



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 23, 2017 – 04:53 PM EST

PDB ID : 1F4A
Title : E. COLI (LACZ) BETA-GALACTOSIDASE (NCS CONSTRAINED MONOMER-ORTHORHOMBIC)
Authors : Juers, D.H.; Jacobson, R.H.; Wigley, D.; Zhang, X.J.; Huber, R.E.; Tronrud, D.E.; Matthews, B.W.
Deposited on : unknown
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

<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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030345

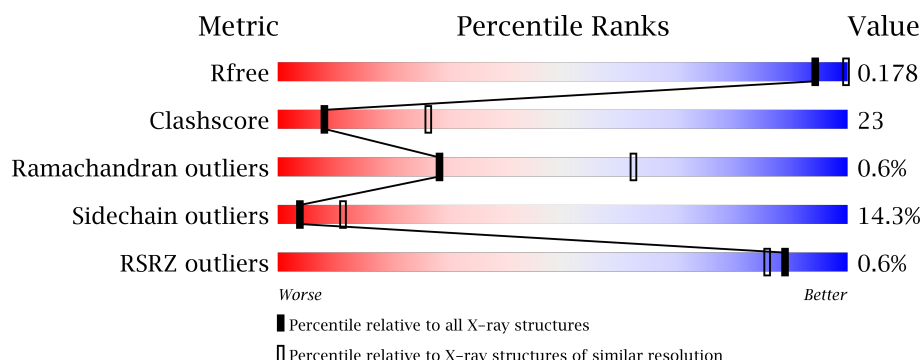
1 Overall quality at a glance

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}	100719	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (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 ≥ 