



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 09:07 am GMT

PDB ID : 3WOE  
Title : Crystal structure of P23-45 gp39 (6-109) bound to Thermus thermophilus RNA polymerase beta-flap domain  
Authors : Tagami, S.; Sekine, S.; Minakhin, L.; Esyunina, D.; Akasaka, R.; Shirouzu, M.; Kulbachinskiy, A.; Severinov, K.; Yokoyama, S.  
Deposited on : 2013-12-26  
Resolution : 2.35 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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	:	trunk28620
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)	:	recalc28949

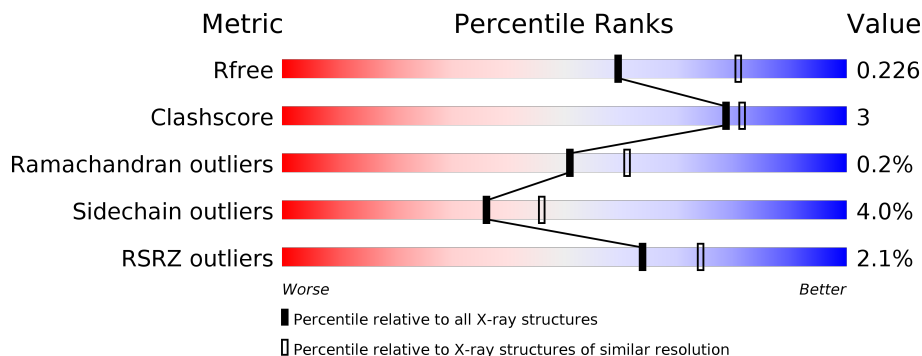
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.35 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	1522 (2.38-2.34)
Clashscore	112137	1626 (2.38-2.34)
Ramachandran outliers	110173	1605 (2.38-2.34)
Sidechain outliers	110143	1606 (2.38-2.34)
RSRZ outliers	101464	1528 (2.38-2.34)

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	130	<div> <div>2%</div> <div> <div></div> <div>85%</div> <div>12%</div> <div>.</div> </div> </div>
1	C	130	<div> <div>3%</div> <div> <div></div> <div>91%</div> <div>8%</div> <div>.</div> </div> </div>
2	B	106	<div> <div>%</div> <div> <div></div> <div>89%</div> <div>8%</div> <div>..</div> </div> </div>
2	D	106	<div> <div>2%</div> <div> <div></div> <div>88%</div> <div>9%</div> <div>..</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3831 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 DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	130	Total	C	N	O	0	0	0
			1018	633	193	192			
1	C	130	Total	C	N	O	0	0	0
			1018	633	193	192			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	701	GLY	-	EXPRESSION TAG	UNP Q8RQE9
A	702	PRO	-	EXPRESSION TAG	UNP Q8RQE9
C	701	GLY	-	EXPRESSION TAG	UNP Q8RQE9
C	702	PRO	-	EXPRESSION TAG	UNP Q8RQE9

- Molecule 2 is a protein called Putative uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	104	Total	C	N	O	Se	0	0	0
			847	553	143	150	1			
2	D	104	Total	C	N	O	Se	0	0	0
			847	553	143	150	1			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	4	GLY	-	EXPRESSION TAG	UNP A7XX65
B	5	PRO	-	EXPRESSION TAG	UNP A7XX65
D	4	GLY	-	EXPRESSION TAG	UNP A7XX65
D	5	PRO	-	EXPRESSION TAG	UNP A7XX65

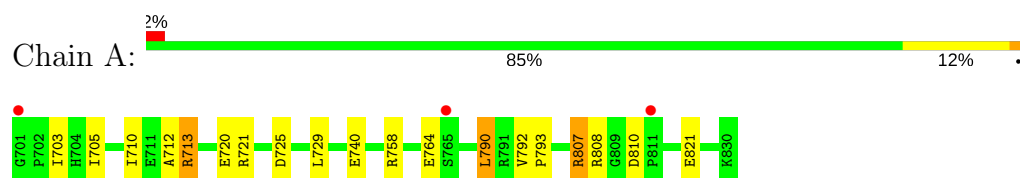
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	32	Total 32	O 32	0	0
3	B	20	Total 20	O 20	0	0
3	C	27	Total 27	O 27	0	0
3	D	22	Total 22	O 22	0	0

