Full wwPDB X-ray Structure Validation Report ⓘ

Feb 18, 2024 – 08:37 AM EST

PDB ID	:	4FB3
Title	:	Polyomavirus T-ag binds symmetrical repeats at the viral origin in an asymmetrical manner
Authors	:	Bohm, A.; Harrison, C.J.; Schaffhausen, B.S.; Jiang, T.
Deposited on	:	2012-05-22
Resolution	:	3.79 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

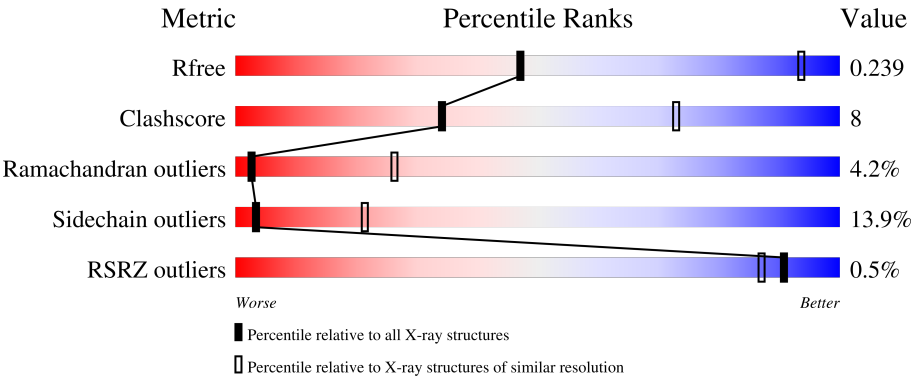
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.79 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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	C	26	54% 38% 8%
2	W	26	69% 31%
3	A	146	% 53% 18% 5% • 21%
3	B	146	53% 21% • • 22%
3	E	146	51% 25% • 22%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3787 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, 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 DNA chain called ORI DNA oligonucleotide-Crick strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	26	Total	C	N	O	P	0	0	0
			530	249	102	154	25			

- Molecule 2 is a DNA chain called ORI DNA oligonucleotide-Watson strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	W	26	Total	C	N	O	P	0	0	0
			530	249	102	154	25			

- Molecule 3 is a protein called Large T antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	115	Total	C	N	O	S	0	0	0
			913	595	145	163	10			
3	B	114	Total	C	N	O	S	0	0	0
			907	591	144	162	10			
3	E	114	Total	C	N	O	S	0	0	0
			907	591	144	162	10			

There are 45 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	283	MET	-	expression tag	UNP P03074
A	284	HIS	-	expression tag	UNP P03074
A	285	HIS	-	expression tag	UNP P03074
A	286	HIS	-	expression tag	UNP P03074
A	287	HIS	-	expression tag	UNP P03074
A	288	HIS	-	expression tag	UNP P03074
A	289	HIS	-	expression tag	UNP P03074
A	421	GLY	-	expression tag	UNP P03074
A	422	SER	-	expression tag	UNP P03074
A	423	HIS	-	expression tag	UNP P03074

Continued on next page...

Continued from previous page...

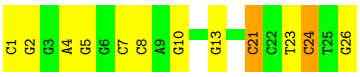
Chain	Residue	Modelled	Actual	Comment	Reference
A	424	HIS	-	expression tag	UNP P03074
A	425	HIS	-	expression tag	UNP P03074
A	426	HIS	-	expression tag	UNP P03074
A	427	HIS	-	expression tag	UNP P03074
A	428	HIS	-	expression tag	UNP P03074
B	283	MET	-	expression tag	UNP P03074
B	284	HIS	-	expression tag	UNP P03074
B	285	HIS	-	expression tag	UNP P03074
B	286	HIS	-	expression tag	UNP P03074
B	287	HIS	-	expression tag	UNP P03074
B	288	HIS	-	expression tag	UNP P03074
B	289	HIS	-	expression tag	UNP P03074
B	421	GLY	-	expression tag	UNP P03074
B	422	SER	-	expression tag	UNP P03074
B	423	HIS	-	expression tag	UNP P03074
B	424	HIS	-	expression tag	UNP P03074
B	425	HIS	-	expression tag	UNP P03074
B	426	HIS	-	expression tag	UNP P03074
B	427	HIS	-	expression tag	UNP P03074
B	428	HIS	-	expression tag	UNP P03074
E	283	MET	-	expression tag	UNP P03074
E	284	HIS	-	expression tag	UNP P03074
E	285	HIS	-	expression tag	UNP P03074
E	286	HIS	-	expression tag	UNP P03074
E	287	HIS	-	expression tag	UNP P03074
E	288	HIS	-	expression tag	UNP P03074
E	289	HIS	-	expression tag	UNP P03074
E	421	GLY	-	expression tag	UNP P03074
E	422	SER	-	expression tag	UNP P03074
E	423	HIS	-	expression tag	UNP P03074
E	424	HIS	-	expression tag	UNP P03074
E	425	HIS	-	expression tag	UNP P03074
E	426	HIS	-	expression tag	UNP P03074
E	427	HIS	-	expression tag	UNP P03074
E	428	HIS	-	expression tag	UNP P03074

