



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 7, 2016 – 07:32 PM EST

PDB ID : 5I97  
Title : Structural analysis and inhibition of TraE from the pKM101 type IV secretion system  
Authors : Casu, B.; Sygusch, J.; Baron, C.  
Deposited on : 2016-02-19  
Resolution : 2.44 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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028320
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)	:	rb-20028320

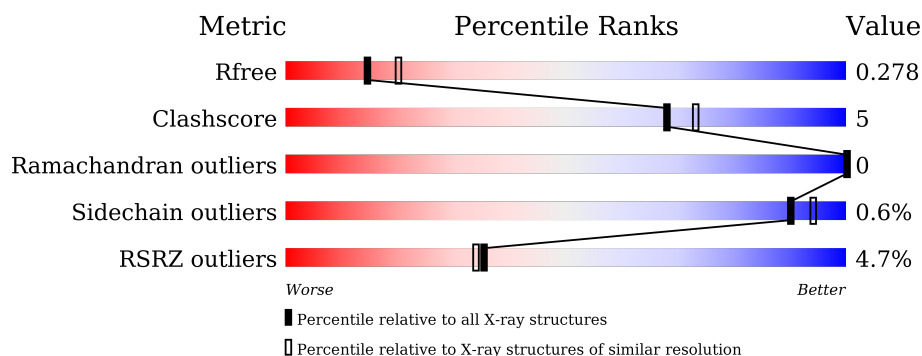
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.44 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	1003 (2.46-2.42)
Clashscore	102246	1071 (2.46-2.42)
Ramachandran outliers	100387	1065 (2.46-2.42)
Sidechain outliers	100360	1065 (2.46-2.42)
RSRZ outliers	91569	1005 (2.46-2.42)

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	163	<div> <div>2%</div> <div> <div></div> <div>82%</div> <div>6%</div> <div>12%</div> </div> </div>
1	B	163	<div> <div>0%</div> <div> <div></div> <div>84%</div> <div>•</div> <div>13%</div> </div> </div>
1	C	163	<div> <div>5%</div> <div> <div></div> <div>69%</div> <div>15%</div> <div>16%</div> </div> </div>
1	D	163	<div> <div>8%</div> <div> <div></div> <div>71%</div> <div>12%</div> <div>•</div> <div>16%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4636 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 Conjugal transfer protein.

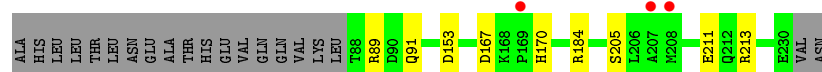
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	143	Total	C	N	O	S	0	0	0
			1168	736	206	223	3			
1	B	142	Total	C	N	O	S	0	0	0
			1161	732	205	221	3			
1	C	137	Total	C	N	O	S	0	0	0
			1120	710	196	211	3			
1	D	137	Total	C	N	O	S	0	0	0
			1120	710	196	211	3			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	28	Total	O	0	0
			28	28		
2	B	30	Total	O	0	0
			30	30		
2	C	5	Total	O	0	0
			5	5		
2	D	4	Total	O	0	0
			4	4		

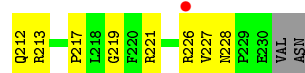


- Molecule 1: Conjugal transfer protein



ALA	HIS	LEU	THR	LEU	ASN	GLU	ALA	THR	HIS	GLU	VAL	GLN	GLN	VAL	LYS	LEU	THR	R89	K100	V164	D167	H170	L206	A207	R213	E230	VAL	ASN
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	-----	-----

ALA	HIS	LIEU	TRR	LIEU	ASN	GLU	ALA	TRR	HIS	GLI	VAL	GLN	GLM	VAL	LVS	LIEU	TRR	ARG	ASP	GLM	TRR	SER	TYR	R110	D121	Y122	V125	Y137	Q138	S139	K142	G143	R144	I175	R176	R182	V188	D189	I196	M199	S205	A207	N208	Z209	L210
-----	-----	------	-----	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------



