



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 11:42 PM GMT

PDB ID : 1YDX  
Title : Crystal structure of Type-I restriction-modification system S subunit from *M. genitalium*  
Authors : Machado, B.; Quijada, O.; Pinol, J.; Fita, I.; Querol, E.; Carpena, X.  
Deposited on : 2004-12-27  
Resolution : 2.30 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

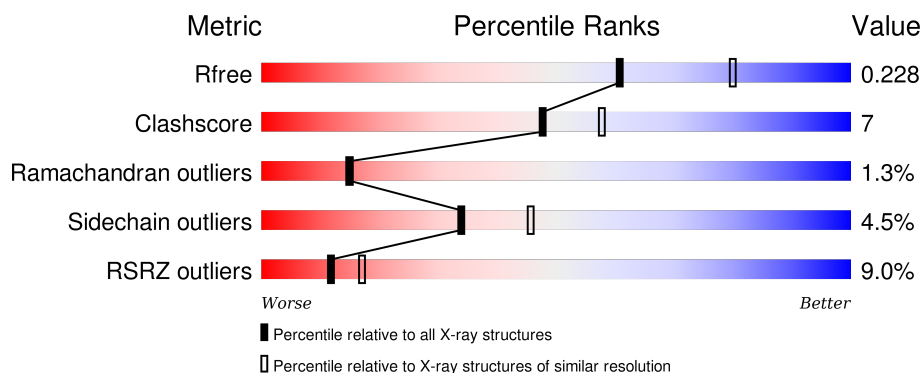
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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))
$R_{free}$	91344	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (2.30-2.30)

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	A	406	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3163 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 type I restriction enzyme specificity protein MG438.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	374	3032	1950	500	572	4	6	0	0	0

There are 29 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-22	MSE	-	CLONING ARTIFACT	UNP Q49434
A	-21	GLY	-	CLONING ARTIFACT	UNP Q49434
A	-20	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-19	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-18	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-17	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-16	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-15	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-14	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-13	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-12	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-11	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-10	SER	-	CLONING ARTIFACT	UNP Q49434
A	-9	SER	-	CLONING ARTIFACT	UNP Q49434
A	-8	GLY	-	CLONING ARTIFACT	UNP Q49434
A	-7	HIS	-	CLONING ARTIFACT	UNP Q49434
A	-6	ILE	-	CLONING ARTIFACT	UNP Q49434
A	-5	ASP	-	CLONING ARTIFACT	UNP Q49434
A	-4	ASP	-	CLONING ARTIFACT	UNP Q49434
A	-3	ASP	-	CLONING ARTIFACT	UNP Q49434
A	-2	ASP	-	CLONING ARTIFACT	UNP Q49434
A	-1	LYS	-	CLONING ARTIFACT	UNP Q49434
A	0	HIS	-	CLONING ARTIFACT	UNP Q49434
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q49434
A	29	MSE	MET	MODIFIED RESIDUE	UNP Q49434
A	201	MSE	MET	MODIFIED RESIDUE	UNP Q49434
A	283	MSE	MET	MODIFIED RESIDUE	UNP Q49434

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	307	MSE	MET	MODIFIED RESIDUE	UNP Q49434
A	373	MSE	MET	MODIFIED RESIDUE	UNP Q49434

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total Cl 2 2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	129	Total O 129 129	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	76.60Å 76.60Å 174.85Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 2.30 24.82 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.8 (20.00-2.30) 99.6 (24.82-2.20)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.96 (at 2.19Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.197 , 0.231 0.194 , 0.228	Depositor DCC
$R_{free}$ test set	1386 reflections (5.39%)	DCC
Wilson B-factor (Å <sup>2</sup> )	43.6	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 57.3	EDS
Estimated twinning fraction	0.033 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 30971 reflections (0.003%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3163	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.11% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: CL

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	1.07	10/3088 (0.3%)	0.92	11/4149 (0.3%)

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	153	LEU	C-N	16.68	1.72	1.34
1	A	223	ARG	CZ-NH1	13.81	1.51	1.33
1	A	223	ARG	NE-CZ	13.40	1.50	1.33
1	A	223	ARG	CZ-NH2	12.73	1.49	1.33
1	A	223	ARG	CD-NE	9.90	1.63	1.46

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	223	ARG	NE-CZ-NH2	-23.69	108.45	120.30
1	A	223	ARG	NE-CZ-NH1	12.61	126.61	120.30
1	A	153	LEU	O-C-N	-11.22	104.75	122.70
1	A	76	ARG	NE-CZ-NH2	-9.85	115.38	120.30
1	A	76	ARG	NE-CZ-NH1	9.77	125.18	120.30

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	3032	0	3051	45	0
2	A	2	0	0	0	0
3	A	129	0	0	2	0
All	All	3163	0	3051	45	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 7.

The worst 5 of 45 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:153:LEU:C	1:A:154:SER:N	1.72	1.41
1:A:217:TYR:HB2	1:A:275:THR:HG23	1.68	0.75
1:A:34:LEU:O	1:A:54:THR:HG22	1.88	0.73
1:A:1:MSE:HE1	1:A:373:MSE:HG2	1.74	0.69
1:A:265:ARG:HD2	1:A:266:TYR:H	1.57	0.67

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	372/406 (92%)	363 (98%)	4 (1%)	5 (1%)	15 15

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	234	ALA
1	A	310	ASP
1	A	313	PRO
1	A	217	TYR
1	A	312	SER



### 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	335/358 (94%)	320 (96%)	15 (4%)	34	46

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	222	GLU
1	A	224	LEU
1	A	265	ARG
1	A	199	SER
1	A	263	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	238	ASN
1	A	273	ASN
1	A	285	ASN
1	A	350	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

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	368/406 (90%)	0.49	33 (8%) 12 17	31, 40, 48, 61	0

The worst 5 of 33 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	312	SER	10.5
1	A	310	ASP	4.3
1	A	311	ASP	4.3
1	A	374	THR	4.0
1	A	313	PRO	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 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(Å <sup>2</sup> )	Q<0.9
2	CL	A	384	1/1	0.96	0.16	0.04	43,43,43,43	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	CL	A	385	1/1	0.97	0.26	-	50,50,50,50	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.