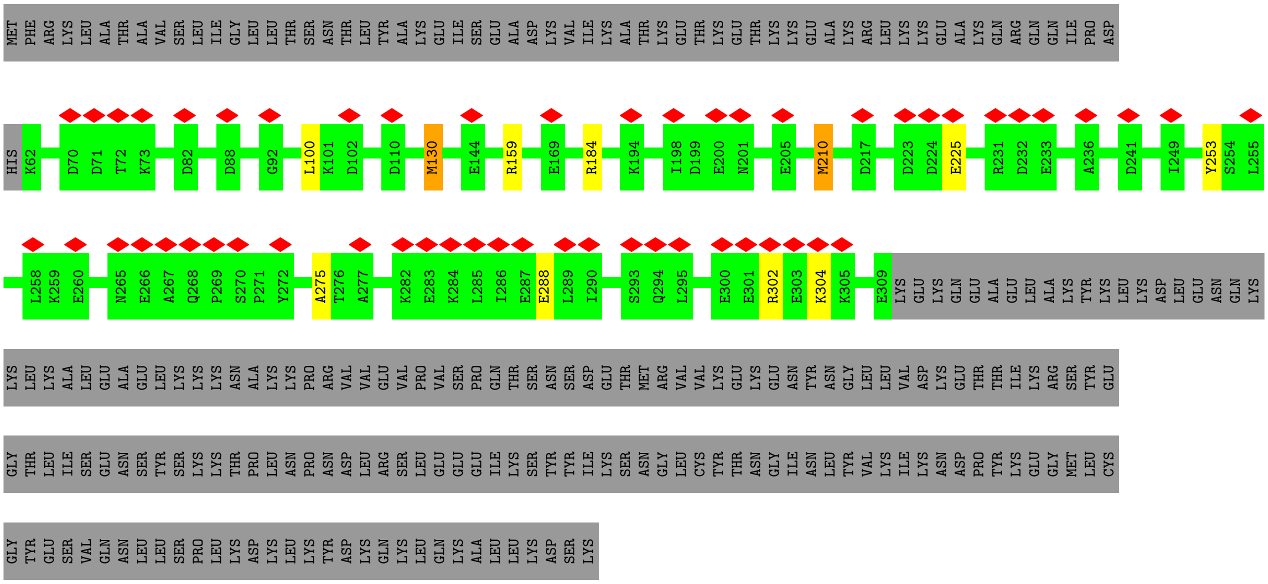




- Molecule 1: Type IV secretion system apparatus protein Cag3

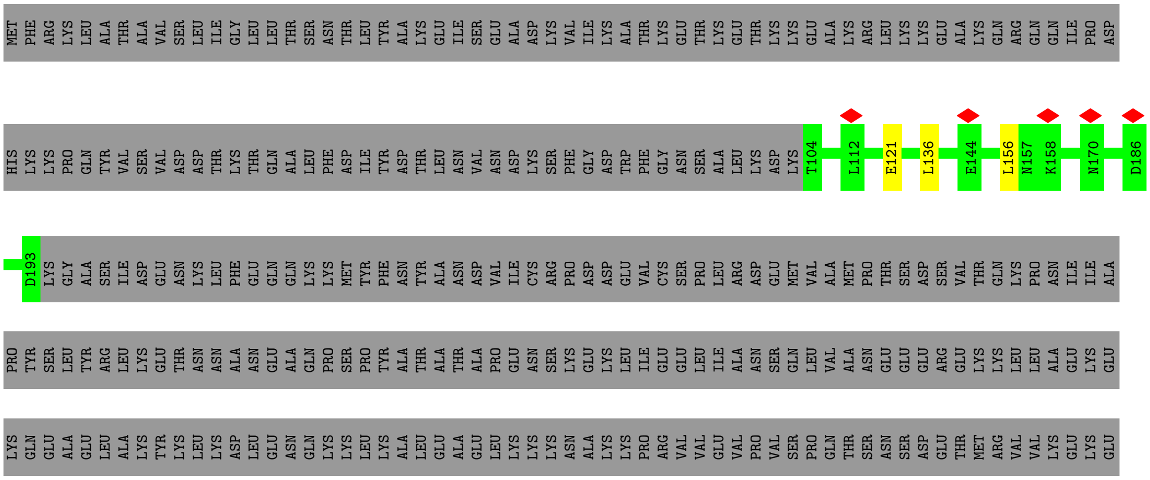
Chain NA:

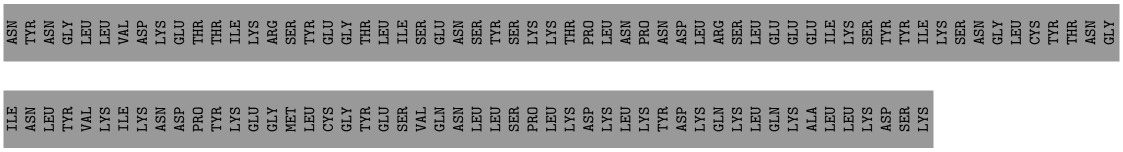




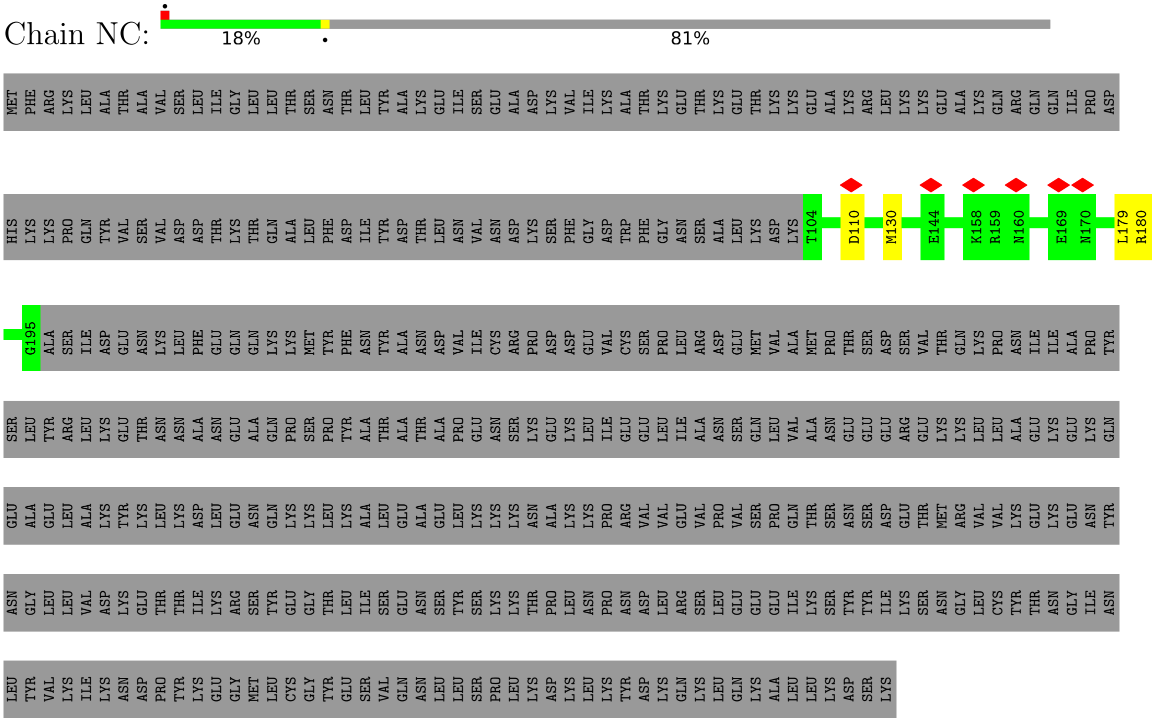
- Molecule 1: Type IV secretion system apparatus protein Cag3

Chain NB:

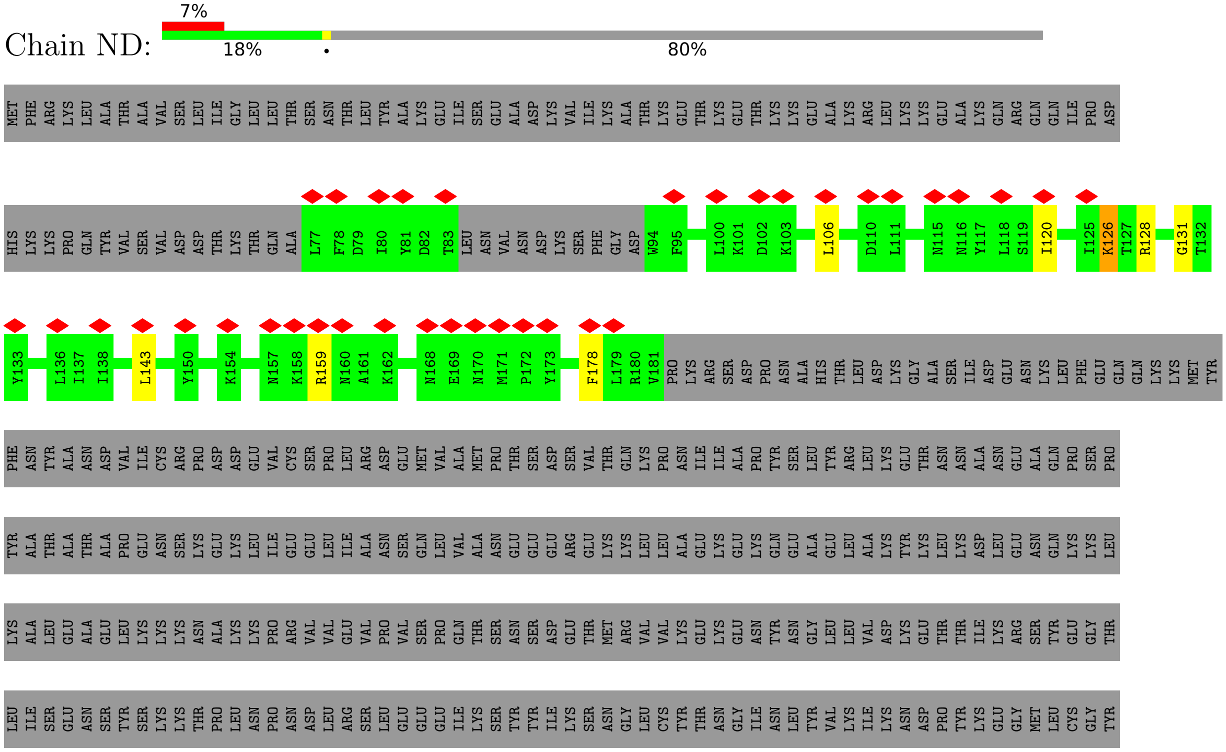




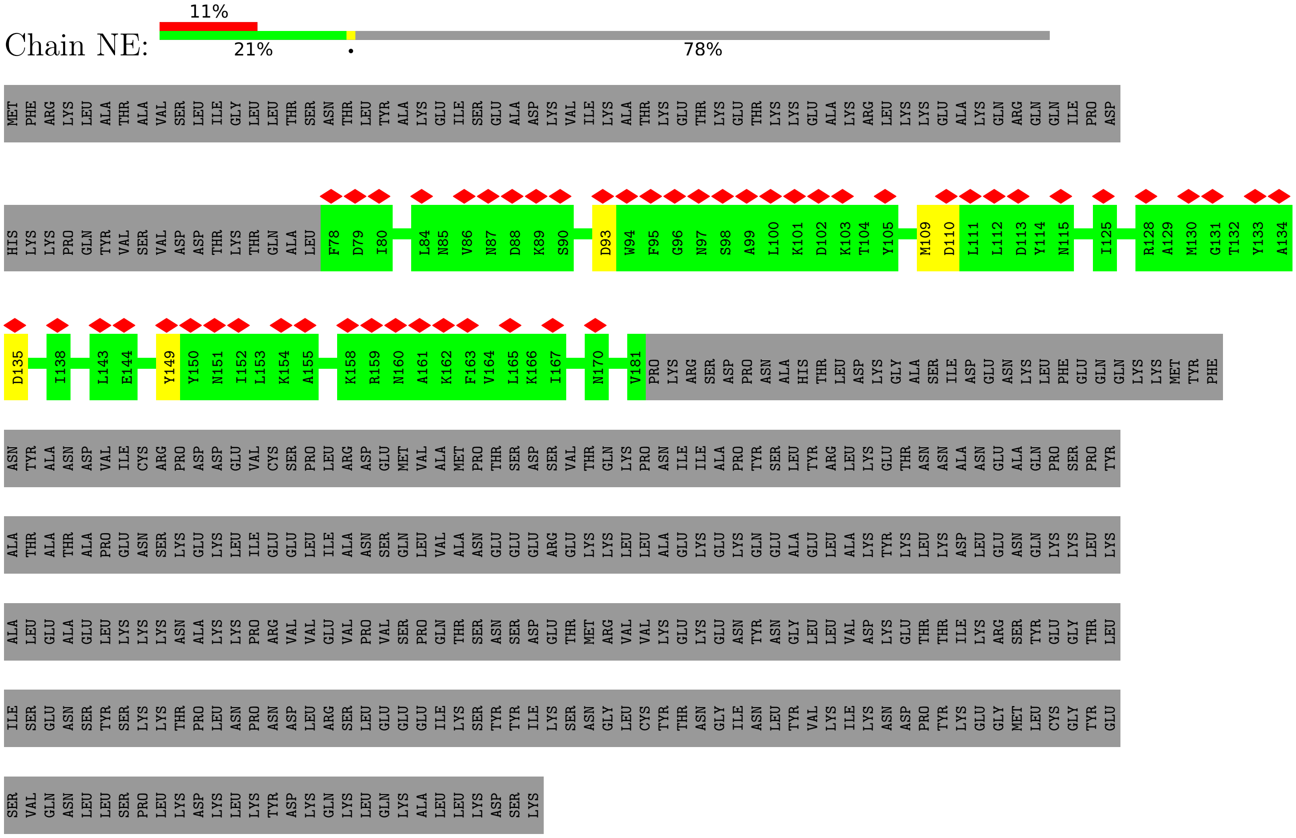
- Molecule 1: Type IV secretion system apparatus protein Cag3



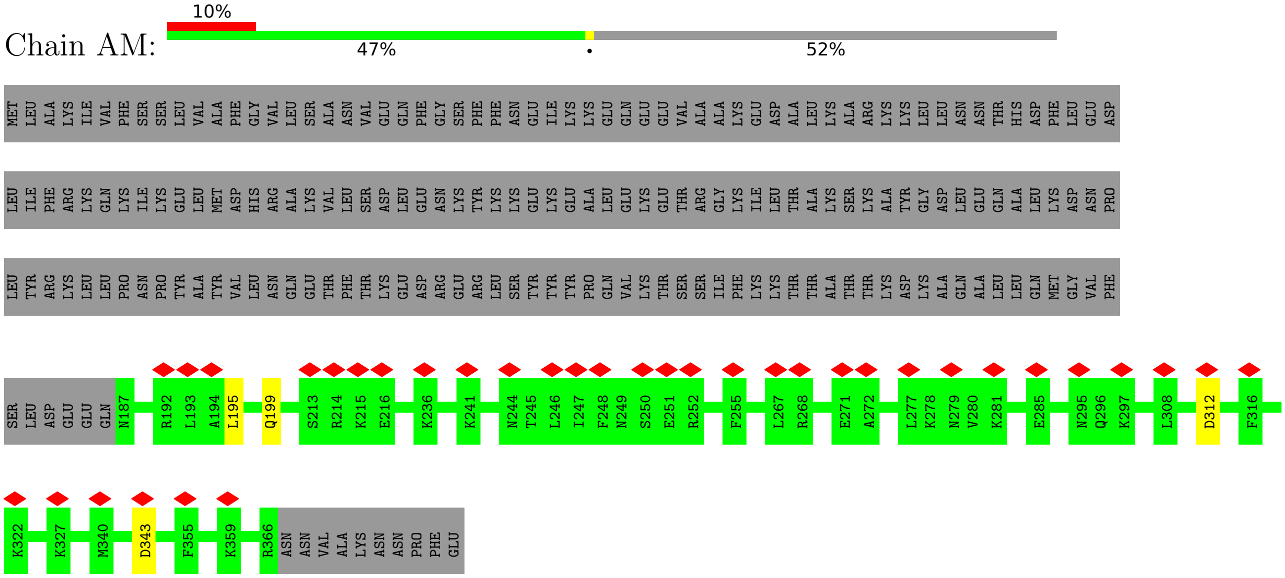
- Molecule 1: Type IV secretion system apparatus protein Cag3



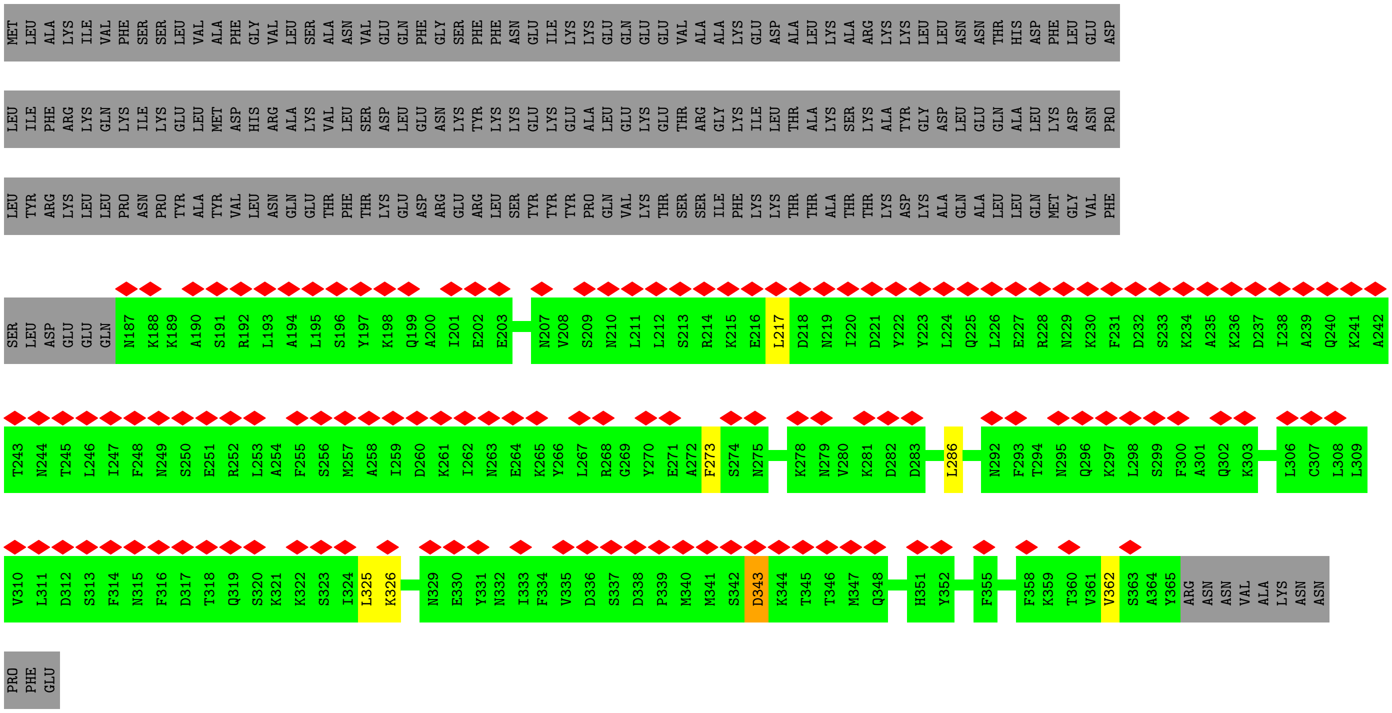
- Molecule 1: Type IV secretion system apparatus protein Cag3



- Molecule 2: Type IV secretion system apparatus protein CagM

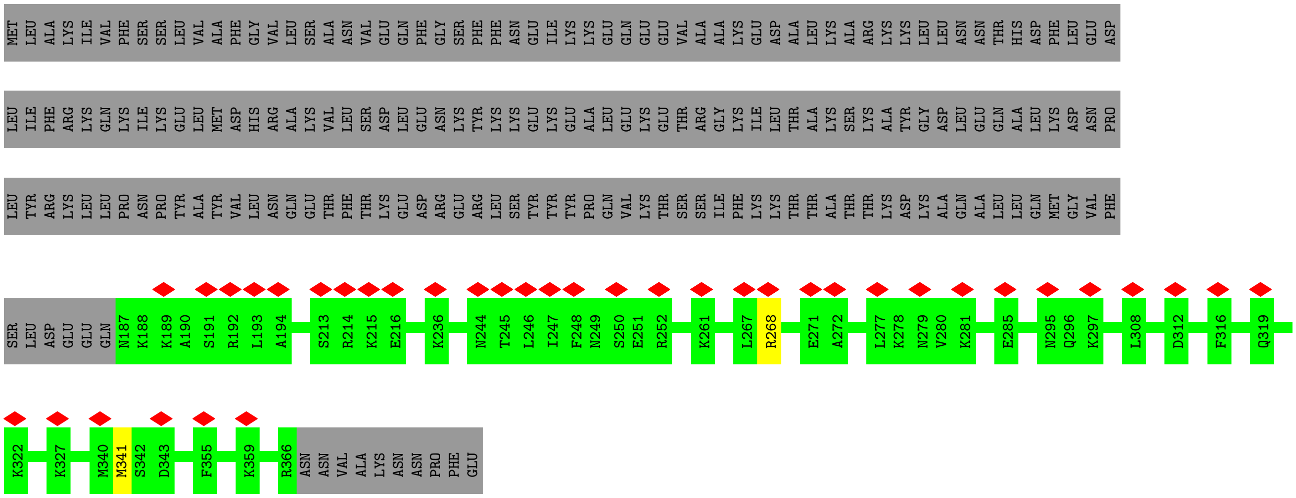


- Molecule 2: Type IV secretion system apparatus protein CagM



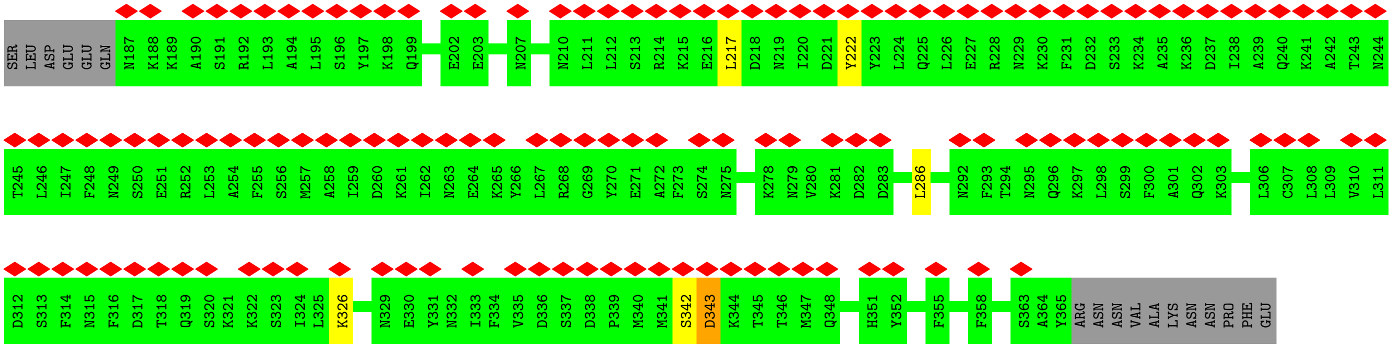
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain BM:

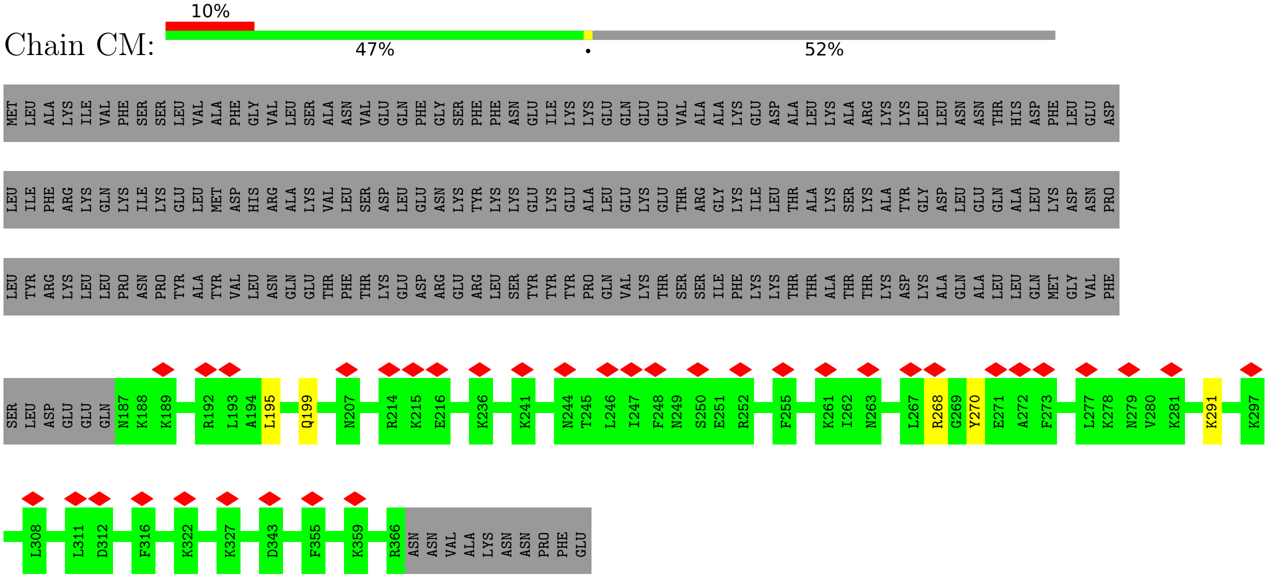


- Molecule 2: Type IV secretion system apparatus protein CagM

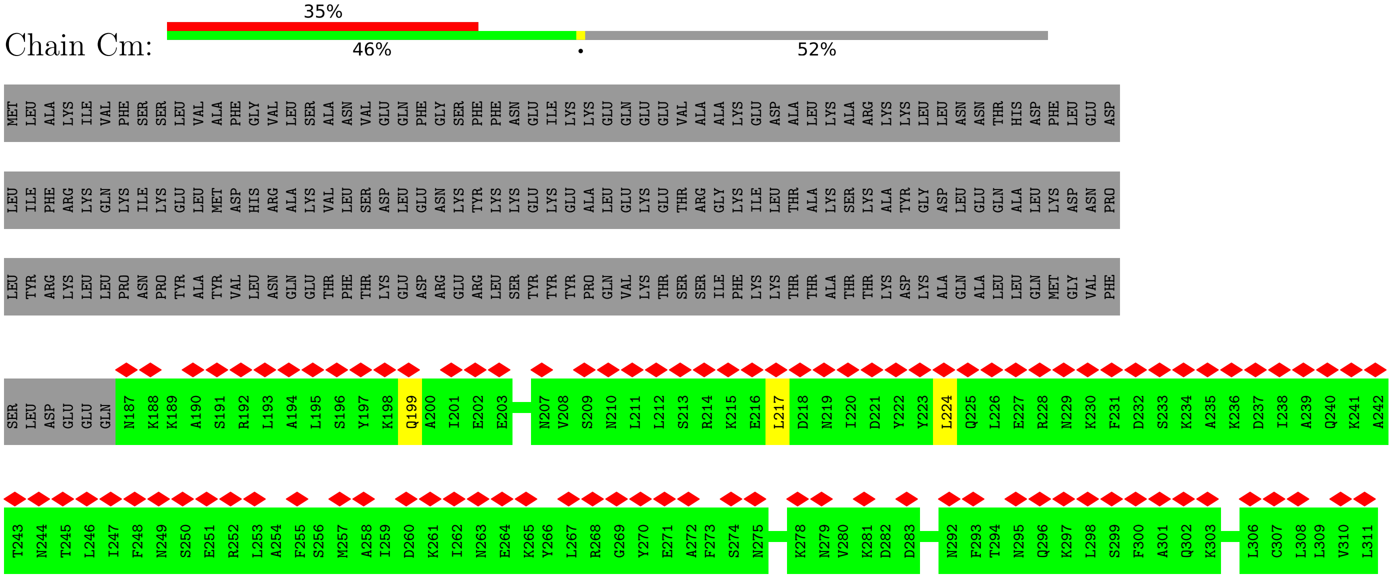
Chain Bm:

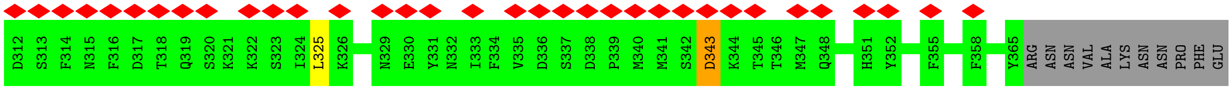


- Molecule 2: Type IV secretion system apparatus protein CagM

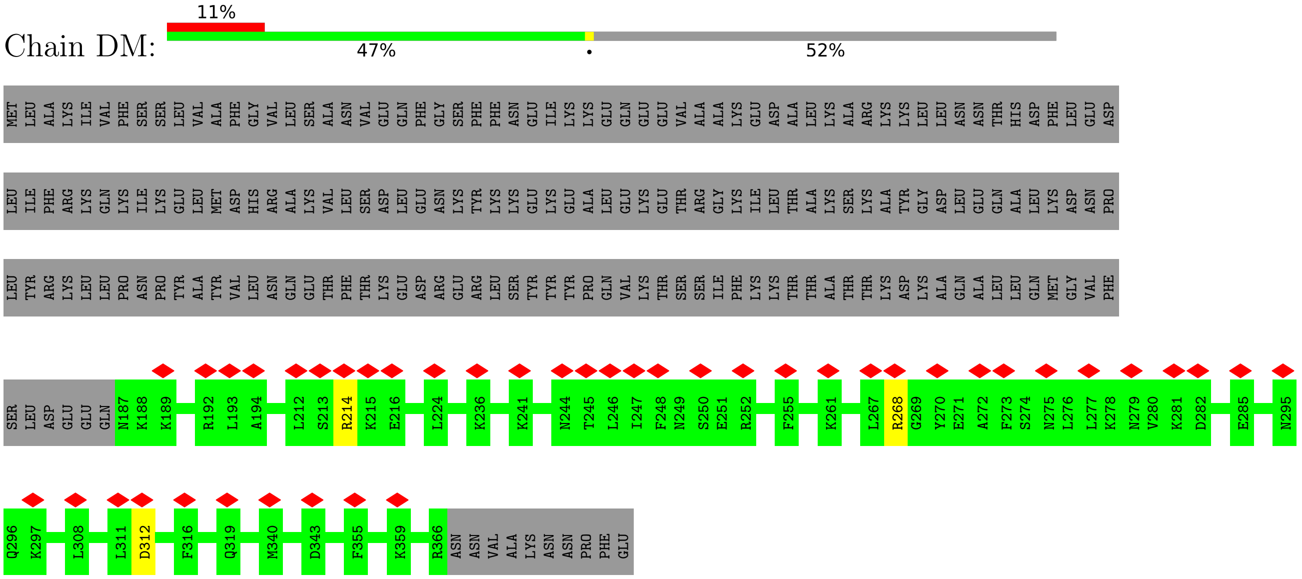


- Molecule 2: Type IV secretion system apparatus protein CagM

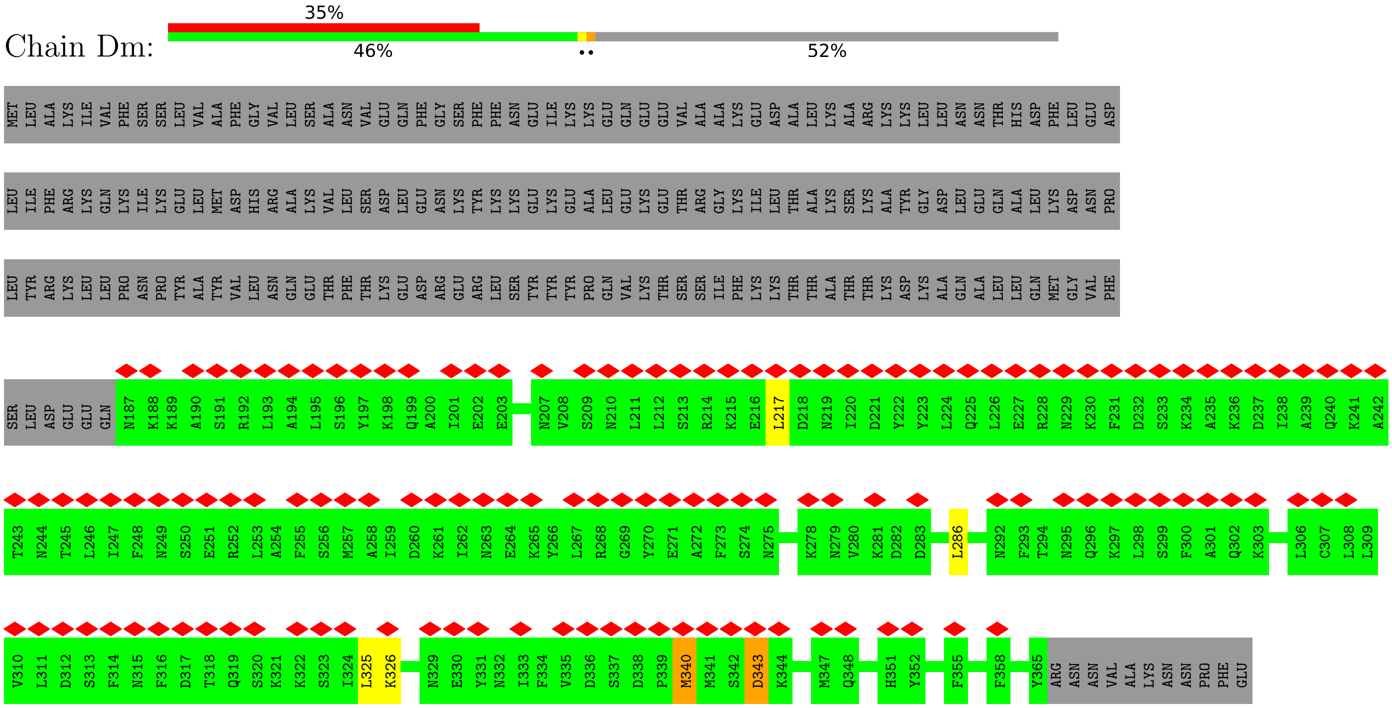




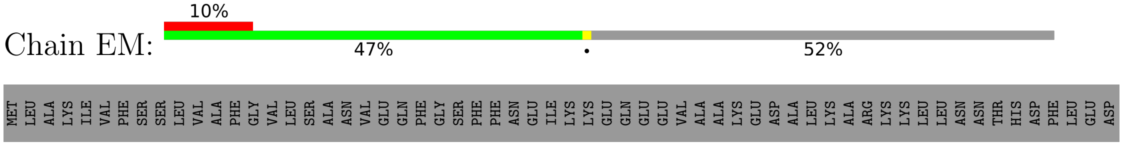
- Molecule 2: Type IV secretion system apparatus protein CagM

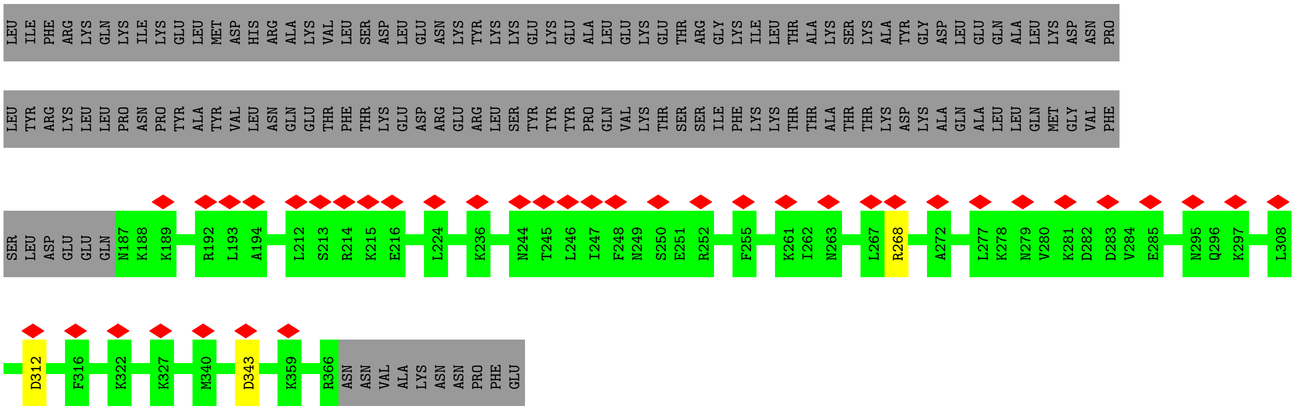


- Molecule 2: Type IV secretion system apparatus protein CagM



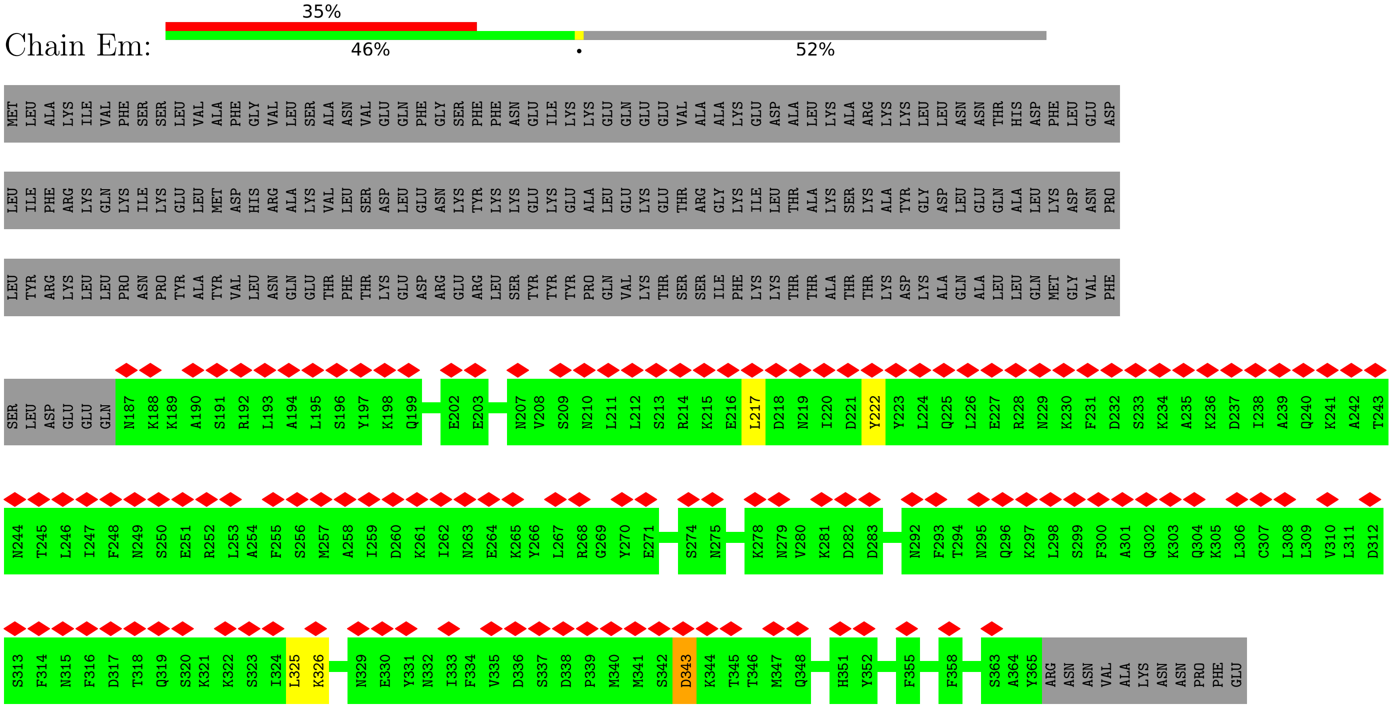
- Molecule 2: Type IV secretion system apparatus protein CagM





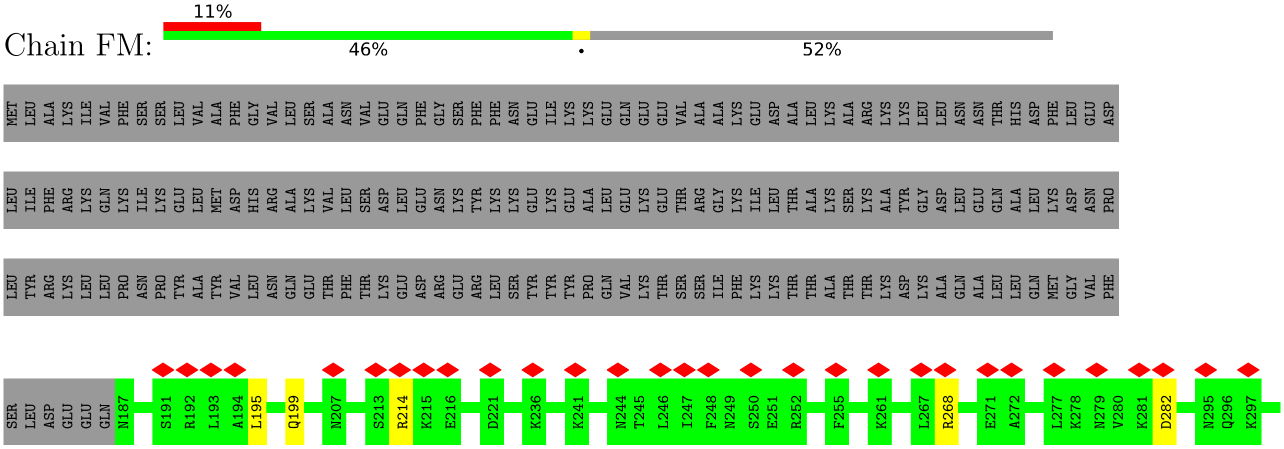
- Molecule 2: Type IV secretion system apparatus protein CagM

