



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 10, 2017 – 10:15 AM EDT

PDB ID : 2OST  
Title : The structure of a bacterial homing endonuclease : I-Ssp6803I  
Authors : Zhao, L.; Bonocora, R.P.; Shub, D.A.; Stoddard, B.L.  
Deposited on : unknown  
Resolution : 3.10 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

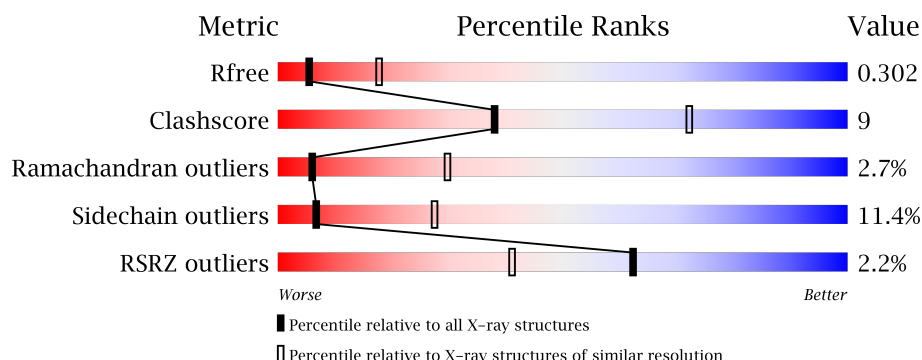
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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))
$R_{free}$	100719	1001 (3.12-3.08)
Clashscore	112137	1099 (3.12-3.08)
Ramachandran outliers	110173	1057 (3.12-3.08)
Sidechain outliers	110143	1057 (3.12-3.08)
RSRZ outliers	101464	1006 (3.12-3.08)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	Y	29	<div> <div>24%</div> <div>48%</div> <div>28%</div> </div>
2	Z	29	<div> <div>3%</div> <div>7%</div> <div>55%</div> <div>38%</div> </div>
3	A	151	<div> <div>%</div> <div>74%</div> <div>21%</div> <div>...</div> </div>
3	B	151	<div> <div>%</div> <div>73%</div> <div>19%</div> <div>5%</div> <div>..</div> </div>
3	C	151	<div> <div>3%</div> <div>70%</div> <div>17%</div> <div>5%</div> <div>9%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	151	<div><div></div><div>4%</div><div>70%</div><div>15%</div><div>• •</div><div>12%</div></div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5747 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 Synthetic DNA 29 MER.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	Y	29	Total	C	N	O	P	0	0	0
			599	282	117	171	29			

- Molecule 2 is a DNA chain called Synthetic DNA 29 MER.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	Z	29	Total	C	N	O	P	0	0	0
			590	280	104	177	29			

- Molecule 3 is a protein called Putative endonuclease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	148	Total	C	N	O	S	0	0	0
			1201	767	215	216	3			
3	B	148	Total	C	N	O	S	0	0	0
			1194	763	212	217	2			
3	C	138	Total	C	N	O	S	0	0	0
			1107	710	191	203	3			
3	D	133	Total	C	N	O	S	0	0	0
			1038	669	170	197	2			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	INITIATING METHIONINE	UNP Q57253
A	1	GLY	-	CLONING ARTIFACT	UNP Q57253
A	11	GLN	GLU	ENGINEERED	UNP Q57253
A	16	MET	LEU	ENGINEERED	UNP Q57253
A	21	MET	LEU	ENGINEERED	UNP Q57253
A	55	LYS	PHE	ENGINEERED	UNP Q57253
B	0	MET	-	INITIATING METHIONINE	UNP Q57253
B	1	GLY	-	CLONING ARTIFACT	UNP Q57253

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	11	GLN	GLU	ENGINEERED	UNP Q57253
B	16	MET	LEU	ENGINEERED	UNP Q57253
B	21	MET	LEU	ENGINEERED	UNP Q57253
B	55	LYS	PHE	ENGINEERED	UNP Q57253
C	0	MET	-	INITIATING METHIONINE	UNP Q57253
C	1	GLY	-	CLONING ARTIFACT	UNP Q57253
C	11	GLN	GLU	ENGINEERED	UNP Q57253
C	16	MET	LEU	ENGINEERED	UNP Q57253
C	21	MET	LEU	ENGINEERED	UNP Q57253
C	55	LYS	PHE	ENGINEERED	UNP Q57253
D	0	MET	-	INITIATING METHIONINE	UNP Q57253
D	1	GLY	-	CLONING ARTIFACT	UNP Q57253
D	11	GLN	GLU	ENGINEERED	UNP Q57253
D	16	MET	LEU	ENGINEERED	UNP Q57253
D	21	MET	LEU	ENGINEERED	UNP Q57253
D	55	LYS	PHE	ENGINEERED	UNP Q57253

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	Y	2	Total O 2 2	0	0
5	Z	3	Total O 3 3	0	0
5	A	2	Total O 2 2	0	0
5	B	4	Total O 4 4	0	0
5	C	1	Total O 1 1	0	0
5	D	4	Total O 4 4	0	0

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

These plots are drawn for all protein, RNA and DNA 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.

