Full wwPDB X-ray Structure Validation Report ⓘ

Oct 2, 2023 – 10:09 pm BST

PDB ID	:	8B6Q
Title	:	X-ray structure of the haloalkane dehalogenase HaloTag7 with an insertion of Calmodulin-M13 fusion at position 154-156 that mimic the structure of CaProLa, an calcium gated protein labeling technology
Authors	:	Tarnawski, M.; Johnsson, K.; Hiblot, J.
Deposited on	:	2022-09-27
Resolution	:	2.60 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:

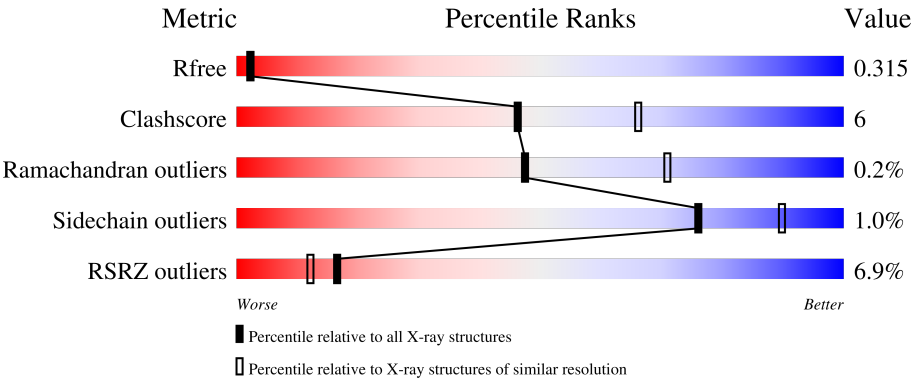
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 <=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	472	7% 82% 17%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3699 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,Calmodulin-1,Haloalkane dehalogenase,Calmodulin-1,M13 peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	466	3689	2346	621	704	18	0	0	0

There are 39 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P0A3G3
A	45	VAL	LEU	engineered mutation	UNP P0A3G3
A	56	THR	SER	engineered mutation	UNP P0A3G3
A	76	GLY	ASP	engineered mutation	UNP P0A3G3
A	85	PHE	TYR	engineered mutation	UNP P0A3G3
A	86	MET	LEU	engineered mutation	UNP P0A3G3
A	126	PHE	CYS	engineered mutation	UNP P0A3G3
A	153	THR	-	linker	UNP P0A3G3
A	154	GLY	-	linker	UNP P0A3G3
A	155	SER	-	linker	UNP P0A3G3
A	214	ASP	ASN	engineered mutation	UNP P0DP23
A	297	VAL	GLN	engineered mutation	UNP P0DP23
A	302	GLY	-	linker	UNP P0DP23
A	303	GLY	-	linker	UNP P0DP23
A	304	THR	-	linker	UNP P0DP23
A	305	GLY	-	linker	UNP P0DP23
A	306	GLY	-	linker	UNP P0DP23
A	329	LEU	-	linker	UNP P0DP23
A	330	GLU	-	linker	UNP P0DP23
A	331	GLY	-	linker	UNP P0DP23
A	332	GLY	-	linker	UNP P0DP23
A	333	SER	-	linker	UNP P0DP23
A	337	LYS	GLU	engineered mutation	UNP P0A3G3
A	344	VAL	ALA	engineered mutation	UNP P0A3G3
A	349	THR	ALA	engineered mutation	UNP P0A3G3
A	352	MET	LYS	engineered mutation	UNP P0A3G3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	353	GLY	CYS	engineered mutation	UNP P0A3G3
A	372	ASN	LYS	engineered mutation	UNP P0A3G3
A	401	GLU	ALA	engineered mutation	UNP P0A3G3
A	404	ASP	ASN	engineered mutation	UNP P0A3G3
A	434	LYS	GLU	engineered mutation	UNP P0A3G3
A	441	ALA	THR	engineered mutation	UNP P0A3G3
A	449	ASN	HIS	engineered mutation	UNP P0A3G3
A	450	LEU	TYR	engineered mutation	UNP P0A3G3
A	468	SER	-	expression tag	UNP P0A3G3
A	469	THR	-	expression tag	UNP P0A3G3
A	470	LEU	-	expression tag	UNP P0A3G3
A	471	GLU	-	expression tag	UNP P0A3G3
A	472	ILE	-	expression tag	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

