W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 20, 2020 – 01:14 PM BST

PDB ID	:	6KOO
Title	:	Mycobacterium tuberculosis initial transcription complex comprising sigma H and 5'-OH RNA of 7 nt
Authors	:	Li, L.; Zhang, Y.
Deposited on	:	2019-08-12
Resolution	:	2.80 Å(reported)

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

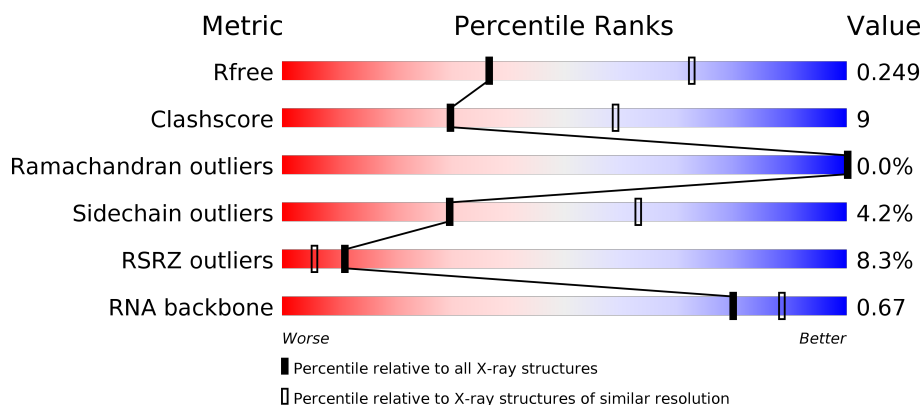|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 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	Percentile Ranks	Value
Rfree		0.249
Clashscore		9
Ramachandran outliers		0.0%
Sidechain outliers		4.2%
RSRZ outliers		8.3%
RNA backbone		0.67

Worse   *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

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)
RNA backbone	3102	1227 (3.10-2.50)

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 >=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	368	49% 9% 41%
1	B	368	10% 46% 15% • 38%
2	C	1174	8% 72% 23% • •
3	D	1317	6% 73% 21% • •

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	110	16% / 53% / 15% / 32%
5	F	218	14% / 67% / 16% / 17%
6	G	23	17% / 70% / 30%
7	H	21	10% / 29% / 67% / 5%
8	I	7	86% / 14%

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 24837 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	217	Total	C	N	O	S	0	0	0
			1646	1035	285	324	2			
1	B	230	Total	C	N	O	S	0	0	0
			1670	1057	283	328	2			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	initiating methionine	UNP P9WGZ1
A	-19	GLY	-	expression tag	UNP P9WGZ1
A	-18	HIS	-	expression tag	UNP P9WGZ1
A	-17	HIS	-	expression tag	UNP P9WGZ1
A	-16	HIS	-	expression tag	UNP P9WGZ1
A	-15	HIS	-	expression tag	UNP P9WGZ1
A	-14	HIS	-	expression tag	UNP P9WGZ1
A	-13	HIS	-	expression tag	UNP P9WGZ1
A	-12	HIS	-	expression tag	UNP P9WGZ1
A	-11	HIS	-	expression tag	UNP P9WGZ1
A	-10	HIS	-	expression tag	UNP P9WGZ1
A	-9	HIS	-	expression tag	UNP P9WGZ1
A	-8	SER	-	expression tag	UNP P9WGZ1
A	-7	SER	-	expression tag	UNP P9WGZ1
A	-6	GLY	-	expression tag	UNP P9WGZ1
A	-5	HIS	-	expression tag	UNP P9WGZ1
A	-4	ILE	-	expression tag	UNP P9WGZ1
A	-3	GLU	-	expression tag	UNP P9WGZ1
A	-2	GLY	-	expression tag	UNP P9WGZ1
A	-1	ARG	-	expression tag	UNP P9WGZ1
A	0	HIS	-	expression tag	UNP P9WGZ1
B	-20	MET	-	initiating methionine	UNP P9WGZ1
B	-19	GLY	-	expression tag	UNP P9WGZ1
B	-18	HIS	-	expression tag	UNP P9WGZ1
B	-17	HIS	-	expression tag	UNP P9WGZ1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	HIS	-	expression tag	UNP P9WGZ1
B	-15	HIS	-	expression tag	UNP P9WGZ1
B	-14	HIS	-	expression tag	UNP P9WGZ1
B	-13	HIS	-	expression tag	UNP P9WGZ1
B	-12	HIS	-	expression tag	UNP P9WGZ1
B	-11	HIS	-	expression tag	UNP P9WGZ1
B	-10	HIS	-	expression tag	UNP P9WGZ1
B	-9	HIS	-	expression tag	UNP P9WGZ1
B	-8	SER	-	expression tag	UNP P9WGZ1
B	-7	SER	-	expression tag	UNP P9WGZ1
B	-6	GLY	-	expression tag	UNP P9WGZ1
B	-5	HIS	-	expression tag	UNP P9WGZ1
B	-4	ILE	-	expression tag	UNP P9WGZ1
B	-3	GLU	-	expression tag	UNP P9WGZ1
B	-2	GLY	-	expression tag	UNP P9WGZ1
B	-1	ARG	-	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
2	C	1135	Total	C	N	O	S	0	0	0
			8673	5428	1519	1687	39			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	MET	-	initiating methionine	UNP P9WGY9
C	0	VAL	-	expression tag	UNP P9WGY9

