



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 23, 2017 – 04:48 PM EST

PDB ID : 1HN1
Title : E. COLI (LAC Z) BETA-GALACTOSIDASE (ORTHORHOMBIC)
Authors : Juers, D.H.; Matthews, B.W.
Deposited on : unknown
Resolution : 3.00 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

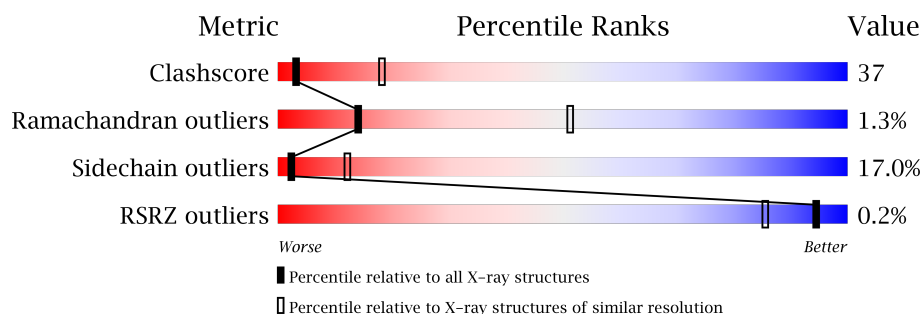
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

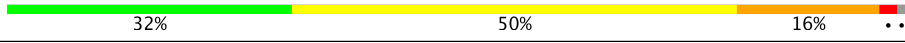
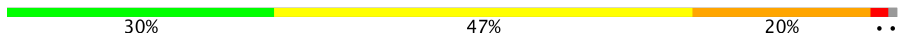
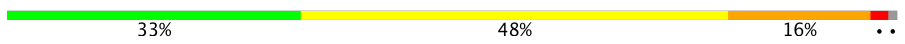

The reported resolution of this entry is 3.00 Å.

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))
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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	1023	
1	B	1023	
1	C	1023	
1	D	1023	

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	MG	A	3001	-	-	-	X
2	MG	C	3001	-	-	-	X
2	MG	D	3001	-	-	-	X
2	MG	D	3002	-	-	-	X
3	NA	A	3101	-	-	-	X
3	NA	A	3102	-	-	-	X
3	NA	B	3102	-	-	-	X
3	NA	C	3101	-	-	-	X
3	NA	C	3102	-	-	-	X
3	NA	D	3102	-	-	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 32954 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	1011	Total	C	N	O	S	0	2	0
			8136	5145	1443	1510	38			
1	B	1011	Total	C	N	O	S	0	2	0
			8136	5145	1443	1510	38			
1	C	1011	Total	C	N	O	S	0	2	0
			8136	5145	1443	1510	38			
1	D	1011	Total	C	N	O	S	0	1	0
			8130	5141	1441	1510	38			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P00722
A	2	SER	-	CLONING ARTIFACT	UNP P00722
A	3	HIS	-	CLONING ARTIFACT	UNP P00722
A	4	MET	-	CLONING ARTIFACT	UNP P00722
A	5	LEU	-	CLONING ARTIFACT	UNP P00722
A	6	GLU	-	CLONING ARTIFACT	UNP P00722
A	7	ASP	-	CLONING ARTIFACT	UNP P00722
A	8	PRO	-	CLONING ARTIFACT	UNP P00722
B	1	GLY	-	CLONING ARTIFACT	UNP P00722
B	2	SER	-	CLONING ARTIFACT	UNP P00722
B	3	HIS	-	CLONING ARTIFACT	UNP P00722
B	4	MET	-	CLONING ARTIFACT	UNP P00722
B	5	LEU	-	CLONING ARTIFACT	UNP P00722
B	6	GLU	-	CLONING ARTIFACT	UNP P00722
B	7	ASP	-	CLONING ARTIFACT	UNP P00722
B	8	PRO	-	CLONING ARTIFACT	UNP P00722
C	1	GLY	-	CLONING ARTIFACT	UNP P00722
C	2	SER	-	CLONING ARTIFACT	UNP P00722
C	3	HIS	-	CLONING ARTIFACT	UNP P00722
C	4	MET	-	CLONING ARTIFACT	UNP P00722
C	5	LEU	-	CLONING ARTIFACT	UNP P00722

