Full wwPDB X-ray Structure Validation Report (i)

Dec 18, 2023 – 10:51 am GMT

PDB ID	:	4CK2
Title	:	Interrogating HIV integrase for compounds that bind- a SAMPL challenge
Authors	:	Peat, T.S.
Deposited on	:	2013-12-24
Resolution	:	1.85 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:

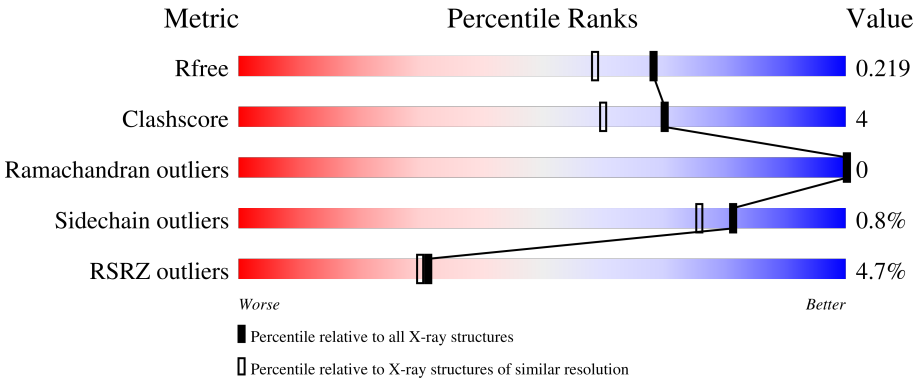
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance

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

The reported resolution of this entry is 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	ACT	A	1213	-	-	X	-
3	ACT	B	1214	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2619 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 INTEGRASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	150	Total	C	N	O	S	0	8	0
			1205	771	206	224	4			
1	B	150	Total	C	N	O	S	0	9	0
			1216	780	207	225	4			

There are 46 discrepancies between the modelled and reference sequences:

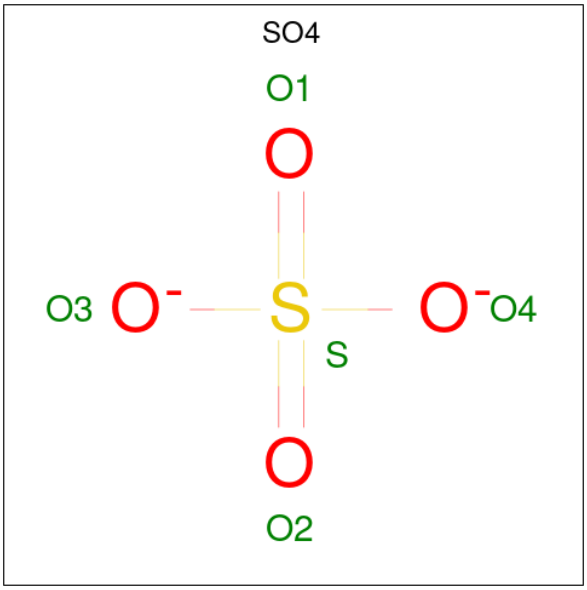
Chain	Residue	Modelled	Actual	Comment	Reference
A	30	MET	-	expression tag	UNP Q76353
A	31	GLY	-	expression tag	UNP Q76353
A	32	SER	-	expression tag	UNP Q76353
A	33	SER	-	expression tag	UNP Q76353
A	34	HIS	-	expression tag	UNP Q76353
A	35	HIS	-	expression tag	UNP Q76353
A	36	HIS	-	expression tag	UNP Q76353
A	37	HIS	-	expression tag	UNP Q76353
A	38	HIS	-	expression tag	UNP Q76353
A	39	HIS	-	expression tag	UNP Q76353
A	40	SER	-	expression tag	UNP Q76353
A	41	SER	-	expression tag	UNP Q76353
A	42	GLY	-	expression tag	UNP Q76353
A	43	LEU	-	expression tag	UNP Q76353
A	44	VAL	-	expression tag	UNP Q76353
A	45	PRO	-	expression tag	UNP Q76353
A	46	ARG	-	expression tag	UNP Q76353
A	47	GLY	-	expression tag	UNP Q76353
A	48	SER	-	expression tag	UNP Q76353
A	49	HIS	-	expression tag	UNP Q76353
A	56	SER	CYS	engineered mutation	UNP Q76353
A	139	ASP	PHE	engineered mutation	UNP Q76353
A	185	HIS	PHE	engineered mutation	UNP Q76353
B	30	MET	-	expression tag	UNP Q76353
B	31	GLY	-	expression tag	UNP Q76353

