



wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Jul 17, 2017 – 10:56 PM EDT

PDB ID : 5HX2
EMDB ID: : EMD-8064
Title : In vitro assembled star-shaped hubless T4 baseplate
Authors : Yap, M.L.; Klose, T.; Fokine, A.; Rossmann, M.G.
Deposited on : unknown
Resolution : 3.80 Å(reported)
Based on PDB ID : 1N80, 3H2T, 2FKK

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report
for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

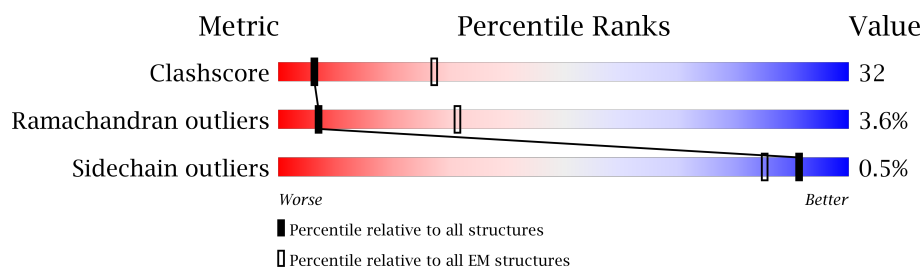
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.80 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	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	1032	
2	B	334	
2	C	334	
3	D	660	
3	E	660	
4	F	196	
5	G	602	
5	H	602	
5	I	602	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 28612 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 Baseplate wedge protein gp7.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1030	Total	C	N	O	S	0	0
			7511	4844	1265	1383	19		

- Molecule 2 is a protein called Baseplate wedge protein gp8.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	334	Total	C	N	O	S	0	0
			2521	1598	423	485	15		
2	C	332	Total	C	N	O	S	0	0
			2585	1650	428	491	16		

- Molecule 3 is a protein called Baseplate wedge protein gp6.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	647	Total	C	N	O	S	0	0
			4845	3072	811	953	9		
3	E	634	Total	C	N	O	S	0	0
			4660	2966	784	902	8		

- Molecule 4 is a protein called Baseplate wedge protein gp53.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	F	192	Total	C	N	O	S	0	0
			1495	972	239	278	6		

- Molecule 5 is a protein called Baseplate wedge protein gp10.

Mol	Chain	Residues	Atoms				AltConf	Trace
5	G	340	Total	C	N	O	0	0
			1665	985	340	340		
5	H	340	Total	C	N	O	0	0
			1665	985	340	340		

Continued on next page...

