wwPDB X-ray Structure Validation Summary Report (i)

Nov 2, 2026 – 07:23 PM JST

PDB ID	:	9XJ4 / pdb_00009xj4
Title	:	Soluble expression of CRM197 in Escherichia coli(Named CRM197-OA)
Authors	:	Hu, E.; Yu, R.; Chen, C.
Deposited on	:	2025-11-04
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 (i) 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 (1)) were used in the production of this report:

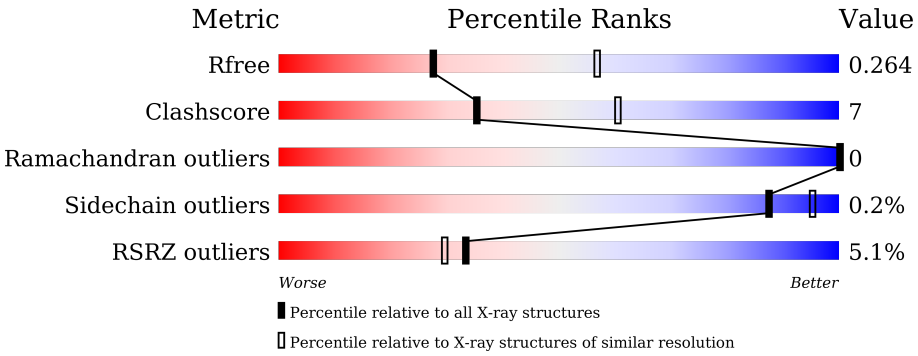
MolProbity	:	4-5-2 with Phenix2.2
Xtriage (Phenix)	:	2.2
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.51

1 Overall quality at a glance (i)

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}	180053	3538 (2.70-2.70)
Clashscore	190562	3843 (2.70-2.70)
Ramachandran outliers	187476	3778 (2.70-2.70)
Sidechain outliers	187428	3778 (2.70-2.70)
RSRZ outliers	180081	3538 (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 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 <=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	542	5% 80% 15% 5%
1	B	542	5% 83% 12% 5%
1	C	542	4% 80% 15% 5%
1	D	542	5% 77% 18% 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 15950 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 Diphtheria toxin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	515	Total	C	N	O	S	0	0	0
			3971	2506	675	778	12			
1	B	515	Total	C	N	O	S	0	0	0
			3968	2503	674	779	12			
1	C	513	Total	C	N	O	S	0	0	0
			3954	2493	670	779	12			
1	D	514	Total	C	N	O	S	0	0	0
			3965	2501	674	778	12			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	53	GLU	GLY	conflict	UNP Q5PY51
A	537	HIS	-	expression tag	UNP Q5PY51
A	538	HIS	-	expression tag	UNP Q5PY51
A	539	HIS	-	expression tag	UNP Q5PY51
A	540	HIS	-	expression tag	UNP Q5PY51
A	541	HIS	-	expression tag	UNP Q5PY51
A	542	HIS	-	expression tag	UNP Q5PY51
B	53	GLU	GLY	conflict	UNP Q5PY51
B	537	HIS	-	expression tag	UNP Q5PY51
B	538	HIS	-	expression tag	UNP Q5PY51
B	539	HIS	-	expression tag	UNP Q5PY51
B	540	HIS	-	expression tag	UNP Q5PY51
B	541	HIS	-	expression tag	UNP Q5PY51
B	542	HIS	-	expression tag	UNP Q5PY51
C	53	GLU	GLY	conflict	UNP Q5PY51
C	537	HIS	-	expression tag	UNP Q5PY51
C	538	HIS	-	expression tag	UNP Q5PY51
C	539	HIS	-	expression tag	UNP Q5PY51
C	540	HIS	-	expression tag	UNP Q5PY51
C	541	HIS	-	expression tag	UNP Q5PY51
C	542	HIS	-	expression tag	UNP Q5PY51

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	53	GLU	GLY	conflict	UNP Q5PY51
D	537	HIS	-	expression tag	UNP Q5PY51
D	538	HIS	-	expression tag	UNP Q5PY51
D	539	HIS	-	expression tag	UNP Q5PY51
D	540	HIS	-	expression tag	UNP Q5PY51
D	541	HIS	-	expression tag	UNP Q5PY51
D	542	HIS	-	expression tag	UNP Q5PY51

- Molecule 2 is water.

