Full wwPDB X-ray Structure Validation Report ⓘ

Oct 23, 2023 – 10:10 pm BST

PDB ID	:	8BE4
Title	:	Crystal structure of SOS1-KRasG12V-Nanobody14
Authors	:	Fischer, B.; Wohlkonig, A.; Steyaert, J.
Deposited on	:	2022-10-21
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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

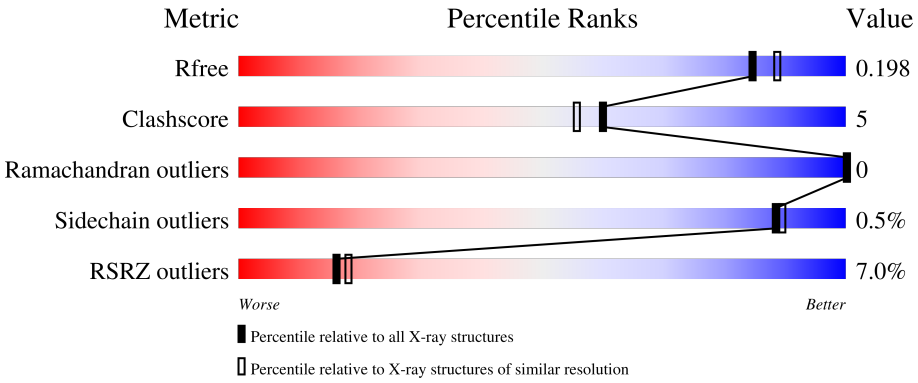
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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	S	507	6% 78% 9% 13%
2	R	190	8% 76% 11% · 13%
3	C000	128	5% 93% 7%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6622 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 Son of sevenless homolog 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	S	440	Total	C	N	O	S	0	12	0
			3768	2421	645	688	14			

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
S	543	MET	-	initiating methionine	UNP Q07889
S	544	GLY	-	expression tag	UNP Q07889
S	545	SER	-	expression tag	UNP Q07889
S	546	SER	-	expression tag	UNP Q07889
S	547	HIS	-	expression tag	UNP Q07889
S	548	HIS	-	expression tag	UNP Q07889
S	549	HIS	-	expression tag	UNP Q07889
S	550	HIS	-	expression tag	UNP Q07889
S	551	HIS	-	expression tag	UNP Q07889
S	552	HIS	-	expression tag	UNP Q07889
S	553	SER	-	expression tag	UNP Q07889
S	554	SER	-	expression tag	UNP Q07889
S	555	GLY	-	expression tag	UNP Q07889
S	556	LEU	-	expression tag	UNP Q07889
S	557	VAL	-	expression tag	UNP Q07889
S	558	PRO	-	expression tag	UNP Q07889
S	559	ARG	-	expression tag	UNP Q07889
S	560	GLY	-	expression tag	UNP Q07889
S	561	SER	-	expression tag	UNP Q07889
S	562	HIS	-	expression tag	UNP Q07889
S	563	MET	-	expression tag	UNP Q07889

- Molecule 2 is a protein called Isoform 2B of GTPase KRas.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	R	165	Total	C	N	O	S	0	7	0
			1385	865	243	271	6			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
R	-20	MET	-	initiating methionine	UNP P01116-2
R	-19	GLY	-	expression tag	UNP P01116-2
R	-18	SER	-	expression tag	UNP P01116-2
R	-17	SER	-	expression tag	UNP P01116-2
R	-16	HIS	-	expression tag	UNP P01116-2
R	-15	HIS	-	expression tag	UNP P01116-2
R	-14	HIS	-	expression tag	UNP P01116-2
R	-13	HIS	-	expression tag	UNP P01116-2
R	-12	HIS	-	expression tag	UNP P01116-2
R	-11	HIS	-	expression tag	UNP P01116-2
R	-10	SER	-	expression tag	UNP P01116-2
R	-9	SER	-	expression tag	UNP P01116-2
R	-8	GLY	-	expression tag	UNP P01116-2
R	-7	GLU	-	expression tag	UNP P01116-2
R	-6	ASN	-	expression tag	UNP P01116-2
R	-5	LEU	-	expression tag	UNP P01116-2
R	-4	TYR	-	expression tag	UNP P01116-2
R	-3	PHE	-	expression tag	UNP P01116-2
R	-2	GLN	-	expression tag	UNP P01116-2
R	-1	GLY	-	expression tag	UNP P01116-2
R	0	SER	-	expression tag	UNP P01116-2
R	12	VAL	GLY	engineered mutation	UNP P01116-2

- Molecule 3 is a protein called Nanobody14.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C000	119	Total	C	N	O	S	0	1	0
			918	566	172	176	4			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	S	335	Total	O	0	0
			335	335		
4	R	108	Total	O	0	0
			108	108		

Continued on next page...

