



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 31, 2020 – 10:12 AM BST

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 : 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/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)	:	1.13
EDS	:	2.13
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

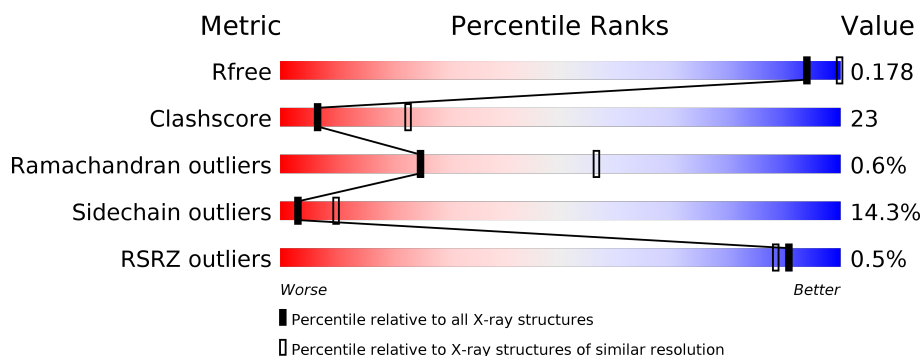
1 Overall quality at a glance ⓘ

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}	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)

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 $\leq 5\%$. 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></div> <div> <div>49%</div> <div>38%</div> <div>11%</div> </div> </div>
1	B	1021	<div> <div>49%</div> <div>38%</div> <div>11%</div> </div>
1	C	1021	<div> <div>49%</div> <div>38%</div> <div>11%</div> </div>
1	D	1021	<div> <div>49%</div> <div>38%</div> <div>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		

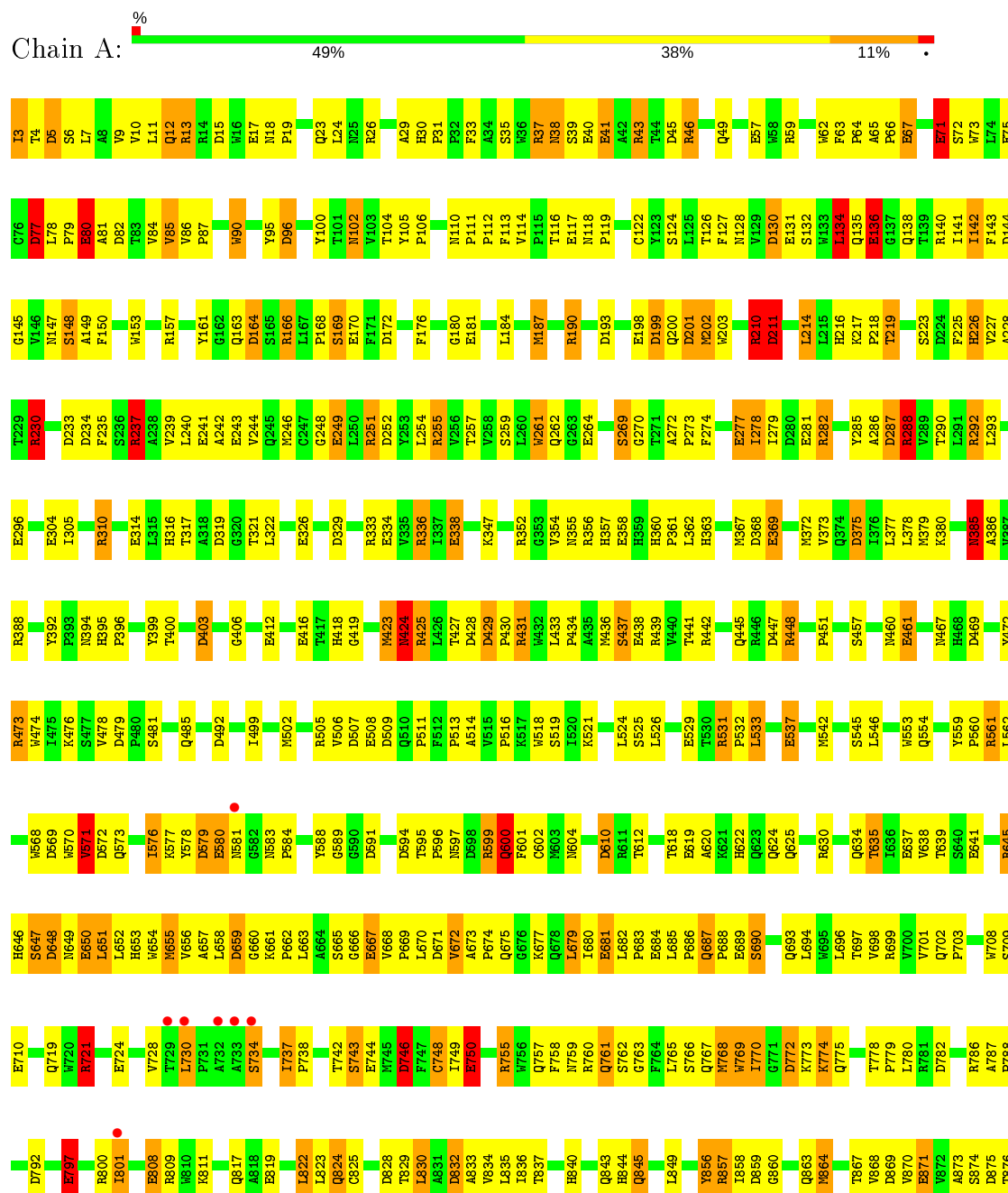
- Molecule 3 is water.

