



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 13, 2017 – 11:25 PM EST

PDB ID : 5UHD  
Title : Crystal structure of Mycobacterium tuberculosis transcription initiation complex containing 4nt RNA in complex with Rifampin  
Authors : Lin, W.; Das, K.; Feng, Y.; Ebright, R.H.  
Deposited on : unknown  
Resolution : 4.01 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
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) : rb-20030345

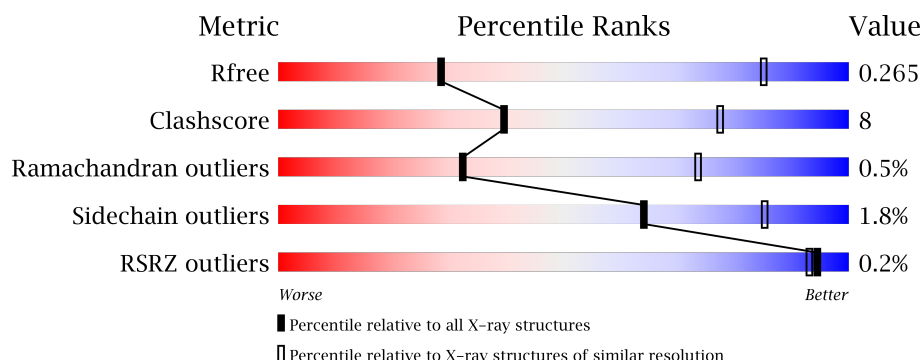
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 4.01 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.




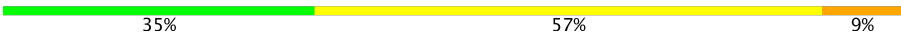

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	1089 (4.42-3.60)
Clashscore	112137	1189 (4.44-3.60)
Ramachandran outliers	110173	1140 (4.42-3.60)
Sidechain outliers	110143	1127 (4.42-3.60)
RSRZ outliers	101464	1100 (4.42-3.60)

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	A	347	
1	B	347	
2	C	1178	
3	D	1316	
4	E	110	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	F	528	 48% 13% 39%
6	H	23	 35% 57% 9%
7	G	16	 13% 56% 25% 6% 13%

## 2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 26022 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 alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	224	Total	C	N	O	S	0	0	0
			1704	1072	295	335	2			
1	B	227	Total	C	N	O	S	0	0	0
			1715	1080	291	342	2			

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1126	Total	C	N	O	S	0	0	0
			8714	5454	1528	1693	39			

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1265	Total	C	N	O	S	0	0	0
			9887	6188	1793	1866	40			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	E	81	Total	C	N	O	0	0	0
			637	408	106	123			

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	320	Total	C	N	O	S	0	0	0
			2543	1583	459	492	9			

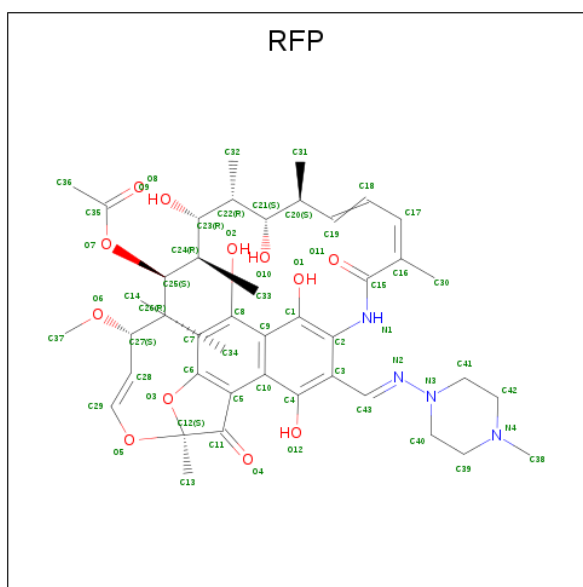
- Molecule 6 is a DNA chain called DNA (5'-D(\*TP\*AP\*TP\*AP\*AP\*TP\*GP\*GP\*GP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	H	23	Total	C	N	O	P	0	0	0
			476	227	91	136	22			

