wwPDB EM Map/Model Validation Report (i)

Oct 6, 2016 – 05:32 PM EDT

PDB ID	:	5GAI
EMDB ID:	:	EMD-8005
Title	:	Probabilistic Structural Models of Mature P22 Bacteriophage Portal, Hub, and Tailspike proteins
Authors	:	Pintilie, G.; Chen, D.H.; Haase-Pettingell, C.A.; King, J.A.; Chiu, W.
Deposited on	:	2015-12-01
Resolution	:	10.50 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

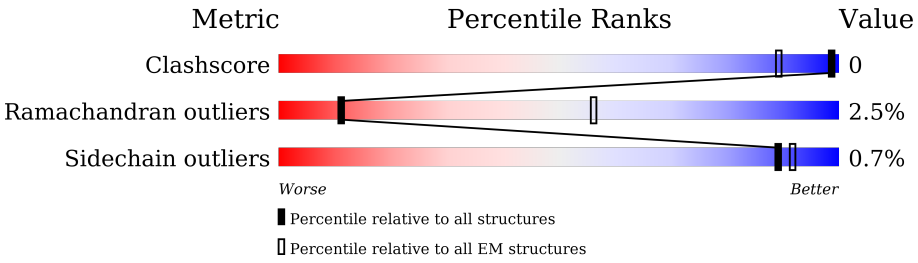
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 10.50 Å.

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



Metric	Whole archive (#Entries)	EM structures (#Entries)
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	721	96%
1	B	721	96%
1	C	721	94% 6%
1	D	721	95% 5%
1	E	721	95%
1	F	721	96%
1	G	721	95% 5%
1	H	721	96%
1	I	721	97%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	J	721	95% 5%
1	W	721	95% 5%
1	X	721	97%
2	K	146	97%
2	L	146	97%
2	M	146	97%
2	N	146	99%
2	O	146	98%
2	P	146	97%
2	Q	146	99%
2	R	146	95% 5%
2	S	146	99%
2	T	146	97%
2	U	146	94% 5%
2	V	146	99%
3	0	662	96%
3	Y	662	94% 6%
3	Z	662	97%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 98043 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 Portal protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	B	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	C	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	D	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	E	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	F	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	G	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	H	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	I	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	J	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	W	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		
1	X	721	Total	C	N	O	S	0	0
			5792	3622	1005	1141	24		

- Molecule 2 is a protein called Peptidoglycan hydrolase gp4.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	K	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	L	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	M	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					AltConf	Trace
2	N	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	O	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	P	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	Q	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	R	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	S	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	T	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	U	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		
2	V	146	Total	C	N	O	S	0	0
			1122	704	192	221	5		

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	150	PRO	ALA	engineered mutation	UNP P26746
L	150	PRO	ALA	engineered mutation	UNP P26746
M	150	PRO	ALA	engineered mutation	UNP P26746
N	150	PRO	ALA	engineered mutation	UNP P26746
O	150	PRO	ALA	engineered mutation	UNP P26746
P	150	PRO	ALA	engineered mutation	UNP P26746
Q	150	PRO	ALA	engineered mutation	UNP P26746
R	150	PRO	ALA	engineered mutation	UNP P26746
S	150	PRO	ALA	engineered mutation	UNP P26746
T	150	PRO	ALA	engineered mutation	UNP P26746
U	150	PRO	ALA	engineered mutation	UNP P26746
V	150	PRO	ALA	engineered mutation	UNP P26746

- Molecule 3 is a protein called Tail fiber protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	Y	662	Total	C	N	O	S	0	0
			5025	3169	855	985	16		
3	Z	662	Total	C	N	O	S	0	0
			5025	3169	855	985	16		

Continued on next page...

Continued from previous page...

