



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 31, 2020 – 10:14 AM BST

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

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

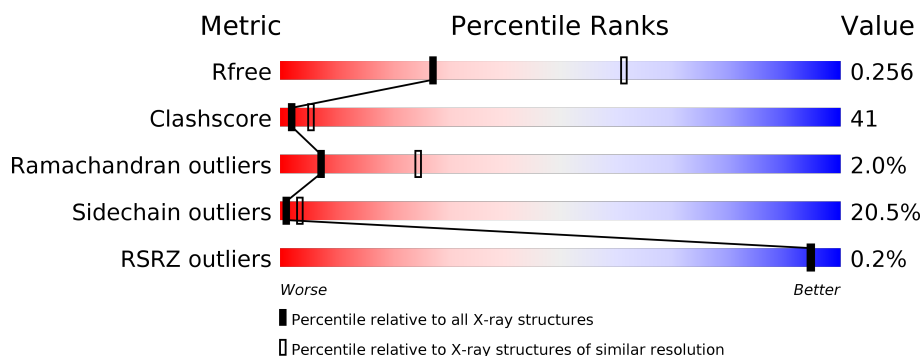
1 Overall quality at a glance

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>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>

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		

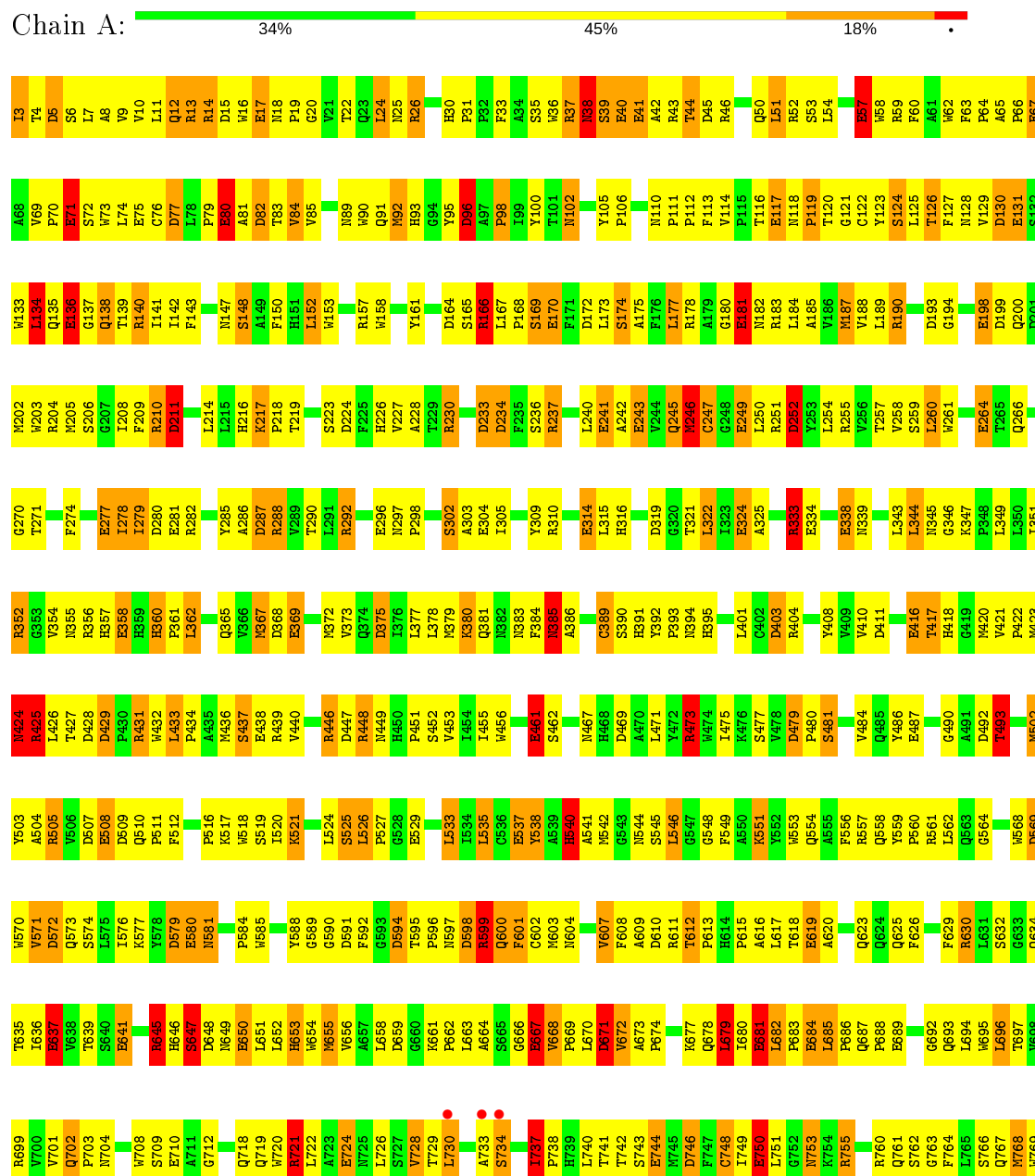
- Molecule 3 is water.

