



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 14, 2018 – 02:12 pm GMT

PDB ID : 5VQZ
Title : Crystal Structure of HIV-1 Reverse Transcriptase (K103N, Y181C) Variant in Complex with 2-chloro-N-(6-cyano-3-(2-(2-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)ethoxy)phenoxy)-4-methylnaphthalen-1-yl)-N-methylacetamide (JLJ686), a Non-nucleoside Inhibitor
Authors : Buckingham, A.B.; Chan, A.H.; Anderson, K.S.
Deposited on : 2017-05-09
Resolution : 2.23 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 ⓘ symbol.

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

MolProbity	:	4.02b-467
Mogul	:	1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix)	:	1.13
EDS	:	trunk31020
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk31020

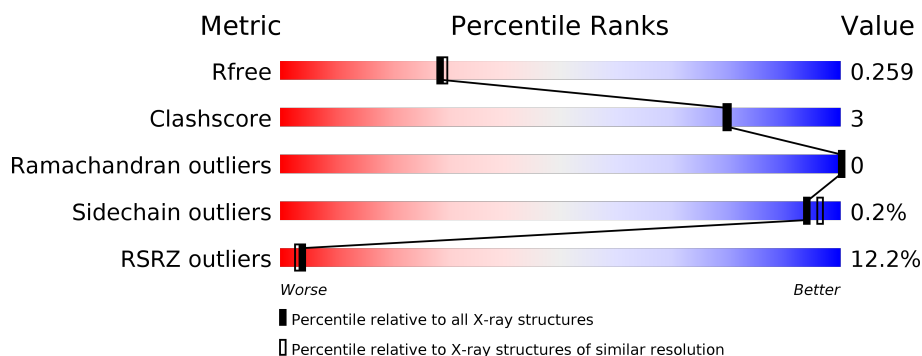
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.23 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}	111664	2027 (2.26-2.22)
Clashscore	122126	2170 (2.26-2.22)
Ramachandran outliers	120053	2129 (2.26-2.22)
Sidechain outliers	120020	2130 (2.26-2.22)
RSRZ outliers	108989	1991 (2.26-2.22)

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	557	<div> <div>9%</div> <div>87%</div> <div>11%</div> <div>.</div> </div>
2	B	428	<div> <div>16%</div> <div>88%</div> <div>6%</div> <div>6%</div> </div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7961 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/ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	547	Total	C	N	O	S	0	0	0
			4412	2856	729	818	9			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	103	ASN	LYS	engineered mutation	UNP P03366
A	172	ALA	LYS	engineered mutation	UNP P03366
A	173	ALA	LYS	engineered mutation	UNP P03366
A	181	CYS	TYR	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366

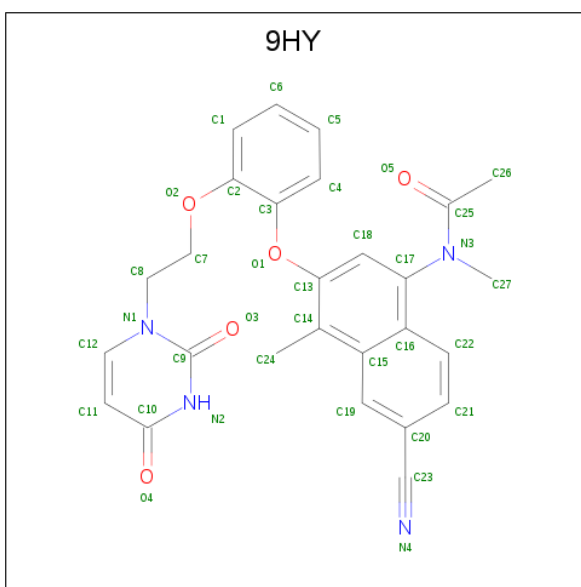
- Molecule 2 is a protein called p51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	401	Total	C	N	O	S	0	0	0
			3323	2163	551	603	6			

There is a discrepancy between the modelled and reference sequences:

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

- Molecule 3 is N-(6-cyano-3-{2-[2-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)ethoxy]phenoxy}-4-methylnaphthalen-1-yl)-N-methylacetamide (three-letter code: 9HY) (formula: C₂₇H₂₄N₄O₅).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			36	27	4	5		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

