



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 12, 2023 – 12:04 PM EDT

PDB ID : 8HKA  
Title : TPA bound-form of Periplasmic terephthalate binding protein (TBP) from Ideonella sakaiensis  
Authors : Lee, S.H.; Seo, H.; Kim, K.-J.  
Deposited on : 2022-11-25  
Resolution : 1.02 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](mailto: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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.33  
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) : 2.33

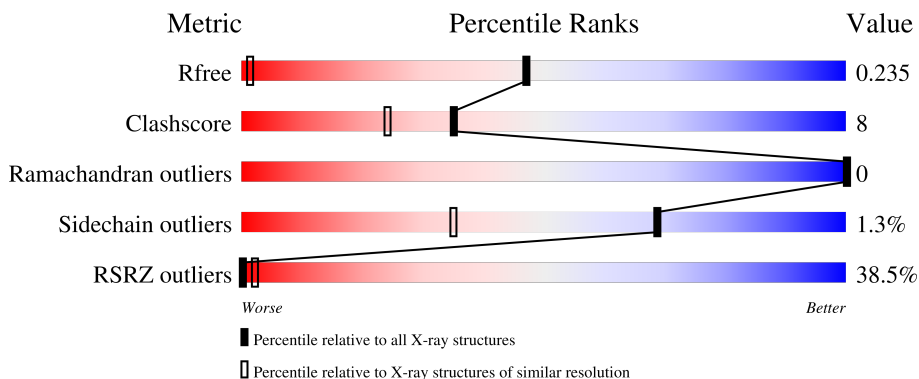
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.02 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	1188 (1.08-0.96)
Clashscore	141614	1253 (1.08-0.96)
Ramachandran outliers	138981	1178 (1.08-0.96)
Sidechain outliers	138945	1180 (1.08-0.96)
RSRZ outliers	127900	1158 (1.08-0.96)

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  $\geq 3$ , 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.

Mol	Chain	Length	Quality of chain
1	A	308	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4821 atoms, of which 2267 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.

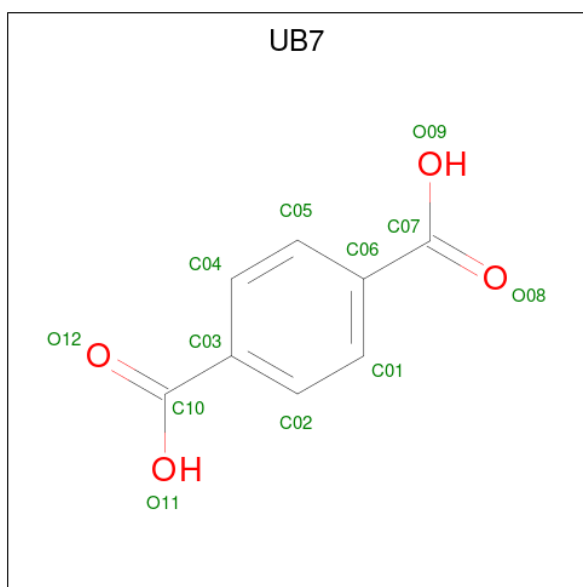
- Molecule 1 is a protein called Periplasmic terephthalate binding protein (TBP).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	299	4517	1437	2267	398	410	5	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	SER	-	expression tag	UNP A0A0K8P8D2
A	24	HIS	-	expression tag	UNP A0A0K8P8D2
A	25	MET	-	expression tag	UNP A0A0K8P8D2
A	323	LEU	-	expression tag	UNP A0A0K8P8D2
A	324	GLU	-	expression tag	UNP A0A0K8P8D2
A	325	HIS	-	expression tag	UNP A0A0K8P8D2
A	326	HIS	-	expression tag	UNP A0A0K8P8D2
A	327	HIS	-	expression tag	UNP A0A0K8P8D2
A	328	HIS	-	expression tag	UNP A0A0K8P8D2
A	329	HIS	-	expression tag	UNP A0A0K8P8D2
A	330	HIS	-	expression tag	UNP A0A0K8P8D2

