Full wwPDB X-ray Structure Validation Report ⓘ

May 15, 2020 – 05:29 am BST

PDB ID	:	4YOV
Title	:	Crystal structure of a trimeric exonuclease PhoExo I from Pyrococcus horikoshii OT3 in complex with poly-dA
Authors	:	Miyazono, K.; Ito, T.; Tanokura, M.
Deposited on	:	2015-03-12
Resolution	:	2.05 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:

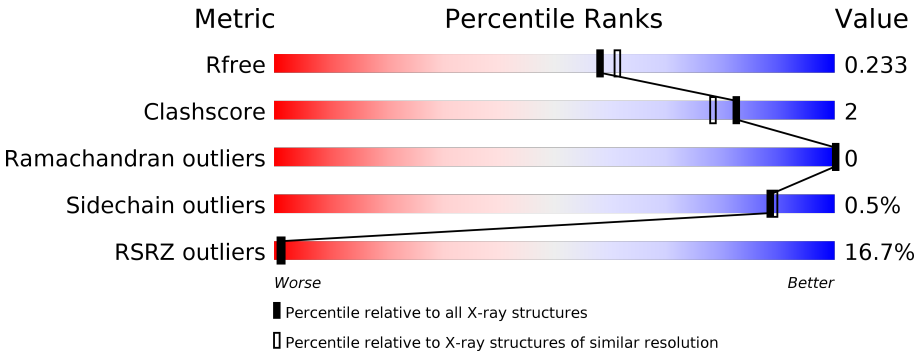
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.05 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	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 on 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	233	11% 91% 6%
1	C	233	12% 92% 5%
1	E	233	27% 87% 9%
2	B	7	43% 14% 43%
2	D	7	57% 43%
2	F	7	29% 29% 43%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5709 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 3-5 exonuclease PhoExo I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	226	Total	C	N	O	S	0	0	0
			1781	1149	296	333	3			
1	C	226	Total	C	N	O	S	0	0	0
			1781	1149	296	333	3			
1	E	224	Total	C	N	O	S	0	0	0
			1766	1140	293	330	3			

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	80	ASN	ASP	engineered mutation	UNP A0A060P168
A	230	HIS	-	expression tag	UNP A0A060P168
A	231	HIS	-	expression tag	UNP A0A060P168
A	232	HIS	-	expression tag	UNP A0A060P168
A	233	HIS	-	expression tag	UNP A0A060P168
C	80	ASN	ASP	engineered mutation	UNP A0A060P168
C	230	HIS	-	expression tag	UNP A0A060P168
C	231	HIS	-	expression tag	UNP A0A060P168
C	232	HIS	-	expression tag	UNP A0A060P168
C	233	HIS	-	expression tag	UNP A0A060P168
E	80	ASN	ASP	engineered mutation	UNP A0A060P168
E	230	HIS	-	expression tag	UNP A0A060P168
E	231	HIS	-	expression tag	UNP A0A060P168
E	232	HIS	-	expression tag	UNP A0A060P168
E	233	HIS	-	expression tag	UNP A0A060P168

- Molecule 2 is a DNA chain called 5'-D(*AP*AP*AP*AP*AP*AP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	4	Total	C	N	O	P	0	0	0
			84	40	20	20	4			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	4	Total	C	N	O	P	0	0	0
			84	40	20	20	4			
2	F	4	Total	C	N	O	P	0	0	0
			84	40	20	20	4			

- Molecule 3 is water.

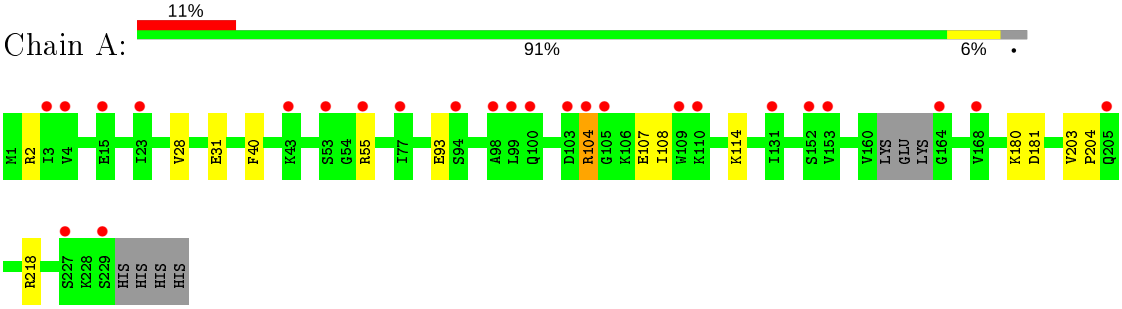
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	52	Total	O	0	0
			52	52		
3	B	2	Total	O	0	0
			2	2		
3	C	58	Total	O	0	0
			58	58		
3	D	1	Total	O	0	0
			1	1		
3	E	16	Total	O	0	0
			16	16		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

