



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 2, 2023 – 11:54 PM EDT

PDB ID : 1HN1
Title : E. COLI (LAC Z) BETA-GALACTOSIDASE (ORTHORHOMBIC)
Authors : Juers, D.H.; Matthews, B.W.
Deposited on : 2000-12-05
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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.34
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) : 2.34

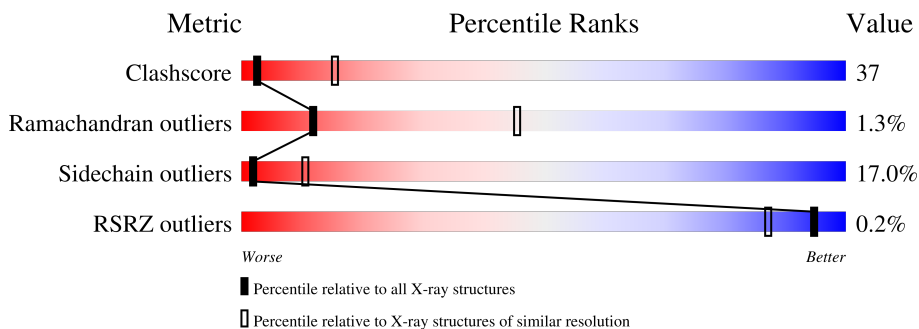
1 Overall quality at a glance

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	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (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 of 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	1023	
1	B	1023	
1	C	1023	
1	D	1023	

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
			Total	C	N	O	S			
1	A	1011	8136	5145	1443	1510	38	0	2	0
1	B	1011	8136	5145	1443	1510	38	0	2	0
1	C	1011	8136	5145	1443	1510	38	0	2	0
1	D	1011	8130	5141	1441	1510	38	0	1	0

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	A	2	Total Mg 2 2	0	0
2	B	2	Total Mg 2 2	0	0
2	C	2	Total Mg 2 2	0	0
2	D	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	A	2	Total Na 2 2	0	0
3	B	2	Total Na 2 2	0	0
3	C	2	Total Na 2 2	0	0
3	D	1	Total Na 1 1	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...

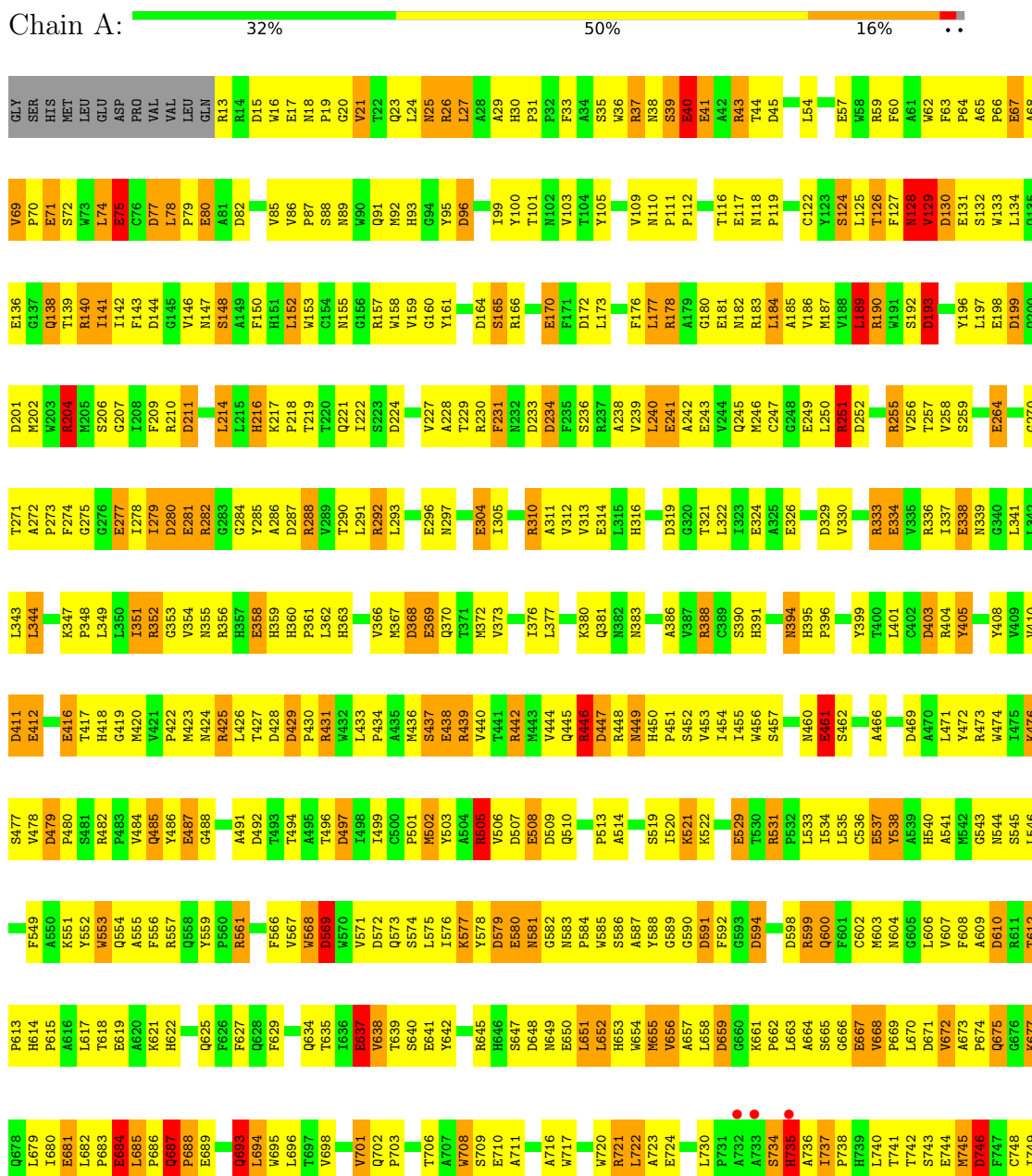
Continued from previous page...

