Full wwPDB X-ray Structure Validation Report (i)

May 15, 2020 – 05:43 pm BST

PDB ID	:	3RPT
Title	:	Crystal structure of the anti-HIV b12 scaffold protein
Authors	:	Chen, L.; Kwong, P.D.
Deposited on	:	2011-04-27
Resolution	:	1.30 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 following versions of software and data (see references (1)) were used in the production of this report:

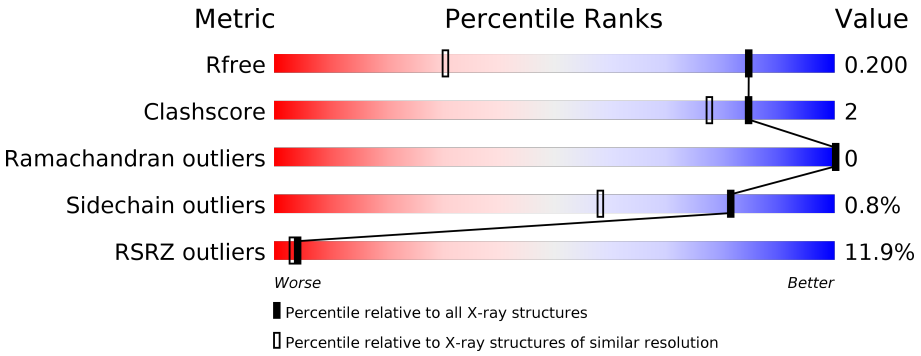
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.30 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	1058 (1.30-1.30)
Clashscore	141614	1101 (1.30-1.30)
Ramachandran outliers	138981	1058 (1.30-1.30)
Sidechain outliers	138945	1058 (1.30-1.30)
RSRZ outliers	127900	1029 (1.30-1.30)

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 on 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	282	13% 86% • 10%
1	X	282	8% 84% 5% 11%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7876 atoms, of which 3680 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 Endoglucanase E-2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	X	252	Total	C	H	N	O	S	6	0	0
			3680	1177	1800	337	359	7			
1	A	253	Total	C	H	N	O	S	22	13	0
			3834	1219	1880	355	373	7			

There are 102 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
X	41	HIS	PHE	SEE REMARK 999	UNP P26222
X	42	ASN	ALA	SEE REMARK 999	UNP P26222
X	43	GLN	HIS	SEE REMARK 999	UNP P26222
X	72	ASN	-	SEE REMARK 999	UNP P26222
X	73	PRO	-	SEE REMARK 999	UNP P26222
X	74	ASP	-	SEE REMARK 999	UNP P26222
X	75	PRO	-	SEE REMARK 999	UNP P26222
X	76	GLY	-	SEE REMARK 999	UNP P26222
X	77	GLY	-	SEE REMARK 999	UNP P26222
X	78	ASP	-	SEE REMARK 999	UNP P26222
X	79	MET	-	SEE REMARK 999	UNP P26222
X	80	ASP	-	SEE REMARK 999	UNP P26222
X	81	ASN	-	SEE REMARK 999	UNP P26222
X	82	GLY	-	SEE REMARK 999	UNP P26222
X	83	PHE	-	SEE REMARK 999	UNP P26222
X	84	GLU	-	SEE REMARK 999	UNP P26222
X	85	GLU	-	SEE REMARK 999	UNP P26222
X	86	GLY	-	SEE REMARK 999	UNP P26222
X	87	LYS	-	SEE REMARK 999	UNP P26222
X	88	GLN	-	SEE REMARK 999	UNP P26222
X	107	ASP	-	SEE REMARK 999	UNP P26222
X	108	PRO	-	SEE REMARK 999	UNP P26222
X	109	GLY	-	SEE REMARK 999	UNP P26222
X	110	GLY	-	SEE REMARK 999	UNP P26222
X	111	SER	-	SEE REMARK 999	UNP P26222

Continued on next page...

Continued from previous page...

