



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 13, 2017 – 11:12 PM EST

PDB ID : 5UHF  
Title : Crystal structure of Mycobacterium tuberculosis transcription initiation complex in complex with D-IX336  
Authors : Lin, W.; Das, K.; Feng, Y.; Ebright, R.H.  
Deposited on : unknown  
Resolution : 4.34 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

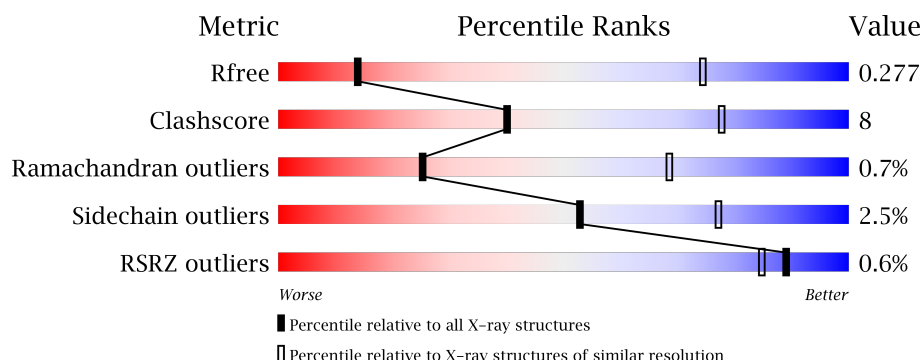
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 4.34 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	1023 (5.06-3.62)
Clashscore	112137	1025 (5.00-3.68)
Ramachandran outliers	110173	1015 (5.02-3.66)
Sidechain outliers	110143	1017 (5.04-3.64)
RSRZ outliers	101464	1005 (5.04-3.64)

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	
6	H	23	
7	G	16	

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	88D	D	1404	-	-	-	X

## 2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 25949 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	225	Total	C	N	O	S	0	0	0
			1706	1075	289	340	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	322	Total	C	N	O	S	0	0	0
			2555	1589	461	496	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\*CP\*AP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	12	Total	C	N	O	P	0	0	0
			241	116	43	71	11			

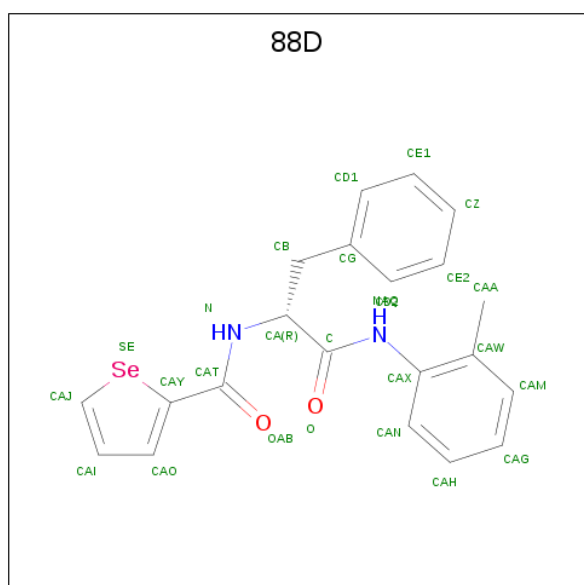
- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).


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

- Molecule 10 is N-(2-methylphenyl)-Nalpha-(selenophene-2-carbonyl)-D-phenylalaninamide (three-letter code: 88D) (formula: C<sub>21</sub>H<sub>20</sub>N<sub>2</sub>O<sub>2</sub>Se).



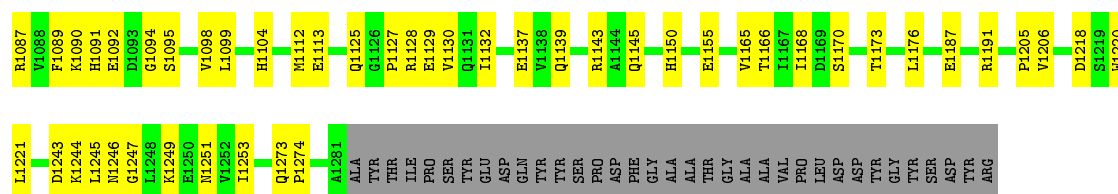
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
10	D	1	Total	C	N	O	Se	0	0
			26	21	2	2	1		

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 ( $\text{RSRZ} > 2$ ). 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.