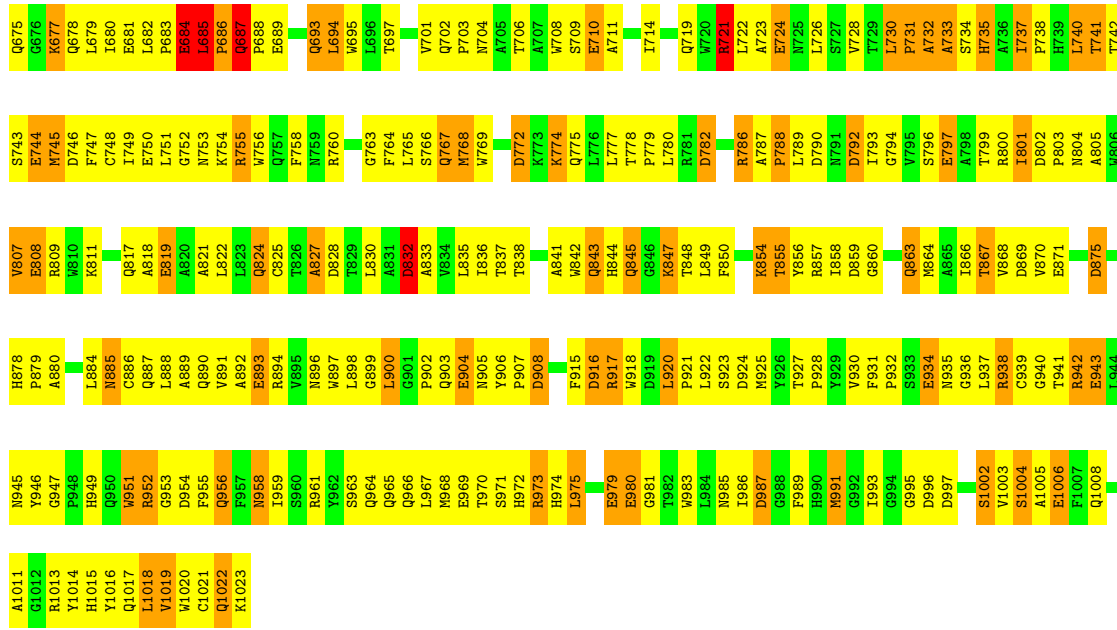
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	92	Total 92	O 92	0	0
4	C	108	Total 108	O 108	0	0
4	D	98	Total 98	O 98	0	0

