Full wwPDB X-ray Structure Validation Report (i)

Jun 25, 2024 – 11:15 PM EDT

PDB ID	:	7B08
Title	:	TgoT apo
Authors	:	Samson, C.; Legrand, P.; Tekpinar, M.; Rozenski, J.; Abramov, M.; Holliger, P.; Pinheiro, V.; Herdewijn, P.; Delarue, M.
Deposited on	:	2020-11-18
Resolution	:	2.39 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 (i)) were used in the production of this report:

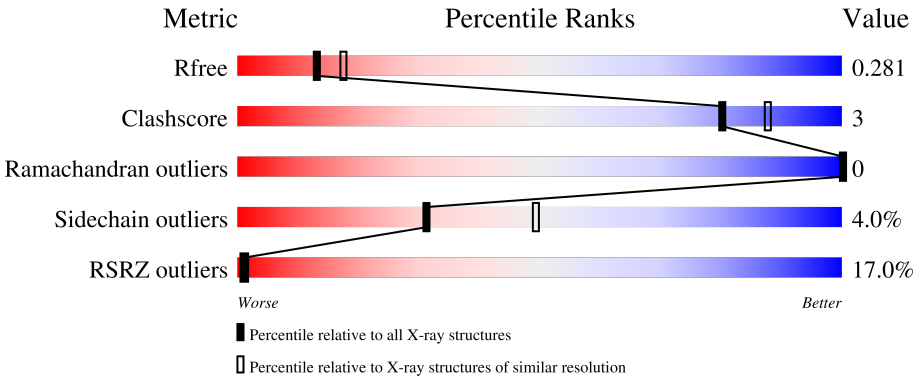
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.39 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	773	16% 87% 9%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	PGE	A	802	-	-	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6174 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.

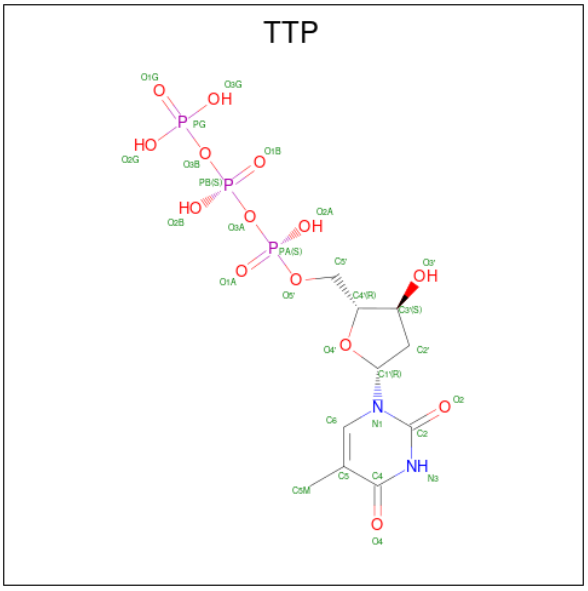
- Molecule 1 is a protein called DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	742	Total	C	N	O	S	0	0	0
			6094	3935	1024	1122	13			

There are 4 discrepancies between the modelled and reference sequences:

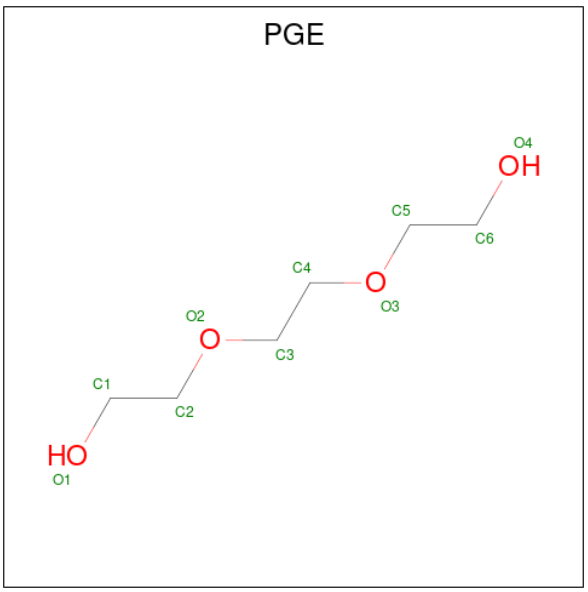
Chain	Residue	Modelled	Actual	Comment	Reference
A	93	GLN	VAL	conflict	UNP P56689
A	141	ALA	ASP	conflict	UNP P56689
A	143	ALA	GLU	conflict	UNP P56689
A	485	LEU	ALA	conflict	UNP P56689

- Molecule 2 is THYMIDINE-5'-TRIPHOSPHATE (three-letter code: TTP) (formula: $C_{10}H_{17}N_2O_{14}P_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	P	0	0
			13	10	3		

- Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 10 6 4	0	0

- Molecule 4 is water.

