



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 12, 2026 – 04:08 PM EST

PDB ID : 9YZO / pdb_00009yzo
Title : Isoreticular co-crystal 1 with asymmetrical expanded duplex (31mer) containing insert sequence TGATGAGCAG and loaded with Even-skipped homeodomain
Authors : Shields, E.T.; Slaughter, C.K.; Magna, E.N.; Snow, C.D.
Deposited on : 2025-10-30
Resolution : 2.96 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
Xtriage (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

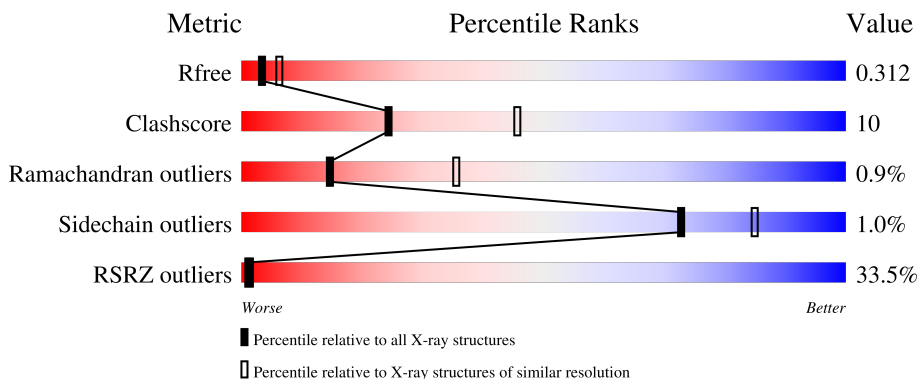
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.96 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	1044 (2.98-2.94)
Clashscore	180529	1097 (2.98-2.94)
Ramachandran outliers	177936	1049 (2.98-2.94)
Sidechain outliers	177891	1049 (2.98-2.94)
RSRZ outliers	164620	1044 (2.98-2.94)

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	
4	D	64	
4	E	64	

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	MG	C	301	-	-	-	X

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3645 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	635	302	115	187	31	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	636	302	118	185	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	217	1667	1068	285	307	7	0	0	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 a protein called Segmentation protein even-skipped.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	56	Total	C	N	O	S	0	0	0
			400	248	77	74	1			
4	E	56	Total	C	N	O		0	0	0
			306	183	61	62				

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	GLY	-	expression tag	UNP P06602
D	2	THR	-	expression tag	UNP P06602
D	3	GLY	-	expression tag	UNP P06602
E	1	GLY	-	expression tag	UNP P06602
E	2	THR	-	expression tag	UNP P06602
E	3	GLY	-	expression tag	UNP P06602

- Molecule 5 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	1	Total	Mg	0	0
			1	1		

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.

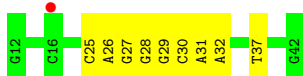
- Molecule 1: DNA (31-MER)

Chain A: 77% 23%



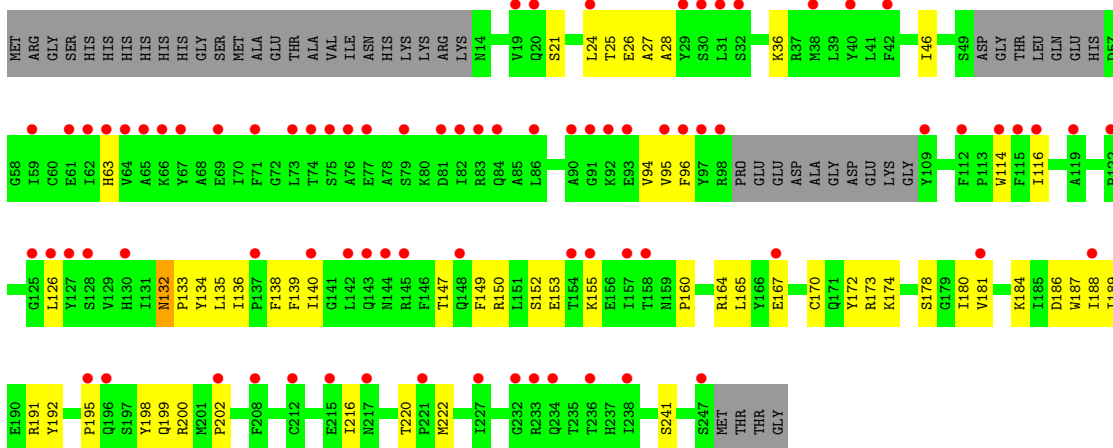
- Molecule 2: DNA (31-MER)

