



wwPDB X-ray Structure Validation Summary Report ⓘ

May 2, 2019 – 08:46 AM EDT

PDB ID : 4C7R
Title : Inward facing conformation of the trimeric betaine transporter BetP in complex with lipids
Authors : Koshy, C.; Yildiz, O.; Ziegler, C.
Deposited on : 2013-09-24
Resolution : 2.70 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.0 (224370), CSD as540be (2019)
Xtriage (Phenix) : 1.13
EDS : rb-20031633
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20031633

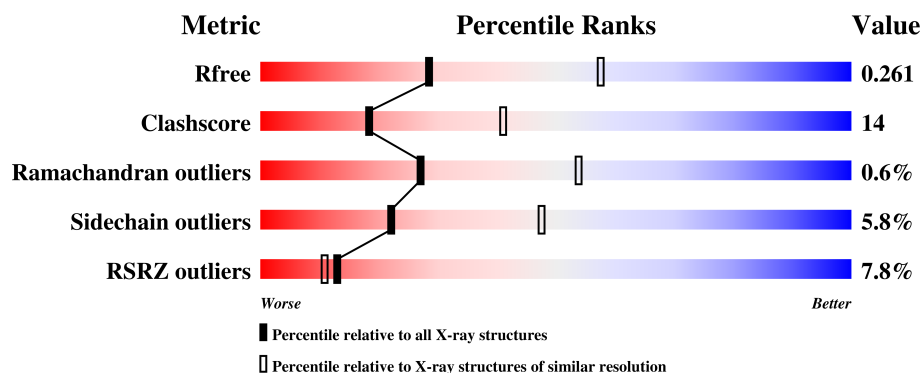
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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}	111664	2449 (2.70-2.70)
Clashscore	122126	2756 (2.70-2.70)
Ramachandran outliers	120053	2716 (2.70-2.70)
Sidechain outliers	120020	2716 (2.70-2.70)
RSRZ outliers	108989	2376 (2.70-2.70)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	566	<div> <div>6%</div> <div> <div></div> <div>67%</div> <div>23%</div> <div>• 6%</div> </div> </div>
1	B	566	<div> <div>10%</div> <div> <div></div> <div>61%</div> <div>25%</div> <div>• 12%</div> </div> </div>
1	C	566	<div> <div>5%</div> <div> <div></div> <div>62%</div> <div>24%</div> <div>• 11%</div> </div> </div>

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	PGT	A	1588	-	-	-	X
4	PGT	A	1589	-	-	-	X
4	PGT	B	1554	-	-	-	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 12208 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.

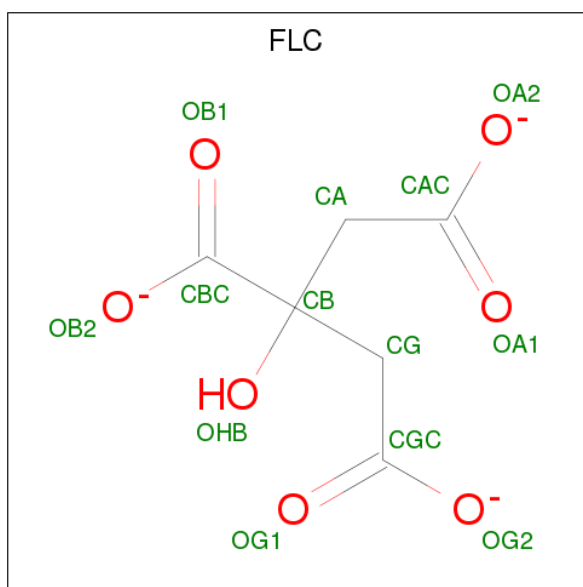
- Molecule 1 is a protein called GLYCINE BETAINES TRANSPORTER BETP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	531	Total	C	N	O	S	0	0	0
			4057	2659	675	707	16			
1	B	497	Total	C	N	O	S	0	0	0
			3742	2473	595	658	16			
1	C	506	Total	C	N	O	S	0	0	0
			3827	2525	613	673	16			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	44	ALA	GLU	engineered mutation	UNP P54582
A	45	ALA	GLU	engineered mutation	UNP P54582
A	46	ALA	GLU	engineered mutation	UNP P54582
B	44	ALA	GLU	engineered mutation	UNP P54582
B	45	ALA	GLU	engineered mutation	UNP P54582
B	46	ALA	GLU	engineered mutation	UNP P54582
C	44	ALA	GLU	engineered mutation	UNP P54582
C	45	ALA	GLU	engineered mutation	UNP P54582
C	46	ALA	GLU	engineered mutation	UNP P54582

- Molecule 2 is CITRATE ANION (three-letter code: FLC) (formula: C₆H₅O₇).

