Full wwPDB X-ray Structure Validation Report (i)

Jul 25, 2022 – 01:07 pm BST

PDB ID	:	7QVH
Title	:	The crystal structure of HotPETase, an evolved thermostable variant of IsPETase
Authors	:	Hardy, F.J.; Levy, C.; Green, A.P.
Deposited on	:	2022-01-21
Resolution	:	2.24 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:

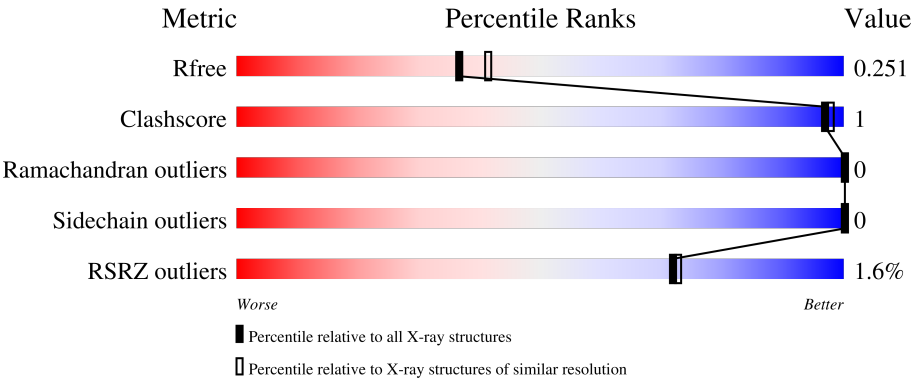
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.29
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 2.24 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	2391 (2.26-2.22)
Clashscore	141614	2539 (2.26-2.22)
Ramachandran outliers	138981	2489 (2.26-2.22)
Sidechain outliers	138945	2490 (2.26-2.22)
RSRZ outliers	127900	2353 (2.26-2.22)

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	272	96%
1	B	272	97%
1	C	272	96%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11922 atoms, of which 5747 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 Poly(ethylene terephthalate) hydrolase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	265	Total	C	H	N	O	S	0	2	0
			3841	1216	1888	341	380	16			
1	B	272	Total	C	H	N	O	S	0	1	0
			3958	1257	1935	360	390	16			
1	C	267	Total	C	H	N	O	S	0	0	0
			3863	1224	1896	345	382	16			

There are 90 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	27	MET	-	initiating methionine	UNP A0A0K8P6T7
A	58	ALA	SER	engineered mutation	UNP A0A0K8P6T7
A	61	VAL	SER	engineered mutation	UNP A0A0K8P6T7
A	90	THR	ARG	engineered mutation	UNP A0A0K8P6T7
A	95	ASN	LYS	engineered mutation	UNP A0A0K8P6T7
A	119	LYS	GLN	engineered mutation	UNP A0A0K8P6T7
A	121	GLU	SER	engineered mutation	UNP A0A0K8P6T7
A	154	GLY	MET	engineered mutation	UNP A0A0K8P6T7
A	181	VAL	PRO	engineered mutation	UNP A0A0K8P6T7
A	182	MET	GLN	engineered mutation	UNP A0A0K8P6T7
A	186	HIS	ASP	engineered mutation	UNP A0A0K8P6T7
A	207	ARG	SER	engineered mutation	UNP A0A0K8P6T7
A	212	LYS	ASN	engineered mutation	UNP A0A0K8P6T7
A	213	GLU	SER	engineered mutation	UNP A0A0K8P6T7
A	214	TYR	SER	engineered mutation	UNP A0A0K8P6T7
A	224	LEU	ARG	engineered mutation	UNP A0A0K8P6T7
A	233	CYS	ASN	engineered mutation	UNP A0A0K8P6T7
A	241	CYS	ASN	engineered mutation	UNP A0A0K8P6T7
A	252	MET	LYS	engineered mutation	UNP A0A0K8P6T7
A	270	GLN	THR	engineered mutation	UNP A0A0K8P6T7
A	280	ALA	ARG	engineered mutation	UNP A0A0K8P6T7
A	282	CYS	SER	engineered mutation	UNP A0A0K8P6T7
A	291	LEU	-	expression tag	UNP A0A0K8P6T7

Continued on next page...

Continued from previous page...

