



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:22 PM GMT

PDB ID : 4I3H  
Title : A three-gate structure of topoisomerase IV from *Streptococcus pneumoniae*  
Authors : Laponogov, I.; Veselkov, D.A.; Pan, X.-S.; Crevel, I.; Fisher, L.M.; Sanderson, M.R.  
Deposited on : 2012-11-26  
Resolution : 3.70 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

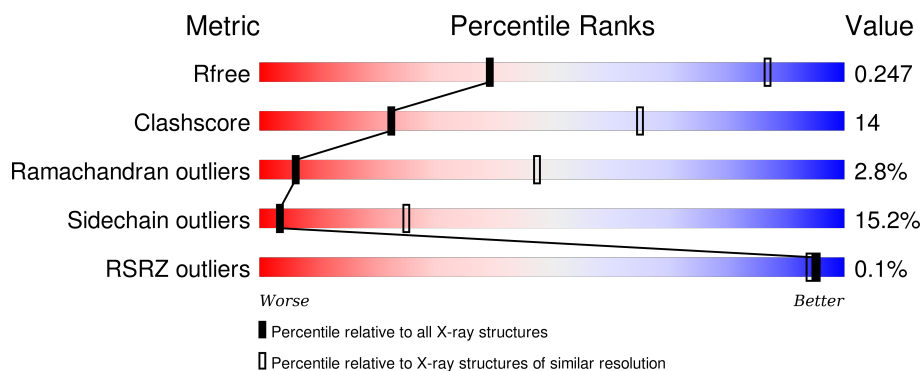
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.70 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))
$R_{free}$	91344	1101 (3.90-3.50)
Clashscore	102246	1224 (3.90-3.50)
Ramachandran outliers	100387	1172 (3.90-3.50)
Sidechain outliers	100360	1170 (3.90-3.50)
RSRZ outliers	91569	1108 (3.90-3.50)

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  $\geq 3$ , 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	E	34	
1	G	34	
2	F	34	
2	H	34	
3	A	1144	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	B	1144	 A horizontal bar chart showing the quality of chain B. The bar is divided into three segments: green (62%), yellow (24%), and grey (10%). The percentages are labeled below the bar. A small black dot is visible on the grey segment.

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 16228 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 DNA chain called DNA (5'-D(\*CP\*AP\*AP\*AP\*GP\*GP\*CP\*GP\*GP\*TP\*AP\*AP\*TP\*AP\*CP\*GP\*GP\*TP\*TP\*AP\*TP\*CP\*CP\*AP\*CP\*AP\*GP\*AP\*AP\*TP\*CP\*AP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	14	Total	C	N	O	P	0	0	0
			282	133	52	83	14			
1	G	20	Total	C	N	O	P	0	0	0
			407	191	77	119	20			

- Molecule 2 is a DNA chain called DNA (5'-D(\*CP\*CP\*TP\*GP\*AP\*TP\*TP\*CP\*TP\*GP\*TP\*GP\*GP\*AP\*TP\*AP\*AP\*CP\*CP\*GP\*TP\*AP\*TP\*TP\*AP\*CP\*CP\*GP\*CP\*CP\*TP\*TP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	14	Total	C	N	O	P	0	0	0
			283	133	51	85	14			
2	H	20	Total	C	N	O	P	0	0	0
			402	189	72	121	20			

- Molecule 3 is a protein called Topoisomerase IV subunit B, DNA topoisomerase 4 subunit A chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	1037	Total	C	N	O	S	0	0	0
			7477	4740	1286	1430	21			
3	B	1035	Total	C	N	O	S	0	0	0
			7372	4657	1280	1413	22			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1000	HIS	-	LINKER	UNP Q3HZ71
B	1000	HIS	-	LINKER	UNP Q3HZ71

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	H	1	Total	Mg	0	0
			1	1		
4	B	1	Total	Mg	0	0
			1	1		
4	A	1	Total	Mg	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	E	2	Total	O	0	0
			2	2		

### 3 Residue-property plots


These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 (5'-D(\*CP\*AP\*AP\*AP\*GP\*GP\*CP\*GP\*GP\*TP\*AP\*AP\*TP\*AP\*CP\*GP\*GP\*TP\*TP\*AP\*TP\*CP\*CP\*AP\*CP\*AP\*GP\*AP\*AP\*TP\*CP\*AP\*GP\*G)-3')

Chain E: 



