



# wwPDB X-ray Structure Validation Summary Report

Feb 15, 2017 – 06:47 am GMT

PDB ID : 4C3H  
Title : Structure of 14-subunit RNA polymerase I at 3.27 Å resolution, crystal form C2-93  
Authors : Fernandez-Tornero, C.; Moreno-Morcillo, M.; Rashid, U.J.; Taylor, N.M.I.; Ruiz, F.M.; Gruene, T.; Legrand, P.; Steuerwald, U.; Muller, C.W.  
Deposited on : 2013-08-24  
Resolution : 3.27 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  
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

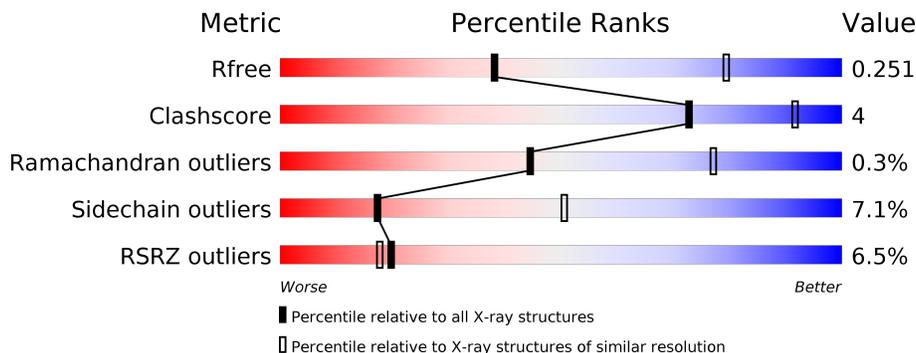
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.27 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	1006 (3.34-3.22)
Clashscore	112137	1070 (3.34-3.22)
Ramachandran outliers	110173	1050 (3.34-3.22)
Sidechain outliers	110143	1049 (3.34-3.22)
RSRZ outliers	101464	1011 (3.34-3.22)

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	1664	
2	B	1203	
3	C	335	
4	D	137	
5	E	215	
6	F	155	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
7	G	326	
8	H	146	
9	I	125	
10	J	70	
11	K	142	
12	L	70	
13	M	415	
14	N	233	

## 2 Entry composition [i](#)

There are 15 unique types of molecules in this entry. The entry contains 34552 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 I SUBUNIT RPA190.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1523	12019	7577	2086	2292	64	0	0	0

- Molecule 2 is a protein called DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA135.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1176	9322	5898	1629	1745	50	0	0	0

- Molecule 3 is a protein called DNA-DIRECTED RNA POLYMERASES I AND III SUBUNIT RPAC1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	304	2418	1536	414	460	8	0	0	0

- Molecule 4 is a protein called DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA14.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
4	D	59	466	292	80	94	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	12	SER	THR	CONFLICT	UNP P50106

- Molecule 5 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	215	1759	1116	310	321	12	0	0	0

- Molecule 6 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	100	823	522	144	154	3	0	0	0

- Molecule 7 is a protein called DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA43.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	G	259	2052	1301	348	398	5	0	0	0

- Molecule 8 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	H	134	1072	676	181	211	4	0	0	0

- Molecule 9 is a protein called DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA12.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	I	124	942	584	160	189	9	0	0	0

- Molecule 10 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
10	J	69	569	362	101	100	6	0	0	0

- Molecule 11 is a protein called DNA-DIRECTED RNA POLYMERASES I AND III SUBUNIT RPAC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
11	K	103	810	506	132	167	5	0	0	0

- Molecule 12 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
12	L	45	359	221	71	63	4	0	0	0

- Molecule 13 is a protein called DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA49.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
13	M	105	831	528	137	166	0	0	0

