wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 11, 2023 – 02:22 PM EDT

PDB ID	:	8EML
Title	:	Crystal Structure of Gsx2 Homeodomain in Complex with DNA
Authors	:	Webb, J.A.; Kovall, R.A.
Deposited on	:	2022-09-28
Resolution	:	2.21 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

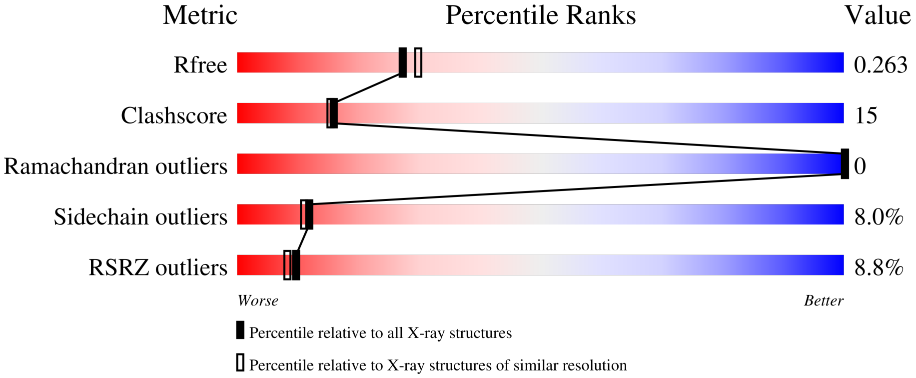
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance ⓘ

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

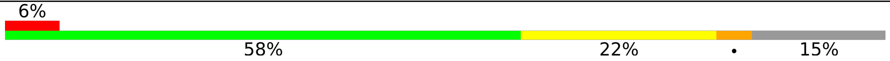
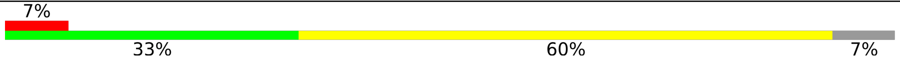
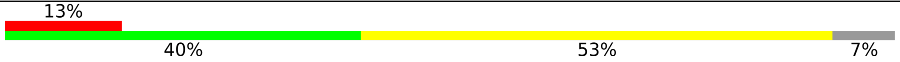
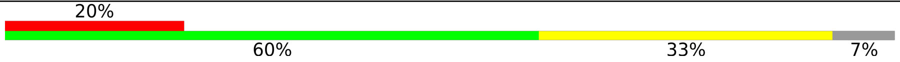
The reported resolution of this entry is 2.21 Å.

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.



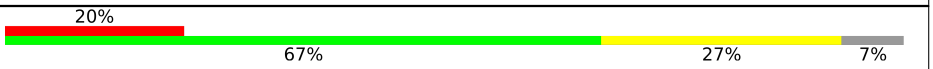
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5912 (2.24-2.20)
Clashscore	141614	6646 (2.24-2.20)
Ramachandran outliers	138981	6543 (2.24-2.20)
Sidechain outliers	138945	6544 (2.24-2.20)
RSRZ outliers	127900	5797 (2.24-2.20)

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	67	
1	B	67	
2	C	15	
2	E	15	
3	D	15	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	15	

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
4	PGE	A	301	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2123 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 GS homeobox 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	57	Total 473	C 300	N 87	O 84	S 2	0	0	0
1	A	58	Total 477	C 302	N 88	O 85	S 2	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	198	GLY	-	expression tag	UNP P31316
B	199	PRO	-	expression tag	UNP P31316
B	200	ALA	-	expression tag	UNP P31316
B	201	ALA	-	expression tag	UNP P31316
B	202	ALA	-	expression tag	UNP P31316
A	198	GLY	-	expression tag	UNP P31316
A	199	PRO	-	expression tag	UNP P31316
A	200	ALA	-	expression tag	UNP P31316
A	201	ALA	-	expression tag	UNP P31316
A	202	ALA	-	expression tag	UNP P31316

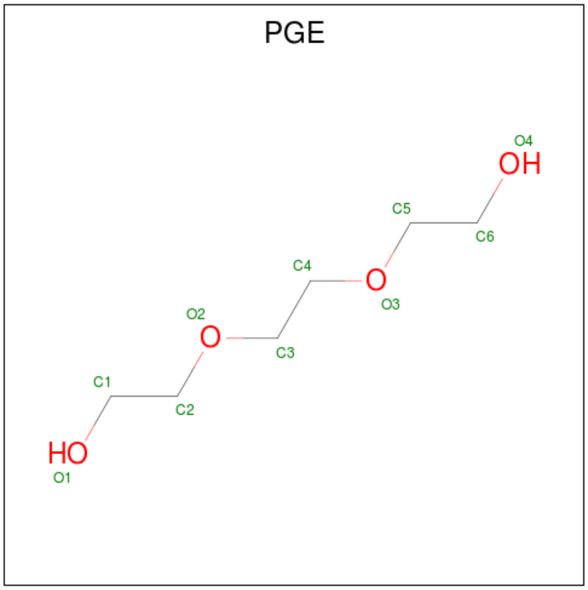
- Molecule 2 is a DNA chain called DNA (5'-D(P*GP*AP*GP*CP*TP*AP*AP*TP*TP*AP *AP*AP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	14	Total 290	C 138	N 57	O 81	P 14	0	0	0
2	C	14	Total 290	C 138	N 57	O 81	P 14	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*CP*TP*TP*TP*AP*AP*TP*TP*AP *GP*CP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	F	14	284	137	46	87	14	0	0	0
3	D	14	284	137	46	87	14	0	0	0

