



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 29, 2020 – 09:36 PM BST

PDB ID : 6DVE  
Title : Crystal structure of Mycobacterium tuberculosis transcription initiation complex (ECF selenomethionine-labelled sigma factor L) with 6 nt spacer  
Authors : Lin, W.; Das, K.; Feng, Y.; Ebright, R.H.  
Deposited on : 2018-06-23  
Resolution : 3.81 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 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 : 2.13  
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) : 2.13

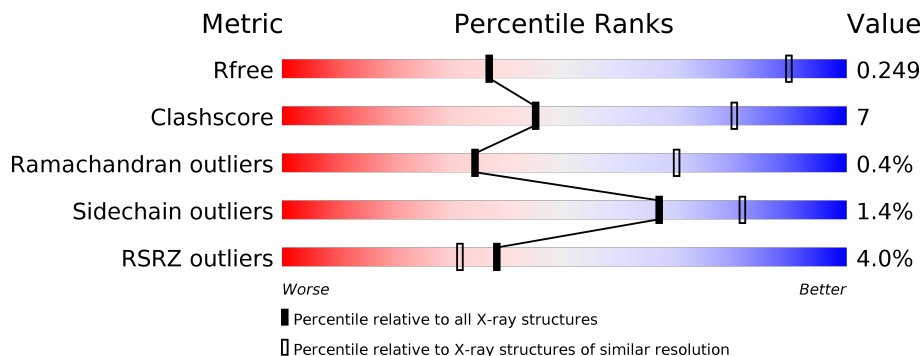
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.81 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.



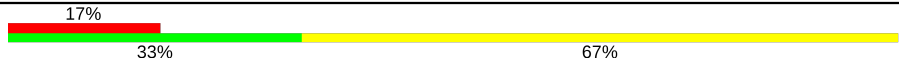

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1231 (4.04-3.60)
Clashscore	141614	1031 (4.02-3.62)
Ramachandran outliers	138981	1261 (4.04-3.60)
Sidechain outliers	138945	1255 (4.04-3.60)
RSRZ outliers	127900	1139 (4.04-3.60)

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 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	A	359	 0% (upper red bar), 50% (green), 12% (yellow), 37% (grey)
1	B	359	 2% (upper red bar), 49% (green), 15% (yellow), 35% (grey)
2	C	1178	 2% (upper red bar), 80% (green), 16% (yellow), 4% (grey)
3	D	1316	 5% (upper red bar), 80% (green), 16% (yellow), 4% (grey)
4	E	110	 7% (upper red bar), 57% (green), 16% (yellow), 26% (grey)
5	F	177	 5% (upper red bar), 80% (green), 18% (yellow), 2% (grey)

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
6	G	12	 17% 33% 67%
7	H	24	 8% 38% 63%

## 2 Entry composition i

There are 8 unique types of molecules in this entry. The entry contains 24785 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
			Total	C	N	O	S			
1	A	225	Total	C	N	O	S	0	0	0
			1716	1080	296	338	2			
1	B	232	Total	C	N	O	S	0	0	0
			1732	1093	296	341	2			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MET	-	initiating methionine	UNP P9WGZ1
A	-10	GLY	-	expression tag	UNP P9WGZ1
A	-9	HIS	-	expression tag	UNP P9WGZ1
A	-8	HIS	-	expression tag	UNP P9WGZ1
A	-7	HIS	-	expression tag	UNP P9WGZ1
A	-6	HIS	-	expression tag	UNP P9WGZ1
A	-5	HIS	-	expression tag	UNP P9WGZ1
A	-4	HIS	-	expression tag	UNP P9WGZ1
A	-3	HIS	-	expression tag	UNP P9WGZ1
A	-2	HIS	-	expression tag	UNP P9WGZ1
A	-1	HIS	-	expression tag	UNP P9WGZ1
A	0	HIS	-	expression tag	UNP P9WGZ1
B	-11	MET	-	initiating methionine	UNP P9WGZ1
B	-10	GLY	-	expression tag	UNP P9WGZ1
B	-9	HIS	-	expression tag	UNP P9WGZ1
B	-8	HIS	-	expression tag	UNP P9WGZ1
B	-7	HIS	-	expression tag	UNP P9WGZ1
B	-6	HIS	-	expression tag	UNP P9WGZ1
B	-5	HIS	-	expression tag	UNP P9WGZ1
B	-4	HIS	-	expression tag	UNP P9WGZ1
B	-3	HIS	-	expression tag	UNP P9WGZ1
B	-2	HIS	-	expression tag	UNP P9WGZ1
B	-1	HIS	-	expression tag	UNP P9WGZ1
B	0	HIS	-	expression tag	UNP P9WGZ1

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	1126	8724	5459	1531	1695	39	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	1265	9895	6195	1794	1866	40	0	0	0

