



## wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Aug 17, 2017 – 11:42 AM EDT

PDB ID : 4BX4  
EMDB ID: : EMD-1300  
Title : Fitting of the bacteriophage Phi8 P1 capsid protein into cryo-EM density  
Authors : El Omari, K.; Sutton, G.; Ravantti, J.J.; Zhang, H.; Walter, T.S.; Grimes, J.M.; Bamford, D.H.; Stuart, D.I.; Mancini, E.J.  
Deposited on : unknown  
Resolution : 8.70 Å(reported)  
Based on PDB ID : 4BTP

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](mailto: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

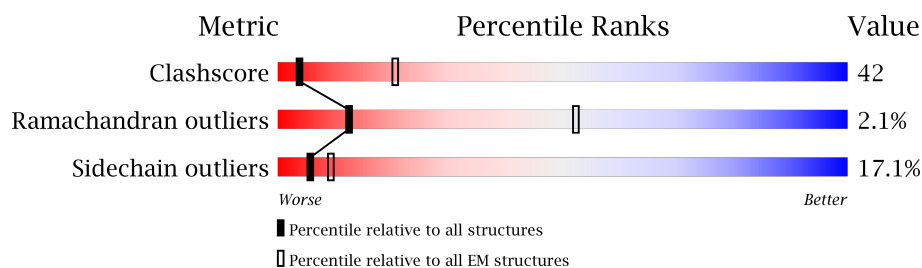
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 8.70 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  $\geq 3$ , 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	792	
1	B	792	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 11384 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 P1.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	738	Total	C	N	O	S	0	0
			5692	3588	984	1096	24		
1	B	738	Total	C	N	O	S	0	0
			5692	3588	984	1096	24		

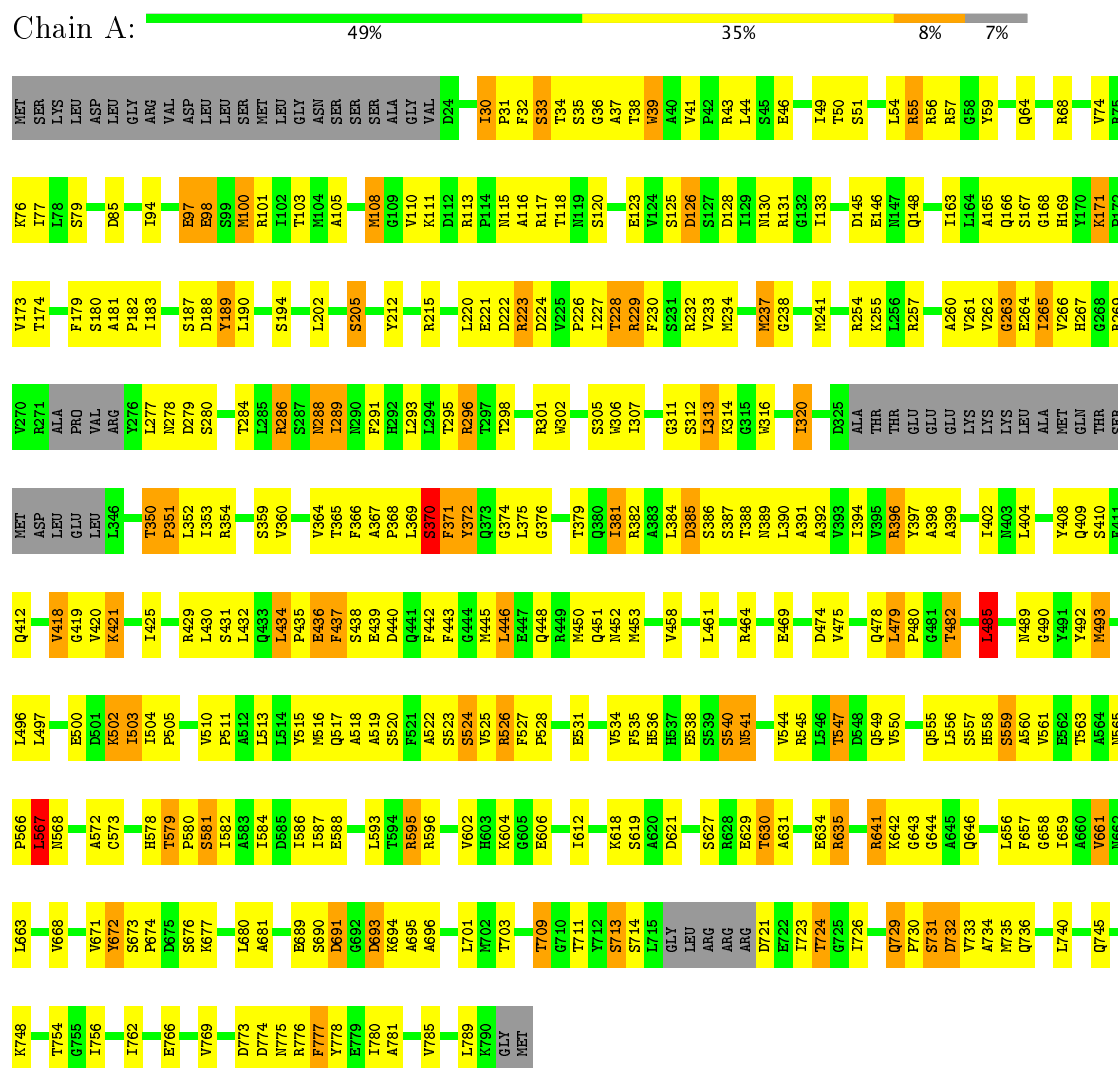
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	691	ASP	GLU	conflict	UNP Q9MC13
B	691	ASP	GLU	conflict	UNP Q9MC13