- Molecule 1: DNA (5'-D(\*CP\*AP\*AP\*AP\*GP\*GP\*CP\*GP\*GP\*TP\*AP\*AP\*TP\*AP\*CP\*GP\*GP\*TP\*TP\*AP\*TP\*CP\*CP\*AP\*CP\*AP\*GP\*AP\*AP\*TP\*CP\*AP\*GP\*G)-3')

Chain G: 




- Molecule 2: DNA (5'-D(\*CP\*CP\*TP\*GP\*AP\*TP\*TP\*CP\*TP\*GP\*TP\*GP\*GP\*AP\*TP\*AP\*AP\*CP\*CP\*GP\*TP\*AP\*TP\*TP\*AP\*CP\*CP\*GP\*CP\*CP\*TP\*TP\*TP\*G)-3')

Chain F: 



- Molecule 2: DNA (5'-D(\*CP\*CP\*TP\*GP\*AP\*TP\*TP\*CP\*TP\*GP\*TP\*GP\*GP\*AP\*TP\*AP\*AP\*CP\*CP\*GP\*TP\*AP\*TP\*TP\*AP\*CP\*CP\*GP\*CP\*CP\*TP\*TP\*TP\*G)-3')

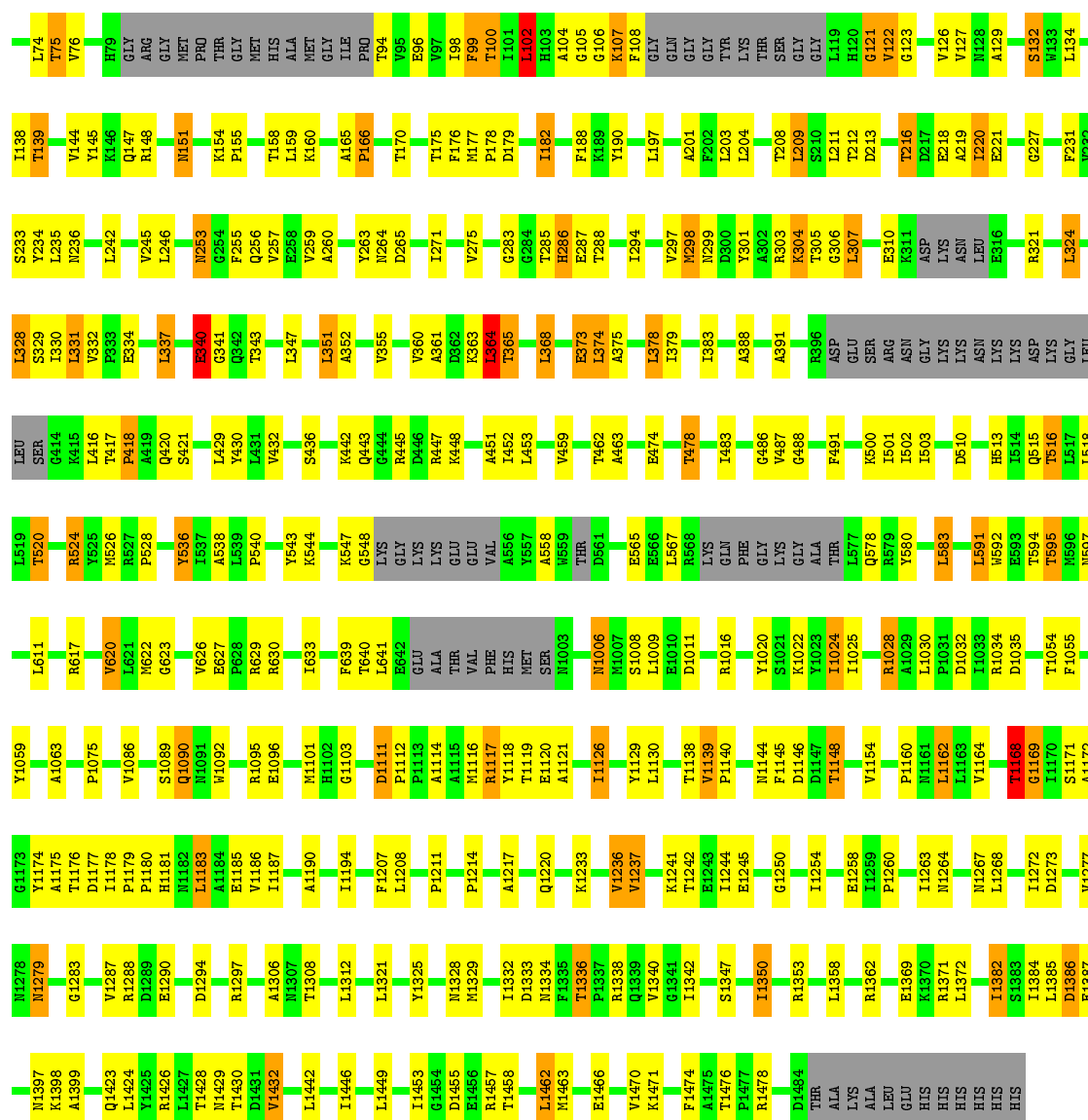
Chain H: 



