wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:50 PM BST

PDB ID	:	3J2M
EMDB ID:	:	EMD-1126
Title	:	The X-ray structure of the gp15 hexamer and the model of the gp18 protein fitted into the cryo-EM reconstruction of the extended T4 tail
Authors	:	Fokine, A.; Zhang, Z.; Kanamaru, S.; Bowman, V.D.; Aksyuk, A.; Arisaka, F.; Rao, V.B.; Rossmann, M.G.
Deposited on	:	2012-11-09
Resolution	:	15.00 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

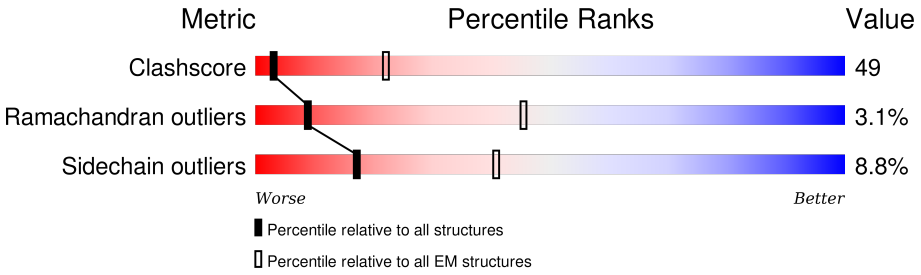
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

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 15.00 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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%

Mol	Chain	Length	Quality of chain
1	A	272	58% 17% • 22%
1	B	272	60% 15% • 22%
1	C	272	57% 17% • 22%
1	D	272	60% 14% • 22%
1	E	272	60% 14% • 22%
1	F	272	60% 15% • 22%
2	U	659	33% 50% 9% 8%
2	V	659	34% 49% 9% 8%
2	W	659	34% 49% 9% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	X	659	34% 49% 9% 8%
2	Y	659	34% 49% 9% 8%
2	Z	659	33% 50% 9% 8%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 38334 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 Tail connector protein Gp15.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	211	Total	C	N	O	S	0	0
			1742	1123	283	328	8		
1	B	211	Total	C	N	O	S	0	0
			1742	1123	283	328	8		
1	C	211	Total	C	N	O	S	0	0
			1742	1123	283	328	8		
1	D	211	Total	C	N	O	S	0	0
			1742	1123	283	328	8		
1	E	211	Total	C	N	O	S	0	0
			1742	1123	283	328	8		
1	F	211	Total	C	N	O	S	0	0
			1742	1123	283	328	8		

- Molecule 2 is a protein called Tail sheath protein Gp18.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	U	609	Total	C	N	O	S	0	0
			4647	2929	785	923	10		
2	V	609	Total	C	N	O	S	0	0
			4647	2929	785	923	10		
2	W	609	Total	C	N	O	S	0	0
			4647	2929	785	923	10		
2	X	609	Total	C	N	O	S	0	0
			4647	2929	785	923	10		
2	Y	609	Total	C	N	O	S	0	0
			4647	2929	785	923	10		
2	Z	609	Total	C	N	O	S	0	0
			4647	2929	785	923	10		

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
U	100	GLU	ASP	SEE REMARK 999	UNP P13332
U	148	ALA	GLY	SEE REMARK 999	UNP P13332

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
U	150	ILE	ASN	SEE REMARK 999	UNP P13332
U	151	ILE	TYR	SEE REMARK 999	UNP P13332
U	301	GLY	GLU	SEE REMARK 999	UNP P13332
U	399	VAL	ALA	SEE REMARK 999	UNP P13332
U	454	TYR	HIS	SEE REMARK 999	UNP P13332
U	510	PRO	ARG	ENGINEERED MUTATION	UNP P13332
V	100	GLU	ASP	SEE REMARK 999	UNP P13332
V	148	ALA	GLY	SEE REMARK 999	UNP P13332
V	150	ILE	ASN	SEE REMARK 999	UNP P13332
V	151	ILE	TYR	SEE REMARK 999	UNP P13332
V	301	GLY	GLU	SEE REMARK 999	UNP P13332
V	399	VAL	ALA	SEE REMARK 999	UNP P13332
V	454	TYR	HIS	SEE REMARK 999	UNP P13332
V	510	PRO	ARG	ENGINEERED MUTATION	UNP P13332
W	100	GLU	ASP	SEE REMARK 999	UNP P13332
W	148	ALA	GLY	SEE REMARK 999	UNP P13332
W	150	ILE	ASN	SEE REMARK 999	UNP P13332
W	151	ILE	TYR	SEE REMARK 999	UNP P13332
W	301	GLY	GLU	SEE REMARK 999	UNP P13332
W	399	VAL	ALA	SEE REMARK 999	UNP P13332
W	454	TYR	HIS	SEE REMARK 999	UNP P13332
W	510	PRO	ARG	ENGINEERED MUTATION	UNP P13332
X	100	GLU	ASP	SEE REMARK 999	UNP P13332
X	148	ALA	GLY	SEE REMARK 999	UNP P13332
X	150	ILE	ASN	SEE REMARK 999	UNP P13332
X	151	ILE	TYR	SEE REMARK 999	UNP P13332
X	301	GLY	GLU	SEE REMARK 999	UNP P13332
X	399	VAL	ALA	SEE REMARK 999	UNP P13332
X	454	TYR	HIS	SEE REMARK 999	UNP P13332
X	510	PRO	ARG	ENGINEERED MUTATION	UNP P13332
Y	100	GLU	ASP	SEE REMARK 999	UNP P13332
Y	148	ALA	GLY	SEE REMARK 999	UNP P13332
Y	150	ILE	ASN	SEE REMARK 999	UNP P13332
Y	151	ILE	TYR	SEE REMARK 999	UNP P13332
Y	301	GLY	GLU	SEE REMARK 999	UNP P13332
Y	399	VAL	ALA	SEE REMARK 999	UNP P13332
Y	454	TYR	HIS	SEE REMARK 999	UNP P13332
Y	510	PRO	ARG	ENGINEERED MUTATION	UNP P13332
Z	100	GLU	ASP	SEE REMARK 999	UNP P13332
Z	148	ALA	GLY	SEE REMARK 999	UNP P13332
Z	150	ILE	ASN	SEE REMARK 999	UNP P13332
Z	151	ILE	TYR	SEE REMARK 999	UNP P13332

Continued on next page...