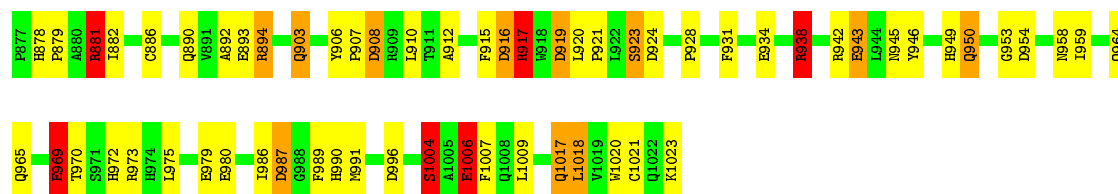
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	365	Total	O	0	0
			365	365		
3	B	366	Total	O	0	0
			366	366		
3	C	367	Total	O	0	0
			367	367		
3	D	366	Total	O	0	0
			366	366		

3 Residue-property plots

These plots are drawn for all protein, RNA, 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.

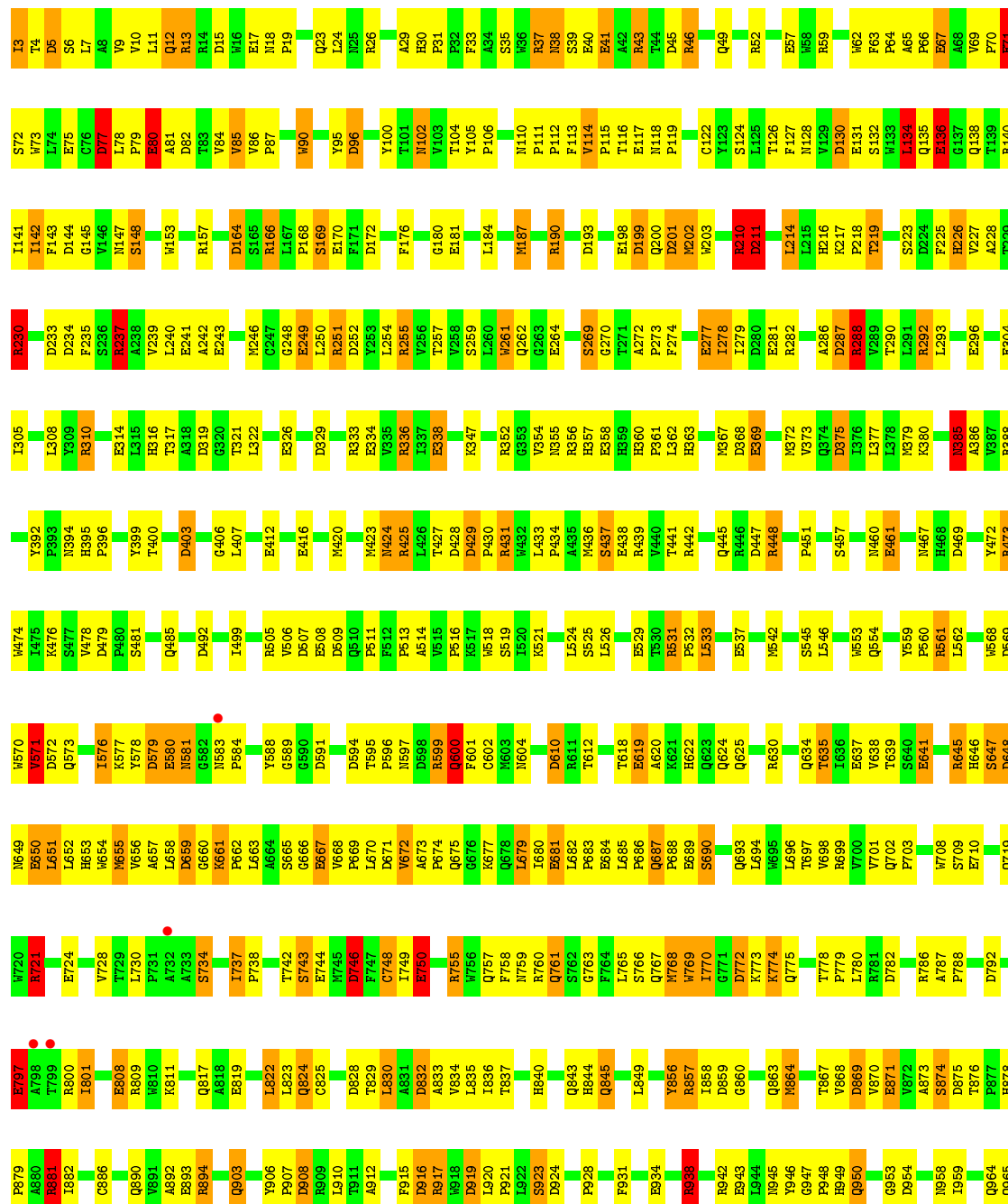
• Molecule 1: BETA-GALACTOSIDASE

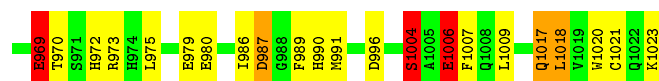




• Molecule 1: BETA-GALACTOSIDASE

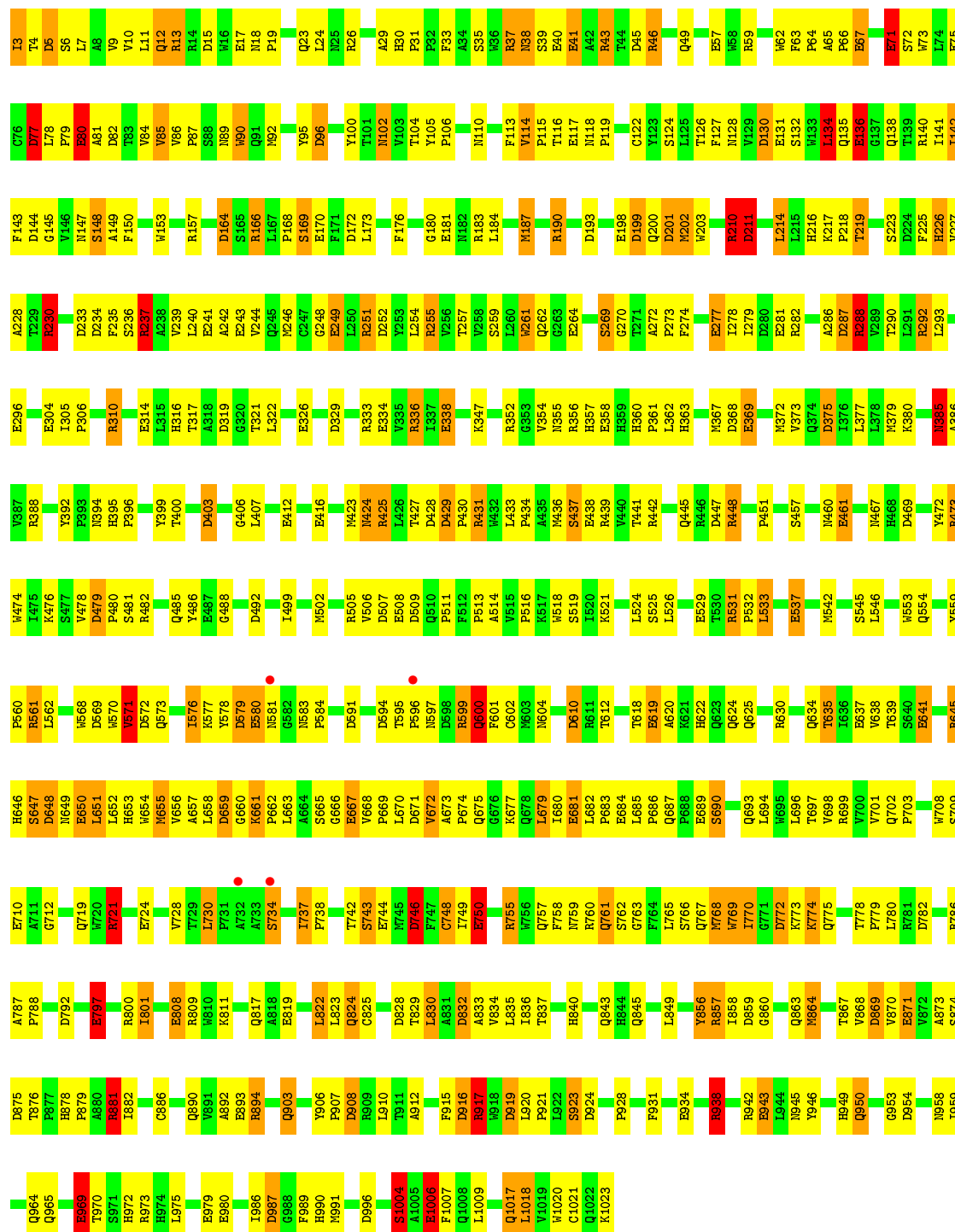
Chain B: 49% 38% 11% •





• Molecule 1: BETA-GALACTOSIDASE

Chain C: 49% 38% 11%



• Molecule 1: BETA-GALACTOSIDASE

E969	E970	S971	H972	R973	H974	L975	E976	E977	H720	L651	D569	R473	R388	E304	T229	I142	L74	I3
E969	E970	E980	R881	E982	C886	E987	E988	E989	R721	L652	D570	H474	R388	E304	T230	I142	L74	I3
S971	E972	E973	E974	E975	E976	E977	E978	E979	E979	H653	D571	H475	E979	E305	E980	I143	E75	T4
E973	E974	E975	E976	E977	E978	E979	E980	E981	E982	H654	D572	H476	E982	E306	D233	I144	E76	D5
E974	E975	E976	E977	E978	E979	E980	E981	E982	E983	H655	D573	H477	E983	E307	D234	I145	E77	S6
E975	E976	E977	E978	E979	E980	E981	E982	E983	E984	H656	D574	H478	E984	E308	F235	I146	E78	L7
