



Full wwPDB X-ray Structure Validation Report i

Jun 16, 2024 – 12:07 PM EDT

PDB ID : 4YOX

Title : Crystal structure of a trimeric exonuclease PhoExo I from Pyrococcus horikoshii OT3 in complex with poly-dT

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 i 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](#) ①) were used in the production of this report:

MolProbity : 4.02b-467

Xtriage (Phenix) : 1.13

EDS : 2.37.1

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) : 2.37.1

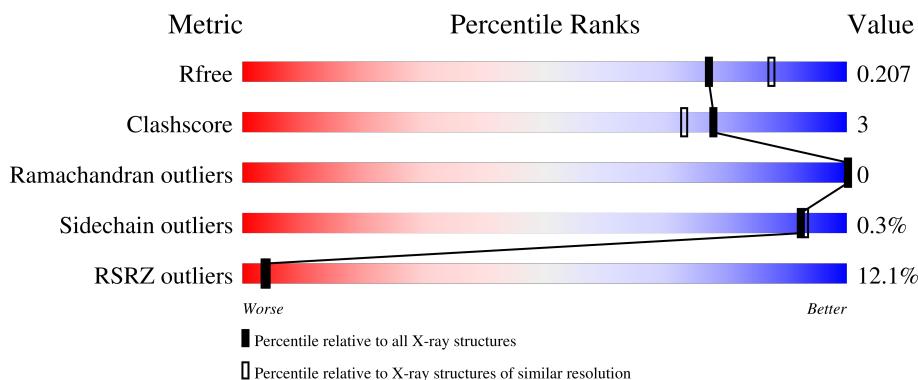
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

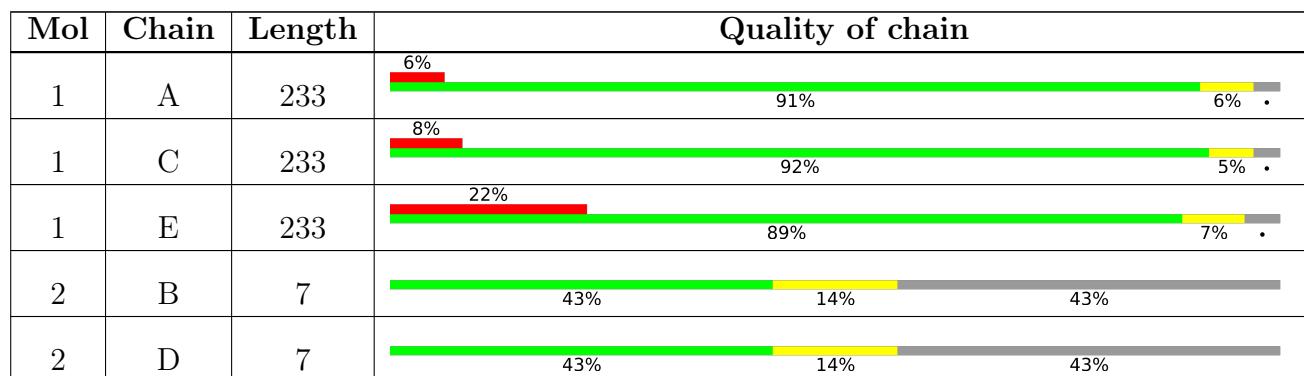
The reported resolution of this entry is 2.05 Å.

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 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 $\leq 5\%$. 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.



Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain	
2	F	7	<div style="width: 57%;">57%</div>	<div style="width: 43%;">43%</div>

2 Entry composition [\(i\)](#)

There are 3 unique types of molecules in this entry. The entry contains 5719 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(*TP*TP*TP*TP*TP*TP*TP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	4	Total	C	N	O	P	0	0	0
			80	40	8	28	4			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	D	4	Total C N O P 80 40 8 28 4	0	0	0
2	F	4	Total C N O P 80 40 8 28 4	0	0	0

