



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 20, 2026 – 12:13 AM UTC

PDB ID : 3P03 / pdb_00003p03
Title : Crystal structure of BetP-G153D with choline bound
Authors : Perez, C.; Ressler, S.; Ziegler, Z.
Deposited on : 2010-09-27
Resolution : 3.35 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
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

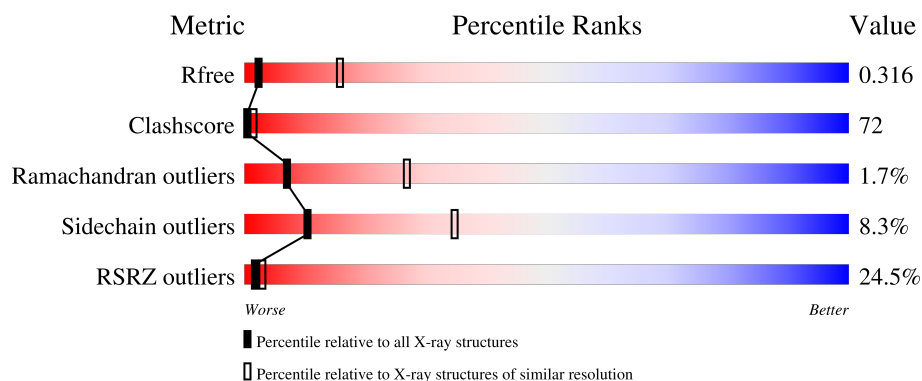
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.35 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	1099 (3.40-3.32)
Clashscore	190562	1116 (3.40-3.32)
Ramachandran outliers	187476	1101 (3.40-3.32)
Sidechain outliers	187428	1101 (3.40-3.32)
RSRZ outliers	180081	1099 (3.40-3.32)

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	25% 26% 56% 6% • 11%
1	B	566	26% 26% 51% 7% • 16%
1	C	566	13% 24% 54% 11% • 10%

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
2	CHT	C	2486	-	-	X	-

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11349 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.

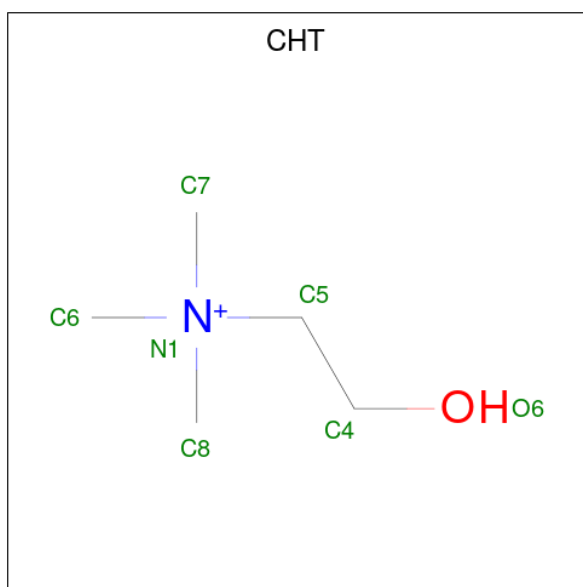
- Molecule 1 is a protein called Glycine betaine transporter BetP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	504	Total	C	N	O	S	0	0	0
			3868	2542	641	669	16			
1	B	476	Total	C	N	O	S	0	0	0
			3612	2391	577	628	16			
1	C	508	Total	C	N	O	S	0	0	0
			3862	2545	627	674	16			

