



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 07:53 am GMT

PDB ID : 3GFH
Title : Crystal structure of EUTL shell protein of the bacterial ethanolamine microcompartment
Authors : Sagermann, M.; Nikolakakis, K.; Ohtaki, A.
Deposited on : 2009-02-26
Resolution : 2.20 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

<http://wwpdb.org/validation/2016/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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

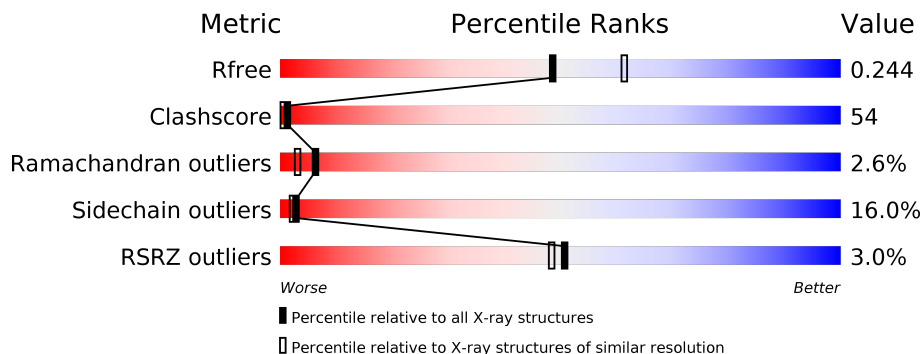
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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}	100719	4002 (2.20-2.20)
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)
RSRZ outliers	101464	4033 (2.20-2.20)

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. 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	225	
1	B	225	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3291 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 Ethanolamine utilization protein eutL.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	214	Total	C	N	O	S	0	0	0
			1559	983	263	306	7			
1	B	214	Total	C	N	O	S	0	0	0
			1559	983	263	306	7			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	220	HIS	-	EXPRESSION TAG	UNP P76541
A	221	HIS	-	EXPRESSION TAG	UNP P76541
A	222	HIS	-	EXPRESSION TAG	UNP P76541
A	223	HIS	-	EXPRESSION TAG	UNP P76541
A	224	HIS	-	EXPRESSION TAG	UNP P76541
A	225	HIS	-	EXPRESSION TAG	UNP P76541
B	220	HIS	-	EXPRESSION TAG	UNP P76541
B	221	HIS	-	EXPRESSION TAG	UNP P76541
B	222	HIS	-	EXPRESSION TAG	UNP P76541
B	223	HIS	-	EXPRESSION TAG	UNP P76541
B	224	HIS	-	EXPRESSION TAG	UNP P76541
B	225	HIS	-	EXPRESSION TAG	UNP P76541

- Molecule 2 is MERCURY (II) ION (three-letter code: HG) (formula: Hg).

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

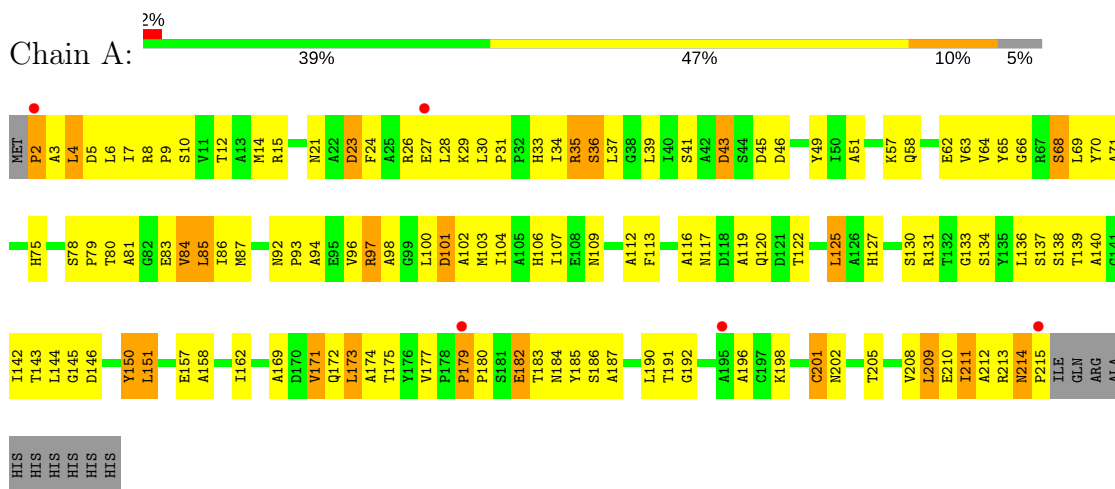
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	85	Total 85	O 85	0	0
3	B	86	Total 86	O 86	0	0