Chain B: 3% 71% 29%



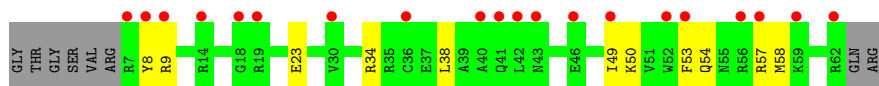
- Molecule 3: Replication initiation protein

Chain C: 30% 61% 21% 17%

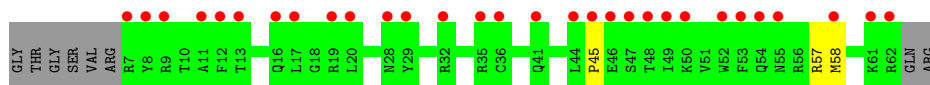
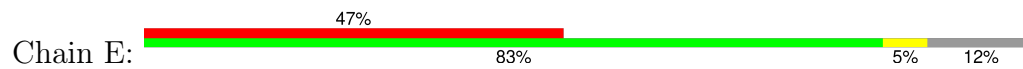


- Molecule 4: Segmentation protein even-skipped

Chain D: 31% 70% 17% 12%



- Molecule 4: Segmentation protein even-skipped



4 Data and refinement statistics

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	74.11Å 118.26Å 141.68Å 90.00° 90.89° 90.00°	Depositor
Resolution (Å)	46.68 – 2.96 46.68 – 2.96	Depositor EDS
% Data completeness (in resolution range)	97.2 (46.68-2.96) 97.1 (46.68-2.96)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.12 (at 2.96Å)	Xtrriage
Refinement program	PHENIX 1.21.2_5419	Depositor
R, R_{free}	0.289 , 0.312 0.289 , 0.312	Depositor DCC
R_{free} test set	2478 reflections (9.70%)	wwPDB-VP
Wilson B-factor (Å ²)	90.4	Xtrriage
Anisotropy	0.179	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 43.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	0.015 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	3645	wwPDB-VP
Average B, all atoms (Å ²)	104.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.03% 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	0/711	0.34	0/1095
2	B	0.19	0/713	0.34	0/1098
3	C	0.28	0/1708	0.68	0/2317
4	D	0.09	0/407	0.24	0/555
4	E	0.08	0/308	0.24	0/428
All	All	0.22	0/3847	0.50	0/5493

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	635	0	350	5	0
2	B	636	0	349	8	0
3	C	1667	0	1519	43	0
4	D	400	0	322	8	0
4	E	306	0	161	0	0
5	C	1	0	0	0	0
All	All	3645	0	2701	62	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 10.

The worst 5 of 62 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:220:THR:HG23	3:C:222:MET:H	1.34	0.92
3:C:28:ALA:HB3	3:C:95:VAL:HG23	1.58	0.83
3:C:135:LEU:HB3	3:C:139:PHE:HE1	1.57	0.68
3:C:132:ASN:HD22	3:C:134:TYR:H	1.41	0.66
2:B:37:DT:H5'	4:D:9:ARG:HD3	1.78	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	
3	C	211/263 (80%)	205 (97%)	6 (3%)	0	100	100
4	D	54/64 (84%)	53 (98%)	1 (2%)	0	100	100
4	E	54/64 (84%)	49 (91%)	2 (4%)	3 (6%)	1	3
All	All	319/391 (82%)	307 (96%)	9 (3%)	3 (1%)	14	35

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	E	45	PRO
4	E	57	ARG
4	E	58	MET

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	
3	C	162/236 (69%)	160 (99%)	2 (1%)	67	82
4	D	30/58 (52%)	30 (100%)	0	100	100
4	E	8/58 (14%)	8 (100%)	0	100	100
All	All	200/352 (57%)	198 (99%)	2 (1%)	73	85

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	132	ASN
3	C	147	THR

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

Mol	Chain	Res	Type
3	C	132	ASN
4	D	28	ASN

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 1 ligands modelled in this entry, 1 is 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%)	0.53	0 100 100	60, 90, 141, 150	0
2	B	31/31 (100%)	0.69	1 (3%) 50 47	67, 94, 146, 155	0
3	C	217/263 (82%)	1.72	80 (36%) 1 1	66, 95, 130, 154	0
4	D	56/64 (87%)	1.62	20 (35%) 1 1	79, 102, 117, 123	0
4	E	56/64 (87%)	2.25	30 (53%) 0 0	127, 167, 180, 181	0
All	All	391/453 (86%)	1.61	131 (33%) 1 1	60, 101, 170, 181	0

The worst 5 of 131 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	43	ASN	6.4
4	D	8	TYR	5.8
4	E	48	THR	5.3
4	E	49	ILE	5.2
4	E	54	GLN	5.2

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(Å ²)	Q<0.9
5	MG	C	301	1/1	0.69	0.86	78,78,78,78	0

6.5 Other polymers [i](#)

There are no such residues in this entry.