



# Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 08:00 am BST

PDB ID : 5NWT  
Title : Crystal Structure of Escherichia coli RNA polymerase - Sigma54 Holoenzyme complex  
Authors : Zhang, X.; Buck, M.; Darbari, V.C.; Yang, Y.; Zhang, N.; Lu, D.; Glyde, R.; Wang, Y.; Winkelman, J.; Gourse, R.L.; Murakami, K.S.  
Deposited on : 2017-05-08  
Resolution : 3.76 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 2.11  
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) : 2.11

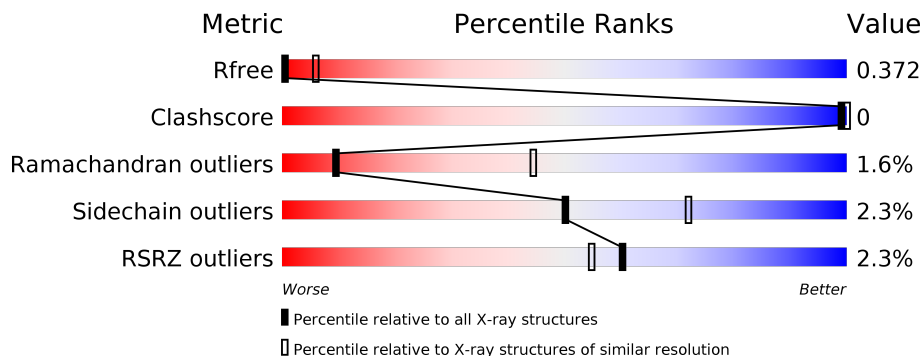
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.76 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	1039 (3.94-3.58)
Clashscore	141614	1051 (3.92-3.60)
Ramachandran outliers	138981	1015 (3.92-3.60)
Sidechain outliers	138945	1011 (3.92-3.60)
RSRZ outliers	127900	1050 (3.96-3.56)

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	329	
1	B	329	
2	C	1342	
3	D	1407	
4	E	91	
5	M	477	

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 22461 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	307	Total	C	N	O	S	0	0	0
			1888	1159	348	375	6			
1	B	215	Total	C	N	O	S	0	0	0
			1272	781	233	254	4			

- 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	1318	Total	C	N	O	S	0	0	0
			8247	5047	1515	1661	24			

- 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	1358	Total	C	N	O	S	0	0	0
			8092	4936	1535	1598	23			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	78	Total	C	N	O	S	0	0	0
			589	358	108	122	1			

- Molecule 5 is a protein called RNA polymerase sigma-54 factor, RNA polymerase sigma-54 factor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
5	M	431	Total	C	N	O	Se	51	0	0
			2370	1442	444	476	8			

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

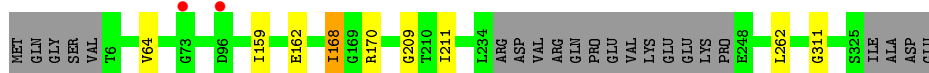
- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

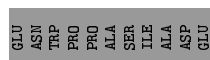
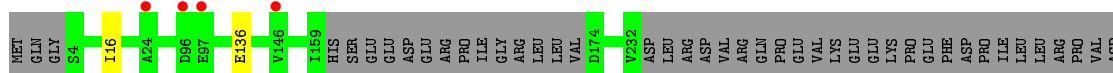
### 3 Residue-property plots [i](#)

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 ( $RSRZ > 2$ ). 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.

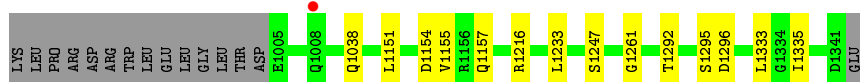
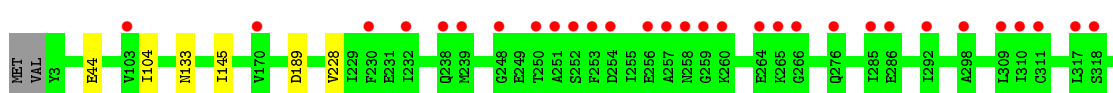
- Molecule 1: DNA-directed RNA polymerase subunit alpha



