Full wwPDB X-ray Structure Validation Report ⓘ

Mar 2, 2021 – 11:41 AM EST

PDB ID	:	5R1O
Title	:	PanDDA analysis group deposition – Auto-refined data of Aar2/RNaseH for ground state model 39, 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	:	1.90 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:

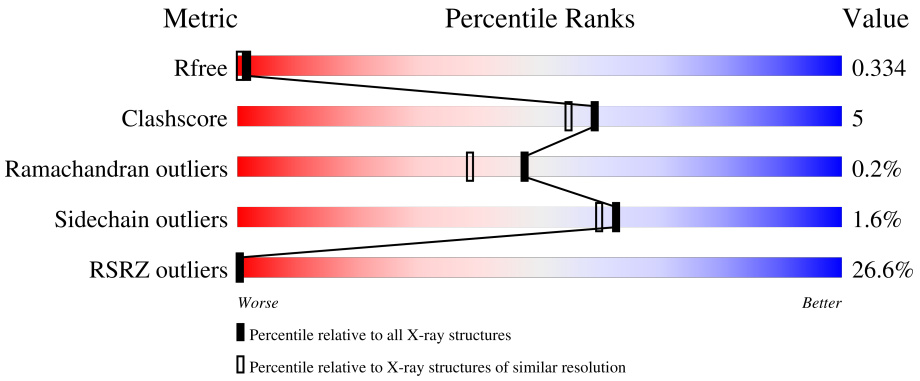
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.17.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.17.1

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	22% 80% 10% · 8%
2	B	308	28% 82% 14% · ·

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4672 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	C	N	O	S	0	12	0
			2002	1283	335	372	12			

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	C	N	O	S	0	9	0
			2580	1654	421	485	20			

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.

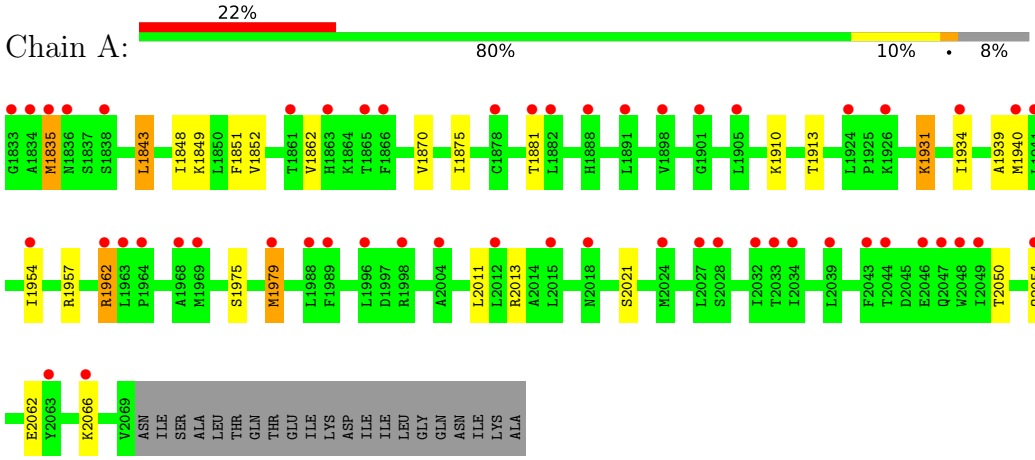
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	55	Total O 55 55	0	0
3	B	35	Total O 35 35	0	0

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.

