wwPDB X-ray Structure Validation Summary Report (i)

Aug 27, 2023 – 02:01 AM EDT

PDB ID	:	3FH6
Title	:	Crystal structure of the resting state maltose transporter from E. coli
Authors	:	Khare, D.; Oldham, M.L.; Orelle, C.; Davidson, A.L.; Chen, J.
Deposited on	:	2008-12-08
Resolution	:	4.50 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:

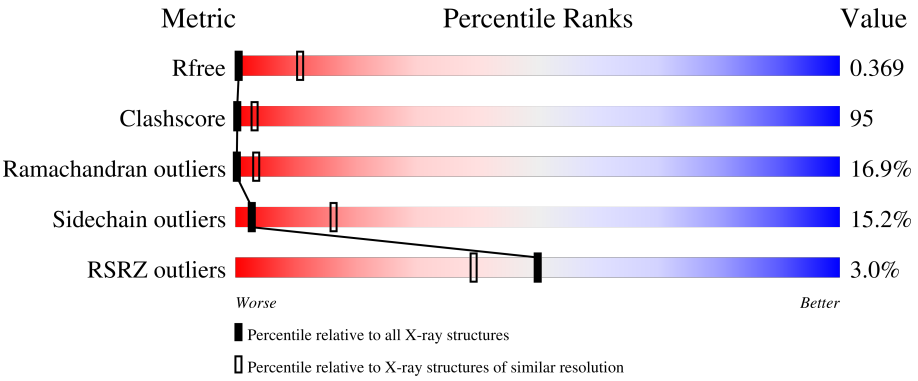
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 4.50 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.



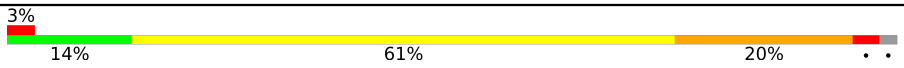
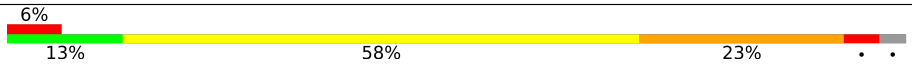
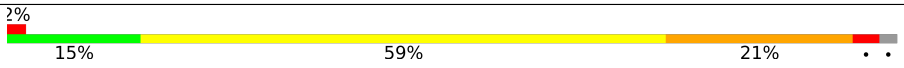
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1055 (5.20-3.80)
Clashscore	141614	1123 (5.20-3.80)
Ramachandran outliers	138981	1069 (5.20-3.80)
Sidechain outliers	138945	1050 (5.20-3.80)
RSRZ outliers	127900	1101 (5.30-3.70)

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	F	480	9% 41% 14% 34%
1	H	480	4% 9% 42% 14% 34%
2	G	296	2% 17% 52% 14% 14%
2	I	296	2% 17% 52% 14% 14%
3	A	381	2% 12% 60% 22%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	B	381	
3	C	381	
3	D	381	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 20236 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 Maltose transport system permease protein malF.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	F	316	Total	C	N	O	S	0	0	0
			2418	1607	378	418	15			
1	H	316	Total	C	N	O	S	0	0	0
			2418	1607	378	418	15			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	35	MET	-	expression tag	UNP P02916
H	35	MET	-	expression tag	UNP P02916

- Molecule 2 is a protein called Maltose transport system permease protein malG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	254	Total	C	N	O	S	0	0	0
			1942	1308	306	319	9			
2	I	254	Total	C	N	O	S	0	0	0
			1942	1308	306	319	9			

- Molecule 3 is a protein called Maltose/maltodextrin import ATP-binding protein malK.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	371	Total	C	N	O	S	0	0	0
			2876	1819	515	529	13			
3	B	372	Total	C	N	O	S	0	0	0
			2882	1822	516	531	13			
3	C	371	Total	C	N	O	S	0	0	0
			2876	1819	515	529	13			
3	D	372	Total	C	N	O	S	0	0	0
			2882	1822	516	531	13			

There are 40 discrepancies between the modelled and reference sequences:

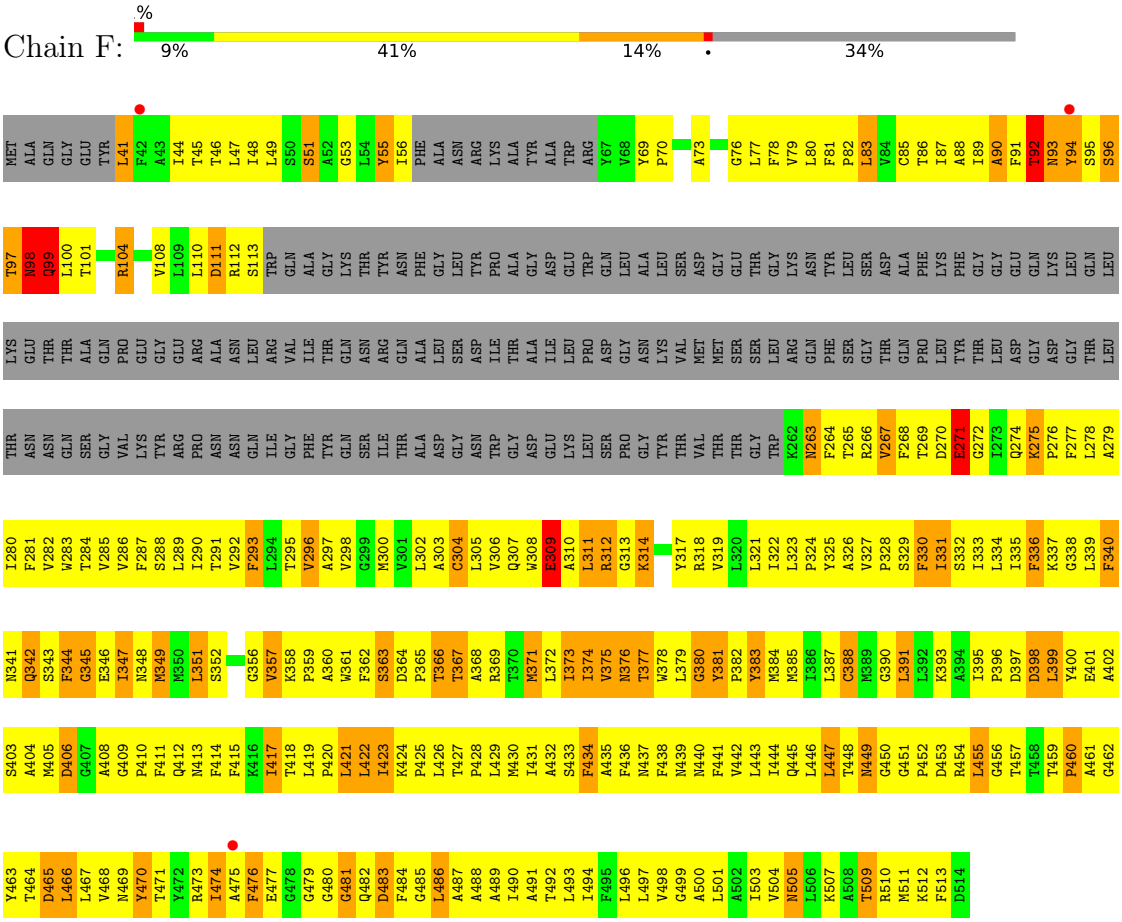
Chain	Residue	Modelled	Actual	Comment	Reference
A	372	ALA	-	expression tag	UNP P68187
A	373	SER	-	expression tag	UNP P68187
A	374	ALA	-	expression tag	UNP P68187
A	375	SER	-	expression tag	UNP P68187
A	376	HIS	-	expression tag	UNP P68187
A	377	HIS	-	expression tag	UNP P68187
A	378	HIS	-	expression tag	UNP P68187
A	379	HIS	-	expression tag	UNP P68187
A	380	HIS	-	expression tag	UNP P68187
A	381	HIS	-	expression tag	UNP P68187
B	372	ALA	-	expression tag	UNP P68187
B	373	SER	-	expression tag	UNP P68187
B	374	ALA	-	expression tag	UNP P68187
B	375	SER	-	expression tag	UNP P68187
B	376	HIS	-	expression tag	UNP P68187
B	377	HIS	-	expression tag	UNP P68187
B	378	HIS	-	expression tag	UNP P68187
B	379	HIS	-	expression tag	UNP P68187
B	380	HIS	-	expression tag	UNP P68187
B	381	HIS	-	expression tag	UNP P68187
C	372	ALA	-	expression tag	UNP P68187
C	373	SER	-	expression tag	UNP P68187
C	374	ALA	-	expression tag	UNP P68187
C	375	SER	-	expression tag	UNP P68187
C	376	HIS	-	expression tag	UNP P68187
C	377	HIS	-	expression tag	UNP P68187
C	378	HIS	-	expression tag	UNP P68187
C	379	HIS	-	expression tag	UNP P68187
C	380	HIS	-	expression tag	UNP P68187
C	381	HIS	-	expression tag	UNP P68187
D	372	ALA	-	expression tag	UNP P68187
D	373	SER	-	expression tag	UNP P68187
D	374	ALA	-	expression tag	UNP P68187
D	375	SER	-	expression tag	UNP P68187
D	376	HIS	-	expression tag	UNP P68187
D	377	HIS	-	expression tag	UNP P68187
D	378	HIS	-	expression tag	UNP P68187
D	379	HIS	-	expression tag	UNP P68187
D	380	HIS	-	expression tag	UNP P68187
D	381	HIS	-	expression tag	UNP P68187

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.