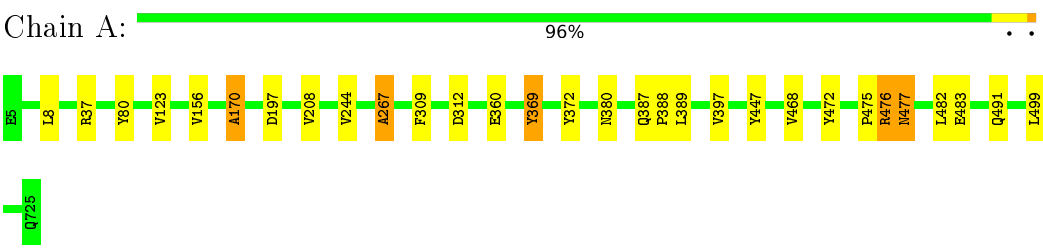
Mol	Chain	Residues	Atoms					AltConf	Trace
3	0	662	Total	C	N	O	S	0	0
			5025	3169	855	985	16		

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.

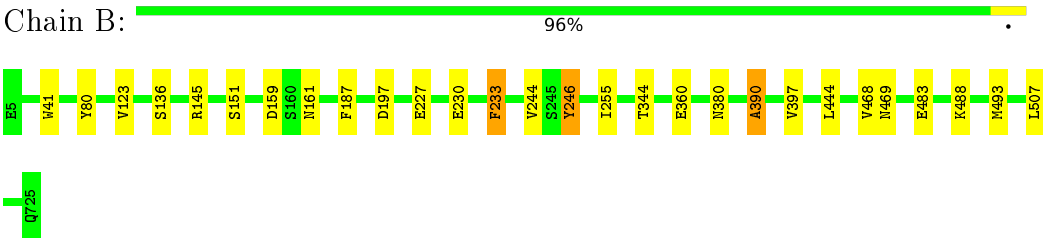
- Molecule 1: Portal protein

Chain A:



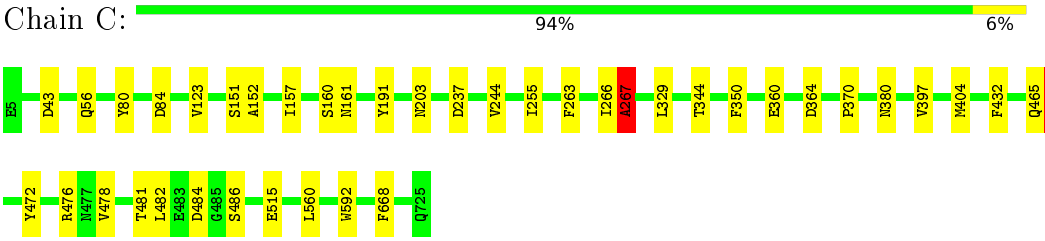
- Molecule 1: Portal protein

Chain B:



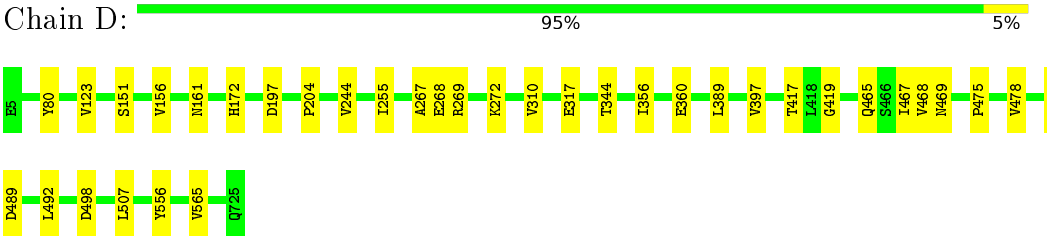
- Molecule 1: Portal protein

Chain C:



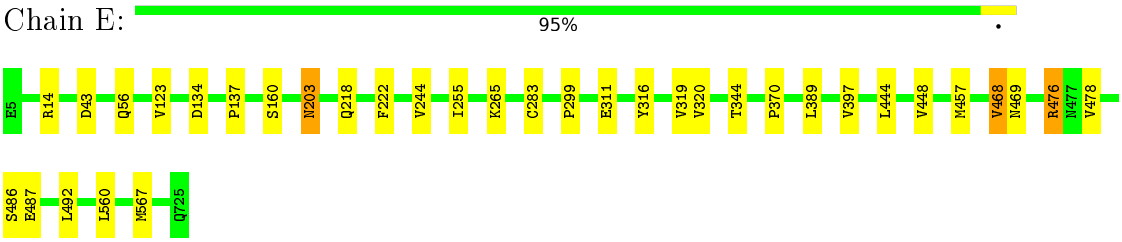
- Molecule 1: Portal protein

Chain D:



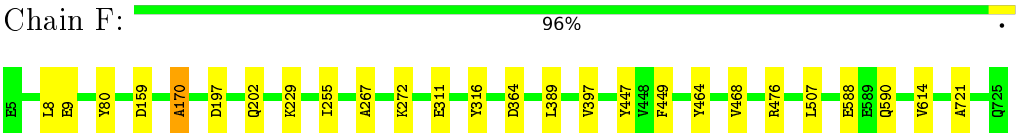
- Molecule 1: Portal protein

Chain E:



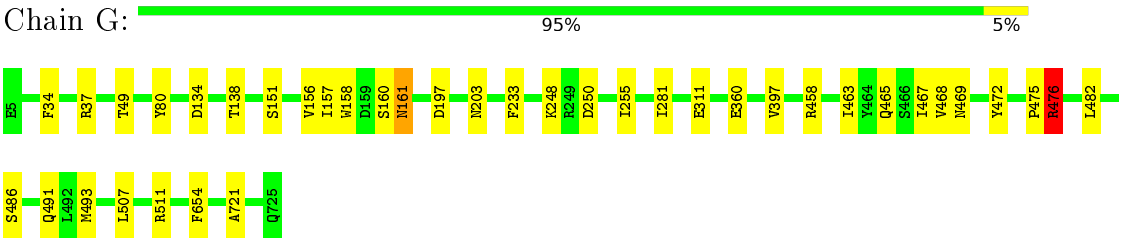
• Molecule 1: Portal protein

Chain F:



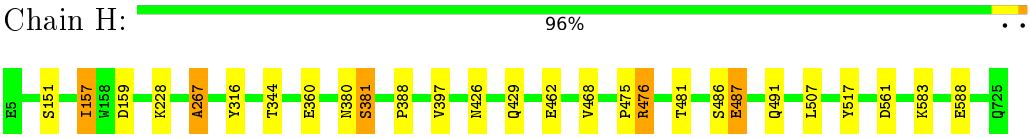
• Molecule 1: Portal protein

Chain G:



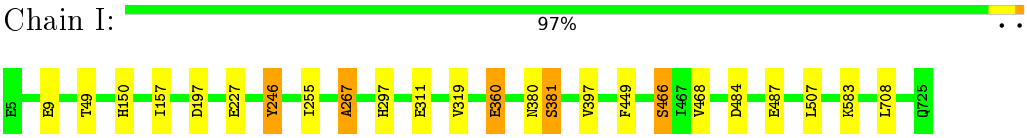
• Molecule 1: Portal protein

Chain H:



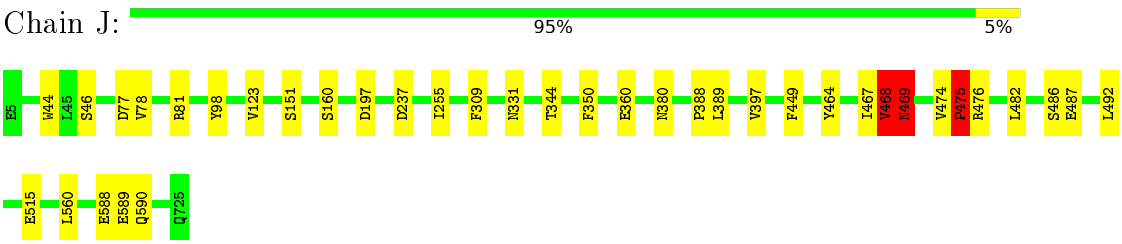
• Molecule 1: Portal protein

Chain I:



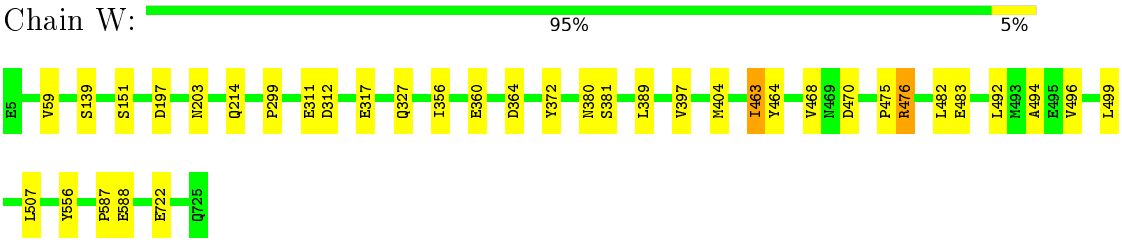
• Molecule 1: Portal protein

Chain J:



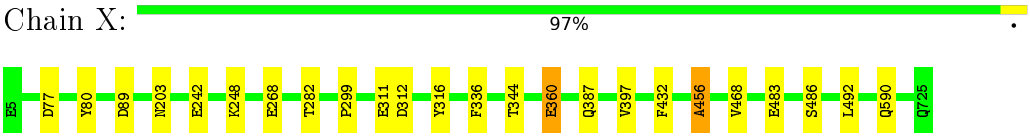
• Molecule 1: Portal protein

Chain W:



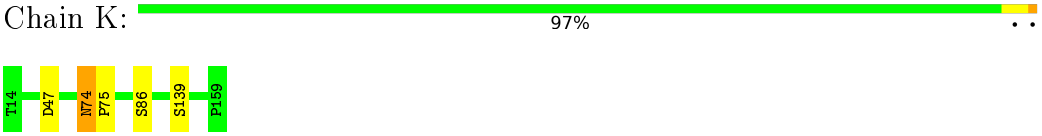
• Molecule 1: Portal protein

Chain X:



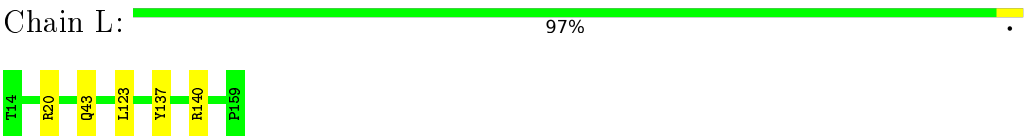
• Molecule 2: Peptidoglycan hydrolase gp4

Chain K:



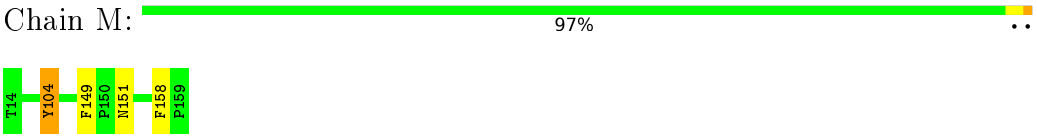
• Molecule 2: Peptidoglycan hydrolase gp4

Chain L:



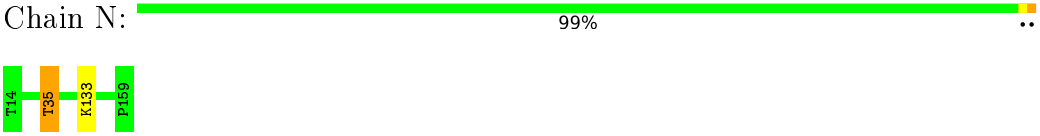
• Molecule 2: Peptidoglycan hydrolase gp4

