Full wwPDB X-ray Structure Validation Report ⓘ

Mar 2, 2021 – 10:25 AM EST

PDB ID	:	5R12
Title	:	PanDDA analysis group deposition – Auto-refined data of Aar2/RNaseH for ground state model 16, 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	:	1.70 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:

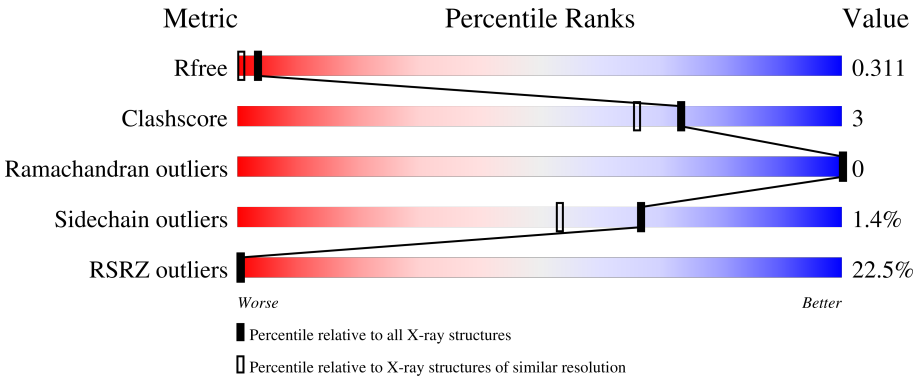
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.17.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 1.70 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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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	20% 82% 9% • 8%
2	B	308	22% 86% 10% • •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4676 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
			Total	C	N	O	S			
1	A	237	1994	1278	334	371	11	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
			Total	C	N	O	S			
2	B	300	2580	1654	421	485	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

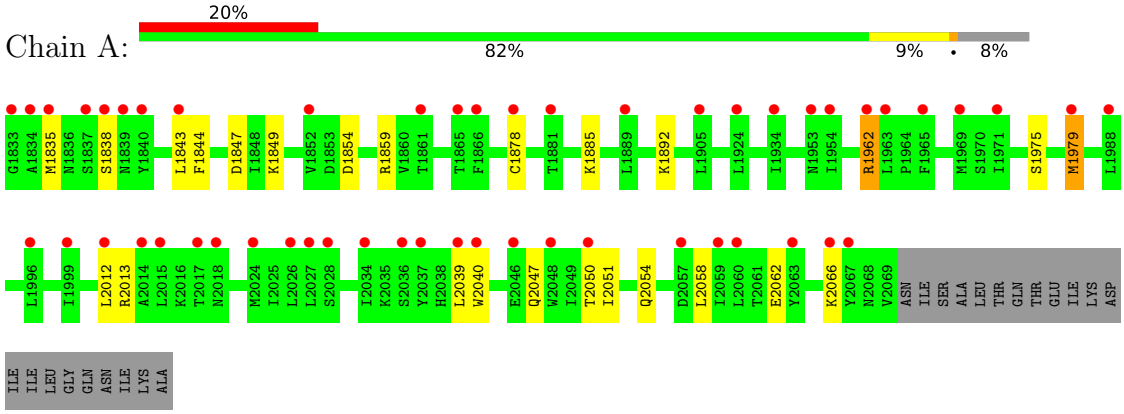
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	69	Total O 69 69	0	0
3	B	33	Total O 33 33	0	0

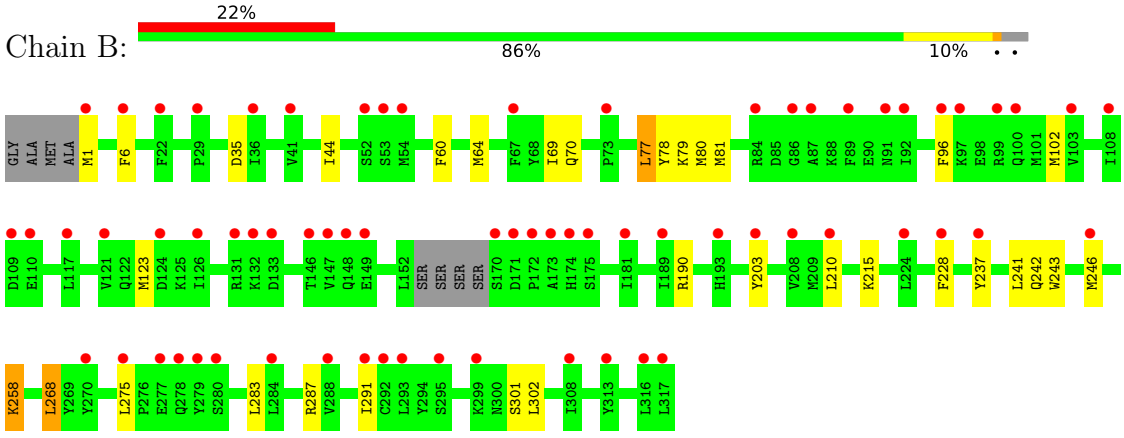
3 Residue-property plots (i)

