



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 04:14 pm GMT

PDB ID : 5C22
Title : Crystal structure of Zn-bound HlyD from E. coli
Authors : Ha, N.C.; Kim, J.S.
Deposited on : 2015-06-15
Resolution : 2.30 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

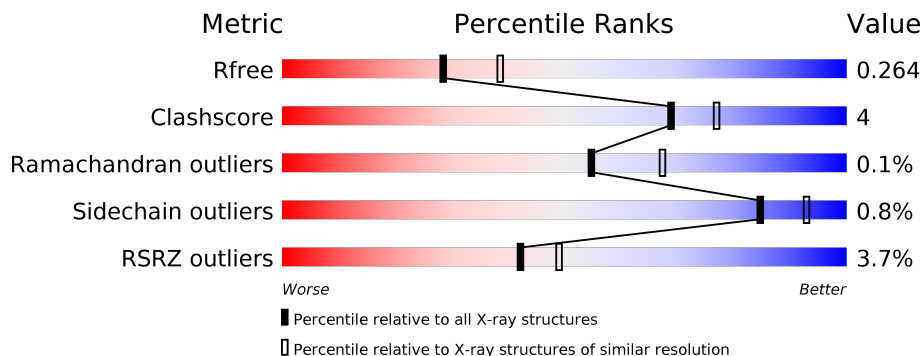
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.30 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	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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	279	<div> <div>4%</div> <div> <div></div> <div>84%</div> <div>11%</div> <div>5%</div> </div> </div>
1	B	279	<div> <div>5%</div> <div> <div></div> <div>81%</div> <div>15%</div> <div>•</div> </div> </div>
1	C	279	<div> <div>3%</div> <div> <div></div> <div>79%</div> <div>11%</div> <div>•</div> <div>9%</div> </div> </div>
1	D	279	<div> <div>2%</div> <div> <div></div> <div>86%</div> <div>9%</div> <div>5%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8740 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 Chromosomal hemolysin D.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	266	Total	C	N	O	S	0	0	0
			2167	1362	380	424	1			
1	B	267	Total	C	N	O	S	0	0	0
			2172	1365	381	425	1			
1	C	254	Total	C	N	O	S	0	0	0
			2078	1308	364	405	1			
1	D	266	Total	C	N	O	S	0	0	0
			2167	1362	380	424	1			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	94	MET	-	expression tag	UNP O87505
A	95	ALA	-	expression tag	UNP O87505
A	294	ARG	GLN	conflict	UNP O87505
B	94	MET	-	expression tag	UNP O87505
B	95	ALA	-	expression tag	UNP O87505
B	294	ARG	GLN	conflict	UNP O87505
C	94	MET	-	expression tag	UNP O87505
C	95	ALA	-	expression tag	UNP O87505
C	294	ARG	GLN	conflict	UNP O87505
D	94	MET	-	expression tag	UNP O87505
D	95	ALA	-	expression tag	UNP O87505
D	294	ARG	GLN	conflict	UNP O87505

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	2	Total 2	Zn 2	0	0

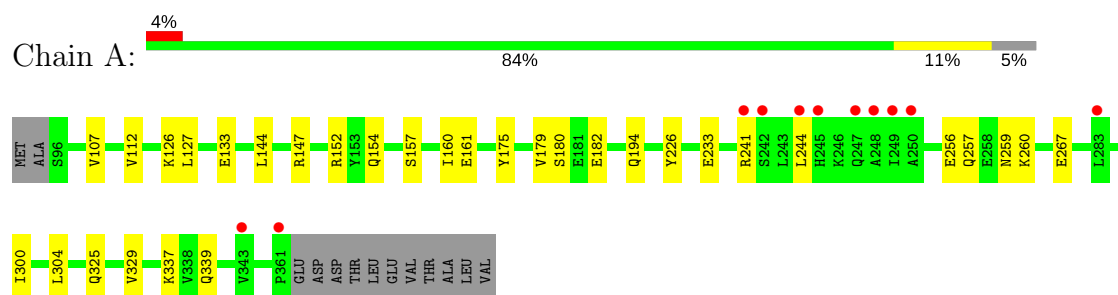
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	30	Total 30	O 30	0	0
3	B	19	Total 19	O 19	0	0
3	C	34	Total 34	O 34	0	0
3	D	68	Total 68	O 68	0	0

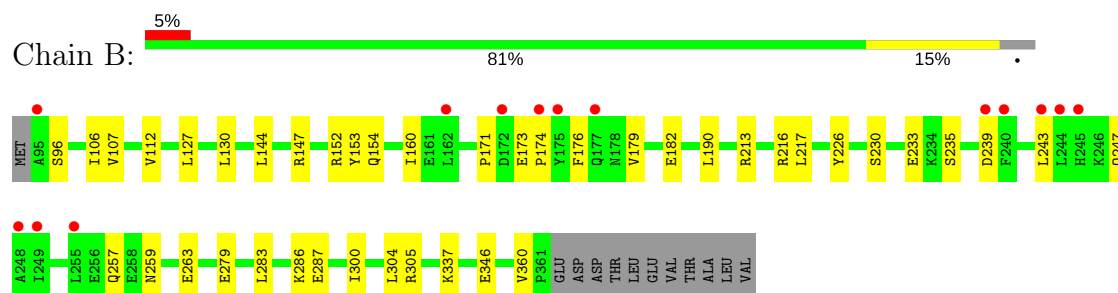
3 Residue-property plots

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.

