wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 08:18 pm BST

PDB ID	:	5MKB
Title	:	Maltodextrin binding protein MalE1 from L. casei BL23 without ligand
Authors	:	Homburg, C.; Bommer, M.; Wuttge, S.; Hobe, C.; Beck, S.; Dobbek, H.; Deutscher, J.; Licht, A.; Schneider, E.
Deposited on	:	2016-12-03
Resolution	:	1.70 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 following versions of software and data (see references (1)) were used in the production of this report:

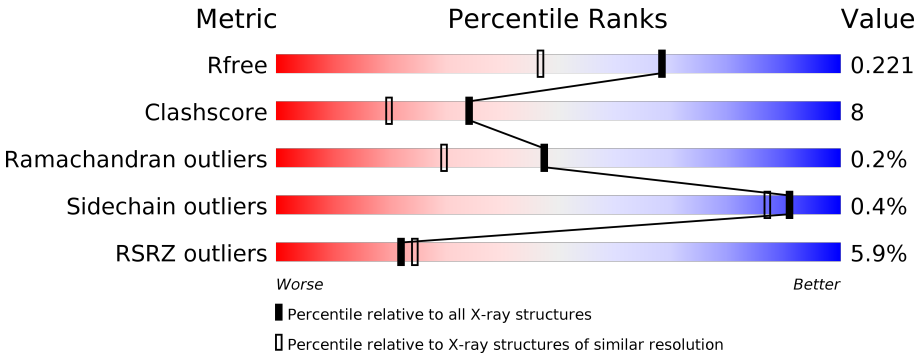
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.70 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.



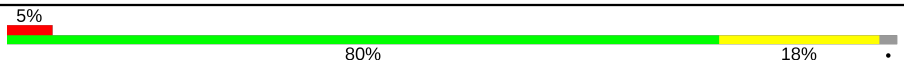
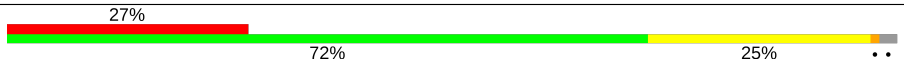
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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 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 <=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	379	3% 82% 16%
1	B	379	% 88% 10%
1	C	379	6% 84% 15%
1	D	379	2% 85% 13%
1	E	379	% 88% 11%
1	F	379	3% 81% 17%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	379	
1	H	379	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 24739 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 MalE1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	373	Total	C	N	O	S	0	0	0
			2803	1779	469	545	10			
1	B	373	Total	C	N	O	S	0	0	0
			2803	1779	469	545	10			
1	C	373	Total	C	N	O	S	0	0	0
			2803	1779	469	545	10			
1	D	373	Total	C	N	O	S	0	0	0
			2803	1779	469	545	10			
1	E	373	Total	C	N	O	S	0	0	0
			2803	1779	469	545	10			
1	F	373	Total	C	N	O	S	0	0	0
			2803	1779	469	545	10			
1	G	373	Total	C	N	O	S	0	0	0
			2803	1779	469	545	10			
1	H	373	Total	C	N	O	S	0	0	0
			2803	1779	469	545	10			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	10	GLY	-	expression tag	UNP B0L7B0
A	11	SER	-	expression tag	UNP B0L7B0
A	12	HIS	-	expression tag	UNP B0L7B0
A	13	MET	-	expression tag	UNP B0L7B0
B	10	GLY	-	expression tag	UNP B0L7B0
B	11	SER	-	expression tag	UNP B0L7B0
B	12	HIS	-	expression tag	UNP B0L7B0
B	13	MET	-	expression tag	UNP B0L7B0
C	10	GLY	-	expression tag	UNP B0L7B0
C	11	SER	-	expression tag	UNP B0L7B0
C	12	HIS	-	expression tag	UNP B0L7B0
C	13	MET	-	expression tag	UNP B0L7B0
D	10	GLY	-	expression tag	UNP B0L7B0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	11	SER	-	expression tag	UNP B0L7B0
D	12	HIS	-	expression tag	UNP B0L7B0
D	13	MET	-	expression tag	UNP B0L7B0
E	10	GLY	-	expression tag	UNP B0L7B0
E	11	SER	-	expression tag	UNP B0L7B0
E	12	HIS	-	expression tag	UNP B0L7B0
E	13	MET	-	expression tag	UNP B0L7B0
F	10	GLY	-	expression tag	UNP B0L7B0
F	11	SER	-	expression tag	UNP B0L7B0
F	12	HIS	-	expression tag	UNP B0L7B0
F	13	MET	-	expression tag	UNP B0L7B0
G	10	GLY	-	expression tag	UNP B0L7B0
G	11	SER	-	expression tag	UNP B0L7B0
G	12	HIS	-	expression tag	UNP B0L7B0
G	13	MET	-	expression tag	UNP B0L7B0
H	10	GLY	-	expression tag	UNP B0L7B0
H	11	SER	-	expression tag	UNP B0L7B0
H	12	HIS	-	expression tag	UNP B0L7B0
H	13	MET	-	expression tag	UNP B0L7B0