- Molecule 2 is terephthalic acid (three-letter code: UB7) (formula: C<sub>8</sub>H<sub>6</sub>O<sub>4</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	8	4		

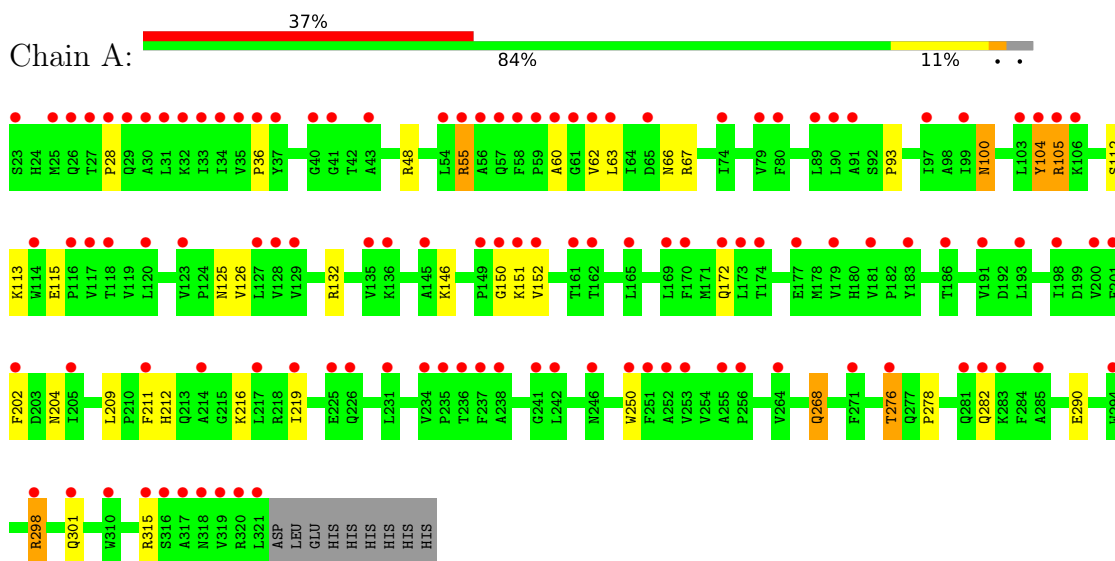
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	292	Total	O	0	0
			292	292		

### 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: Periplasmic terephthalate binding protein (TBP)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	46.18Å 54.10Å 50.60Å 90.00° 104.55° 90.00°	Depositor
Resolution (Å)	20.67 – 1.02 20.66 – 1.02	Depositor EDS
% Data completeness (in resolution range)	99.4 (20.67-1.02) 99.4 (20.66-1.02)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.48 (at 1.02Å)	Xtrriage
Refinement program	REFMAC 8.0	Depositor
R, $R_{free}$	0.204 , 0.229 0.211 , 0.235	Depositor DCC
$R_{free}$ test set	6204 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	9.1	Xtrriage
Anisotropy	0.489	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.57 , 64.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4821	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	12.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.40% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: UB7

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	0.60	1/2308 (0.0%)	0.91	8/3153 (0.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	290	GLU	CD-OE1	-5.06	1.20	1.25

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	276	THR	CB-CA-C	-9.18	86.83	111.60
1	A	298	ARG	NE-CZ-NH2	6.19	123.40	120.30
1	A	132	ARG	NE-CZ-NH1	6.17	123.38	120.30
1	A	105	ARG	NE-CZ-NH2	-6.16	117.22	120.30
1	A	105	ARG	NE-CZ-NH1	5.82	123.21	120.30
1	A	100	ASN	CB-CA-C	5.73	121.87	110.40
1	A	276	THR	CA-CB-OG1	5.57	120.70	109.00
1	A	104	TYR	CB-CG-CD1	5.16	124.10	121.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	55	ARG	Sidechain

