Full wwPDB X-ray Structure Validation Report (i)

Jun 23, 2024 – 05:49 AM EDT

PDB ID	:	5EZN
Title	:	Crystal Structure of PfCyRPA
Authors	:	Favuzza, P.; Pluschke, G.; Rudolph, M.G.
Deposited on	:	2015-11-26
Resolution	:	2.51 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 (i) 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 (1)) were used in the production of this report:

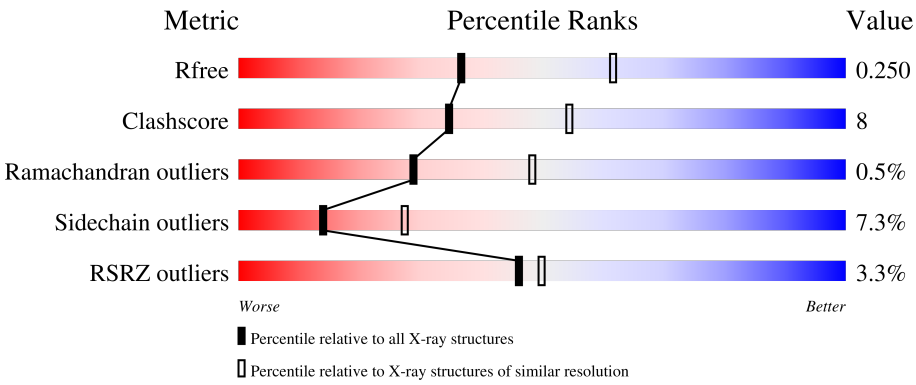
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.51 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	A	157	7%, 68%, 24%, 6%
2	E	176	2%, 67%, 28%
3	B	161	3%, 69%, 23%, 5%
4	G	170	%, 80%, 16%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5318 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 protein called Cysteine-rich protective antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	148	1250	810	198	235	7	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	145	GLN	ASN	engineered mutation	UNP Q8IFM8
A	186	LYS	-	expression tag	UNP Q8IFM8
A	187	PHE	-	expression tag	UNP Q8IFM8
A	188	LYS	-	expression tag	UNP Q8IFM8

- Molecule 2 is a protein called Cysteine-rich protective antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	170	1404	901	224	273	6	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	322	GLN	ASN	engineered mutation	UNP Q8IFM8
E	338	GLN	ASN	engineered mutation	UNP Q8IFM8
E	363	GLU	-	expression tag	UNP Q8IFM8
E	364	PHE	-	expression tag	UNP Q8IFM8

- Molecule 3 is a protein called Cysteine-rich protective antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	B	153	1287	828	207	245	7	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	122	ASP	GLU	conflict	UNP Q8IFM8
B	145	GLN	ASN	engineered mutation	UNP Q8IFM8

- Molecule 4 is a protein called Cysteine-rich protective antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	G	166	1372	881	219	266	6	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	338	GLN	ASN	engineered mutation	UNP Q8IFM8

- Molecule 5 is water.

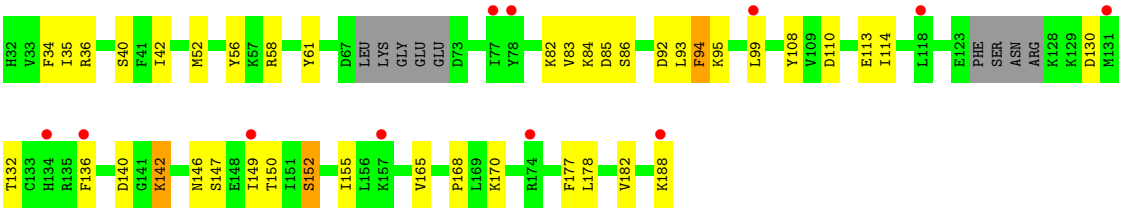
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	E	2	2	2	0	0
			Total	O		
5	G	3	3	3	0	0

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.