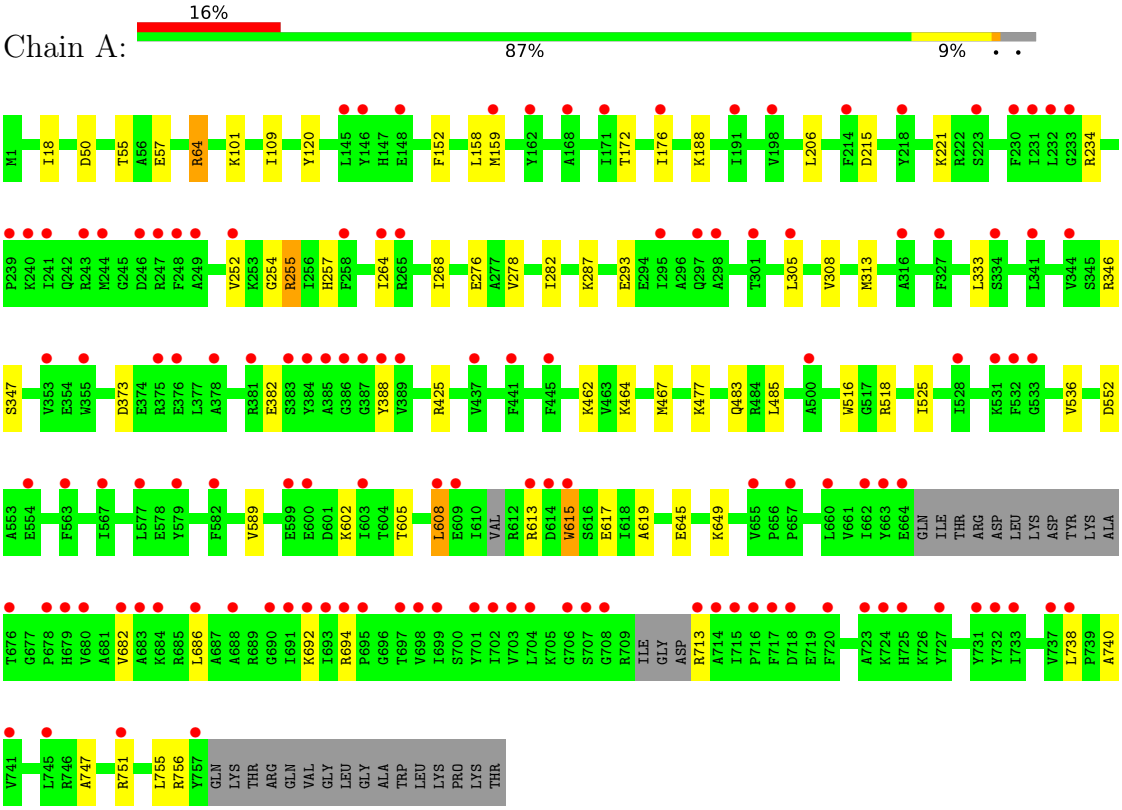
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	57	Total O 57 57	0	0

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: DNA polymerase

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	65.06Å 105.22Å 152.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.86 – 2.39 45.86 – 2.39	Depositor EDS
% Data completeness (in resolution range)	99.5 (45.86-2.39) 99.6 (45.86-2.39)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.25 (at 2.39Å)	Xtriage
Refinement program	BUSTER 2.10.3 (20-MAY-2020)	Depositor
R, R_{free}	0.241 , 0.261 0.262 , 0.281	Depositor DCC
R_{free} test set	2098 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	71.4	Xtriage
Anisotropy	0.384	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 50.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6174	wwPDB-VP
Average B, all atoms (Å^2)	100.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.71% 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