- Molecule 2 is water.

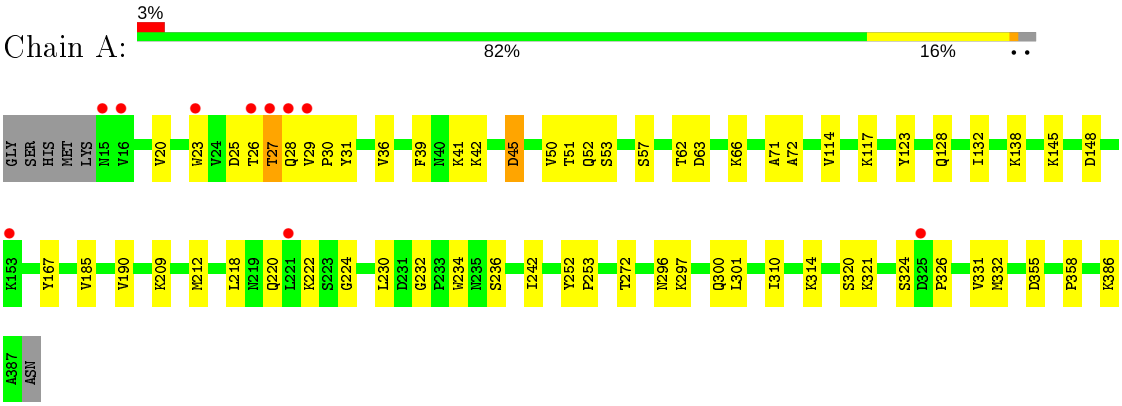
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	336	Total O 336 336	0	0
2	B	312	Total O 312 312	0	0
2	C	302	Total O 302 302	0	0
2	D	290	Total O 290 290	0	0
2	E	338	Total O 338 338	0	0
2	F	310	Total O 310 310	0	0
2	G	223	Total O 223 223	0	0
2	H	204	Total O 204 204	0	0

3 Residue-property plots (i)

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.

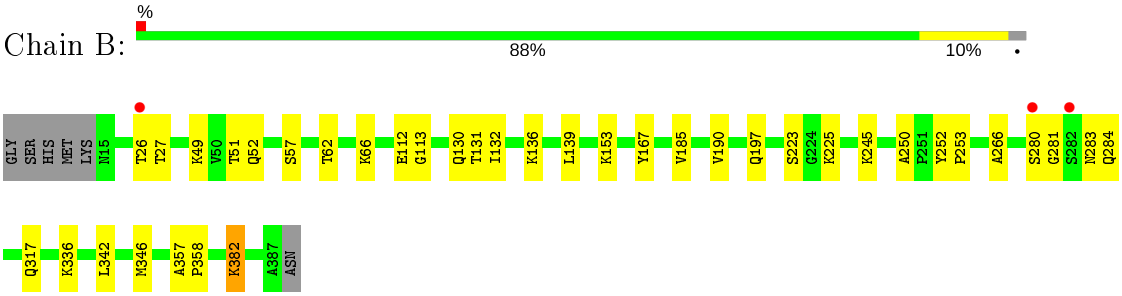
- Molecule 1: MalE1

Chain A:



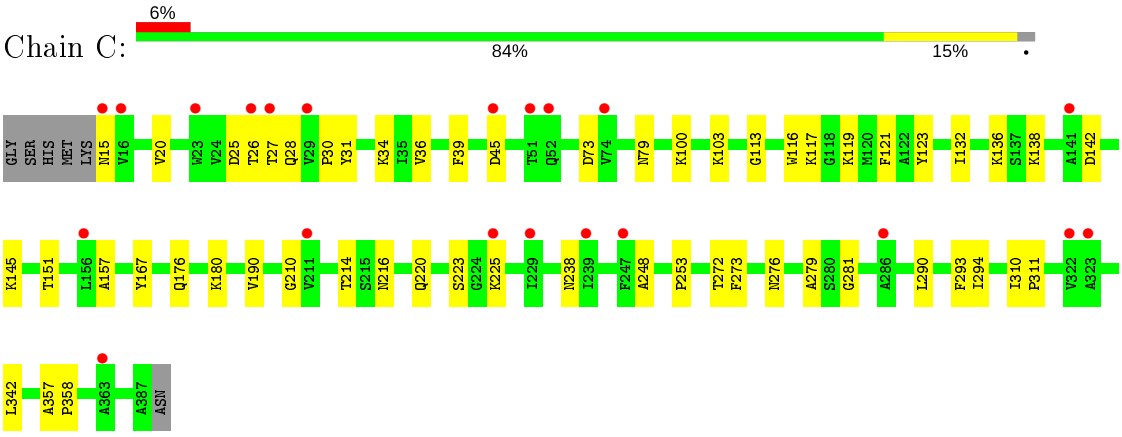
- Molecule 1: MalE1

Chain B:



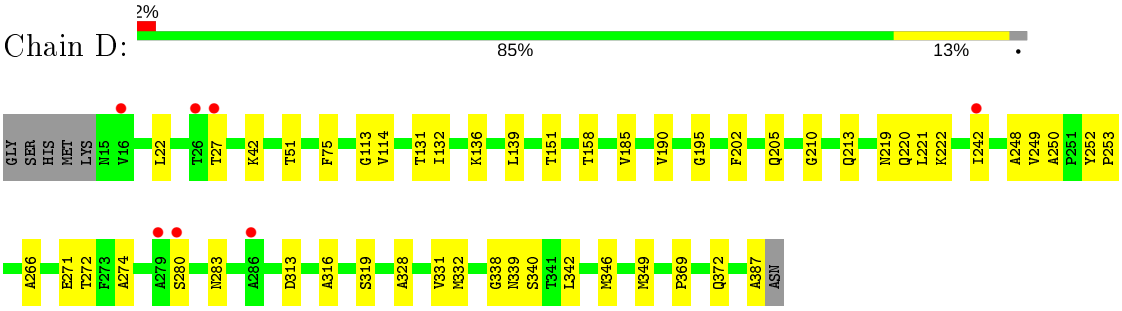
- Molecule 1: MalE1

Chain C:



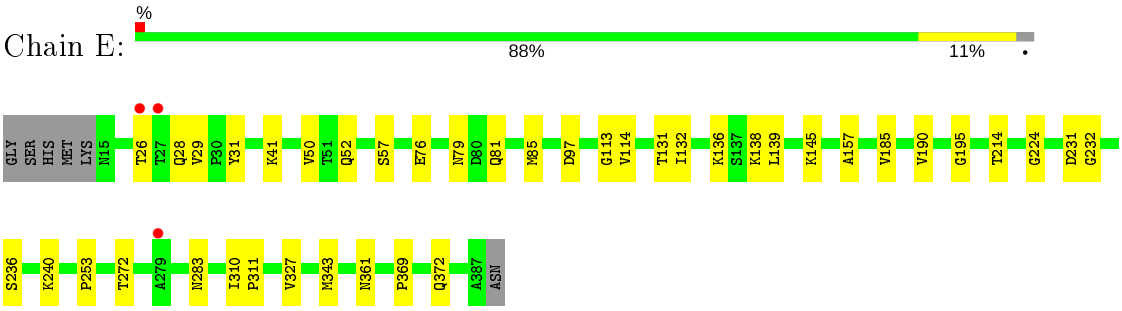
- Molecule 1: MalE1

Chain D:



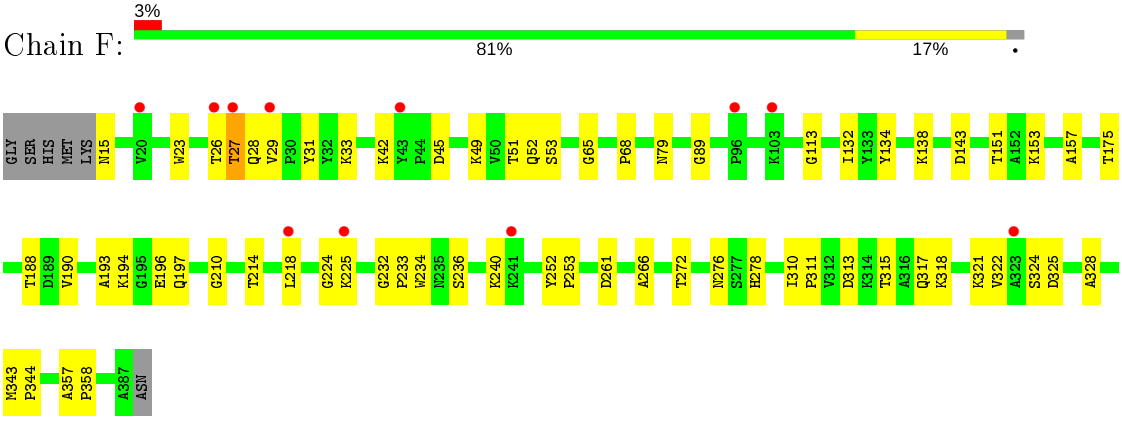
- Molecule 1: MalE1

Chain E:



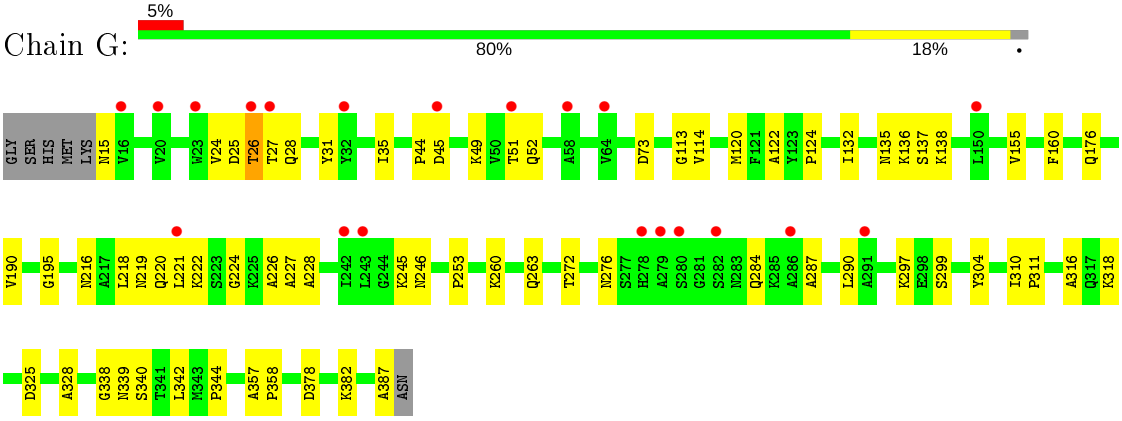
- Molecule 1: MalE1

Chain F:



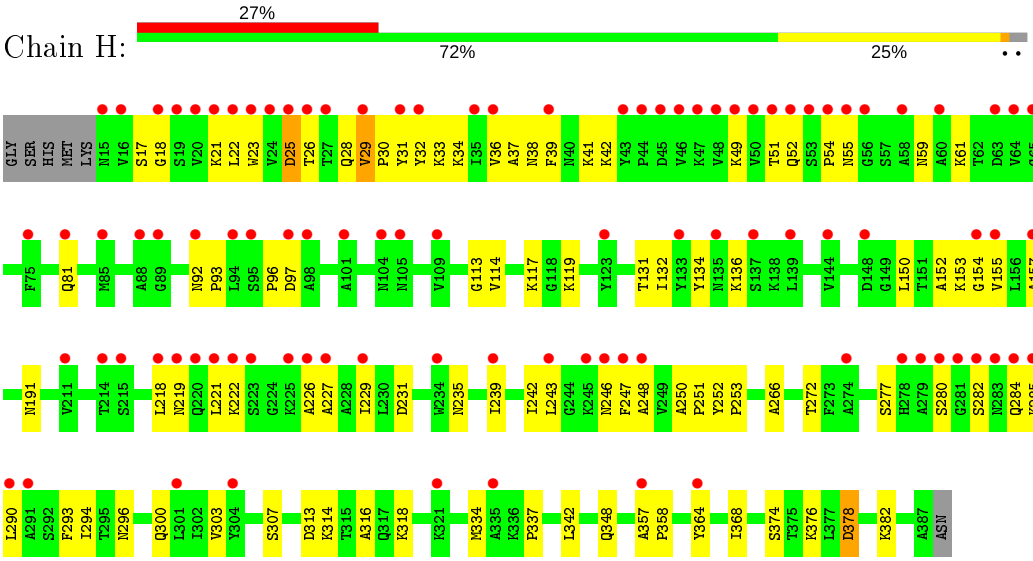
- Molecule 1: MalE1

Chain G:



- Molecule 1: MalE1

Chain H:


27%
72%
25%
GLY SER HIS MET LYS N15 V16 S17 G18 S19 V20 K21 L22 W23 V24 D25 T26 T27 Q28 V29 P30 Y31 Y32 K33 K34 I35 V36 A37 N38 F39 N40 K41 K42 Y43 P44 D45 V46 K47 V48 K49 V50 T51 Q52 S53 P54 N55 G56 S57 A58 N59 A60 K61 T62 D63 V64 G65 K70 A71 A72
F75 Q81 M85 A88 G89 N92 P93 L94 S95 P96 D97 A98 A101 N104 N105 V109 G113 V114 K117 G118 K119 Y123 T131 I132 Y133 Y134 N135 K136 S137 K138 L139 V144 D148 G149 L150 T151 A152 K153 G154 V155 L156 A157 T161 N165
N191 V211 T214 S215 L218 N219 Q220 L221 K222 S223 G224 K225 A226 A227 A228 I229 L230 D231 W234 N235 I239 I242 L243 G244 K245 N246 F247 A248 V249 A250 P251 Y252 P253 A266 T272 F273 A274 S277 H278 A279 S280 G281 S282 N283 Q284 K285 A286 A287 A288 T289
L290 A291 S292 F293 I294 T295 N296 Q300 L301 I302 V303 Y304 S307 D313 K314 T315 A316 Q317 K318 K321 M334 A335 K336 P337 L342 Q348 A357 P358 Y364 I368 S374 T375 K376 L377 D378 K382 A387 ASN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	41.04Å 124.94Å 143.74Å 74.51° 89.61° 89.86°	Depositor
Resolution (Å)	46.17 – 1.70 46.17 – 1.70	Depositor EDS
% Data completeness (in resolution range)	96.9 (46.17-1.70) 95.2 (46.17-1.70)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.81 (at 1.70Å)	Xtriage
Refinement program	PHENIX 1.9_1690	Depositor
R, R_{free}	0.178 , 0.218 0.181 , 0.221	Depositor DCC
R_{free} test set	14709 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	18.2	Xtriage
Anisotropy	0.545	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 33.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	0.437 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	24739	wwPDB-VP
Average B, all atoms (Å^2)	24.0	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 44.47 % 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 1.5194e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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

5.1 Standard geometry

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	0.25	0/2862	0.43	0/3885
1	B	0.24	0/2862	0.43	0/3885
1	C	0.24	0/2862	0.42	0/3885
1	D	0.24	0/2862	0.43	0/3885
1	E	0.24	0/2862	0.43	0/3885
1	F	0.25	0/2862	0.44	0/3885
1	G	0.24	0/2862	0.43	0/3885
1	H	0.27	0/2862	0.48	0/3885
All	All	0.24	0/22896	0.44	0/31080

There are no bond length outliers.

There are no bond angle outliers.

