WORLDWIDE PDB

PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

May 14, 2020 – 10:24 am BST

PDB ID	:	4Y9S
Title	:	structure of an H300N mutant of potato epoxide hydrolase, StEH1
Authors	:	Naworyta, A.; Mowbray, S.L.; Widersten, M.
Deposited on	:	2015-02-17
Resolution	:	2.00 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

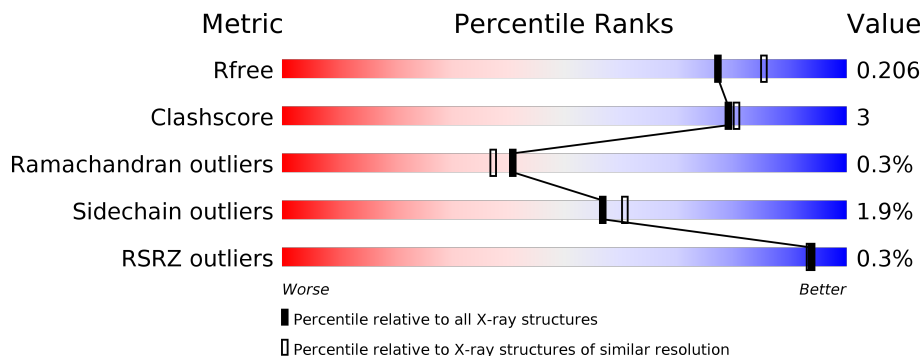|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.00 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	Percentile Ranks	Value
Rfree		0.206
Clashscore		3
Ramachandran outliers		0.3%
Sidechain outliers		1.9%
RSRZ outliers		0.3%

Worse ← → Better

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 <=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	328	88% 9% ··
1	B	328	% 90% 6% ··

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5593 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 Epoxide hydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	320	Total	C	N	O	S	0	0	0
			2560	1679	424	452	5			
1	B	319	Total	C	N	O	S	0	0	0
			2551	1673	422	451	5			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	THR	-	expression tag	UNP Q41415
A	-5	SER	-	expression tag	UNP Q41415
A	-4	HIS	-	expression tag	UNP Q41415
A	-3	HIS	-	expression tag	UNP Q41415
A	-2	HIS	-	expression tag	UNP Q41415
A	-1	HIS	-	expression tag	UNP Q41415
A	0	HIS	-	expression tag	UNP Q41415
A	1	MET	-	expression tag	UNP Q41415
A	2	LYS	-	expression tag	UNP Q41415
A	300	ASN	HIS	engineered mutation	UNP Q41415
B	-6	THR	-	expression tag	UNP Q41415
B	-5	SER	-	expression tag	UNP Q41415
B	-4	HIS	-	expression tag	UNP Q41415
B	-3	HIS	-	expression tag	UNP Q41415
B	-2	HIS	-	expression tag	UNP Q41415
B	-1	HIS	-	expression tag	UNP Q41415
B	0	HIS	-	expression tag	UNP Q41415
B	1	MET	-	expression tag	UNP Q41415
B	2	LYS	-	expression tag	UNP Q41415
B	300	ASN	HIS	engineered mutation	UNP Q41415

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	268	Total 268	O 268	0	0
2	B	214	Total 214	O 214	0	0

3 Residue-property plots (i)

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: Epoxide hydrolase
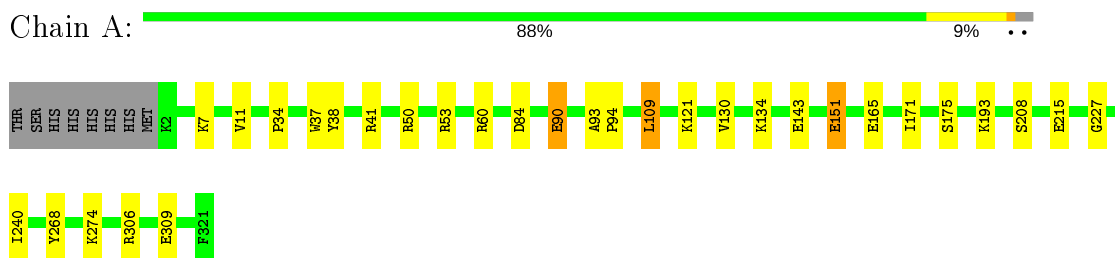
• Molecule 1: Epoxide hydrolase
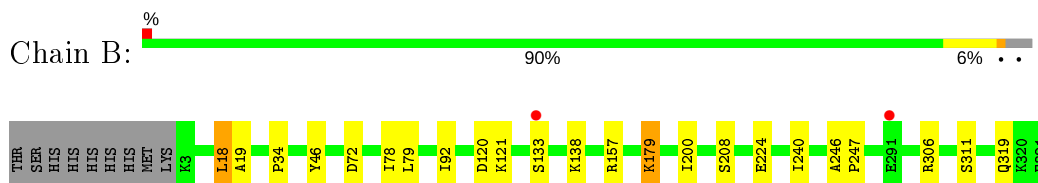
4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	56.02Å 96.04Å 121.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	51.36 – 2.00 51.36 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (51.36-2.00) 100.0 (51.36-2.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.61 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.147 , 0.199 0.160 , 0.206	Depositor DCC
R_{free} test set	2276 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	18.7	Xtriage
Anisotropy	0.287	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 51.5	EDS
L-test for twinning[2]	$< |L| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5593	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.97% 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 (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	1.01	5/2640 (0.2%)	0.94	5/3591 (0.1%)
1	B	0.91	0/2631	0.91	4/3580 (0.1%)
All	All	0.96	5/5271 (0.1%)	0.93	9/7171 (0.1%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	151	GLU	CD-OE1	6.73	1.33	1.25
1	A	227	GLY	CA-C	6.12	1.61	1.51
1	A	151	GLU	CG-CD	5.90	1.60	1.51
1	A	143	GLU	CD-OE1	5.44	1.31	1.25
1	A	165	GLU	CD-OE2	5.29	1.31	1.25