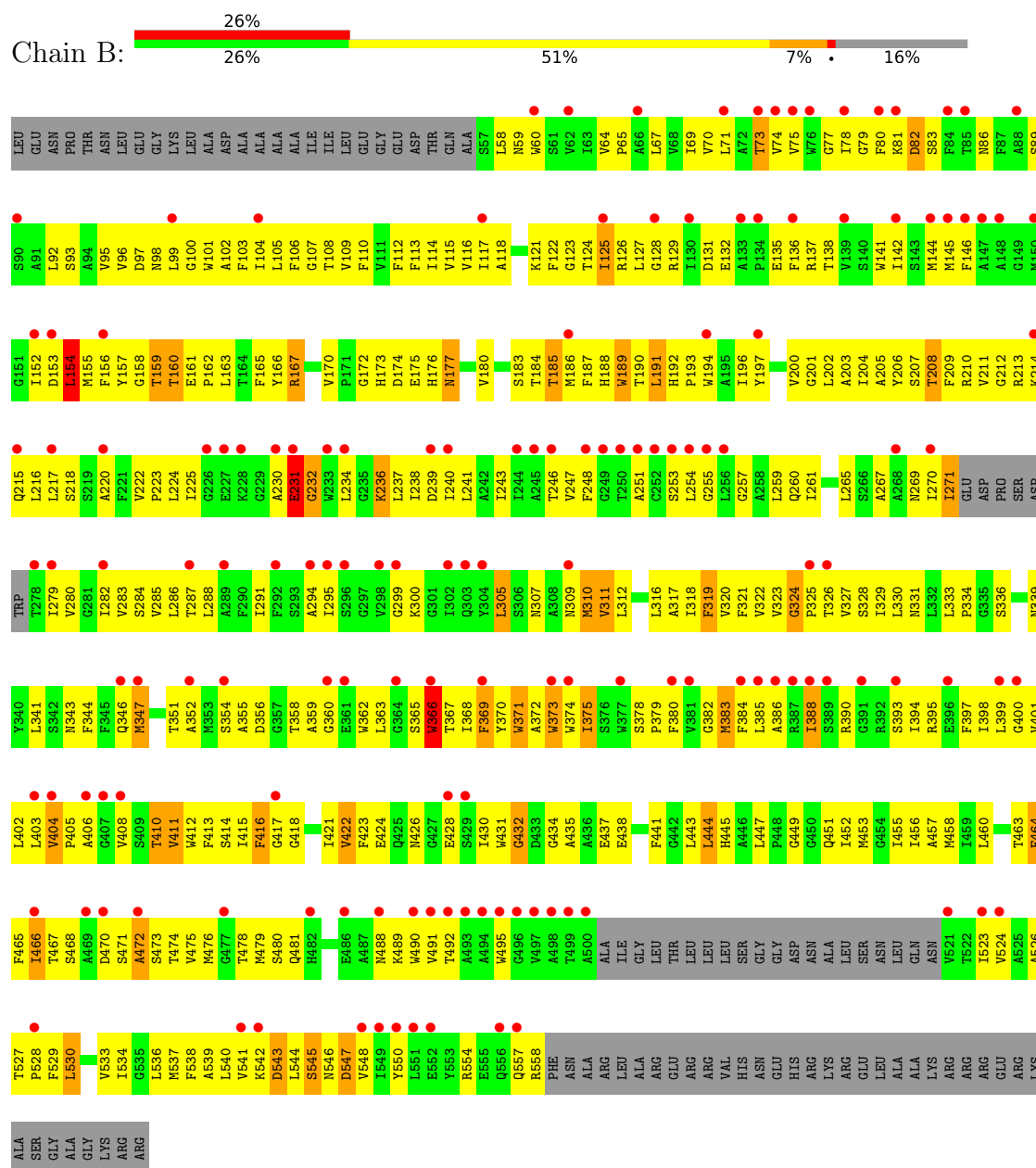
There are 12 discrepancies between the modelled and reference sequences:

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
A	153	ASP	GLY	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
B	153	ASP	GLY	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
C	153	ASP	GLY	engineered mutation	UNP P54582

