



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2026 – 08:10 AM UTC

PDB ID : 1P8K / pdb_00001p8k
Title : The structure and DNA recognition of a bifunctional homing endonuclease and group I intron splicing factor
Authors : Stoddard, B.L.; Bolduc, J.M.
Deposited on : 2003-05-07
Resolution : 2.60 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
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
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) : 2.49

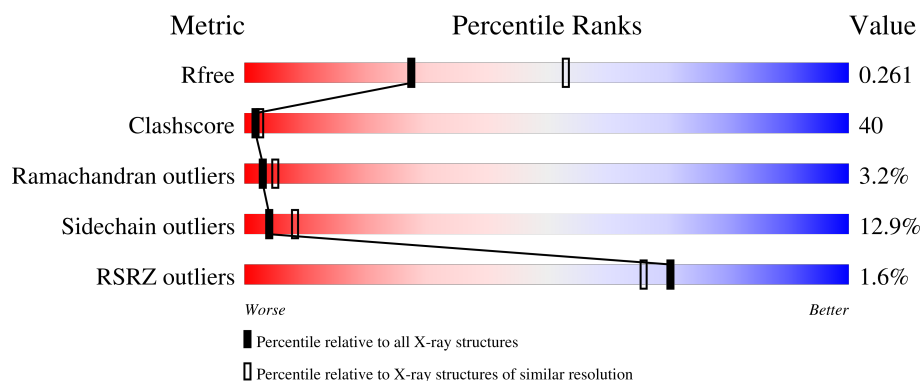
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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	4008 (2.60-2.60)
Clashscore	190562	4347 (2.60-2.60)
Ramachandran outliers	187476	4277 (2.60-2.60)
Sidechain outliers	187428	4277 (2.60-2.60)
RSRZ outliers	180081	4008 (2.60-2.60)

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	18	7% 44% 50% 6%
2	B	13	92% 8%
3	C	16	38% 56% 6%
4	D	15	7% 47% 53%
5	Z	254	2% 41% 44% 13% .

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 3362 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 5'-D(P*GP*CP*GP*CP*GP*CP*TP*GP*AP*GP*GP*AP*GP*GP*TP*TP*TP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	18	Total	C	N	O	P	0	0	0
			375	176	70	111	18			

- Molecule 2 is a DNA chain called 5'-D(P*TP*CP*TP*GP*TP*AP*AP*AP*GP*CP*GP*CP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	13	Total	C	N	O	P	0	0	0
			268	127	50	78	13			

- Molecule 3 is a DNA chain called 5'-D(P*GP*CP*GP*CP*TP*TP*TP*AP*CP*AP*GP*AP*GP*AP*AP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	16	Total	C	N	O	P	0	0	0
			332	157	65	94	16			

- Molecule 4 is a DNA chain called 5'-D(P*CP*CP*TP*CP*CP*TP*CP*AP*GP*CP*GP*CP*GP*CP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	15	Total	C	N	O	P	0	0	0
			300	142	50	93	15			

- Molecule 5 is a protein called Intron-encoded endonuclease I-AniI.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
5	Z	254	Total	C	N	O	S	Se	0	0	0
			2085	1363	339	379	2	2			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Z	1	GLY	-	cloning artifact	UNP P03880
Z	2	SER	-	cloning artifact	UNP P03880
Z	61	ARG	ILE	SEE REMARK 999	UNP P03880
Z	66	MSE	MET	modified residue	UNP P03880
Z	90	MSE	MET	modified residue	UNP P03880

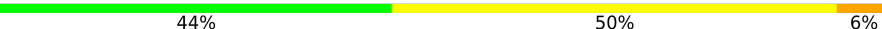
- Molecule 6 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	1	Total 1	Mg 1	0	0
6	Z	1	Total 1	Mg 1	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.

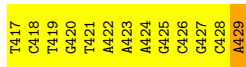
- Molecule 1: 5'-D(P*GP*CP*GP*CP*GP*CP*TP*GP*AP*GP*GP*AP*GP*GP*TP*TP*TP*C)-3'

Chain A: 

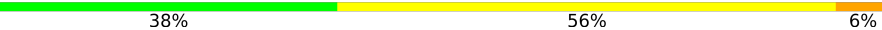


- Molecule 2: 5'-D(P*TP*CP*TP*GP*TP*AP*AP*AP*GP*CP*GP*CP*A)-3'

Chain B: 



- Molecule 3: 5'-D(P*GP*CP*GP*CP*TP*TP*TP*AP*CP*AP*GP*AP*GP*AP*AP*A)-3',

Chain C: 



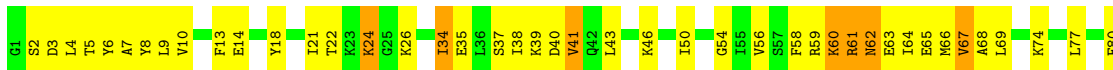
- Molecule 4: 5'-D(P*CP*CP*TP*CP*CP*TP*CP*AP*GP*CP*GP*CP*GP*CP*T)-3'

