



Full wwPDB X-ray Structure Validation Report ⓘ

Oct 2, 2023 – 10:14 pm BST

PDB ID : 8B6P
Title : X-ray structure of the haloalkane dehalogenase HaloTag7 circular permuted at positions 154-156 (cpHaloTag7_154-156)
Authors : Tarnawski, M.; Johnsson, K.; Hiblot, J.
Deposited on : 2022-09-27
Resolution : 1.10 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 4.02b-467
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

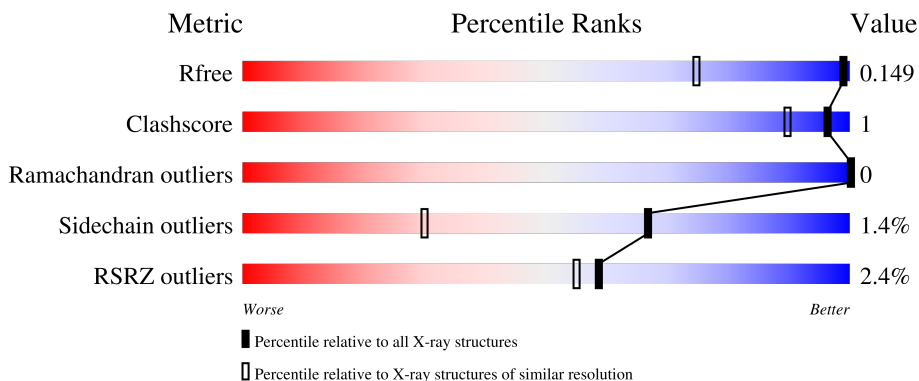
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.10 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	1619 (1.14-1.06)
Clashscore	141614	1671 (1.14-1.06)
Ramachandran outliers	138981	1615 (1.14-1.06)
Sidechain outliers	138945	1613 (1.14-1.06)
RSRZ outliers	127900	1588 (1.14-1.06)

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	309	 3% 92% 5%
1	B	309	 % 94%

2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 5418 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 Haloalkane dehalogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	294	2361	1535	397	419	10	0	2	0
1	B	296	2375	1544	400	421	10	0	3	0

There are 80 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP P0A3G3
A	0	GLY	-	expression tag	UNP P0A3G3
A	1	GLY	-	expression tag	UNP P0A3G3
A	6	LYS	GLU	conflict	UNP P0A3G3
A	13	VAL	ALA	conflict	UNP P0A3G3
A	18	THR	ALA	conflict	UNP P0A3G3
A	21	MET	LYS	conflict	UNP P0A3G3
A	22	GLY	CYS	conflict	UNP P0A3G3
A	41	ASN	LYS	conflict	UNP P0A3G3
A	70	GLU	ALA	conflict	UNP P0A3G3
A	73	ASP	ASN	conflict	UNP P0A3G3
A	103	LYS	GLU	conflict	UNP P0A3G3
A	110	ALA	THR	conflict	UNP P0A3G3
A	118	ASN	HIS	conflict	UNP P0A3G3
A	119	LEU	TYR	conflict	UNP P0A3G3
A	137	SER	PRO	conflict	UNP P0A3G3
A	138	THR	ALA	conflict	UNP P0A3G3
A	140	GLU	-	linker	UNP P0A3G3
A	141	ILE	-	linker	UNP P0A3G3
A	142	GLY	-	linker	UNP P0A3G3
A	143	GLY	-	linker	UNP P0A3G3
A	144	THR	-	linker	UNP P0A3G3
A	145	GLY	-	linker	UNP P0A3G3
A	146	GLY	-	linker	UNP P0A3G3
A	147	SER	-	linker	UNP P0A3G3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	148	GLY	-	linker	UNP P0A3G3
A	149	GLY	-	linker	UNP P0A3G3
A	150	THR	-	linker	UNP P0A3G3
A	151	GLY	-	linker	UNP P0A3G3
A	152	GLY	-	linker	UNP P0A3G3
A	153	SER	-	linker	UNP P0A3G3
A	154	GLY	-	linker	UNP P0A3G3
A	155	GLY	-	linker	UNP P0A3G3
A	156	SER	-	linker	UNP P0A3G3
A	200	VAL	LEU	conflict	UNP P0A3G3
A	211	THR	SER	conflict	UNP P0A3G3
A	231	GLY	ASP	conflict	UNP P0A3G3
A	240	PHE	TYR	conflict	UNP P0A3G3
A	241	MET	LEU	conflict	UNP P0A3G3
A	281	PHE	CYS	conflict	UNP P0A3G3
B	-1	GLY	-	expression tag	UNP P0A3G3
B	0	GLY	-	expression tag	UNP P0A3G3
B	1	GLY	-	expression tag	UNP P0A3G3
B	6	LYS	GLU	conflict	UNP P0A3G3
B	13	VAL	ALA	conflict	UNP P0A3G3
B	18	THR	ALA	conflict	UNP P0A3G3
B	21	MET	LYS	conflict	UNP P0A3G3
B	22	GLY	CYS	conflict	UNP P0A3G3
B	41	ASN	LYS	conflict	UNP P0A3G3
B	70	GLU	ALA	conflict	UNP P0A3G3
B	73	ASP	ASN	conflict	UNP P0A3G3
B	103	LYS	GLU	conflict	UNP P0A3G3
B	110	ALA	THR	conflict	UNP P0A3G3
B	118	ASN	HIS	conflict	UNP P0A3G3
B	119	LEU	TYR	conflict	UNP P0A3G3
B	137	SER	PRO	conflict	UNP P0A3G3
B	138	THR	ALA	conflict	UNP P0A3G3
B	140	GLU	-	linker	UNP P0A3G3
B	141	ILE	-	linker	UNP P0A3G3
B	142	GLY	-	linker	UNP P0A3G3
B	143	GLY	-	linker	UNP P0A3G3
B	144	THR	-	linker	UNP P0A3G3
B	145	GLY	-	linker	UNP P0A3G3
B	146	GLY	-	linker	UNP P0A3G3
B	147	SER	-	linker	UNP P0A3G3
B	148	GLY	-	linker	UNP P0A3G3
B	149	GLY	-	linker	UNP P0A3G3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	150	THR	-	linker	UNP P0A3G3
B	151	GLY	-	linker	UNP P0A3G3
B	152	GLY	-	linker	UNP P0A3G3
B	153	SER	-	linker	UNP P0A3G3
B	154	GLY	-	linker	UNP P0A3G3
B	155	GLY	-	linker	UNP P0A3G3
B	156	SER	-	linker	UNP P0A3G3
B	200	VAL	LEU	conflict	UNP P0A3G3
B	211	THR	SER	conflict	UNP P0A3G3
B	231	GLY	ASP	conflict	UNP P0A3G3
B	240	PHE	TYR	conflict	UNP P0A3G3
B	241	MET	LEU	conflict	UNP P0A3G3
B	281	PHE	CYS	conflict	UNP P0A3G3

