



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:44 AM GMT

PDB ID : 2DQB  
Title : Crystal structure of dNTP triphosphohydrolase from *Thermus thermophilus* HB8, which is homologous to dGTP triphosphohydrolase  
Authors : Kondo, N.; Nakagawa, N.; Ebihara, A.; Chen, L.; Liu, Z.-J.; Wang, B.-C.; Yokoyama, S.; Kuramitsu, S.; Masui, R.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2006-05-25  
Resolution : 2.20 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

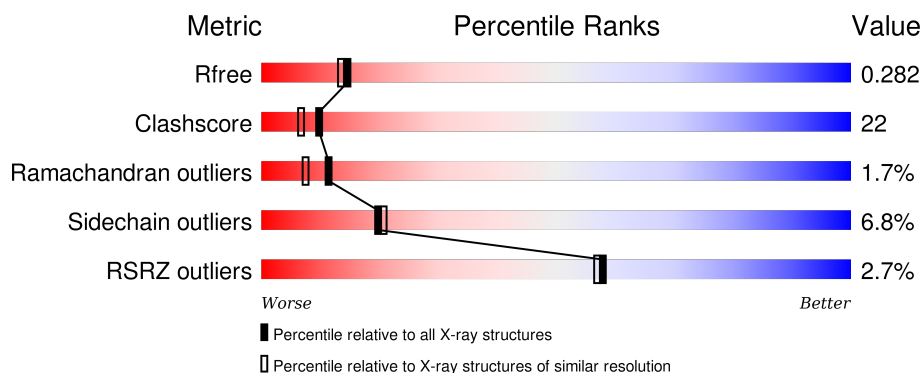
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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.



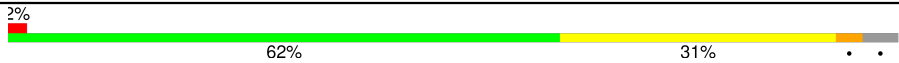
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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	376	
1	B	376	
1	C	376	
1	D	376	
1	E	376	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	376	 A horizontal bar chart showing the quality of chain F. The bar is divided into four segments: a small red segment at the beginning labeled '2%', a large green segment labeled '62%', a yellow segment labeled '31%', and a small grey segment at the end. Two small black dots are visible at the far right end of the bar.

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 17546 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 Deoxyguanosinetriphosphate triphosphohydrolase, putative.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	363	Total	C	N	O	S	Se	0	0	0
			2889	1820	536	528	1	4			
1	B	350	Total	C	N	O	S	Se	0	0	0
			2806	1763	521	517	1	4			
1	C	364	Total	C	N	O	S	Se	0	0	0
			2901	1826	536	534	1	4			
1	D	361	Total	C	N	O	S	Se	0	0	0
			2882	1816	525	536	1	4			
1	E	360	Total	C	N	O	S	Se	0	0	0
			2856	1796	522	533	1	4			
1	F	362	Total	C	N	O	S	Se	0	0	0
			2901	1827	534	535	1	4			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
A	127	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
A	209	PHE	LEU	ENGINEERED	UNP Q5SL81
A	304	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
A	362	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
B	127	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
B	209	PHE	LEU	ENGINEERED	UNP Q5SL81
B	304	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
B	362	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
C	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
C	127	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
C	209	PHE	LEU	ENGINEERED	UNP Q5SL81
C	304	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
C	362	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
D	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
D	127	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	209	PHE	LEU	ENGINEERED	UNP Q5SL81
D	304	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
D	362	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
E	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
E	127	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
E	209	PHE	LEU	ENGINEERED	UNP Q5SL81
E	304	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
E	362	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
F	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
F	127	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
F	209	PHE	LEU	ENGINEERED	UNP Q5SL81
F	304	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81
F	362	MSE	MET	MODIFIED RESIDUE	UNP Q5SL81