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



• Molecule 2: A1 cistron-splicing factor AAR2



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	88.48Å 82.61Å 92.85Å 90.00° 107.88° 90.00°	Depositor
Resolution (Å)	22.32 – 1.70 44.80 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.0 (22.32-1.70) 99.0 (44.80-1.70)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.99 (at 1.70Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.266 , 0.298 0.273 , 0.311	Depositor DCC
R_{free} test set	2101 reflections (3.02%)	wwPDB-VP
Wilson B-factor (Å^2)	42.4	Xtriage
Anisotropy	0.250	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 55.1	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	4676	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.69% 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	2/2041 (0.1%)	0.92	7/2765 (0.3%)
2	B	0.68	2/2651 (0.1%)	0.76	1/3581 (0.0%)
All	All	0.71	4/4692 (0.1%)	0.83	8/6346 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	60	PHE	CE1-CZ	6.22	1.49	1.37
1	A	1878	CYS	C-N	6.04	1.48	1.34
2	B	78	TYR	CD2-CE2	5.97	1.48	1.39
1	A	2040	TRP	CB-CG	5.68	1.60	1.50

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1847	ASP	CB-CG-OD1	6.61	124.25	118.30
1	A	2012	LEU	CB-CG-CD2	6.34	121.78	111.00
1	A	1847	ASP	CB-CG-OD2	-5.91	112.98	118.30
1	A	1854	ASP	CB-CG-OD2	5.82	123.54	118.30
2	B	268	LEU	CB-CG-CD2	-5.63	101.43	111.00
1	A	2039	LEU	CA-CB-CG	5.58	128.14	115.30
1	A	2039	LEU	CB-CG-CD2	-5.56	101.56	111.00
1	A	1859	ARG	NE-CZ-NH1	-5.44	117.58	120.30

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	1994	0	2020	9	0
2	B	2580	0	2450	21	0
3	A	69	0	0	0	0
3	B	33	0	0	1	0
All	All	4676	0	4470	30	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 (30) 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:2062:GLU:O	1:A:2066:LYS:HG2	1.93	0.68
1:A:1962:ARG:O	1:A:2013:ARG:NH1	2.29	0.66
2:B:1:MET:N	3:B:401:HOH:O	2.33	0.61
1:A:1844:PHE:O	1:A:1885:LYS:NZ	2.23	0.55
2:B:287:ARG:O	2:B:291:ILE:HD13	2.07	0.55
1:A:2062:GLU:HB3	1:A:2066:LYS:HE3	1.88	0.54
2:B:258:LYS:HD2	2:B:258:LYS:H	1.74	0.53
2:B:70:GLN:HB3	2:B:81:MET:HE1	1.91	0.51
2:B:69:ILE:HD13	2:B:80:MET:HA	1.92	0.51
2:B:6:PHE:HZ	2:B:44[A]:ILE:HD11	1.75	0.50
2:B:1:MET:HB3	2:B:35:ASP:HA	1.94	0.49
2:B:275:LEU:HD22	2:B:283:LEU:HD13	1.96	0.48
2:B:190:ARG:HG3	2:B:203[B]:TYR:CZ	2.49	0.48
1:A:2047:GLN:O	1:A:2051:ILE:HG12	2.15	0.46
2:B:77:LEU:HD21	2:B:79:LYS:HE3	1.97	0.46
2:B:190:ARG:HG3	2:B:203[B]:TYR:CE2	2.52	0.45
2:B:64[A]:MET:SD	2:B:123:MET:HG2	2.56	0.44
1:A:1843:LEU:HA	1:A:1849:LYS:HD2	2.00	0.44
2:B:96:PHE:HB2	2:B:102:MET:HE3	1.98	0.43
2:B:242:GLN:O	2:B:246:MET:HG3	2.18	0.43
1:A:2058:LEU:C	1:A:2058:LEU:HD23	2.39	0.42
2:B:6:PHE:CZ	2:B:44[A]:ILE:HD11	2.52	0.42
2:B:210:LEU:O	2:B:215:LYS:N	2.52	0.42
2:B:301:SER:O	2:B:302:LEU:HD23	2.20	0.42
1:A:2050:THR:HG22	1:A:2054:GLN:NE2	2.36	0.41
2:B:70:GLN:HB3	2:B:81:MET:CE	2.50	0.41
2:B:237:TYR:CE2	2:B:241:LEU:HD11	2.56	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:268:LEU:HA	2:B:268:LEU:HD23	1.82	0.40
1:A:1975:SER:O	1:A:1979[A]:MET:SD	2.80	0.40
2:B:228:PHE:HB2	2:B:243:TRP:CD1	2.56	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	247/258 (96%)	242 (98%)	5 (2%)	0	100	100
2	B	306/308 (99%)	293 (96%)	13 (4%)	0	100	100
All	All	553/566 (98%)	535 (97%)	18 (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	225/233 (97%)	219 (97%)	6 (3%)	44	26
2	B	287/284 (101%)	285 (99%)	2 (1%)	84	77
All	All	512/517 (99%)	504 (98%)	8 (2%)	67	48