- Chain C: 

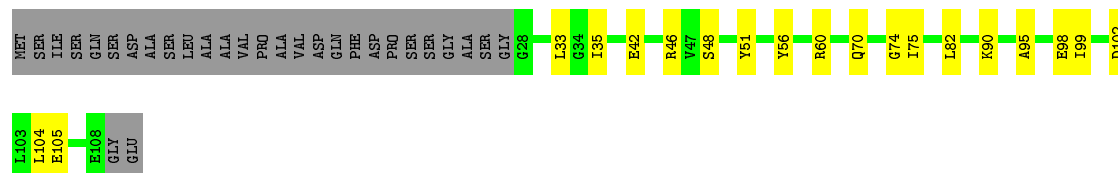




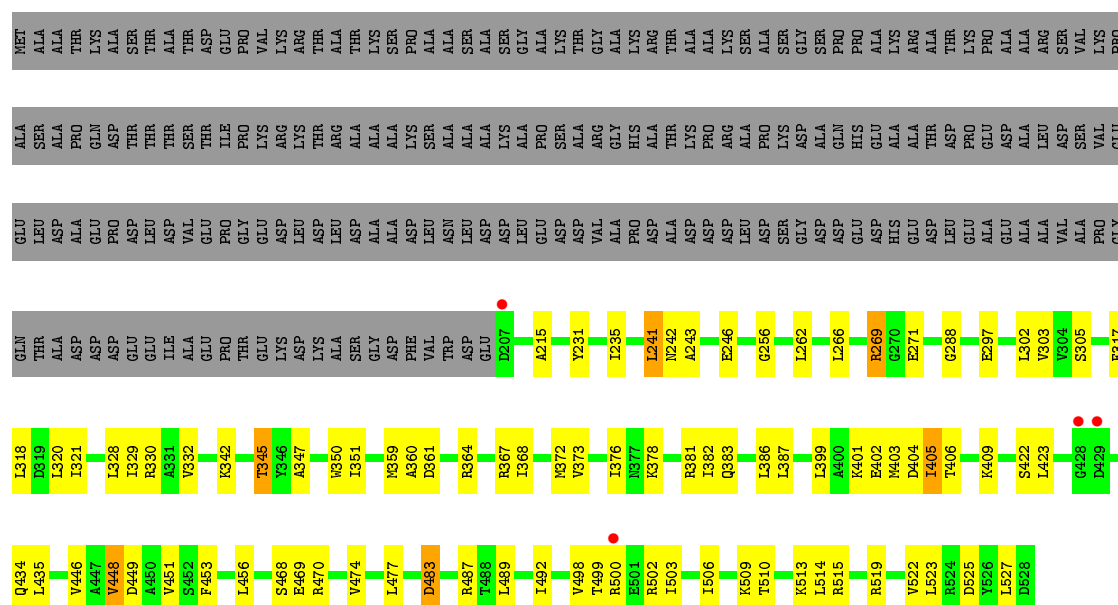


• Molecule 4: DNA-directed RNA polymerase subunit omega

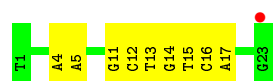
Chain E:



• Molecule 5: RNA polymerase sigma factor SigA

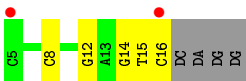


• Molecule 6: DNA (5'-D(\*TP\*AP\*TP\*AP\*AP\*TP\*GP\*GP\*GP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*G)-3')



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





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	152.37Å 163.46Å 197.38Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.50 – 4.34 49.80 – 4.34	Depositor EDS
% Data completeness (in resolution range)	71.6 (48.50-4.34) 86.3 (49.80-4.34)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.58 (at 4.29Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.198 , 0.267 0.211 , 0.277	Depositor DCC
$R_{free}$ test set	2000 reflections (7.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	52.4	Xtriage
Anisotropy	0.747	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 32.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.81	EDS
Total number of atoms	25949	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.34% 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, MG, 88D

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	0.22	0/1730	0.44	0/2354
1	B	0.22	0/1731	0.44	0/2356
2	C	0.23	0/8873	0.42	1/12031 (0.0%)
3	D	0.24	0/10052	0.42	0/13591
4	E	0.23	0/650	0.41	0/886
5	F	0.24	0/2585	0.41	0/3485
6	H	0.56	0/535	0.89	0/826
7	G	0.58	0/269	0.90	0/413
All	All	0.25	0/26425	0.45	1/35942 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	48	LEU	CA-CB-CG	5.30	127.49	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	25	0
1	B	1706	0	1730	28	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	8714	0	8636	160	0
3	D	9887	0	9943	181	0
4	E	637	0	635	14	0
5	F	2555	0	2579	56	0
6	H	476	0	261	14	0
7	G	241	0	137	4	0
8	D	2	0	0	0	0
9	D	1	0	0	0	0
10	D	26	0	0	1	0
All	All	25949	0	25662	428	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 428 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)
3:D:891:CYS:SG	3:D:975:CYS:HB3	2.07	0.95
3:D:902:ALA:HB3	3:D:958:THR:HG22	1.53	0.90
6:H:16:DC:O2	7:G:12:DG:N2	2.06	0.89
3:D:900:GLU:OE1	3:D:959:GLN:NE2	2.06	0.87
3:D:901:LEU:O	3:D:901:LEU:HD22	1.85	0.77