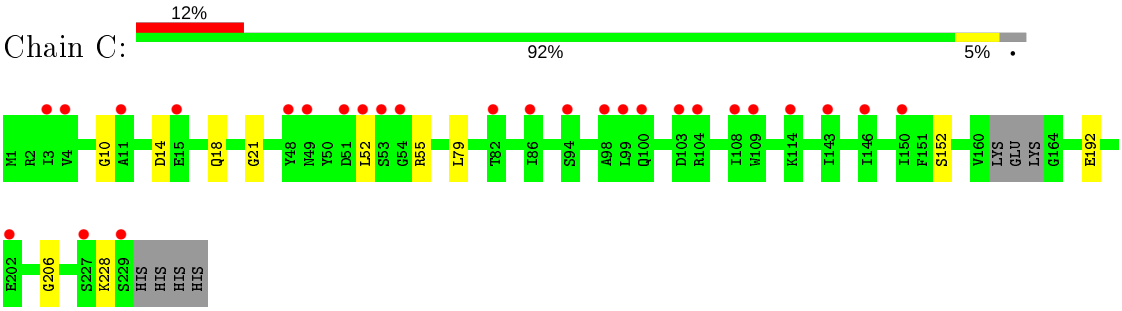
- Molecule 1: 3-5 exonuclease PhoExo I

Chain A:



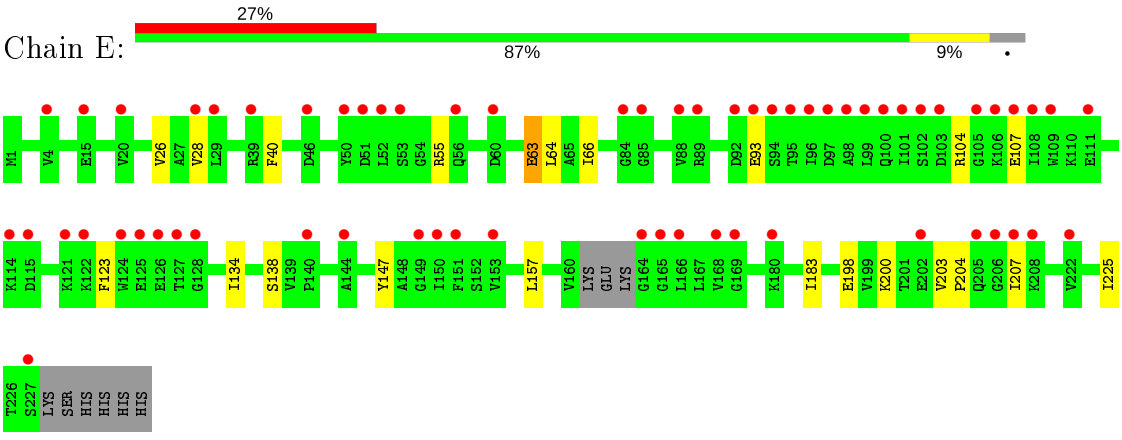
- Molecule 1: 3-5 exonuclease PhoExo I

Chain C:



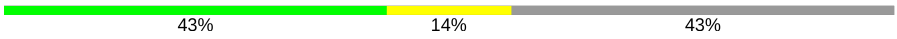
- Molecule 1: 3-5 exonuclease PhoExo I

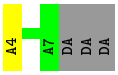
Chain E:



- Molecule 2: 5'-D(*AP*AP*AP*AP*AP*AP*A)-3'

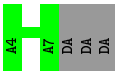
Chain B:





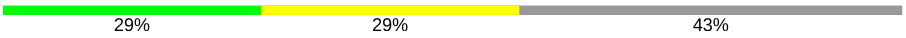
- Molecule 2: 5'-D(*AP*AP*AP*AP*AP*AP*A)-3'

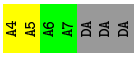
Chain D:



- Molecule 2: 5'-D(*AP*AP*AP*AP*AP*AP*A)-3'

