



# wwPDB X-ray Structure Validation Summary Report

Mar 10, 2026 – 06:09 AM UTC

PDB ID : 7S8D / pdb\_00007s8d  
Title : Structure of DNA-free SgrAI  
Authors : Horton, N.C.  
Deposited on : 2021-09-17  
Resolution : 2.02 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](mailto: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 : 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

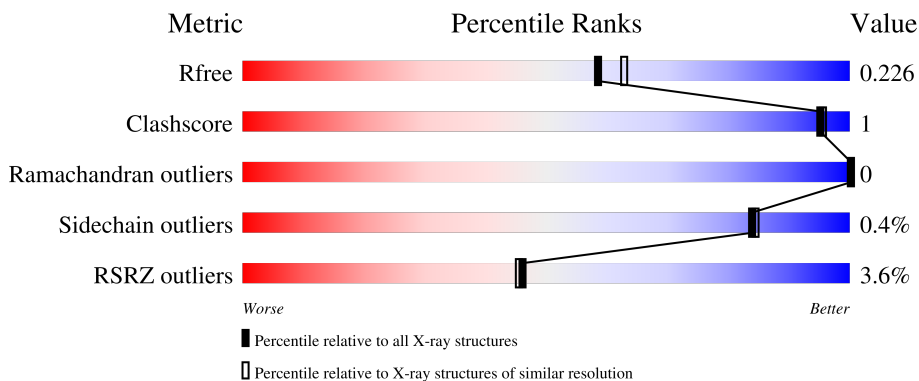
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.02 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	13299 (2.04-2.00)
Clashscore	190562	1022 (2.02-2.02)
Ramachandran outliers	187476	1014 (2.02-2.02)
Sidechain outliers	187428	1014 (2.02-2.02)
RSRZ outliers	180081	13314 (2.04-2.00)

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  $\geq 3$ , 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	352	
1	B	352	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 5471 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 SgraIR restriction enzyme.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	325	2477	1590	426	452	9	0	0	0
1	B	318	2431	1559	413	450	9	0	0	0

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	63	ASP	ASN	conflict	UNP Q9F6L0
A	340	GLU	-	expression tag	UNP Q9F6L0
A	341	ASN	-	expression tag	UNP Q9F6L0
A	342	LEU	-	expression tag	UNP Q9F6L0
A	343	TYR	-	expression tag	UNP Q9F6L0
A	344	PHE	-	expression tag	UNP Q9F6L0
A	345	GLN	-	expression tag	UNP Q9F6L0
A	346	SER	-	expression tag	UNP Q9F6L0
A	347	HIS	-	expression tag	UNP Q9F6L0
A	348	HIS	-	expression tag	UNP Q9F6L0
A	349	HIS	-	expression tag	UNP Q9F6L0
A	350	HIS	-	expression tag	UNP Q9F6L0
A	351	HIS	-	expression tag	UNP Q9F6L0
A	352	HIS	-	expression tag	UNP Q9F6L0
B	63	ASP	ASN	conflict	UNP Q9F6L0
B	340	GLU	-	expression tag	UNP Q9F6L0
B	341	ASN	-	expression tag	UNP Q9F6L0
B	342	LEU	-	expression tag	UNP Q9F6L0
B	343	TYR	-	expression tag	UNP Q9F6L0
B	344	PHE	-	expression tag	UNP Q9F6L0
B	345	GLN	-	expression tag	UNP Q9F6L0
B	346	SER	-	expression tag	UNP Q9F6L0
B	347	HIS	-	expression tag	UNP Q9F6L0
B	348	HIS	-	expression tag	UNP Q9F6L0
B	349	HIS	-	expression tag	UNP Q9F6L0

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	350	HIS	-	expression tag	UNP Q9F6L0
B	351	HIS	-	expression tag	UNP Q9F6L0
B	352	HIS	-	expression tag	UNP Q9F6L0

- Molecule 2 is CALCIUM ION (CCD ID: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Ca 1 1	0	0
2	B	1	Total Ca 1 1	0	0

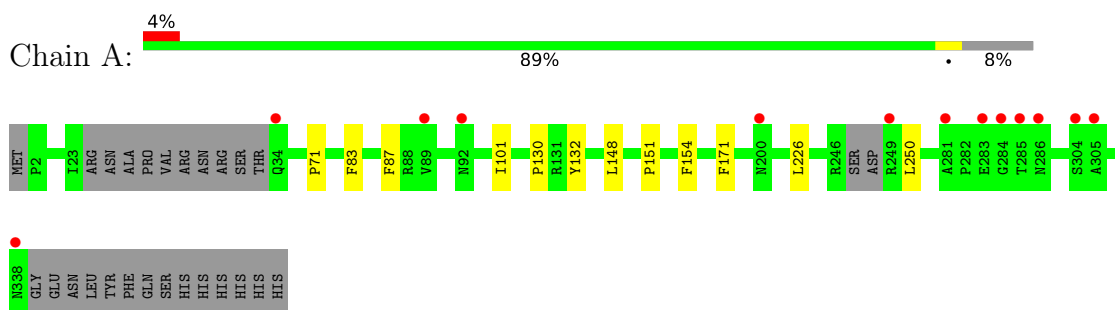
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	285	Total O 285 285	0	0
3	B	276	Total O 276 276	0	0