Chain M:



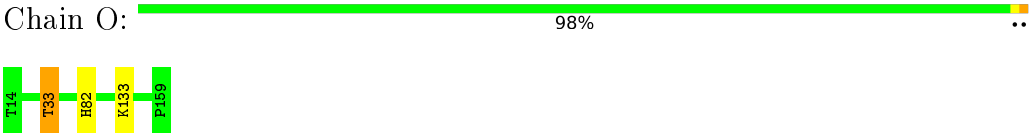
• Molecule 2: Peptidoglycan hydrolase gp4

Chain N:



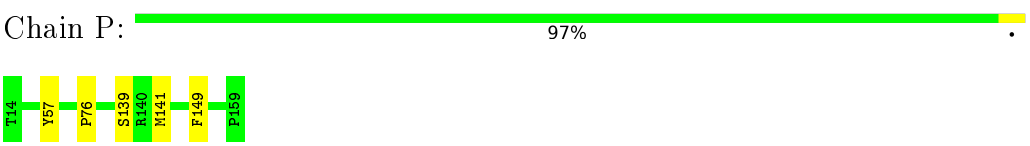
• Molecule 2: Peptidoglycan hydrolase gp4

Chain O:



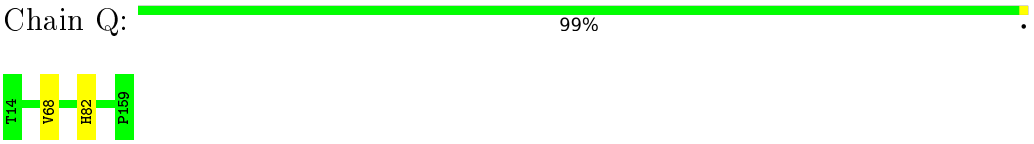
• Molecule 2: Peptidoglycan hydrolase gp4

Chain P:



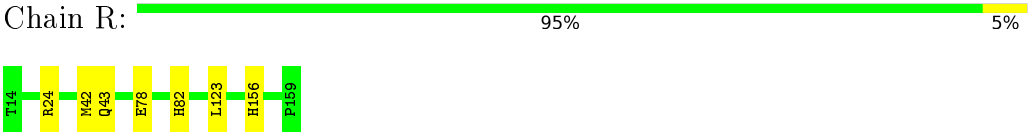
• Molecule 2: Peptidoglycan hydrolase gp4

Chain Q:



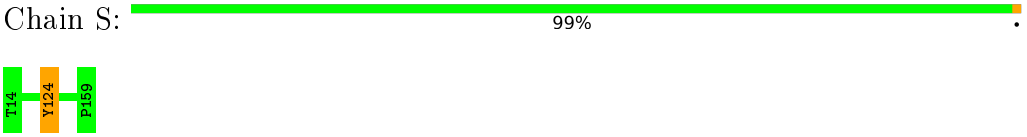
• Molecule 2: Peptidoglycan hydrolase gp4

Chain R:



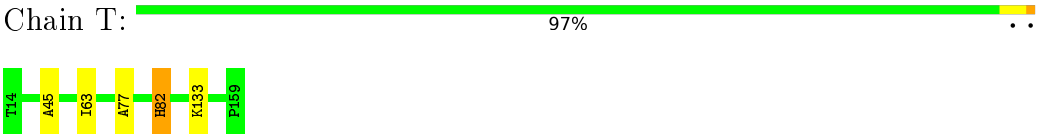
• Molecule 2: Peptidoglycan hydrolase gp4

Chain S:



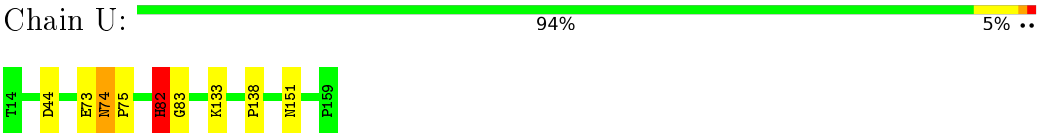
• Molecule 2: Peptidoglycan hydrolase gp4

