Full wwPDB X-ray Structure Validation Report ⓘ

Aug 7, 2023 – 06:03 PM EDT

PDB ID	:	8E1E
Title	:	Scaffolding protein functional sites using deep learning
Authors	:	Bera, A.K.; Gerben, S.; Baker, D.
Deposited on	:	2022-08-10
Resolution	:	4.27 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:

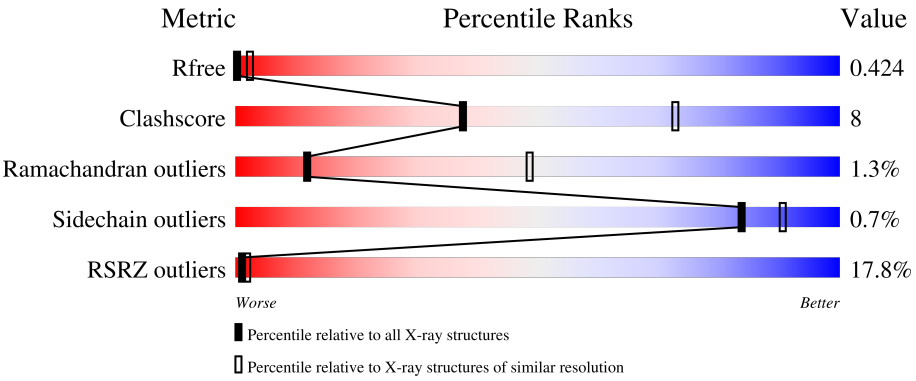
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 4.27 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	1017 (4.72-3.78)
Clashscore	141614	1059 (4.72-3.80)
Ramachandran outliers	138981	1014 (4.72-3.80)
Sidechain outliers	138945	1018 (4.72-3.78)
RSRZ outliers	127900	1072 (4.80-3.70)

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	243	22% 67% 28% • 5%
1	B	243	26% 77% 18% 5%
1	C	243	3% 81% 14% 5%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 5622 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 SG122_C3.

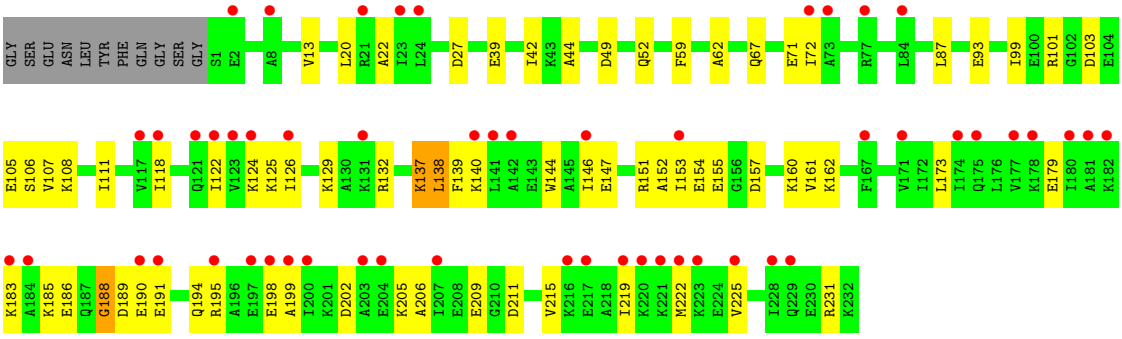
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	232	Total	C	N	O	S	0	0	0
			1874	1197	320	353	4			
1	B	232	Total	C	N	O	S	0	0	0
			1874	1197	320	353	4			
1	C	232	Total	C	N	O	S	0	0	0
			1874	1197	320	353	4			

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.