Chain F:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	94.68Å 95.10Å 105.27Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.80 – 2.05 19.80 – 2.05	Depositor EDS
% Data completeness (in resolution range)	98.7 (19.80-2.05) 99.2 (19.80-2.05)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	1.70 (at 2.06Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.191 , 0.233 0.194 , 0.233	Depositor DCC
R_{free} test set	3012 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	36.7	Xtriage
Anisotropy	0.798	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 61.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.039 for k,h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5709	wwPDB-VP
Average B, all atoms (Å^2)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.54% 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	0.42	0/1813	0.57	0/2450
1	C	0.45	0/1813	0.56	0/2450
1	E	0.34	0/1798	0.52	0/2431
2	B	0.87	0/95	0.70	0/144
2	D	0.68	0/95	0.68	0/144
2	F	0.68	0/95	0.71	0/144
All	All	0.43	0/5709	0.56	0/7763

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	A	1781	0	1842	10	0
1	C	1781	0	1842	6	0
1	E	1766	0	1824	13	0
2	B	84	0	45	1	0
2	D	84	0	45	0	0
2	F	84	0	45	2	0
3	A	52	0	0	0	0
3	B	2	0	0	0	0
3	C	58	0	0	0	0
3	D	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	E	16	0	0	0	0
All	All	5709	0	5643	28	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 2.

All (28) 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:218:ARG:NH2	1:C:192:GLU:OE1	2.28	0.66
1:A:55:ARG:NH2	2:B:4:DA:N1	2.45	0.65
1:E:63:GLU:HG3	1:E:64:LEU:N	2.18	0.57
1:E:198:GLU:OE2	1:E:200:LYS:NZ	2.39	0.56
1:E:207:ILE:HD12	1:E:225:ILE:HD11	1.91	0.53
1:E:28:VAL:HG12	1:E:40:PHE:HB3	1.93	0.50
1:E:134:ILE:HD13	1:E:138:SER:HB2	1.94	0.49
1:E:55:ARG:NH2	2:F:5:DA:H1'	2.27	0.49
1:A:104:ARG:HD3	1:A:107:GLU:OE2	2.13	0.48
1:E:26:VAL:HG11	1:E:147:TYR:HB3	1.97	0.46
1:A:181:ASP:N	1:A:181:ASP:OD1	2.46	0.46
1:C:10:GLY:HA3	1:C:152:SER:OG	2.17	0.45
1:C:79:LEU:HA	1:C:79:LEU:HD12	1.89	0.44
1:A:2:ARG:HG2	1:A:31:GLU:HG2	2.00	0.44
1:E:104:ARG:O	1:E:107:GLU:HG2	2.18	0.43
1:E:66:ILE:HG23	1:E:123:PHE:CE1	2.54	0.43
1:A:104:ARG:CZ	1:A:108:ILE:HD11	2.48	0.43
1:E:157:LEU:HD12	1:E:183:ILE:HD11	1.99	0.43
1:A:180:LYS:HE2	1:A:180:LYS:HB3	1.77	0.42
1:C:206:GLY:HA2	1:C:228:LYS:HD2	2.01	0.42
1:C:21:GLY:HA2	1:C:52:LEU:HD21	2.02	0.41
1:A:93:GLU:OE1	1:A:114:LYS:NZ	2.53	0.41
1:E:93:GLU:OE1	1:E:93:GLU:N	2.47	0.41
1:E:55:ARG:HD2	2:F:4:DA:H2	1.85	0.41
1:A:203:VAL:HA	1:A:204:PRO:HD2	1.89	0.41
1:C:14:ASP:OD2	1:C:18:GLN:HB2	2.22	0.40
1:E:203:VAL:HA	1:E:204:PRO:HD2	1.97	0.40
1:A:28:VAL:HG12	1:A:40:PHE:HB3	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	222/233 (95%)	217 (98%)	5 (2%)	0	100	100
1	C	222/233 (95%)	217 (98%)	5 (2%)	0	100	100
1	E	220/233 (94%)	213 (97%)	7 (3%)	0	100	100
All	All	664/699 (95%)	647 (97%)	17 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	193/200 (96%)	192 (100%)	1 (0%)	88	89
1	C	193/200 (96%)	192 (100%)	1 (0%)	88	89
1	E	191/200 (96%)	190 (100%)	1 (0%)	88	89
All	All	577/600 (96%)	574 (100%)	3 (0%)	88	89

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

Mol	Chain	Res	Type
1	A	104	ARG
1	C	55	ARG
1	E	63	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no

such sidechains identified.