Chain T:



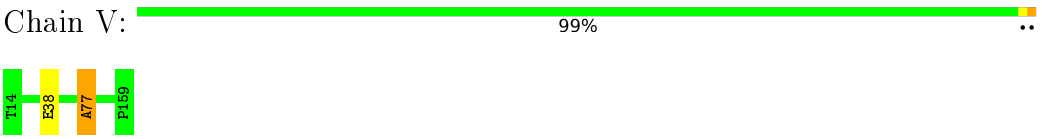
• Molecule 2: Peptidoglycan hydrolase gp4

Chain U:



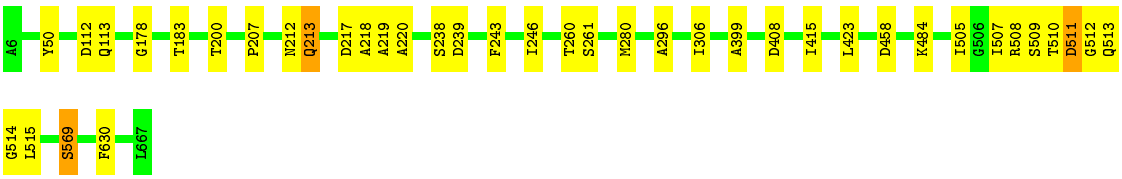
• Molecule 2: Peptidoglycan hydrolase gp4

Chain V:



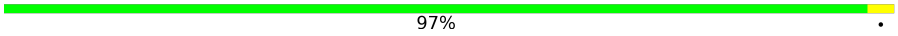
• Molecule 3: Tail fiber protein

Chain Y:



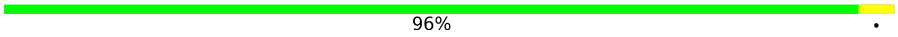
- Molecule 3: Tail fiber protein

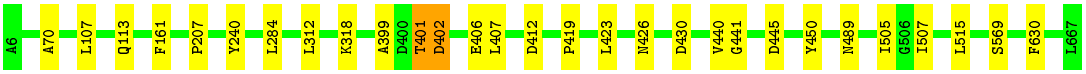
Chain Z:



- Molecule 3: Tail fiber protein

Chain 0:





4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	79731	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Not provided	Depositor
Microscope	JEOL 3200FSC	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	Not provided	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	4000	Depositor
Magnification	40000	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	0.94	0/5902	0.96	4/7997 (0.1%)
1	B	0.94	0/5902	0.96	9/7997 (0.1%)
1	C	0.94	0/5902	0.97	9/7997 (0.1%)
1	D	0.94	0/5902	0.94	2/7997 (0.0%)
1	E	0.94	0/5902	0.94	2/7997 (0.0%)
1	F	0.94	0/5902	0.95	3/7997 (0.0%)
1	G	0.94	0/5902	0.94	6/7997 (0.1%)
1	H	0.94	0/5902	0.94	2/7997 (0.0%)
1	I	0.94	0/5902	0.95	7/7997 (0.1%)
1	J	0.94	0/5902	0.95	6/7997 (0.1%)
1	W	0.94	0/5902	0.95	5/7997 (0.1%)
1	X	0.94	0/5902	0.95	5/7997 (0.1%)
2	K	0.92	0/1148	0.92	0/1558
2	L	0.92	0/1148	0.99	2/1558 (0.1%)
2	M	0.91	0/1148	0.97	4/1558 (0.3%)
2	N	0.92	0/1148	0.99	0/1558
2	O	0.91	0/1148	0.95	0/1558
2	P	0.92	0/1148	0.99	2/1558 (0.1%)
2	Q	0.91	0/1148	0.95	0/1558
2	R	0.92	0/1148	0.96	0/1558
2	S	0.92	0/1148	0.98	2/1558 (0.1%)
2	T	0.92	0/1148	0.98	0/1558
2	U	0.92	0/1148	1.01	0/1558
2	V	0.92	0/1148	0.99	1/1558 (0.1%)
3	0	0.92	0/5128	1.02	6/6966 (0.1%)
3	Y	0.93	0/5128	1.02	5/6966 (0.1%)
3	Z	0.92	0/5128	0.98	6/6966 (0.1%)
All	All	0.93	0/99984	0.96	88/135558 (0.1%)