- Molecule 7 is a DNA chain called DNA (5'-D(\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*CP\*GP\*AP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	14	Total	C	N	O	P	0	0	0
			284	136	53	82	13			

- Molecule 8 is RIFAMPICIN (three-letter code: RFP) (formula: C<sub>43</sub>H<sub>58</sub>N<sub>4</sub>O<sub>12</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	C	1	Total	C	N	O	0	0
			59	43	4	12		

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

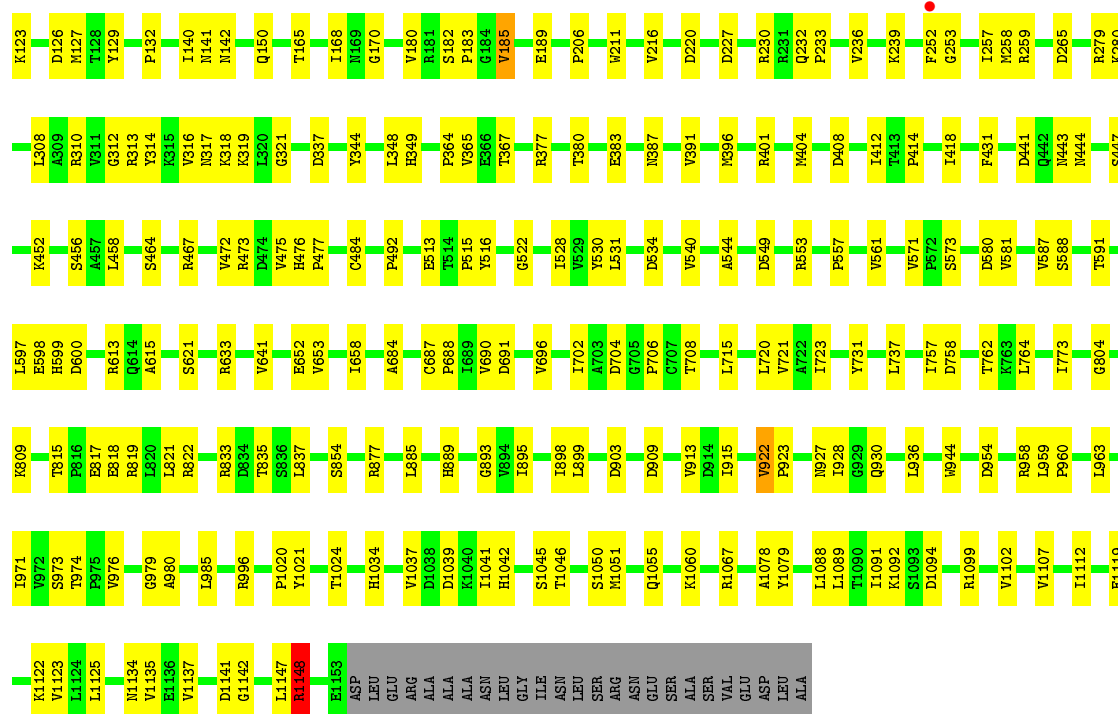
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Zn	0	0
			2	2		

- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	D	1	Total	Mg	0	0
			1	1		

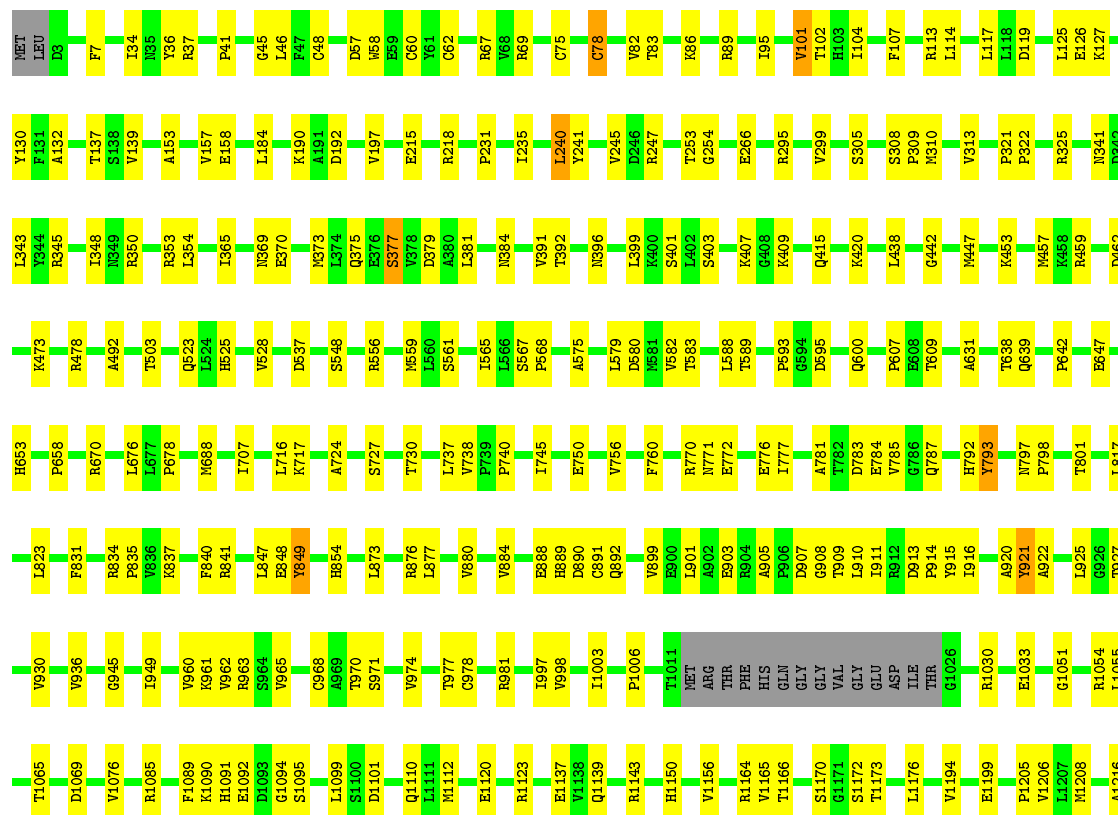


WORLD WIDE  
PDB  
PROTEIN DATA BANK



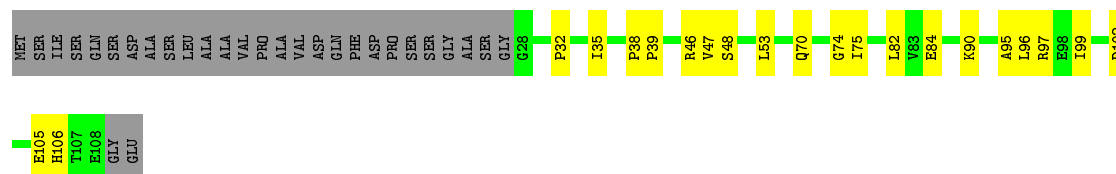
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D: 75% 20%



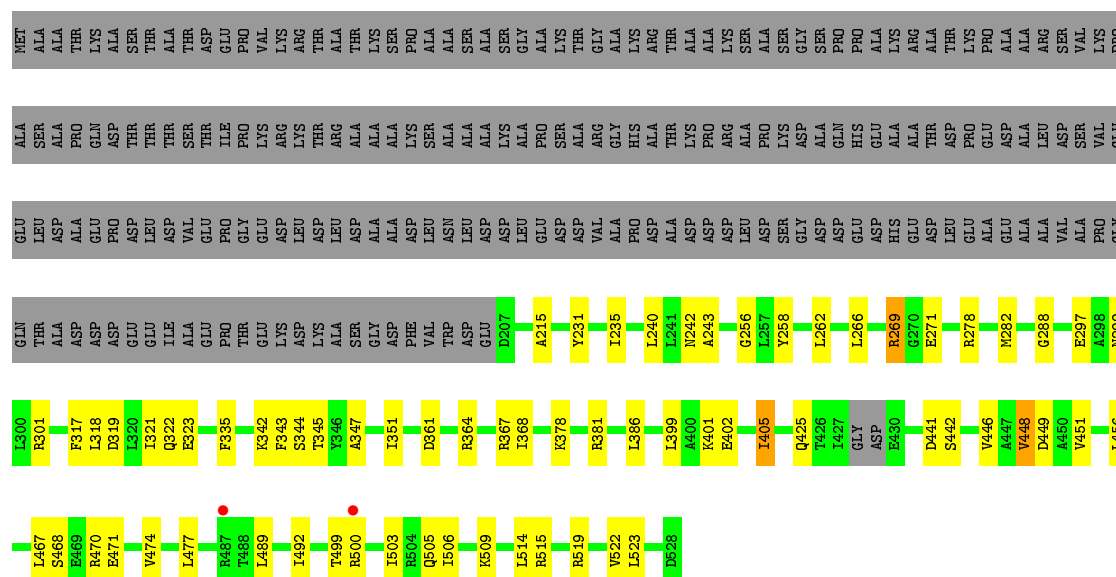
- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:  55% 19% 26%