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

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

- Molecule 3 is water.

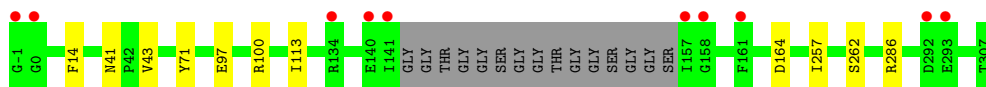
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	332	Total O 332 332	0	0
3	B	348	Total O 348 348	0	0

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.

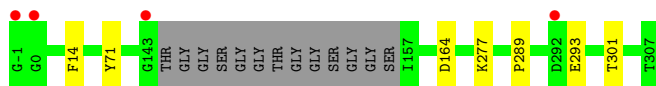
- Molecule 1: Haloalkane dehalogenase

Chain A: 



- Molecule 1: Haloalkane dehalogenase

Chain B: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	68.87Å 94.66Å 100.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.79 – 1.10 42.79 – 1.10	Depositor EDS
% Data completeness (in resolution range)	95.4 (42.79-1.10) 95.4 (42.79-1.10)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.79 (at 1.10Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.137 , 0.149 0.137 , 0.149	Depositor DCC
R_{free} test set	12592 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	8.6	Xtrriage
Anisotropy	0.278	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 44.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	5418	wwPDB-VP
Average B, all atoms (Å ²)	11.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 56.24 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.8167e-05.*

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: 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	0.47	0/2445	0.75	0/3338
1	B	0.46	0/2462	0.72	0/3359
All	All	0.47	0/4907	0.74	0/6697

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 [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	2361	0	2307	6	0
1	B	2375	0	2326	3	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	332	0	0	1	0
3	B	348	0	0	2	0
All	All	5418	0	4633	9	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 1.

All (9) 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:97:GLU:OE2	1:A:100:ARG:NH2	2.22	0.72
1:B:277[A]:LYS:HE3	3:B:510:HOH:O	1.92	0.69
1:A:286:ARG:HG2	3:A:569:HOH:O	1.97	0.65
1:B:289:PRO:HG2	1:B:293:GLU:OE2	1.97	0.64
1:A:97:GLU:CD	1:A:100:ARG:HH21	2.10	0.55
1:B:301:THR:HG21	3:B:625:HOH:O	2.09	0.52
1:A:97:GLU:CD	1:A:100:ARG:NH2	2.68	0.47
1:A:41:ASN:OD1	1:A:43:VAL:HG22	2.21	0.40
1:A:257:ILE:HB	1:A:262:SER:HA	2.04	0.40

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	292/309 (94%)	283 (97%)	9 (3%)	0	100	100
1	B	295/309 (96%)	287 (97%)	8 (3%)	0	100	100
All	All	587/618 (95%)	570 (97%)	17 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

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	253/256 (99%)	249 (98%)	4 (2%)	62	25
1	B	254/256 (99%)	251 (99%)	3 (1%)	71	36
All	All	507/512 (99%)	500 (99%)	7 (1%)	67	30

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	14	PHE
1	A	71	TYR
1	A	113	ILE
1	A	164	ASP
1	B	14	PHE
1	B	71	TYR
1	B	164	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 [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	294/309 (95%)	-0.15	10 (3%) 45 42	5, 10, 18, 39	0
1	B	296/309 (95%)	-0.20	4 (1%) 75 71	6, 10, 19, 30	0
All	All	590/618 (95%)	-0.18	14 (2%) 59 55	5, 10, 19, 39	0

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	0	GLY	4.8
1	B	143	GLY	4.7
1	B	-1	GLY	4.3
1	B	0	GLY	3.8
1	A	157	ILE	3.6
1	A	140	GLU	3.5
1	A	-1	GLY	3.3
1	A	292	ASP	3.1
1	A	134	ARG	3.0
1	A	161	PHE	2.7
1	A	158	GLY	2.4
1	B	292	ASP	2.3
1	A	293	GLU	2.2
1	A	141	ILE	2.0

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 monosaccharides 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(Å ²)	Q<0.9
2	CL	A	400	1/1	0.99	0.03	8,8,8,8	0
2	CL	B	400	1/1	1.00	0.02	9,9,9,9	0

6.5 Other polymers [i](#)

There are no such residues in this entry.