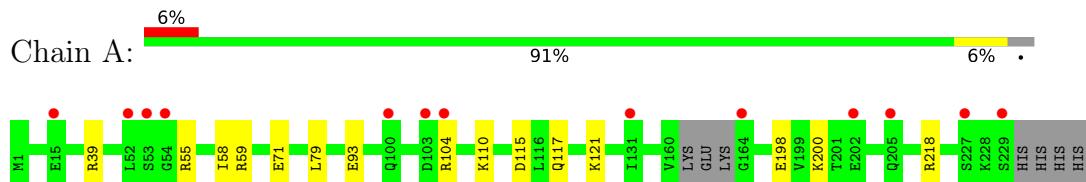
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	58	Total O 58 58	0	0
3	B	2	Total O 2 2	0	0
3	C	63	Total O 63 63	0	0
3	D	2	Total O 2 2	0	0
3	E	23	Total O 23 23	0	0
3	F	3	Total O 3 3	0	0

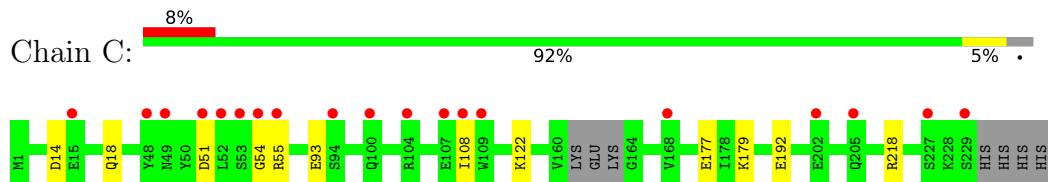
3 Residue-property plots [\(i\)](#)

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.

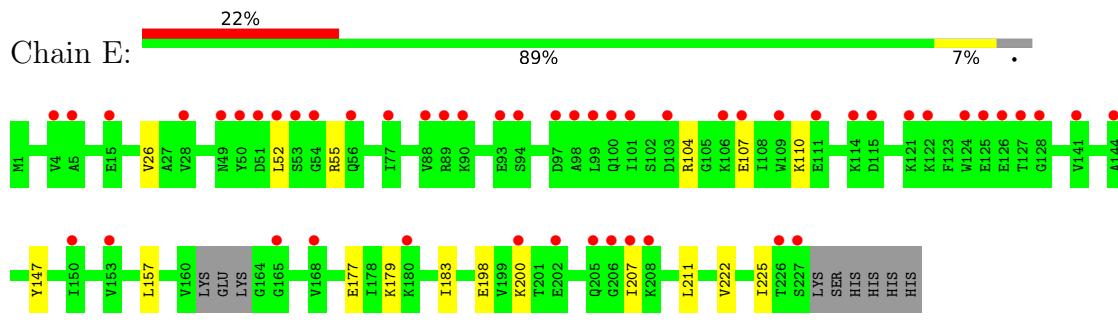
- Molecule 1: 3-5 exonuclease PhoExo I



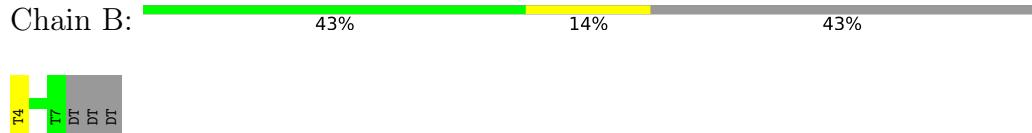
- Molecule 1: 3-5 exonuclease PhoExo II



- Molecule 1: 3'-5' exonuclease PhoExo I



- Molecule 2: 5'-D(*TP*TP*TP*TP*TP*TP*TP*T)-3'



- Molecule 2: 5'-D(*TP*TP*TP*TP*TP*TP*T)-3'



- Molecule 2: 5'-D(*TP*TP*TP*TP*TP*TP*T)-3'

Chain F:  57% 43%



4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	94.74 Å 94.75 Å 105.15 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.76 – 2.05 19.76 – 2.05	Depositor EDS
% Data completeness (in resolution range)	99.9 (19.76-2.05) 100.0 (19.76-2.05)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ ¹	1.83 (at 2.06 Å)	Xtriage
Refinement program	PHENIX	Depositor
R , R_{free}	0.179 , 0.207 0.180 , 0.207	Depositor DCC
R_{free} test set	3027 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	35.6	Xtriage
Anisotropy	0.689	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 59.6	EDS
L-test for twinning ²	$< L > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.034 for k,h,-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	5719	wwPDB-VP
Average B, all atoms (Å ²)	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.52% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [\(i\)](#)