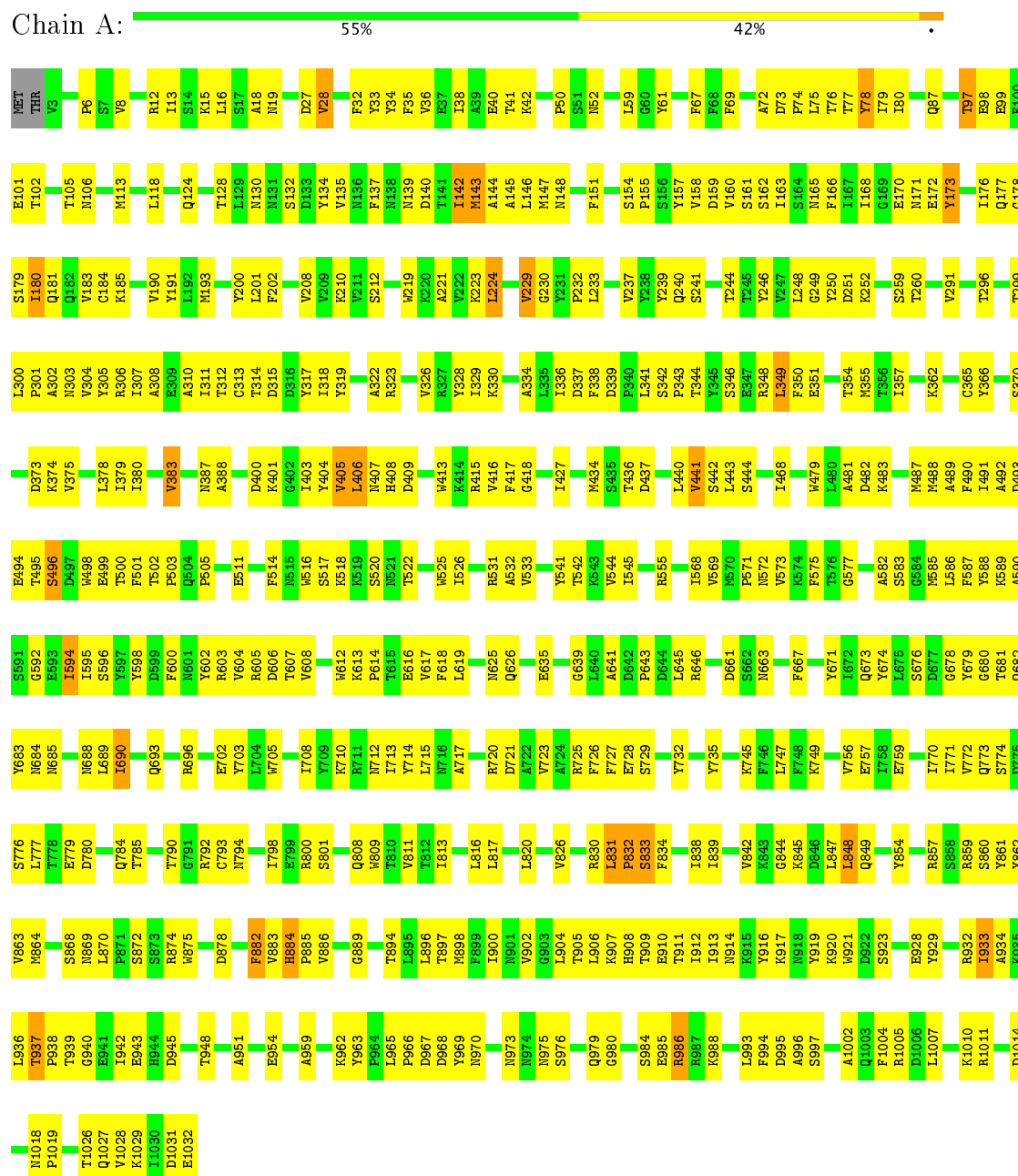
Continued from previous page...

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
5	I	340	1665	985	340	340	0	0

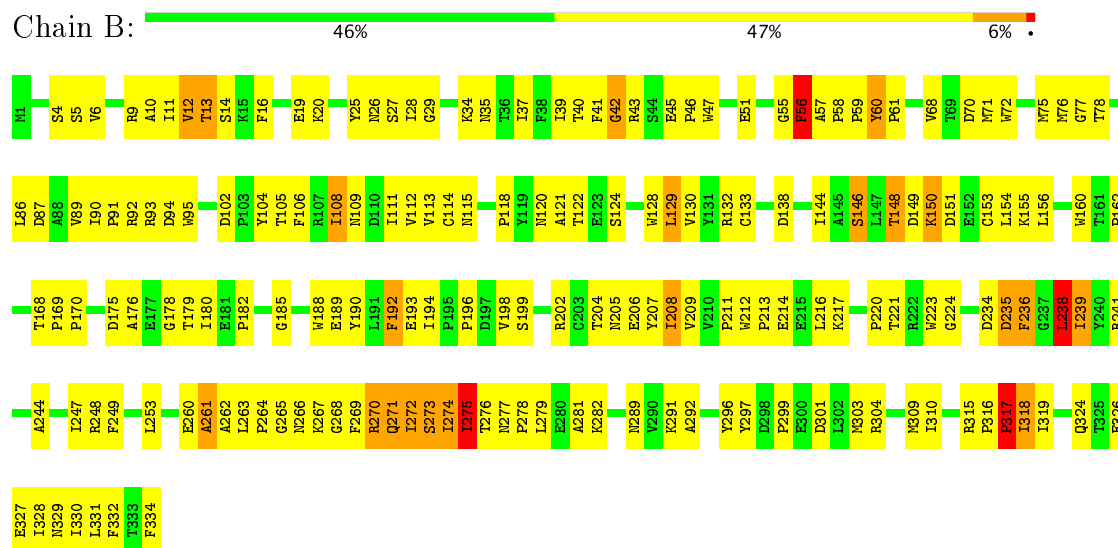
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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. 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.