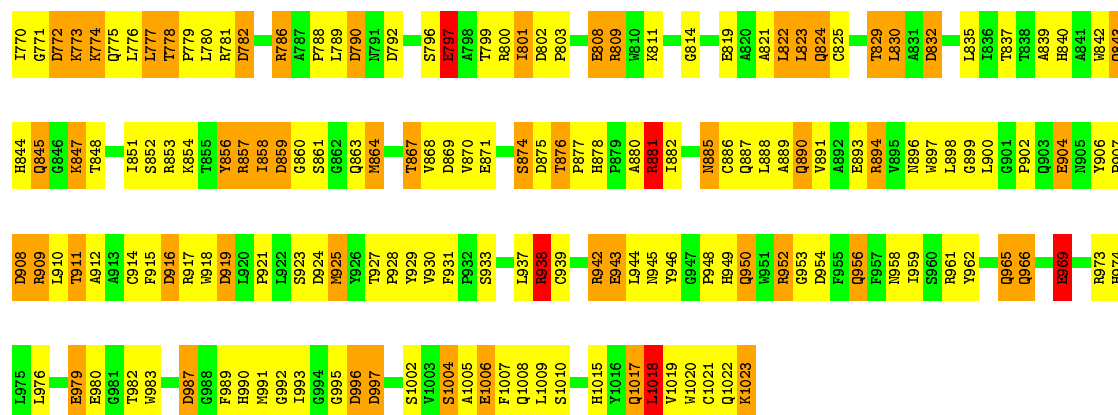
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	211	Total	O	0	0
			211	211		
3	B	202	Total	O	0	0
			202	202		
3	C	220	Total	O	0	0
			220	220		
3	D	212	Total	O	0	0
			212	212		

3 Residue-property plots

These plots are drawn for all protein, RNA, 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.