Continued from previous page...

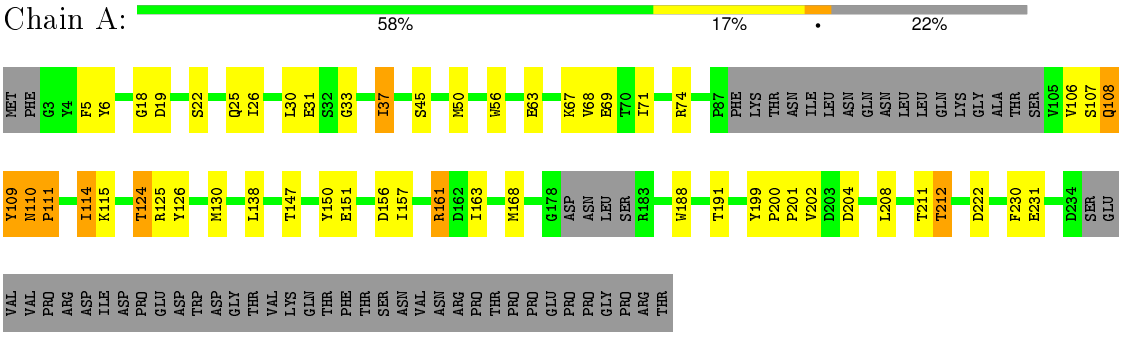
Chain	Residue	Modelled	Actual	Comment	Reference
Z	301	GLY	GLU	SEE REMARK 999	UNP P13332
Z	399	VAL	ALA	SEE REMARK 999	UNP P13332
Z	454	TYR	HIS	SEE REMARK 999	UNP P13332
Z	510	PRO	ARG	ENGINEERED MUTATION	UNP P13332

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

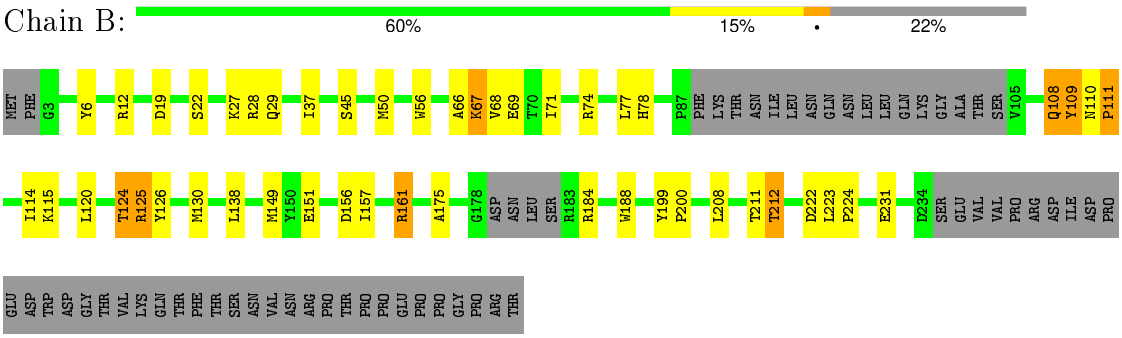
- Molecule 1: Tail connector protein Gp15

Chain A:



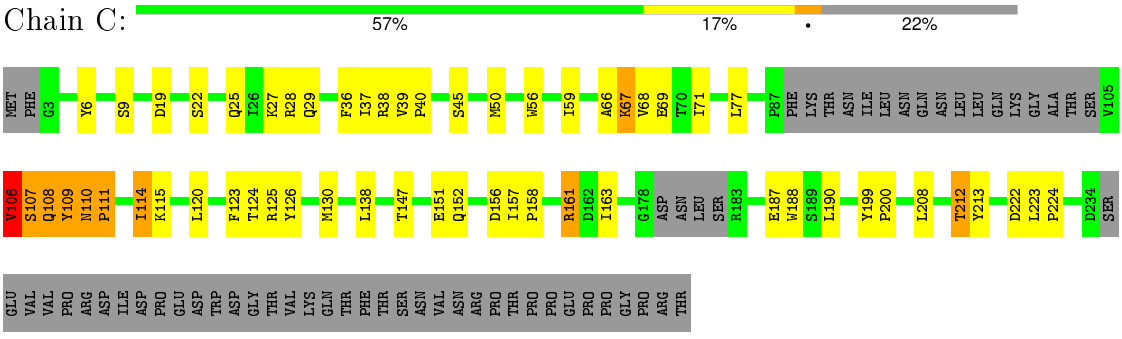
- Molecule 1: Tail connector protein Gp15

Chain B:



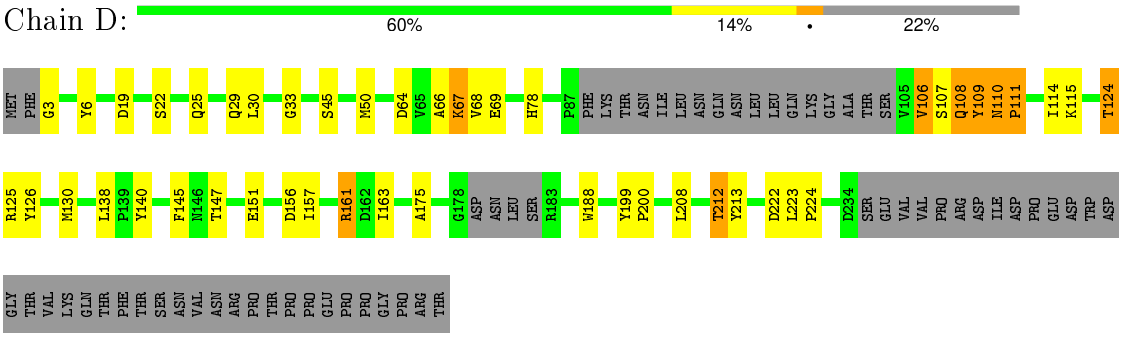
- Molecule 1: Tail connector protein Gp15

Chain C:



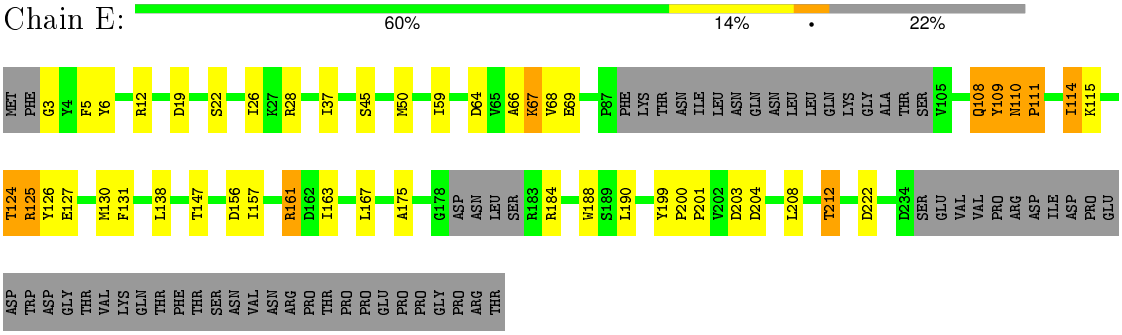
- Molecule 1: Tail connector protein Gp15

Chain D:



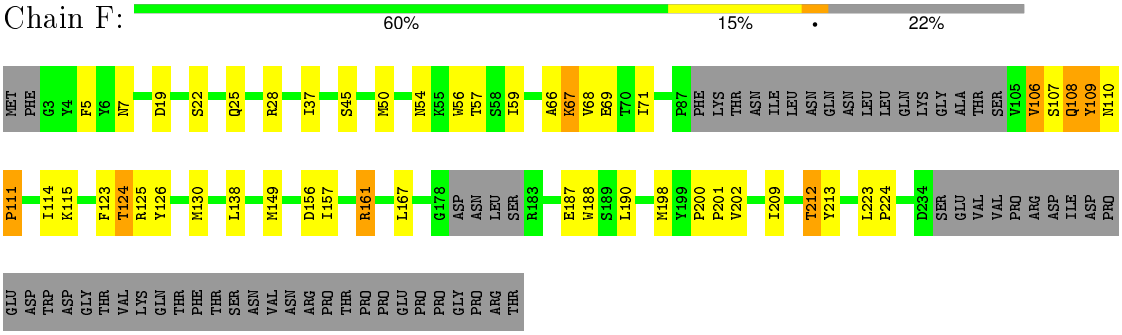
- Molecule 1: Tail connector protein Gp15

Chain E:



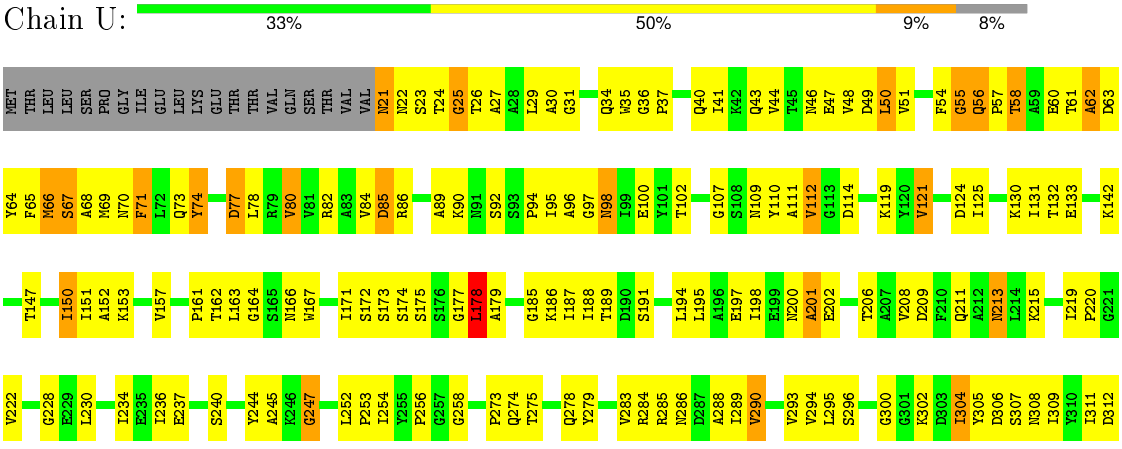
- Molecule 1: Tail connector protein Gp15

