



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 23, 2024 – 09:44 AM EDT

PDB ID : 6I9U  
Title : Crystal structure of the halohydrin dehalogenase HheG T123W mutant  
Authors : Klunenmann, T.; Blankenfeldt, W.; Schallmeyer, A.  
Deposited on : 2018-11-26  
Resolution : 2.40 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](mailto: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)	:	1.20.1
EDS	:	2.37.1
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)	:	2.37.1

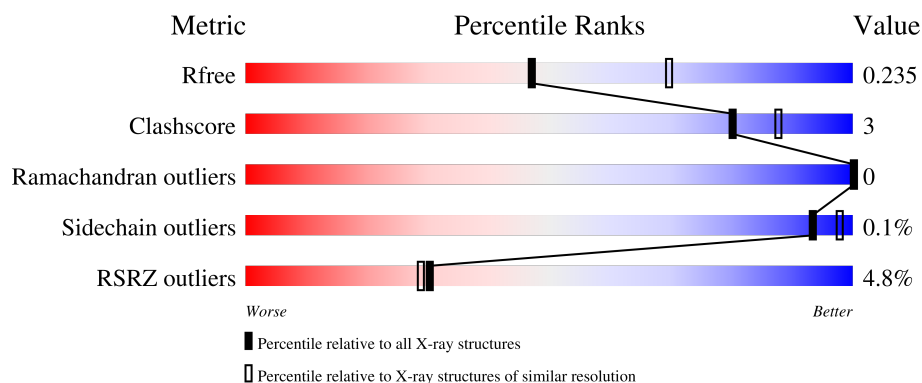
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

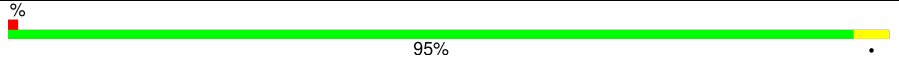
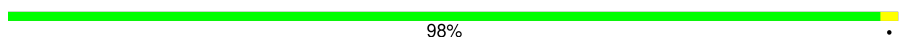
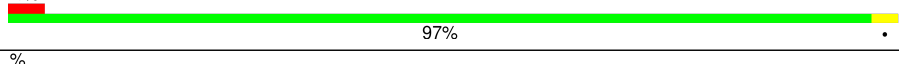
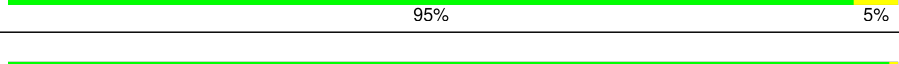
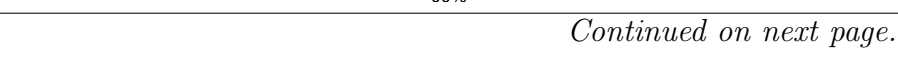
The reported resolution of this entry is 2.40 Å.

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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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  $\geq 3$ , 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	258	
1	B	258	
1	C	258	
1	D	258	
1	E	258	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	258	<div><div></div><div>96%</div><div></div></div>
1	G	258	<div><div>7%</div><div></div><div>90%</div><div></div><div>9%</div></div>
1	H	258	<div><div>12%</div><div></div><div>90%</div><div></div><div>10%</div></div>
1	I	258	<div><div>9%</div><div></div><div>90%</div><div></div><div>10%</div></div>
1	J	258	<div><div>13%</div><div></div><div>93%</div><div></div><div>7%</div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 38098 atoms, of which 18633 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 Putative oxidoreductase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	257	Total	C	H	N	O	S	0	4	0
			3778	1192	1870	337	362	17			
1	B	258	Total	C	H	N	O	S	0	2	0
			3794	1195	1879	340	365	15			
1	C	257	Total	C	H	N	O	S	0	4	0
			3805	1197	1886	344	364	14			
1	D	257	Total	C	H	N	O	S	0	2	0
			3769	1188	1864	338	365	14			
1	E	257	Total	C	H	N	O	S	0	3	0
			3810	1198	1891	343	364	14			
1	F	258	Total	C	H	N	O	S	0	2	0
			3826	1201	1901	344	365	15			
1	G	257	Total	C	H	N	O	S	0	2	0
			3728	1181	1836	333	365	13			
1	H	257	Total	C	H	N	O	S	0	1	0
			3731	1180	1842	337	359	13			
1	I	257	Total	C	H	N	O	S	0	0	0
			3698	1172	1823	335	356	12			
1	J	257	Total	C	H	N	O	S	0	0	0
			3731	1180	1841	335	363	12			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	123	TRP	THR	engineered mutation	UNP M5A5Y8
B	123	TRP	THR	engineered mutation	UNP M5A5Y8
C	123	TRP	THR	engineered mutation	UNP M5A5Y8
D	123	TRP	THR	engineered mutation	UNP M5A5Y8
E	123	TRP	THR	engineered mutation	UNP M5A5Y8
F	123	TRP	THR	engineered mutation	UNP M5A5Y8
G	123	TRP	THR	engineered mutation	UNP M5A5Y8
H	123	TRP	THR	engineered mutation	UNP M5A5Y8
I	123	TRP	THR	engineered mutation	UNP M5A5Y8