3 Residue-property plots

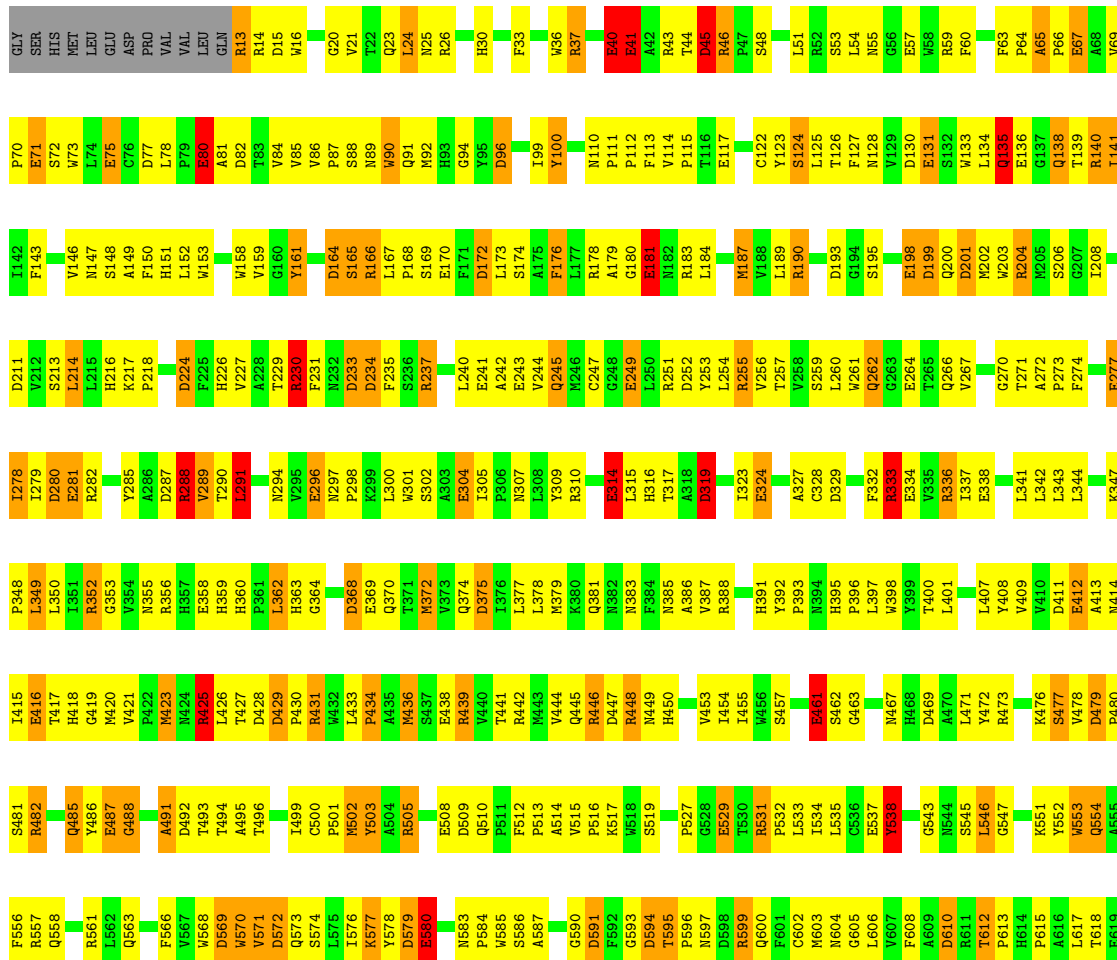
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.