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
4	E	81	630	403	106	121	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
5	F	174	1352	840	256	254	2	0	0	0

- Molecule 6 is a DNA chain called DNA (5'-D(\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	G	12	244	117	45	71	11	0	0	0

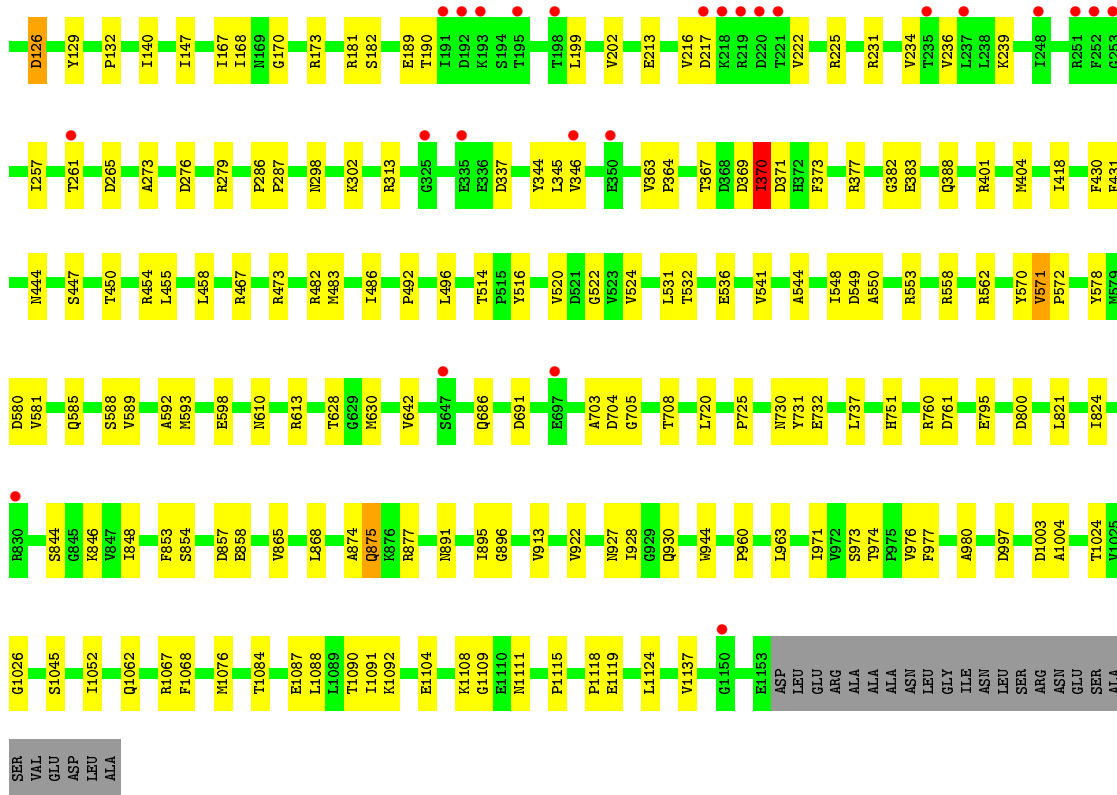
- Molecule 7 is a DNA chain called DNA (5'-D(P\*CP\*GP\*TP\*GP\*TP\*GP\*AP\*GP\*TP\*AP\*AP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
7	H	24	490	234	90	143	23	0	0	0

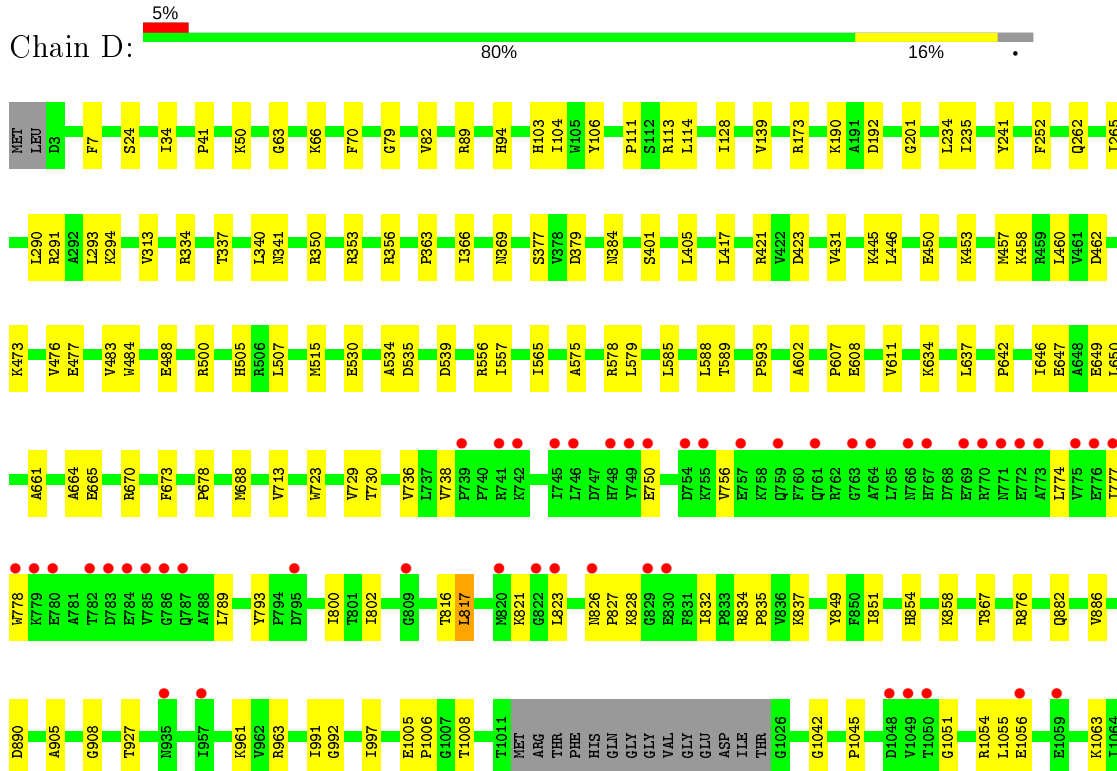
- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