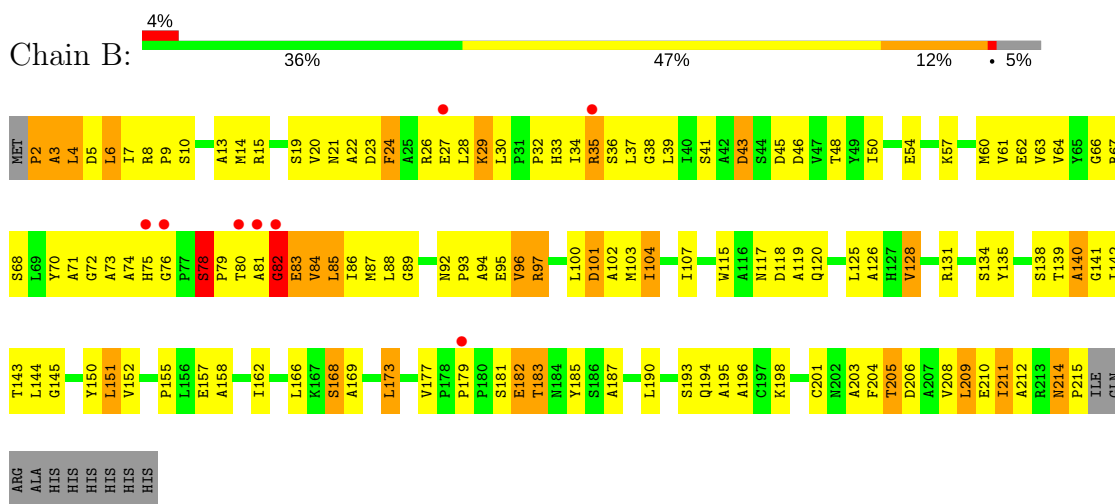
3 Residue-property plots

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: Ethanolamine utilization protein eutL



• Molecule 1: Ethanolamine utilization protein eutL



4 Data and refinement statistics

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, α , β , γ	67.38Å 67.38Å 79.66Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.65 – 2.20 19.64 – 1.96	Depositor EDS
% Data completeness (in resolution range)	100.0 (19.65-2.20) 96.9 (19.64-1.96)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.26 (at 1.96Å)	Xtriage
Refinement program	REFMAC 5.5.0066	Depositor
R, R_{free}	0.223 , 0.278 0.238 , 0.244	Depositor DCC
R_{free} test set	1475 reflections (7.85%)	DCC
Wilson B-factor (Å ²)	16.8	Xtriage
Anisotropy	0.631	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 34.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.048 for -h,-k,l 0.457 for h,-h-k,-l 0.054 for -k,-h,-l	Xtriage
Reported twinning fraction	0.575 for H, K, L 0.425 for K, H, -L	Depositor
Outliers	0 of 51500 reflections	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	3291	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 7.04% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HG

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	0.99	2/1592 (0.1%)	0.96	2/2180 (0.1%)
1	B	0.95	0/1592	1.07	6/2180 (0.3%)
All	All	0.97	2/3184 (0.1%)	1.02	8/4360 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	201	CYS	CB-SG	17.80	2.12	1.82
1	A	49	TYR	CD1-CE1	-5.49	1.31	1.39

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	78	SER	C-N-CD	-10.40	97.73	120.60
1	B	118	ASP	CB-CG-OD2	-6.37	112.57	118.30
1	A	23	ASP	CB-CG-OD1	-6.03	112.88	118.30
1	B	118	ASP	CB-CG-OD1	5.82	123.54	118.30
1	B	209	LEU	CB-CG-CD2	5.11	119.69	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	140	ALA	Peptide
1	B	78	SER	Peptide
1	B	82	GLY	Peptide

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	1559	0	1537	173	8
1	B	1559	0	1536	160	9
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	85	0	0	22	1
3	B	86	0	0	14	1
All	All	3291	0	3073	332	16

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 54.

The worst 5 of 332 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:201:CYS:SG	1:A:201:CYS:CB	2.12	1.37
1:A:2:PRO:HB3	1:A:5:ASP:OD1	1.26	1.30
1:A:205:THR:O	1:A:209:LEU:HD23	1.38	1.22
1:A:117:ASN:ND2	1:A:119:ALA:HB3	1.55	1.19
1:B:100:LEU:O	1:B:104:ILE:HD13	1.45	1.13

The worst 5 of 16 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)
1:A:27:GLU:OE1	1:A:173:LEU:CD1[2_665]	1.09	1.11
1:B:27:GLU:OE1	1:B:173:LEU:CD1[3_665]	1.38	0.82
1:B:27:GLU:CD	1:B:173:LEU:CD1[3_665]	1.61	0.59
1:B:15:ARG:NH1	1:B:211:ILE:CG1[3_665]	1.61	0.59

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:15:ARG:NH1	1:A:211:ILE:CG1[2_665]	1.65	0.55

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	212/225 (94%)	186 (88%)	23 (11%)	3 (1%)	13	10
1	B	212/225 (94%)	188 (89%)	16 (8%)	8 (4%)	4	1
All	All	424/450 (94%)	374 (88%)	39 (9%)	11 (3%)	6	3

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	3	ALA
1	A	179	PRO
1	A	214	ASN
1	B	3	ALA
1	B	78	SER

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	159/169 (94%)	135 (85%)	24 (15%)	3	2
1	B	159/169 (94%)	132 (83%)	27 (17%)	2	2

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	318/338 (94%)	267 (84%)	51 (16%)	3 2

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	211	ILE
1	B	24	PHE
1	B	182	GLU
1	B	4	LEU
1	B	33	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	120	GLN
1	B	120	GLN
1	B	75	HIS
1	A	75	HIS
1	B	21	ASN

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](#)

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	214/225 (95%)	-0.14	5 (2%) 61 58	6, 20, 34, 37	0
1	B	214/225 (95%)	-0.14	8 (3%) 42 40	7, 17, 33, 42	0
All	All	428/450 (95%)	-0.14	13 (3%) 51 48	6, 19, 34, 42	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	179	PRO	4.4
1	A	27	GLU	3.4
1	B	27	GLU	3.1
1	B	81	ALA	3.0
1	B	179	PRO	2.9

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
2	HG	B	800	1/1	0.41	0.17	0.78	44,44,44,44	1
2	HG	A	800	1/1	0.60	0.11	-0.79	59,59,59,59	1

6.5 Other polymers [i](#)

There are no such residues in this entry.