5.1 Standard geometry [\(i\)](#)

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	0.41	0/1813	0.59	0/2450
1	C	0.44	0/1813	0.57	0/2450
1	E	0.32	0/1798	0.51	0/2431
2	B	0.93	0/87	1.37	0/132
2	D	0.73	0/87	1.37	0/132
2	F	0.68	0/87	1.42	0/132
All	All	0.42	0/5685	0.63	0/7727

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 [\(i\)](#)

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	11	0
1	C	1781	0	1842	12	0
1	E	1766	0	1824	10	0
2	B	80	0	49	2	0
2	D	80	0	49	1	0
2	F	80	0	49	0	0
3	A	58	0	0	0	0
3	B	2	0	0	0	0
3	C	63	0	0	1	0
3	D	2	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	23	0	0	0	0
3	F	3	0	0	0	0
All	All	5719	0	5655	32	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 (32) 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:55:ARG:HH22	2:B:4:DT:H3	1.20	0.90
1:E:107:GLU:HA	1:E:110:LYS:HD2	1.68	0.74
1:A:55:ARG:NH2	2:B:4:DT:H3	1.89	0.68
1:E:177:GLU:OE2	1:E:179:LYS:NZ	2.28	0.67
1:C:55:ARG:NH2	1:C:108:ILE:HD13	2.11	0.65
1:C:177:GLU:OE2	1:C:179:LYS:NZ	2.35	0.59
1:A:218:ARG:NH2	1:C:192:GLU:OE1	2.34	0.59
1:E:198:GLU:OE2	1:E:200:LYS:NZ	2.36	0.58
1:A:104:ARG:HH21	1:A:104:ARG:HG2	1.74	0.53
1:A:39:ARG:NH2	1:A:71:GLU:OE1	2.41	0.52
1:C:55:ARG:NH2	1:C:108:ILE:HG21	2.27	0.50
1:E:211:LEU:HD12	1:E:222:VAL:O	2.12	0.49
1:A:39:ARG:HH21	1:A:71:GLU:CD	2.18	0.48
1:A:59:ARG:NH2	1:A:115:ASP:OD2	2.44	0.47
1:C:55:ARG:N	1:C:55:ARG:HD3	2.28	0.46
1:E:26:VAL:HG11	1:E:147:TYR:HB3	1.97	0.46
1:C:55:ARG:NH2	2:D:4:DT:N3	2.64	0.46
1:C:55:ARG:HH22	1:C:108:ILE:HD13	1.80	0.45
1:E:55:ARG:HD3	1:E:55:ARG:HA	1.61	0.45
1:E:104:ARG:NH2	1:E:107:GLU:OE2	2.41	0.45
1:E:52:LEU:O	1:E:55:ARG:NH2	2.29	0.44
1:C:51:ASP:O	1:C:54:GLY:HA2	2.18	0.44
1:E:157:LEU:HD12	1:E:183:ILE:HD11	1.99	0.43
1:C:55:ARG:CZ	1:C:108:ILE:HG21	2.49	0.42
1:A:117:GLN:HE22	1:A:121:LYS:HE3	1.85	0.41
1:C:122:LYS:HA	1:C:122:LYS:HD2	1.84	0.41
1:A:79:LEU:HD12	1:A:79:LEU:HA	1.91	0.41
1:A:93:GLU:HG3	1:A:110:LYS:HG2	2.03	0.41
1:C:14:ASP:OD2	1:C:18:GLN:HB2	2.21	0.41
1:C:218:ARG:NH2	3:C:303:HOH:O	2.50	0.41
1:E:207:ILE:HD12	1:E:225:ILE:HD11	2.03	0.40