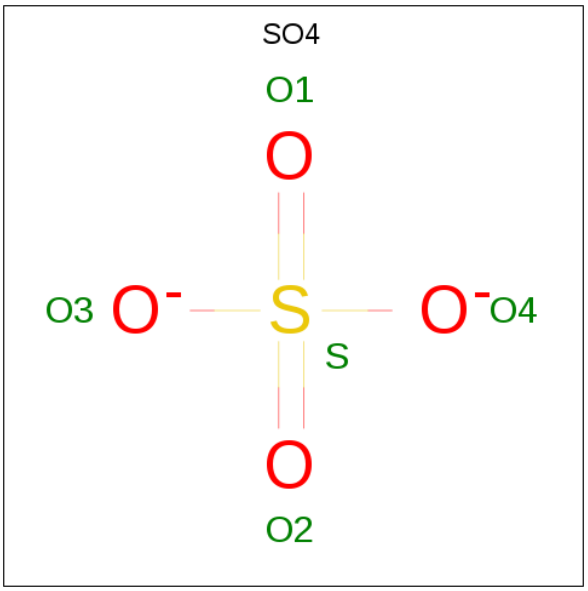
Chain	Residue	Modelled	Actual	Comment	Reference
X	112	GLY	-	SEE REMARK 999	UNP P26222
X	113	GLY	-	SEE REMARK 999	UNP P26222
X	114	ASP	-	SEE REMARK 999	UNP P26222
X	115	PRO	-	SEE REMARK 999	UNP P26222
X	116	GLU	-	SEE REMARK 999	UNP P26222
X	117	ILE	-	SEE REMARK 999	UNP P26222
X	118	ALA	-	SEE REMARK 999	UNP P26222
X	119	GLU	-	SEE REMARK 999	UNP P26222
X	120	ALA	-	SEE REMARK 999	UNP P26222
X	121	ALA	-	SEE REMARK 999	UNP P26222
X	122	TRP	-	SEE REMARK 999	UNP P26222
X	123	ARG	-	SEE REMARK 999	UNP P26222
X	124	PHE	-	SEE REMARK 999	UNP P26222
X	125	ALA	-	SEE REMARK 999	UNP P26222
X	158	ALA	SER	SEE REMARK 999	UNP P26222
X	159	ALA	TRP	SEE REMARK 999	UNP P26222
X	162	ARG	GLN	SEE REMARK 999	UNP P26222
X	219	SER	TRP	SEE REMARK 999	UNP P26222
X	275	LEU	-	EXPRESSION TAG	UNP P26222
X	276	GLU	-	EXPRESSION TAG	UNP P26222
X	277	HIS	-	EXPRESSION TAG	UNP P26222
X	278	HIS	-	EXPRESSION TAG	UNP P26222
X	279	HIS	-	EXPRESSION TAG	UNP P26222
X	280	HIS	-	EXPRESSION TAG	UNP P26222
X	281	HIS	-	EXPRESSION TAG	UNP P26222
X	282	HIS	-	EXPRESSION TAG	UNP P26222
A	41	HIS	PHE	SEE REMARK 999	UNP P26222
A	42	ASN	ALA	SEE REMARK 999	UNP P26222
A	43	GLN	HIS	SEE REMARK 999	UNP P26222
A	72	ASN	-	SEE REMARK 999	UNP P26222
A	73	PRO	-	SEE REMARK 999	UNP P26222
A	74	ASP	-	SEE REMARK 999	UNP P26222
A	75	PRO	-	SEE REMARK 999	UNP P26222
A	76	GLY	-	SEE REMARK 999	UNP P26222
A	77	GLY	-	SEE REMARK 999	UNP P26222
A	78	ASP	-	SEE REMARK 999	UNP P26222
A	79	MET	-	SEE REMARK 999	UNP P26222
A	80	ASP	-	SEE REMARK 999	UNP P26222
A	81	ASN	-	SEE REMARK 999	UNP P26222
A	82	GLY	-	SEE REMARK 999	UNP P26222
A	83	PHE	-	SEE REMARK 999	UNP P26222
A	84	GLU	-	SEE REMARK 999	UNP P26222

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	85	GLU	-	SEE REMARK 999	UNP P26222
A	86	GLY	-	SEE REMARK 999	UNP P26222
A	87	LYS	-	SEE REMARK 999	UNP P26222
A	88	GLN	-	SEE REMARK 999	UNP P26222
A	107	ASP	-	SEE REMARK 999	UNP P26222
A	108	PRO	-	SEE REMARK 999	UNP P26222
A	109	GLY	-	SEE REMARK 999	UNP P26222
A	110	GLY	-	SEE REMARK 999	UNP P26222
A	111	SER	-	SEE REMARK 999	UNP P26222
A	112	GLY	-	SEE REMARK 999	UNP P26222
A	113	GLY	-	SEE REMARK 999	UNP P26222
A	114	ASP	-	SEE REMARK 999	UNP P26222
A	115	PRO	-	SEE REMARK 999	UNP P26222
A	116	GLU	-	SEE REMARK 999	UNP P26222
A	117	ILE	-	SEE REMARK 999	UNP P26222
A	118	ALA	-	SEE REMARK 999	UNP P26222
A	119	GLU	-	SEE REMARK 999	UNP P26222
A	120	ALA	-	SEE REMARK 999	UNP P26222
A	121	ALA	-	SEE REMARK 999	UNP P26222
A	122	TRP	-	SEE REMARK 999	UNP P26222
A	123	ARG	-	SEE REMARK 999	UNP P26222
A	124	PHE	-	SEE REMARK 999	UNP P26222
A	125	ALA	-	SEE REMARK 999	UNP P26222
A	158	ALA	SER	SEE REMARK 999	UNP P26222
A	159	ALA	TRP	SEE REMARK 999	UNP P26222
A	162	ARG	GLN	SEE REMARK 999	UNP P26222
A	219	SER	TRP	SEE REMARK 999	UNP P26222
A	275	LEU	-	EXPRESSION TAG	UNP P26222
A	276	GLU	-	EXPRESSION TAG	UNP P26222
A	277	HIS	-	EXPRESSION TAG	UNP P26222
A	278	HIS	-	EXPRESSION TAG	UNP P26222
A	279	HIS	-	EXPRESSION TAG	UNP P26222
A	280	HIS	-	EXPRESSION TAG	UNP P26222
A	281	HIS	-	EXPRESSION TAG	UNP P26222
A	282	HIS	-	EXPRESSION TAG	UNP P26222