E976	E977	E978	E979	E980	E981	E982	E983	E984	E985	H657	D575	H479	E985	E309	E236	I147	E79	E6
E977	E978	E979	E980	E981	E982	E983	E984	E985	E986	H658	D576	H480	E986	E310	E237	I148	E80	E9
E978	E979	E980	E981	E982	E983	E984	E985	E986	E987	H659	D577	H481	E987	E311	E238	I149	E81	V10
E979	E980	E981	E982	E983	E984	E985	E986	E987	E988	H660	D578	H482	E988	E312	E239	I150	E82	L11
E980	E981	E982	E983	E984	E985	E986	E987	E988	E989	H661	D579	H483	E989	E313	E240	I151	E83	Q12
E981	E982	E983	E984	E985	E986	E987	E988	E989	E990	H662	D580	H484	E990	E314	E241	I152	E84	R13
E982	E983	E984	E985	E986	E987	E988	E989	E990	E991	H663	D581	H485	E991	E315	E242	I153	E85	E14
E983	E984	E985	E986	E987	E988	E989	E990	E991	E992	H664	D582	H486	E992	E316	E243	I154	E86	D15
E984	E985	E986	E987	E988	E989	E990	E991	E992	E993	H665	D583	H487	E993	E317	E244	I155	E87	E16
E985	E986	E987	E988	E989	E990	E991	E992	E993	E994	H666	D584	H488	E994	E318	E245	I156	E88	N18
E986	E987	E988	E989	E990	E991	E992	E993	E994	E995	H667	D585	H489	E995	E319	E246	I157	E89	P19
E987	E988	E989	E990	E991	E992	E993	E994	E995	E996	H668	D586	H490	E996	E320	E247	I158	E90	Q23
E988	E989	E990	E991	E992	E993	E994	E995	E996	E997	H669	D587	H491	E997	E321	E248	I159	E91	L24