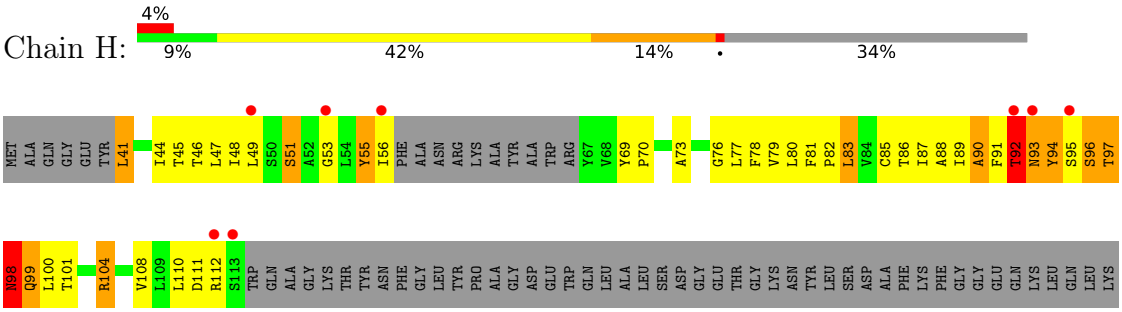
• Molecule 1: Maltose transport system permease protein malF

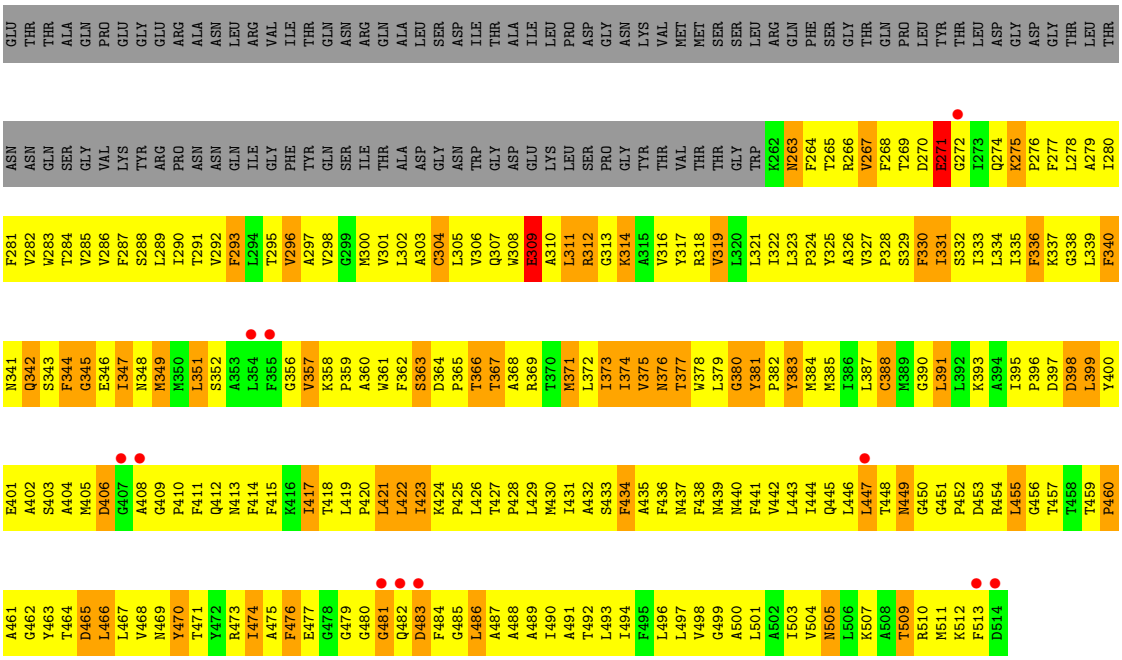
Chain F:



• Molecule 1: Maltose transport system permease protein malF

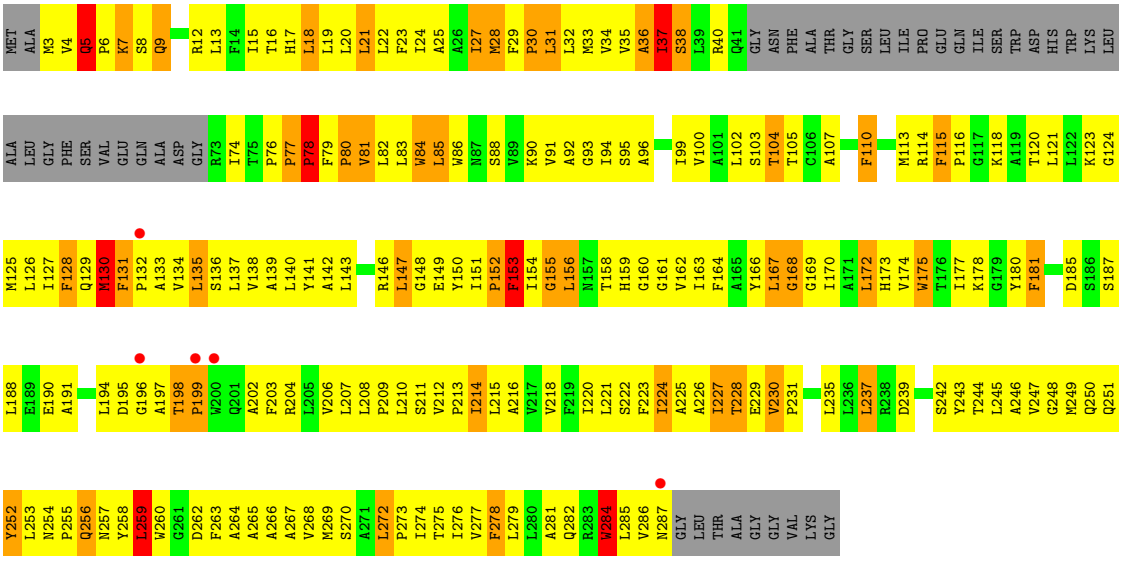
Chain H:





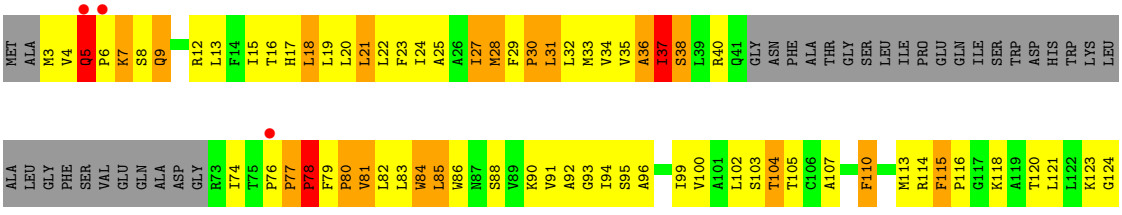
• Molecule 2: Maltose transport system permease protein malG

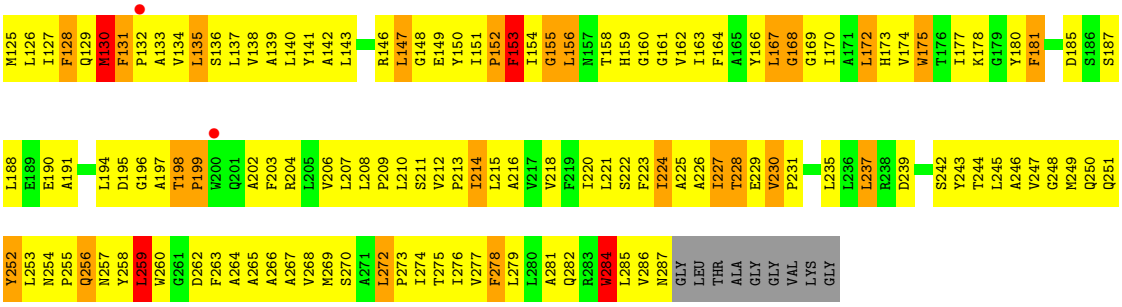
Chain G:

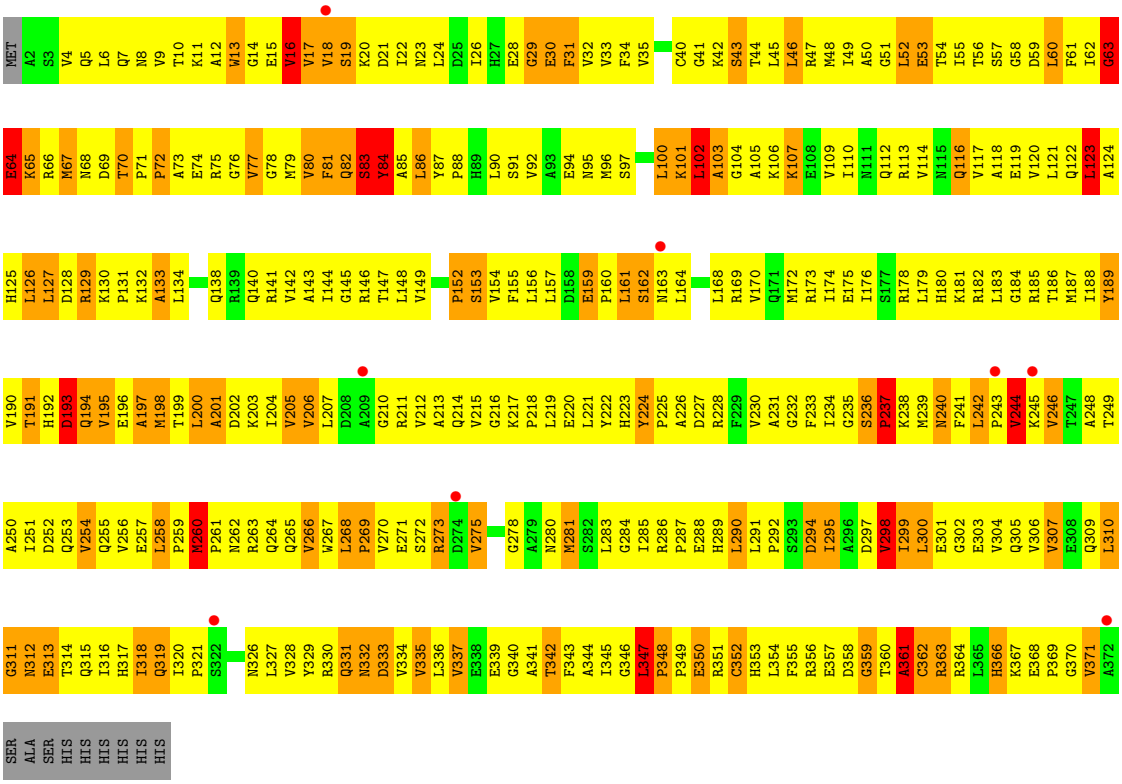


• Molecule 2: Maltose transport system permease protein malG

Chain I:





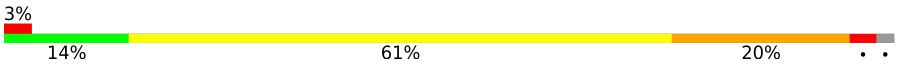


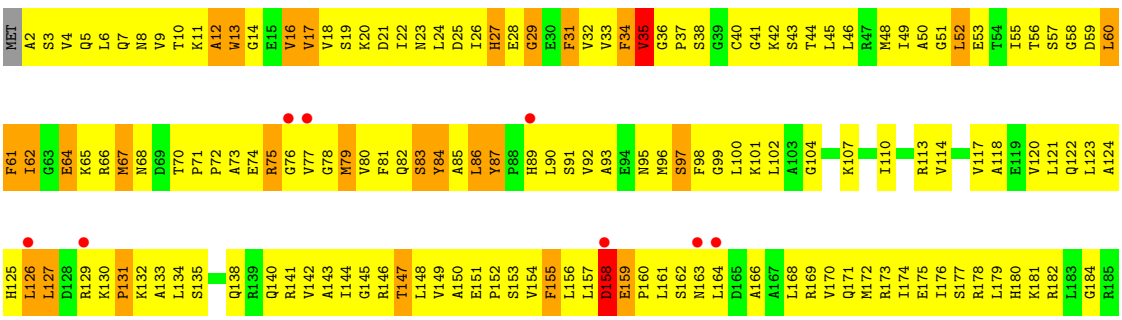
• Molecule 3: Maltose/maltodextrin import ATP-binding protein malK

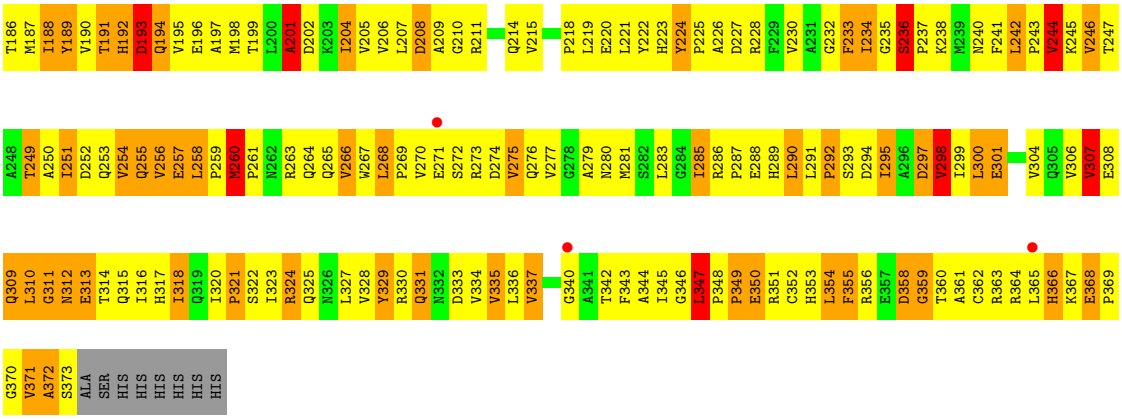
Chain A:

• Molecule 3: Maltose/maltodextrin import ATP-binding protein malK

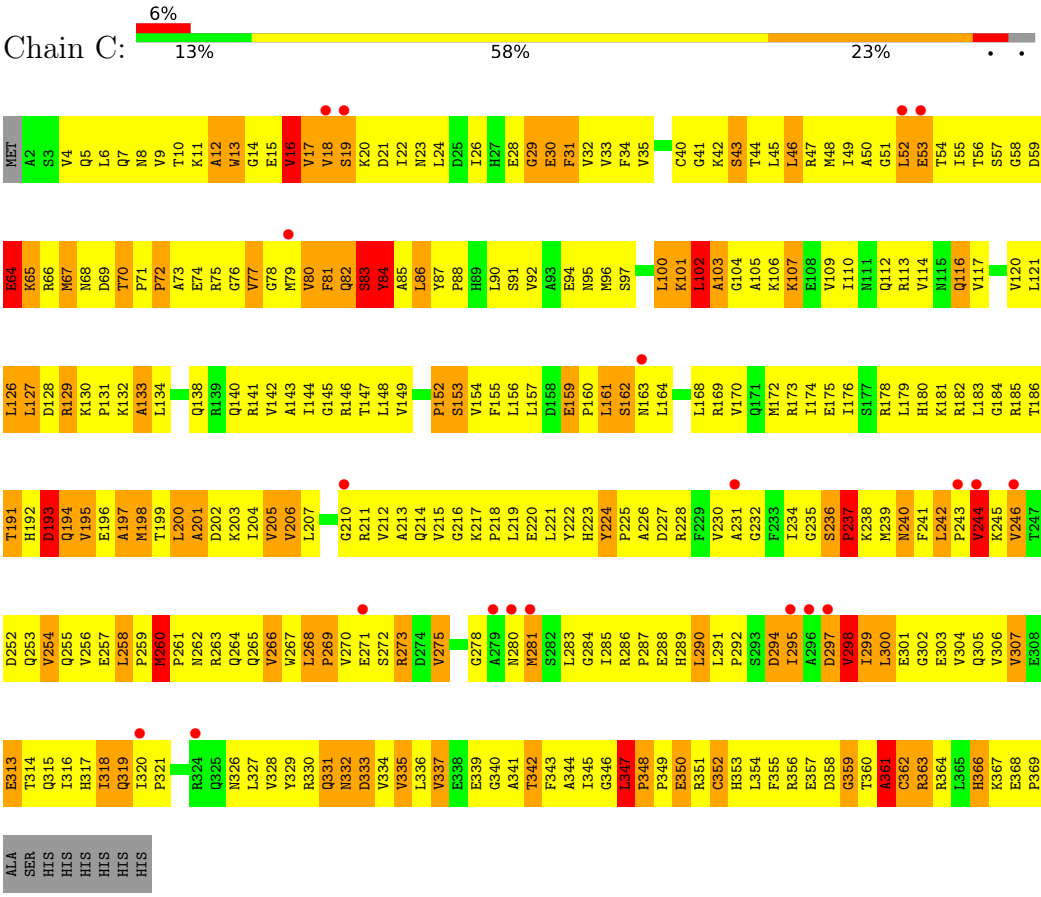
Chain B:



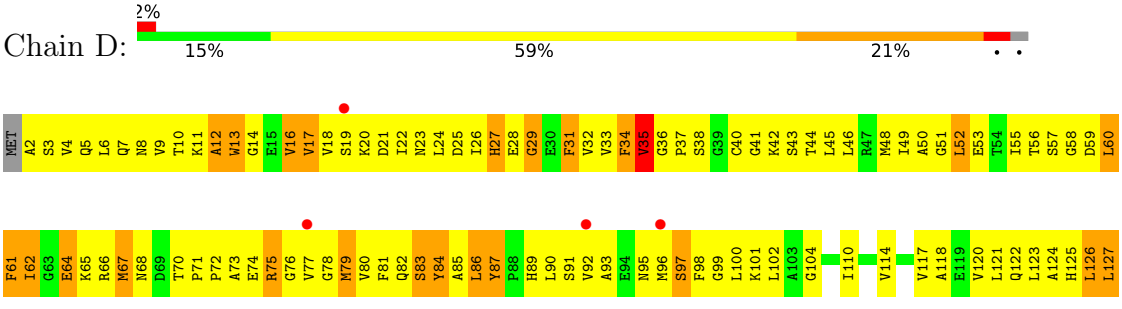


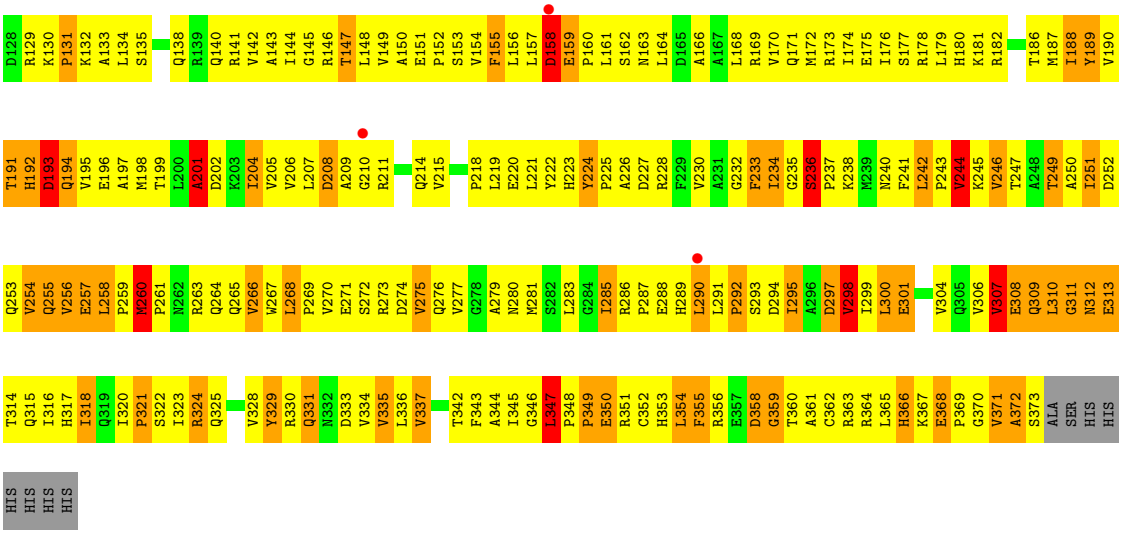


• Molecule 3: Maltose/maltodextrin import ATP-binding protein malK



• Molecule 3: Maltose/maltodextrin import ATP-binding protein malK




D128 R129 K130 P131 K132 A133 L134 S135 Q138 R139 Q140 R141 V142 A143 I144 G145 R146 T147 L148 V149 A150 E151 P152 S153 V154 F155 L156 L157 D158 E159 P160 L161 S162 N163 L164 D165 A166 A167 L168 R169 V170 Q171 M172 R173 I174 E175 I176 S177 R178 L179 H180 K181 R182 T186 M187 I188 Y189 V190
T191 H192 D193 Q194 V195 E196 A197 M198 T199 L200 A201 D202 K203 I204 V205 V206 L207 D208 A209 G210 R211 Q214 V215 P218 L219 E220 L221 Y222 H223 Y224 P225 A226 D227 R228 F229 V230 A231 G232 F233 I234 G235 S236 P237 K238 M239 N240 F241 L242 P243 V244 K245 V246 T247 A248 T249 A250 I251 D252
Q253 V254 Q255 V256 E257 L258 P259 M260 P261 N262 R263 Q264 Q265 V266 W267 L268 P269 V270 E271 S272 R273 D274 V275 Q276 V277 Q278 A279 N280 M281 S282 L283 Q284 I285 R286 P287 E288 H289 L290 L291 P292 S293 D294 I295 A296 D297 V298 I299 L300 E301 V304 Q305 V306 W307 E308 Q309 L310 G311 N312 E313
T314 Q315 I316 H317 I318 Q319 I320 P321 S322 I323 R324 Q325 V328 Y329 R330 Q331 N332 D333 V334 V335 L336 V337 T342 F343 A344 I345 G346 L347 P348 P349 E350 R351 C352 H353 L354 F355 R356 E357 D358 G359 T360 A361 C362 R363 R364 L365 H366 K367 E368 P369 G370 V371 A372 S373 ALA SER HIS HIS
HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	171.10Å 209.48Å 438.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 4.50 49.19 – 4.50	Depositor EDS
% Data completeness (in resolution range)	85.2 (50.00-4.50) 85.2 (49.19-4.50)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.54 (at 4.45Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.340 , 0.363 0.354 , 0.369	Depositor DCC
R_{free} test set	2038 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	216.6	Xtriage
Anisotropy	0.633	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.18 , 150.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	20236	wwPDB-VP
Average B, all atoms (Å^2)	300.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.84% 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

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	F	0.51	0/2473	0.85	3/3365 (0.1%)
1	H	0.51	0/2473	0.85	2/3365 (0.1%)
2	G	0.49	0/1992	0.87	5/2724 (0.2%)
2	I	0.50	0/1992	0.87	5/2724 (0.2%)
3	A	0.46	1/2926 (0.0%)	0.90	6/3968 (0.2%)
3	B	0.49	1/2932 (0.0%)	0.92	2/3976 (0.1%)
3	C	0.46	1/2926 (0.0%)	0.90	6/3968 (0.2%)
3	D	0.49	1/2932 (0.0%)	0.92	2/3976 (0.1%)
All	All	0.49	4/20646 (0.0%)	0.89	31/28066 (0.1%)

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
3	A	0	1
3	C	0	1
All	All	0	2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	159	GLU	CD-OE2	6.81	1.33	1.25
3	A	159	GLU	CD-OE2	6.79	1.33	1.25
3	D	159	GLU	CD-OE2	6.50	1.32	1.25
3	B	159	GLU	CD-OE2	6.49	1.32	1.25

