



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2026 – 04:36 PM UTC

PDB ID : 4DOJ / pdb_00004doj
Title : Crystal structure of BetP in outward-facing conformation
Authors : Perez, C.; Ziegler, C.
Deposited on : 2012-02-09
Resolution : 3.25 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Buster-report : wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

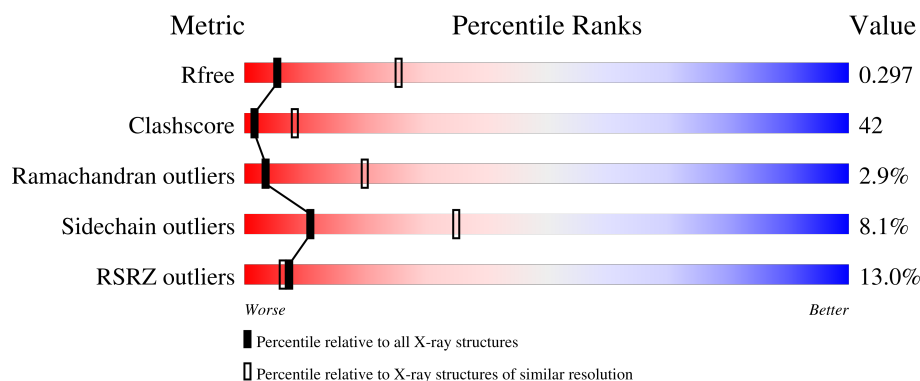
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.25 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}	180053	1605 (3.30-3.22)
Clashscore	190562	1660 (3.30-3.22)
Ramachandran outliers	187476	1630 (3.30-3.22)
Sidechain outliers	187428	1629 (3.30-3.22)
RSRZ outliers	180081	1605 (3.30-3.22)

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 $\leq 5\%$. 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	14% 39% 47% 7% • 7%
1	B	566	16% 36% 44% 7% • 11%
1	C	566	5% 36% 50% • 10%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11763 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 betaine transporter BetP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	526	Total	C	N	O	S	0	0	0
			4017	2635	664	702	16			
1	B	501	Total	C	N	O	S	0	0	0
			3794	2502	607	669	16			
1	C	507	Total	C	N	O	S	0	0	0
			3860	2542	626	676	16			

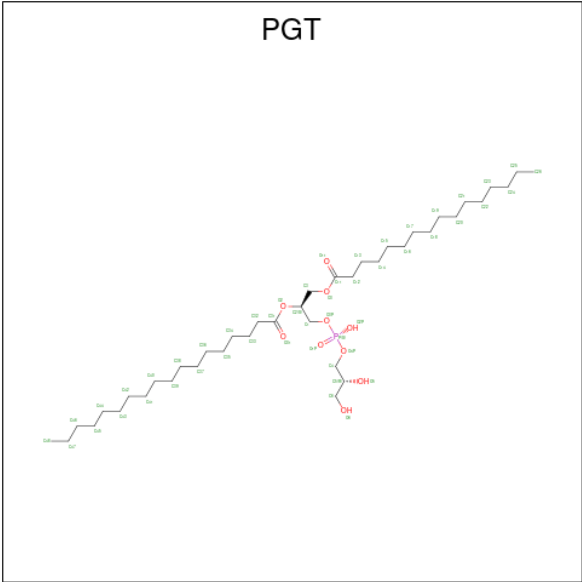
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	153	ASP	GLY	engineered mutation	UNP P54582
B	153	ASP	GLY	engineered mutation	UNP P54582
C	153	ASP	GLY	engineered mutation	UNP P54582

- Molecule 2 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

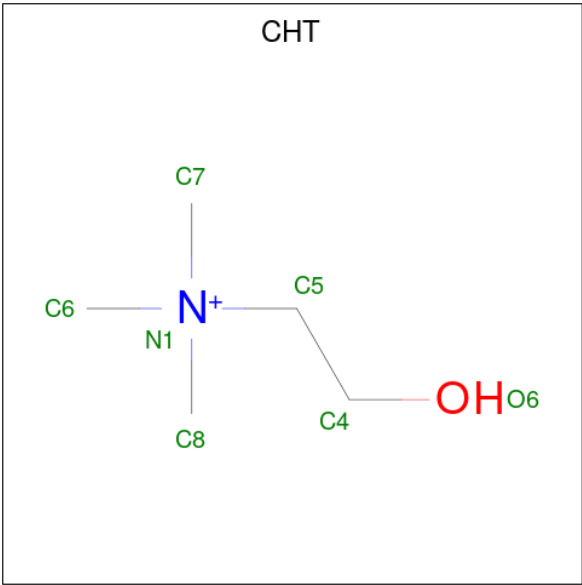
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Cl	0	0
			1	1		
2	C	1	Total	Cl	0	0
			1	1		