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 carbohydrates 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, 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	226/233 (96%)	0.74	25 (11%) 5 5	33, 53, 109, 131	0
1	C	226/233 (96%)	0.81	27 (11%) 4 4	33, 54, 115, 146	0
1	E	224/233 (96%)	1.42	63 (28%) 0 0	39, 80, 141, 171	0
2	B	4/7 (57%)	0.43	0 100 100	74, 84, 90, 102	0
2	D	4/7 (57%)	1.41	0 100 100	95, 112, 114, 142	0
2	F	4/7 (57%)	0.72	0 100 100	85, 102, 112, 122	0
All	All	688/720 (95%)	0.99	115 (16%) 1 1	33, 62, 128, 171	0

All (115) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	103	ASP	8.6
1	C	53	SER	8.3
1	E	100	GLN	6.0
1	E	122	LYS	5.9
1	A	100	GLN	5.8
1	C	52	LEU	5.7
1	E	126	GLU	5.7
1	E	206	GLY	5.5
1	C	94	SER	5.4
1	C	99	LEU	5.4
1	C	100	GLN	5.3
1	E	98	ALA	5.1
1	E	124	TRP	4.8
1	E	52	LEU	4.7
1	C	15	GLU	4.6
1	C	98	ALA	4.4
1	C	227	SER	4.2
1	E	101	ILE	4.2
1	E	128	GLY	4.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	E	97	ASP	4.1
1	E	105	GLY	4.1
1	C	229	SER	4.0
1	E	93	GLU	4.0
1	A	103	ASP	3.8
1	E	127	THR	3.8
1	E	56	GLN	3.8
1	E	102	SER	3.8
1	E	202	GLU	3.7
1	A	98	ALA	3.7
1	A	229	SER	3.7
1	E	4	VAL	3.6
1	E	99	LEU	3.6
1	E	28	VAL	3.5
1	E	115	ASP	3.5
1	C	51	ASP	3.4
1	E	95	THR	3.4
1	C	49	ASN	3.4
1	E	108	ILE	3.4
1	E	168	VAL	3.3
1	A	227	SER	3.3
1	A	99	LEU	3.2
1	E	208	LYS	3.2
1	E	88	VAL	3.2
1	E	111	GLU	3.2
1	E	165	GLY	3.2
1	E	53	SER	3.2
1	E	205	GLN	3.2
1	E	164	GLY	3.1
1	C	103	ASP	3.1
1	E	207	ILE	3.1
1	E	222	VAL	3.0
1	E	107	GLU	3.0
1	E	89	ARG	3.0
1	C	150	ILE	3.0
1	A	94	SER	3.0
1	A	109	TRP	2.9
1	C	143	ILE	2.9
1	E	227	SER	2.9
1	A	55	ARG	2.9
1	E	109	TRP	2.9
1	C	48	TYR	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	82	THR	2.8
1	A	105	GLY	2.8
1	A	131	ILE	2.8
1	E	51	ASP	2.8
1	E	180	LYS	2.8
1	E	144	ALA	2.8
1	E	153	VAL	2.8
1	E	121	LYS	2.8
1	A	53	SER	2.8
1	E	151	PHE	2.7
1	C	4	VAL	2.7
1	C	104	ARG	2.7
1	A	4	VAL	2.6
1	E	114	LYS	2.6
1	C	202	GLU	2.6
1	E	20	VAL	2.6
1	E	94	SER	2.6
1	E	85	GLY	2.6
1	E	92	ASP	2.6
1	C	109	TRP	2.6
1	C	54	GLY	2.5
1	C	114	LYS	2.5
1	E	84	GLY	2.5
1	A	15	GLU	2.5
1	E	29	LEU	2.4
1	A	104	ARG	2.4
1	A	3	ILE	2.4
1	A	77	ILE	2.4
1	A	153	VAL	2.4
1	E	39	ARG	2.4
1	A	23	ILE	2.3
1	A	43	LYS	2.3
1	E	149	GLY	2.3
1	C	86	ILE	2.3
1	E	96	ILE	2.2
1	C	11	ALA	2.2
1	A	110	LYS	2.2
1	E	169	GLY	2.2
1	A	168	VAL	2.2
1	E	15	GLU	2.2
1	E	50	TYR	2.2
1	E	125	GLU	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	152	SER	2.2
1	E	60	ASP	2.1
1	A	205	GLN	2.1
1	E	140	PRO	2.1
1	E	150	ILE	2.1
1	A	164	GLY	2.1
1	C	146	ILE	2.0
1	E	166	LEU	2.0
1	E	106	LYS	2.0
1	C	3	ILE	2.0
1	C	108	ILE	2.0
1	E	46	ASP	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 carbohydrates 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.