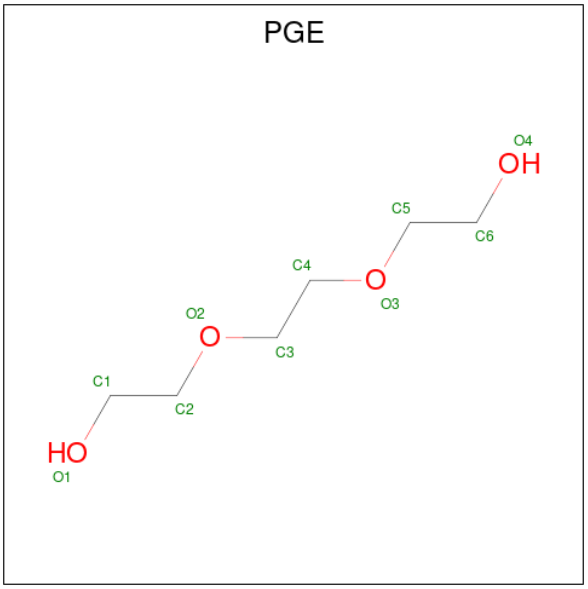
Chain	Residue	Modelled	Actual	Comment	Reference
A	292	GLU	-	expression tag	UNP A0A0K8P6T7
A	293	HIS	-	expression tag	UNP A0A0K8P6T7
A	294	HIS	-	expression tag	UNP A0A0K8P6T7
A	295	HIS	-	expression tag	UNP A0A0K8P6T7
A	296	HIS	-	expression tag	UNP A0A0K8P6T7
A	297	HIS	-	expression tag	UNP A0A0K8P6T7
A	298	HIS	-	expression tag	UNP A0A0K8P6T7
B	27	MET	-	initiating methionine	UNP A0A0K8P6T7
B	58	ALA	SER	engineered mutation	UNP A0A0K8P6T7
B	61	VAL	SER	engineered mutation	UNP A0A0K8P6T7
B	90	THR	ARG	engineered mutation	UNP A0A0K8P6T7
B	95	ASN	LYS	engineered mutation	UNP A0A0K8P6T7
B	119	LYS	GLN	engineered mutation	UNP A0A0K8P6T7
B	121	GLU	SER	engineered mutation	UNP A0A0K8P6T7
B	154	GLY	MET	engineered mutation	UNP A0A0K8P6T7
B	181	VAL	PRO	engineered mutation	UNP A0A0K8P6T7
B	182	MET	GLN	engineered mutation	UNP A0A0K8P6T7
B	186	HIS	ASP	engineered mutation	UNP A0A0K8P6T7
B	207	ARG	SER	engineered mutation	UNP A0A0K8P6T7
B	212	LYS	ASN	engineered mutation	UNP A0A0K8P6T7
B	213	GLU	SER	engineered mutation	UNP A0A0K8P6T7
B	214	TYR	SER	engineered mutation	UNP A0A0K8P6T7
B	224	LEU	ARG	engineered mutation	UNP A0A0K8P6T7
B	233	CYS	ASN	engineered mutation	UNP A0A0K8P6T7
B	241	CYS	ASN	engineered mutation	UNP A0A0K8P6T7
B	252	MET	LYS	engineered mutation	UNP A0A0K8P6T7
B	270	GLN	THR	engineered mutation	UNP A0A0K8P6T7
B	280	ALA	ARG	engineered mutation	UNP A0A0K8P6T7
B	282	CYS	SER	engineered mutation	UNP A0A0K8P6T7
B	291	LEU	-	expression tag	UNP A0A0K8P6T7
B	292	GLU	-	expression tag	UNP A0A0K8P6T7
B	293	HIS	-	expression tag	UNP A0A0K8P6T7
B	294	HIS	-	expression tag	UNP A0A0K8P6T7
B	295	HIS	-	expression tag	UNP A0A0K8P6T7
B	296	HIS	-	expression tag	UNP A0A0K8P6T7
B	297	HIS	-	expression tag	UNP A0A0K8P6T7
B	298	HIS	-	expression tag	UNP A0A0K8P6T7
C	27	MET	-	initiating methionine	UNP A0A0K8P6T7
C	58	ALA	SER	engineered mutation	UNP A0A0K8P6T7
C	61	VAL	SER	engineered mutation	UNP A0A0K8P6T7
C	90	THR	ARG	engineered mutation	UNP A0A0K8P6T7
C	95	ASN	LYS	engineered mutation	UNP A0A0K8P6T7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	119	LYS	GLN	engineered mutation	UNP A0A0K8P6T7
C	121	GLU	SER	engineered mutation	UNP A0A0K8P6T7
C	154	GLY	MET	engineered mutation	UNP A0A0K8P6T7
C	181	VAL	PRO	engineered mutation	UNP A0A0K8P6T7
C	182	MET	GLN	engineered mutation	UNP A0A0K8P6T7
C	186	HIS	ASP	engineered mutation	UNP A0A0K8P6T7
C	207	ARG	SER	engineered mutation	UNP A0A0K8P6T7
C	212	LYS	ASN	engineered mutation	UNP A0A0K8P6T7
C	213	GLU	SER	engineered mutation	UNP A0A0K8P6T7
C	214	TYR	SER	engineered mutation	UNP A0A0K8P6T7
C	224	LEU	ARG	engineered mutation	UNP A0A0K8P6T7
C	233	CYS	ASN	engineered mutation	UNP A0A0K8P6T7
C	241	CYS	ASN	engineered mutation	UNP A0A0K8P6T7
C	252	MET	LYS	engineered mutation	UNP A0A0K8P6T7
C	270	GLN	THR	engineered mutation	UNP A0A0K8P6T7
C	280	ALA	ARG	engineered mutation	UNP A0A0K8P6T7
C	282	CYS	SER	engineered mutation	UNP A0A0K8P6T7
C	291	LEU	-	expression tag	UNP A0A0K8P6T7
C	292	GLU	-	expression tag	UNP A0A0K8P6T7
C	293	HIS	-	expression tag	UNP A0A0K8P6T7
C	294	HIS	-	expression tag	UNP A0A0K8P6T7
C	295	HIS	-	expression tag	UNP A0A0K8P6T7
C	296	HIS	-	expression tag	UNP A0A0K8P6T7
C	297	HIS	-	expression tag	UNP A0A0K8P6T7
C	298	HIS	-	expression tag	UNP A0A0K8P6T7