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1258	Total	C	N	O	S	0	0	0
			9763	6118	1766	1838	41			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	MET	-	initiating methionine	UNP P9WGY7
D	1	VAL	-	expression tag	UNP P9WGY7

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	E	75	Total	C	N	O	0	0	0
			592	378	99	115			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	182	Total	C	N	O	S	0	0	0
			1432	896	251	280	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	GLY	-	expression tag	UNP P9WGH9
F	0	ALA	-	expression tag	UNP P9WGH9

- Molecule 6 is a DNA chain called DNA (5'-D(*TP*TP*GP*TP*GP*GP*GP*AP*GP*CP* TP*GP*TP*CP*AP*CP*GP*GP*AP*TP*GP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	23	Total	C	N	O	P	0	0	0
			475	226	89	138	22			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	20	Total	C	N	O	P	0	0	0
			412	196	77	120	19			

- Molecule 8 is a RNA chain called RNA (5'-R(*CP*CP*CP*UP*CP*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	7	Total	C	N	O	P	0	0	0
			142	65	24	47	6			

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

• Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

• Molecule 11 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	C	8	Total 8	O 8	0	0
11	D	12	Total 12	O 12	0	0
11	F	4	Total 4	O 4	0	0
11	H	3	Total 3	O 3	0	0
11	I	2	Total 2	O 2	0	0

3 Residue-property plots ⓘ

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.

