



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 07:21 PM GMT

PDB ID : 1F4H  
Title : E. COLI (LACZ) BETA-GALACTOSIDASE (ORTHORHOMBIC)  
Authors : Juers, D.H.; Jacobson, R.H.; Wigley, D.; Zhang, X.J.; Huber, R.E.; Tronrud, D.E.; Matthews, B.W.  
Deposited on : 2000-06-07  
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](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

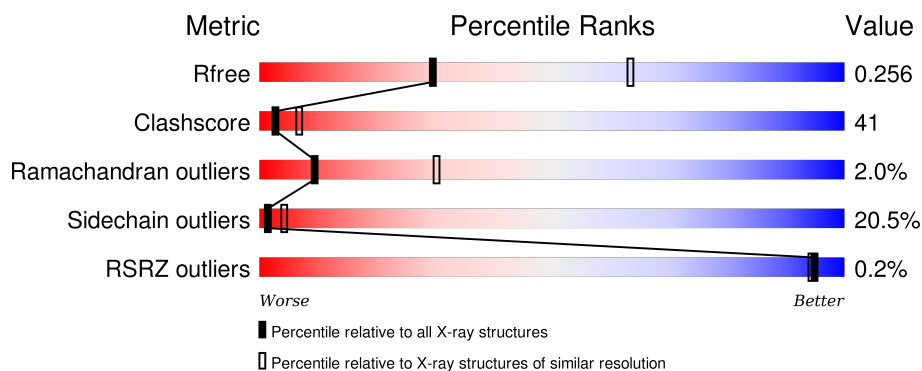
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (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 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	1021	<div> <div>34%</div> <div>45%</div> <div>18%</div> <div>.</div> </div>
1	B	1021	<div> <div>30%</div> <div>48%</div> <div>19%</div> <div>.</div> </div>
1	C	1021	<div> <div>32%</div> <div>47%</div> <div>18%</div> <div>.</div> </div>
1	D	1021	<div> <div>35%</div> <div>43%</div> <div>18%</div> <div>.</div> </div>

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	MG	A	3002	-	-	-	X
2	MG	D	3002	-	-	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 33805 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 BETA-GALACTOSIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1021	Total	C	N	O	S	84	7	0
			8238	5209	1466	1525	38			
1	B	1021	Total	C	N	O	S	84	7	0
			8238	5209	1466	1525	38			
1	C	1021	Total	C	N	O	S	84	7	0
			8238	5209	1466	1525	38			
1	D	1021	Total	C	N	O	S	84	7	0
			8238	5209	1466	1525	38			

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

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

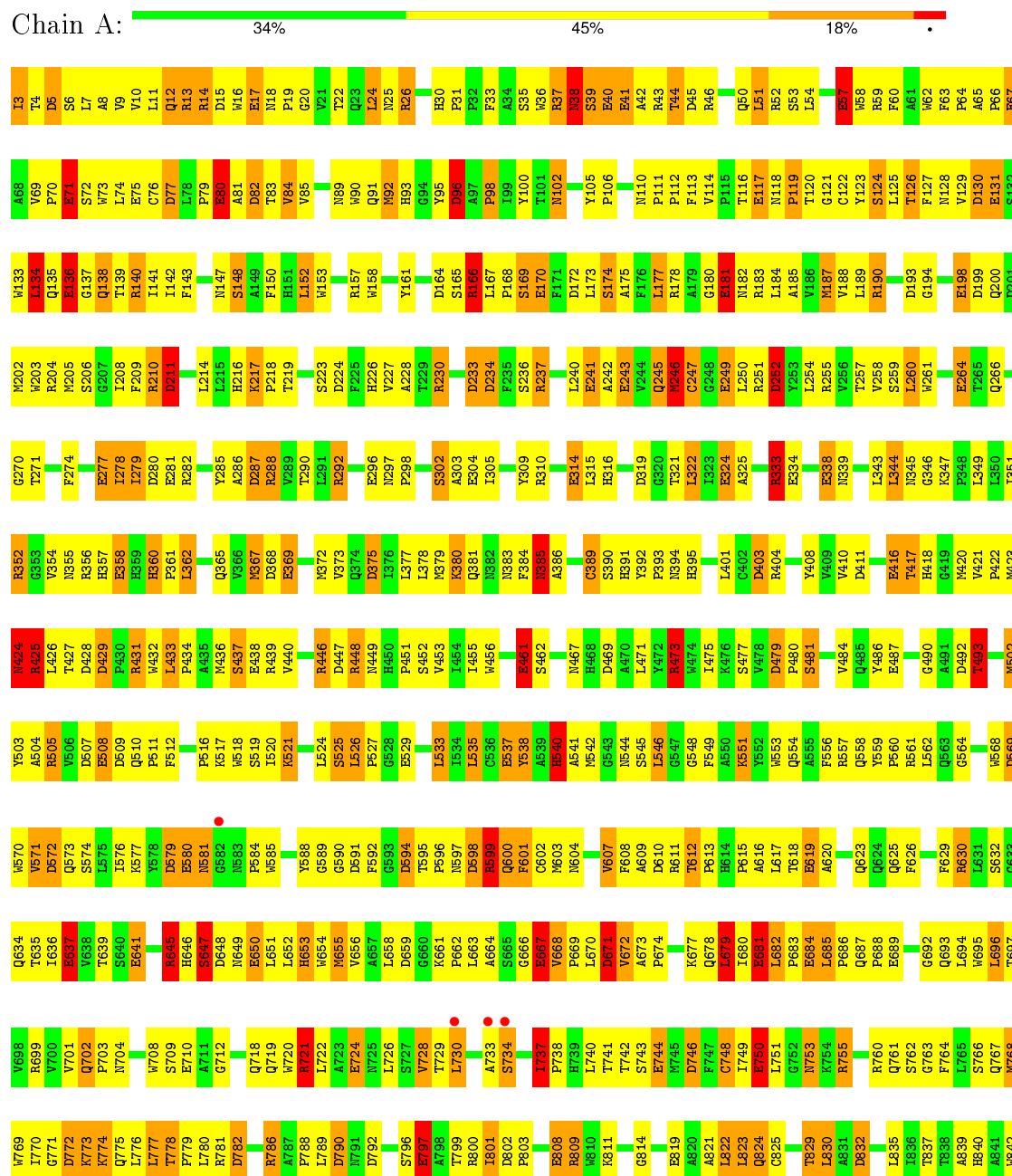
- Molecule 3 is water.