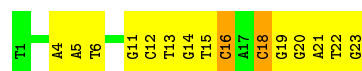
- Molecule 5: RNA polymerase sigma factor SigA

Chain F:  48% 13% . 39%



- Molecule 6: DNA (5'-D(\*TP\*AP\*TP\*AP\*AP\*TP\*GP\*GP\*GP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*G)-3')

Chain H:  35% 57% 9%



- Molecule 7: DNA (5'-D(\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*CP\*GP\*AP\*GP\*G)-3')

Chain G: 





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	152.08 Å 163.10 Å 197.90 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.48 – 4.01 49.83 – 4.01	Depositor EDS
% Data completeness (in resolution range)	90.8 (49.48-4.01) 90.9 (49.83-4.01)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.97 (at 4.00 Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.211 , 0.264 0.213 , 0.265	Depositor DCC
$R_{free}$ test set	1988 reflections (5.22%)	DCC
Wilson B-factor (Å <sup>2</sup> )	30.9	Xtriage
Anisotropy	0.345	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , -13.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.83	EDS
Total number of atoms	26022	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.01% 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 ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, RFP, MG

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	0.23	0/1730	0.44	0/2354
1	B	0.23	0/1741	0.45	0/2371
2	C	0.25	0/8873	0.43	1/12031 (0.0%)
3	D	0.26	0/10052	0.43	0/13591
4	E	0.30	0/650	0.43	0/886
5	F	0.25	0/2572	0.41	0/3466
6	H	0.69	1/535 (0.2%)	1.03	2/826 (0.2%)
7	G	0.61	0/318	1.01	1/489 (0.2%)
All	All	0.27	1/26471 (0.0%)	0.46	4/36014 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	H	16	DC	O3'-P	8.16	1.71	1.61

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	H	18	DC	P-O3'-C3'	10.69	132.53	119.70
7	G	8	DC	P-O3'-C3'	10.23	131.97	119.70
6	H	16	DC	OP1-P-O3'	5.57	117.44	105.20
2	C	48	LEU	CA-CB-CG	5.25	127.38	115.30

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	A	1704	0	1741	26	0
1	B	1715	0	1739	23	0
2	C	8714	0	8636	143	0
3	D	9887	0	9943	188	0
4	E	637	0	635	13	0
5	F	2543	0	2571	54	0
6	H	476	0	261	32	0
7	G	284	0	159	11	0
8	C	59	0	58	6	0
9	D	2	0	0	0	0
10	D	1	0	0	0	0
All	All	26022	0	25743	432	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 8.

The worst 5 of 432 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:H:20:DG:O6	7:G:8:DC:N4	1.99	0.95
6:H:20:DG:N1	7:G:8:DC:N3	2.18	0.90
3:D:353:ARG:HH21	5:F:323:GLU:HG2	1.38	0.88
3:D:353:ARG:HH21	5:F:323:GLU:CG	1.88	0.86
6:H:15:DT:C2	6:H:16:DC:C5	2.63	0.86