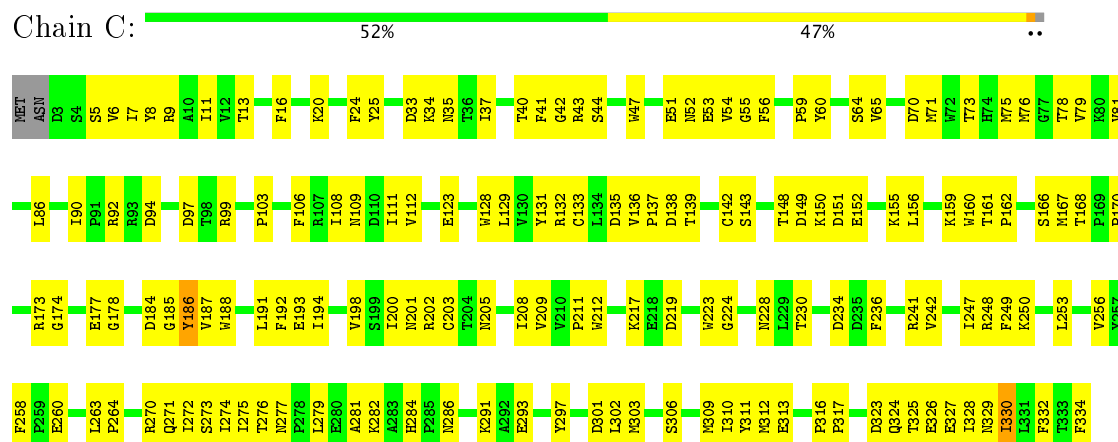
• Molecule 1: Baseplate wedge protein gp7



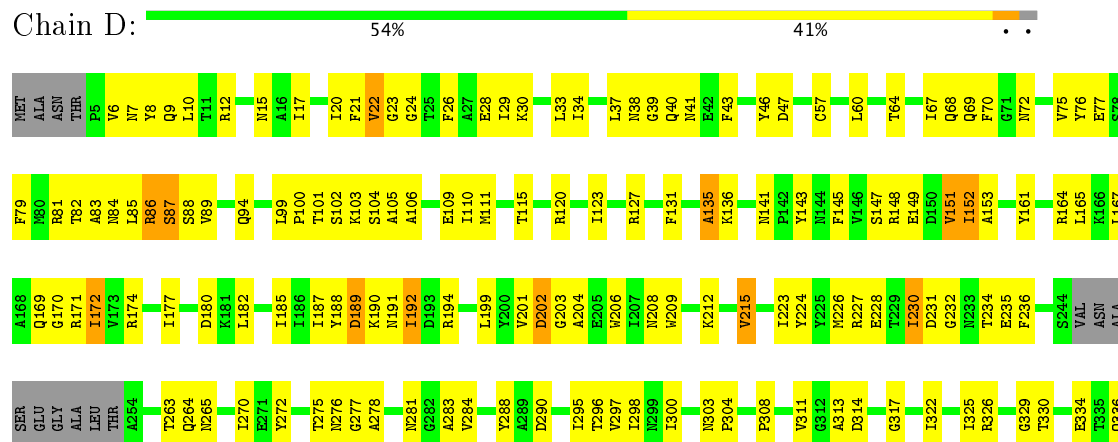
- Molecule 2: Baseplate wedge protein gp8

