



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 10:00 AM UTC

PDB ID : 2QOJ / pdb_00002qoj
Title : Coevolution of a homing endonuclease and its host target sequence
Authors : Scalley-Kim, M.; McConnell Smith, A.; Stoddard, B.L.
Deposited on : 2007-07-20
Resolution : 2.40 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

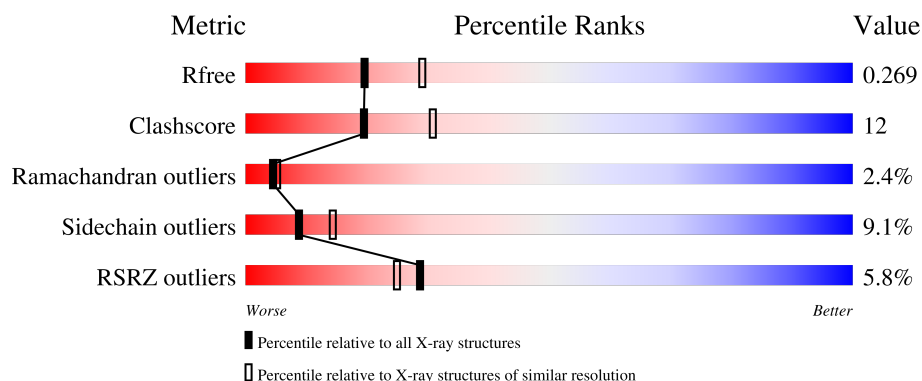
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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	4912 (2.40-2.40)
Clashscore	190562	5391 (2.40-2.40)
Ramachandran outliers	187476	5320 (2.40-2.40)
Sidechain outliers	187428	5321 (2.40-2.40)
RSRZ outliers	180081	4916 (2.40-2.40)

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	Z	254	4% 70% 24% 5%
2	X	31	3% 48% 45% 6%
3	Y	31	19% 52% 42% • •

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3344 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 LAGLIDADG endonuclease.

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

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Z	1	GLY	ALA	conflict	UNP H9D0N7
Z	2	SER	GLY	conflict	UNP H9D0N7

- Molecule 2 is a DNA chain called I-AniI DNA target seq1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	29	Total	C	N	O	P	0	0	0
			600	284	112	175	29			

- Molecule 3 is a DNA chain called I-AniI DNA target seq2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	Y	30	Total	C	N	O	P	0	0	0
			610	289	113	178	30			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	Z	3	Total	Mg	0	0
			3	3		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	Z	38	Total	O	0	0
			38	38		

Continued on next page...

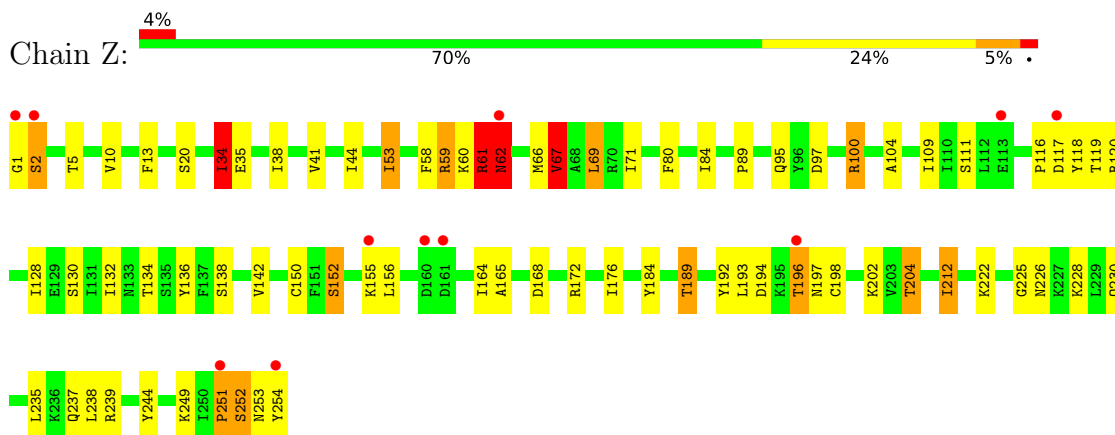
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	X	4	Total	O	0	0
			4	4		
5	Y	4	Total	O	0	0
			4	4		