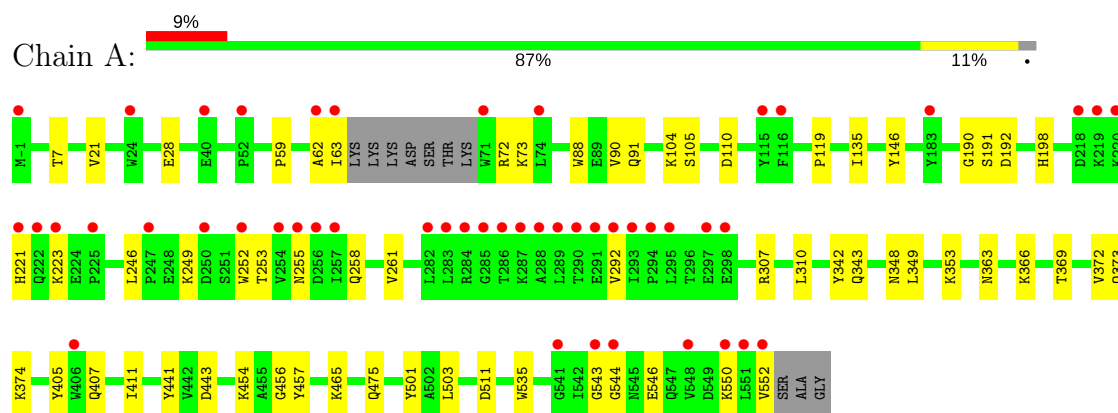
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	115	Total 115	O 115	0	0
5	B	70	Total 70	O 70	0	0

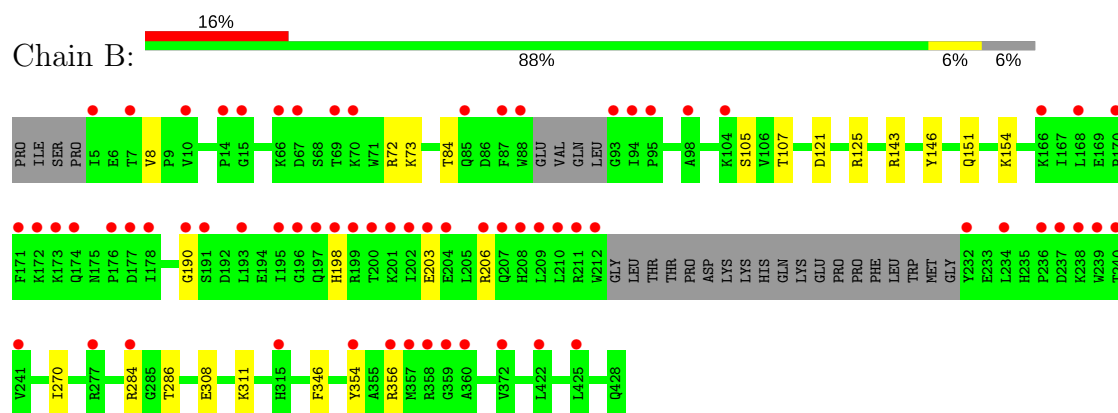
3 Residue-property plots [i](#)

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: Reverse transcriptase/ribonuclease H



- Molecule 2: p51 RT



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	162.12Å 73.33Å 108.69Å 90.00° 99.45° 90.00°	Depositor
Resolution (Å)	43.12 – 2.23 43.12 – 2.23	Depositor EDS
% Data completeness (in resolution range)	98.9 (43.12-2.23) 94.1 (43.12-2.23)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.21 (at 2.22Å)	Xtriage
Refinement program	PHENIX 1.11.1 _2575	Depositor
R, R_{free}	0.227 , 0.259 0.227 , 0.259	Depositor DCC
R_{free} test set	2000 reflections (3.29%)	wwPDB-VP
Wilson B-factor (Å ²)	43.1	Xtriage
Anisotropy	0.029	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 43.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	7961	wwPDB-VP
Average B, all atoms (Å ²)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.11% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 9HY, 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	0.24	0/4528	0.42	0/6166
2	B	0.24	0/3417	0.42	0/4639
All	All	0.24	0/7945	0.42	0/10805

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 [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	4412	0	4418	36	0
2	B	3323	0	3352	13	0
3	A	36	0	0	0	0
4	A	5	0	0	0	0
5	A	115	0	0	0	0
5	B	70	0	0	0	0
All	All	7961	0	7770	46	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 3.