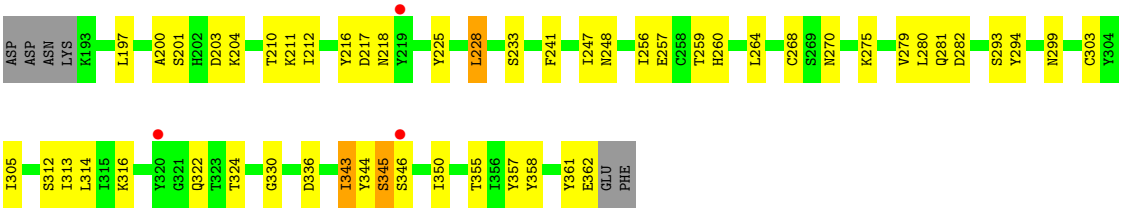
- Molecule 1: Cysteine-rich protective antigen

Chain A:



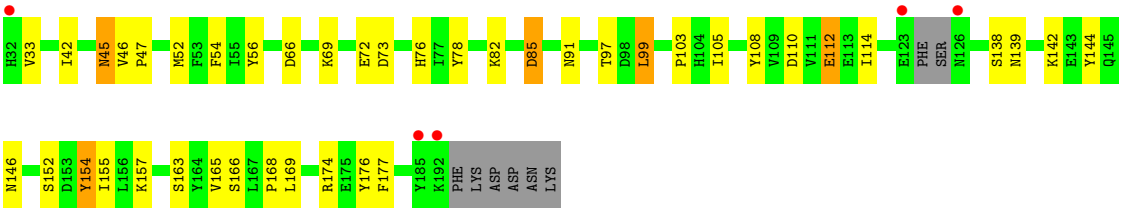
- Molecule 2: Cysteine-rich protective antigen

Chain E:



- Molecule 3: Cysteine-rich protective antigen

Chain B:



- Molecule 4: Cysteine-rich protective antigen

Chain G:

K193
D194
L197
C198
S201
K215
Y219
F226
F227
L228
R229
I232
N235
F241
Y242
V243
N246
I247
N248
N252
I256
V267
N270
K275
I305
F306
N309
I313
E318
K319
TYR
GLY
ASN
THR
T324
G330
T331
I343
S346

I350
Y351
E362

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	46.24Å 78.50Å 97.04Å 90.00° 94.85° 90.00°	Depositor
Resolution (Å)	43.00 – 2.51 46.07 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.7 (43.00-2.51) 89.7 (46.07-2.50)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.82 (at 2.51Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.183 , 0.250 0.183 , 0.250	Depositor DCC
R_{free} test set	1179 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å^2)	71.5	Xtriage
Anisotropy	0.145	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 66.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5318	wwPDB-VP
Average B, all atoms (Å^2)	103.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.47% 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 (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	$\#\|Z\| >5$	RMSZ	$\#\|Z\| >5$
1	A	0.40	0/1276	0.58	0/1718
2	E	0.49	0/1438	0.61	0/1945
3	B	0.50	1/1313 (0.1%)	0.62	0/1769
4	G	0.49	0/1404	0.64	0/1897
All	All	0.47	1/5431 (0.0%)	0.61	0/7329

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	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	46	VAL	C-N	-5.73	1.23	1.34

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	83	VAL	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	1250	0	1230	26	0
2	E	1404	0	1319	30	0
3	B	1287	0	1265	21	0
4	G	1372	0	1291	19	0
5	E	2	0	0	0	0
5	G	3	0	0	0	0
All	All	5318	0	5105	86	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 (86) 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:168:PRO:HB3	1:A:177:PHE:HE1	1.55	0.71
3:B:69:LYS:NZ	3:B:72:GLU:OE1	2.19	0.70
3:B:52:MET:HG2	3:B:165:VAL:HG23	1.78	0.65
2:E:281:GLN:OE1	2:E:293:SER:OG	2.14	0.64
3:B:56:TYR:OH	3:B:112:GLU:O	2.15	0.64
4:G:346:SER:HB2	4:G:350:ILE:HB	1.82	0.61
3:B:45:ASN:N	3:B:45:ASN:OD1	2.32	0.60
3:B:91:ASN:HD22	3:B:144:TYR:HD1	1.50	0.59
1:A:56:TYR:HB3	1:A:61:TYR:CE2	2.37	0.59
1:A:130:ASP:HB3	1:A:152:SER:HA	1.86	0.58
1:A:113:GLU:OE1	2:E:204:LYS:NZ	2.24	0.58
1:A:42:ILE:HD11	1:A:82:LYS:HD2	1.87	0.57
1:A:56:TYR:CG	1:A:114:ILE:HD11	2.39	0.57
1:A:94:PHE:HE1	1:A:99:LEU:HB3	1.71	0.56
2:E:197:LEU:HD11	2:E:211:LYS:HE3	1.86	0.56
1:A:130:ASP:OD1	1:A:130:ASP:N	2.37	0.54
2:E:212:ILE:HG12	2:E:264:LEU:HD22	1.88	0.54
4:G:226:PHE:CZ	4:G:246:ASN:HA	2.42	0.53
1:A:56:TYR:CD2	1:A:114:ILE:HD11	2.44	0.53
2:E:330:GLY:HA2	2:E:343:ILE:O	2.09	0.53
1:A:130:ASP:CB	1:A:152:SER:HA	2.39	0.53
4:G:215:LYS:HB2	4:G:267:VAL:HG22	1.91	0.52
3:B:97:THR:OG1	3:B:99:LEU:HB2	2.10	0.52
1:A:140:ASP:HB2	1:A:142:LYS:HE3	1.91	0.51
1:A:84:LYS:O	1:A:85:ASP:HB2	2.09	0.51
1:A:92:ASP:OD2	1:A:95:LYS:HG3	2.10	0.51

