wwPDB X-ray Structure Validation Summary Report (i)

Jun 16, 2024 – 08:41 PM EDT

PDB ID	:	3MI6
Title	:	Crystal structure of the alpha-galactosidase from Lactobacillus brevis, Northeast Structural Genomics Consortium Target LbR11.
Authors	:	Vorobiev, S.; Chen, Y.; Seetharaman, J.; Belote, R.; Sahdev, S.; Xiao, R.; Acton, T.B.; Everett, J.K.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2010-04-09
Resolution	:	2.70 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 (i) 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 (i)) were used in the production of this report:

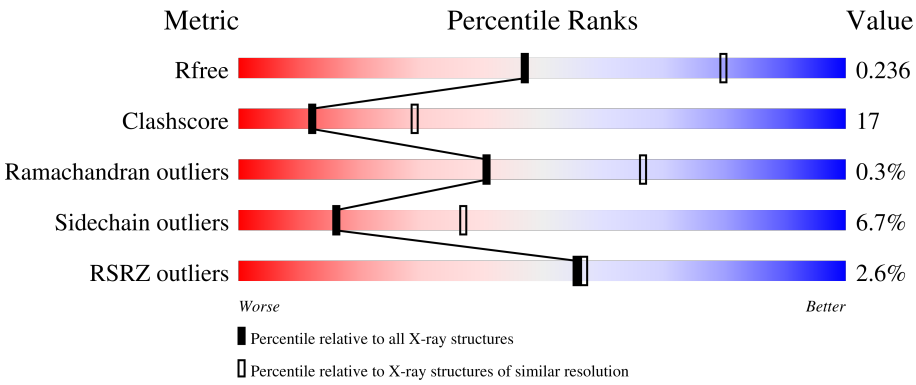
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 <=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	745	2% 65% 32% •
1	B	745	71% 25% • •
1	C	745	5% 64% 31% • •
1	D	745	2% 71% 25% •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 24417 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 Alpha-galactosidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	742	6007	3822	1022	1137	2	24	0	0	0
1	B	736	5947	3786	1008	1127	2	24	0	0	0
1	C	737	5946	3782	1008	1130	2	24	0	0	0
1	D	743	6018	3829	1027	1136	2	24	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	738	LEU	-	expression tag	UNP Q03PP7
A	739	GLU	-	expression tag	UNP Q03PP7
A	740	HIS	-	expression tag	UNP Q03PP7
A	741	HIS	-	expression tag	UNP Q03PP7
A	742	HIS	-	expression tag	UNP Q03PP7
A	743	HIS	-	expression tag	UNP Q03PP7
A	744	HIS	-	expression tag	UNP Q03PP7
A	745	HIS	-	expression tag	UNP Q03PP7
B	738	LEU	-	expression tag	UNP Q03PP7
B	739	GLU	-	expression tag	UNP Q03PP7
B	740	HIS	-	expression tag	UNP Q03PP7
B	741	HIS	-	expression tag	UNP Q03PP7
B	742	HIS	-	expression tag	UNP Q03PP7
B	743	HIS	-	expression tag	UNP Q03PP7
B	744	HIS	-	expression tag	UNP Q03PP7
B	745	HIS	-	expression tag	UNP Q03PP7
C	738	LEU	-	expression tag	UNP Q03PP7
C	739	GLU	-	expression tag	UNP Q03PP7
C	740	HIS	-	expression tag	UNP Q03PP7
C	741	HIS	-	expression tag	UNP Q03PP7
C	742	HIS	-	expression tag	UNP Q03PP7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	743	HIS	-	expression tag	UNP Q03PP7
C	744	HIS	-	expression tag	UNP Q03PP7
C	745	HIS	-	expression tag	UNP Q03PP7
D	738	LEU	-	expression tag	UNP Q03PP7
D	739	GLU	-	expression tag	UNP Q03PP7
D	740	HIS	-	expression tag	UNP Q03PP7
D	741	HIS	-	expression tag	UNP Q03PP7
D	742	HIS	-	expression tag	UNP Q03PP7
D	743	HIS	-	expression tag	UNP Q03PP7
D	744	HIS	-	expression tag	UNP Q03PP7
D	745	HIS	-	expression tag	UNP Q03PP7

- Molecule 2 is water.

