wwPDB X-ray Structure Validation Summary Report (i)

Sep 25, 2023 – 05:42 AM EDT

PDB ID	:	5VOI
Title	:	X-ray crystal structure of bacterial RNA polymerase and pyrG promoter complex
Authors	:	Murakami, K.S.; Shin, Y.; Turnbough Jr, C.L.; Molodtsov, V.
Deposited on	:	2017-05-02
Resolution	:	2.80 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

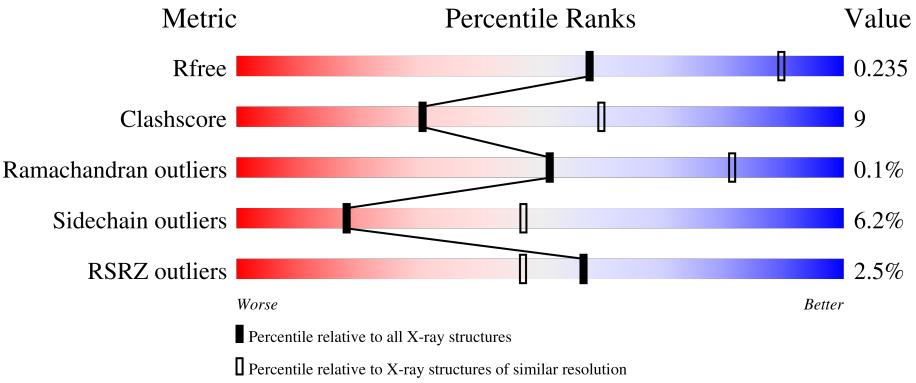
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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.



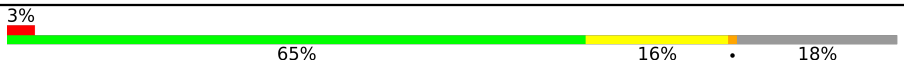
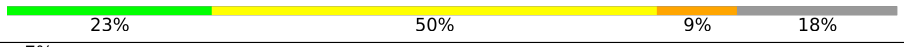
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	315	54% 17% 28%
1	B	315	% 55% 14% 30%
2	C	1119	2% 73% 23%
3	D	1524	3% 73% 22%
4	E	99	84% 11% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	423	
6	G	22	
7	H	27	

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
8	ZN	D	2002	-	-	X	-

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 28432 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	226	Total	C	N	O	S	0	0	0
			1782	1138	310	332	2			
1	B	222	Total	C	N	O	S	0	0	0
			1750	1118	304	326	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1111	Total	C	N	O	S	0	0	0
			8770	5548	1564	1634	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1486	Total	C	N	O	S	0	0	0
			11738	7440	2067	2195	36			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	94	Total	C	N	O	S	0	0	0
			761	486	132	139	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	346	Total	C	N	O	S	0	0	0
			2807	1770	509	524	4			

- Molecule 6 is a DNA chain called PyrG promoter.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	18	Total	C	N	O	P	0	0	0
			368	175	71	104	18			

- Molecule 7 is a DNA chain called PyrG promoter.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	22	Total	C	N	O	P	0	0	0
			451	216	84	130	21			

- 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).

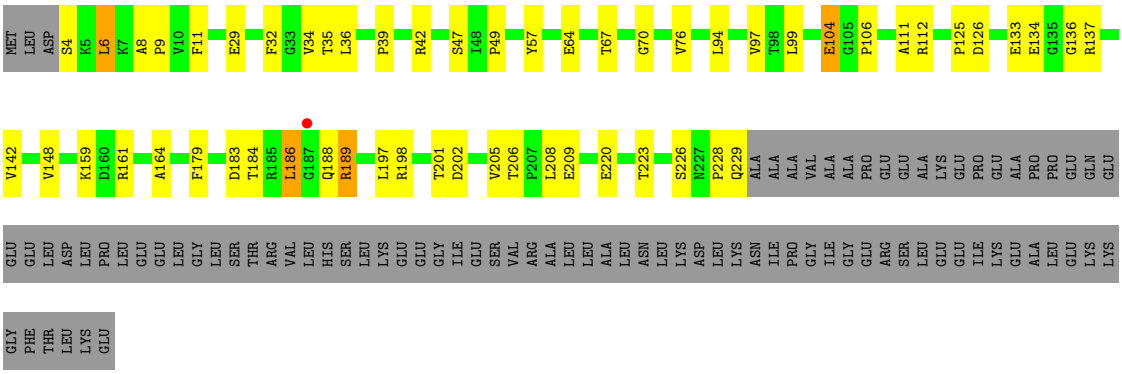
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Mg	0	0
			2	2		
9	G	1	Total	Mg	0	0
			1	1		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