- Molecule 3 is (1S)-2-{{[(2R)-2,3-DIHYDROXYPROPYL]OXY}(HYDROXY)PHOSPHORYL]OXY}-1-[(PALMITOYL OXY)METHYL]ETHYL STEARATE (CCD ID: PGT) (formula: C₄₀H₇₉O₁₀P).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	O	P	5	0
			51	40	10	1		

- Molecule 4 is CHOLINE ION (CCD ID: CHT) (formula: C₅H₁₄NO).



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

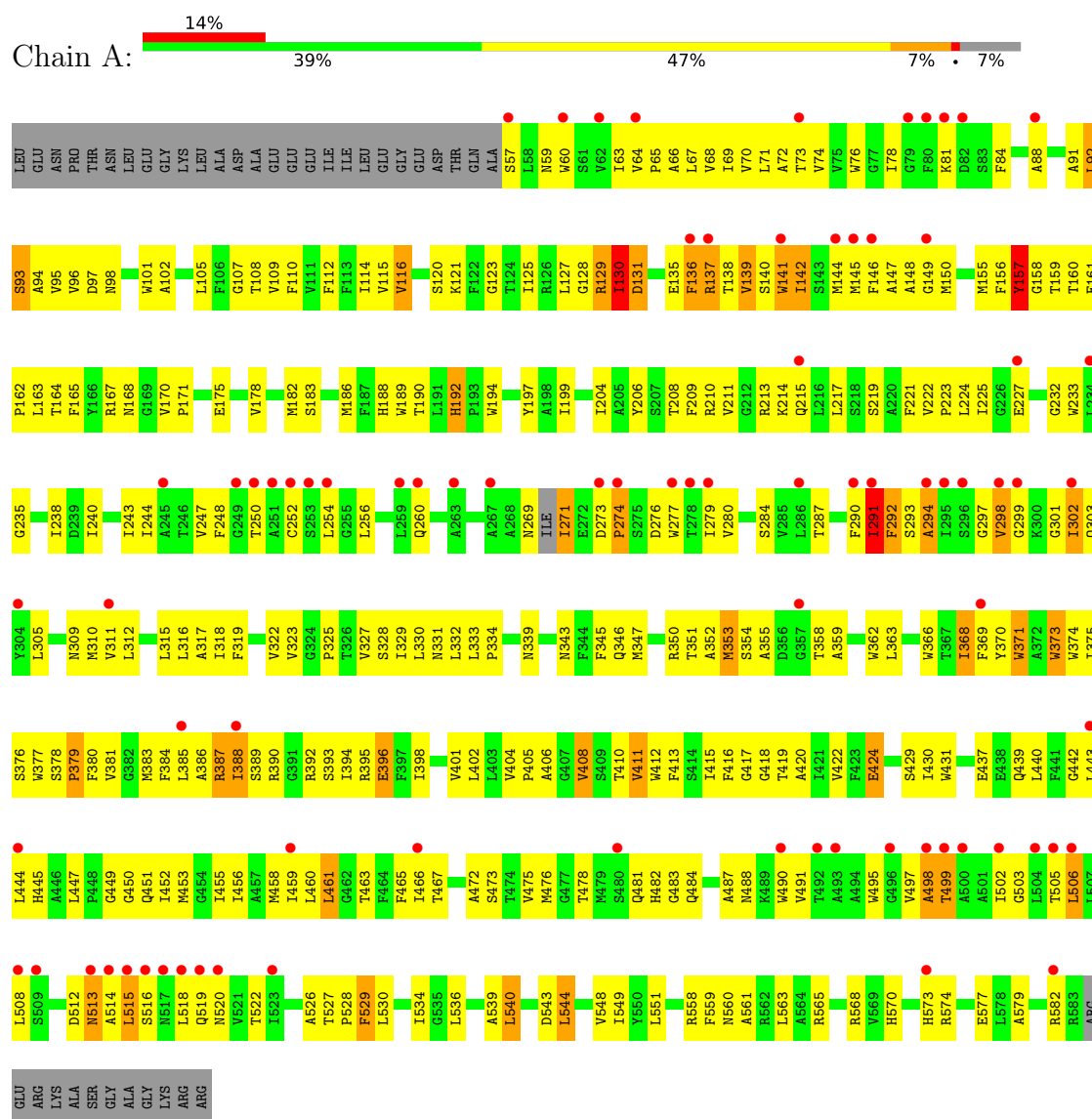
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	7	Total 7	O 7	0	0
5	B	12	Total 12	O 12	0	0
5	C	13	Total 13	O 13	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: Glycine betaine transporter BetP



