



## wwPDB EM Map/Model Validation Report i

Apr 10, 2016 – 01:56 PM BST

PDB ID : 4BTG  
EMDB ID: : EMD-2364  
Title : Coordinates of the bacteriophage phi6 capsid subunits (P1A and P1B) fitted into the cryoEM reconstruction of the procapsid at 4.4 Å resolution  
Authors : Nemecek, D.; Boura, E.; Wu, W.; Cheng, N.; Plevka, P.; Qiao, J.; Mindich, L.; Heymann, J.B.; Hurley, J.H.; Steven, A.C.  
Deposited on : 2013-06-17  
Resolution : 4.40 Å (reported)  
Based on PDB ID : 4K7H

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>

---

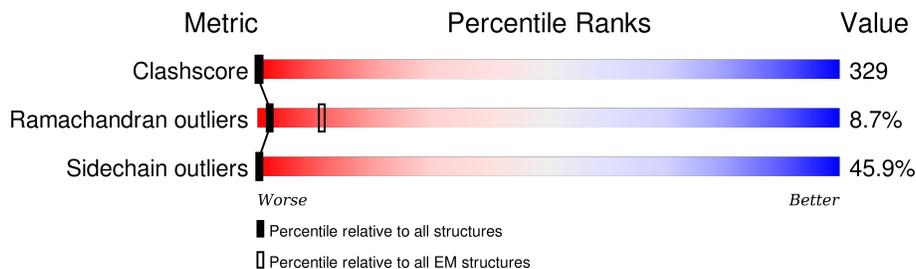
MolProbity : 4.02b-467  
Mogul : unknown  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk27241

# 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 4.40 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	761	5% 40% 41% 14%
1	B	761	. 48% 38% 10%

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 11840 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 MAJOR INNER PROTEIN P1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	761	Total	C	N	O	S	0	0
			5920	3741	1048	1109	22		
1	B	761	Total	C	N	O	S	0	0
			5920	3741	1048	1109	22		

