



## wwPDB EM Map/Model Validation Report ⓘ

Apr 10, 2016 – 02:10 PM BST

PDB ID : 5A20  
EMDB ID: : EMD-2993  
Title : Structure of bacteriophage SPP1 head-to-tail interface filled with 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 : 2015-05-06  
Resolution : 7.60 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](mailto: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

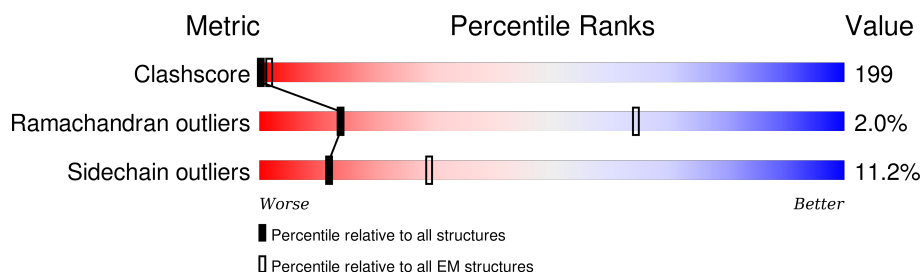
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 7.60 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  $\geq 3$ , 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	17% 62% 8% • 13%
1	B	503	17% 62% 8% • 13%
2	C	102	10% 75% 12% •
2	D	102	10% 74% 14% •
3	E	109	8% 78% 12% •
3	F	109	7% 73% 19%
4	G	134	15% 69% 14% ••
5	H	177	19% 55% 15% • 9%

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 12799 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	440	Total	C	N	O	S	0	1
			3573	2259	589	711	14		
1	B	440	Total	C	N	O	S	0	1
			3573	2259	589	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
			891	569	145	175	2		
3	F	109	Total	C	N	O	S	0	0
			891	569	145	175	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		

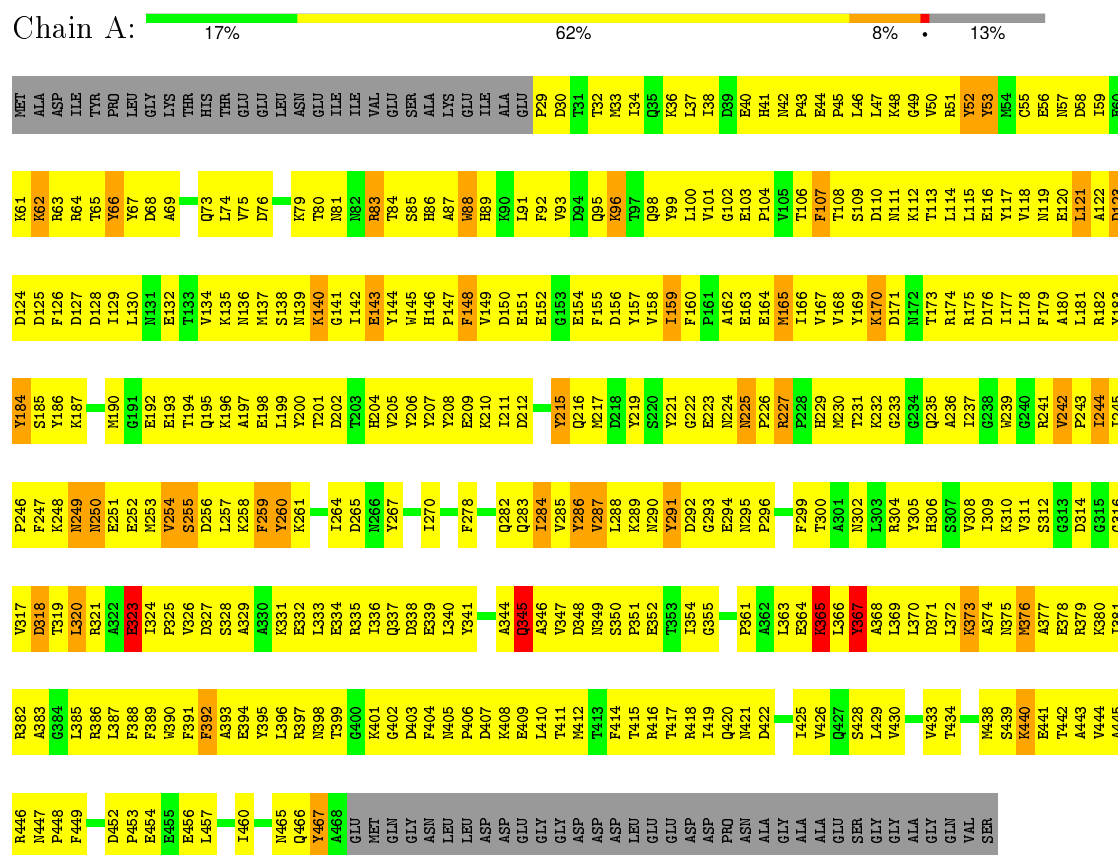
- Molecule 5 is a protein called MAJOR TAIL PROTEIN GP17.1.

