



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Aug 17, 2017 – 04:29 PM EDT

PDB ID : 5A21
EMDB ID: : EMD-2994
Title : Structure of bacteriophage SPP1 head-to-tail interface without DNA and tape measure protein
Authors : Chaban, Y.; Lurz, R.; Brasiles, S.; Cornilleau, C.; Karreman, M.; Zinn-Justin, S.; Tavares, P.; Orlova, E.V.
Deposited on : unknown
Resolution : 7.20 Å(reported)

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

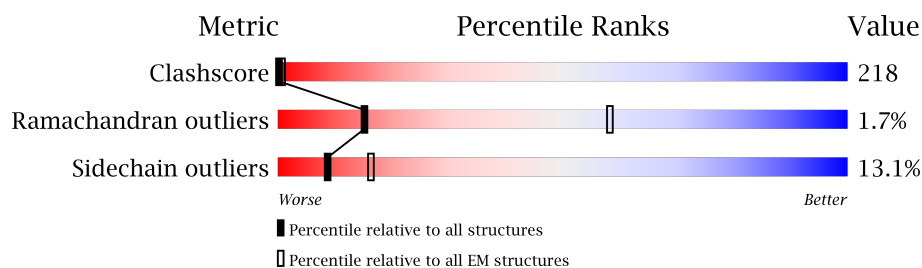
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 7.20 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	503	14% 65% 9% 13%
1	B	503	13% 66% 8% 13%
2	C	102	11% 73% 14% .
2	D	102	12% 69% 17% .
3	E	109	6% 72% 20% .
3	F	109	7% 68% 23% .
4	G	134	13% 70% 14% ..
5	H	177	16% 58% 15% . 9%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 12789 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	439	Total	C	N	O	S	0	0
			3572	2259	588	711	14		
1	B	439	Total	C	N	O	S	0	0
			3572	2259	588	711	14		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	365	LYS	ASN	conflict	UNP P54309
B	365	LYS	ASN	conflict	UNP P54309

- Molecule 2 is a protein called 15 PROTEIN.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	99	Total	C	N	O	S	0	0
			790	502	134	149	5		
2	D	99	Total	C	N	O	S	0	0
			790	502	134	149	5		

- Molecule 3 is a protein called HEAD COMPLETION PROTEIN GP16.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	E	109	Total	C	N	O	S	0	0
			887	567	145	173	2		
3	F	109	Total	C	N	O	S	0	0
			887	567	145	173	2		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	6	ARG	PRO	conflict	UNP O48446
F	6	ARG	PRO	conflict	UNP O48446

- Molecule 4 is a protein called TAIL-TO-HEAD JOINING PROTEIN GP17.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	G	133	Total	C	N	O	S	0	0
			1047	667	172	204	4		