- Molecule 3: Topoisomerase IV subunit B, DNA topoisomerase 4 subunit A chimera

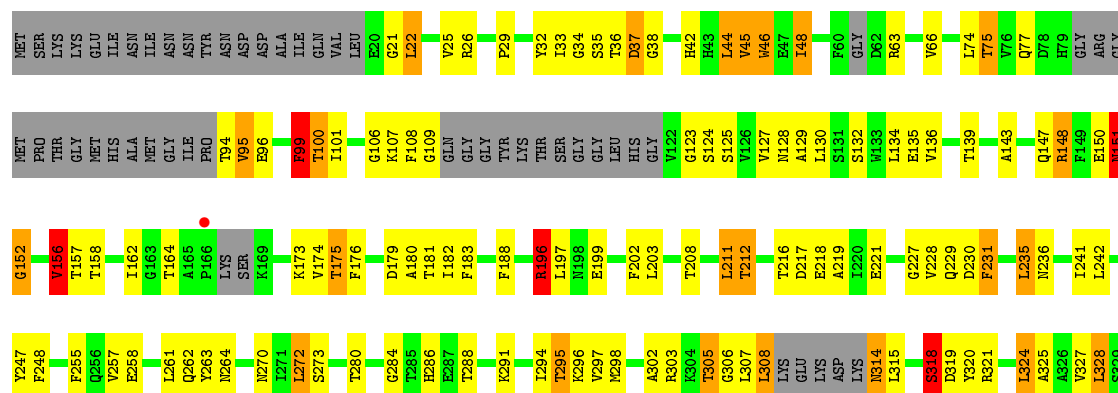
Chain A: 





- Molecule 3: Topoisomerase IV subunit B, DNA topoisomerase 4 subunit A chimera

Chain B: 62% 24% 10%



L1405	L1406	S1407	S1408	T1412	E1413	E1414	Q1415	I1419	L1422	L1442	M1448	L1449	A1450	A1451	I1452	I1453	L1462	M1463	K1464	R1468	R1478	L1482	E1483	D1484	THR	ALA	LYS	ALA	LEU	GLU	HIS	HIS	HIS	HIS	HIS									
V1287	E1290	S1291	D1292	R1293	D1294	G1295	L1296	R1297	I1300	E1301	L1302	T1308	V1311	L1312	L1321	M1329	R1338	Q1339	L1346	Y1349	I1350	A1351	H1352	R1353	R1354	E1355	R1362	A1368	R1371	L1372	H1373	I1374	G1377	L1378	I1379	R1380	V1381	I1382	I1383	I1384	I1389	R1393	D1400	
N1161	V1164	S1167	T1168	R1169	G1170	T1171	A1172	G1173	T1176	D1177	I1178	H1181	N1182	L1183	A1184	E1185	A1190	V1191	I1194	F1207	L1208	P1211	A1217	I1218	I1219	Q1220	G1221	K1233	V1236	K1248	Q1253	I1254	I1256	K1265	L1268	I1272	R1276	V1277	I1284					
L1030	D1035	Q1041	R1042	R1043	S1052	N1053	S1058	Y1059	R1060	K1061	S1062	M1070	F1073	H1074	S1079	S1080	I1081	V1086	E1096	M1101	H1102	G1103	N1104	D1111	A1114	Y1118	T1119	E1120	I1126	L1130	E1135	K1136	K1137	T1138	V1139	N1144	F1145	D1146	D1147	T1149	P1160			
L452	L453	V459	T462	T478	G488	F491	E494	N497	M504	T505	D506	T509	T516	L517	L518	L519	T520	F521	R524	Y525	M526	R527	P528	G533	Y536	I537	A538	M545	SER	LYS	GLY	GLY	LYS	GLY	GLY	GLU	VAL	ALA	TYR	A558	L567	ARG	LYS	
I330	P333	H336	L337	T343	L351	A352	V355	V356	I359	V360	K363	L364	T365	F366	L368	K369	E370	S399	ARG	ASN	GLY	LYS	LYS	ASN	LYS	LYS	ASP	LYS	GLY	LEU	LEU	SER	G414	T417	P424	Y430	A437	A441	K442	Q443	G444	R445	D446	R447