• Molecule 1: Glycine betaine transporter BetP



ALA	L518	Q439	Y370	I302	●
LYS	Q519	L440	W371	Q303	
ARG	N520	L443	W374	Y304	
ARG	V521		L305	L306	
GLU	T522		I375	S306	
ARG	I523		S376	N307	
ARG	V524		W377	A308	
LYS			S378	N309	
ALA	T527		P379	N310	
ALA	P528		F380	V311	
SER	Q451	●	F381	V312	
GLY	F529	●	V381	L312	
GLY	L530		G382	A313	
ALA	F531		M383	A314	
GLY	V532		F384	L315	
LYS	V533		L385	L316	
ARG	V534		A386	A317	
ARG	I534		R387	I318	
	Q535		I388	F319	
	L536		S389	V320	
	M537		R390		
	F538		G391		
	A539			V323	
	L540			G324	
	V541			P325	
				T326	
	R542			V327	
	D543			S328	
	L544			I329	
	S545			L330	
	N546			N331	
				L332	
				I333	
	I549			P334	
	Y550				
	L551			I337	
	E552				
	Y553			Y340	
	R554			L341	
	E555			S342	
	Q556	●		N343	
	Q557			F344	
	R558			F345	
	F559			Q346	
	N560			M347	
	A561				
	R562			A352	
	L563			N353	
	A564			S354	
	R565			A355	
	E566			D356	
	R567			G357	
	R568	●		T358	
	VAL			A359	
HIS					
ASN	I502			L363	
GLU	G503			G364	
GLU	L504			S365	
HIS	T505			V366	
ARG	L506			T367	
LYS	L507			I368	
LYS	ARG			F369	
ARG	S509	●			
GLU					
LEU					
ALA	N517				

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	117.44Å 129.32Å 184.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.80 – 3.25 29.80 – 3.25	Depositor EDS
% Data completeness (in resolution range)	86.0 (29.80-3.25) 85.8 (29.80-3.25)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	14.40 (at 3.19Å)	Xtriage
Refinement program	PHENIX 1.7_650	Depositor
R, R_{free}	0.249 , 0.297 0.248 , 0.297	Depositor DCC
R_{free} test set	3848 reflections (9.92%)	wwPDB-VP
Wilson B-factor (Å ²)	79.9	Xtriage
Anisotropy	0.018	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 121.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	11763	wwPDB-VP
Average B, all atoms (Å ²)	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.91% 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 [i](#)

5.1 Standard geometry [i](#)

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

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	0.30	0/4119	0.71	0/5609
1	B	0.34	0/3892	0.72	3/5308 (0.1%)
1	C	0.32	0/3958	0.74	2/5393 (0.0%)
All	All	0.32	0/11969	0.72	5/16310 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	155	MET	N-CA-C	-7.21	103.59	113.18
1	C	90	SER	N-CA-C	-5.67	106.53	113.50
1	C	337	ILE	N-CA-C	-5.43	107.07	113.42
1	B	273	ASP	CA-C-N	5.04	126.14	119.84
1	B	273	ASP	C-N-CA	5.04	126.14	119.84

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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	4017	0	4051	378	0
1	B	3794	0	3816	334	0
1	C	3860	0	3888	288	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	1	0	0	0	0
2	C	1	0	0	1	0
3	A	51	0	78	10	0
4	C	7	0	14	2	0
5	A	7	0	0	1	0
5	B	12	0	0	2	0
5	C	13	0	0	2	0
All	All	11763	0	11847	979	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 42.