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	6	GLU	-	CLONING ARTIFACT	UNP P00722
C	7	ASP	-	CLONING ARTIFACT	UNP P00722
C	8	PRO	-	CLONING ARTIFACT	UNP P00722
D	1	GLY	-	CLONING ARTIFACT	UNP P00722
D	2	SER	-	CLONING ARTIFACT	UNP P00722
D	3	HIS	-	CLONING ARTIFACT	UNP P00722
D	4	MET	-	CLONING ARTIFACT	UNP P00722
D	5	LEU	-	CLONING ARTIFACT	UNP P00722
D	6	GLU	-	CLONING ARTIFACT	UNP P00722
D	7	ASP	-	CLONING ARTIFACT	UNP P00722
D	8	PRO	-	CLONING ARTIFACT	UNP P00722

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

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

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	2	Total Na 2 2	0	0
3	A	2	Total Na 2 2	0	0
3	D	1	Total Na 1 1	0	0
3	C	2	Total Na 2 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	103	Total O 103 103	0	0

Continued on next page...

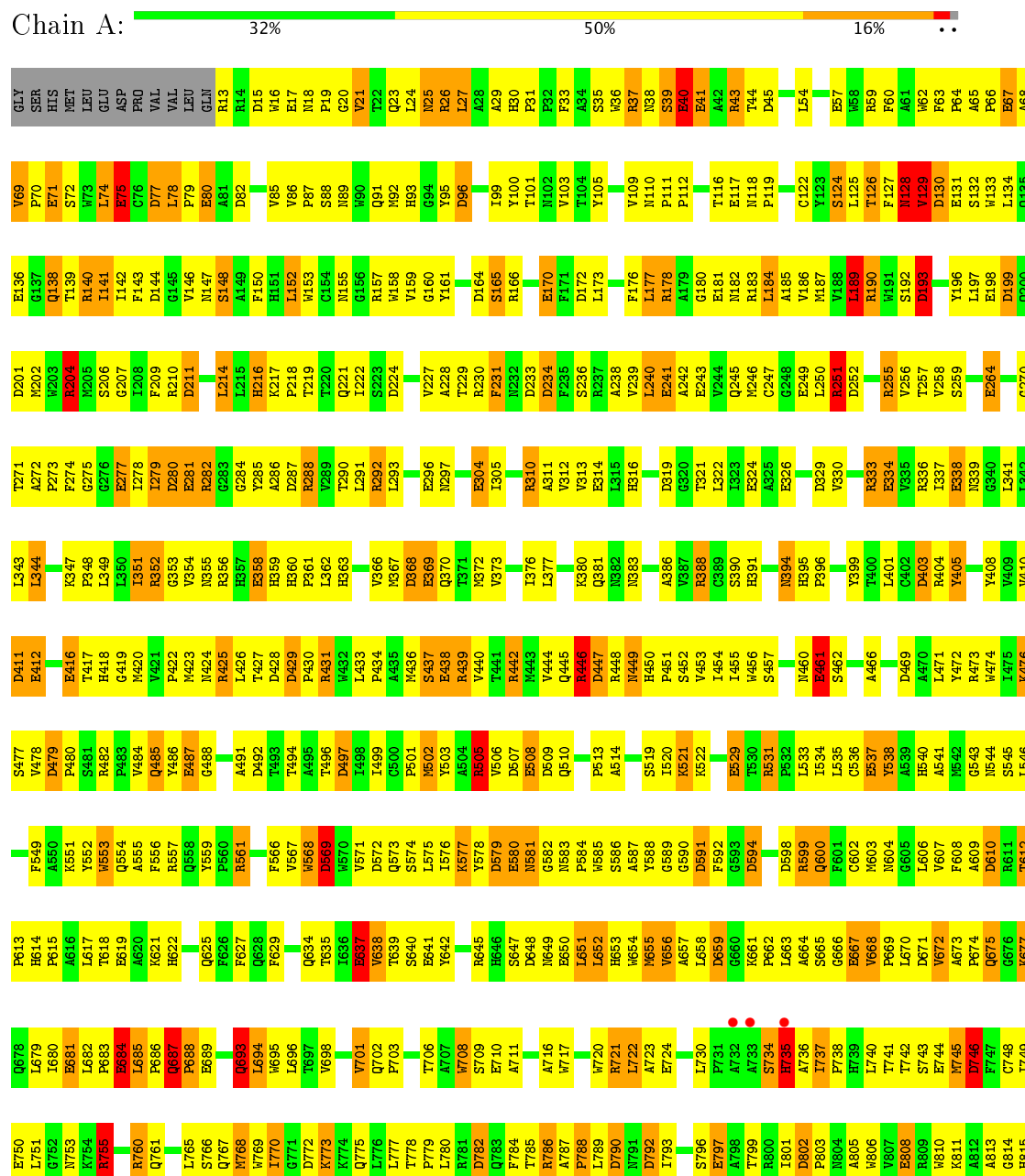
Continued from previous page...

