



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 26, 2024 – 02:12 AM EDT

PDB ID : 7EU9
Title : Crystal structure of the selenomethionine(SeMet)-derived Cas12i1 R-loop complex before target DNA cleavage
Authors : Zhang, B.; Luo, D.Y.; Li, Y.; OuYang, S.Y.
Deposited on : 2021-05-16
Resolution : 2.35 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 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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
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

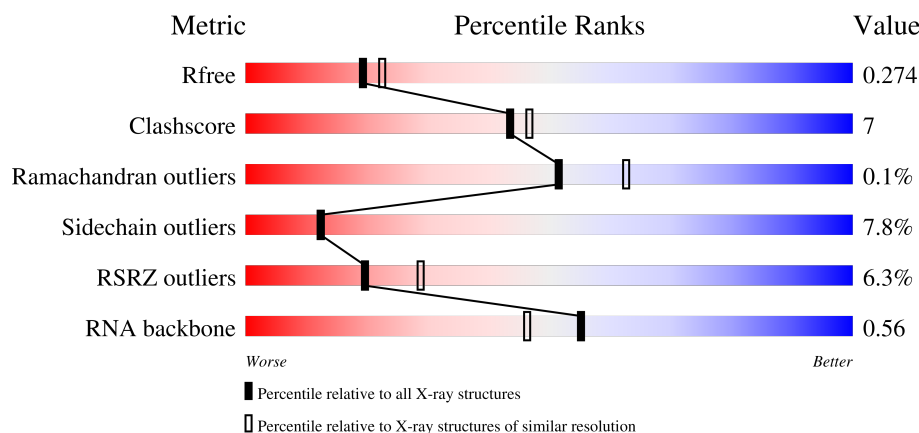
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)
RNA backbone	3102	1006 (2.74-1.98)

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 ≥ 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	1101	<div> <div>6%</div> <div>79%</div> <div>16%</div> <div>••</div> </div>
2	B	43	<div> <div>7%</div> <div>56%</div> <div>28%</div> <div>16%</div> </div>
3	C	40	<div> <div>8%</div> <div>48%</div> <div>30%</div> <div>22%</div> </div>
4	D	40	<div> <div>8%</div> <div>42%</div> <div>18%</div> <div>40%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	CIT	A	1201	-	X	-	-

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 10849 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 Cas12i1 D647A mutant.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1078	Total	C	N	O	S	Se	0	0	0
			8720	5549	1508	1623	14	26			

- Molecule 2 is a RNA chain called RNA (43-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	43	Total	C	N	O	P	0	0	0
			914	409	158	304	43			

- Molecule 3 is a DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	31	Total	C	N	O	P	0	0	0
			629	303	108	187	31			

- Molecule 4 is a DNA chain called DNA (24-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	24	Total	C	N	O	P	0	0	0
			494	235	89	146	24			

- Molecule 5 is CITRIC ACID (three-letter code: CIT) (formula: C₆H₈O₇).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			13	6	7		

- Molecule 6 is water.

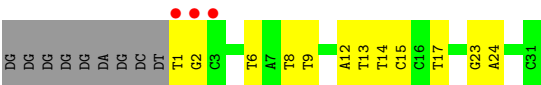
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	69	Total	O	0	0
			69	69		
6	B	3	Total	O	0	0
			3	3		
6	C	3	Total	O	0	0
			3	3		
6	D	4	Total	O	0	0
			4	4		

- Molecule 1: Cas12i1 D647A mutant

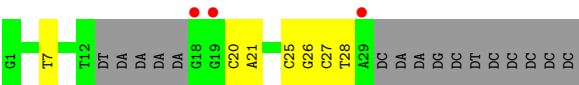




● Molecule 3: DNA (31-MER)



● Molecule 4: DNA (24-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	127.06Å 141.58Å 208.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	58.63 – 2.35 58.56 – 2.35	Depositor EDS
% Data completeness (in resolution range)	99.6 (58.63-2.35) 99.6 (58.56-2.35)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.09 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0222	Depositor
R, R_{free}	0.228 , 0.268 0.237 , 0.274	Depositor DCC
R_{free} test set	3923 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	50.8	Xtriage
Anisotropy	0.054	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 32.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	10849	wwPDB-VP
Average B, all atoms (Å ²)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.79% 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: CIT

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	0.54	0/8872	0.72	0/11896
2	B	0.91	6/1021 (0.6%)	0.84	0/1588
3	C	0.87	4/703 (0.6%)	0.87	0/1081
4	D	0.58	0/552	0.90	0/848
All	All	0.61	10/11148 (0.1%)	0.76	0/15413

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	A	0	8

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	15	DC	O3'-P	-7.08	1.52	1.61
2	B	29	A	O3'-P	-6.93	1.52	1.61
3	C	24	DA	O3'-P	-6.22	1.53	1.61
3	C	23	DG	O3'-P	-6.16	1.53	1.61
3	C	12	DA	O3'-P	-6.07	1.53	1.61

There are no bond angle outliers.