E989	E990	E991	E992	E993	E994	E995	E996	E997	E998	H670	D588	H492	E998	E322	E249	I160	E92	E25
E990	E991	E992	E993	E994	E995	E996	E997	E998	E999	H671	D589	H493	E999	E323	E250	I161	E93	R26
E991	E992	E993	E994	E995	E996	E997	E998	E999	E1000	H672	D590	H494	E1000	E324	E251	I162	E94	E26
E992	E993	E994	E995	E996	E997	E998	E999	E1000	E1001	H673	D591	H495	E1001	E325	E252	I163	E95	E27
E993	E994	E995	E996	E997	E998	E999	E1000	E1001	E1002	H674	D592	H496	E1002	E326	E253	I164	E96	E28
E994	E995	E996	E997	E998	E999	E1000	E1001	E1002	E1003	H675	D593	H497	E1003	E327	E254	I165	E97	E29
E995	E996	E997	E998	E999	E1000	E1001	E1002	E1003	E1004	H676	D594	H498	E1004	E328	E255	I166	E98	E30
E996	E997	E998	E999	E1000	E1001	E1002	E1003	E1004										

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 (1.50%)	wwPDB-VP
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 , 100.6	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%)	25	56
1	B	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	25	56
1	C	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	25	56
1	D	1026/1021 (100%)	953 (93%)	67 (6%)	6 (1%)	25	56
All	All	4104/4084 (100%)	3812 (93%)	268 (6%)	24 (1%)	25	56

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%)	3	10