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	160.60 Å   160.60 Å   280.56 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	71.82 – 3.70 71.82 – 3.70	Depositor EDS
% Data completeness (in resolution range)	99.7 (71.82-3.70) 99.7 (71.82-3.70)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.10 (at 3.67 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.2_869)	Depositor
R, $R_{free}$	0.185   ,   0.248 0.186   ,   0.247	Depositor DCC
$R_{free}$ test set	2006 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	95.5	Xtriage
Anisotropy	0.205	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 82.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 39830 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	16228	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.52% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: 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	E	0.83	1/316 (0.3%)	1.85	12/483 (2.5%)
1	G	0.96	0/457	1.94	12/701 (1.7%)
2	F	0.80	0/317	1.91	10/485 (2.1%)
2	H	0.86	0/450	1.83	13/689 (1.9%)
3	A	0.45	0/7600	0.67	2/10359 (0.0%)
3	B	0.43	0/7491	0.65	2/10214 (0.0%)
All	All	0.50	1/16631 (0.0%)	0.87	51/22931 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
3	A	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	4	DT	C1'-N1	5.18	1.55	1.49

The worst 5 of 51 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	20	DA	O4'-C1'-N9	15.76	119.03	108.00
2	H	16	DA	O4'-C4'-C3'	-10.97	99.42	106.00
1	G	24	DA	O4'-C1'-N9	-8.97	101.72	108.00
2	H	21	DT	O4'-C4'-C3'	-8.15	101.11	106.00
2	H	24	DT	O4'-C1'-N1	-8.15	102.30	108.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	102	LEU	Peptide

## 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	E	282	0	147	8	0
1	G	407	0	211	10	0
2	F	283	0	145	11	0
2	H	402	0	209	8	0
3	A	7477	0	6788	217	0
3	B	7372	0	6577	186	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	H	1	0	0	0	0
5	E	2	0	0	0	0
All	All	16228	0	14077	414	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 14.

The worst 5 of 414 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:E:3:DA:N6	2:F:13:DA:N6	2.20	0.89
1:E:3:DA:N1	2:F:13:DA:N1	2.21	0.89
3:A:520:THR:HG21	3:A:622:MET:HG3	1.56	0.88
3:B:48:ILE:HG12	3:B:127:VAL:HG21	1.60	0.82
1:E:3:DA:N6	2:F:13:DA:C6	2.47	0.82

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	
3	A	1017/1144 (89%)	891 (88%)	101 (10%)	25 (2%)	7	49
3	B	1015/1144 (89%)	868 (86%)	115 (11%)	32 (3%)	5	44
All	All	2032/2288 (89%)	1759 (87%)	216 (11%)	57 (3%)	6	47

5 of 57 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	305	THR
3	A	373	GLU
3	A	1398	LYS
3	B	35	SER
3	B	99	PHE

### 5.3.2 Protein sidechains

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	
3	A	672/969 (69%)	565 (84%)	107 (16%)	3	22
3	B	640/969 (66%)	548 (86%)	92 (14%)	4	28
All	All	1312/1938 (68%)	1113 (85%)	199 (15%)	3	25

5 of 199 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	1279	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	B	75	THR
3	B	1296	LEU
3	A	1308	THR
3	A	1387	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	51	ASN
3	A	264	ASN
3	B	264	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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, 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	E	14/34 (41%)	-0.57	0	100	100	144, 201, 233, 259	14 (100%)
1	G	20/34 (58%)	-0.48	0	100	100	66, 82, 159, 192	0
2	F	14/34 (41%)	-0.60	0	100	100	138, 196, 225, 254	14 (100%)
2	H	20/34 (58%)	-0.54	0	100	100	60, 85, 168, 199	0
3	A	1037/1144 (90%)	-0.48	1 (0%)	95	94	35, 83, 129, 225	0
3	B	1035/1144 (90%)	-0.46	1 (0%)	95	94	44, 89, 136, 212	0
All	All	2140/2424 (88%)	-0.47	2 (0%)	95	94	35, 86, 142, 259	28 (1%)

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	B	166	PRO	3.4
3	A	37	ASP	3.0

### 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. 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
4	MG	A	1700	1/1	0.92	0.24	0.28	94,94,94,94	0
4	MG	B	1700	1/1	0.85	0.07	-1.50	109,109,109,109	0
4	MG	H	1901	1/1	0.91	0.14	-	72,72,72,72	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.