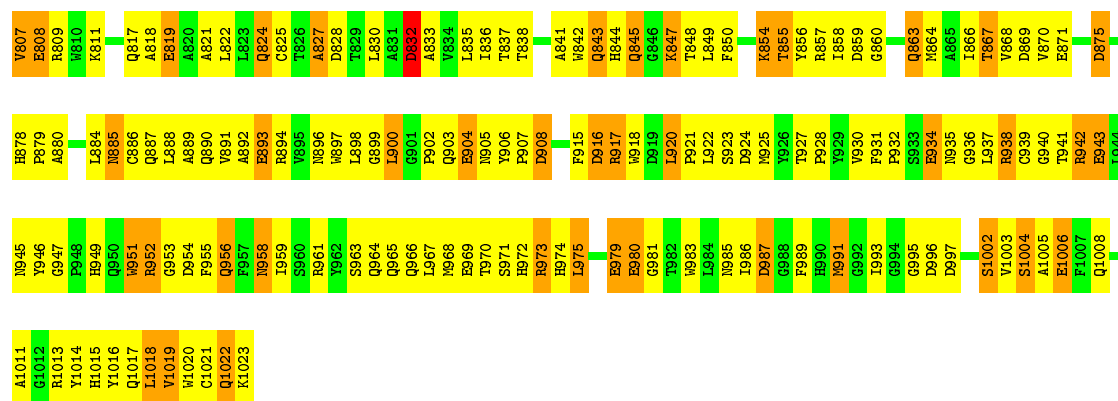
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	92	Total 92	O 92	0	0
4	C	108	Total 108	O 108	0	0
4	D	98	Total 98	O 98	0	0

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.