- Molecule 2 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			24	6	14	4		
2	C	1	Total	C	H	O	0	0
			24	6	14	4		

- Molecule 3 is water.

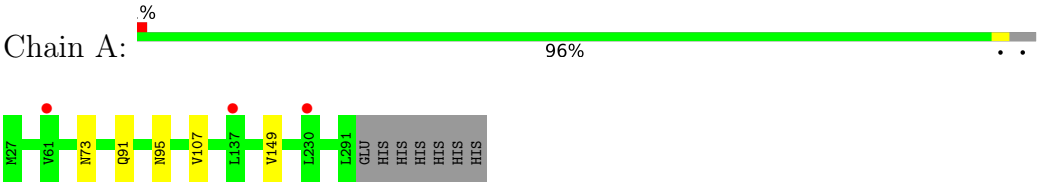
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	79	Total	O	0	0
			79	79		
3	B	71	Total	O	0	0
			71	71		
3	C	62	Total	O	0	0
			62	62		

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.

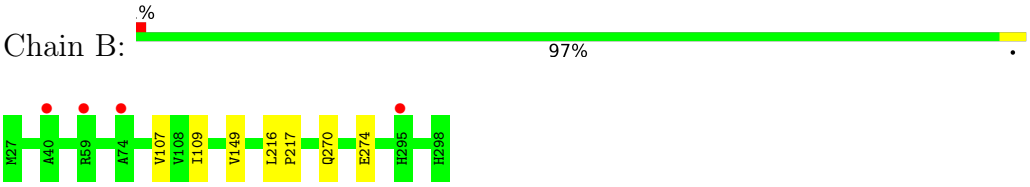
• Molecule 1: Poly(ethylene terephthalate) hydrolase

Chain A:



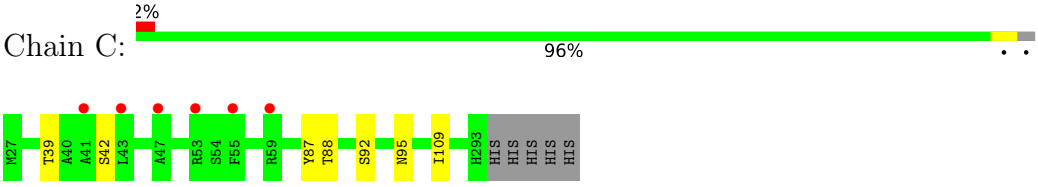
• Molecule 1: Poly(ethylene terephthalate) hydrolase

Chain B:



• Molecule 1: Poly(ethylene terephthalate) hydrolase

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	81.70Å 74.86Å 82.70Å 90.00° 118.57° 90.00°	Depositor
Resolution (Å)	72.63 – 2.24 72.63 – 2.24	Depositor EDS
% Data completeness (in resolution range)	99.8 (72.63-2.24) 91.8 (72.63-2.24)	Depositor EDS
R_{merge}	0.22	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.82 (at 2.25Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.198 , 0.251 0.198 , 0.251	Depositor DCC
R_{free} test set	1926 reflections (4.55%)	wwPDB-VP
Wilson B-factor (Å^2)	30.3	Xtriage
Anisotropy	0.572	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.097 for -h-l,k,h 0.097 for l,k,-h-l 0.038 for h,-k,-h-l 0.034 for -h-l,-k,l 0.035 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11922	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.12% 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: 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	0.26	0/2011	0.45	0/2740
1	B	0.26	0/2078	0.47	0/2830
1	C	0.26	0/2014	0.46	0/2744
All	All	0.26	0/6103	0.46	0/8314

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	1953	1888	1878	3	0
1	B	2023	1935	1932	4	0
1	C	1967	1896	1893	4	0
2	A	10	14	14	0	0
2	C	10	14	14	0	0
3	A	79	0	0	0	0
3	B	71	0	0	0	0
3	C	62	0	0	0	0
All	All	6175	5747	5731	11	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 1.

All (11) 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:C:39:THR:HG23	1:C:42:SER:H	1.57	0.69
1:A:91:GLN:HB3	1:A:95:ASN:HD21	1.71	0.55
1:A:107:VAL:HG21	1:A:149:VAL:CG2	2.41	0.51
1:C:87:TYR:CD2	1:C:88:THR:HG23	2.46	0.50
1:B:216:LEU:HB3	1:B:217:PRO:HD3	1.98	0.46
1:C:92:SER:HA	1:C:95:ASN:ND2	2.30	0.46
1:A:73:ASN:OD1	1:A:73:ASN:N	2.52	0.43
1:C:109:ILE:O	1:C:109:ILE:HG23	2.20	0.42
1:B:109:ILE:O	1:B:109:ILE:HG23	2.20	0.41
1:B:270:GLN:HG3	1:B:274:GLU:OE2	2.21	0.41
1:B:107:VAL:HG21	1:B:149:VAL:HG22	2.03	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	265/272 (97%)	251 (95%)	14 (5%)	0	100	100
1	B	271/272 (100%)	263 (97%)	8 (3%)	0	100	100
1	C	265/272 (97%)	258 (97%)	7 (3%)	0	100	100
All	All	801/816 (98%)	772 (96%)	29 (4%)	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	210/215 (98%)	210 (100%)	0	100	100
1	B	216/215 (100%)	216 (100%)	0	100	100
1	C	210/215 (98%)	210 (100%)	0	100	100
All	All	636/645 (99%)	636 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	95	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry

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	$\#\|Z\| > 2$
2	PGE	C	401	-	9,9,9	0.31	0	8,8,8	0.29	0
2	PGE	A	401	-	9,9,9	0.31	0	8,8,8	0.31	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	PGE	C	401	-	-	3/7/7/7	-
2	PGE	A	401	-	-	3/7/7/7	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	401	PGE	O2-C3-C4-O3
2	A	401	PGE	O2-C3-C4-O3
2	A	401	PGE	O1-C1-C2-O2
2	C	401	PGE	O1-C1-C2-O2
2	A	401	PGE	O3-C5-C6-O4
2	C	401	PGE	C4-C3-O2-C2

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^{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	265/272 (97%)	0.32	3 (1%)	80	81	32, 39, 50, 61	0
1	B	272/272 (100%)	0.29	4 (1%)	73	74	29, 38, 51, 61	0
1	C	267/272 (98%)	0.30	6 (2%)	62	63	31, 39, 49, 64	0
All	All	804/816 (98%)	0.30	13 (1%)	72	73	29, 39, 50, 64	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	59	ARG	3.8
1	A	61	VAL	3.1
1	B	40	ALA	3.0
1	B	295	HIS	2.7
1	B	59	ARG	2.6
1	B	74	ALA	2.5
1	A	230	LEU	2.5
1	A	137	LEU	2.5
1	C	47	ALA	2.3
1	C	55	PHE	2.2
1	C	41	ALA	2.1
1	C	43	LEU	2.1
1	C	53	ARG	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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
2	PGE	C	401	10/10	0.77	0.30	48,58,65,70	0
2	PGE	A	401	10/10	0.85	0.24	43,52,58,59	0

6.5 Other polymers

There are no such residues in this entry.