



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 12:40 AM GMT

PDB ID : 2BE2
Title : Crystal structure of HIV-1 reverse transcriptase (RT) in complex with R221239
Authors : Himmel, D.M.; Das, K.; Clark Jr., A.D.; Hughes, S.H.; Benjahad, A.; Oumouch, S.; Guillemont, J.; Coupa, S.; Poncelet, A.; Csoka, I.; Meyer, C.; Andries, K.; Nguyen, C.H.; Grierson, D.S.; Arnold, E.
Deposited on : 2005-10-21
Resolution : 2.43 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

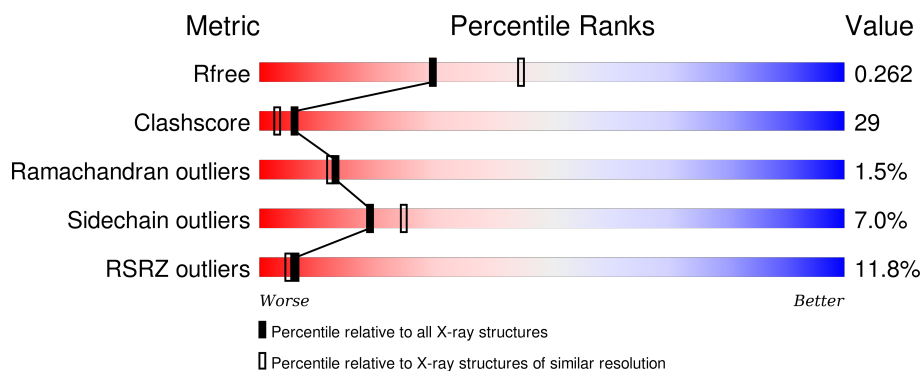
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION



The reported resolution of this entry is 2.43 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}	91344	1003 (2.46-2.42)
Clashscore	102246	1071 (2.46-2.42)
Ramachandran outliers	100387	1065 (2.46-2.42)
Sidechain outliers	100360	1065 (2.46-2.42)
RSRZ outliers	91569	1005 (2.46-2.42)

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. 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 $\leq 5\%$. 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	560	 10% 55% 38% 5%
2	B	430	 13% 59% 33% 6%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	SUC	A	2000	-	-	-	X
6	GOL	A	3001	-	-	X	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 8262 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 REVERSE TRANSCRIPTASE P66 SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	552	Total	C	N	O	S	0	0	0
			4498	2913	748	830	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	SER	CYS	ENGINEERED	UNP P03366

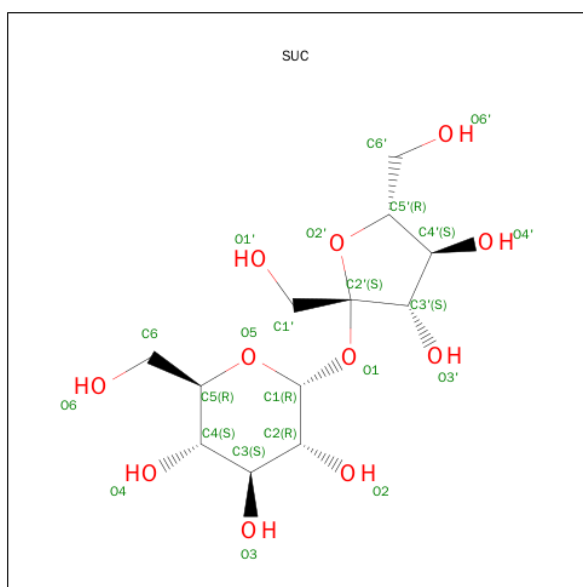
- Molecule 2 is a protein called REVERSE TRANSCRIPTASE P51 SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	426	Total	C	N	O	S	0	0	0
			3525	2297	584	637	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 3 is SUGAR (SUCROSE) (three-letter code: SUC) (formula: C₁₂H₂₂O₁₁).