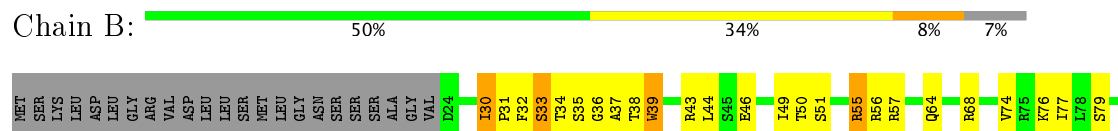
### 3 Residue-property plots

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: P1



#### • Molecule 1: P1



D774	D85
M775	T94
F776	E97
F777	E98
Y778	S99
E779	M100
I780	R101
A781	I102
	T103
	M104
	A105
	M108
	G109
	V110
	K111
	D112
	R113
	P114
	T118
	N119
	S120
	I121
	E122
	E123
	V124
	S125
	D126
	S127
	V233
	D128
	I129
	N130
	R131
	G132
	I133
	D145
	E146
	N147
	Q148
	I163
	L164
	A165
	Q166
	S167
	G168
	H169
	G268
	K171
	P172
	V173
	T174
	F179
	S180
	A181
	P182
	I183
	S187
	D188
	Y189
	L190
	S194
	L202
	S205
	Y212
	R215
	L220
	E221
	A116
	D222
	R223
	D224
	E225
	P226
	I227
	T228
	R229
	F230
	S231
	R232
	V233
	M234
	M237
	G238
	M241
	R254
	R257
	A260
	V261
	V262
	G263
	E264
	I265
	V266
	H267
	G268
	R269
	K171
	V270
	R271
	ALA
	PRO
	VAL
	ARG
	T276
	L277
	N278
	S280
	T284
	L285
	R286
	S287
	N288
	A367
	P368
	L369
	S370
	F291
	F371
	Y372
	R296
	T297
	T298
	R301
	K302
	S305
	K306
	I307
	G311
	S312
	L313
	K314
	G315
	K316
	I320
	P325
	ALA
	THR
	THR
	GLU
	GLU
	GLU
	LYS
	LYS
	LYS
	LEU
	ALA
	NET
	GLN
	THR
	SER
	NET
	ASP
	LEU
	GLU
	LEU
	L346
	T350
	P351
	L352
	I353
	R354
	V360
	V364
	T365
	F366
	A367
	P368
	L369
	S370
	F371
	Y372
	R373
	G374
	L375
	G376
	T379
	K380
	I381
	R382
	D385
	S386
	S387
	T388
	N389
	L390
	A391
	A392
	V393
	I394
	V395
	R396
	Y397
	A398
	A399
	I402
	M403
	L404
	Y408
	Q409
	S410
	F411
	Q412
	V413
	P414
	T415
	V416
	D417
	V418
	G419
	V420
	K421
	K422
	L425
	R429
	L430
	S431
	L432
	Q433
	L434
	P435
	E436
	F437
	S438
	E439
	D440
	Q441
	F442
	F443
	L446
	N450
	N453
	V458
	L461
	R464
	E469
	D474
	V475
	Q478
	L479
	P480
	Q481
	T482
	L483
	N489
	Q490
	Y491
	Y492
	N493
	L496
	L497
	N568
	F569
	A572
	C573
	H578
	P505
	V510
	P511
	A512
	L513
	L514
	Y515
	M516
	Q517
	A518
	A519
	S520
	F521
	A522
	S523
	S524
	V525
	E526
	F527
	Q528
	E531
	V534
	F535
	H536
	H537
	S538
	S539
	S540
	N541
	V544
	R545
	L546
	T547
	D548
	Q549
	V550
	Q555
	L556
	S557
	H558
	S559
	A560
	V561
	E562
	T563
	A564
	N565
	P566
	L567
	N568
	F569
	A572
	C573
	H578
	T579
	P580
	S581
	L582
	A583
	L584
	D585
	L586
	L587
	L593
	T594
	B595
	B596
	V602
	H603
	K604
	L612
	K618
	S619
	A620
	D621
	B629
	T630
	A631
	B634
	R635
	V636
	L637
	V638
	R641
	K642
	G643
	G644
	A645
	Q646
	V647
	L656
	F657
	G658
	V661
	N662
	L663
	V668
	V671
	V672
	S673
	P674
	D675
	S676
	D773
	L680
	A681
	E689
	S690
	D691
	G692
	D693
	K694
	A695
	A696
	L701
	F702
	T703
	T709
	G710
	T711
	Y712
	S713
	S714
	L715
	GLY
	LEU
	ARG
	ARG
	ARG
	D721
	E722
	T723
	L724
	G725
	I726
	Q729
	P730
	D732
	V733
	A734
	K735
	Q736
	L740
	Q745
	K748
	I756
	I762
	E766
	V769
	S676
	D773

## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of particles used	12867	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI TECNAI F20	Depositor
Voltage (kV)	200	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	Not provided	Depositor
Minimum defocus (nm)	700	Depositor
Maximum defocus (nm)	3300	Depositor
Magnification	49300	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	3.90	1/5791 (0.0%)	0.94	4/7850 (0.1%)
1	B	0.58	0/5790	0.84	2/7847 (0.0%)
All	All	2.79	1/11581 (0.0%)	0.89	6/15697 (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	1	3
1	B	0	3
All	All	1	6

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	370	SER	N-CA	293.68	7.33	1.46

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	370	SER	N-CA-CB	33.33	160.49	110.50
1	A	370	SER	N-CA-C	-17.18	64.62	111.00
1	A	567	LEU	CA-CB-CG	6.07	129.26	115.30
1	B	567	LEU	CA-CB-CG	6.06	129.24	115.30
1	A	485	LEU	CA-CB-CG	5.47	127.88	115.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	370	SER	CA

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	263	GLY	Peptide
1	A	595	ARG	Peptide
1	A	631	ALA	Peptide
1	B	263	GLY	Peptide
1	B	595	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	5692	0	5684	516	0
1	B	5692	0	5694	519	0
All	All	11384	0	11378	960	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 42.

The worst 5 of 960 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:39:TRP:CH2	1:A:515:TYR:CD2	1.78	1.69
1:A:289:ILE:HD13	1:A:515:TYR:CE2	1.22	1.63
1:A:39:TRP:CE3	1:A:515:TYR:CD2	1.87	1.63
1:A:39:TRP:HH2	1:A:515:TYR:CB	1.01	1.62
1:A:39:TRP:CZ3	1:A:515:TYR:CG	1.85	1.61

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	730/792 (92%)	640 (88%)	74 (10%)	16 (2%)	8	44
1	B	728/792 (92%)	641 (88%)	72 (10%)	15 (2%)	8	45
All	All	1458/1584 (92%)	1281 (88%)	146 (10%)	31 (2%)	12	45

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	370	SER
1	A	503	ILE
1	B	503	ILE
1	A	559	SER
1	A	643	GLY

### 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	617/663 (93%)	511 (83%)	106 (17%)	2	14
1	B	617/663 (93%)	512 (83%)	105 (17%)	2	15
All	All	1234/1326 (93%)	1023 (83%)	211 (17%)	6	15

5 of 211 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	703	THR
1	B	115	ASN
1	B	672	TYR
1	A	723	ILE
1	B	33	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	729	GLN
1	A	741	GLN
1	B	478	GLN
1	A	478	GLN
1	A	517	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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	369:LEU	C	370:SER	N	10.38