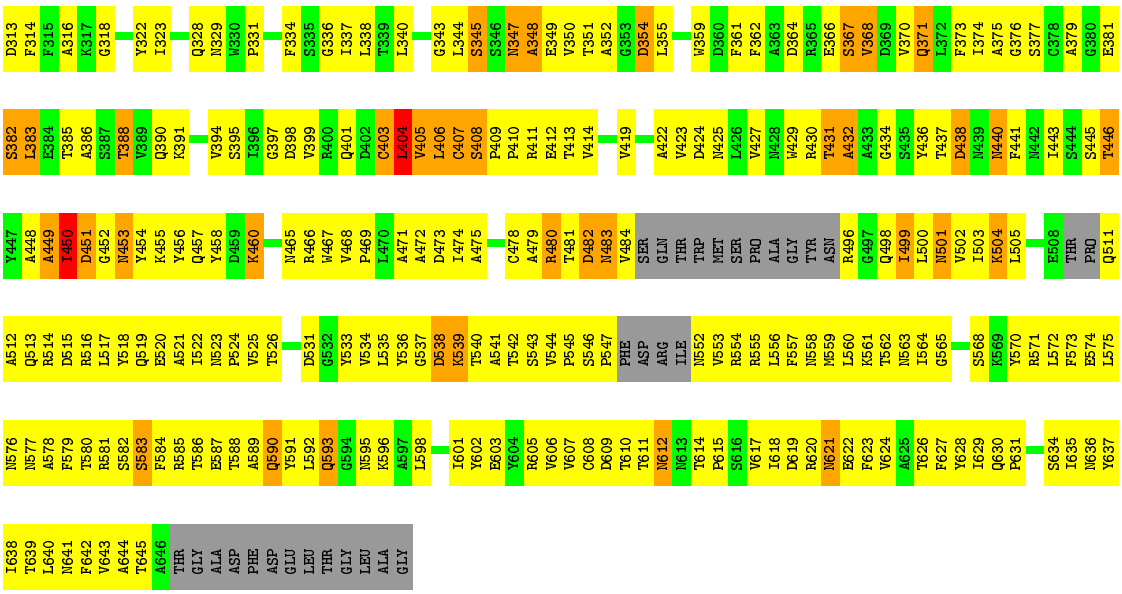
Chain F:



- Molecule 2: Tail sheath protein Gp18

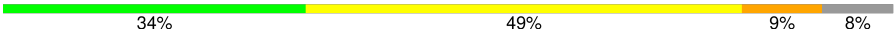
Chain U:

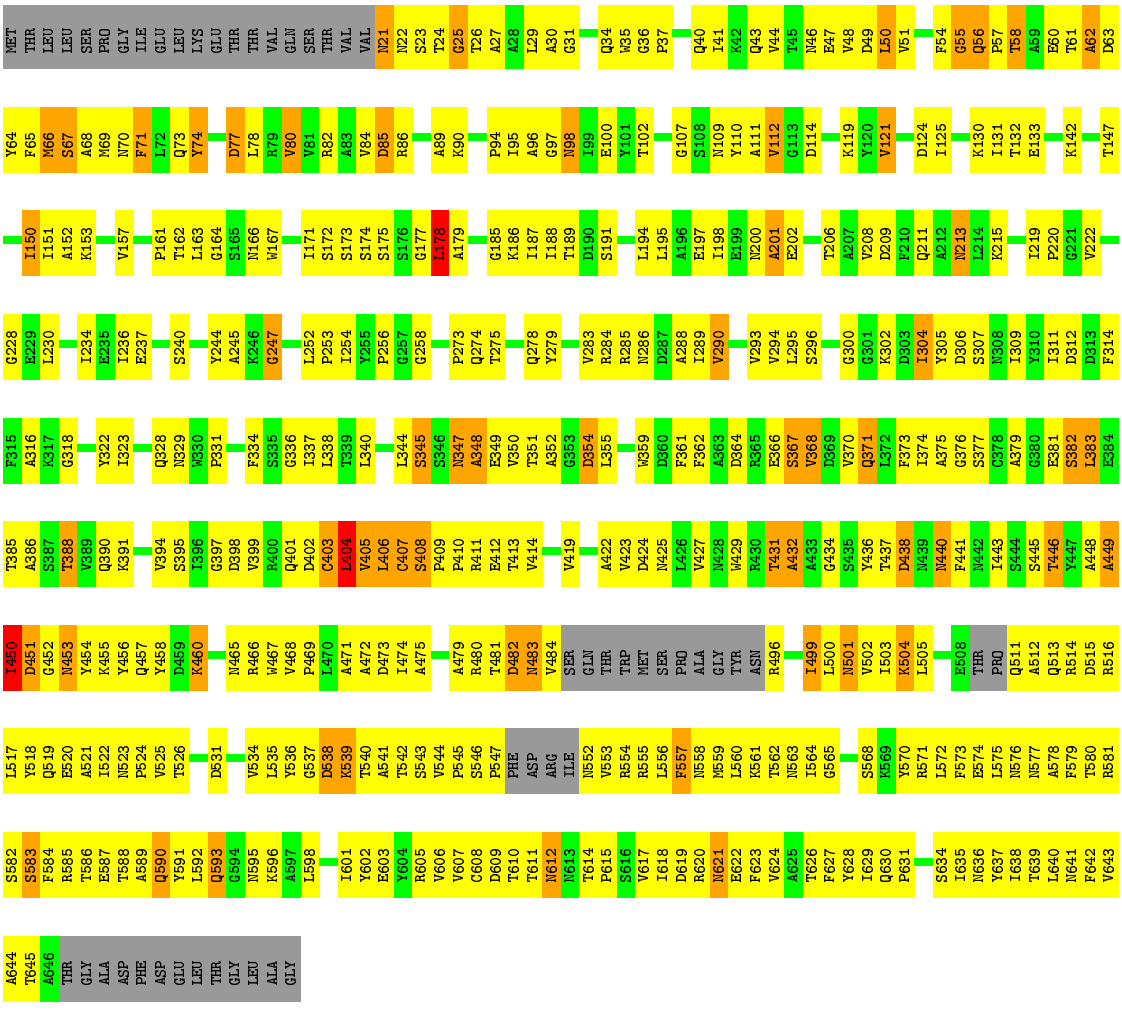




• Molecule 2: Tail sheath protein Gp18

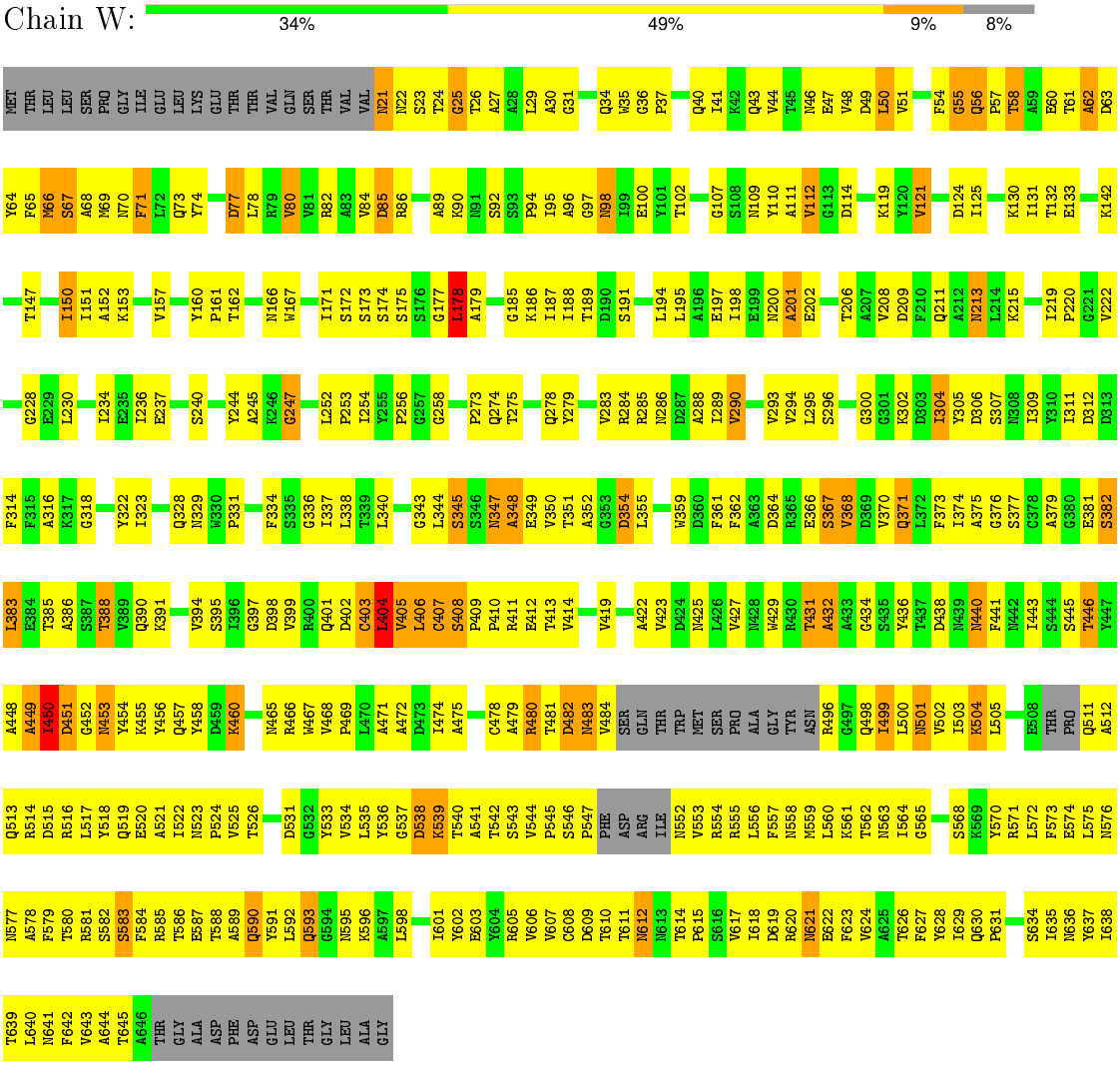
Chain V:





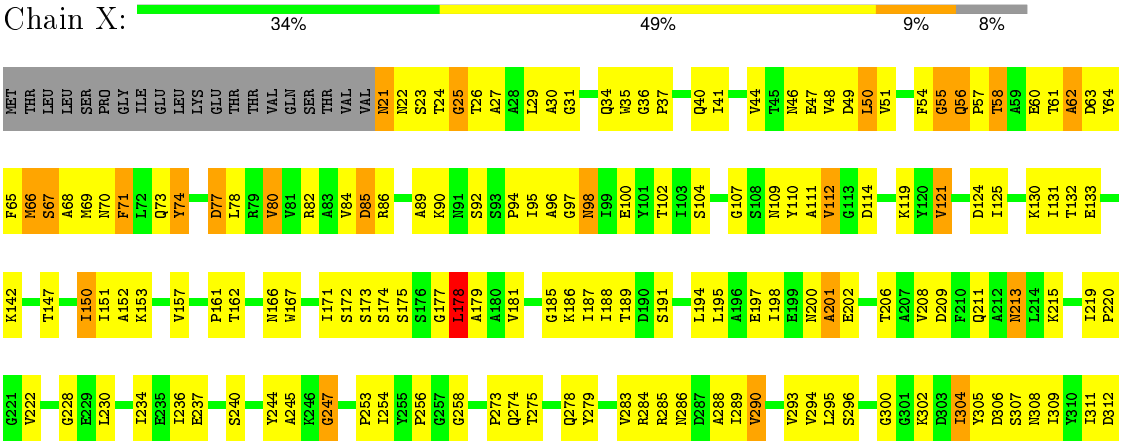
- Molecule 2: Tail sheath protein Gp18

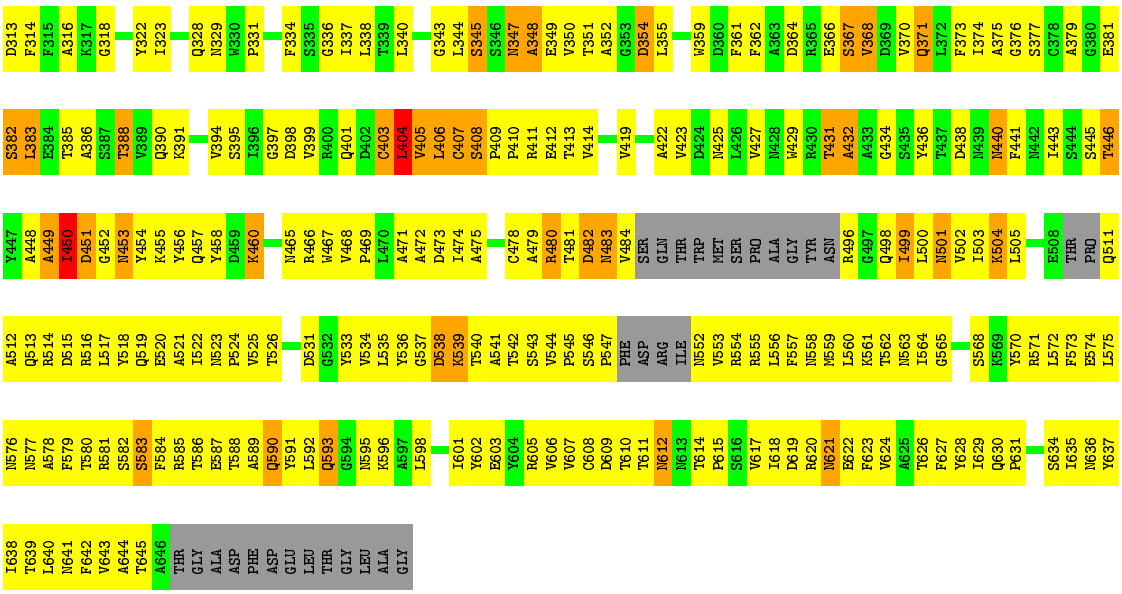
Chain W:



- Molecule 2: Tail sheath protein Gp18

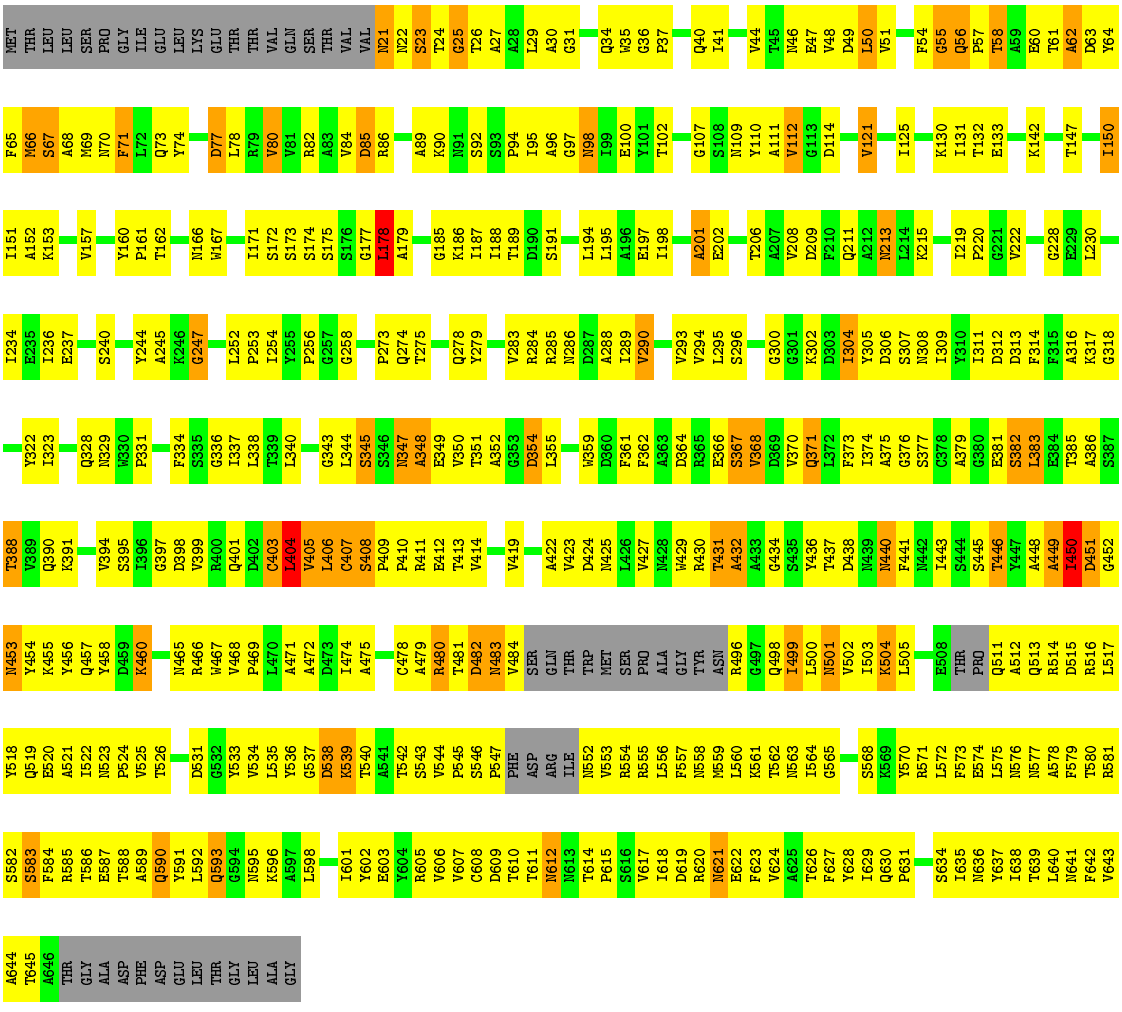
Chain X:





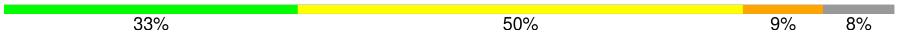
• Molecule 2: Tail sheath protein Gp18

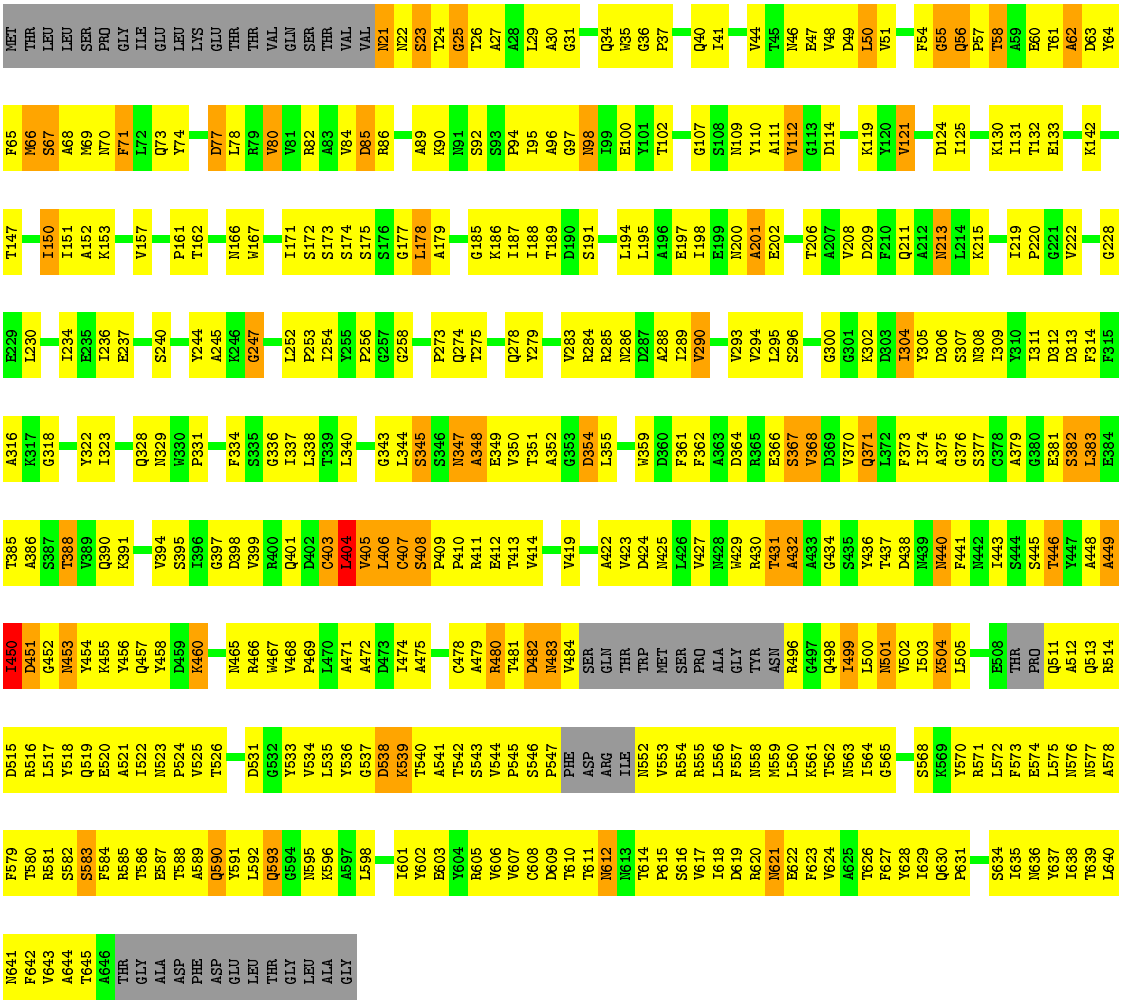
Chain Y:



- Molecule 2: Tail sheath protein Gp18

Chain Z:





4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	Not provided	Depositor
Resolution determination method	FSC at 0.5 cut-off	Depositor
CTF correction method	each particle image	Depositor
Microscope	FEI/PHILIPS CM300FEG/ST	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	Not provided	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	4000	Depositor
Magnification	45000	Depositor
Image detector	Not provided	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	#$\|Z\|>2$	RMSZ	#$\|Z\|>2$
1	A	0.49	0/1787	0.64	0/2421
1	B	0.47	0/1787	0.63	0/2421
1	C	0.48	0/1787	0.64	1/2421 (0.0%)
1	D	0.49	0/1787	0.63	0/2421
1	E	0.49	0/1787	0.63	0/2421
1	F	0.50	0/1787	0.65	0/2421
2	U	0.60	0/4729	0.89	21/6427 (0.3%)
2	V	0.60	0/4729	0.90	21/6427 (0.3%)
2	W	0.60	0/4729	0.89	21/6427 (0.3%)
2	X	0.60	0/4729	0.90	21/6427 (0.3%)
2	Y	0.60	1/4729 (0.0%)	0.89	20/6427 (0.3%)
2	Z	0.60	1/4729 (0.0%)	0.89	19/6427 (0.3%)
All	All	0.57	2/39096 (0.0%)	0.83	124/53088 (0.2%)

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
2	U	0	4
2	V	0	4
2	W	0	4
2	X	0	4
2	Y	0	4
2	Z	0	4
All	All	0	24

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	Z	23	SER	C-O	5.16	1.33	1.23
2	Y	23	SER	C-O	5.03	1.32	1.23

The worst 5 of 124 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	V	55	GLY	N-CA-C	18.09	158.33	113.10
2	U	55	GLY	N-CA-C	18.03	158.18	113.10
2	Y	55	GLY	N-CA-C	18.01	158.12	113.10
2	W	55	GLY	N-CA-C	17.99	158.08	113.10
2	X	55	GLY	N-CA-C	17.93	157.92	113.10

There are no chirality outliers.

5 of 24 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	U	21	ASN	Peptide
2	U	450	ILE	Peptide
2	U	451	ASP	Peptide
2	U	452	GLY	Peptide
2	V	21	ASN	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	A	1742	0	1673	54	0
1	B	1742	0	1673	59	0
1	C	1742	0	1673	49	0
1	D	1742	0	1673	44	0
1	E	1742	0	1673	46	0
1	F	1742	0	1673	36	0
2	U	4647	0	4564	595	0
2	V	4647	0	4564	592	0
2	W	4647	0	4564	588	0
2	X	4647	0	4564	587	0
2	Y	4647	0	4562	576	0
2	Z	4647	0	4564	577	0
All	All	38334	0	37420	3690	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 49.

The worst 5 of 3690 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:156:ASP:HB3	2:U:579:PHE:CE1	1.23	1.68
1:B:157:ILE:CD1	2:V:579:PHE:HB3	1.20	1.63
1:B:157:ILE:HG13	2:V:579:PHE:CB	1.21	1.57
1:B:156:ASP:HB3	2:V:579:PHE:CE1	1.39	1.54
1:A:157:ILE:HG13	2:U:579:PHE:CB	1.22	1.54

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	A	205/272 (75%)	184 (90%)	18 (9%)	3 (2%)	13	57
1	B	205/272 (75%)	184 (90%)	17 (8%)	4 (2%)	9	51
1	C	205/272 (75%)	184 (90%)	17 (8%)	4 (2%)	9	51
1	D	205/272 (75%)	184 (90%)	17 (8%)	4 (2%)	9	51
1	E	205/272 (75%)	183 (89%)	18 (9%)	4 (2%)	9	51
1	F	205/272 (75%)	184 (90%)	18 (9%)	3 (2%)	13	57
2	U	601/659 (91%)	479 (80%)	101 (17%)	21 (4%)	4	39
2	V	601/659 (91%)	480 (80%)	100 (17%)	21 (4%)	4	39
2	W	601/659 (91%)	479 (80%)	101 (17%)	21 (4%)	4	39
2	X	601/659 (91%)	479 (80%)	101 (17%)	21 (4%)	4	39
2	Y	601/659 (91%)	479 (80%)	101 (17%)	21 (4%)	4	39
2	Z	601/659 (91%)	480 (80%)	100 (17%)	21 (4%)	4	39
All	All	4836/5586 (87%)	3979 (82%)	709 (15%)	148 (3%)	9	42

5 of 148 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	108	GLN
1	B	67	LYS
1	B	108	GLN
1	C	67	LYS
1	C	108	GLN

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	A	192/250 (77%)	176 (92%)	16 (8%)	14	49
1	B	192/250 (77%)	181 (94%)	11 (6%)	25	62
1	C	192/250 (77%)	176 (92%)	16 (8%)	14	49
1	D	192/250 (77%)	180 (94%)	12 (6%)	22	59
1	E	192/250 (77%)	180 (94%)	12 (6%)	22	59
1	F	192/250 (77%)	179 (93%)	13 (7%)	20	57
2	U	494/536 (92%)	446 (90%)	48 (10%)	10	40
2	V	494/536 (92%)	447 (90%)	47 (10%)	11	41
2	W	494/536 (92%)	448 (91%)	46 (9%)	11	42
2	X	494/536 (92%)	448 (91%)	46 (9%)	11	42
2	Y	494/536 (92%)	446 (90%)	48 (10%)	10	40
2	Z	494/536 (92%)	446 (90%)	48 (10%)	10	40
All	All	4116/4716 (87%)	3753 (91%)	363 (9%)	17	45

5 of 363 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	V	446	THR
2	W	388	THR
2	Z	347	ASN
2	V	499	ILE
2	W	95	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 100 such sidechains are listed below:

Mol	Chain	Res	Type
2	W	523	ASN
2	X	390	GLN
2	Z	511	GLN
2	W	590	GLN
2	W	641	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 carbohydrates 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

There are no chain breaks in this entry.