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	32	SER	-	expression tag	UNP Q76353
B	33	SER	-	expression tag	UNP Q76353
B	34	HIS	-	expression tag	UNP Q76353
B	35	HIS	-	expression tag	UNP Q76353
B	36	HIS	-	expression tag	UNP Q76353
B	37	HIS	-	expression tag	UNP Q76353
B	38	HIS	-	expression tag	UNP Q76353
B	39	HIS	-	expression tag	UNP Q76353
B	40	SER	-	expression tag	UNP Q76353
B	41	SER	-	expression tag	UNP Q76353
B	42	GLY	-	expression tag	UNP Q76353
B	43	LEU	-	expression tag	UNP Q76353
B	44	VAL	-	expression tag	UNP Q76353
B	45	PRO	-	expression tag	UNP Q76353
B	46	ARG	-	expression tag	UNP Q76353
B	47	GLY	-	expression tag	UNP Q76353
B	48	SER	-	expression tag	UNP Q76353
B	49	HIS	-	expression tag	UNP Q76353
B	56	SER	CYS	engineered mutation	UNP Q76353
B	139	ASP	PHE	engineered mutation	UNP Q76353
B	185	HIS	PHE	engineered mutation	UNP Q76353

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



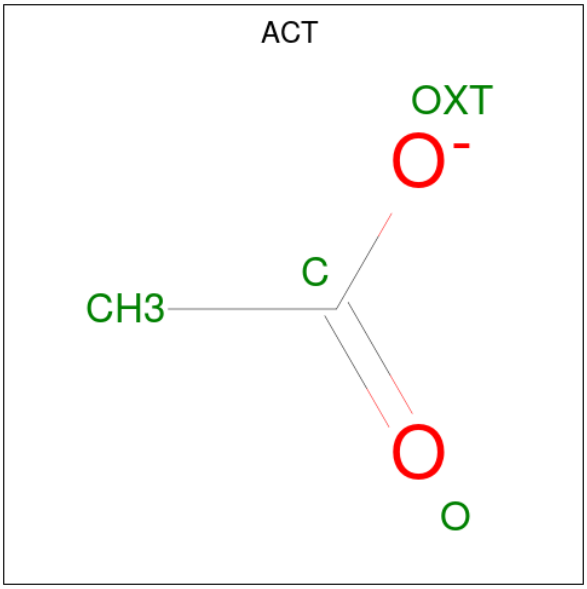
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0

Continued on next page...

Continued from previous page...

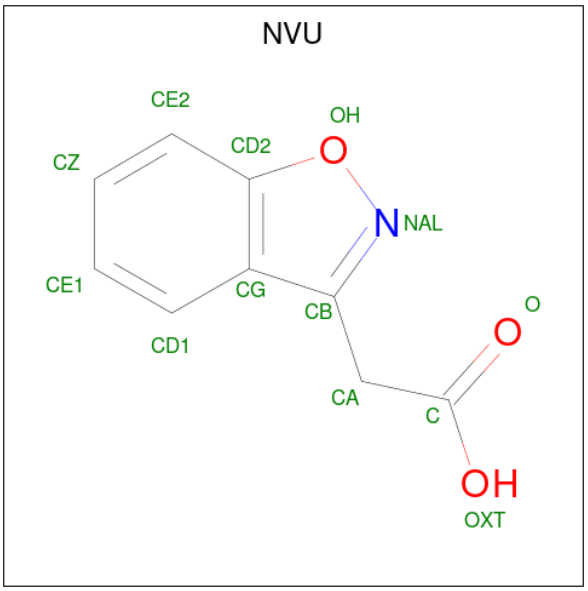
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0

- Molecule 4 is 2-(1,2-benzoxazol-3-yl)ethanoic acid (three-letter code: NVU) (formula: $C_9H_7NO_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C N O 13 9 1 3	0	0
4	A	1	Total C N O 13 9 1 3	0	0
4	B	1	Total C N O 13 9 1 3	0	0
4	B	1	Total C N O 13 9 1 3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	60	Total O 60 60	0	0
5	B	40	Total O 40 40	0	0

SEQUENCE-PLOTS INFOmissingINFO

3 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	71.48Å 71.48Å 66.77Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.41 – 1.85 45.40 – 1.85	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.41-1.85) 100.0 (45.40-1.85)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.02 (at 1.86Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.186 , 0.211 0.194 , 0.219	Depositor DCC
R_{free} test set	1647 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	25.2	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 29.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.030 for -h,-k,l 0.487 for h,-h-k,-l 0.031 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2619	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.08% 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.