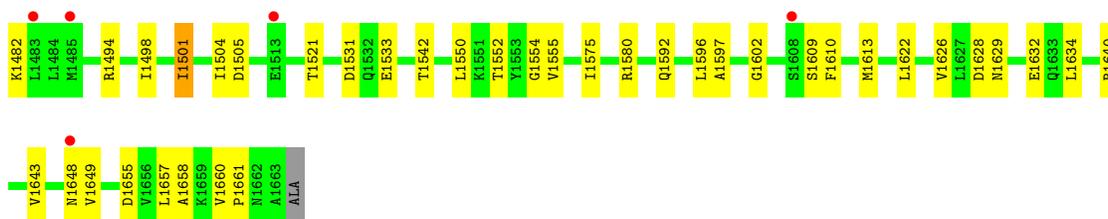
- Molecule 14 is a protein called DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA34.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
14	N	139	1103	706	179	214	4	0	0	0

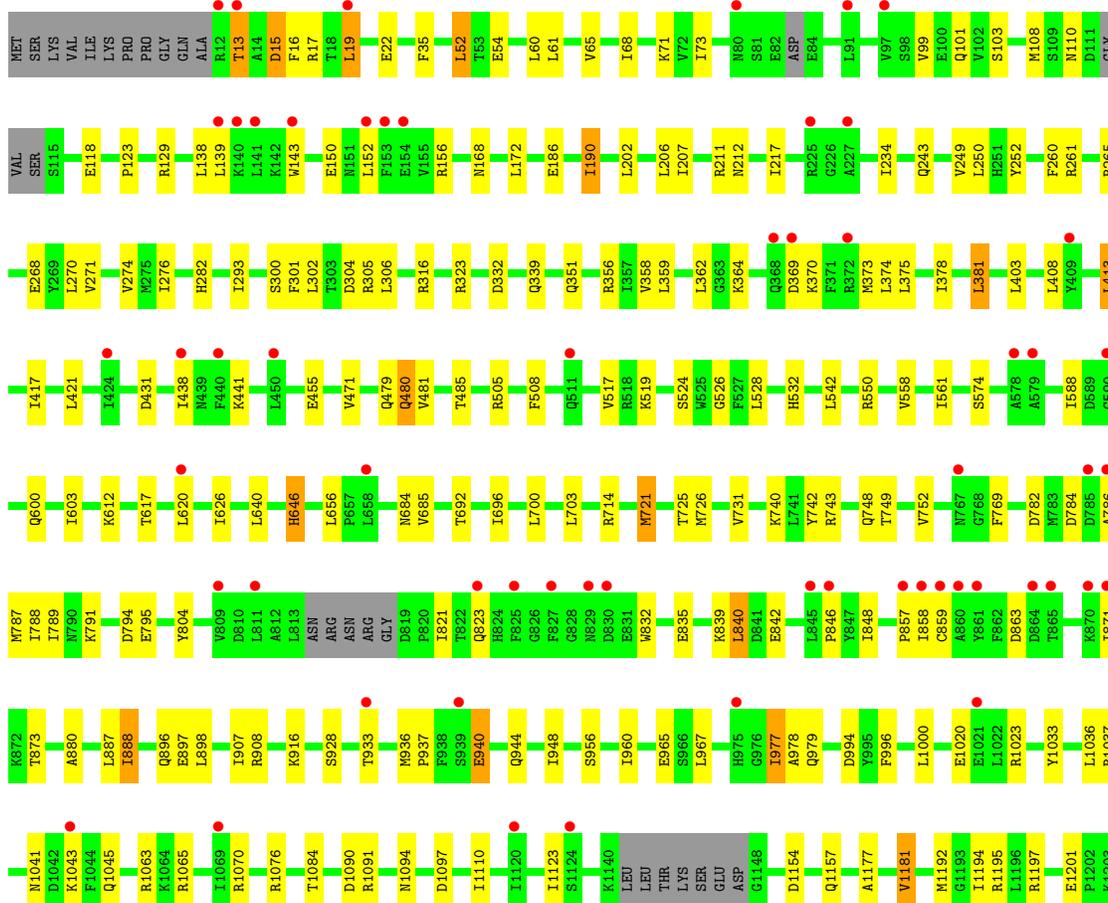
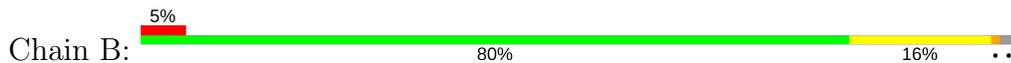
- Molecule 15 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
15	B	1	Total	Zn	0	0
			1	1		
15	A	2	Total	Zn	0	0
			2	2		
15	L	1	Total	Zn	0	0
			1	1		
15	J	1	Total	Zn	0	0
			1	1		
15	I	2	Total	Zn	0	0
			2	2		