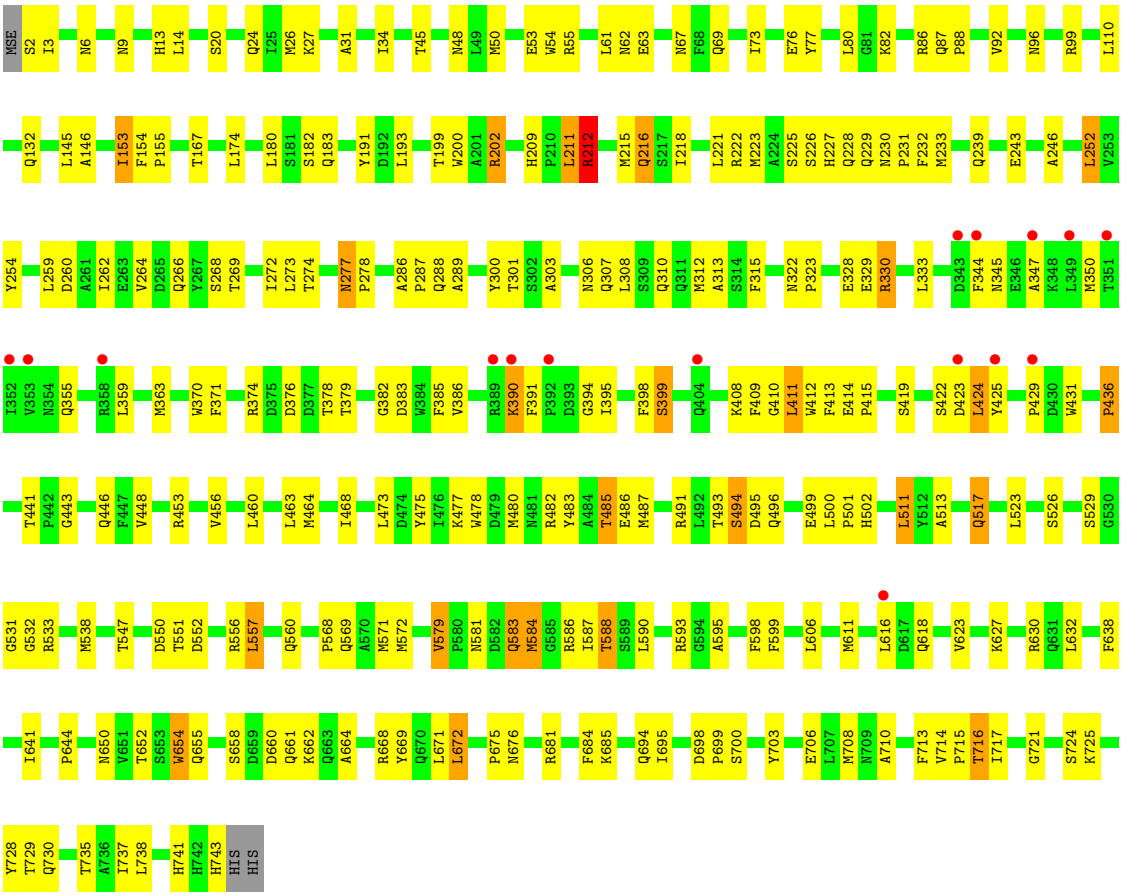
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	117	117	117	0	0
			Total	O		
2	B	179	179	179	0	0
			Total	O		
2	C	84	84	84	0	0
			Total	O		
2	D	119	119	119	0	0

3 Residue-property plots (i)

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.