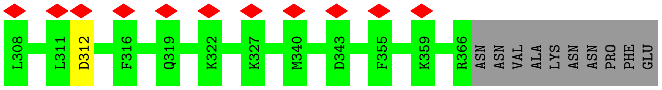
Chain Em:



- Molecule 2: Type IV secretion system apparatus protein CagM

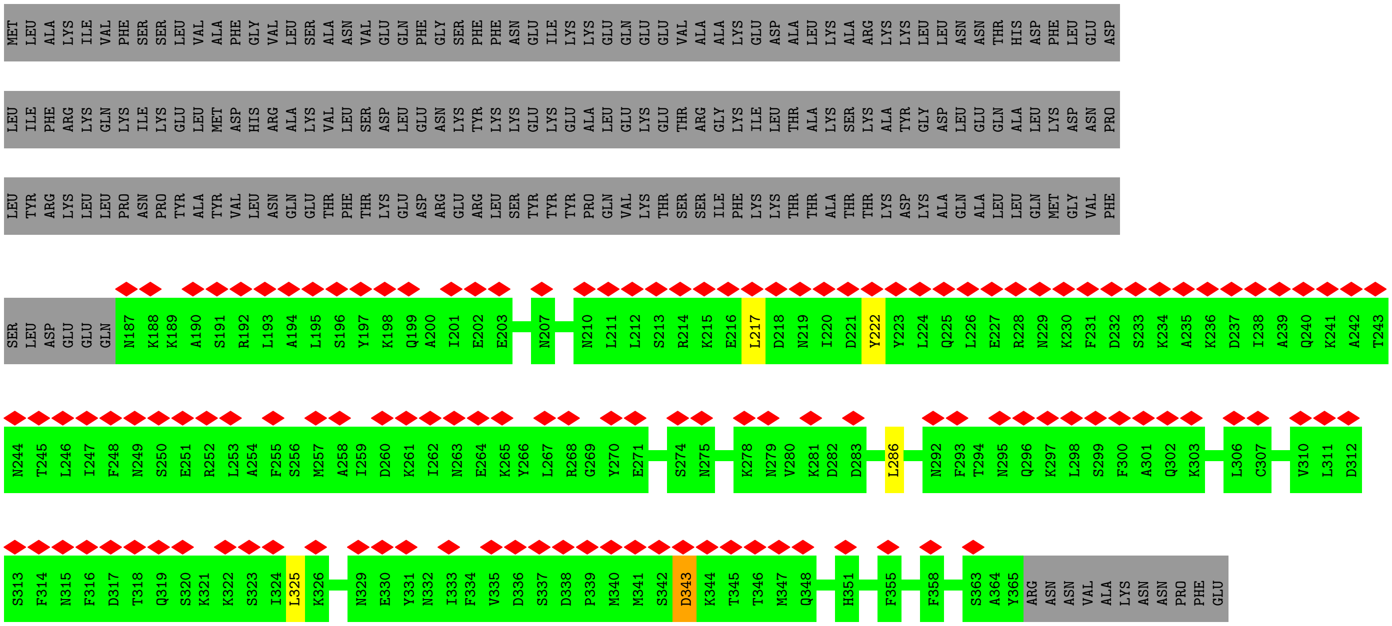
Chain FM:





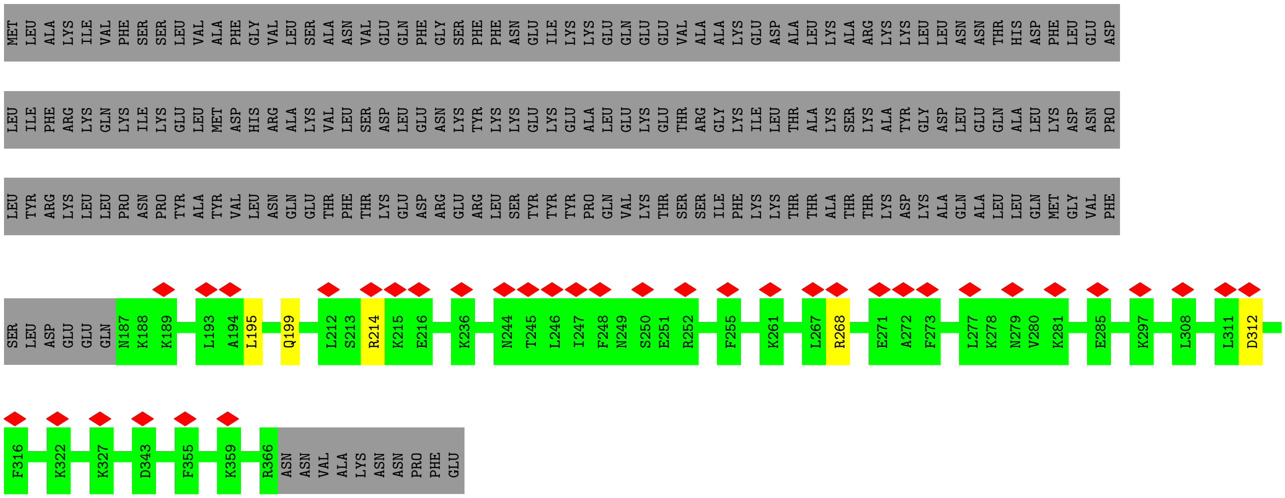
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain Fm:



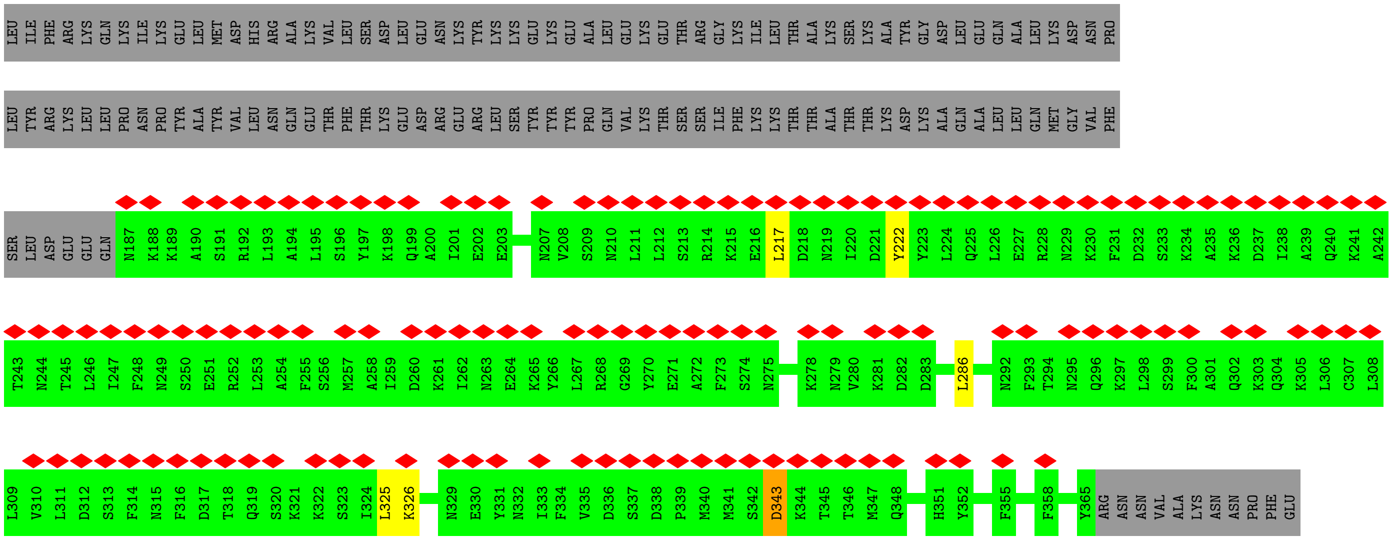
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain GM:



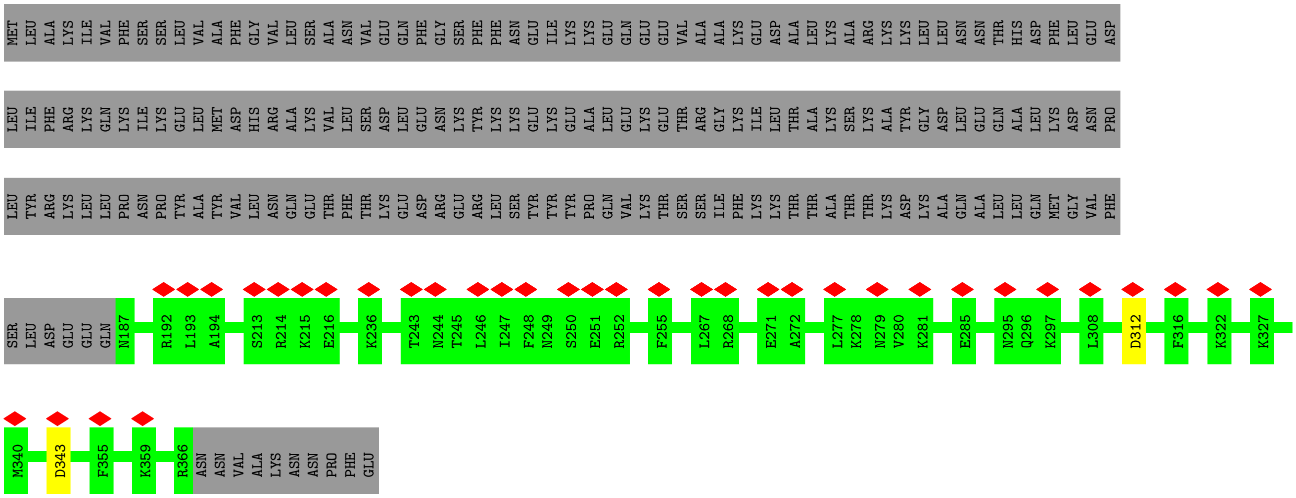
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain Gm:



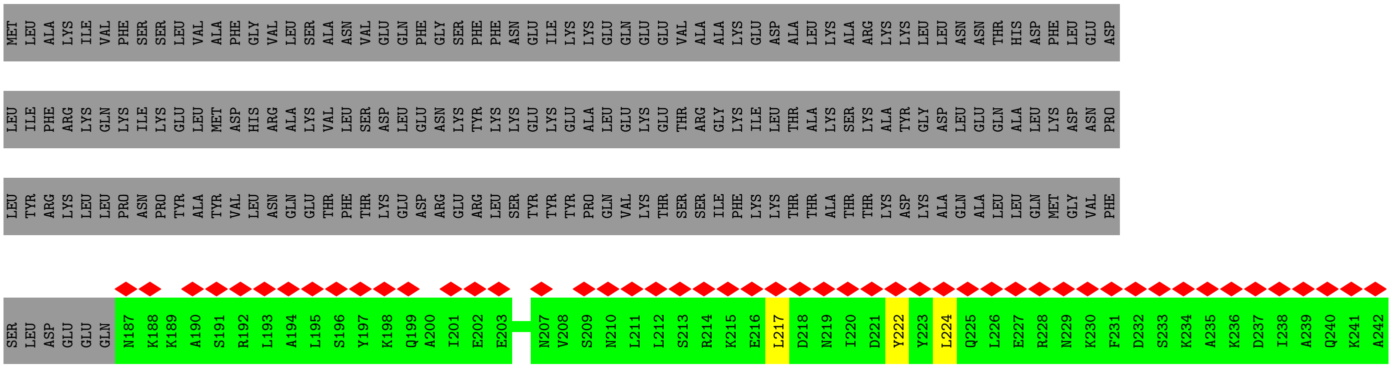
- Molecule 2: Type IV secretion system apparatus protein CagM

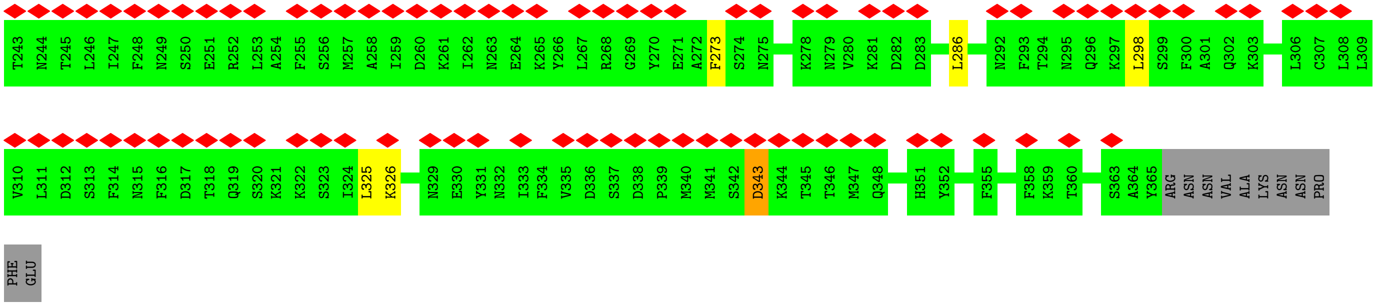
Chain HM:



- Molecule 2: Type IV secretion system apparatus protein CagM

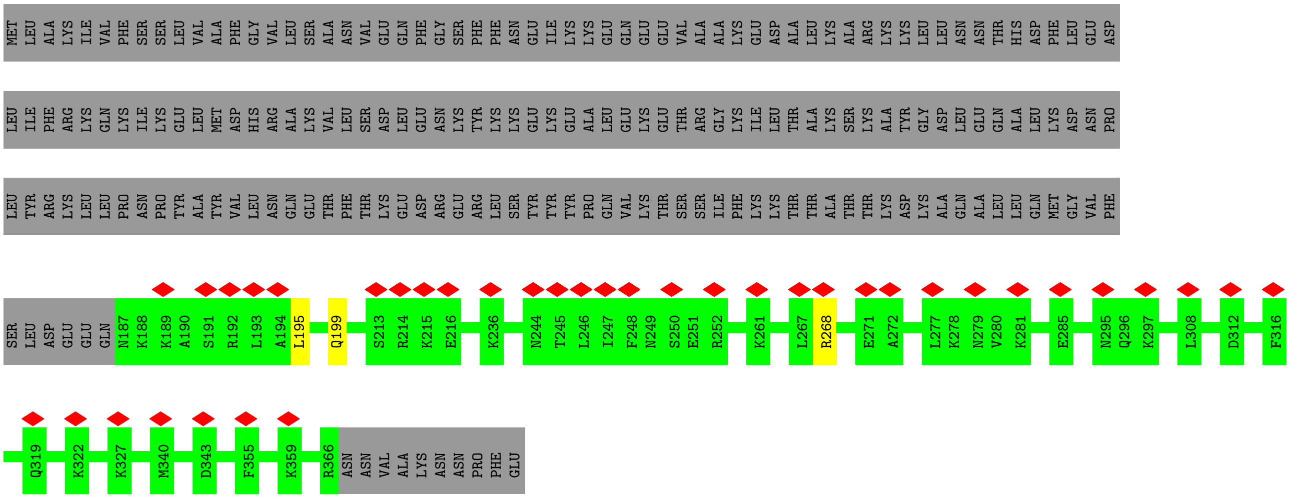
Chain Hm:





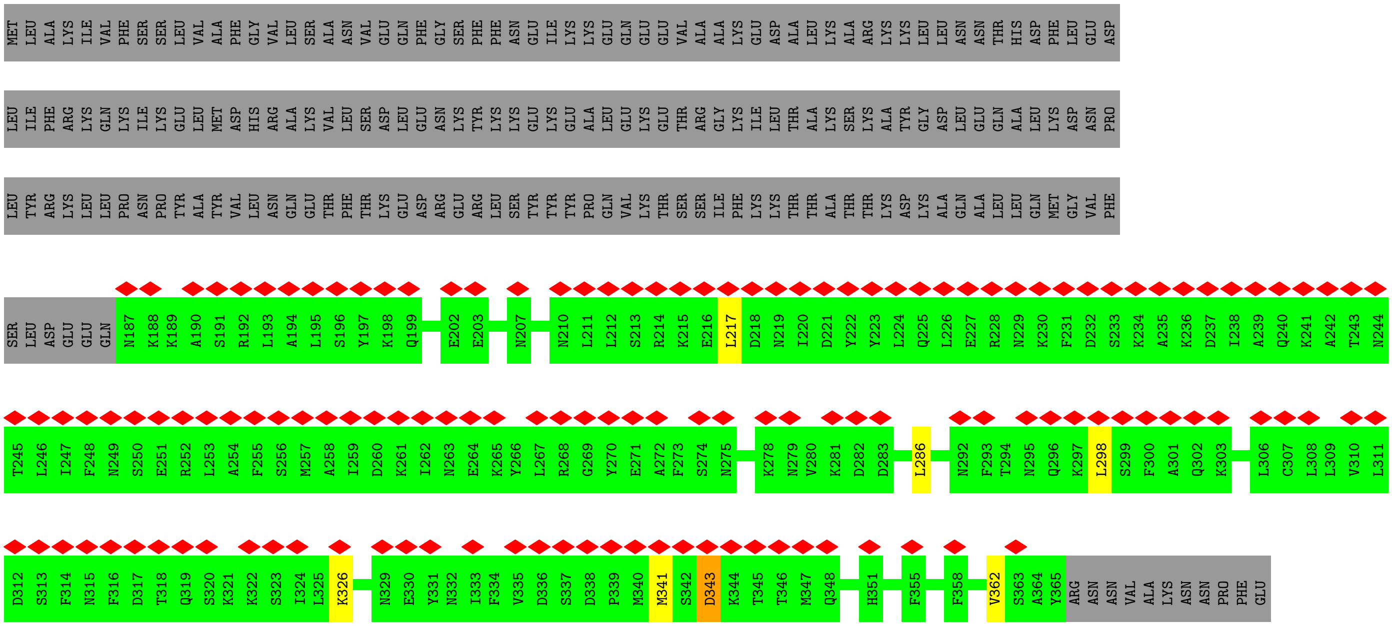
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain IM:



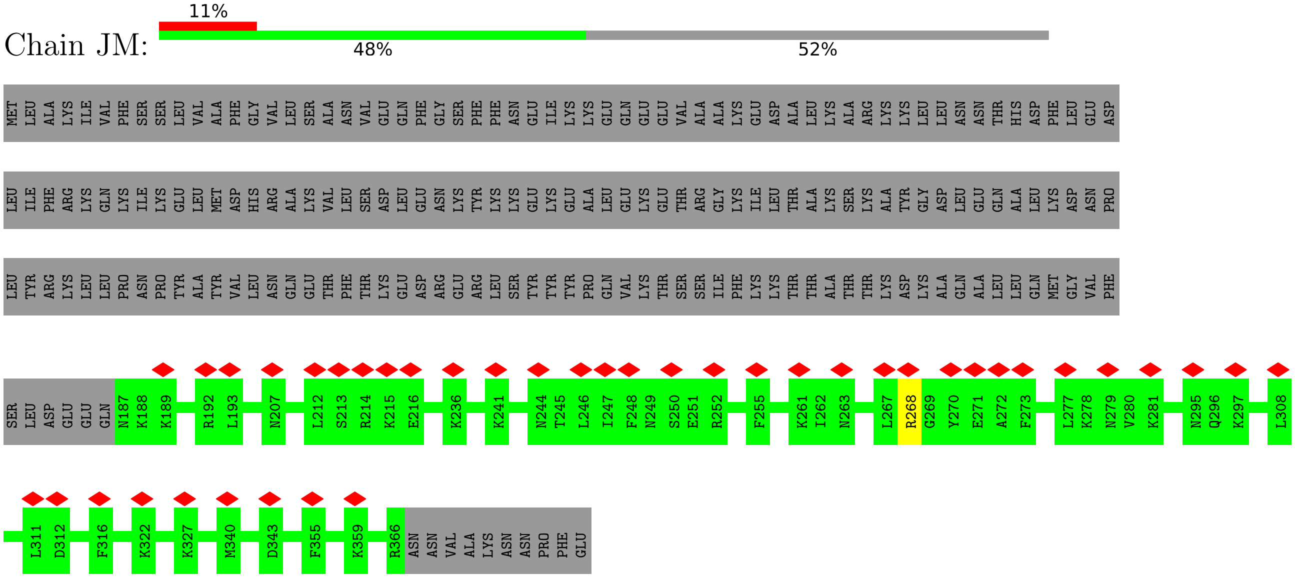
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain Im:



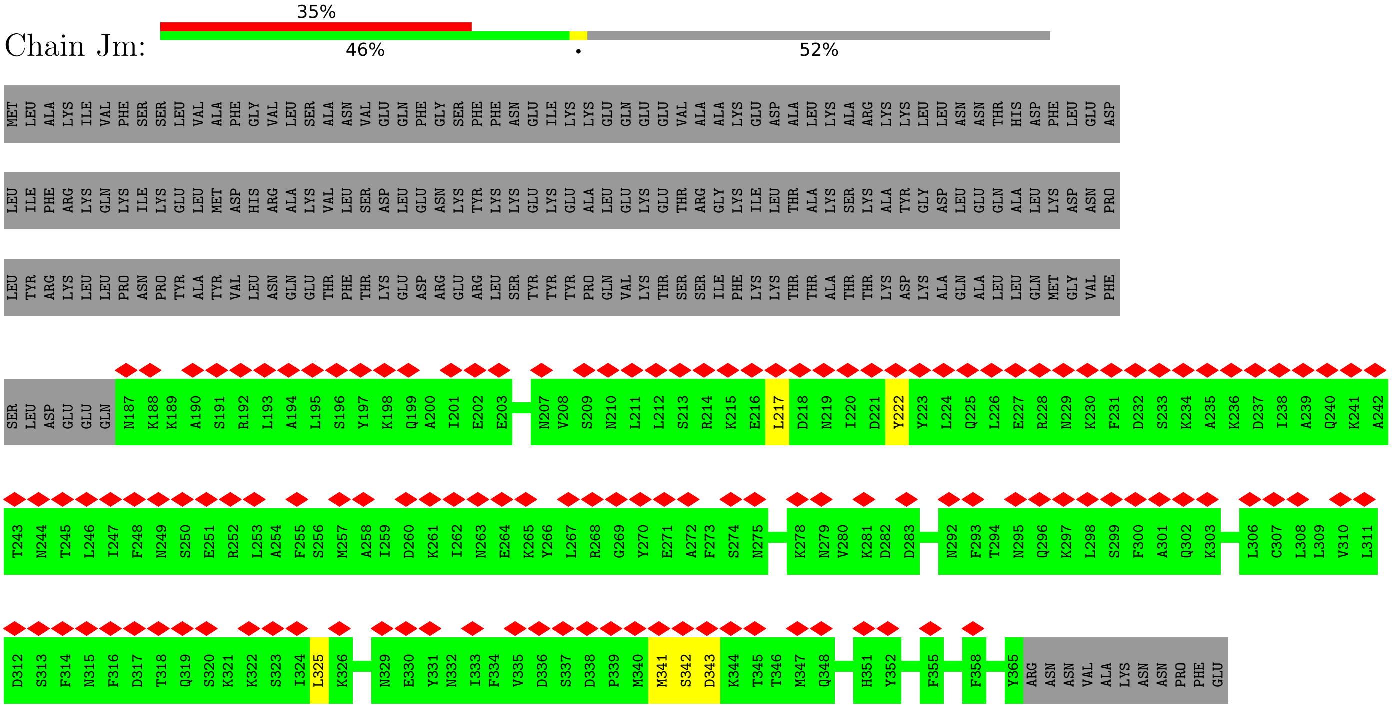
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain JM:



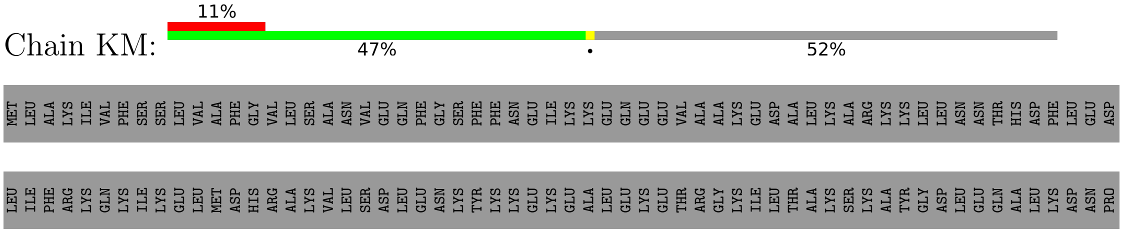
- Molecule 2: Type IV secretion system apparatus protein CagM

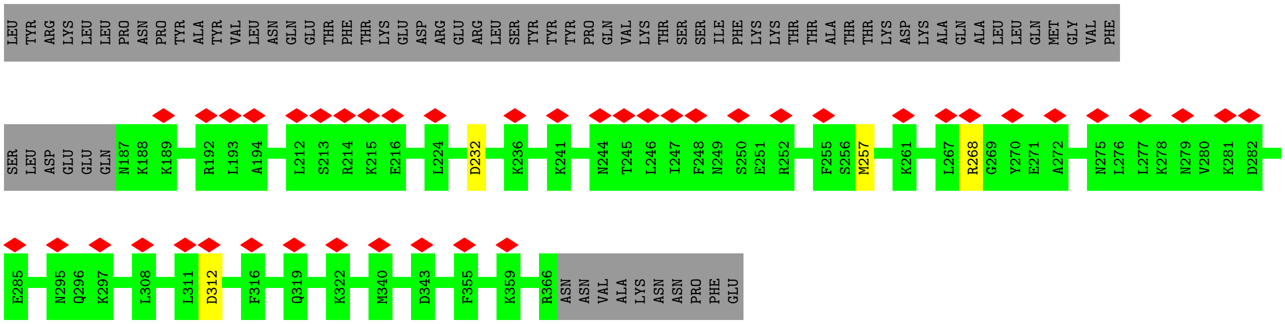
Chain Jm:



- Molecule 2: Type IV secretion system apparatus protein CagM

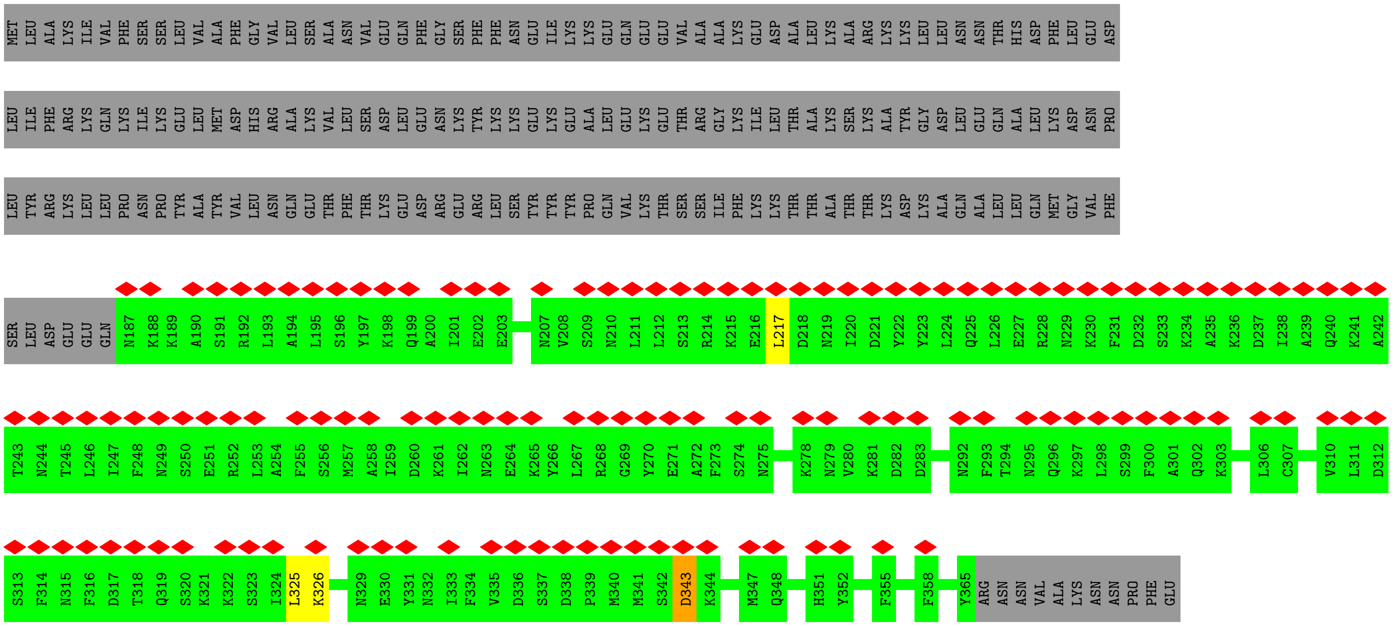
Chain KM:





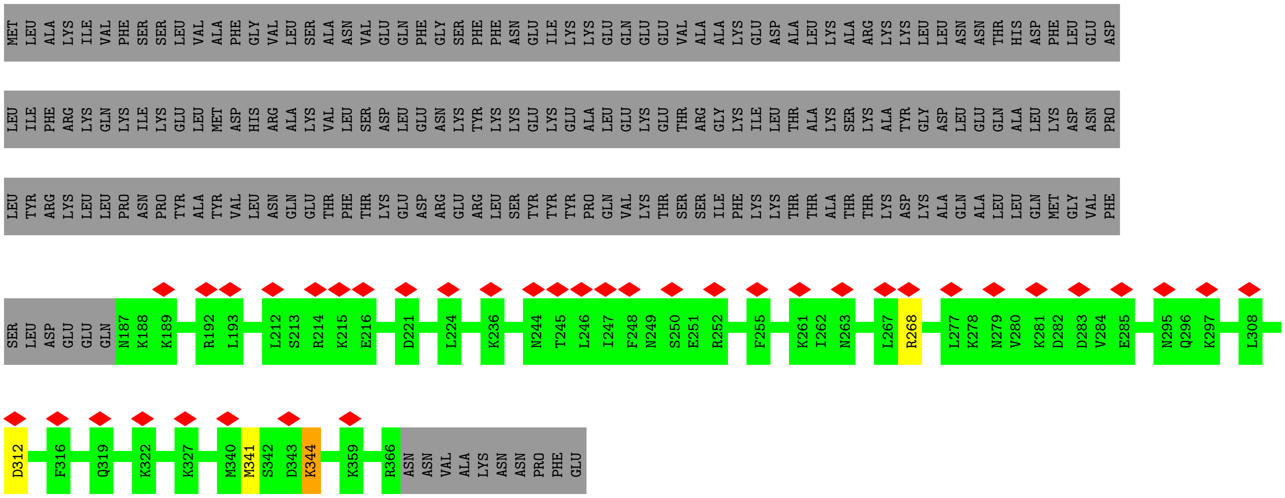
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain Km:



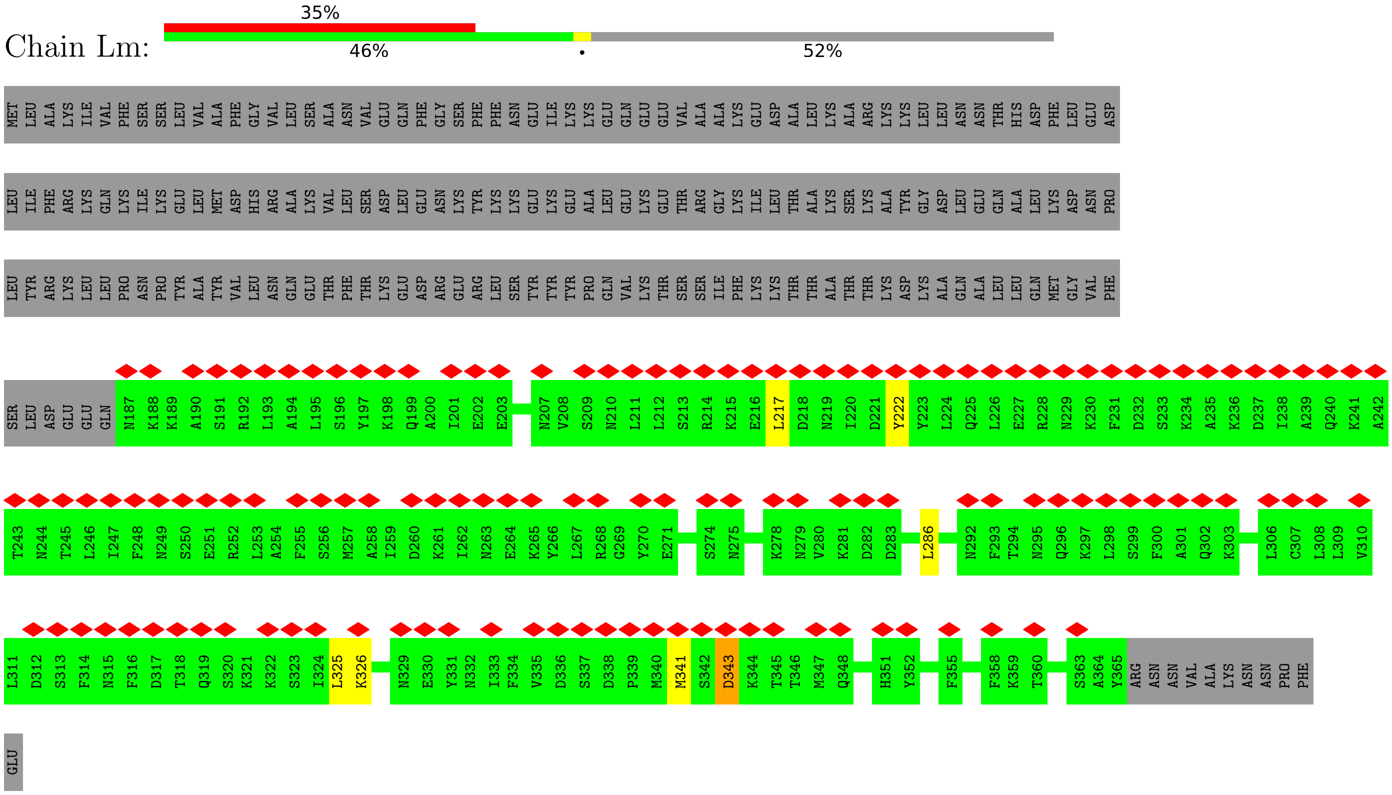
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain LM:



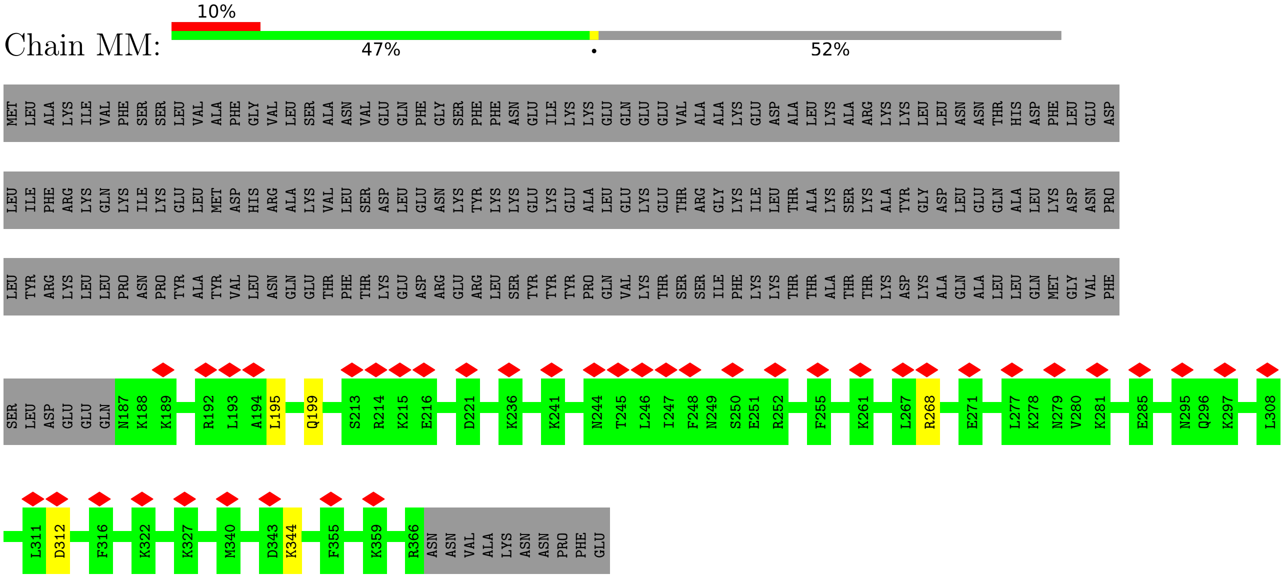
- Molecule 2: Type IV secretion system apparatus protein CagM

Chain Lm:



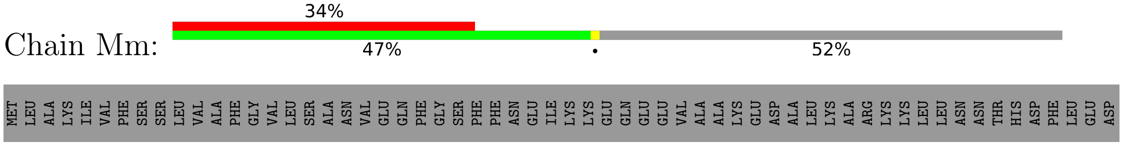
- Molecule 2: Type IV secretion system apparatus protein CagM

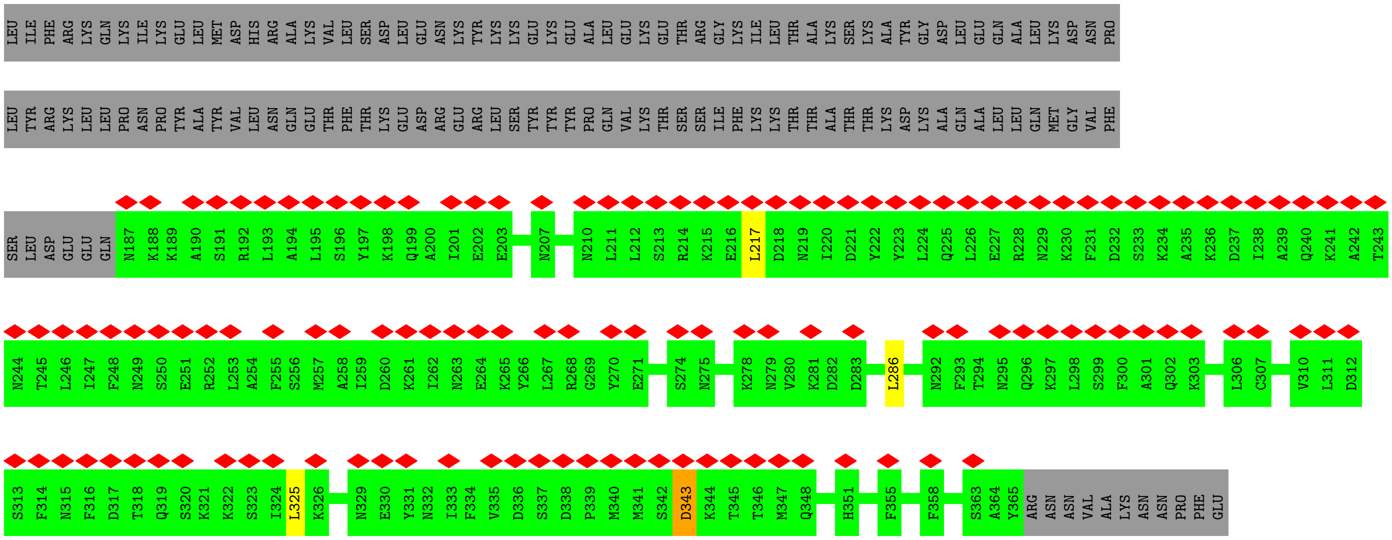
Chain MM:



- Molecule 2: Type IV secretion system apparatus protein CagM

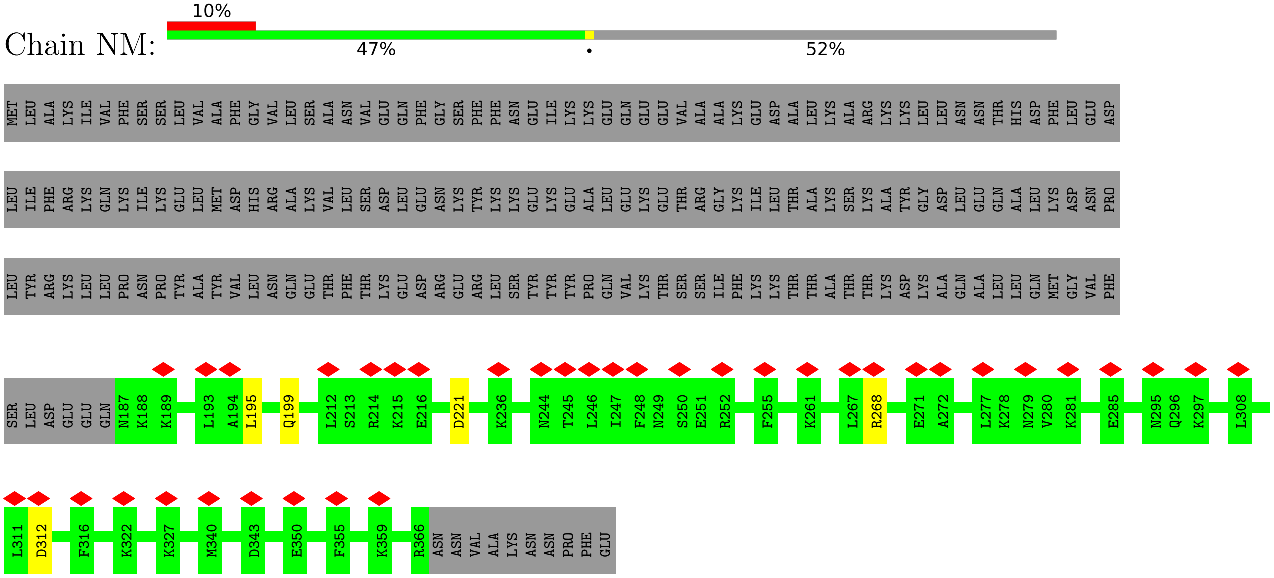
Chain Mm:





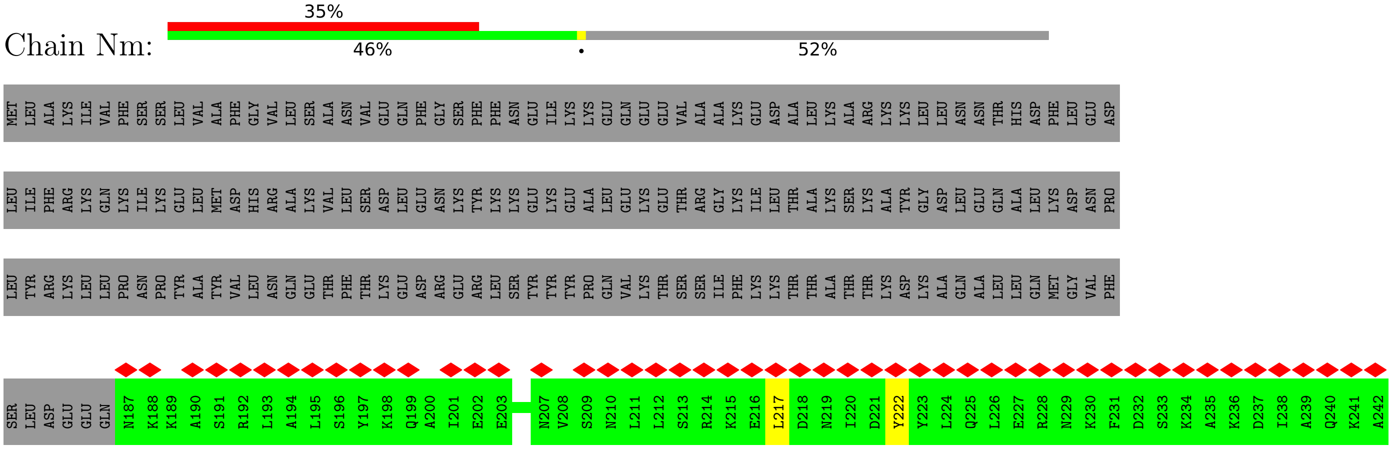
• Molecule 2: Type IV secretion system apparatus protein CagM

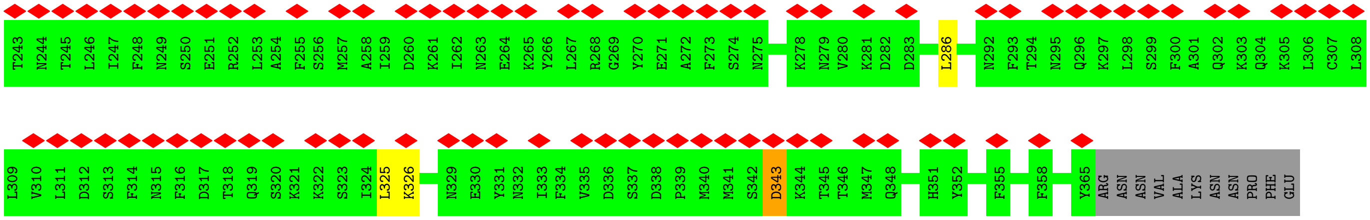
Chain NM:



• Molecule 2: Type IV secretion system apparatus protein CagM

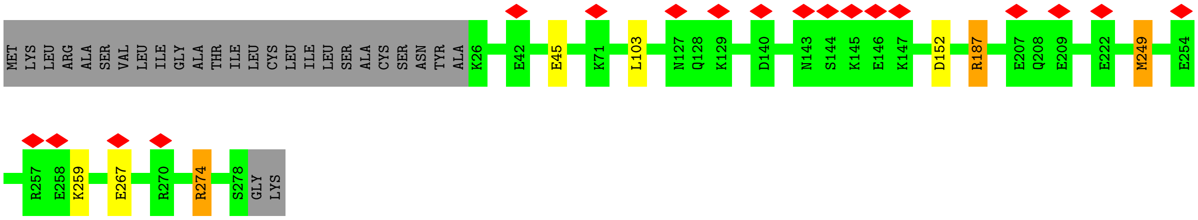
Chain Nm:





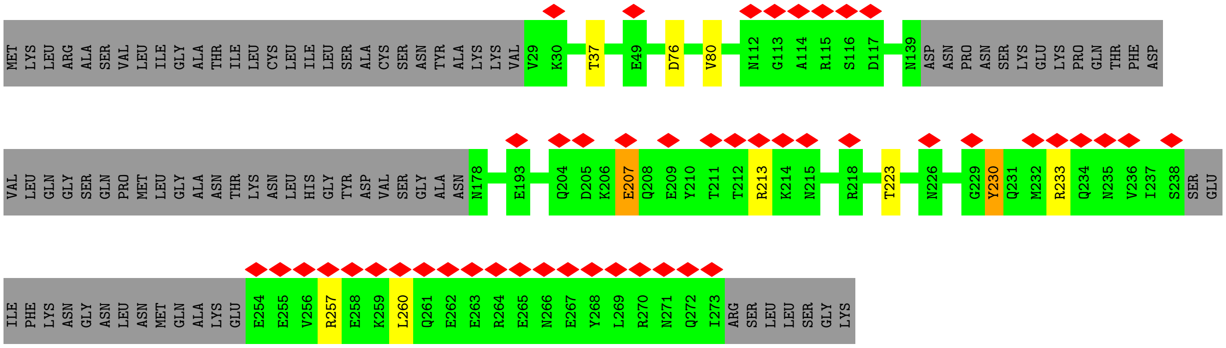
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain AT:



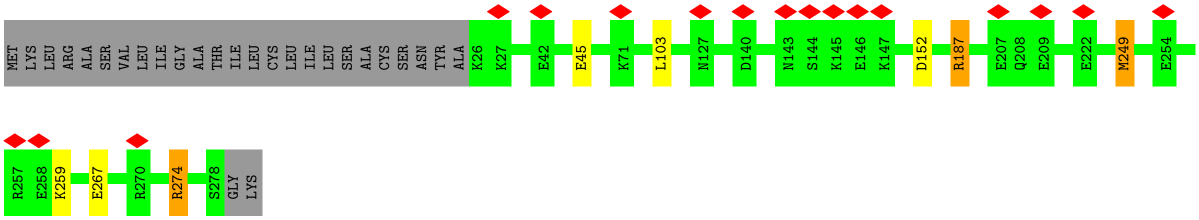
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain At:



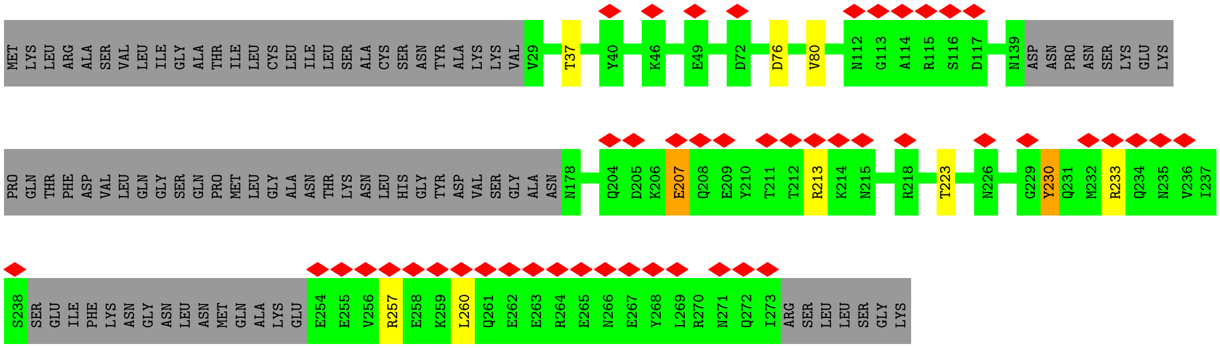
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain BT:



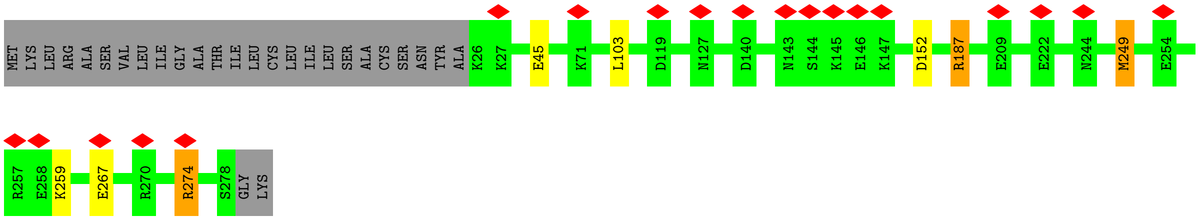
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain Bt:



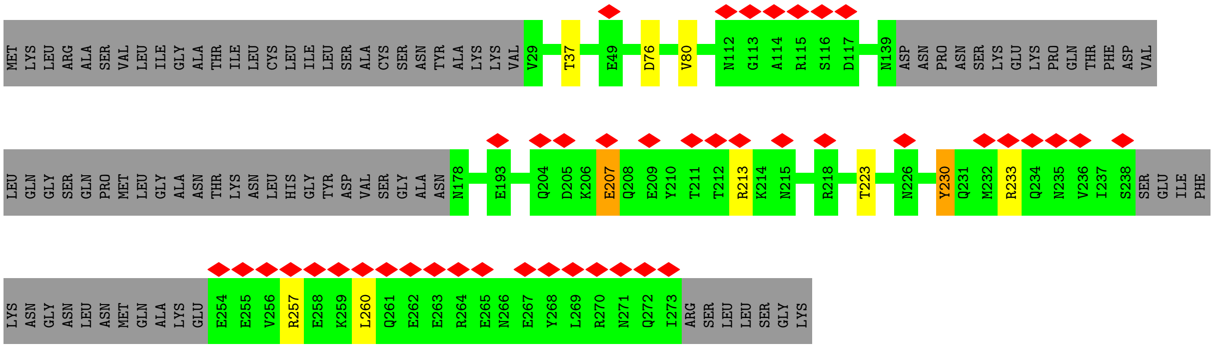
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain CT:



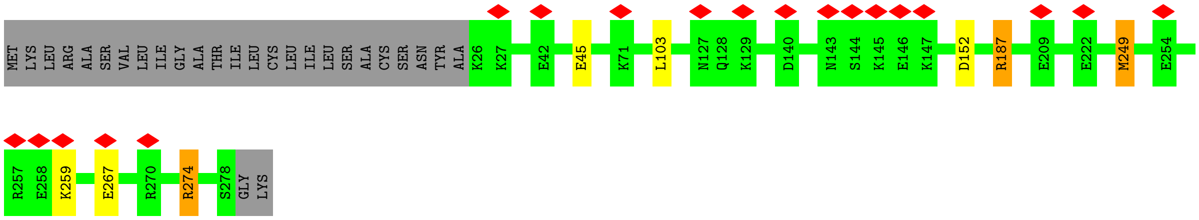
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain Ct:



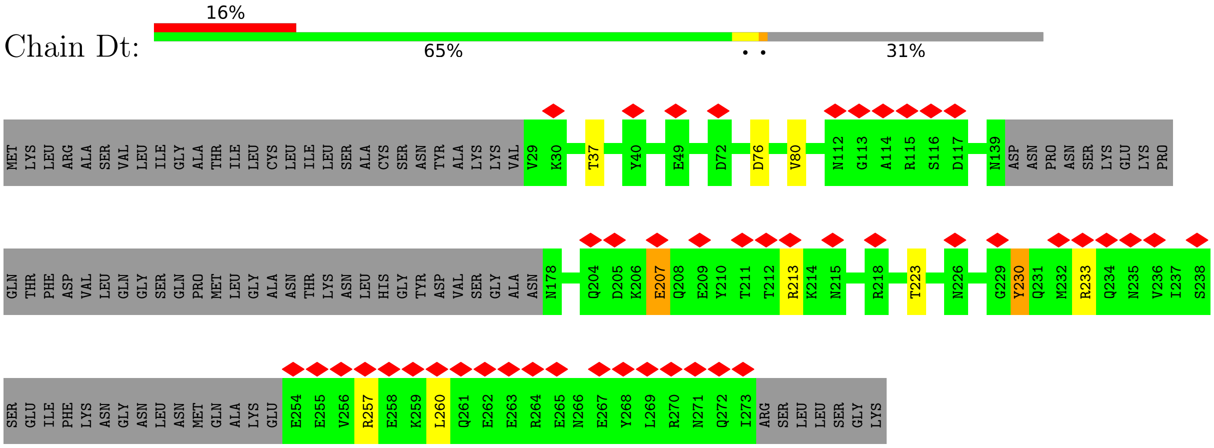
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain DT:



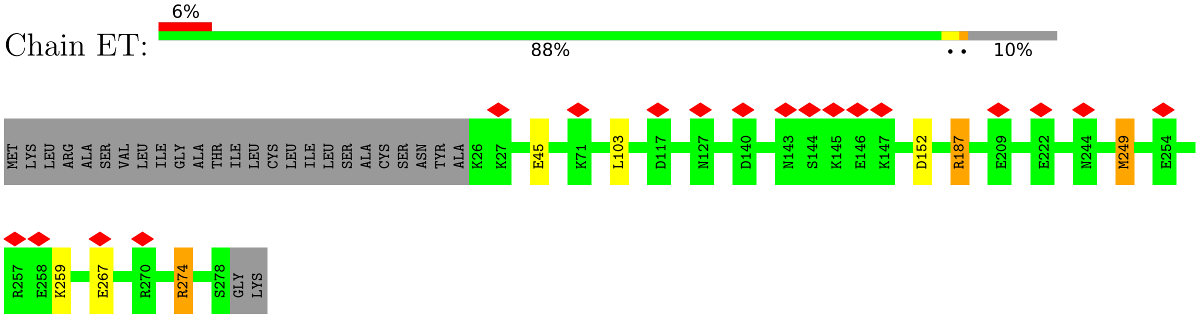
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain Dt:



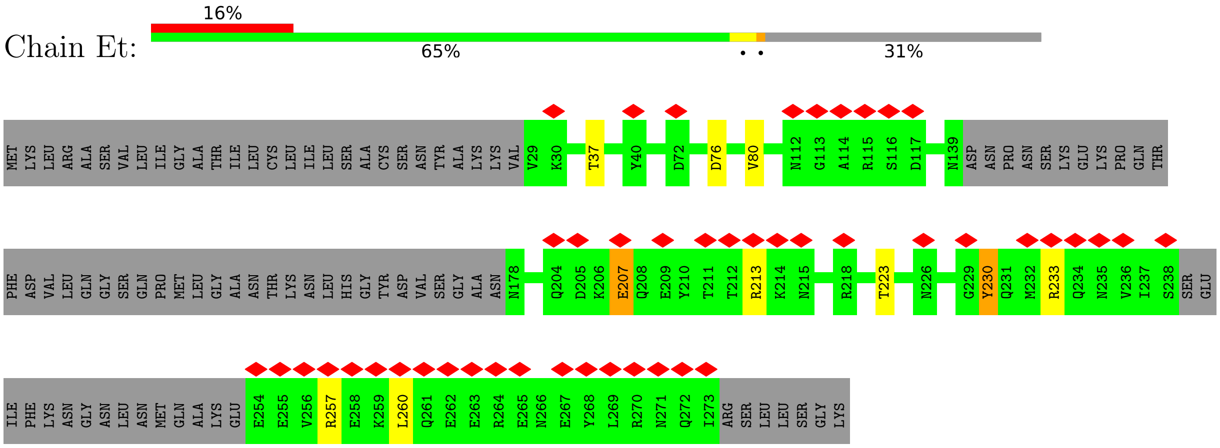
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain ET:



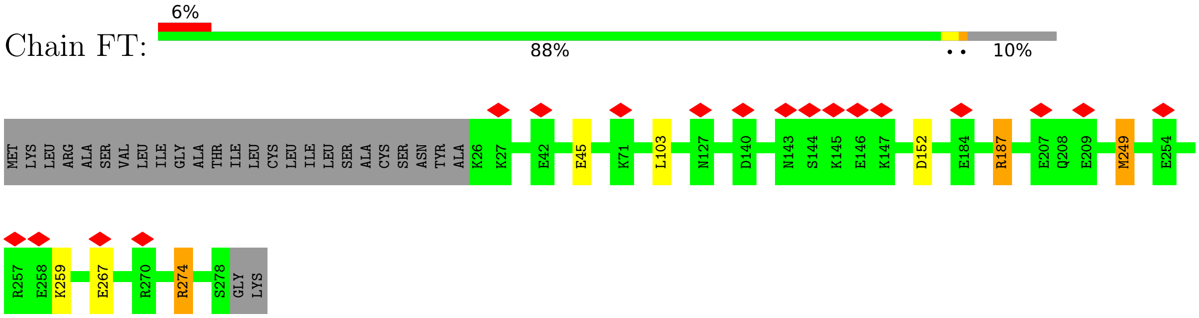
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain Et:

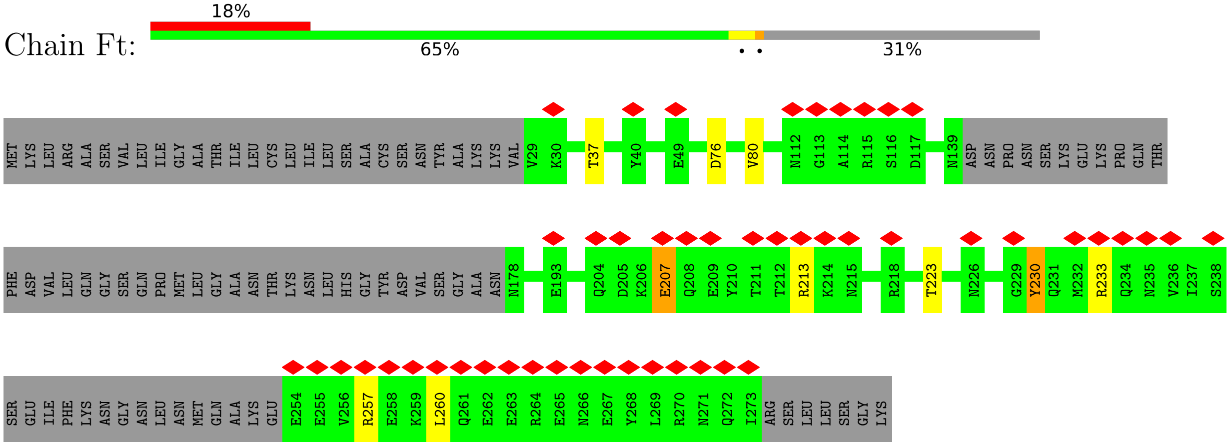


- Molecule 3: Type IV secretion system apparatus protein CagT

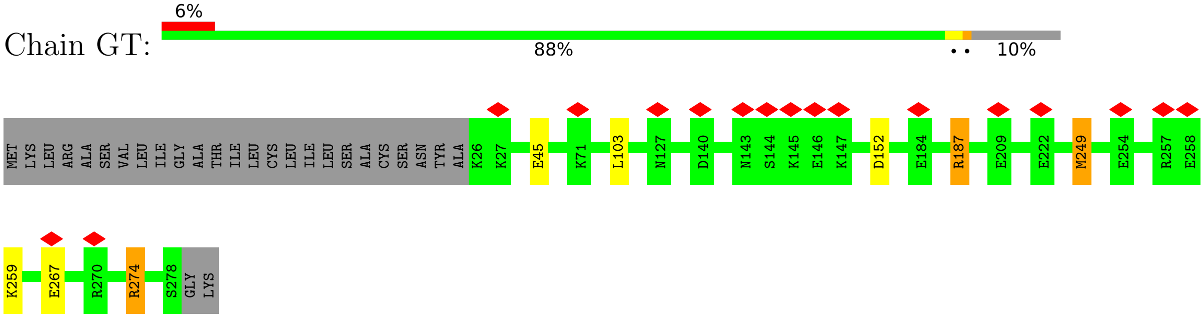
Chain FT:



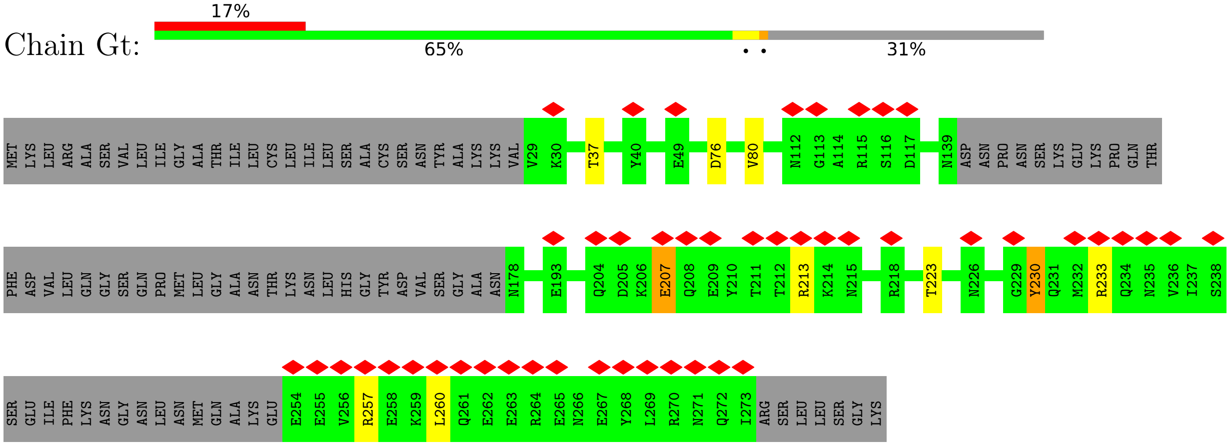
- Molecule 3: Type IV secretion system apparatus protein CagT



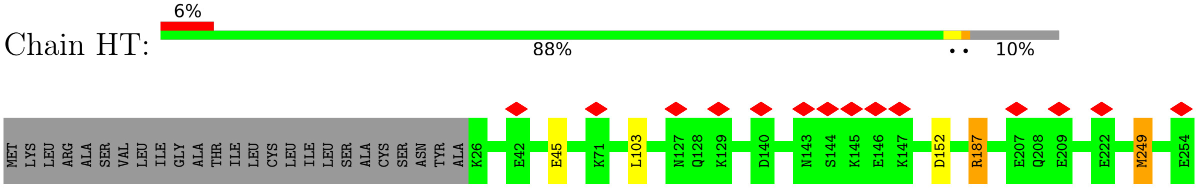
- Molecule 3: Type IV secretion system apparatus protein CagT

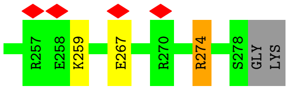


- Molecule 3: Type IV secretion system apparatus protein CagT



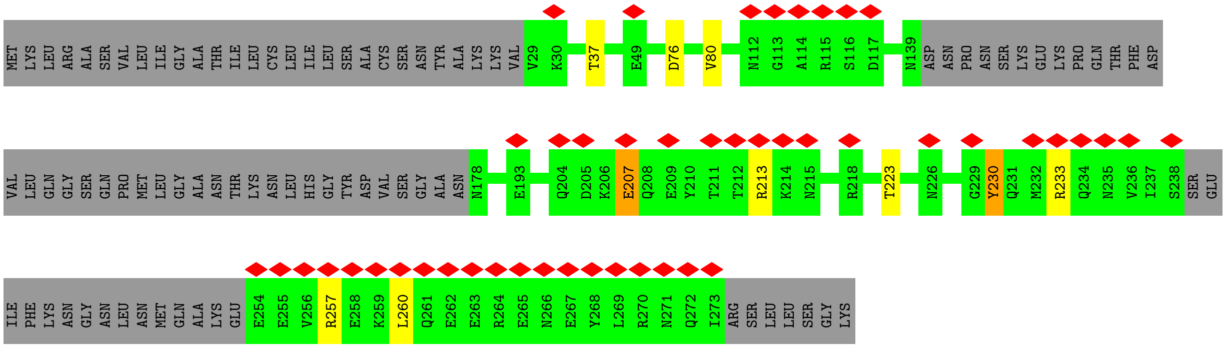
- Molecule 3: Type IV secretion system apparatus protein CagT





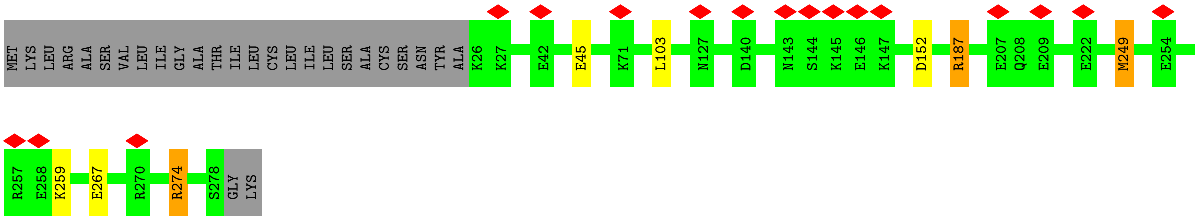
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain Ht:



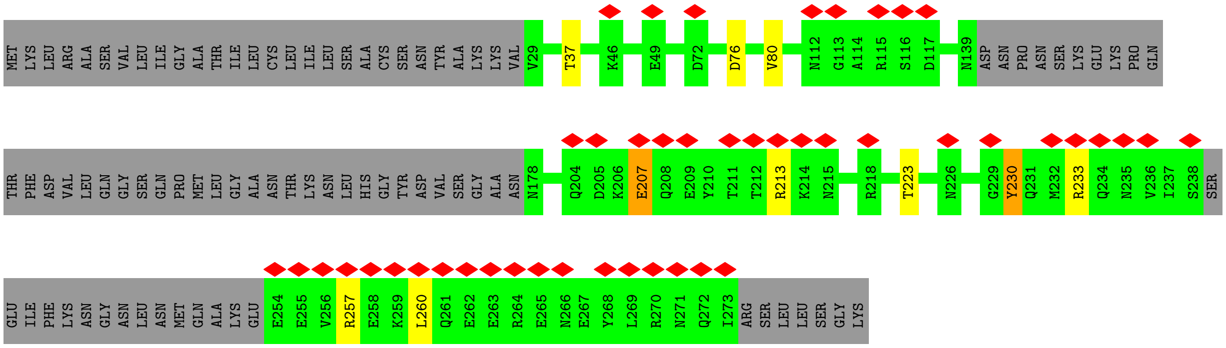
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain IT:



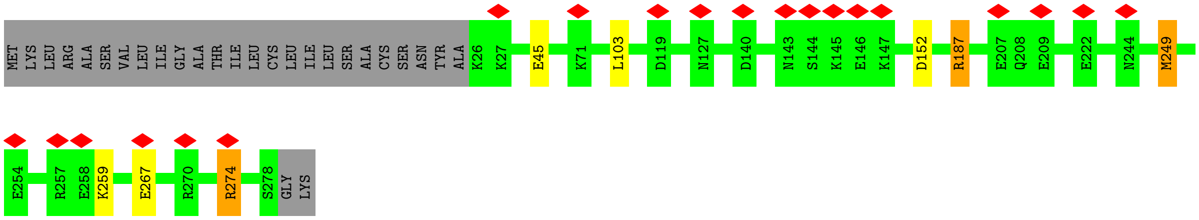
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain It:

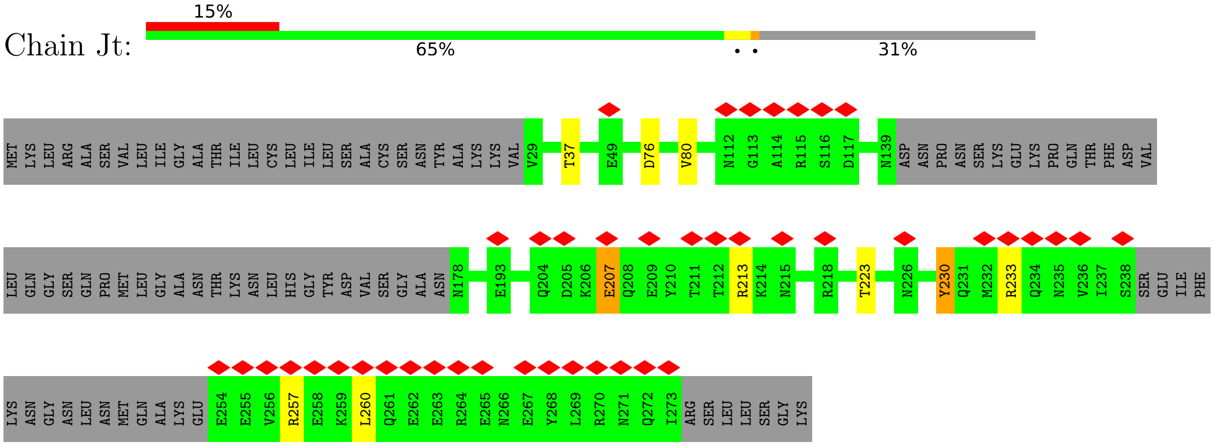


- Molecule 3: Type IV secretion system apparatus protein CagT

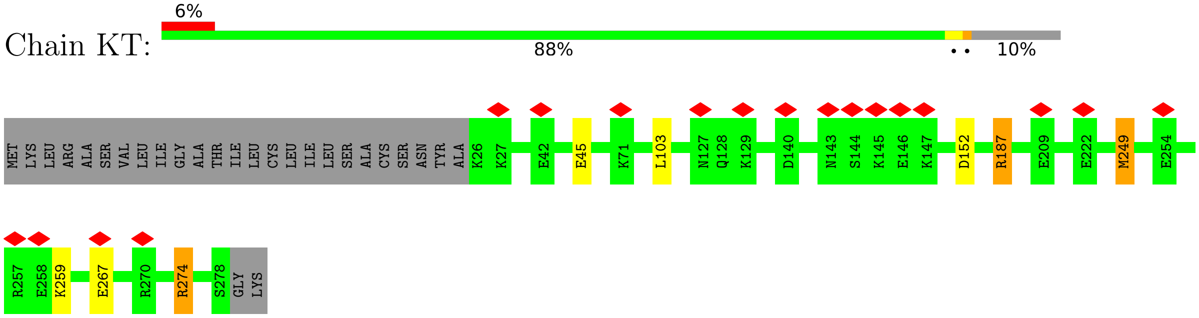
Chain JT:



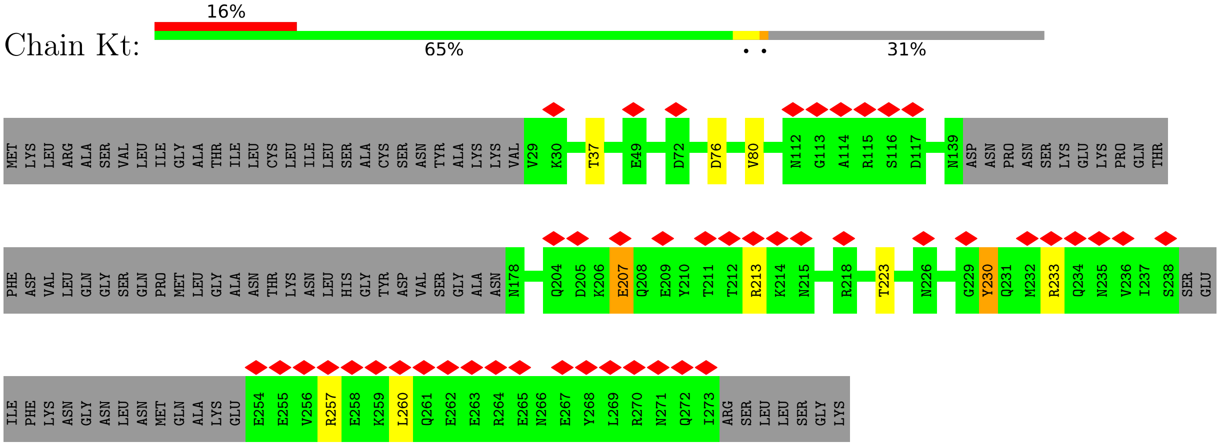
- Molecule 3: Type IV secretion system apparatus protein CagT



- Molecule 3: Type IV secretion system apparatus protein CagT

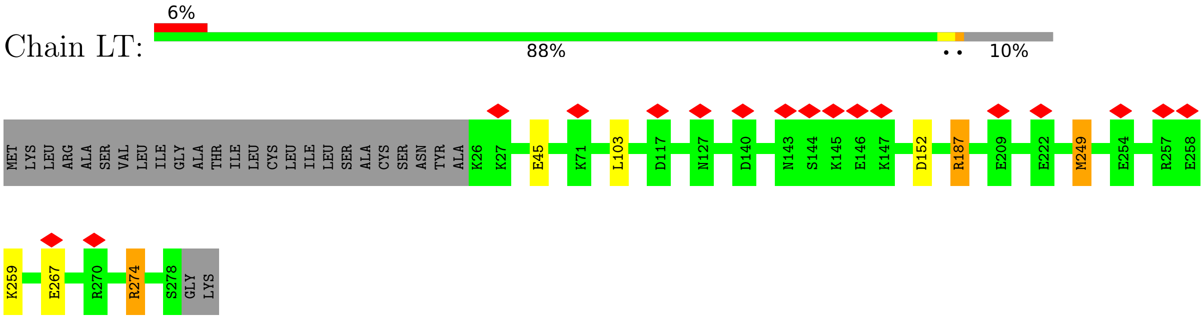


- Molecule 3: Type IV secretion system apparatus protein CagT



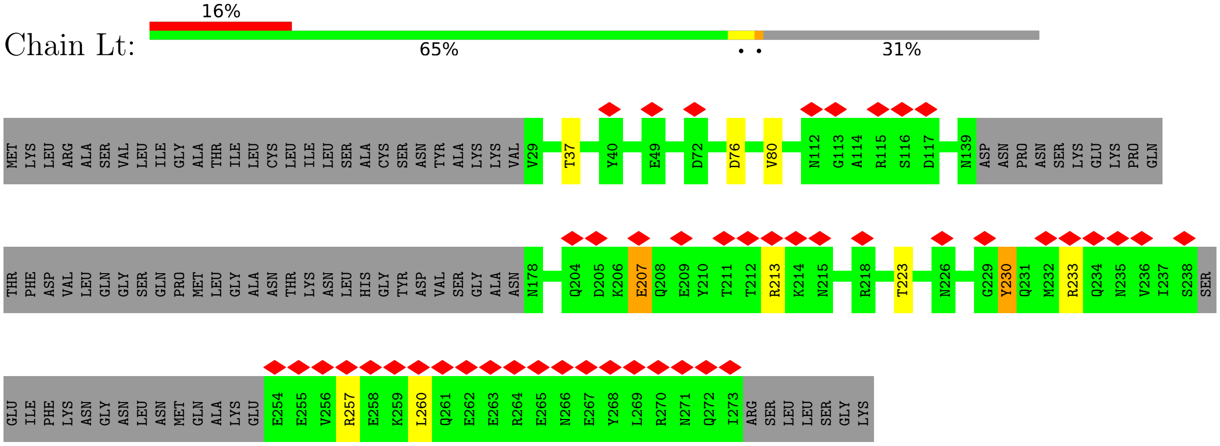
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain LT:



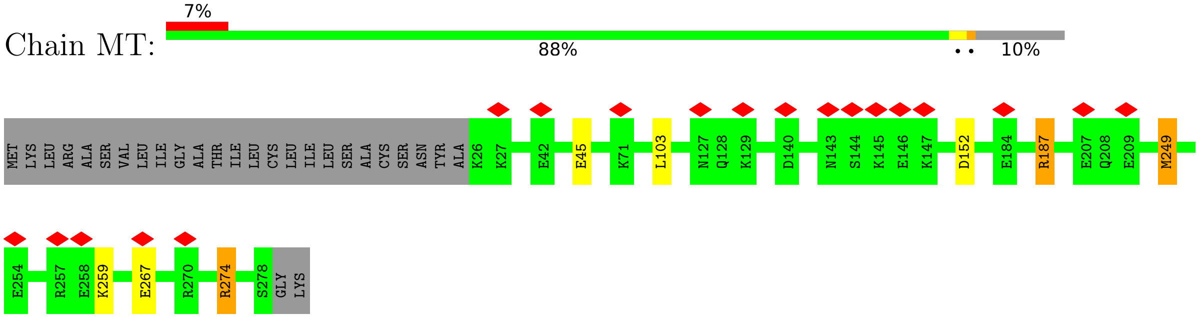
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain Lt:



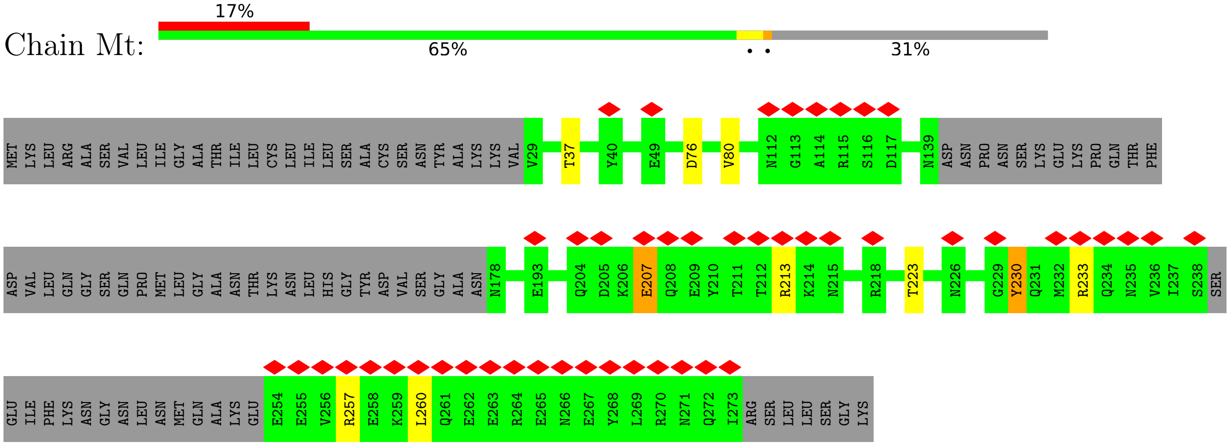
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain MT:



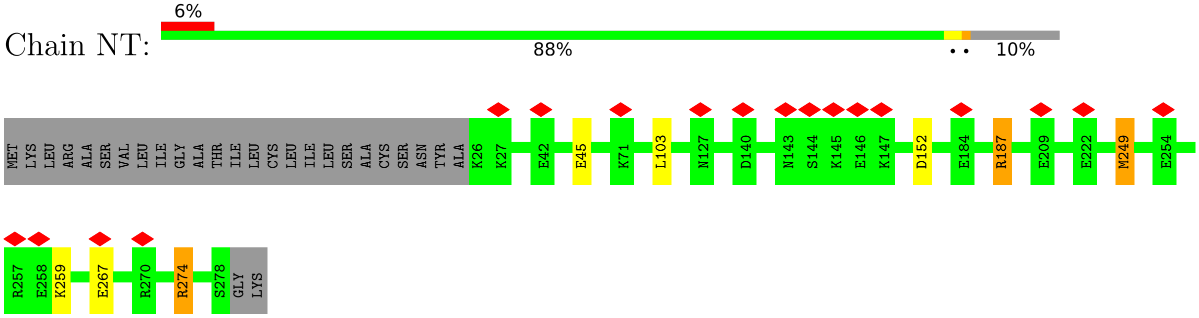
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain Mt:



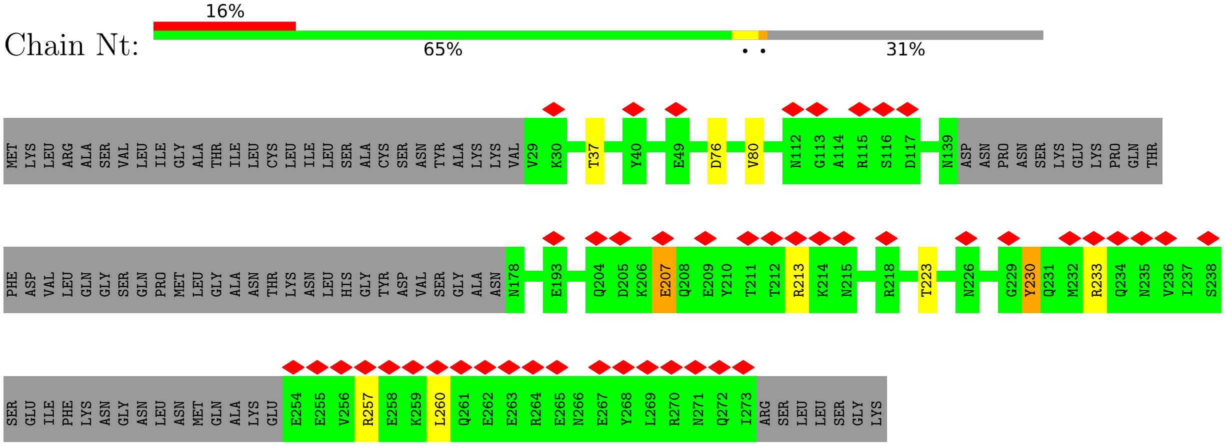
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain NT:



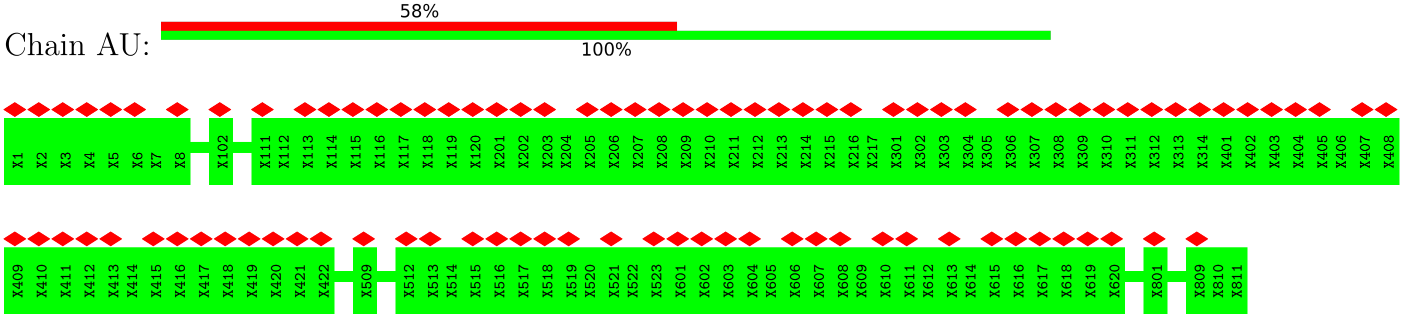
- Molecule 3: Type IV secretion system apparatus protein CagT

Chain Nt:



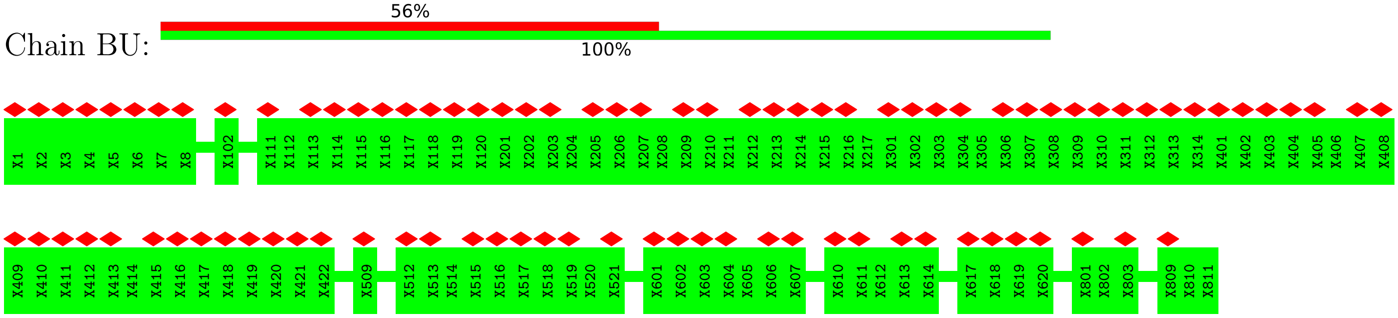
- Molecule 4: Unknown protein fragment

Chain AU:



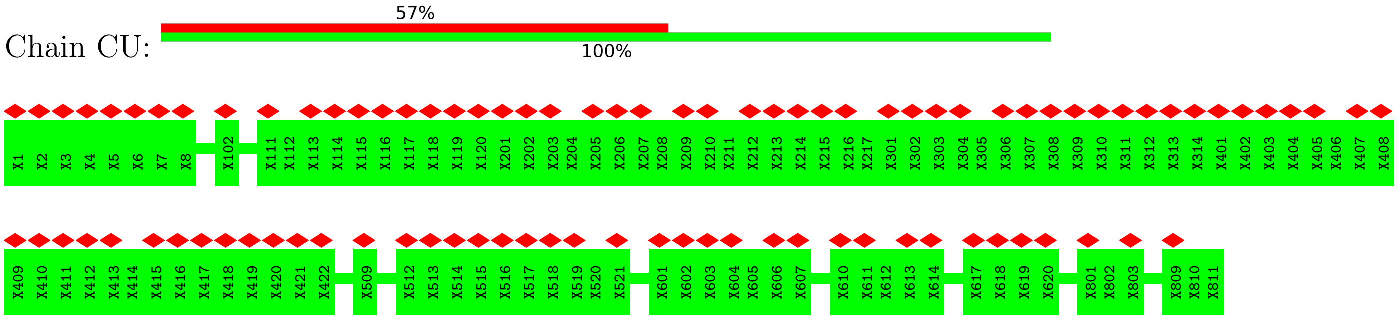
- Molecule 4: Unknown protein fragment

Chain BU:



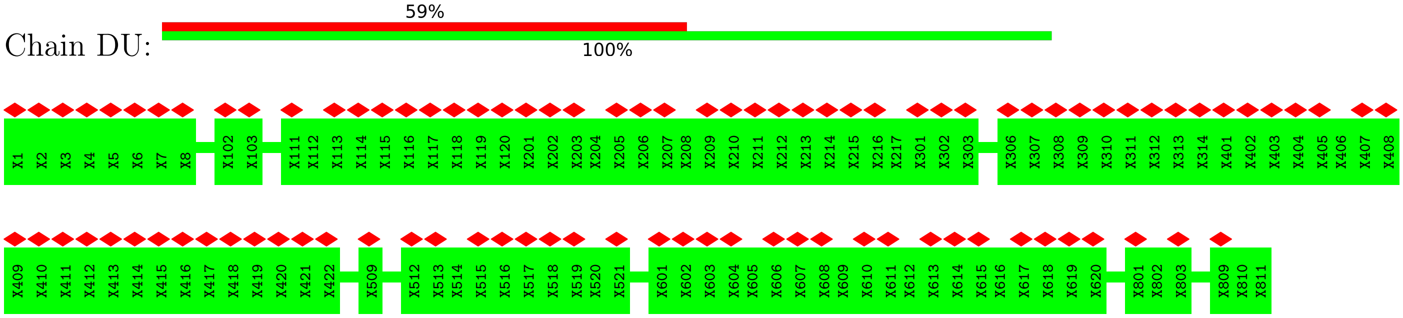
- Molecule 4: Unknown protein fragment

Chain CU:



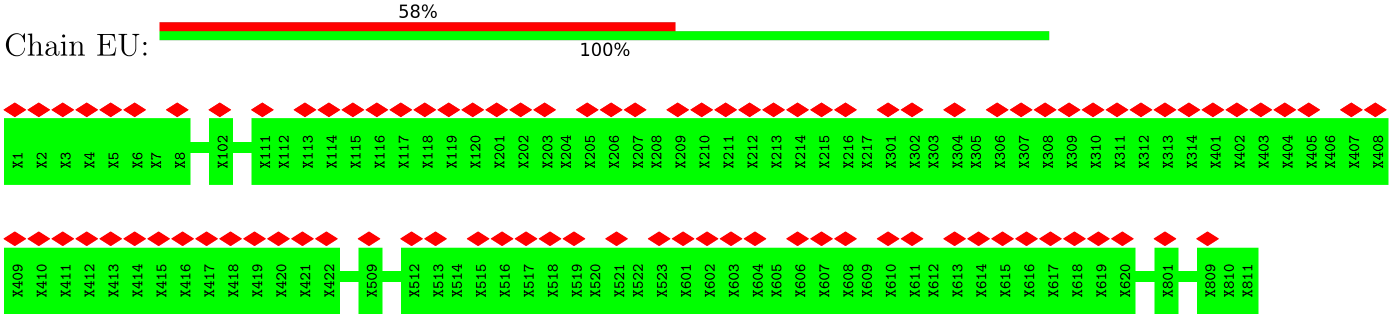
- Molecule 4: Unknown protein fragment

Chain DU:



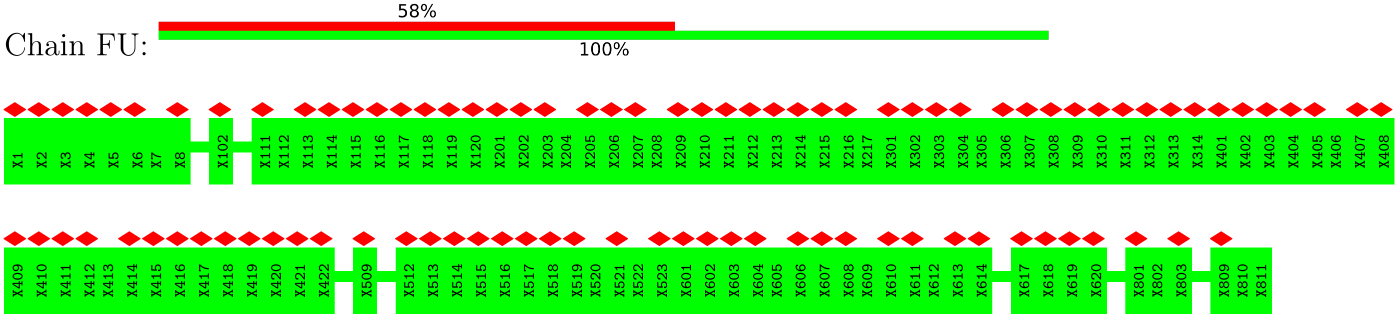
- Molecule 4: Unknown protein fragment

Chain EU:



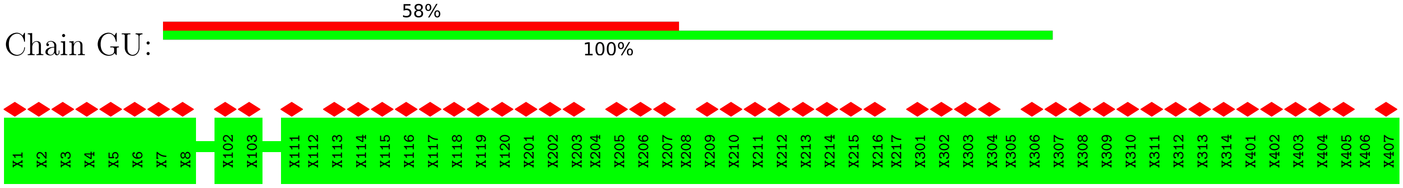
- Molecule 4: Unknown protein fragment

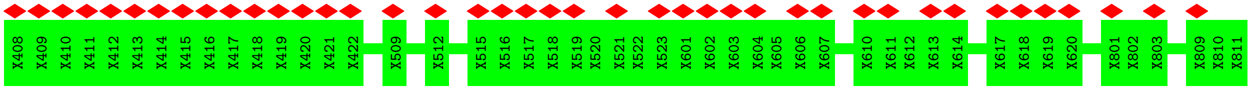
Chain FU:



- Molecule 4: Unknown protein fragment

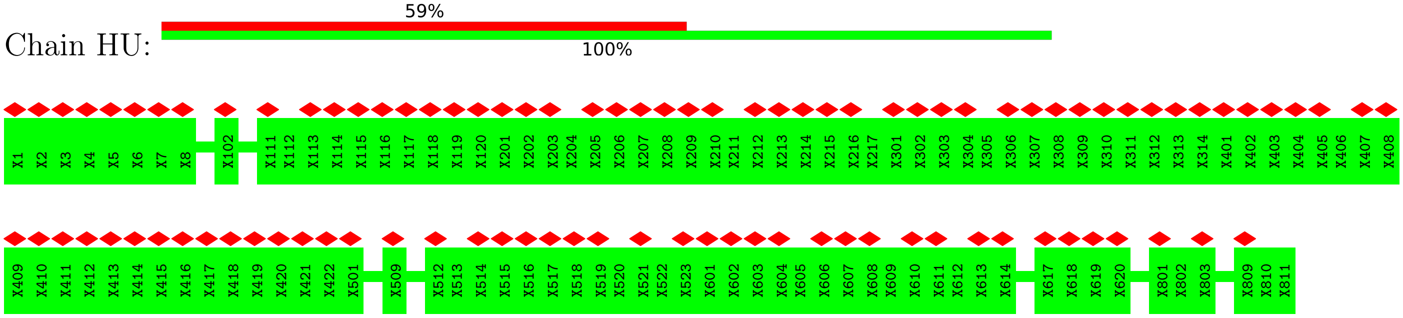
Chain GU:





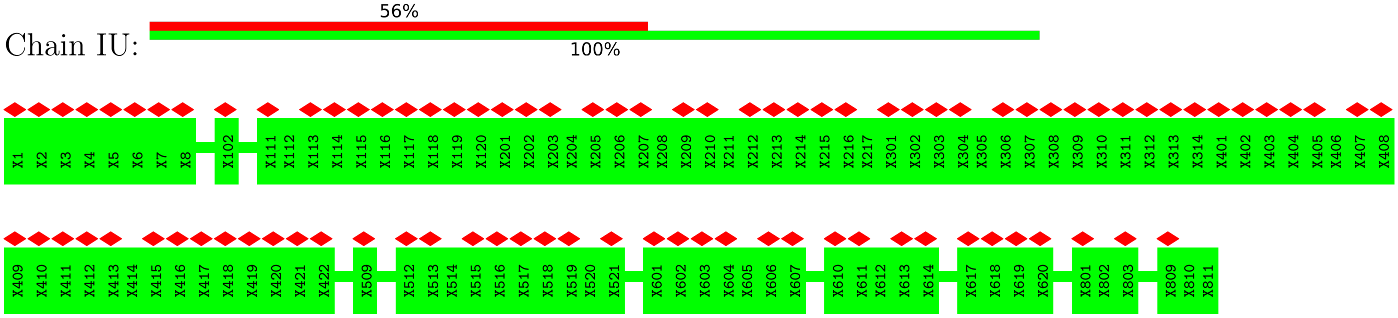
- Molecule 4: Unknown protein fragment

Chain HU:



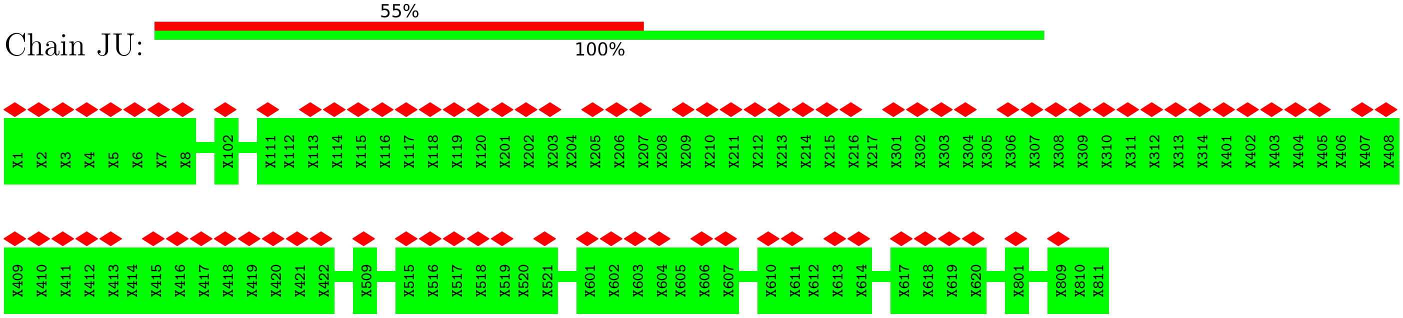
- Molecule 4: Unknown protein fragment

Chain IU:



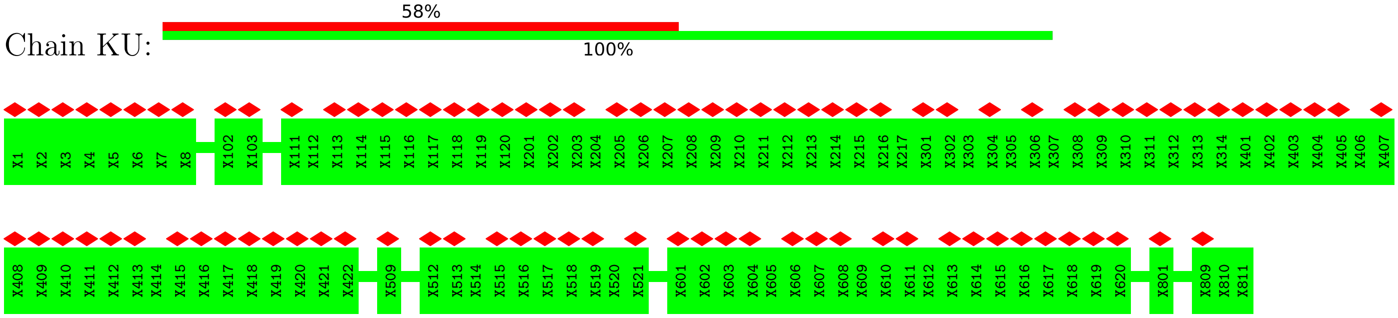
- Molecule 4: Unknown protein fragment

Chain JU:



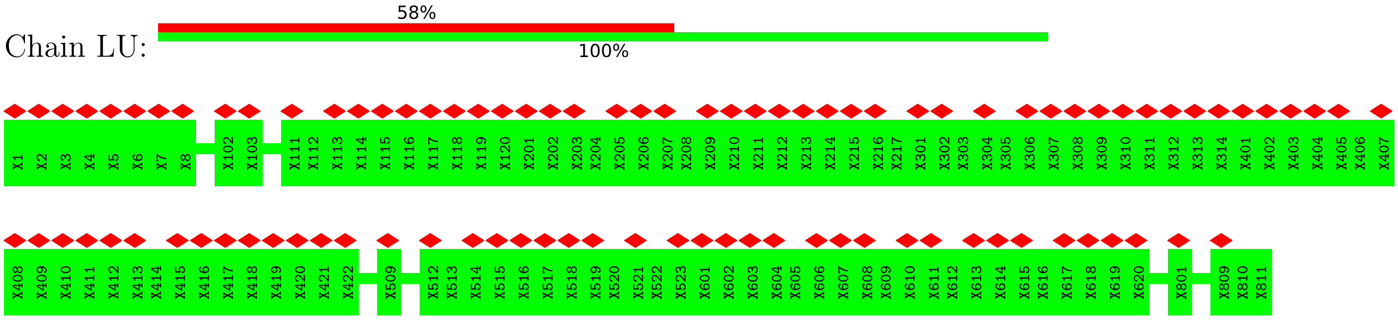
- Molecule 4: Unknown protein fragment

Chain KU:



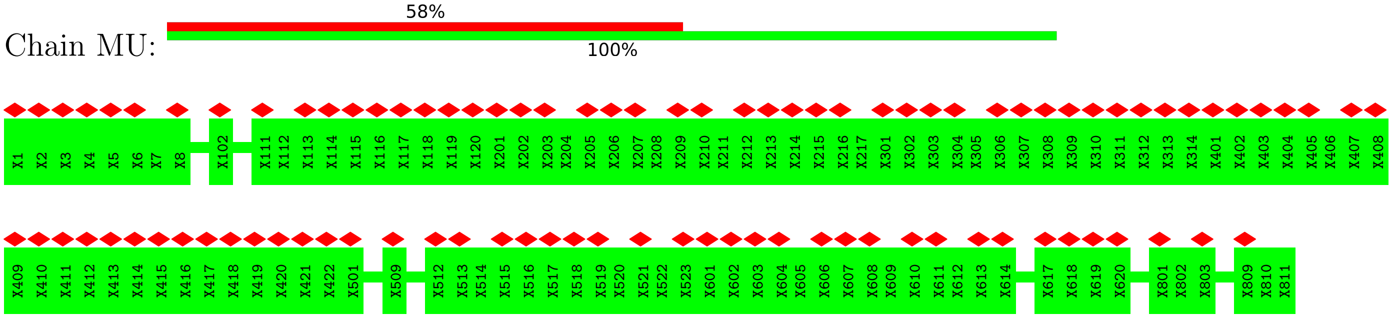
- Molecule 4: Unknown protein fragment

Chain LU:



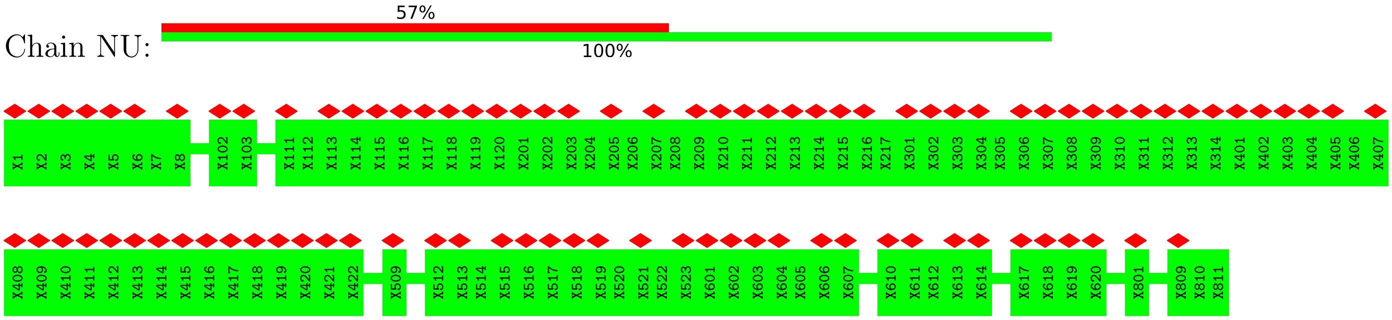
- Molecule 4: Unknown protein fragment

Chain MU:



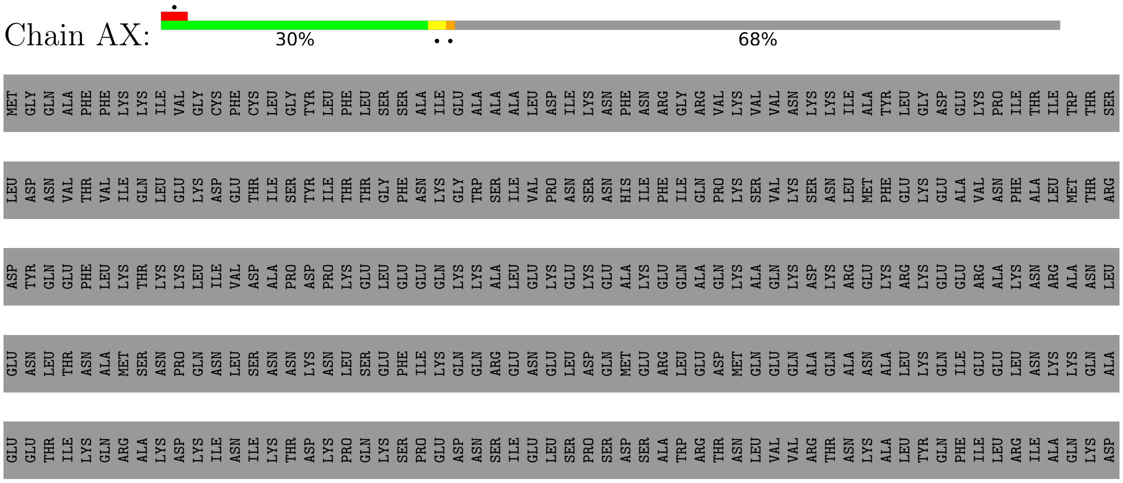
- Molecule 4: Unknown protein fragment

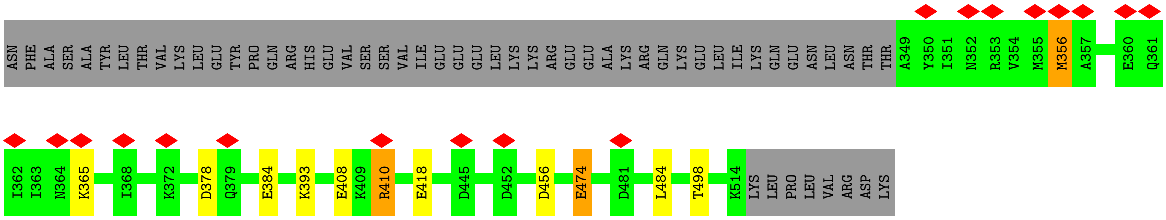
Chain NU:



- Molecule 5: Type IV secretion system apparatus protein CagX

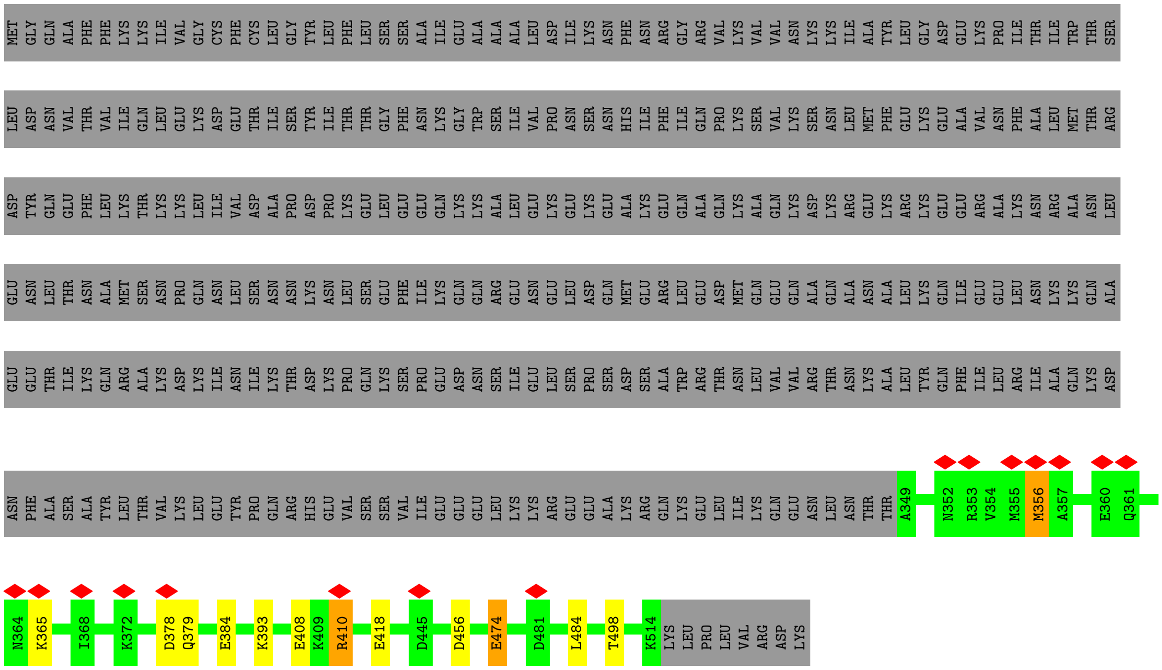
Chain AX:





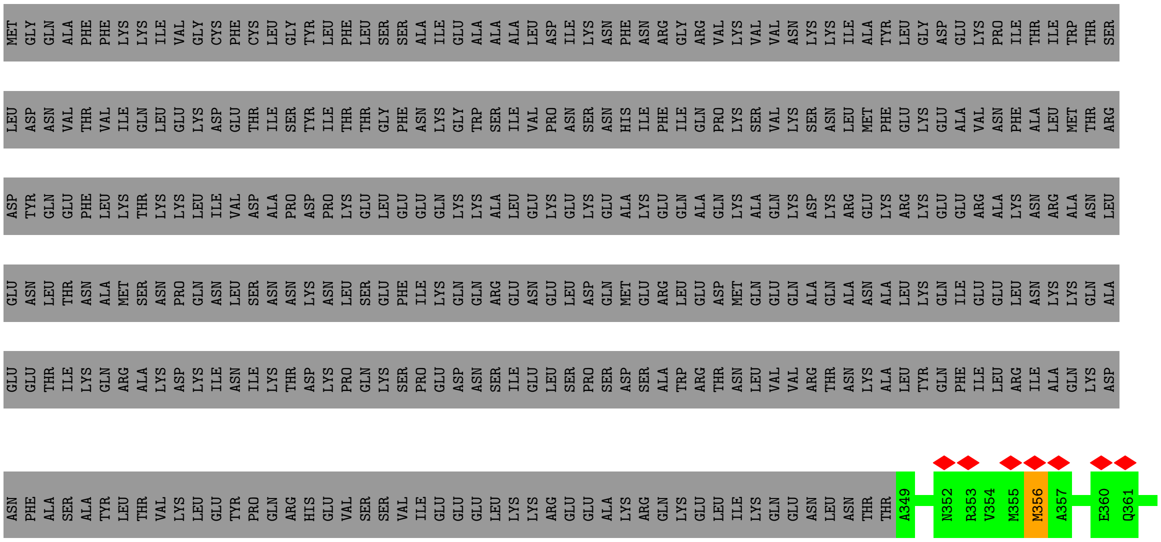
- Molecule 5: Type IV secretion system apparatus protein CagX

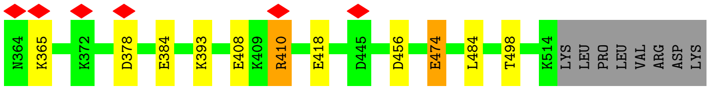
Chain BX:



- Molecule 5: Type IV secretion system apparatus protein CagX

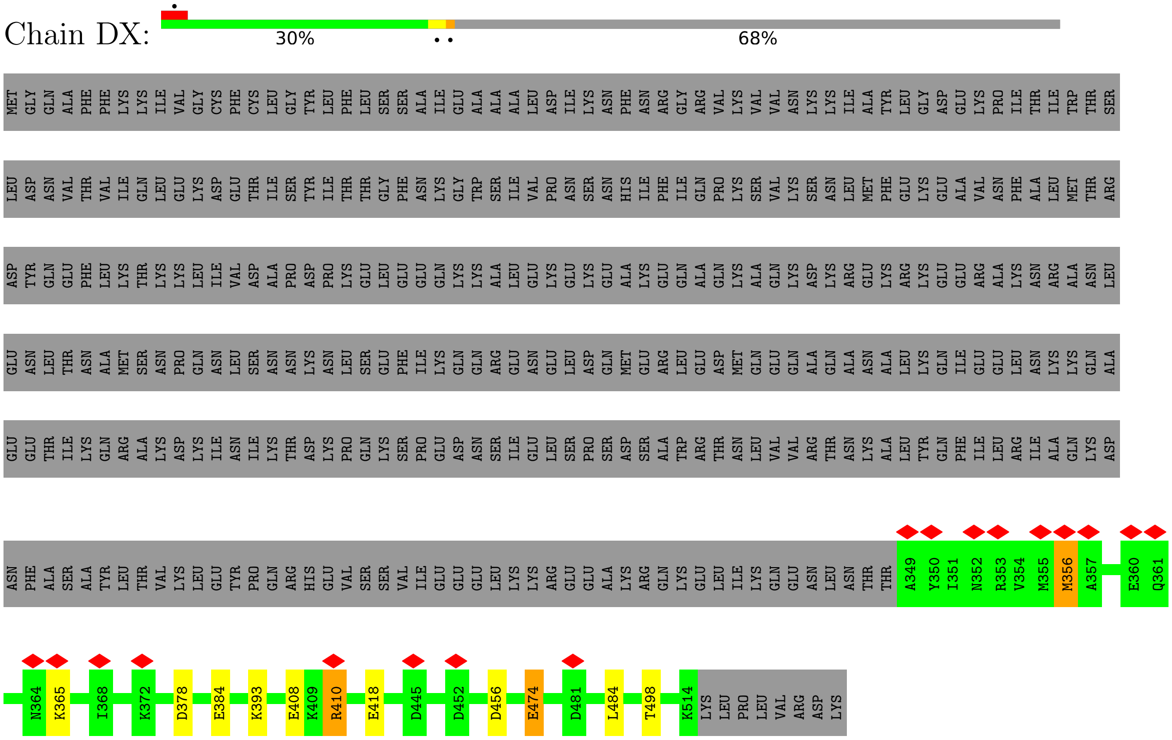
Chain CX:





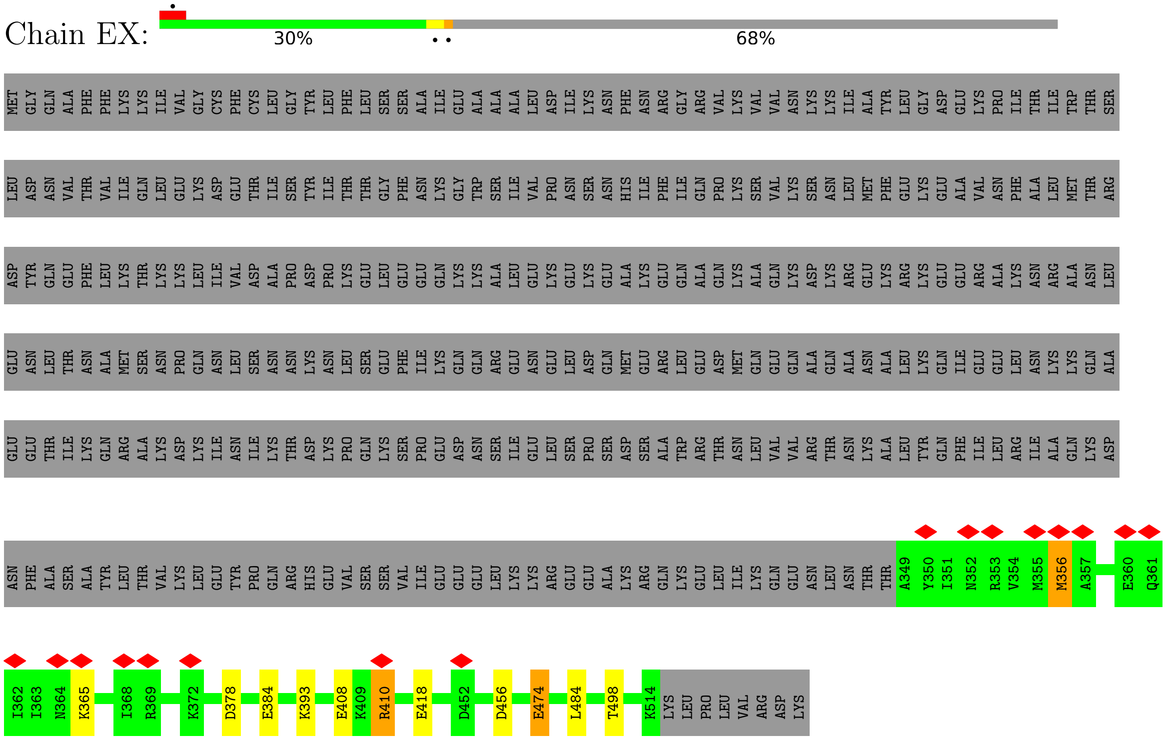
- Molecule 5: Type IV secretion system apparatus protein CagX

Chain DX:



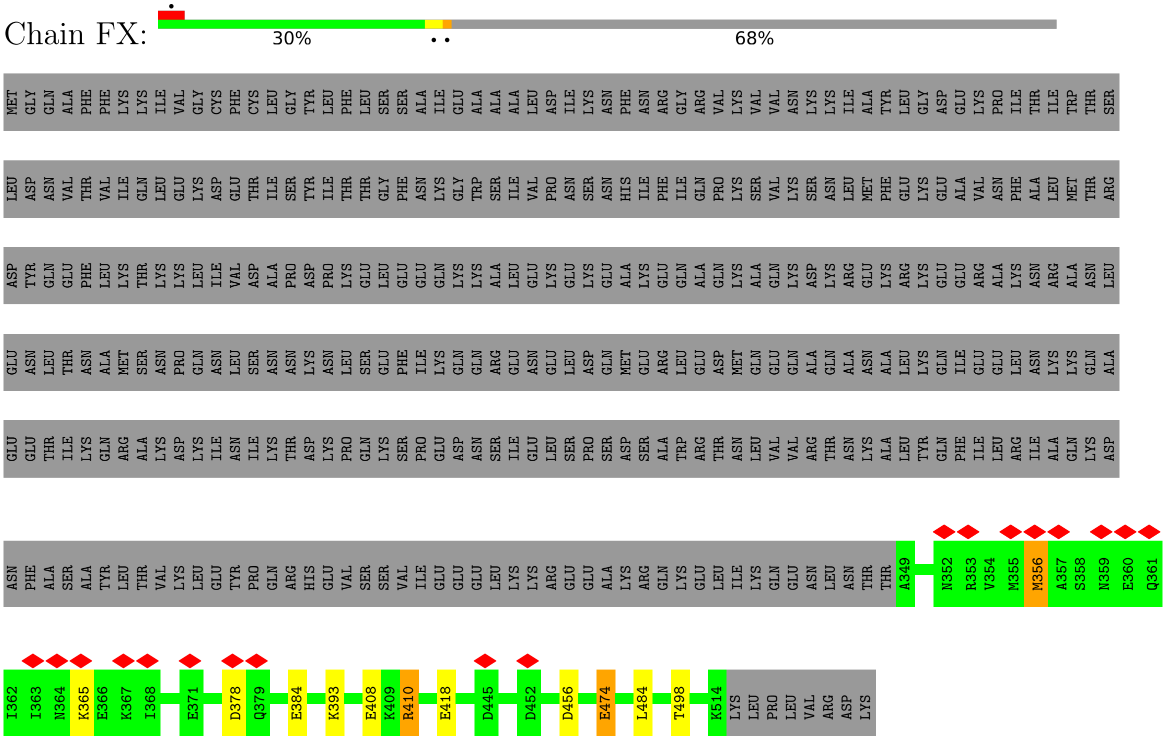
- Molecule 5: Type IV secretion system apparatus protein CagX

Chain EX:



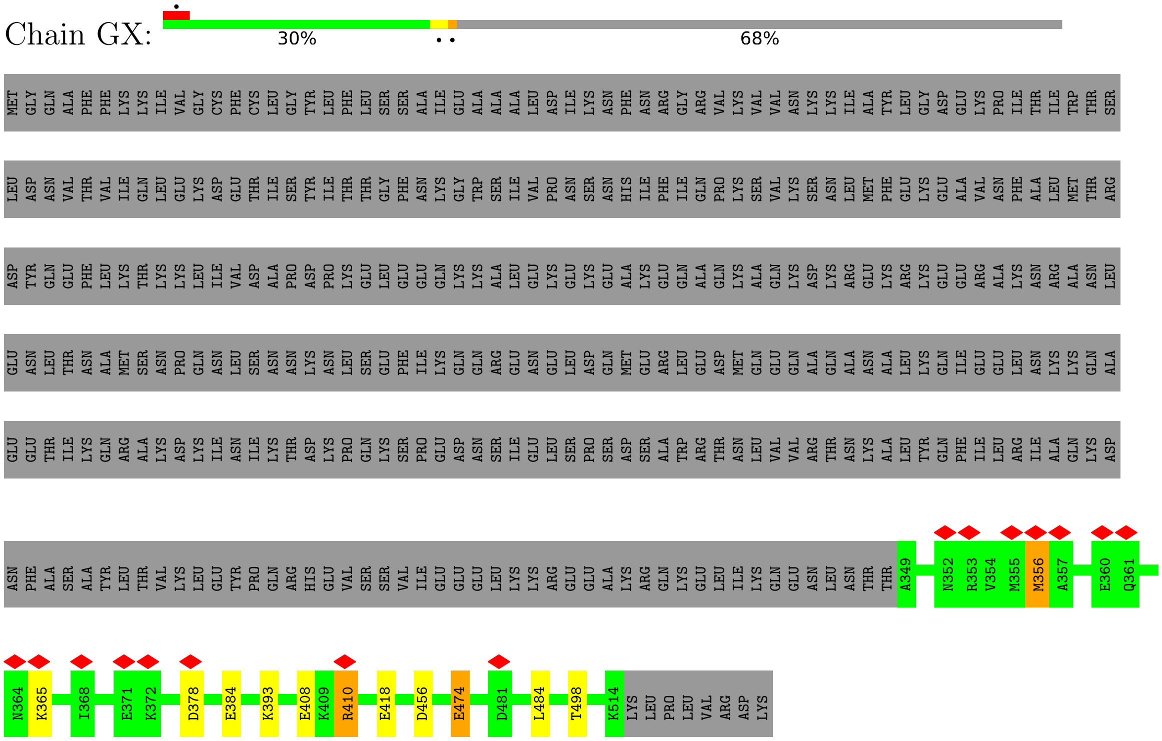
- Molecule 5: Type IV secretion system apparatus protein CagX

