



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 25, 2023 – 10:22 AM EDT

PDB ID : 5VI5  
Title : Structure of Mycobacterium smegmatis transcription initiation complex with a full transcription bubble  
Authors : Darst, S.A.; Campbell, E.A.; Lilic, M.  
Deposited on : 2017-04-14  
Resolution : 3.20 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35.1  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.35.1

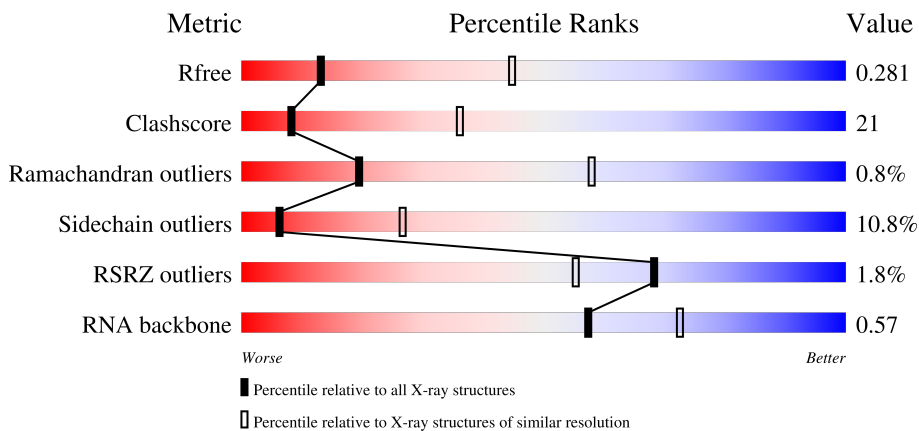
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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))
$R_{free}$	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)
RNA backbone	3102	1010 (3.50-2.90)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	O	50	 2% 22% 76%
2	P	50	 2% 40% 48% 12%
3	Q	4	 50% 50%
4	A	350	 2% 37% 21% 37%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	B	350	
5	C	1169	
6	D	1317	
7	E	107	
8	F	466	
9	J	114	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
10	SO4	C	1201	-	-	X	X
10	SO4	F	502	-	-	X	-

## 2 Entry composition [i](#)

There are 13 unique types of molecules in this entry. The entry contains 26574 atoms, of which 16 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 DNA (49-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	O	49	1009	481	188	292	48	0	0	0

- Molecule 2 is a DNA chain called DNA (44-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	P	44	903	430	167	262	44	0	0	0

- Molecule 3 is a RNA chain called RNA (5'-R(\*UP\*CP\*GP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	Q	4	82	38	15	26	3	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	A	220	1617	1022	278	315	2	0	0	0
4	B	233	1660	1050	287	321	2	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	C	1113	8138	5103	1427	1573	35	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	238	ASN	GLN	conflict	UNP P60281

- Molecule 6 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	D	1238	9391	5885	1699	1767	40	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	663	GLU	ALA	conflict	UNP A0QS66
D	1272	ASN	GLN	conflict	UNP A0QS66

- Molecule 7 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
7	E	76	592	378	100	114	0	0	0

- Molecule 8 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	F	314	2461	1542	442	470	7	0	0	0

- Molecule 9 is a protein called RNA polymerase-binding protein RbpA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	J	86	663	414	122	125	2	0	0	0

- Molecule 10 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
10	C	1	Total O S 5 4 1	0	0
10	D	1	Total O S 5 4 1	0	0
10	D	1	Total O S 5 4 1	0	0
10	D	1	Total O S 5 4 1	0	0
10	F	1	Total O S 5 4 1	0	0
10	F	1	Total O S 5 4 1	0	0

- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
11	D	2	Total Zn 2 2	0	0

- Molecule 12 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
12	D	1	Total	C	H	O	0	0
			10	2	6	2		
12	F	1	Total	C	H	O	0	0
			10	2	6	2		

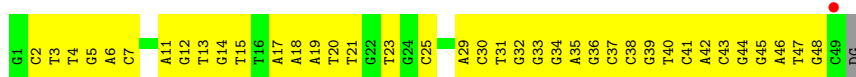
- Molecule 13 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
13	C	1	Total	H	O	0	0
			3	2	1		
13	D	1	Total	H	O	0	0
			3	2	1		

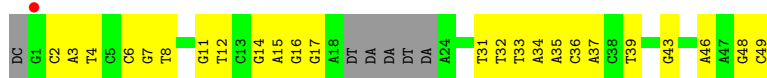
### 3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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 (49-MER)