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 style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 49% 38% 11% </div> </div>
1	B	1021	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 49% 38% 11% </div> </div>
1	C	1021	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 49% 38% 11% </div> </div>
1	D	1021	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 49% 38% 11% </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 34424 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		

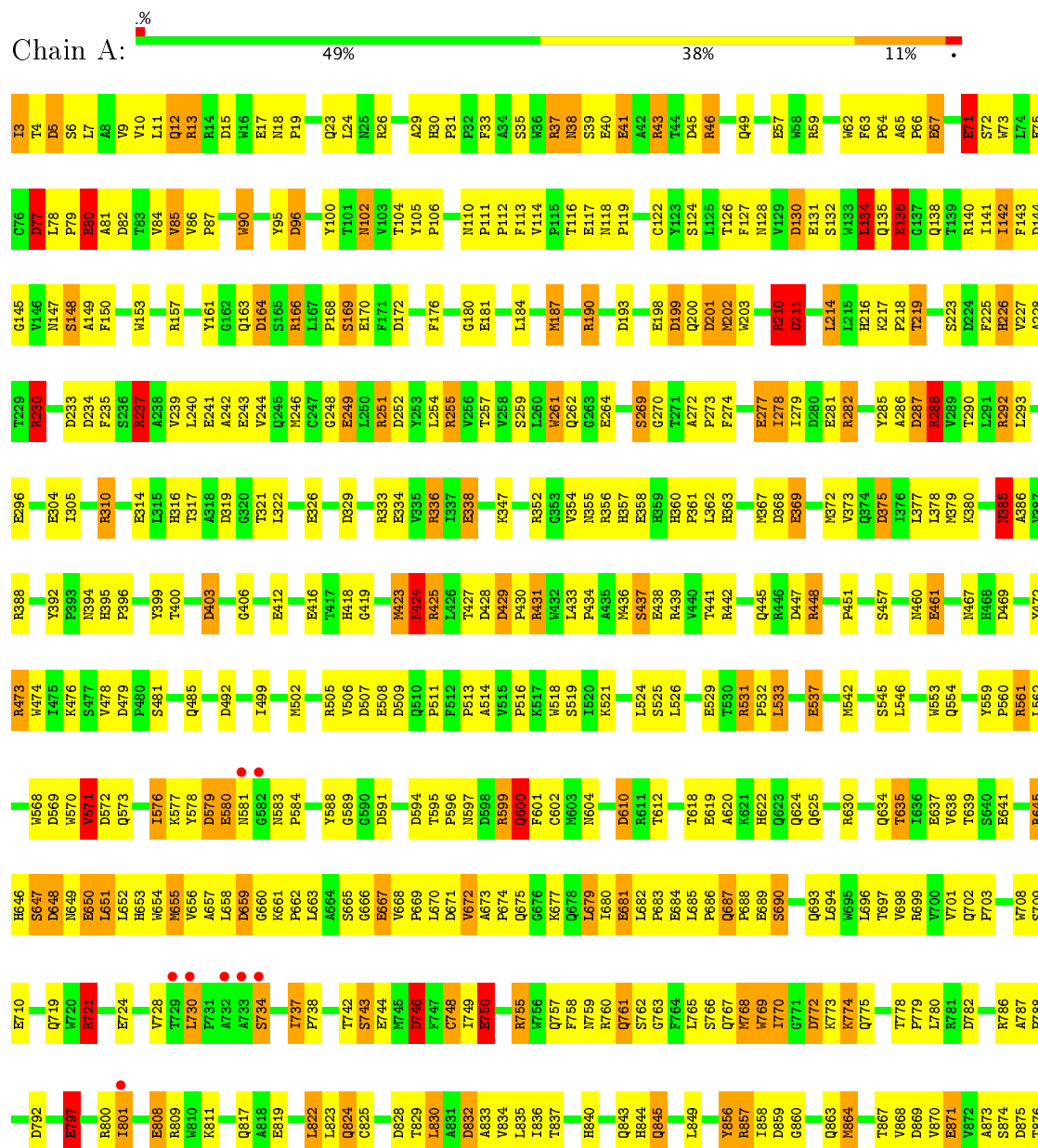
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	365	Total	O	0	0
			365	365		
3	B	366	Total	O	0	0
			366	366		
3	C	367	Total	O	0	0
			367	367		
3	D	366	Total	O	0	0
			366	366		