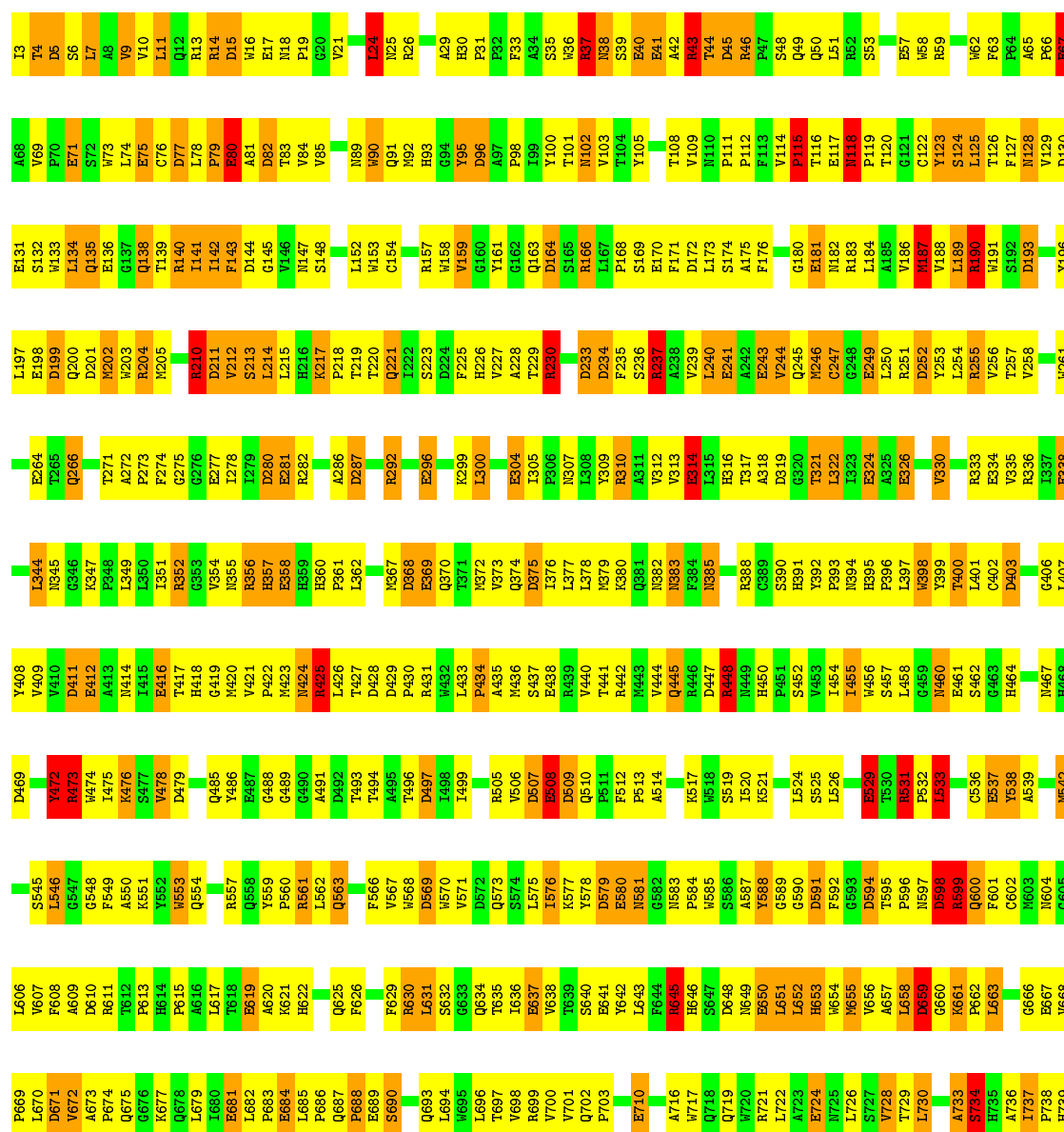
• Molecule 1: BETA-GALACTOSIDASE

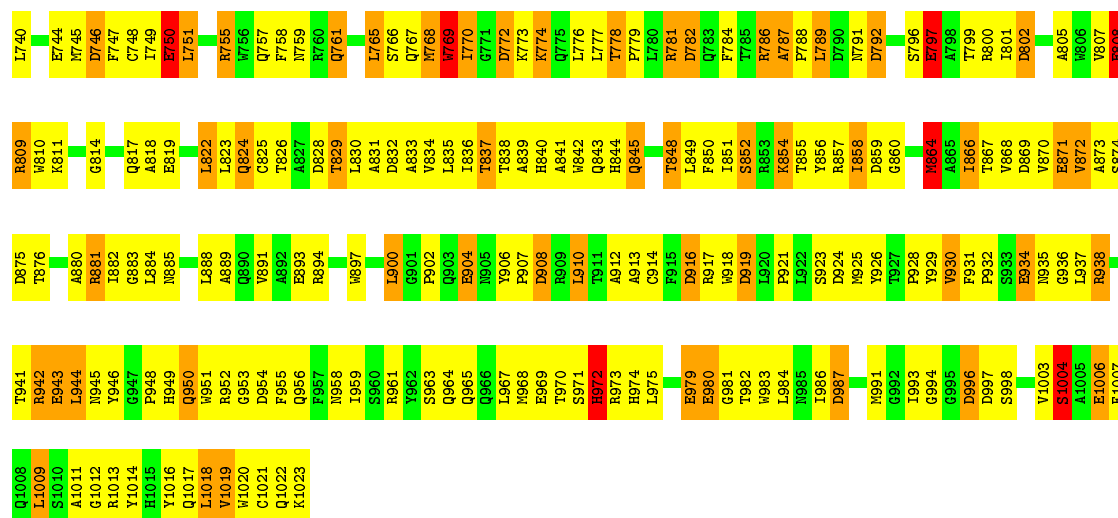




• Molecule 1: BETA-GALACTOSIDASE

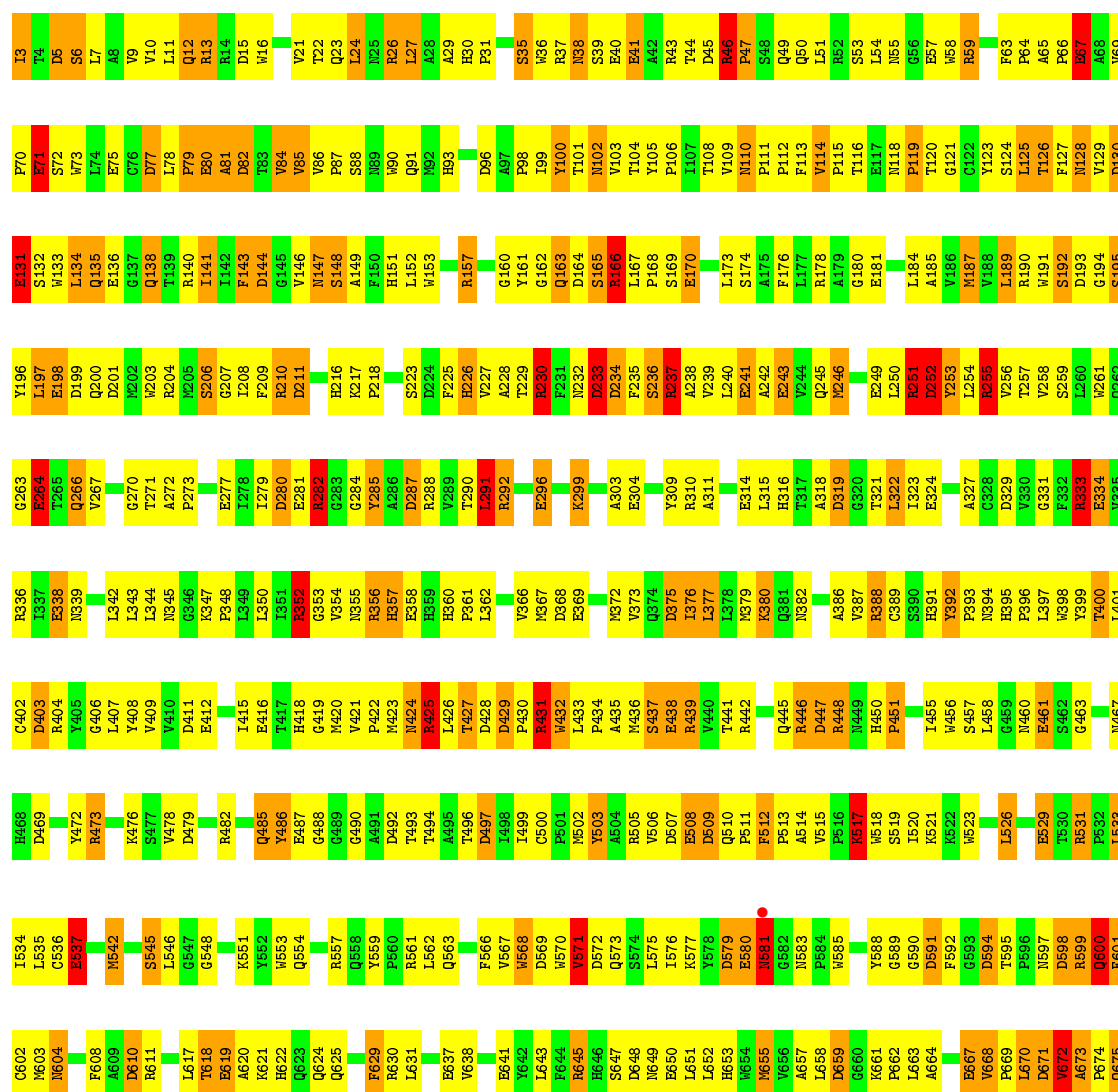
Chain B: 30% 48% 19%





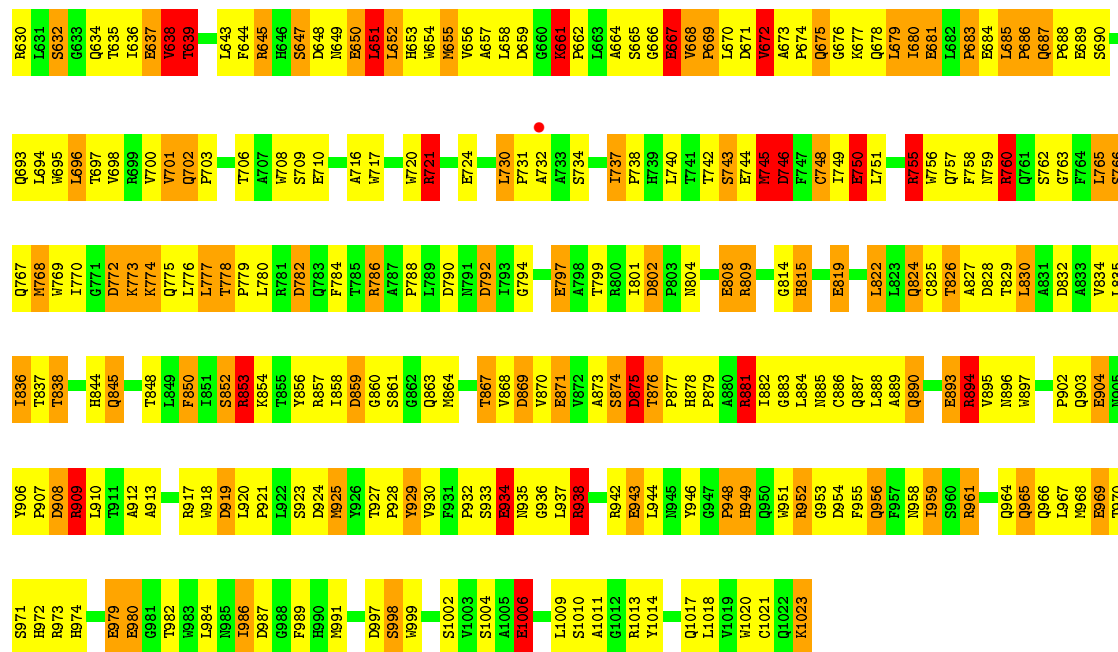
● Molecule 1: BETA-GALACTOSIDASE

Chain C: 32% 47% 18%





Y486 Y487 Y488 Y489 Y490 Y500 Y501 Y502 Y503 Y504 Y505 Y506 Y507 Y508 Y509 Y510 Y511 Y512 Y513 Y514 Y515 Y516 Y517 Y518 Y519 Y520 Y521 Y522 Y523 Y524 Y525 Y526 Y527 Y528 Y529 Y530 Y531 Y532 Y533 Y534 Y535 Y536 Y537 Y540 Y541 Y542 Y545 Y546 Y547 Y550 Y551 Y552	Y420	Y421	Y422	Y423	Y424	Y425	Y426	Y427	Y428	Y429	Y430	Y431	Y432	Y433	Y434	Y435	Y436	Y437	Y438	Y439	Y440	Y441	Y442	Y443	Y444	Y445	Y446	Y447	Y448	Y449	Y450	Y451	Y452	Y453	Y454	Y455	Y456	Y457	Y458	Y459	Y460	Y461	Y462	Y463	Y464	Y465	Y466	Y467	Y468	Y469	Y470	Y471	Y472	Y473	Y474	Y475	Y476	Y477	Y478	Y479	Y480	Y481	Y482	Y483	Y484	Y485	Y486	Y487	Y488	Y489	Y490	Y491	Y492	Y493	Y494	Y495	Y496	Y497	Y498	Y499	Y500	Y501	Y502	Y503	Y504	Y505	Y506	Y507	Y508	Y509	Y510	Y511	Y512	Y513	Y514	Y515	