## 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	2250	2267	2265	37	0
2	A	12	0	0	0	0
3	A	292	0	0	21	0
All	All	2554	2267	2265	37	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 (37) 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:315:ARG:HD3	3:A:678:HOH:O	1.63	0.98
1:A:150:GLY:HA2	3:A:663:HOH:O	1.64	0.97
1:A:115:GLU:OE2	3:A:501:HOH:O	1.86	0.93
1:A:212:HIS:HD2	1:A:219:ILE:H	1.23	0.84
1:A:276:THR:HG22	3:A:698:HOH:O	1.80	0.81
1:A:301:GLN:OE1	3:A:502:HOH:O	1.98	0.81
1:A:146:LYS:HE3	3:A:633:HOH:O	1.80	0.79
1:A:146:LYS:HD3	3:A:571:HOH:O	1.83	0.79
1:A:172:GLN:HG2	3:A:627:HOH:O	1.87	0.75
1:A:48:ARG:HH21	1:A:66:ASN:HD21	1.37	0.71
1:A:212:HIS:CD2	1:A:219:ILE:H	2.07	0.69
1:A:315:ARG:HB2	3:A:686:HOH:O	1.95	0.67
1:A:315:ARG:CD	3:A:678:HOH:O	2.31	0.63
1:A:150:GLY:CA	3:A:504:HOH:O	2.47	0.62
1:A:28:PRO:HB2	1:A:60:ALA:HB2	1.83	0.59
1:A:146:LYS:CE	3:A:633:HOH:O	2.44	0.58
1:A:115:GLU:CD	3:A:501:HOH:O	2.35	0.57
1:A:315:ARG:HG3	1:A:315:ARG:HH21	1.70	0.56
1:A:100:ASN:HB2	1:A:104:TYR:CE1	2.43	0.54
1:A:150:GLY:N	3:A:504:HOH:O	2.42	0.53

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:151:LYS:N	3:A:504:HOH:O	2.41	0.52
1:A:276:THR:HA	3:A:698:HOH:O	2.09	0.52
1:A:62:VAL:O	1:A:63:LEU:HD12	2.10	0.52
1:A:105:ARG:CZ	3:A:513:HOH:O	2.58	0.52
1:A:150:GLY:C	3:A:504:HOH:O	2.49	0.51
1:A:211:PHE:CD2	1:A:216:LYS:HE2	2.46	0.51
1:A:48:ARG:HH21	1:A:66:ASN:ND2	2.09	0.49
1:A:55:ARG:NH1	1:A:62:VAL:O	2.42	0.48
1:A:209:LEU:C	1:A:209:LEU:HD23	2.35	0.47
1:A:112:SER:OG	1:A:113:LYS:HE3	2.16	0.45
1:A:126:VAL:O	1:A:202:PHE:HA	2.17	0.45
1:A:93:PRO:HA	1:A:250:TRP:O	2.17	0.44
1:A:268:GLN:NE2	3:A:511:HOH:O	2.52	0.43
1:A:36:PRO:HA	1:A:67:ARG:O	2.19	0.42
1:A:152:VAL:HB	3:A:605:HOH:O	2.21	0.41
1:A:125:ASN:HA	1:A:204:ASN:HA	2.03	0.41
1:A:282:GLN:HB2	3:A:682:HOH:O	2.21	0.41

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	297/308 (96%)	292 (98%)	5 (2%)	0	<b>100</b> <b>100</b>

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	234/243 (96%)	231 (99%)	3 (1%)	69 35

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

Mol	Chain	Res	Type
1	A	268	GLN
1	A	278	PRO
1	A	298	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	26	GLN
1	A	57	GLN
1	A	66	ASN
1	A	87	ASN
1	A	125	ASN
1	A	212	HIS
1	A	268	GLN
1	A	281	GLN
1	A	287	GLN
1	A	301	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](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	UB7	A	401	-	12,12,12	0.94	0	16,16,16	0.73	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	UB7	A	401	-	-	0/8/8/8	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	299/308 (97%)	1.96	115 (38%) <b>0</b> <b>2</b>	5, 10, 22, 33	0