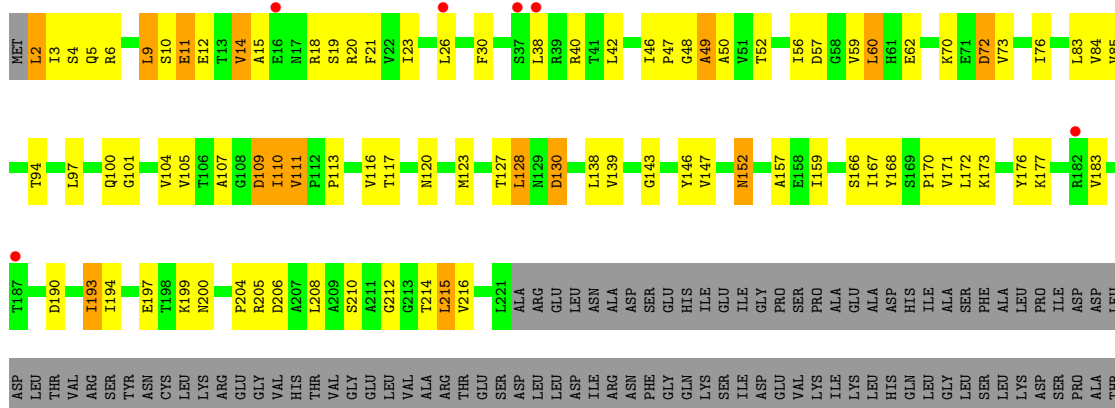
- Molecule 2: DNA (44-MER)



- Molecule 3: RNA (5'-R(\*UP\*CP\*GP\*A)-3')