There are no chirality outliers.

5 of 8 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	132	ARG	Sidechain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	228	ARG	Sidechain
1	A	57	ARG	Sidechain
1	A	606	ARG	Sidechain
1	A	723	ARG	Sidechain

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	8720	0	8784	118	0
2	B	914	0	459	10	0
3	C	629	0	353	8	0
4	D	494	0	273	8	0
5	A	13	0	5	3	0
6	A	69	0	0	0	0
6	B	3	0	0	0	0
6	C	3	0	0	0	0
6	D	4	0	0	0	0
All	All	10849	0	9874	137	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 137 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:574:GLN:HG2	1:A:593:MSE:HE2	1.52	0.90
1:A:966:CYS:SG	1:A:971:LEU:CD2	2.60	0.90
3:C:6:DT:H5"	3:C:6:DT:H6	1.38	0.89
1:A:483:LYS:NZ	1:A:614:ASN:HD22	1.75	0.83
1:A:225:ARG:O	1:A:229:ILE:HD12	1.80	0.81

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	1074/1101 (98%)	1025 (95%)	48 (4%)	1 (0%)	51	63

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	666	ASP

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	959/956 (100%)	884 (92%)	75 (8%)	12	12

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	849	MSE
1	A	1030	LEU
1	A	913	MSE
1	A	985	ASP
1	A	397	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	672	ASN
1	A	843	GLN
1	A	711	HIS
1	A	946	HIS
1	A	505	GLN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	42/43 (97%)	9 (21%)	2 (4%)

5 of 9 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	2	U
2	B	4	U
2	B	5	U
2	B	7	G
2	B	14	U

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	4	U
2	B	14	U

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](#)

1 ligand is 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
5	CIT	A	1201	-	12,12,12	1.95	4 (33%)	17,17,17	2.77	8 (47%)

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
5	CIT	A	1201	-	-	7/16/16/16	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1201	CIT	C3-C6	4.31	1.57	1.53
5	A	1201	CIT	C2-C3	-3.08	1.50	1.53
5	A	1201	CIT	O1-C1	2.17	1.29	1.22
5	A	1201	CIT	O2-C1	-2.01	1.24	1.30

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1201	CIT	O6-C6-C3	7.33	125.78	113.05
5	A	1201	CIT	C4-C3-C6	4.68	120.16	110.11
5	A	1201	CIT	C3-C2-C1	-4.41	103.14	113.81
5	A	1201	CIT	O7-C3-C2	-2.76	102.95	109.40
5	A	1201	CIT	O6-C6-O5	-2.72	115.18	123.82

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	1201	CIT	C2-C3-C6-O5
5	A	1201	CIT	C2-C3-C6-O6
5	A	1201	CIT	O7-C3-C6-O5
5	A	1201	CIT	O7-C3-C6-O6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
5	A	1201	CIT	C3-C4-C5-O4

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1201	CIT	3	0

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	1052/1101 (95%)	0.52	63 (5%) 21 32	31, 54, 100, 166	0
2	B	43/43 (100%)	0.37	3 (6%) 16 24	38, 50, 81, 140	0
3	C	31/40 (77%)	0.43	3 (9%) 7 12	39, 50, 175, 187	0
4	D	24/40 (60%)	0.74	3 (12%) 3 6	40, 73, 151, 173	0
All	All	1150/1224 (93%)	0.51	72 (6%) 20 29	31, 54, 104, 187	0

The worst 5 of 72 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	732	GLY	9.4
1	A	734	ASN	8.6
1	A	733	ASN	8.5
1	A	730	ASN	7.8
1	A	729	ASP	7.5

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](#)

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(\AA^2)	Q<0.9
5	CIT	A	1201	13/13	0.86	0.15	42,65,76,78	0

6.5 Other polymers [i](#)

There are no such residues in this entry.