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	G	259	LEU	N-CA-C	-7.22	91.50	111.00
2	I	259	LEU	N-CA-C	-7.22	91.51	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	G	196	GLY	N-CA-C	6.62	129.66	113.10
2	I	196	GLY	N-CA-C	6.62	129.66	113.10
1	H	380	GLY	N-CA-C	6.46	129.24	113.10

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	84	TYR	Sidechain
3	C	84	TYR	Sidechain

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	F	2418	0	2476	555	0
1	H	2418	0	2476	557	0
2	G	1942	0	2008	362	0
2	I	1942	0	2008	365	0
3	A	2876	0	2942	582	15
3	B	2882	0	2947	557	12
3	C	2876	0	2942	571	9
3	D	2882	0	2947	549	5
All	All	20236	0	20746	3892	31

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 95.

The worst 5 of 3892 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:F:347:ILE:HG23	1:F:348:ASN:H	1.03	1.15
3:D:157:LEU:HD23	3:D:160:PRO:HG3	1.30	1.13
1:F:275:LYS:H	1:F:276:PRO:HD2	1.07	1.13
1:F:387:LEU:HD21	1:F:429:LEU:HD13	1.29	1.12
3:C:79:MET:HG2	3:C:80:VAL:H	1.14	1.12

The worst 5 of 31 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)
3:A:182:ARG:O	3:B:324:ARG:NH1[2_565]	0.90	1.30
3:C:267:TRP:NE1	3:C:267:TRP:NE1[3_555]	1.15	1.05
3:B:340:GLY:O	3:B:340:GLY:O[2_565]	1.19	1.01
3:B:276:GLN:OE1	3:D:274:ASP:N[2_565]	1.26	0.94
3:A:119:GLU:CG	3:A:119:GLU:CD[2_565]	1.27	0.93

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	F	310/480 (65%)	209 (67%)	61 (20%)	40 (13%)	0 5
1	H	310/480 (65%)	209 (67%)	61 (20%)	40 (13%)	0 5
2	G	250/296 (84%)	179 (72%)	38 (15%)	33 (13%)	0 4
2	I	250/296 (84%)	179 (72%)	38 (15%)	33 (13%)	0 4
3	A	369/381 (97%)	198 (54%)	97 (26%)	74 (20%)	0 2
3	B	370/381 (97%)	203 (55%)	95 (26%)	72 (20%)	0 2
3	C	369/381 (97%)	198 (54%)	96 (26%)	75 (20%)	0 2
3	D	370/381 (97%)	203 (55%)	95 (26%)	72 (20%)	0 2
All	All	2598/3076 (84%)	1578 (61%)	581 (22%)	439 (17%)	0 3

5 of 439 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	90	ALA
1	F	92	THR
1	F	94	TYR
1	F	96	SER
1	F	98	ASN

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	F	253/394 (64%)	201 (79%)	52 (21%)	1 7
1	H	253/394 (64%)	201 (79%)	52 (21%)	1 7
2	G	198/237 (84%)	166 (84%)	32 (16%)	2 15
2	I	198/237 (84%)	166 (84%)	32 (16%)	2 15
3	A	314/323 (97%)	273 (87%)	41 (13%)	4 20
3	B	315/323 (98%)	276 (88%)	39 (12%)	4 22
3	C	314/323 (97%)	273 (87%)	41 (13%)	4 20
3	D	315/323 (98%)	276 (88%)	39 (12%)	4 22
All	All	2160/2554 (85%)	1832 (85%)	328 (15%)	3 16

5 of 328 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	I	130	MET
3	C	342	THR
2	I	181	PHE
3	C	83	SER
3	D	127	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 59 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	309	GLN
3	D	276	GLN
1	H	376	ASN
3	D	264	GLN
3	D	82	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)

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	F	316/480 (65%)	-0.31	3 (0%)	84	77	187, 310, 422, 493	0
1	H	316/480 (65%)	0.04	19 (6%)	21	18	187, 310, 422, 493	0
2	G	254/296 (85%)	-0.19	5 (1%)	65	56	204, 296, 401, 510	0
2	I	254/296 (85%)	-0.29	5 (1%)	65	56	204, 296, 401, 510	0
3	A	371/381 (97%)	-0.08	8 (2%)	62	52	195, 296, 384, 474	0
3	B	372/381 (97%)	-0.06	11 (2%)	50	39	150, 281, 361, 454	0
3	C	371/381 (97%)	0.15	22 (5%)	22	19	195, 296, 384, 474	0
3	D	372/381 (97%)	-0.15	7 (1%)	66	58	150, 281, 361, 454	0
All	All	2626/3076 (85%)	-0.10	80 (3%)	50	39	150, 294, 398, 510	0

The worst 5 of 80 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	280	ASN	6.5
1	H	56	ILE	6.2
1	H	483	ASP	5.7
3	C	281	MET	5.5
3	C	244	VAL	5.4

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 monosaccharides 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.