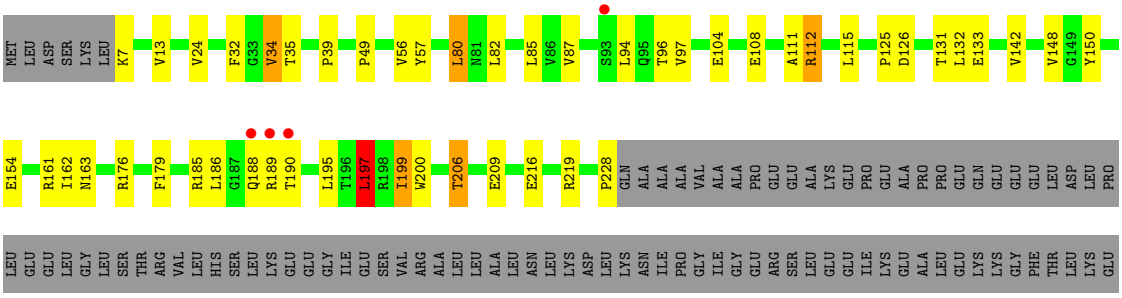
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



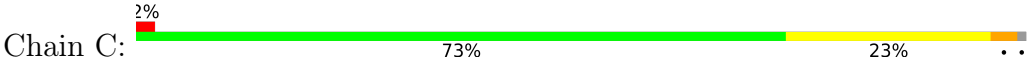
• Molecule 1: DNA-directed RNA polymerase subunit alpha

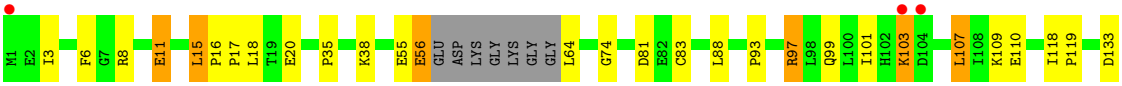
Chain B:

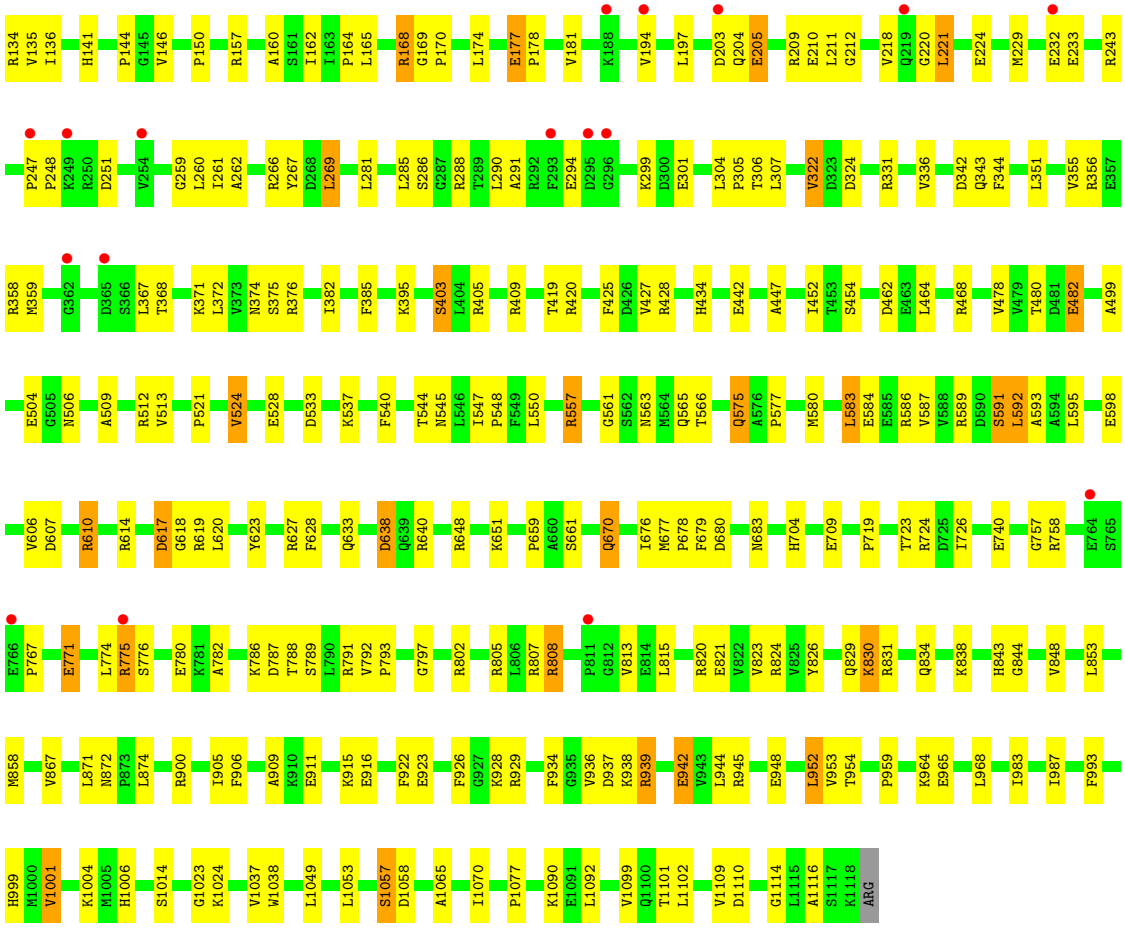


• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain C:

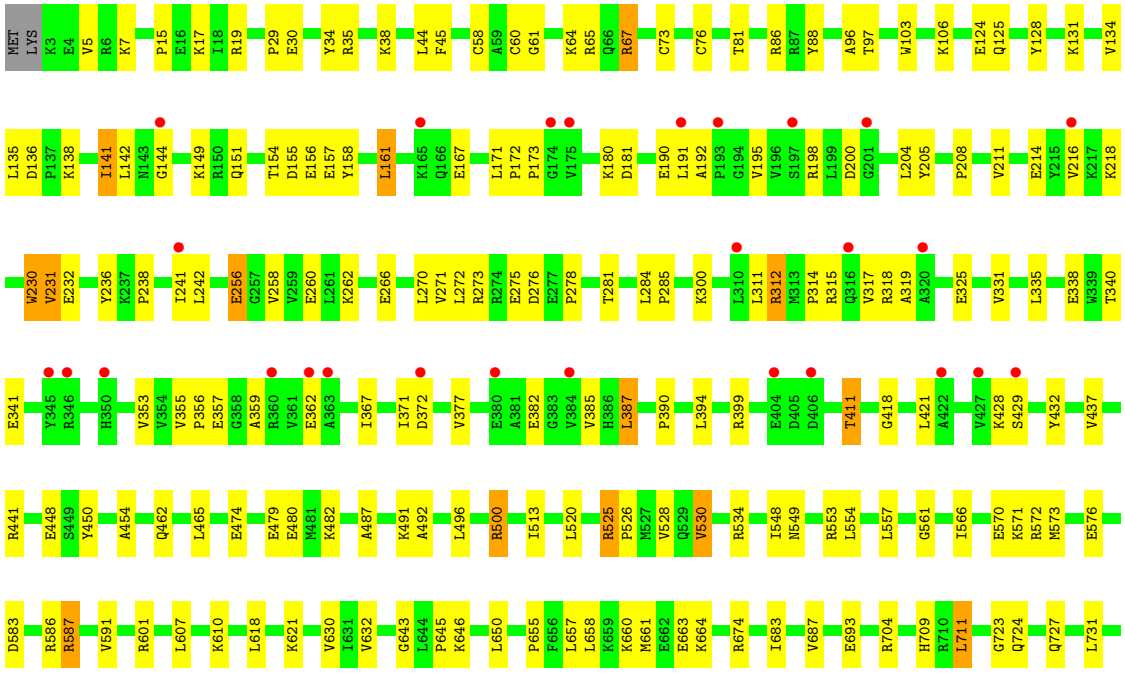


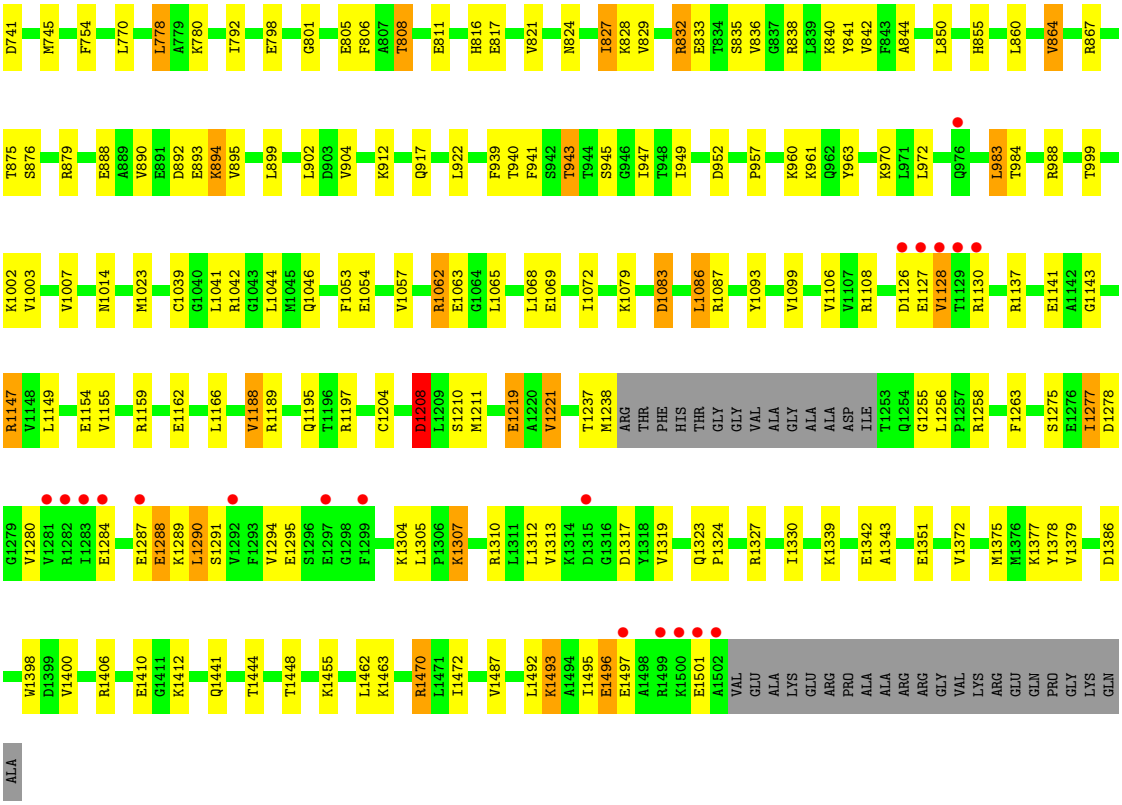




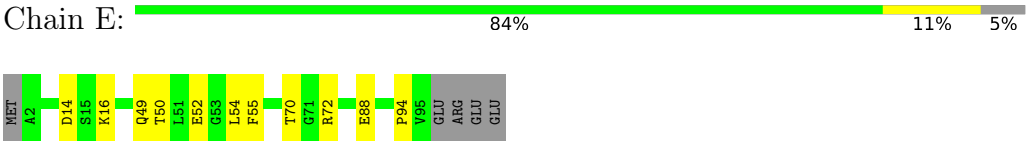
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:

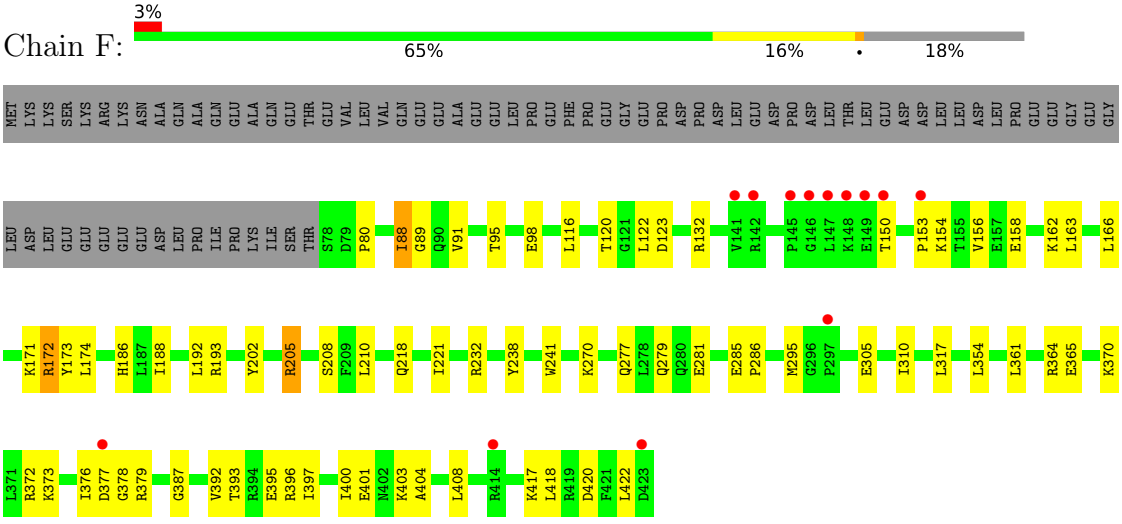




• Molecule 4: DNA-directed RNA polymerase subunit omega

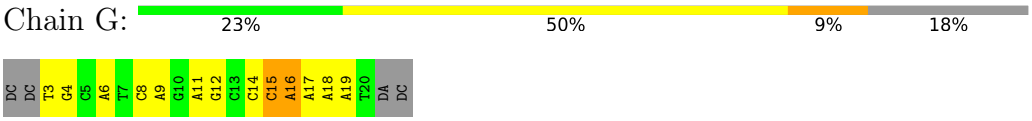


• Molecule 5: RNA polymerase sigma factor SigA



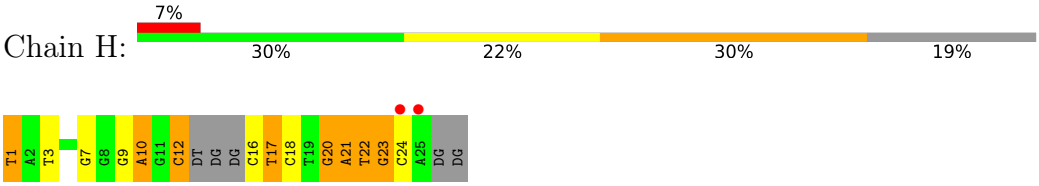
• Molecule 6: PyrG promoter

Chain G:



• Molecule 7: PyrG promoter

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	183.18Å 103.47Å 294.44Å 90.00° 99.14° 90.00°	Depositor
Resolution (Å)	29.74 – 2.80 29.74 – 2.80	Depositor EDS
% Data completeness (in resolution range)	98.2 (29.74-2.80) 98.1 (29.74-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.30 (at 2.80Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.192 , 0.236 0.193 , 0.235	Depositor DCC
R_{free} test set	2000 reflections (1.52%)	wwPDB-VP
Wilson B-factor (Å^2)	56.5	Xtriage
Anisotropy	0.372	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 49.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.014 for 1/2*h-3/2*k,-1/2*h-1/2*k,-1/2*h+1/2*k-l 0.014 for 1/2*h+3/2*k,1/2*h-1/2*k,-1/2*h-1/2*k-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	28432	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.90% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: MG, 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	0.55	0/1814	0.78	0/2466
1	B	0.55	1/1782 (0.1%)	0.83	1/2424 (0.0%)
2	C	0.58	0/8937	0.82	5/12087 (0.0%)
3	D	0.59	1/11944 (0.0%)	0.81	6/16148 (0.0%)
4	E	0.55	0/775	0.77	0/1045
5	F	0.51	0/2852	0.73	0/3837
6	G	1.62	7/413 (1.7%)	1.24	3/634 (0.5%)
7	H	1.49	8/505 (1.6%)	1.45	8/776 (1.0%)
All	All	0.63	17/29022 (0.1%)	0.83	23/39417 (0.1%)

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	H	1	DT	C1'-N1	9.36	1.61	1.49
6	G	15	DC	N1-C2	8.71	1.48	1.40
6	G	12	DG	C3'-O3'	-8.08	1.33	1.44
7	H	17	DT	C1'-N1	7.82	1.59	1.49
6	G	15	DC	C4-C5	7.56	1.49	1.43

