



wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 08:42 pm BST

PDB ID : 5J43
Title : CdiA-CT from uropathogenic Escherichia coli in complex with CysK
Authors : Morse, R.P.; Goulding, C.W.; Johnson, P.M.
Deposited on : 2016-03-31
Resolution : 2.70 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

<https://www.wwpdb.org/validation/2017/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
Mogul : 1.8.5 (274361), CSD as541be (2020)
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

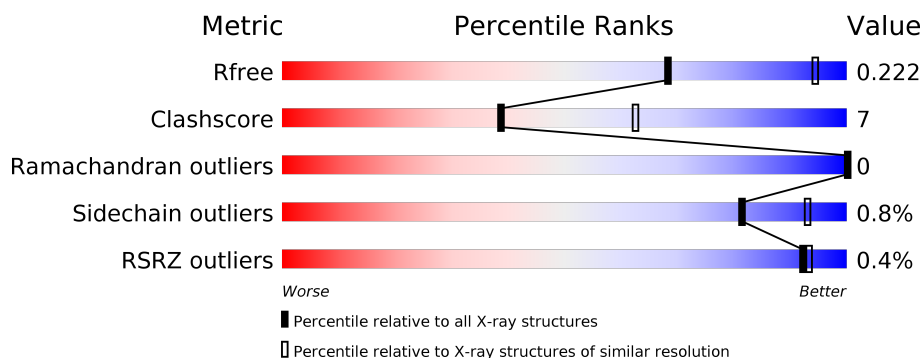
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 $\leq 5\%$. 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	323	<div> <div style="width: 80%;"></div> <div style="width: 17%;"></div> <div style="width: 3%;"></div> </div> <div>80% 17% .</div>
1	E	323	<div> <div style="width: 78%;"></div> <div style="width: 19%;"></div> <div style="width: 3%;"></div> </div> <div>78% 19% .</div>
2	B	239	<div> <div style="width: 36%;"></div> <div style="width: 6%;"></div> <div style="width: 58%;"></div> </div> <div>36% 6% 58%</div>
2	F	239	<div> <div style="width: 37%;"></div> <div style="width: 5%;"></div> <div style="width: 58%;"></div> </div> <div>37% 5% 58%</div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6328 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 Cysteine synthase A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	E	314	Total	C	N	O	P	S	0	0	0
			2351	1483	405	456	1	6			
1	A	313	Total	C	N	O	P	S	0	0	0
			2346	1480	404	455	1	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	2	GLY	SER	conflict	UNP P0ABK6
A	2	GLY	SER	conflict	UNP P0ABK6

- Molecule 2 is a protein called tRNA nuclease CdiA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	101	Total	C	N	O	S	0	0	0
			765	469	141	152	3			
2	B	101	Total	C	N	O	S	0	0	0
			756	461	140	152	3			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-11	MET	-	initiating methionine	UNP Q0T963
F	-10	ALA	-	expression tag	UNP Q0T963
F	-9	LYS	-	expression tag	UNP Q0T963
F	-8	SER	-	expression tag	UNP Q0T963
F	-7	HIS	-	expression tag	UNP Q0T963
F	-6	HIS	-	expression tag	UNP Q0T963
F	-5	HIS	-	expression tag	UNP Q0T963
F	-4	HIS	-	expression tag	UNP Q0T963
F	-3	HIS	-	expression tag	UNP Q0T963
F	-2	HIS	-	expression tag	UNP Q0T963

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	THR	-	expression tag	UNP Q0T963
F	0	SER	-	expression tag	UNP Q0T963
F	178	ALA	HIS	engineered mutation	UNP Q0T963
B	-11	MET	-	initiating methionine	UNP Q0T963
B	-10	ALA	-	expression tag	UNP Q0T963
B	-9	LYS	-	expression tag	UNP Q0T963
B	-8	SER	-	expression tag	UNP Q0T963
B	-7	HIS	-	expression tag	UNP Q0T963
B	-6	HIS	-	expression tag	UNP Q0T963
B	-5	HIS	-	expression tag	UNP Q0T963
B	-4	HIS	-	expression tag	UNP Q0T963
B	-3	HIS	-	expression tag	UNP Q0T963
B	-2	HIS	-	expression tag	UNP Q0T963
B	-1	THR	-	expression tag	UNP Q0T963
B	0	SER	-	expression tag	UNP Q0T963
B	178	ALA	HIS	engineered mutation	UNP Q0T963

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	E	40	Total O 40 40	0	0
3	F	9	Total O 9 9	0	0
3	A	48	Total O 48 48	0	0
3	B	13	Total O 13 13	0	0

- Molecule 1: Cysteine synthase A



ASP	ARG	MET	THR	SER	THR	ASP	GLU	LEU	GLU	HIS	LEU	ASN	ASP	GLU	ILE	THR	THR	THR	LYS	TYR	LEU	SER	SER	LEU	HIS	ASP	LYS	THR	GLY	SER	GLY	ALA	ALA	SER	ASN	ASN	PRO	ASN	ASN	ILE	GLY	LYS	ASP	LEU	THR	ASP	ALA	GLU	LYS	VAL	GLU	LEU	GLY	GLY	GLY
SER	GLY	THR	GLY	THR	THR	PRO	PRO	PRO	SER	GLU	ASN	ASP	PRO	LYS	GLN	GLN	ASN	GLU	ASN	GLU	K127	L132	N133	Q134	K135	Q136	T140	K141	K142	T146	L151	K166	K170	E171	N172	G173	G174	Y175	H176	D177	N201	P202	I218	E219	I227										

