



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:02 AM GMT

PDB ID : 2P5Z  
Title : The E. coli c3393 protein is a component of the type VI secretion system and exhibits structural similarity to T4 bacteriophage tail proteins gp27 and gp5  
Authors : Ramagopal, U.A.; Bonanno, J.B.; Sridhar, V.; Lau, C.; Toro, R.; Gheyi, T.; Maletic, M.; Freeman, J.C.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2007-03-16  
Resolution : 2.60 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

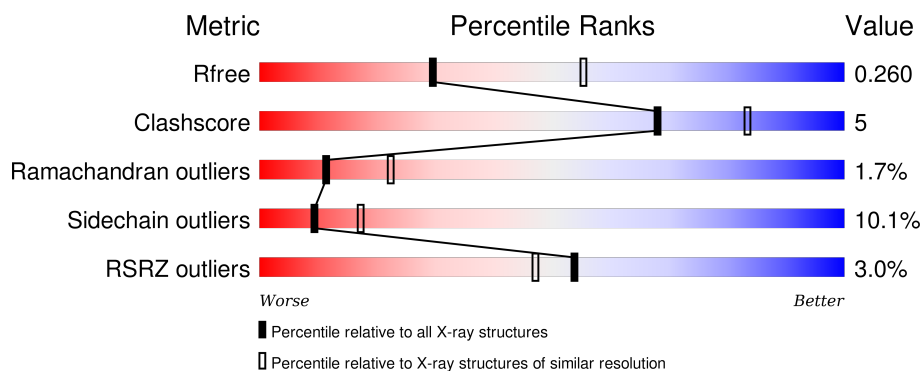
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*


The reported resolution of this entry is 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	X	491	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2962 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 VI secretion system component.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	X	361	Total	C	N	O	S	0	1	0
			2900	1842	508	544	6			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
X	-1	MET	-	CLONING ARTIFACT	UNP Q8FED2
X	0	SER	-	CLONING ARTIFACT	UNP Q8FED2
X	1	LEU	-	CLONING ARTIFACT	UNP Q8FED2
X	482	GLU	-	CLONING ARTIFACT	UNP Q8FED2
X	483	GLY	-	CLONING ARTIFACT	UNP Q8FED2
X	484	HIS	-	CLONING ARTIFACT	UNP Q8FED2
X	485	HIS	-	CLONING ARTIFACT	UNP Q8FED2
X	486	HIS	-	CLONING ARTIFACT	UNP Q8FED2
X	487	HIS	-	CLONING ARTIFACT	UNP Q8FED2
X	488	HIS	-	CLONING ARTIFACT	UNP Q8FED2
X	489	HIS	-	CLONING ARTIFACT	UNP Q8FED2

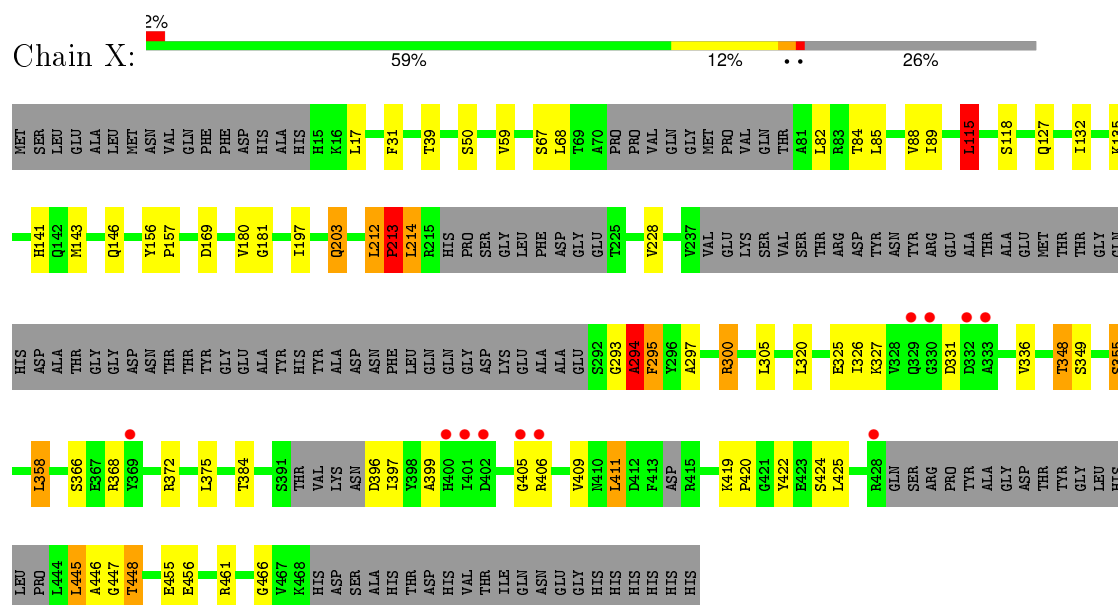
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	X	62	Total	O	0	0
			62	62		

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Type VI secretion system component



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	116.35Å 116.35Å 80.58Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.19 – 2.60 47.17 – 2.61	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.19-2.60) 99.9 (47.17-2.61)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.31 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.217 , 0.264 0.215 , 0.260	Depositor DCC
$R_{free}$ test set	983 reflections (5.43%)	DCC
Wilson B-factor (Å <sup>2</sup> )	61.4	Xtriage
Anisotropy	0.172	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 50.5	EDS
Estimated twinning fraction	0.046 for h,-h-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 19100 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2962	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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

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	X	0.52	0/2958	0.85	10/3998 (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	X	0	4

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	X	212	LEU	C-N-CD	-12.72	92.62	120.60
1	X	212	LEU	C-N-CA	8.23	156.58	122.00
1	X	358	LEU	CA-CB-CG	7.71	133.04	115.30
1	X	411	LEU	CA-CB-CG	6.80	130.95	115.30
1	X	115	LEU	CA-CB-CG	5.97	129.02	115.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	X	212	LEU	Peptide
1	X	213	PRO	Peptide
1	X	294	ALA	Peptide
1	X	355	SER	Peptide

## 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	X	2900	0	2858	27	0
2	X	62	0	0	1	0
All	All	2962	0	2858	27	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 27 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:X:135:LYS:NZ	1:X:422:TYR:OH	2.23	0.69
1:X:293:GLY:HA3	1:X:294:ALA:HB3	1.87	0.56
1:X:293:GLY:CA	1:X:294:ALA:HB3	2.39	0.52
1:X:181:GLY:O	1:X:372:ARG:NH1	2.45	0.49
1:X:213:PRO:O	1:X:214:LEU:HB2	2.11	0.49

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	X	348/491 (71%)	320 (92%)	22 (6%)	6 (2%)	11 22

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	X	213	PRO
1	X	214	LEU
1	X	294	ALA
1	X	399	ALA
1	X	295	PHE

### 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
1	X	307/417 (74%)	276 (90%)	31 (10%)	9 17

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	X	300	ARG
1	X	327	LYS
1	X	424	SER
1	X	305	LEU
1	X	331	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	X	308	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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	X	361/491 (73%)	0.30	11 (3%) 54 47	34, 55, 95, 112	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	X	333	ALA	4.3
1	X	330	GLY	3.2
1	X	332	ASP	2.7
1	X	400	HIS	2.7
1	X	428	ARG	2.6

### 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.