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	H	23	DG	O4'-C4'-C3'	-9.00	100.60	106.00
6	G	15	DC	O4'-C1'-N1	7.94	113.56	108.00
7	H	20	DG	O4'-C1'-N9	-7.58	102.70	108.00
7	H	10	DA	P-O3'-C3'	7.33	128.49	119.70
1	B	197	LEU	CA-CB-CG	7.17	131.79	115.30

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	1782	0	1834	41	0
1	B	1750	0	1797	38	0
2	C	8770	0	8874	187	0
3	D	11738	0	11971	233	0
4	E	761	0	778	7	0
5	F	2807	0	2882	46	0
6	G	368	0	202	6	0
7	H	451	0	251	15	0
8	D	2	0	0	2	0
9	D	2	0	0	0	0
9	G	1	0	0	0	0
All	All	28432	0	28589	520	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 9.

The worst 5 of 520 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)
1:B:206:THR:HG22	1:B:209:GLU:H	1.29	0.98
2:C:628:PHE:H	2:C:638:ASP:HB2	1.34	0.92
3:D:76:CYS:HG	8:D:2002:ZN:ZN	0.82	0.91
1:B:112:ARG:NH1	1:B:126:ASP:OD1	2.03	0.90
3:D:61:GLY:O	3:D:64:LYS:NZ	2.08	0.85

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	224/315 (71%)	222 (99%)	2 (1%)	0	100 100
1	B	220/315 (70%)	215 (98%)	5 (2%)	0	100 100
2	C	1107/1119 (99%)	1084 (98%)	23 (2%)	0	100 100
3	D	1482/1524 (97%)	1457 (98%)	24 (2%)	1 (0%)	51 81
4	E	92/99 (93%)	89 (97%)	2 (2%)	1 (1%)	14 41
5	F	344/423 (81%)	341 (99%)	3 (1%)	0	100 100
All	All	3469/3795 (91%)	3408 (98%)	59 (2%)	2 (0%)	51 81

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	E	94	PRO
3	D	530	VAL

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	199/273 (73%)	189 (95%)	10 (5%)	24 56
1	B	195/273 (71%)	186 (95%)	9 (5%)	27 60
2	C	936/941 (100%)	872 (93%)	64 (7%)	16 42
3	D	1253/1279 (98%)	1167 (93%)	86 (7%)	15 41
4	E	83/88 (94%)	82 (99%)	1 (1%)	71 92
5	F	301/371 (81%)	286 (95%)	15 (5%)	24 56
All	All	2967/3225 (92%)	2782 (94%)	185 (6%)	18 47

5 of 185 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	572	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	D	1083	ASP
3	D	610	LYS
3	D	832	ARG
3	D	1195	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1195	GLN
3	D	1359	GLN
5	F	279	GLN
1	B	81	ASN
1	A	128	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 5 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

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^{th} 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^2)	Q<0.9
1	A	226/315 (71%)	-0.51	1 (0%) 92 91	43, 64, 95, 117	0
1	B	222/315 (70%)	-0.35	4 (1%) 68 61	43, 76, 114, 141	0
2	C	1111/1119 (99%)	-0.34	20 (1%) 68 61	23, 56, 125, 169	0
3	D	1486/1524 (97%)	-0.16	47 (3%) 47 37	23, 64, 127, 154	0
4	E	94/99 (94%)	-0.53	0 100 100	31, 58, 96, 129	0
5	F	346/423 (81%)	-0.10	13 (3%) 40 30	38, 86, 133, 171	0
6	G	18/22 (81%)	0.03	0 100 100	46, 75, 164, 175	0
7	H	22/27 (81%)	-0.06	2 (9%) 9 5	58, 91, 140, 171	0
All	All	3525/3844 (91%)	-0.25	87 (2%) 57 47	23, 65, 125, 175	0

The worst 5 of 87 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	1128	VAL	5.4
3	D	1297	GLU	5.1
5	F	147	LEU	4.8
3	D	144	GLY	4.7
5	F	149	GLU	4.6

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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^{th} 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(Å^2)	Q<0.9
9	MG	G	101	1/1	0.86	0.58	91,91,91,91	0
9	MG	D	2004	1/1	0.93	0.48	52,52,52,52	0
8	ZN	D	2002	1/1	0.94	0.33	238,238,238,238	0
9	MG	D	2003	1/1	0.97	0.19	44,44,44,44	0
8	ZN	D	2001	1/1	0.98	0.18	77,77,77,77	0

6.5 Other polymers

There are no such residues in this entry.