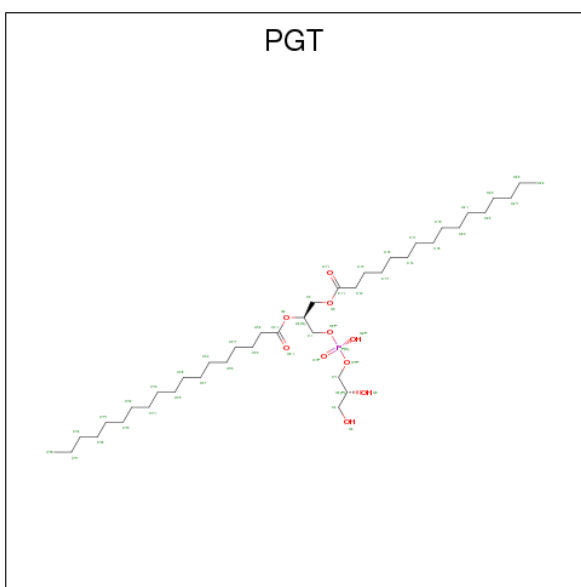


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			13	6	7		
2	B	1	Total	C	O	0	0
			13	6	7		
2	C	1	Total	C	O	0	0
			13	6	7		
2	C	1	Total	C	O	0	0
			13	6	7		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

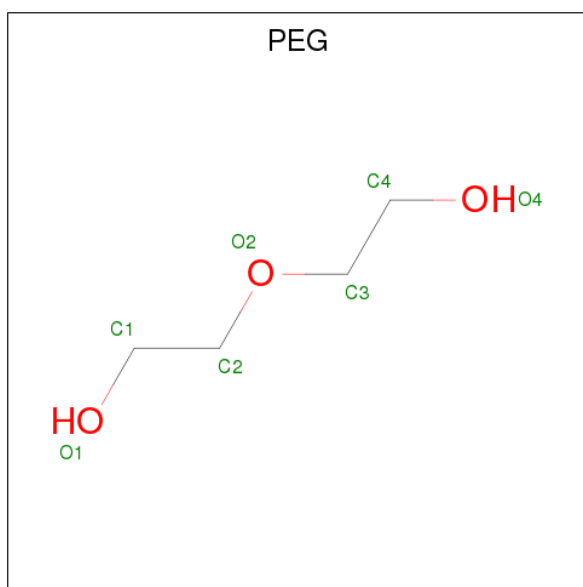
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Cl	0	0
			1	1		
3	A	1	Total	Cl	0	0
			1	1		
3	C	3	Total	Cl	0	0
			3	3		

- Molecule 4 is (1S)-2-{{[(2R)-2,3-DIHYDROXYPROPYL]OXY}(HYDROXY)PHOSPHORYL]OXY}-1-[(PALMITOYLOXY)METHYL]ETHYL STEARATE (three-letter code: PGT) (formula: C₄₀H₇₉O₁₀P).



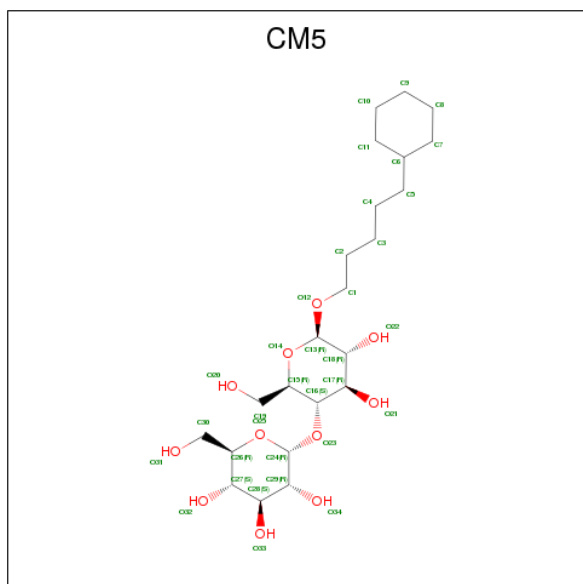
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total 51	C 40	O 10	P 1	0	0
4	A	1	Total 51	C 40	O 10	P 1	0	0
4	A	1	Total 51	C 40	O 10	P 1	0	0
4	A	1	Total 18	C 18			0	0
4	A	1	Total 51	C 40	O 10	P 1	0	0
4	B	1	Total 51	C 40	O 10	P 1	0	0
4	C	1	Total 51	C 40	O 10	P 1	0	0
4	C	1	Total 51	C 40	O 10	P 1	0	0
4	C	1	Total 51	C 40	O 10	P 1	0	0

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $\text{C}_4\text{H}_{10}\text{O}_3$).