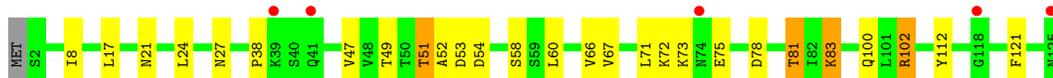
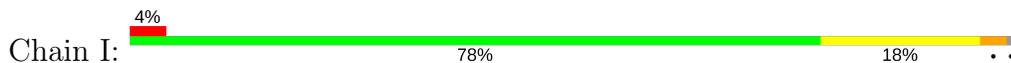
● Molecule 2: DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA135







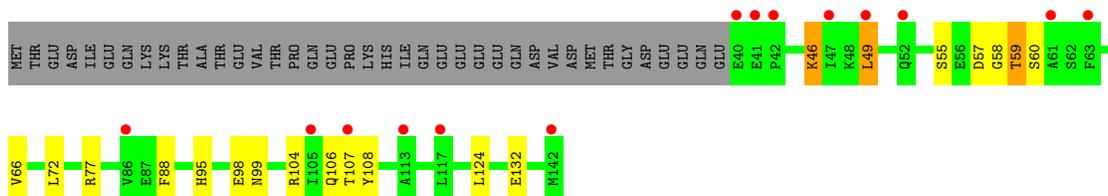
- Molecule 9: DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA12



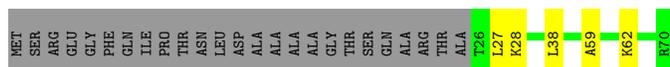
- Molecule 10: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 5



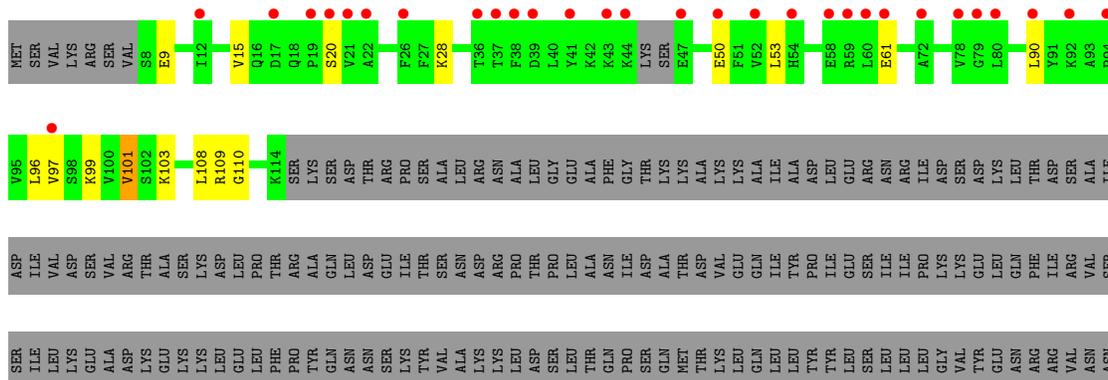
- Molecule 11: DNA-DIRECTED RNA POLYMERASES I AND III SUBUNIT RPAC2