ALA	HIS	LEU	THR	LEU	ASN	GLU	ALA	THR	HIS	GLU	VAL	GLN	VAL	LYS	LEU	THR	ARG	ASP	GLN	THR	THR	SER	Y94	L103	Y116	A134	Q138	K139	K140	F141	G143	R144	N145	G146	L147	I175	R176	R181	R182	M187	P188	V188	D189	R194	S205	L206	L207
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	112.05Å 123.36Å 109.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.60 – 2.44 41.01 – 2.44	Depositor EDS
% Data completeness (in resolution range)	96.5 (38.60-2.44) 95.8 (41.01-2.44)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.25 (at 2.45Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.240 , 0.280 0.245 , 0.278	Depositor DCC
$R_{free}$ test set	1905 reflections (6.96%)	DCC
Wilson B-factor (Å <sup>2</sup> )	46.5	Xtriage
Anisotropy	0.454	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 20.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	4636	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.49% 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](#)

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	0.39	0/1194	0.54	0/1619
1	B	0.39	0/1187	0.54	0/1609
1	C	0.48	0/1146	0.60	0/1554
1	D	0.39	0/1146	0.58	0/1554
All	All	0.41	0/4673	0.57	0/6336

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	B	0	1
1	C	0	1
1	D	0	1
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	213	ARG	Sidechain
1	C	176	ARG	Sidechain
1	D	181	ARG	Sidechain

### 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	A	1168	0	1139	7	0
1	B	1161	0	1132	2	0
1	C	1120	0	1095	22	0
1	D	1120	0	1095	12	0
2	A	28	0	0	1	0
2	B	30	0	0	0	0
2	C	5	0	0	0	0
2	D	4	0	0	0	0
All	All	4636	0	4461	42	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 5.

The worst 5 of 42 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:C:226:ARG:HH22	1:C:228:ASN:ND2	1.70	0.88
1:C:209:ASN:HD21	1:C:212:GLN:HB2	1.47	0.79
1:C:209:ASN:OD1	1:C:212:GLN:N	2.17	0.76
1:C:213:ARG:O	1:C:217:PRO:HG3	1.92	0.70
1:C:226:ARG:HH22	1:C:228:ASN:HD22	1.40	0.68

There are no symmetry-related clashes.

## 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	
1	A	141/163 (86%)	138 (98%)	3 (2%)	0	100	100
1	B	140/163 (86%)	138 (99%)	2 (1%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	135/163 (83%)	133 (98%)	2 (2%)	0	100	100
1	D	135/163 (83%)	133 (98%)	2 (2%)	0	100	100
All	All	551/652 (84%)	542 (98%)	9 (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	129/147 (88%)	129 (100%)	0	100	100
1	B	128/147 (87%)	128 (100%)	0	100	100
1	C	123/147 (84%)	122 (99%)	1 (1%)	86	92
1	D	123/147 (84%)	121 (98%)	2 (2%)	70	82
All	All	503/588 (86%)	500 (99%)	3 (1%)	90	94

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	139	SER
1	D	103	LEU
1	D	206	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	91	GLN
1	C	228	ASN

### 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](#)

There are no ligands in this entry.

## 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, 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	143/163 (87%)	0.18	3 (2%) 67 67	22, 29, 41, 52	0
1	B	142/163 (87%)	0.12	2 (1%) 78 79	22, 29, 41, 50	0
1	C	137/163 (84%)	0.37	8 (5%) 26 24	31, 43, 58, 69	0
1	D	137/163 (84%)	0.59	13 (9%) 10 9	28, 44, 69, 78	0
All	All	559/652 (85%)	0.31	26 (4%) 35 34	22, 36, 59, 78	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	207	ALA	12.1
1	D	207	ALA	11.7
1	C	206	LEU	9.4
1	C	210	ALA	8.3
1	D	206	LEU	6.2

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.