



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 11, 2024 – 07:39 PM JST

PDB ID : 6JOO  
Title : Crystal structure of *Corynebacterium diphtheriae* Cas9 in complex with sgRNA and target DNA  
Authors : Hirano, S.; Ishitani, R.; Nishimasu, H.; Nureki, O.  
Deposited on : 2019-03-22  
Resolution : 2.90 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](mailto: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	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

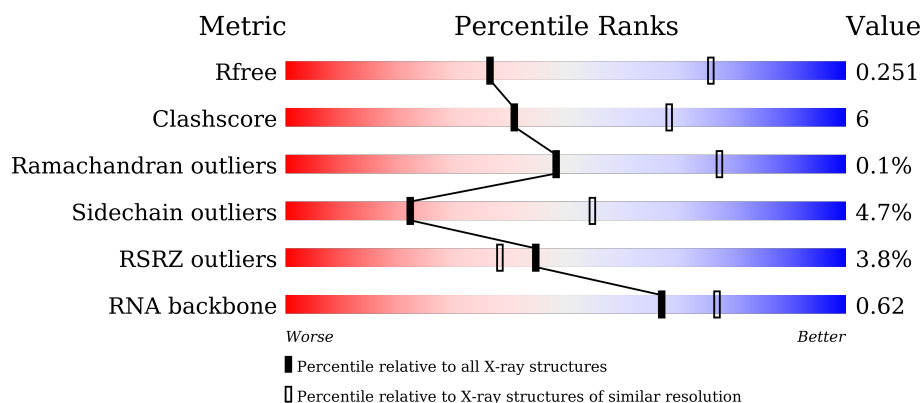
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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))
$R_{free}$	164625	2335 (2.90-2.90)
Clashscore	180529	2564 (2.90-2.90)
Ramachandran outliers	177936	2514 (2.90-2.90)
Sidechain outliers	177891	2516 (2.90-2.90)
RSRZ outliers	164620	2337 (2.90-2.90)
RNA backbone	3690	1039 (3.10-2.70)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	927	<div> <div>4%</div> <div> <div></div> <div>72%</div> <div>16%</div> <div>•</div> <div>10%</div> </div> </div>
2	B	113	<div> <div>%</div> <div> <div></div> <div>62%</div> <div>22%</div> <div>•</div> <div>15%</div> </div> </div>
3	D	8	<div> <div></div> <div> <div>75%</div> <div>25%</div> </div> </div>
4	C	28	<div> <div></div> <div> <div>71%</div> <div>29%</div> </div> </div>

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 9059 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 CRISPR-associated protein, CRISPR-associated endonuclease Cas9.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	834	Total	C	N	O	S	Se	0	0	0
			6292	3952	1147	1176	4	13			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP A0A2T1BQ76
A	-1	SER	-	expression tag	UNP A0A2T1BQ76
A	0	HIS	-	expression tag	UNP A0A2T1BQ76
A	658	GLY	-	linker	UNP A0A2T1BQ76
A	659	GLY	-	linker	UNP A0A2T1BQ76
A	660	GLY	-	linker	UNP A0A2T1BQ76
A	661	SER	-	linker	UNP A0A2T1BQ76
A	662	GLY	-	linker	UNP A0A2T1BQ76
A	663	GLY	-	linker	UNP A0A2T1BQ76

- Molecule 2 is a RNA chain called Guide RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	96	Total	C	N	O	P	0	0	0
			2030	903	357	674	96			

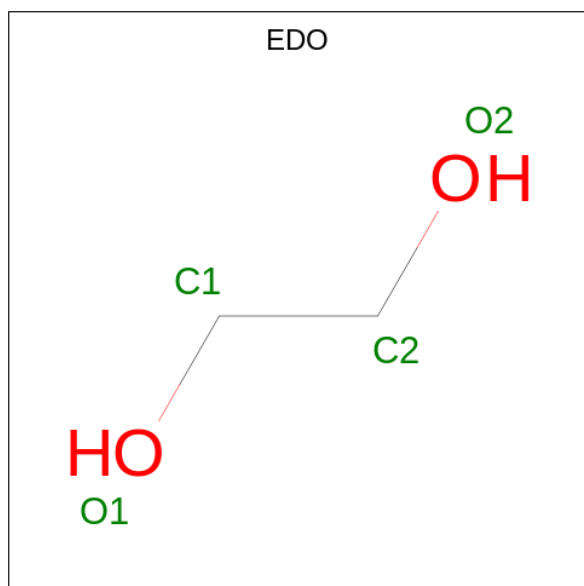
- Molecule 3 is a DNA chain called Non-Target DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	8	Total	C	N	O	P	0	0	0
			167	80	34	46	7			