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%)	205 (92%)	16 (7%)	1 (0%)	32 74
1	B	221/347 (64%)	201 (91%)	17 (8%)	3 (1%)	13 54

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	1124/1178 (95%)	1042 (93%)	70 (6%)	12 (1%)	17	60
3	D	1261/1316 (96%)	1187 (94%)	67 (5%)	7 (1%)	28	71
4	E	79/110 (72%)	75 (95%)	4 (5%)	0	100	100
5	F	320/528 (61%)	301 (94%)	18 (6%)	1 (0%)	44	81
All	All	3227/3826 (84%)	3011 (93%)	192 (6%)	24 (1%)	25	68

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	227	VAL
3	D	593	PRO
3	D	678	PRO
3	D	971	SER
5	F	405	ILE

### 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%)	187 (97%)	5 (3%)	51	76
1	B	192/297 (65%)	191 (100%)	1 (0%)	91	95
2	C	948/998 (95%)	928 (98%)	20 (2%)	59	81
3	D	1048/1095 (96%)	1018 (97%)	30 (3%)	48	73
4	E	68/90 (76%)	66 (97%)	2 (3%)	48	73
5	F	271/427 (64%)	261 (96%)	10 (4%)	39	69
All	All	2719/3204 (85%)	2651 (98%)	68 (2%)	53	77

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	105	TRP
3	D	580	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
5	F	367	ARG
3	D	110	VAL
3	D	297	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	141	ASN
3	D	375	GLN
3	D	510	GLN
5	F	415	GLN
5	F	505	GLN

### 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	$\# Z  > 2$
10	88D	D	1404	-	25,28,28	2.65	5 (20%)	31,37,37	0.89	2 (6%)

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
10	88D	D	1404	-	-	0/17/20/20	0/3/3/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
10	D	1404	88D	CAY-CAT	-9.71	1.33	1.49
10	D	1404	88D	CAA-CAW	-5.94	1.39	1.51
10	D	1404	88D	CB-CG	-4.68	1.40	1.51
10	D	1404	88D	CAX-NAQ	-3.82	1.34	1.41
10	D	1404	88D	CAI-CAJ	-2.10	1.35	1.39

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
10	D	1404	88D	CAX-NAQ-C	-2.75	119.61	126.92
10	D	1404	88D	CAM-CAW-CAX	2.22	119.21	117.42

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	D	1404	88D	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	A	224/347 (64%)	-0.16	1 (0%) 92 88	31, 66, 148, 176	0
1	B	225/347 (64%)	0.25	4 (1%) 69 61	70, 115, 167, 184	0
2	C	1126/1178 (95%)	-0.31	8 (0%) 87 82	24, 47, 139, 169	0
3	D	1265/1316 (96%)	-0.39	0 100 100	22, 43, 112, 152	0
4	E	81/110 (73%)	-0.20	0 100 100	29, 49, 84, 105	0
5	F	322/528 (60%)	-0.24	4 (1%) 79 71	25, 66, 171, 204	0
6	H	23/23 (100%)	0.38	1 (4%) 36 29	37, 134, 181, 268	0
7	G	12/16 (75%)	1.12	2 (16%) 2 4	128, 176, 352, 369	0
All	All	3278/3865 (84%)	-0.27	20 (0%) 89 85	22, 52, 148, 369	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	500	ARG	4.0
7	G	5	DC	3.5
1	B	159	ILE	3.2
6	H	23	DG	2.9
2	C	215	ASP	2.9

### 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(Å <sup>2</sup> )	Q<0.9
10	88D	D	1404	26/26	0.89	0.45	2.08	46,64,136,144	0
8	ZN	D	1402	1/1	0.95	0.06	-2.23	35,35,35,35	0
8	ZN	D	1401	1/1	0.95	0.08	-2.28	21,21,21,21	0
9	MG	D	1403	1/1	0.95	0.30	-	28,28,28,28	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.