WORLDWIDE
wwPDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report ⓘ

Mar 2, 2021 – 11:53 AM EST

PDB ID : 5R1G

Title : PanDDA analysis group deposition – Auto-refined data of Aar2/RNaseH for
ground state model 31, DMSO-free

Authors : Wollenhaupt, J.; Metz, A.; Barthel, T.; Lima, G.M.A.; Heine, A.; Mueller, U.;
Klebe, G.; Weiss, M.S.

Deposited on : 2020-02-12

Resolution : 1.81 Å(reported)

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

|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.17.1
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)	:	2.17.1

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 1.81 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	Percentile Ranks	Value
Rfree		0.307
Clashscore		6
Ramachandran outliers		0.2%
Sidechain outliers		2.0%
RSRZ outliers		11.5%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
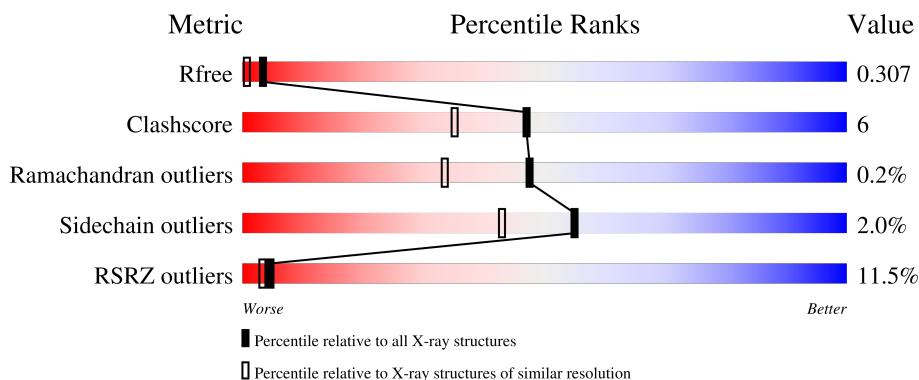
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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	258	9% 75% 16% • 8%
2	B	308	13% 81% 15% • •
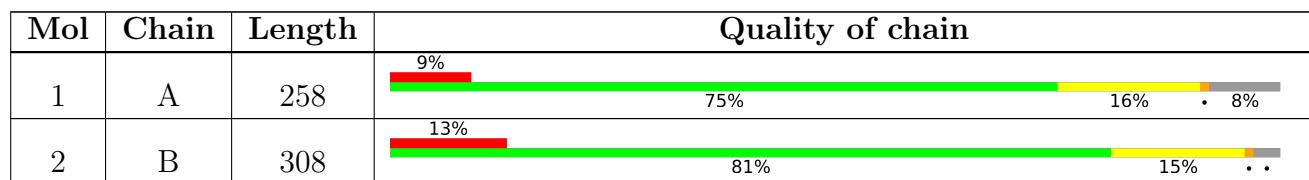
2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 4679 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 Pre-mRNA-splicing factor 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	237	Total 2002	C 1283	N 335	O 372	S 12	0	12	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1833	GLY	-	expression tag	UNP P33334
A	1834	ALA	-	expression tag	UNP P33334
A	1835	MET	-	expression tag	UNP P33334

- Molecule 2 is a protein called A1 cistron-splicing factor AAR2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	300	Total 2580	C 1654	N 421	O 485	S 20	0	9	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	expression tag	UNP P32357
B	-2	ALA	-	expression tag	UNP P32357
B	-1	MET	-	expression tag	UNP P32357
B	0	ALA	-	expression tag	UNP P32357
B	166	SER	LEU	conflict	UNP P32357
B	167	SER	LYS	conflict	UNP P32357
B	170	SER	LEU	conflict	UNP P32357
B	?	-	GLN	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357
B	?	-	ALA	deletion	UNP P32357
B	?	-	GLY	deletion	UNP P32357
B	?	-	SER	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	MET	deletion	UNP P32357
B	?	-	GLU	deletion	UNP P32357
B	?	-	ALA	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357
B	?	-	ASN	deletion	UNP P32357
B	?	-	GLU	deletion	UNP P32357
B	?	-	ASP	deletion	UNP P32357

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	65	Total 65	O 65	0	0
3	B	32	Total 32	O 32	0	0