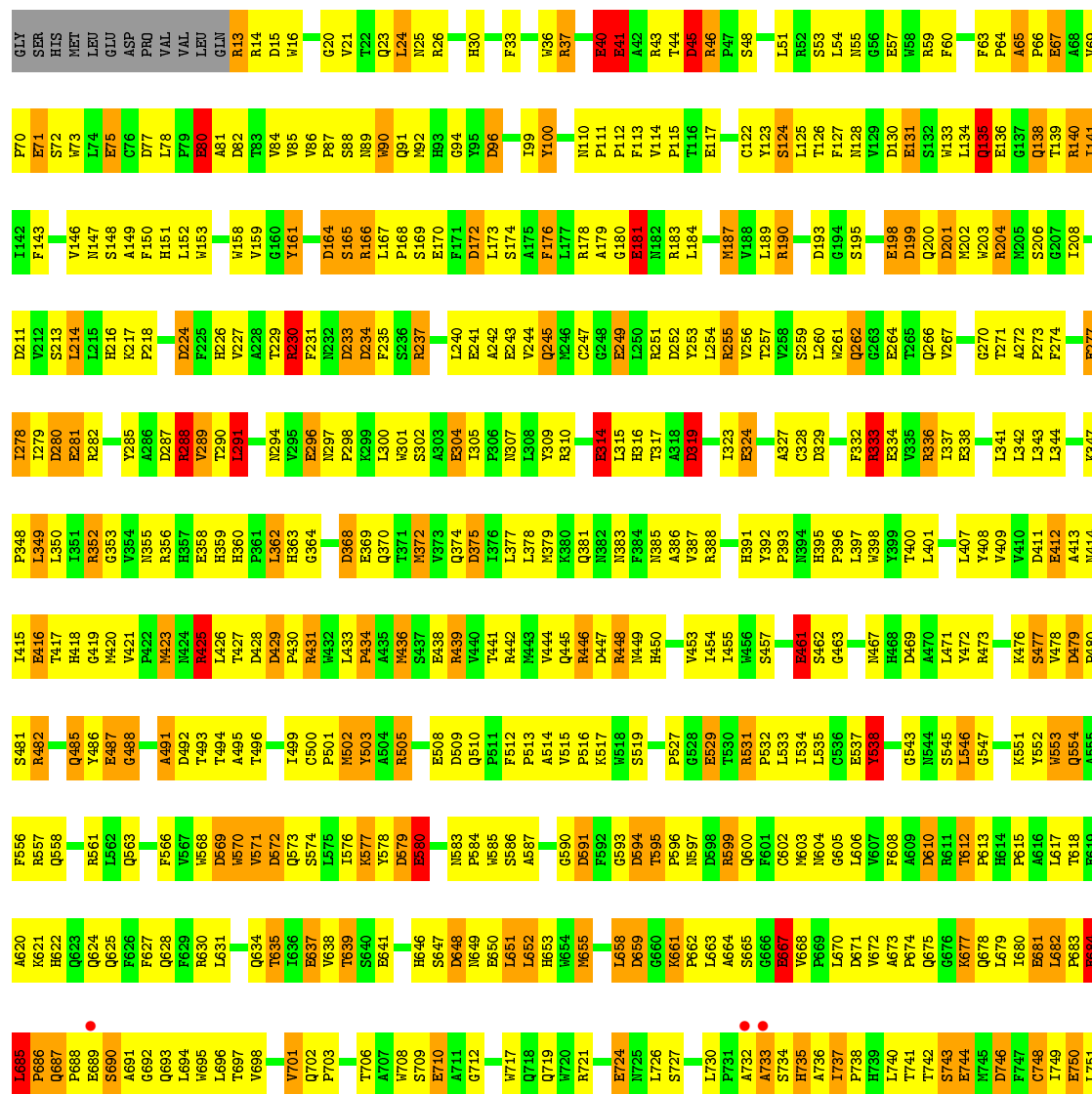
• Molecule 1: BETA-GALACTOSIDASE

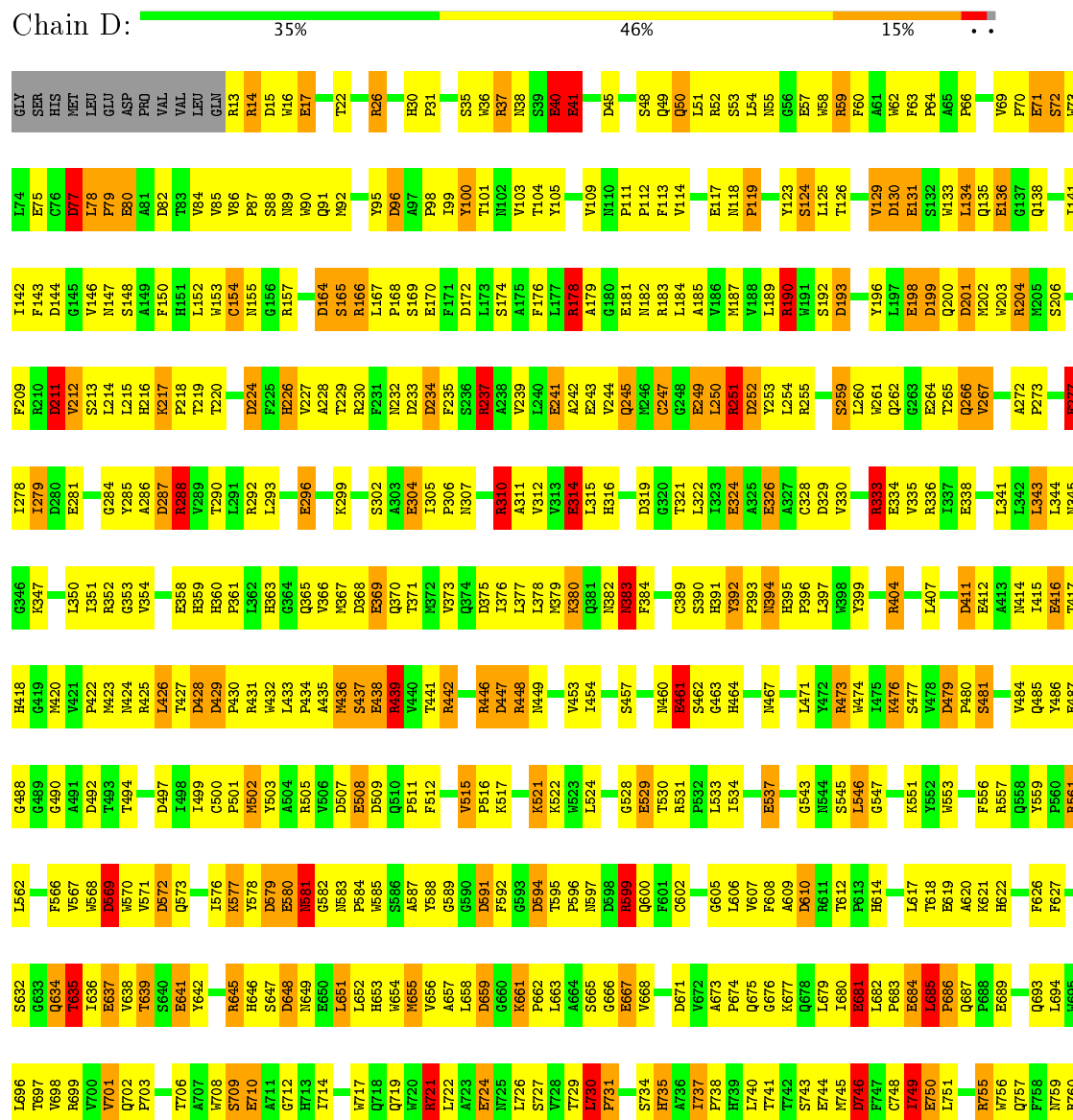




• Molecule 1: BETA-GALACTOSIDASE

Chain C: 33% 48% 16%





S960	S961	E969	T970	H971	H972	H973	H974	L975	A976	E979	E980	T981	T982	H983	L984	H985	T986	D987	H988	H989	H990	H991	D996	D997	S998	S1004	A1005	E1006	F1007	A1011	G1012	R1013	Y1014	Q1017	L1018	Y1019	H1020	C1021	Q1022	K1023	L898	G899	L900	P901	P902	Q903	E904	H905	Y906	P907	D908	H909	L910	T911	A912	A913	C914	H915	D916	H917	H918	D919	L920	P921	D924	H925	Y926	T927	P928	Y929	Y930	P931	P932	S933	E934	H935	G936	L937	H938	C939	T941	G940	T942	E943	Y946	G947	P948	H949	Q950	H951	H952	G953	D954	H955	Q956	F957	H958	I959	H815	L816	Q817	A818	E819	A820	A821	Q824	C825	T826	A827	D828	T829	L830	A831	D832	H833	A836	T837	T838	A839	H840	A841	H842	Q843	H844	H845	F850	R853	K854	T855	Y856	R857	I858	D859	G860	I866	T867	V868	D869	V870	E871	V872	D875	T876	P877	H878	P879	A880	R881	I882	G883	L884	Q887	L888	A889	Q890	V891	A892	E893	R894	V895	H896	W897	L765	S766	Q767	H768	L770	G771	D772	K773	K774	Q775	L776	L777	L778	H779	L780	K781	D782	Q783	F784	T785	H786	A787	P788	L789	D790	H791	D792	L793	G794	H795	S796	E797	A798	T799	R800	I801	D802	P803	H804	A805	H806	H807	E808	R809	H815	L816	Q817	A818	E819	A820	A821	Q824	C825	T826
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	153.90 Å 171.40 Å 204.50 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 3.00 29.96 – 2.99	Depositor EDS
% Data completeness (in resolution range)	94.0 (15.00-3.00) 88.8 (29.96-2.99)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.47 (at 3.00 Å)	Xtriage
Refinement program	TNT	Depositor
R, R_{free}	0.148 , 0.299 0.133 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å ²)	31.4	Xtriage