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	53	ARG	NE-CZ-NH1	8.08	124.34	120.30
1	A	53	ARG	NE-CZ-NH2	-7.76	116.42	120.30
1	B	157	ARG	NE-CZ-NH2	-6.52	117.04	120.30
1	A	60	ARG	NE-CZ-NH2	-6.41	117.10	120.30
1	B	120	ASP	CB-CG-OD1	6.36	124.03	118.30
1	B	306	ARG	NE-CZ-NH2	-5.84	117.38	120.30
1	A	50	ARG	NE-CZ-NH1	5.79	123.19	120.30
1	B	72	ASP	CB-CG-OD2	-5.74	113.13	118.30
1	A	84	ASP	CB-CG-OD1	5.16	122.94	118.30

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	2560	0	2533	13	0
1	B	2551	0	2520	15	0
2	A	268	0	0	2	1
2	B	214	0	0	5	1
All	All	5593	0	5053	28	1

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.

All (28) 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:306:ARG:NH2	1:A:309:GLU:OE2	2.18	0.77
1:B:224:GLU:OE1	2:B:514:HOH:O	2.03	0.76
1:B:208:SER:HB3	2:B:543:HOH:O	1.90	0.71
1:B:121:LYS:NZ	2:B:529:HOH:O	2.12	0.57
1:B:179:LYS:HE3	1:B:200:ILE:O	2.05	0.57
1:B:121:LYS:HE2	2:B:417:HOH:O	2.05	0.56
1:B:18:LEU:HD23	1:B:18:LEU:C	2.28	0.53
1:A:11:VAL:HG23	1:A:11:VAL:O	2.10	0.51
1:A:41:ARG:HH21	1:A:215:GLU:CD	2.12	0.51
1:B:319:GLN:HA	1:B:319:GLN:OE1	2.11	0.50
1:B:179:LYS:HE2	1:B:200:ILE:HB	1.96	0.47
1:A:208:SER:HB3	2:A:546:HOH:O	2.15	0.46
1:A:151:GLU:CD	1:A:151:GLU:N	2.69	0.46
1:B:18:LEU:CD1	1:B:92:ILE:HG21	2.47	0.45
1:B:246:ALA:N	1:B:247:PRO:CD	2.80	0.44
1:B:121:LYS:CE	2:B:417:HOH:O	2.63	0.44
1:A:93:ALA:N	1:A:94:PRO:CD	2.80	0.44
1:B:18:LEU:HD23	1:B:19:ALA:N	2.33	0.43
1:B:46:TYR:CD2	1:B:311:SER:HB3	2.53	0.43
1:A:151:GLU:H	1:A:151:GLU:CD	2.22	0.43
1:A:240:ILE:HG21	1:A:240:ILE:HD13	1.84	0.42
1:A:109:LEU:HD13	1:A:130:VAL:HG11	2.02	0.42
1:A:37:TRP:CG	1:A:38:TYR:N	2.87	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:268:TYR:CE2	1:A:274:LYS:HE2	2.55	0.41
1:A:171:ILE:HG22	1:A:175:SER:HB2	2.02	0.40
1:A:90:GLU:CG	2:A:414:HOH:O	2.68	0.40
1:B:240:ILE:HA	1:B:240:ILE:HD12	1.92	0.40
1:B:78:ILE:HG23	1:B:79:LEU:HD22	2.03	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:409:HOH:O	2:B:426:HOH:O[2_674]	2.15	0.05

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	318/328 (97%)	311 (98%)	6 (2%)	1 (0%)	41	37
1	B	317/328 (97%)	310 (98%)	6 (2%)	1 (0%)	41	37
All	All	635/656 (97%)	621 (98%)	12 (2%)	2 (0%)	41	37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	34	PRO
1	A	34	PRO

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	267/275 (97%)	261 (98%)	6 (2%)	52	55
1	B	266/275 (97%)	262 (98%)	4 (2%)	65	69
All	All	533/550 (97%)	523 (98%)	10 (2%)	57	61

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	7	LYS
1	A	90	GLU
1	A	109	LEU
1	A	121	LYS
1	A	134	LYS
1	A	193	LYS
1	B	18	LEU
1	B	133	SER
1	B	138	LYS
1	B	179	LYS

Some 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 carbohydrates 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, 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	320/328 (97%)	-0.61	0	100	100	9, 15, 28, 48	0
1	B	319/328 (97%)	-0.50	2 (0%)	89	88	9, 19, 36, 52	0
All	All	639/656 (97%)	-0.55	2 (0%)	94	93	9, 17, 34, 52	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	291	GLU	2.5
1	B	133	SER	2.2

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.