3 Residue-property plots ⓘ

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: Pre-mRNA-splicing factor 8
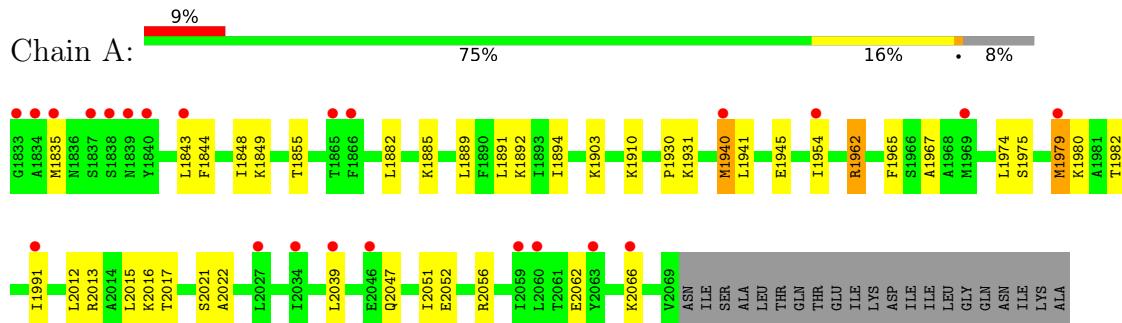
• Molecule 2: A1 cistron-splicing factor AAR2
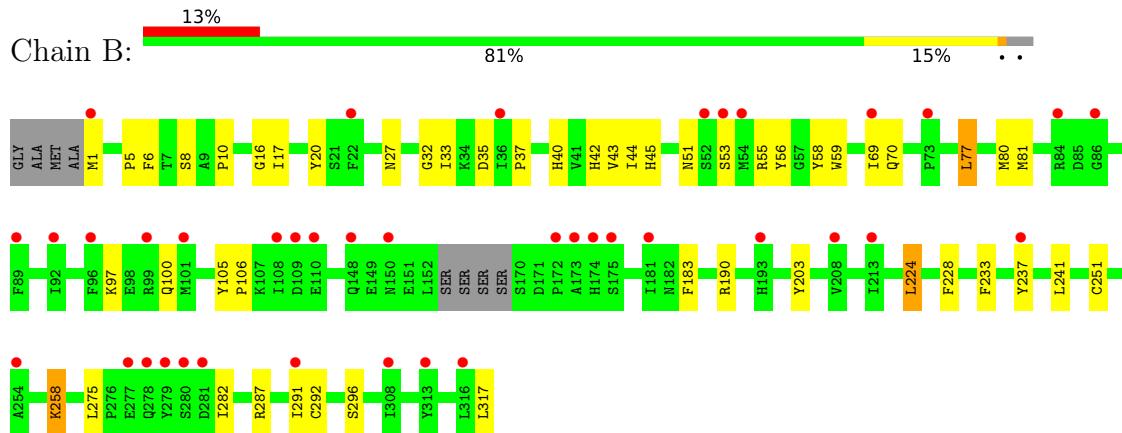
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	89.14Å 82.38Å 92.64Å 90.00° 107.99° 90.00°	Depositor
Resolution (Å)	23.48 – 1.81 44.81 – 1.81	Depositor EDS
% Data completeness (in resolution range)	98.1 (23.48-1.81) 98.2 (44.81-1.81)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.96 (at 1.81Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.252 , 0.303 0.262 , 0.307	Depositor DCC
R_{free} test set	2100 reflections (3.66%)	wwPDB-VP
Wilson B-factor (Å^2)	47.2	Xtriage
Anisotropy	0.289	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 60.6	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4679	wwPDB-VP
Average B, all atoms (Å^2)	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.03% 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 ⓘ

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	0.85	1/2049 (0.0%)	0.97	6/2775 (0.2%)
2	B	0.74	3/2651 (0.1%)	0.83	3/3581 (0.1%)
All	All	0.79	4/4700 (0.1%)	0.89	9/6356 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	183	PHE	CD1-CE1	6.20	1.51	1.39
1	A	1965	PHE	CD1-CE1	-5.83	1.27	1.39
2	B	56	TYR	CD1-CE1	-5.53	1.31	1.39
2	B	56	TYR	CD2-CE2	-5.52	1.31	1.39

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	224	LEU	CB-CG-CD2	-8.05	97.32	111.00
1	A	2012	LEU	CB-CG-CD1	-6.35	100.20	111.00
1	A	1940	MET	CG-SD-CE	-5.94	90.70	100.20
2	B	77	LEU	CA-CB-CG	5.59	128.16	115.30
2	B	224	LEU	CA-CB-CG	5.55	128.07	115.30
1	A	1982	THR	CA-CB-CG2	-5.31	104.97	112.40
1	A	1889	LEU	CB-CG-CD2	-5.29	102.01	111.00
1	A	1882	LEU	CB-CG-CD2	-5.28	102.02	111.00
1	A	1889	LEU	CB-CG-CD1	5.22	119.88	111.00

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	2002	0	2029	21	0
2	B	2580	0	2450	35	0
3	A	65	0	0	2	0
3	B	32	0	0	4	0
All	All	4679	0	4479	56	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 6.