- Molecule 12: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 4



- Molecule 13: DNA-DIRECTED RNA POLYMERASE I SUBUNIT RPA49





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	425.24Å 140.62Å 139.72Å 90.00° 93.35° 90.00°	Depositor
Resolution (Å)	47.88 – 3.27 47.78 – 3.27	Depositor EDS
% Data completeness (in resolution range)	97.7 (47.88-3.27) 97.7 (47.78-3.27)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.41 (at 3.25Å)	Xtrriage
Refinement program	BUSTER 2.11.5	Depositor
R, $R_{free}$	0.221 , 0.226 0.248 , 0.251	Depositor DCC
$R_{free}$ test set	6200 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	135.5	Xtrriage
Anisotropy	0.369	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 101.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	34552	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	171.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.44% 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

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	0.41	0/12233	0.58	0/16523
2	B	0.41	0/9527	0.60	0/12879
3	C	0.40	0/2469	0.60	0/3347
4	D	0.39	0/472	0.52	0/639
5	E	0.41	0/1795	0.56	0/2416
6	F	0.40	0/838	0.54	0/1129
7	G	0.39	0/2094	0.58	0/2843
8	H	0.39	0/1090	0.57	0/1476
9	I	0.39	0/953	0.56	0/1282
10	J	0.41	0/578	0.62	0/775
11	K	0.39	0/821	0.59	0/1108
12	L	0.39	0/361	0.60	0/478
13	M	0.39	0/846	0.53	0/1136
14	N	0.39	0/1124	0.52	0/1512
All	All	0.40	0/35201	0.58	0/47543

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

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	12019	0	12051	127	0
2	B	9322	0	9187	92	0
3	C	2418	0	2401	21	0
4	D	466	0	466	3	0
5	E	1759	0	1788	8	0
6	F	823	0	841	6	0
7	G	2052	0	2016	15	0
8	H	1072	0	1042	6	0
9	I	942	0	928	14	0
10	J	569	0	585	12	0
11	K	810	0	801	12	0
12	L	359	0	381	1	0
13	M	831	0	820	10	0
14	N	1103	0	1106	6	0
15	A	2	0	0	0	0
15	B	1	0	0	0	0
15	I	2	0	0	0	0
15	J	1	0	0	0	0
15	L	1	0	0	0	0
All	All	34552	0	34413	272	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 4.

The worst 5 of 272 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)
11:K:59:THR:HG22	11:K:107:THR:OG1	1.46	1.13
11:K:60:SER:OG	11:K:106:GLN:HG2	1.62	0.99
1:A:1382:VAL:HA	2:B:1070:ARG:NH1	1.78	0.97
2:B:99:VAL:HG21	2:B:417:ILE:HD11	1.58	0.85
3:C:222:VAL:HG21	3:C:225:ALA:HB2	1.68	0.75

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	1503/1664 (90%)	1421 (94%)	78 (5%)	4 (0%)	44	77
2	B	1166/1203 (97%)	1095 (94%)	67 (6%)	4 (0%)	44	77
3	C	300/335 (90%)	283 (94%)	17 (6%)	0	100	100
4	D	55/137 (40%)	52 (94%)	3 (6%)	0	100	100
5	E	213/215 (99%)	202 (95%)	11 (5%)	0	100	100
6	F	98/155 (63%)	96 (98%)	2 (2%)	0	100	100
7	G	251/326 (77%)	232 (92%)	17 (7%)	2 (1%)	22	58
8	H	130/146 (89%)	119 (92%)	11 (8%)	0	100	100
9	I	118/125 (94%)	104 (88%)	12 (10%)	2 (2%)	11	42
10	J	67/70 (96%)	62 (92%)	5 (8%)	0	100	100
11	K	101/142 (71%)	95 (94%)	5 (5%)	1 (1%)	18	54
12	L	43/70 (61%)	40 (93%)	3 (7%)	0	100	100
13	M	101/415 (24%)	94 (93%)	7 (7%)	0	100	100
14	N	131/233 (56%)	122 (93%)	9 (7%)	0	100	100
All	All	4277/5236 (82%)	4017 (94%)	247 (6%)	13 (0%)	44	77