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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 and electron density. 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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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.

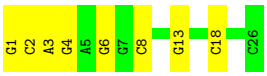
• Molecule 1: ORI DNA oligonucleotide-Crick strand

Chain C:



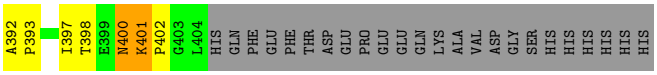
• Molecule 2: ORI DNA oligonucleotide-Watson strand

Chain W:



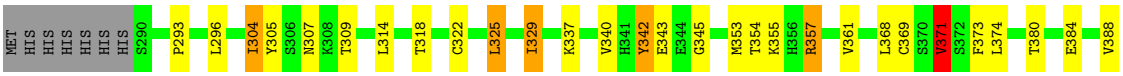
• Molecule 3: Large T antigen

Chain A:



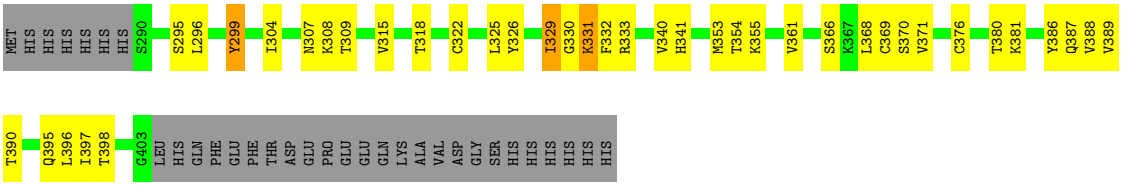
• Molecule 3: Large T antigen

Chain B:



• Molecule 3: Large T antigen

Chain E:


MET HIS HIS HIS HIS HIS HIS S290 S295 L296 Y299 I304 M307 K308 T309 V315 T318 C322 L325 Y326 I329 G330 K331 F332 R333 V340 H341 M353 T354 K355 V361 S366 K367 L368 C369 S370 V371 C376 T380 K381 Y386 Q387 V388 V389
T390 Q395 L396 I397 T398 G403 LEU HIS GLN PHE GLU PHE THR ASP GLU PRO GLU GLU GLN LYS ALA VAL ASP GLY SER HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	165.15Å 167.92Å 77.73Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.93 – 3.79 46.93 – 3.79	Depositor EDS
% Data completeness (in resolution range)	99.9 (46.93-3.79) 99.9 (46.93-3.79)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.39 (at 3.77Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.2_869)	Depositor
R, R_{free}	0.222 , 0.244 0.220 , 0.239	Depositor DCC
R_{free} test set	531 reflections (4.79%)	wwPDB-VP
Wilson B-factor (Å^2)	91.4	Xtriage
Anisotropy	0.315	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 39.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.011 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	3787	wwPDB-VP
Average B, all atoms (Å^2)	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.00% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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	#$\|Z\|$ >5	RMSZ	#$\|Z\|$ >5
1	C	0.51	0/594	1.21	4/915 (0.4%)
2	W	0.52	0/594	1.22	6/915 (0.7%)
3	A	0.33	0/937	0.55	0/1263
3	B	0.27	0/931	0.47	0/1255
3	E	0.29	0/931	0.50	0/1255
All	All	0.38	0/3987	0.81	10/5603 (0.2%)

There are no bond length outliers.

