



wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2020 – 10:21 am BST

PDB ID : 3P03
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 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	:	2.11
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)	:	2.11

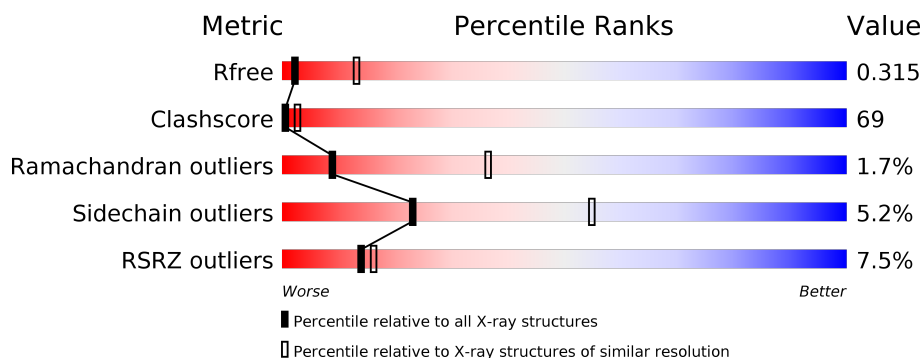
1 Overall quality at a glance

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}	130704	1558 (3.42-3.30)
Clashscore	141614	1627 (3.42-3.30)
Ramachandran outliers	138981	1599 (3.42-3.30)
Sidechain outliers	138945	1598 (3.42-3.30)
RSRZ outliers	127900	1507 (3.42-3.30)

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 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	<div> <div>8%</div> <div>30%</div> <div>55%</div> <div>11%</div> </div>
1	B	566	<div> <div>8%</div> <div>28%</div> <div>52%</div> <div>16%</div> </div>
1	C	566	<div> <div>4%</div> <div>27%</div> <div>57%</div> <div>10%</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
2	CHT	C	2486	-	-	X	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.

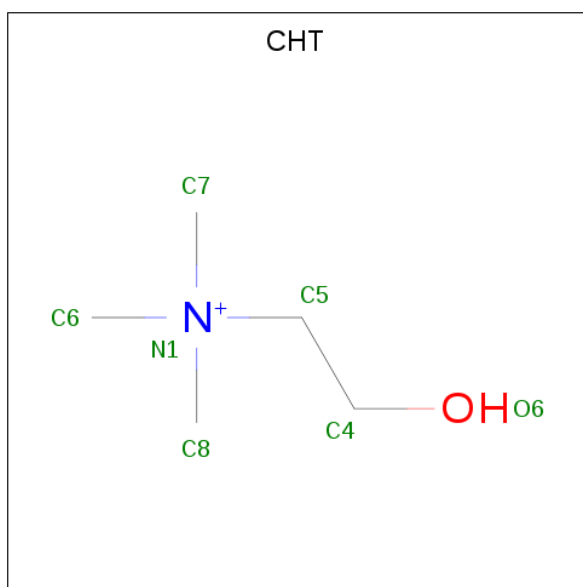
- Molecule 1 is a protein called Glycine betaine transporter BetP.

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 (three-letter code: CHT) (formula: C₅H₁₄NO).