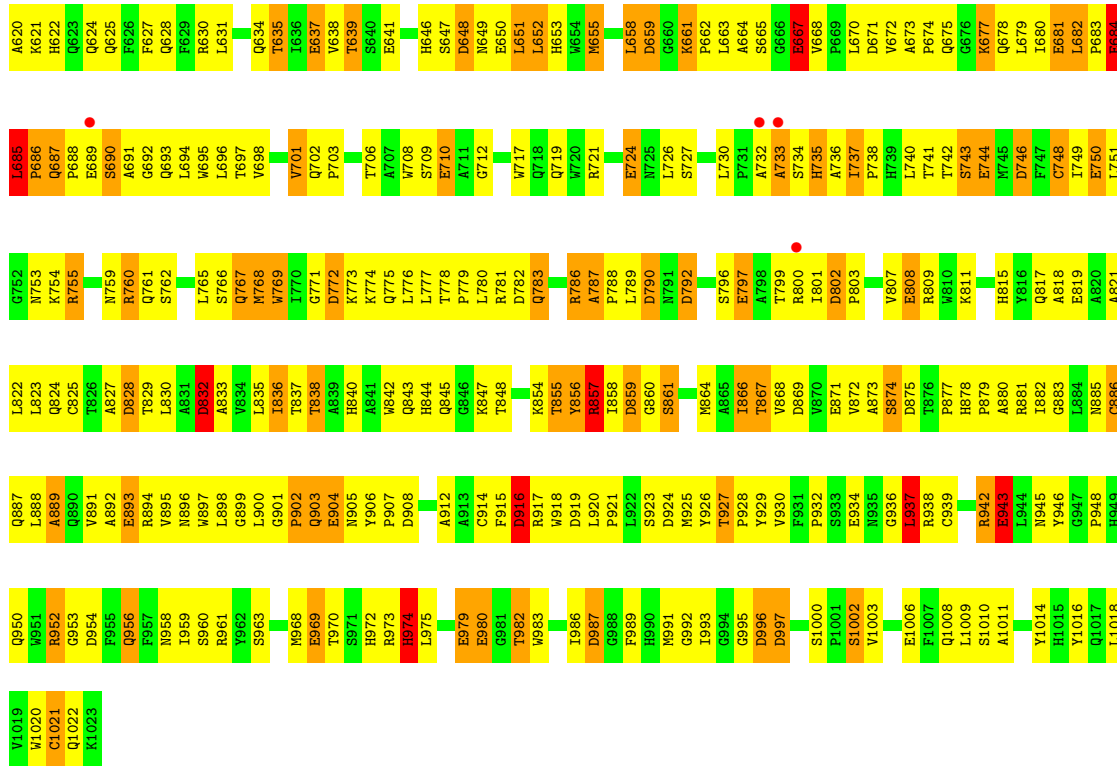
• Molecule 1: BETA-GALACTOSIDASE



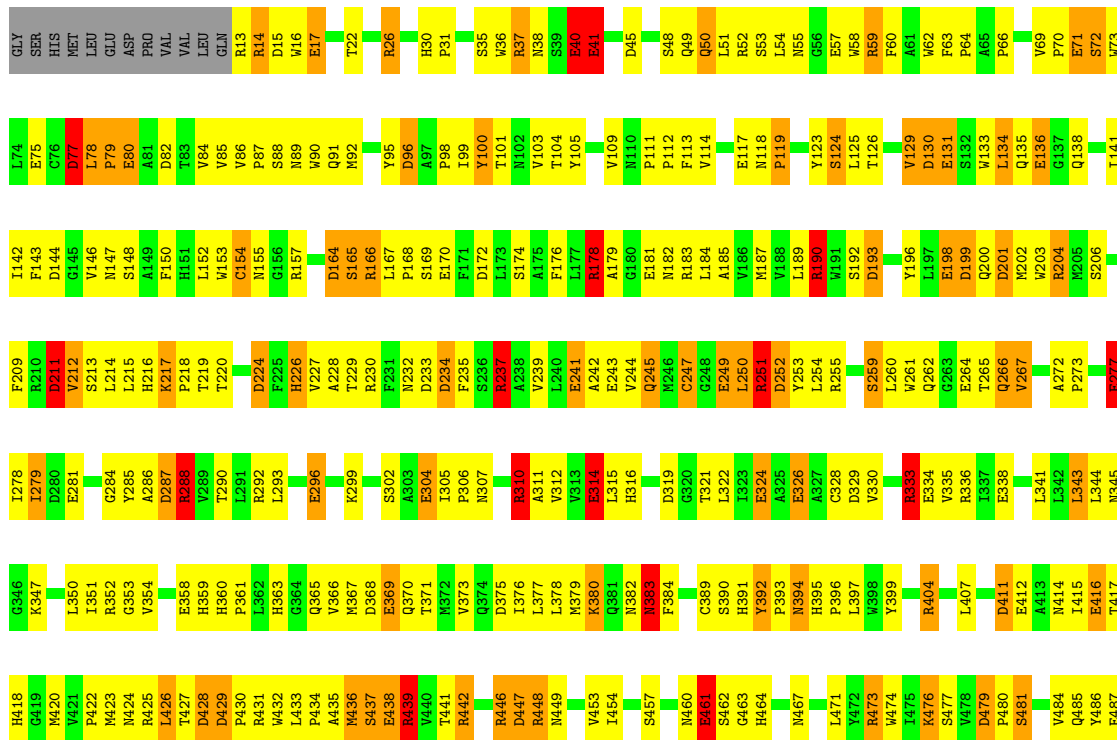
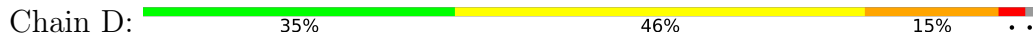