Chain FX:



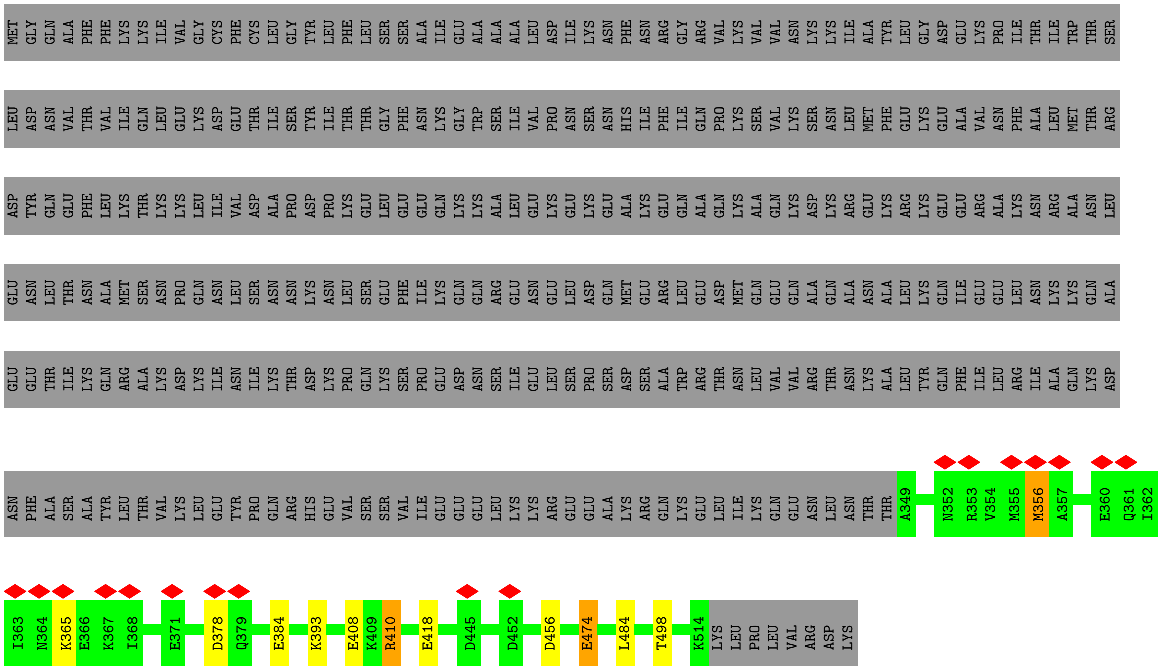
- Molecule 5: Type IV secretion system apparatus protein CagX

Chain GX:



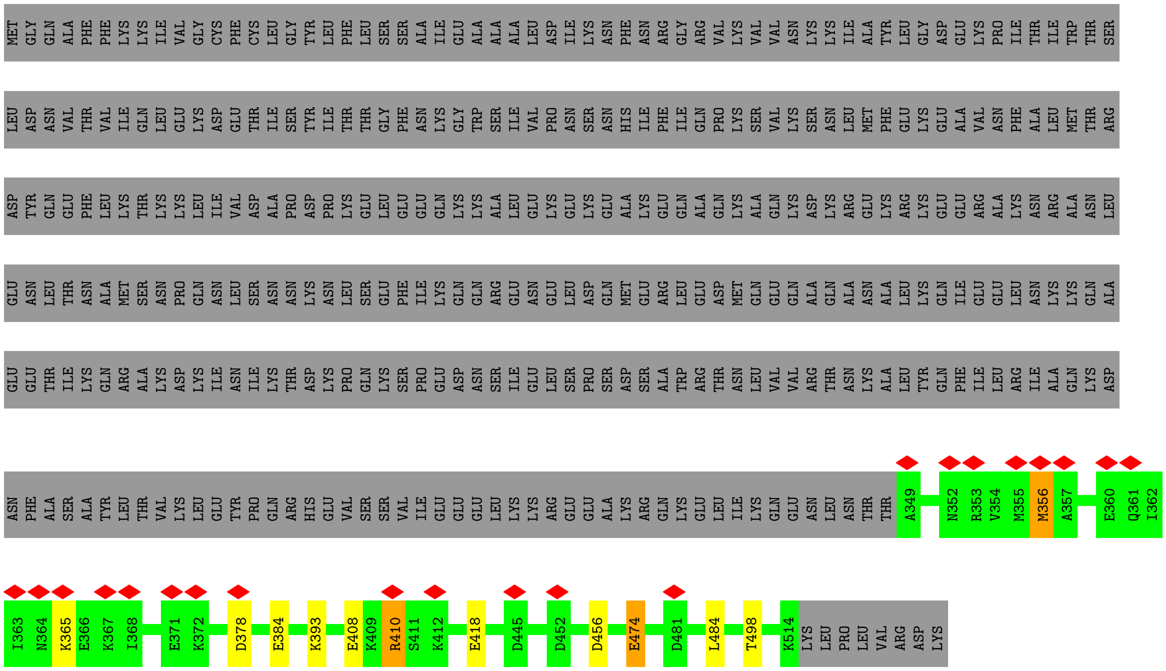
- Molecule 5: Type IV secretion system apparatus protein CagX

Chain HX:



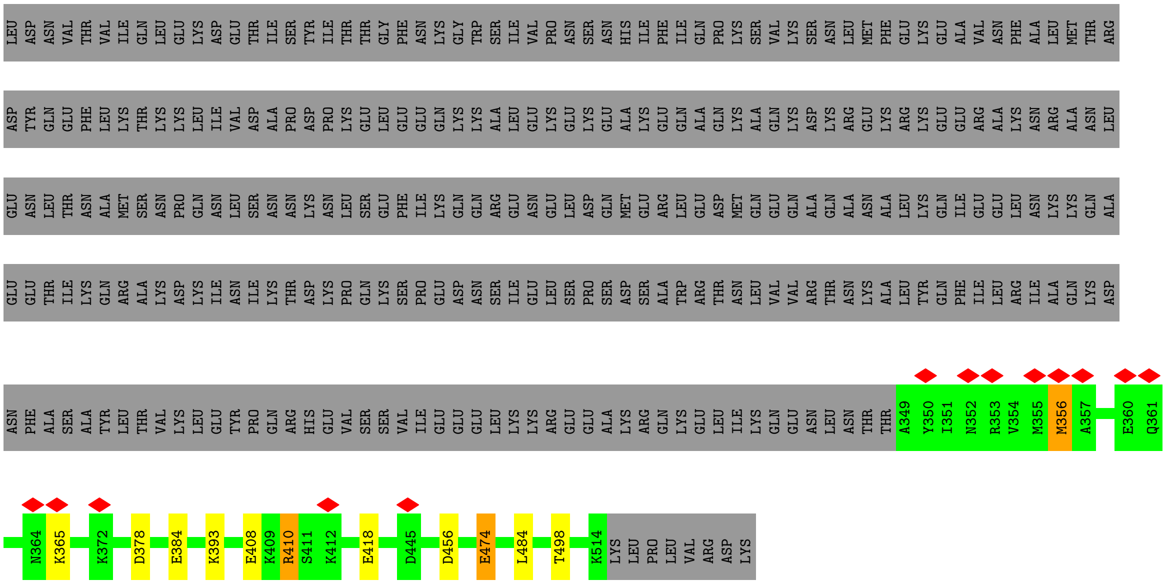
- Molecule 5: Type IV secretion system apparatus protein CagX

Chain IX:



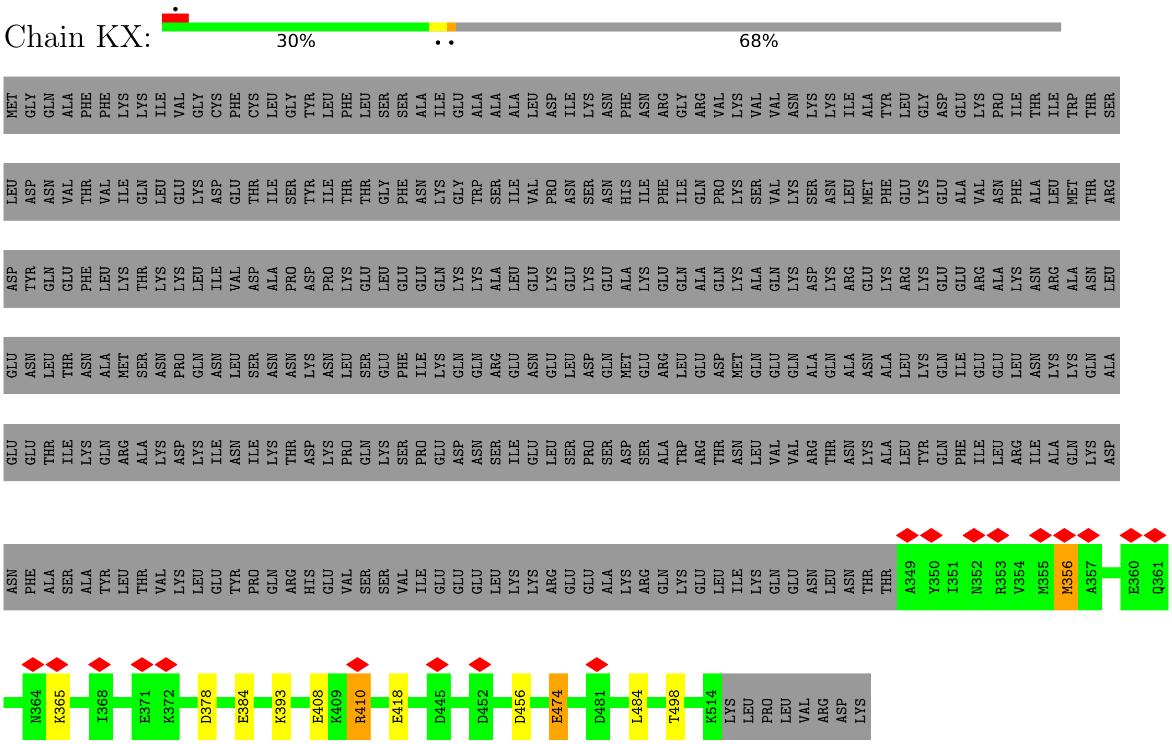
- Molecule 5: Type IV secretion system apparatus protein CagX

Chain JX:



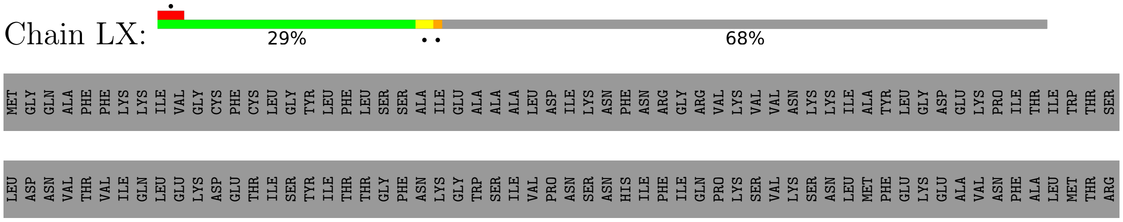
- Molecule 5: Type IV secretion system apparatus protein CagX

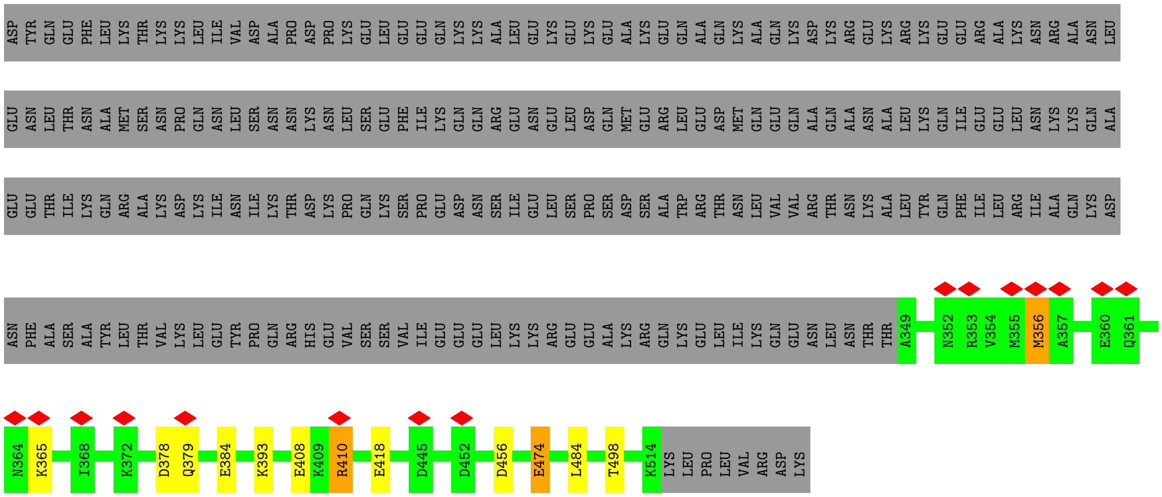
Chain KX:



- Molecule 5: Type IV secretion system apparatus protein CagX

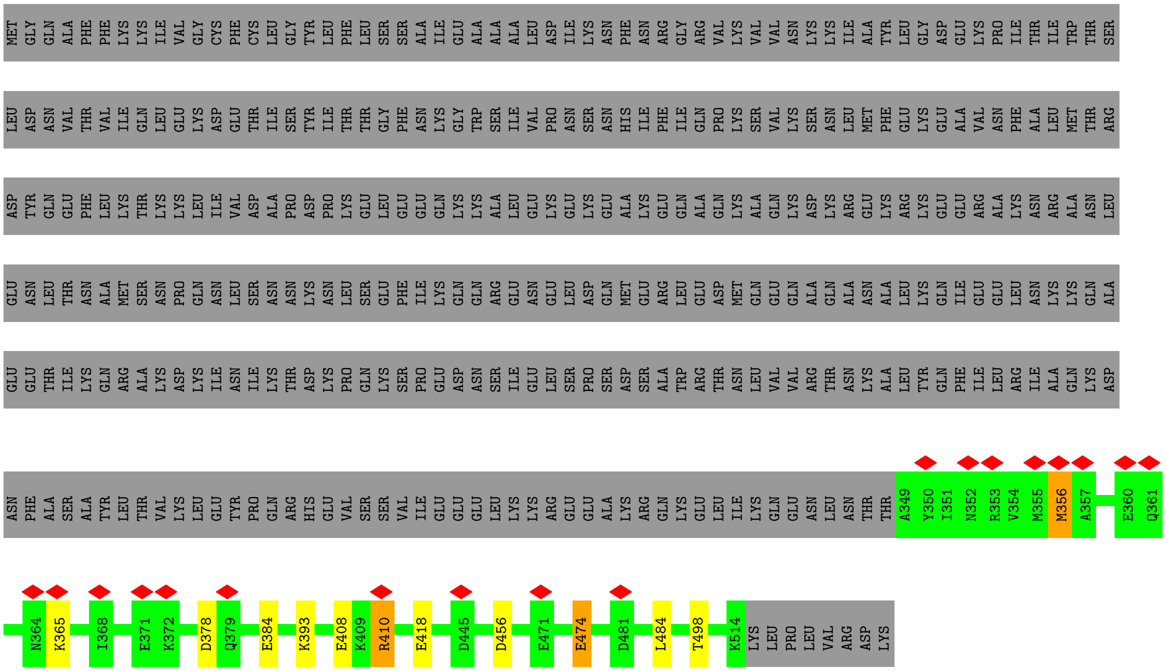
Chain LX:





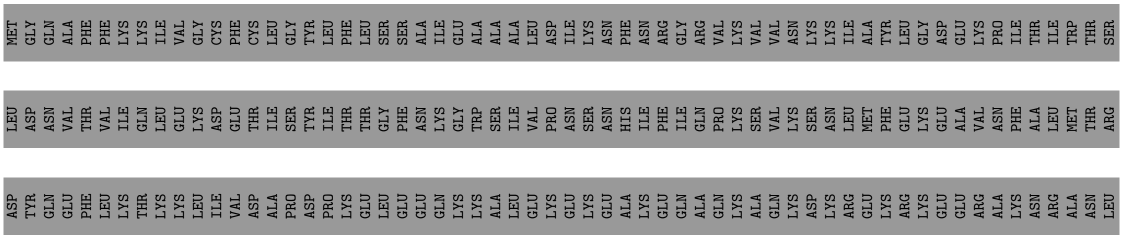
- Molecule 5: Type IV secretion system apparatus protein CagX

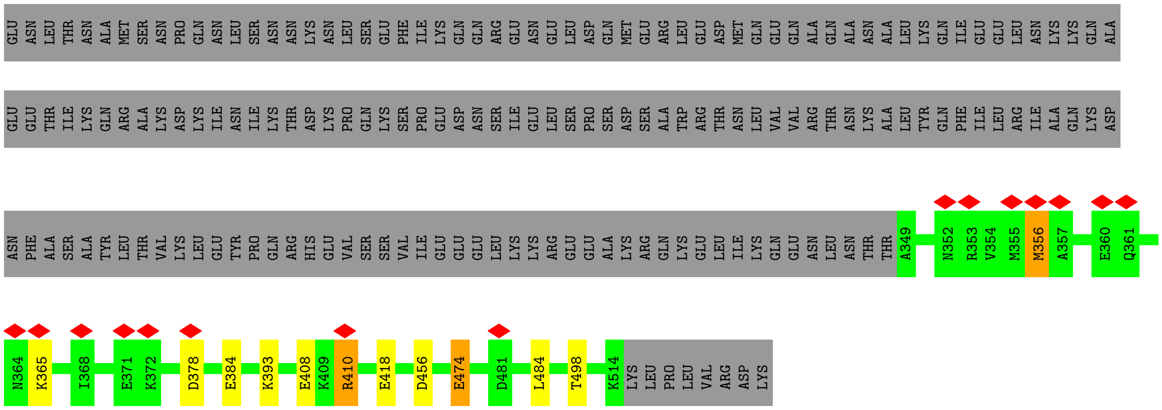
Chain MX:



- Molecule 5: Type IV secretion system apparatus protein CagX

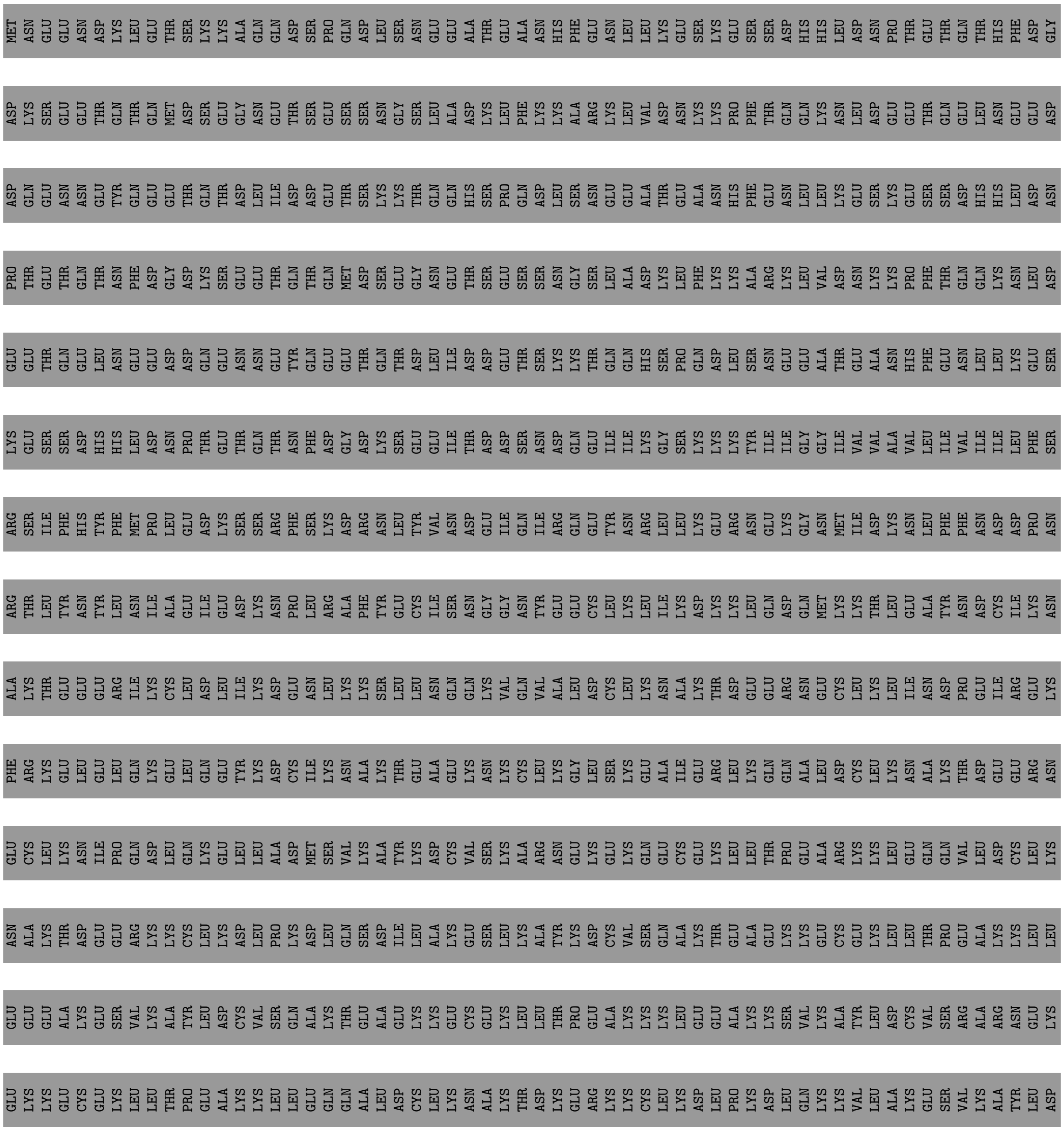
Chain NX:

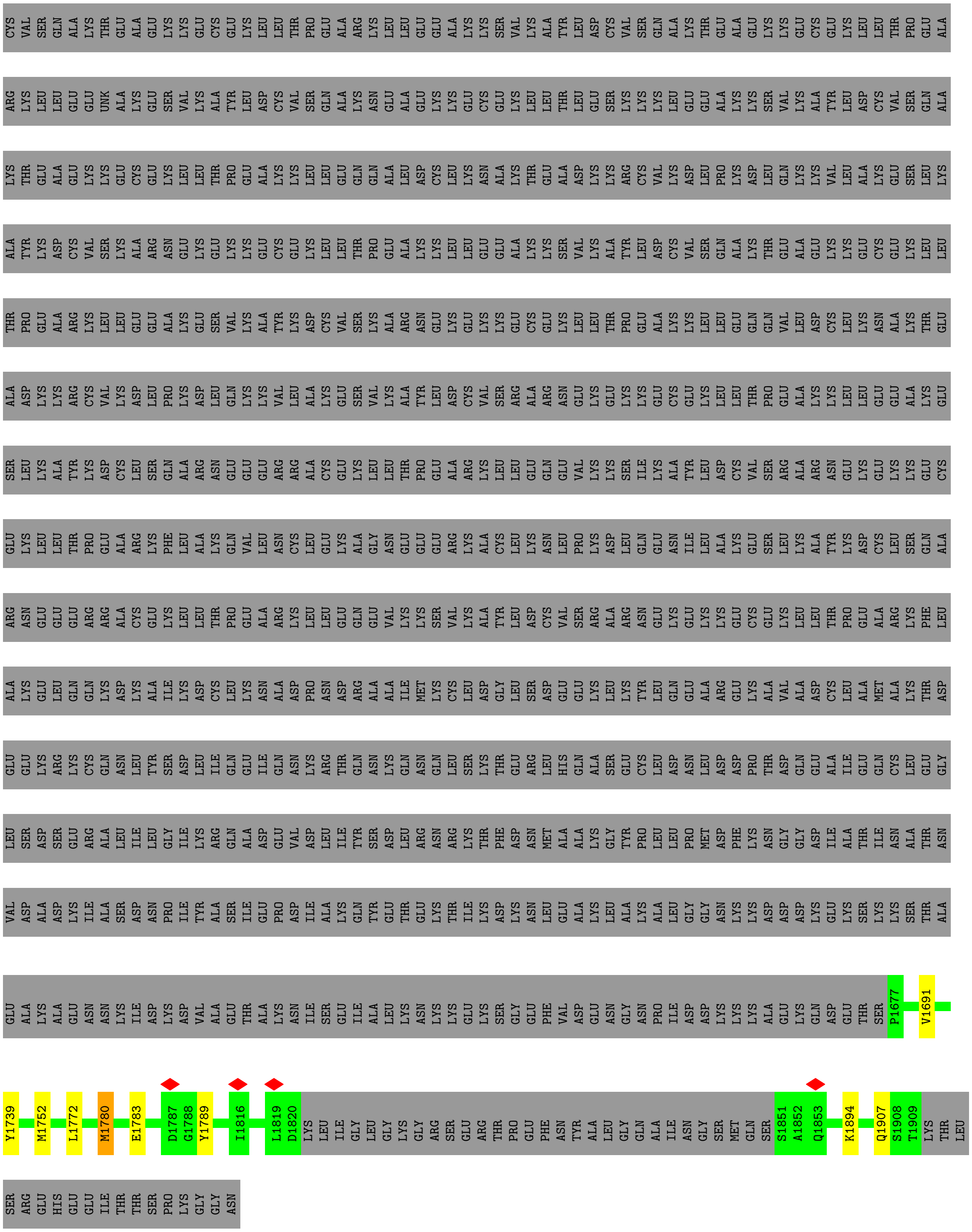




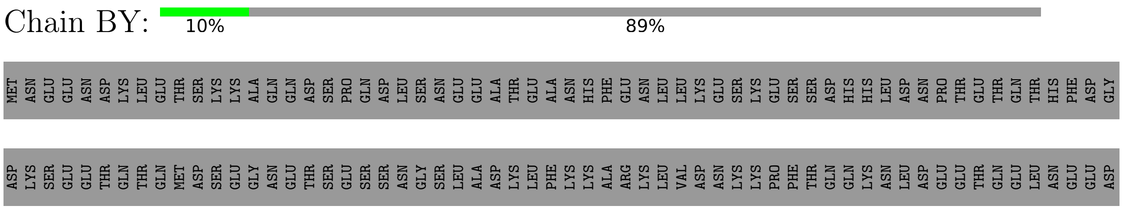
- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain AY:





- Molecule 6: Cag pathogenicity island protein (Cag7)



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 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

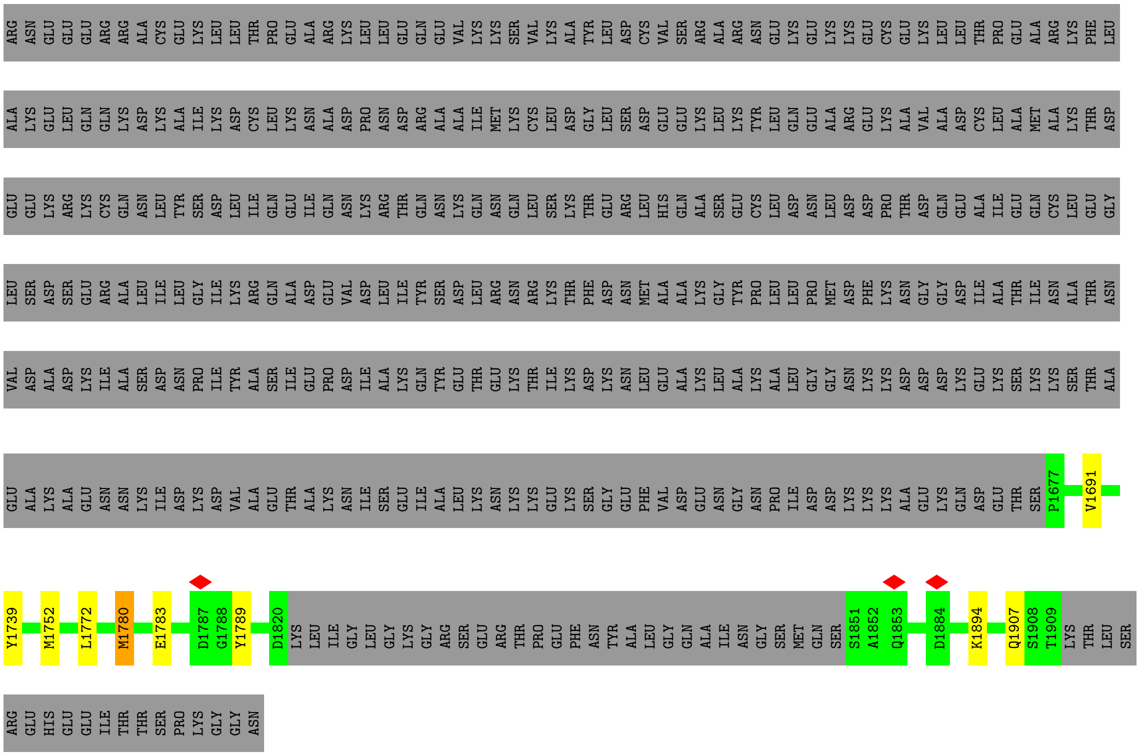
ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU

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

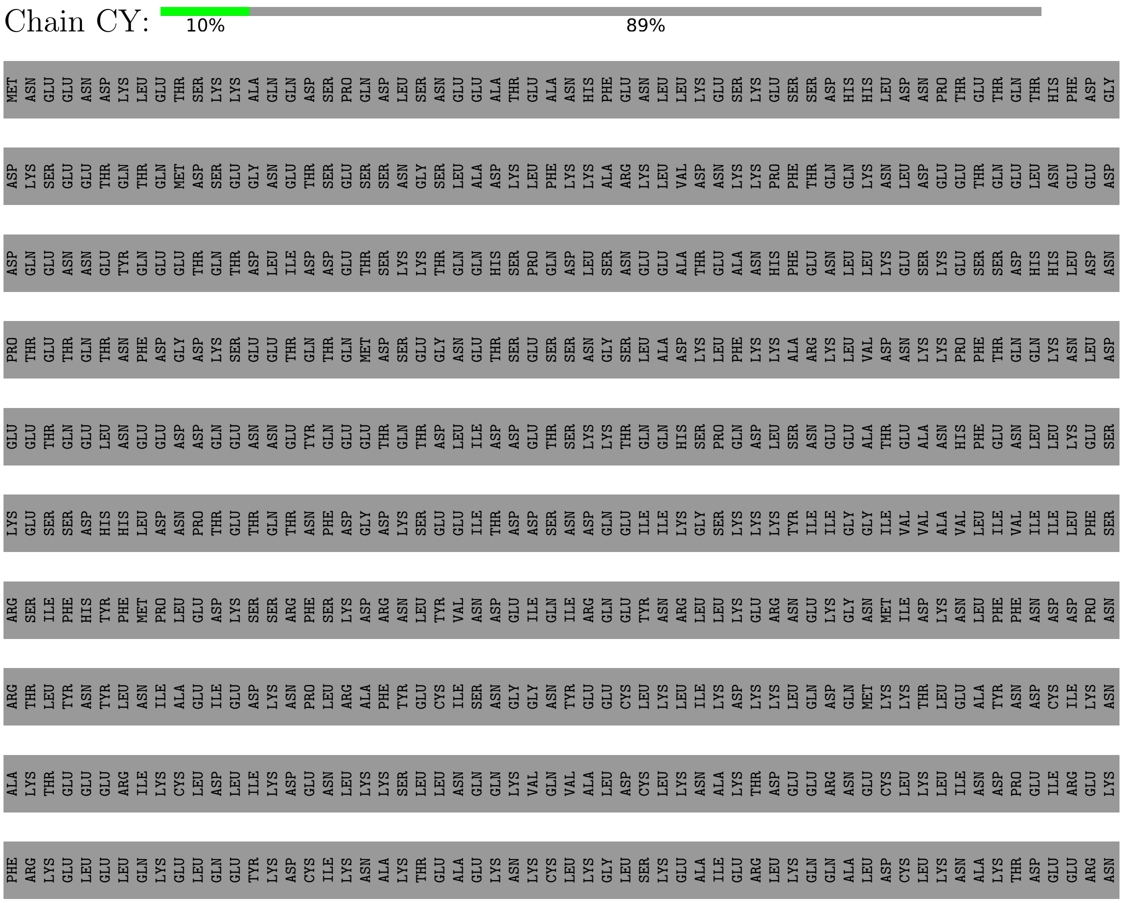
ALA ASP LYS LYS ARG CYS VAL LYS ASP LEU PRO LYS ASP LEU GLN LYS LYS VAL LEU ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS GLU

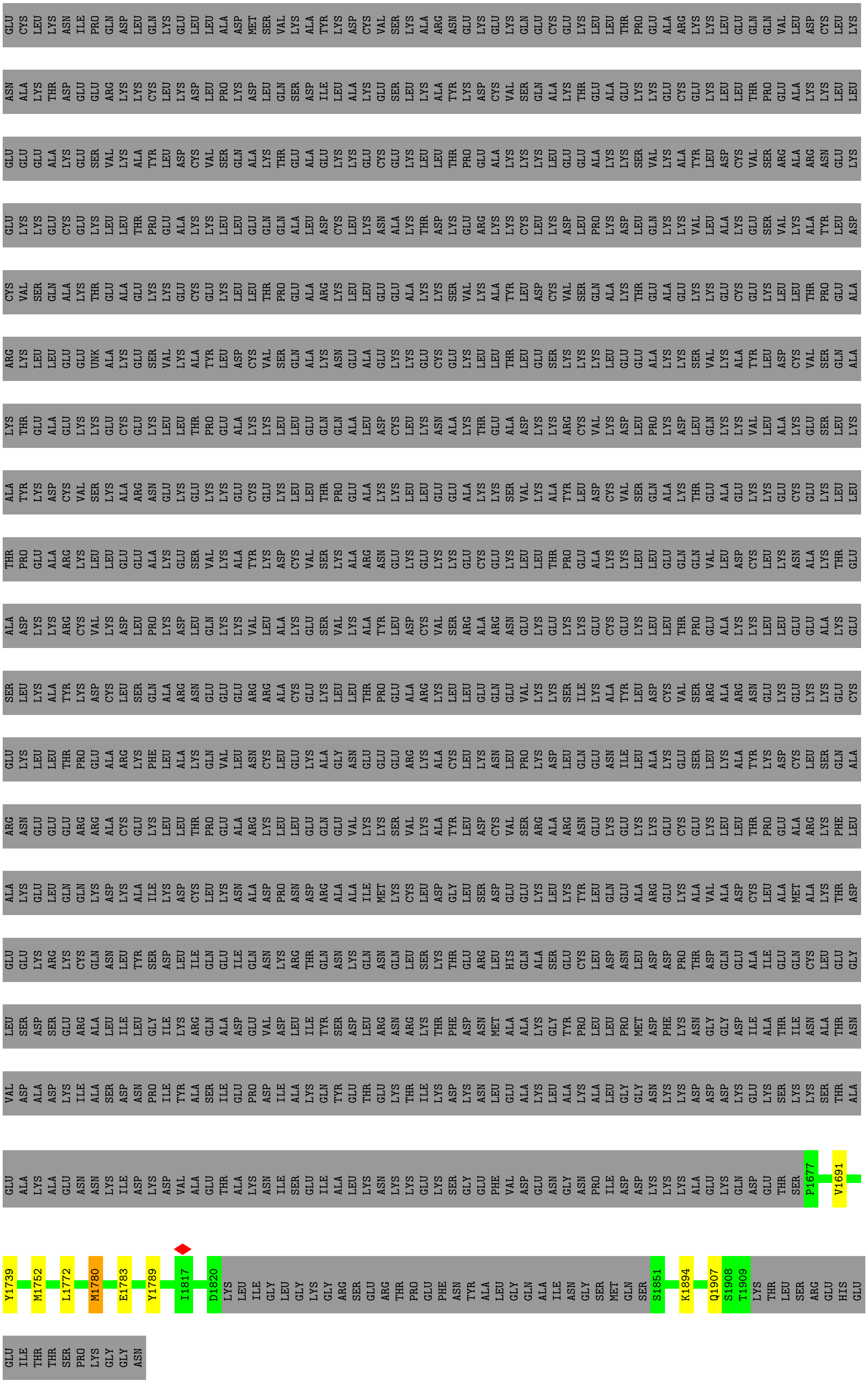
SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER ILE LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS

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 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



- Molecule 6: Cag pathogenicity island protein (Cag7)




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 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
ALA LYS GLU LEU GLN GLN LYS ASP LYS ALA ILE LYS ASP CYS LEU LYS ASN ALA ASP PRO ASN 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 ASP CYS LEU 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 ARG THR GLN ASN LYS GLN ASN GLN LEU SER LYS THR GLU ARG LEU HIS GLN ALA SER GLU CYS LEU ASP ASN LEU ASP ASP PRO THR ASP GLN GLU ALA ILE GLU GLN CYS LEU GLU GLY
LEU SER ASP SER GLU ARG ALA LEU ILE LEU GLY ILE LYS ARG GLN ALA ASP GLU VAL ASP LEU ILE TYR SER ASP LEU ARG ASN ARG LYS THR PHE ASP ASN MET ALA ALA LYS GLY TYR PRO LEU LEU PRO MET ASP PHE LYS ASN GLY GLY ASP ILE ALA THR ILE ASN ALA THR ASN
VAL ASP ALA ASP LYS ILE ALA SER ASP ASN PRO ILE TYR ALA SER ILE GLU PRO ASP ILE ALA LYS GLN TYR GLU THR GLU LYS THR ILE LYS ASP LYS ASN LEU GLU ALA LYS LEU ALA LYS ALA LEU 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 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 P1677 V1691
Y1739 M1752 L1772 M1780 E1783 Y1789 I1817 D1820 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 S1851 K1894 Q1907 S1908 T1909 LYS THR LEU SER ARG GLU HIS GLU
GLU ILE THR THR SER PRO LYS GLY GLY ASN

- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain DY:

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 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 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

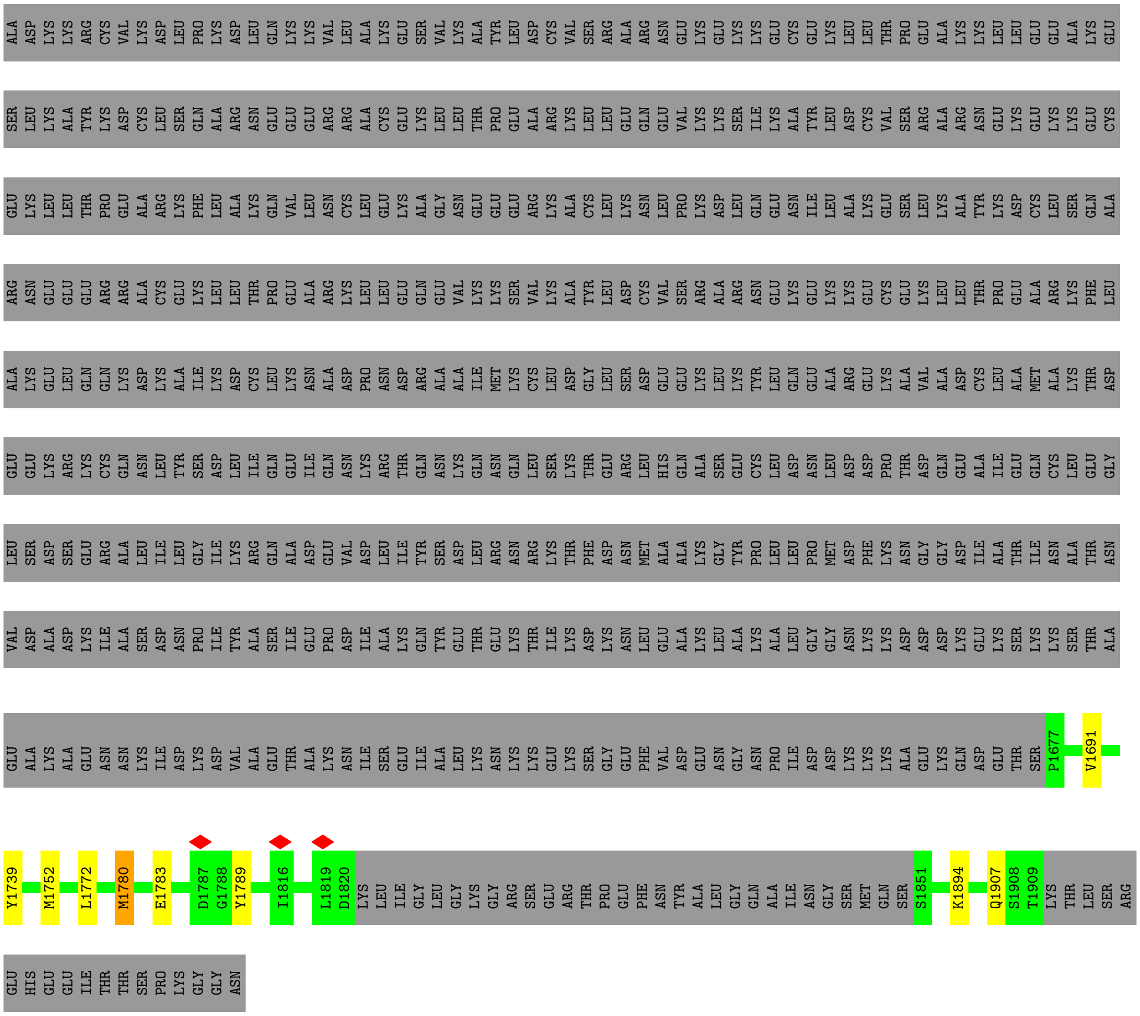
CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA

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

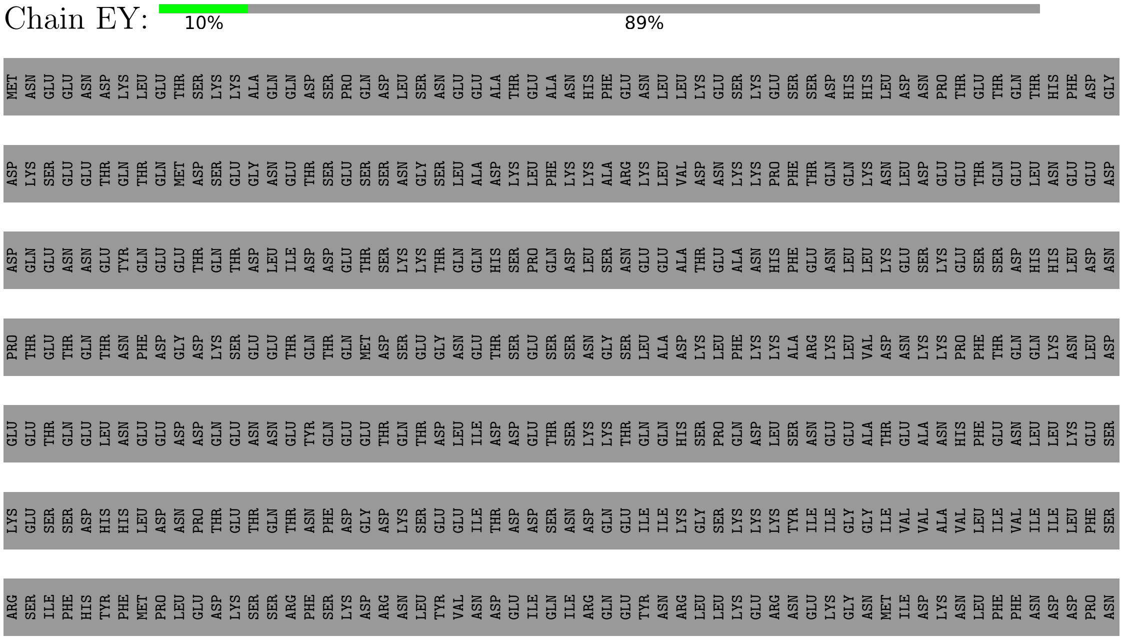
ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS THR GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU

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



- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain EY:



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 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 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

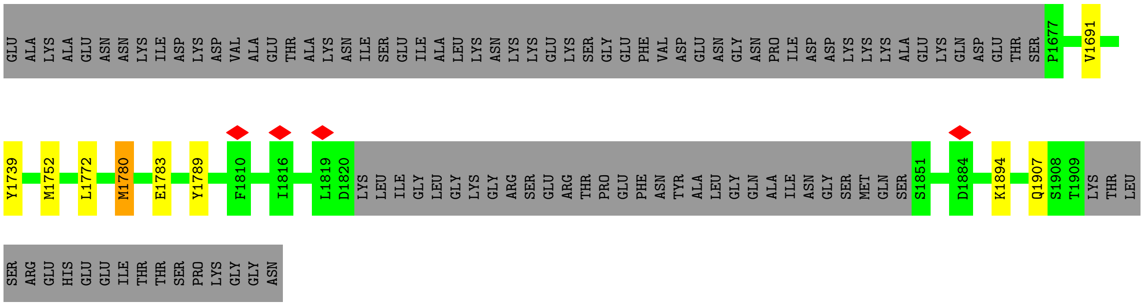
ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS PHE LEU

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

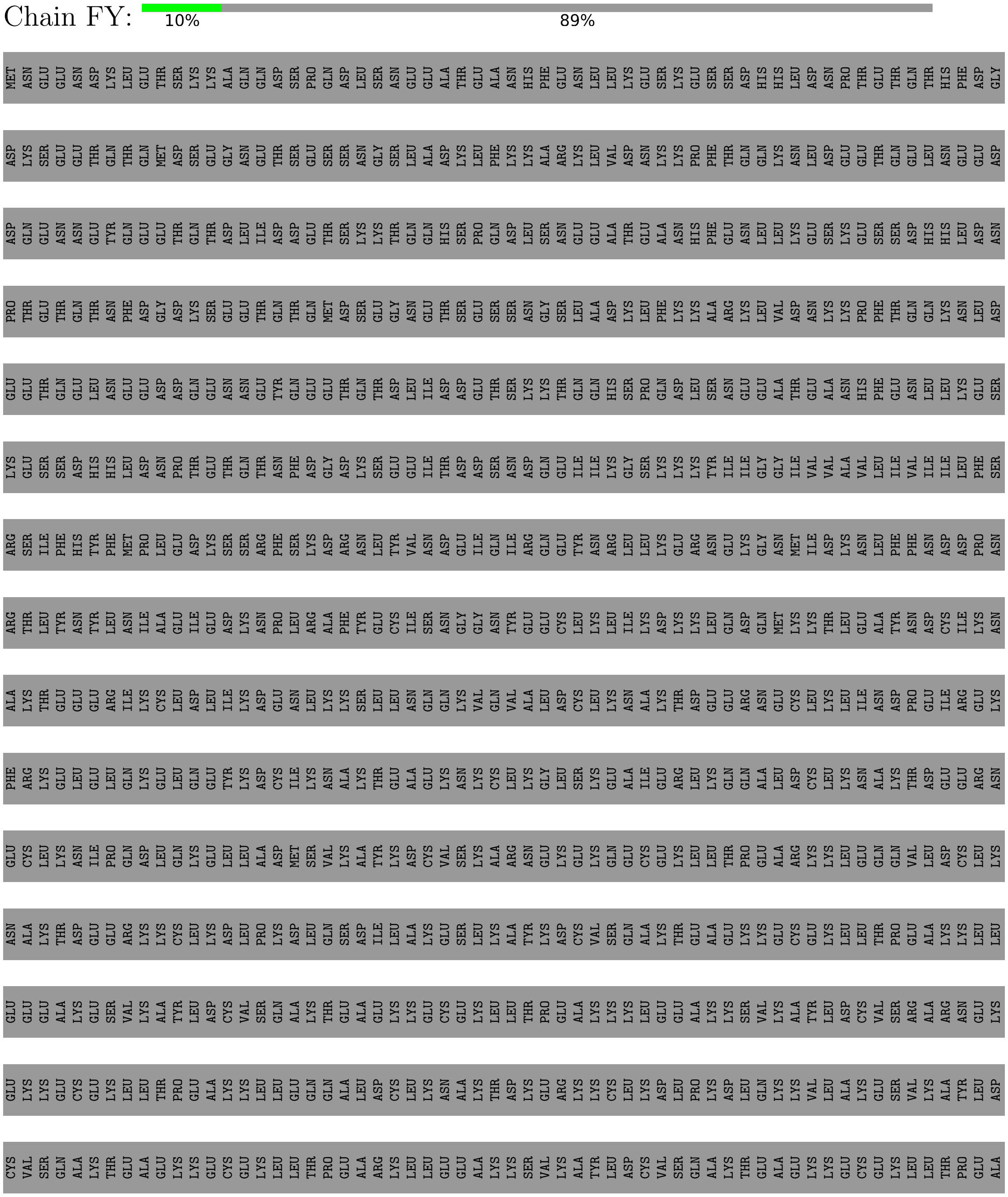
LEU SER ASP SER GLU ARG ALA LEU ILE LEU GLY ILE LYS ARG GLN ALA ASP GLU VAL ASP LEU ILE TYR SER ASP LEU ARG ASN ARG LYS THR PHE ASP ASN MET ALA ALA LYS GLY TYR PRO LEU LEU PRO MET ASP PHE LYS ASN GLY GLY ASP ILE ALA THR ILE ASN ALA THR ASN

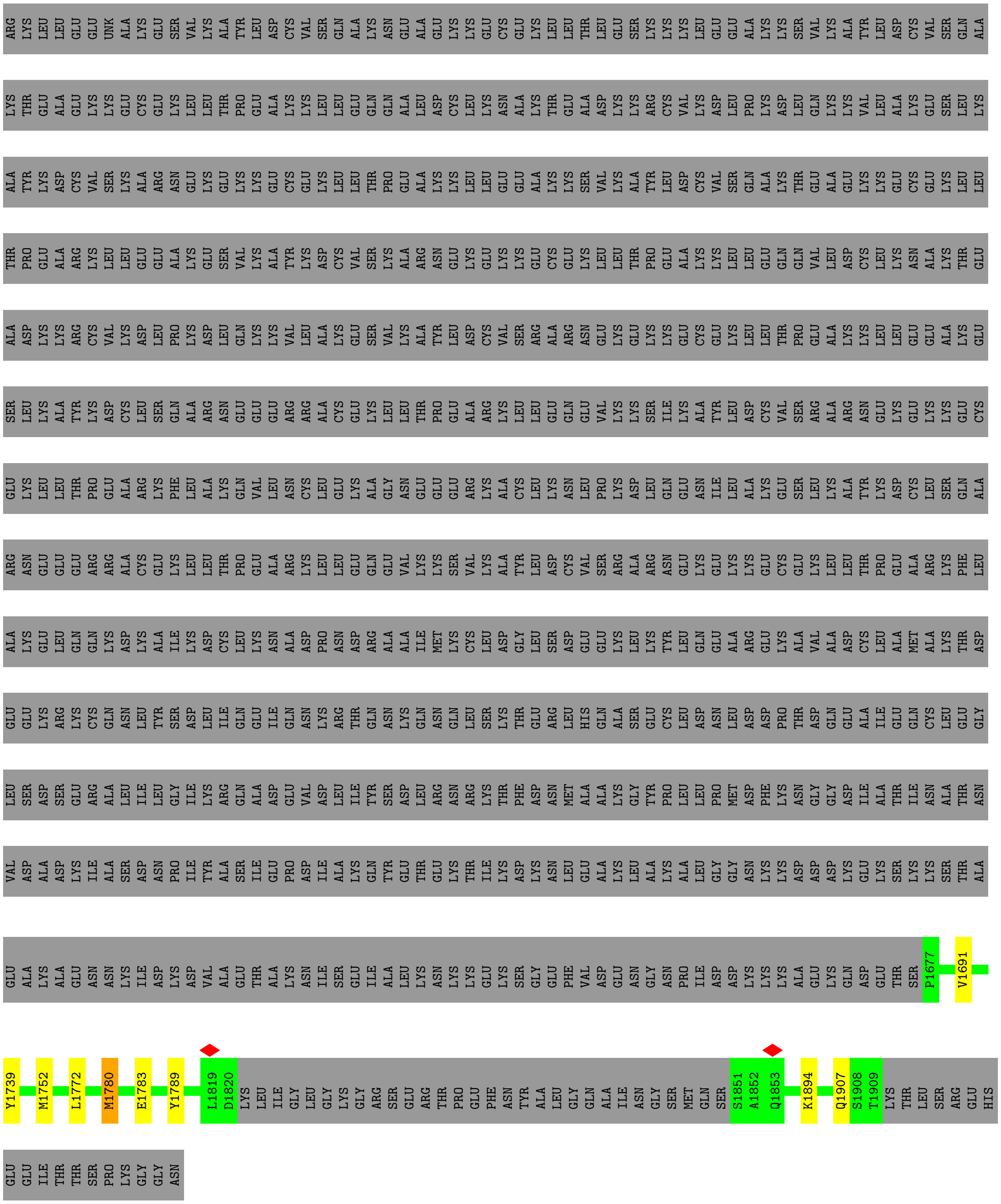
VAL ASP ALA ASP LYS ILE ALA SER ASP ASN PRO ILE TYR ALA SER ILE GLU PRO ASP ILE ALA LYS GLN TYR GLU THR GLU LYS THR ILE LYS ASP LYS ASN LEU GLU ALA LYS LEU ALA LYS ALA LEU GLY GLY ASN LYS LYS ASP ASP ASP LYS GLU LYS SER LYS LYS SER THR ALA



- Molecule 6: Cag pathogenicity island protein (Cag7)

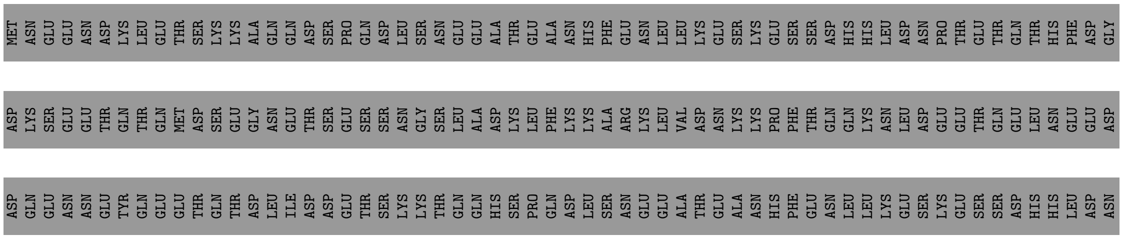
Chain FY:





- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain GY:



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 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

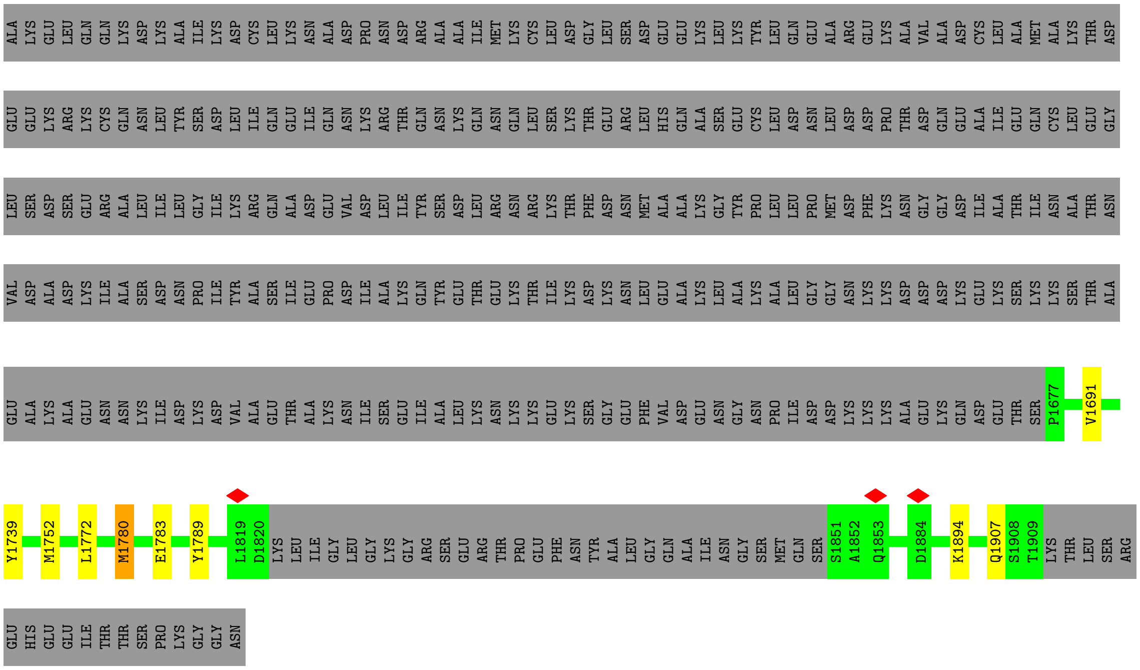
THR PRO GLU ALA ARG LYS LEU LEU GLU GLU ALA LYS GLU SER VAL LYS ALA TYR LYS ASP CYS VAL SER LYS ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS GLU LYS LEU LEU THR PRO GLU ALA LYS LYS LEU LEU GLU GLN GLN VAL LEU ASP CYS LEU LYS ASN ALA LYS THR GLU

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

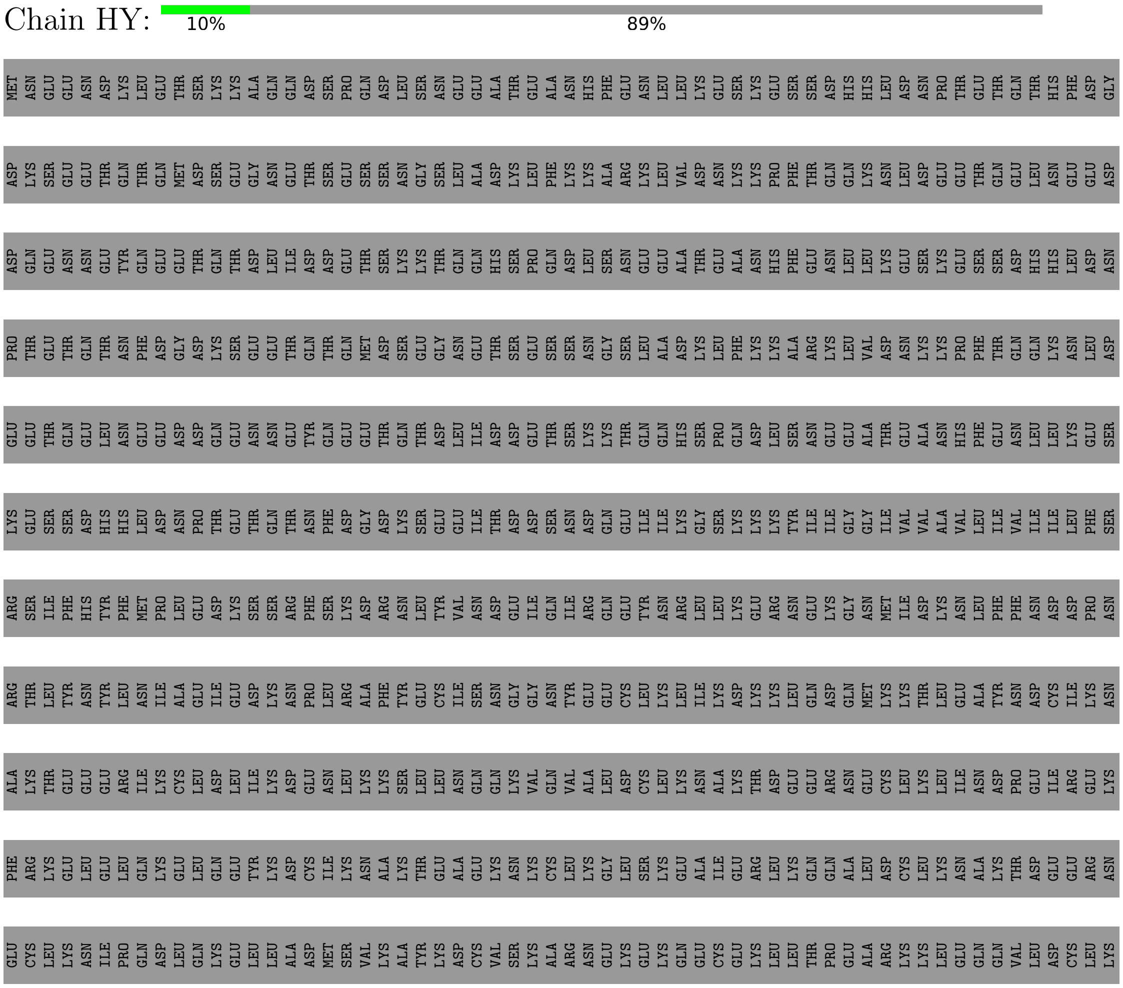
SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER ILE LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS

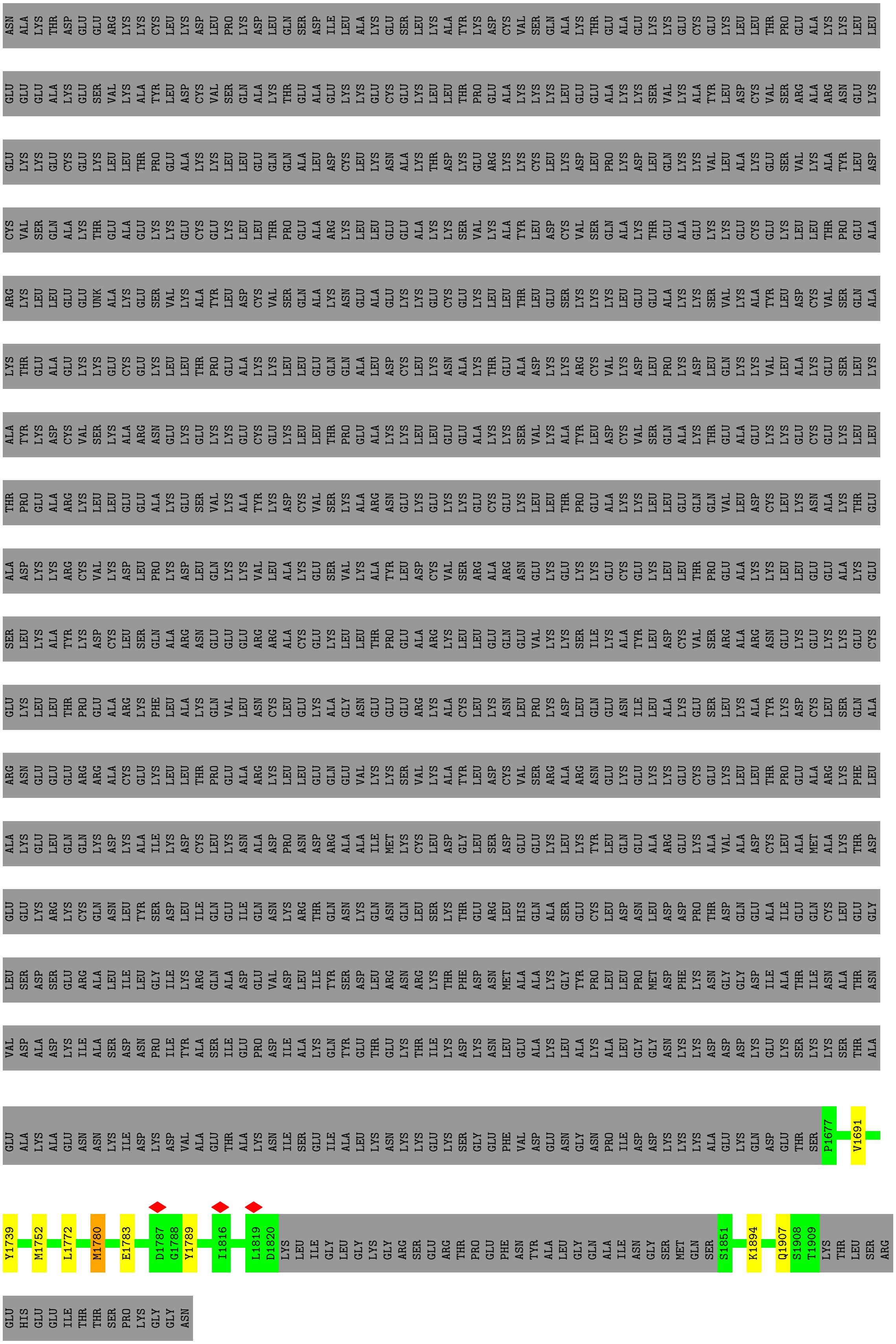
GLU LYS LEU LEU THR PRO GLU ALA ARG LYS PHE LEU ALA LYS GLN VAL LEU ASN CYS LEU GLU LYS ALA GLY ASN GLU GLU GLU ARG LYS ALA CYS LEU LYS ASN LEU PRO LYS ASP LEU GLN GLU ASN ILE LEU ALA LYS GLU SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA

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 6: Cag pathogenicity island protein (Cag7)





- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain IY:

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 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 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

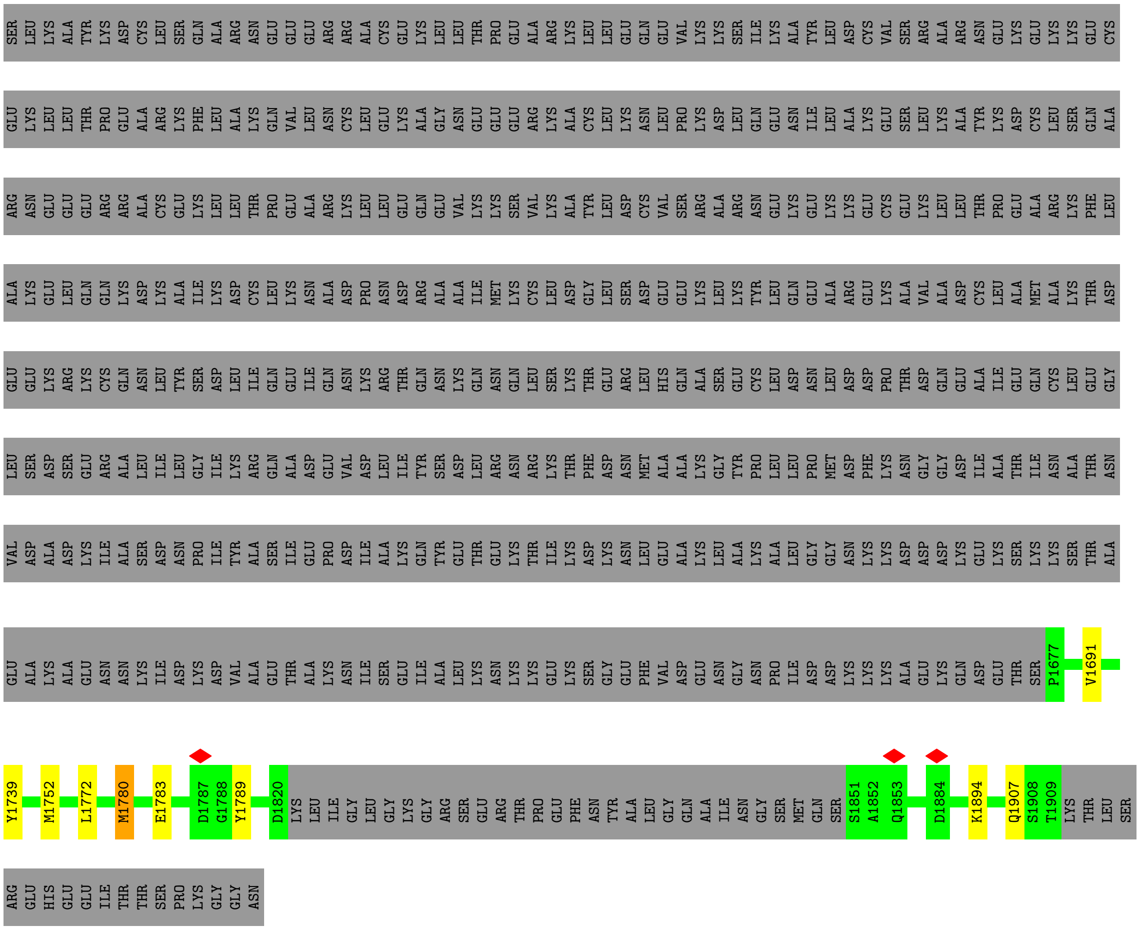
ARG LYS LEU LEU GLU GLU UNK ALA LYS GLU SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA LYS ASN GLU ALA GLU LYS LYS GLU CYS GLU LYS LEU LEU THR LEU GLU SER LYS LYS LYS LEU GLU GLU ALA LYS LYS SER VAL LYS ALA TYR LEU ASP CYS VAL SER GLN ALA

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

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

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

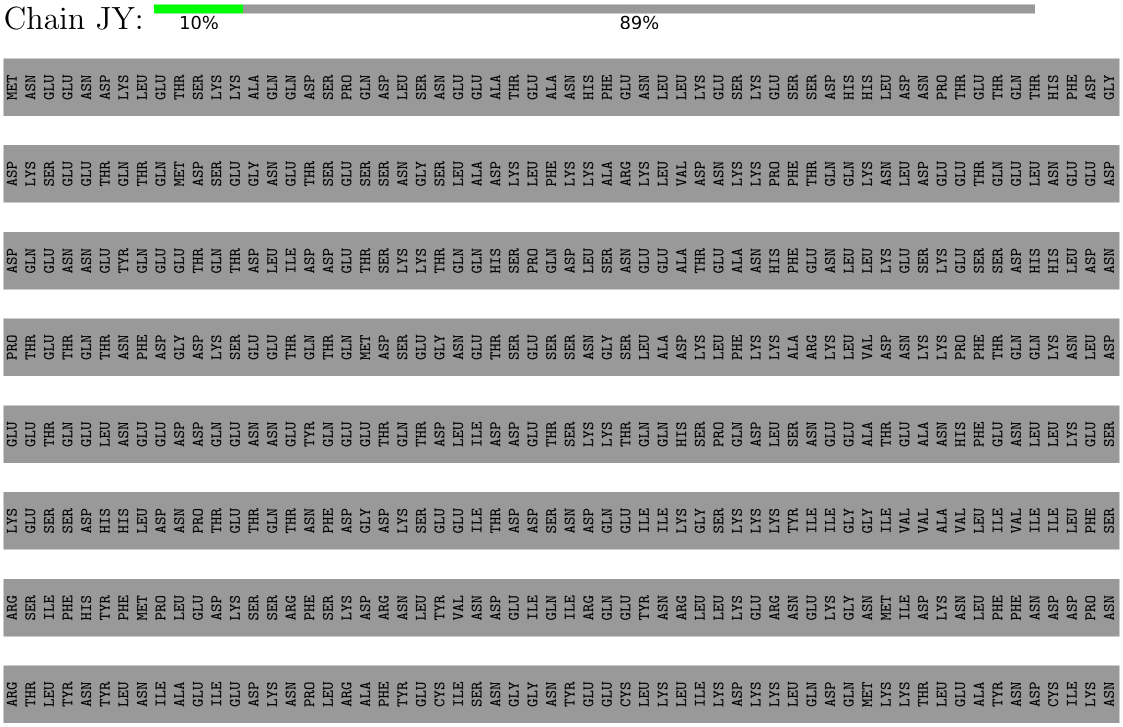
GLU ALA LYS ALA GLU ASN ASN LYS ILE ASP LYS ASP VAL ALA GLU THR ALA LYS ASN ILE SER GLU ILE ALA LEU LYS ASN LYS LYS GLU LYS SER GLY GLU PHE VAL ASP GLU ASN GLY ASN PRO ILE ASP ASP LYS LYS LYS ALA GLU LYS GLN ASP GLU THR SER P1677 V1691

Y1739 M1752 L1772 M1780 E1783 D1787 G1788 Y1789 D1820 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 S1851 A1852 Q1853 D1884 K1894 Q1907 S1908 T1909 LYS THR LEU SER

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

- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain JY: 10% 89%



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

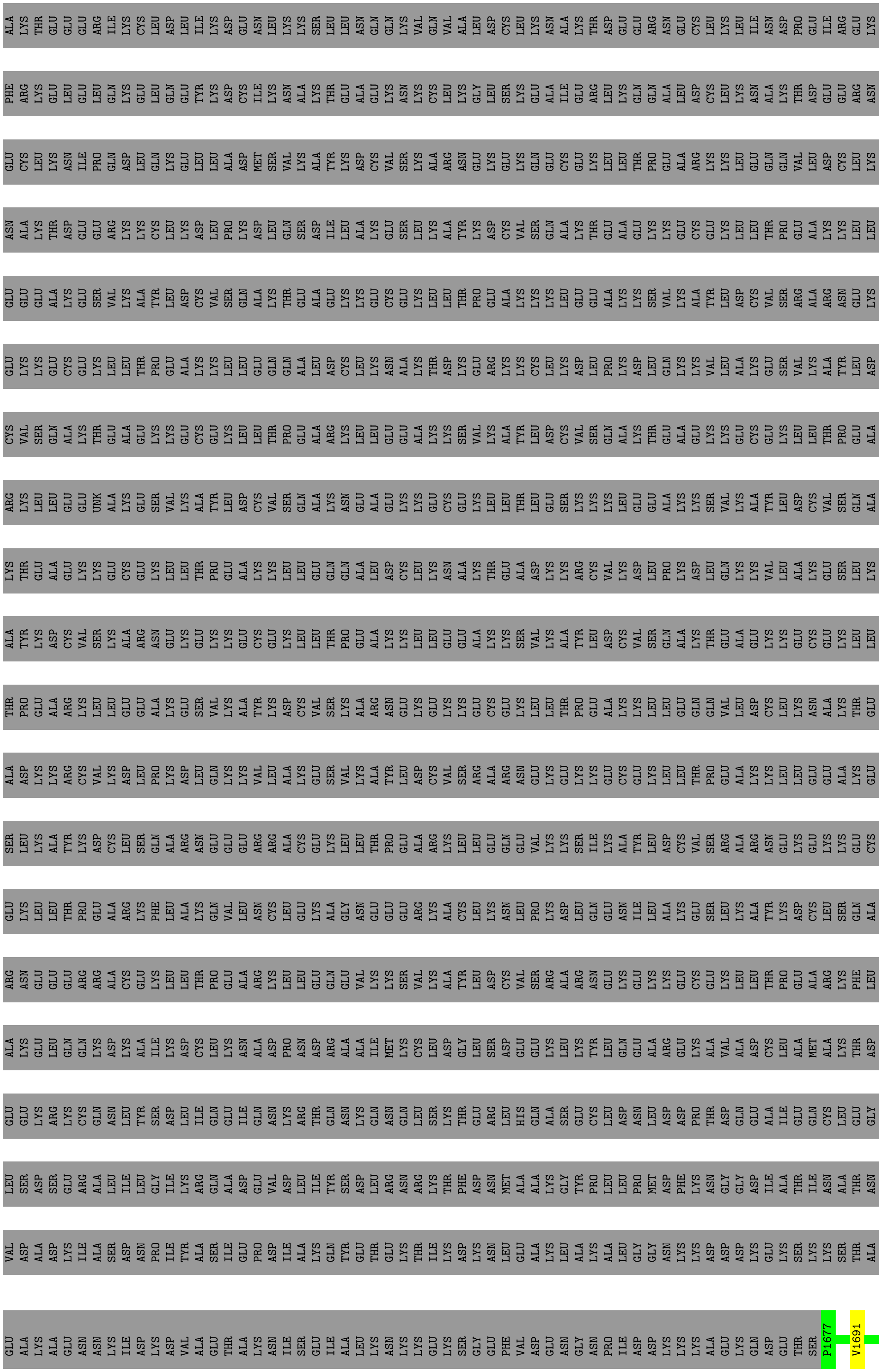
PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS ASN LEU ASP

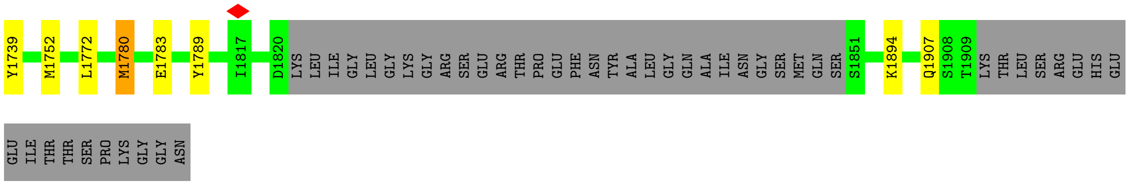
GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER

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

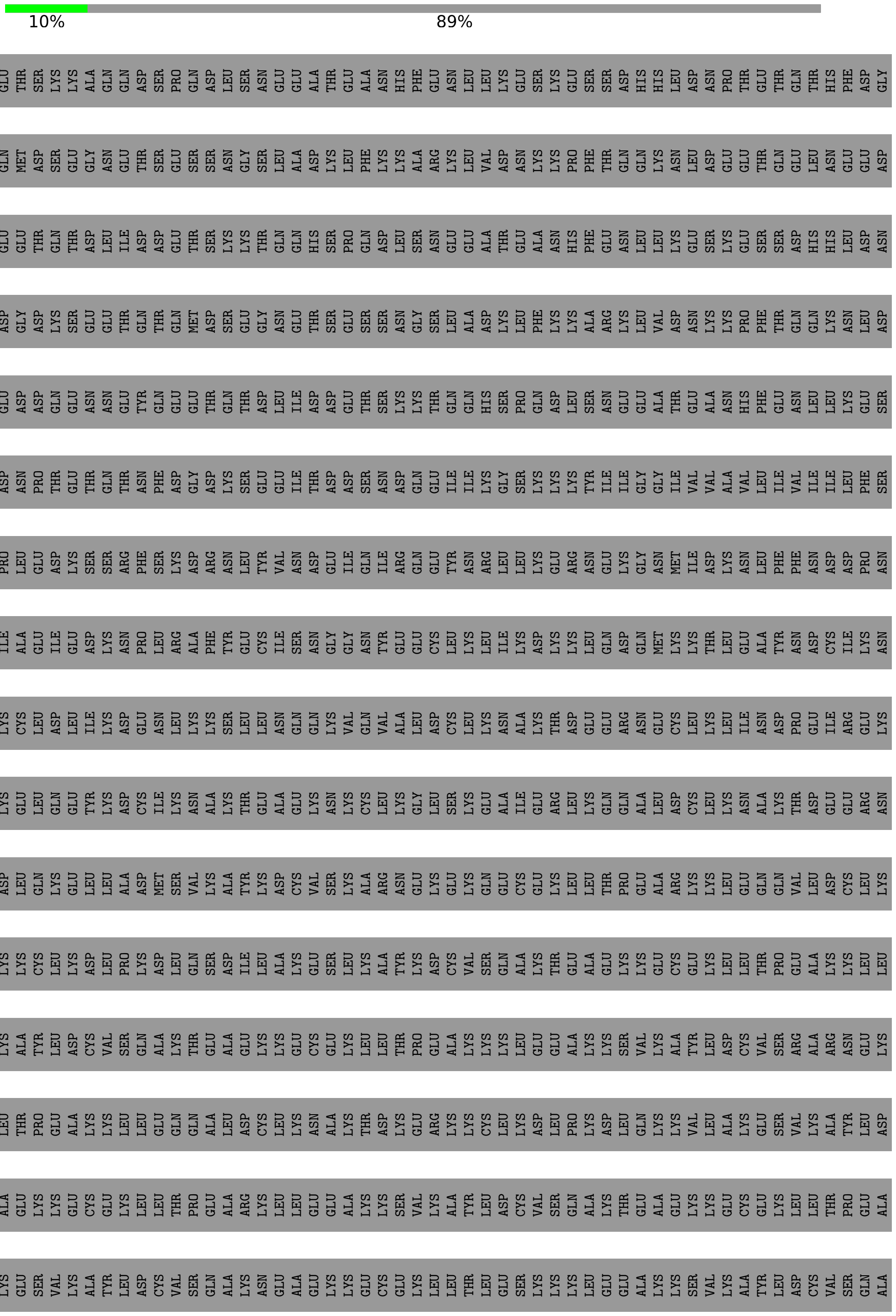
ARG THR LEU TYR ASN TYR LEU ASN ILE ALA GLU ILE GLU ASP LYS ASN PRO LEU ARG ALA PHE TYR GLU CYS ILE SER ASN GLY GLY ASN TYR GLU CYS LEU LYS LEU ILE LYS ASP LYS LYS LEU GLN ASP GLN MET LYS LYS THR LEU GLU ALA TYR ASN ASP CYS ILE LYS ASN


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 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 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 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 ASP CYS LEU 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 ARG THR GLN ASN LYS GLN ASN GLN LEU SER LYS THR GLU ARG LEU HIS GLN ALA SER GLU CYS LEU ASP ASN LEU ASP ASP PRO THR ASP GLN GLU ALA ILE GLU GLN CYS LEU GLU GLY
LEU SER ASP SER GLU ARG ALA LEU ILE LEU GLY ILE LYS ARG GLN ALA ASP GLU VAL ASP LEU ILE TYR SER ASP LEU ARG ASN ARG LYS THR PHE ASP ASN MET ALA ALA LYS GLY TYR PRO LEU LEU PRO MET ASP PHE LYS ASN GLY GLY ASP ILE ALA THR ILE ASN ALA THR ASN
VAL ASP ALA ASP LYS ILE ALA SER ASP ASN PRO ILE TYR ALA SER ILE GLU PRO ASP ILE ALA LYS GLN TYR GLU THR GLU LYS THR ILE LYS ASP LYS ASN LEU GLU ALA LYS LEU ALA LYS ALA LEU 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 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 P1677 V1691



- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain KY:



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

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

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

VAL ASP ALA ASP LYS ILE ALA SER ASP ASN PRO ILE TYR ALA SER ILE GLU PRO ASP ILE ALA LYS GLN TYR GLU THR GLU LYS THR ILE LYS ASP LYS ASN LEU GLU ALA LYS LEU ALA LYS ALA LEU 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 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 P1677 V1691