All (115) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	30	ALA	7.4
1	A	56	ALA	6.5
1	A	318	ASN	6.4
1	A	55	ARG	6.1
1	A	59	PRO	6.1
1	A	25	MET	6.1
1	A	320	ARG	5.8
1	A	321	LEU	5.7
1	A	63	LEU	5.5
1	A	317	ALA	4.9
1	A	28	PRO	4.9
1	A	282	GLN	4.5
1	A	315	ARG	4.4
1	A	27	THR	4.3
1	A	264	VAL	3.6
1	A	151	LYS	3.3
1	A	177	GLU	3.3
1	A	170	PHE	3.3
1	A	150	GLY	3.3
1	A	31	LEU	3.3
1	A	103	LEU	3.2
1	A	90	LEU	3.2
1	A	105	ARG	3.2
1	A	58	PHE	3.1
1	A	149	PRO	3.1
1	A	238	ALA	3.1
1	A	23	SER	3.1

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	226	GLN	3.0
1	A	276	THR	3.0
1	A	106	LYS	2.9
1	A	29	GLN	2.9
1	A	61	GLY	2.8
1	A	198	ILE	2.8
1	A	89	LEU	2.8
1	A	201	PHE	2.8
1	A	237	PHE	2.8
1	A	316	SER	2.8
1	A	319	VAL	2.7
1	A	231	LEU	2.7
1	A	57	GLN	2.7
1	A	202	PHE	2.7
1	A	127	LEU	2.7
1	A	91	ALA	2.7
1	A	200	VAL	2.6
1	A	253	VAL	2.6
1	A	294	TRP	2.6
1	A	310	TRP	2.6
1	A	250	TRP	2.6
1	A	26	GLN	2.6
1	A	33	ILE	2.5
1	A	32	LYS	2.5
1	A	60	ALA	2.5
1	A	251	PHE	2.5
1	A	114	TRP	2.5
1	A	225	GLU	2.5
1	A	79	VAL	2.5
1	A	285	ALA	2.5
1	A	129	VAL	2.5
1	A	120	LEU	2.4
1	A	104	TYR	2.4
1	A	172	GLN	2.4
1	A	65	ASP	2.4
1	A	179	VAL	2.4
1	A	191	VAL	2.4
1	A	116	PRO	2.4
1	A	161	THR	2.3
1	A	62	VAL	2.3
1	A	241	GLY	2.3
1	A	97	ILE	2.3

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	236	THR	2.3
1	A	283	LYS	2.3
1	A	169	LEU	2.3
1	A	128	VAL	2.3
1	A	234	VAL	2.3
1	A	183	TYR	2.3
1	A	214	ALA	2.2
1	A	246	ASN	2.2
1	A	173	LEU	2.2
1	A	193	LEU	2.2
1	A	217	LEU	2.2
1	A	74	ILE	2.2
1	A	162	THR	2.2
1	A	281	GLN	2.2
1	A	123	VAL	2.2
1	A	43	ALA	2.2
1	A	255	ALA	2.2
1	A	80	PHE	2.2
1	A	135	VAL	2.2
1	A	252	ALA	2.1
1	A	186	THR	2.1
1	A	54	LEU	2.1
1	A	211	PHE	2.1
1	A	271	PHE	2.1
1	A	99	ILE	2.1
1	A	145	ALA	2.1
1	A	298	ARG	2.1
1	A	41	GLY	2.1
1	A	136	LYS	2.1
1	A	235	PRO	2.1
1	A	256	PRO	2.1
1	A	34	ILE	2.1
1	A	205	ILE	2.1
1	A	35	VAL	2.1
1	A	152	VAL	2.1
1	A	37	TYR	2.1
1	A	165	LEU	2.1
1	A	219	ILE	2.0
1	A	40	GLY	2.0
1	A	118	THR	2.0
1	A	181	VAL	2.0
1	A	242	LEU	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	301	GLN	2.0
1	A	174	THR	2.0
1	A	36	PRO	2.0
1	A	117	VAL	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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	UB7	A	401	12/12	0.97	0.16	4,5,5,5	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.