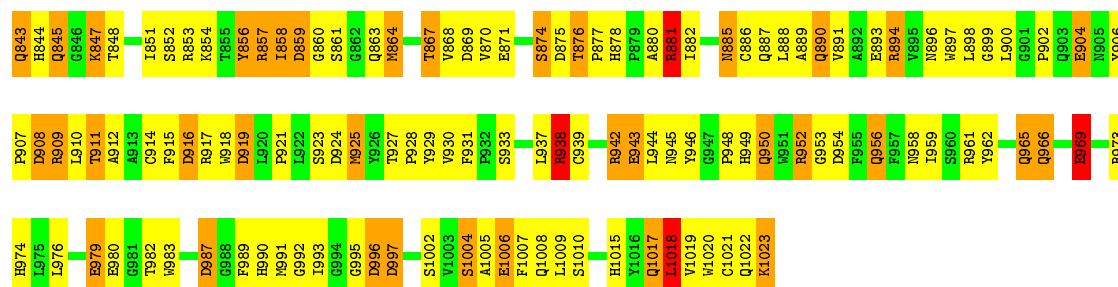
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	211	Total	O	0	0
			211	211		
3	B	202	Total	O	0	0
			202	202		
3	C	220	Total	O	0	0
			220	220		
3	D	212	Total	O	0	0
			212	212		