- Molecule 4 is a DNA chain called Target DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	C	28	Total	C	N	O	P	0	0	0
			558	268	98	165	27			

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			4	2	2		
5	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula:  $Zn$ ).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Zn	0	0
			1	1		

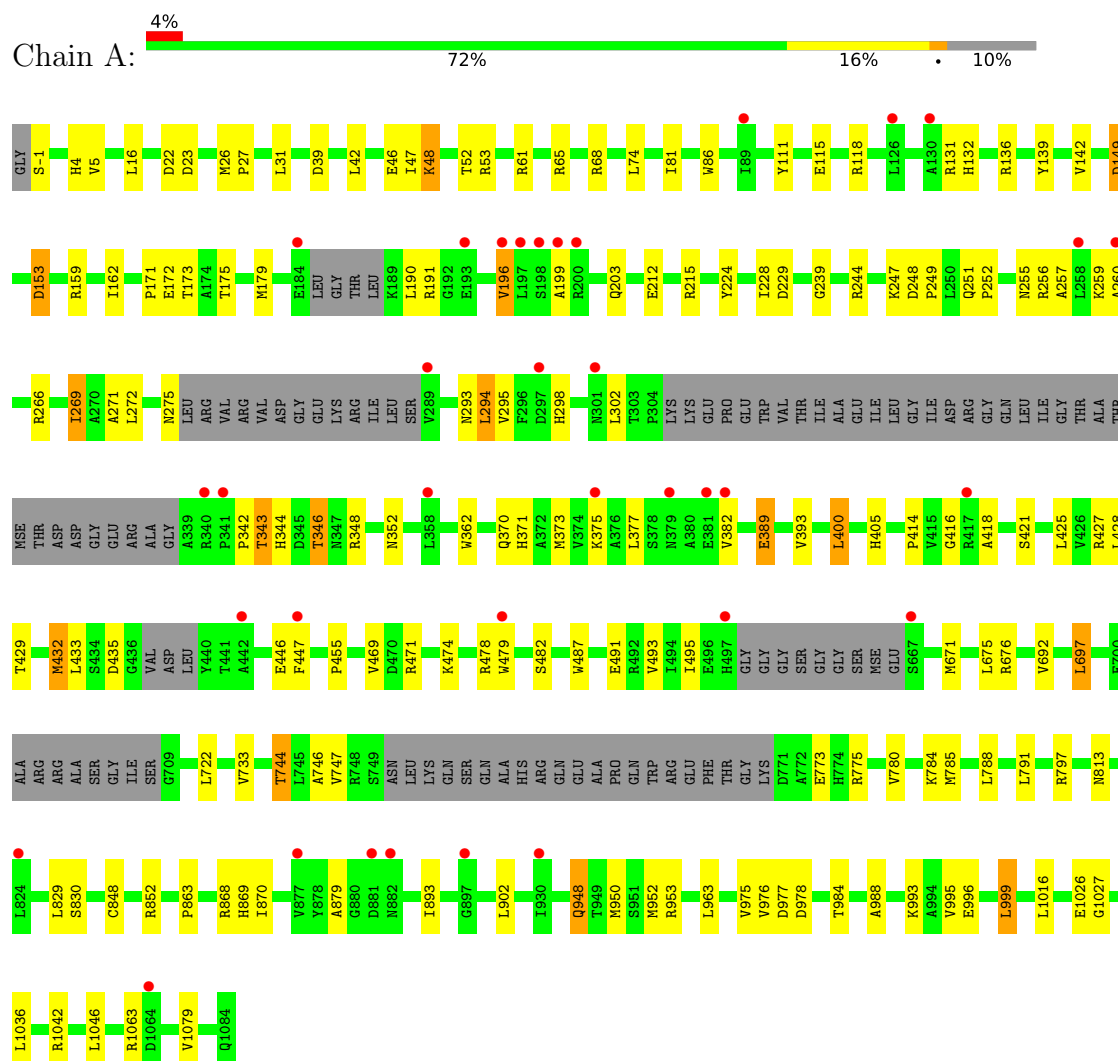
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	3	Total	O	0	0
			3	3		

### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: CRISPR-associated protein, CRISPR-associated endonuclease Cas9





● Molecule 3: Non-Target DNA



● Molecule 4: Target DNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	139.01Å 119.00Å 116.27Å 90.00° 113.59° 90.00°	Depositor
Resolution (Å)	67.94 – 2.90 67.94 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.6 (67.94-2.90) 99.6 (67.94-2.90)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.53 (at 2.91Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, $R_{free}$	0.221 , 0.254 0.220 , 0.251	Depositor DCC
$R_{free}$ test set	2000 reflections (5.19%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	82.0	Xtriage
Anisotropy	0.463	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 83.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9059	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	114.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.68% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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

### 5.1 Standard geometry

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

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	0.25	0/6402	0.43	0/8680
2	B	0.21	0/2266	0.75	0/3529
3	D	0.61	0/188	0.85	0/290
4	C	0.55	0/623	0.88	0/956
All	All	0.28	0/9479	0.58	0/13455

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

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	6292	0	6114	95	0
2	B	2030	0	1022	14	0
3	D	167	0	92	2	0
4	C	558	0	316	9	0
5	A	4	0	6	1	0
5	B	4	0	6	0	0
6	A	1	0	0	0	0
7	A	3	0	0	0	0
All	All	9059	0	7556	106	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 6.

The worst 5 of 106 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:239:GLY:HA3	1:A:382:VAL:HG22	1.67	0.76
1:A:697:LEU:HB2	1:A:746:ALA:HB2	1.72	0.72
2:B:9:G:H1	4:C:20:DC:H42	1.38	0.70
1:A:142:VAL:HG11	1:A:229:ASP:HA	1.75	0.69
1:A:427:ARG:NH2	1:A:446:GLU:OE2	2.26	0.69

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	818/927 (88%)	793 (97%)	24 (3%)	1 (0%)	48 77

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	47	ILE

### 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	640/756 (85%)	610 (95%)	30 (5%)	22 55

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	429	THR
1	A	978	ASP
1	A	435	ASP
1	A	1079	VAL
1	A	869	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	4	HIS
1	A	293	ASN
1	A	298	HIS
1	A	873	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	93/113 (82%)	12 (12%)	0

5 of 12 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	19	G
2	B	20	C
2	B	32	G
2	B	38	U
2	B	42	C

There are no RNA pucker outliers to report.

## 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 oligosaccharides in this entry.

## 5.6 Ligand geometry

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 for Mogul analysis.

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	EDO	B	201	-	3,3,3	0.48	0	2,2,2	0.31	0
5	EDO	A	1101	-	3,3,3	0.48	0	2,2,2	0.29	0

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	EDO	B	201	-	-	0/1/1/1	-
5	EDO	A	1101	-	-	0/1/1/1	-

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1101	EDO	1	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	821/927 (88%)	0.08	35 (4%) 40 34	52, 111, 200, 336	0
2	B	96/113 (84%)	-0.39	1 (1%) 79 74	66, 103, 173, 210	0
3	D	8/8 (100%)	-0.91	0 100 100	63, 72, 83, 94	0
4	C	28/28 (100%)	-0.68	0 100 100	76, 104, 195, 211	0
All	All	953/1076 (88%)	-0.00	36 (3%) 44 38	52, 109, 198, 336	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	196	VAL	6.1
1	A	375	LYS	5.3
1	A	258	LEU	5.2
1	A	381	GLU	4.4
1	A	881	ASP	4.1

### 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
5	EDO	B	201	4/4	0.50	0.18	90,91,92,92	0
5	EDO	A	1101	4/4	0.81	0.15	77,77,78,79	0
6	ZN	A	1102	1/1	0.95	0.08	121,121,121,121	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.