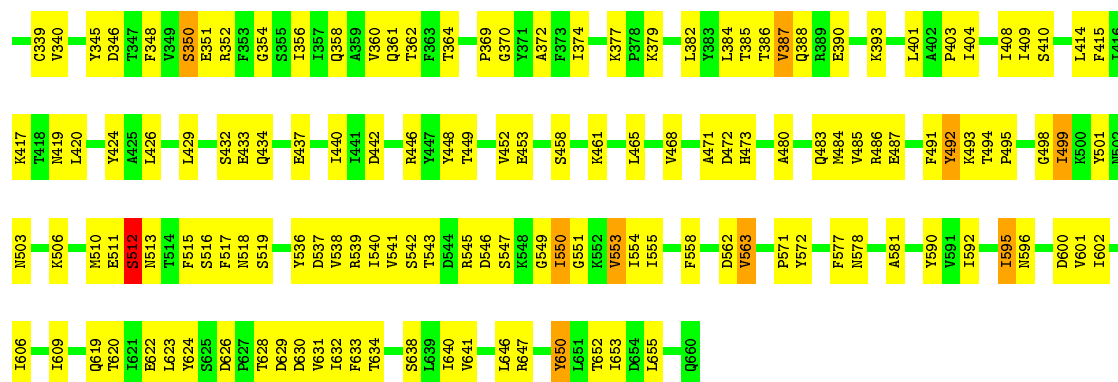


- Molecule 2: Baseplate wedge protein gp8

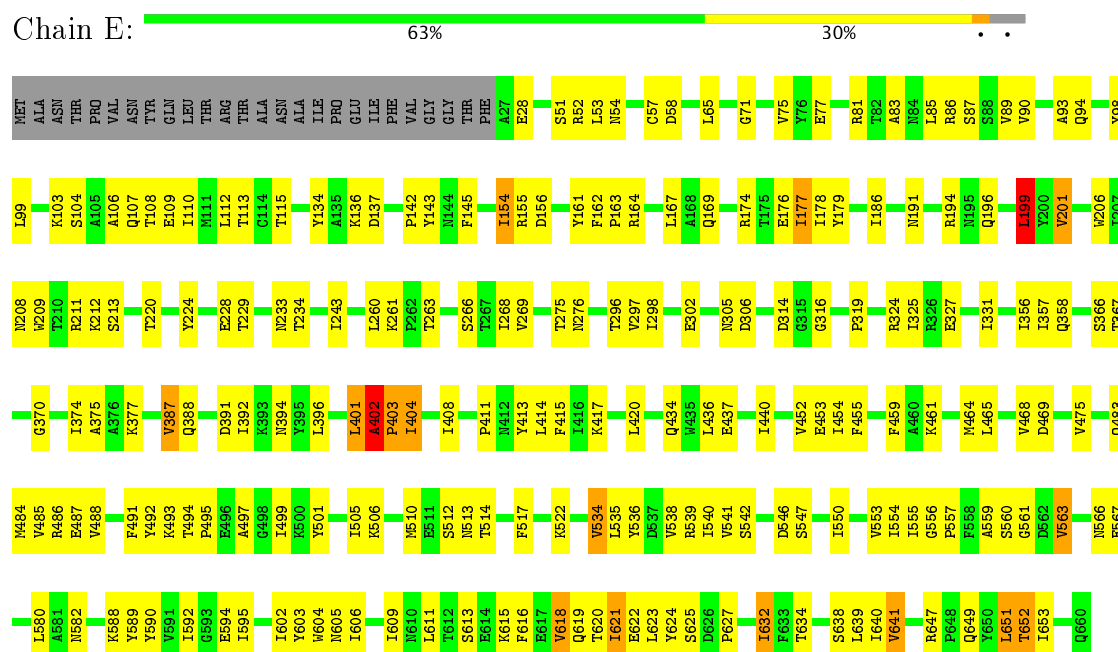


- Molecule 3: Baseplate wedge protein gp6





• Molecule 3: Baseplate wedge protein gp6

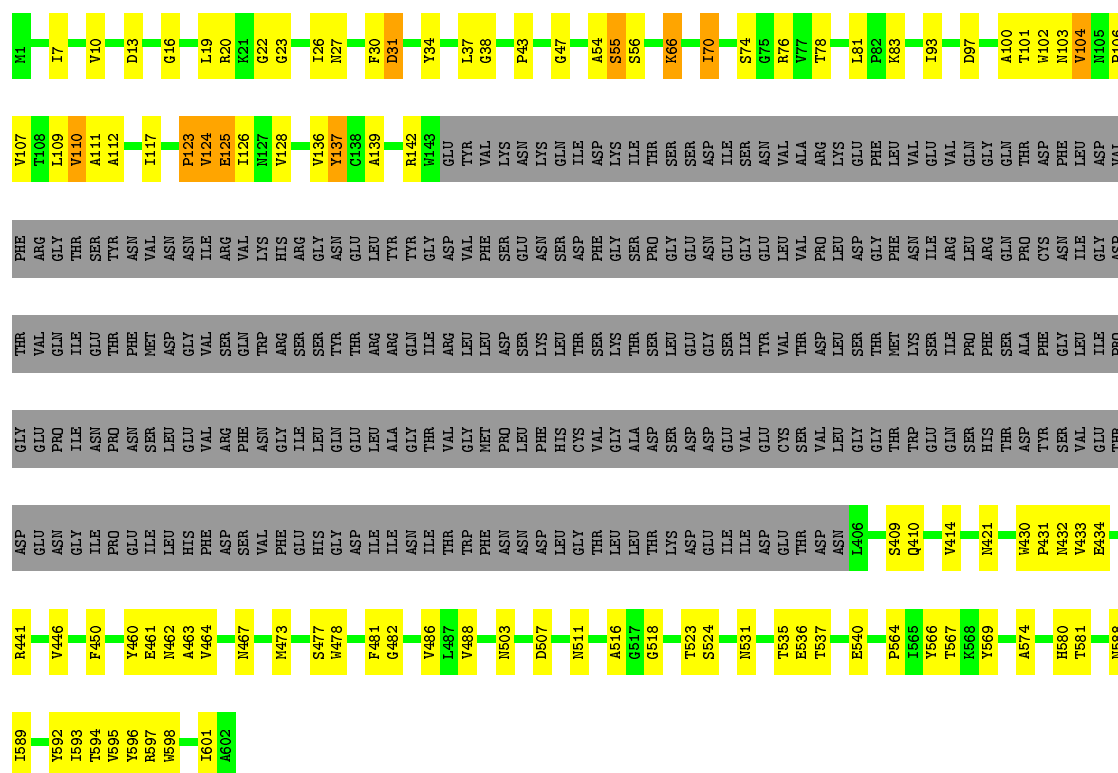


• Molecule 4: Baseplate wedge protein gp53

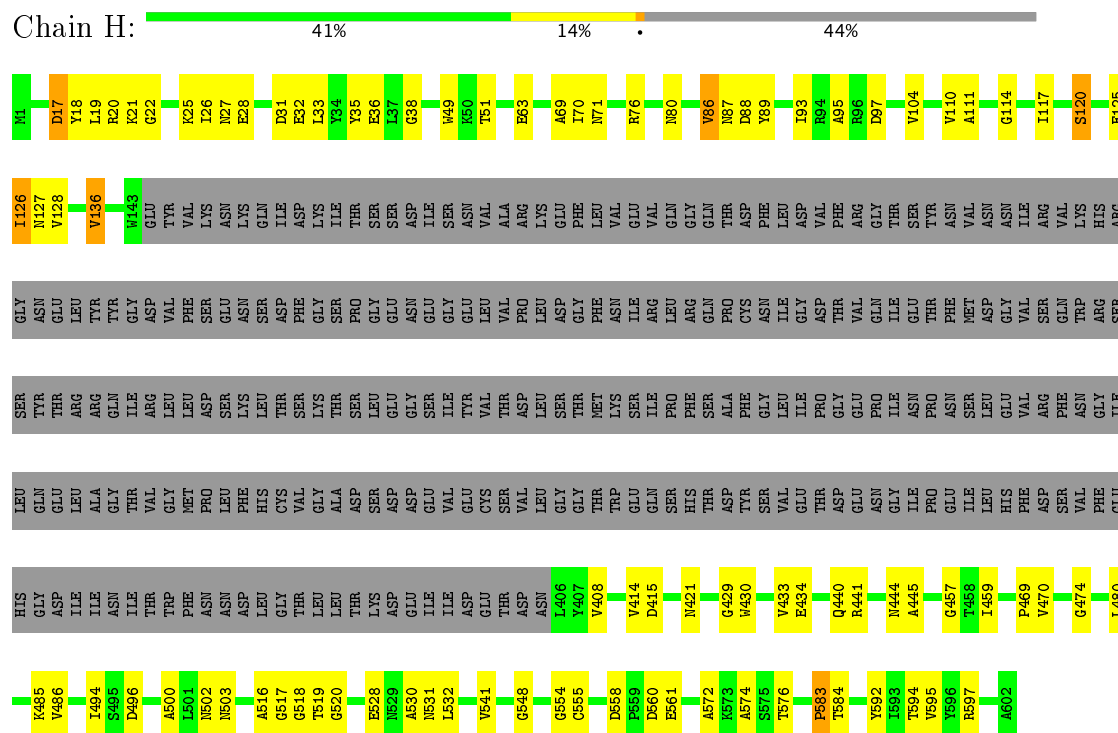


• Molecule 5: Baseplate wedge protein gp10