- Molecule 1: DNA-directed RNA polymerase subunit alpha

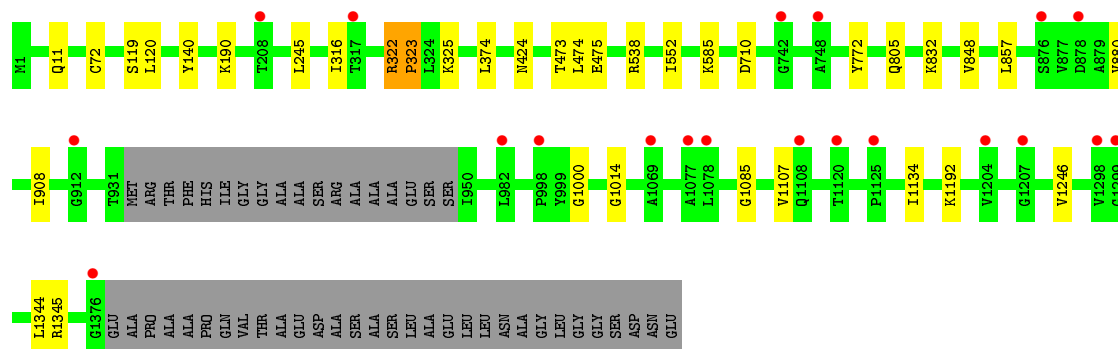


- Molecule 2: DNA-directed RNA polymerase subunit beta

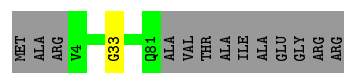
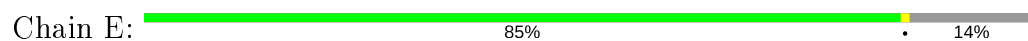


- Molecule 3: DNA-directed RNA polymerase subunit beta'

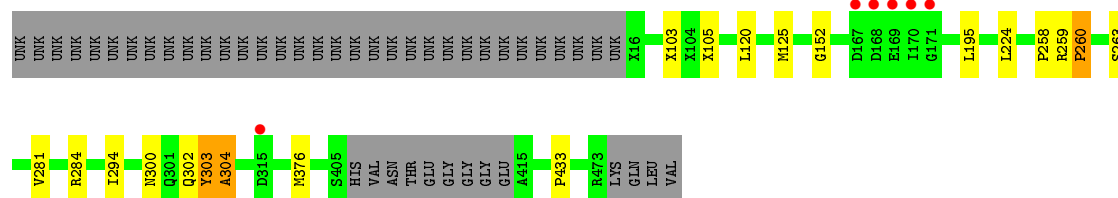
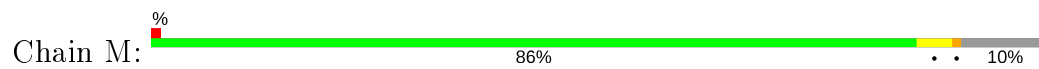




• Molecule 4: DNA-directed RNA polymerase subunit omega



• Molecule 5: RNA polymerase sigma-54 factor, RNA polymerase sigma-54 factor



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	208.48Å 151.53Å 195.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	195.28 – 3.76 29.64 – 3.76	Depositor EDS
% Data completeness (in resolution range)	98.2 (195.28-3.76) 98.5 (29.64-3.76)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.26 (at 3.75Å)	Xtrriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.313 , 0.370 0.316 , 0.372	Depositor DCC
$R_{free}$ test set	3077 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	142.0	Xtrriage
Anisotropy	0.740	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.22 , 166.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	22461	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	194.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.33% 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

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	0.33	0/1913	0.48	0/2643
1	B	0.32	0/1284	0.45	0/1779
2	C	0.34	0/8363	0.47	0/11499
3	D	0.33	0/8201	0.48	0/11314
4	E	0.37	0/591	0.51	0/801
5	M	0.34	0/1986	0.47	0/2746
All	All	0.34	0/22338	0.48	0/30782

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	1888	0	1450	1	0
1	B	1272	0	946	0	0
2	C	8247	0	6196	2	0
3	D	8092	0	5804	4	0
4	E	589	0	559	0	0
5	M	2370	0	1363	6	0
6	D	2	0	0	1	0

*Continued on next page...*



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	D	1	0	0	0	0
All	All	22461	0	16318	13	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 0.