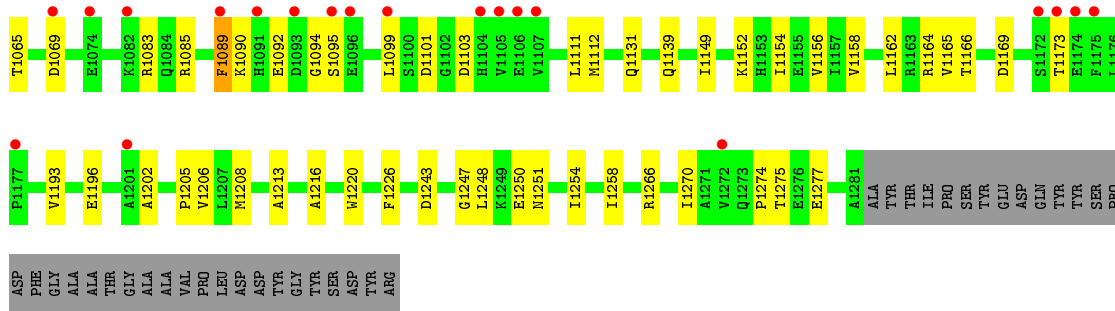
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
8	D	2	2	2	0	0



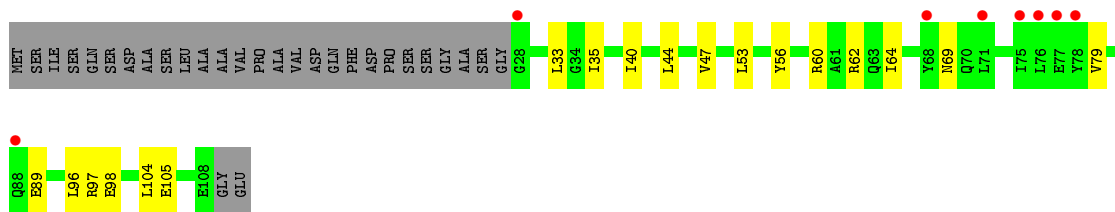


• Molecule 3: DNA-directed RNA polymerase subunit beta'

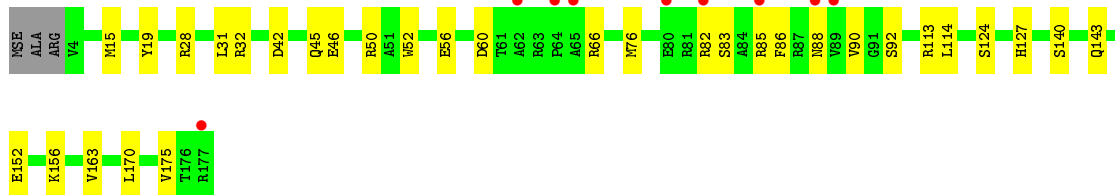
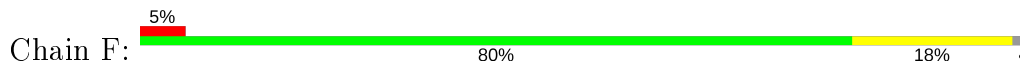




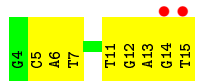
- Molecule 4: DNA-directed RNA polymerase subunit omega



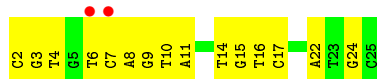
- Molecule 5: ECF RNA polymerase sigma factor SigL



