wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 02:11 PM BST

PDB ID	:	5FJ7
EMDB ID:	:	EMD-3187
Title	:	Structure of the P2 polymerase inside in vitro assembled bacteriophage phi6 polymerase complex, with P1 included
Authors	:	Ilca, S.; Kotecha, A.; Sun, X.; Poranen, M.P.; Stuart, D.I.; Huiskonen, J.T.
Deposited on	:	2015-10-06
Resolution	:	7.90 Å(reported)
Based on PDB ID	:	4K7H,1HHS

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

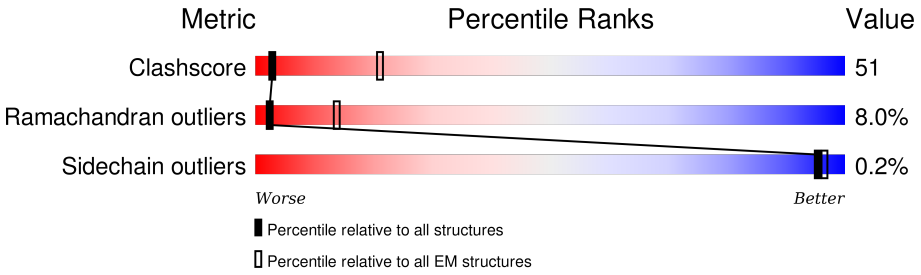
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

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 7.90 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	761	33% 58% 7% .
1	B	761	33% 59% 7% .
2	C	664	23% 74% .

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 17106 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
1	A	761	Total 5920	C 3741	N 1048	O 1109	S 22	0	0
1	B	761	Total 5920	C 3741	N 1048	O 1109	S 22	0	0

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

- Molecule 2 is a protein called RNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	664	Total 5265	C 3342	N 914	O 977	S 32	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	456	MET	ILE	CONFLICT	UNP P11124

- Molecule 3 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

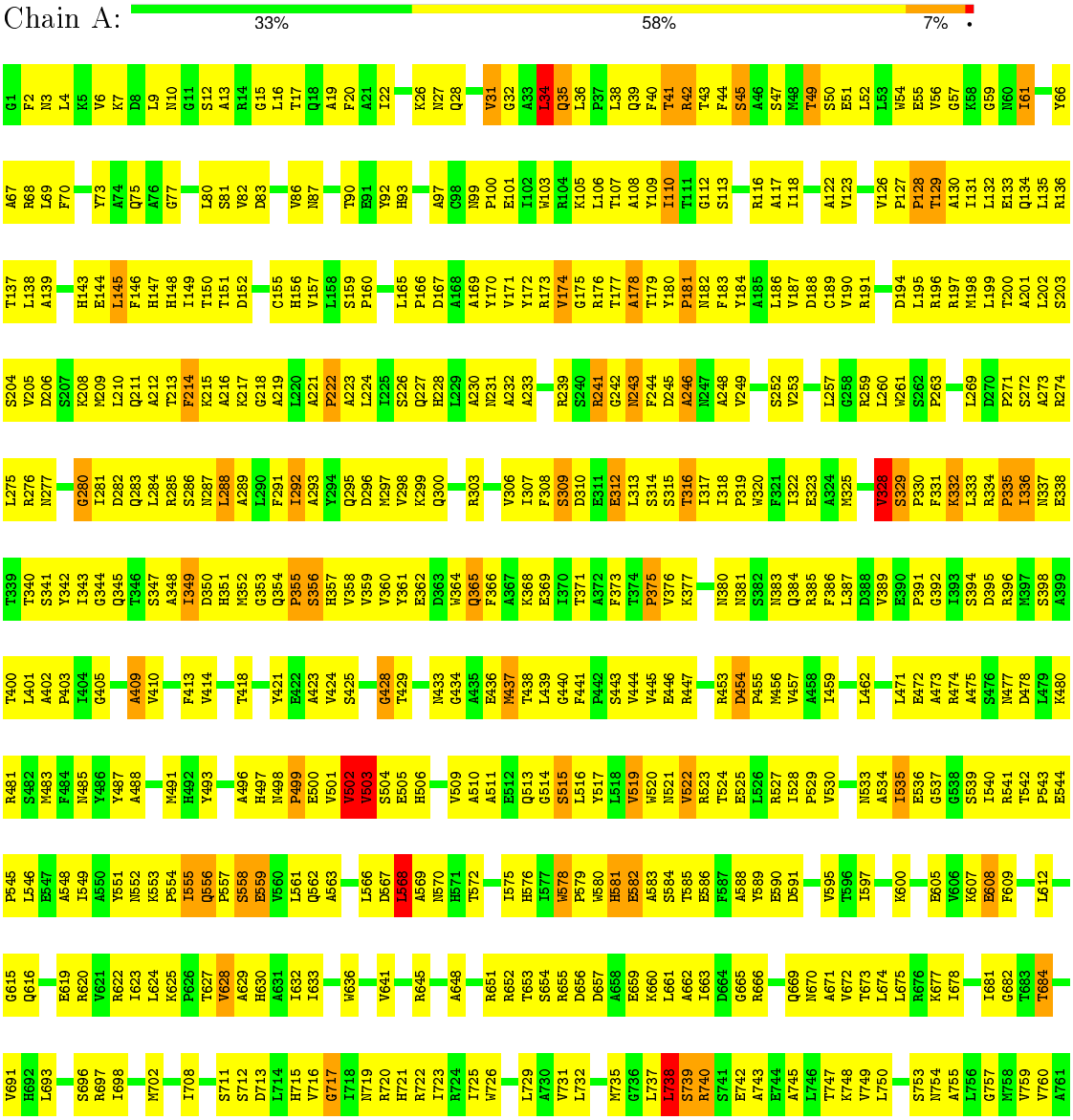
Mol	Chain	Residues	Atoms		AltConf
3	C	1	Total 1	Mn 1	0