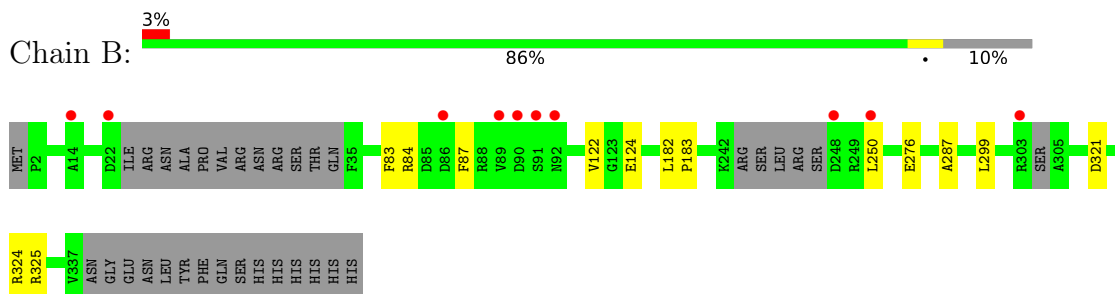
### 3 Residue-property plots [i](#)

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: SgraIR restriction enzyme



- Molecule 1: SgraIR restriction enzyme



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.11Å 72.38Å 78.64Å 90.00° 107.95° 90.00°	Depositor
Resolution (Å)	38.00 – 2.02 38.00 – 2.02	Depositor EDS
% Data completeness (in resolution range)	97.5 (38.00-2.02) 97.5 (38.00-2.02)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.13 (at 2.03Å)	Xtrriage
Refinement program	PHENIX 1.19_4080	Depositor
R, $R_{free}$	0.202 , 0.246 (Not available) , 0.226	Depositor DCC
$R_{free}$ test set	2000 reflections (4.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.1	Xtrriage
Anisotropy	0.549	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 43.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5471	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.53% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: CA

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	0.07	0/2533	0.22	0/3449
1	B	0.07	0/2486	0.23	0/3385
All	All	0.07	0/5019	0.22	0/6834

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

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	2477	0	2394	6	0
1	B	2431	0	2337	7	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	285	0	0	0	0
3	B	276	0	0	0	0
All	All	5471	0	4731	12	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 1.

The worst 5 of 12 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:A:148:LEU:HD11	1:A:226:LEU:HD12	1.85	0.57
1:A:151:PRO:HG2	1:A:154:PHE:HB2	1.93	0.49
1:B:250:LEU:HD21	1:B:287:ALA:HA	1.96	0.47
1:B:122:VAL:HG23	1:B:124:GLU:HG3	1.98	0.45
1:B:83:PHE:HB2	1:B:87:PHE:HB2	1.99	0.44

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	
1	A	319/352 (91%)	315 (99%)	4 (1%)	0	100	100
1	B	310/352 (88%)	304 (98%)	6 (2%)	0	100	100
All	All	629/704 (89%)	619 (98%)	10 (2%)	0	100	100

There are no Ramachandran outliers to report.

#### 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	A	246/296 (83%)	245 (100%)	1 (0%)	84	84
1	B	244/296 (82%)	243 (100%)	1 (0%)	84	84

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	490/592 (83%)	488 (100%)	2 (0%)	84	84

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

Mol	Chain	Res	Type
1	A	250	LEU
1	B	84	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	21	GLN
1	A	39	GLN
1	B	199	GLN
1	B	224	GLN
1	B	227	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

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	325/352 (92%)	0.25	13 (4%) 42 41	14, 24, 41, 56	0
1	B	318/352 (90%)	0.20	10 (3%) 51 51	14, 22, 41, 57	0
All	All	643/704 (91%)	0.22	23 (3%) 46 45	14, 23, 41, 57	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	22	ASP	3.7
1	B	303	ARG	3.3
1	B	250	LEU	3.0
1	A	338	ASN	2.9
1	B	92	ASN	2.9

### 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, 95<sup>th</sup> 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
2	CA	A	401	1/1	0.91	0.07	26,26,26,26	1
2	CA	B	401	1/1	0.92	0.07	31,31,31,31	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.