wwPDB EM Validation Summary Report ⓘ

Nov 8, 2022 – 05:06 PM EST

PDB ID	:	6ODI
EMDB ID	:	EMD-20020
Title	:	Structure of CagY from a cryo-EM reconstruction of a T4SS
Authors	:	Chung, J.M.; Sheedlo, M.J.; Campbell, A.; Sawhney, N.; Frick-Cheng, A.E.; Lacy, D.B.; Cover, T.L.; Ohi, M.D.
Deposited on	:	2019-03-26
Resolution	:	3.80 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 ⓘ 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 ⓘ) were used in the production of this report:

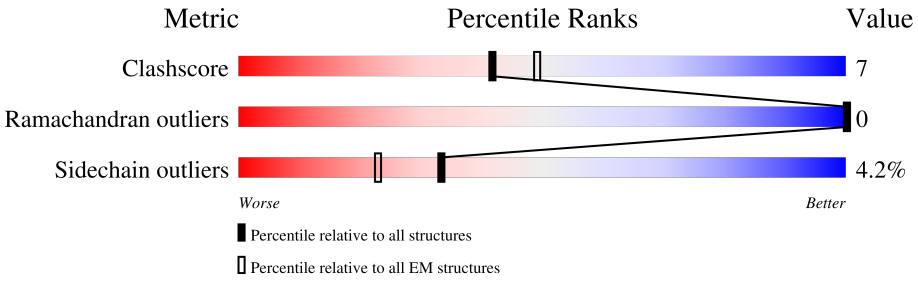
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.80 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)
Clashscore	158937	4297
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	b	1927	10% 90%
1	c	1927	10% 90%
1	d	1927	10% 90%
1	e	1927	10% 90%
1	f	1927	10% 90%
1	g	1927	10% 90%
1	h	1927	10% 90%
1	i	1927	10% 90%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	j	1927	10% 90%
1	k	1927	10% 90%
1	l	1927	10% 90%
1	m	1927	10% 90%
1	n	1927	10% 90%
1	o	1927	10% 90%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 20902 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 CagY.

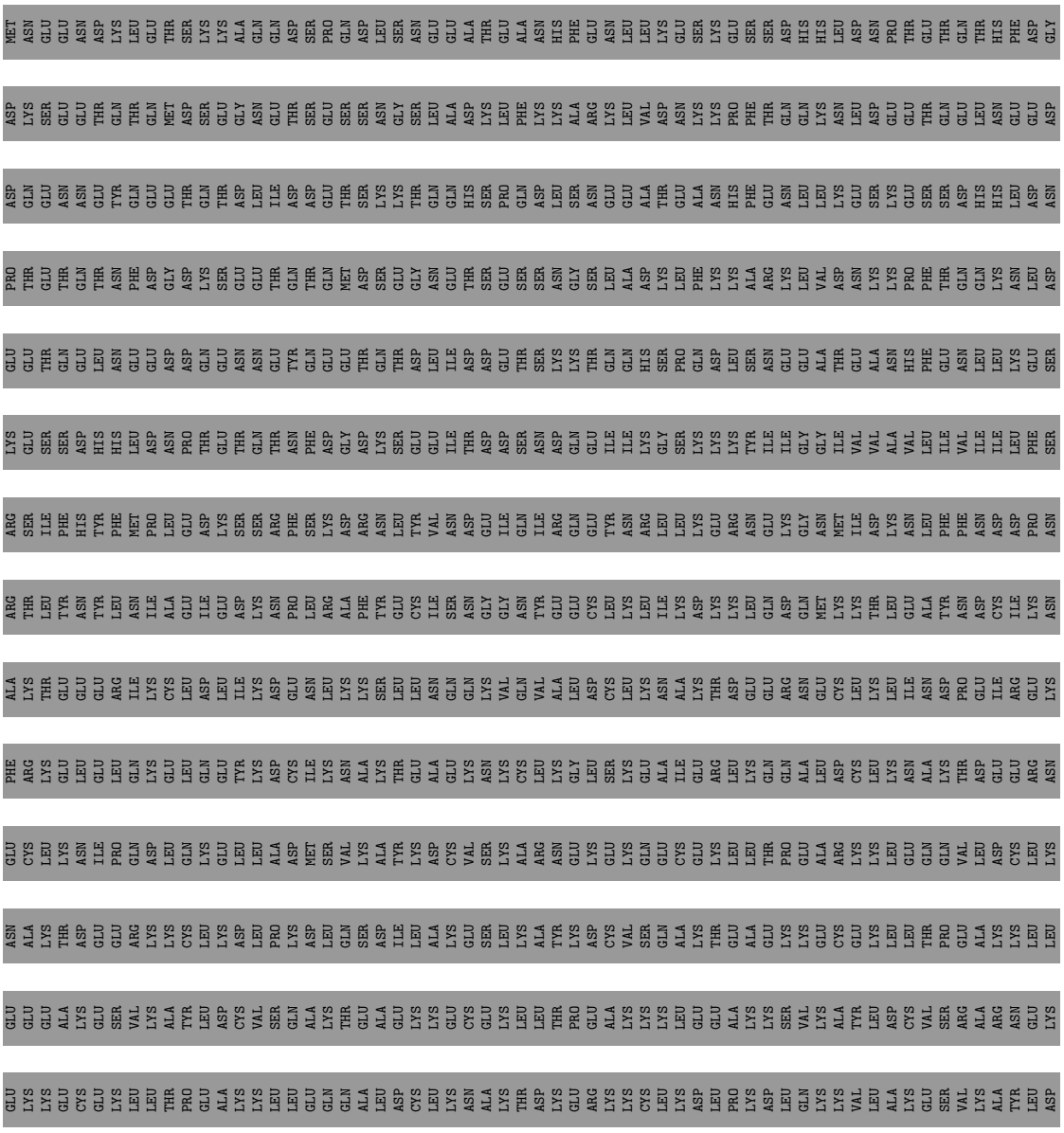
Mol	Chain	Residues	Atoms					AltConf	Trace
1	b	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	c	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	d	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	e	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	f	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	g	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	h	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	i	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	j	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	k	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	l	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	m	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	n	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		
1	o	198	Total	C	N	O	S	0	0
			1493	959	241	284	9		

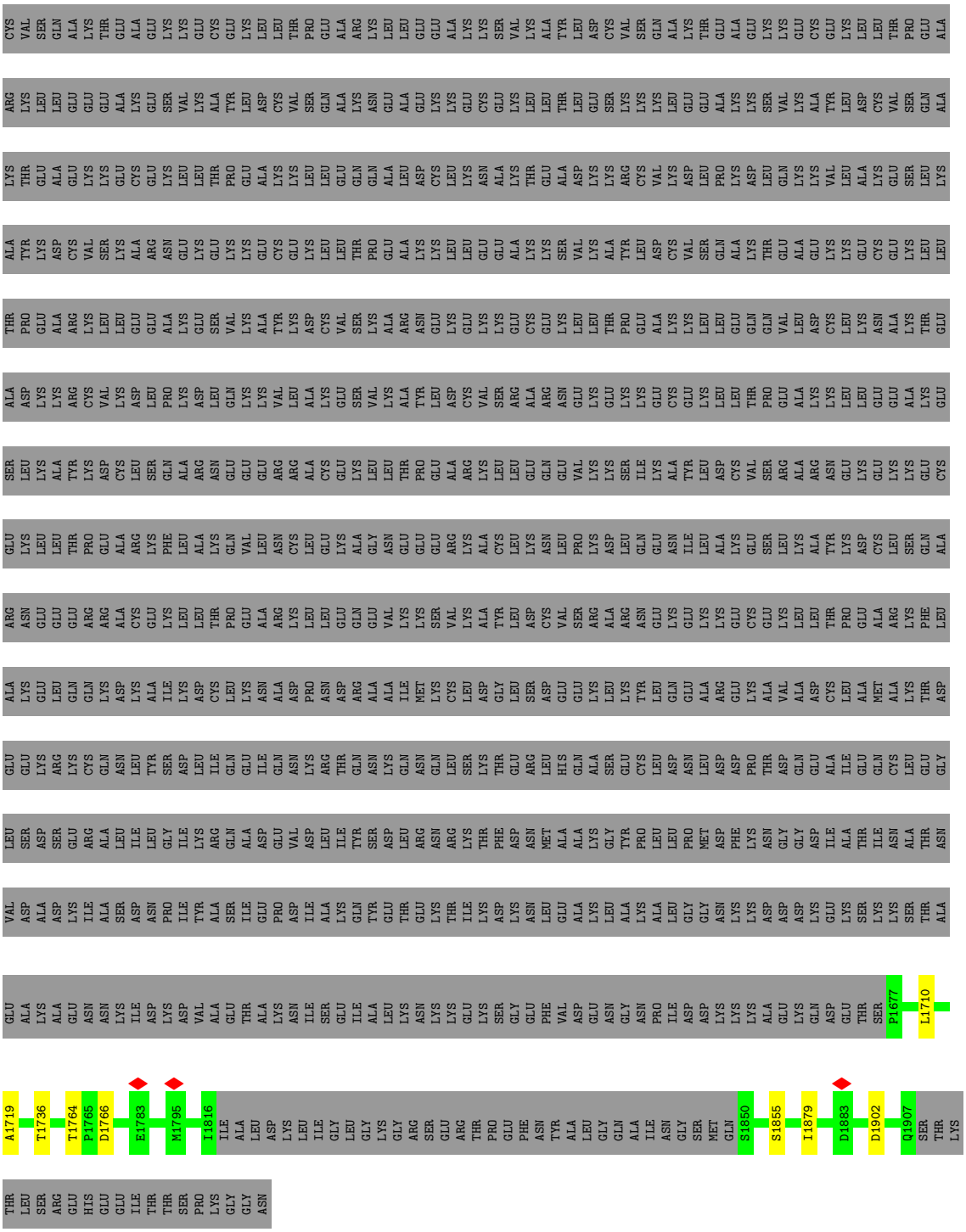
3 Residue-property plots (i)

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

- Molecule 1: Type IV secretion system apparatus protein CagY

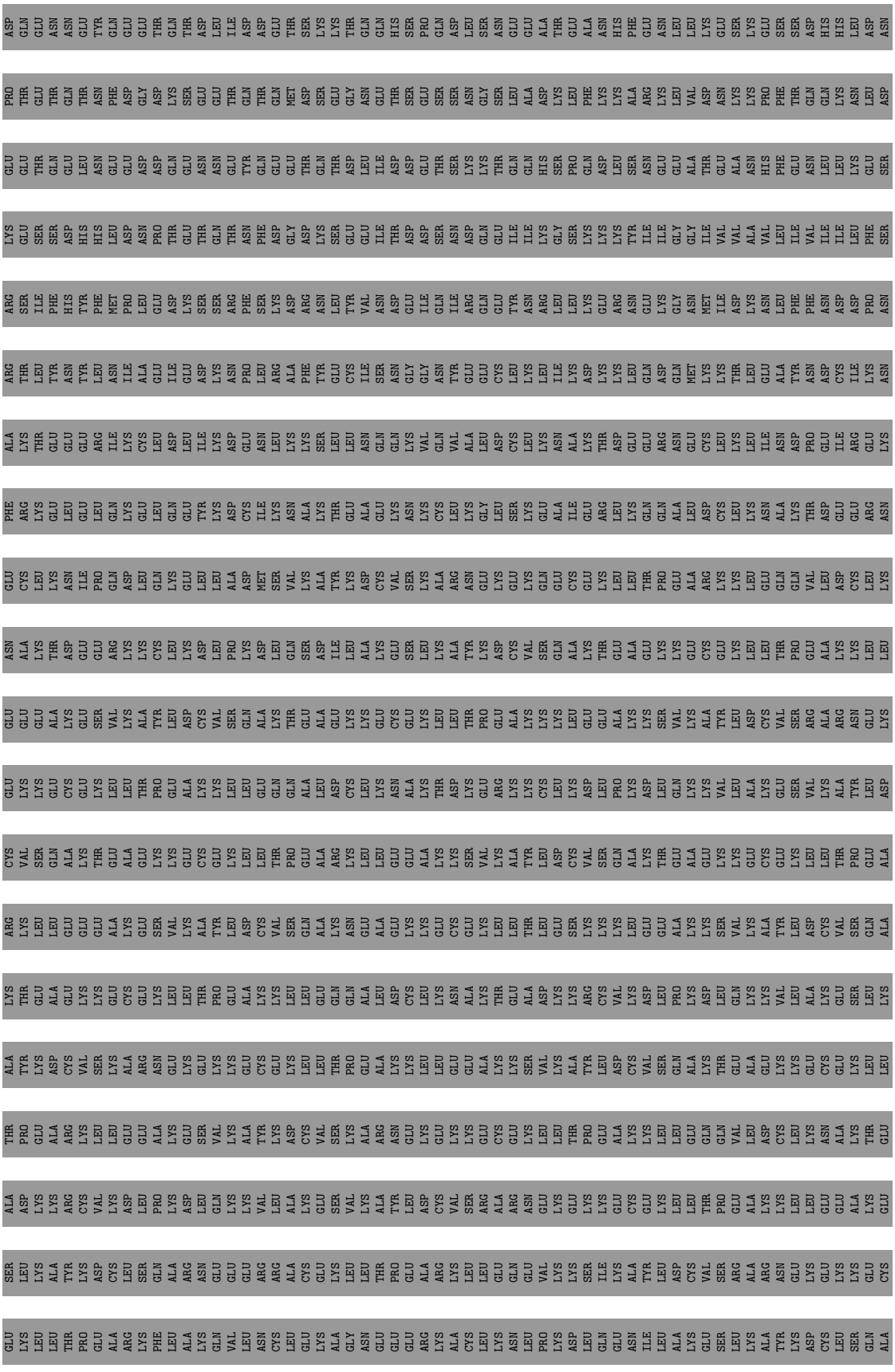
Chain b:

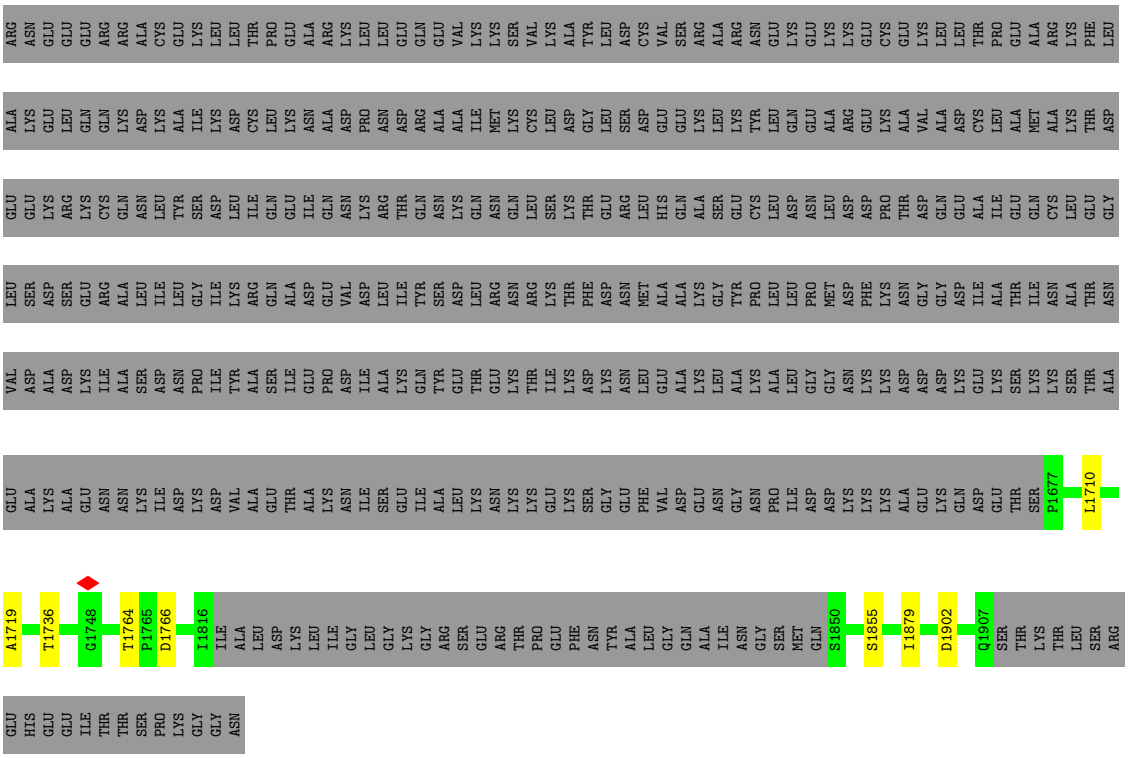




• Molecule 1: Type IV secretion system apparatus protein CagY

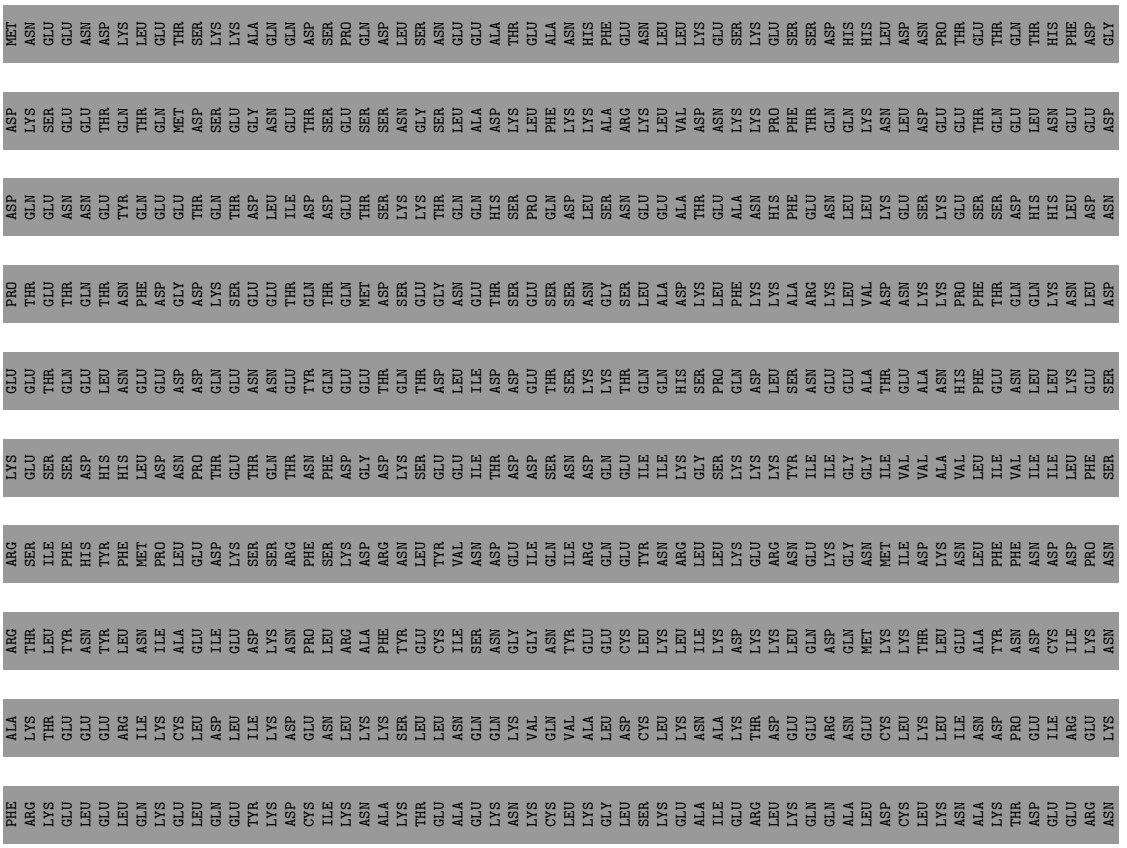
Chain c:

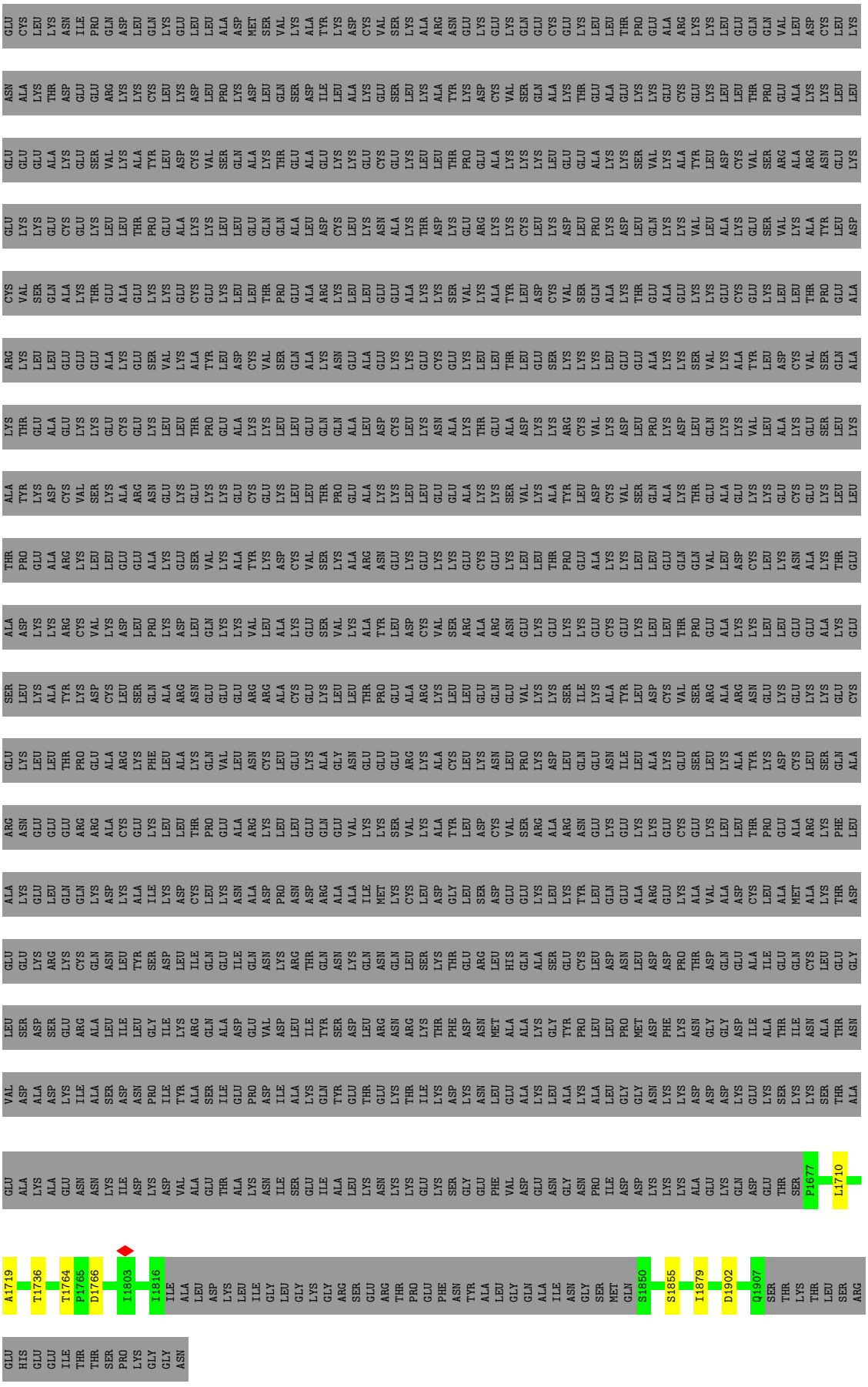

ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN
PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN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



• Molecule 1: Type IV secretion system apparatus protein CagY

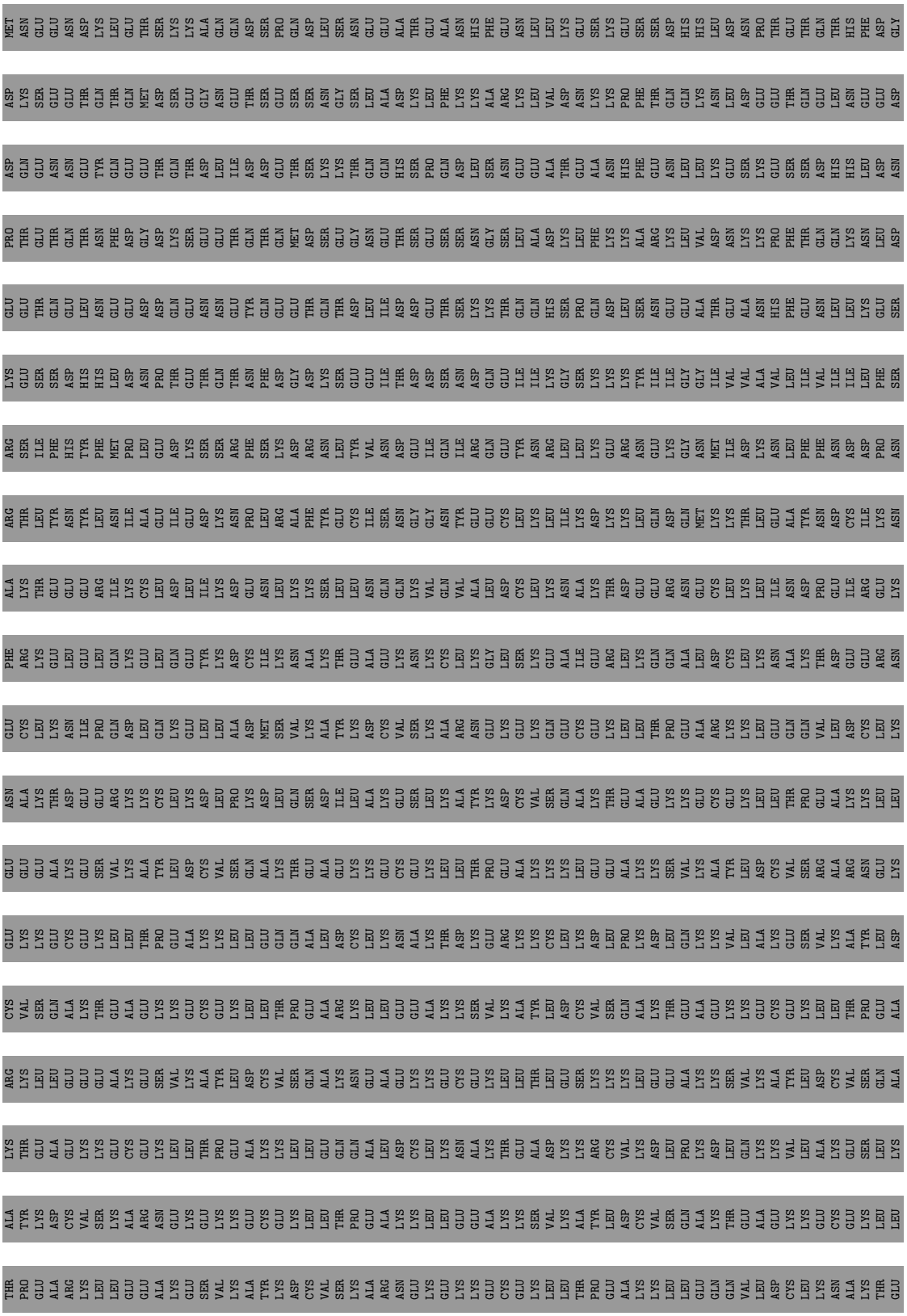
Chain d:

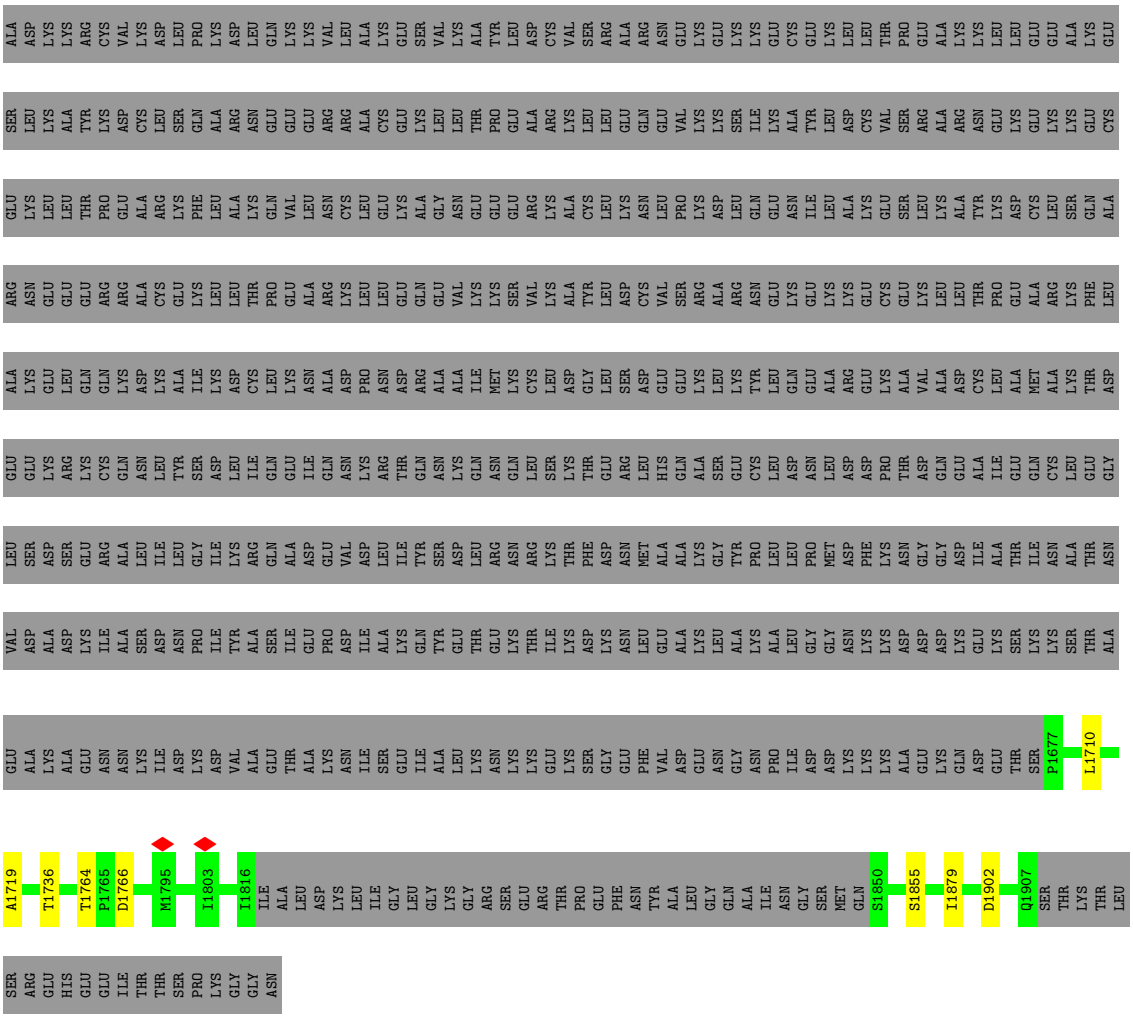



P1677
L1710
A1719
T1736
T1764
P1765
D1766
I1803
I1816
S1850
S1855
I1879
D1902
Q1907

• Molecule 1: Type IV secretion system apparatus protein CagY

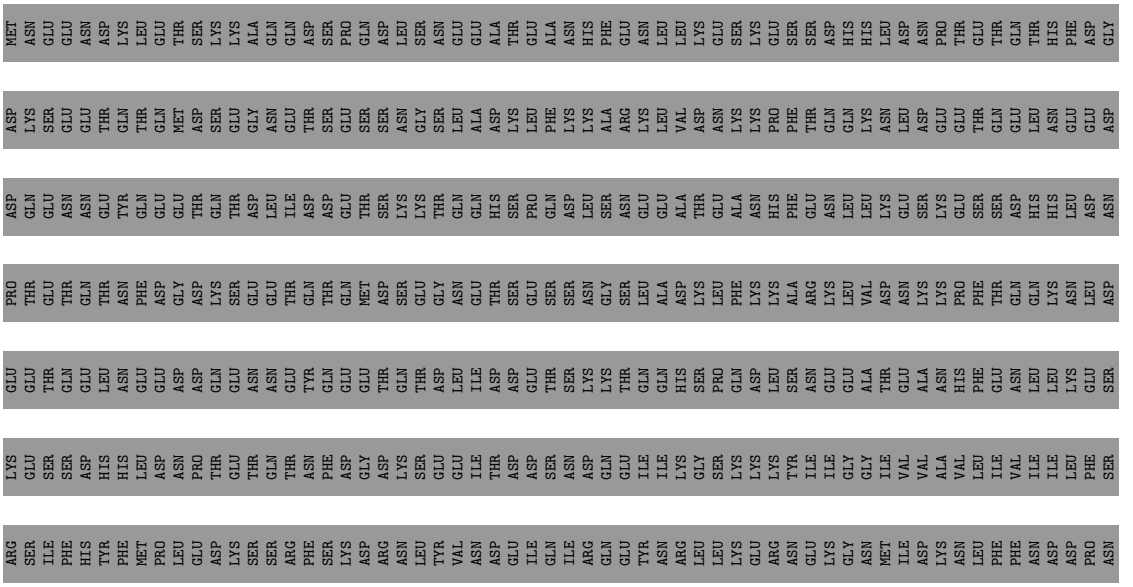
Chain e: 10% 90%

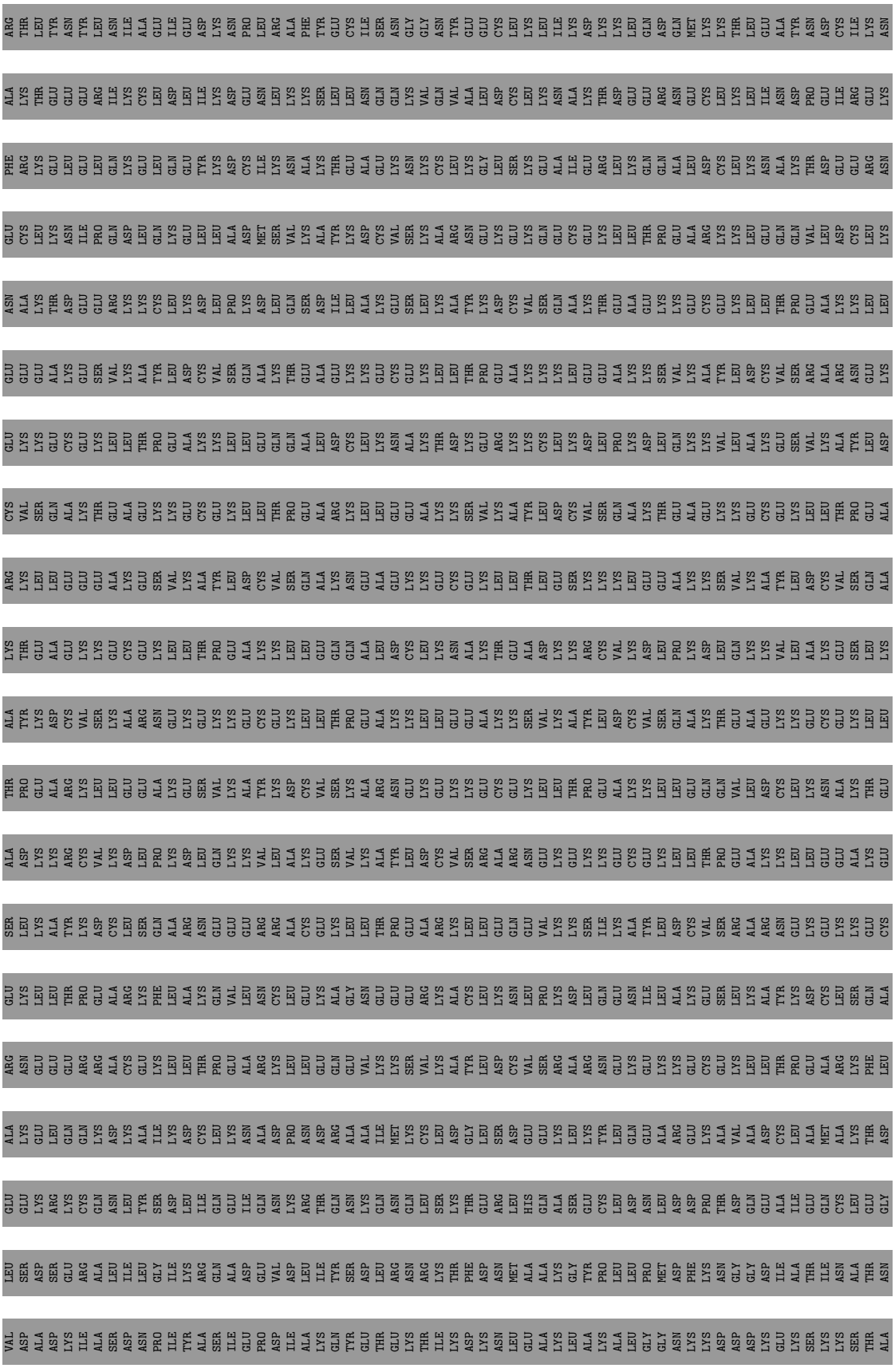

MET ASN GLU GLU ASN ASP LYS LEU GLU THR SER LYS LYS ALA GLN GLN ASP SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN THR HIS PHE ASP GLY
ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU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

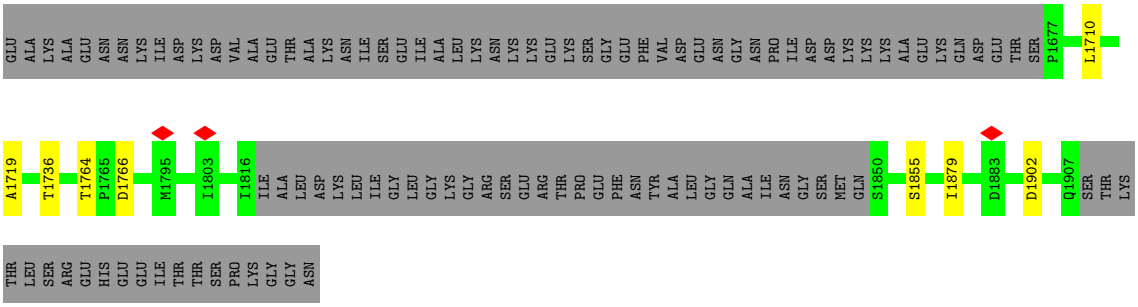


• Molecule 1: Type IV secretion system apparatus protein CagY

Chain f:

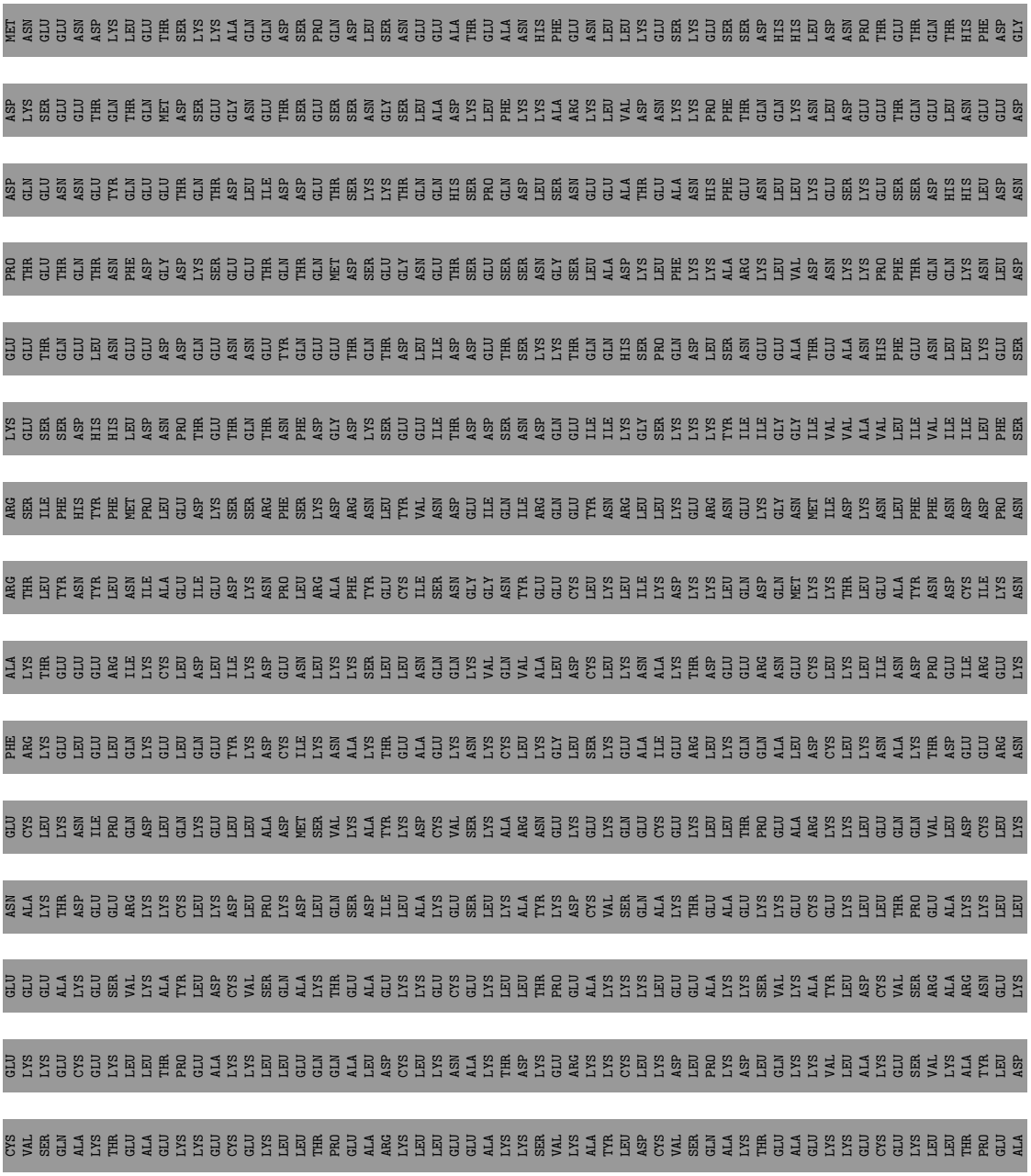


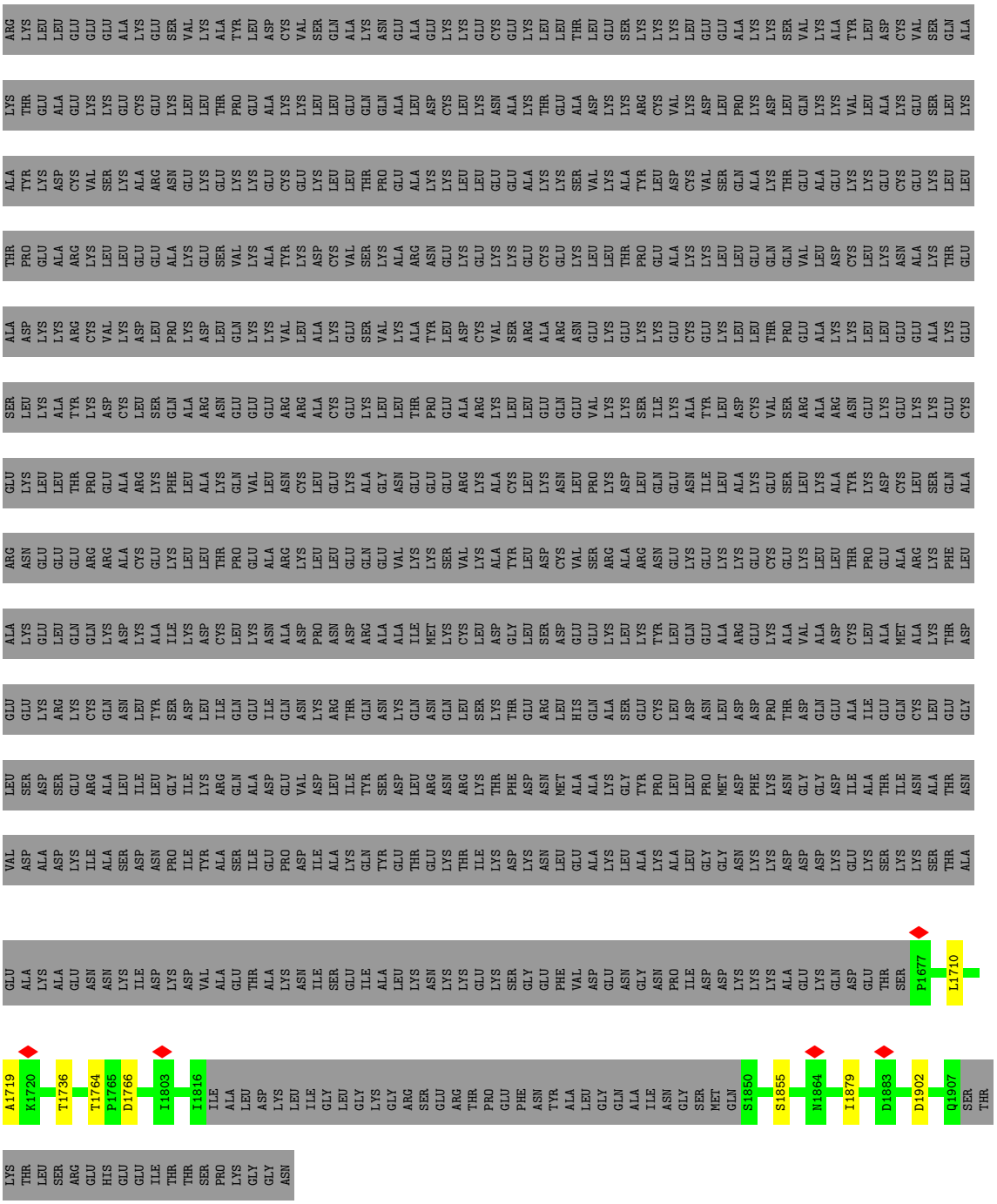




• Molecule 1: Type IV secretion system apparatus protein CagY

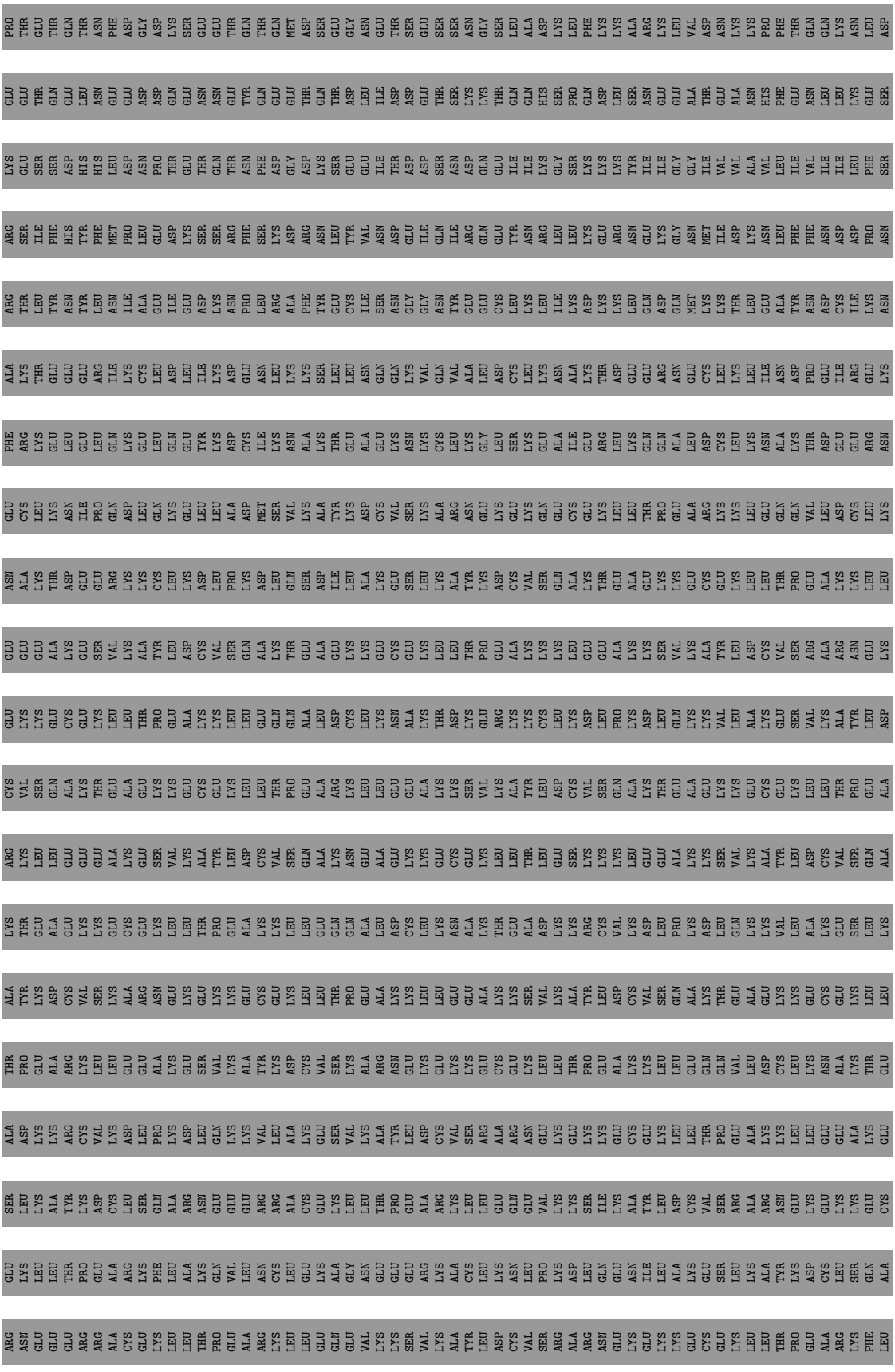
Chain g:

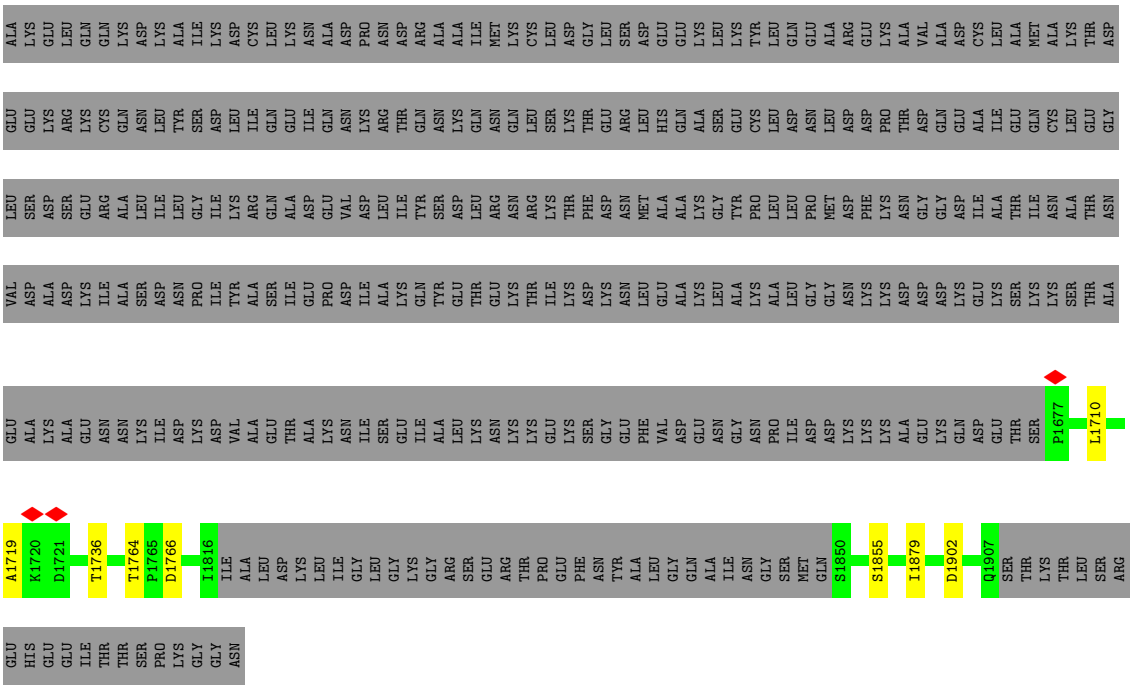




• Molecule 1: Type IV secretion system apparatus protein CagY

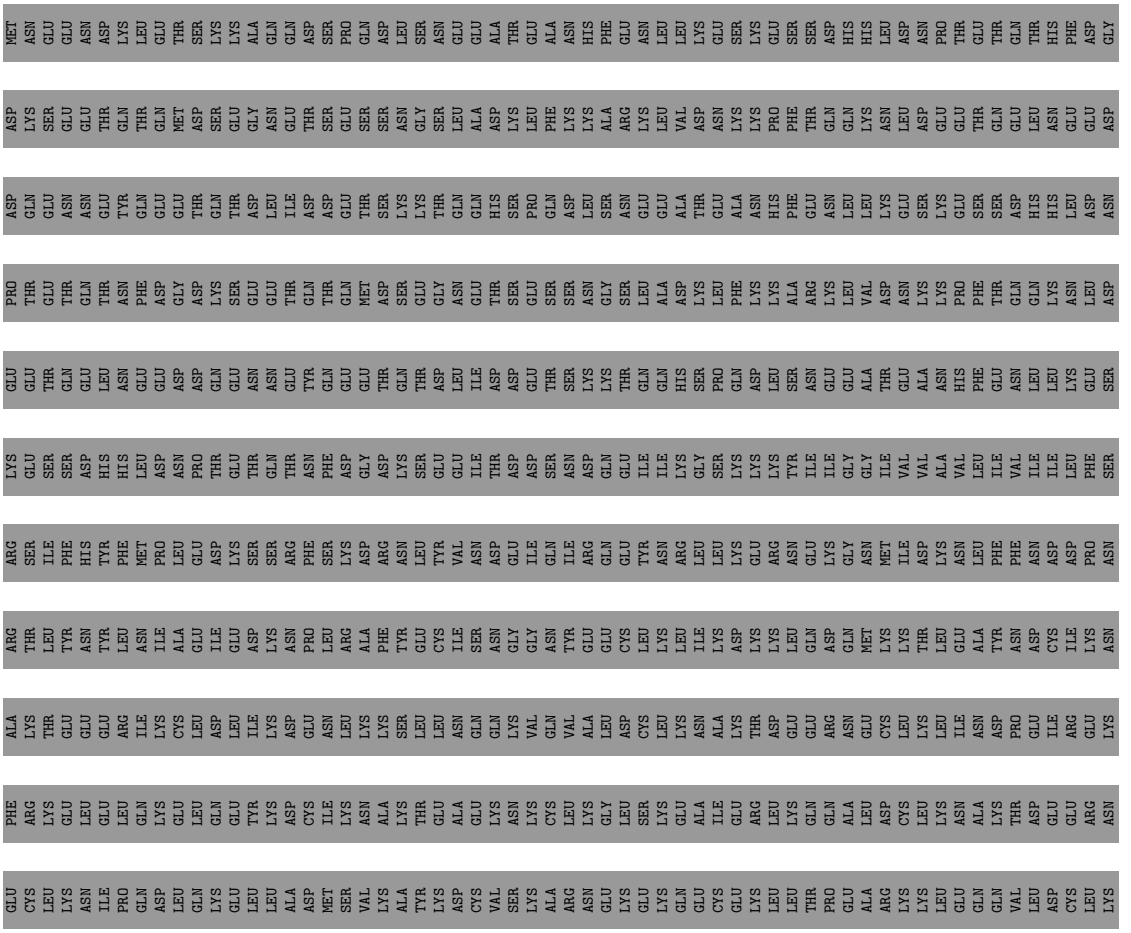
Chain h:

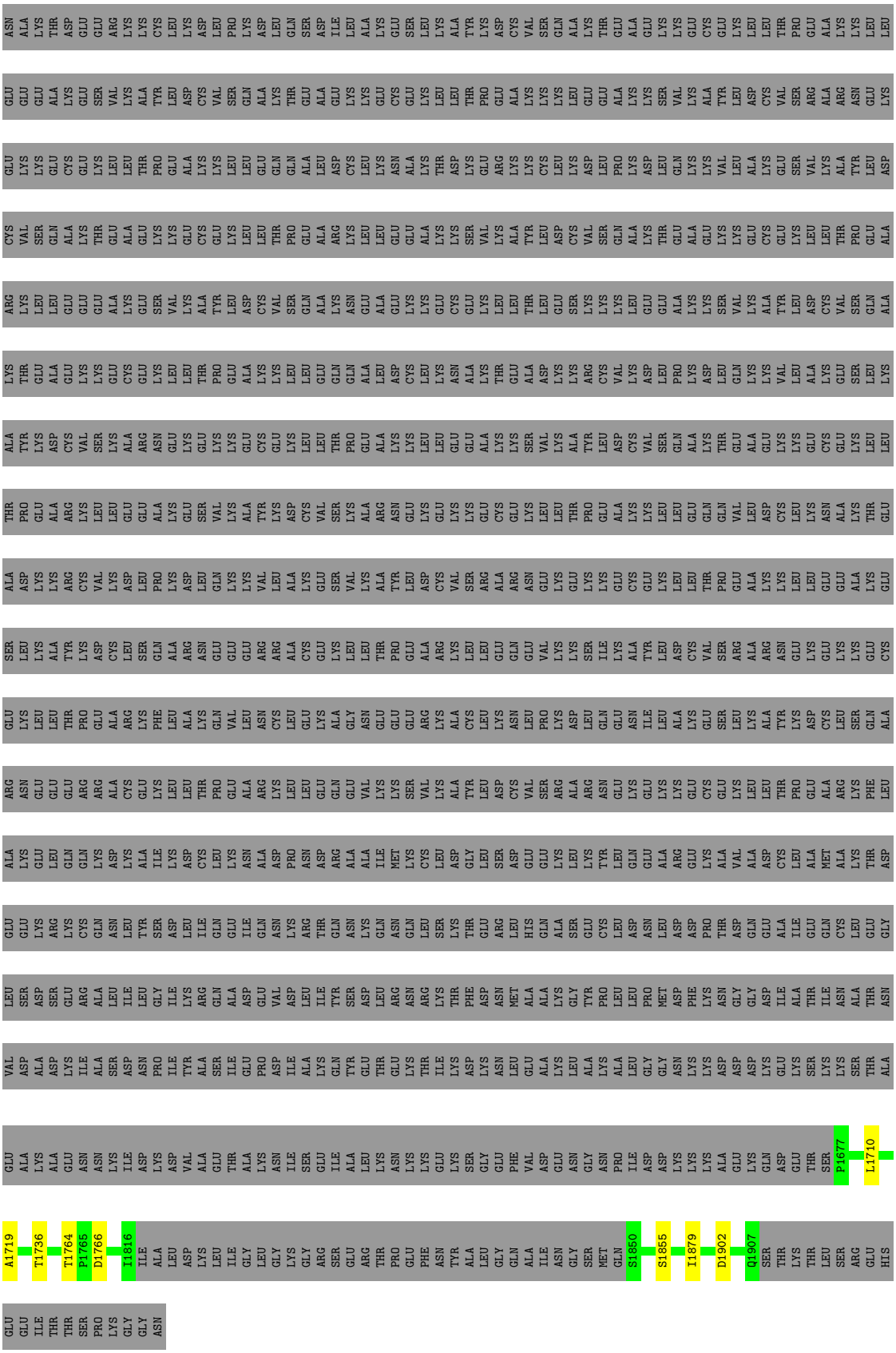

PRO THR GLU THR GLN THR ASN PHE ASP GLY ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU THR SER GLU SER SER ASN GLY SER LEU ALA ASP LYS LEU PHE LYS LYS ALA ARG LYS LEU VAL ASP ASN LYS LYS PRO PHE THR GLN GLN LYS SER GLU ASN LEU ASP
GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN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



• Molecule 1: Type IV secretion system apparatus protein CagY

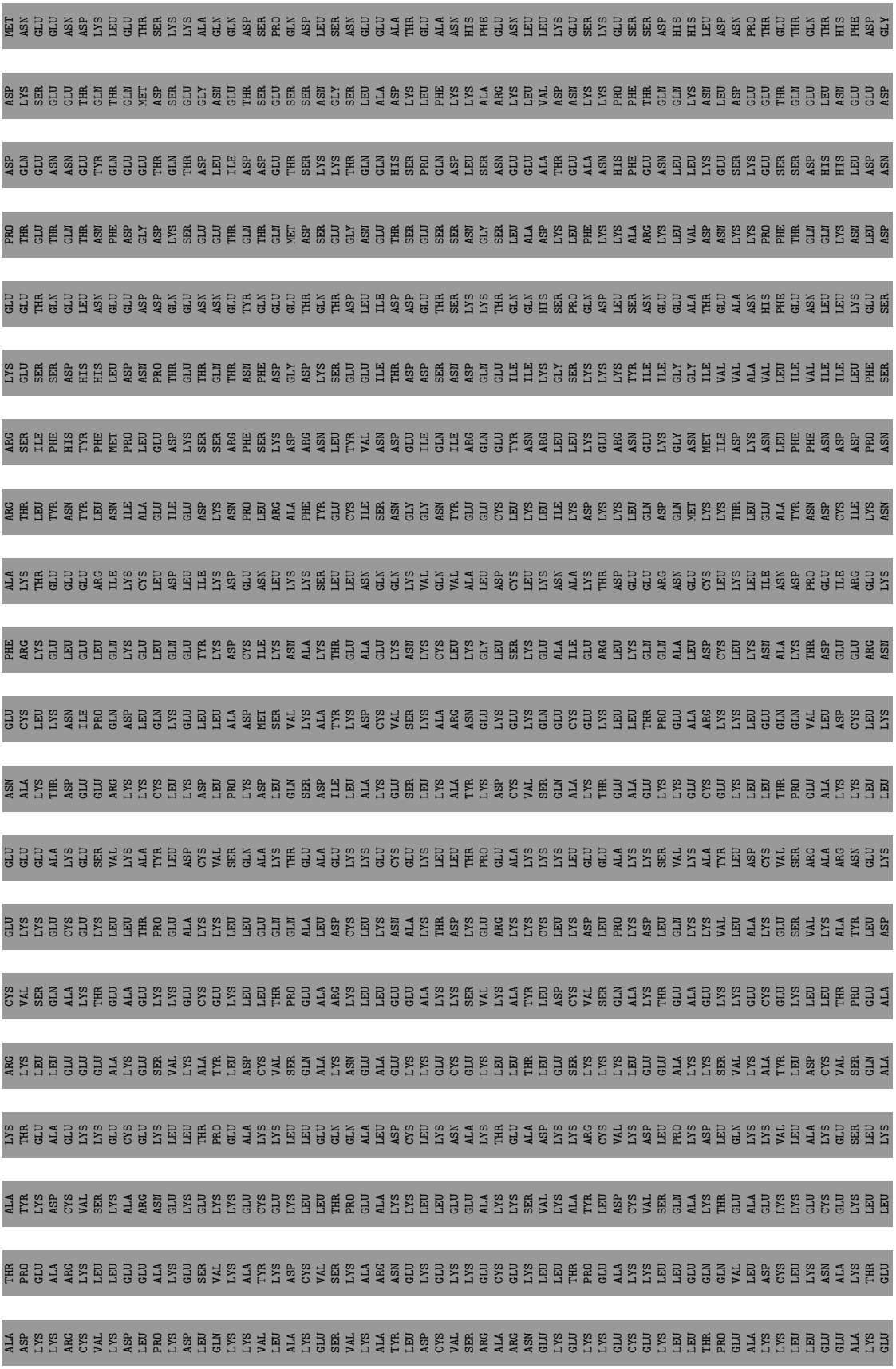
Chain i:

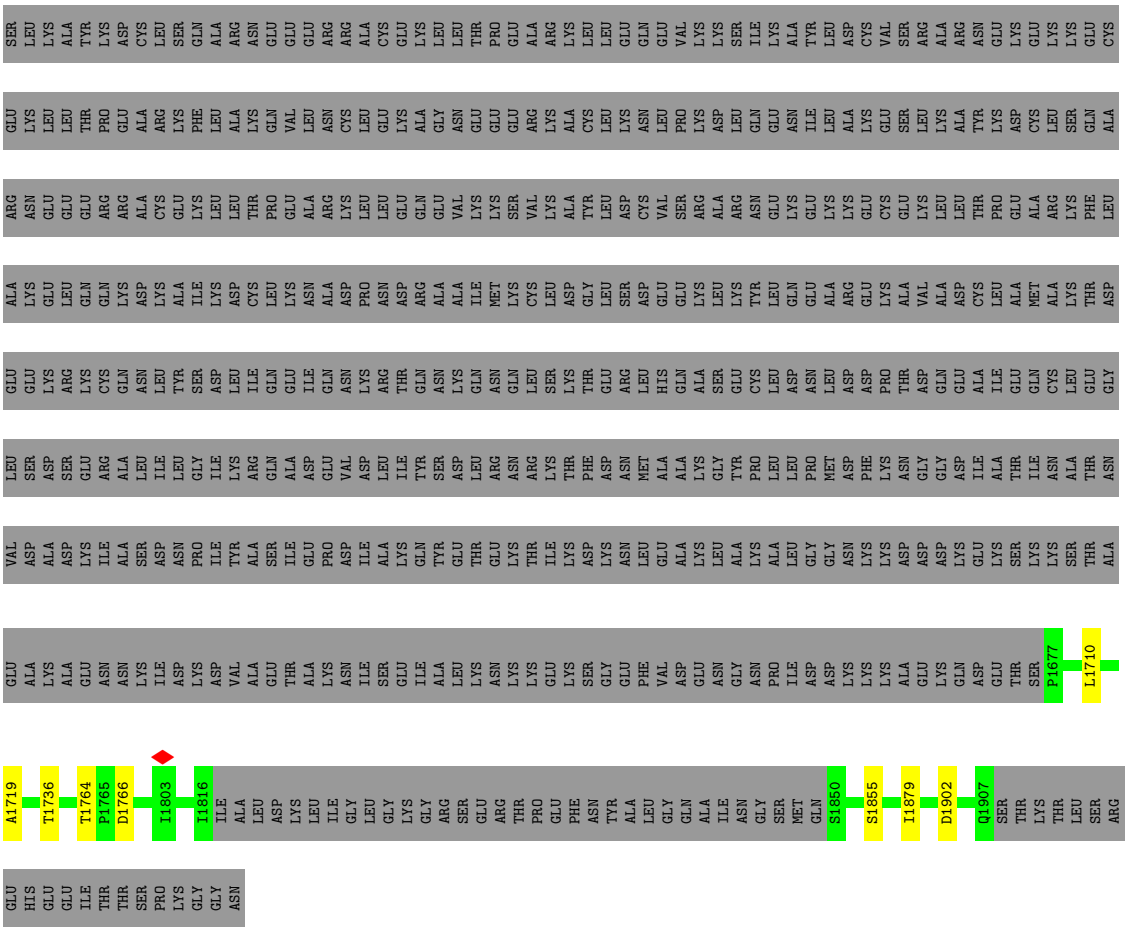




• Molecule 1: Type IV secretion system apparatus protein CagY

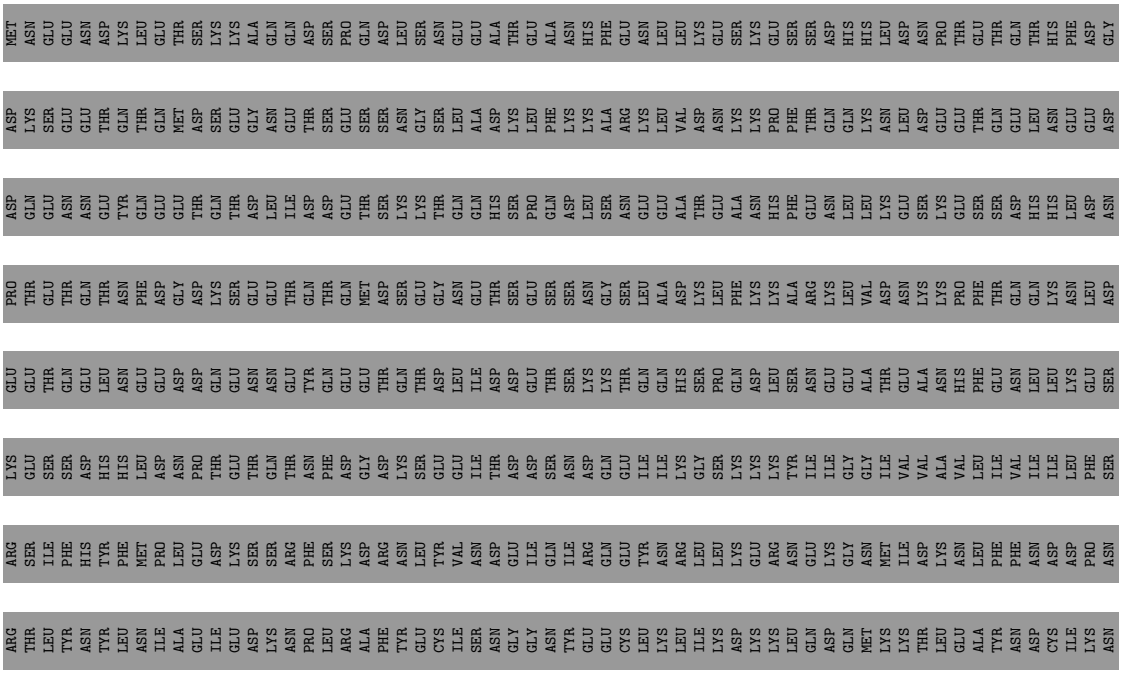
Chain j: 10% 90%

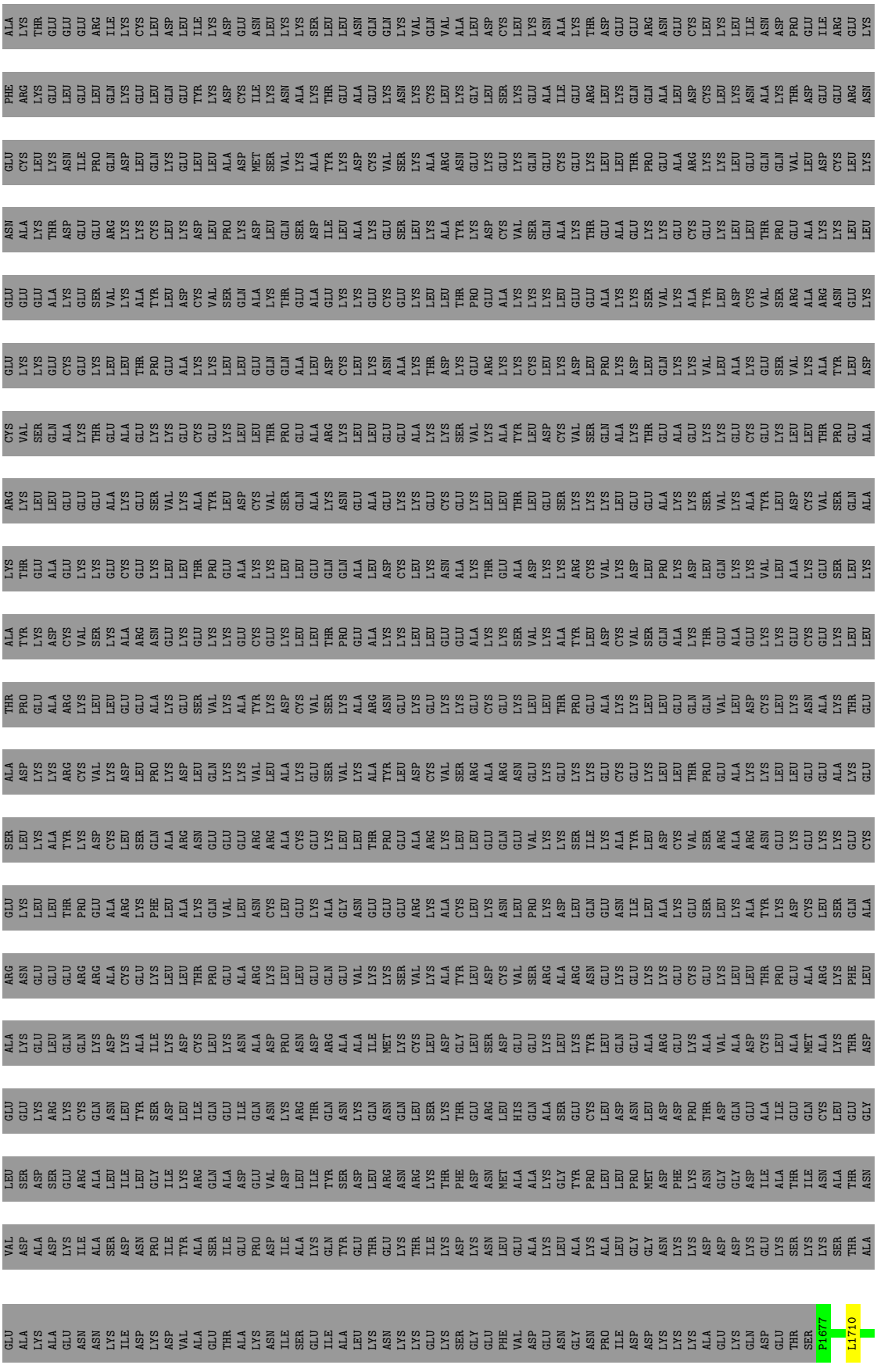

MET ASN GLU GLU ASN ASP LYS LEU GLU THR SER LYS LYS ALA GLN GLN ASP SER PRO GLN ASP LEU SER SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN LEU LEU LYS GLU SER LYS GLU SER SER ASP HIS HIS LEU ASP PRO THR GLU THR GLN THR HIS PHE ASP GLY
ASP LYS SER GLU GLU THR GLN THR GLN MET ASP SER GLU GLY ASN GLU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

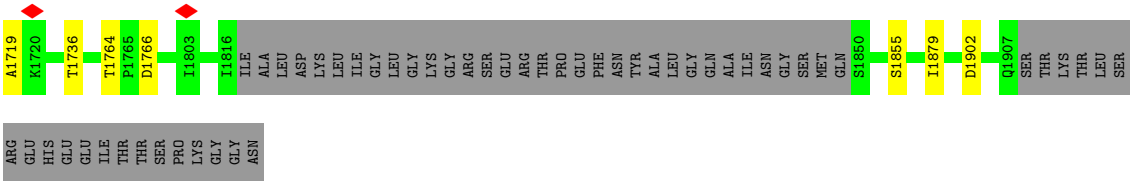


• Molecule 1: Type IV secretion system apparatus protein CagY

Chain k:

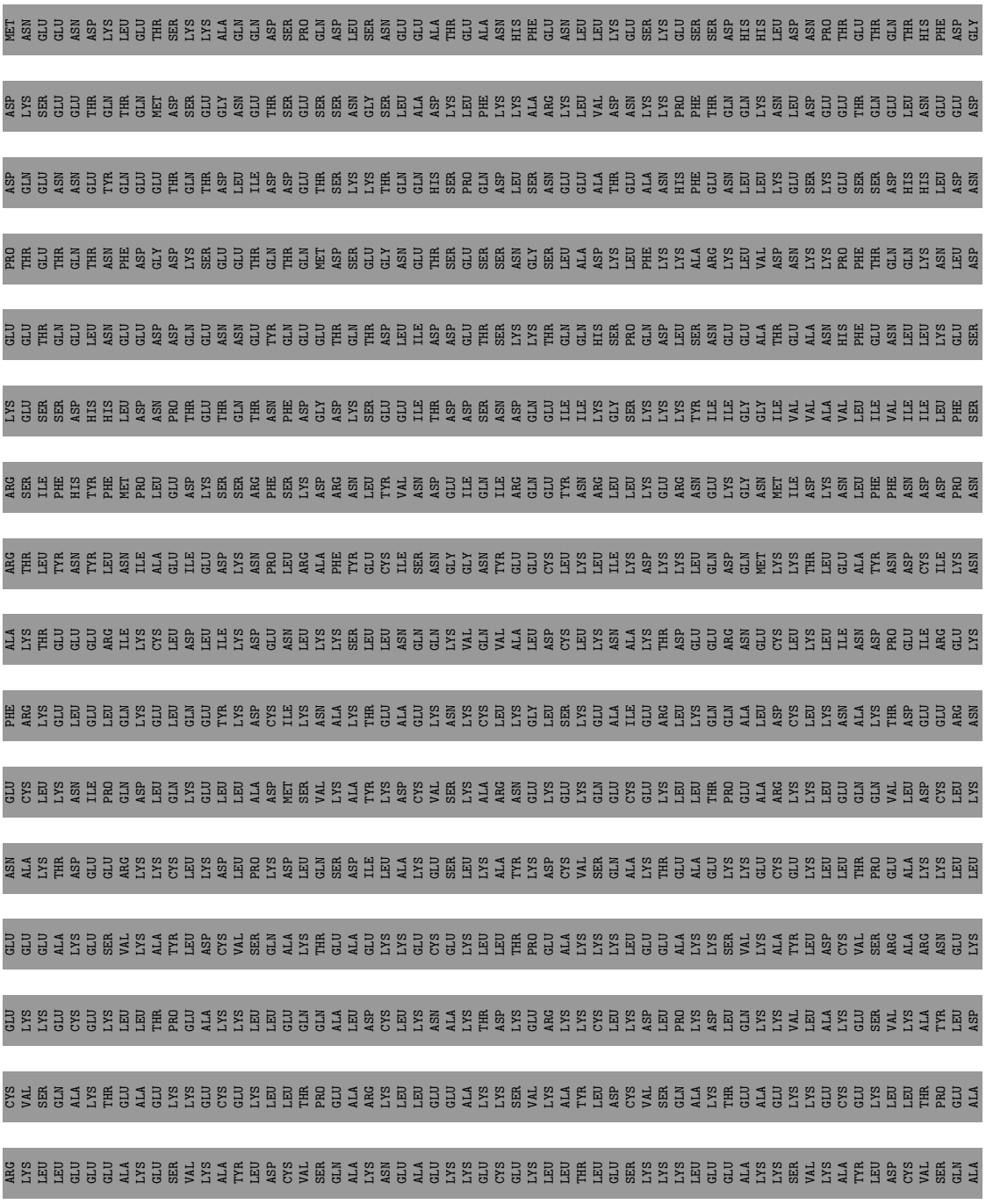


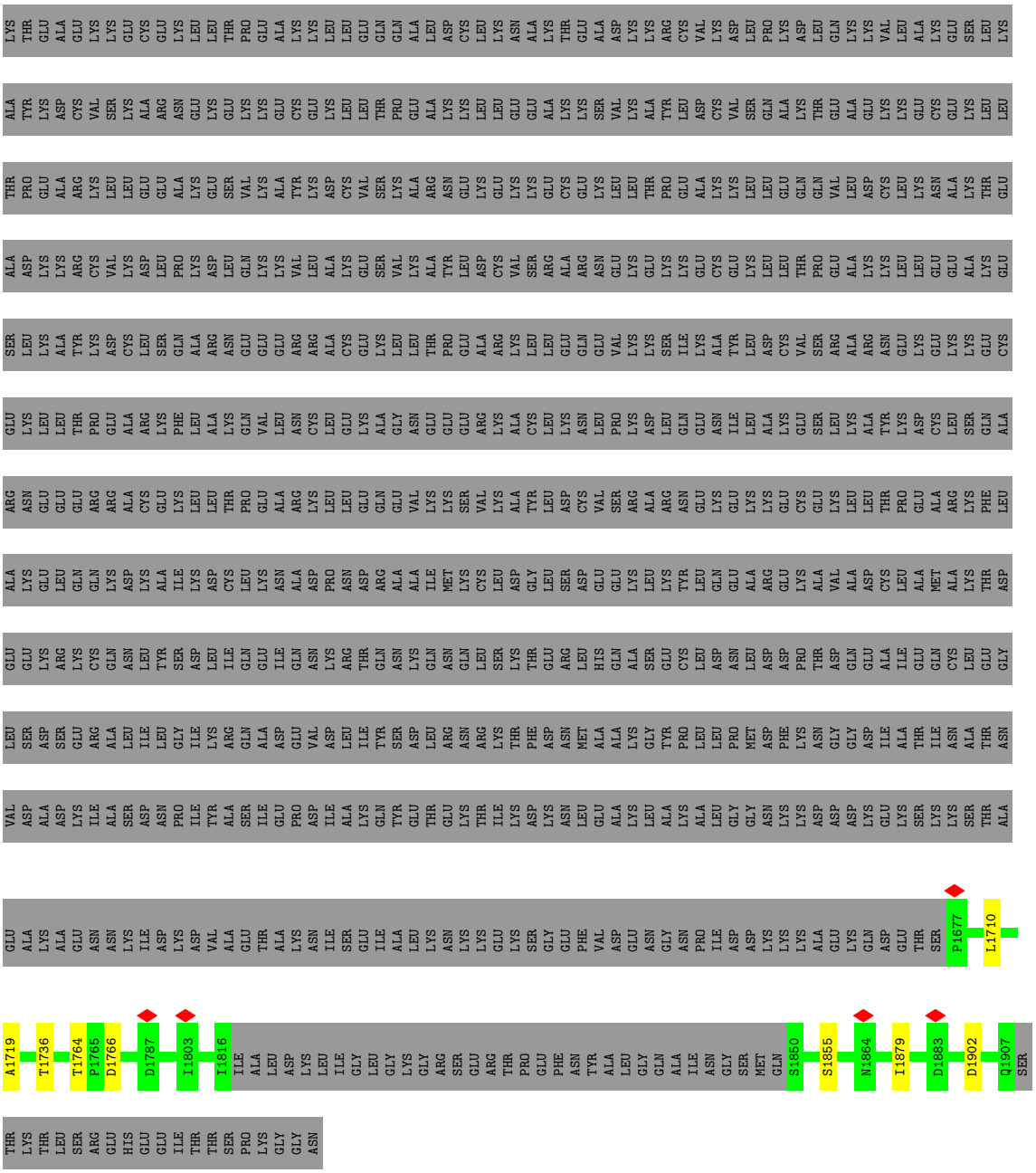

P1677
L1710



• Molecule 1: Type IV secretion system apparatus protein CagY

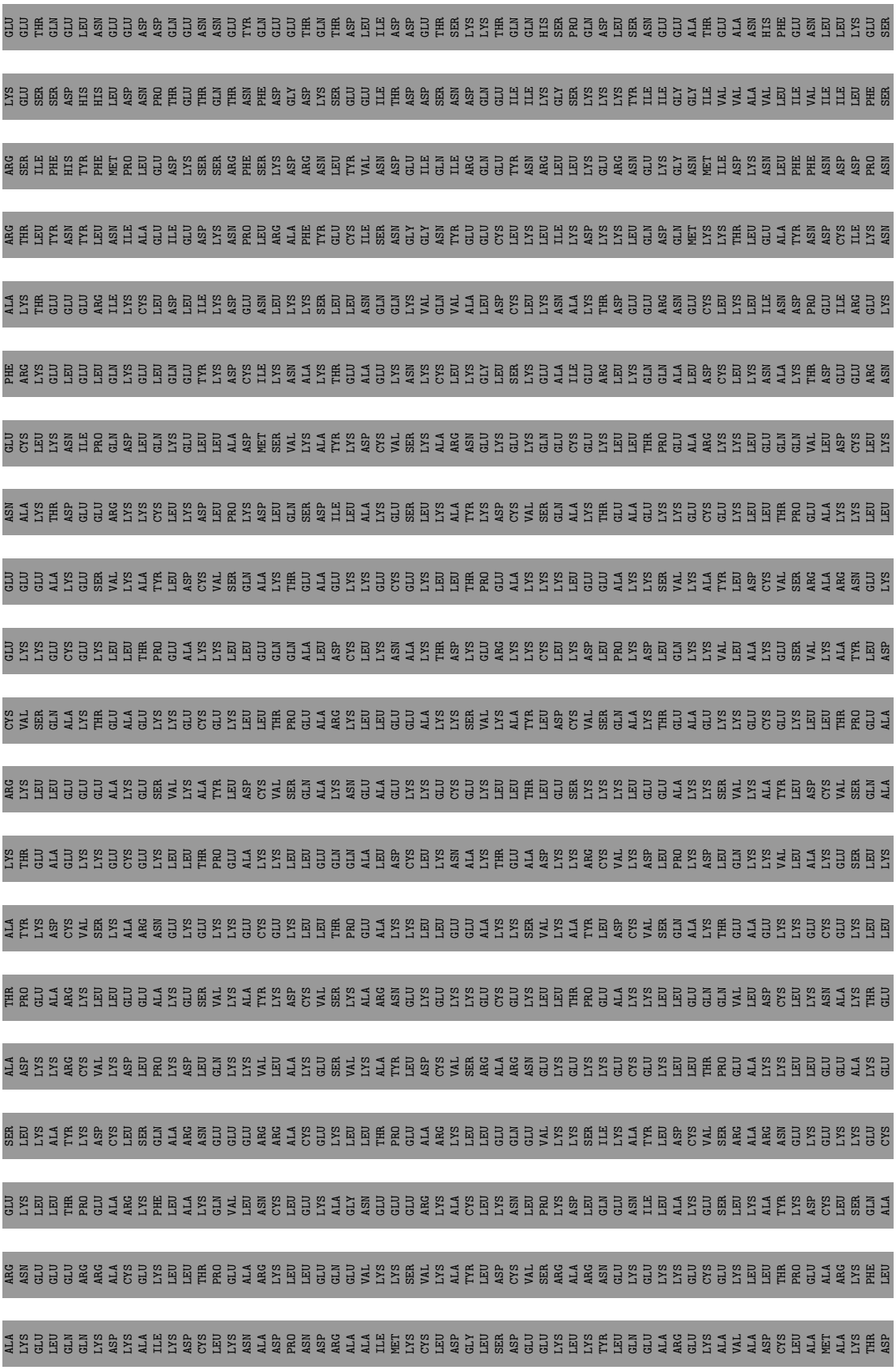
Chain l:

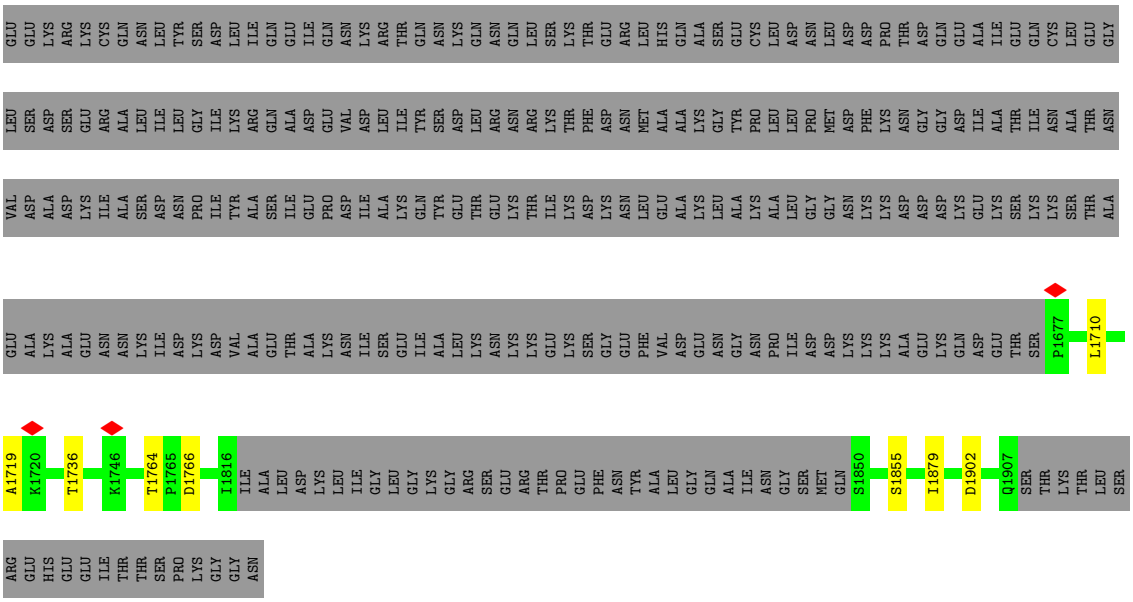




• Molecule 1: Type IV secretion system apparatus protein CagY

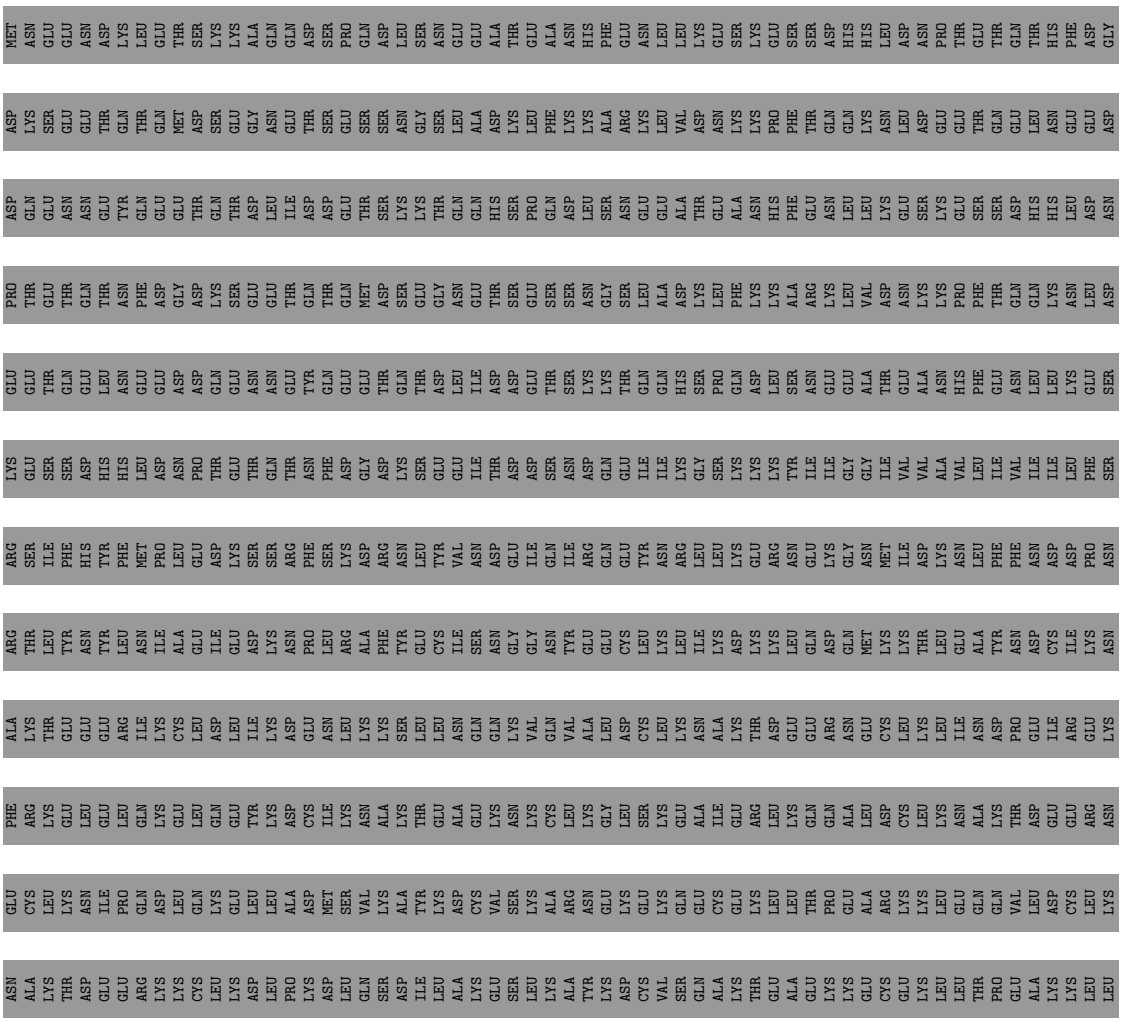
Chain m:

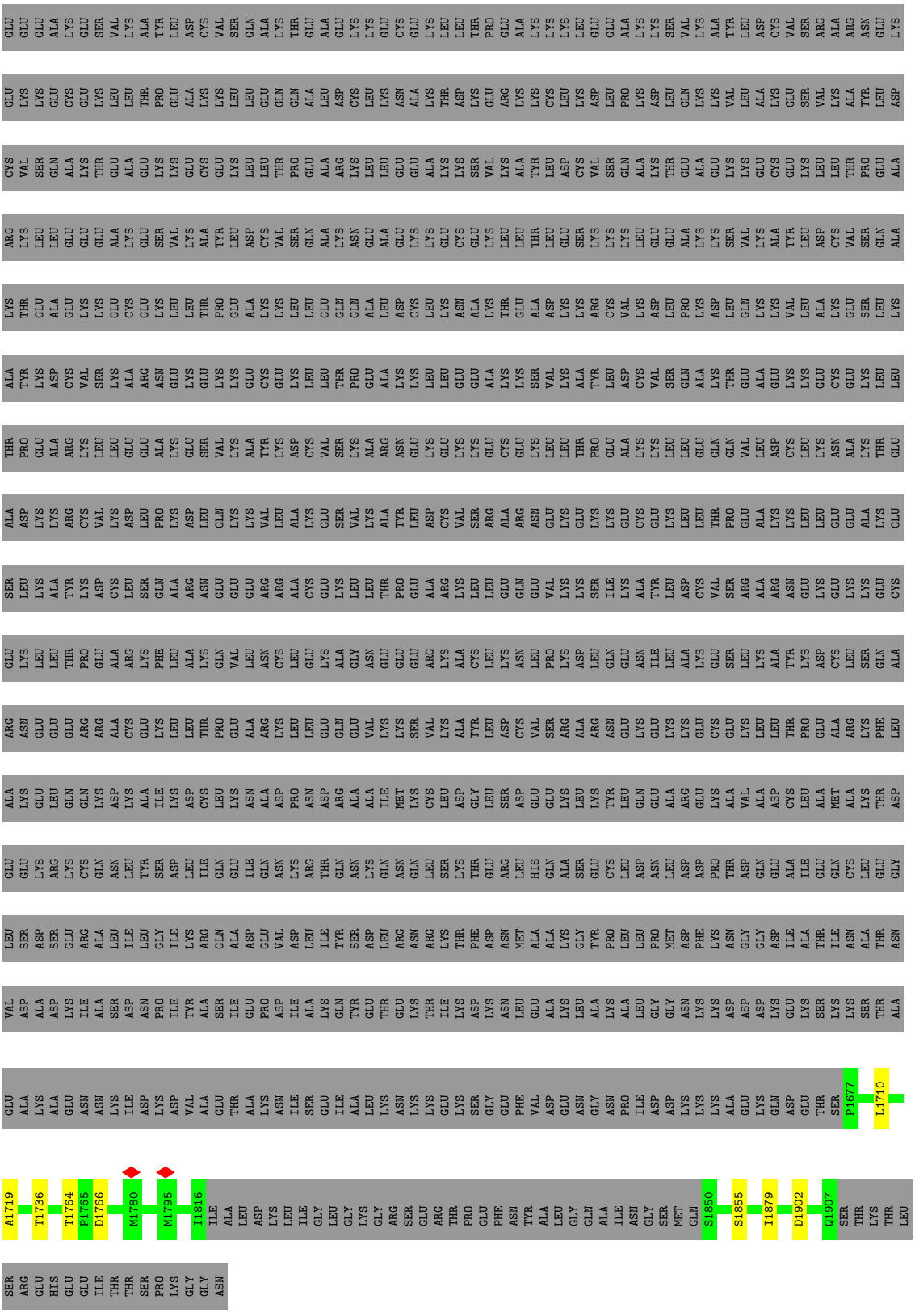

GLU GLU THR GLN GLU LEU ASN GLU GLU ASP ASP GLN GLU ASN ASN GLU TYR GLN GLU GLU THR GLN THR ASP LEU ILE ASP ASP GLU THR SER LYS LYS THR GLN GLN HIS SER PRO GLN ASP LEU SER ASN GLU GLU ALA THR GLU ALA ASN HIS PHE GLU ASN GLU TYR LYS GLU SER
LYS GLU SER SER ASP HIS HIS LEU ASP ASN PRO THR GLU THR GLN ASP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



• Molecule 1: Type IV secretion system apparatus protein CagY

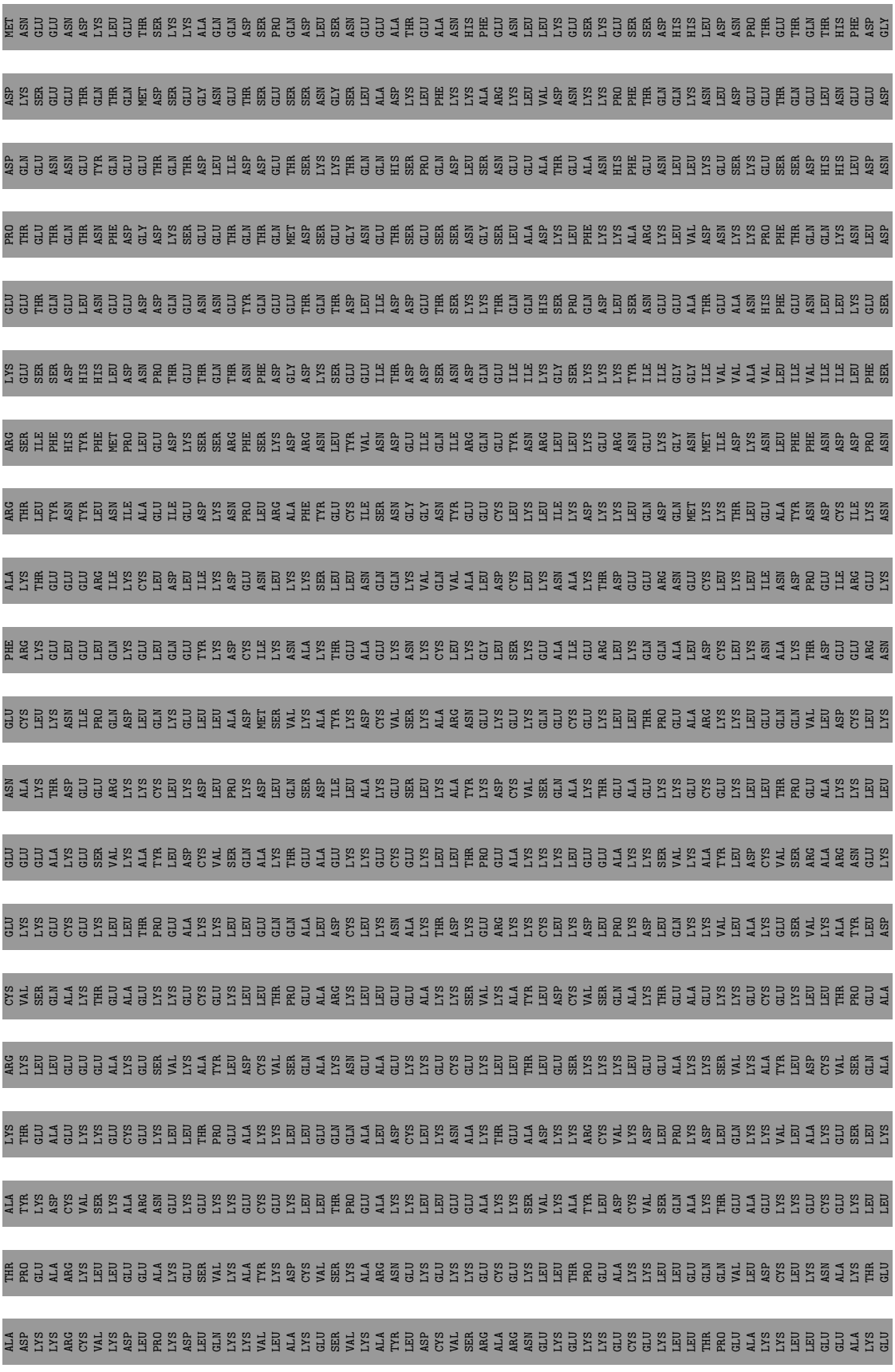
Chain n:

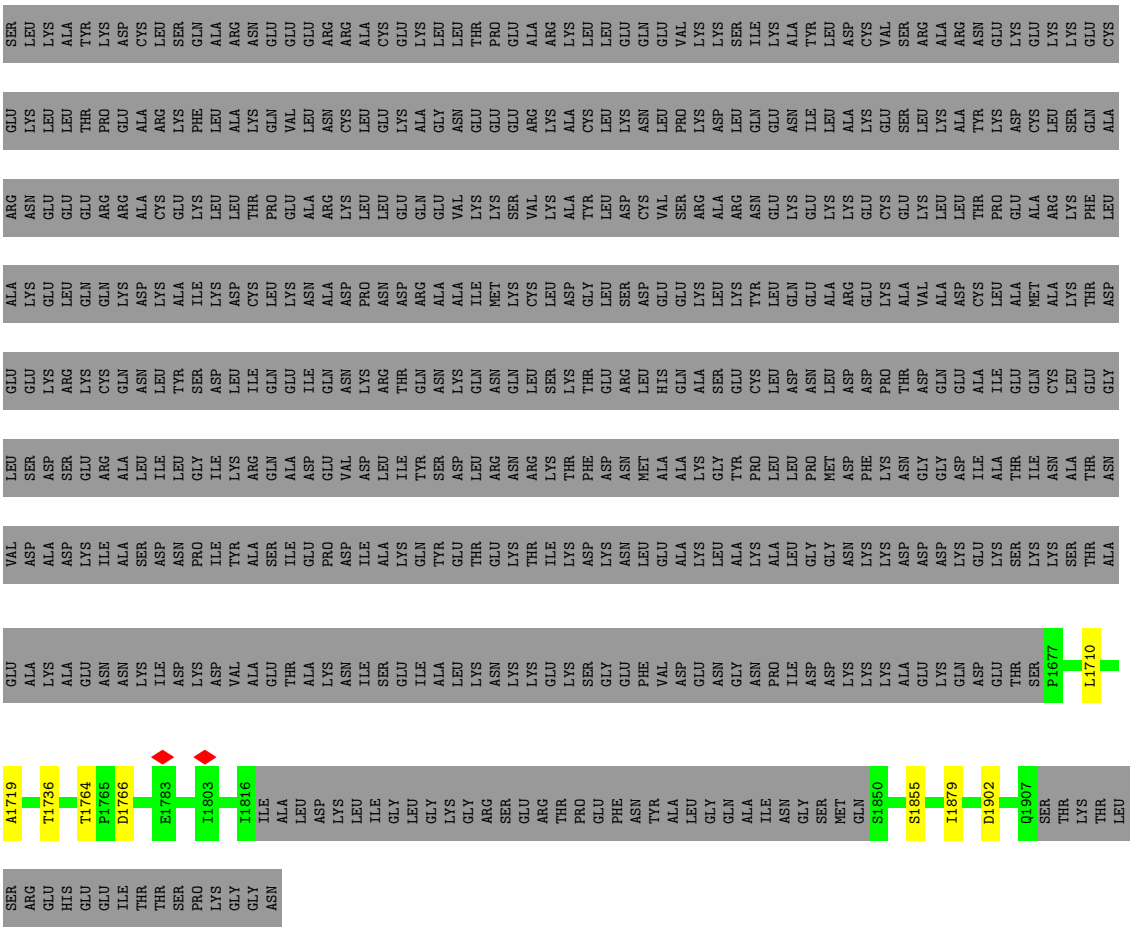




• Molecule 1: Type IV secretion system apparatus protein CagY

Chain o:


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


P1677
L1710
A1719
T1736
T1764
P1765
D1766
E1783
I1803
I1816
S1850
S1855
I1879
D1902
Q1907

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C14	Depositor
Number of particles used	17159	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	59.2	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.136	Depositor
Minimum map value	-0.067	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.027	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	b	0.41	0/1517	0.58	0/2056
1	c	0.41	0/1517	0.58	0/2056
1	d	0.41	0/1517	0.58	0/2056
1	e	0.41	0/1517	0.58	0/2056
1	f	0.41	0/1517	0.58	0/2056
1	g	0.41	0/1517	0.58	0/2056
1	h	0.41	0/1517	0.58	0/2056
1	i	0.41	0/1517	0.58	0/2056
1	j	0.41	0/1517	0.58	0/2056
1	k	0.41	0/1517	0.58	0/2056
1	l	0.41	0/1517	0.58	0/2056
1	m	0.41	0/1517	0.58	0/2056
1	n	0.41	0/1517	0.58	0/2056
1	o	0.41	0/1517	0.58	0/2056
All	All	0.41	0/21238	0.58	0/28784

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	b	0	1
1	c	0	1
1	d	0	1
1	e	0	1
1	f	0	1
1	g	0	1
1	h	0	1
1	i	0	1
1	j	0	1
1	k	0	1
1	l	0	1

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
1	m	0	1
1	n	0	1
1	o	0	1
All	All	0	14

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 14 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	b	1719	ALA	Peptide
1	c	1719	ALA	Peptide
1	d	1719	ALA	Peptide
1	e	1719	ALA	Peptide
1	f	1719	ALA	Peptide

5.2 Too-close contacts (i)

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	b	1493	0	1538	0	0
1	c	1493	0	1538	0	0
1	d	1493	0	1538	0	0
1	e	1493	0	1538	0	0
1	f	1493	0	1538	0	0
1	g	1493	0	1538	0	0
1	h	1493	0	1538	0	0
1	i	1493	0	1538	0	0
1	j	1493	0	1538	0	0
1	k	1493	0	1538	0	0
1	l	1493	0	1538	0	0
1	m	1493	0	1538	0	0
1	n	1493	0	1538	0	0
1	o	1493	0	1538	0	0
All	All	20902	0	21532	0	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 7.

There are no clashes within the asymmetric unit.

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

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

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

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	b	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	c	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	d	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	e	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	f	194/1927 (10%)	176 (91%)	18 (9%)	0	100	100
1	g	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	h	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	i	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	j	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	k	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	l	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	m	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	n	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
1	o	194/1927 (10%)	175 (90%)	19 (10%)	0	100	100
All	All	2716/26978 (10%)	2451 (90%)	265 (10%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

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

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

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	b	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	c	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	d	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	e	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	f	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	g	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	h	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	i	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	j	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	k	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	l	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	m	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	n	167/1725 (10%)	160 (96%)	7 (4%)	30	58
1	o	167/1725 (10%)	160 (96%)	7 (4%)	30	58
All	All	2338/24150 (10%)	2240 (96%)	98 (4%)	33	58

5 of 98 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	j	1736	THR
1	l	1736	THR
1	j	1766	ASP
1	k	1764	THR
1	l	1879	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

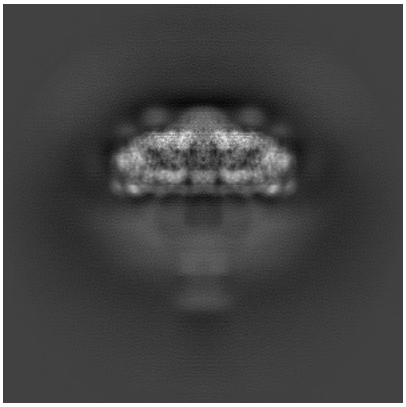
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-20020. These allow visual inspection of the internal detail of the map and identification of artifacts.

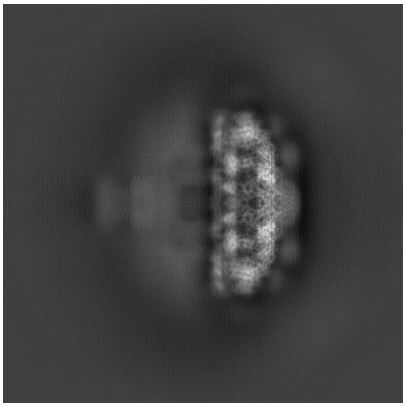
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

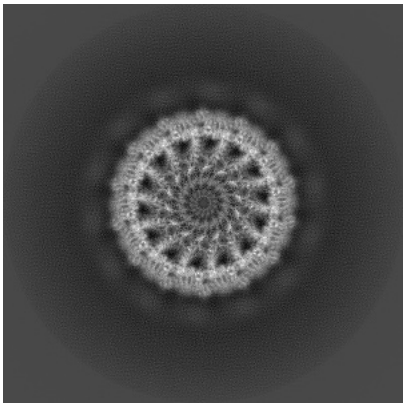
6.1.1 Primary map



X



Y

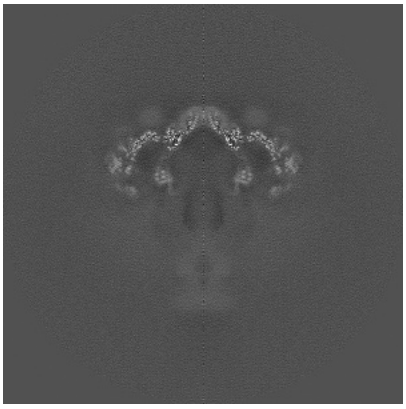


Z

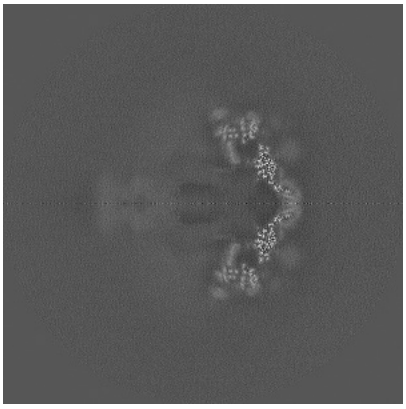
The images above show the map projected in three orthogonal directions.

6.2 Central slices

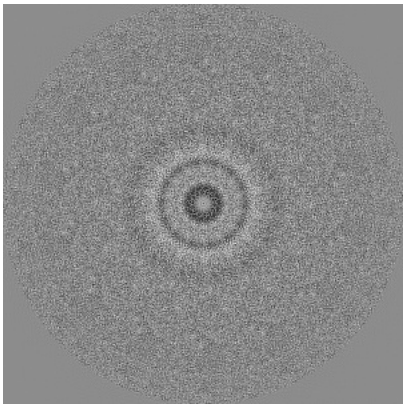
6.2.1 Primary map



X Index: 255



Y Index: 255

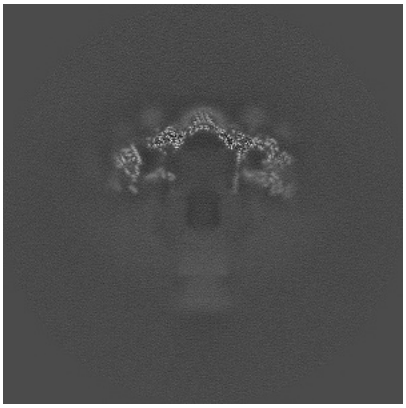


Z Index: 255

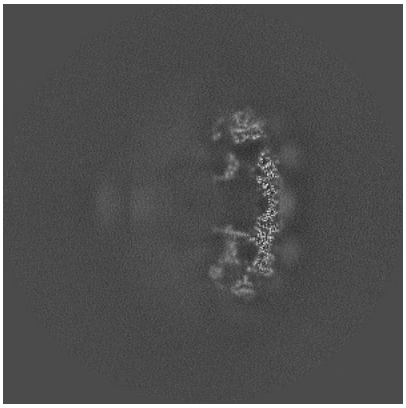
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

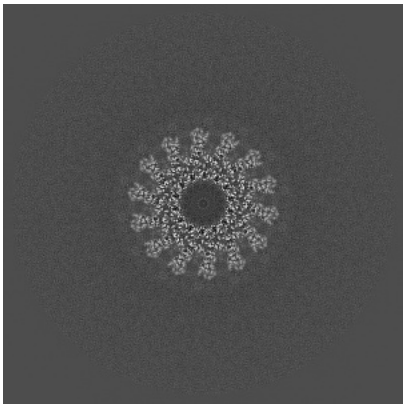
6.3.1 Primary map



X Index: 241



Y Index: 228



Z Index: 335

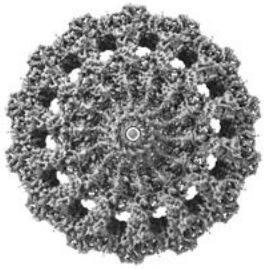
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal surface views (i)

6.4.1 Primary map

X

Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.027. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

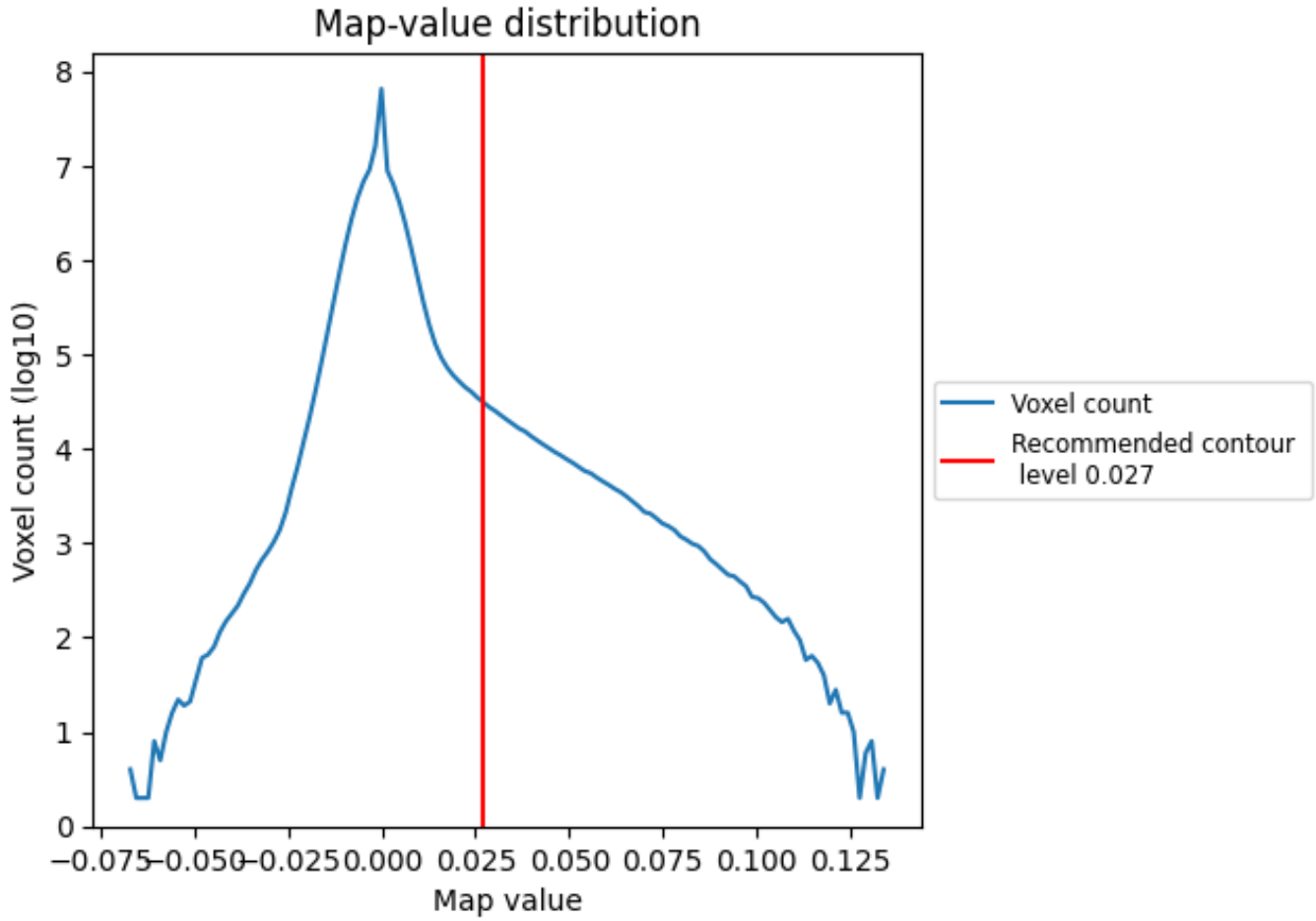
6.5 Mask visualisation (i)

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

7 Map analysis

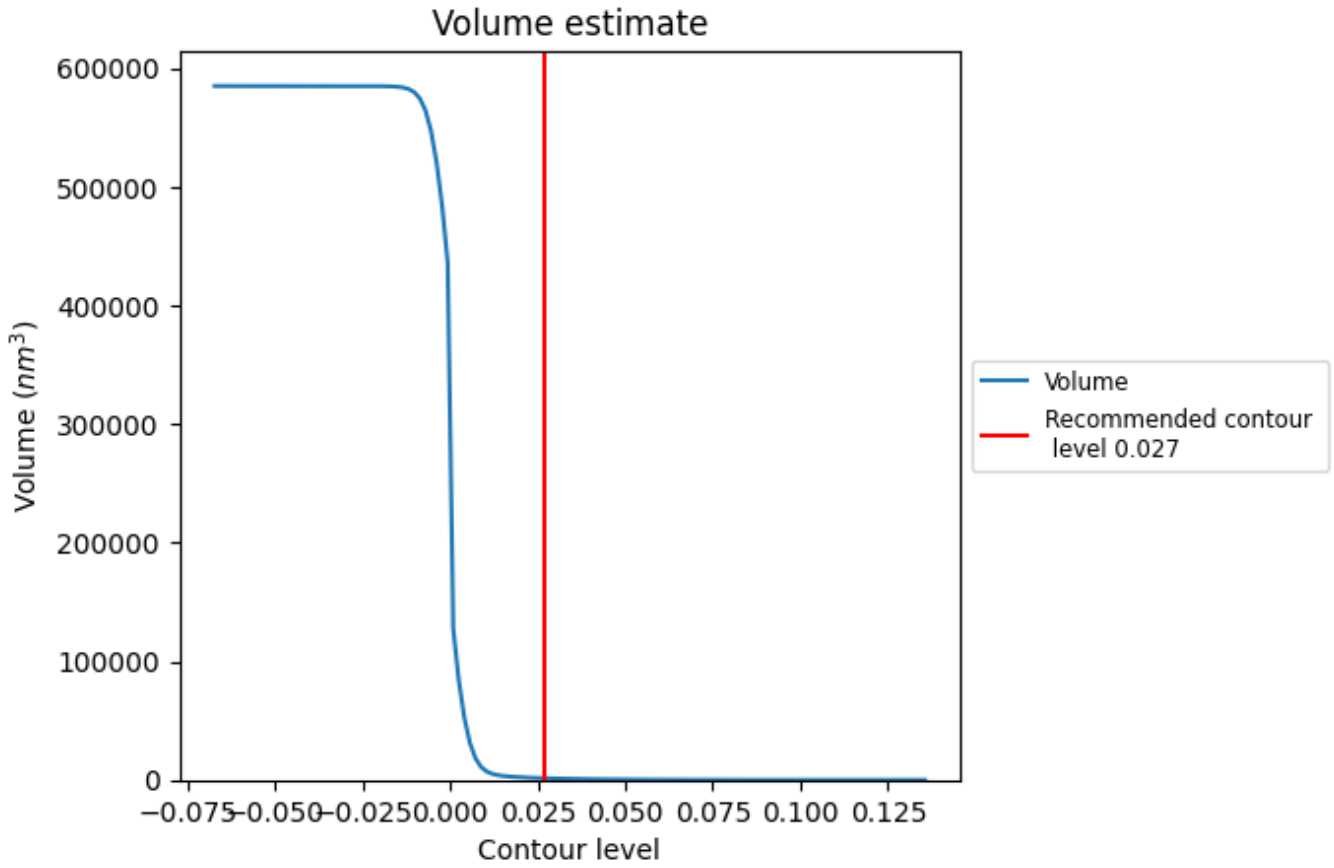
This section contains the results of statistical analysis of the map.