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
J	123	TRP	THR	engineered mutation	UNP M5A5Y8

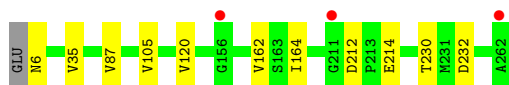
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	56	Total O 56 56	0	0
2	B	66	Total O 66 66	0	0
2	C	43	Total O 43 43	0	0
2	D	55	Total O 55 55	0	0
2	E	82	Total O 82 82	0	0
2	F	83	Total O 83 83	0	0
2	G	15	Total O 15 15	0	0
2	H	4	Total O 4 4	0	0
2	I	5	Total O 5 5	0	0
2	J	19	Total O 19 19	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.

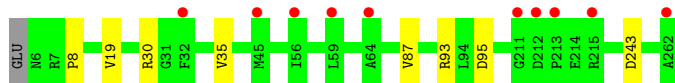
- Molecule 1: Putative oxidoreductase



- Molecule 1: Putative oxidoreductase



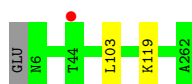
- Molecule 1: Putative oxidoreductase



- Molecule 1: Putative oxidoreductase



- Molecule 1: Putative oxidoreductase




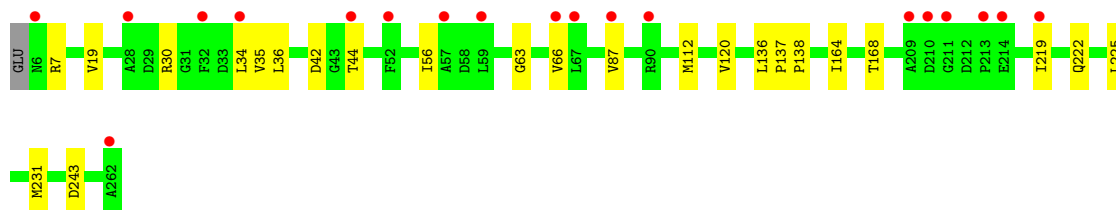
- Molecule 1: Putative oxidoreductase

Chain F:  96%



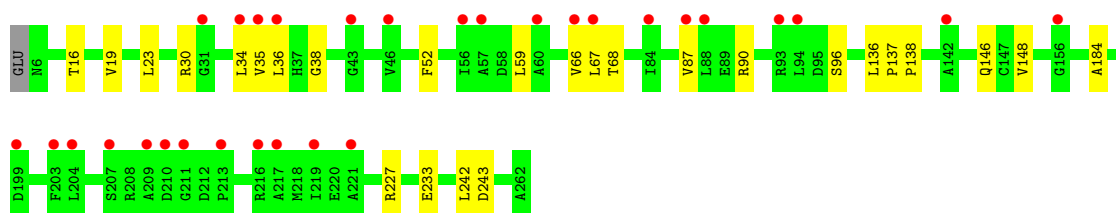
● Molecule 1: Putative oxidoreductase

Chain G:  7% 90% 9%




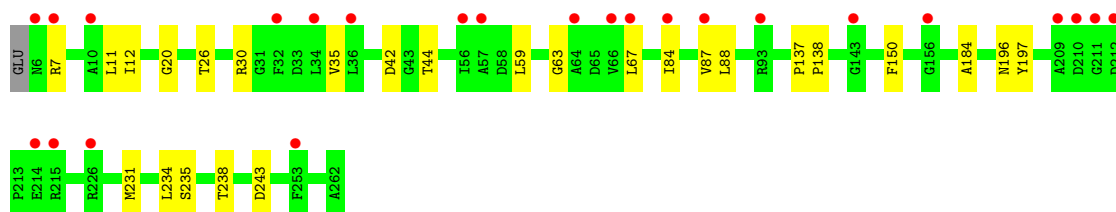
● Molecule 1: Putative oxidoreductase

Chain H:  12% 90% 10%



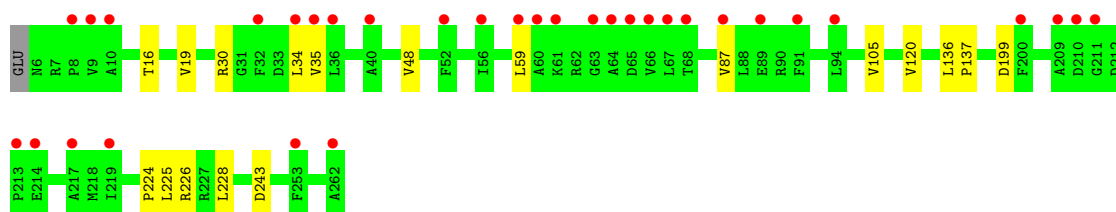
● Molecule 1: Putative oxidoreductase

Chain I:  9% 90% 10%



● Molecule 1: Putative oxidoreductase

Chain J:  13% 93% 7%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	197.25Å 197.25Å 197.72Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.34 – 2.40 49.34 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.34-2.40) 92.8 (49.34-2.40)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.29 (at 2.39Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.208 , 0.229 0.215 , 0.235	Depositor DCC
$R_{free}$ test set	8507 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	51.4	Xtriage
Anisotropy	0.238	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 60.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.022 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	38098	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.06% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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](#)