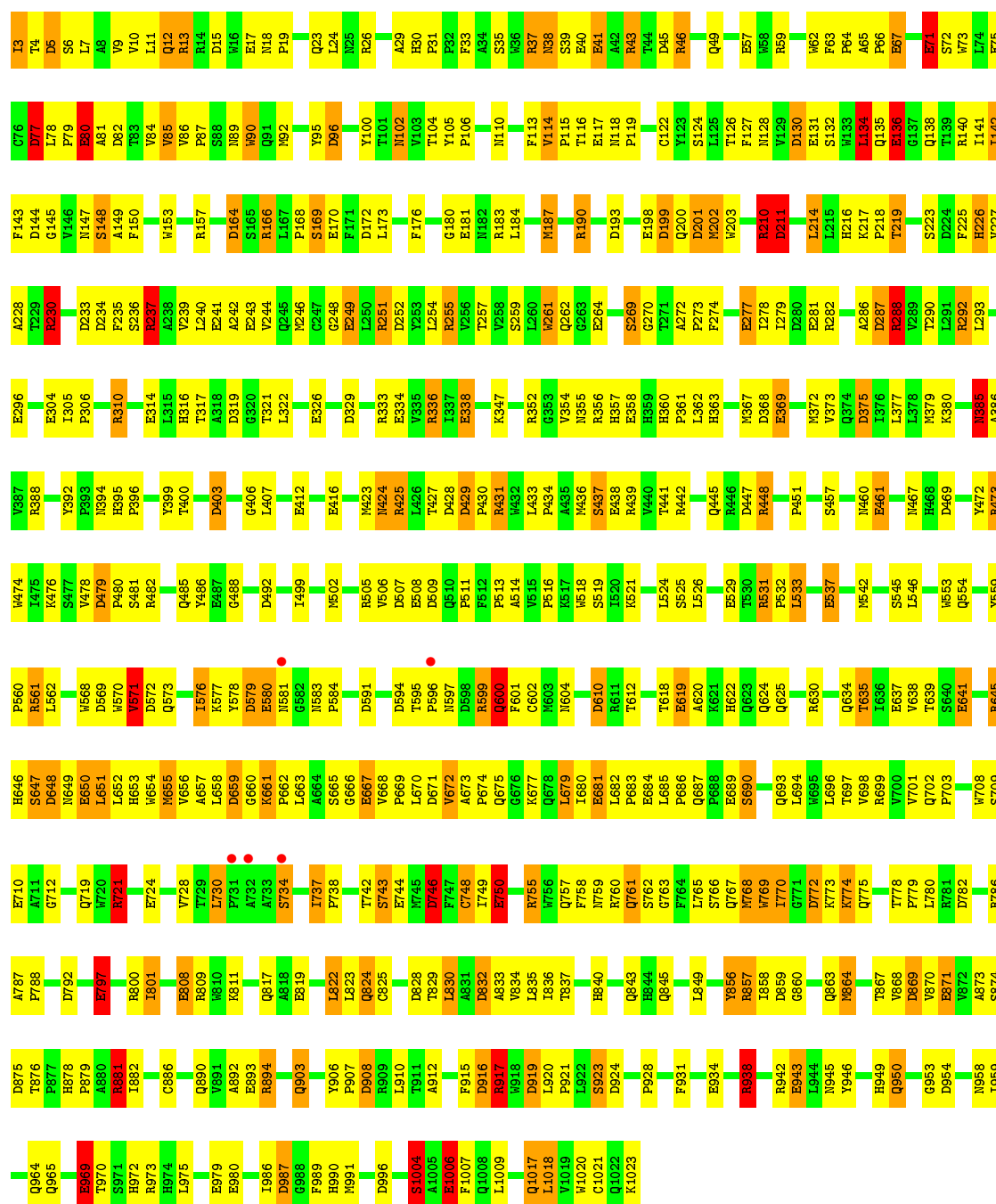
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: BETA-GALACTOSIDASE