All (13) 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)
5:M:152:GLY:O	5:M:260:PRO:HG3	1.65	0.96
3:D:72:CYS:SG	6:D:1501:ZN:ZN	1.81	0.70
5:M:152:GLY:HA3	5:M:260:PRO:HG2	1.81	0.61
5:M:152:GLY:HA3	5:M:260:PRO:CG	2.34	0.57
3:D:322:ARG:CB	3:D:323:PRO:CD	2.84	0.56
5:M:303:TYR:O	5:M:304:ALA:HB3	2.09	0.53
5:M:103:UNK:C	5:M:105:UNK:N	2.77	0.47
5:M:259:ARG:HA	5:M:260:PRO:HD2	1.68	0.45
2:C:145:ILE:HD12	2:C:145:ILE:N	2.33	0.44
1:A:159:ILE:O	1:A:159:ILE:HG23	2.17	0.43
3:D:322:ARG:CB	3:D:323:PRO:HD2	2.49	0.42
2:C:1333:LEU:CD1	2:C:1335:ILE:HD12	2.51	0.41
3:D:848:VAL:HG11	3:D:880:VAL:HG13	2.03	0.40

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	303/329 (92%)	260 (86%)	36 (12%)	7 (2%)	<b>6</b> 38
1	B	211/329 (64%)	186 (88%)	23 (11%)	2 (1%)	17 <b>54</b>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	1310/1342 (98%)	1156 (88%)	140 (11%)	14 (1%)	14	51
3	D	1354/1407 (96%)	1182 (87%)	152 (11%)	20 (2%)	10	45
4	E	76/91 (84%)	69 (91%)	6 (8%)	1 (1%)	12	48
5	M	349/477 (73%)	276 (79%)	59 (17%)	14 (4%)	3	28
All	All	3603/3975 (91%)	3129 (87%)	416 (12%)	58 (2%)	9	44

All (58) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	119	SER
1	B	16	ILE
2	C	333	ILE
3	D	120	LEU
3	D	322	ARG
3	D	323	PRO
3	D	325	LYS
3	D	1000	GLY
3	D	1085	GLY
3	D	1107	VAL
3	D	1192	LYS
4	E	33	GLY
5	M	258	PRO
5	M	263	SER
5	M	281	VAL
5	M	294	ILE
5	M	302	GLN
5	M	303	TYR
2	C	747	GLY
2	C	813	GLU
2	C	1157	GLN
2	C	1292	THR
3	D	805	GLN
3	D	1344	LEU
5	M	120	LEU
5	M	224	LEU
5	M	300	ASN
1	A	162	GLU
1	A	168	ILE
1	A	211	ILE
2	C	1154	ASP
2	C	1295	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	D	140	TYR
3	D	1246	VAL
3	D	1345	ARG
5	M	304	ALA
1	A	170	ARG
1	A	311	GLY
1	B	136	GLU
2	C	555	TYR
2	C	851	THR
2	C	1151	LEU
2	C	1155	VAL
2	C	1247	SER
3	D	585	LYS
3	D	908	ILE
3	D	1134	ILE
5	M	195	LEU
5	M	284	ARG
5	M	433	PRO
3	D	316	ILE
2	C	228	VAL
1	A	64	VAL
1	A	209	GLY
2	C	1261	GLY
3	D	1014	GLY
3	D	552	ILE
5	M	260	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	127/286 (44%)	125 (98%)	2 (2%)	62 80
1	B	77/286 (27%)	77 (100%)	0	100 100
2	C	546/1157 (47%)	532 (97%)	14 (3%)	46 70
3	D	439/1168 (38%)	426 (97%)	13 (3%)	41 66

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	E	59/75 (79%)	59 (100%)	0	100	100
5	M	82/314 (26%)	80 (98%)	2 (2%)	49	71
All	All	1330/3286 (40%)	1299 (98%)	31 (2%)	50	72

All (31) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	168	ILE
1	A	262	LEU
2	C	44	GLU
2	C	104	ILE
2	C	133	ASN
2	C	189	ASP
2	C	487	LEU
2	C	553	THR
2	C	554	HIS
2	C	816	ILE
2	C	912	ASP
2	C	965	GLN
2	C	1038	GLN
2	C	1216	ARG
2	C	1233	LEU
2	C	1296	ASP
3	D	11	GLN
3	D	190	LYS
3	D	245	LEU
3	D	374	LEU
3	D	424	ASN
3	D	473	THR
3	D	474	LEU
3	D	475	GLU
3	D	538	ARG
3	D	710	ASP
3	D	772	TYR
3	D	832	LYS
3	D	857	LEU
5	M	125	MSE
5	M	376	MSE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	450	ASN
3	D	11	GLN
3	D	861	ASN
3	D	1367	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 3 ligands modelled in this entry, 3 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	C	2
5	M	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	M	66:UNK	C	85:UNK	N	16.93
1	C	336:LEU	C	337:PHE	N	3.66
1	C	225:PHE	C	226:GLU	N	3.29