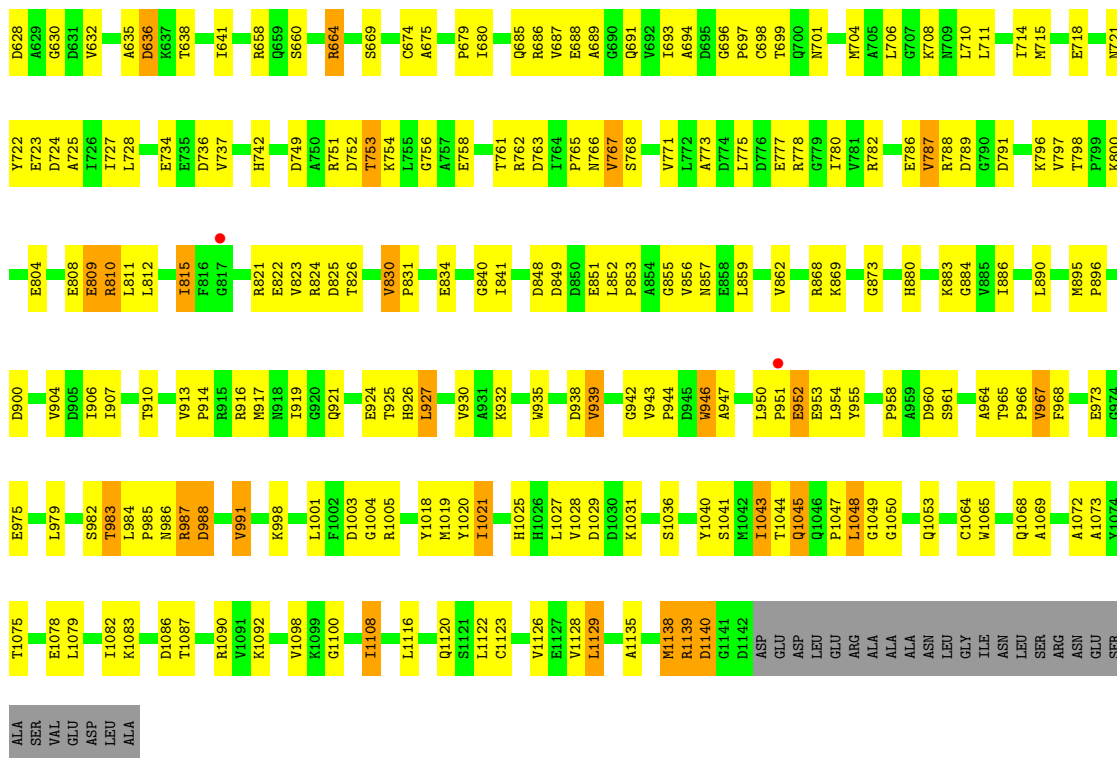


- Molecule 4: DNA-directed RNA polymerase subunit alpha

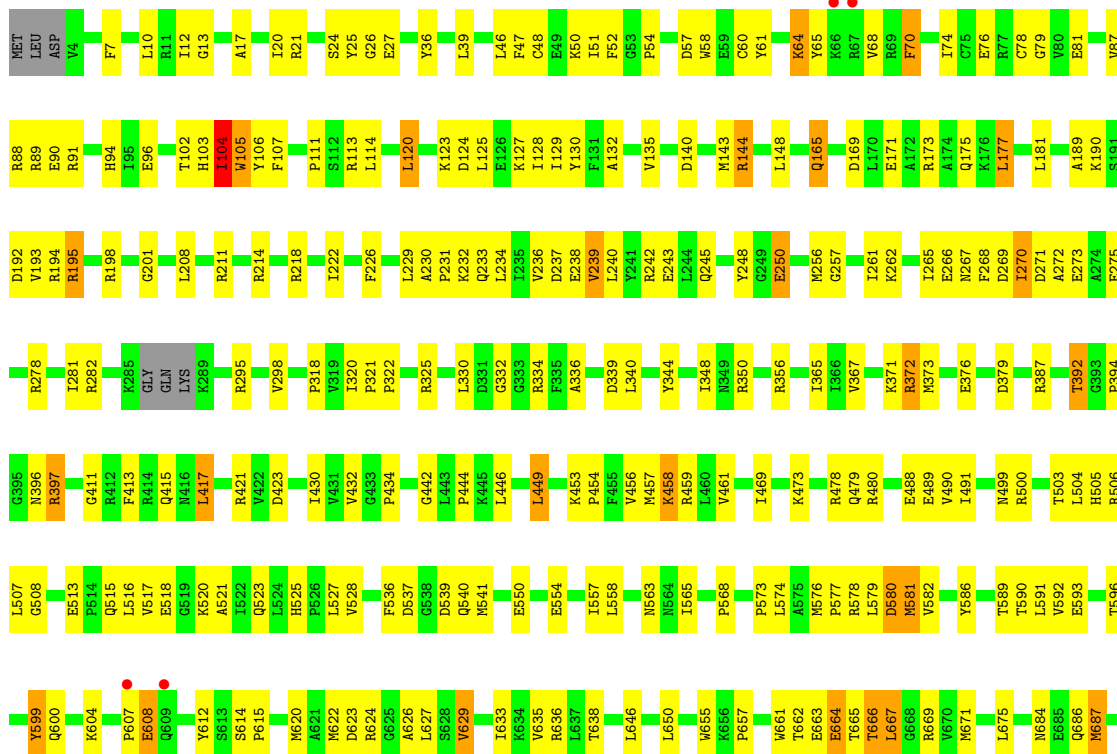