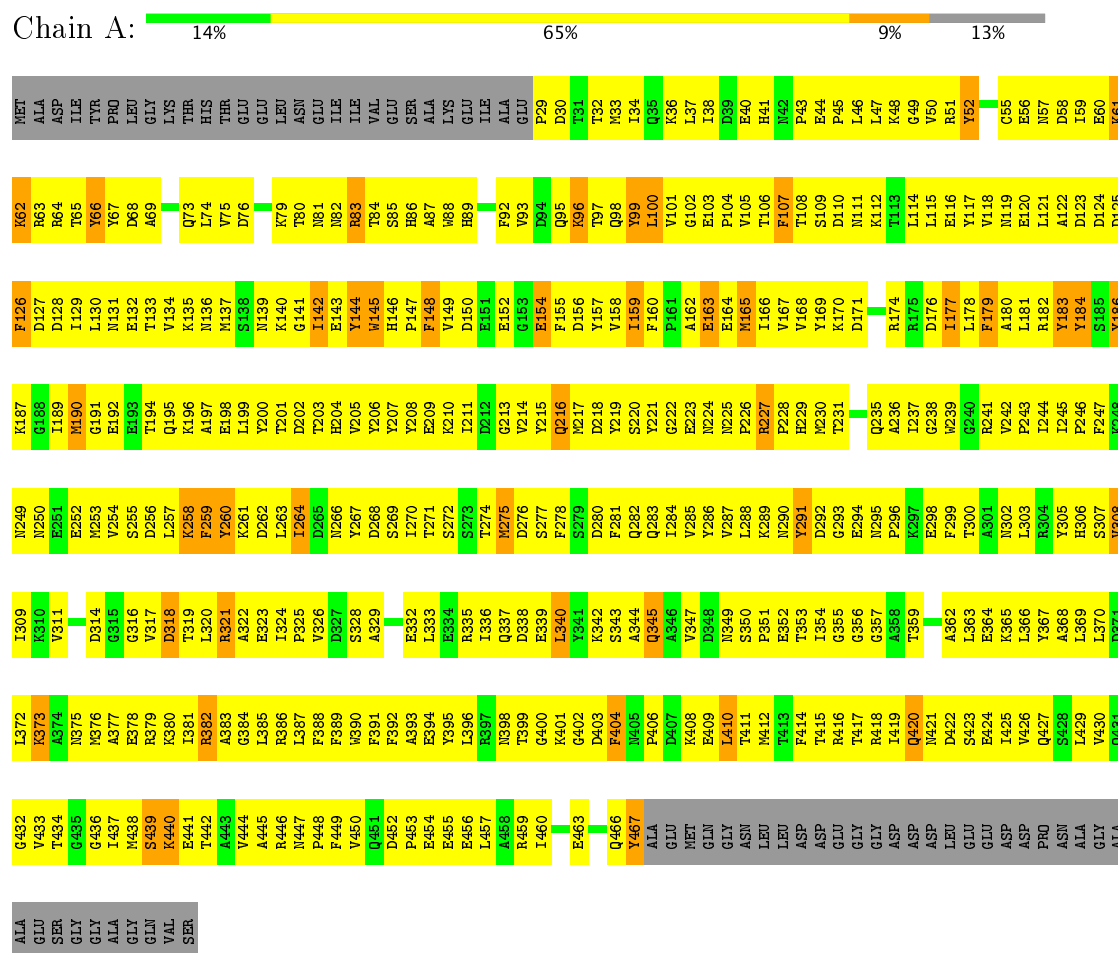
- Molecule 5 is a protein called MAJOR TAIL PROTEIN 17.1.

Mol	Chain	Residues	Atoms				AltConf	Trace
5	H	161	Total	C	N	O	0	0
			1244	776	205	263		

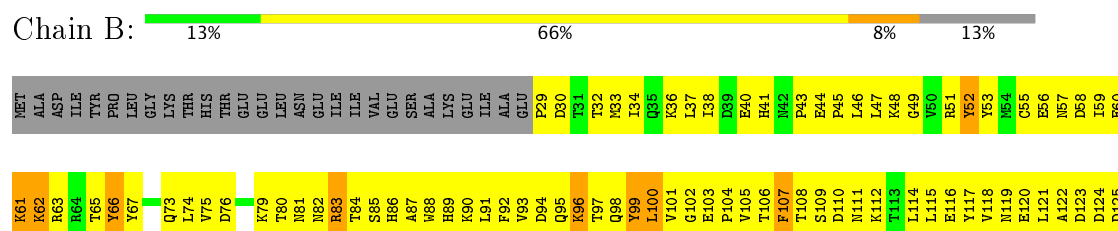
3 Residue-property plots

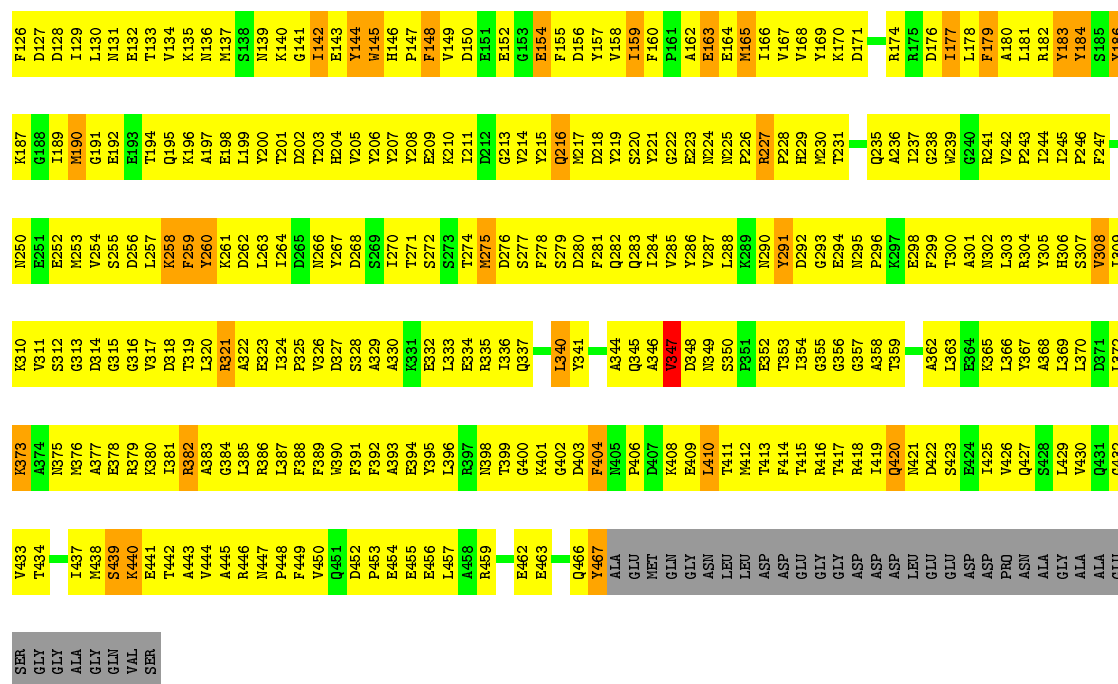
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according 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: PORTAL PROTEIN