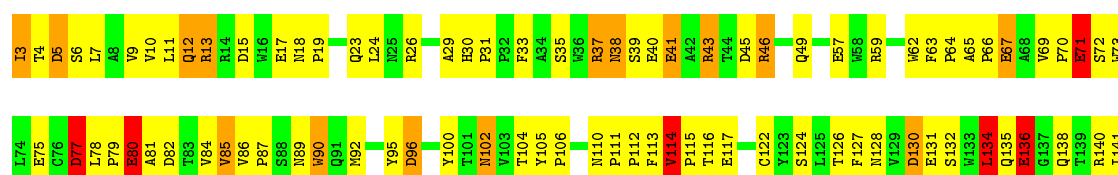


Chain C: 



• Molecule 1: BETA-GALACTOSIDASE

Chain D: 



E969	H878	E797	V720	L651	D569	R473	R388	E304	T229	I142
E970	P879	R800	R721	L652	H570	H474	Y392	L305	R230	F143
S971	A880	I801	E724	H653	Y571	I475	Y393	L308	D233	D144
H972	R881	E808	V728	H654	D572	K476	N394	R310	D234	G145
R973	L882	E809	V729	M655	Q573	S477	H395	E314	P235	N147
H974	C886	R809	L730	V656	Y576	D479	H396	L315	S236	S148
L975	Q890	H810	P731	A657	K577	P480	Y399	H316	R237	A149
E979	H891	K811	A732	L658	Y578	S481	T400	H317	A238	F150
E980	A892	Q817	A733	G660	D579	Q485	T403	A318	E241	W153
	E893	A818	S734	K661	E580	D492	D403	D319	A242	R157
	E894	E819	I737	L662	N581	D499	L407	G320	E243	
	Q903	L822	P738	A664	N583	Y499	E412	L322	M246	D164
	Y906	L823	T742	S665	P584	R505	E416	E326	C247	S165
	P907	Q824	S743	G666	D591	V506	T417	E269	Q248	R166
	D908	C825	E744	V668	D594	D607	H418	D250	E249	L167
	G909	D828	H745	P669	T595	E508	H419	R251	E260	P168
	L910	T829	D746	L670	P596	D609	G419	D252	R252	S169
	P911	L830	F747	V672	N597	Q510	M423	Y253	F171	E170
	A912	A831	C748	A673	D598	P611	R424	E254	D172	
		D832	I749	P674	R599	F512	R425	R355	L254	
	F915	A833	E750	Q675	Q600	P513	L436	R356	R255	F176
	D916	V834	R755	G676	F601	A514	T427	L337	V256	
	E918	L835	H756	Q677	C602	V515	D428	E338	T257	G180
	R918	I836	Q757	M603	P604	P516	M429	R347	V258	E181
	L919	T837	F758	N604	D610	R517	P430	R352	S259	L184
	P921	H840	N759	E681	R611	N518	H431	G353	W261	
	L922		R760	L882	D611	T620	M432	Q262	Q262	H187
	S923	Q843	Q761	P683	T612	R521	L433	G263	E264	
	D924	H844	S762	E884		L524	A435	N355		R190
		Q845	G763	L685	T618	S525	R357	R356	S269	D193
	P928	L849	F764	P686	E619	L526	S437	E358	G270	
			L765	Q687	A620		E438	R359	T271	E198
	F931		S766	P688	K621	E529	R439	H360	A272	D199
	E934	K854	Q767	E889	H622	T630	V440	L361	P273	Q200
		T855	N768	S890	Q623	N531	T441	L362	F274	D201
	Y856	R856	N769		Q624	P532	R442	H363		M202
	R857	R857	I770		Q625	L533			E277	W203
	I858	I858	Q693		L630		Q445	R367	L278	R210
	D859	Q860	L694			E537	R446	D368	I279	D211
	Q863		L695		Q634	N642	D447	E369	D280	
	M864		L696		T635		R448		E281	L214
			V700		I636	S545		K372	R282	L215
	T867		V701		V638	L546	P451	K373	A286	H216
	V868		Q702		T639		S457	Q374	D287	K217
	V870		P703		E641	N553	N460	D375	R288	P218
	E871		W708			Q554	E461	L377	V289	T219
	V872		S709		R645			L378	T290	
	A873		E710		H646	Y559	N467	K379	L291	S223
	S874		A711		S647	R561	H468	K380	R292	D224
	D875		G712		D648	L562	D469	L293	F225	
	T876		D792		N649		A386	E296	H226	
	P877				E650	N568	Y472		V227	A228

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	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
$\langle I/\sigma(I) \rangle$ ¹	1.68 (at 2.80 Å)	Xtriage
Refinement program	TNT 5E	Depositor
R, R_{free}	0.167 , 0.198 0.149 , 0.178	Depositor DCC
R_{free} test set	1590 reflections (1.53%)	DCC
Wilson B-factor (Å ²)	40.7	Xtriage
Anisotropy	0.148	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 100.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	34424	wwPDB-VP
Average B, all atoms (Å ²)	34.0	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.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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

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	1.08	52/8515 (0.6%)	1.61	173/11615 (1.5%)
1	B	1.08	52/8515 (0.6%)	1.61	175/11615 (1.5%)
1	C	1.08	52/8515 (0.6%)	1.61	174/11615 (1.5%)
1	D	1.08	52/8515 (0.6%)	1.61	176/11615 (1.5%)
All	All	1.08	208/34060 (0.6%)	1.61	698/46460 (1.5%)

The worst 5 of 208 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	75	GLU	CD-OE2	9.53	1.36	1.25
1	C	75	GLU	CD-OE2	9.53	1.36	1.25
1	D	75	GLU	CD-OE2	9.49	1.36	1.25
1	B	75	GLU	CD-OE2	9.46	1.36	1.25
1	D	710	GLU	CD-OE2	7.62	1.34	1.25