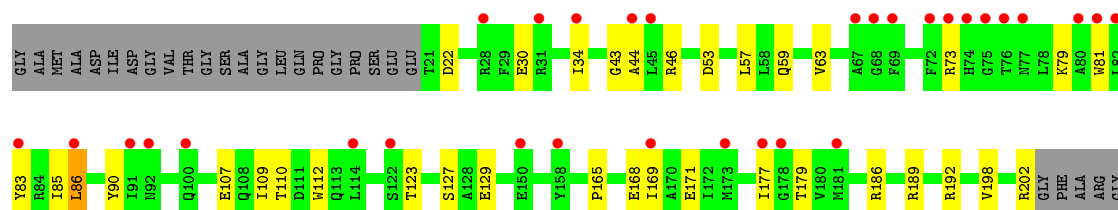- Molecule 1: DNA-directed RNA polymerase subunit alpha
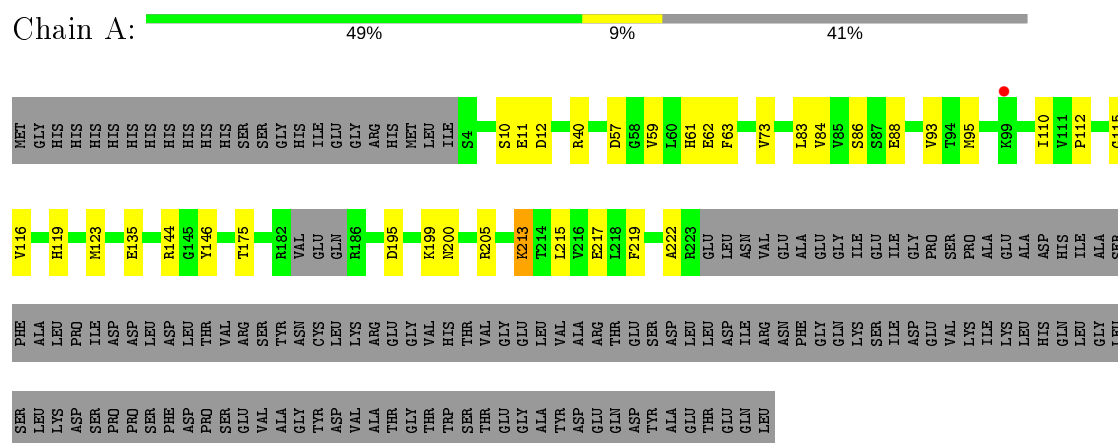
- Molecule 1: DNA-directed RNA polymerase subunit alpha
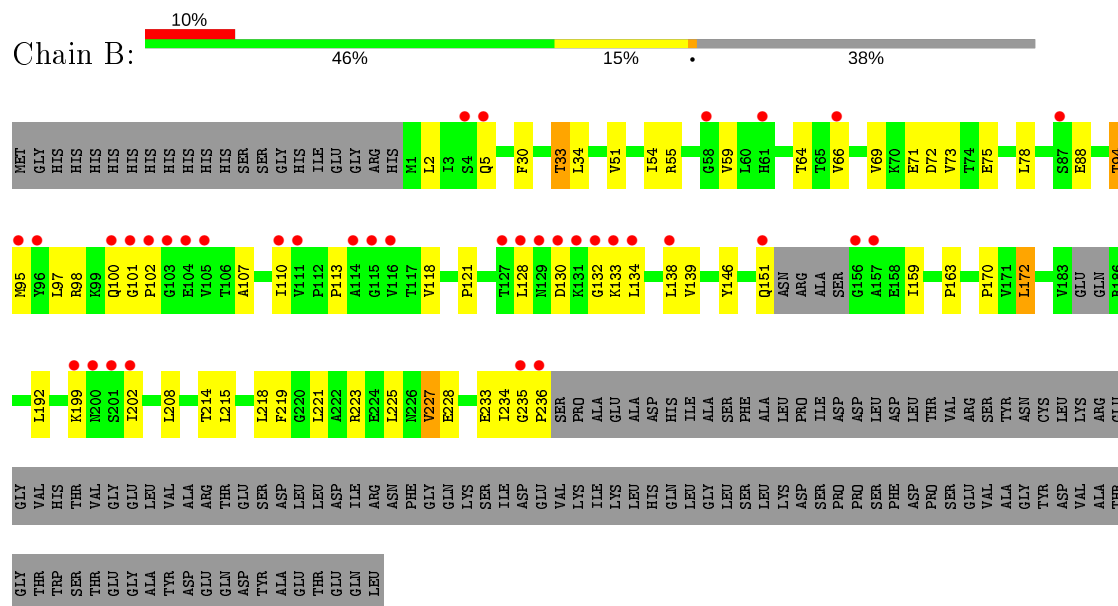
- Molecule 2: DNA-directed RNA polymerase subunit beta

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

Chain D:

• Molecule 4: DNA-directed RNA polymerase subunit omega

• Molecule 5: ECF RNA polymerase sigma factor SigH
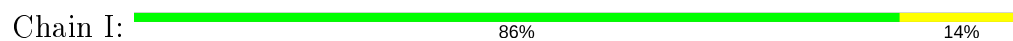
GLU GLN ALA HIS GLU GLY VAL SER SER

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

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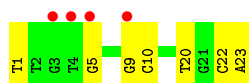
• Molecule 7: DNA (5'-D(*TP*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*GP*TP*CP*GP*AP *GP*GP*GP*TP*G)-3')

Chain H:

• Molecule 8: RNA (5'-R(*CP*CP*CP*UP*CP*GP*A)-3')

Chain I:

4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	125.03Å 162.31Å 128.48Å 90.00° 117.03° 90.00°	Depositor
Resolution (Å)	31.38 – 2.80 48.67 – 2.79	Depositor EDS
% Data completeness (in resolution range)	96.9 (31.38-2.80) 97.1 (48.67-2.79)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.77 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.206 , 0.248 0.207 , 0.249	Depositor DCC
R_{free} test set	2398 reflections (2.19%)	wwPDB-VP
Wilson B-factor (Å^2)	77.6	Xtriage
Anisotropy	0.424	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 50.8	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.016 for l,-k,h	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	24837	wwPDB-VP
Average B, all atoms (Å^2)	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.35% 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 (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	0.32	0/1671	0.54	0/2273
1	B	0.32	0/1695	0.51	0/2314
2	C	0.35	0/8830	0.51	1/11988 (0.0%)
3	D	0.33	0/9923	0.50	0/13425
4	E	0.27	0/604	0.48	0/822
5	F	0.29	0/1460	0.43	0/1983
6	G	0.65	0/533	0.92	0/823
7	H	0.67	0/462	0.81	0/713
8	I	0.48	0/157	1.07	0/242
All	All	0.35	0/25335	0.53	1/34583 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	75	ASN	C-N-CD	5.77	140.51	128.40