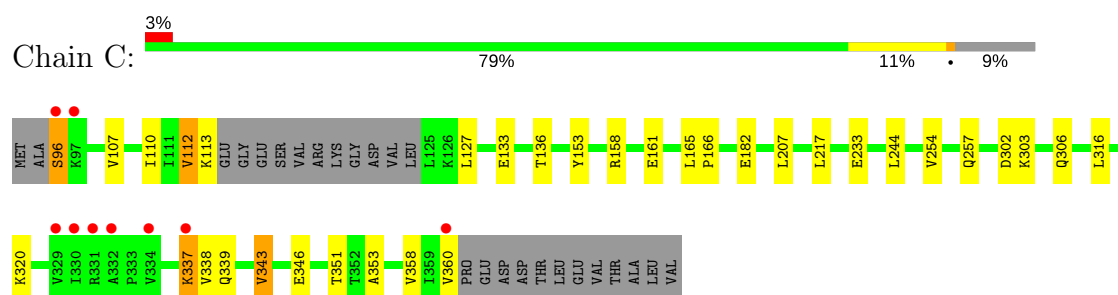
• Molecule 1: Chromosomal hemolysin D



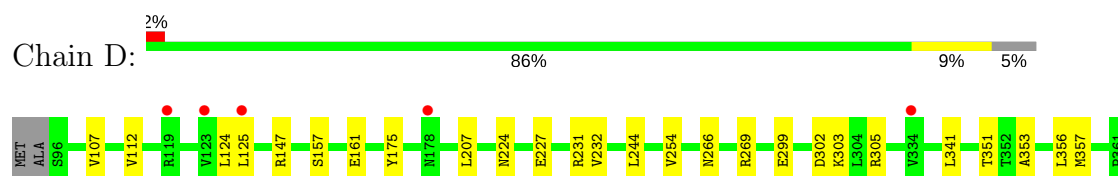
• Molecule 1: Chromosomal hemolysin D



• Molecule 1: Chromosomal hemolysin D



• Molecule 1: Chromosomal hemolysin D



GLU
ASP
ASP
THR
LEU
LEU
GLU
VAL
THR
ALA
LEU
VAL

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	80.86Å 94.48Å 181.38Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.20 – 2.30 37.22 – 2.30	Depositor EDS
% Data completeness (in resolution range)	91.3 (37.20-2.30) 90.9 (37.22-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.52 (at 2.31Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.227 , 0.264 0.227 , 0.264	Depositor DCC
R_{free} test set	3614 reflections (3.33%)	DCC
Wilson B-factor (Å ²)	25.3	Xtriage
Anisotropy	0.528	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 42.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	8740	wwPDB-VP
Average B, all atoms (Å ²)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 25.94 % 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 2.8854e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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: ZN

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	0/2188	0.45	0/2945
1	B	0.29	0/2193	0.48	0/2952
1	C	0.27	0/2097	0.44	0/2821
1	D	0.25	0/2188	0.41	0/2945
All	All	0.27	0/8666	0.44	0/11663

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	2167	0	2258	17	0
1	B	2172	0	2263	25	0
1	C	2078	0	2168	21	0
1	D	2167	0	2258	15	0
2	A	1	0	0	0	0
2	B	2	0	0	0	0
2	D	2	0	0	0	0
3	A	30	0	0	0	0
3	B	19	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	34	0	0	0	0
3	D	68	0	0	2	0
All	All	8740	0	8947	76	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 4.

The worst 5 of 76 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:C:158:ARG:NH1	1:C:161:GLU:OE2	1.77	1.17
1:D:303:LYS:NZ	3:D:501:HOH:O	2.16	0.78
1:D:351:THR:HG23	1:D:353:ALA:H	1.51	0.75
1:B:226:TYR:O	1:B:230:SER:OG	2.05	0.74
1:C:351:THR:HG23	1:C:353:ALA:H	1.51	0.74

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	264/279 (95%)	259 (98%)	5 (2%)	0	100	100
1	B	265/279 (95%)	264 (100%)	1 (0%)	0	100	100
1	C	250/279 (90%)	248 (99%)	1 (0%)	1 (0%)	38	47
1	D	264/279 (95%)	264 (100%)	0	0	100	100
All	All	1043/1116 (94%)	1035 (99%)	7 (1%)	1 (0%)	55	67

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	343	VAL

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	246/257 (96%)	245 (100%)	1 (0%)	93	97
1	B	246/257 (96%)	243 (99%)	3 (1%)	75	87
1	C	236/257 (92%)	233 (99%)	3 (1%)	73	86
1	D	246/257 (96%)	245 (100%)	1 (0%)	93	97
All	All	974/1028 (95%)	966 (99%)	8 (1%)	85	93

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	247	GLN
1	D	112	VAL
1	C	112	VAL
1	B	235	SER
1	C	96	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	339	GLN
1	B	339	GLN
1	C	339	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 5 ligands modelled in this entry, 5 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	266/279 (95%)	0.40	11 (4%) 38 45	21, 41, 77, 118	0
1	B	267/279 (95%)	0.45	14 (5%) 28 35	29, 46, 77, 101	0
1	C	254/279 (91%)	0.26	9 (3%) 44 51	21, 40, 67, 81	0
1	D	266/279 (95%)	0.13	5 (1%) 67 73	19, 36, 57, 73	0
All	All	1053/1116 (94%)	0.31	39 (3%) 42 49	19, 41, 71, 118	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	248	ALA	5.5
1	B	244	LEU	4.9
1	C	329	VAL	4.7
1	B	177	GLN	4.6
1	A	244	LEU	4.5

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 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	ZN	B	401	1/1	0.95	0.08	-	71,71,71,71	0
2	ZN	D	402	1/1	0.97	0.12	-	63,63,63,63	0
2	ZN	D	401	1/1	0.98	0.06	-	56,56,56,56	0
2	ZN	A	401	1/1	0.90	0.12	-	147,147,147,147	0
2	ZN	B	402	1/1	0.98	0.05	-	63,63,63,63	0

6.5 Other polymers [i](#)

There are no such residues in this entry.