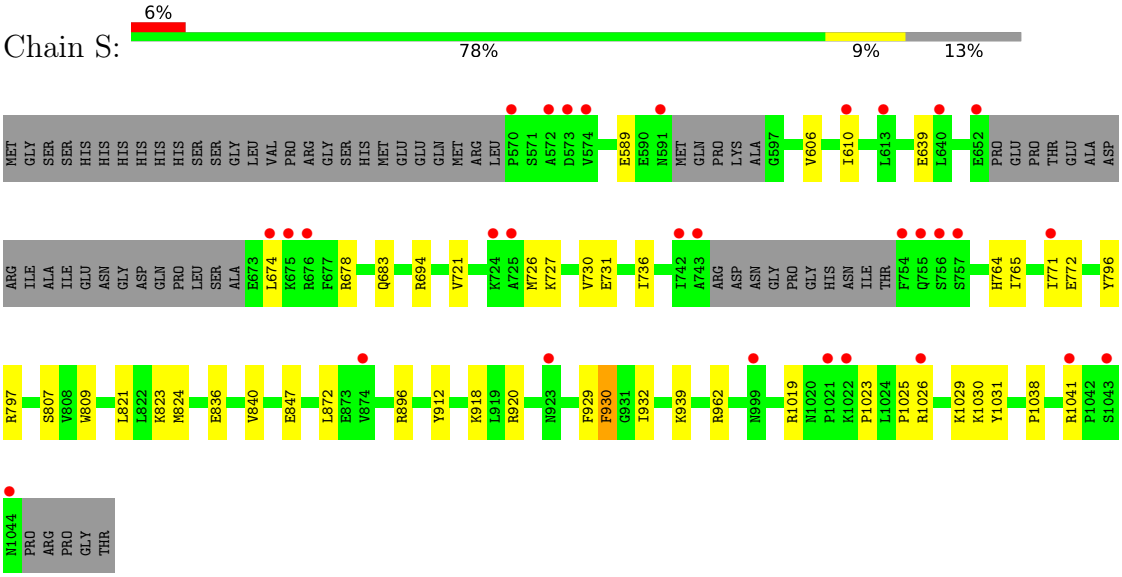
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C000	108	Total 108	O 108	0	0

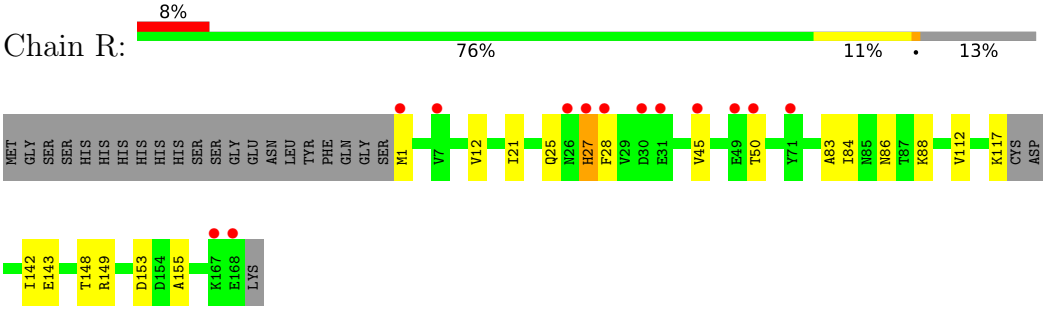
3 Residue-property plots (i)

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.

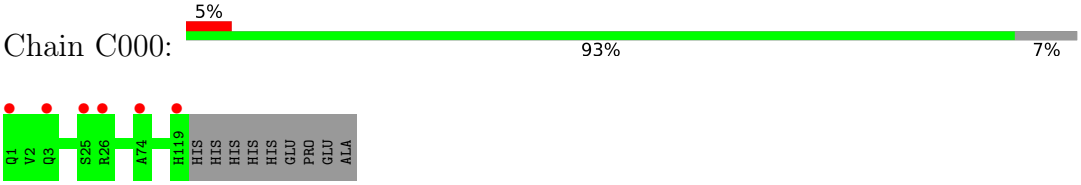
• Molecule 1: Son of sevenless homolog 1