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	1646	0	1676	28	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	1670	0	1651	43	0
2	C	8673	0	8508	177	0
3	D	9763	0	9772	195	0
4	E	592	0	587	11	0
5	F	1432	0	1359	22	0
6	G	475	0	261	6	0
7	H	412	0	227	22	0
8	I	142	0	78	1	0
9	D	2	0	0	0	0
10	D	1	0	0	0	0
11	C	8	0	0	0	0
11	D	12	0	0	0	0
11	F	4	0	0	0	0
11	H	3	0	0	0	0
11	I	2	0	0	0	0
All	All	24837	0	24119	458	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 458 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)
7:H:14:DC:H2'	7:H:15:DG:H5"	1.20	1.11
7:H:14:DC:C2'	7:H:15:DG:H5"	1.96	0.95
3:D:1090:LYS:HG3	3:D:1095:SER:O	1.75	0.87
3:D:1090:LYS:HG3	3:D:1095:SER:C	1.96	0.86
7:H:19:DG:C2'	7:H:20:DT:H5'	2.06	0.86

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	213/368 (58%)	208 (98%)	5 (2%)	0	100	100
1	B	224/368 (61%)	211 (94%)	13 (6%)	0	100	100
2	C	1131/1174 (96%)	1082 (96%)	48 (4%)	1 (0%)	51	81
3	D	1250/1317 (95%)	1211 (97%)	39 (3%)	0	100	100
4	E	71/110 (64%)	68 (96%)	3 (4%)	0	100	100
5	F	180/218 (83%)	177 (98%)	3 (2%)	0	100	100
All	All	3069/3555 (86%)	2957 (96%)	111 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	76	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	185/315 (59%)	182 (98%)	3 (2%)	62	88
1	B	177/315 (56%)	170 (96%)	7 (4%)	31	65
2	C	929/995 (93%)	888 (96%)	41 (4%)	28	61
3	D	1026/1096 (94%)	981 (96%)	45 (4%)	28	61
4	E	64/90 (71%)	63 (98%)	1 (2%)	62	88
5	F	143/175 (82%)	135 (94%)	8 (6%)	21	51
All	All	2524/2986 (84%)	2419 (96%)	105 (4%)	30	63

5 of 105 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	1088	ASP
3	D	308	SER
5	F	22	ASP
2	C	1134	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	D	150	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	144	GLN
3	D	854	HIS

5.3.3   RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	6/7 (85%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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 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)

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, 95th 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	217/368 (58%)	0.03	1 (0%)	91	88	53, 75, 108, 128	0
1	B	230/368 (62%)	0.76	37 (16%)	1	1	75, 112, 147, 156	0
2	C	1135/1174 (96%)	0.42	93 (8%)	11	6	47, 75, 147, 169	0
3	D	1258/1317 (95%)	0.26	76 (6%)	21	14	48, 79, 122, 153	0
4	E	75/110 (68%)	1.03	18 (24%)	0	0	73, 99, 138, 175	0
5	F	182/218 (83%)	1.02	31 (17%)	1	1	52, 103, 132, 147	12 (6%)
6	G	23/23 (100%)	0.68	4 (17%)	1	1	87, 118, 154, 161	3 (13%)
7	H	20/21 (95%)	0.27	2 (10%)	7	4	54, 78, 136, 137	0
8	I	7/7 (100%)	-0.32	0	100	100	53, 54, 81, 87	0
All	All	3147/3606 (87%)	0.40	262 (8%)	11	6	47, 81, 139, 175	15 (0%)

The worst 5 of 262 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	156	GLY	11.8
1	B	236	PRO	9.8
2	C	262	VAL	8.7
3	D	65	TYR	7.4
5	F	76	THR	7.4

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, 95th 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	ZN	D	2002	1/1	0.94	0.08	105,105,105,105	0
10	MG	D	2003	1/1	0.94	0.25	54,54,54,54	0
9	ZN	D	2001	1/1	0.99	0.12	87,87,87,87	0

6.5 Other polymers (i)

There are no such residues in this entry.