



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 12, 2026 – 06:12 PM EST

PDB ID : 9YZJ / pdb_00009yzj
Title : Isoreticular co-crystal 1 with asymmetrical expanded duplex (31mer) containing insert sequence CGTAATTAGG
Authors : Shields, E.T.; Slaughter, C.K.; Magna, E.N.; Snow, C.D.
Deposited on : 2025-10-30
Resolution : 2.97 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 : 4-5-2 with Phenix2.0
Xtrriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.48

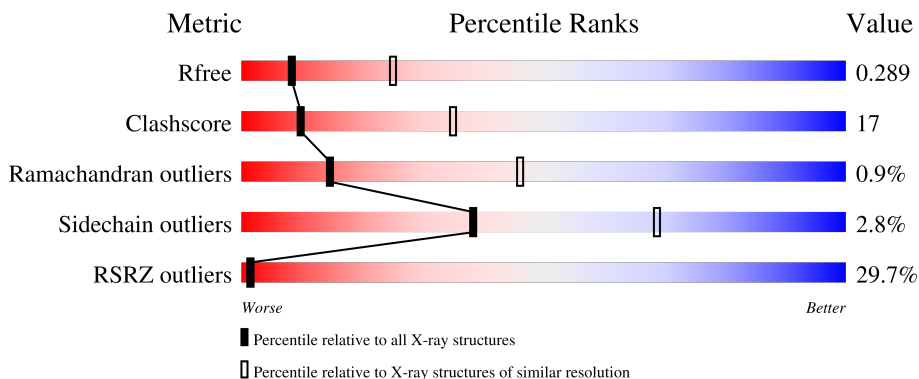
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.97 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	3360 (3.00-2.96)
Clashscore	180529	3751 (3.00-2.96)
Ramachandran outliers	177936	3628 (3.00-2.96)
Sidechain outliers	177891	3631 (3.00-2.96)
RSRZ outliers	164620	3372 (3.00-2.96)

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	31	
2	B	31	
3	C	263	

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 2896 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 DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	31	631	302	115	184	30	0	0	0

- Molecule 2 is a DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	31	637	303	117	186	31	0	0	0

- Molecule 3 is a protein called Replication initiation protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	221	1624	1041	275	302	6	0	3	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-11	MET	-	initiating methionine	UNP P03856
C	-10	ARG	-	expression tag	UNP P03856
C	-9	GLY	-	expression tag	UNP P03856
C	-8	SER	-	expression tag	UNP P03856
C	-7	HIS	-	expression tag	UNP P03856
C	-6	HIS	-	expression tag	UNP P03856
C	-5	HIS	-	expression tag	UNP P03856
C	-4	HIS	-	expression tag	UNP P03856
C	-3	HIS	-	expression tag	UNP P03856
C	-2	HIS	-	expression tag	UNP P03856
C	-1	GLY	-	expression tag	UNP P03856
C	0	SER	-	expression tag	UNP P03856
C	118	PRO	ARG	conflict	UNP P03856

- Molecule 4 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

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

4 Data and refinement statistics i

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	73.78Å 132.65Å 126.77Å 90.00° 90.26° 90.00°	Depositor
Resolution (Å)	46.01 – 2.97 46.01 – 3.09	Depositor EDS
% Data completeness (in resolution range)	97.1 (46.01-2.97) 97.1 (46.01-3.09)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.38 (at 3.06Å)	Xtrriage
Refinement program	PHENIX 1.21.2_5419	Depositor
R, R_{free}	0.272 , 0.292 0.272 , 0.289	Depositor DCC
R_{free} test set	2200 reflections (9.82%)	wwPDB-VP
Wilson B-factor (Å ²)	100.2	Xtrriage
Anisotropy	0.217	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 226.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtrriage
Estimated twinning fraction	0.026 for -h,-l,-k 0.000 for -h,l,k 0.000 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	2896	wwPDB-VP
Average B, all atoms (Å ²)	156.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.74% 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 [i](#)

5.1 Standard geometry [i](#)

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

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	0.18	0/707	0.37	0/1089
2	B	0.27	0/714	0.52	0/1100
3	C	0.57	0/1673	0.99	0/2286
All	All	0.45	0/3094	0.78	0/4475

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	631	0	350	12	0
2	B	637	0	350	9	0
3	C	1624	0	1387	67	0
4	A	2	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
All	All	2896	0	2087	86	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 17.

The worst 5 of 86 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)
3:C:136:ILE:HA	3:C:139:PHE:CE2	2.03	0.93
3:C:209:LEU:O	3:C:213:VAL:HG23	1.85	0.77
3:C:136:ILE:HA	3:C:139:PHE:HE2	1.47	0.75
3:C:26:GLU:HB3	3:C:187:TRP:CZ2	2.24	0.73
3:C:26:GLU:HB3	3:C:187:TRP:HZ2	1.54	0.72

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
3	C	218/263 (83%)	210 (96%)	6 (3%)	2 (1%)	14 46

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	81	ASP
3	C	48	LYS

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
-----	-------	----------	-----------	----------	-------------

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
3	C	144/236 (61%)	140 (97%)	4 (3%)	38 69

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

Mol	Chain	Res	Type
3	C	45	GLN
3	C	95	VAL
3	C	226	TYR
3	C	234	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	C	234	GLN

5.3.3 RNA [i](#)

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](#)

Of 4 ligands modelled in this entry, 4 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	31/31 (100%)	0.34	0 100 100	116, 168, 204, 215	0
2	B	31/31 (100%)	0.44	1 (3%) 50 34	123, 175, 210, 228	0
3	C	221/263 (84%)	1.77	83 (37%) 1 1	67, 133, 210, 282	3 (1%)
All	All	283/325 (87%)	1.47	84 (29%) 1 1	67, 141, 210, 282	3 (1%)

The worst 5 of 84 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	166	TYR	6.6
3	C	93	GLU	6.2
3	C	140	ILE	5.8
3	C	91	GLY	5.3
3	C	142	LEU	5.3

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 oligosaccharides 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
4	MG	A	102	1/1	0.79	0.27	2,2,2,2	0
4	MG	C	301	1/1	0.89	0.10	36,36,36,36	0
4	MG	B	101	1/1	0.92	0.24	16,16,16,16	0
4	MG	A	101	1/1	0.93	0.13	44,44,44,44	0

6.5 Other polymers [i](#)

There are no such residues in this entry.