Y1739 M1752 L1772 M1780 E1783 ◆ D1787 G1788 Y1789 ◆ I1816 D1820 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 S1851 K1894 Q1907 S1908 T1909 LYS THR LEU SER ARG GLU

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

- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain LY: 10% 89%

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 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 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

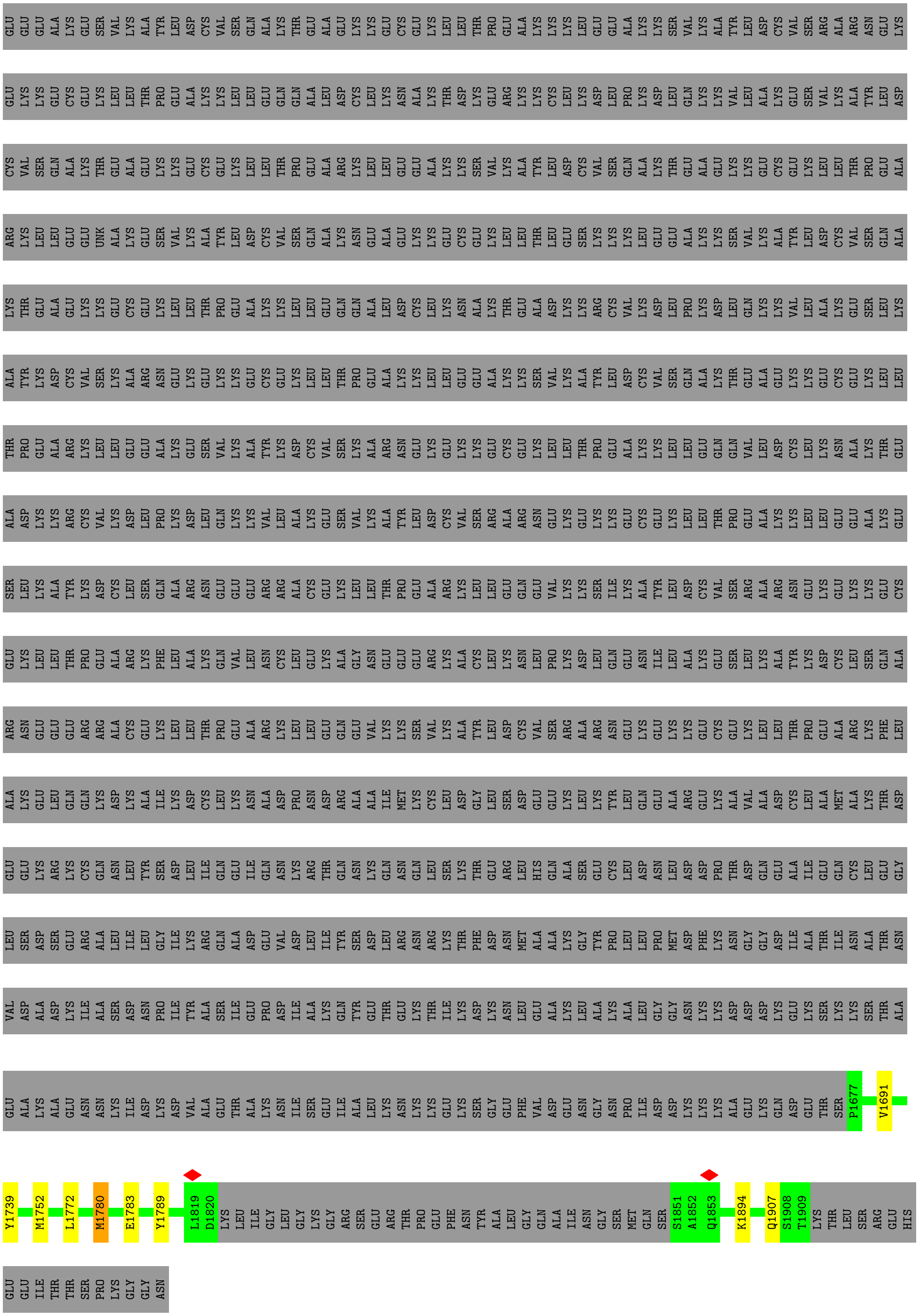
SER LEU LYS ALA TYR LYS ASP CYS LEU SER GLN ALA ARG ASN GLU GLU GLU ARG ARG ALA CYS GLU LYS LEU LEU THR PRO GLU ALA ARG LYS LEU LEU GLU GLN GLU VAL LYS LYS SER ILE LYS ALA TYR LEU ASP CYS VAL SER ARG ALA ARG ASN GLU LYS GLU LYS LYS GLU CYS

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 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

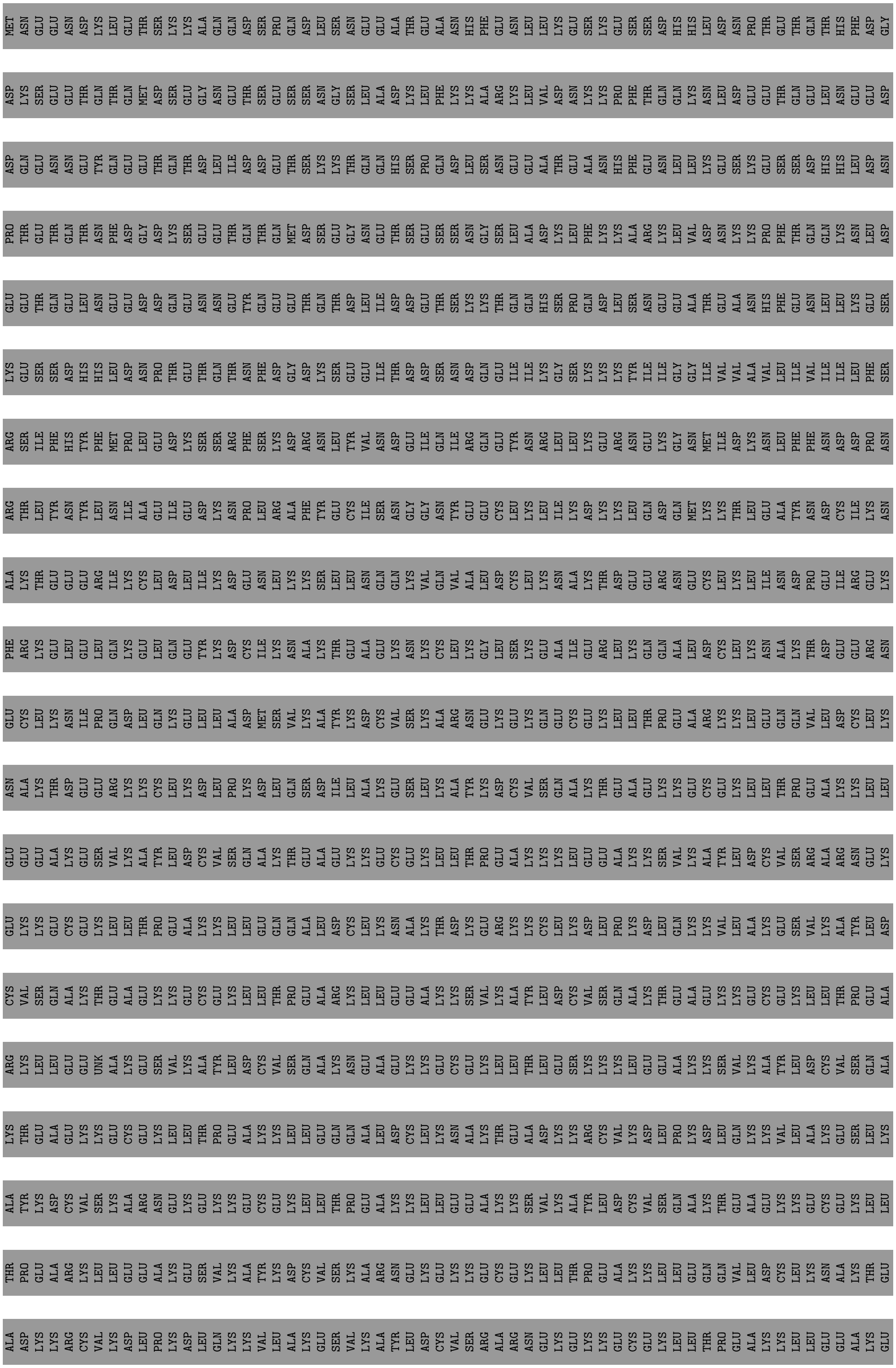
ALA LYS GLU LEU GLN GLN LYS ASP LYS ALA ILE LYS ASP CYS LEU LYS ASN ALA ASP PRO ASN ASP ARG ALA ALA ILE MET LYS CYS LEU ASP GLY LEU SER ASP GLU GLU LYS LEU LYS TYR LEU GLN GLU ALA ARG GLU LYS ALA VAL ALA ASP CYS LEU ALA MET ALA LYS THR ASP

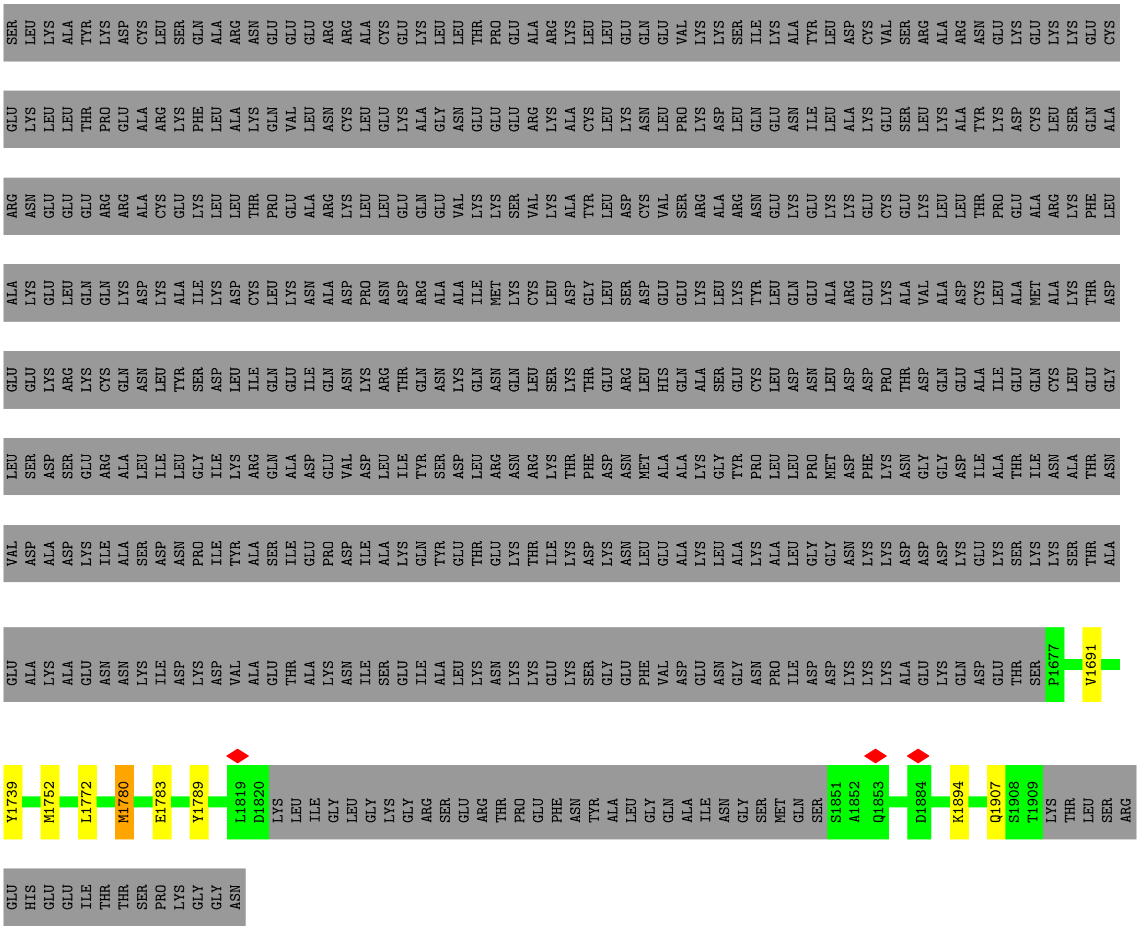
- Molecule 6: Cag pathogenicity island protein (Cag7)



- Molecule 6: Cag pathogenicity island protein (Cag7)

Chain NY:


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 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 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 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 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 ASP CYS LEU 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 ARG THR GLN ASN LYS GLN ASN GLN LEU SER LYS THR GLU ARG LEU HIS GLN ALA SER GLU CYS LEU ASP ASN LEU ASP ASP PRO THR ASP GLN GLU ALA ILE GLU GLN CYS LEU GLU GLY
LEU SER ASP SER GLU ARG ALA LEU ILE LEU GLY ILE LYS ARG GLN ALA ASP GLU VAL ASP LEU ILE TYR SER ASP LEU ARG ASN ARG LYS THR PHE ASP ASN MET ALA ALA LYS GLY TYR PRO LEU LEU PRO MET ASP PHE LYS ASN GLY GLY ASP ILE ALA THR ILE ASN ALA THR ASN
VAL ASP ALA ASP LYS ILE ALA SER ASP ASN PRO ILE TYR ALA SER ILE GLU PRO ASP ILE ALA LYS GLN TYR GLU THR GLU LYS THR ILE LYS ASP LYS ASN LEU GLU ALA LYS LEU ALA LYS ALA LEU 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 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 P1677 V1691
Y1739 M1752 L1772 M1780 E1783 Y1789 L1819 D1820 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 S1851 A1852 Q1853 D1884 K1894 Q1907 S1908 T1909 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, C14	Depositor
Number of particles used	47440	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 ($e^-/\text{Å}^2$)	59.2, 59.0	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k), GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.139	Depositor
Minimum map value	-0.056	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.004	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

5.1 Standard geometry

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	AA	0.50	0/2023	0.83	6/2742 (0.2%)
1	AB	0.56	1/742 (0.1%)	0.87	2/1007 (0.2%)
1	AC	0.58	0/755	0.90	1/1023 (0.1%)
1	AD	0.62	0/793	0.99	2/1073 (0.2%)
1	AE	0.59	1/864 (0.1%)	0.94	3/1170 (0.3%)
1	BA	0.50	0/2023	0.83	6/2742 (0.2%)
1	BB	0.54	1/742 (0.1%)	0.82	1/1007 (0.1%)
1	BC	0.54	0/755	0.90	2/1023 (0.2%)
1	BD	0.69	2/793 (0.3%)	1.11	7/1073 (0.7%)
1	BE	0.52	1/864 (0.1%)	0.83	2/1170 (0.2%)
1	CA	0.50	0/2023	0.83	6/2742 (0.2%)
1	CB	0.73	3/742 (0.4%)	0.92	4/1007 (0.4%)
1	CC	0.55	0/755	0.94	1/1023 (0.1%)
1	CD	0.56	0/793	1.01	4/1073 (0.4%)
1	CE	0.44	0/864	0.92	4/1170 (0.3%)
1	DA	0.50	0/2023	0.83	6/2742 (0.2%)
1	DB	0.81	3/742 (0.4%)	1.04	4/1007 (0.4%)
1	DC	0.56	0/755	0.96	3/1023 (0.3%)
1	DD	0.64	0/793	0.98	4/1073 (0.4%)
1	DE	0.46	0/864	1.01	9/1170 (0.8%)
1	EA	0.50	0/2023	0.83	6/2742 (0.2%)
1	EB	0.73	2/742 (0.3%)	0.99	4/1007 (0.4%)
1	EC	0.55	0/755	0.95	2/1023 (0.2%)
1	ED	0.64	0/793	1.14	7/1073 (0.7%)
1	EE	0.48	0/864	0.93	5/1170 (0.4%)
1	FA	0.50	0/2023	0.83	6/2742 (0.2%)
1	FB	0.71	3/742 (0.4%)	0.95	4/1007 (0.4%)
1	FC	0.65	1/755 (0.1%)	1.03	5/1023 (0.5%)
1	FD	0.68	0/793	1.03	5/1073 (0.5%)
1	FE	0.49	0/864	0.87	5/1170 (0.4%)
1	GA	0.50	0/2023	0.83	6/2742 (0.2%)
1	GB	0.58	1/742 (0.1%)	0.85	0/1007
1	GC	0.63	1/755 (0.1%)	0.95	2/1023 (0.2%)
1	GD	0.57	0/793	1.05	5/1073 (0.5%)

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#|Z| >5	RMSZ	#|Z| >5
1	GE	0.58	1/864 (0.1%)	0.90	5/1170 (0.4%)
1	HA	0.50	0/2023	0.83	6/2742 (0.2%)
1	HB	0.73	2/742 (0.3%)	0.89	3/1007 (0.3%)
1	HC	0.61	1/755 (0.1%)	0.90	1/1023 (0.1%)
1	HD	0.60	1/793 (0.1%)	1.05	5/1073 (0.5%)
1	HE	0.53	1/864 (0.1%)	0.98	7/1170 (0.6%)
1	IA	0.50	0/2023	0.83	6/2742 (0.2%)
1	IB	0.53	1/742 (0.1%)	0.81	1/1007 (0.1%)
1	IC	0.51	0/755	0.85	0/1023
1	ID	0.63	0/793	1.12	8/1073 (0.7%)
1	IE	0.50	0/864	0.85	3/1170 (0.3%)
1	JA	0.50	0/2023	0.83	6/2742 (0.2%)
1	JB	0.75	3/742 (0.4%)	0.93	4/1007 (0.4%)
1	JC	0.56	0/755	0.98	2/1023 (0.2%)
1	JD	0.67	1/793 (0.1%)	1.13	6/1073 (0.6%)
1	JE	0.50	0/864	0.87	4/1170 (0.3%)
1	KA	0.50	0/2023	0.83	6/2742 (0.2%)
1	KB	0.76	2/742 (0.3%)	0.96	2/1007 (0.2%)
1	KC	0.56	1/755 (0.1%)	0.93	4/1023 (0.4%)
1	KD	1.61	7/793 (0.9%)	1.45	10/1073 (0.9%)
1	KE	1.00	4/864 (0.5%)	1.32	14/1170 (1.2%)
1	LA	0.50	0/2023	0.83	6/2742 (0.2%)
1	LB	0.71	2/742 (0.3%)	0.93	2/1007 (0.2%)
1	LC	0.59	0/755	0.98	5/1023 (0.5%)
1	LD	0.60	0/793	1.10	4/1073 (0.4%)
1	LE	0.46	0/864	0.94	6/1170 (0.5%)
1	MA	0.50	0/2023	0.83	6/2742 (0.2%)
1	MB	0.76	2/742 (0.3%)	0.92	3/1007 (0.3%)
1	MC	0.57	0/755	0.95	2/1023 (0.2%)
1	MD	0.71	2/793 (0.3%)	1.10	5/1073 (0.5%)
1	ME	0.61	2/864 (0.2%)	0.91	5/1170 (0.4%)
1	NA	0.50	0/2023	0.83	6/2742 (0.2%)
1	NB	0.62	1/742 (0.1%)	0.90	2/1007 (0.2%)
1	NC	0.55	0/755	0.91	2/1023 (0.2%)
1	ND	0.72	3/793 (0.4%)	1.11	6/1073 (0.6%)
1	NE	0.52	0/864	0.90	4/1170 (0.3%)
2	AM	0.43	0/1507	0.72	4/2019 (0.2%)
2	Am	0.48	2/1496 (0.1%)	0.77	4/2005 (0.2%)
2	BM	0.41	0/1507	0.67	1/2019 (0.0%)
2	Bm	0.41	1/1496 (0.1%)	0.75	4/2005 (0.2%)
2	CM	0.42	0/1507	0.68	3/2019 (0.1%)
2	Cm	0.37	0/1496	0.73	4/2005 (0.2%)
2	DM	0.45	0/1507	0.76	2/2019 (0.1%)

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#|Z| >5	RMSZ	#|Z| >5
2	Dm	0.47	0/1496	0.83	6/2005 (0.3%)
2	EM	0.44	0/1507	0.75	2/2019 (0.1%)
2	Em	0.45	1/1496 (0.1%)	0.78	5/2005 (0.2%)
2	FM	0.47	0/1507	0.79	5/2019 (0.2%)
2	Fm	0.49	1/1496 (0.1%)	0.79	5/2005 (0.2%)
2	GM	0.43	0/1507	0.72	5/2019 (0.2%)
2	Gm	0.44	1/1496 (0.1%)	0.83	6/2005 (0.3%)
2	HM	0.42	0/1507	0.71	2/2019 (0.1%)
2	Hm	0.51	2/1496 (0.1%)	0.83	8/2005 (0.4%)
2	IM	0.42	0/1507	0.67	2/2019 (0.1%)
2	Im	0.43	1/1496 (0.1%)	0.77	5/2005 (0.2%)
2	JM	0.42	0/1507	0.68	0/2019
2	Jm	0.41	1/1496 (0.1%)	0.78	4/2005 (0.2%)
2	KM	0.41	0/1507	0.72	3/2019 (0.1%)
2	Km	0.45	0/1496	0.77	3/2005 (0.1%)
2	LM	0.52	2/1507 (0.1%)	0.82	3/2019 (0.1%)
2	Lm	0.46	0/1496	0.81	7/2005 (0.3%)
2	MM	0.49	1/1507 (0.1%)	0.76	3/2019 (0.1%)
2	Mm	0.45	0/1496	0.76	4/2005 (0.2%)
2	NM	0.44	0/1507	0.72	4/2019 (0.2%)
2	Nm	0.47	1/1496 (0.1%)	0.77	5/2005 (0.2%)
3	AT	0.58	0/2115	0.80	10/2847 (0.4%)
3	At	0.49	1/1638 (0.1%)	0.78	5/2203 (0.2%)
3	BT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Bt	0.49	1/1638 (0.1%)	0.78	6/2203 (0.3%)
3	CT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Ct	0.49	1/1638 (0.1%)	0.79	6/2203 (0.3%)
3	DT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Dt	0.49	1/1638 (0.1%)	0.79	6/2203 (0.3%)
3	ET	0.58	0/2115	0.80	10/2847 (0.4%)
3	Et	0.49	1/1638 (0.1%)	0.78	6/2203 (0.3%)
3	FT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Ft	0.49	1/1638 (0.1%)	0.79	6/2203 (0.3%)
3	GT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Gt	0.49	1/1638 (0.1%)	0.78	5/2203 (0.2%)
3	HT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Ht	0.49	1/1638 (0.1%)	0.78	6/2203 (0.3%)
3	IT	0.58	0/2115	0.80	10/2847 (0.4%)
3	It	0.49	1/1638 (0.1%)	0.79	6/2203 (0.3%)
3	JT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Jt	0.49	1/1638 (0.1%)	0.79	6/2203 (0.3%)
3	KT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Kt	0.49	1/1638 (0.1%)	0.78	6/2203 (0.3%)

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#|Z| >5	RMSZ	#|Z| >5
3	LT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Lt	0.49	1/1638 (0.1%)	0.79	6/2203 (0.3%)
3	MT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Mt	0.49	1/1638 (0.1%)	0.78	6/2203 (0.3%)
3	NT	0.58	0/2115	0.80	10/2847 (0.4%)
3	Nt	0.49	1/1638 (0.1%)	0.78	6/2203 (0.3%)
5	AX	0.67	0/1388	0.89	9/1870 (0.5%)
5	BX	0.67	0/1388	0.89	9/1870 (0.5%)
5	CX	0.67	0/1388	0.89	9/1870 (0.5%)
5	DX	0.67	0/1388	0.89	9/1870 (0.5%)
5	EX	0.67	0/1388	0.89	9/1870 (0.5%)
5	FX	0.67	0/1388	0.89	9/1870 (0.5%)
5	GX	0.67	0/1388	0.89	9/1870 (0.5%)
5	HX	0.67	0/1388	0.89	9/1870 (0.5%)
5	IX	0.67	0/1388	0.89	9/1870 (0.5%)
5	JX	0.67	0/1388	0.89	9/1870 (0.5%)
5	KX	0.67	0/1388	0.89	9/1870 (0.5%)
5	LX	0.67	0/1388	0.89	9/1870 (0.5%)
5	MX	0.67	0/1388	0.89	9/1870 (0.5%)
5	NX	0.67	0/1388	0.89	9/1870 (0.5%)
6	AY	0.64	0/1553	0.75	3/2106 (0.1%)
6	BY	0.64	0/1553	0.75	3/2106 (0.1%)
6	CY	0.64	0/1553	0.76	3/2106 (0.1%)
6	DY	0.64	0/1553	0.75	3/2106 (0.1%)
6	EY	0.64	0/1553	0.75	3/2106 (0.1%)
6	FY	0.64	0/1553	0.75	3/2106 (0.1%)
6	GY	0.64	0/1553	0.75	3/2106 (0.1%)
6	HY	0.64	0/1553	0.75	3/2106 (0.1%)
6	IY	0.64	0/1553	0.75	3/2106 (0.1%)
6	JY	0.64	0/1553	0.75	3/2106 (0.1%)
6	KY	0.64	0/1553	0.75	3/2106 (0.1%)
6	LY	0.64	0/1553	0.76	3/2106 (0.1%)
6	MY	0.64	0/1553	0.75	3/2106 (0.1%)
6	NY	0.64	0/1553	0.75	3/2106 (0.1%)
All	All	0.57	85/208236 (0.0%)	0.84	805/280910 (0.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	AA	0	3

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
1	AD	0	2
1	AE	0	1
1	BA	0	3
1	BD	0	2
1	CA	0	3
1	CB	0	1
1	CD	0	1
1	DA	0	3
1	DB	0	1
1	EA	0	3
1	ED	0	1
1	EE	0	1
1	FA	0	3
1	FB	0	1
1	FE	0	1
1	GA	0	3
1	GC	0	1
1	GD	0	2
1	GE	0	1
1	HA	0	3
1	HB	0	1
1	HD	0	2
1	IA	0	3
1	JA	0	3
1	JB	0	1
1	JD	0	2
1	KA	0	3
1	KB	0	1
1	KD	0	2
1	KE	0	2
1	LA	0	3
1	LB	0	1
1	LD	0	1
1	MA	0	3
1	MB	0	1
1	MD	0	1
1	ME	0	1
1	NA	0	3
1	NC	0	1
1	ND	0	1
1	NE	0	1
2	Am	0	1

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
2	Bm	0	2
2	Cm	0	1
2	Dm	0	1
2	Em	0	1
2	Fm	0	1
2	Gm	0	1
2	Hm	0	1
2	Im	0	1
2	Jm	0	1
2	Km	0	1
2	LM	0	1
2	Lm	0	1
2	Mm	0	1
2	Nm	0	1
5	AX	0	1
5	BX	0	1
5	CX	0	1
5	DX	0	1
5	EX	0	1
5	FX	0	1
5	GX	0	1
5	HX	0	1
5	IX	0	1
5	JX	0	1
5	KX	0	1
5	LX	0	1
5	MX	0	1
5	NX	0	1
All	All	0	107

The worst 5 of 85 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	KD	81	TYR	CE2-CZ	21.33	1.66	1.38
1	KD	81	TYR	CD2-CE2	21.19	1.71	1.39
1	KD	81	TYR	CG-CD2	18.52	1.63	1.39
1	KE	107	TYR	CB-CG	16.40	1.76	1.51
1	KD	81	TYR	CG-CD1	-14.30	1.20	1.39

The worst 5 of 805 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	KE	107	TYR	CB-CG-CD1	19.20	132.52	121.00
1	KD	81	TYR	CB-CG-CD2	16.50	130.90	121.00
2	LM	344	LYS	CD-CE-NZ	-15.77	75.43	111.70
1	KD	81	TYR	CZ-CE2-CD2	-14.49	106.76	119.80
1	KD	81	TYR	CD1-CE1-CZ	14.21	132.59	119.80

There are no chirality outliers.

5 of 107 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AA	253	TYR	Peptide
1	AA	275	ALA	Peptide
1	AA	304	LYS	Peptide
1	AD	131	GLY	Peptide
1	AD	94	TRP	Peptide

5.2 Too-close contacts (i)

Due to software issues we are unable to calculate clashes - this section is therefore empty.

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	AA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	AB	88/481 (18%)	81 (92%)	7 (8%)	0	100	100
1	AC	90/481 (19%)	85 (94%)	5 (6%)	0	100	100
1	AD	91/481 (19%)	80 (88%)	11 (12%)	0	100	100
1	AE	102/481 (21%)	97 (95%)	5 (5%)	0	100	100
1	BA	246/481 (51%)	217 (88%)	29 (12%)	0	100	100
1	BB	88/481 (18%)	84 (96%)	4 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	BC	90/481 (19%)	84 (93%)	6 (7%)	0	100	100
1	BD	91/481 (19%)	79 (87%)	11 (12%)	1 (1%)	14	44
1	BE	102/481 (21%)	96 (94%)	6 (6%)	0	100	100
1	CA	246/481 (51%)	217 (88%)	29 (12%)	0	100	100
1	CB	88/481 (18%)	83 (94%)	5 (6%)	0	100	100
1	CC	90/481 (19%)	85 (94%)	5 (6%)	0	100	100
1	CD	91/481 (19%)	77 (85%)	14 (15%)	0	100	100
1	CE	102/481 (21%)	99 (97%)	3 (3%)	0	100	100
1	DA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	DB	88/481 (18%)	81 (92%)	7 (8%)	0	100	100
1	DC	90/481 (19%)	86 (96%)	4 (4%)	0	100	100
1	DD	91/481 (19%)	76 (84%)	15 (16%)	0	100	100
1	DE	102/481 (21%)	94 (92%)	8 (8%)	0	100	100
1	EA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	EB	88/481 (18%)	81 (92%)	7 (8%)	0	100	100
1	EC	90/481 (19%)	85 (94%)	5 (6%)	0	100	100
1	ED	91/481 (19%)	78 (86%)	13 (14%)	0	100	100
1	EE	102/481 (21%)	98 (96%)	4 (4%)	0	100	100
1	FA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	FB	88/481 (18%)	82 (93%)	6 (7%)	0	100	100
1	FC	90/481 (19%)	84 (93%)	6 (7%)	0	100	100
1	FD	91/481 (19%)	77 (85%)	14 (15%)	0	100	100
1	FE	102/481 (21%)	96 (94%)	6 (6%)	0	100	100
1	GA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	GB	88/481 (18%)	84 (96%)	4 (4%)	0	100	100
1	GC	90/481 (19%)	84 (93%)	6 (7%)	0	100	100
1	GD	91/481 (19%)	79 (87%)	12 (13%)	0	100	100
1	GE	102/481 (21%)	95 (93%)	7 (7%)	0	100	100
1	HA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	HB	88/481 (18%)	83 (94%)	5 (6%)	0	100	100
1	HC	90/481 (19%)	85 (94%)	5 (6%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	HD	91/481 (19%)	77 (85%)	14 (15%)	0	100	100
1	HE	102/481 (21%)	96 (94%)	6 (6%)	0	100	100
1	IA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	IB	88/481 (18%)	84 (96%)	4 (4%)	0	100	100
1	IC	90/481 (19%)	86 (96%)	4 (4%)	0	100	100
1	ID	91/481 (19%)	78 (86%)	12 (13%)	1 (1%)	14	44
1	IE	102/481 (21%)	96 (94%)	6 (6%)	0	100	100
1	JA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	JB	88/481 (18%)	83 (94%)	5 (6%)	0	100	100
1	JC	90/481 (19%)	85 (94%)	5 (6%)	0	100	100
1	JD	91/481 (19%)	79 (87%)	12 (13%)	0	100	100
1	JE	102/481 (21%)	97 (95%)	5 (5%)	0	100	100
1	KA	246/481 (51%)	217 (88%)	29 (12%)	0	100	100
1	KB	88/481 (18%)	81 (92%)	7 (8%)	0	100	100
1	KC	90/481 (19%)	85 (94%)	5 (6%)	0	100	100
1	KD	91/481 (19%)	80 (88%)	11 (12%)	0	100	100
1	KE	102/481 (21%)	96 (94%)	6 (6%)	0	100	100
1	LA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	LB	88/481 (18%)	79 (90%)	9 (10%)	0	100	100
1	LC	90/481 (19%)	84 (93%)	6 (7%)	0	100	100
1	LD	91/481 (19%)	76 (84%)	15 (16%)	0	100	100
1	LE	102/481 (21%)	96 (94%)	6 (6%)	0	100	100
1	MA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	MB	88/481 (18%)	84 (96%)	4 (4%)	0	100	100
1	MC	90/481 (19%)	85 (94%)	5 (6%)	0	100	100
1	MD	91/481 (19%)	79 (87%)	12 (13%)	0	100	100
1	ME	102/481 (21%)	97 (95%)	5 (5%)	0	100	100
1	NA	246/481 (51%)	216 (88%)	30 (12%)	0	100	100
1	NB	88/481 (18%)	81 (92%)	7 (8%)	0	100	100
1	NC	90/481 (19%)	85 (94%)	5 (6%)	0	100	100
1	ND	91/481 (19%)	80 (88%)	11 (12%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	NE	102/481 (21%)	97 (95%)	5 (5%)	0	100	100
2	AM	178/376 (47%)	170 (96%)	8 (4%)	0	100	100
2	Am	177/376 (47%)	167 (94%)	10 (6%)	0	100	100
2	BM	178/376 (47%)	170 (96%)	8 (4%)	0	100	100
2	Bm	177/376 (47%)	169 (96%)	8 (4%)	0	100	100
2	CM	178/376 (47%)	170 (96%)	8 (4%)	0	100	100
2	Cm	177/376 (47%)	166 (94%)	11 (6%)	0	100	100
2	DM	178/376 (47%)	172 (97%)	6 (3%)	0	100	100
2	Dm	177/376 (47%)	168 (95%)	9 (5%)	0	100	100
2	EM	178/376 (47%)	169 (95%)	9 (5%)	0	100	100
2	Em	177/376 (47%)	169 (96%)	8 (4%)	0	100	100
2	FM	178/376 (47%)	171 (96%)	7 (4%)	0	100	100
2	Fm	177/376 (47%)	166 (94%)	11 (6%)	0	100	100
2	GM	178/376 (47%)	167 (94%)	11 (6%)	0	100	100
2	Gm	177/376 (47%)	167 (94%)	10 (6%)	0	100	100
2	HM	178/376 (47%)	169 (95%)	9 (5%)	0	100	100
2	Hm	177/376 (47%)	167 (94%)	10 (6%)	0	100	100
2	IM	178/376 (47%)	171 (96%)	7 (4%)	0	100	100
2	Im	177/376 (47%)	168 (95%)	9 (5%)	0	100	100
2	JM	178/376 (47%)	170 (96%)	8 (4%)	0	100	100
2	Jm	177/376 (47%)	168 (95%)	9 (5%)	0	100	100
2	KM	178/376 (47%)	172 (97%)	6 (3%)	0	100	100
2	Km	177/376 (47%)	169 (96%)	8 (4%)	0	100	100
2	LM	178/376 (47%)	168 (94%)	10 (6%)	0	100	100
2	Lm	177/376 (47%)	170 (96%)	7 (4%)	0	100	100
2	MM	178/376 (47%)	170 (96%)	8 (4%)	0	100	100
2	Mm	177/376 (47%)	168 (95%)	9 (5%)	0	100	100
2	NM	178/376 (47%)	169 (95%)	9 (5%)	0	100	100
2	Nm	177/376 (47%)	168 (95%)	9 (5%)	0	100	100
3	AT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	At	186/280 (66%)	179 (96%)	7 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	BT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Bt	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
3	CT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Ct	186/280 (66%)	179 (96%)	7 (4%)	0	100	100
3	DT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Dt	186/280 (66%)	179 (96%)	7 (4%)	0	100	100
3	ET	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Et	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
3	FT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Ft	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
3	GT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Gt	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
3	HT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Ht	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
3	IT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	It	186/280 (66%)	179 (96%)	7 (4%)	0	100	100
3	JT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Jt	186/280 (66%)	179 (96%)	7 (4%)	0	100	100
3	KT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Kt	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
3	LT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Lt	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
3	MT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Mt	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
3	NT	251/280 (90%)	242 (96%)	9 (4%)	0	100	100
3	Nt	186/280 (66%)	178 (96%)	8 (4%)	0	100	100
5	AX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	BX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	CX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	DX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	EX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
5	FX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	GX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	HX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	IX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	JX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	KX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	LX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	MX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
5	NX	164/522 (31%)	156 (95%)	8 (5%)	0	100	100
6	AY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	BY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	CY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	DY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	EY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	FY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	GY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	HY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	IY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	JY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	KY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	LY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	MY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
6	NY	199/1927 (10%)	190 (96%)	9 (4%)	0	100	100
All	All	24808/86324 (29%)	23268 (94%)	1538 (6%)	2 (0%)	100	100

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	ID	160	ASN
1	BD	160	ASN

5.3.2 Protein sidechains ⓘ

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	AA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	AB	79/433 (18%)	79 (100%)	0	100	100
1	AC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	AD	83/433 (19%)	82 (99%)	1 (1%)	71	85
1	AE	91/433 (21%)	91 (100%)	0	100	100
1	BA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	BB	79/433 (18%)	79 (100%)	0	100	100
1	BC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	BD	83/433 (19%)	80 (96%)	3 (4%)	35	63
1	BE	91/433 (21%)	91 (100%)	0	100	100
1	CA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	CB	79/433 (18%)	79 (100%)	0	100	100
1	CC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	CD	83/433 (19%)	82 (99%)	1 (1%)	71	85
1	CE	91/433 (21%)	91 (100%)	0	100	100
1	DA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	DB	79/433 (18%)	79 (100%)	0	100	100
1	DC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	DD	83/433 (19%)	81 (98%)	2 (2%)	49	74
1	DE	91/433 (21%)	91 (100%)	0	100	100
1	EA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	EB	79/433 (18%)	79 (100%)	0	100	100
1	EC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	ED	83/433 (19%)	81 (98%)	2 (2%)	49	74
1	EE	91/433 (21%)	91 (100%)	0	100	100
1	FA	220/433 (51%)	214 (97%)	6 (3%)	44	70

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	FB	79/433 (18%)	79 (100%)	0	100	100
1	FC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	FD	83/433 (19%)	81 (98%)	2 (2%)	49	74
1	FE	91/433 (21%)	91 (100%)	0	100	100
1	GA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	GB	79/433 (18%)	79 (100%)	0	100	100
1	GC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	GD	83/433 (19%)	82 (99%)	1 (1%)	71	85
1	GE	91/433 (21%)	91 (100%)	0	100	100
1	HA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	HB	79/433 (18%)	79 (100%)	0	100	100
1	HC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	HD	83/433 (19%)	82 (99%)	1 (1%)	71	85
1	HE	91/433 (21%)	91 (100%)	0	100	100
1	IA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	IB	79/433 (18%)	79 (100%)	0	100	100
1	IC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	ID	83/433 (19%)	82 (99%)	1 (1%)	71	85
1	IE	91/433 (21%)	91 (100%)	0	100	100
1	JA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	JB	79/433 (18%)	79 (100%)	0	100	100
1	JC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	JD	83/433 (19%)	82 (99%)	1 (1%)	71	85
1	JE	91/433 (21%)	91 (100%)	0	100	100
1	KA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	KB	79/433 (18%)	79 (100%)	0	100	100
1	KC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	KD	83/433 (19%)	82 (99%)	1 (1%)	71	85
1	KE	91/433 (21%)	90 (99%)	1 (1%)	73	86
1	LA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	LB	79/433 (18%)	79 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	LC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	LD	83/433 (19%)	81 (98%)	2 (2%)	49	74
1	LE	91/433 (21%)	91 (100%)	0	100	100
1	MA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	MB	79/433 (18%)	79 (100%)	0	100	100
1	MC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	MD	83/433 (19%)	82 (99%)	1 (1%)	71	85
1	ME	91/433 (21%)	91 (100%)	0	100	100
1	NA	220/433 (51%)	214 (97%)	6 (3%)	44	70
1	NB	79/433 (18%)	79 (100%)	0	100	100
1	NC	80/433 (18%)	79 (99%)	1 (1%)	69	84
1	ND	83/433 (19%)	81 (98%)	2 (2%)	49	74
1	NE	91/433 (21%)	91 (100%)	0	100	100
2	AM	168/342 (49%)	168 (100%)	0	100	100
2	Am	167/342 (49%)	166 (99%)	1 (1%)	86	94
2	BM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Bm	167/342 (49%)	166 (99%)	1 (1%)	86	94
2	CM	168/342 (49%)	166 (99%)	2 (1%)	71	85
2	Cm	167/342 (49%)	166 (99%)	1 (1%)	86	94
2	DM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Dm	167/342 (49%)	165 (99%)	2 (1%)	71	85
2	EM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Em	167/342 (49%)	166 (99%)	1 (1%)	86	94
2	FM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Fm	167/342 (49%)	167 (100%)	0	100	100
2	GM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Gm	167/342 (49%)	166 (99%)	1 (1%)	86	94
2	HM	168/342 (49%)	168 (100%)	0	100	100
2	Hm	167/342 (49%)	166 (99%)	1 (1%)	86	94
2	IM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Im	167/342 (49%)	166 (99%)	1 (1%)	86	94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	JM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Jm	167/342 (49%)	167 (100%)	0	100	100
2	KM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Km	167/342 (49%)	166 (99%)	1 (1%)	86	94
2	LM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Lm	167/342 (49%)	166 (99%)	1 (1%)	86	94
2	MM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Mm	167/342 (49%)	167 (100%)	0	100	100
2	NM	168/342 (49%)	167 (99%)	1 (1%)	86	94
2	Nm	167/342 (49%)	166 (99%)	1 (1%)	86	94
3	AT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	At	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	BT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Bt	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	CT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Ct	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	DT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Dt	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	ET	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Et	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	FT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Ft	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	GT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Gt	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	HT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Ht	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	IT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	It	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	JT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Jt	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	KT	234/255 (92%)	227 (97%)	7 (3%)	41	68

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	Kt	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	LT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Lt	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	MT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Mt	181/255 (71%)	175 (97%)	6 (3%)	38	66
3	NT	234/255 (92%)	227 (97%)	7 (3%)	41	68
3	Nt	181/255 (71%)	175 (97%)	6 (3%)	38	66
5	AX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	BX	148/470 (32%)	138 (93%)	10 (7%)	16	45
5	CX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	DX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	EX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	FX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	GX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	HX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	IX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	JX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	KX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	LX	148/470 (32%)	138 (93%)	10 (7%)	16	45
5	MX	148/470 (32%)	139 (94%)	9 (6%)	18	48
5	NX	148/470 (32%)	139 (94%)	9 (6%)	18	48
6	AY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	BY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	CY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	DY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	EY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	FY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	GY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	HY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	IY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	JY	171/1724 (10%)	163 (95%)	8 (5%)	26	57

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
6	KY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	LY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	MY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
6	NY	171/1724 (10%)	163 (95%)	8 (5%)	26	57
All	All	22708/77742 (29%)	22141 (98%)	567 (2%)	50	72

5 of 567 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	LX	365	LYS
6	LY	1783	GLU
5	LX	356	MET
3	Mt	37	THR
6	EY	1780	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 71 such sidechains are listed below:

Mol	Chain	Res	Type
2	LM	244	ASN
2	Lm	207	ASN
2	Mm	207	ASN
2	EM	244	ASN
1	EE	87	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

The following chains have linkage breaks:

Mol	Chain	Number of breaks
4	AU	8
4	BU	8
4	CU	8
4	DU	8
4	EU	8
4	FU	8
4	GU	8
4	HU	8
4	IU	8
4	JU	8
4	KU	8
4	LU	8
4	MU	8
4	NU	8

The worst 5 of 112 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	AU	723:UNK	C	801:UNK	N	98.58
1	BU	723:UNK	C	801:UNK	N	98.58
1	CU	723:UNK	C	801:UNK	N	98.58
1	DU	723:UNK	C	801:UNK	N	98.58
1	EU	723:UNK	C	801:UNK	N	98.58

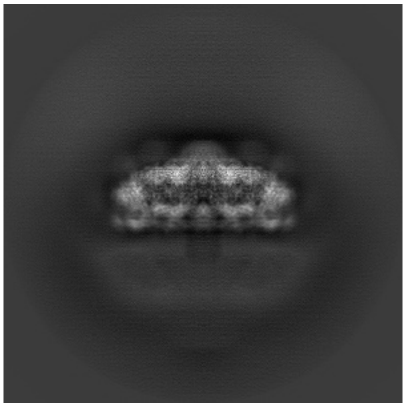
6 Map visualisation (i)

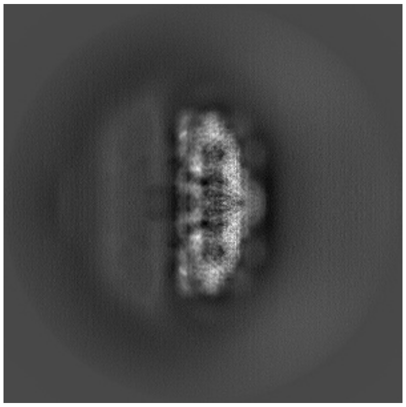
This section contains visualisations of the EMDB entry EMD-22081. These allow visual inspection of the internal detail of the map and identification of artifacts.