4 Model quality

4.1 Standard geometry

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

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	0.31	0/1250	0.51	0/1695
1	B	0.32	0/1261	0.51	0/1710
All	All	0.32	0/2511	0.51	0/3405

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

4.2 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	1205	0	1218	6	0
1	B	1216	0	1237	8	0
2	A	15	0	0	0	0
2	B	15	0	0	0	0
3	A	8	0	6	5	0
3	B	8	0	6	5	0
4	A	26	0	12	0	0
4	B	26	0	12	0	0
5	A	60	0	0	2	0
5	B	40	0	0	1	0
All	All	2619	0	2491	18	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 4.

All (18) 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:A:1213:ACT:H2	5:B:2034:HOH:O	1.99	0.61
1:B:187:ARG:NH1	1:B:198:GLU:OE1	2.37	0.57
1:B:62:GLN:NE2	1:B:151[A]:ILE:HD11	2.20	0.57
1:B:141:ILE:HG23	1:B:143:TYR:CE1	2.40	0.57
1:B:115[A]:THR:HG21	1:B:137:GLN:HE21	1.72	0.54
5:A:2046:HOH:O	3:B:1214:ACT:H3	2.08	0.54
3:A:1213:ACT:H3	3:B:1214:ACT:C	2.40	0.52
1:A:141:ILE:HD12	1:A:142:PRO:HD2	1.93	0.50
1:B:141:ILE:HD12	1:B:142:PRO:HD2	1.93	0.49
1:A:115[A]:THR:HG21	1:A:137:GLN:OE1	2.13	0.48
3:A:1213:ACT:H3	3:B:1214:ACT:CH3	2.46	0.46
3:A:1213:ACT:O	3:B:1214:ACT:H1	2.17	0.44
1:A:145:PRO:O	1:A:148:GLN:HG3	2.18	0.44
1:A:103[B]:LYS:NZ	1:B:87[B]:GLU:OE2	2.42	0.43
1:A:87[B]:GLU:OE2	1:B:103[B]:LYS:NZ	2.43	0.42
1:A:127:LYS:HE3	5:A:2032:HOH:O	2.21	0.41
3:A:1213:ACT:H3	3:B:1214:ACT:OXT	2.21	0.41
1:B:181:PHE:O	1:B:185:HIS:HD2	2.03	0.41

There are no symmetry-related clashes.

4.3 Torsion angles (i)

4.3.1 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	154/183 (84%)	152 (99%)	2 (1%)	0	100	100
1	B	155/183 (85%)	152 (98%)	3 (2%)	0	100	100
All	All	309/366 (84%)	304 (98%)	5 (2%)	0	100	100

There are no Ramachandran outliers to report.

4.3.2 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	130/150 (87%)	129 (99%)	1 (1%)	81	76
1	B	132/150 (88%)	131 (99%)	1 (1%)	81	76
All	All	262/300 (87%)	260 (99%)	2 (1%)	81	76

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	141	ILE
1	B	141	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	155	ASN
1	B	137	GLN
1	B	155	ASN
1	B	185	HIS

4.3.3 RNA

There are no RNA molecules in this entry.

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

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

4.5 Carbohydrates

There are no monosaccharides in this entry.

4.6 Ligand geometry (i)

14 ligands are 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	$\#\|Z\| > 2$
3	ACT	B	1214	-	3,3,3	0.71	0	3,3,3	0.67	0
2	SO4	A	1210	-	4,4,4	0.24	0	6,6,6	0.05	0
2	SO4	B	1212	-	4,4,4	0.32	0	6,6,6	0.13	0
3	ACT	A	1213	-	3,3,3	0.78	0	3,3,3	0.88	0
4	NVU	B	1217	-	9,14,14	1.23	1 (11%)	7,19,19	2.07	3 (42%)
2	SO4	A	1211	-	4,4,4	0.30	0	6,6,6	0.07	0
3	ACT	A	1214	-	3,3,3	0.86	0	3,3,3	0.56	0
4	NVU	A	1215	-	9,14,14	1.24	1 (11%)	7,19,19	2.39	3 (42%)
3	ACT	B	1215	-	3,3,3	0.85	0	3,3,3	0.58	0
2	SO4	B	1213	-	4,4,4	0.32	0	6,6,6	0.14	0
4	NVU	A	1216	-	9,14,14	1.25	1 (11%)	7,19,19	2.01	4 (57%)
2	SO4	A	1212	-	4,4,4	0.32	0	6,6,6	0.11	0
2	SO4	B	1211	-	4,4,4	0.21	0	6,6,6	0.13	0
4	NVU	B	1216	-	9,14,14	1.23	1 (11%)	7,19,19	2.28	3 (42%)