Anisotropy	0.253	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 132.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	32954	wwPDB-VP
Average B, all atoms (Å ²)	42.0	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 41.90 % 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 2.2078e-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: NA, 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	58/8387 (0.7%)	1.66	166/11442 (1.5%)
1	B	1.07	55/8387 (0.7%)	1.65	170/11442 (1.5%)
1	C	1.07	56/8387 (0.7%)	1.65	157/11442 (1.4%)
1	D	1.09	58/8376 (0.7%)	1.65	168/11427 (1.5%)
All	All	1.08	227/33537 (0.7%)	1.65	661/45753 (1.4%)

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	B	1	0
1	D	1	0
All	All	2	0

The worst 5 of 227 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	314	GLU	CD-OE2	8.73	1.35	1.25
1	D	304	GLU	CD-OE2	8.61	1.35	1.25
1	D	334	GLU	CD-OE2	8.13	1.34	1.25
1	A	304	GLU	CD-OE2	8.00	1.34	1.25
1	C	461	GLU	CD-OE2	7.99	1.34	1.25

The worst 5 of 661 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	685	LEU	C-N-CD	-21.41	73.50	120.60
1	B	730	LEU	C-N-CD	-21.06	74.26	120.60

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	987	ASP	CB-CG-OD2	-13.47	106.18	118.30
1	A	687	GLN	C-N-CD	-12.61	92.86	120.60
1	D	166	ARG	NE-CZ-NH2	-12.33	114.13	120.30

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	B	37	ARG	CA
1	D	951	TRP	CA

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	8136	0	7723	593	0
1	B	8136	0	7723	660	0
1	C	8136	0	7723	548	0
1	D	8130	0	7720	550	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	2	0	0	0	0
3	B	2	0	0	0	0
3	C	2	0	0	0	0
3	D	1	0	0	0	0
4	A	103	0	0	2	0
4	B	92	0	0	9	0
4	C	108	0	0	8	0
4	D	98	0	0	6	0
All	All	32954	0	30889	2319	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 37.

