Full wwPDB X-ray Structure Validation Report (i)

Sep 4, 2022 – 10:21 pm BST

PDB ID	:	5LK4
Title	:	Structure of the Red Fluorescent Protein mScarlet at pH 7.8
Authors	:	Aumonier, S.; Gotthard, G.; Royant, A.
Deposited on	:	2016-07-20
Resolution	:	1.47 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:

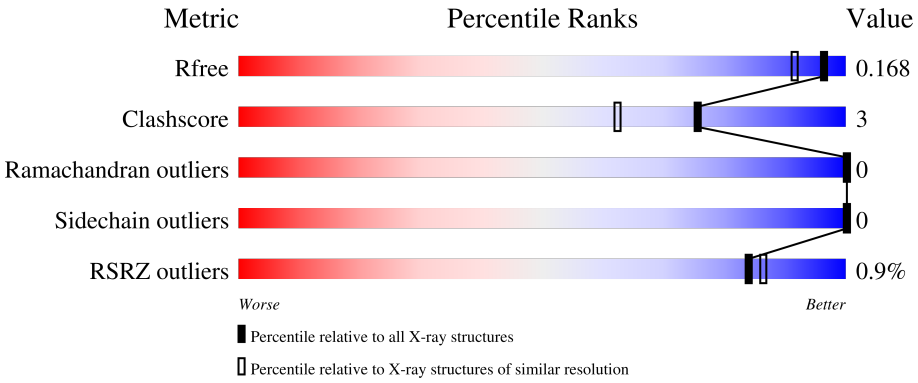
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.30
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.30

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 1.47 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	4690 (1.50-1.46)
Clashscore	141614	4955 (1.50-1.46)
Ramachandran outliers	138981	4846 (1.50-1.46)
Sidechain outliers	138945	4844 (1.50-1.46)
RSRZ outliers	127900	4614 (1.50-1.46)

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	217	% 93% 6% ·

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2128 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 mScarlet.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	217	1865	1204	308	342	11	0	24	0

There are 46 discrepancies between the modelled and reference sequences:

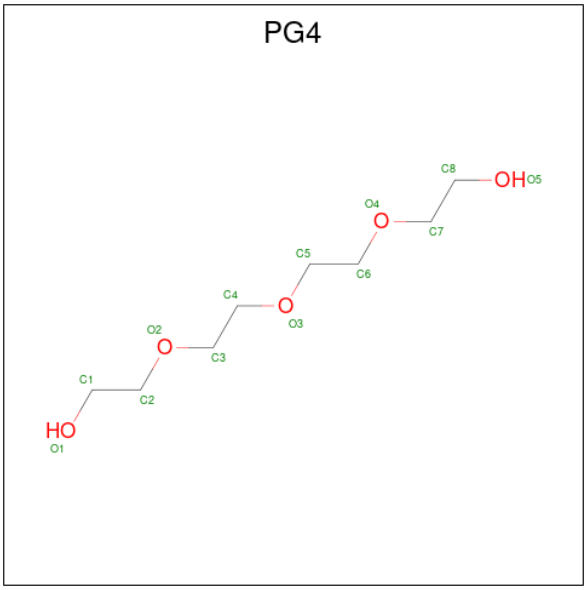
Chain	Residue	Modelled	Actual	Comment	Reference
A	7	ALA	-	expression tag	UNP Q9U6Y8
A	18	HIS	ARG	engineered mutation	UNP Q9U6Y8
A	22	SER	THR	engineered mutation	UNP Q9U6Y8
A	23	MET	VAL	engineered mutation	UNP Q9U6Y8
A	42	THR	HIS	engineered mutation	UNP Q9U6Y8
A	43	GLN	ASN	engineered mutation	UNP Q9U6Y8
A	45	ALA	VAL	engineered mutation	UNP Q9U6Y8
A	58	SER	ALA	engineered mutation	UNP Q9U6Y8
A	67	NRQ	GLN	chromophore	UNP Q9U6Y8
A	?	-	TYR	chromophore	UNP Q9U6Y8
A	?	-	GLY	chromophore	UNP Q9U6Y8
A	71	ARG	LYS	engineered mutation	UNP Q9U6Y8
A	72	ALA	VAL	engineered mutation	UNP Q9U6Y8
A	73	PHE	TYR	engineered mutation	UNP Q9U6Y8
A	74	THR	VAL	engineered mutation	UNP Q9U6Y8
A	84	TYR	LYS	engineered mutation	UNP Q9U6Y8
A	86	GLN	LEU	engineered mutation	UNP Q9U6Y8
A	105	ALA	VAL	engineered mutation	UNP Q9U6Y8
A	112	THR	SER	engineered mutation	UNP Q9U6Y8
A	115	GLU	GLN	engineered mutation	UNP Q9U6Y8
A	118	THR	CYS	engineered mutation	UNP Q9U6Y8
A	119	LEU	PHE	engineered mutation	UNP Q9U6Y8
A	125	LEU	PHE	engineered mutation	UNP Q9U6Y8
A	126	ARG	ILE	engineered mutation	UNP Q9U6Y8
A	128	THR	VAL	engineered mutation	UNP Q9U6Y8
A	132	PRO	SER	engineered mutation	UNP Q9U6Y8
A	154	GLU	ARG	engineered mutation	UNP Q9U6Y8