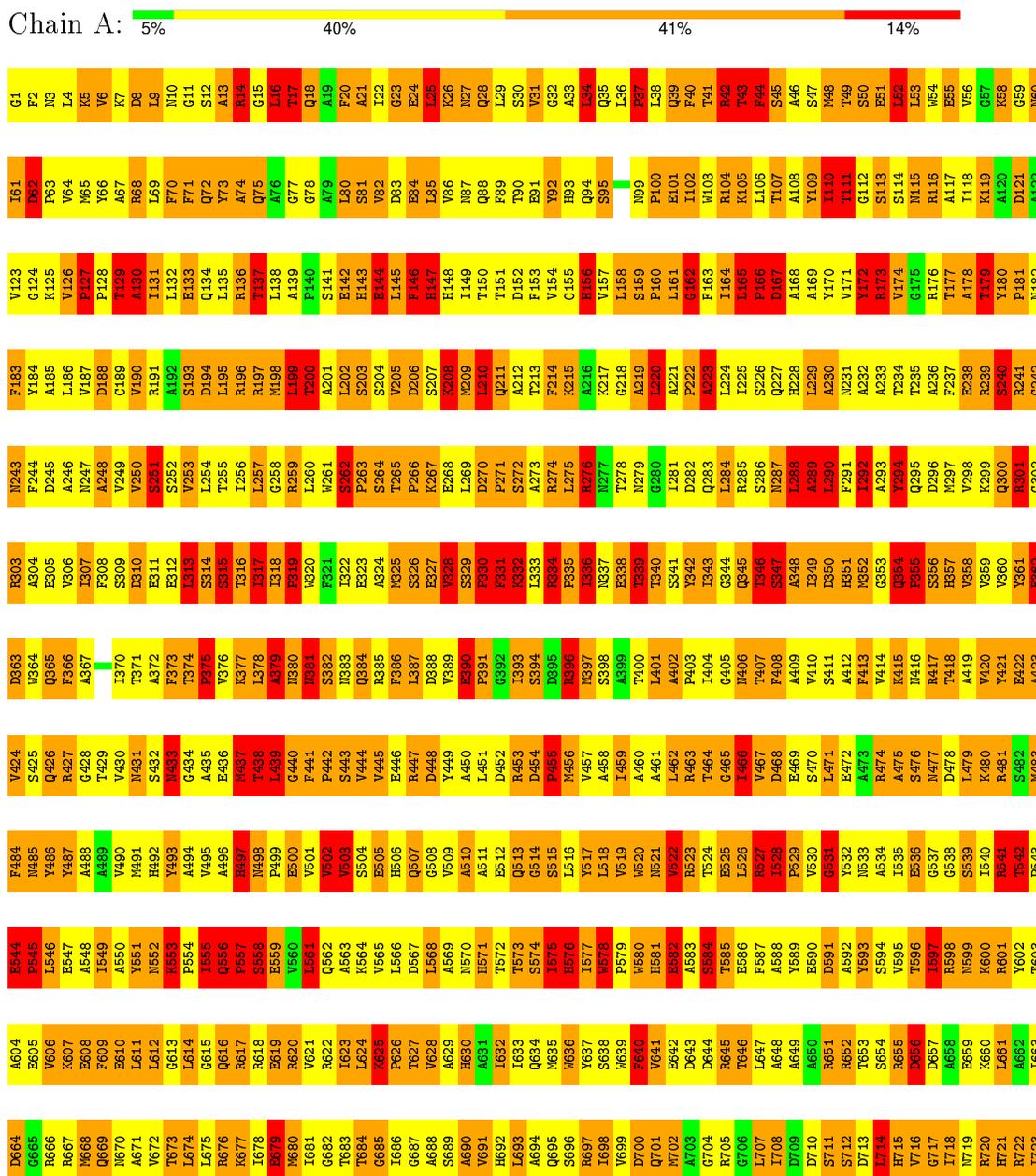
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP P11126
B	1	GLY	-	EXPRESSION TAG	UNP P11126

### 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: MAJOR INNER PROTEIN P1



R724	R725	R726	R727	R728	R729	R730	R731	R732	R733	R734	R735	R736	R737	R738	R739	R740	R741	R742	R743	R744	R745	R746	R747	R748	R749	R750	R751	R752	R753	R754	R755	R756	R757	R758	R759	R760	R761
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