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	6	Total Ca 6 6	0	0

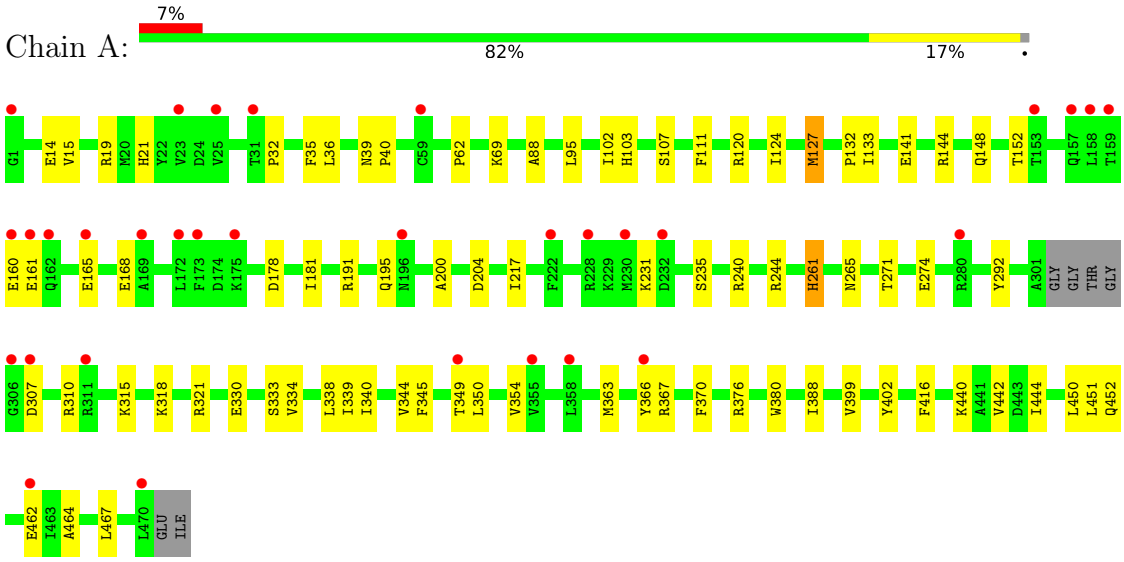
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	3	Total O 3 3	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,Calmodulin-1,Haloalkane dehalogenase,Calmodulin-1,M13 peptide



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	74.53Å 74.53Å 177.84Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.39 – 2.60 46.39 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.7 (46.39-2.60) 99.7 (46.39-2.60)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.66 (at 2.61Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.281 , 0.318 0.278 , 0.315	Depositor DCC
R_{free} test set	806 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	65.1	Xtriage
Anisotropy	0.121	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 45.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	3699	wwPDB-VP
Average B, all atoms (Å^2)	89.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.25% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: CL, CA

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	0/3781	0.47	0/5131

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	3689	0	3572	46	0
2	A	1	0	0	0	0
3	A	6	0	0	0	0
4	A	3	0	0	0	0
All	All	3699	0	3572	46	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 6.