Continued on next page...

Continued from previous page...

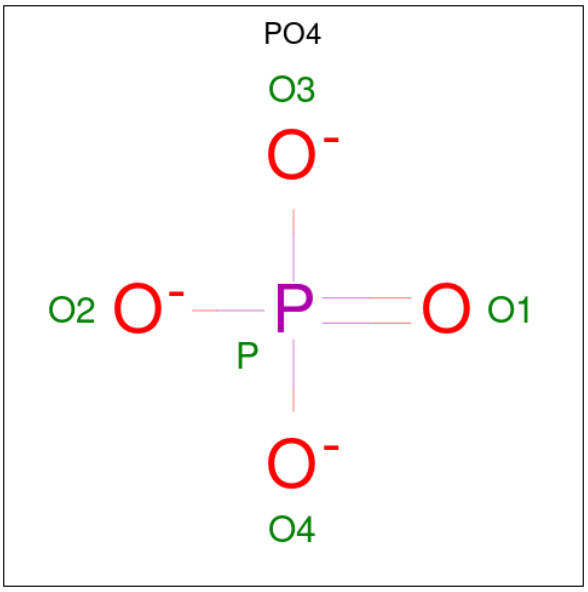
Chain	Residue	Modelled	Actual	Comment	Reference
A	161	ASP	GLU	engineered mutation	UNP Q9U6Y8
A	163	LYS	HIS	engineered mutation	UNP Q9U6Y8
A	164	MET	LYS	engineered mutation	UNP Q9U6Y8
A	167	ARG	LYS	engineered mutation	UNP Q9U6Y8
A	173	ARG	HIS	engineered mutation	UNP Q9U6Y8
A	176	ALA	VAL	engineered mutation	UNP Q9U6Y8
A	177	ASP	GLU	engineered mutation	UNP Q9U6Y8
A	180	THR	SER	engineered mutation	UNP Q9U6Y8
A	181	THR	ILE	engineered mutation	UNP Q9U6Y8
A	183	LYS	MET	engineered mutation	UNP Q9U6Y8
A	190	MET	LEU	engineered mutation	UNP Q9U6Y8
A	193	ALA	TYR	engineered mutation	UNP Q9U6Y8
A	195	ASN	TYR	engineered mutation	UNP Q9U6Y8
A	198	ARG	SER	engineered mutation	UNP Q9U6Y8
A	211	VAL	ILE	engineered mutation	UNP Q9U6Y8
A	218	SER	THR	engineered mutation	UNP Q9U6Y8
A	223	SER	-	expression tag	UNP Q9U6Y8
A	224	THR	-	expression tag	UNP Q9U6Y8
A	225	GLY	-	expression tag	UNP Q9U6Y8

- Molecule 2 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).



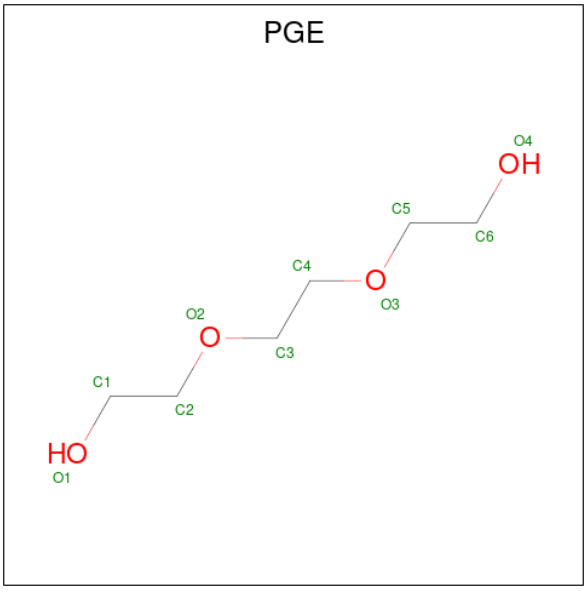
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 13 8 5	0	0