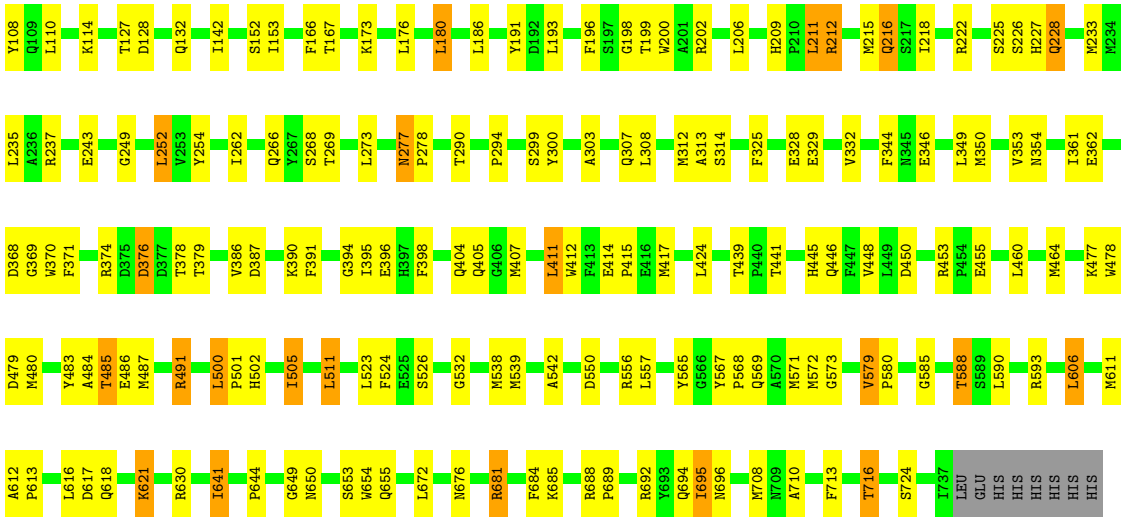
• Molecule 1: Alpha-galactosidase

Chain A:



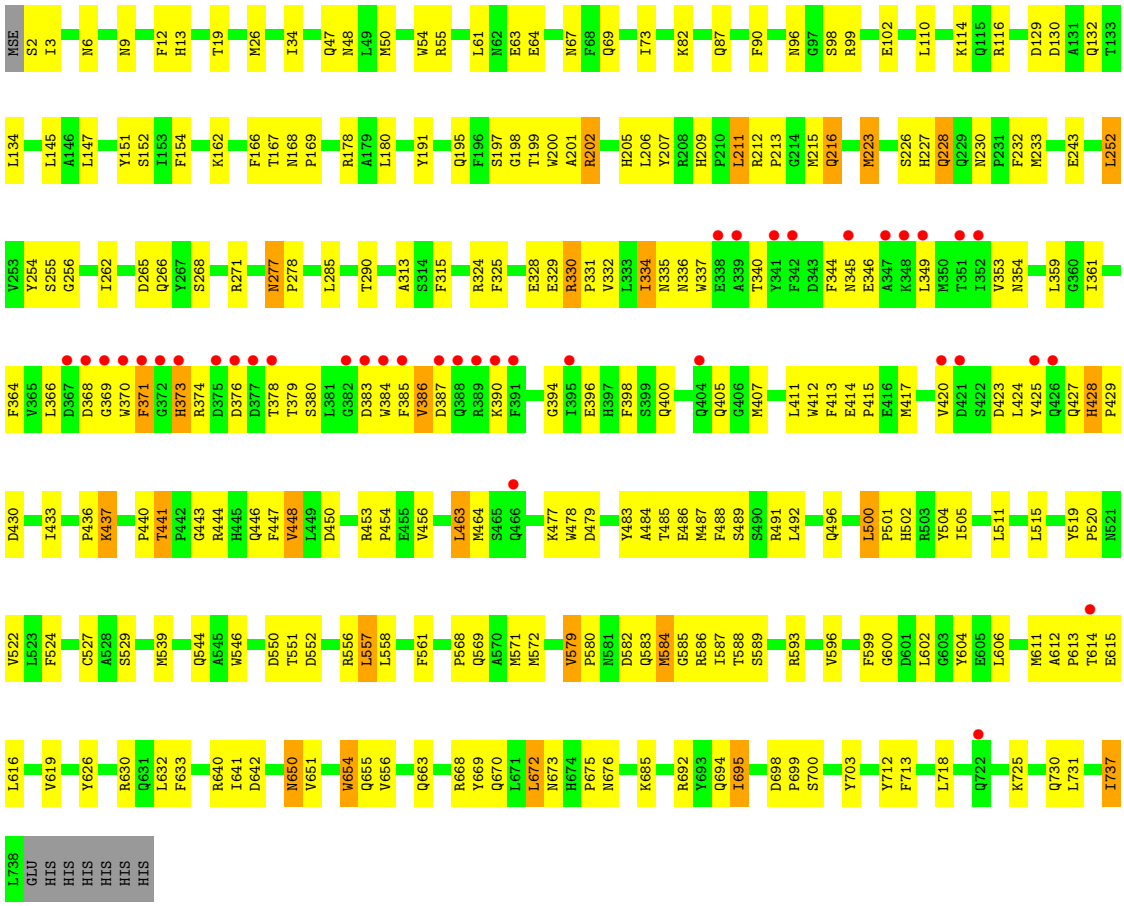
• Molecule 1: Alpha-galactosidase

Chain B:



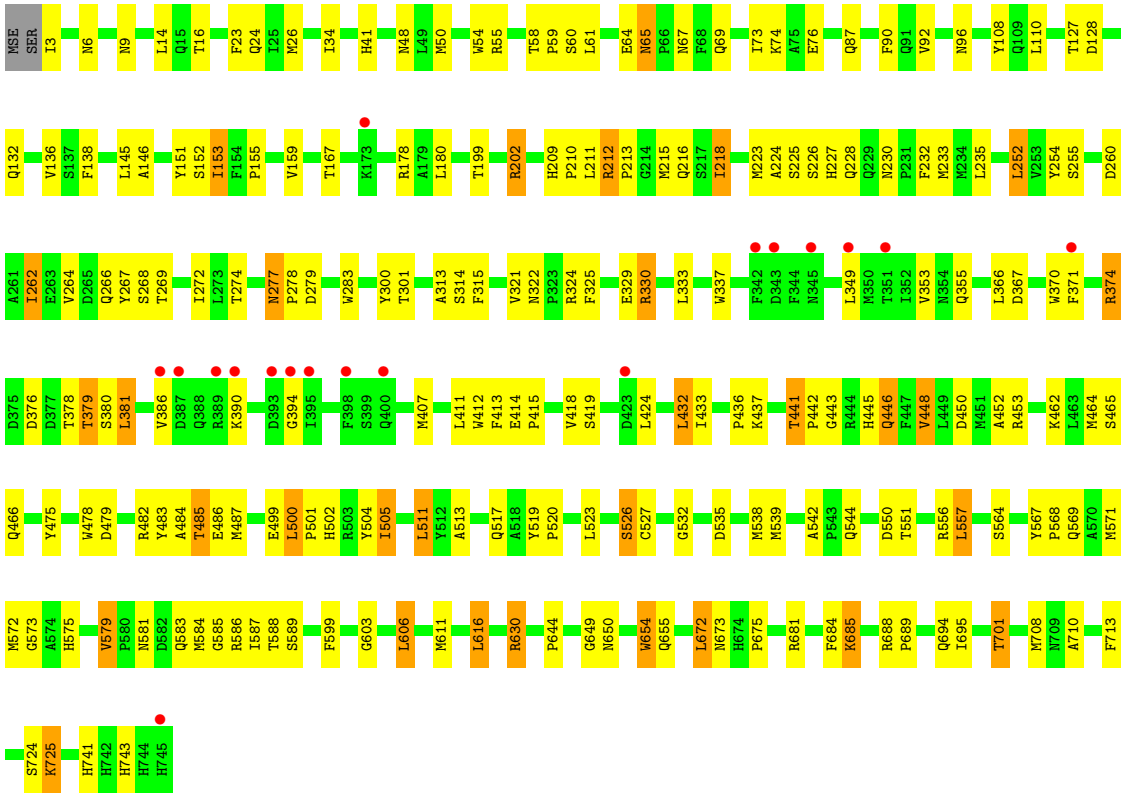
- Molecule 1: Alpha-galactosidase

Chain C:



- Molecule 1: Alpha-galactosidase

Chain D:


MSE SER I3 N6 N9 L14 Q15 T16 F23 Q24 I25 M26 I34 H41 N48 L49 M50 W54 R55 T58 P59 S60 L61 E64 M65 P66 N67 F68 Q69 I73 K74 A75 E76 Q87 F90 Q91 V92 N96 Y108 Q109 L110 T127 D128
Q132 V136 S137 F138 L145 A146 Y151 S152 I153 F154 P155 V159 T167 K173 R178 A179 L180 T199 R202 H209 P210 L211 R212 P213 G214 M215 Q216 S217 I218 M223 A224 S225 S226 H227 Q228 Q229 N230 P231 F232 M233 M234 L235 L252 V253 Y254 S255 D260
A261 I262 E263 V264 D265 Q266 Y267 S268 T269 I272 L273 T274 N277 P278 D279 W283 Y300 T301 A313 S314 F315 V321 N322 P323 R324 F325 E329 R330 L333 W337 F342 D343 F344 N345 L349 M350 T351 I352 V353 M354 Q355 L366 D367 W370 F371 R374
D375 D376 D377 T378 T379 S380 L381 V386 D387 Q388 R389 K390 D393 G394 I395 F398 S399 Q400 M407 L411 W412 F413 E414 P415 V418 S419 D423 L424 L432 I433 P436 K437 T441 P442 G443 R444 H445 Q446 F447 V448 L449 D450 M451 A452 R453 K462 L463 M464 S465
Q466 Y475 W478 D479 R482 Y483 A484 T485 E486 M487 E499 L500 P501 H502 R503 Y504 I505 L511 Y512 A513 Q517 A518 Y519 P520 L523 S526 C527 G532 D535 M538 M539 A542 P543 Q544 D550 T551 R556 L557 S564 Y567 P568 Q569 A570 M571
M572 G573 A574 H575 V579 P580 N581 D582 Q583 M584 G585 R586 I587 T588 S589 F599 G603 L606 M611 L616 R630 P644 G649 N650 W654 Q655 L672 N673 H674 P675 R681 F684 K685 R688 P689 Q694 I695 T701 M708 N709 A710 F713
S724 K725 H741 H742 H743 H744 H745