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	1	Total Mg 1 1	0	0
2	E	1	Total Mg 1 1	0	0
2	B	1	Total Mg 1 1	0	0
2	C	1	Total Mg 1 1	0	0
2	A	1	Total Mg 1 1	0	0
2	F	1	Total Mg 1 1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	63	Total O 63 63	0	0
3	B	41	Total O 41 41	0	0
3	C	60	Total O 60 60	0	0
3	D	47	Total O 47 47	0	0
3	E	46	Total O 46 46	0	0

*Continued on next page...*

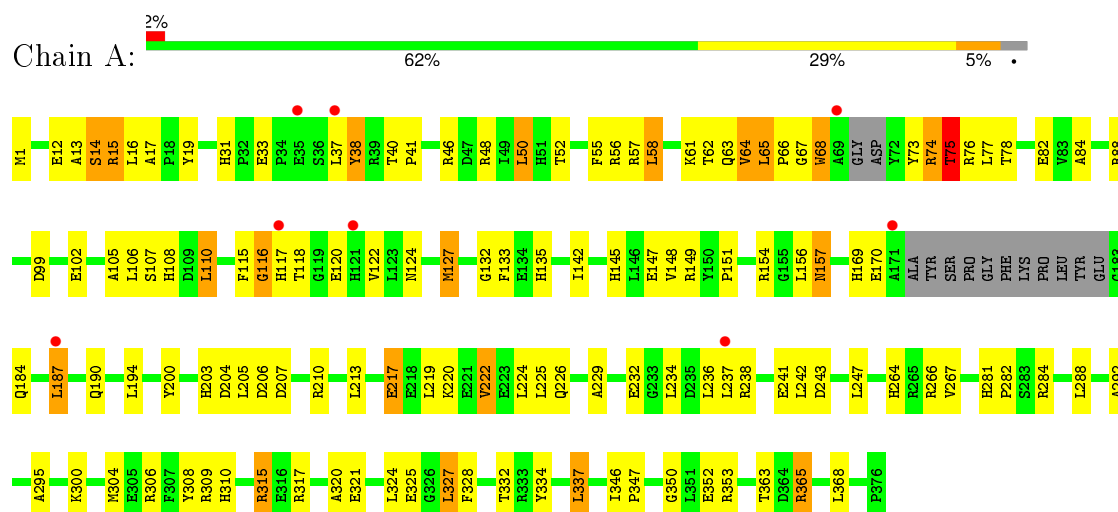
*Continued from previous page...*

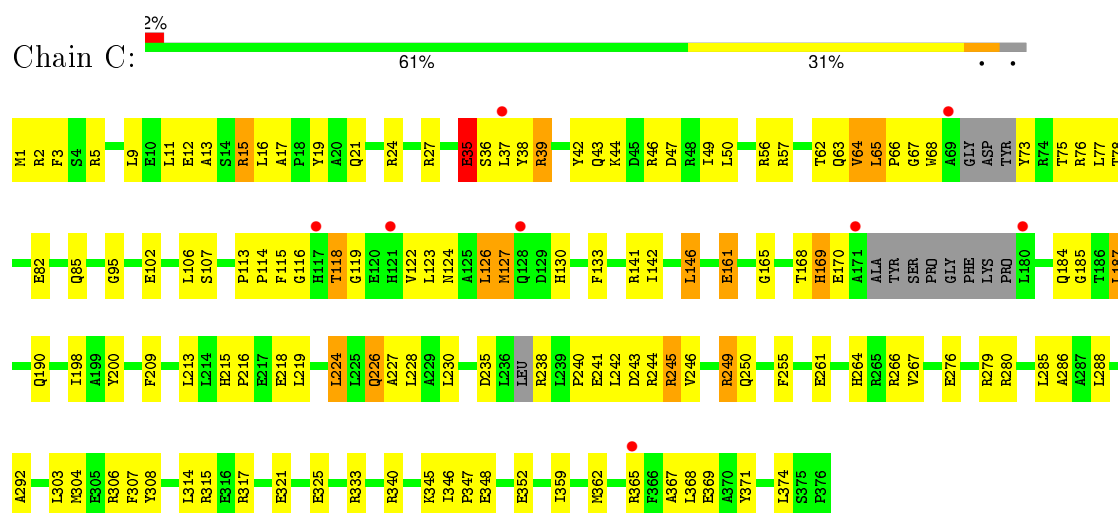
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	F	48	Total	O	0	0
			48	48		