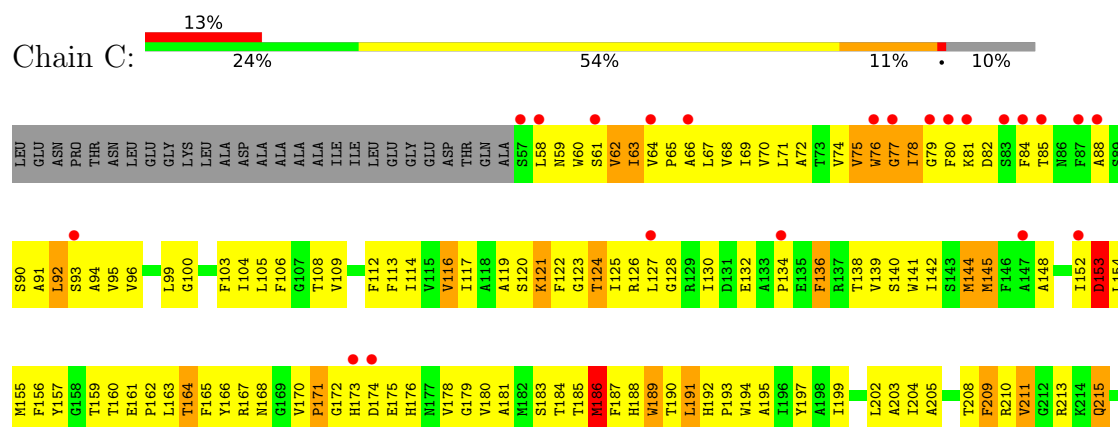
- Molecule 2 is CHOLINE ION (CCD ID: CHT) (formula: C₅H₁₄NO).