• Molecule 1: PORTAL PROTEIN





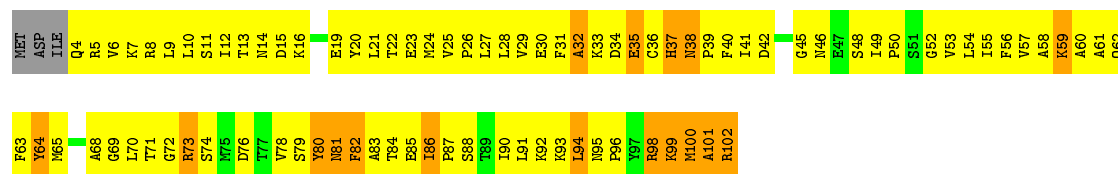
• Molecule 2: 15 PROTEIN

Chain C: 11% 73% 14%



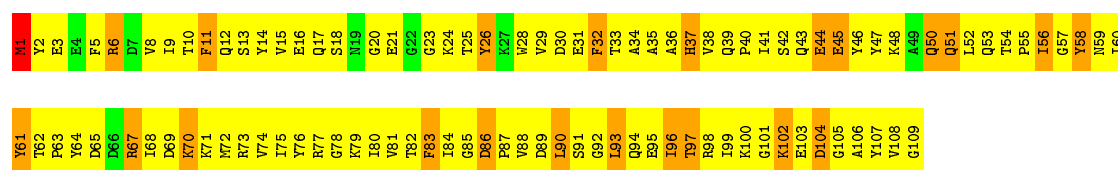
• Molecule 2: 15 PROTEIN

Chain D: 12% 69% 17%

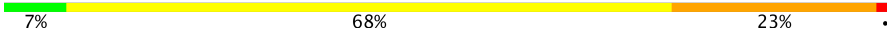


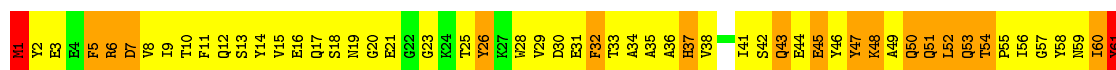
• Molecule 3: HEAD COMPLETION PROTEIN GP16

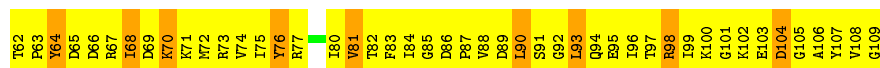
Chain E: 6% 72% 20%



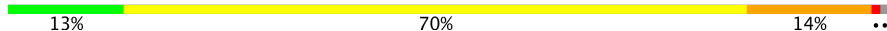
• Molecule 3: HEAD COMPLETION PROTEIN GP16

Chain F:  7% 68% 23% .

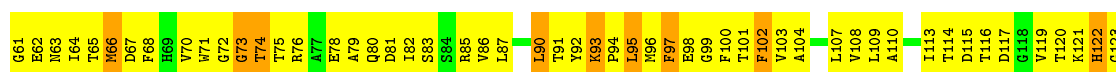




• Molecule 4: TAIL-TO-HEAD JOINING PROTEIN GP17


Chain G:  13% 70% 14% ..