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	W	6	DG	O4'-C1'-N9	7.43	113.20	108.00
2	W	6	DG	C1'-O4'-C4'	-6.61	103.49	110.10
2	W	6	DG	C3'-C2'-C1'	-6.24	95.01	102.50
1	C	24	DC	O4'-C1'-N1	6.06	112.24	108.00
1	C	5	DG	C3'-C2'-C1'	-5.65	95.72	102.50
2	W	8	DC	O4'-C1'-N1	5.21	111.65	108.00
1	C	21	DC	O4'-C4'-C3'	-5.09	102.46	104.50
1	C	24	DC	C1'-O4'-C4'	-5.09	105.01	110.10
2	W	13	DG	C3'-C2'-C1'	-5.07	96.42	102.50
2	W	6	DG	O4'-C1'-C2'	-5.01	101.89	105.90

There are no chirality outliers.

There are no planarity outliers.

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	C	530	0	290	9	0
2	W	530	0	290	5	0
3	A	913	0	917	20	0
3	B	907	0	913	17	0
3	E	907	0	913	15	0
All	All	3787	0	3323	60	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 8.

All (60) 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)
3:B:397:ILE:HG22	3:B:398:THR:HG23	1.65	0.76
1:C:21:DC:N4	3:E:307:ASN:O	2.19	0.74
3:B:369:CYS:HB3	3:B:371:VAL:HG13	1.73	0.71
3:B:337:LYS:H	3:B:400:ASN:HD21	1.39	0.68
1:C:26:DG:N2	2:W:2:DC:O2	2.26	0.67
3:A:400:ASN:O	3:A:401:LYS:HG2	1.96	0.65
3:A:370:SER:N	3:A:371:VAL:HB	2.12	0.65
2:W:3:DA:H5'	3:E:380:THR:HG21	1.79	0.65
2:W:4:DG:N7	3:E:308:LYS:NZ	2.41	0.61
3:A:341:HIS:CE1	3:A:397:ILE:HG13	2.35	0.61
3:B:392:ALA:HA	3:B:394:PHE:H	1.66	0.61
3:E:299:TYR:HB3	3:E:381:LYS:HB2	1.84	0.60
3:E:397:ILE:HG22	3:E:398:THR:HG22	1.84	0.59
3:E:341:HIS:HB2	3:E:395:GLN:HG3	1.86	0.58
3:A:329:ILE:HD13	3:A:334:PRO:HG3	1.87	0.56
3:A:296:LEU:O	3:A:298:GLY:N	2.40	0.54
3:B:293:PRO:HG2	3:B:296:LEU:HD12	1.88	0.54
3:B:322:CYS:HA	3:B:325:LEU:HD22	1.90	0.53
3:B:340:VAL:HG12	3:B:396:LEU:HA	1.90	0.53
3:E:388:VAL:HG13	3:E:389:VAL:HG23	1.92	0.52
1:C:10:DG:H1	2:W:18:DC:H42	1.59	0.51
3:B:304:ILE:HG23	3:B:305:TYR:CD2	2.46	0.51
3:A:362:LYS:NZ	3:A:375:MET:HA	2.27	0.49
3:E:322:CYS:O	3:E:326:TYR:N	2.46	0.48
3:A:321:LYS:HE2	3:A:325:LEU:HG	1.96	0.48
1:C:4:DA:C8	1:C:4:DA:H5'	2.49	0.48
3:A:321:LYS:HD3	3:A:369:CYS:HB2	1.96	0.48
3:B:343:GLU:C	3:B:345:GLY:H	2.16	0.47
3:B:388:VAL:HG13	3:B:389:VAL:HG23	1.96	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:E:329:ILE:O	3:E:331:LYS:N	2.47	0.47
3:B:337:LYS:HG2	3:B:400:ASN:ND2	2.30	0.47
1:C:1:DC:H2'	1:C:2:DG:C8	2.50	0.47
3:A:362:LYS:HZ3	3:A:375:MET:HA	1.80	0.46
3:B:353:MET:HG3	3:B:354:THR:HG23	1.96	0.46
3:A:374:LEU:H	3:A:374:LEU:HD13	1.81	0.46
3:A:369:CYS:HB3	3:A:371:VAL:HB	1.98	0.45
3:A:352:THR:HG21	3:A:356:HIS:CD2	2.53	0.44
1:C:13:DG:N7	3:A:357:ARG:NH2	2.53	0.44
1:C:23:DT:H2"	1:C:24:DC:O5'	2.17	0.44
3:A:401:LYS:HA	3:A:402:PRO:HD3	1.53	0.44
3:E:340:VAL:HG21	3:E:389:VAL:CG1	2.48	0.43
3:A:329:ILE:HD11	3:A:337:LYS:NZ	2.34	0.43
3:E:309:THR:HB	3:E:355:LYS:HB3	2.00	0.43
2:W:1:DG:H2"	2:W:2:DC:OP1	2.18	0.43
3:A:304:ILE:HG23	3:A:305:TYR:HD2	1.82	0.43
3:A:392:ALA:HA	3:A:393:PRO:HA	1.80	0.42
3:B:343:GLU:O	3:B:345:GLY:N	2.44	0.42
3:E:333:ARG:NH2	3:E:353:MET:SD	2.92	0.42
3:B:309:THR:HB	3:B:355:LYS:HB3	2.01	0.42
1:C:7:DC:H2'	1:C:8:DC:C6	2.55	0.42
3:B:329:ILE:H	3:B:329:ILE:HG13	1.61	0.41
3:E:386:TYR:O	3:E:390:THR:HG23	2.20	0.41
3:B:304:ILE:HG23	3:B:305:TYR:HD2	1.85	0.41
3:B:307:ASN:CG	3:B:357:ARG:HH21	2.23	0.41
3:E:332:PHE:CE2	3:E:361:VAL:HG12	2.55	0.41
3:E:340:VAL:HG21	3:E:389:VAL:HG12	2.02	0.41
1:C:23:DT:H2'	1:C:24:DC:C6	2.56	0.41
3:A:385:CYS:O	3:A:388:VAL:HG13	2.21	0.41
3:A:296:LEU:C	3:A:298:GLY:H	2.24	0.41
3:A:384:GLU:OE1	3:A:384:GLU:N	2.43	0.40