The worst 5 of 979 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:297:GLY:H	1:A:298:VAL:CG2	1.52	1.21
1:A:506:LEU:O	1:A:506:LEU:HD23	1.37	1.18
1:C:226:GLY:HA2	1:C:227:GLU:HB3	1.29	1.07
1:A:292:PHE:H	1:A:293:SER:HB2	1.17	1.05
1:A:297:GLY:H	1:A:298:VAL:HG22	1.12	1.04

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	524/566 (93%)	414 (79%)	90 (17%)	20 (4%)	2	15
1	B	497/566 (88%)	406 (82%)	76 (15%)	15 (3%)	3	19
1	C	501/566 (88%)	411 (82%)	81 (16%)	9 (2%)	6	29
All	All	1522/1698 (90%)	1231 (81%)	247 (16%)	44 (3%)	3	20

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	137	ARG
1	A	498	ALA
1	A	499	THR
1	A	513	ASN
1	B	134	PRO

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	412/443 (93%)	385 (93%)	27 (7%)	15	42
1	B	391/443 (88%)	348 (89%)	43 (11%)	6	23
1	C	397/443 (90%)	370 (93%)	27 (7%)	14	41
All	All	1200/1329 (90%)	1103 (92%)	97 (8%)	11	35

5 of 97 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	402	LEU
1	C	92	LEU
1	B	409	SER
1	B	470	ASP
1	C	117	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	426	ASN
1	C	519	GLN
1	B	482	HIS
1	C	556	GLN
1	C	439	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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	$\# Z > 2$
3	PGT	A	602	-	50,50,50	0.86	2 (4%)	53,56,56	1.12	3 (5%)
4	CHT	C	601	-	6,6,6	0.86	0	8,8,8	0.37	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	PGT	A	602	-	-	27/55/55/55	-
4	CHT	C	601	-	-	0/4/4/4	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	602	PGT	O2-C2	-3.03	1.39	1.46

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	602	PGT	O3-C3	-2.41	1.39	1.45

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	602	PGT	O2-C31-C32	4.31	120.81	111.48
3	A	602	PGT	O3-C3-C2	3.03	117.14	108.40
3	A	602	PGT	O3-C11-C12	2.34	118.97	111.83

There are no chirality outliers.

5 of 27 torsion outliers are listed below:

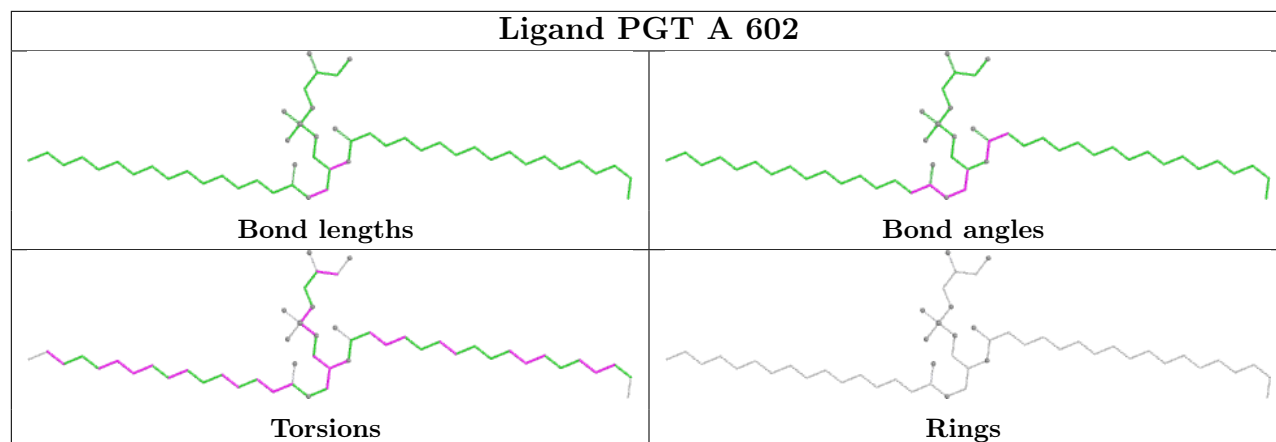
Mol	Chain	Res	Type	Atoms
3	A	602	PGT	C1-O3P-P-O1P
3	A	602	PGT	C1-O3P-P-O2P
3	A	602	PGT	C1-O3P-P-O4P
3	A	602	PGT	C20-C21-C22-C23
3	A	602	PGT	C4-C5-C6-O6

There are no ring outliers.