All (56) 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)
2:B:1:MET:N	3:B:401:HOH:O	2.18	0.77
1:A:1910:LYS:HG2	1:A:1940:MET:SD	2.29	0.73
1:A:2062:GLU:O	1:A:2066:LYS:HG2	1.89	0.71
2:B:70:GLN:HB3	2:B:81:MET:HE1	1.75	0.67
2:B:287:ARG:O	2:B:291:ILE:HD13	1.97	0.65
2:B:70:GLN:HB3	2:B:81:MET:CE	2.31	0.61
2:B:6:PHE:HZ	2:B:44[A]:ILE:HD11	1.69	0.58
2:B:251:CYS:O	2:B:296:SER:HB2	2.03	0.57
2:B:190:ARG:HG3	2:B:203[B]:TYR:CZ	2.41	0.56
1:A:1980:LYS:NZ	3:A:2101:HOH:O	2.28	0.55
2:B:69:ILE:HD13	2:B:80:MET:HA	1.88	0.54
1:A:1843:LEU:HA	1:A:1849:LYS:HD2	1.89	0.54
1:A:1962:ARG:O	1:A:2013:ARG:NH1	2.38	0.54
2:B:53:SER:HA	3:B:404:HOH:O	2.09	0.52
2:B:33:ILE:HD12	2:B:44[A]:ILE:HD12	1.93	0.50
1:A:2052:GLU:O	1:A:2056[B]:ARG:HG3	2.12	0.50
2:B:20:TYR:CD1	2:B:106:PRO:HG2	2.47	0.48
2:B:37:PRO:HD3	2:B:105:TYR:CD1	2.49	0.48
1:A:1967:ALA:HB2	1:A:2016:LYS:HB2	1.95	0.48
1:A:1844:PHE:HA	1:A:1885:LYS:HD3	1.95	0.47
2:B:6:PHE:CZ	2:B:44[A]:ILE:HD11	2.49	0.47
1:A:2047:GLN:O	1:A:2051:ILE:HG12	2.16	0.46
1:A:1941:LEU:O	1:A:1945:GLU:HB2	2.15	0.46
2:B:237:TYR:CE2	2:B:241:LEU:HD11	2.51	0.46
2:B:17:ILE:O	2:B:17:ILE:HG23	2.15	0.46
1:A:1979[C]:MET:HE3	3:A:2114:HOH:O	2.16	0.45
2:B:42:HIS:O	2:B:59:TRP:HA	2.17	0.45

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:2015:LEU:HD23	1:A:2022:ALA:HB3	1.99	0.45
2:B:258:LYS:HD2	2:B:258:LYS:H	1.82	0.45
1:A:1844:PHE:O	1:A:1885:LYS:HD3	2.17	0.44
1:A:1855[A]:THR:O	1:A:1855[A]:THR:HG22	2.17	0.44
2:B:5:PRO:HA	2:B:32:GLY:HA3	2.00	0.44
2:B:224:LEU:C	2:B:224:LEU:HD23	2.38	0.44
2:B:16:GLY:HA3	2:B:45:HIS:CE1	2.52	0.44
1:A:1848:ILE:H	1:A:1931[A]:LYS:HZ2	1.66	0.44
2:B:1:MET:HB3	2:B:35:ASP:HA	2.00	0.44
2:B:228:PHE:CE1	2:B:275:LEU:HD21	2.53	0.43
1:A:1891:LEU:HD22	1:A:1974:LEU:HD22	1.99	0.43
2:B:40:HIS:HD2	3:B:429:HOH:O	2.02	0.43
2:B:251:CYS:HB2	2:B:292:CYS:HG	1.84	0.42
1:A:2013:ARG:O	1:A:2017[A]:THR:HG23	2.19	0.42
2:B:190:ARG:HG3	2:B:203[B]:TYR:CE2	2.54	0.42
2:B:8:SER:C	2:B:10:PRO:HD3	2.40	0.42
2:B:43:VAL:HA	2:B:58:TYR:O	2.20	0.42
2:B:97:LYS:O	2:B:100:GLN:N	2.52	0.41
2:B:70:GLN:CB	2:B:81:MET:CE	2.98	0.41
2:B:70:GLN:CB	2:B:81:MET:HE1	2.45	0.41
2:B:251:CYS:SG	2:B:292:CYS:HB3	2.60	0.41
2:B:6:PHE:CD1	2:B:32:GLY:HA2	2.55	0.41
1:A:1975:SER:O	1:A:1979[A]:MET:SD	2.79	0.41
1:A:1894:ILE:HA	1:A:1894:ILE:HD13	1.88	0.41
2:B:55:ARG:HD2	2:B:233:PHE:O	2.21	0.41
1:A:1930:PRO:O	1:A:1954:ILE:HG12	2.22	0.40
2:B:51:ASN:C	2:B:53:SER:H	2.25	0.40
2:B:237:TYR:HB2	3:B:431:HOH:O	2.22	0.40
1:A:1991:ILE:HA	1:A:2039:LEU:HD12	2.03	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	248/258 (96%)	240 (97%)	8 (3%)	0	100	100
2	B	306/308 (99%)	287 (94%)	18 (6%)	1 (0%)	41	27
All	All	554/566 (98%)	527 (95%)	26 (5%)	1 (0%)	47	33