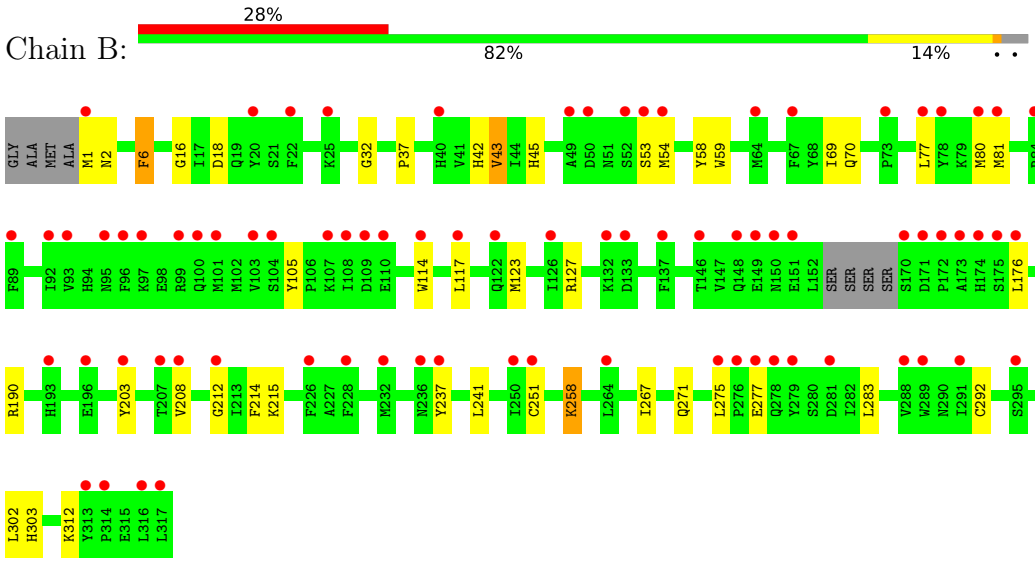
• Molecule 1: Pre-mRNA-splicing factor 8

Chain A:



• Molecule 2: A1 cistron-splicing factor AAR2

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	88.95Å 82.32Å 92.48Å 90.00° 107.71° 90.00°	Depositor
Resolution (Å)	22.55 – 1.90 44.85 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.7 (22.55-1.90) 99.8 (44.85-1.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.99 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.290 , 0.334 0.295 , 0.334	Depositor DCC
R_{free} test set	2100 reflections (4.19%)	wwPDB-VP
Wilson B-factor (Å^2)	46.8	Xtriage
Anisotropy	0.205	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 63.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	4672	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.75% 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	0.74	1/2049 (0.0%)	0.91	3/2775 (0.1%)
2	B	0.73	1/2651 (0.0%)	0.85	3/3581 (0.1%)
All	All	0.73	2/4700 (0.0%)	0.88	6/6356 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	43	VAL	CB-CG2	5.29	1.64	1.52
1	A	1852	VAL	CB-CG1	5.28	1.64	1.52

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	2011	LEU	CB-CG-CD1	-6.78	99.47	111.00
2	B	176	LEU	CB-CG-CD1	-6.75	99.53	111.00
2	B	6	PHE	CB-CG-CD2	-6.35	116.36	120.80
2	B	117	LEU	CB-CG-CD2	6.07	121.32	111.00
1	A	1843	LEU	CB-CG-CD2	-5.94	100.89	111.00
1	A	1881	THR	CA-CB-CG2	-5.24	105.07	112.40

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	17	0
2	B	2580	0	2450	26	0
3	A	55	0	0	0	0
3	B	35	0	0	2	0
All	All	4672	0	4479	43	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 5.