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 X-ray entries followed by that with respect to entries of similar resolution.

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
3	A	113/146 (77%)	92 (81%)	14 (12%)	7 (6%)	1 20
3	B	112/146 (77%)	99 (88%)	9 (8%)	4 (4%)	3 30
3	E	112/146 (77%)	95 (85%)	14 (12%)	3 (3%)	5 35
All	All	337/438 (77%)	286 (85%)	37 (11%)	14 (4%)	3 26

All (14) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	297	THR
3	A	369	CYS
3	A	370	SER
3	A	371	VAL
3	E	330	GLY
3	B	304	ILE
3	B	329	ILE
3	A	329	ILE
3	A	401	LYS
3	B	342	TYR
3	E	370	SER
3	E	371	VAL
3	B	371	VAL
3	A	304	ILE

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 X-ray entries followed by that with respect to entries of similar resolution.

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
3	A	103/133 (77%)	88 (85%)	15 (15%)	3 19
3	B	103/133 (77%)	91 (88%)	12 (12%)	5 27
3	E	103/133 (77%)	87 (84%)	16 (16%)	2 17
All	All	309/399 (77%)	266 (86%)	43 (14%)	3 21

All (43) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	304	ILE
3	A	319	LYS
3	A	328	THR
3	A	335	GLU
3	A	344	GLU
3	A	350	PHE
3	A	352	THR
3	A	354	THR
3	A	367	LYS
3	A	369	CYS
3	A	371	VAL
3	A	374	LEU
3	A	388	VAL
3	A	398	THR
3	A	400	ASN
3	B	314	LEU
3	B	318	THR
3	B	325	LEU
3	B	342	TYR
3	B	357	ARG
3	B	361	VAL
3	B	368	LEU
3	B	371	VAL
3	B	373	PHE
3	B	374	LEU
3	B	380	THR
3	B	384	GLU
3	E	295	SER
3	E	296	LEU
3	E	299	TYR
3	E	304	ILE
3	E	315	VAL
3	E	318	THR
3	E	325	LEU
3	E	329	ILE
3	E	331	LYS
3	E	354	THR
3	E	366	SER
3	E	368	LEU
3	E	369	CYS
3	E	376	CYS
3	E	387	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	E	396	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	324	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 monosaccharides 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.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	C	26/26 (100%)	-0.59	0 100 100	77, 107, 133, 142	0
2	W	26/26 (100%)	-0.57	0 100 100	83, 100, 137, 140	0
3	A	115/146 (78%)	0.03	2 (1%) 70 62	61, 91, 129, 191	0
3	B	114/146 (78%)	0.21	0 100 100	66, 95, 125, 145	0
3	E	114/146 (78%)	0.11	0 100 100	60, 91, 132, 187	0
All	All	395/490 (80%)	0.03	2 (0%) 91 87	60, 94, 133, 191	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	372	SER	6.1
3	A	373	PHE	3.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.