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:36:ARG:HB2	2:E:357:TYR:HB3	1.93	0.50
3:B:108:TYR:HA	3:B:166:SER:O	2.11	0.50
2:E:294:TYR:O	2:E:303:CYS:HA	2.12	0.50
1:A:108:TYR:CE2	1:A:177:PHE:HB3	2.46	0.50
1:A:132:THR:HG22	1:A:150:THR:HG23	1.93	0.50
3:B:176:TYR:HA	4:G:201:SER:O	2.12	0.50
4:G:275:LYS:NZ	4:G:309:ASN:OD1	2.43	0.50
4:G:256:ILE:HG22	4:G:270:ASN:HA	1.92	0.50
4:G:226:PHE:CE1	4:G:247:ILE:HG12	2.47	0.49
1:A:155:ILE:HG22	2:E:197:LEU:HD21	1.93	0.49
1:A:168:PRO:HB3	1:A:177:PHE:CE1	2.41	0.49
4:G:330:GLY:HA2	4:G:343:ILE:O	2.12	0.49
2:E:299:ASN:OD1	3:B:47:PRO:HG3	2.13	0.48
4:G:275:LYS:HE2	4:G:306:PHE:CZ	2.48	0.48
4:G:226:PHE:HZ	4:G:246:ASN:HA	1.78	0.48
3:B:42:ILE:HD11	3:B:82:LYS:HD2	1.95	0.48
2:E:256:ILE:HG22	2:E:270:ASN:HA	1.96	0.48
2:E:343:ILE:HD12	2:E:343:ILE:H	1.78	0.47
2:E:345:SER:O	2:E:345:SER:OG	2.31	0.47
1:A:58:ARG:NH2	2:E:336:ASP:OD2	2.48	0.47
1:A:177:PHE:C	1:A:178:LEU:HD23	2.34	0.47
4:G:243:VAL:HG22	4:G:252:ASN:HB2	1.97	0.47
3:B:138:SER:HB2	3:B:144:TYR:CD2	2.50	0.47
1:A:35:ILE:HG12	2:E:358:TYR:HD1	1.80	0.47
2:E:280:LEU:HD12	2:E:281:GLN:N	2.29	0.47
4:G:194:ASP:OD2	4:G:215:LYS:NZ	2.29	0.47
3:B:168:PRO:HB3	3:B:177:PHE:HE1	1.80	0.46
3:B:174:ARG:HD2	3:B:176:TYR:OH	2.14	0.46
2:E:275:LYS:HD2	2:E:294:TYR:CZ	2.50	0.46
3:B:155:ILE:HG22	4:G:197:LEU:HD21	1.98	0.46
2:E:257:GLU:O	2:E:268:CYS:HA	2.16	0.45
3:B:66:ASP:OD1	3:B:76:HIS:HB2	2.17	0.45
2:E:264:LEU:HA	2:E:264:LEU:HD23	1.79	0.45
3:B:78:TYR:HA	3:B:91:ASN:O	2.18	0.44
1:A:93:LEU:HB3	1:A:136:PHE:CZ	2.51	0.44
4:G:313:ILE:HD12	4:G:313:ILE:HA	1.77	0.44
2:E:201:SER:OG	2:E:203:ASP:OD1	2.34	0.44
2:E:228:LEU:HA	2:E:241:PHE:HB3	2.00	0.44
3:B:154:TYR:CE1	3:B:157:LYS:HA	2.53	0.44
1:A:182:VAL:HG11	2:E:225:TYR:HB2	2.00	0.43
2:E:280:LEU:HD12	2:E:281:GLN:H	1.83	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:E:330:GLY:HA2	2:E:344:TYR:HB3	2.00	0.43
4:G:197:LEU:HD12	4:G:198:CYS:H	1.83	0.43
1:A:52:MET:HG2	1:A:165:VAL:CG2	2.47	0.43
2:E:211:LYS:O	2:E:264:LEU:HD13	2.18	0.43
2:E:200:ALA:O	2:E:210:THR:HB	2.18	0.43
2:E:361:TYR:HB3	2:E:362:GLU:H	1.71	0.42
1:A:146:ASN:O	1:A:147:SER:HB2	2.19	0.42
3:B:169:LEU:HD13	4:G:232:ILE:HG23	2.00	0.42
3:B:85:ASP:N	3:B:85:ASP:OD1	2.53	0.41
3:B:54:PHE:CE1	3:B:114:ILE:HD13	2.56	0.41
2:E:314:LEU:HD21	2:E:316:LYS:HG3	2.02	0.41
4:G:228:LEU:HA	4:G:241:PHE:HB3	2.01	0.41
4:G:243:VAL:CG2	4:G:252:ASN:HB2	2.51	0.41
4:G:228:LEU:HG	4:G:229:ARG:N	2.37	0.40
2:E:216:TYR:O	2:E:218:ASN:N	2.53	0.40
2:E:305:ILE:O	2:E:312:SER:HA	2.22	0.40
2:E:282:ASP:H	2:E:293:SER:HB3	1.86	0.40
3:B:105:ILE:O	3:B:163:SER:HB3	2.21	0.40
1:A:170:LYS:HE3	1:A:170:LYS:HB2	1.92	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	
1	A	142/157 (90%)	129 (91%)	13 (9%)	0	100	100
2	E	168/176 (96%)	155 (92%)	12 (7%)	1 (1%)	25	43
3	B	149/161 (92%)	137 (92%)	11 (7%)	1 (1%)	22	39
4	G	162/170 (95%)	150 (93%)	11 (7%)	1 (1%)	25	43
All	All	621/664 (94%)	571 (92%)	47 (8%)	3 (0%)	29	48