Y516	Y517	Y518	Y519	Y520	Y521	Y522	Y523	Y524	Y525	Y526	Y527	Y528	Y529	Y530	Y531	Y532	Y533	Y534	Y535	Y536	Y537	Y540	Y541	Y542	Y545	Y546	Y547	Y550	Y551	Y552																																																																				
	M420	M421	M422	M423	M424	M425	M426	M427	M428	M429	M430	M431	M432	M433	M434	M435	M436	M437	M438	M439	M440	M441	M442	M443	M444	M445	M446	M447	M448	M449	M450	M451	M452	M453	M454	M455	M456	M457	M458	M459	M460	M461	M462	M463	M464	M465	M466	M467	M468	M469	M470	M471	M472	M473	M474	M475	M476	M477	M478	M479	M480	M481	M482	M483	M484	M485	M486	M487	M488	M489	M490	M491	M492	M493	M494	M495	M496	M497	M498	M499	M500	M501	M502	M503	M504	M505	M506	M507	M508	M509	M510	M511	M512	M513	M514	M515	M516	M517	M518	M519	M520	M521	M522	M523	M524	M525	M526	M527	M528	M529	M530	M531	M532	M533	M534	M535	M536	M537	M540	M541	M542	M545	M546	M547	M550	M551	M552																																																																				