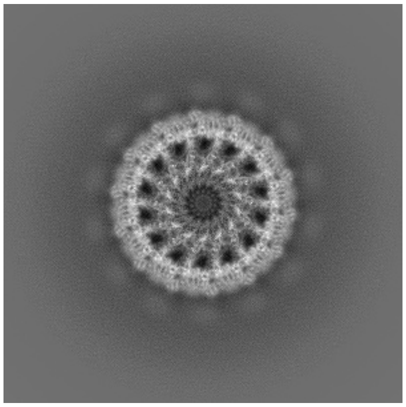
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

6.1.1 Primary map


X

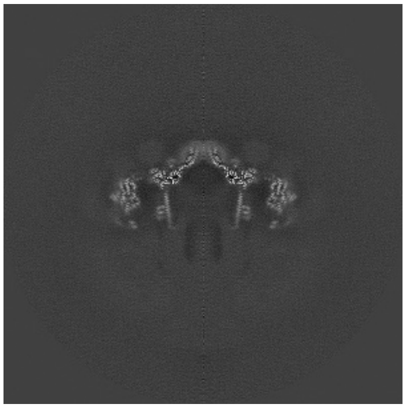

Y

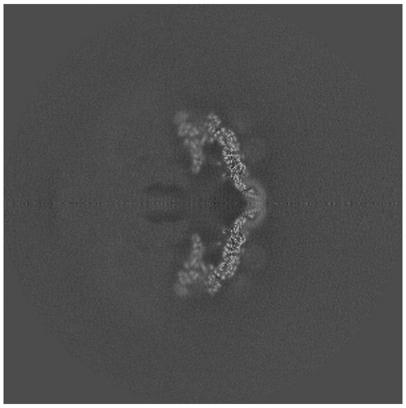

Z

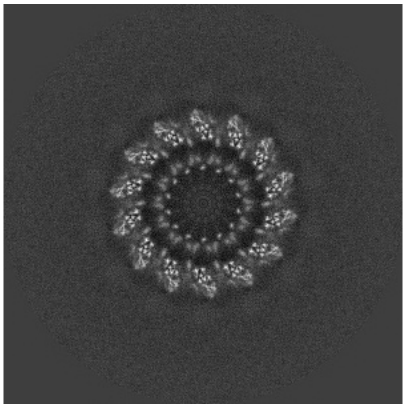
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

6.2.1 Primary map


X Index: 255

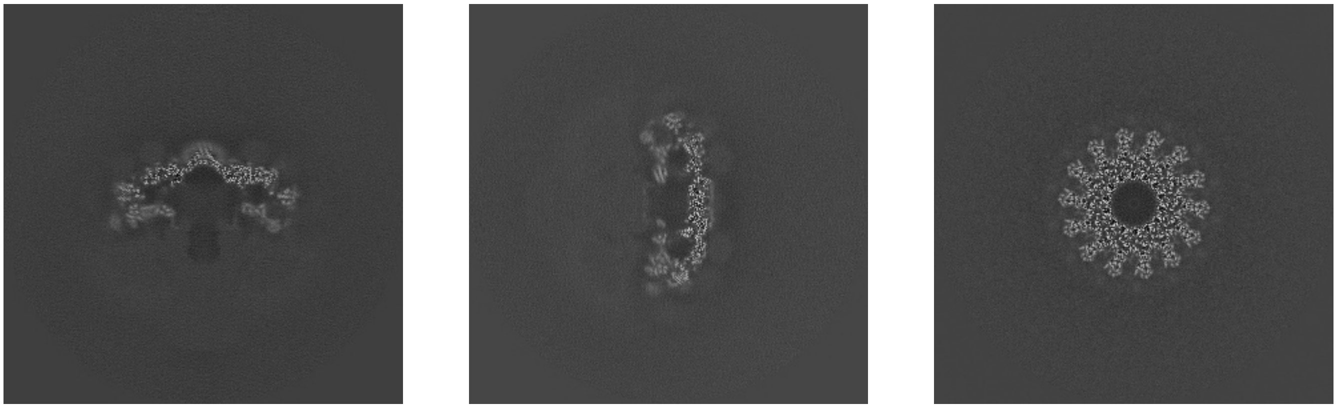

Y Index: 255


Z Index: 255

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

6.3 Largest variance slices ⓘ

6.3.1 Primary map

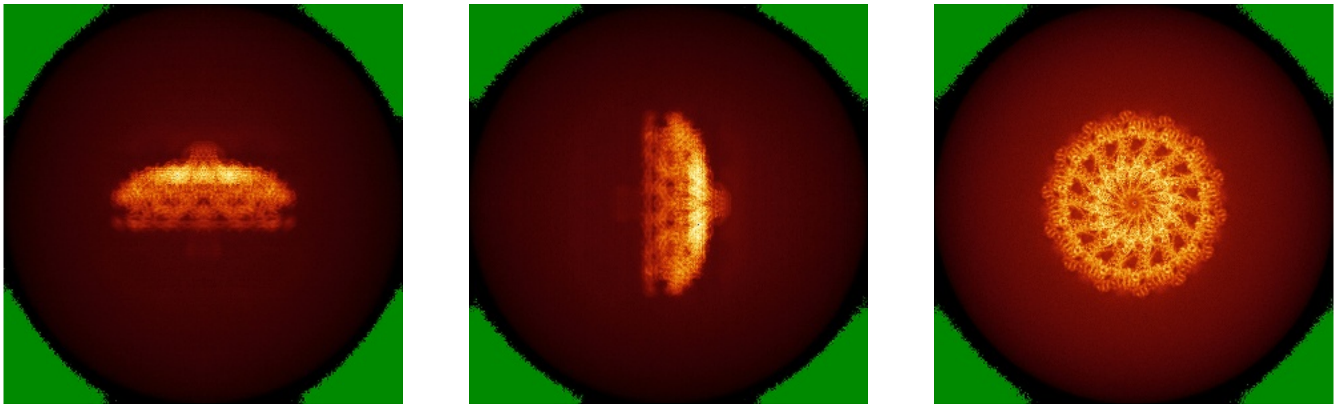


X Index: 241 Y Index: 289 Z Index: 291

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

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

6.4.1 Primary map

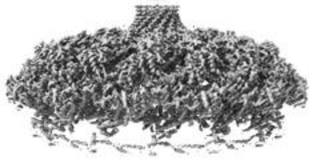


X Y Z

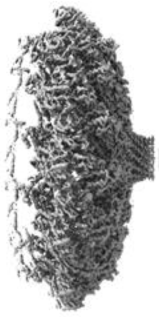
The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views ⓘ

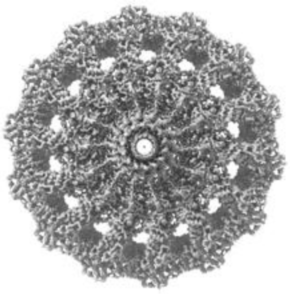
6.5.1 Primary map



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.

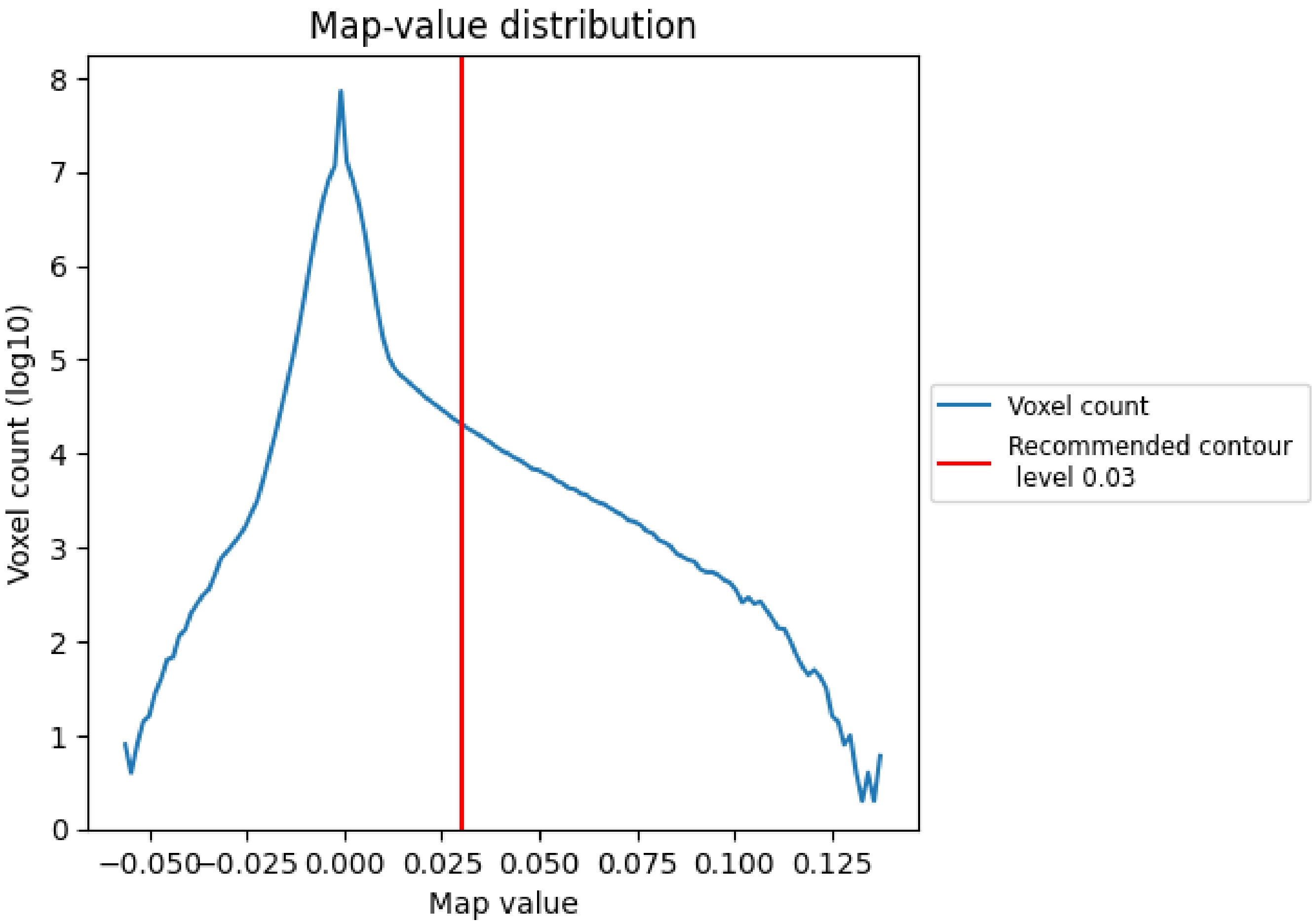
6.6 Mask visualisation ⓘ

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

7 Map analysis ⓘ

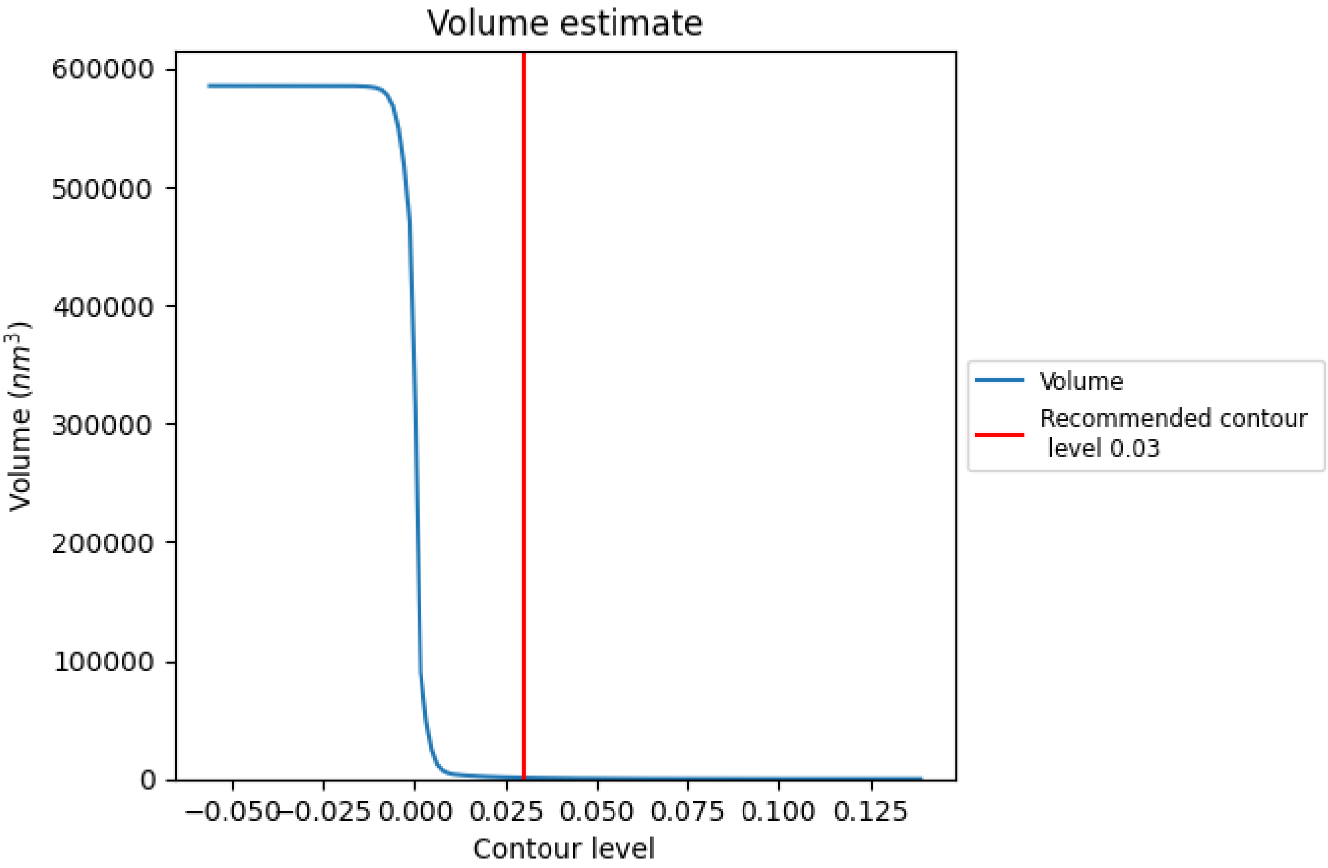
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution ⓘ



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.

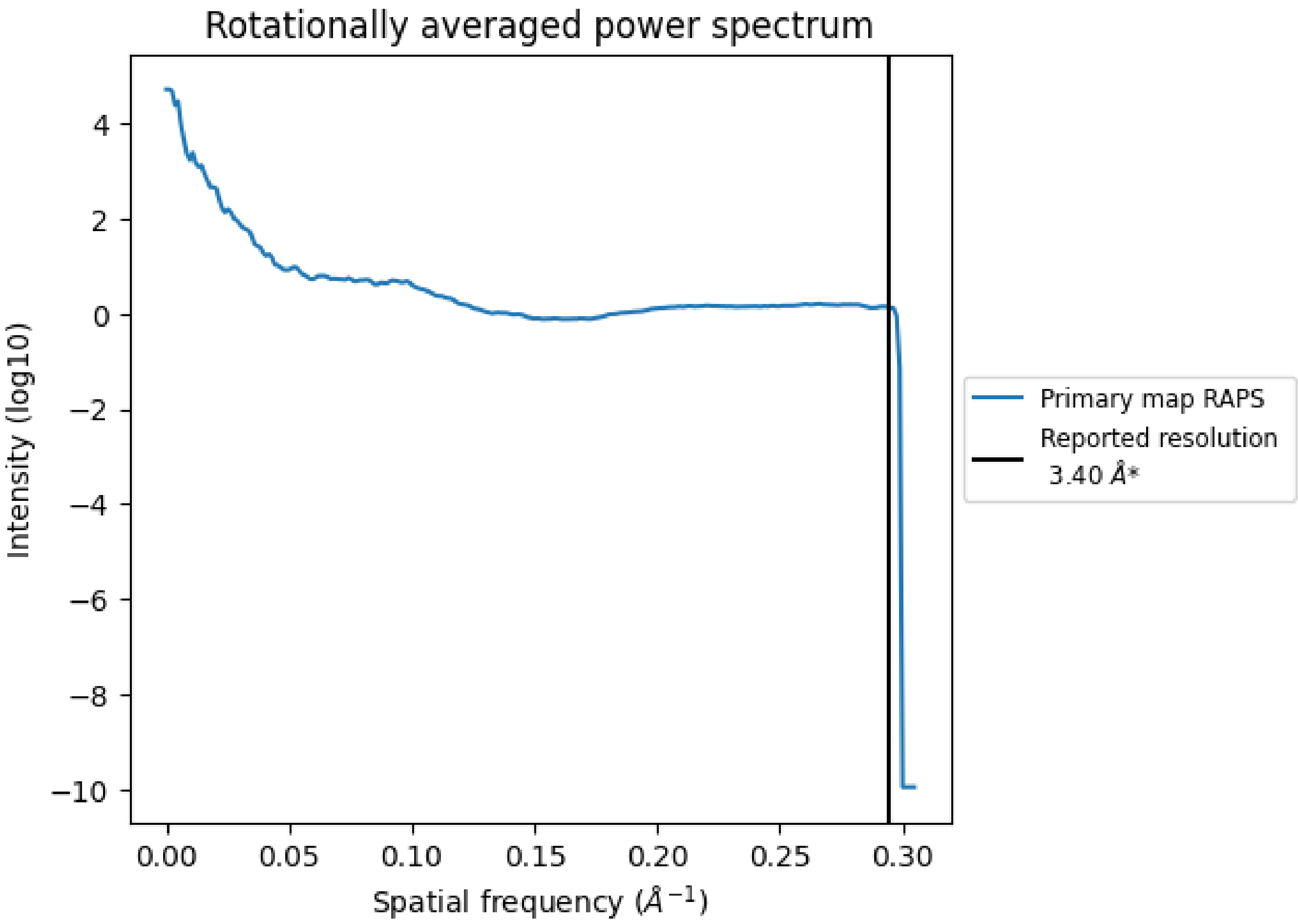
7.2 Volume estimate ⓘ



The volume at the recommended contour level is 1078 nm^3; this corresponds to an approximate mass of 974 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 ⓘ

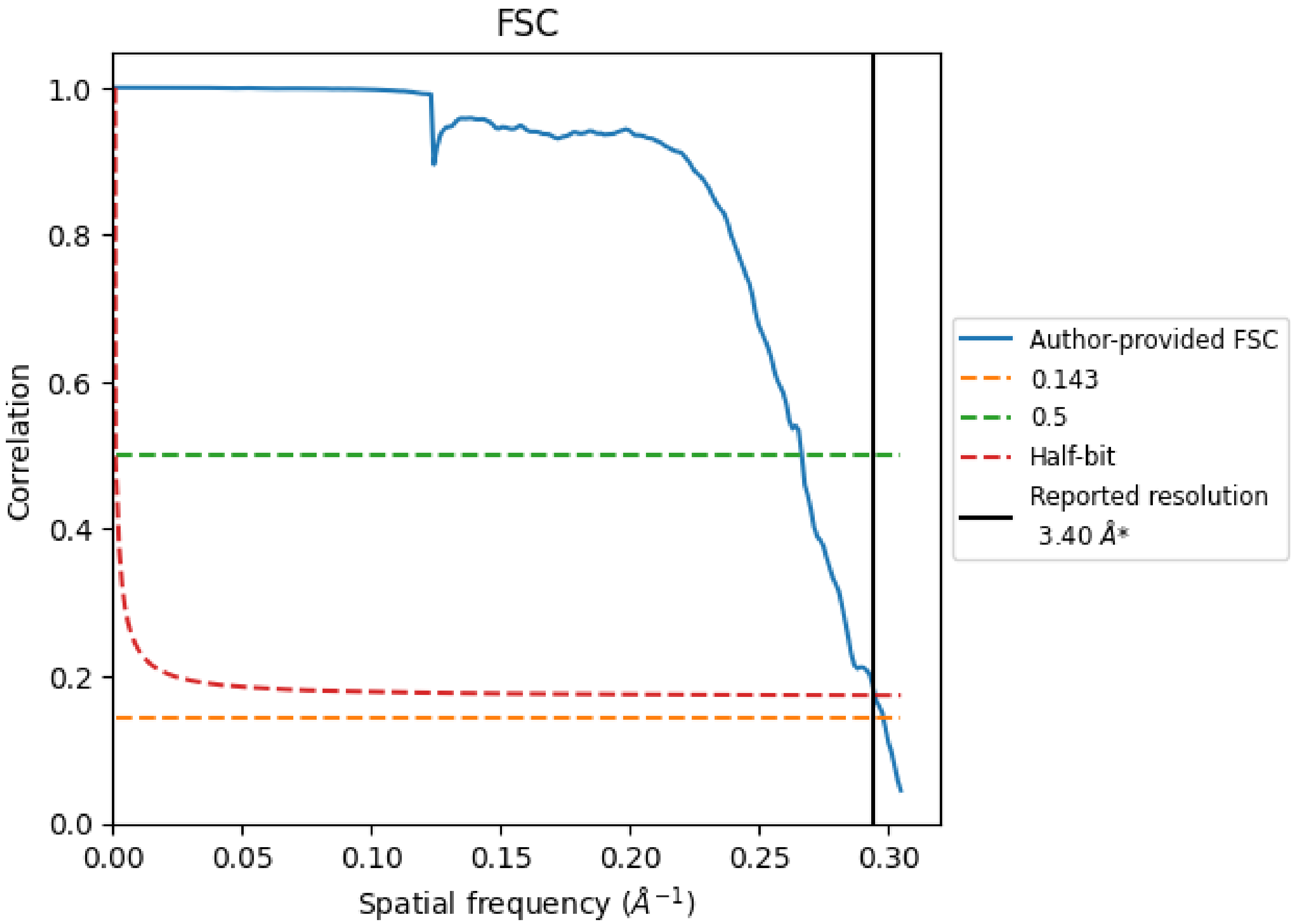


*Reported resolution corresponds to spatial frequency of 0.294 Å^{-1}

8 Fourier-Shell correlation ⓘ

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 ⓘ



*Reported resolution corresponds to spatial frequency of 0.294 $Å^{-1}$

8.2 Resolution estimates ⓘ

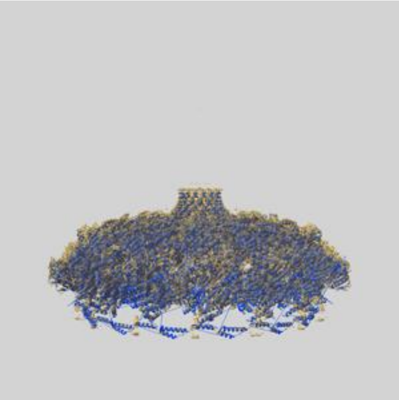
Resolution estimate (Å)	**Estimation criterion (FSC cut-off)**		
	0.143	0.5	Half-bit
Reported by author	3.40	-	-
Author-provided FSC curve	3.35	3.75	3.39
Unmasked-calculated*	-	-	-

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

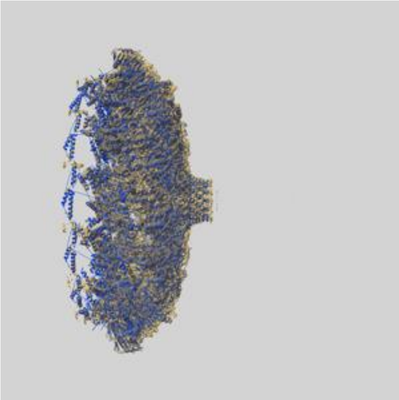
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-22081 and PDB model 6X6S. Per-residue inclusion information can be found in section 3 on page 21.

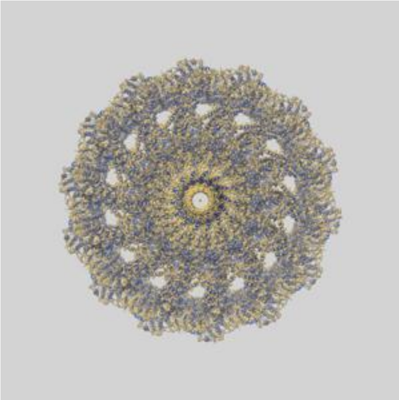
9.1 Map-model overlay



X



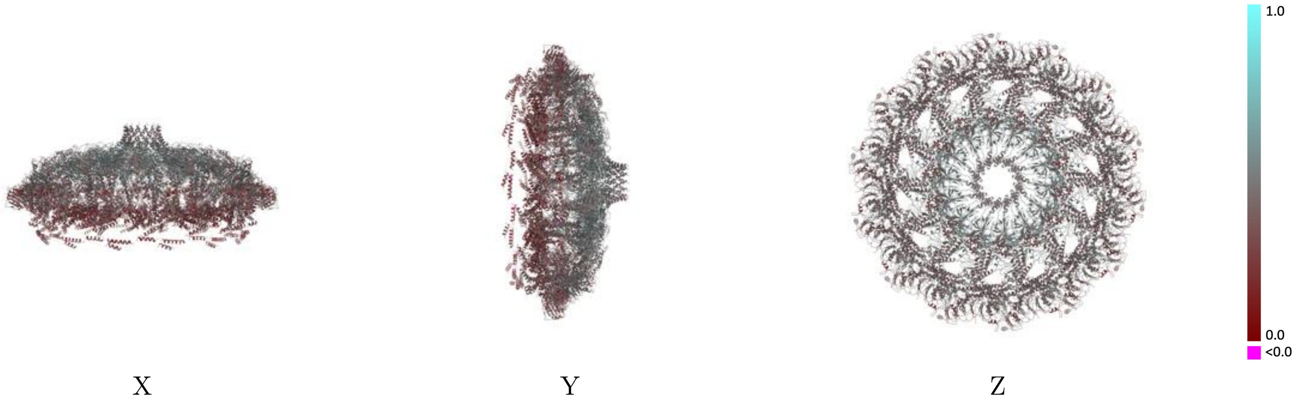
Y



Z

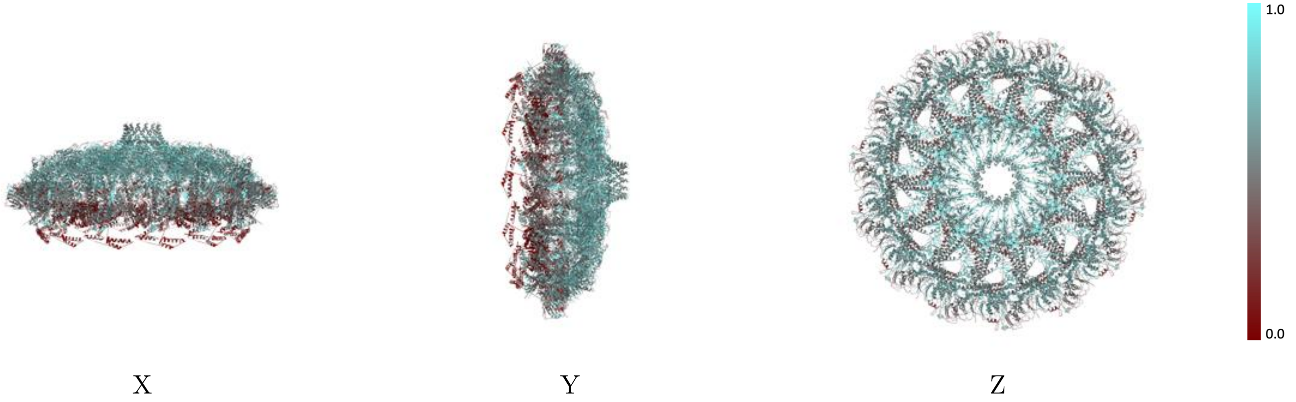
The images above show the 3D surface view of the map at the recommended contour level 0.03 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2 Q-score mapped to coordinate model ⓘ



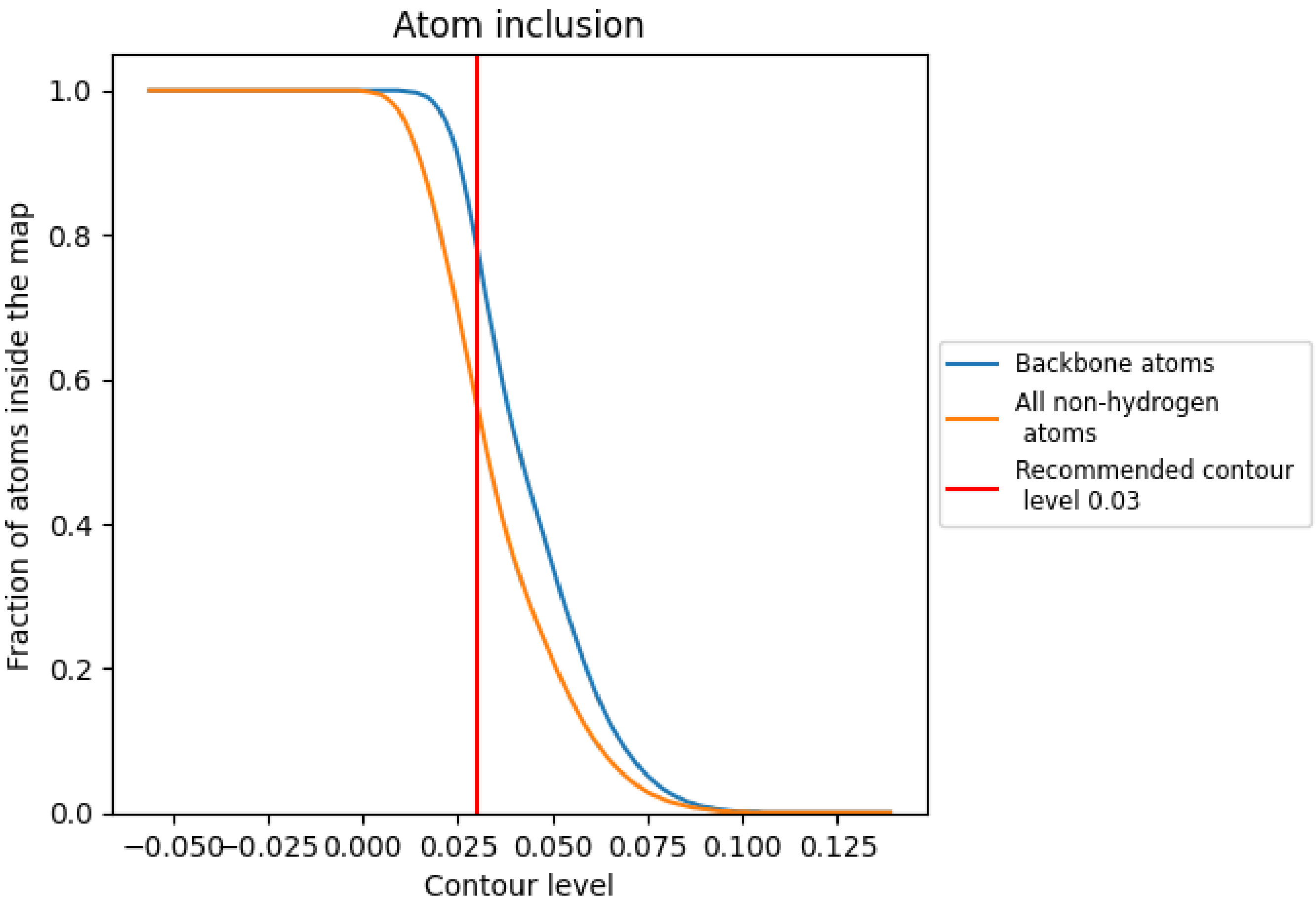
The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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 ⓘ



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 ⓘ



At the recommended contour level, 79% of all backbone atoms, 57% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary ⓘ

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	0.5680	0.4030
AA	0.5830	0.4110
AB	0.6710	0.4540
AC	0.6500	0.4320
AD	0.4750	0.3500
AE	0.4190	0.3380
AM	0.5260	0.3640
AT	0.6750	0.4620
AU	0.4120	0.3410
AX	0.7080	0.4760
AY	0.7460	0.5050
Am	0.2610	0.2510
At	0.5610	0.4120
BA	0.5780	0.4100
BB	0.6740	0.4410
BC	0.6360	0.4250
BD	0.4840	0.3550
BE	0.4220	0.3390
BM	0.5250	0.3670
BT	0.6780	0.4630
BU	0.4110	0.3360
BX	0.7080	0.4750
BY	0.7390	0.5020
Bm	0.2600	0.2460
Bt	0.5670	0.4100
CA	0.5820	0.4110
CB	0.6660	0.4380
CC	0.6370	0.4250
CD	0.4650	0.3530
CE	0.4260	0.3360
CM	0.5210	0.3620
CT	0.6750	0.4620
CU	0.4170	0.3380
CX	0.7070	0.4750
CY	0.7480	0.5020

Continued on next page...

Continued from previous page...

Chain	Atom inclusion	Q-score
Cm	0.2610	0.2470
Ct	0.5660	0.4090
DA	0.5840	0.4090
DB	0.6590	0.4340
DC	0.6410	0.4330
DD	0.4790	0.3550
DE	0.4300	0.3380
DM	0.5150	0.3640
DT	0.6770	0.4600
DU	0.4120	0.3350
DX	0.7100	0.4760
DY	0.7420	0.5020
Dm	0.2570	0.2480
Dt	0.5650	0.4100
EA	0.5780	0.4050
EB	0.6600	0.4390
EC	0.6500	0.4230
ED	0.4720	0.3520
EE	0.4220	0.3280
EM	0.5190	0.3620
ET	0.6810	0.4620
EU	0.4140	0.3370
EX	0.7000	0.4730
EY	0.7410	0.5030
Em	0.2600	0.2470
Et	0.5740	0.4100
FA	0.5790	0.4050
FB	0.6740	0.4380
FC	0.6430	0.4240
FD	0.4840	0.3660
FE	0.4260	0.3340
FM	0.5220	0.3640
FT	0.6830	0.4610
FU	0.4040	0.3350
FX	0.7060	0.4730
FY	0.7390	0.4990
Fm	0.2650	0.2470
Ft	0.5700	0.4110
GA	0.5770	0.4050
GB	0.6640	0.4380
GC	0.6440	0.4280
GD	0.4710	0.3550

Continued on next page...

Continued from previous page...

Chain	Atom inclusion	Q-score
GE	0.4240	0.3330
GM	0.5190	0.3660
GT	0.6810	0.4620
GU	0.4110	0.3310
GX	0.7080	0.4750
GY	0.7440	0.4980
Gm	0.2620	0.2470
Gt	0.5700	0.4140
HA	0.5830	0.4120
HB	0.6740	0.4520
HC	0.6510	0.4250
HD	0.4740	0.3540
HE	0.4200	0.3390
HM	0.5260	0.3650
HT	0.6750	0.4610
HU	0.4030	0.3370
HX	0.7050	0.4760
HY	0.7480	0.5030
Hm	0.2570	0.2480
Ht	0.5620	0.4100
IA	0.5780	0.4090
IB	0.6680	0.4390
IC	0.6410	0.4290
ID	0.4790	0.3520
IE	0.4220	0.3440
IM	0.5240	0.3650
IT	0.6770	0.4640
IU	0.4110	0.3360
IX	0.7040	0.4770
IY	0.7390	0.5000
Im	0.2610	0.2450
It	0.5640	0.4050
JA	0.5830	0.4130
JB	0.6700	0.4380
JC	0.6450	0.4210
JD	0.4690	0.3580
JE	0.4210	0.3330
JM	0.5200	0.3610
JT	0.6750	0.4620
JU	0.4220	0.3370
JX	0.7080	0.4750
JY	0.7470	0.5010

Continued on next page...

Continued from previous page...

Chain	Atom inclusion	Q-score
Jm	0.2630	0.2460
Jt	0.5670	0.4080
KA	0.5820	0.4100
KB	0.6610	0.4390
KC	0.6480	0.4310
KD	0.4750	0.3550
KE	0.4270	0.3370
KM	0.5120	0.3630
KT	0.6780	0.4600
KU	0.4070	0.3370
KX	0.7170	0.4770
KY	0.7420	0.5020
Km	0.2580	0.2480
Kt	0.5670	0.4090
LA	0.5770	0.4090
LB	0.6660	0.4440
LC	0.6540	0.4280
LD	0.4740	0.3490
LE	0.4210	0.3410
LM	0.5200	0.3600
LT	0.6800	0.4640
LU	0.4070	0.3330
LX	0.7080	0.4750
LY	0.7370	0.5050
Lm	0.2580	0.2450
Lt	0.5720	0.4110
MA	0.5810	0.4080
MB	0.6670	0.4320
MC	0.6400	0.4310
MD	0.4820	0.3650
ME	0.4250	0.3410
MM	0.5200	0.3620
MT	0.6810	0.4630
MU	0.3990	0.3380
MX	0.7080	0.4720
MY	0.7390	0.5000
Mm	0.2630	0.2500
Mt	0.5690	0.4120
NA	0.5810	0.4120
NB	0.6700	0.4380
NC	0.6470	0.4280
ND	0.4820	0.3560

Continued on next page...

Continued from previous page...

Chain	Atom inclusion	Q-score
NE	0.4250	0.3380
NM	0.5190	0.3630
NT	0.6750	0.4620
NU	0.4170	0.3350
NX	0.7080	0.4750
NY	0.7420	0.5010
Nm	0.2630	0.2460
Nt	0.5710	0.4130