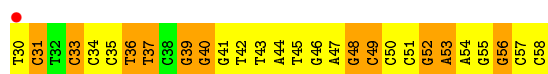
- Molecule 1: Synthetic DNA 29 MER

Chain Y: 




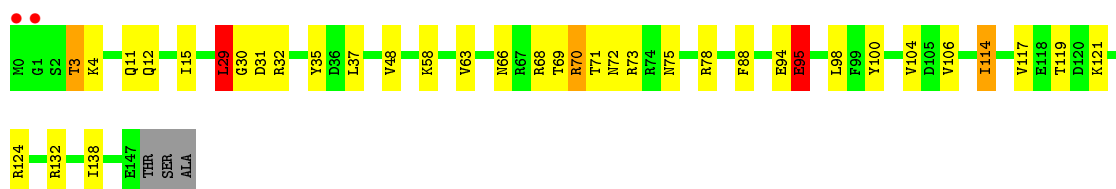
- Molecule 2: Synthetic DNA 29 MER

Chain Z: 




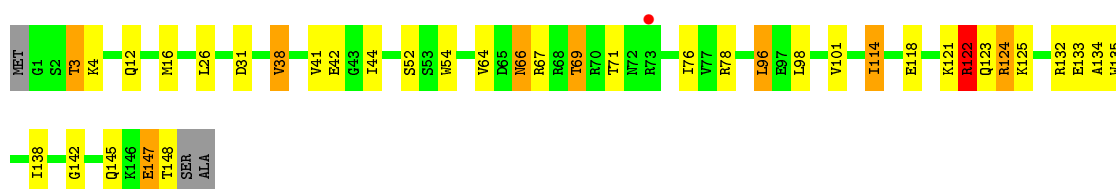
- Molecule 3: Putative endonuclease

Chain A: 



- Molecule 3: Putative endonuclease

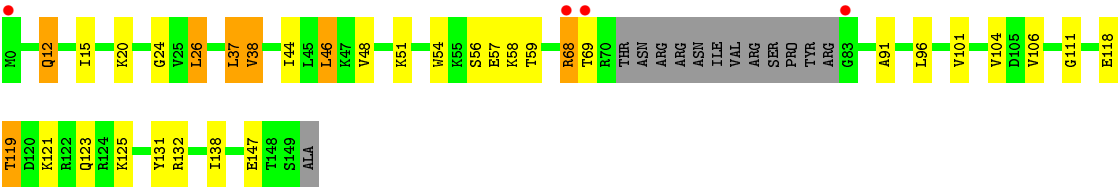
Chain B: 



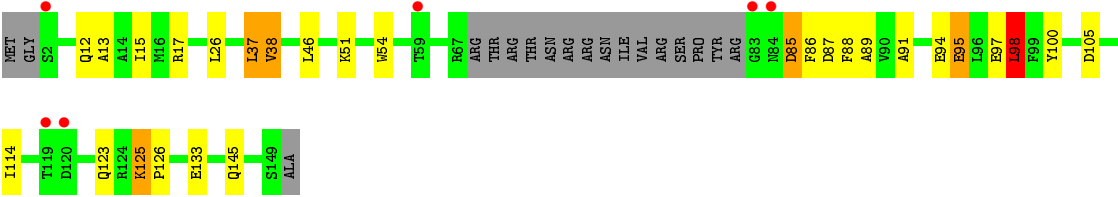
- Molecule 3: Putative endonuclease

Chain C: 





● Molecule 3: Putative endonuclease



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	143.78Å 143.78Å 319.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	160.13 – 3.10 48.44 – 3.10	Depositor EDS
% Data completeness (in resolution range)	94.0 (160.13-3.10) 94.0 (48.44-3.10)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.67 (at 3.12Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.278 , 0.328 0.261 , 0.302	Depositor DCC
$R_{free}$ test set	1459 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	73.4	Xtriage
Anisotropy	0.739	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 37.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	5747	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	Y	0.72	0/673	1.96	40/1037 (3.9%)
2	Z	0.71	0/659	2.27	50/1014 (4.9%)
3	A	0.34	0/1226	0.55	0/1652
3	B	0.37	0/1219	0.52	0/1645
3	C	0.36	0/1129	0.56	0/1520
3	D	0.38	0/1060	0.54	1/1437 (0.1%)
All	All	0.47	0/5966	1.15	91/8305 (1.1%)

There are no bond length outliers.