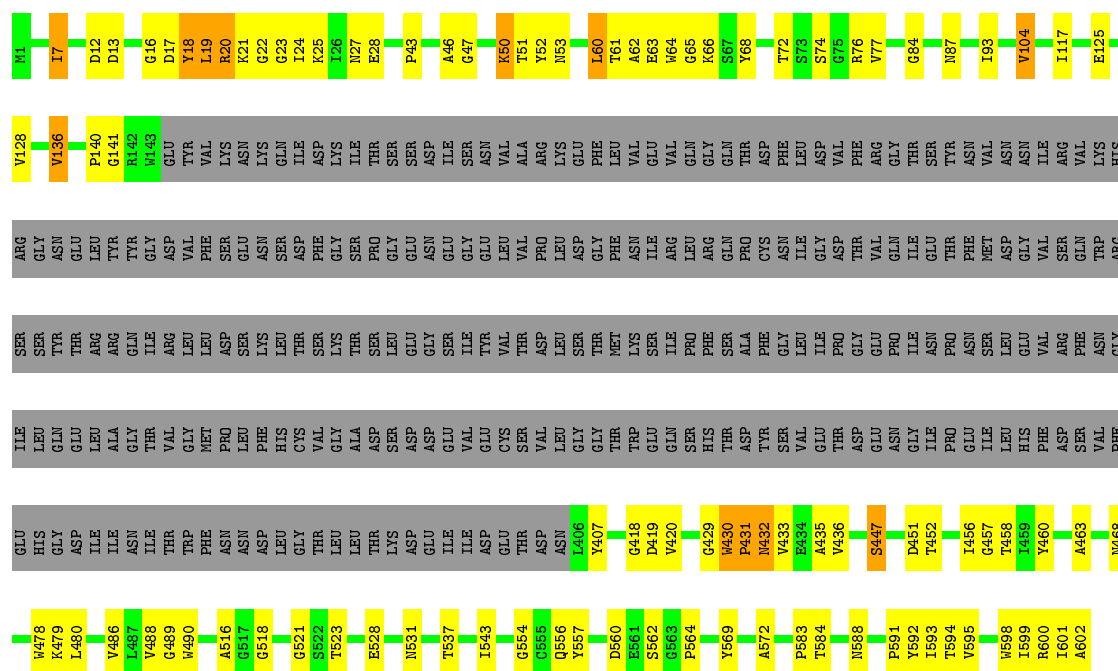


• Molecule 5: Baseplate wedge protein gp10



• Molecule 5: Baseplate wedge protein gp10





4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C6	Depositor
Number of particles used	45607	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 ($e^-/\text{\AA}^2$)	35	Depositor
Minimum defocus (nm)	400	Depositor
Maximum defocus (nm)	3740	Depositor
Magnification	38168	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	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	0.37	0/7697	0.62	2/10550 (0.0%)
2	B	0.51	0/2591	0.72	2/3546 (0.1%)
2	C	0.55	0/2662	0.65	0/3640
3	D	0.30	0/4935	0.61	0/6735
3	E	0.28	0/4750	0.56	0/6485
4	F	0.52	0/1536	0.64	0/2095
5	G	0.33	0/1663	0.68	0/2304
5	H	0.34	0/1663	0.72	0/2304
5	I	0.35	0/1663	0.68	0/2304
All	All	0.39	0/29160	0.64	4/39963 (0.0%)