### 3 Residue-property plots

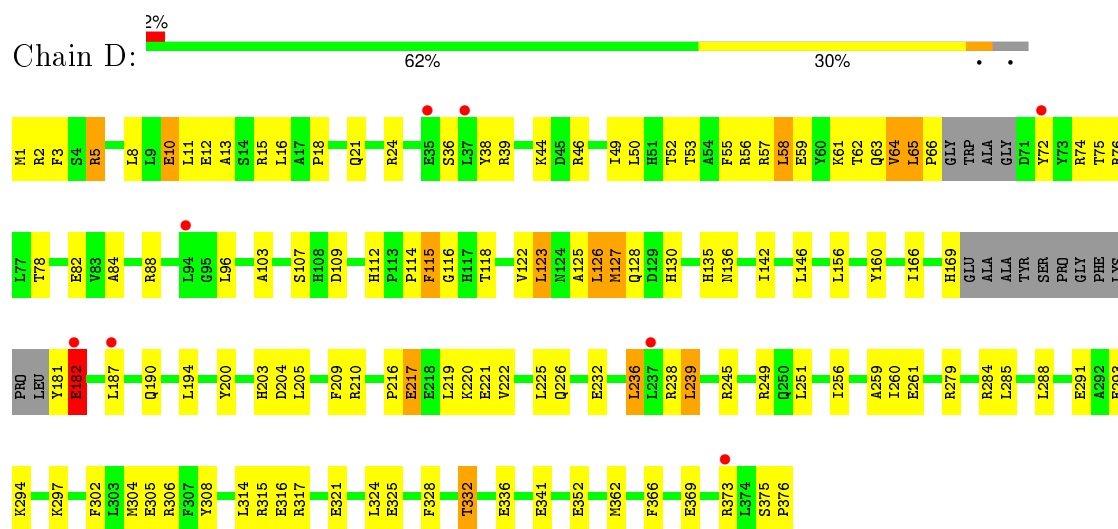
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Deoxyguanosinetriphosphate triphosphohydrolase, putative