The worst 5 of 91 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	Z	52	DG	O3'-P-O5'	-12.35	80.55	104.00
2	Z	39	DG	O3'-P-O5'	-11.52	82.11	104.00
1	Y	3	DG	O3'-P-O5'	-10.97	83.16	104.00
2	Z	34	DC	O3'-P-O5'	-10.82	83.45	104.00
1	Y	21	DC	O3'-P-O5'	-10.73	83.61	104.00

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	Y	599	0	324	11	0
2	Z	590	0	327	19	0
3	A	1201	0	1202	24	0
3	B	1194	0	1189	22	0
3	C	1107	0	1101	17	0
3	D	1038	0	987	13	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	2	0	0	2	0
5	B	4	0	0	0	0
5	C	1	0	0	1	0
5	D	4	0	0	2	0
5	Y	2	0	0	1	0
5	Z	3	0	0	2	0
All	All	5747	0	5130	93	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 9.

The worst 5 of 93 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)
2:Z:53:DA:H2''	2:Z:54:DA:H5''	1.45	0.95
3:C:58:LYS:HB2	3:C:59:THR:HA	1.53	0.88
3:A:12:GLN:HE22	3:D:12:GLN:HE21	1.27	0.82
3:D:86:PHE:HE2	3:D:89:ALA:HB2	1.44	0.81
1:Y:8:DT:H2'	1:Y:9:DC:C6	2.17	0.80

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	A	146/151 (97%)	126 (86%)	16 (11%)	4 (3%)	6	30
3	B	146/151 (97%)	130 (89%)	12 (8%)	4 (3%)	6	30
3	C	134/151 (89%)	118 (88%)	13 (10%)	3 (2%)	8	35
3	D	129/151 (85%)	110 (85%)	15 (12%)	4 (3%)	5	26
All	All	555/604 (92%)	484 (87%)	56 (10%)	15 (3%)	6	30

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	29	LEU
3	B	3	THR
3	C	119	THR
3	D	87	ASP
3	A	3	THR

### 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	
3	A	126/130 (97%)	110 (87%)	16 (13%)	5	21
3	B	125/130 (96%)	115 (92%)	10 (8%)	14	47
3	C	116/130 (89%)	101 (87%)	15 (13%)	5	21
3	D	106/130 (82%)	93 (88%)	13 (12%)	5	23
All	All	473/520 (91%)	419 (89%)	54 (11%)	7	27

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	145	GLN
3	C	38	VAL
3	D	114	ILE
3	B	147	GLU
3	C	20	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	115	HIS
3	B	145	GLN
3	C	145	GLN
3	B	66	ASN
3	C	140	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 carbohydrates 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 ⓘ

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	Y	29/29 (100%)	-0.35	0 <span>100</span> <span>100</span>	62, 78, 101, 112	0
2	Z	29/29 (100%)	-0.17	1 (3%) <span>46</span> <span>23</span>	53, 77, 143, 159	0
3	A	148/151 (98%)	0.24	2 (1%) <span>75</span> <span>57</span>	66, 82, 100, 101	0
3	B	148/151 (98%)	0.08	1 (0%) <span>87</span> <span>75</span>	67, 84, 102, 105	0
3	C	138/151 (91%)	0.20	4 (2%) <span>52</span> <span>28</span>	64, 83, 102, 109	0
3	D	133/151 (88%)	0.25	6 (4%) <span>34</span> <span>16</span>	65, 84, 102, 104	0
All	All	625/662 (94%)	0.15	14 (2%) <span>62</span> <span>41</span>	53, 83, 102, 159	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	Z	30	DT	6.1
3	A	0	MET	5.9
3	A	1	GLY	3.6
3	D	2	SER	3.0
3	C	83	GLY	2.8

### 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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	CA	A	501	1/1	0.96	0.29	0.95	54,54,54,54	0
4	CA	B	502	1/1	0.95	0.25	-0.07	59,59,59,59	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.