All (46) 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:253:THR:HG22	1:A:292:VAL:HG22	1.73	0.70
2:B:73:LYS:NZ	2:B:146:TYR:OH	2.26	0.67
1:A:544:GLY:HA2	2:B:286:THR:HG22	1.78	0.66
1:A:543:GLY:HA3	2:B:284:ARG:HA	1.80	0.62
2:B:308:GLU:HA	2:B:311:LYS:HE2	1.82	0.61
1:A:110:ASP:HB2	1:A:221:HIS:HE1	1.64	0.60
1:A:443:ASP:HB3	1:A:550:LYS:HD3	1.84	0.60
1:A:28:GLU:HG3	1:A:135:ILE:HD12	1.84	0.58
2:B:72:ARG:NH2	2:B:151:GLN:OE1	2.37	0.58
1:A:191:SER:OG	1:A:198:HIS:ND1	2.37	0.54
1:A:110:ASP:HB2	1:A:221:HIS:CE1	2.42	0.54
1:A:90:VAL:HG23	1:A:91:GLN:HG2	1.89	0.54
2:B:203:GLU:HA	2:B:206:ARG:HG2	1.89	0.54
1:A:105:SER:O	1:A:190:GLY:HA2	2.09	0.53
1:A:342:TYR:HB3	1:A:348:ASN:HA	1.92	0.52
1:A:343:GLN:HG3	1:A:349:LEU:HD11	1.93	0.51
1:A:372:VAL:HG11	1:A:411:ILE:HG13	1.92	0.51
1:A:21:VAL:HB	1:A:59:PRO:HD3	1.93	0.51
1:A:7:THR:HG22	1:A:119:PRO:HG2	1.94	0.50
1:A:88:TRP:CD1	2:B:143:ARG:HD2	2.47	0.50
1:A:73:LYS:NZ	1:A:146:TYR:OH	2.40	0.49
1:A:405:TYR:CE2	1:A:407:GLN:HB2	2.47	0.49
1:A:454:LYS:HZ2	1:A:552:VAL:HB	1.77	0.48
1:A:353:LYS:O	1:A:374:LYS:NZ	2.44	0.48
1:A:249:LYS:HB2	1:A:252:TRP:CE2	2.48	0.47
2:B:84:THR:HB	2:B:154:LYS:HE2	1.96	0.47
2:B:354:TYR:CE1	2:B:356:ARG:HB3	2.51	0.46
1:A:104:LYS:HB2	1:A:192:ASP:HA	1.97	0.46
1:A:475:GLN:HB3	1:A:501:TYR:CE2	2.51	0.46
1:A:363:ASN:ND2	1:A:366:LYS:HE3	2.31	0.45
2:B:105:SER:O	2:B:190:GLY:HA2	2.15	0.45
1:A:363:ASN:HA	1:A:511:ASP:OD1	2.16	0.45
1:A:221:HIS:HD2	1:A:223:LYS:H	1.64	0.44
1:A:253:THR:OG1	1:A:255:ASN:OD1	2.33	0.44
1:A:246:LEU:HD11	1:A:310:LEU:HD12	2.00	0.43
1:A:441:TYR:O	1:A:457:TYR:HA	2.19	0.43
1:A:62:ALA:HA	1:A:72:ARG:O	2.18	0.43
1:A:258:GLN:HA	1:A:261:VAL:HG12	2.01	0.43
1:A:544:GLY:O	1:A:546:GLU:HG2	2.19	0.43
2:B:270:ILE:HG12	2:B:346:PHE:HB3	1.99	0.42
1:A:246:LEU:HD12	1:A:307:ARG:HG2	2.02	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:503:LEU:HD22	1:A:535:TRP:HB2	2.01	0.42
1:A:369:THR:O	1:A:373:GLN:HG2	2.19	0.42
2:B:107:THR:OG1	2:B:198:HIS:NE2	2.41	0.41
2:B:121:ASP:O	2:B:125:ARG:HG3	2.20	0.41
1:A:456:GLY:HA3	1:A:465:LYS:O	2.21	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	543/557 (98%)	533 (98%)	10 (2%)	0	100	100
2	B	395/428 (92%)	384 (97%)	11 (3%)	0	100	100
All	All	938/985 (95%)	917 (98%)	21 (2%)	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	480/495 (97%)	479 (100%)	1 (0%)	94	96
2	B	365/390 (94%)	364 (100%)	1 (0%)	93	95
All	All	845/885 (96%)	843 (100%)	2 (0%)	94	96

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

Mol	Chain	Res	Type
1	A	63	ILE
2	B	8	VAL

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	221	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

2 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	9HY	A	601	1	34,39,39	1.13	3 (8%)	41,55,55	2.41	5 (12%)
4	SO4	A	602	-	4,4,4	0.18	0	6,6,6	0.07	0

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	9HY	A	601	1	-	0/20/20/20	0/4/4/4
4	SO4	A	602	-	-	0/0/0/0	0/0/0/0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	601	9HY	C8-N1	-2.22	1.45	1.48
3	A	601	9HY	C22-C21	2.12	1.41	1.36
3	A	601	9HY	C10-N2	3.57	1.39	1.33

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	601	9HY	C11-C10-N2	-3.48	115.10	123.17
3	A	601	9HY	C18-C13-C14	-2.17	119.77	122.39
3	A	601	9HY	C8-N1-C12	2.25	122.32	118.18
3	A	601	9HY	C16-C17-N3	3.76	122.55	119.00
3	A	601	9HY	C10-N2-C9	13.49	125.75	114.14