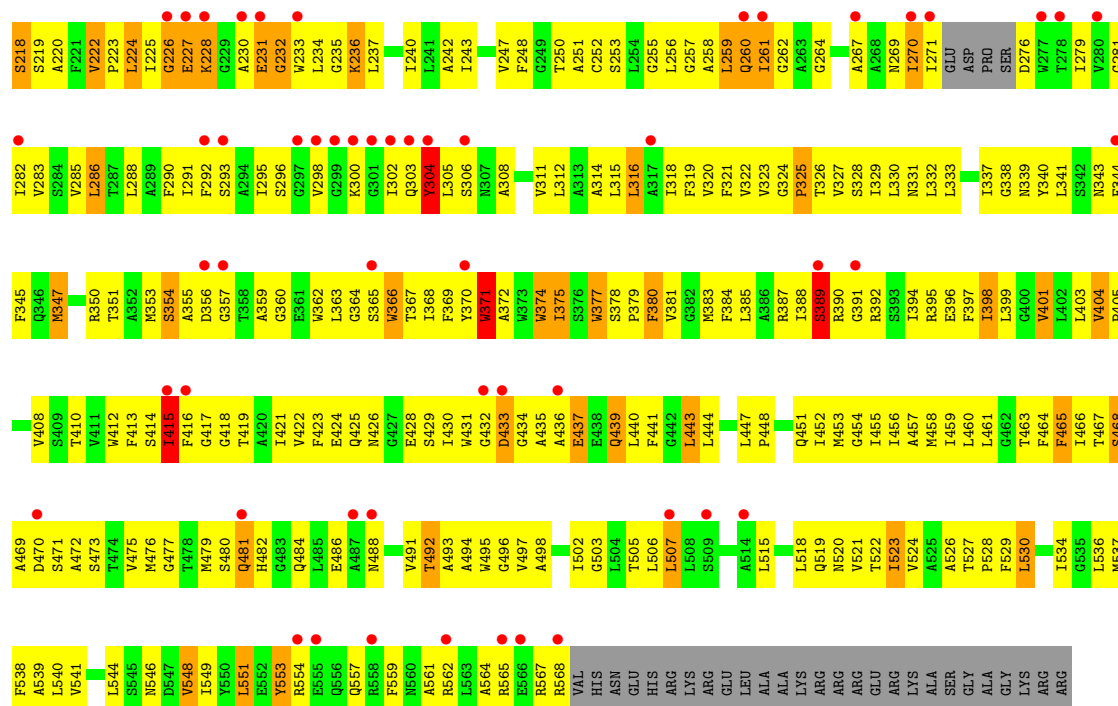


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	C	1	7	5	1	1	0	0



• Molecule 1: Glycine betaine 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	117.56Å 129.31Å 183.14Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.20 – 3.35 46.20 – 3.35	Depositor EDS
% Data completeness (in resolution range)	91.2 (46.20-3.35) 91.1 (46.20-3.35)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	15.91 (at 3.33Å)	Xtriage
Refinement program	PHENIX 1.6.2_432	Depositor
R, R_{free}	0.245 , 0.300 0.278 , 0.316	Depositor DCC
R_{free} test set	3712 reflections (9.95%)	wwPDB-VP
Wilson B-factor (Å ²)	70.2	Xtriage
Anisotropy	0.012	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 115.1	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	0.83	EDS
Total number of atoms	11349	wwPDB-VP
Average B, all atoms (Å ²)	107.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.96% 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: 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	0.43	1/3967 (0.0%)	0.96	15/5397 (0.3%)
1	B	0.47	0/3706	1.03	23/5051 (0.5%)
1	C	0.57	2/3960 (0.1%)	1.13	37/5396 (0.7%)
All	All	0.50	3/11633 (0.0%)	1.04	75/15844 (0.5%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	368	ILE	CA-CB	-7.01	1.44	1.54
1	C	366	TRP	CA-C	-6.50	1.48	1.52
1	C	371	TRP	C-O	-5.21	1.18	1.24

The worst 5 of 75 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	77	GLY	N-CA-C	10.54	125.37	112.73
1	C	415	ILE	N-CA-C	9.99	119.92	110.53
1	A	155	MET	N-CA-C	9.41	121.30	111.14
1	B	160	THR	N-CA-C	9.01	121.18	111.36
1	C	63	ILE	N-CA-C	8.88	118.94	110.42

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	3868	0	3897	529	0
1	B	3612	0	3647	560	0
1	C	3862	0	3899	600	0
2	C	7	0	14	6	0
All	All	11349	0	11457	1649	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 72.

The worst 5 of 1649 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:170:VAL:CG1	1:C:171:PRO:HD2	1.50	1.41
1:C:226:GLY:HA2	1:C:227:GLU:CB	1.47	1.40
1:C:247:VAL:HG12	1:C:502:ILE:CD1	1.54	1.35
1:B:69:ILE:O	1:B:73:THR:HG23	1.33	1.29
1:B:208:THR:HG22	1:B:213:ARG:O	1.32	1.27

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	498/566 (88%)	426 (86%)	62 (12%)	10 (2%)	6	24
1	B	470/566 (83%)	403 (86%)	61 (13%)	6 (1%)	9	32
1	C	504/566 (89%)	435 (86%)	60 (12%)	9 (2%)	6	26
All	All	1472/1698 (87%)	1264 (86%)	183 (12%)	25 (2%)	7	27

5 of 25 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	377	TRP
1	A	403	LEU
1	B	231	GLU
1	B	432	GLY
1	C	231	GLU

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	396/440 (90%)	367 (93%)	29 (7%)	13	39
1	B	371/440 (84%)	345 (93%)	26 (7%)	14	40
1	C	396/440 (90%)	354 (89%)	42 (11%)	6	23
All	All	1163/1320 (88%)	1066 (92%)	97 (8%)	10	35

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

Mol	Chain	Res	Type
1	C	92	LEU
1	C	259	LEU
1	C	116	VAL
1	C	186	MET
1	C	347	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	59	ASN
1	C	260	GLN
1	C	557	GLN
1	C	192	HIS
1	C	307	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 oligosaccharides 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	$\# Z > 2$
2	CHT	C	2486	-	6,6,6	0.81	0	8,8,8	0.28	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	CHT	C	2486	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	2486	CHT	C4-C5-N1-C6
2	C	2486	CHT	C4-C5-N1-C8
2	C	2486	CHT	C4-C5-N1-C7
2	C	2486	CHT	O6-C4-C5-N1

There are no ring outliers.

1 monomer is involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	2486	CHT	6	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	504/566 (89%)	1.42	143 (28%) 1 2	28, 107, 309, 534	0
1	B	476/566 (84%)	1.59	147 (30%) 1 1	14, 97, 292, 557	0
1	C	508/566 (89%)	0.77	75 (14%) 5 6	7, 62, 194, 353	0
All	All	1488/1698 (87%)	1.25	365 (24%) 2 3	7, 86, 284, 557	0

The worst 5 of 365 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	496	GLY	11.2
1	A	296	SER	10.3
1	B	500	ALA	10.3
1	B	499	THR	9.7
1	A	490	TRP	9.0

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 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
2	CHT	C	2486	7/7	0.76	0.36	85,89,97,97	0

6.5 Other polymers [i](#)

There are no such residues in this entry.