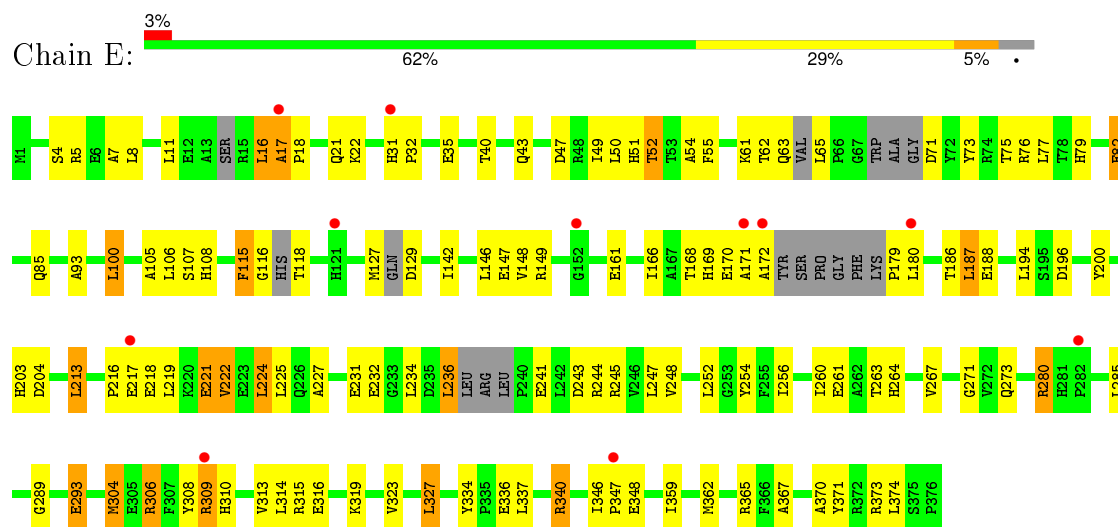




- Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase, putative

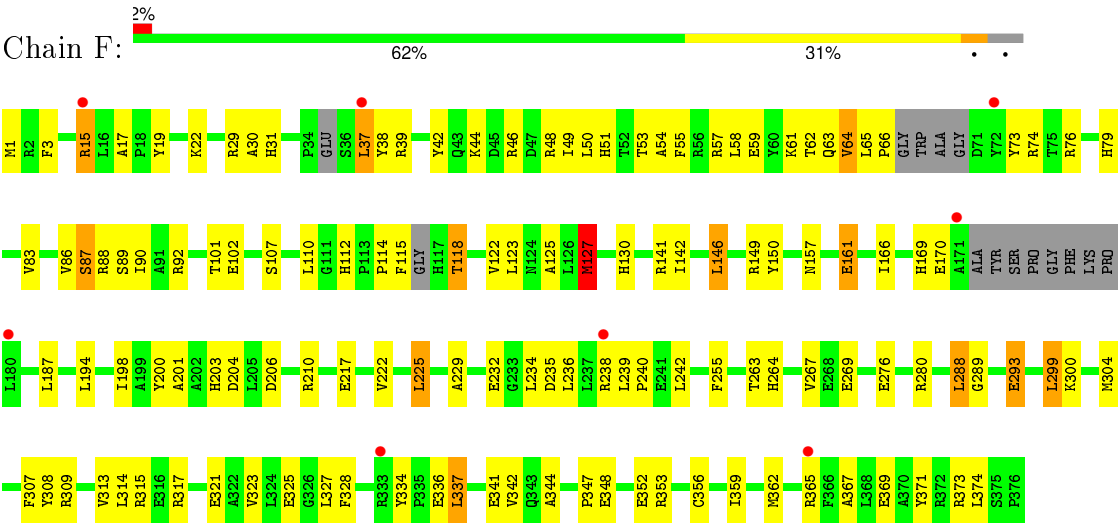


- Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase, putative



- Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase, putative





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.10Å 84.20Å 108.00Å 110.20° 109.70° 94.40°	Depositor
Resolution (Å)	20.00 – 2.20 19.98 – 2.20	Depositor EDS
% Data completeness (in resolution range)	90.1 (20.00-2.20) 79.4 (19.98-2.20)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.79 (at 2.19Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.223 , 0.286 0.221 , 0.282	Depositor DCC
$R_{free}$ test set	10257 reflections (9.98%)	DCC
Wilson B-factor (Å <sup>2</sup> )	30.4	Xtriage
Anisotropy	0.472	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 55.8	EDS
Estimated twinning fraction	0.010 for -h,-k,h+k+l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 102946 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	17546	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.57% 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	A	0.44	0/2942	0.66	0/3973
1	B	0.47	0/2854	0.69	1/3847 (0.0%)
1	C	0.46	0/2954	0.66	0/3988
1	D	0.48	0/2933	0.67	1/3961 (0.0%)
1	E	0.43	0/2903	0.67	1/3911 (0.0%)
1	F	0.46	1/2951 (0.0%)	0.67	1/3981 (0.0%)
All	All	0.46	1/17537 (0.0%)	0.67	4/23661 (0.0%)

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
1	F	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	127	MSE	C-N	7.43	1.51	1.34

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	129	ASP	N-CA-C	-6.53	93.37	111.00
1	D	126	LEU	N-CA-C	-5.99	94.84	111.00
1	E	213	LEU	CA-CB-CG	-5.86	101.81	115.30
1	F	125	ALA	O-C-N	-5.49	113.92	122.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	F	127	MSE	Mainchain

## 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	2889	0	2875	130	0
1	B	2806	0	2790	165	0
1	C	2901	0	2870	153	0
1	D	2882	0	2860	114	0
1	E	2856	0	2810	137	0
1	F	2901	0	2884	142	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	A	63	0	0	5	0
3	B	41	0	0	3	0
3	C	60	0	0	3	0
3	D	47	0	0	4	0
3	E	46	0	0	5	0
3	F	48	0	0	0	0
All	All	17546	0	17089	763	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 22.