All (46) 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:133:ILE:HD11	1:A:388:ILE:HG23	1.73	0.71
1:A:36:LEU:HB2	1:A:102:ILE:HG22	1.77	0.66
1:A:464:ALA:HA	1:A:467:LEU:HD12	1.82	0.60
1:A:132:PRO:HG3	1:A:399:VAL:HG23	1.85	0.58
1:A:168:GLU:HB3	1:A:315:LYS:HG3	1.86	0.58
1:A:161:GLU:O	1:A:165:GLU:N	2.38	0.56
1:A:32:PRO:HG3	1:A:95:LEU:HD22	1.86	0.56
1:A:178:ASP:N	1:A:178:ASP:OD1	2.40	0.53
1:A:340:ILE:O	1:A:376:ARG:NH1	2.39	0.52
1:A:318:LYS:HD3	1:A:321:ARG:HH12	1.76	0.51
1:A:399:VAL:HA	1:A:402:TYR:CE2	2.46	0.50
1:A:442:VAL:HG11	1:A:462:GLU:HG3	1.93	0.49
1:A:35:PHE:O	1:A:62:PRO:HD2	2.14	0.48
1:A:318:LYS:HD3	1:A:321:ARG:NH1	2.29	0.47
1:A:191:ARG:HA	1:A:195:GLN:O	2.14	0.47
1:A:102:ILE:HD12	1:A:107:SER:HA	1.97	0.46
1:A:339:ILE:HD12	1:A:380:TRP:HD1	1.81	0.45
1:A:307:ASP:HA	1:A:310:ARG:HB3	1.97	0.45
1:A:370:PHE:HB3	1:A:376:ARG:HG2	1.99	0.45
1:A:240:ARG:O	1:A:244:ARG:HG2	2.17	0.45
1:A:141:GLU:HA	1:A:144:ARG:HB2	1.99	0.45
1:A:444:ILE:HD11	1:A:451:LEU:HD22	1.99	0.44
1:A:111:PHE:CE1	1:A:124:ILE:HG21	2.53	0.44
1:A:127:MET:HB3	1:A:416:PHE:HB2	2.00	0.43
1:A:14:GLU:HA	1:A:19:ARG:HA	1.99	0.43
1:A:261:HIS:O	1:A:265:ASN:N	2.48	0.43
1:A:339:ILE:HD12	1:A:380:TRP:CD1	2.54	0.43
1:A:334:VAL:O	1:A:338:LEU:HG	2.19	0.43
1:A:95:LEU:O	1:A:120:ARG:NH2	2.52	0.42
1:A:338:LEU:O	1:A:344:VAL:HG13	2.19	0.42
1:A:148:GLN:O	1:A:152:THR:N	2.51	0.42
1:A:440:LYS:HE3	1:A:440:LYS:HB2	1.84	0.42
1:A:15:VAL:HG13	1:A:88:ALA:HB3	2.01	0.42
1:A:350:LEU:HD21	1:A:366:TYR:CD2	2.55	0.42
1:A:231:LYS:O	1:A:235:SER:OG	2.29	0.42
1:A:345:PHE:HA	1:A:349:THR:HB	2.02	0.41
1:A:354:VAL:HG22	1:A:450:LEU:HB2	2.02	0.41
1:A:102:ILE:HB	1:A:107:SER:HA	2.01	0.41
1:A:181:ILE:HB	1:A:217:ILE:HB	2.01	0.41
1:A:200:ALA:O	1:A:204:ASP:HB2	2.20	0.41
1:A:240:ARG:HG2	1:A:292:TYR:CE1	2.55	0.41
1:A:363:MET:O	1:A:367:ARG:HG3	2.19	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:103:HIS:CG	1:A:452:GLN:HE22	2.38	0.41
1:A:271:THR:OG1	1:A:274:GLU:HG2	2.21	0.41
1:A:39:ASN:HA	1:A:40:PRO:HA	1.90	0.41
1:A:21:HIS:CG	1:A:69:LYS:HB2	2.56	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	462/472 (98%)	430 (93%)	31 (7%)	1 (0%)	47 71

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	160	GLU

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	396/399 (99%)	392 (99%)	4 (1%)	76 90

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	127	MET
1	A	261	HIS
1	A	330	GLU
1	A	333	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	112	HIS
1	A	342	GLN

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 7 ligands modelled in this entry, 7 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 (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

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	466/472 (98%)	0.38	32 (6%)	16	12	37, 87, 141, 194	0

All (32) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	230	MET	9.8
1	A	228	ARG	5.8
1	A	172	LEU	4.4
1	A	153	THR	3.9
1	A	355	VAL	3.6
1	A	158	LEU	3.6
1	A	157	GLN	3.6
1	A	232	ASP	3.4
1	A	173	PHE	3.4
1	A	165	GLU	3.2
1	A	169	ALA	3.1
1	A	162	GLN	3.1
1	A	175	LYS	3.1
1	A	59	CYS	2.8
1	A	306	GLY	2.7
1	A	358	LEU	2.6
1	A	307	ASP	2.6
1	A	23	VAL	2.6
1	A	366	TYR	2.5
1	A	161	GLU	2.5
1	A	222	PHE	2.5
1	A	462	GLU	2.3
1	A	160	GLU	2.3
1	A	196	ASN	2.2
1	A	349	THR	2.2
1	A	25	VAL	2.2
1	A	280	ARG	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	470	LEU	2.1
1	A	1	GLY	2.1
1	A	311	ARG	2.1
1	A	159	THR	2.0
1	A	31	THR	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, 95^{th} 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
3	CA	A	505	1/1	0.45	0.16	132,132,132,132	0
3	CA	A	506	1/1	0.80	0.21	119,119,119,119	0
3	CA	A	507	1/1	0.84	0.14	123,123,123,123	0
3	CA	A	504	1/1	0.91	0.22	94,94,94,94	0
3	CA	A	503	1/1	0.95	0.16	61,61,61,61	0
3	CA	A	502	1/1	0.96	0.14	62,62,62,62	0
2	CL	A	501	1/1	0.98	0.15	64,64,64,64	0

6.5 Other polymers (i)

There are no such residues in this entry.