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O P 5 4 1	0	0

- Molecule 4 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



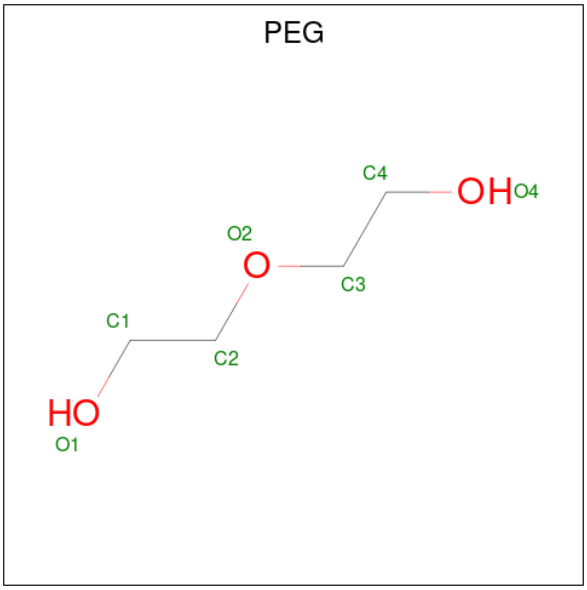
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 10 6 4	0	0
4	A	1	Total C O 10 6 4	0	0
4	A	1	Total C O 10 6 4	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 7 4 3	0	0

- Molecule 6 is water.

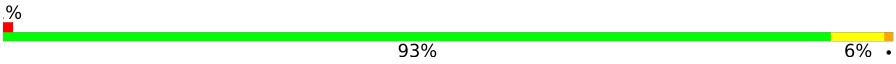
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	196	Total O 198 198	0	2

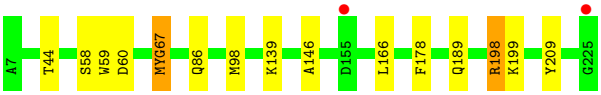
3 Residue-property plots (i)

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: mScarlet

Chain A:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	84.20Å 35.64Å 86.01Å 90.00° 109.69° 90.00°	Depositor
Resolution (Å)	80.98 – 1.47 39.64 – 1.47	Depositor EDS
% Data completeness (in resolution range)	98.7 (80.98-1.47) 98.7 (39.64-1.47)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.54 (at 1.47Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.137 , 0.167 0.137 , 0.168	Depositor DCC
R_{free} test set	2008 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å^2)	15.1	Xtriage
Anisotropy	0.021	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	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	0.97	EDS
Total number of atoms	2128	wwPDB-VP
Average B, all atoms (Å^2)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.33% 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: PEG, PG4, PO4, NRQ, 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	0.48	0/1954	0.76	2/2627 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	198	ARG	NE-CZ-NH2	-7.78	116.41	120.30
1	A	198	ARG	NE-CZ-NH1	7.60	124.10	120.30

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	1865	0	1876	12	0
2	A	13	0	18	0	0
3	A	5	0	0	0	0
4	A	40	0	56	6	0
5	A	7	0	10	0	0
6	A	198	0	0	1	0
All	All	2128	0	1960	12	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 (12) 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:146:ALA:HB1	4:A:306:PGE:H42	1.79	0.63
1:A:86:GLN:HE22	1:A:189:GLN:H	1.51	0.58
1:A:44[B]:THR:HG23	6:A:401:HOH:O	2.09	0.53
1:A:59:TRP:H	4:A:304:PGE:C5	2.24	0.50
1:A:59:TRP:H	4:A:304:PGE:H52	1.78	0.48
1:A:146:ALA:HB2	1:A:199[B]:LYS:HD2	1.95	0.48
1:A:139:LYS:HD3	4:A:304:PGE:H6	1.96	0.47
1:A:98[B]:MET:SD	1:A:178:PHE:CZ	3.09	0.46
1:A:209:TYR:CE2	4:A:304:PGE:H32	2.51	0.45
1:A:58:SER:HA	4:A:304:PGE:H52	1.99	0.44
1:A:67:NRQ:CD2	1:A:198:ARG:HD3	2.48	0.44
1:A:60:ASP:HB3	1:A:166:LEU:HD21	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	236/217 (109%)	235 (100%)	1 (0%)	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	200/184 (109%)	200 (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	86	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)