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



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

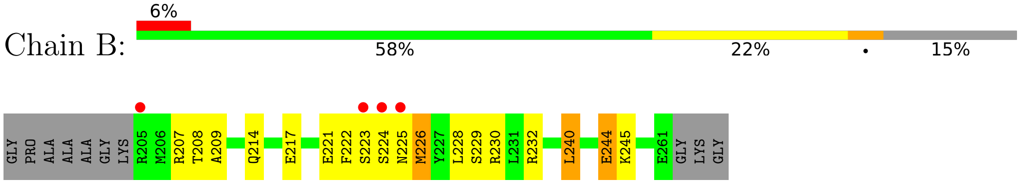
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	B	4	4	4	0	0
5	E	1	1	1	0	0
5	F	3	3	3	0	0
5	A	3	3	3	0	0
5	C	2	2	2	0	0
5	D	2	2	2	0	0

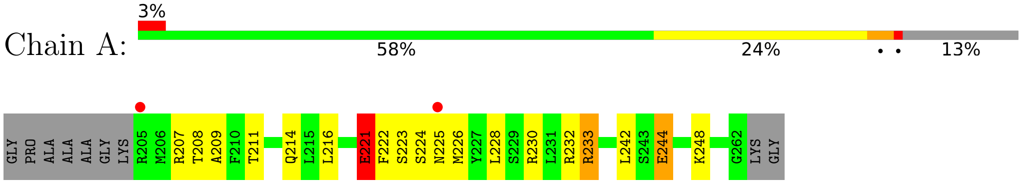
3 Residue-property plots (i)

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.

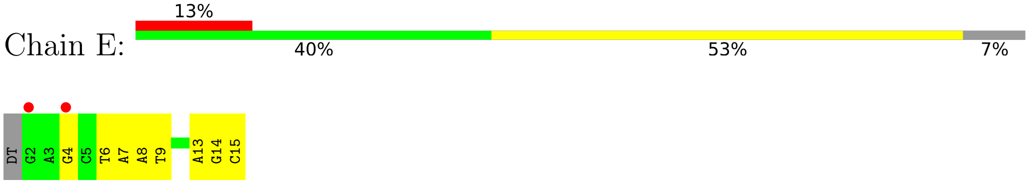
- Molecule 1: GS homeobox 2



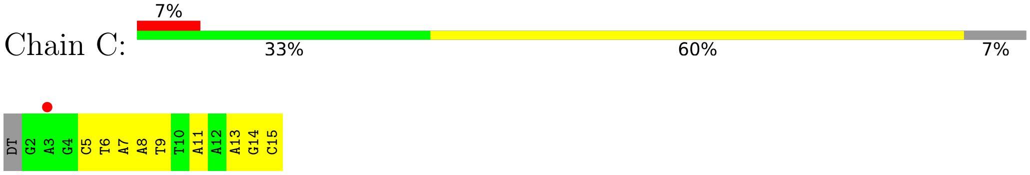
- Molecule 1: GS homeobox 2



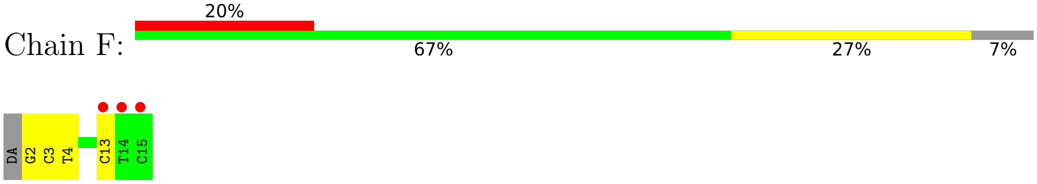
- Molecule 2: DNA (5'-D(P*GP*AP*GP*CP*TP*AP*AP*TP*TP*AP*AP*AP*GP*C)-3')



- Molecule 2: DNA (5'-D(P*GP*AP*GP*CP*TP*AP*AP*TP*TP*AP*AP*AP*GP*C)-3')