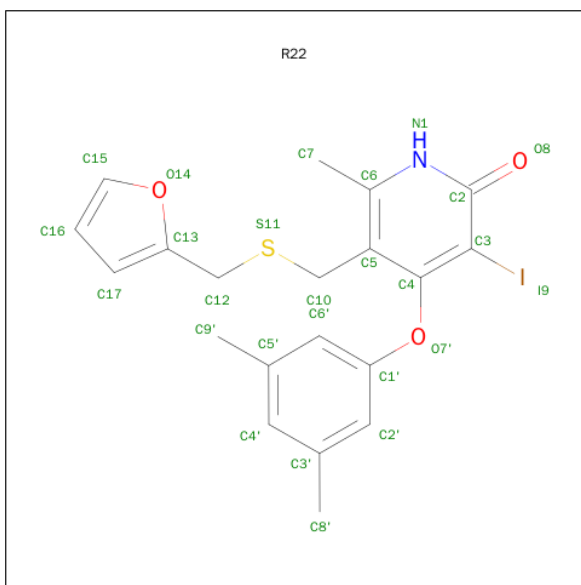


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			23	12	11		
3	B	1	Total	C	O	0	0
			23	12	11		

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Mn	0	0
			1	1		

- Molecule 5 is 4-(3,5-DIMETHYLPHENOXY)-5-(FURAN-2-YLMETHYLSULFANYLM ETHYL)-3-iodo-6-METHYLPYRIDIN-2(1H)-ONE (three-letter code: R22) (formula: C₂₀H₂₀INO₃S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	I	N	O	S	
			26	20	1	1	3	1	

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	C O		
			6	3 3	0	0
6	A	1	Total	C O		
			6	3 3	0	0

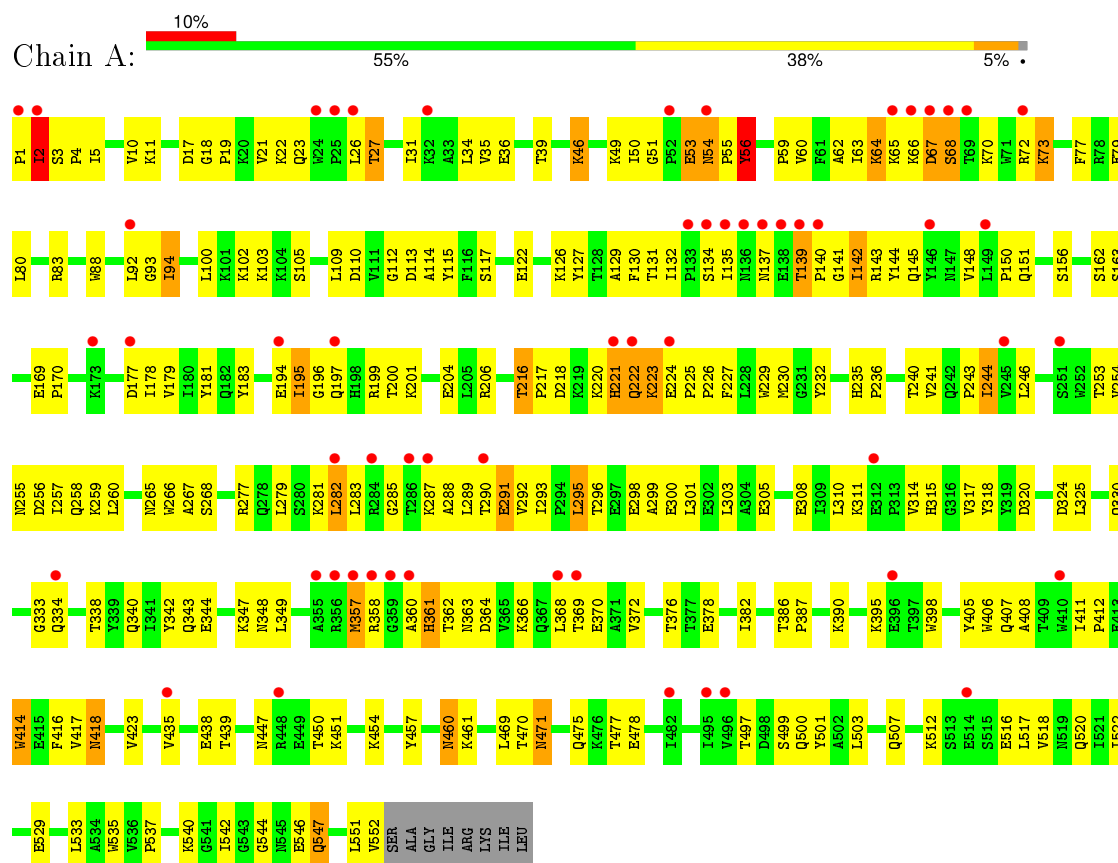
- Molecule 7 is water.