All (43) 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	1.96	0.98
1:A:1848:ILE:H	1:A:1931[A]:LYS:HZ2	1.32	0.77
1:A:2062:GLU:O	1:A:2066:LYS:HG2	1.87	0.74
2:B:70:GLN:HB3	2:B:81:MET:HE1	1.78	0.65
2:B:258:LYS:HD2	2:B:258:LYS:H	1.63	0.63
2:B:251:CYS:SG	2:B:292:CYS:SG	3.01	0.59
2:B:301:SER:O	2:B:302:LEU:HD23	2.02	0.59
1:A:1962:ARG:O	1:A:2013:ARG:NH1	2.33	0.57
1:A:2058:LEU:C	1:A:2058:LEU:HD23	2.25	0.56
1:A:2062:GLU:HB3	1:A:2066:LYS:HE3	1.88	0.55
1:A:1848:ILE:H	1:A:1931[A]:LYS:NZ	2.04	0.55
1:A:1913:THR:HB	1:A:1940:MET:HE1	1.89	0.53
2:B:70:GLN:HB3	2:B:81:MET:CE	2.39	0.53
2:B:42:HIS:O	2:B:59:TRP:HA	2.10	0.52
2:B:300:ASN:O	2:B:303:HIS:NE2	2.42	0.51
1:A:1875:ILE:HD12	1:A:1979[B]:MET:HE1	1.93	0.50
1:A:1843:LEU:HA	1:A:1849:LYS:HD2	1.95	0.49
1:A:2050:THR:HG22	1:A:2054:GLN:NE2	2.29	0.48
1:A:1951:PHE:HB3	1:A:1954:ILE:HD12	1.97	0.46
2:B:251:CYS:SG	2:B:292:CYS:HB3	2.56	0.46
2:B:275:LEU:HD22	2:B:283:LEU:HD13	1.98	0.46
1:A:1848:ILE:N	1:A:1931[A]:LYS:HZ2	2.09	0.45
2:B:237:TYR:CE2	2:B:241:LEU:HD11	2.52	0.45
2:B:277:GLU:HG3	2:B:312:LYS:HE2	1.97	0.45
2:B:53:SER:HA	3:B:405:HOH:O	2.18	0.44
2:B:267:ILE:O	2:B:271:GLN:HG3	2.18	0.44
2:B:208:VAL:O	2:B:212:GLY:HA3	2.18	0.44
2:B:43:VAL:HA	2:B:58:TYR:O	2.17	0.43
2:B:298:GLN:HA	2:B:298:GLN:OE1	2.18	0.43
2:B:37:PRO:HD3	2:B:105:TYR:CD1	2.53	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1862:VAL:HG13	1:A:1870:VAL:CG1	2.49	0.42
1:A:1910:LYS:HD3	1:A:1939:ALA:HB1	2.00	0.42
1:A:1934:ILE:HA	1:A:1957:ARG:O	2.18	0.42
2:B:6:PHE:CD1	2:B:32:GLY:HA2	2.55	0.42
2:B:18:ASP:HB3	2:B:114:TRP:CE3	2.54	0.42
2:B:190:ARG:HG3	2:B:203[B]:TYR:CE2	2.55	0.41
2:B:69:ILE:HD13	2:B:80:MET:HA	2.01	0.41
2:B:123:MET:O	2:B:127:ARG:HG3	2.21	0.41
1:A:1835:MET:HE1	1:A:1851:PHE:HZ	1.86	0.41
2:B:16:GLY:HA3	2:B:45:HIS:CE1	2.55	0.41
2:B:275:LEU:CD2	2:B:283:LEU:HD13	2.51	0.41
2:B:214:PHE:O	2:B:215:LYS:HB2	2.20	0.41
1:A:1975:SER:O	1:A:1979[A]:MET:SD	2.79	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	248/258 (96%)	241 (97%)	7 (3%)	0	100 100
2	B	306/308 (99%)	291 (95%)	13 (4%)	2 (1%)	22 12
All	All	554/566 (98%)	532 (96%)	20 (4%)	2 (0%)	47 24

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	54[A]	MET
2	B	54[B]	MET

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	226/233 (97%)	218 (96%)	8 (4%)	36 27
2	B	287/284 (101%)	284 (99%)	3 (1%)	76 76
All	All	513/517 (99%)	502 (98%)	11 (2%)	62 48

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1835	MET
1	A	1931[A]	LYS
1	A	1931[B]	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	2	ASN
2	B	77	LEU
2	B	258	LYS

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

Mol	Chain	Res	Type
1	A	1907	GLN

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)

There are no ligands in this entry.

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	237/258 (91%)	1.38	57 (24%)	0	0	41, 76, 124, 198	0
2	B	300/308 (97%)	1.63	86 (28%)	0	0	42, 83, 143, 199	0
All	All	537/566 (94%)	1.52	143 (26%)	0	0	41, 81, 137, 199	0

