Full wwPDB X-ray Structure Validation Report (i)

Jun 17, 2021 – 07:11 PM EDT

PDB ID	:	5RPI
Title	:	PanDDA analysis group deposition – Proteinase K crystal structure Apo26
Authors	:	Lima, G.M.A.; Talibov, V.; Benz, L.S.; Jagudin, E.; Mueller, U.
Deposited on	:	2020-09-23
Resolution	:	1.03 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:

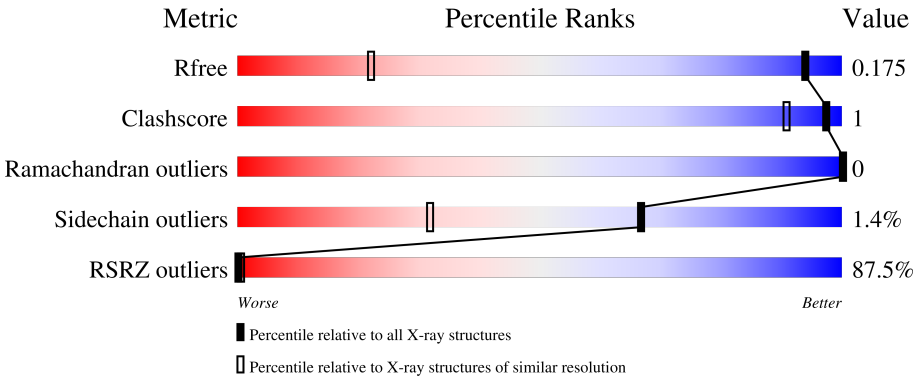
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.20
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.20

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 1.03 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	1596 (1.10-0.98)
Clashscore	141614	1677 (1.10-0.98)
Ramachandran outliers	138981	1591 (1.10-0.98)
Sidechain outliers	138945	1589 (1.10-0.98)
RSRZ outliers	127900	1557 (1.10-0.98)

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	279	87% 97% .

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2339 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.

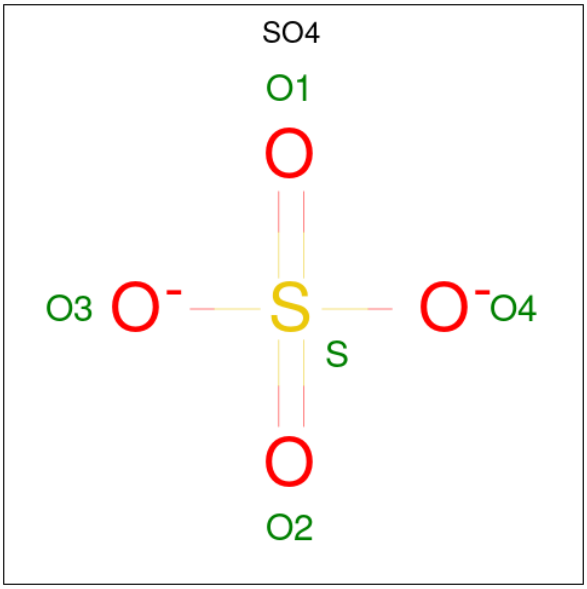
- Molecule 1 is a protein called Proteinase K.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	279	2023	1244	355	414	10	0	17	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	207	ASP	SER	conflict	UNP P06873

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



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

- Molecule 3 is water.

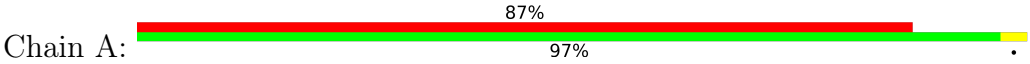
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	311	Total O 311 311	0	0