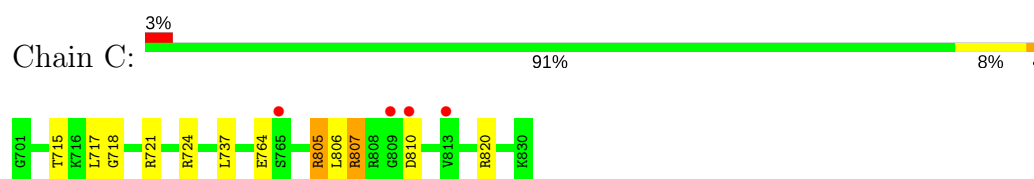
### 3 Residue-property plots [i](#)

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.

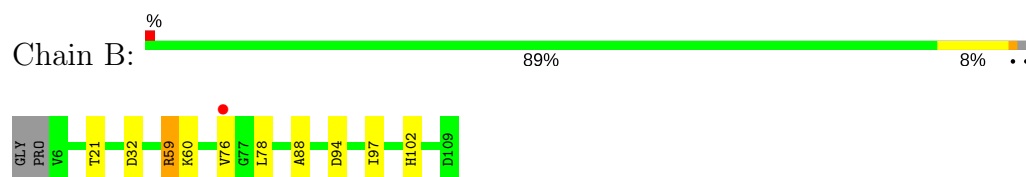
- Molecule 1: DNA-directed RNA polymerase subunit beta



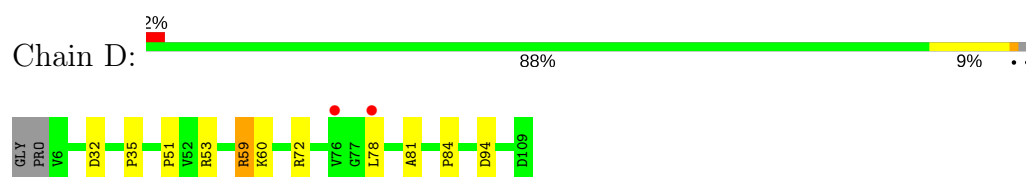
- Molecule 1: DNA-directed RNA polymerase subunit beta



- Molecule 2: Putative uncharacterized protein



- Molecule 2: Putative uncharacterized protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.23Å 99.23Å 117.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.62 – 2.35 49.62 – 2.35	Depositor EDS
% Data completeness (in resolution range)	99.6 (49.62-2.35) 99.5 (49.62-2.35)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.14	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.77 (at 2.34Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.4_1496)	Depositor
R, $R_{free}$	0.181 , 0.225 0.180 , 0.226	Depositor DCC
$R_{free}$ test set	2377 reflections (5.13%)	DCC
Wilson B-factor (Å <sup>2</sup> )	31.8	Xtriage
Anisotropy	0.716	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 39.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3831	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 19.90 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 9.7850e-03.*

<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	0/1033	0.56	0/1393
1	C	0.36	0/1033	0.55	0/1393
2	B	0.38	0/872	0.55	0/1189
2	D	0.38	0/872	0.53	0/1189
All	All	0.38	0/3810	0.55	0/5164

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 [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	1018	0	1050	14	0
1	C	1018	0	1050	7	0
2	B	847	0	839	4	0
2	D	847	0	839	5	0
3	A	32	0	0	1	0
3	B	20	0	0	0	0
3	C	27	0	0	1	0
3	D	22	0	0	0	0
All	All	3831	0	3778	25	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 3.