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	2803	0	2800	51	0
1	B	2803	0	2800	24	0
1	C	2803	0	2800	39	0
1	D	2803	0	2800	49	0
1	E	2803	0	2800	27	0
1	F	2803	0	2800	50	0
1	G	2803	0	2800	57	0
1	H	2803	0	2800	81	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	336	0	0	23	1
2	B	312	0	0	9	0
2	C	302	0	0	19	1
2	D	290	0	0	31	0
2	E	338	0	0	10	1
2	F	310	0	0	25	0
2	G	223	0	0	30	0
2	H	204	0	0	31	0
All	All	24739	0	22400	375	2

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 8.

The worst 5 of 375 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:G:113:GLY:HA2	1:G:342:LEU:HD21	1.54	0.90
1:H:36:VAL:HA	1:H:39:PHE:HB2	1.56	0.85
1:E:132:ILE:HD11	1:E:253:PRO:HD3	1.60	0.84
1:B:113:GLY:HA2	1:B:342:LEU:HD21	1.58	0.83
1:D:132:ILE:HD11	1:D:253:PRO:HD3	1.59	0.83

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:403:HOH:O	2:A:530:HOH:O[1_455]	2.16	0.04
2:C:691:HOH:O	2:E:622:HOH:O[1_655]	2.16	0.04

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	371/379 (98%)	363 (98%)	8 (2%)	0	100	100
1	B	371/379 (98%)	363 (98%)	8 (2%)	0	100	100
1	C	371/379 (98%)	363 (98%)	7 (2%)	1 (0%)	41	24
1	D	371/379 (98%)	365 (98%)	6 (2%)	0	100	100
1	E	371/379 (98%)	363 (98%)	8 (2%)	0	100	100
1	F	371/379 (98%)	365 (98%)	5 (1%)	1 (0%)	41	24
1	G	371/379 (98%)	360 (97%)	10 (3%)	1 (0%)	41	24
1	H	371/379 (98%)	357 (96%)	12 (3%)	2 (0%)	29	13
All	All	2968/3032 (98%)	2899 (98%)	64 (2%)	5 (0%)	47	30

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	27	THR
1	G	27	THR
1	C	25	ASP
1	H	25	ASP
1	H	96	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	297/302 (98%)	293 (99%)	4 (1%)	69	56
1	B	297/302 (98%)	296 (100%)	1 (0%)	92	89
1	C	297/302 (98%)	296 (100%)	1 (0%)	92	89
1	D	297/302 (98%)	297 (100%)	0	100	100
1	E	297/302 (98%)	297 (100%)	0	100	100
1	F	297/302 (98%)	296 (100%)	1 (0%)	92	89
1	G	297/302 (98%)	296 (100%)	1 (0%)	92	89
1	H	297/302 (98%)	295 (99%)	2 (1%)	84	77

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2376/2416 (98%)	2366 (100%)	10 (0%)	91 87

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	382	LYS
1	C	45	ASP
1	G	26	THR
1	A	148	ASP
1	F	321	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	219	ASN
1	E	128	GLN
1	G	219	ASN
1	D	361	ASN
1	E	52	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 carbohydrates 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

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	373/379 (98%)	0.24	10 (2%) 54 58	11, 19, 35, 93	0
1	B	373/379 (98%)	0.10	3 (0%) 86 88	10, 18, 33, 56	0
1	C	373/379 (98%)	0.49	21 (5%) 24 27	13, 24, 41, 84	0
1	D	373/379 (98%)	0.27	7 (1%) 66 70	11, 18, 29, 87	0
1	E	373/379 (98%)	0.15	3 (0%) 86 88	11, 19, 33, 64	0
1	F	373/379 (98%)	0.22	11 (2%) 51 56	12, 20, 33, 100	0
1	G	373/379 (98%)	0.62	20 (5%) 25 28	12, 27, 50, 75	0
1	H	373/379 (98%)	1.47	102 (27%) 0 0	16, 35, 79, 122	0
All	All	2984/3032 (98%)	0.45	177 (5%) 22 24	10, 21, 47, 122	0

The worst 5 of 177 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	43	TYR	13.6
1	H	15	ASN	12.5
1	H	39	PHE	12.1
1	H	32	TYR	11.4
1	A	27	THR	9.9

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.