The worst 5 of 763 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:F:127:MSE:HE1	1:F:356:CYS:SG	1.54	1.46
1:B:359:ILE:HA	1:B:362:MSE:HE2	1.30	1.14
1:A:61:LYS:HB3	1:A:76:ARG:HH21	1.19	1.07

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:17:ALA:HA	1:D:1:MSE:HE2	1.39	1.04
1:F:127:MSE:CE	1:F:356:CYS:SG	2.46	1.02

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	357/376 (95%)	339 (95%)	11 (3%)	7 (2%)	9	5
1	B	340/376 (90%)	318 (94%)	15 (4%)	7 (2%)	9	5
1	C	356/376 (95%)	336 (94%)	14 (4%)	6 (2%)	11	7
1	D	355/376 (94%)	336 (95%)	12 (3%)	7 (2%)	9	5
1	E	344/376 (92%)	327 (95%)	11 (3%)	6 (2%)	11	7
1	F	352/376 (94%)	337 (96%)	12 (3%)	3 (1%)	21	19
All	All	2104/2256 (93%)	1993 (95%)	75 (4%)	36 (2%)	11	7

5 of 36 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	38	TYR
1	A	68	TRP
1	A	75	THR
1	B	217	GLU
1	D	182	GLU

### 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	289/300 (96%)	264 (91%)	25 (9%)	13	12
1	B	284/300 (95%)	267 (94%)	17 (6%)	24	26
1	C	290/300 (97%)	270 (93%)	20 (7%)	19	20
1	D	291/300 (97%)	276 (95%)	15 (5%)	29	33
1	E	285/300 (95%)	264 (93%)	21 (7%)	17	17
1	F	293/300 (98%)	273 (93%)	20 (7%)	20	21
All	All	1732/1800 (96%)	1614 (93%)	118 (7%)	20	21

5 of 118 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	187	LEU
1	D	72	TYR
1	F	200	TYR
1	C	200	TYR
1	C	245	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 60 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	63	GLN
1	C	215	HIS
1	F	108	HIS
1	C	136	ASN
1	D	43	GLN

### 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 [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 6 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	359/376 (95%)	0.06	8 (2%) 65 64	20, 35, 54, 65	0
1	B	346/376 (92%)	0.05	15 (4%) 39 38	18, 36, 54, 71	0
1	C	360/376 (95%)	0.01	8 (2%) 65 64	17, 33, 57, 80	0
1	D	357/376 (94%)	-0.04	8 (2%) 65 64	17, 34, 54, 71	0
1	E	356/376 (94%)	0.10	11 (3%) 52 51	21, 37, 55, 74	0
1	F	358/376 (95%)	0.04	8 (2%) 65 64	21, 36, 57, 70	0
All	All	2136/2256 (94%)	0.04	58 (2%) 58 57	17, 35, 55, 80	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	72	TYR	6.7
1	A	69	ALA	6.0
1	C	171	ALA	5.9
1	A	237	LEU	5.2
1	C	37	LEU	5.1

### 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(Å <sup>2</sup> )	Q<0.9
2	MG	D	504	1/1	0.92	0.18	1.52	39,39,39,39	0
2	MG	E	501	1/1	0.90	0.16	1.37	34,34,34,34	0
2	MG	F	505	1/1	0.95	0.15	1.33	42,42,42,42	0
2	MG	A	506	1/1	0.77	0.12	-0.32	40,40,40,40	0
2	MG	B	502	1/1	0.96	0.08	-1.20	27,27,27,27	0
2	MG	C	503	1/1	0.97	0.05	-1.81	36,36,36,36	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.