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	6
1	B	0	1
1	C	0	5
1	D	0	5
1	E	0	5
1	F	0	3
1	G	0	4
1	H	0	5
1	J	0	8
1	W	0	4
1	X	0	2
2	K	0	1
2	U	0	1
2	V	0	1
3	0	0	9
3	Y	0	17
3	Z	0	1
All	All	0	78

There are no bond length outliers.

The worst 5 of 88 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	W	404	MET	CG-SD-CE	-7.64	87.97	100.20
1	W	556	TYR	CB-CG-CD2	-7.22	116.67	121.00
1	C	432	PHE	CB-CG-CD2	6.80	125.56	120.80
1	I	246	TYR	CB-CG-CD2	-6.80	116.92	121.00
1	I	246	TYR	CB-CG-CD1	6.71	125.03	121.00

There are no chirality outliers.

5 of 78 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	369	TYR	Sidechain
1	A	372	TYR	Sidechain
1	A	447	TYR	Sidechain
1	A	475	PRO	Peptide
1	A	476	ARG	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	5792	0	5623	1	0
1	B	5792	0	5623	1	0
1	C	5792	0	5623	5	0
1	D	5792	0	5623	4	0
1	E	5792	0	5623	4	0
1	F	5792	0	5623	2	0
1	G	5792	0	5623	3	0
1	H	5792	0	5623	1	0
1	I	5792	0	5623	1	0
1	J	5792	0	5623	3	0
1	W	5792	0	5623	3	0
1	X	5792	0	5623	0	0
2	K	1122	0	1082	1	0
2	L	1122	0	1082	1	0
2	M	1122	0	1082	0	0
2	N	1122	0	1082	0	0
2	O	1122	0	1082	0	0
2	P	1122	0	1082	0	0
2	Q	1122	0	1082	0	0
2	R	1122	0	1082	1	0
2	S	1122	0	1082	0	0
2	T	1122	0	1082	2	0
2	U	1122	0	1082	2	0
2	V	1122	0	1082	0	0
3	0	5025	0	4930	2	0
3	Y	5025	0	4930	2	0
3	Z	5025	0	4930	3	0
All	All	98043	0	95250	39	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 0.

The worst 5 of 39 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:L:20:ARG:HH22	3:Z:25:ALA:H	1.55	0.54
2:R:24:ARG:HH12	2:R:82:HIS:CE1	2.28	0.52
2:T:82:HIS:H	2:T:82:HIS:CD2	2.28	0.52
1:W:59:VAL:HG12	1:W:327:GLN:HE21	1.74	0.50
3:Y:260:THR:HG22	3:Y:261:SER:H	1.76	0.50

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	719/721 (100%)	646 (90%)	53 (7%)	20 (3%)	6	44
1	B	719/721 (100%)	639 (89%)	60 (8%)	20 (3%)	6	44
1	C	719/721 (100%)	630 (88%)	67 (9%)	22 (3%)	5	42
1	D	719/721 (100%)	633 (88%)	67 (9%)	19 (3%)	7	45
1	E	719/721 (100%)	648 (90%)	51 (7%)	20 (3%)	6	44
1	F	719/721 (100%)	643 (89%)	59 (8%)	17 (2%)	7	47
1	G	719/721 (100%)	642 (89%)	54 (8%)	23 (3%)	5	41
1	H	719/721 (100%)	645 (90%)	56 (8%)	18 (2%)	7	46
1	I	719/721 (100%)	641 (89%)	60 (8%)	18 (2%)	7	46
1	J	719/721 (100%)	637 (89%)	62 (9%)	20 (3%)	6	44
1	W	719/721 (100%)	644 (90%)	52 (7%)	23 (3%)	5	41
1	X	719/721 (100%)	646 (90%)	56 (8%)	17 (2%)	7	47
2	K	144/146 (99%)	126 (88%)	16 (11%)	2 (1%)	14	58
2	L	144/146 (99%)	122 (85%)	20 (14%)	2 (1%)	14	58
2	M	144/146 (99%)	125 (87%)	17 (12%)	2 (1%)	14	58
2	N	144/146 (99%)	125 (87%)	17 (12%)	2 (1%)	14	58