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 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.

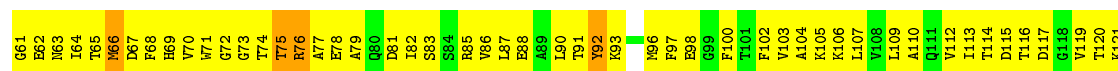
#### • Molecule 1: PORTAL PROTEIN







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



## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	Not provided	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	EACH PARTICLE, Not provided	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	0/3651	1.07	0/4941
1	B	0.99	0/3651	1.07	0/4941
2	C	0.99	0/805	1.09	0/1085
2	D	0.99	0/805	1.09	0/1085
3	E	1.02	0/911	1.14	0/1229
3	F	1.02	0/911	1.11	0/1229
4	G	0.97	0/1069	1.11	1/1451 (0.1%)
5	H	1.21	2/1264 (0.2%)	1.36	3/1703 (0.2%)
All	All	1.01	2/13067 (0.0%)	1.11	4/17664 (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	6
1	B	0	6
2	C	0	4
2	D	0	4
3	E	0	2
3	F	0	2
4	G	0	3
5	H	0	5
All	All	0	32

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	H	141	GLY	C-N	18.96	1.77	1.34
5	H	53	ILE	C-N	16.64	1.72	1.34

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	H	141	GLY	CA-C-N	-22.88	66.86	117.20
5	H	53	ILE	O-C-N	-19.53	91.46	122.70
5	H	141	GLY	C-N-CA	-11.11	93.92	121.70
4	G	44	VAL	CB-CA-C	-5.01	101.89	111.40

There are no chirality outliers.

5 of 32 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	123	ASP	Peptide
1	A	255	SER	Peptide
1	A	282	GLN	Peptide
1	A	323	GLU	Peptide
1	A	365	LYS	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	3573	0	3416	1326	0
1	B	3573	0	3414	1396	0
2	C	790	0	795	628	0
2	D	790	0	794	554	0
3	E	891	0	855	606	0
3	F	891	0	858	600	0
4	G	1047	0	1027	550	0
5	H	1244	0	1160	571	0
All	All	12799	0	12319	5006	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 199.

The worst 5 of 5006 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:438:MET:SD	1:B:434:THR:HA	1.26	1.74

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:99:LYS:CA	2:D:101:ALA:HA	1.22	1.69
4:G:100:PHE:CE1	5:H:45:ASP:HB3	1.33	1.62
1:A:259:PHE:CZ	1:B:90:LYS:HD3	1.32	1.61
4:G:59:SER:CB	5:H:138:LEU:HD12	1.11	1.58

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	438/503 (87%)	426 (97%)	7 (2%)	5 (1%)	17	63
1	B	438/503 (87%)	426 (97%)	7 (2%)	5 (1%)	17	63
2	C	97/102 (95%)	90 (93%)	4 (4%)	3 (3%)	5	42
2	D	97/102 (95%)	90 (93%)	4 (4%)	3 (3%)	5	42
3	E	107/109 (98%)	98 (92%)	7 (6%)	2 (2%)	10	52
3	F	107/109 (98%)	98 (92%)	7 (6%)	2 (2%)	10	52
4	G	131/134 (98%)	122 (93%)	6 (5%)	3 (2%)	8	48
5	H	155/177 (88%)	140 (90%)	7 (4%)	8 (5%)	2	30
All	All	1570/1739 (90%)	1490 (95%)	49 (3%)	31 (2%)	14	51

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	284	ILE
1	A	287	VAL
1	B	284	ILE
1	B	287	VAL
2	C	95	ASN

### 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	388/436 (89%)	349 (90%)	39 (10%)	9	38
1	B	388/436 (89%)	349 (90%)	39 (10%)	9	38
2	C	88/91 (97%)	81 (92%)	7 (8%)	15	50
2	D	88/91 (97%)	81 (92%)	7 (8%)	15	50
3	E	94/94 (100%)	81 (86%)	13 (14%)	4	27
3	F	94/94 (100%)	77 (82%)	17 (18%)	2	15
4	G	117/118 (99%)	102 (87%)	15 (13%)	5	29
5	H	130/142 (92%)	111 (85%)	19 (15%)	4	24
All	All	1387/1502 (92%)	1231 (89%)	156 (11%)	12	33

5 of 156 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	365	LYS
2	D	80	TYR
5	H	43	LEU
1	B	373	LYS
2	C	64	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 53 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	282	GLN
1	B	420	GLN
4	G	49	GLN
1	B	302	ASN
1	B	345	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.