### 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.

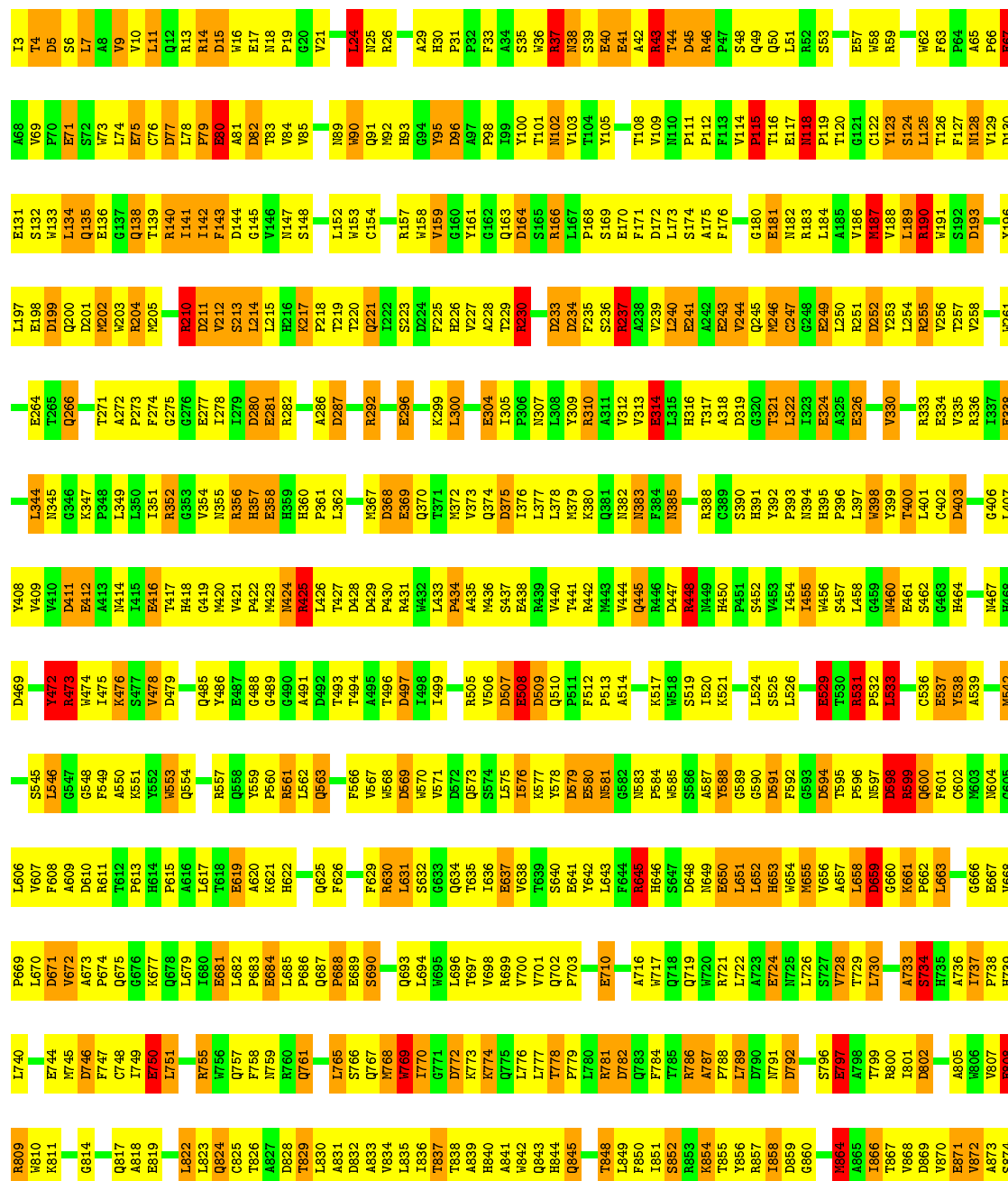
#### • Molecule 1: BETA-GALACTOSIDASE





# Molecule 1: BETA-GALACTOSIDASE

Chain B:  30%  48%  19%  5%



Chain C:  32% 47% 18% 3%







S971	H972	R973	H974	E979	E980	G981	T982	W983	L984	W985	I986	D987	G988	F989	H990	M991	D997	S998	W999	S1002	Y1003	S1004	A1005	E1006	L1009	S1010	A1011	G1012	R1013	Y1014	Q1017	L1018	Y1019	W1020	C1021	Q1022	K1023																						
Y906	P907	D908	R909	L910	T911	A912	A913	R917	W918	D919	L920	P921	L922	S923	D924	M925	Y926	T927	P928	Y929	W930	F931	P932	S933	E934	G936	L937	R938	R942	E943	L944	N945	Y946	G947	P948	H949	W950	N951	R952	G953	D954	F955	Q956	F957	N958	I959	S960	R961	Q964	Q965	Q966	L967	M968	E969	T970				
I836	T837	D838		H844	Q845			T848	L849	F850	I851	S852	R853	K854	T855	R856	I857	I858	D859	G860	S861	G862	Q863	H864		T867	W868	D869	W870	E871	W872	A873	S874	D875	T876	P877	H878	P879	A880	I882	G883	L884	I885	C886	Q887	L888	A889	Q890		E893	R894	W895	W896	W897		P902	Q903	E904	W905

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	153.40 Å   173.40 Å   204.40 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	25.00 – 2.80 24.97 – 2.82	Depositor EDS
% Data completeness (in resolution range)	88.0 (25.00-2.80) 80.5 (24.97-2.82)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.68 (at 2.80 Å)	Xtriage
Refinement program	TNT V. 5-E	Depositor
R, $R_{free}$	0.137   ,   0.279 0.125   ,   0.256	Depositor DCC
$R_{free}$ test set	1590 reflections (1.53%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.7	Xtriage
Anisotropy	0.148	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 133.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 105768 reflections	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	33805	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 41.86 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 2.2179e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 ⓘ

### 5.1 Standard geometry ⓘ

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	1.11	56/8515 (0.7%)	1.72	186/11615 (1.6%)
1	B	1.09	53/8515 (0.6%)	1.69	174/11615 (1.5%)
1	C	1.08	49/8515 (0.6%)	1.69	187/11615 (1.6%)
1	D	1.10	54/8515 (0.6%)	1.70	183/11615 (1.6%)
All	All	1.09	212/34060 (0.6%)	1.70	730/46460 (1.6%)