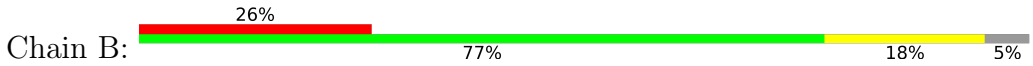
• Molecule 1: SG122_C3

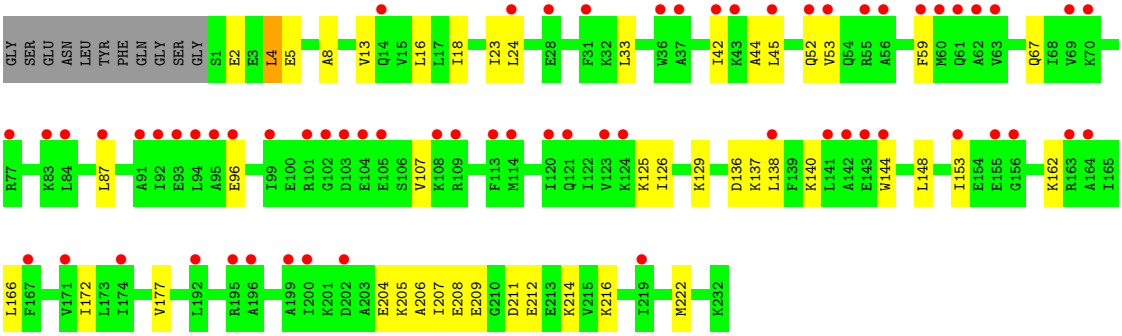
Chain A:



• Molecule 1: SG122_C3

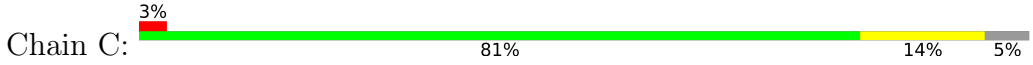
Chain B:

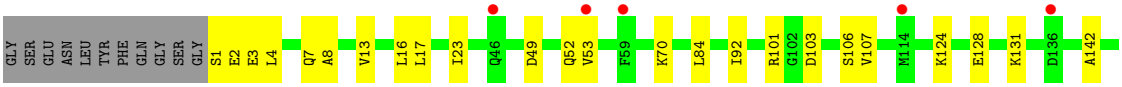


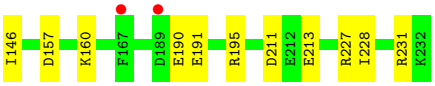


• Molecule 1: SG122_C3

Chain C:






I146
D157
K160
F167
D189
E190
E191
R195
D211
E212
E213
R227
I228
R231
K232

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, α, β, γ	234.09Å 234.09Å 234.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	95.57 – 4.27 95.57 – 4.27	Depositor EDS
% Data completeness (in resolution range)	99.7 (95.57-4.27) 99.8 (95.57-4.27)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.49 (at 4.30Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.405 , 0.427 0.403 , 0.424	Depositor DCC
R_{free} test set	741 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	282.1	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 238.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.37$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	0.147 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.82	EDS
Total number of atoms	5622	wwPDB-VP
Average B, all atoms (Å^2)	249.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.11% 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	0.27	0/1887	0.51	1/2517 (0.0%)
1	B	0.26	0/1887	0.51	1/2517 (0.0%)
1	C	0.26	0/1887	0.47	0/2517
All	All	0.26	0/5661	0.50	2/7551 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	138	LEU	CA-CB-CG	6.71	130.72	115.30
1	B	4	LEU	CA-CB-CG	5.12	127.09	115.30

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	1874	0	1984	47	0
1	B	1874	0	1984	27	0
1	C	1874	0	1984	23	0
All	All	5622	0	5952	89	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 8.