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	A	0	5
2	B	0	3
2	C	0	2
3	D	0	9
3	E	0	5
4	F	0	1
5	G	0	9
5	H	0	10
5	I	0	11
All	All	0	55

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	848	LEU	CA-CB-CG	5.42	127.77	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	274	ILE	N-CA-C	-5.34	96.59	111.00
2	B	42	GLY	N-CA-C	5.29	126.31	113.10
1	A	349	LEU	CA-CB-CG	5.08	126.98	115.30

There are no chirality outliers.

5 of 55 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	496	SER	Peptide
1	A	831	LEU	Peptide
1	A	832	PRO	Peptide
1	A	917	LYS	Peptide
1	A	97	THR	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	7511	0	6716	677	0
2	B	2521	0	2282	345	0
2	C	2585	0	2410	232	0
3	D	4845	0	4477	346	0
3	E	4660	0	4258	182	0
4	F	1495	0	1361	121	0
5	G	1665	0	764	81	0
5	H	1665	0	764	81	0
5	I	1665	0	764	86	0
All	All	28612	0	23796	1674	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 32.

The worst 5 of 1674 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)
2:B:47:TRP:CZ2	2:B:59:PRO:HD2	1.18	1.67

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:902:VAL:HG21	2:B:16:PHE:CE2	1.20	1.67
1:A:685:ASN:ND2	3:D:21:PHE:CE1	1.68	1.58
1:A:663:ASN:CB	3:D:46:TYR:HA	1.11	1.57
1:A:898:MET:HE3	2:C:272:ILE:CD1	1.33	1.56

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	1028/1032 (100%)	803 (78%)	198 (19%)	27 (3%)	6	43
2	B	332/334 (99%)	266 (80%)	47 (14%)	19 (6%)	2	27
2	C	330/334 (99%)	286 (87%)	41 (12%)	3 (1%)	20	62
3	D	643/660 (97%)	454 (71%)	162 (25%)	27 (4%)	3	33
3	E	632/660 (96%)	483 (76%)	131 (21%)	18 (3%)	6	42
4	F	190/196 (97%)	151 (80%)	32 (17%)	7 (4%)	4	36
5	G	336/602 (56%)	250 (74%)	69 (20%)	17 (5%)	2	29
5	H	336/602 (56%)	253 (75%)	71 (21%)	12 (4%)	4	37
5	I	336/602 (56%)	254 (76%)	64 (19%)	18 (5%)	2	28
All	All	4163/5022 (83%)	3200 (77%)	815 (20%)	148 (4%)	7	37

5 of 148 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	28	VAL
1	A	383	VAL
1	A	468	ILE
1	A	690	ILE
1	A	833	SER

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	A	683/921 (74%)	681 (100%)	2 (0%)	94	97
2	B	252/295 (85%)	245 (97%)	7 (3%)	49	76
2	C	272/295 (92%)	272 (100%)	0	100	100
3	D	479/578 (83%)	477 (100%)	2 (0%)	93	97
3	E	440/578 (76%)	439 (100%)	1 (0%)	94	98
4	F	140/169 (83%)	140 (100%)	0	100	100
All	All	2266/2836 (80%)	2254 (100%)	12 (0%)	91	96

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	238	LEU
2	B	239	ILE
3	D	86	ARG
2	B	146	SER
2	B	275	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	26	ASN
2	B	305	HIS
3	E	582	ASN
2	B	205	ASN
2	B	233	GLN

5.3.3 RNA ⓘ

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 carbohydrates 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.