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:198:GLU:OE1	1:A:200:LYS:NZ	2.54	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	222/233 (95%)	218 (98%)	4 (2%)	0	100 100
1	C	222/233 (95%)	217 (98%)	5 (2%)	0	100 100
1	E	220/233 (94%)	216 (98%)	4 (2%)	0	100 100
All	All	664/699 (95%)	651 (98%)	13 (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	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%)	191 (100%)	0	100 100
All	All	577/600 (96%)	575 (100%)	2 (0%)	92 93

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	58	ILE
1	C	93	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 i

6.1 Protein, DNA and RNA chains i

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, 95th 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%)	0.39	13 (5%) 23 25	31, 49, 92, 117	0
1	C	226/233 (96%)	0.43	19 (8%) 11 11	32, 52, 96, 122	0
1	E	224/233 (96%)	1.09	51 (22%) 0 0	40, 78, 132, 158	0
2	B	4/7 (57%)	-0.39	0 100 100	48, 60, 63, 84	0
2	D	4/7 (57%)	-0.06	0 100 100	60, 73, 83, 100	0
2	F	4/7 (57%)	-0.07	0 100 100	68, 82, 92, 101	0
All	All	688/720 (95%)	0.62	83 (12%) 4 4	31, 59, 119, 158	0

All (83) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	53	SER	9.7
1	E	103	ASP	7.1
1	E	52	LEU	6.2
1	E	53	SER	5.8
1	E	100	GLN	5.6
1	A	53	SER	5.2
1	C	52	LEU	5.2
1	E	54	GLY	5.0
1	C	54	GLY	4.9
1	E	206	GLY	4.8
1	E	124	TRP	4.7
1	E	93	GLU	4.5
1	E	114	LYS	4.3
1	E	122	LYS	4.1
1	E	202	GLU	4.0
1	C	51	ASP	4.0
1	A	229	SER	4.0
1	C	229	SER	4.0
1	A	100	GLN	4.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	E	227	SER	3.8
1	E	126	GLU	3.7
1	C	100	GLN	3.7
1	E	168	VAL	3.7
1	C	55	ARG	3.6
1	E	97	ASP	3.5
1	E	208	LYS	3.4
1	E	51	ASP	3.4
1	E	128	GLY	3.4
1	E	226	THR	3.3
1	C	227	SER	3.3
1	E	28	VAL	3.3
1	A	54	GLY	3.3
1	E	107	GLU	3.3
1	C	49	ASN	3.2
1	E	101	ILE	3.2
1	E	94	SER	3.1
1	E	50	TYR	3.0
1	E	205	GLN	2.9
1	E	98	ALA	2.9
1	C	15	GLU	2.9
1	E	89	ARG	2.8
1	E	109	TRP	2.8
1	A	227	SER	2.8
1	E	144	ALA	2.8
1	E	180	LYS	2.8
1	A	52	LEU	2.8
1	E	56	GLN	2.8
1	E	88	VAL	2.8
1	A	205	GLN	2.8
1	E	111	GLU	2.6
1	C	94	SER	2.6
1	A	103	ASP	2.6
1	E	121	LYS	2.5
1	E	153	VAL	2.5
1	E	106	LYS	2.5
1	C	202	GLU	2.5
1	E	125	GLU	2.5
1	E	115	ASP	2.4
1	A	164	GLY	2.4
1	E	4	VAL	2.4
1	C	205	GLN	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	107	GLU	2.4
1	C	109	TRP	2.4
1	E	5	ALA	2.3
1	A	15	GLU	2.3
1	C	108	ILE	2.2
1	E	127	THR	2.2
1	E	90	LYS	2.2
1	E	150	ILE	2.2
1	E	99	LEU	2.2
1	C	48	TYR	2.2
1	E	165	GLY	2.1
1	A	104	ARG	2.1
1	E	200	LYS	2.1
1	A	131	ILE	2.1
1	A	202	GLU	2.1
1	E	49	ASN	2.0
1	E	207	ILE	2.0
1	C	168	VAL	2.0
1	E	141	VAL	2.0
1	E	15	GLU	2.0
1	E	77	ILE	2.0
1	C	104	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.