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	0.26	0/1953	0.45	0/2646
1	B	0.26	0/1954	0.44	0/2646
1	C	0.25	0/1964	0.44	0/2660
1	D	0.26	0/1944	0.45	0/2635
1	E	0.26	0/1961	0.45	0/2654
1	F	0.26	0/1964	0.45	0/2657
1	G	0.25	0/1931	0.43	0/2620
1	H	0.25	0/1925	0.44	0/2610
1	I	0.25	0/1908	0.44	0/2589
1	J	0.25	0/1923	0.43	0/2607
All	All	0.25	0/19427	0.44	0/26324

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	1908	1870	1869	8	0
1	B	1915	1879	1879	3	0
1	C	1919	1886	1885	5	0
1	D	1905	1864	1864	8	0
1	E	1919	1891	1894	1	0
1	F	1925	1901	1901	7	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	1892	1836	1836	16	0
1	H	1889	1842	1842	23	0
1	I	1875	1823	1822	21	0
1	J	1890	1841	1841	12	0
2	A	56	0	0	1	0
2	B	66	0	0	0	0
2	C	43	0	0	0	0
2	D	55	0	0	0	0
2	E	82	0	0	0	0
2	F	83	0	0	0	0
2	G	15	0	0	0	0
2	H	4	0	0	0	0
2	I	5	0	0	0	0
2	J	19	0	0	0	0
All	All	19465	18633	18633	102	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 3.

The worst 5 of 102 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:G:7:ARG:NH1	1:G:63:GLY:O	2.10	0.83
1:F:7:ARG:NH1	1:F:63:GLY:O	2.21	0.72
1:J:30:ARG:NH1	1:J:243:ASP:OD2	2.24	0.71
1:I:84:ILE:O	1:I:88:LEU:HD12	1.90	0.71
1:I:84:ILE:HG23	1:I:88:LEU:HD11	1.74	0.70

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	259/258 (100%)	251 (97%)	8 (3%)	0	100	100
1	B	258/258 (100%)	249 (96%)	9 (4%)	0	100	100
1	C	259/258 (100%)	251 (97%)	8 (3%)	0	100	100
1	D	257/258 (100%)	249 (97%)	8 (3%)	0	100	100
1	E	258/258 (100%)	250 (97%)	8 (3%)	0	100	100
1	F	258/258 (100%)	250 (97%)	8 (3%)	0	100	100
1	G	257/258 (100%)	247 (96%)	10 (4%)	0	100	100
1	H	256/258 (99%)	242 (94%)	14 (6%)	0	100	100
1	I	255/258 (99%)	239 (94%)	16 (6%)	0	100	100
1	J	255/258 (99%)	245 (96%)	10 (4%)	0	100	100
All	All	2572/2580 (100%)	2473 (96%)	99 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	195/197 (99%)	195 (100%)	0	100	100
1	B	196/197 (100%)	196 (100%)	0	100	100
1	C	196/197 (100%)	196 (100%)	0	100	100
1	D	195/197 (99%)	195 (100%)	0	100	100
1	E	197/197 (100%)	197 (100%)	0	100	100
1	F	198/197 (100%)	198 (100%)	0	100	100
1	G	192/197 (98%)	192 (100%)	0	100	100
1	H	190/197 (96%)	190 (100%)	0	100	100
1	I	187/197 (95%)	186 (100%)	1 (0%)	88	95
1	J	191/197 (97%)	191 (100%)	0	100	100
All	All	1937/1970 (98%)	1936 (100%)	1 (0%)	93	98

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	59	LEU

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](#)

There are no ligands in this entry.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	257/258 (99%)	0.00	3 (1%) 79 77	49, 67, 114, 186	0
1	B	258/258 (100%)	-0.09	1 (0%) 92 91	47, 62, 94, 150	0
1	C	257/258 (99%)	0.15	10 (3%) 39 38	50, 77, 120, 175	0
1	D	257/258 (99%)	0.04	3 (1%) 79 77	49, 74, 112, 144	0
1	E	257/258 (99%)	-0.19	1 (0%) 92 91	45, 57, 91, 116	0
1	F	258/258 (100%)	-0.23	0 100 100	46, 58, 94, 120	0
1	G	257/258 (99%)	0.58	19 (7%) 14 13	63, 99, 146, 188	0
1	H	257/258 (99%)	0.53	30 (11%) 4 4	74, 106, 150, 181	0
1	I	257/258 (99%)	0.45	24 (9%) 8 8	72, 106, 166, 188	0
1	J	257/258 (99%)	0.65	33 (12%) 3 3	61, 100, 158, 185	0
All	All	2572/2580 (99%)	0.19	124 (4%) 30 29	45, 78, 140, 188	0

The worst 5 of 124 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	66	VAL	9.6
1	J	34	LEU	6.4
1	C	211	GLY	6.1
1	I	211	GLY	5.7
1	G	34	LEU	5.5

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.