	L350	L351	L352	L355	L356	L357	L358	L359	L360	L361	L362	L363	L364	L365	L366	L367	L368	L369	L370	L371	L372	L373	L374	L375	L376	L377	L378	L379	L380	L381	L382	L383	L384	L385	L386	L387	L388	L389	L390	L391	L392	L393	L394	L395	L396	L397	L398	L399	L400	L401	L402	L403	L404	L405	L406	L407	L408	L409	L410	L411	L412	L413	L414	L415	L416	L417	L418	L419	L420	L421	L422	L423	L424	L425	L426	L427	L428	L429	L430	L431	L432	L433	L434	L435	L436	L437	L438	L439	L440	L441	L442	L443	L444	L445	L446	L447	L448	L449	L450	L451	L452	L453	L454	L455	L456	L457	L458	L459	L460	L461	L462	L463	L464	L465	L466	L467	L468	L469	L470	L471	L472	L473	L474	L475	L476	L477	L478	L479	L480	L481	L482	L483	L484	L485	L486	L487	L488	L489	L490	L491	L492	L493	L494	L495	L496	L497	L498	L499	L500	L501	L502	L503	L504	L505	L506	L507	L508	L509	L510	L511	L512	L513	L514	L515	L516	L517	L518	L519	L520	L521	L522	L523	L524	L525	L526	L527	L528	L529	L530	L531	L532	L533	L534	L535	L536	L537	L540	L541	L542	L545	L546	L547	L550	L551	L552
	I278	I279	I280	I281	I282	I																																																																																																																																																																																													