4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α , β , γ	64.01Å 64.01Å 365.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.92 – 2.70 44.92 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.0 (44.92-2.70) 99.3 (44.92-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.43 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.8.2 _1309	Depositor
R, R_{free}	0.200 , 0.224 0.199 , 0.222	Depositor DCC
R_{free} test set	1987 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	47.5	Xtriage
Anisotropy	0.717	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 18.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	0.499 for h,-k,-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6328	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.99% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² 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](#)

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

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	0.46	0/2354	0.59	0/3185
1	E	0.47	0/2359	0.62	0/3192
2	B	0.45	0/762	0.56	0/1022
2	F	0.41	0/773	0.58	0/1038
All	All	0.46	0/6248	0.60	0/8437

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	2346	0	2412	36	0
1	E	2351	0	2414	45	0
2	B	756	0	749	11	0
2	F	765	0	757	7	0
3	A	48	0	0	2	0
3	B	13	0	0	0	0
3	E	40	0	0	0	0
3	F	9	0	0	0	0
All	All	6328	0	6332	94	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 7.

The worst 5 of 94 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:E:48:ASN:HD22	1:E:152:ILE:HG22	1.56	0.70
2:F:167:PRO:HD2	2:F:169:PRO:HD3	1.74	0.69
1:E:65:LEU:HD22	1:E:141:LEU:HD21	1.76	0.67
1:A:65:LEU:HD22	1:A:141:LEU:HD21	1.78	0.66
1:E:152:ILE:HG12	1:E:153:HIS:N	2.11	0.65

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	310/323 (96%)	300 (97%)	10 (3%)	0	100	100
1	E	311/323 (96%)	300 (96%)	11 (4%)	0	100	100
2	B	99/239 (41%)	98 (99%)	1 (1%)	0	100	100
2	F	99/239 (41%)	97 (98%)	2 (2%)	0	100	100
All	All	819/1124 (73%)	795 (97%)	24 (3%)	0	100	100

There are no Ramachandran outliers to report.

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%)	91	97
1	E	246/257 (96%)	244 (99%)	2 (1%)	81	93
2	B	77/190 (40%)	76 (99%)	1 (1%)	69	87
2	F	78/190 (41%)	77 (99%)	1 (1%)	69	87
All	All	647/894 (72%)	642 (99%)	5 (1%)	81	93

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	92	MET
1	E	152	ILE
2	F	212	THR
1	A	92	MET
2	B	177	ASP

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

Mol	Chain	Res	Type
1	E	48	ASN
1	A	14	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
1	LLP	E	42	1	23,24,25	1.77	4 (17%)	25,32,34	2.36	4 (16%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
1	LLP	A	42	1	23,24,25	1.83	5 (21%)	25,32,34	2.24	4 (16%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	LLP	E	42	1	-	3/16/17/19	0/1/1/1
1	LLP	A	42	1	-	5/16/17/19	0/1/1/1

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	42	LLP	O3-C3	-5.53	1.24	1.37
1	E	42	LLP	O3-C3	-5.27	1.24	1.37
1	A	42	LLP	C4-C4'	3.04	1.52	1.46
1	A	42	LLP	C2-N1	2.92	1.39	1.33
1	E	42	LLP	C4-C4'	2.90	1.52	1.46

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	42	LLP	OP4-C5'-C5	7.52	123.68	109.35
1	A	42	LLP	OP4-C5'-C5	7.07	122.83	109.35
1	E	42	LLP	C5-C4-C4'	-5.29	112.86	121.56
1	A	42	LLP	C5-C4-C4'	-5.15	113.08	121.56
1	E	42	LLP	CE-NZ-C4'	-4.22	105.94	118.90

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	42	LLP	C4-C4'-NZ-CE
1	E	42	LLP	CD-CE-NZ-C4'
1	E	42	LLP	CE-CD-CG-CB
1	A	42	LLP	CE-CD-CG-CB
1	A	42	LLP	CA-CB-CG-CD

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	E	42	LLP	2	0
1	A	42	LLP	3	0

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	312/323 (96%)	-0.09	0 100 100	28, 38, 59, 74	0
1	E	313/323 (96%)	-0.09	0 100 100	28, 39, 60, 73	0
2	B	101/239 (42%)	0.08	3 (2%) 50 51	33, 45, 98, 128	0
2	F	101/239 (42%)	-0.03	0 100 100	30, 44, 96, 125	0
All	All	827/1124 (73%)	-0.06	3 (0%) 92 93	28, 40, 62, 128	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	172	ASN	4.1
2	B	173	GLY	3.4
2	B	175	TYR	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [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, 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	B-factors(Å ²)	Q<0.9
1	LLP	A	42	24/25	0.96	0.17	29,36,42,45	0
1	LLP	E	42	24/25	0.97	0.18	28,35,42,44	0

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.