1	B	880/873 (101%)	755 (86%)	125 (14%)	3	10
1	C	880/873 (101%)	755 (86%)	125 (14%)	3	10
1	D	880/873 (101%)	755 (86%)	125 (14%)	3	10
All	All	3520/3492 (101%)	3020 (86%)	500 (14%)	3	10

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 monosaccharides 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, 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%)	-0.82	7 (0%) 87 84	8, 29, 70, 100	19 (1%)
1	B	1021/1021 (100%)	-0.82	4 (0%) 92 91	8, 29, 70, 100	19 (1%)
1	C	1021/1021 (100%)	-0.85	4 (0%) 92 91	8, 29, 70, 100	19 (1%)
1	D	1021/1021 (100%)	-0.82	5 (0%) 91 88	8, 29, 70, 100	19 (1%)
All	All	4084/4084 (100%)	-0.83	20 (0%) 91 88	8, 29, 71, 100	76 (1%)

The worst 5 of 20 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	3.7
1	A	581	ASN	3.3
1	B	732	ALA	3.2

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(\AA^2)	Q<0.9
2	MG	B	3002	1/1	0.93	0.09	31,31,31,31	0
2	MG	A	3001	1/1	0.96	0.08	28,28,28,28	0
2	MG	D	3002	1/1	0.96	0.12	31,31,31,31	0
2	MG	C	3002	1/1	0.96	0.09	31,31,31,31	0
2	MG	A	3002	1/1	0.98	0.11	31,31,31,31	0
2	MG	D	3001	1/1	0.99	0.08	28,28,28,28	0
2	MG	C	3001	1/1	0.99	0.05	28,28,28,28	0
2	MG	B	3001	1/1	0.99	0.08	28,28,28,28	0

6.5 Other polymers [i](#)

There are no such residues in this entry.