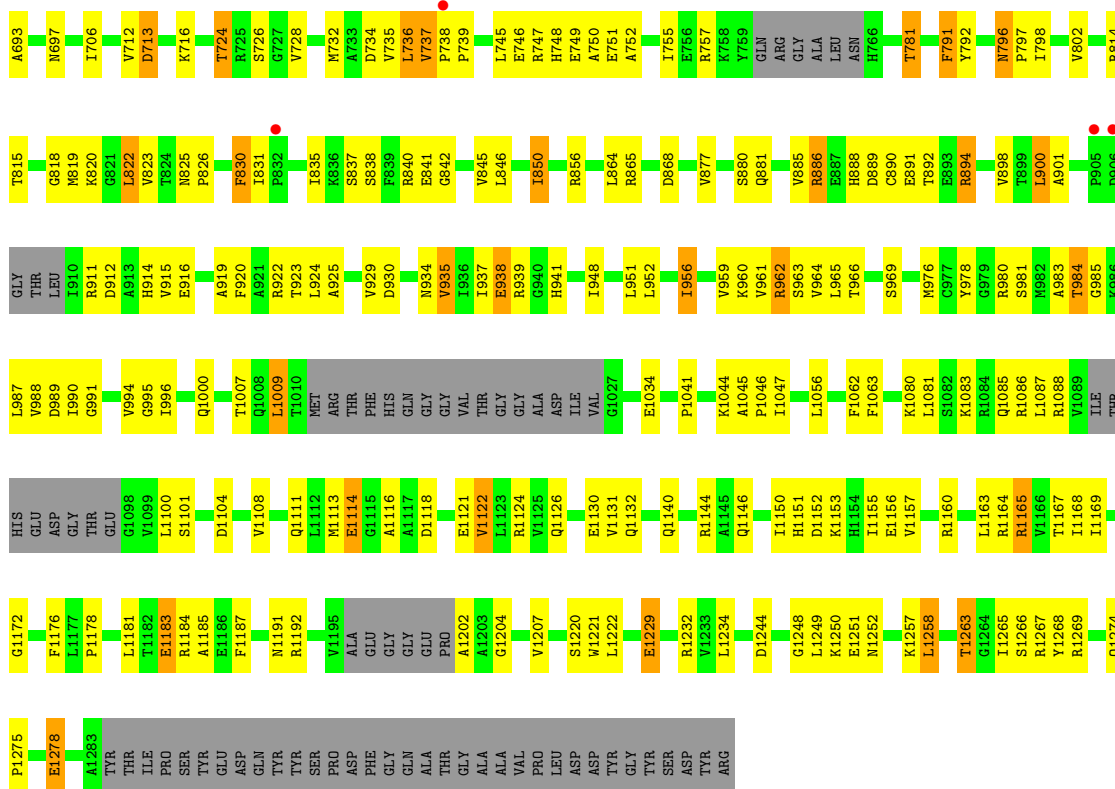




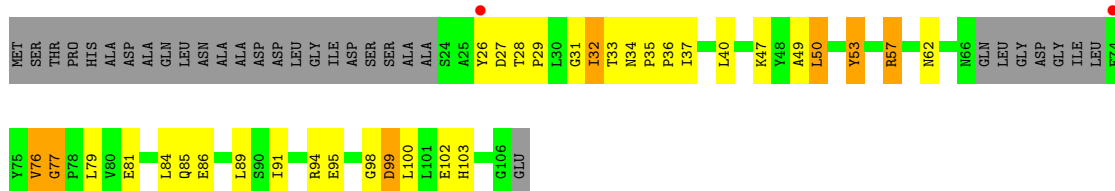


• Molecule 6: DNA-directed RNA polymerase subunit beta'