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	222/347 (64%)	210 (95%)	11 (5%)	1 (0%)	32	73
1	B	225/347 (65%)	208 (92%)	16 (7%)	1 (0%)	38	77
2	C	1124/1178 (95%)	1050 (93%)	68 (6%)	6 (0%)	32	73
3	D	1261/1316 (96%)	1194 (95%)	61 (5%)	6 (0%)	32	73
4	E	79/110 (72%)	76 (96%)	3 (4%)	0	100	100
5	F	316/528 (60%)	299 (95%)	16 (5%)	1 (0%)	44	80
All	All	3227/3826 (84%)	3037 (94%)	175 (5%)	15 (0%)	32	73

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	678	PRO
2	C	47	PRO
3	D	971	SER
2	C	922	VAL
2	C	1134	ASN

### 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	192/297 (65%)	189 (98%)	3 (2%)	68	86
1	B	192/297 (65%)	192 (100%)	0	100	100
2	C	948/998 (95%)	934 (98%)	14 (2%)	70	87
3	D	1048/1095 (96%)	1027 (98%)	21 (2%)	60	83
4	E	68/90 (76%)	65 (96%)	3 (4%)	33	67
5	F	270/427 (63%)	262 (97%)	8 (3%)	46	75
All	All	2718/3204 (85%)	2669 (98%)	49 (2%)	64	85

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	83	THR
3	D	503	THR
5	F	402	GLU
3	D	240	LEU
3	D	582	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	941	HIS
5	F	415	GLN
2	C	1034	HIS
2	C	543	GLN
2	C	969	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](#)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 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
8	RFP	C	1201	-	62,63,63	2.63	10 (16%)	91,94,94	1.98	14 (15%)

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
8	RFP	C	1201	-	-	0/60/85/85	0/1/5/5

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	C	1201	RFP	O7-C25	-4.91	1.37	1.44
8	C	1201	RFP	C12-C11	-2.89	1.40	1.54
8	C	1201	RFP	O9-C23	-2.28	1.37	1.43
8	C	1201	RFP	C6-C7	3.06	1.45	1.39
8	C	1201	RFP	C18-C17	3.40	1.54	1.43

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	C	1201	RFP	C30-C16-C17	-6.95	105.12	123.44
8	C	1201	RFP	C43-N2-N3	-5.92	109.85	120.40
8	C	1201	RFP	C2-C3-C43	-4.61	116.77	123.29
8	C	1201	RFP	O4-C11-C5	-3.73	123.30	131.84
8	C	1201	RFP	C33-C24-C25	-2.51	106.81	111.43

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	C	1201	RFP	6	0

## 5.7 Other polymers [i](#)

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	224/347 (64%)	-0.26	0 <span>100</span> <span>100</span>	18, 62, 126, 158	0
1	B	227/347 (65%)	0.26	3 (1%) <span>77</span> <span>68</span>	59, 102, 143, 167	0
2	C	1126/1178 (95%)	-0.41	1 (0%) <span>95</span> <span>94</span>	5, 38, 126, 161	0
3	D	1265/1316 (96%)	-0.53	0 <span>100</span> <span>100</span>	3, 38, 109, 144	0
4	E	81/110 (73%)	-0.43	0 <span>100</span> <span>100</span>	24, 53, 89, 110	0
5	F	320/528 (60%)	-0.35	2 (0%) <span>89</span> <span>84</span>	7, 63, 157, 185	0
6	H	23/23 (100%)	0.18	0 <span>100</span> <span>100</span>	33, 122, 194, 200	0
7	G	14/16 (87%)	0.96	2 (14%) <span>3</span> <span>4</span>	104, 162, 240, 304	0
All	All	3280/3865 (84%)	-0.38	8 (0%) <span>94</span> <span>93</span>	3, 48, 131, 304	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
7	G	6	DA	3.5
5	F	487	ARG	3.1
7	G	5	DC	2.8
1	B	135	GLU	2.7
5	F	500	ARG	2.5

### 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( $\text{\AA}^2$ )	Q<0.9
8	RFP	C	1201	59/59	0.96	0.24	0.71	12,23,53,58	0
9	ZN	D	1402	1/1	0.99	0.10	-2.12	44,44,44,44	0
9	ZN	D	1401	1/1	0.99	0.07	-2.52	32,32,32,32	0
10	MG	D	1403	1/1	0.91	0.10	-	3,3,3,3	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.