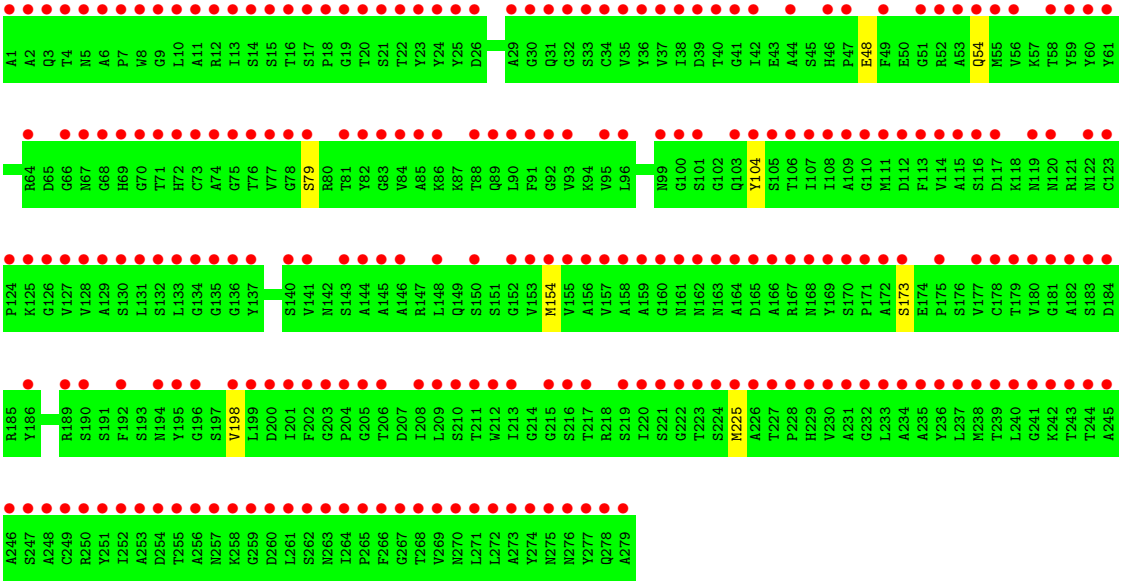
3 Residue-property plots (i)

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: Proteinase K

Chain A:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	67.97Å 67.97Å 102.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.06 – 1.03 56.61 – 1.03	Depositor EDS
% Data completeness (in resolution range)	90.3 (48.06-1.03) 90.3 (56.61-1.03)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 1.03Å)	Xtriage
Refinement program	PHENIX 1.19.1	Depositor
R, R_{free}	0.170 , 0.179 0.168 , 0.175	Depositor DCC
R_{free} test set	5281 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	6.3	Xtriage
Anisotropy	0.001	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 39.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2339	wwPDB-VP
Average B, all atoms (Å^2)	8.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.15% 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	0.30	0/2062	0.60	0/2804

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	2023	0	1900	3	0
2	A	5	0	0	0	0
3	A	311	0	0	1	1
All	All	2339	0	1900	3	1

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 1.

All (3) 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:54[A]:GLN:HG2	3:A:1111:HOH:O	1.99	0.62

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:173:SER:HA	1:A:198:VAL:HG21	1.94	0.50
1:A:48:GLU:HB3	1:A:79:SER:HB2	2.01	0.42

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:1118:HOH:O	3:A:1207:HOH:O[5_655]	1.81	0.39

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	277/279 (99%)	269 (97%)	8 (3%)	0	100 100

There are no Ramachandran outliers to report.

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	210/213 (99%)	207 (99%)	3 (1%)	67 32

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

Mol	Chain	Res	Type
1	A	104	TYR
1	A	154[A]	MET
1	A	225	MET

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)

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	1001	-	4,4,4	0.17	0	6,6,6	0.18	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 (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	A	279/279 (100%)	3.00	244 (87%)	0	0	4, 6, 10, 21	14 (5%)