All (25) 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:820:ARG:NH1	3:C:926:HOH:O	2.08	0.87
1:A:810:ASP:OD2	3:A:927:HOH:O	2.14	0.65
1:C:724:ARG:NH1	1:C:737:LEU:O	2.33	0.61
2:D:35:PRO:HG3	2:D:51:PRO:HA	1.85	0.59
1:A:703:ILE:HG22	2:D:81:ALA:HA	1.89	0.53
1:A:740:GLU:OE2	1:A:807:ARG:NH2	2.38	0.51
1:A:712:ALA:HB3	1:A:821:GLU:HG3	1.95	0.49
1:A:721:ARG:HH22	2:B:94:ASP:CG	2.18	0.47
1:C:721:ARG:HH22	2:D:94:ASP:CG	2.18	0.47
1:C:807:ARG:HG3	1:C:810:ASP:OD2	2.15	0.47
1:A:713:ARG:HG2	1:A:720:GLU:OE1	2.15	0.46
2:D:59:ARG:HD3	2:D:60:LYS:O	2.16	0.46
1:A:705:ILE:O	2:D:84:PRO:HD3	2.17	0.45
1:A:713:ARG:HH21	1:A:758:ARG:HH12	1.65	0.43
1:A:713:ARG:HH21	1:A:758:ARG:NH1	2.16	0.43
1:C:805:ARG:O	1:C:806:LEU:HD23	2.18	0.43
1:A:710:ILE:HB	1:A:790:LEU:HG	2.01	0.43
1:C:715:THR:HB	1:C:718:GLY:O	2.19	0.43
2:B:59:ARG:HD3	2:B:60:LYS:O	2.19	0.42
1:C:715:THR:HG22	1:C:717:LEU:H	1.85	0.41
1:A:808:ARG:HH11	1:A:808:ARG:HG2	1.84	0.41
1:A:725:ASP:OD1	2:B:97:ILE:HB	2.20	0.41
1:A:729:LEU:HD23	1:A:729:LEU:HA	1.88	0.41
1:A:792:VAL:HA	1:A:793:PRO:HD3	1.91	0.41
2:B:88:ALA:HA	2:B:102:HIS:CE1	2.56	0.40

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	128/130 (98%)	127 (99%)	1 (1%)	0	100	100
1	C	128/130 (98%)	126 (98%)	2 (2%)	0	100	100
2	B	102/106 (96%)	99 (97%)	2 (2%)	1 (1%)	18	18
2	D	102/106 (96%)	100 (98%)	2 (2%)	0	100	100
All	All	460/472 (98%)	452 (98%)	7 (2%)	1 (0%)	51	61

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	76	VAL

### 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	A	110/110 (100%)	106 (96%)	4 (4%)	40	51
1	C	110/110 (100%)	107 (97%)	3 (3%)	50	63
2	B	88/88 (100%)	84 (96%)	4 (4%)	32	40
2	D	88/88 (100%)	83 (94%)	5 (6%)	24	28
All	All	396/396 (100%)	380 (96%)	16 (4%)	36	46

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	713	ARG
1	A	764	GLU
1	A	790	LEU
1	A	807	ARG
2	B	21	THR
2	B	32	ASP
2	B	59	ARG
2	B	78	LEU
1	C	764	GLU
1	C	805	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	807	ARG
2	D	32	ASP
2	D	53	ARG
2	D	59	ARG
2	D	72	ARG
2	D	78	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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	130/130 (100%)	-0.14	3 (2%) 61 71	20, 36, 71, 83	0
1	C	130/130 (100%)	-0.10	4 (3%) 49 61	19, 36, 71, 83	0
2	B	103/106 (97%)	-0.07	1 (0%) 82 90	19, 32, 66, 80	0
2	D	103/106 (97%)	-0.12	2 (1%) 67 76	15, 32, 65, 91	0
All	All	466/472 (98%)	-0.11	10 (2%) 64 74	15, 33, 71, 91	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	813	VAL	3.9
2	D	76	VAL	3.7
1	C	809	GLY	3.2
1	C	765	SER	2.9
1	A	811	PRO	2.8
1	A	701	GLY	2.4
2	B	76	VAL	2.2
1	A	765	SER	2.2
1	C	810	ASP	2.2
2	D	78	LEU	2.1

### 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.