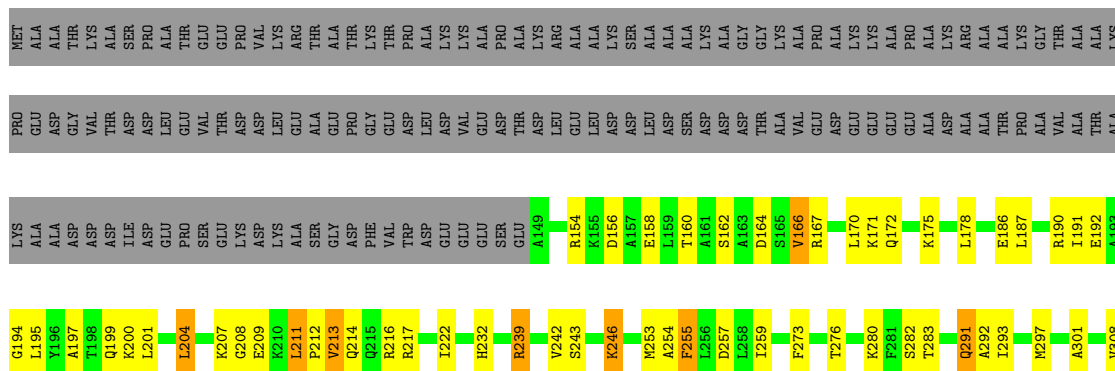


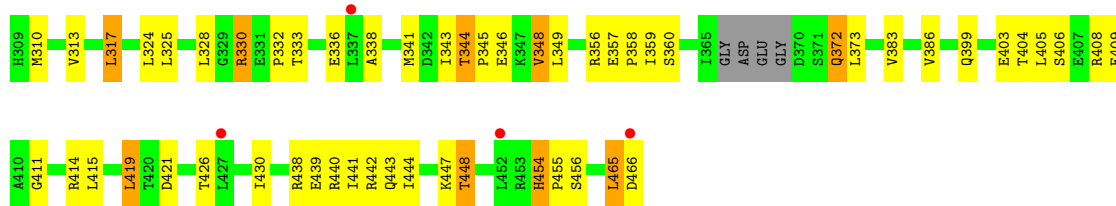


• Molecule 7: DNA-directed RNA polymerase subunit omega

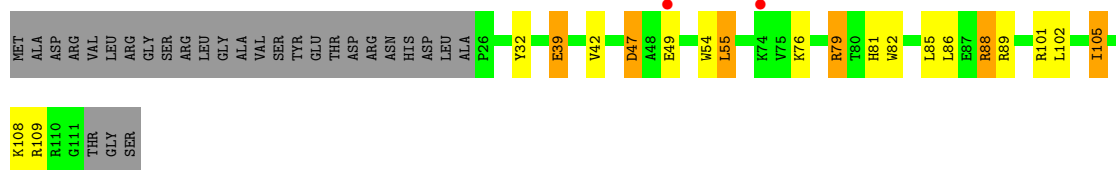


• Molecule 8: RNA polymerase sigma factor SigA





● Molecule 9: RNA polymerase-binding protein RbpA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	132.06Å 163.56Å 139.96Å 90.00° 107.90° 90.00°	Depositor
Resolution (Å)	50.02 – 3.20 50.01 – 3.20	Depositor EDS
% Data completeness (in resolution range)	93.5 (50.02-3.20) 93.5 (50.01-3.20)	Depositor EDS
$R_{merge}$	0.22	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.08 (at 3.19Å)	Xtrriage
Refinement program	PHENIX (1.10.1_2155)	Depositor
R, $R_{free}$	0.253 , 0.280 0.254 , 0.281	Depositor DCC
$R_{free}$ test set	1974 reflections (2.26%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	85.8	Xtrriage
Anisotropy	0.156	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 40.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	26574	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	88.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.48% 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: ZN, EDO, SO4

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	O	0.51	0/1133	0.92	0/1749
2	P	0.51	0/1012	0.89	0/1557
3	Q	0.11	0/91	0.60	0/140
4	A	0.23	0/1643	0.45	0/2242
4	B	0.23	0/1686	0.44	0/2308
5	C	0.26	0/8285	0.44	1/11285 (0.0%)
6	D	0.24	0/9540	0.41	0/12930
7	E	0.23	0/604	0.41	0/822
8	F	0.25	0/2491	0.40	0/3365
9	J	0.23	0/677	0.39	0/920
All	All	0.28	0/27162	0.48	1/37318 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	C	67	ASP	CB-CG-OD2	5.17	122.95	118.30

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	O	1009	0	554	84	0
2	P	903	0	497	29	0
3	Q	82	0	45	1	0
4	A	1617	0	1616	85	0
4	B	1660	0	1624	97	0
5	C	8138	0	7718	388	1
6	D	9391	0	9221	376	1
7	E	592	0	583	31	0
8	F	2461	0	2467	92	0
9	J	663	0	619	17	0
10	C	5	0	0	2	0
10	D	15	0	0	1	0
10	F	10	0	0	3	0
11	D	2	0	0	0	0
12	D	4	6	6	3	0
12	F	4	6	6	0	0
13	C	1	2	0	1	0
13	D	1	2	0	0	0
All	All	26558	16	24956	1084	1

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 21.

The worst 5 of 1084 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)
6:D:732:MET:O	6:D:840:ARG:NH1	1.85	1.10
6:D:981:SER:HB3	6:D:984:THR:OG1	1.53	1.08
4:A:48:GLY:HA2	4:A:49:ALA:HB3	1.39	1.04
5:C:935:TRP:CH2	5:C:954:LEU:CD2	2.40	1.04
5:C:935:TRP:HH2	5:C:954:LEU:HD22	1.18	1.03

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:C:782:ARG:NH1	6:D:171:GLU:OE2[2_356]	1.98	0.22

## 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	
4	A	218/350 (62%)	203 (93%)	13 (6%)	2 (1%)	17	56
4	B	231/350 (66%)	217 (94%)	13 (6%)	1 (0%)	34	69
5	C	1109/1169 (95%)	1031 (93%)	66 (6%)	12 (1%)	14	51
6	D	1224/1317 (93%)	1165 (95%)	53 (4%)	6 (0%)	29	67
7	E	72/107 (67%)	64 (89%)	5 (7%)	3 (4%)	3	20
8	F	310/466 (66%)	303 (98%)	5 (2%)	2 (1%)	25	64
9	J	84/114 (74%)	81 (96%)	3 (4%)	0	100	100
All	All	3248/3873 (84%)	3064 (94%)	158 (5%)	26 (1%)	19	58