- Molecule 6: DNA (5'-D(\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*T)-3')



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





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	142.76Å 160.61Å 240.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.68 – 3.81 50.13 – 3.81	Depositor EDS
% Data completeness (in resolution range)	58.5 (46.68-3.81) 75.0 (50.13-3.81)	Depositor EDS
$R_{merge}$	0.18	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.84 (at 3.77Å)	Xtrriage
Refinement program	PHENIX 1.12_2829	Depositor
R, $R_{free}$	0.200 , 0.248 0.200 , 0.249	Depositor DCC
$R_{free}$ test set	2000 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	33.4	Xtrriage
Anisotropy	1.117	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 48.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.83	EDS
Total number of atoms	24785	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	79.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.98% 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

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	0/1742	0.44	0/2370
1	B	0.24	0/1758	0.44	0/2397
2	C	0.24	0/8883	0.42	0/12043
3	D	0.24	0/10061	0.40	0/13600
4	E	0.24	0/643	0.38	0/877
5	F	0.24	0/1372	0.40	0/1863
6	G	0.59	0/273	0.96	0/420
7	H	0.60	0/549	0.98	0/846
All	All	0.26	0/25281	0.45	0/34416

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	1716	0	1756	31	0
1	B	1732	0	1754	37	0
2	C	8724	0	8651	111	0
3	D	9895	0	9953	131	0
4	E	630	0	622	15	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	F	1352	0	1346	22	0
6	G	244	0	137	6	0
7	H	490	0	272	18	0
8	D	2	0	0	0	0
All	All	24785	0	24491	320	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 7.

The worst 5 of 320 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:334:ARG:HD3	5:F:90:VAL:HG21	1.52	0.90
1:A:40:ARG:HE	1:B:33:THR:HG22	1.42	0.84
6:G:11:DT:H2'	6:G:12:DG:C8	2.16	0.80
7:H:2:DC:H2'	7:H:3:DG:C8	2.18	0.78
2:C:1024:THR:H	3:D:730:THR:HG21	1.50	0.76

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	223/359 (62%)	211 (95%)	12 (5%)	0	100	100
1	B	230/359 (64%)	208 (90%)	19 (8%)	3 (1%)	12	48
2	C	1124/1178 (95%)	1076 (96%)	43 (4%)	5 (0%)	34	70
3	D	1261/1316 (96%)	1200 (95%)	57 (4%)	4 (0%)	41	74
4	E	79/110 (72%)	75 (95%)	4 (5%)	0	100	100
5	F	172/177 (97%)	166 (96%)	6 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	3089/3499 (88%)	2936 (95%)	141 (5%)	12 (0%)	34 70

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	159	ILE
2	C	370	ILE
3	D	678	PRO
1	B	150	VAL
3	D	593	PRO

### 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
1	A	194/308 (63%)	190 (98%)	4 (2%)	53 73
1	B	191/308 (62%)	189 (99%)	2 (1%)	76 86
2	C	950/998 (95%)	935 (98%)	15 (2%)	62 79
3	D	1050/1095 (96%)	1040 (99%)	10 (1%)	76 86
4	E	66/90 (73%)	66 (100%)	0	100 100
5	F	134/133 (101%)	128 (96%)	6 (4%)	27 57
All	All	2585/2932 (88%)	2548 (99%)	37 (1%)	67 81

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	800	ASP
2	C	1137	VAL
5	F	66	ARG
2	C	858	GLU
2	C	875	GLN

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 monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

### 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	225/359 (62%)	-0.42	2 (0%) 84 78	17, 47, 117, 159	0
1	B	232/359 (64%)	-0.16	7 (3%) 50 39	23, 68, 131, 190	0
2	C	1126/1178 (95%)	-0.29	28 (2%) 57 48	10, 58, 145, 190	0
3	D	1265/1316 (96%)	-0.12	69 (5%) 25 21	9, 70, 174, 237	0
4	E	81/110 (73%)	0.48	8 (9%) 7 6	50, 83, 153, 208	0
5	F	172/177 (97%)	0.19	9 (5%) 27 23	34, 86, 160, 190	0
6	G	12/12 (100%)	0.14	2 (16%) 1 2	109, 142, 163, 165	0
7	H	24/24 (100%)	0.74	2 (8%) 11 9	113, 153, 212, 257	0
All	All	3137/3535 (88%)	-0.17	127 (4%) 38 31	9, 66, 163, 257	0

The worst 5 of 127 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	1099	LEU	6.3
5	F	177	ARG	5.7
3	D	1105	VAL	5.2
7	H	6	DT	4.9
3	D	1106	GLU	4.8

### 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
8	ZN	D	2002	1/1	0.98	0.08	82,82,82,82	0
8	ZN	D	2001	1/1	0.99	0.10	57,57,57,57	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.