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

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	0/6229	0.47	0/8395

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	6094	0	6127	33	0
2	A	13	0	0	0	0
3	A	10	0	14	2	0
4	A	57	0	0	0	0
All	All	6174	0	6141	33	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 (33) 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:605:THR:HB	1:A:608:LEU:HD21	1.63	0.78
1:A:188:LYS:NZ	3:A:802:PGE:H32	2.01	0.75
1:A:608:LEU:H	1:A:608:LEU:HD22	1.51	0.75
1:A:589:VAL:HG13	1:A:747:ALA:HB2	1.72	0.71
1:A:615:TRP:HD1	1:A:619:ALA:HB3	1.62	0.65
1:A:333:LEU:HD13	1:A:485:LEU:HD22	1.81	0.63
1:A:608:LEU:HD22	1:A:608:LEU:N	2.18	0.58
1:A:172:THR:HG21	1:A:176:ILE:HD12	1.87	0.57
1:A:738:LEU:HD11	1:A:756:ARG:HB3	1.87	0.55
1:A:305:LEU:HA	1:A:308:VAL:HG22	1.88	0.55
1:A:264:ILE:HD13	1:A:278:VAL:HG11	1.88	0.55
1:A:18:ILE:HD11	1:A:120:TYR:CE2	2.41	0.55
1:A:252:VAL:HG12	1:A:257:HIS:HB2	1.92	0.52
1:A:615:TRP:CD1	1:A:619:ALA:HB3	2.43	0.51
1:A:257:HIS:O	1:A:346:ARG:NH2	2.44	0.50
1:A:293:GLU:HG2	1:A:692:LYS:HE3	1.93	0.50
1:A:268:ILE:HD13	1:A:282:ILE:HD11	1.95	0.48
1:A:276:GLU:HG3	1:A:287:LYS:HG2	1.94	0.48
1:A:525:ILE:HG23	1:A:536:VAL:HG21	1.95	0.48
1:A:252:VAL:HG22	1:A:255:ARG:HG3	1.96	0.47
1:A:188:LYS:HZ2	3:A:802:PGE:H32	1.77	0.46
1:A:615:TRP:CD1	1:A:619:ALA:CB	2.99	0.45
1:A:55:THR:HG22	1:A:64:ARG:HB3	1.99	0.44
1:A:234:ARG:HB2	1:A:254:GLY:HA3	2.00	0.44
1:A:152:PHE:HD2	1:A:221:LYS:HD2	1.82	0.44
1:A:101:LYS:HE3	1:A:109:ILE:HG12	1.99	0.44
1:A:615:TRP:HE1	1:A:740:ALA:HB2	1.83	0.43
1:A:645:GLU:O	1:A:649:LYS:HG2	2.19	0.42
1:A:206:LEU:HD12	1:A:252:VAL:HG11	2.03	0.41
1:A:462:LYS:HE2	1:A:462:LYS:HB3	1.86	0.41
1:A:464:LYS:HA	1:A:467:MET:HE3	2.03	0.41
1:A:388:TYR:HB2	1:A:518:ARG:HG2	2.03	0.41
1:A:159:MET:HG2	1:A:172:THR:HB	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	A	733/773 (95%)	715 (98%)	18 (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	642/669 (96%)	616 (96%)	26 (4%)	31 49

All (26) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	50	ASP
1	A	57	GLU
1	A	64	ARG
1	A	158	LEU
1	A	215	ASP
1	A	255	ARG
1	A	313	MET
1	A	347	SER
1	A	373	ASP
1	A	382	GLU
1	A	425	ARG
1	A	477	LYS
1	A	483	GLN
1	A	516	TRP
1	A	552	ASP
1	A	602	LYS
1	A	608	LEU
1	A	613	ARG
1	A	615	TRP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	617	GLU
1	A	682	VAL
1	A	686	LEU
1	A	694	ARG
1	A	713	ARG
1	A	751	ARG
1	A	755	LEU

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)

2 ligands are 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	$\#\|Z\| > 2$
3	PGE	A	802	-	9,9,9	0.10	0	8,8,8	0.05	0
2	TTP	A	801	-	8,12,30	0.73	0	15,20,47	0.88	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
3	PGE	A	802	-	-	1/7/7/7	-
2	TTP	A	801	-	-	0/12/12/34	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	802	PGE	C1-C2-O2-C3

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	802	PGE	2	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	756:ARG	C	757:TYR	N	3.82

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	742/773 (95%)	1.12	126 (16%) 1 1	63, 96, 144, 188	0