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
4	NVU	B	1217	-	-	0/3/4/4	0/2/2/2
4	NVU	A	1216	-	-	0/3/4/4	0/2/2/2
4	NVU	B	1216	-	-	2/3/4/4	0/2/2/2
4	NVU	A	1215	-	-	2/3/4/4	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1215	NVU	CG-CD2	-2.53	1.38	1.43

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1216	NVU	CG-CD2	-2.53	1.38	1.43
4	B	1217	NVU	CG-CD2	-2.53	1.38	1.43
4	B	1216	NVU	CG-CD2	-2.47	1.38	1.43

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	1215	NVU	O-C-CA	-4.21	111.03	123.04
4	B	1216	NVU	O-C-CA	-3.87	112.00	123.04
4	A	1215	NVU	CE1-CD1-CG	-3.07	116.64	120.89
4	B	1217	NVU	O-C-CA	-3.01	114.45	123.04
4	B	1216	NVU	CE1-CD1-CG	-2.91	116.86	120.89
4	B	1217	NVU	CE1-CD1-CG	-2.75	117.08	120.89
4	A	1216	NVU	OXT-C-O	2.72	130.07	123.30
4	A	1216	NVU	O-C-CA	-2.60	115.62	123.04
4	B	1217	NVU	OXT-C-O	2.60	129.77	123.30
4	A	1216	NVU	CE1-CD1-CG	-2.59	117.30	120.89
4	B	1216	NVU	OXT-C-O	2.44	129.38	123.30
4	A	1215	NVU	OXT-C-O	2.38	129.23	123.30
4	A	1216	NVU	CD1-CG-CD2	-2.29	118.94	120.38

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	1216	NVU	OXT-C-CA-CB
4	B	1216	NVU	O-C-CA-CB
4	A	1215	NVU	O-C-CA-CB
4	A	1215	NVU	OXT-C-CA-CB

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	1214	ACT	5	0
3	A	1213	ACT	5	0

4.7 Other polymers (i)

There are no such residues in this entry.

4.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

5 Fit of model and data

5.1 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	150/183 (81%)	-0.11	7 (4%)	31	30	15, 24, 56, 82	0
1	B	150/183 (81%)	-0.09	7 (4%)	31	30	15, 24, 60, 83	1 (0%)
All	All	300/366 (81%)	-0.10	14 (4%)	31	30	15, 24, 60, 83	1 (0%)

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	143	TYR	9.6
1	A	143	TYR	8.0
1	A	142	PRO	5.4
1	B	142	PRO	5.2
1	A	194	TYR	3.5
1	B	144	ASN	3.5
1	A	144	ASN	3.3
1	B	194	TYR	3.1
1	B	209	GLN	2.6
1	B	56	SER	2.4
1	A	147	SER	2.3
1	A	209	GLN	2.3
1	B	147	SER	2.3
1	A	56	SER	2.0

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

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

5.3 Carbohydrates

There are no monosaccharides in this entry.

5.4 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
4	NVU	A	1216	13/13	0.56	0.23	43,48,50,52	0
4	NVU	B	1217	13/13	0.62	0.25	45,49,52,53	0
4	NVU	A	1215	13/13	0.85	0.15	34,35,40,44	0
4	NVU	B	1216	13/13	0.87	0.16	35,37,41,46	0
3	ACT	B	1214	4/4	0.91	0.10	25,28,29,31	4
3	ACT	B	1215	4/4	0.91	0.12	38,40,41,41	0
2	SO4	A	1211	5/5	0.91	0.15	44,48,50,50	5
2	SO4	B	1212	5/5	0.92	0.13	46,50,53,54	5
3	ACT	A	1213	4/4	0.92	0.12	27,29,29,32	4
3	ACT	A	1214	4/4	0.92	0.14	37,40,41,41	0
2	SO4	A	1212	5/5	0.95	0.11	34,38,40,42	5
2	SO4	B	1213	5/5	0.97	0.11	35,40,40,41	5
2	SO4	A	1210	5/5	0.98	0.14	30,31,33,35	5
2	SO4	B	1211	5/5	0.98	0.15	30,31,32,35	5

5.5 Other polymers (i)

There are no such residues in this entry.