All (244) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	279	ALA	9.7
1	A	198	VAL	5.6
1	A	13[A]	ILE	5.6
1	A	34[A]	CYS	5.3
1	A	64	ARG	5.3
1	A	278	GLN	5.2
1	A	201	ILE	4.9
1	A	212	TRP	4.9
1	A	8	TRP	4.7
1	A	178	CYS	4.7
1	A	84	VAL	4.7
1	A	38	ILE	4.7
1	A	93	VAL	4.7
1	A	73	CYS	4.7
1	A	249	CYS	4.7
1	A	252	ILE	4.6
1	A	237	LEU	4.6
1	A	77	VAL	4.6
1	A	180	VAL	4.5
1	A	123	CYS	4.5
1	A	99	ASN	4.3
1	A	90	LEU	4.3
1	A	127	VAL	4.3
1	A	82	TYR	4.2
1	A	213	ILE	4.2
1	A	114	VAL	4.2
1	A	36	TYR	4.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	61	TYR	4.2
1	A	272	LEU	4.2
1	A	42	ILE	4.2
1	A	266	PHE	4.2
1	A	169	TYR	4.1
1	A	202	PHE	4.1
1	A	24	TYR	4.1
1	A	37	VAL	4.1
1	A	59	TYR	4.0
1	A	131[A]	LEU	4.0
1	A	236	TYR	4.0
1	A	251	TYR	4.0
1	A	271	LEU	4.0
1	A	107	ILE	4.0
1	A	240	LEU	3.9
1	A	85	ALA	3.9
1	A	208	ILE	3.9
1	A	91[A]	PHE	3.9
1	A	25	TYR	3.8
1	A	35	VAL	3.8
1	A	95	VAL	3.8
1	A	155	VAL	3.8
1	A	128	VAL	3.8
1	A	141	VAL	3.8
1	A	113	PHE	3.8
1	A	60	TYR	3.8
1	A	115	ALA	3.8
1	A	21	SER	3.7
1	A	230	VAL	3.7
1	A	10	LEU	3.7
1	A	49	PHE	3.7
1	A	153	VAL	3.7
1	A	148	LEU	3.7
1	A	5	ASN	3.6
1	A	233	LEU	3.6
1	A	261	LEU	3.6
1	A	106	THR	3.6
1	A	182	ALA	3.6
1	A	177	VAL	3.6
1	A	220	ILE	3.6
1	A	104	TYR	3.6
1	A	269	VAL	3.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	23	TYR	3.5
1	A	186	TYR	3.5
1	A	192	PHE	3.5
1	A	54[A]	GLN	3.5
1	A	158	ALA	3.5
1	A	18	PRO	3.4
1	A	56	VAL	3.4
1	A	195	TYR	3.4
1	A	74	ALA	3.4
1	A	129	ALA	3.4
1	A	156	ALA	3.4
1	A	76	THR	3.4
1	A	6	ALA	3.4
1	A	164	ALA	3.4
1	A	16	THR	3.4
1	A	157	VAL	3.4
1	A	231	ALA	3.3
1	A	204	PRO	3.3
1	A	31[A]	GLN	3.3
1	A	96	LEU	3.3
1	A	274	TYR	3.3
1	A	133	LEU	3.3
1	A	206	THR	3.3
1	A	211	THR	3.3
1	A	227	THR	3.3
1	A	238[A]	MET	3.3
1	A	172	ALA	3.3
1	A	226	ALA	3.3
1	A	30	GLY	3.3
1	A	179	THR	3.3
1	A	89	GLN	3.2
1	A	29	ALA	3.2
1	A	152	GLY	3.2
1	A	234	ALA	3.2
1	A	246	ALA	3.2
1	A	256	ALA	3.2
1	A	81	THR	3.2
1	A	245	ALA	3.1
1	A	88	THR	3.1
1	A	273	ALA	3.1
1	A	22	THR	3.1
1	A	39	ASP	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	2	ALA	3.1
1	A	11	ALA	3.1
1	A	235	ALA	3.1
1	A	108	ILE	3.0
1	A	75	GLY	3.0
1	A	134	GLY	3.0
1	A	1	ALA	3.0
1	A	137	TYR	3.0
1	A	276	ASN	3.0
1	A	209	LEU	3.0
1	A	66	GLY	3.0
1	A	203	GLY	3.0
1	A	71	THR	3.0
1	A	112	ASP	3.0
1	A	103	GLN	3.0
1	A	126	GLY	3.0