All (143) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	53	SER	11.7
2	B	172	PRO	7.4
1	A	2027	LEU	7.1
2	B	99	ARG	6.9
1	A	1878	CYS	6.3
2	B	101	MET	6.2
2	B	103	VAL	6.0
2	B	316	LEU	5.9
2	B	107	LYS	5.8
2	B	1	MET	5.8
2	B	175	SER	5.5
1	A	1833	GLY	5.4
2	B	54[A]	MET	5.4
2	B	52	SER	5.2
2	B	174	HIS	5.0
2	B	86	GLY	4.9
2	B	173	ALA	4.9
2	B	171	ASP	4.7
2	B	251	CYS	4.7
2	B	109	ASP	4.5
2	B	108	ILE	4.5
1	A	1838	SER	4.5
1	A	1940	MET	4.4
2	B	148	GLN	4.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	117	LEU	4.1
1	A	1996	LEU	4.0
1	A	1882	LEU	3.8
2	B	317	LEU	3.8
2	B	193	HIS	3.7
2	B	313	TYR	3.7
2	B	181	ILE	3.7
2	B	146	THR	3.6
2	B	110	GLU	3.6
1	A	2044	THR	3.6
2	B	89	PHE	3.6
1	A	1881	THR	3.5
2	B	73	PRO	3.5
2	B	170	SER	3.4
1	A	1989	PHE	3.3
1	A	1865	THR	3.3
2	B	87	ALA	3.3
2	B	22	PHE	3.3
2	B	150	ASN	3.3
1	A	2046	GLU	3.2
2	B	80	MET	3.1
1	A	1888	HIS	3.1
2	B	208	VAL	3.0
1	A	2012	LEU	3.0
1	A	2049	ILE	3.0
1	A	2063	TYR	3.0
2	B	277	GLU	2.9
2	B	84	ARG	2.9
1	A	2039	LEU	2.9
2	B	281	ASP	2.9
1	A	2034	ILE	2.9
2	B	20	TYR	2.9
1	A	1924	LEU	2.9
2	B	132	LYS	2.8
2	B	189	ILE	2.8
1	A	1834	ALA	2.8
2	B	188	ALA	2.8
1	A	1979[A]	MET	2.8
1	A	1954	ILE	2.7
2	B	149	GLU	2.7
1	A	2057[A]	ASP	2.7
2	B	232	MET	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	1934	ILE	2.7
1	A	1891	LEU	2.7
2	B	264	LEU	2.7
2	B	64[A]	MET	2.6
2	B	151	GLU	2.6
1	A	2048	TRP	2.6
2	B	88	LYS	2.5
2	B	276	PRO	2.5
1	A	1836	ASN	2.5
1	A	2018	ASN	2.5
1	A	1863	HIS	2.5
2	B	289	TRP	2.5
2	B	228	PHE	2.5
1	A	1988	LEU	2.5
2	B	291	ILE	2.5
2	B	278	GLN	2.5
2	B	295	SER	2.5
1	A	1964	PRO	2.4
2	B	104	SER	2.4
1	A	1835	MET	2.4
2	B	78	TYR	2.4
2	B	226	PHE	2.4
2	B	212	GLY	2.4
2	B	176	LEU	2.4
1	A	2043	PHE	2.4
2	B	67	PHE	2.4
1	A	1948	MET	2.4
1	A	2024[A]	MET	2.4
1	A	1901	GLY	2.4
1	A	2066	LYS	2.4
2	B	77	LEU	2.4
1	A	1968	ALA	2.4
1	A	1969	MET	2.4
2	B	50	ASP	2.3
1	A	2032	ILE	2.3
1	A	1926	LYS	2.3
2	B	236	ASN	2.3
2	B	250	ILE	2.3
2	B	100	GLN	2.3
1	A	1898[A]	VAL	2.3
1	A	1962	ARG	2.3
1	A	1861	THR	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	95	ASN	2.3
2	B	279	TYR	2.2
2	B	96	PHE	2.2
1	A	1866	PHE	2.2
2	B	81	MET	2.2
1	A	2033	THR	2.2
2	B	133	ASP	2.2
2	B	40	HIS	2.2
2	B	122[A]	GLN	2.2
1	A	1905	LEU	2.2
2	B	288	VAL	2.2
2	B	92	ILE	2.2
2	B	25	LYS	2.2
2	B	203[A]	TYR	2.2
2	B	137	PHE	2.2
2	B	93	VAL	2.2
2	B	126	ILE	2.1
2	B	314	PRO	2.1
1	A	2047	GLN	2.1
1	A	2056[A]	ARG	2.1
2	B	237	TYR	2.1
1	A	2054	GLN	2.1
1	A	2004	ALA	2.1
2	B	49	ALA	2.1
2	B	275	LEU	2.1
2	B	207	THR	2.1
1	A	1963	LEU	2.1
2	B	196	GLU	2.1
1	A	1941	LEU	2.1
1	A	1951	PHE	2.0
2	B	114	TRP	2.0
1	A	2015	LEU	2.0
1	A	2028	SER	2.0
2	B	97	LYS	2.0
1	A	1998	ARG	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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.