All (126) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	697	THR	12.3
1	A	686	LEU	12.2
1	A	615	TRP	9.2
1	A	693	ILE	8.8
1	A	295	ILE	7.9
1	A	692	LYS	7.2
1	A	714	ALA	7.2
1	A	694	ARG	7.0
1	A	699	ILE	6.7
1	A	717	PHE	6.4
1	A	725	HIS	6.0
1	A	613	ARG	5.8
1	A	609	GLU	5.7
1	A	723	ALA	5.7
1	A	176	ILE	5.6
1	A	384	TYR	5.6
1	A	706	GLY	5.6
1	A	243	ARG	5.5
1	A	246	ASP	5.4
1	A	679	HIS	5.4
1	A	381	ARG	5.3
1	A	691	ILE	5.2
1	A	614	ASP	4.9
1	A	698	VAL	4.6
1	A	704	LEU	4.6
1	A	703	VAL	4.5
1	A	688	ALA	4.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	244	MET	4.4
1	A	386	GLY	4.3
1	A	690	GLY	4.3
1	A	664	GLU	4.2
1	A	732	TYR	4.1
1	A	720	PHE	4.0
1	A	684	LYS	4.0
1	A	724	LYS	3.9
1	A	383	SER	3.9
1	A	701	TYR	3.8
1	A	500	ALA	3.8
1	A	662	ILE	3.7
1	A	738	LEU	3.7
1	A	707	SER	3.7
1	A	168	ALA	3.7
1	A	682	VAL	3.7
1	A	248	PHE	3.6
1	A	713	ARG	3.6
1	A	385	ALA	3.5
1	A	305	LEU	3.5
1	A	145	LEU	3.5
1	A	533	GLY	3.4
1	A	146	TYR	3.4
1	A	678	PRO	3.4
1	A	702	ILE	3.4
1	A	718	ASP	3.4
1	A	600	GLU	3.4
1	A	676	THR	3.3
1	A	297	GLN	3.3
1	A	241	ILE	3.3
1	A	230	PHE	3.3
1	A	301	THR	3.3
1	A	695	PRO	3.2
1	A	258	PHE	3.2
1	A	252	VAL	3.2
1	A	387	GLY	3.2
1	A	663	TYR	3.1
1	A	657	PRO	3.1
1	A	214	PHE	3.1
1	A	660	LEU	3.1
1	A	247	ARG	3.1
1	A	375	ARG	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	445	PHE	3.0
1	A	751	ARG	3.0
1	A	715	ILE	3.0
1	A	757	TYR	2.9
1	A	655	VAL	2.9
1	A	603	ILE	2.9
1	A	741	VAL	2.9
1	A	727	TYR	2.9
1	A	265	ARG	2.8
1	A	162	TYR	2.8
1	A	231	ILE	2.7
1	A	239	PRO	2.7
1	A	582	PHE	2.7
1	A	599	GLU	2.7
1	A	731	TYR	2.7
1	A	577	LEU	2.7
1	A	608	LEU	2.7
1	A	316	ALA	2.6
1	A	683	ALA	2.6
1	A	249	ALA	2.6
1	A	716	PRO	2.5
1	A	680	VAL	2.5
1	A	378	ALA	2.4
1	A	233	GLY	2.4
1	A	264	ILE	2.4
1	A	733	ILE	2.4
1	A	159	MET	2.4
1	A	737	VAL	2.4
1	A	198	VAL	2.3
1	A	344	VAL	2.3
1	A	353	VAL	2.3
1	A	388	TYR	2.3
1	A	223	SER	2.3
1	A	563	PHE	2.3
1	A	240	LYS	2.3
1	A	567	ILE	2.3
1	A	579	TYR	2.2
1	A	327	PHE	2.2
1	A	355	TRP	2.2
1	A	148	GLU	2.2
1	A	298	ALA	2.2
1	A	218	TYR	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	554	GLU	2.2
1	A	171	ILE	2.2
1	A	232	LEU	2.2
1	A	376	GLU	2.2
1	A	745	LEU	2.2
1	A	708	GLY	2.2
1	A	532	PHE	2.1
1	A	191	ILE	2.1
1	A	341	LEU	2.1
1	A	437	VAL	2.0
1	A	441	PHE	2.0
1	A	528	ILE	2.0
1	A	389	VAL	2.0
1	A	531	LYS	2.0
1	A	334	SER	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^{th} 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(Å^2)	Q<0.9
3	PGE	A	802	10/10	0.74	0.41	134,134,135,135	0
2	TTP	A	801	13/29	0.78	0.24	172,172,173,173	0

6.5 Other polymers (i)

There are no such residues in this entry.