• Molecule 2: Isoform 2B of GTPase KRas



• Molecule 3: Nanobody14



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, α, β, γ	127.67Å 127.67Å 152.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.59 – 1.90 29.59 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.6 (29.59-1.90) 99.7 (29.59-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.33 (at 1.89Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.175 , 0.199 0.174 , 0.198	Depositor DCC
R_{free} test set	4795 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	35.8	Xtriage
Anisotropy	0.048	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 45.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.021 for -h,k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	6622	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.87% 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	S	0.89	0/3856	0.78	0/5209
2	R	0.84	0/1406	0.80	0/1893
3	C000	0.84	0/935	0.79	0/1265
All	All	0.87	0/6197	0.78	0/8367

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	S	3768	0	3773	38	0
2	R	1385	0	1362	16	0
3	C000	918	0	0	0	0
4	C000	108	0	0	0	0
4	R	108	0	0	3	0
4	S	335	0	0	11	0
All	All	6622	0	5135	54	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 (54) 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:S:847:GLU:OE2	1:S:1029:LYS:HE2	1.85	0.75
2:R:142:ILE:HD13	2:R:155:ALA:HA	1.76	0.67
1:S:1030:LYS:HD2	1:S:1031:TYR:CZ	2.32	0.65
1:S:764:HIS:CE1	1:S:1038:PRO:HD3	2.33	0.64
1:S:797:ARG:NH1	4:S:1104:HOH:O	2.31	0.63
1:S:727:LYS:HE2	1:S:731:GLU:OE2	1.99	0.62
2:R:27:HIS:NE2	2:R:149:ARG:HD3	2.16	0.61
1:S:1026:ARG:HG3	1:S:1026:ARG:HH11	1.69	0.57
1:S:847:GLU:OE2	1:S:1029:LYS:CE	2.52	0.56
2:R:149:ARG:NH2	2:R:153:ASP:OD1	2.37	0.56
1:S:896:ARG:HG3	4:S:1335:HOH:O	2.04	0.56
2:R:1:MET:HA	4:R:297:HOH:O	2.06	0.56
1:S:939:LYS:HE2	4:R:248:HOH:O	2.06	0.56
1:S:1030:LYS:HD2	1:S:1031:TYR:CE1	2.41	0.55
1:S:771:ILE:HG13	1:S:772:GLU:N	2.21	0.55
1:S:1023:PRO:O	1:S:1025:PRO:HD3	2.08	0.54
2:R:123:ARG:NH2	2:R:143:GLU:OE1	2.40	0.54
1:S:920:ARG:NH1	4:S:1101:HOH:O	2.20	0.53
1:S:589:GLU:OE2	1:S:962:ARG:NH2	2.39	0.51
2:R:123:ARG:HH22	2:R:143:GLU:CD	2.15	0.49
2:R:21:ILE:O	2:R:25[A]:GLN:HG2	2.12	0.49
1:S:694:ARG:NH1	1:S:736:ILE:HD11	2.28	0.49
2:R:45:VAL:HG22	2:R:50:THR:HG22	1.94	0.48
1:S:872[A]:LEU:HD12	1:S:929:PHE:HB2	1.95	0.48
1:S:606:VAL:O	1:S:610[B]:ILE:HG12	2.14	0.47
1:S:796:TYR:HB2	1:S:824[A]:MET:HE3	1.96	0.46
1:S:1030:LYS:HE2	4:S:1307:HOH:O	2.14	0.46
1:S:912:TYR:CG	1:S:932:ILE:HD11	2.51	0.45
1:S:1019:ARG:HD2	4:S:1261:HOH:O	2.15	0.45
1:S:1041:ARG:HB2	1:S:1041:ARG:NH1	2.32	0.45
1:S:674:LEU:O	1:S:678[B]:ARG:HG3	2.16	0.45
1:S:721:VAL:HG22	1:S:730:VAL:CG2	2.47	0.44
2:R:12[B]:VAL:HG23	4:R:280:HOH:O	2.17	0.44
1:S:962:ARG:NH2	4:S:1107:HOH:O	2.44	0.44
1:S:639[B]:GLU:HG2	4:S:1151:HOH:O	2.17	0.44
2:R:88:LYS:HE2	2:R:88:LYS:HB2	1.85	0.44
1:S:824[A]:MET:HE2	4:S:1206:HOH:O	2.17	0.43
1:S:823:LYS:HD3	4:S:1434:HOH:O	2.19	0.43
1:S:872[B]:LEU:HD13	1:S:872[B]:LEU:HA	1.87	0.43
1:S:1029:LYS:HE2	1:S:1029:LYS:HB3	1.86	0.43
1:S:764:HIS:CD2	1:S:765:ILE:HG13	2.54	0.43
1:S:807:SER:HA	1:S:809:TRP:CZ3	2.53	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:R:27:HIS:HB3	2:R:28:PHE:CD2	2.56	0.41
2:R:83:ALA:HB3	2:R:86:ASN:HB3	2.03	0.41
2:R:112:VAL:HG11	2:R:142:ILE:HD12	2.03	0.41
1:S:726:MET:HB2	1:S:726:MET:HE2	1.71	0.41
1:S:821:LEU:HD12	1:S:824[B]:MET:HE2	2.03	0.41
1:S:824[A]:MET:HE1	1:S:930:PHE:HD2	1.85	0.41
1:S:918:LYS:HE3	4:S:1396:HOH:O	2.20	0.41
2:R:148:THR:O	2:R:149:ARG:HB2	2.20	0.41
1:S:836:GLU:O	1:S:840:VAL:HG22	2.21	0.41
2:R:25[B]:GLN:HE21	2:R:25[B]:GLN:HB2	1.66	0.41
1:S:824[A]:MET:CE	4:S:1206:HOH:O	2.69	0.40
2:R:84:ILE:HG12	2:R:117:LYS:HA	2.03	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	S	444/507 (88%)	437 (98%)	7 (2%)	0	100	100
2	R	168/190 (88%)	167 (99%)	1 (1%)	0	100	100
3	C000	118/128 (92%)	115 (98%)	3 (2%)	0	100	100
All	All	730/825 (88%)	719 (98%)	11 (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	S	424/469 (90%)	422 (100%)	2 (0%)	88 89
2	R	153/168 (91%)	152 (99%)	1 (1%)	84 84
3	C000	96/103 (93%)	96 (100%)	0	100 100
All	All	673/740 (91%)	670 (100%)	3 (0%)	88 91