• Molecule 1: MAJOR INNER PROTEIN P1



G1	F2	M3	L4	K5	V6	K7	D8	L9	N10	G11	S12	A13	R14	G15	L16	T17	Q18	A19	F20	A21	E22	D23	E24	L25	K26	N27	Q28	L29	S30	V31	G32	A33	L34	Q35	S36	P37	L38	Q39	F40	T41	E42	I43	R44	F45	A46	S47	N48	T49	S50	E51	L52	L53	W54	S55	E56	G57	R58	G59	N60	
I61	D62	P63	V64	M65	V66	A67	R68	L69	F70	F71	Q72	Y73	A74	G75	Q76	G77	G78	A79	L80	A81	V82	D83	E84	L85	V86	N87	Q88	L89	T90	E91	Y92	Q93	C94	T95	L96	A97	C98	R99	P100	E101	I102	M103	R104	S105	A106	T107	A108	Y109	I110	T111	G112	L113	S114	W115	E116	R117	A118	T119	G120	
D121	A122	G123	V124	M125	V126	P127	F128	L129	A130	I131	Q132	E133	Q134	L135	L136	T137	L138	A139	F140	A141	V142	E143	D144	L145	V146	H147	Q148	L149	T150	T151	D152	F153	V154	C155	H156	A157	L158	S159	P160	L161	G162	P163	L164	L165	P166	L167	A168	A169	Y170	V171	Y172	G173	S174	W175	E176	R177	T178	T179	R239	R240
P181	N182	F183	Y184	A185	L186	A187	D188	C189	V190	A191	S192	S193	D194	L195	A196	L197	Q198	M199	T200	A201	V202	L203	S204	V205	D206	S207	K208	M209	L210	Q211	A212	T213	F214	L215	L216	K217	G218	A219	L220	A221	P222	A223	L224	L225	S226	Q227	D228	A229	L230	F231	M232	A233	Q234	T235	A236	F237	E238	R239	S240	
R241	G242	N243	F244	D245	A246	N247	A248	V249	W250	S251	S252	V253	L254	T255	L256	L257	Q258	A259	L260	M261	S262	P263	S264	T265	P266	K267	E268	L269	D270	P271	S272	R273	R274	L275	R276	N277	T278	N279	G280	I281	D282	D283	L284	R285	S286	L287	A288	A289	L290	F291	I292	A293	Q294	W295	D296	K297	V298	R299	Q300	
R301	G302	R303	W304	E305	V306	L307	F308	S309	A310	E311	S312	L313	S314	L315	S316	I317	Q318	P319	M320	F321	I322	E323	A324	M325	S326	E327	V328	S329	P330	F331	K332	L333	R334	L335	I336	N337	E338	L339	T340	S341	Y342	I343	G344	Q345	T346	S347	F408	A409	V410	S411	A412	F413	W414	Q415	D416	R417	T418	V419	S420	
Y361	E362	R363	W364	D365	F366	A367	K368	E369	L370	T371	S372	R373	S374	P375	W376	L378	A379	N380	R381	I382	S383	Q384	R385	F386	L387	D388	V389	E390	P391	S392	L393	S394	R395	R396	N397	S398	A399	T400	L401	A402	P403	L404	G405	M406	T407	F408	A409	V410	S411	A412	F413	W414	Q415	D416	R417	T418	V419	S420		
Y421	E422	A423	W424	S425	Q426	E427	G428	R429	M430	N431	S432	M433	G434	A435	W436	M437	L438	L439	G440	F441	P442	S443	V444	V445	E446	R447	D448	Y449	A450	H451	D452	R453	D454	P455	M456	V457	A458	L459	A460	A461	L462	R463	T464	G465	I466	V467	D468	E469	S470	L471	E472	A473	R474	A475	S476	M477	D478	L479	K480	
R481	S482	M483	F484	N485	Y486	Y487	A488	A489	W490	M491	H492	Y493	A494	V495	A496	H497	N498	P499	E500	V501	S502	V503	S504	E505	H506	Q507	S508	V509	A510	A511	E512	Q513	E514	S515	L516	Y517	L518	V519	W520	N521	V522	S523	T524	E525	L526	R527	S528	I529	P529	V530	G531	Y532	N533	S534	A535	E536	G537	G538	S539	L540
R541	T542	P543	E544	P545	L546	E547	A548	L549	A550	Y551	K552	K553	L554	S555	Q556	P557	S558	E559	V560	L561	Q562	A563	K564	V565	L566	D567	L568	A569	N570	H571	T572	T573	S574	L575	H576	I577	H578	P579	W580	R581	E582	A583	S584	S585	E586	F587	A588	Y589	E590	D591	A592	Y593	S594	W595	L596	I597	R598	N599	K600	
R601	V602	A603	A604	E605	V606	R607	E608	F609	N610	L611	L612	G613	L614	G615	Q616	R617	R618	E619	R620	V621	R622	L623	L624	V625	P626	T627	V628	A629	H630	A631	L632	L633	Q634	Q635	H636	V637	S638	P639	F640	V641	E642	D643	G644	L645	T646	L647	A648	A649	A650	R651	R652	T653	L654	H655	L656	V657	G658	L659	R660	
L661	A662	I663	D664	G665	R666	R667	M668	Q669	N670	A671	V672	T673	L674	L675	R676	K677	I678	E679	M680	V681	L682	G683	T684	G685	L686	G687	A688	S689	A690	V691	H692	L693	A694	Q695	S696	R697	L698	V699	D700	Q701	E702	A703	G704	R705	L706	I707	T708	D709	D710	S711	S712	D713	L714	H715	V716	G717	L718	N719	E720	
R721	R722	R723	R724	R725	R726	R727	R728	R729	R730	R731	R732	R733	R734	R735	R736	R737	R738	R739	R740	R741	R742	R743	R744	R745	R746	R747	R748	R749	R750	R751	R752	R753	R754	R755	R756	R757	R758	R759	R760	R761																				

## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of images	Not provided	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	PARTICLES FROM EACH MICRO- GRAPH	Depositor
Microscope	OTHER	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	15	Depositor
Minimum defocus (nm)	800	Depositor
Maximum defocus (nm)	2200	Depositor
Magnification	46000	Depositor
Image detector	KODAK SO163 FILM	Depositor

## 5 Model quality i

### 5.1 Standard geometry i

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	1.22	32/6038 (0.5%)	1.56	144/8200 (1.8%)
1	B	0.96	27/6039 (0.4%)	1.41	88/8203 (1.1%)
All	All	1.10	59/12077 (0.5%)	1.49	232/16403 (1.4%)

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	67
1	B	0	45
All	All	0	112

The worst 5 of 59 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	466	ILE	C-N	-36.28	0.50	1.34
1	B	127	PRO	C-N	30.25	1.91	1.34
1	A	437	MET	C-N	11.92	1.61	1.34
1	A	75	GLN	CA-CB	-11.01	1.29	1.53
1	B	124	GLY	C-N	9.05	1.54	1.34

The worst 5 of 232 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	441	PHE	C-N-CD	-46.69	17.89	120.60
1	A	542	THR	C-N-CD	15.13	160.17	128.40
1	A	290	LEU	CB-CG-CD1	-13.07	88.79	111.00
1	A	16	LEU	CB-CG-CD1	-12.81	89.22	111.00
1	A	739	SER	O-C-N	-11.26	104.68	122.70

There are no chirality outliers.

5 of 112 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	23	GLY	Peptide
1	A	25	LEU	Peptide
1	A	34	LEU	Peptide
1	A	52	LEU	Peptide
1	A	60	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	5920	0	5870	3751	0
1	B	5920	0	5883	4050	0
All	All	11840	0	11753	7764	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 329.

The worst 5 of 7764 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:61:ILE:HG22	1:A:156:HIS:CD2	1.18	1.68
1:A:732:LEU:CD1	1:A:738:LEU:HD22	1.23	1.67
1:B:373:PHE:CD2	1:B:583:ALA:HB1	1.31	1.66
1:B:670:ASN:HD21	1:B:749:VAL:CG1	1.07	1.65
1:B:156:HIS:CD2	1:B:158:LEU:H	1.14	1.65

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	755/761 (99%)	575 (76%)	109 (14%)	71 (9%)	1	16
1	B	757/761 (100%)	591 (78%)	105 (14%)	61 (8%)	1	18
All	All	1512/1522 (99%)	1166 (77%)	214 (14%)	132 (9%)	2	17

5 of 132 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	34	LEU
1	A	61	ILE
1	A	119	LYS
1	A	123	VAL
1	A	144	GLU

### 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	629/629 (100%)	341 (54%)	288 (46%)	0	0
1	B	629/629 (100%)	340 (54%)	289 (46%)	0	0
All	All	1258/1258 (100%)	681 (54%)	577 (46%)	0	0

5 of 577 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	698	ILE
1	B	91	GLU
1	B	642	GLU
1	A	714	LEU
1	B	26	LYS

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	72	GLN
1	B	182	ASN
1	B	581	HIS
1	B	75	GLN
1	B	88	GLN

### 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.