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

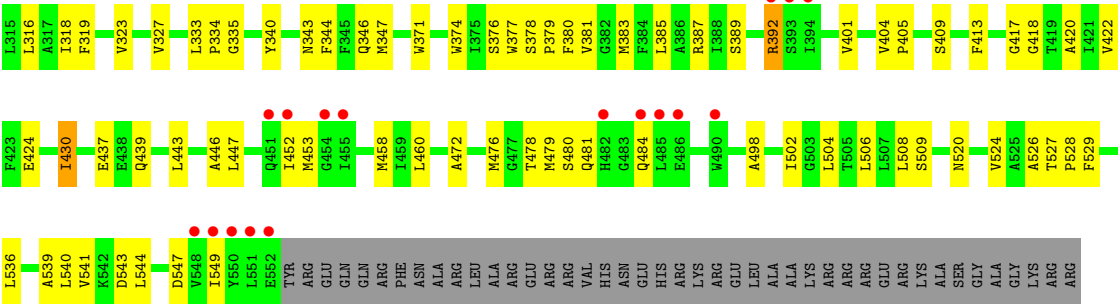
- Molecule 6 is 5-CYCLOHEXYL-1-PENTYL-BETA-D-MALTOSE (three-letter code: CM5) (formula: $C_{23}H_{42}O_{11}$).



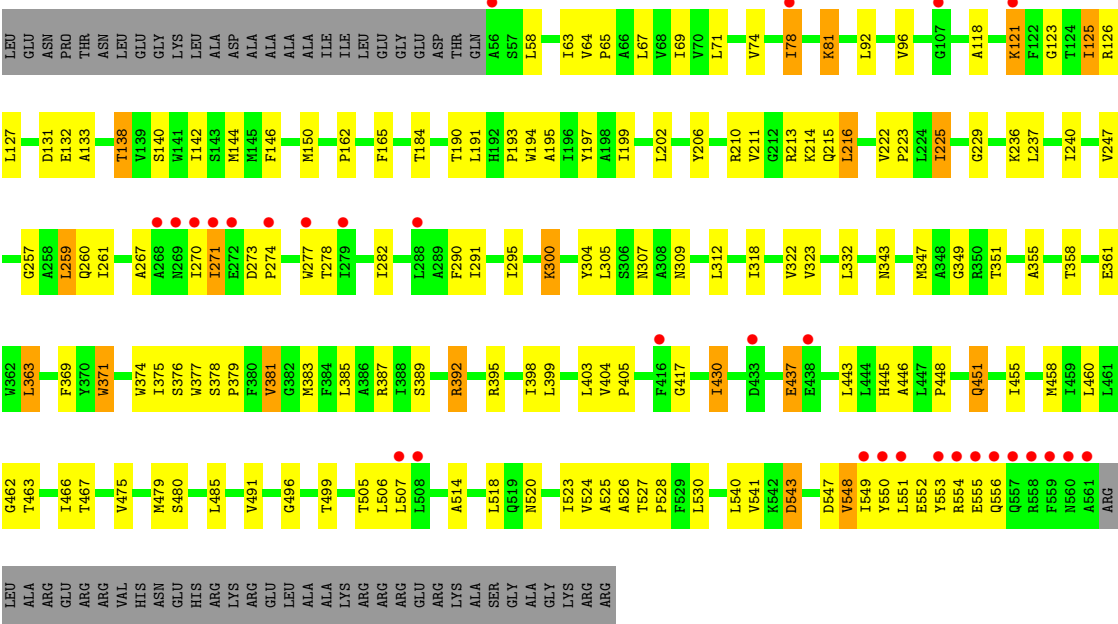
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	C	1	Total	C	O	0	0
			34	23	11		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	13	Total 13	O 13	0	0
7	B	23	Total 23	O 23	0	0
7	C	22	Total 22	O 22	0	0



• Molecule 1: GLYCINE BETAINES TRANSPORTER BETP



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	115.56Å 129.50Å 167.95Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.42 – 2.70 47.60 – 2.65	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.42-2.70) 99.7 (47.60-2.65)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.77 (at 2.65Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.212 , 0.266 0.208 , 0.261	Depositor DCC
R_{free} test set	3678 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	67.4	Xtriage
Anisotropy	0.592	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 83.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	12208	wwPDB-VP
Average B, all atoms (Å ²)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.96% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PGT, PEG, CM5, FLC, CL