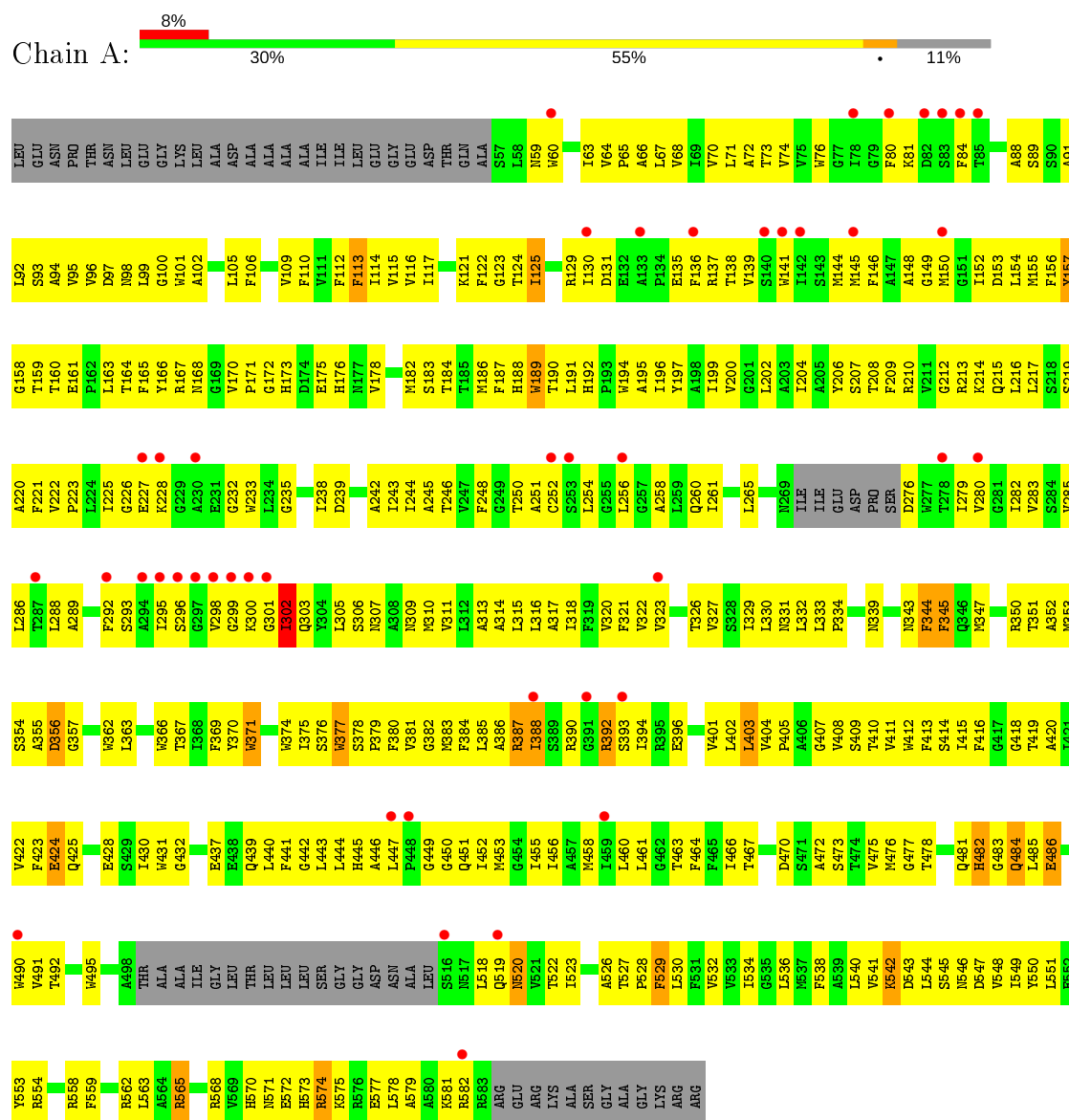


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	C	1	7	5	1	1	0	0

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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





Y553	Y554	E555	Q556	Q557	R558	F559	R560	A561	R562	L563	A564	R565	E566	R567	R568	VAL	HIS	ASN	GLU	HIS	ARG	LYS	ARG	GLU	LEU	ALA	ALA	LYS	ARG	ARG	ARG	GLU	ALA	LYS	ASP	GLY	ALA	LYS	ARG																	
L288	A289	F290	I291	F292	I295	S296	G297	V298	G299	K300	G301	I302	Q303	Y304	L305	S306	R307	A308	V311	L312	A313	A314	L315	L316	A317	I318	F319	V320	F321	V322	V323	G324	P325	T326	V327	S328	I329	L330	N331	L332	L333															
N353	S354	A355	D356	I357	T358	A359	W362	L363	G364	S365	W366	T367	I368	F369	Y370	W371	A372	W373	W374	I375	S376	W377	S378	P379	F380	V381	G382	N383	F384	L385	A386	I388	S389	R390	G391	R392	S393	I394	R395	E396	F397	I398	L399	G400	V401	V404	P405	T410	W411	W412	P413	S414	I415	F416	T351	A352
G418	I421	F423	F424	Q425	W426	G427	E428	S429	I430	W431	G432	D433	G434	A435	A436	E437	E438	Q439	L440	F441	G442	L443	L444	L447	P448	Q451	I452	M453	G454	I455	I456	A457	M458	I459	L460	L461	G462	T463	F464	F465	T466	T467	S468	A469	D470	S471	A472	S473	T474	V475	M476	G477	T478	M479	S480	
Q481	E482	G483	Q484	L485	E486	A487	M488	V491	T492	A493	A494	W495	G496	V497	A498	I502	G503	L504	T505	L506	L507	L515	L518	Q519	N520	V521	T522	L523	V524	A525	A526	T527	P528	F529	L530	I534	G535	L536	M537	F538	A539	L540	V541	K542	D543	L544	S545	N546	D547	Y548	I549	Y550	L551	E552		

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 48.18 – 3.30	Depositor EDS
% Data completeness (in resolution range)	91.2 (46.20-3.35) 87.6 (48.18-3.30)	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 0.276 , 0.315	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)	0.27 , 105.3	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 [i](#)

5.1 Standard geometry [i](#)

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	0.33	0/3967	0.56	1/5397 (0.0%)
1	B	0.37	0/3706	0.59	0/5051
1	C	0.44	0/3960	0.64	0/5396
All	All	0.38	0/11633	0.59	1/15844 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	367	THR	C-N-CA	-5.06	109.06	121.70

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	3868	0	3897	495	0
1	B	3612	0	3647	542	0
1	C	3862	0	3899	579	0
2	C	7	0	14	6	0
All	All	11349	0	11457	1576	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 69.

The worst 5 of 1576 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%)	7	33
1	B	470/566 (83%)	403 (86%)	61 (13%)	6 (1%)	12	42
1	C	504/566 (89%)	435 (86%)	60 (12%)	9 (2%)	8	35
All	All	1472/1698 (87%)	1264 (86%)	183 (12%)	25 (2%)	9	36

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%)	378 (96%)	18 (4%)	27	59
1	B	371/440 (84%)	356 (96%)	15 (4%)	31	62
1	C	396/440 (90%)	369 (93%)	27 (7%)	16	46
All	All	1163/1320 (88%)	1103 (95%)	60 (5%)	23	55

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	371	TRP
1	C	76	TRP
1	C	465	PHE
1	B	397	PHE
1	C	121	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	445	HIS
1	C	173	HIS
1	C	517	ASN
1	B	343	ASN
1	B	426	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 carbohydrates 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	0.78	0	8,8,8	0.26	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%)	0.37	44 (8%) 10 12	28, 107, 309, 534	0
1	B	476/566 (84%)	0.40	47 (9%) 7 8	14, 97, 292, 557	0
1	C	508/566 (89%)	-0.02	20 (3%) 39 41	7, 62, 194, 353	0
All	All	1488/1698 (87%)	0.25	111 (7%) 14 16	7, 86, 284, 557	0

The worst 5 of 111 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	296	SER	15.2
1	B	296	SER	8.9
1	A	297	GLY	8.3
1	A	299	GLY	7.1
1	C	299	GLY	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
2	CHT	C	2486	7/7	0.70	0.46	85,89,97,97	0

6.5 Other polymers ⓘ

There are no such residues in this entry.