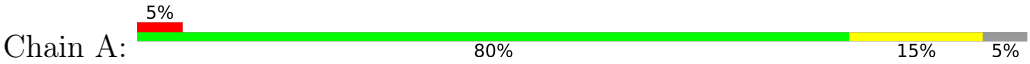
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	26	Total O 26 26	0	0
2	B	17	Total O 17 17	0	0
2	C	30	Total O 30 30	0	0
2	D	19	Total O 19 19	0	0

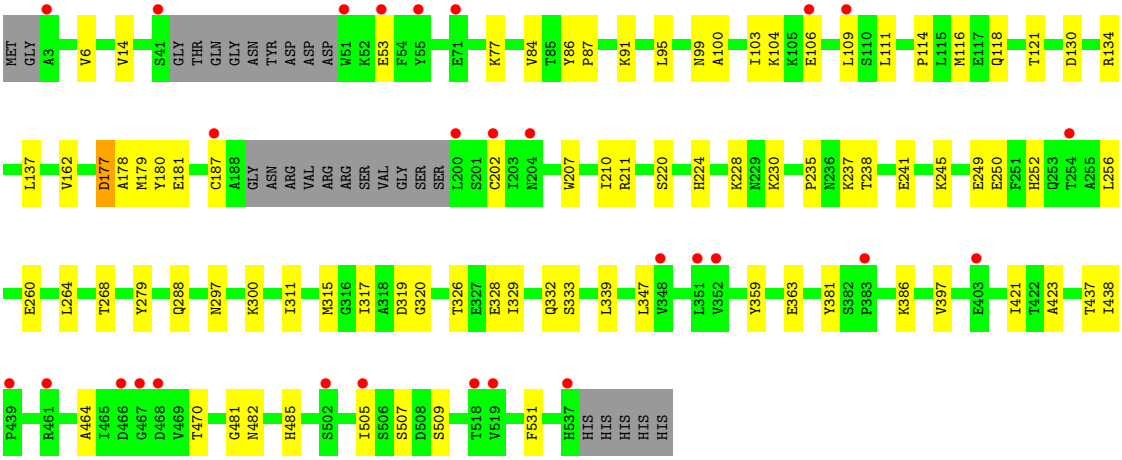
3 Residue-property plots (i)

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: Diphtheria toxin

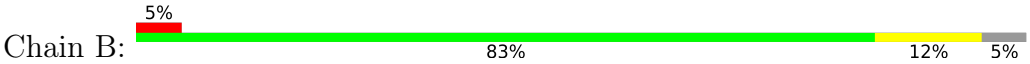
Chain A:

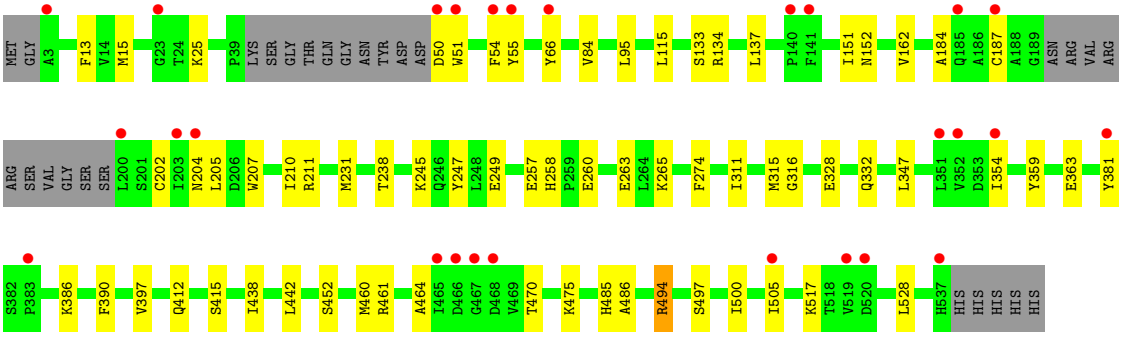




• Molecule 1: Diphtheria toxin

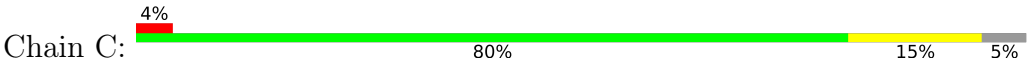
Chain B:

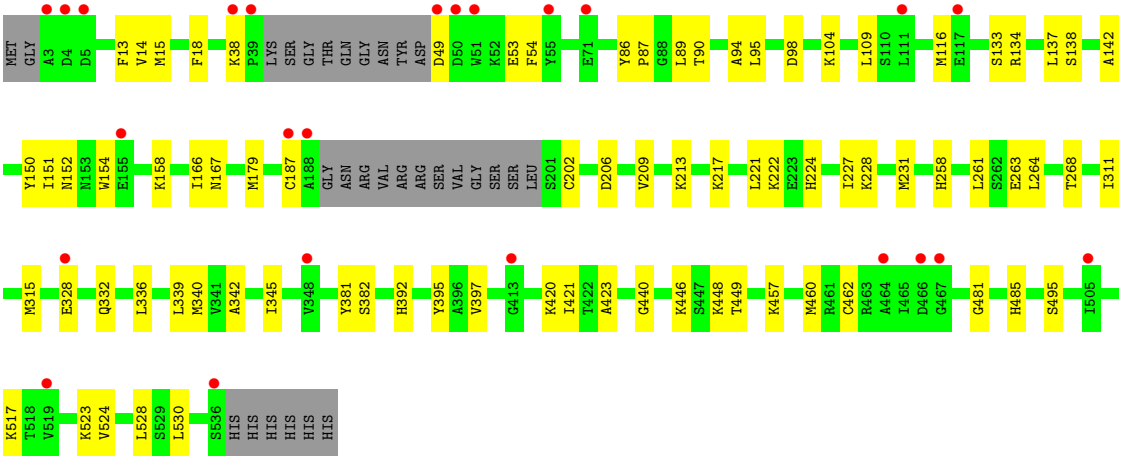




• Molecule 1: Diphtheria toxin

Chain C:

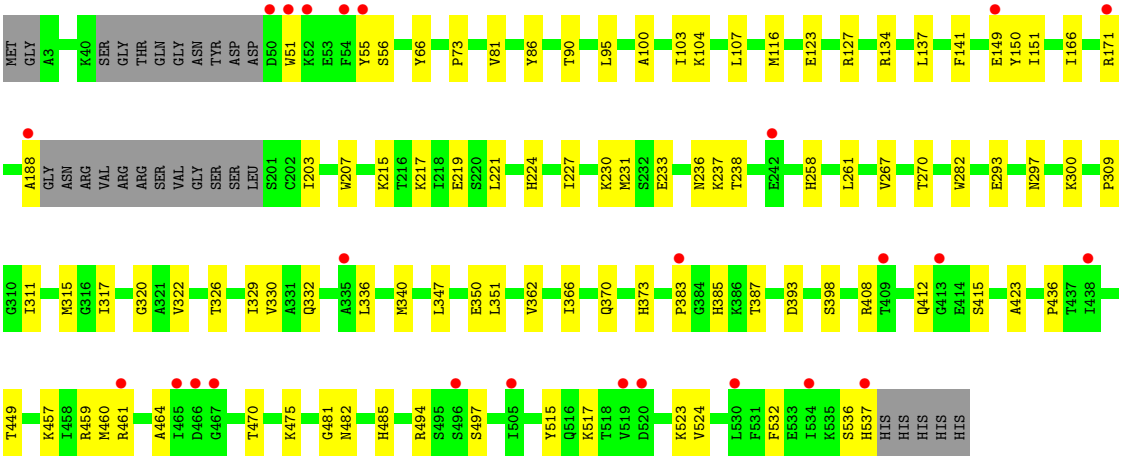




• Molecule 1: Diphtheria toxin

Chain D:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	47.38Å 170.01Å 143.70Å 90.00° 96.04° 90.00°	Depositor
Resolution (Å)	47.63 – 2.70 47.63 – 2.70	Depositor EDS
% Data completeness (in resolution range)	91.8 (47.63-2.70) 84.1 (47.63-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.89 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.224 , 0.268 0.223 , 0.264	Depositor DCC
R_{free} test set	1978 reflections (3.42%)	wwPDB-VP
Wilson B-factor (Å^2)	41.9	Xtriage
Anisotropy	0.312	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 47.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	15950	wwPDB-VP
Average B, all atoms (Å^2)	56.0	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 31.21 % 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 1.1499e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	0.36	0/4051	0.51	1/5487 (0.0%)
1	B	0.27	0/4048	0.41	0/5484
1	C	0.26	0/4033	0.43	0/5464
1	D	0.28	0/4045	0.44	0/5479
All	All	0.30	0/16177	0.45	1/21914 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	177	ASP	CA-C-O	-6.25	114.44	121.00

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	3971	0	3917	57	1
1	B	3968	0	3906	43	1
1	C	3954	0	3889	57	0
1	D	3965	0	3905	79	0
2	A	26	0	0	5	0
2	B	17	0	0	1	0
2	C	30	0	0	1	0
2	D	19	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	15950	0	15617	234	2

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 234 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:D:55:TYR:OH	1:D:141:PHE:CE1	1.79	1.33
1:D:141:PHE:HD1	1:D:149:GLU:CG	1.52	1.21
1:D:55:TYR:OH	1:D:141:PHE:HE1	0.88	1.21
1:D:188:ALA:HB3	2:D:601:HOH:O	1.43	1.14
1:D:141:PHE:HD1	1:D:149:GLU:HG3	1.08	1.11

All (2) 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:B:359:TYR:OH	1:B:386:LYS:O[1_655]	2.08	0.12
1:A:359:TYR:OH	1:A:386:LYS:O[1_655]	2.10	0.10

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	509/542 (94%)	495 (97%)	14 (3%)	0	100	100
1	B	509/542 (94%)	496 (97%)	13 (3%)	0	100	100
1	C	507/542 (94%)	493 (97%)	14 (3%)	0	100	100
1	D	508/542 (94%)	490 (96%)	18 (4%)	0	100	100
All	All	2033/2168 (94%)	1974 (97%)	59 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains

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	438/460 (95%)	437 (100%)	1 (0%)	87 95
1	B	437/460 (95%)	436 (100%)	1 (0%)	87 95
1	C	436/460 (95%)	434 (100%)	2 (0%)	81 92
1	D	437/460 (95%)	437 (100%)	0	100 100
All	All	1748/1840 (95%)	1744 (100%)	4 (0%)	87 95

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	339	LEU
1	B	494	ARG
1	C	133	SER
1	C	517	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	258	HIS
1	D	525	ASN
1	C	167	ASN
1	C	176	GLN
1	C	285	ASN

5.3.3 RNA

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 oligosaccharides 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

6.1 Protein, DNA and RNA chains

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^{th} 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	515/542 (95%)	0.41	28 (5%)	31	27	26, 48, 88, 138	0
1	B	515/542 (95%)	0.49	27 (5%)	33	29	31, 51, 85, 135	0
1	C	513/542 (94%)	0.28	24 (4%)	36	33	24, 45, 91, 159	0
1	D	514/542 (94%)	0.68	25 (4%)	35	31	35, 60, 104, 140	0
All	All	2057/2168 (94%)	0.46	104 (5%)	33	29	24, 51, 94, 159	0

The worst 5 of 104 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	50	ASP	8.2
1	D	50	ASP	5.9
1	C	55	TYR	5.0
1	B	204	ASN	4.8
1	C	188	ALA	4.5

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no oligosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.