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0

- Molecule 3 is water.

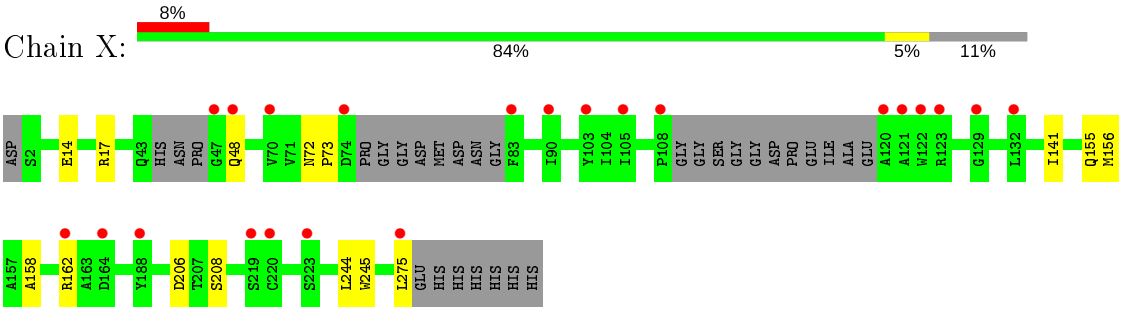
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	X	185	Total O 185 185	0	0
3	A	172	Total O 172 172	0	0

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 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.

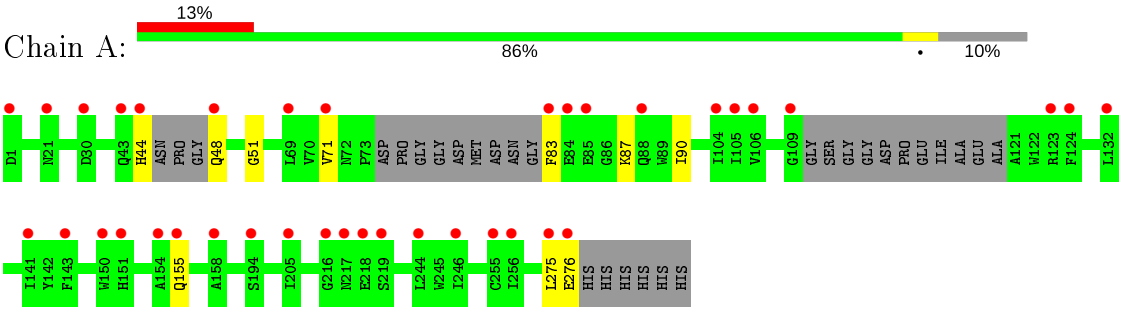
- Molecule 1: Endoglucanase E-2

Chain X:



- Molecule 1: Endoglucanase E-2

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	45.87Å 46.24Å 62.28Å 78.57° 81.80° 61.11°	Depositor
Resolution (Å)	26.21 – 1.30 26.68 – 1.30	Depositor EDS
% Data completeness (in resolution range)	75.5 (26.21-1.30) 75.5 (26.68-1.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.39 (at 1.30Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, R_{free}	0.173 , 0.202 0.170 , 0.200	Depositor DCC
R_{free} test set	4319 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	11.6	Xtriage
Anisotropy	0.681	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.46 , 62.5	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	0.95	EDS
Total number of atoms	7876	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.70% 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

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.49	0/2048	0.62	0/2794
1	X	0.52	0/1926	0.63	0/2627
All	All	0.51	0/3974	0.62	0/5421

There are no bond length outliers.

There are no bond angle outliers.

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	A	1954	1880	1836	7	0
1	X	1880	1800	1800	10	0
2	A	5	0	0	0	0
3	A	172	0	0	0	0
3	X	185	0	0	1	0
All	All	4196	3680	3636	17	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 2.