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	O	144/146 (99%)	123 (85%)	19 (13%)	2 (1%)	14 58
2	P	144/146 (99%)	123 (85%)	17 (12%)	4 (3%)	6 44
2	Q	144/146 (99%)	131 (91%)	12 (8%)	1 (1%)	26 71
2	R	144/146 (99%)	129 (90%)	11 (8%)	4 (3%)	6 44
2	S	144/146 (99%)	130 (90%)	14 (10%)	0	100 100
2	T	144/146 (99%)	130 (90%)	11 (8%)	3 (2%)	9 50
2	U	144/146 (99%)	124 (86%)	14 (10%)	6 (4%)	3 34
2	V	144/146 (99%)	127 (88%)	16 (11%)	1 (1%)	26 71
3	0	660/662 (100%)	603 (91%)	44 (7%)	13 (2%)	9 51
3	Y	660/662 (100%)	606 (92%)	36 (6%)	18 (3%)	6 45
3	Z	660/662 (100%)	621 (94%)	32 (5%)	7 (1%)	17 63
All	All	12336/12390 (100%)	11039 (90%)	993 (8%)	304 (2%)	11 46

5 of 304 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	170	ALA
1	A	267	ALA
1	A	476	ARG
1	A	477	ASN
1	A	501	THR

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	627/627 (100%)	622 (99%)	5 (1%)	86 94
1	B	627/627 (100%)	622 (99%)	5 (1%)	86 94
1	C	627/627 (100%)	620 (99%)	7 (1%)	80 91
1	D	627/627 (100%)	621 (99%)	6 (1%)	82 92
1	E	627/627 (100%)	622 (99%)	5 (1%)	86 94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	F	627/627 (100%)	624 (100%)	3 (0%)	92 96
1	G	627/627 (100%)	620 (99%)	7 (1%)	80 91
1	H	627/627 (100%)	622 (99%)	5 (1%)	86 94
1	I	627/627 (100%)	622 (99%)	5 (1%)	86 94
1	J	627/627 (100%)	620 (99%)	7 (1%)	80 91
1	W	627/627 (100%)	624 (100%)	3 (0%)	92 96
1	X	627/627 (100%)	623 (99%)	4 (1%)	90 95
2	K	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	L	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	M	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	N	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	O	116/116 (100%)	114 (98%)	2 (2%)	68 87
2	P	116/116 (100%)	116 (100%)	0	100 100
2	Q	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	R	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	S	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	T	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	U	116/116 (100%)	115 (99%)	1 (1%)	84 93
2	V	116/116 (100%)	116 (100%)	0	100 100
3	0	543/543 (100%)	540 (99%)	3 (1%)	90 95
3	Y	543/543 (100%)	542 (100%)	1 (0%)	95 97
3	Z	543/543 (100%)	542 (100%)	1 (0%)	95 97
All	All	10545/10545 (100%)	10467 (99%)	78 (1%)	89 94

5 of 78 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	281	ILE
1	I	150	HIS
1	X	360	GLU
1	G	360	GLU
1	H	316	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 70 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	150	HIS
1	I	105	ASN
3	Z	212	ASN
1	G	172	HIS
1	H	135	GLN

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.