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1835	MET
1	A	1838	SER
1	A	1892	LYS
1	A	1962	ARG
1	A	1979[A]	MET
1	A	1979[C]	MET
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%)	1.27	52 (21%)	0 0	38, 68, 108, 139	0
2	B	300/308 (97%)	1.40	69 (23%)	0 0	42, 78, 117, 204	0
All	All	537/566 (94%)	1.34	121 (22%)	0 0	38, 72, 114, 204	0

All (121) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	53	SER	11.7
1	A	1833	GLY	9.5
2	B	1	MET	9.2
2	B	52	SER	8.8
2	B	54[A]	MET	6.2
1	A	2034	ILE	5.9
2	B	87	ALA	5.6
2	B	174	HIS	5.5
2	B	173	ALA	5.2
1	A	2027	LEU	5.1
1	A	1878	CYS	5.0
1	A	2039	LEU	4.7
2	B	279	TYR	4.7
2	B	171	ASP	4.7
2	B	109	ASP	4.6
1	A	1979[A]	MET	4.6
2	B	277	GLU	4.4
2	B	96	PHE	4.4
1	A	1934	ILE	4.4
2	B	181	ILE	4.3
1	A	2059	ILE	4.3
1	A	2048	TRP	4.2
2	B	89	PHE	4.2
2	B	73	PRO	4.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	108	ILE	4.1
2	B	208	VAL	4.1
2	B	210	LEU	3.9
2	B	189	ILE	3.8
1	A	1889	LEU	3.7
1	A	2028	SER	3.6
2	B	170	SER	3.6
2	B	193	HIS	3.6
1	A	1996	LEU	3.5
2	B	228	PHE	3.5
1	A	1963	LEU	3.5
2	B	172	PRO	3.4
2	B	146	THR	3.4
1	A	1838	SER	3.4
2	B	22	PHE	3.3
1	A	1924	LEU	3.3
2	B	316	LEU	3.3
1	A	2063	TYR	3.3
2	B	175	SER	3.2
1	A	1905	LEU	3.2
2	B	317	LEU	3.2
2	B	126	ILE	3.1
2	B	29	PRO	3.1
2	B	237	TYR	3.1
1	A	2015	LEU	3.1
1	A	1965	PHE	3.0
1	A	1840	TYR	3.0
2	B	103	VAL	3.0
1	A	2066	LYS	3.0
2	B	84	ARG	3.0
1	A	1837	SER	2.9
2	B	147	VAL	2.9
1	A	1969	MET	2.8
2	B	313	TYR	2.8
1	A	1953	ASN	2.8
1	A	1843	LEU	2.8
1	A	2012	LEU	2.8
1	A	1962	ARG	2.8
2	B	278	GLN	2.8
1	A	2036	SER	2.8
2	B	295	SER	2.7
1	A	1865	THR	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	110	GLU	2.7
1	A	1881	THR	2.7
1	A	2046	GLU	2.6
1	A	1861	THR	2.6
2	B	149	GLU	2.6
1	A	1834	ALA	2.6
1	A	1852	VAL	2.6
2	B	246	MET	2.6
2	B	86	GLY	2.5
1	A	2040	TRP	2.5
1	A	1866	PHE	2.5
1	A	2060	LEU	2.5
2	B	6	PHE	2.5
1	A	1954	ILE	2.5
2	B	291	ILE	2.5
2	B	293	LEU	2.5
2	B	97	LYS	2.4
1	A	1988	LEU	2.4
1	A	2026	LEU	2.4
2	B	121	VAL	2.4
2	B	131	ARG	2.4
2	B	299	LYS	2.4
2	B	92	ILE	2.4
2	B	100	GLN	2.4
1	A	2017[A]	THR	2.4
2	B	148	GLN	2.3
2	B	117	LEU	2.3
1	A	1999	ILE	2.3
1	A	2067	TYR	2.3
2	B	275	LEU	2.3
1	A	2050	THR	2.3
2	B	280	SER	2.3
2	B	288	VAL	2.2
1	A	2024[A]	MET	2.2
2	B	132	LYS	2.2
2	B	292	CYS	2.2
2	B	308	ILE	2.2
1	A	2018	ASN	2.2
1	A	2037	TYR	2.2
2	B	36	ILE	2.2
1	A	2057[A]	ASP	2.2
2	B	124	ASP	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	284	LEU	2.1
2	B	99	ARG	2.1
2	B	203[A]	TYR	2.1
2	B	91	ASN	2.1
2	B	224	LEU	2.1
2	B	270	TYR	2.1
1	A	2014	ALA	2.1
2	B	41[A]	VAL	2.1
2	B	133	ASP	2.1
1	A	1971	ILE	2.1
1	A	1835	MET	2.0
2	B	67	PHE	2.0
1	A	1839	ASN	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.