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, 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	547/557 (98%)	0.65	49 (8%)	9 9	30, 52, 92, 134	0
2	B	401/428 (93%)	1.05	67 (16%)	1 1	31, 53, 105, 122	0
All	All	948/985 (96%)	0.82	116 (12%)	4 3	30, 52, 99, 134	0

All (116) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	289	LEU	10.3
2	B	241	VAL	10.1
2	B	358	ARG	9.8
1	A	288	ALA	7.8
1	A	287	LYS	7.8
2	B	357	MET	7.7
1	A	290	THR	7.5
2	B	240	THR	7.1
1	A	24	TRP	6.9
1	A	548	VAL	6.5
2	B	356	ARG	6.3
1	A	292	VAL	6.2
1	A	286	THR	6.1
2	B	200	THR	6.1
1	A	221	HIS	6.1
2	B	207	GLN	5.9
2	B	359	GLY	5.9
1	A	552	VAL	5.7
2	B	94	ILE	5.6
2	B	88	TRP	5.5
2	B	168	LEU	5.5
2	B	212	TRP	5.5
1	A	219	LYS	5.4
2	B	206	ARG	5.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	198	HIS	5.3
1	A	116	PHE	5.2
1	A	285	GLY	5.2
2	B	238	LYS	5.0
1	A	220	LYS	4.9
2	B	202	ILE	4.8
2	B	69	THR	4.7
1	A	551	LEU	4.7
2	B	67	ASP	4.6
2	B	284	ARG	4.6
2	B	196	GLY	4.6
2	B	85	GLN	4.5
2	B	173	LYS	4.4
1	A	295	LEU	4.4
2	B	195	ILE	4.3
1	A	257	ILE	4.2
1	A	541	GLY	4.2
2	B	211	ARG	4.1
1	A	218	ASP	4.0
2	B	237	ASP	3.9
2	B	87	PHE	3.9
2	B	170	PRO	3.7
2	B	178	ILE	3.7
2	B	174	GLN	3.7
1	A	550	LYS	3.6
2	B	166	LYS	3.6
2	B	197	GLN	3.5
2	B	209	LEU	3.5
1	A	71	TRP	3.5
2	B	239	TRP	3.5
2	B	354	TYR	3.5
1	A	256	ASP	3.5
2	B	93	GLY	3.4
2	B	14	PRO	3.4
2	B	95	PRO	3.4
2	B	7	THR	3.3
1	A	250	ASP	3.3
2	B	210	LEU	3.1
2	B	232	TYR	3.1
2	B	204	GLU	3.0
2	B	5	ILE	3.0
1	A	255	ASN	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	172	LYS	3.0
1	A	225	PRO	3.0
1	A	297	GLU	3.0
1	A	222	GLN	3.0
1	A	115	TYR	2.9
2	B	193	LEU	2.9
1	A	543	GLY	2.9
1	A	291	GLU	2.9
1	A	63	ILE	2.9
2	B	10	VAL	2.8
2	B	372	VAL	2.8
2	B	360	ALA	2.8
2	B	190	GLY	2.8
2	B	15	GLY	2.8
2	B	104	LYS	2.7
2	B	203	GLU	2.7
1	A	62	ALA	2.7
1	A	406	TRP	2.6
1	A	293	ILE	2.6
2	B	315	HIS	2.6
2	B	425	LEU	2.6
2	B	422	LEU	2.5
2	B	208	HIS	2.5
1	A	40	GLU	2.5
2	B	236	PRO	2.5
1	A	283	LEU	2.5
1	A	223	LYS	2.5
1	A	544	GLY	2.5
2	B	177	ASP	2.5
1	A	252	TRP	2.4
1	A	298	GLU	2.4
1	A	74	LEU	2.4
2	B	234	LEU	2.4
2	B	171	PHE	2.3
2	B	277	ARG	2.3
1	A	282	LEU	2.3
1	A	284	ARG	2.2
2	B	176	PRO	2.2
1	A	294	PRO	2.2
2	B	191	SER	2.2
2	B	98	ALA	2.2
1	A	-1	MET	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	52	PRO	2.2
2	B	199	ARG	2.2
2	B	66	LYS	2.1
1	A	183	TYR	2.1
1	A	254	VAL	2.1
2	B	70	LYS	2.1
2	B	201	LYS	2.1
1	A	247	PRO	2.1

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, 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	B-factors(\AA^2)	Q<0.9
4	SO4	A	602	5/5	0.94	0.14	68,71,76,82	0
3	9HY	A	601	36/36	0.94	0.15	39,43,53,57	0

6.5 Other polymers [i](#)

There are no such residues in this entry.