2 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	602	PGT	10	0
4	C	601	CHT	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



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	526/566 (92%)	0.86	80 (15%) 5 5	22, 120, 299, 416	0
1	B	501/566 (88%)	0.87	89 (17%) 4 3	13, 116, 284, 438	0
1	C	507/566 (89%)	0.15	30 (5%) 28 19	9, 75, 181, 389	0
All	All	1534/1698 (90%)	0.63	199 (12%) 7 6	9, 98, 275, 438	0

The worst 5 of 199 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	253	SER	12.5
1	B	250	THR	7.7
1	C	270	ILE	6.4
1	B	254	LEU	6.2
1	A	517	ASN	6.2

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](#)

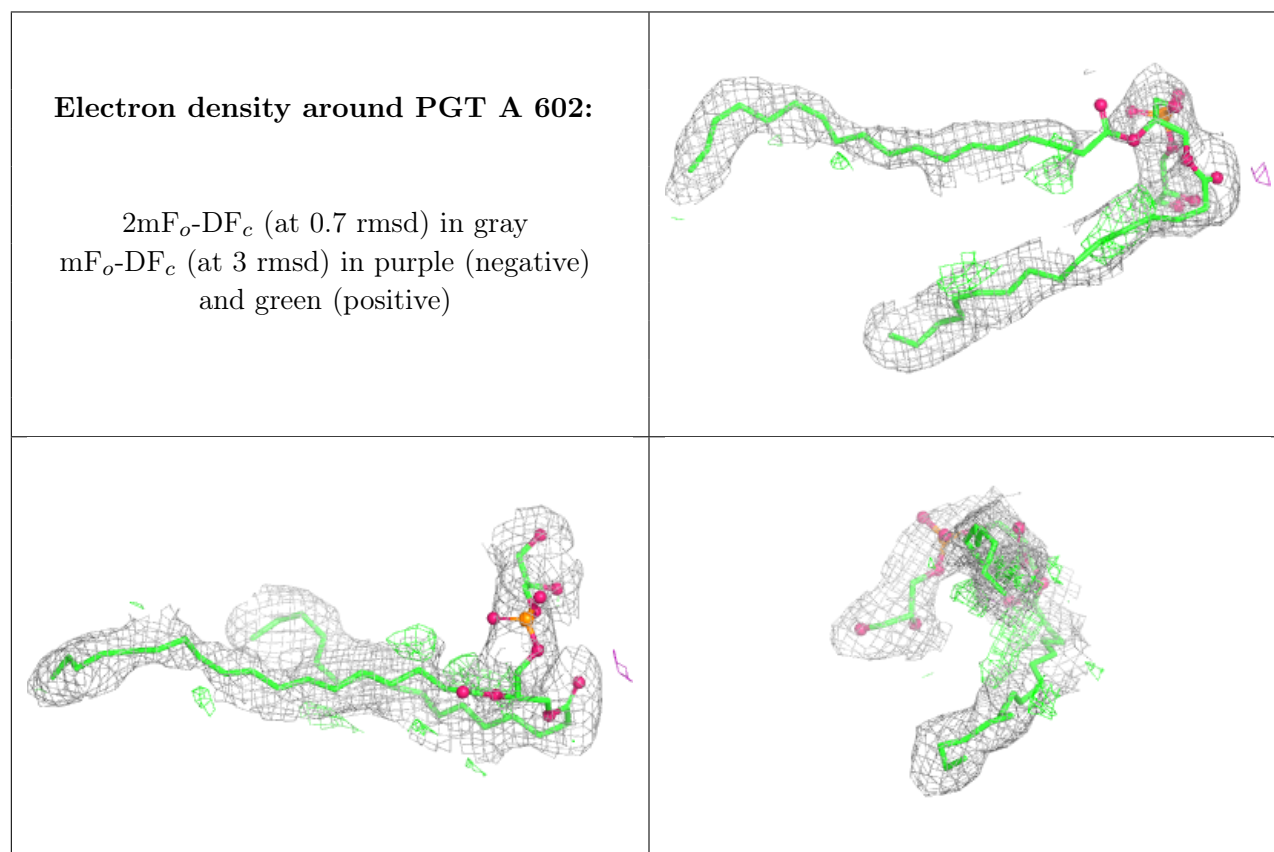
There are no oligosaccharides in this entry.

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($\text{\AA}^2$)	Q<0.9
4	CHT	C	601	7/7	0.84	0.27	99,103,112,121	0
2	CL	A	601	1/1	0.85	0.21	105,105,105,105	0
3	PGT	A	602	51/51	0.91	0.13	24,47,72,73	30
2	CL	C	602	1/1	0.98	0.06	57,57,57,57	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



6.5 Other polymers [i](#)

There are no such residues in this entry.