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}	0.137 , 0.279 0.125 , 0.256	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)	0.25 , 133.6	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	0.97	EDS
Total number of atoms	33805	wwPDB-VP
Average B, all atoms (Å ²)	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.*

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	1.11	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

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%)	9	29
1	B	1026/1021 (100%)	917 (89%)	89 (9%)	20 (2%)	8	26
1	C	1026/1021 (100%)	918 (90%)	82 (8%)	26 (2%)	5	19
1	D	1026/1021 (100%)	921 (90%)	85 (8%)	20 (2%)	8	26
All	All	4104/4084 (100%)	3679 (90%)	342 (8%)	83 (2%)	7	24

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 [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	880/873 (101%)	700 (80%)	180 (20%)	1	3
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%)	1	4
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 [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%)	-1.15	3 (0%) 94 93	2, 26, 68, 100	19 (1%)
1	B	1021/1021 (100%)	-1.05	0 100 100	4, 31, 72, 100	19 (1%)
1	C	1021/1021 (100%)	-1.10	3 (0%) 94 93	6, 28, 65, 100	19 (1%)
1	D	1021/1021 (100%)	-1.14	1 (0%) 95 95	3, 26, 65, 100	19 (1%)
All	All	4084/4084 (100%)	-1.11	7 (0%) 95 94	2, 28, 68, 100	76 (1%)

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	733	ALA	3.0
1	A	734	SER	2.5
1	D	732	ALA	2.4
1	A	730	LEU	2.3
1	C	581	ASN	2.2

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	A	3001	1/1	0.97	0.13	28,28,28,28	0
2	MG	B	3002	1/1	0.97	0.08	36,36,36,36	0
2	MG	C	3002	1/1	0.98	0.07	31,31,31,31	0
2	MG	D	3001	1/1	0.98	0.07	40,40,40,40	0
2	MG	C	3001	1/1	0.99	0.04	15,15,15,15	0
2	MG	B	3001	1/1	0.99	0.05	20,20,20,20	0
2	MG	D	3002	1/1	0.99	0.11	25,25,25,25	0
2	MG	A	3002	1/1	0.99	0.15	31,31,31,31	0

6.5 Other polymers [i](#)

There are no such residues in this entry.