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	27	ASN

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	226/233 (97%)	218 (96%)	8 (4%)	36	20
2	B	287/284 (101%)	283 (99%)	4 (1%)	67	58
All	All	513/517 (99%)	501 (98%)	12 (2%)	55	37

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1835	MET
1	A	1892	LYS
1	A	1903	LYS
1	A	1962	ARG
1	A	1979[A]	MET
1	A	1979[B]	MET
1	A	1979[C]	MET
1	A	2021	SER
2	B	77	LEU
2	B	258	LYS
2	B	282	ILE
2	B	317	LEU

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

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Res	Type
1	A	1907	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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, 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	237/258 (91%)	0.79	23 (9%)	7	5	43, 68, 113, 196	0
2	B	300/308 (97%)	0.90	39 (13%)	3	2	45, 74, 129, 240	0
All	All	537/566 (94%)	0.85	62 (11%)	4	3	43, 71, 121, 240	0

All (62) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	109	ASP	9.5
2	B	53	SER	8.6
2	B	52	SER	6.6
2	B	1	MET	5.4
1	A	1833	GLY	5.1
2	B	54[A]	MET	4.9
2	B	108	ILE	4.7
1	A	2060	LEU	4.4
2	B	279	TYR	4.2
2	B	175	SER	3.6
2	B	181	ILE	3.6
1	A	1838	SER	3.5
2	B	101	MET	3.5
2	B	280	SER	3.5
2	B	277	GLU	3.4
2	B	174	HIS	3.4
1	A	2027	LEU	3.4
1	A	1969	MET	3.4
2	B	110	GLU	3.4
1	A	2063	TYR	3.3
2	B	150	ASN	3.1
2	B	193	HIS	3.1
2	B	148	GLN	3.1
2	B	254	ALA	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	237	TYR	2.9
1	A	1940	MET	2.9
2	B	208	VAL	2.8
2	B	308	ILE	2.8
1	A	2034	ILE	2.8
1	A	1979[A]	MET	2.7
2	B	36	ILE	2.6
2	B	92	ILE	2.6
2	B	213	ILE	2.6
2	B	278	GLN	2.6
2	B	89	PHE	2.6
1	A	1834	ALA	2.5
2	B	96	PHE	2.5
1	A	2039	LEU	2.4
2	B	316	LEU	2.4
2	B	84	ARG	2.4
2	B	22	PHE	2.4
2	B	172	PRO	2.3
1	A	1837	SER	2.3
2	B	99	ARG	2.3
1	A	1866	PHE	2.2
1	A	1835	MET	2.2
2	B	73	PRO	2.2
1	A	1865	THR	2.2
2	B	281	ASP	2.2
2	B	69	ILE	2.1
1	A	1954	ILE	2.1
2	B	291	ILE	2.1
1	A	1840	TYR	2.1
2	B	86	GLY	2.1
1	A	1991	ILE	2.1
1	A	2059	ILE	2.1
1	A	2066	LYS	2.1
1	A	1843	LEU	2.0
1	A	2046	GLU	2.0
2	B	313	TYR	2.0
2	B	173	ALA	2.0
1	A	1839	ASN	2.0

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.