● Molecule 1: BETA-GALACTOSIDASE





● Molecule 1: BETA-GALACTOSIDASE



S960	S961	L888	A827	Q761	L696	S632	L562	G488
G899	L900	D828	D828	S762	T697	G633	F566	G489
G901	G901	L830	L830	L765	V698	Q634	V567	G490
P902	P902	A831	L831	S766	R699	L835	A491	A491
Q903	Q903	D832	D832	Q767	V700	E636	W568	D492
E904	E904	A833	A833	M768	Q702	E637	D569	T493
N905	N905	W769	W769	P703	P703	W638	W570	T494
Y906	Y906	I770	I770	T706	D572	S640	D497	D497
P907	P907	G771	G771	A707	Q573	E641	I498	I498
D908	D908	K773	K773	W708	I576	E642	I499	I499
R909	R909	A839	A839	S709	K577	R645	C500	C500
L910	L910	H840	H840	E710	Y578	H646	P501	P501
T911	T911	A841	A841	A711	D579	S647	M502	M502
A912	A912	W842	W842	L776	E580	H646	Y503	Y503
A913	A913	Q843	Q843	L777	N649	D648	A504	A504
Q914	Q914	H844	H844	H713	N649	D648	R581	R581
F915	F915	Q845	Q845	T778	E650	N649	R505	R505
D916	D916	W779	W779	I714	E650	N649	V566	V566
A917	A917	L780	L780	W174	L651	L651	D507	D507
W918	W918	R781	R781	W177	L652	L652	E508	E508
L919	L919	D782	D782	Q718	H653	H653	D509	D509
L920	L920	Q783	Q783	Q719	W554	W554	Q510	Q510
P921	P921	F784	F784	W720	M655	M655	P511	P511
D924	D924	T855	T855	R721	V556	V556	F512	F512
N925	N925	Y856	Y856	L722	A657	A657	V515	V515
Y926	Y926	A857	A857	L723	L658	L658	P516	P516
P927	P927	I858	I858	E724	D659	D659	K517	K517
P928	P928	D859	D859	N725	F592	F592	K521	K521
Y929	Y929	G860	G860	D790	K661	K661	K522	K522
V930	V930	I866	I866	N791	D662	D662	W523	W523
F931	F931	T867	T867	D792	T595	T595	L524	L524
S932	S932	W868	W868	W798	N596	N596	G528	G528
S933	S933	E869	E869	L729	D598	D598	E529	E529
N934	N934	V870	V870	P731	R599	R599	Q530	Q530
G935	G935	E871	E871	S734	F601	F601	T530	T530
L937	L937	W872	W872	H735	G602	G602	R531	R531
C939	C939	D875	D875	A736	D671	D671	P532	P532
G940	G940	T876	T876	I737	V672	V672	L533	L533
T941	T941	P877	P877	P738	A673	A673	I534	I534
R942	R942	H878	H878	H739	L606	L606	E537	E537
E943	E943	P879	P879	L740	V607	V607	G543	G543
Y946	Y946	A880	A880	T742	F608	F608	M544	M544
P948	P948	R881	R881	S743	D610	D610	S545	S545
Q950	Q950	I882	I882	E744	R611	R611	L546	L546
R951	R951	G883	G883	M745	T612	T612	G547	G547
G952	G952	L884	L884	D746	P613	P613	K551	K551
G953	G953	Q887	Q887	F747	L682	L682	Y552	Y552
F954	F954	L888	L888	C748	E684	E684	W553	W553
F955	F955	Q889	Q889	I749	E684	E684	F556	F556
Q956	Q956	A889	A889	E750	L685	L685	R557	R557
P957	P957	R890	R890	L751	P686	P686	Q558	Q558
N958	N958	E891	E891	W755	Q687	Q687	Y559	Y559
F959	F959	A892	A892	W756	P688	P688	P560	P560
N959	N959	R893	R893	Q757	E689	E689	F626	F626
F960	F960	R894	R894	P758	Q693	Q693	L627	L627
G961	G961	W895	W895	N759	L694	L694	W561	W561
G962	G962	N896	N896	R760	W695	W695		
G963	G963	W897	W897					