All (89) 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:B:2:GLU:OE2	1:B:4:LEU:HG	1.79	0.83
1:A:49:ASP:HB3	1:A:52:GLN:HG2	1.61	0.82
1:C:227:ARG:HH12	1:C:231:ARG:HB3	1.47	0.80
1:A:138:LEU:HD12	1:A:139:PHE:H	1.49	0.77
1:A:231:ARG:O	1:A:231:ARG:NH1	2.17	0.75
1:A:129:LYS:HE2	1:A:132:ARG:HH11	1.55	0.70
1:B:2:GLU:OE1	1:B:5:GLU:OE1	2.11	0.69
1:C:227:ARG:NH1	1:C:231:ARG:HB3	2.08	0.68
1:A:206:ALA:HB1	1:A:215:VAL:HG22	1.75	0.68
1:A:160:LYS:HG2	1:A:161:VAL:H	1.58	0.68
1:B:24:LEU:HD11	1:C:23:ILE:HG12	1.75	0.67
1:A:39:GLU:HG2	1:C:17:LEU:HD13	1.76	0.67
1:A:179:GLU:O	1:A:183:LYS:NZ	2.29	0.65
1:A:42:ILE:HD11	1:C:13:VAL:HG12	1.82	0.62
1:C:228:ILE:O	1:C:231:ARG:HG2	2.01	0.61
1:C:190:GLU:OE1	1:C:190:GLU:N	2.25	0.60
1:A:147:GLU:O	1:A:151:ARG:HG2	2.03	0.59
1:B:5:GLU:HB2	1:B:45:LEU:HD22	1.85	0.58
1:B:24:LEU:HD21	1:C:23:ILE:HG23	1.85	0.57
1:A:124:LYS:HG3	1:A:146:ILE:HD13	1.86	0.57
1:A:101:ARG:NH2	1:A:106:SER:OG	2.37	0.57
1:B:33:LEU:HD11	1:B:96:GLU:HG2	1.86	0.57
1:A:129:LYS:HE2	1:A:132:ARG:NH1	2.20	0.56
1:A:137:LYS:HG3	1:A:138:LEU:H	1.70	0.56
1:A:151:ARG:HA	1:A:154:GLU:HG2	1.86	0.56
1:A:122:ILE:O	1:A:126:ILE:HD12	2.06	0.56
1:A:194:GLN:O	1:A:198:GLU:HG2	2.05	0.56
1:A:211:ASP:HB3	1:A:215:VAL:HG23	1.87	0.56
1:C:157:ASP:HB3	1:C:160:LYS:HB2	1.88	0.55
1:C:8:ALA:HB2	1:C:53:VAL:HG13	1.86	0.55
1:C:191:GLU:OE2	1:C:195:ARG:NH2	2.40	0.54
1:C:49:ASP:OD2	1:C:52:GLN:NE2	2.41	0.53
1:C:128:GLU:HA	1:C:131:LYS:HG2	1.90	0.53
1:A:67:GLN:NE2	1:A:71:GLU:OE2	2.36	0.52
1:B:177:VAL:HG21	1:B:222:MET:HG2	1.90	0.52
1:A:87:LEU:HD22	1:A:153:ILE:HD13	1.92	0.52
1:B:13:VAL:HG13	1:C:16:LEU:HD11	1.91	0.51
1:A:188:GLY:O	1:A:190:GLU:N	2.44	0.51
1:B:125:LYS:HE2	1:B:129:LYS:HE3	1.93	0.51
1:B:204:GLU:O	1:B:207:ILE:HG23	2.10	0.51
1:A:103:ASP:O	1:A:107:VAL:HG23	2.11	0.51