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	0.24	0/4159	0.39	0/5661
1	B	0.24	0/3840	0.40	0/5241
1	C	0.26	0/3927	0.41	0/5357
All	All	0.25	0/11926	0.40	0/16259

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	4057	0	4098	111	0
1	B	3742	0	3775	100	0
1	C	3827	0	3852	118	0
2	A	13	0	5	0	0
2	B	13	0	5	3	0
2	C	26	0	10	2	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	3	0	0	0	0
4	A	222	0	347	28	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	51	0	78	5	0
4	C	153	0	234	15	0
5	B	7	0	10	0	0
6	C	34	0	42	3	0
7	A	13	0	0	0	0
7	B	23	0	0	2	0
7	C	22	0	0	0	0
All	All	12208	0	12456	352	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 14.

The worst 5 of 352 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:252:CYS:HA	1:A:518:LEU:HD11	1.49	0.92
4:A:1589:PGT:H121	1:C:395:ARG:HD2	1.50	0.90
1:C:81:LYS:HD2	1:C:81:LYS:H	1.49	0.78
1:C:553:TYR:HA	1:C:556:GLN:HG2	1.67	0.76
4:A:1588:PGT:H342	4:A:1592:PGT:H131	1.67	0.76

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	529/566 (94%)	496 (94%)	28 (5%)	5 (1%)	19	44
1	B	495/566 (88%)	458 (92%)	35 (7%)	2 (0%)	36	64
1	C	504/566 (89%)	476 (94%)	26 (5%)	2 (0%)	36	64
All	All	1528/1698 (90%)	1430 (94%)	89 (6%)	9 (1%)	27	54

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	583	ARG
1	A	431	TRP
1	B	81	LYS
1	A	585	GLU
1	C	131	ASP

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	415/439 (94%)	388 (94%)	27 (6%)	19	42
1	B	385/439 (88%)	370 (96%)	15 (4%)	35	65
1	C	393/439 (90%)	366 (93%)	27 (7%)	17	39
All	All	1193/1317 (91%)	1124 (94%)	69 (6%)	22	47

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	167	ARG
1	B	377	TRP
1	C	437	GLU
1	B	228	LYS
1	B	327	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	303	GLN
1	B	439	GLN
1	C	215	GLN
1	B	215	GLN
1	C	481	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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 20 ligands modelled in this entry, 5 are monoatomic - leaving 15 for Mogul analysis.

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	FLC	A	1001	-	3,12,12	1.07	0	3,17,17	1.51	1 (33%)
4	PGT	A	1588	-	50,50,50	0.92	2 (4%)	53,56,56	1.06	3 (5%)
4	PGT	A	1589	-	50,50,50	0.96	2 (4%)	53,56,56	0.96	2 (3%)
4	PGT	A	1590	-	50,50,50	0.93	2 (4%)	53,56,56	1.02	3 (5%)
4	PGT	A	1591	-	17,17,50	0.23	0	16,16,56	0.60	0
4	PGT	A	1592	-	50,50,50	0.91	2 (4%)	53,56,56	1.01	2 (3%)
2	FLC	B	1001	-	3,12,12	1.03	0	3,17,17	1.78	2 (66%)
4	PGT	B	1554	-	50,50,50	0.92	2 (4%)	53,56,56	0.97	2 (3%)
5	PEG	B	1555	-	6,6,6	0.45	0	5,5,5	0.26	0
2	FLC	C	1001	-	3,12,12	1.16	0	3,17,17	2.40	2 (66%)
4	PGT	C	1565	-	50,50,50	0.91	2 (4%)	53,56,56	1.03	3 (5%)
4	PGT	C	1566	-	50,50,50	0.92	2 (4%)	53,56,56	1.02	3 (5%)
4	PGT	C	1567	-	50,50,50	0.93	2 (4%)	53,56,56	0.96	3 (5%)
6	CM5	C	1568	-	36,36,36	0.40	0	49,49,49	0.79	1 (2%)
2	FLC	C	1569	-	3,12,12	1.14	0	3,17,17	2.26	2 (66%)