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.

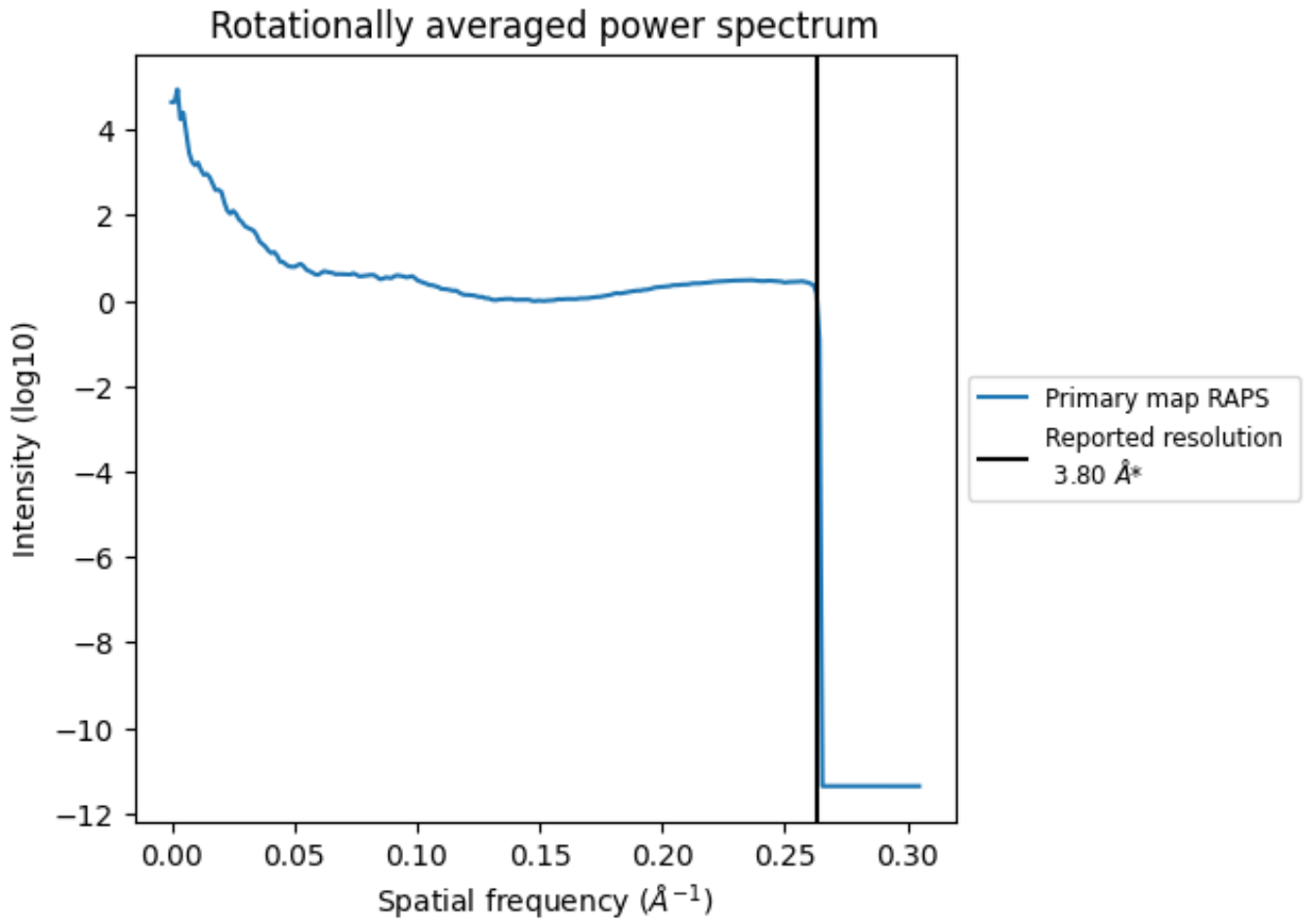
7.2 Volume estimate (i)



The volume at the recommended contour level is 1439 nm^3; this corresponds to an approximate mass of 1300 kDa.

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

7.3 Rotationally averaged power spectrum (i)

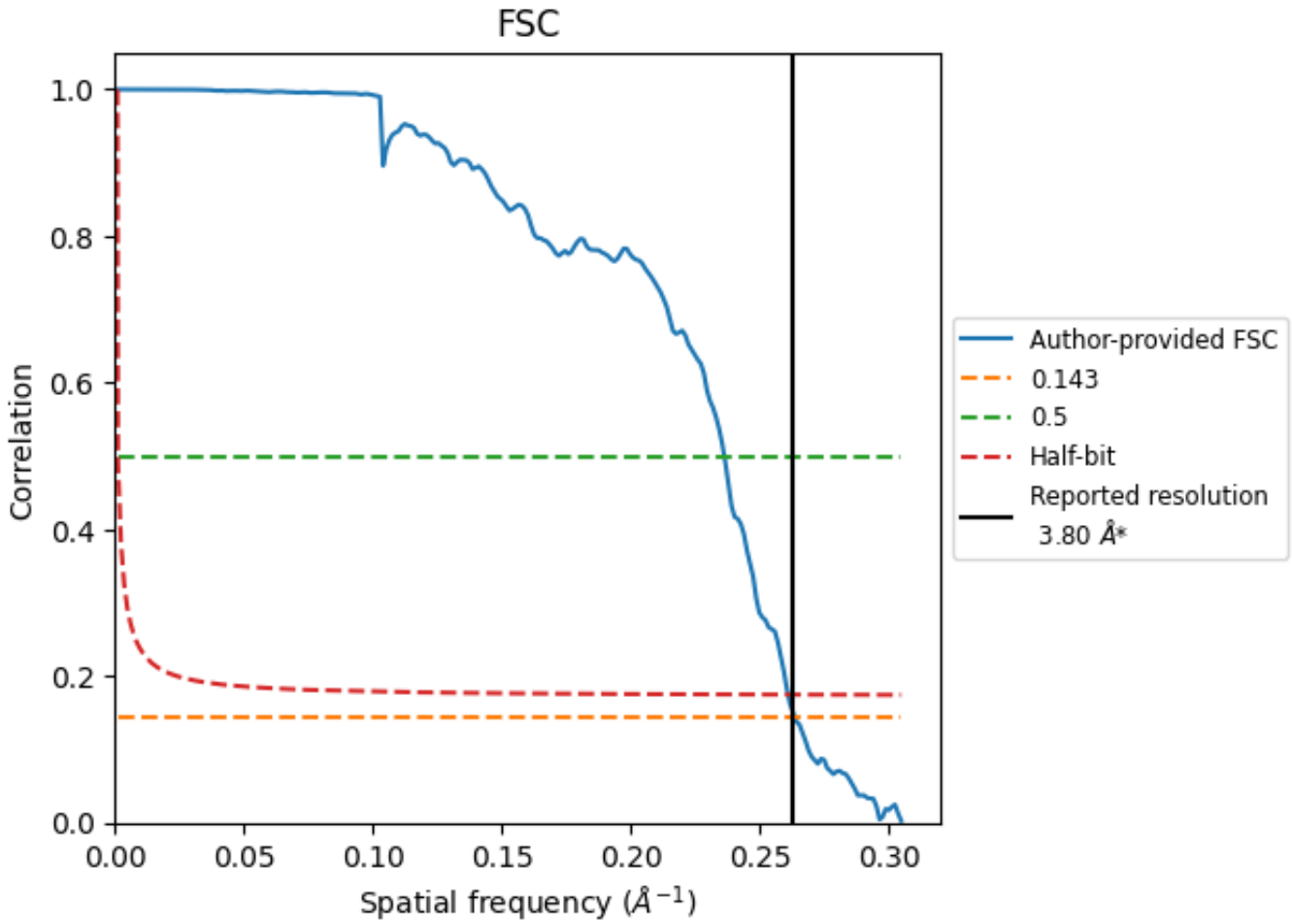


*Reported resolution corresponds to spatial frequency of 0.263 Å^{-1}

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 0.263 Å^{-1}

8.2 Resolution estimates (i)

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.80	-	-
Author-provided FSC curve	3.79	4.23	3.83
Unmasked-calculated*	-	-	-

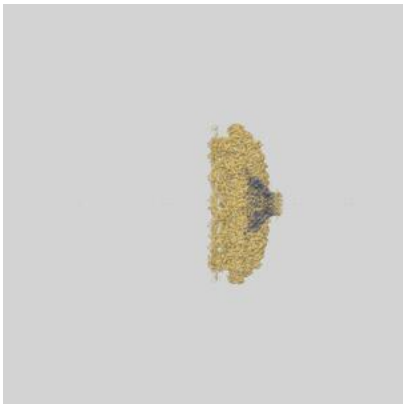
*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

9 Map-model fit

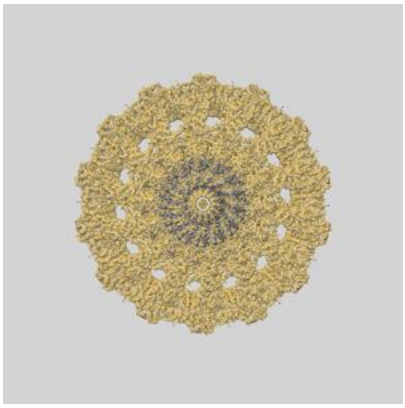
This section contains information regarding the fit between EMDB map EMD-20020 and PDB model 6ODI. Per-residue inclusion information can be found in section 3 on page 5.

9.1 Map-model overlay

X



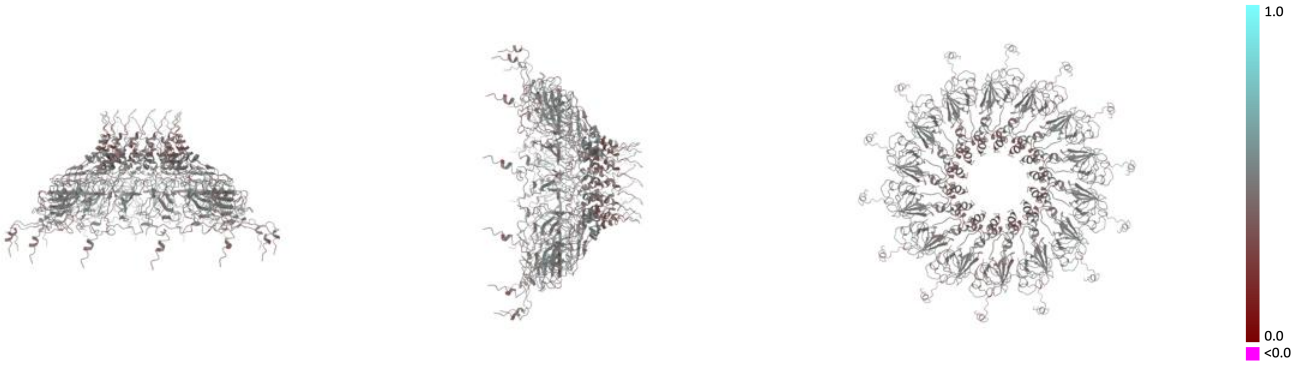
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.027 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

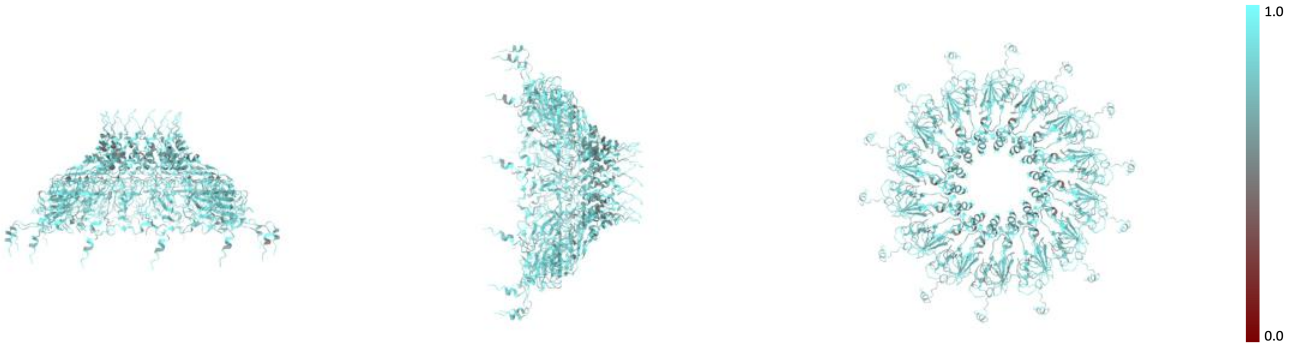
9.2 Q-score mapped to coordinate model (i)



X Y Z

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

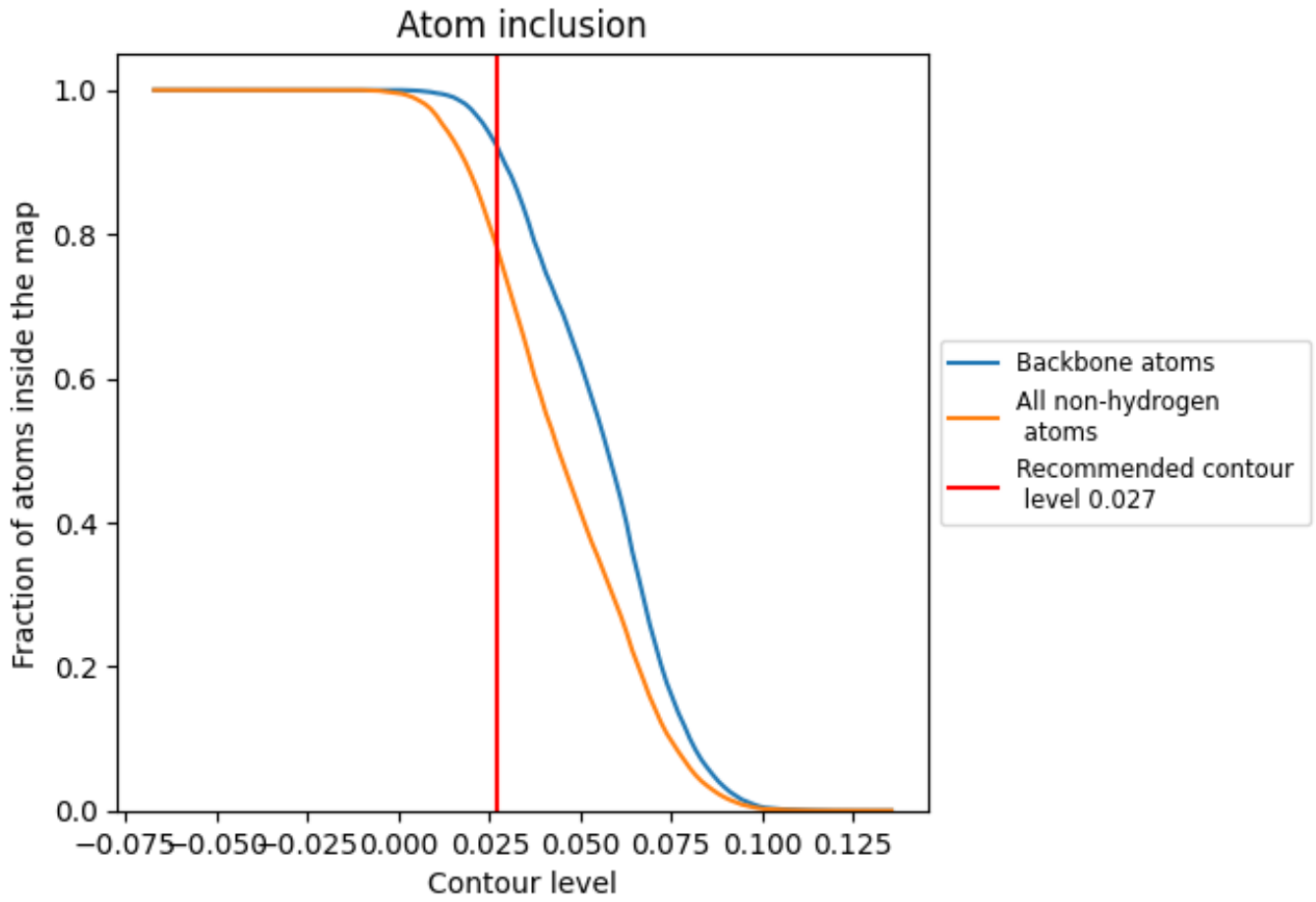
9.3 Atom inclusion mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.027).

9.4 Atom inclusion (i)



At the recommended contour level, 92% of all backbone atoms, 78% 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 (0.027) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7825	0.4500
b	0.7713	0.4370
c	0.7800	0.4500
d	0.7935	0.4580
e	0.7895	0.4600
f	0.7888	0.4580
g	0.7780	0.4510
h	0.7530	0.4180
i	0.7976	0.4550
j	0.7969	0.4580
k	0.7935	0.4590
l	0.7699	0.4460
m	0.7699	0.4450
n	0.7827	0.4450
o	0.7901	0.4570