1 non-standard protein/DNA/RNA residue 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	$\#\|Z\| > 2$
1	NRQ	A	67	1	23,24,25	2.86	4 (17%)	23,32,34	4.18	7 (30%)

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
1	NRQ	A	67	1	-	1/9/31/32	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	67	NRQ	CB2-CA2	11.33	1.44	1.35
1	A	67	NRQ	CA2-C2	-5.39	1.43	1.48
1	A	67	NRQ	C2-N3	-3.72	1.31	1.39
1	A	67	NRQ	C1-N2	2.77	1.39	1.33

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	67	NRQ	CA2-C2-N3	14.56	110.26	103.37
1	A	67	NRQ	O2-C2-CA2	-11.85	124.31	130.96
1	A	67	NRQ	O3-C3-CA3	-3.91	114.59	126.39
1	A	67	NRQ	CA3-N3-C1	2.63	133.37	128.22
1	A	67	NRQ	CD1-CG2-CB2	-2.43	112.95	121.22
1	A	67	NRQ	CE2-CD2-CG2	-2.25	118.32	121.25
1	A	67	NRQ	CD1-CG2-CD2	2.20	120.90	117.64

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	67	NRQ	CB1-CG1-SD-CE

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	67	NRQ	1	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

7 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	PO4	A	302	-	4,4,4	0.75	0	6,6,6	0.59	0
4	PGE	A	305	-	9,9,9	0.50	0	8,8,8	0.30	0
4	PGE	A	304	-	9,9,9	0.33	0	8,8,8	0.69	0
2	PG4	A	301	-	12,12,12	0.48	0	11,11,11	0.30	0
4	PGE	A	303	-	9,9,9	0.50	0	8,8,8	0.25	0
4	PGE	A	306	-	9,9,9	0.55	0	8,8,8	0.45	0
5	PEG	A	307	-	6,6,6	0.32	0	5,5,5	0.59	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
4	PGE	A	305	-	-	2/7/7/7	-
4	PGE	A	304	-	-	5/7/7/7	-
2	PG4	A	301	-	-	1/10/10/10	-
4	PGE	A	303	-	-	1/7/7/7	-
4	PGE	A	306	-	-	4/7/7/7	-
5	PEG	A	307	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (15) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	304	PGE	O2-C3-C4-O3
4	A	306	PGE	O3-C5-C6-O4
4	A	304	PGE	C1-C2-O2-C3
4	A	306	PGE	C3-C4-O3-C5
4	A	306	PGE	C4-C3-O2-C2
4	A	306	PGE	O2-C3-C4-O3
4	A	304	PGE	C4-C3-O2-C2
4	A	304	PGE	O1-C1-C2-O2
5	A	307	PEG	O1-C1-C2-O2
5	A	307	PEG	O2-C3-C4-O4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	A	301	PG4	C6-C5-O3-C4
4	A	305	PGE	O1-C1-C2-O2
4	A	303	PGE	O1-C1-C2-O2
4	A	304	PGE	O3-C5-C6-O4
4	A	305	PGE	C4-C3-O2-C2

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	304	PGE	5	0
4	A	306	PGE	1	0

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	216/217 (99%)	0.09	2 (0%)	84	86	10, 14, 24, 59	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	155	ASP	3.4
1	A	225	GLY	2.3

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

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
1	NRQ	A	67	23/24	0.93	0.11	12,14,17,18	0

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
4	PGE	A	306	10/10	0.62	0.20	31,37,40,47	0
4	PGE	A	303	10/10	0.72	0.21	28,34,39,44	0
4	PGE	A	304	10/10	0.84	0.20	23,30,37,39	0
4	PGE	A	305	10/10	0.86	0.12	31,32,37,37	0
5	PEG	A	307	7/7	0.89	0.14	25,26,33,39	0
2	PG4	A	301	13/13	0.90	0.11	26,32,37,41	0
3	PO4	A	302	5/5	0.96	0.18	26,28,32,33	0

6.5 Other polymers (i)

There are no such residues in this entry.