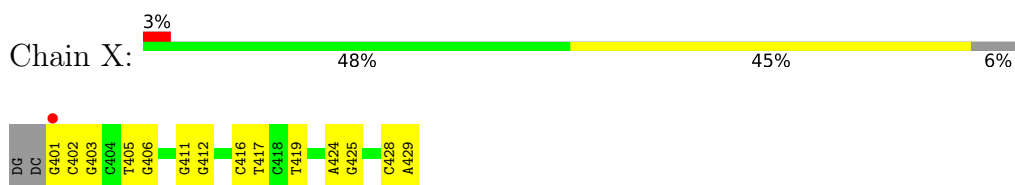
3 Residue-property plots

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: LAGLIDADG endonuclease



- Molecule 2: I-AniI DNA target seq1



- Molecule 3: I-AniI DNA target seq2



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	60.35Å 72.70Å 61.10Å 90.00° 103.88° 90.00°	Depositor
Resolution (Å)	29.30 – 2.40 29.30 – 2.40	Depositor EDS
% Data completeness (in resolution range)	92.1 (29.30-2.40) 76.7 (29.30-2.40)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	11.96 (at 2.42Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.216 , 0.276 0.212 , 0.269	Depositor DCC
R_{free} test set	778 reflections (3.87%)	wwPDB-VP
Wilson B-factor (Å ²)	32.6	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 42.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.056 for l,-k,h	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	3344	wwPDB-VP
Average B, all atoms (Å ²)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.38% 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	Z	0.81	1/2124 (0.0%)	1.06	6/2854 (0.2%)
2	X	0.45	0/673	1.21	2/1038 (0.2%)
3	Y	0.45	0/683	1.24	5/1050 (0.5%)
All	All	0.69	1/3480 (0.0%)	1.13	13/4942 (0.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	Z	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	Z	44	ILE	CA-CB	6.36	1.61	1.54

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	Z	62	ASN	N-CA-C	8.27	128.42	110.80
3	Y	523	DC	O3'-P-O5'	-6.80	93.80	104.00
2	X	416	DC	N1-C1'-C2'	6.40	123.10	113.50
2	X	417	DT	P-O5'-C5'	5.83	128.75	120.00
1	Z	67	VAL	CB-CA-C	-5.70	101.72	110.50

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	Z	61	ARG	Peptide

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	Z	2085	0	2164	61	0
2	X	600	0	327	8	0
3	Y	610	0	336	8	0
4	Z	3	0	0	0	0
5	X	4	0	0	0	0
5	Y	4	0	0	2	0
5	Z	38	0	0	4	1
All	All	3344	0	2827	71	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 12.

The worst 5 of 71 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:Z:239:ARG:HH21	1:Z:251:PRO:HB3	1.33	0.89
1:Z:196:THR:O	1:Z:198:CYS:N	2.04	0.88
1:Z:5:THR:HG22	5:Z:629:HOH:O	1.83	0.77
1:Z:34:ILE:HD12	1:Z:69:LEU:HB3	1.70	0.73
1:Z:60:LYS:O	1:Z:61:ARG:C	2.31	0.73

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)
5:Z:610:HOH:O	5:Z:630:HOH:O[2_545]	1.29	0.91

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	Z	252/254 (99%)	236 (94%)	10 (4%)	6 (2%)	4 5

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	Z	2	SER
1	Z	62	ASN
1	Z	197	ASN
1	Z	252	SER
1	Z	61	ARG

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	Z	232/230 (101%)	211 (91%)	21 (9%)	9 14

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	Z	193	LEU
1	Z	212	ILE
1	Z	238	LEU
1	Z	235	LEU
1	Z	204	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such

sidechains are listed below:

Mol	Chain	Res	Type
1	Z	95	GLN
1	Z	211	ASN
1	Z	237	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 3 ligands modelled in this entry, 3 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	Z	252/254 (99%)	0.01	11 (4%) 39 34	12, 24, 41, 52	0
2	X	29/31 (93%)	0.33	1 (3%) 48 44	16, 31, 55, 61	0
3	Y	30/31 (96%)	0.63	6 (20%) 3 2	16, 37, 75, 87	0
All	All	311/316 (98%)	0.10	18 (5%) 29 25	12, 25, 49, 87	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	Y	501	DG	6.1
2	X	401	DG	5.0
3	Y	530	DC	4.8
1	Z	62	ASN	4.3
1	Z	1	GLY	3.4

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
4	MG	Z	602	1/1	0.81	0.14	47,47,47,47	0
4	MG	Z	601	1/1	0.95	0.13	22,22,22,22	0
4	MG	Z	603	1/1	0.96	0.05	37,37,37,37	0

6.5 Other polymers [i](#)

There are no such residues in this entry.