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

Mol	Chain	Res	Type
1	S	683	GLN
1	S	930	PHE
2	R	27	HIS

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

Mol	Chain	Res	Type
1	S	764	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 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	S	440/507 (86%)	0.28	30 (6%) 17 19	25, 37, 68, 99	0
2	R	165/190 (86%)	0.40	15 (9%) 9 10	26, 38, 62, 85	0
3	C000	119/128 (92%)	0.04	6 (5%) 28 32	29, 36, 67, 89	0
All	All	724/825 (87%)	0.27	51 (7%) 16 18	25, 37, 68, 99	0

All (51) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	S	755	GLN	5.8
1	S	754	PHE	5.2
1	S	674	LEU	5.2
1	S	756	SER	4.8
2	R	27	HIS	4.8
2	R	168	GLU	4.7
3	C000	119	HIS	4.6
2	R	122	SER	4.4
1	S	572	ALA	4.4
1	S	591	ASN	4.3
1	S	1044	ASN	4.3
2	R	121	PRO	4.2
1	S	1043	SER	4.1
1	S	757	SER	4.0
2	R	1	MET	3.9
3	C000	1	GLN	3.6
2	R	30	ASP	3.6
1	S	1021	PRO	3.4
1	S	725	ALA	3.4
1	S	743	ALA	3.4
2	R	7	VAL	3.3
1	S	923	ASN	3.3
2	R	26	ASN	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	S	675	LYS	3.2
1	S	1041	ARG	3.1
1	S	742	ILE	2.9
1	S	676	ARG	2.9
1	S	573	ASP	2.8
1	S	1026	ARG	2.8
1	S	570	PRO	2.8
1	S	574	VAL	2.8
2	R	50	THR	2.7
1	S	771	ILE	2.7
2	R	49	GLU	2.7
1	S	724	LYS	2.6
1	S	652	GLU	2.5
3	C000	3	GLN	2.5
3	C000	26	ARG	2.4
1	S	1022	LYS	2.4
3	C000	25	SER	2.4
2	R	31	GLU	2.3
1	S	874	VAL	2.3
1	S	640	LEU	2.3
2	R	167	LYS	2.3
2	R	28	PHE	2.3
1	S	999	ASN	2.2
2	R	45	VAL	2.2
3	C000	74	ALA	2.2
1	S	610[A]	ILE	2.2
2	R	71	TYR	2.1
1	S	613	LEU	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 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.