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	153.90Å 171.40Å 204.50Å 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 3.00Å)	Xtrriage
Refinement program	TNT	Depositor
R, R_{free}	0.148 , 0.299 0.134 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å ²)	31.4	Xtrriage
Anisotropy	0.253	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 132.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
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

Xtrriage'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.*

¹Intensities estimated from amplitudes.

²Theoretical 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 i

5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: MG, NA

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 [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	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%)	15	53
1	B	1011/1023 (99%)	901 (89%)	96 (10%)	14 (1%)	11	43
1	C	1011/1023 (99%)	904 (89%)	93 (9%)	14 (1%)	11	43
1	D	1010/1023 (99%)	904 (90%)	93 (9%)	13 (1%)	12	45
All	All	4043/4092 (99%)	3611 (89%)	381 (9%)	51 (1%)	12	45

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	11
1	B	866/875 (99%)	699 (81%)	167 (19%)	1	8
1	C	866/875 (99%)	732 (84%)	134 (16%)	2	13
1	D	865/875 (99%)	721 (83%)	144 (17%)	2	11
All	All	3463/3500 (99%)	2875 (83%)	588 (17%)	2	10

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

Mol	Chain	Res	Type
1	D	152	LEU
1	D	914	CYS
1	D	251	ARG
1	D	138	GLN
1	D	645	ARG

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

Mol	Chain	Res	Type
1	C	687	GLN
1	D	394	ASN
1	C	704	ASN
1	C	878	HIS
1	D	702	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 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%) 94 84	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 79	14, 34, 90, 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%) 95 87	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	732	ALA	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 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
3	NA	B	3102	1/1	0.93	0.21	32,32,32,32	0
2	MG	B	3002	1/1	0.94	0.12	30,30,30,30	0
3	NA	C	3101	1/1	0.94	0.43	49,49,49,49	0
3	NA	A	3101	1/1	0.95	0.25	52,52,52,52	0
2	MG	A	3001	1/1	0.96	0.26	29,29,29,29	0
3	NA	B	3103	1/1	0.96	0.13	45,45,45,45	0
3	NA	A	3102	1/1	0.96	0.38	24,24,24,24	0
3	NA	C	3102	1/1	0.96	0.23	24,24,24,24	0
2	MG	C	3001	1/1	0.97	0.26	25,25,25,25	0
2	MG	A	3002	1/1	0.98	0.08	29,29,29,29	0
2	MG	D	3001	1/1	0.98	0.28	26,26,26,26	0
2	MG	D	3002	1/1	0.98	0.22	30,30,30,30	0
3	NA	D	3102	1/1	0.98	0.30	47,47,47,47	0
2	MG	C	3002	1/1	0.99	0.18	25,25,25,25	0
2	MG	B	3001	1/1	0.99	0.15	13,13,13,13	0

6.5 Other polymers [i](#)

There are no such residues in this entry.