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.

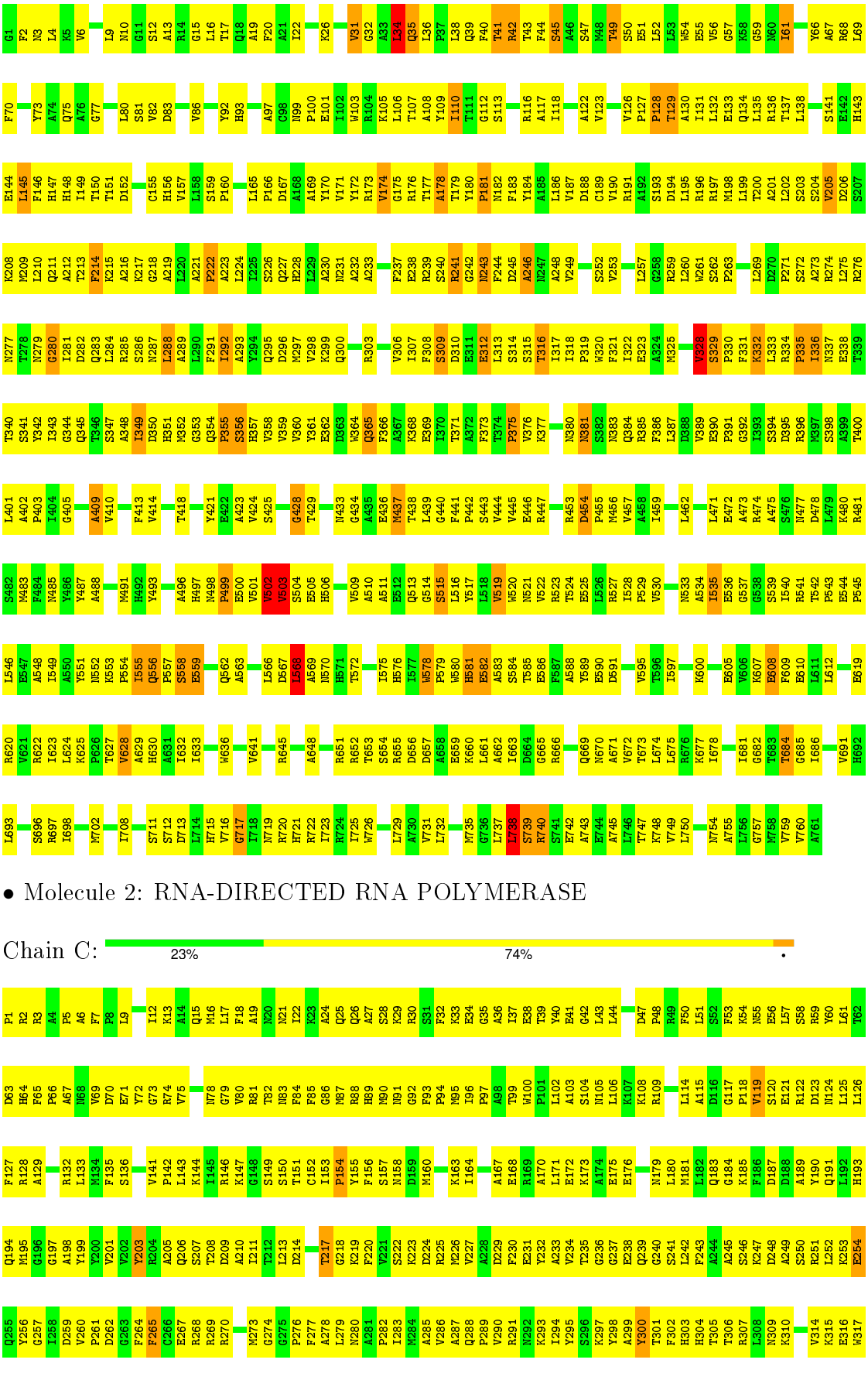
- Molecule 1: MAJOR INNER PROTEIN P1

Chain A:



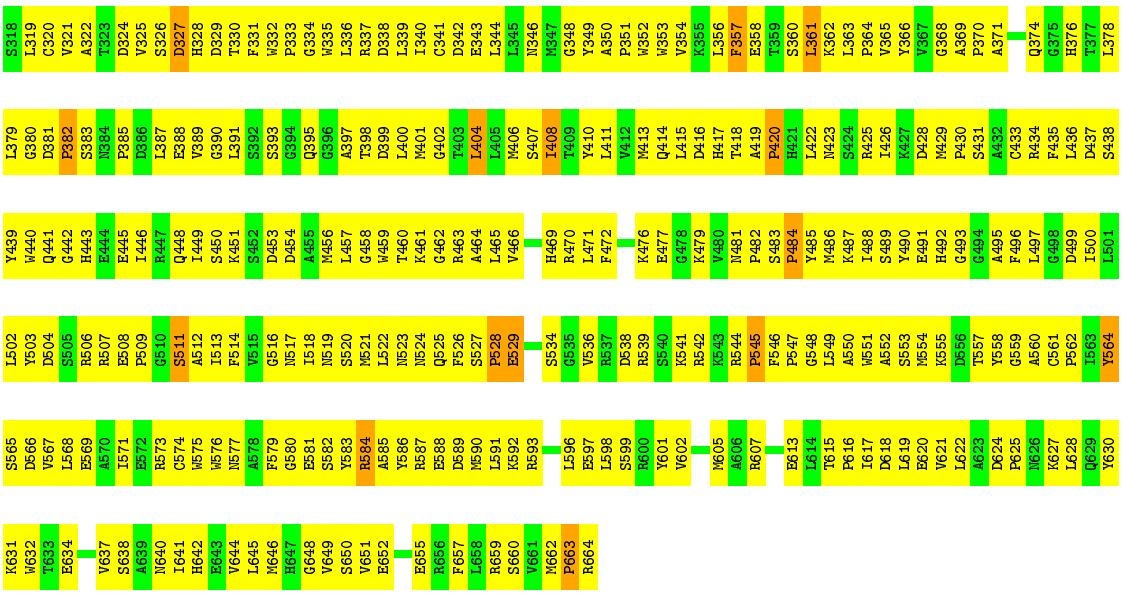
- Molecule 1: MAJOR INNER PROTEIN P1

Chain B:



• Molecule 2: RNA-DIRECTED RNA POLYMERASE

Chain C:


S318 L319 C320 V321 A322 T323 D324 V325 S326 D327 H328 D329 T330 F331 W332 P333 G334 W335 L336 R337 D338 L339 I340 C341 D342 E343 L344 L345 N346 M347 G348 Y349 A350 P351 W352 W353 V354 K355 L356 F357 E358 T359 S360 L361 K362 L363 P364 V365 Y366 V367 G368 A369 P370 A371 Q374 G375 H376 T377 L378
L379 G380 D381 P382 S383 N384 P385 D386 L387 E388 V389 G390 L391 S392 S393 G394 Q395 G396 A397 T398 D399 L400 M401 G402 T403 L404 L405 M406 S407 I408 T409 Y410 L411 V412 M413 Q414 L415 D416 H417 T418 A419 P420 H421 L422 N423 S424 R425 I426 K427 D428 M429 P430 S431 A432 C433 R434 F435 L436 D437 S438
Y439 W440 Q441 G442 H443 E444 E445 I446 R447 Q448 I449 S450 K451 S452 D453 D454 A455 M456 L457 G458 W459 T460 K461 G462 R463 A464 L465 V466 H469 R470 L471 F472 K476 E477 G478 K479 V480 M481 P482 S483 P484 Y485 M486 K487 I488 S489 Y490 E491 H492 G493 G494 A495 F496 L497 G498 D499 I500 L501
L502 Y503 D504 S505 R506 R507 E508 P509 G510 S511 A512 I513 P514 V515 G516 N517 I518 N519 S520 M521 L522 N523 N524 Q525 F526 S527 P528 E529 S534 G535 V536 R537 D538 R539 S540 K541 R542 K543 R544 P545 F546 P547 G548 L549 A550 W551 A552 S553 M554 K555 D556 T557 Y558 G559 A560 C561 P562 I563 Y564
S565 D566 V567 L568 E569 A570 I571 E572 R573 C574 W575 W576 N577 A578 F579 G580 E581 S582 Y583 R584 A585 Y586 R587 E588 D589 M590 L591 K592 R593 L596 E597 L598 S599 R600 Y601 V602 M605 A606 R607 E613 L614 T615 P616 I617 D618 L619 E620 V621 L622 A623 D624 P625 M626 K627 L628 Q629 Y630
K631 W632 T633 E634 V637 S638 A639 M640 I641 H642 E643 V644 L645 M646 H647 G648 V649 S650 V651 E652 E655 R656 F657 L658 R659 S660 V661 M662 P663 R664

4 Experimental information (i)

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 ($e^-/Å^2$)	0.16	Depositor
Minimum defocus (nm)	1100	Depositor
Maximum defocus (nm)	2600	Depositor
Magnification	160000	Depositor
Image detector	GATAN K2 (4K X 4K)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MN

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	0.43	0/6040	0.70	7/8206 (0.1%)
1	B	0.43	0/6040	0.70	7/8206 (0.1%)
2	C	1.17	7/5396 (0.1%)	0.66	2/7297 (0.0%)
All	All	0.74	7/17476 (0.0%)	0.69	16/23709 (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	5
1	B	0	5
2	C	0	1
All	All	0	11

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	361	LEU	CG-CD1	52.88	3.47	1.51
2	C	357	PHE	CD1-CE1	26.88	1.93	1.39
2	C	357	PHE	CD2-CE2	26.52	1.92	1.39
2	C	357	PHE	CE2-CZ	25.99	1.86	1.37
2	C	357	PHE	CE1-CZ	25.62	1.86	1.37

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	361	LEU	CB-CG-CD1	11.31	130.23	111.00
1	B	128	PRO	CA-C-N	6.76	132.06	117.20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	128	PRO	CA-C-N	6.75	132.05	117.20
1	A	128	PRO	C-N-CA	6.09	136.93	121.70
1	B	128	PRO	C-N-CA	6.09	136.93	121.70

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	328	VAL	Peptide
1	A	34	LEU	Peptide
1	A	35	GLN	Peptide
1	A	365	GLN	Peptide
1	A	437	MET	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	5913	504	0
1	B	5920	0	5907	641	0
2	C	5265	0	5154	718	0
3	C	1	0	0	0	0
All	All	17106	0	16974	1735	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 51.

The worst 5 of 1735 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:C:357:PHE:CZ	2:C:357:PHE:CE1	1.86	1.58
2:C:357:PHE:CE2	2:C:357:PHE:CD2	1.92	1.58
2:C:357:PHE:CE2	2:C:357:PHE:CZ	1.86	1.56
2:C:357:PHE:CD1	2:C:357:PHE:CE1	1.93	1.53
1:B:237:PHE:O	2:C:425:ARG:NH2	1.58	1.36

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	759/761 (100%)	594 (78%)	91 (12%)	74 (10%)	1 14
1	B	759/761 (100%)	594 (78%)	91 (12%)	74 (10%)	1 14
2	C	662/664 (100%)	575 (87%)	61 (9%)	26 (4%)	4 36
All	All	2180/2186 (100%)	1763 (81%)	243 (11%)	174 (8%)	2 19

5 of 174 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	31	VAL
1	A	41	THR
1	A	129	THR
1	A	145	LEU
1	A	174	VAL

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%)	627 (100%)	2 (0%)	94 96
1	B	629/629 (100%)	627 (100%)	2 (0%)	94 96
2	C	557/557 (100%)	557 (100%)	0	100 100
All	All	1815/1815 (100%)	1811 (100%)	4 (0%)	95 97

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	503	VAL
1	A	519	VAL
1	B	503	VAL
1	B	519	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	345	GLN
1	B	670	ASN
2	C	448	GLN
1	B	485	ASN
1	B	492	HIS

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

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.