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	C	57	GLY
5	C	274	PRO
5	C	939	VAL
6	D	104	ILE
6	D	802	VAL

### 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	
4	A	174/297 (59%)	154 (88%)	20 (12%)	5	24
4	B	170/297 (57%)	141 (83%)	29 (17%)	2	10

*Continued on next page...*



Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
5	C	812/984 (82%)	718 (88%)	94 (12%)	5	24
6	D	953/1096 (87%)	866 (91%)	87 (9%)	9	34
7	E	62/86 (72%)	54 (87%)	8 (13%)	4	19
8	F	255/379 (67%)	233 (91%)	22 (9%)	10	38
9	J	65/98 (66%)	56 (86%)	9 (14%)	3	16
All	All	2491/3237 (77%)	2222 (89%)	269 (11%)	6	27

5 of 269 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	E	50	LEU
8	F	166	VAL
9	J	39	GLU
5	C	568	ASP
5	C	545	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
5	C	576	GLN
6	D	165	GLN
8	F	372	GLN
6	D	1000	GLN
5	C	534	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	Q	3/4 (75%)	1 (33%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	Q	2	C

There are no RNA pucker outliers to report.

## 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 2 are monoatomic - leaving 8 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
10	SO4	D	2005	6	4,4,4	0.14	0	6,6,6	0.05	0
10	SO4	C	1201	-	4,4,4	0.14	0	6,6,6	0.05	0
10	SO4	D	2004	-	4,4,4	0.14	0	6,6,6	0.05	0
12	EDO	D	2006	-	3,3,3	0.45	0	2,2,2	0.29	0
10	SO4	F	502	-	4,4,4	0.14	0	6,6,6	0.05	0
12	EDO	F	503	-	3,3,3	0.45	0	2,2,2	0.31	0
10	SO4	D	2003	6	4,4,4	0.13	0	6,6,6	0.05	0
10	SO4	F	501	8	4,4,4	0.14	0	6,6,6	0.07	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
12	EDO	D	2006	-	-	0/1/1/1	-
12	EDO	F	503	-	-	0/1/1/1	-

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.

5 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	C	1201	SO4	2	0
10	D	2004	SO4	1	0
12	D	2006	EDO	3	0
10	F	502	SO4	2	0
10	F	501	SO4	1	0

## 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	O	49/50 (98%)	-0.33	1 (2%) 65 51	53, 109, 179, 196	0
2	P	44/50 (88%)	-0.29	1 (2%) 60 47	75, 107, 177, 193	0
3	Q	4/4 (100%)	0.08	0 100 100	119, 123, 126, 143	0
4	A	220/350 (62%)	0.03	6 (2%) 54 39	56, 91, 125, 159	0
4	B	233/350 (66%)	0.15	6 (2%) 56 40	78, 111, 137, 145	0
5	C	1113/1169 (95%)	0.04	32 (2%) 51 36	33, 84, 146, 168	0
6	D	1238/1317 (94%)	-0.09	8 (0%) 89 83	32, 76, 123, 152	0
7	E	76/107 (71%)	0.15	2 (2%) 56 40	55, 86, 121, 132	0
8	F	314/466 (67%)	0.01	4 (1%) 77 65	37, 80, 132, 154	0
9	J	86/114 (75%)	0.01	2 (2%) 60 47	58, 101, 138, 169	0
All	All	3377/3977 (84%)	-0.01	62 (1%) 68 55	32, 85, 138, 196	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
7	E	74	GLU	5.5
4	B	1	MET	5.0
4	A	187	THR	5.0
2	P	1	DG	4.6
6	D	609	GLN	4.4

### 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 monosaccharides 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
10	SO4	C	1201	5/5	0.60	0.51	120,128,134,230	0
10	SO4	F	502	5/5	0.78	0.15	73,78,115,167	0
10	SO4	F	501	5/5	0.86	0.17	72,80,89,163	0
12	EDO	D	2006	4/4	0.90	0.21	68,85,104,104	0
10	SO4	D	2005	5/5	0.91	0.17	85,92,116,121	0
12	EDO	F	503	4/4	0.91	0.19	76,91,105,111	0
10	SO4	D	2003	5/5	0.94	0.12	91,97,128,130	0
11	ZN	D	2002	1/1	0.95	0.18	100,100,100,100	0
10	SO4	D	2004	5/5	0.95	0.18	83,90,99,120	0
11	ZN	D	2001	1/1	0.95	0.29	116,116,116,116	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.