1	A	196	GLY	3.0
1	A	244	THR	3.0
1	A	219[A]	SER	3.0
1	A	53	ALA	2.9
1	A	146	ALA	2.9
1	A	253	ALA	2.9
1	A	264	ILE	2.9
1	A	83	GLY	2.9
1	A	144	ALA	2.9
1	A	171	PRO	2.9
1	A	259	GLY	2.9
1	A	277	TYR	2.9
1	A	47	PRO	2.8
1	A	132	SER	2.8
1	A	168	ASN	2.8
1	A	143[A]	SER	2.8
1	A	7	PRO	2.8
1	A	109	ALA	2.8
1	A	228	PRO	2.8
1	A	163	ASN	2.8
1	A	239	THR	2.7
1	A	145	ALA	2.7
1	A	262	SER	2.7
1	A	32	GLY	2.7
1	A	181	GLY	2.7
1	A	268	THR	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	9	GLY	2.7
1	A	70	GLY	2.7
1	A	159	ALA	2.7
1	A	140[A]	SER	2.7
1	A	120	ASN	2.7
1	A	175	PRO	2.7
1	A	4	THR	2.7
1	A	68	GLY	2.7
1	A	217	THR	2.7
1	A	189	ARG	2.7
1	A	254	ASP	2.6
1	A	135	GLY	2.6
1	A	222	GLY	2.6
1	A	232	GLY	2.6
1	A	101	SER	2.6
1	A	40	THR	2.6
1	A	255	THR	2.6
1	A	154[A]	MET	2.6
1	A	166	ALA	2.6
1	A	248	ALA	2.6
1	A	161[A]	ASN	2.6
1	A	223	THR	2.6
1	A	14	SER	2.6
1	A	105	SER	2.6
1	A	221	SER	2.6
1	A	78	GLY	2.6
1	A	110	GLY	2.6
1	A	160	GLY	2.6
1	A	205	GLY	2.6
1	A	58	THR	2.6
1	A	199	LEU	2.6
1	A	210	SER	2.6
1	A	258	LYS	2.6
1	A	116	SER	2.5
1	A	183	SER	2.5
1	A	100	GLY	2.5
1	A	136	GLY	2.5
1	A	44	ALA	2.5
1	A	250[A]	ARG	2.5
1	A	184	ASP	2.5
1	A	19	GLY	2.5
1	A	130	SER	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	15	SER	2.4
1	A	79	SER	2.4
1	A	167	ARG	2.4
1	A	242[A]	LYS	2.4
1	A	20	THR	2.4
1	A	55	MET	2.4
1	A	225	MET	2.4
1	A	162	ASN	2.4
1	A	270	ASN	2.4
1	A	124	PRO	2.4
1	A	173	SER	2.4
1	A	216[A]	SER	2.4
1	A	224	SER	2.4
1	A	200	ASP	2.3
1	A	92	GLY	2.3
1	A	215	GLY	2.3
1	A	122	ASN	2.3
1	A	51	GLY	2.3
1	A	125	LYS	2.3
1	A	119	ASN	2.3
1	A	33	SER	2.3
1	A	275	ASN	2.3
1	A	17	SER	2.2
1	A	190	SER	2.2
1	A	41	GLY	2.2
1	A	72	HIS	2.2
1	A	229	HIS	2.2
1	A	265	PRO	2.2
1	A	170	SER	2.2
1	A	194	ASN	2.2
1	A	69	HIS	2.2
1	A	111	MET	2.2
1	A	243	THR	2.2
1	A	150	SER	2.2
1	A	247	SER	2.2
1	A	3	GLN	2.1
1	A	267	GLY	2.1
1	A	260	ASP	2.1
1	A	46	HIS	2.1
1	A	26	ASP	2.1
1	A	67	ASN	2.1
1	A	257	ASN	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	52	ARG	2.1
1	A	165	ASP	2.1
1	A	263	ASN	2.1
1	A	86	LYS	2.0
1	A	241	GLY	2.0
1	A	12	ARG	2.0
1	A	117	ASP	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)

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	1001	5/5	0.93	0.17	7,7,11,13	0

6.5 Other polymers (i)

There are no such residues in this entry.