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:13:VAL:HG13	1:B:16:LEU:HD11	1.93	0.50
1:A:103:ASP:HB3	1:A:106:SER:HB2	1.93	0.50
1:A:199:ALA:HB1	1:A:222:MET:HG2	1.92	0.50
1:B:211:ASP:OD1	1:B:214:LYS:HG2	2.12	0.50
1:A:138:LEU:CD1	1:A:139:PHE:H	2.22	0.49
1:B:205:LYS:C	1:B:207:ILE:H	2.15	0.48
1:A:138:LEU:O	1:A:140:LYS:N	2.43	0.48
1:A:155:GLU:HG3	1:A:157:ASP:HB2	1.94	0.48
1:B:166:LEU:HD21	1:B:216:LYS:HD3	1.95	0.47
1:A:195:ARG:HG3	1:A:225:VAL:HG22	1.97	0.47
1:A:13:VAL:HG12	1:B:42:ILE:HD11	1.96	0.47
1:C:211:ASP:OD1	1:C:213:GLU:HG2	2.15	0.47
1:B:18:ILE:HG12	1:B:67:GLN:HG3	1.97	0.46
1:C:103:ASP:O	1:C:107:VAL:HG23	2.16	0.46
1:A:44:ALA:HA	1:A:52:GLN:NE2	2.31	0.46
1:B:2:GLU:CD	1:B:4:LEU:HG	2.35	0.45
1:C:4:LEU:HA	1:C:7:GLN:OE1	2.16	0.45
1:A:188:GLY:O	1:A:191:GLU:N	2.44	0.45
1:B:59:PHE:CE1	1:B:107:VAL:HG21	2.51	0.45
1:A:140:LYS:O	1:A:144:TRP:N	2.43	0.45
1:A:22:ALA:HB1	1:A:27:ASP:O	2.17	0.45
1:A:160:LYS:HB3	1:A:160:LYS:HE2	1.56	0.44
1:C:142:ALA:O	1:C:146:ILE:HG12	2.17	0.44
1:A:152:ALA:HB1	1:A:161:VAL:HG22	1.98	0.44
1:A:125:LYS:O	1:A:129:LYS:N	2.45	0.44
1:A:59:PHE:HE1	1:A:99:ILE:HA	1.83	0.44
1:A:185:LYS:NZ	1:A:186:GLU:HB2	2.33	0.44
1:C:101:ARG:HD3	1:C:103:ASP:HB2	2.00	0.44
1:C:103:ASP:HB3	1:C:106:SER:HB2	2.00	0.43
1:A:160:LYS:O	1:A:162:LYS:N	2.46	0.43
1:B:144:TRP:CZ2	1:B:148:LEU:HD22	2.52	0.43
1:A:152:ALA:HB1	1:A:160:LYS:HE3	2.00	0.43
1:A:62:ALA:HB2	1:A:111:ILE:HD11	2.00	0.43
1:B:137:LYS:O	1:B:140:LYS:N	2.52	0.42
1:A:105:GLU:O	1:A:108:LYS:HG2	2.19	0.42
1:B:126:ILE:HD12	1:B:172:ILE:HG12	2.02	0.42
1:B:8:ALA:HB2	1:B:53:VAL:HG13	2.02	0.42
1:B:44:ALA:HB1	1:B:52:GLN:HB3	2.02	0.42
1:B:205:LYS:O	1:B:207:ILE:N	2.53	0.42
1:C:84:LEU:HB2	1:C:124:LYS:NZ	2.35	0.42
1:A:20:LEU:HB3	1:B:23:ILE:HD11	2.01	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:87:LEU:HD22	1:B:153:ILE:HG22	2.03	0.41
1:C:1:SER:C	1:C:3:GLU:H	2.24	0.41
1:C:70:LYS:HG2	1:C:92:ILE:HD13	2.02	0.41
1:A:173:LEU:HD11	1:A:219:ILE:HG23	2.02	0.41
1:B:166:LEU:HD21	1:B:216:LYS:CD	2.51	0.41
1:A:72:ILE:HD12	1:A:118:ILE:HG12	2.03	0.40
1:A:205:LYS:O	1:A:209:GLU:HG3	2.22	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	230/243 (95%)	216 (94%)	11 (5%)	3 (1%)	12 49
1	B	230/243 (95%)	220 (96%)	5 (2%)	5 (2%)	6 38
1	C	230/243 (95%)	226 (98%)	3 (1%)	1 (0%)	34 72
All	All	690/729 (95%)	662 (96%)	19 (3%)	9 (1%)	12 49

All (9) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	189	ASP
1	B	138	LEU
1	A	188	GLY
1	B	162	LYS
1	B	206	ALA
1	A	137	LYS
1	B	212	GLU
1	C	2	GLU
1	B	208	GLU

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	191/199 (96%)	189 (99%)	2 (1%)	76 86
1	B	191/199 (96%)	189 (99%)	2 (1%)	76 86
1	C	191/199 (96%)	191 (100%)	0	100 100
All	All	573/597 (96%)	569 (99%)	4 (1%)	84 90

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	93	GLU
1	A	202	ASP
1	B	136	ASP
1	B	209	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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	232/243 (95%)	1.18	53 (22%) 0 1	232, 252, 262, 266	0
1	B	232/243 (95%)	1.36	64 (27%) 0 1	241, 253, 261, 264	0
1	C	232/243 (95%)	0.53	7 (3%) 50 39	240, 245, 252, 258	0
All	All	696/729 (95%)	1.02	124 (17%) 1 2	232, 249, 260, 266	0