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	G	247	ILE
2	E	217	ASP
3	B	103	PRO

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
1	A	144/152 (95%)	135 (94%)	9 (6%)	18 34
2	E	156/162 (96%)	141 (90%)	15 (10%)	8 16
3	B	148/156 (95%)	136 (92%)	12 (8%)	11 23
4	G	153/156 (98%)	145 (95%)	8 (5%)	23 44
All	All	601/626 (96%)	557 (93%)	44 (7%)	14 27

All (44) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	34	PHE
1	A	40	SER
1	A	86	SER
1	A	94	PHE
1	A	110	ASP
1	A	142	LYS
1	A	149	ILE
1	A	152	SER
1	A	188	LYS
2	E	228	LEU
2	E	233	SER
2	E	247	ILE
2	E	248	ASN
2	E	259	THR
2	E	260	HIS
2	E	279	VAL
2	E	313	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	E	322	GLN
2	E	324	THR
2	E	343	ILE
2	E	345	SER
2	E	346	SER
2	E	350	ILE
2	E	355	THR
3	B	33	VAL
3	B	45	ASN
3	B	73	ASP
3	B	85	ASP
3	B	99	LEU
3	B	110	ASP
3	B	112	GLU
3	B	139	ASN
3	B	142	LYS
3	B	146	ASN
3	B	152	SER
3	B	154	TYR
4	G	228	LEU
4	G	235	ASN
4	G	247	ILE
4	G	248	ASN
4	G	305	ILE
4	G	313	ILE
4	G	318	GLU
4	G	331	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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	A	148/157 (94%)	0.30	11 (7%)	14	15	67, 125, 173, 195	0
2	E	170/176 (96%)	-0.19	3 (1%)	68	71	53, 90, 155, 199	0
3	B	153/161 (95%)	-0.09	5 (3%)	46	50	62, 93, 160, 220	0
4	G	166/170 (97%)	-0.29	2 (1%)	79	80	56, 83, 146, 196	0
All	All	637/664 (95%)	-0.08	21 (3%)	46	50	53, 96, 163, 220	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	134	HIS	5.3
3	B	126	ASN	4.3
2	E	320	TYR	3.8
1	A	118	LEU	3.6
3	B	123	GLU	3.2
3	B	32	HIS	3.1
3	B	185	TYR	3.0
2	E	219	TYR	2.9
2	E	346	SER	2.9
3	B	192	LYS	2.8
1	A	131	MET	2.7
1	A	99	LEU	2.4
4	G	219	TYR	2.4
1	A	78	TYR	2.4
1	A	188	LYS	2.4
1	A	149	ILE	2.2
1	A	136	PHE	2.2
1	A	174	ARG	2.1
1	A	157	LYS	2.1
1	A	77	ILE	2.0
4	G	351	TYR	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.