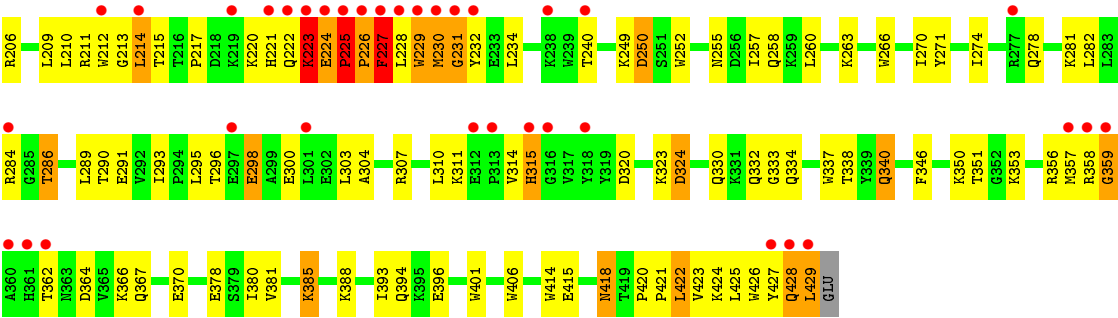
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	94	Total 94	O 94	0	0
7	B	60	Total 60	O 60	0	0

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 errors 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: REVERSE TRANSCRIPTASE P66 SUBUNIT





4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	223.85Å 68.80Å 104.43Å 90.00° 107.22° 90.00°	Depositor
Resolution (Å)	22.95 – 2.43 22.94 – 2.43	Depositor EDS
% Data completeness (in resolution range)	88.4 (22.95-2.43) 88.4 (22.94-2.43)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.73 (at 2.44Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.235 , 0.265 0.233 , 0.262	Depositor DCC
R_{free} test set	2554 reflections (5.29%)	DCC
Wilson B-factor (Å ²)	48.8	Xtriage
Anisotropy	0.108	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 69.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 53398 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	8262	wwPDB-VP
Average B, all atoms (Å ²)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: GOL, MN, SUC, R22

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	0.44	0/4616	0.75	5/6271 (0.1%)
2	B	0.48	0/3629	0.89	21/4932 (0.4%)
All	All	0.46	0/8245	0.81	26/11203 (0.2%)

There are no bond length outliers.

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	67	ASP	N-CA-C	9.43	136.46	111.00
2	B	359	GLY	N-CA-C	-8.98	90.64	113.10
2	B	89	GLU	N-CA-C	-8.94	86.86	111.00
2	B	214	LEU	CB-CA-C	8.07	125.53	110.20
1	A	285	GLY	N-CA-C	-7.93	93.28	113.10

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	4498	0	4560	277	0
2	B	3525	0	3562	198	0
3	A	23	0	22	2	0
3	B	23	0	22	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	1	0	0	0	0
5	A	26	0	20	3	0
6	A	12	0	16	5	0
7	A	94	0	0	5	0
7	B	60	0	0	3	0
All	All	8262	0	8202	465	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 29.