The worst 5 of 698 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	809[A]	ARG	NE-CZ-NH1	11.76	126.18	120.30
1	B	809[B]	ARG	NE-CZ-NH1	11.76	126.18	120.30
1	A	809[A]	ARG	NE-CZ-NH1	11.62	126.11	120.30
1	A	809[B]	ARG	NE-CZ-NH1	11.62	126.11	120.30
1	D	809[A]	ARG	NE-CZ-NH1	11.61	126.11	120.30

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	8238	0	7824	394	0
1	B	8238	0	7824	374	0
1	C	8238	0	7824	379	0
1	D	8238	0	7824	374	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	365	0	0	10	0
3	B	366	0	0	10	0
3	C	367	0	0	10	0
3	D	366	0	0	10	0
All	All	34424	0	31296	1496	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 23.

The worst 5 of 1496 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:427:THR:HA	1:B:436:MET:HE1	1.21	1.11
1:B:746:ASP:HA	1:B:760:ARG:HG3	1.34	1.10
1:A:746:ASP:HA	1:A:760:ARG:HG3	1.34	1.09
1:D:746:ASP:HA	1:D:760:ARG:HG3	1.34	1.09
1:C:427:THR:HA	1:C:436:MET:HE1	1.28	1.08

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	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	28	62
1	B	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	28	62
1	C	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	28	62
1	D	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	28	62
All	All	4104/4084 (100%)	3812 (93%)	268 (6%)	24 (1%)	28	62

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	647	SER
1	B	647	SER
1	C	647	SER
1	D	647	SER
1	A	77	ASP

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%)	755 (86%)	125 (14%)	4	12
1	B	880/873 (101%)	755 (86%)	125 (14%)	4	12
1	C	880/873 (101%)	755 (86%)	125 (14%)	4	12
1	D	880/873 (101%)	755 (86%)	125 (14%)	4	12
All	All	3520/3492 (101%)	3020 (86%)	500 (14%)	4	12

5 of 500 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	822	LEU
1	C	223	SER
1	D	755	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	857	ARG
1	C	37	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 94 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	634	GLN
1	C	216	HIS
1	D	622	HIS
1	B	817	GLN
1	B	1017	GLN

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 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 ⓘ

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, 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	1021/1021 (100%)	-0.81	8 (0%) 86 81	8, 29, 70, 100	19 (1%)
1	B	1021/1021 (100%)	-0.80	5 (0%) 90 88	8, 29, 70, 100	19 (1%)
1	C	1021/1021 (100%)	-0.84	5 (0%) 90 88	8, 29, 70, 100	19 (1%)
1	D	1021/1021 (100%)	-0.81	6 (0%) 89 86	8, 29, 70, 100	19 (1%)
All	All	4084/4084 (100%)	-0.81	24 (0%) 89 86	8, 29, 71, 100	76 (1%)

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	732	ALA	4.5
1	A	732	ALA	4.1
1	C	581	ASN	3.6
1	A	581	ASN	3.4
1	A	801	ILE	3.2

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 carbohydrates 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. 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, 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	LLDF	B-factors(\AA^2)	Q<0.9
2	MG	D	3002	1/1	0.96	0.12	1.95	31,31,31,31	0
2	MG	B	3002	1/1	0.94	0.11	1.60	31,31,31,31	0
2	MG	A	3002	1/1	0.98	0.11	0.98	31,31,31,31	0
2	MG	C	3002	1/1	0.96	0.12	0.97	31,31,31,31	0
2	MG	B	3001	1/1	0.99	0.06	-1.48	28,28,28,28	0
2	MG	C	3001	1/1	0.99	0.04	-2.05	28,28,28,28	0
2	MG	A	3001	1/1	0.96	0.08	-2.15	28,28,28,28	0
2	MG	D	3001	1/1	0.99	0.06	-2.52	28,28,28,28	0

6.5 Other polymers [i](#)

There are no such residues in this entry.