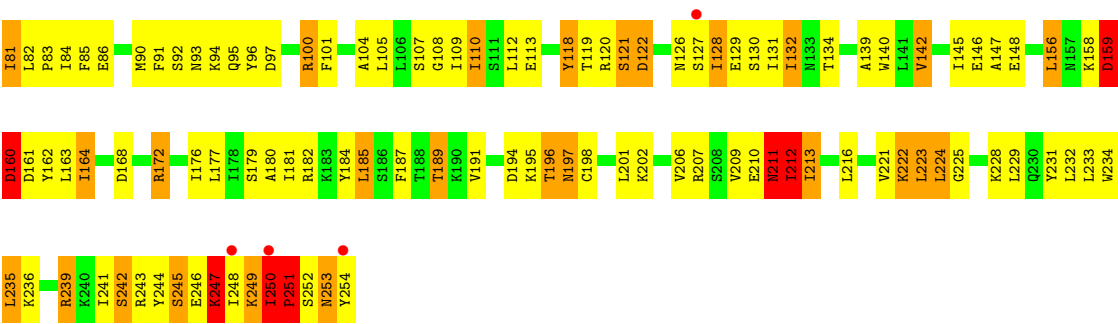
Chain D: 



- Molecule 5: Intron-encoded endonuclease I-AniI

Chain Z: 





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	60.46Å 72.84Å 65.23Å 90.00° 108.21° 90.00°	Depositor
Resolution (Å)	50.00 – 2.60 50.00 – 2.60	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.60) 85.6 (50.00-2.60)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.67 (at 2.61Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.239 , 0.264 0.242 , 0.261	Depositor DCC
R_{free} test set	2799 reflections (9.50%)	wwPDB-VP
Wilson B-factor (Å ²)	37.2	Xtriage
Anisotropy	0.855	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 55.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	3362	wwPDB-VP
Average B, all atoms (Å ²)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.90% 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: 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	0.69	0/420	0.89	1/646 (0.2%)
2	B	0.68	0/300	0.87	1/459 (0.2%)
3	C	0.72	1/373 (0.3%)	0.79	0/572
4	D	0.77	0/333	0.92	0/508
5	Z	0.83	4/2124 (0.2%)	1.47	39/2854 (1.4%)
All	All	0.79	5/3550 (0.1%)	1.25	41/5039 (0.8%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	Z	221	VAL	CA-C	-6.34	1.45	1.53
5	Z	61	ARG	N-CA	6.13	1.54	1.46
5	Z	251	PRO	CB-CG	5.70	1.78	1.49
5	Z	142	VAL	CA-CB	-5.12	1.48	1.54
3	C	501	DG	OP3-P	5.05	1.58	1.48

The worst 5 of 41 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	Z	251	PRO	N-CA-C	13.58	140.44	112.47
5	Z	3	ASP	N-CA-C	-13.15	96.94	111.28
5	Z	62	ASN	N-CA-C	12.55	137.53	110.80
5	Z	250	ILE	N-CA-C	11.24	133.15	108.88
5	Z	62	ASN	CA-C-N	-9.73	106.15	122.11

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	375	0	203	14	1
2	B	268	0	147	21	0
3	C	332	0	180	17	1
4	D	300	0	169	18	0
5	Z	2085	0	2164	185	0
6	B	1	0	0	0	0
6	Z	1	0	0	0	0
All	All	3362	0	2863	245	1

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 40.

The worst 5 of 245 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)
5:Z:251:PRO:CG	5:Z:251:PRO:CB	1.78	1.52
5:Z:222:LYS:HA	5:Z:254:TYR:CE1	1.61	1.36
5:Z:222:LYS:CA	5:Z:254:TYR:HE1	1.41	1.32
2:B:420:DG:H2'	2:B:421:DT:H71	1.31	1.11
2:B:425:DG:H2''	2:B:426:DC:H5'	1.36	1.08

All (1) 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:A:399:DG:OP3	3:C:501:DG:N2[1_456]	2.19	0.01

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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
5	Z	252/254 (99%)	215 (85%)	29 (12%)	8 (3%)	3 5

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	Z	62	ASN
5	Z	128	ILE
5	Z	159	ASP
5	Z	93	ASN
5	Z	250	ILE

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
5	Z	232/230 (101%)	202 (87%)	30 (13%)	4 8

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

Mol	Chain	Res	Type
5	Z	164	ILE
5	Z	246	GLU
5	Z	189	THR
5	Z	251	PRO
5	Z	224	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
5	Z	171	GLN
5	Z	211	ASN
5	Z	230	GLN
5	Z	126	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
5	Z	95	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 2 ligands modelled in this entry, 2 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 [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	18/18 (100%)	-0.17	0	100 100	60, 72, 100, 105	0
2	B	13/13 (100%)	0.23	0	100 100	65, 88, 95, 99	0
3	C	16/16 (100%)	0.09	0	100 100	62, 78, 104, 108	0
4	D	15/15 (100%)	0.25	1 (6%)	24 19	64, 76, 97, 99	0
5	Z	252/254 (99%)	0.15	4 (1%)	70 66	46, 63, 88, 106	0
All	All	314/316 (99%)	0.14	5 (1%)	70 66	46, 65, 93, 108	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	Z	254	TYR	5.9
5	Z	250	ILE	5.0
4	D	517	DC	2.6
5	Z	248	ILE	2.3
5	Z	127	SER	2.0

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($\text{\AA}^2$)	Q<0.9
6	MG	Z	601	1/1	0.83	0.08	60,60,60,60	0
6	MG	B	602	1/1	0.92	0.07	77,77,77,77	0

6.5 Other polymers [i](#)

There are no such residues in this entry.