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	A	2	0
1	B	1	0
1	C	4	1
1	D	1	0
All	All	8	1

The worst 5 of 212 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	249	GLU	CD-OE2	8.77	1.35	1.25
1	B	241	GLU	CD-OE2	8.07	1.34	1.25
1	C	508	GLU	CD-OE2	8.02	1.34	1.25
1	D	181	GLU	CD-OE2	7.94	1.34	1.25
1	A	198	GLU	CD-OE2	7.86	1.34	1.25

The worst 5 of 730 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	251	ARG	NE-CZ-NH2	-13.90	113.35	120.30
1	D	210	ARG	NE-CZ-NH1	13.54	127.07	120.30
1	D	425	ARG	NE-CZ-NH1	13.14	126.87	120.30
1	D	668	VAL	C-N-CD	-13.14	91.69	120.60
1	A	356	ARG	NE-CZ-NH1	13.13	126.87	120.30

5 of 8 chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	166	ARG	CA
1	A	187	MET	CA
1	B	118	ASN	CA
1	C	291	LEU	CA
1	C	503	TYR	CA

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	352	ARG	Mainchain

## 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	8238	0	7824	643	0
1	B	8238	0	7824	695	0
1	C	8238	0	7824	667	0
1	D	8238	0	7823	632	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
3	A	211	0	0	13	0
3	B	202	0	0	12	0
3	C	220	0	0	24	0
3	D	212	0	0	13	0
All	All	33805	0	31295	2582	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 41.

The worst 5 of 2582 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:A:173:LEU:HB3	1:A:177:LEU:HD21	1.21	1.17
1:B:427:THR:HA	1:B:436:MET:HE1	1.26	1.14
1:B:734:SER:HB3	1:B:860:GLY:HA3	1.24	1.12
1:D:427:THR:HA	1:D:436:MET:HE1	1.32	1.11
1:A:777:LEU:HD21	1:A:889:ALA:HB2	1.28	1.10

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	1026/1021 (100%)	923 (90%)	86 (8%)	17 (2%)	11	36
1	B	1026/1021 (100%)	917 (89%)	89 (9%)	20 (2%)	10	32
1	C	1026/1021 (100%)	918 (90%)	82 (8%)	26 (2%)	7	24
1	D	1026/1021 (100%)	921 (90%)	85 (8%)	20 (2%)	10	32
All	All	4104/4084 (100%)	3679 (90%)	342 (8%)	83 (2%)	9	30

5 of 83 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	647	SER
1	A	733	ALA
1	A	734	SER
1	A	1005	ALA
1	B	79	PRO

### 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	
1	A	880/873 (101%)	700 (80%)	180 (20%)	1	4
1	B	880/873 (101%)	690 (78%)	190 (22%)	1	3
1	C	880/873 (101%)	703 (80%)	177 (20%)	1	4
1	D	880/873 (101%)	709 (81%)	171 (19%)	2	5
All	All	3520/3492 (101%)	2802 (80%)	718 (20%)	1	4

5 of 718 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	789	LEU
1	C	165	SER
1	D	746	ASP
1	B	843	GLN
1	B	1004	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 102 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	597	ASN
1	C	221	GLN
1	D	693	GLN
1	B	622	HIS
1	B	965	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 8 ligands modelled in this entry, 8 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	1021/1021 (100%)	-1.15	4 (0%) 93 90	2, 26, 68, 100	19 (1%)
1	B	1021/1021 (100%)	-1.04	0 100 100	4, 31, 72, 100	19 (1%)
1	C	1021/1021 (100%)	-1.09	3 (0%) 94 92	6, 28, 65, 100	19 (1%)
1	D	1021/1021 (100%)	-1.13	1 (0%) 95 95	3, 26, 65, 100	19 (1%)
All	All	4084/4084 (100%)	-1.10	8 (0%) 95 94	2, 28, 68, 100	76 (1%)

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	733	ALA	2.7
1	A	734	SER	2.7
1	D	732	ALA	2.6
1	C	581	ASN	2.3
1	A	730	LEU	2.3

### 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
2	MG	A	3002	1/1	0.99	0.16	4.20	31,31,31,31	0
2	MG	D	3002	1/1	0.99	0.11	2.10	25,25,25,25	0
2	MG	B	3002	1/1	0.97	0.09	0.45	36,36,36,36	0
2	MG	A	3001	1/1	0.96	0.12	0.45	28,28,28,28	0
2	MG	C	3002	1/1	0.98	0.08	-0.74	31,31,31,31	0
2	MG	D	3001	1/1	0.98	0.06	-1.20	40,40,40,40	0
2	MG	B	3001	1/1	0.99	0.04	-1.58	20,20,20,20	0
2	MG	C	3001	1/1	0.99	0.03	-2.49	15,15,15,15	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.