All (124) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	103	ASP	7.2
1	B	59	PHE	6.4
1	B	94	LEU	6.2
1	A	121	GLN	6.0
1	B	104	GLU	5.3
1	A	221	LYS	5.3
1	B	113	PHE	4.9
1	A	220	LYS	4.7
1	B	84	LEU	4.7
1	A	167	PHE	4.6
1	B	61	GLN	4.5
1	A	222	MET	4.4
1	A	181	ALA	4.4
1	A	174	ILE	4.4
1	B	83	LYS	4.4
1	B	95	ALA	4.3
1	B	60	MET	4.1
1	B	52	GLN	4.0
1	B	102	GLY	4.0
1	A	123	VAL	4.0
1	B	101	ARG	3.9
1	A	182	LYS	3.9
1	B	153	ILE	3.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	174	ILE	3.7
1	B	144	TRP	3.7
1	A	195	ARG	3.6
1	B	62	ALA	3.6
1	B	192	LEU	3.6
1	A	177	VAL	3.6
1	A	153	ILE	3.5
1	B	56	ALA	3.5
1	A	198	GLU	3.3
1	B	63	VAL	3.3
1	A	191	GLU	3.3
1	A	142	ALA	3.2
1	B	202	ASP	3.2
1	A	204	GLU	3.2
1	A	207	ILE	3.2
1	B	142	ALA	3.2
1	C	59	PHE	3.1
1	A	73	ALA	3.1
1	A	228	ILE	3.1
1	B	155	GLU	3.0
1	A	225	VAL	3.0
1	B	171	VAL	3.0
1	B	99	ILE	3.0
1	A	141	LEU	3.0
1	B	196	ALA	2.9
1	B	105	GLU	2.9
1	A	140	LYS	2.9
1	A	175	GLN	2.9
1	B	199	ALA	2.9
1	B	28	GLU	2.8
1	A	122	ILE	2.8
1	A	117	VAL	2.8
1	B	91	ALA	2.8
1	A	180	ILE	2.8
1	A	84	LEU	2.8
1	B	96	GLU	2.7
1	A	219	ILE	2.7
1	B	53	VAL	2.7
1	A	223	LYS	2.7
1	A	203	ALA	2.7
1	B	163	ARG	2.7
1	B	167	PHE	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	31	PHE	2.7
1	A	23	ILE	2.6
1	A	24	LEU	2.6
1	A	124	LYS	2.6
1	B	156	GLY	2.6
1	B	121	GLN	2.6
1	A	183	LYS	2.6
1	A	184	ALA	2.6
1	A	197	GLU	2.6
1	B	108	LYS	2.5
1	A	171	VAL	2.5
1	B	138	LEU	2.5
1	A	199	ALA	2.5
1	B	70	LYS	2.5
1	B	69	VAL	2.5
1	B	109	ARG	2.5
1	B	45	LEU	2.5
1	A	146	ILE	2.5
1	A	216	LYS	2.4
1	C	114	MET	2.4
1	B	55	ARG	2.4
1	B	114	MET	2.4
1	B	43	LYS	2.4
1	C	189	ASP	2.4
1	B	93	GLU	2.4
1	A	126	ILE	2.3
1	B	200	ILE	2.3
1	B	124	LYS	2.3
1	B	24	LEU	2.3
1	A	217	GLU	2.3
1	A	21	ARG	2.3
1	B	42	ILE	2.3
1	B	14	GLN	2.3
1	A	178	LYS	2.3
1	A	200	ILE	2.2
1	B	141	LEU	2.2
1	B	87	LEU	2.2
1	B	143	GLU	2.2
1	B	92	ILE	2.2
1	A	190	GLU	2.2
1	B	36	TRP	2.2
1	B	123	VAL	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	118	ILE	2.2
1	B	120	ILE	2.2
1	C	167	PHE	2.1
1	A	77	ARG	2.1
1	C	46	GLN	2.1
1	A	8	ALA	2.1
1	B	37	ALA	2.1
1	B	195	ARG	2.1
1	A	72	ILE	2.1
1	A	131	LYS	2.1
1	A	229	GLN	2.1
1	B	219	ILE	2.0
1	C	136	ASP	2.0
1	A	2	GLU	2.0
1	B	164	ALA	2.0
1	C	53	VAL	2.0
1	B	77	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.