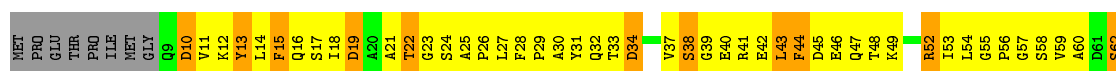






• Molecule 5: MAJOR TAIL PROTEIN 17.1

Chain H:  16% 58% 15% . 9%







4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C6	Depositor
Number of particles used	18000	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	EACH PARTICLE	Depositor
Microscope	FEI TECNAI F30	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	20	Depositor
Minimum defocus (nm)	900	Depositor
Maximum defocus (nm)	3600	Depositor
Magnification	39000	Depositor
Image detector	KODAK SO-163 FILM	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	0.99	0/3650	1.08	0/4939
1	B	0.99	0/3650	1.08	0/4939
2	C	0.99	0/805	1.08	0/1085
2	D	0.99	0/805	1.08	0/1085
3	E	1.13	1/907 (0.1%)	1.37	2/1224 (0.2%)
3	F	1.03	0/907	1.25	4/1224 (0.3%)
4	G	0.97	0/1069	1.14	0/1451
5	H	0.97	0/1266	1.34	11/1709 (0.6%)
All	All	1.00	1/13059 (0.0%)	1.15	17/17656 (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	2
1	B	0	2
2	C	0	3
2	D	0	3
3	E	0	3
3	F	0	1
4	G	0	3
5	H	0	5
All	All	0	22

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	1	MET	N-CA	-13.33	1.19	1.46

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
3	E	1	MET	N-CA-CB	-21.00	72.80	110.60
3	E	1	MET	N-CA-C	-14.94	70.66	111.00
5	H	92	TRP	CA-CB-CG	13.92	140.15	113.70
3	F	61	TYR	CB-CG-CD1	8.99	126.39	121.00
5	H	155	LYS	C-N-CA	8.68	140.53	122.30

There are no chirality outliers.

5 of 22 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	439	SER	Peptide
1	A	61	LYS	Peptide
1	B	439	SER	Peptide
1	B	61	LYS	Peptide
2	C	32	ALA	Peptide

5.2 Too-close contacts

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	3572	0	3417	1587	0
1	B	3572	0	3418	1681	0
2	C	790	0	792	647	0
2	D	790	0	791	607	0
3	E	887	0	853	703	0
3	F	887	0	855	650	0
4	G	1047	0	1028	553	0
5	H	1244	0	1162	676	0
All	All	12789	0	12316	5484	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 218.

The worst 5 of 5484 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:336:ILE:CG1	1:B:330:ALA:HB1	1.20	1.61

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:F:108:VAL:HA	4:G:37:LYS:CB	1.13	1.60
2:C:82:PHE:CD1	2:D:84:THR:HG22	1.36	1.59
1:A:336:ILE:HG12	1:B:330:ALA:CB	1.30	1.59
2:C:54:LEU:HG	2:D:28:LEU:CD2	1.32	1.59

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	437/503 (87%)	429 (98%)	6 (1%)	2 (0%)	32	74
1	B	437/503 (87%)	429 (98%)	6 (1%)	2 (0%)	32	74
2	C	97/102 (95%)	93 (96%)	2 (2%)	2 (2%)	8	45
2	D	97/102 (95%)	93 (96%)	2 (2%)	2 (2%)	8	45
3	E	107/109 (98%)	98 (92%)	8 (8%)	1 (1%)	20	63
3	F	107/109 (98%)	94 (88%)	8 (8%)	5 (5%)	3	28
4	G	131/134 (98%)	121 (92%)	4 (3%)	6 (5%)	3	28
5	H	159/177 (90%)	145 (91%)	8 (5%)	6 (4%)	4	32
All	All	1572/1739 (90%)	1502 (96%)	44 (3%)	26 (2%)	15	50

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	308	VAL
1	B	308	VAL
1	B	347	VAL
2	C	98	ARG
2	C	101	ALA

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	388/436 (89%)	346 (89%)	42 (11%)	7	31
1	B	388/436 (89%)	348 (90%)	40 (10%)	8	32
2	C	88/91 (97%)	76 (86%)	12 (14%)	4	23
2	D	88/91 (97%)	76 (86%)	12 (14%)	4	23
3	E	93/94 (99%)	72 (77%)	21 (23%)	1	7
3	F	93/94 (99%)	71 (76%)	22 (24%)	1	5
4	G	117/118 (99%)	102 (87%)	15 (13%)	5	25
5	H	130/142 (92%)	112 (86%)	18 (14%)	4	23
All	All	1385/1502 (92%)	1203 (87%)	182 (13%)	9	24

5 of 182 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	35	GLU
2	D	82	PHE
5	H	44	PHE
2	C	59	LYS
2	C	100	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 60 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	131	ASN
1	B	283	GLN
5	H	85	ASN
1	B	216	GLN
1	B	302	ASN

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 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.