The worst 5 of 465 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:333:GLY:O	1:A:334:GLN:HG2	1.20	1.31
2:B:88:TRP:H	2:B:90:VAL:HG23	1.08	1.13
1:A:296:THR:HG23	1:A:299:ALA:H	1.09	1.10
1:A:255:ASN:OD1	1:A:259:LYS:HE3	1.52	1.09
1:A:26:LEU:HG	1:A:27:THR:H	1.10	1.08

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	550/560 (98%)	497 (90%)	45 (8%)	8 (2%)	13	12
2	B	424/430 (99%)	371 (88%)	46 (11%)	7 (2%)	11	10
All	All	974/990 (98%)	868 (89%)	91 (9%)	15 (2%)	13	12

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	53	GLU
1	A	54	ASN
1	A	2	ILE
1	A	195	ILE
2	B	221	HIS

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	494/500 (99%)	461 (93%)	33 (7%)	20	26
2	B	388/392 (99%)	359 (92%)	29 (8%)	17	22
All	All	882/892 (99%)	820 (93%)	62 (7%)	19	25

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	418	ASN
2	B	55	PRO
2	B	414	TRP
1	A	547	GLN
2	B	86	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	418	ASN
1	A	460	ASN
2	B	394	GLN
1	A	428	GLN
1	A	509	GLN

5.3.3 RNA ⓘ

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](#)

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 for Mogul analysis.

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	$\# Z > 2$
3	SUC	A	2000	-	24,24,24	1.42	5 (20%)	36,36,36	0.90	1 (2%)
6	GOL	A	3000	-	5,5,5	0.54	0	5,5,5	0.39	0
6	GOL	A	3001	-	5,5,5	0.54	0	5,5,5	0.39	0
5	R22	A	4002	-	25,28,28	1.99	10 (40%)	29,39,39	2.09	6 (20%)
3	SUC	B	2001	-	24,24,24	1.01	2 (8%)	36,36,36	1.10	1 (2%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	SUC	A	2000	-	-	0/12/51/51	0/2/2/2
6	GOL	A	3000	-	-	0/4/4/4	0/0/0/0
6	GOL	A	3001	-	-	0/4/4/4	0/0/0/0
5	R22	A	4002	-	-	0/9/10/10	0/2/3/3
3	SUC	B	2001	-	-	0/12/51/51	0/2/2/2

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	4002	R22	C3-I9	-3.45	2.02	2.10
5	A	4002	R22	C12-S11	-2.48	1.78	1.82
5	A	4002	R22	C5-C6	-2.20	1.35	1.38
3	A	2000	SUC	C1-C2	2.11	1.58	1.52
5	A	4002	R22	C2'-C3'	2.13	1.42	1.38

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	4002	R22	C3-C2-N1	-7.54	115.43	123.79
3	B	2001	SUC	C1-O5-C5	-4.61	104.80	113.75
5	A	4002	R22	C7-C6-N1	-2.79	111.92	116.34
3	A	2000	SUC	C2'-O1-C1	2.13	123.15	117.53
5	A	4002	R22	C12-C13-C17	2.30	135.06	129.15

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	2000	SUC	2	0
6	A	3001	GOL	5	0
5	A	4002	R22	3	0

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	552/560 (98%)	0.50	57 (10%) 9 7	28, 58, 106, 141	0
2	B	426/430 (99%)	0.73	58 (13%) 4 3	24, 55, 124, 150	0
All	All	978/990 (98%)	0.60	115 (11%) 6 5	24, 57, 118, 150	0

The worst 5 of 115 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	227	PHE	16.2
2	B	230	MET	12.3
2	B	358	ARG	9.8
2	B	4	PRO	9.8
2	B	229	TRP	9.5

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(\AA^2)	Q<0.9
6	GOL	A	3001	6/6	0.78	0.31	3.67	80,88,96,97	0
3	SUC	A	2000	23/23	0.77	0.28	2.31	92,102,106,114	0
3	SUC	B	2001	23/23	0.94	0.15	-0.23	29,47,53,68	0
5	R22	A	4002	26/26	0.96	0.17	-0.49	33,45,81,88	0
6	GOL	A	3000	6/6	0.64	0.20	-	76,80,88,90	0
4	MN	A	4001	1/1	0.97	0.12	-	40,40,40,40	0

6.5 Other polymers [i](#)

There are no such residues in this entry.