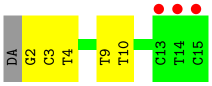


- Molecule 3: DNA (5'-D(P*GP*CP*TP*TP*TP*AP*AP*TP*TP*AP*GP*CP*TP*C)-3')



- Molecule 3: DNA (5'-D(P*GP*CP*TP*TP*TP*AP*AP*TP*TP*AP*GP*CP*TP*C)-3')

Chain D:



4 Data and refinement statistics ⓘ

Property	**Value**	**Source**
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	37.70Å 37.65Å 107.87Å 90.00° 93.99° 90.00°	Depositor
Resolution (Å)	36.30 – 2.21 36.30 – 2.21	Depositor EDS
% Data completeness (in resolution range)	98.9 (36.30-2.21) 98.9 (36.30-2.21)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.36 (at 2.22Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.220 , 0.265 0.217 , 0.263	Depositor DCC
R_{free} test set	1498 reflections (9.86%)	wwPDB-VP
Wilson B-factor (Å^2)	54.8	Xtriage
Anisotropy	0.146	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 63.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2123	wwPDB-VP
Average B, all atoms (Å^2)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.76% 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	0.73	0/484	0.95	3/649 (0.5%)
1	B	0.76	0/480	1.04	4/644 (0.6%)
2	C	1.04	0/326	1.16	4/501 (0.8%)
2	E	1.06	0/326	1.16	2/501 (0.4%)
3	D	0.98	0/316	1.10	0/485
3	F	1.01	0/316	1.13	0/485
All	All	0.92	0/2248	1.08	13/3265 (0.4%)

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

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	226	MET	CG-SD-CE	8.11	113.17	100.20
1	B	244	GLU	OE1-CD-OE2	-7.50	114.30	123.30
1	A	244	GLU	OE1-CD-OE2	-6.87	115.06	123.30
2	C	14	DG	O5'-P-OP2	-6.33	100.00	105.70
2	E	14	DG	O5'-P-OP2	-6.30	100.03	105.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	233	ARG	Sidechain

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	477	0	472	20	0
1	B	473	0	469	19	0
2	C	290	0	158	3	0
2	E	290	0	158	6	0
3	D	284	0	161	4	0
3	F	284	0	161	5	0
4	A	10	0	14	17	0
5	A	3	0	0	0	0
5	B	4	0	0	0	0
5	C	2	0	0	0	0
5	D	2	0	0	0	0
5	E	1	0	0	0	0
5	F	3	0	0	0	0
All	All	2123	0	1593	52	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 15.

The worst 5 of 52 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:233:ARG:HH11	1:A:248:LYS:HG3	1.16	1.10
1:A:233:ARG:NH1	1:A:248:LYS:HG3	1.75	1.00
1:A:221:GLU:OE2	1:A:232:ARG:NH2	2.01	0.94
2:C:8:DA:H2"	2:C:9:DT:H5"	1.64	0.80
1:B:209:ALA:H	4:A:301:PGE:H52	1.47	0.79

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	56/67 (84%)	53 (95%)	3 (5%)	0	100	100
1	B	55/67 (82%)	53 (96%)	2 (4%)	0	100	100
All	All	111/134 (83%)	106 (96%)	5 (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	50/58 (86%)	46 (92%)	4 (8%)	12	11
1	B	50/58 (86%)	46 (92%)	4 (8%)	12	11
All	All	100/116 (86%)	92 (92%)	8 (8%)	12	11

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	244	GLU
1	A	224	SER
1	A	221	GLU
1	B	240	LEU
1	A	223	SER

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	252	GLN

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 ⓘ

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	$\#\|Z\| > 2$
4	PGE	A	301	-	9,9,9	0.83	0	8,8,8	1.38	1 (12%)

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	301	-	-	6/7/7/7	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	301	PGE	O2-C3-C4	-2.40	99.58	110.39

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

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

There are no ring outliers.

1 monomer is involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	301	PGE	17	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 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	58/67 (86%)	0.51	2 (3%)	45	43	37, 61, 105, 128	0
1	B	57/67 (85%)	0.63	4 (7%)	16	15	36, 60, 99, 110	0
2	C	14/15 (93%)	0.48	1 (7%)	16	14	56, 77, 130, 144	0
2	E	14/15 (93%)	0.56	2 (14%)	2	2	56, 77, 129, 146	0
3	D	14/15 (93%)	1.21	3 (21%)	0	0	45, 75, 127, 160	0
3	F	14/15 (93%)	0.85	3 (21%)	0	0	46, 75, 128, 157	0
All	All	171/194 (88%)	0.64	15 (8%)	10	8	36, 68, 120, 160	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	15	DC	8.5
3	F	15	DC	5.2
3	D	13	DC	3.8
3	D	14	DT	3.5
2	C	3	DA	3.0

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, 95th 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	301	10/10	0.84	0.21	37,45,56,58	0

6.5 Other polymers ⓘ

There are no such residues in this entry.