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

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	307/329 (93%)	-0.45	2 (0%) 87 86	91, 180, 277, 375	0
1	B	215/329 (65%)	-0.26	4 (1%) 66 62	141, 222, 314, 352	0
2	C	1318/1342 (98%)	-0.16	51 (3%) 39 34	78, 189, 358, 534	0
3	D	1358/1407 (96%)	-0.31	20 (1%) 73 69	79, 189, 316, 397	0
4	E	78/91 (85%)	-0.23	0 100 100	127, 187, 317, 381	0
5	M	341/477 (71%)	-0.64	6 (1%) 68 64	30, 141, 223, 314	6 (1%)
All	All	3617/3975 (90%)	-0.30	83 (2%) 60 55	30, 185, 323, 534	6 (0%)

All (83) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	252	SER	18.1
2	C	253	PHE	13.4
2	C	265	LYS	8.4
2	C	254	ASP	7.6
2	C	251	ALA	7.5
2	C	332	ARG	7.4
2	C	333	ILE	7.3
2	C	257	ALA	7.0
2	C	318	SER	6.4
2	C	238	GLN	6.0
2	C	310	ILE	5.2
1	B	96	ASP	5.1
2	C	334	GLU	5.1
3	D	317	THR	4.9
3	D	1299	GLY	4.4
2	C	350	THR	4.2
2	C	726	TYR	4.2
1	A	96	ASP	3.7
2	C	317	LEU	3.7

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
3	D	878	ASP	3.6
2	C	103	VAL	3.6
2	C	342	ASP	3.6
5	M	169	GLU	3.5
3	D	1120	THR	3.5
2	C	264	GLU	3.5
2	C	309	LEU	3.5
5	M	168	ASP	3.4
2	C	391	SER	3.3
2	C	276	GLN	3.3
3	D	876	SER	3.3
1	B	97	GLU	3.2
3	D	1376	GLY	3.1
2	C	239	MET	3.1
2	C	421	SER	3.1
2	C	266	GLY	3.1
2	C	311	CYS	3.1
2	C	727	VAL	3.0
2	C	259	GLY	3.0
2	C	170	VAL	3.0
3	D	1078	LEU	3.0
3	D	1108	GLN	2.9
3	D	1298	VAL	2.9
2	C	258	ASN	2.8
2	C	341	LEU	2.8
2	C	614	TYR	2.8
2	C	248	GLY	2.7
2	C	256	GLU	2.7
2	C	230	PHE	2.7
3	D	998	PRO	2.7
2	C	285	ILE	2.6
2	C	250	THR	2.6
5	M	171	GLY	2.6
1	B	146	VAL	2.6
2	C	323	ALA	2.6
2	C	298	ALA	2.6
3	D	748	ALA	2.5
2	C	292	ILE	2.5
3	D	912	GLY	2.5
2	C	652	TYR	2.5
2	C	345	PRO	2.4
2	C	344	GLY	2.4

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
3	D	1125	PRO	2.4
5	M	170	ILE	2.4
3	D	1069	ALA	2.3
2	C	655	VAL	2.2
2	C	232	ILE	2.2
5	M	315	ASP	2.2
1	B	24	ALA	2.2
2	C	1008	GLN	2.2
1	A	73	GLY	2.2
2	C	338	THR	2.2
3	D	1207	GLY	2.1
5	M	167	ASP	2.1
3	D	982	LEU	2.1
2	C	260	LYS	2.1
2	C	420	LEU	2.1
3	D	1204	VAL	2.1
3	D	208	THR	2.1
3	D	1077	ALA	2.1
2	C	286	GLU	2.1
2	C	604	HIS	2.0
3	D	742	GLY	2.0
2	C	347	ILE	2.0

## 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. 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
7	MG	D	1502	1/1	0.85	0.38	163,163,163,163	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
6	ZN	D	1503	1/1	0.88	0.17	156,156,156,156	0
6	ZN	D	1501	1/1	0.91	0.12	120,120,120,120	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.