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	115.94Å 158.82Å 166.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.03 – 2.70 45.53 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.7 (40.03-2.70) 99.7 (45.53-2.70)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.11 (at 2.69Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.5_2), CNS	Depositor
R, R_{free}	0.173 , 0.241 0.173 , 0.236	Depositor DCC
R_{free} test set	4195 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	41.8	Xtriage
Anisotropy	0.370	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 31.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.011 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	24417	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.85% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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)

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	0.46	0/6150	0.62	2/8319 (0.0%)
1	B	0.49	0/6086	0.65	2/8232 (0.0%)
1	C	0.44	0/6083	0.60	2/8227 (0.0%)
1	D	0.47	0/6163	0.63	2/8336 (0.0%)
All	All	0.46	0/24482	0.62	8/33114 (0.0%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	695	ILE	CB-CA-C	-5.93	99.75	111.60
1	C	371	PHE	CB-CA-C	-5.66	99.09	110.40
1	A	695	ILE	CB-CA-C	-5.53	100.54	111.60
1	B	237	ARG	NE-CZ-NH1	-5.43	117.58	120.30
1	B	695	ILE	CB-CA-C	-5.42	100.75	111.60

There are no chirality outliers.

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	6007	0	5701	230	0
1	B	5947	0	5654	192	0
1	C	5946	0	5650	237	0
1	D	6018	0	5706	187	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	117	0	0	9	0
2	B	179	0	0	13	0
2	C	84	0	0	4	0
2	D	119	0	0	10	0
All	All	24417	0	22711	799	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 17.

The worst 5 of 799 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:233:MSE:HE1	1:A:262:ILE:HD11	1.27	1.11
1:B:299:SER:HB3	1:B:312:MSE:HE2	1.10	1.05
1:C:584:MSE:HE1	1:C:586:ARG:HB2	1.41	1.00
1:D:226:SER:HB2	1:D:228:GLN:NE2	1.80	0.97
1:C:167:THR:HG23	1:C:290:THR:HG22	1.47	0.96

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	740/745 (99%)	698 (94%)	38 (5%)	4 (0%)	29	54
1	B	734/745 (98%)	697 (95%)	36 (5%)	1 (0%)	51	78
1	C	735/745 (99%)	677 (92%)	55 (8%)	3 (0%)	34	60
1	D	741/745 (100%)	691 (93%)	48 (6%)	2 (0%)	41	66
All	All	2950/2980 (99%)	2763 (94%)	177 (6%)	10 (0%)	41	66

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	386	VAL
1	D	374	ARG
1	A	77	TYR
1	A	480	MSE
1	C	337	TRP

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	645/624 (103%)	605 (94%)	40 (6%)	18	40
1	B	638/624 (102%)	597 (94%)	41 (6%)	17	39
1	C	638/624 (102%)	592 (93%)	46 (7%)	14	34
1	D	645/624 (103%)	600 (93%)	45 (7%)	15	35
All	All	2566/2496 (103%)	2394 (93%)	172 (7%)	16	37

5 of 172 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	557	LEU
1	D	314	SER
1	C	614	THR
1	D	110	LEU
1	D	419	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 117 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	13	HIS
1	D	481	ASN
1	C	307	GLN
1	D	446	GLN
1	D	277	ASN

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)

There are no ligands in this entry.

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	718/745 (96%)	-0.31	16 (2%)	62	63	24, 38, 84, 103	0
1	B	712/745 (95%)	-0.47	1 (0%)	95	96	23, 35, 49, 80	0
1	C	713/745 (95%)	-0.03	39 (5%)	25	24	25, 44, 97, 104	0
1	D	719/745 (96%)	-0.28	18 (2%)	57	59	25, 39, 67, 94	0
All	All	2862/2980 (96%)	-0.27	74 (2%)	56	57	23, 38, 81, 104	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	389	ARG	7.3
1	C	349	LEU	5.5
1	C	391	PHE	5.4
1	C	342	PHE	5.3
1	C	376	ASP	5.1

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.