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	FLC	A	1001	-	-	0/6/16/16	0/0/0/0
4	PGT	A	1588	-	-	0/55/55/55	0/0/0/0
4	PGT	A	1589	-	-	0/55/55/55	0/0/0/0
4	PGT	A	1590	-	-	0/55/55/55	0/0/0/0
4	PGT	A	1591	-	-	0/15/15/55	0/0/0/0
4	PGT	A	1592	-	-	0/55/55/55	0/0/0/0
2	FLC	B	1001	-	-	0/6/16/16	0/0/0/0
4	PGT	B	1554	-	-	0/55/55/55	0/0/0/0
5	PEG	B	1555	-	-	0/4/4/4	0/0/0/0
2	FLC	C	1001	-	-	0/6/16/16	0/0/0/0
4	PGT	C	1565	-	-	0/55/55/55	0/0/0/0
4	PGT	C	1566	-	-	0/55/55/55	0/0/0/0
4	PGT	C	1567	-	-	0/55/55/55	0/0/0/0
6	CM5	C	1568	-	-	0/17/65/65	0/3/3/3
2	FLC	C	1569	-	-	0/6/16/16	0/0/0/0

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1590	PGT	O2-C31	3.96	1.45	1.34
4	C	1565	PGT	O2-C31	3.99	1.45	1.34
4	A	1592	PGT	O2-C31	4.02	1.45	1.34
4	C	1566	PGT	O2-C31	4.08	1.45	1.34
4	C	1567	PGT	O2-C31	4.14	1.46	1.34

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1590	PGT	C2-O2-C31	-3.43	109.26	117.82
4	C	1565	PGT	C2-O2-C31	-3.39	109.35	117.82
2	C	1001	FLC	CB-CA-CAC	-3.20	109.87	114.98
6	C	1568	CM5	C24-O23-C16	-2.83	110.92	117.97
2	C	1569	FLC	CB-CA-CAC	-2.80	110.50	114.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

12 monomers are involved in 53 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1588	PGT	8	0
4	A	1589	PGT	16	0
4	A	1590	PGT	4	0
4	A	1591	PGT	1	0
4	A	1592	PGT	6	0
2	B	1001	FLC	3	0
4	B	1554	PGT	5	0
2	C	1001	FLC	2	0
4	C	1565	PGT	8	0
4	C	1566	PGT	2	0
4	C	1567	PGT	6	0
6	C	1568	CM5	3	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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	531/566 (93%)	0.15	32 (6%) 22 20	58, 102, 170, 244	1 (0%)
1	B	497/566 (87%)	0.39	57 (11%) 5 4	48, 104, 176, 248	0
1	C	506/566 (89%)	0.19	30 (5%) 22 21	48, 86, 159, 215	0
All	All	1534/1698 (90%)	0.24	119 (7%) 13 11	48, 98, 170, 248	1 (0%)

The worst 5 of 119 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	550	TYR	9.7
1	C	559	PHE	7.9
1	C	561	ALA	7.6
1	B	549	ILE	7.2
1	A	274	PRO	7.0

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 carbohydrates 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, 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(\AA^2)	Q<0.9
4	PGT	B	1554	51/51	0.37	0.51	72,136,208,240	0
4	PGT	A	1588	51/51	0.47	0.46	85,136,175,236	0
4	PGT	A	1589	51/51	0.53	0.49	74,132,232,385	0
4	PGT	A	1592	51/51	0.54	0.34	69,150,203,314	0
4	PGT	C	1567	51/51	0.59	0.36	107,157,233,319	0
4	PGT	C	1565	51/51	0.66	0.33	106,176,230,291	0
2	FLC	C	1569	13/13	0.70	0.39	188,193,208,210	0
4	PGT	C	1566	51/51	0.72	0.20	109,144,184,223	0
2	FLC	B	1001	13/13	0.74	0.33	140,150,161,162	0
6	CM5	C	1568	34/34	0.75	0.28	112,182,209,211	0
4	PGT	A	1590	51/51	0.76	0.28	72,99,130,231	0
5	PEG	B	1555	7/7	0.77	0.26	146,152,153,155	0
4	PGT	A	1591	18/51	0.84	0.77	86,92,114,114	0
2	FLC	A	1001	13/13	0.85	0.27	147,163,169,170	0
2	FLC	C	1001	13/13	0.91	0.36	133,140,145,145	0
3	CL	B	1553	1/1	0.94	0.49	109,109,109,109	0
3	CL	A	1587	1/1	0.95	0.35	105,105,105,105	0
3	CL	C	1563	1/1	0.97	0.09	64,64,64,64	0
3	CL	C	1564	1/1	0.98	0.16	85,85,85,85	0
3	CL	C	1562	1/1	0.98	0.10	74,74,74,74	0

6.5 Other polymers [i](#)

There are no such residues in this entry.