The worst 5 of 2319 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:D:502:MET:HB2	1:D:537:GLU:HB2	1.25	1.18
1:D:734:SER:HB3	1:D:860:GLY:HA3	1.20	1.11
1:B:18:ASN:HD22	1:B:21:VAL:HG23	1.09	1.10
1:B:737:ILE:HD12	1:B:738:PRO:HD2	1.26	1.09
1:A:737:ILE:HG13	1:A:832:ASP:HA	1.35	1.08

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	1011/1023 (99%)	902 (89%)	99 (10%)	10 (1%)	18	59
1	B	1011/1023 (99%)	901 (89%)	96 (10%)	14 (1%)	13	49
1	C	1011/1023 (99%)	904 (89%)	93 (9%)	14 (1%)	13	49
1	D	1010/1023 (99%)	904 (90%)	93 (9%)	13 (1%)	14	51
All	All	4043/4092 (99%)	3611 (89%)	381 (9%)	51 (1%)	14	51

5 of 51 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	129	VAL
1	A	425	ARG
1	A	688	PRO
1	B	201	ASP
1	B	647	SER

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	866/875 (99%)	723 (84%)	143 (16%)	2	13
1	B	866/875 (99%)	699 (81%)	167 (19%)	1	9
1	C	866/875 (99%)	732 (84%)	134 (16%)	3	15
1	D	865/875 (99%)	721 (83%)	144 (17%)	2	13
All	All	3463/3500 (99%)	2875 (83%)	588 (17%)	2	12

5 of 588 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	766	SER
1	C	189	LEU
1	D	749	ILE
1	B	797	GLU
1	B	952	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 85 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	687	GLN
1	B	1022	GLN
1	D	804	ASN
1	B	702	GLN
1	B	735	HIS

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 15 ligands modelled in this entry, 15 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	1011/1023 (98%)	-0.82	3 (0%) 93 82	10, 35, 92, 211	0
1	B	1011/1023 (98%)	-0.78	0 100 100	8, 38, 101, 212	0
1	C	1011/1023 (98%)	-0.79	4 (0%) 92 77	14, 34, 91, 214	0
1	D	1011/1023 (98%)	-0.83	0 100 100	5, 34, 93, 210	0
All	All	4044/4092 (98%)	-0.81	7 (0%) 94 85	5, 35, 94, 214	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	689	GLU	3.0
1	A	735	HIS	2.7
1	A	733	ALA	2.6
1	C	733	ALA	2.4
1	C	800	ARG	2.1

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, 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
3	NA	C	3101	1/1	0.94	0.43	10.24	49,49,49,49	0
2	MG	D	3001	1/1	0.98	0.29	10.20	26,26,26,26	0
2	MG	A	3001	1/1	0.96	0.26	8.12	29,29,29,29	0
2	MG	C	3001	1/1	0.97	0.26	8.10	25,25,25,25	0
3	NA	A	3102	1/1	0.96	0.38	7.18	24,24,24,24	0
3	NA	A	3101	1/1	0.94	0.25	6.95	52,52,52,52	0
3	NA	D	3102	1/1	0.99	0.30	6.92	47,47,47,47	0
3	NA	C	3102	1/1	0.96	0.23	3.45	24,24,24,24	0
3	NA	B	3102	1/1	0.93	0.21	2.92	32,32,32,32	0
2	MG	D	3002	1/1	0.98	0.22	2.09	30,30,30,30	0
2	MG	C	3002	1/1	0.99	0.18	1.68	25,25,25,25	0
2	MG	B	3001	1/1	0.99	0.15	0.58	13,13,13,13	0
2	MG	B	3002	1/1	0.94	0.12	0.12	30,30,30,30	0
3	NA	B	3103	1/1	0.96	0.13	0.02	45,45,45,45	0
2	MG	A	3002	1/1	0.98	0.08	-1.89	29,29,29,29	0

6.5 Other polymers [i](#)

There are no such residues in this entry.