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1385	ASP
1	A	237	GLY
1	A	1389	GLU
2	B	1154	ASP
9	I	78	ASP

### 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	1343/1465 (92%)	1271 (95%)	72 (5%)	26	61
2	B	1024/1053 (97%)	927 (90%)	97 (10%)	10	35
3	C	269/296 (91%)	242 (90%)	27 (10%)	9	33
4	D	56/116 (48%)	50 (89%)	6 (11%)	8	30
5	E	197/197 (100%)	192 (98%)	5 (2%)	53	78
6	F	90/137 (66%)	86 (96%)	4 (4%)	33	67
7	G	234/291 (80%)	219 (94%)	15 (6%)	20	55
8	H	116/128 (91%)	111 (96%)	5 (4%)	33	67
9	I	109/110 (99%)	99 (91%)	10 (9%)	11	37
10	J	64/65 (98%)	56 (88%)	8 (12%)	5	23
11	K	93/130 (72%)	86 (92%)	7 (8%)	16	47
12	L	40/57 (70%)	37 (92%)	3 (8%)	16	47
13	M	94/371 (25%)	90 (96%)	4 (4%)	33	67
14	N	128/220 (58%)	117 (91%)	11 (9%)	12	41
All	All	3857/4636 (83%)	3583 (93%)	274 (7%)	17	50

5 of 274 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	588	ILE
2	B	965	GLU
11	K	99	ASN
2	B	684	ASN
2	B	840	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 62 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	361	HIS
2	B	755	ASN
9	I	100	GLN
2	B	646	HIS
2	B	975	HIS

### 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 7 ligands modelled in this entry, 7 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	1523/1664 (91%)	0.28	54 (3%) 44 40	110, 151, 210, 241	0
2	B	1176/1203 (97%)	0.46	58 (4%) 30 27	111, 155, 215, 239	0
3	C	304/335 (90%)	0.38	19 (6%) 21 19	147, 180, 216, 228	0
4	D	59/137 (43%)	0.10	4 (6%) 18 16	154, 230, 241, 253	0
5	E	215/215 (100%)	0.26	13 (6%) 23 20	131, 191, 235, 241	0
6	F	100/155 (64%)	0.01	0 100 100	121, 155, 195, 205	0
7	G	259/326 (79%)	0.57	30 (11%) 5 4	135, 213, 249, 261	0
8	H	134/146 (91%)	0.80	25 (18%) 1 1	153, 190, 223, 233	0
9	I	124/125 (99%)	0.18	5 (4%) 39 35	130, 171, 227, 234	0
10	J	69/70 (98%)	0.16	0 100 100	144, 160, 191, 218	0
11	K	103/142 (72%)	0.50	14 (13%) 3 3	133, 180, 208, 231	0
12	L	45/70 (64%)	-0.12	0 100 100	157, 192, 219, 224	0
13	M	105/415 (25%)	1.30	30 (28%) 1 1	214, 235, 251, 259	0
14	N	139/233 (59%)	1.14	33 (23%) 1 1	152, 236, 266, 275	0
All	All	4355/5236 (83%)	0.41	285 (6%) 20 18	110, 165, 235, 275	0

The worst 5 of 285 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
14	N	27	ASP	7.7
1	A	706	HIS	7.6
13	M	22	ALA	6.6
13	M	17	ASP	6.2
14	N	121	ILE	6.2

## 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
15	ZN	J	1070	1/1	1.00	0.26	0.14	148,148,148,148	0
15	ZN	L	1071	1/1	0.98	0.14	0.11	178,178,178,178	0
15	ZN	I	1126	1/1	0.95	0.16	-0.12	195,195,195,195	0
15	ZN	B	2204	1/1	0.98	0.17	-0.61	144,144,144,144	0
15	ZN	A	2665	1/1	0.98	0.13	-0.69	137,137,137,137	0
15	ZN	I	1127	1/1	0.98	0.12	-0.82	136,136,136,136	0
15	ZN	A	2664	1/1	0.96	0.12	-0.99	182,182,182,182	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.