All (17) 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:83:PHE:CE2	1:A:87:LYS:HE3	2.25	0.71
1:A:48:GLN:HG2	1:A:51:GLY:H	1.54	0.71
1:X:156:MET:HE2	1:X:156:MET:HA	1.74	0.68
1:A:275:LEU:O	1:A:276:GLU:HB3	2.09	0.52
1:X:275:LEU:C	1:X:275:LEU:HD12	2.29	0.52
1:X:14:GLU:CD	1:X:17:ARG:NH2	2.64	0.51
1:X:158:ALA:O	1:X:162:ARG:HG2	2.14	0.48
1:X:156:MET:HE2	1:X:156:MET:CA	2.38	0.48
1:A:83:PHE:CE2	1:A:87:LYS:CE	2.95	0.47
1:X:72:ASN:N	1:X:73:PRO:CD	2.78	0.45
1:X:141:ILE:HD12	1:X:141:ILE:N	2.35	0.42
1:A:83:PHE:CZ	1:A:87:LYS:HE3	2.55	0.41
1:X:155:GLN:NE2	3:X:331:HOH:O	2.54	0.41
1:X:206:ASP:HA	1:X:244:LEU:O	2.21	0.41
1:A:71:VAL:HG11	1:A:90:ILE:HD13	2.03	0.40
1:X:208:SER:HA	1:X:245:TRP:CD1	2.56	0.40
1:A:48:GLN:O	1:A:48:GLN:HG2	2.20	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	258/282 (92%)	254 (98%)	4 (2%)	0	100	100
1	X	244/282 (86%)	238 (98%)	6 (2%)	0	100	100
All	All	502/564 (89%)	492 (98%)	10 (2%)	0	100	100

There are no Ramachandran outliers to report.

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 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	200/209 (96%)	198 (99%)	2 (1%)	76 48
1	X	187/209 (90%)	186 (100%)	1 (0%)	88 69
All	All	387/418 (93%)	384 (99%)	3 (1%)	81 58

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	X	48	GLN
1	A	44	HIS
1	A	155	GLN

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

Mol	Chain	Res	Type
1	A	72	ASN

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

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	SO4	A	283	-	4,4,4	0.25	0	6,6,6	0.44	0

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.

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	A	253/282 (89%)	0.85	38 (15%)	2	2	13, 21, 43, 61	6 (2%)
1	X	252/282 (89%)	0.66	22 (8%)	10	7	13, 20, 44, 58	6 (2%)
All	All	505/564 (89%)	0.76	60 (11%)	4	3	13, 20, 44, 61	12 (2%)

All (60) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	83	PHE	13.7
1	X	275	LEU	8.2
1	X	120	ALA	6.6
1	A	216	GLY	6.5
1	X	83	PHE	6.3
1	A	109	GLY	5.2
1	X	47	GLY	5.1
1	A	276	GLU	5.0
1	A	85	GLU	4.6
1	A	84	GLU	4.1
1	X	123	ARG	4.0
1	A	48	GLN	3.9
1	A	275	LEU	3.9
1	A	154	ALA	3.8
1	A	143	PHE	3.8
1	X	103	TYR	3.7
1	A	123	ARG	3.5
1	A	43	GLN	3.3
1	A	219	SER	3.3
1	A	255	CYS	3.3
1	X	48	GLN	3.2
1	A	155	GLN	3.1
1	A	158	ALA	3.1
1	X	162	ARG	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	218	GLU	3.0
1	X	105	ILE	3.0
1	X	132	LEU	3.0
1	A	217	ASN	3.0
1	A	106	VAL	2.9
1	A	1	ASP	2.8
1	A	194	SER	2.8
1	A	88	GLN	2.8
1	X	121	ALA	2.7
1	A	30	ASP	2.6
1	X	223	SER	2.6
1	A	44	HIS	2.6
1	X	74	ASP	2.6
1	A	132	LEU	2.5
1	X	122	TRP	2.5
1	A	71	VAL	2.5
1	A	151	HIS	2.4
1	X	220	CYS	2.4
1	A	150	TRP	2.3
1	X	164	ASP	2.3
1	A	256	ILE	2.2
1	A	69	LEU	2.2
1	A	105	ILE	2.2
1	A	21[A]	ASN	2.2
1	A	244	LEU	2.2
1	A	104	ILE	2.2
1	A	205	ILE	2.2
1	X	219	SER	2.1
1	A	141[A]	ILE	2.1
1	A	246	ILE	2.1
1	X	129	GLY	2.1
1	X	70	VAL	2.1
1	X	188	TYR	2.0
1	A	124	PHE	2.0
1	X	108	PRO	2.0
1	X	90	ILE	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 carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	SO4	A	283	5/5	0.91	0.15	32,35,45,52	0

6.5 Other polymers (i)

There are no such residues in this entry.