



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 08:32 PM GMT

PDB ID : 4TPH
Title : Selectivity mechanism of a bacterial homologue of the human drug peptide transporters PepT1 and PepT2
Authors : Guettou, F.; Quistgaard, E.; Raba, M.; Moberg, P.; Low, C.; Nordlund, P.
Deposited on : 2014-06-07
Resolution : 3.15 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
<http://wwpdb.org/validation/2016/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
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

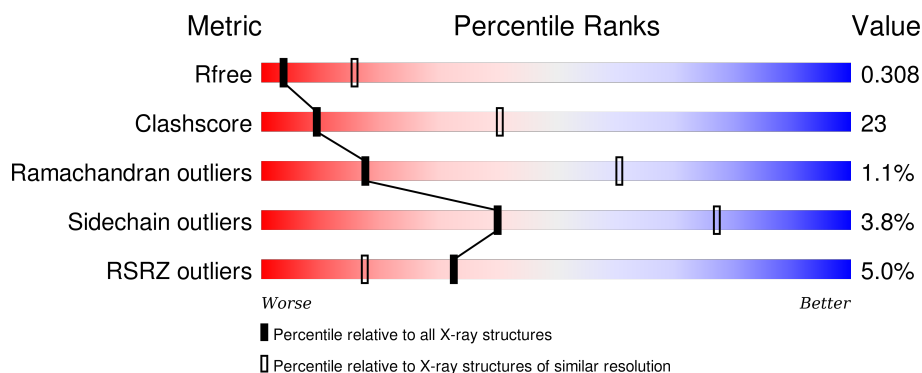
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.15 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}	91344	1112 (3.20-3.12)
Clashscore	102246	1249 (3.20-3.12)
Ramachandran outliers	100387	1222 (3.20-3.12)
Sidechain outliers	100360	1221 (3.20-3.12)
RSRZ outliers	91569	1117 (3.20-3.12)

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. 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	523	<div> <div>5%</div> <div>45%</div> <div>39%</div> <div>•</div> <div>13%</div> </div>
1	B	523	<div> <div>3%</div> <div>49%</div> <div>36%</div> <div>•</div> <div>13%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	LMT	A	602	-	-	-	X
3	LMT	B	602	-	-	-	X
4	ALA	A	603	-	-	-	X
5	DBY	A	604	-	-	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6970 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 Proton:oligopeptide symporter POT family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	453	Total	C	N	O	S	0	0	0
			3422	2277	543	582	20			
1	B	453	Total	C	N	O	S	0	0	0
			3422	2277	543	582	20			

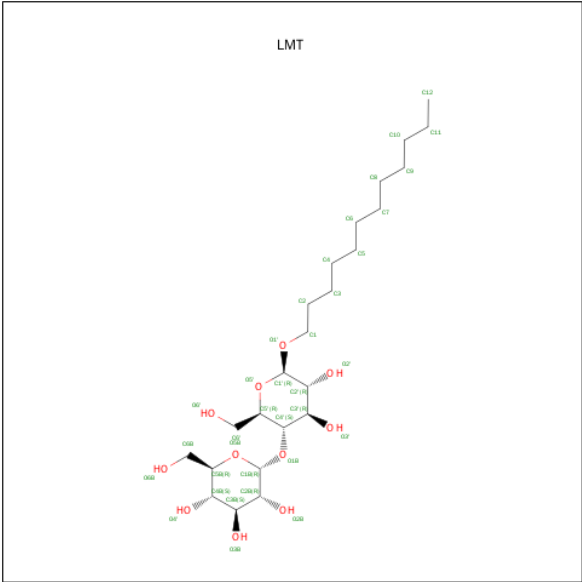
There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	517	ALA	-	expression tag	UNP Q8EHE6
A	518	GLU	-	expression tag	UNP Q8EHE6
A	519	ASN	-	expression tag	UNP Q8EHE6
A	520	LEU	-	expression tag	UNP Q8EHE6
A	521	TYR	-	expression tag	UNP Q8EHE6
A	522	PHE	-	expression tag	UNP Q8EHE6
A	523	GLN	-	expression tag	UNP Q8EHE6
B	517	ALA	-	expression tag	UNP Q8EHE6
B	518	GLU	-	expression tag	UNP Q8EHE6
B	519	ASN	-	expression tag	UNP Q8EHE6
B	520	LEU	-	expression tag	UNP Q8EHE6
B	521	TYR	-	expression tag	UNP Q8EHE6
B	522	PHE	-	expression tag	UNP Q8EHE6
B	523	GLN	-	expression tag	UNP Q8EHE6

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

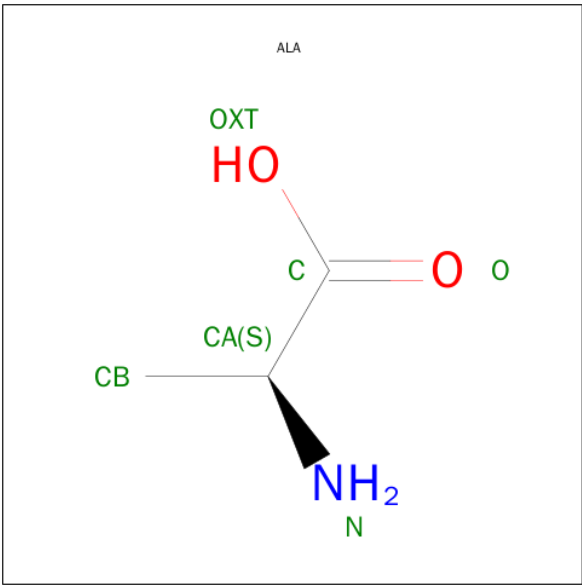
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is DODECYL-BETA-D-MALTOSIDE (three-letter code: LMT) (formula: C₂₄H₄₆O₁₁).



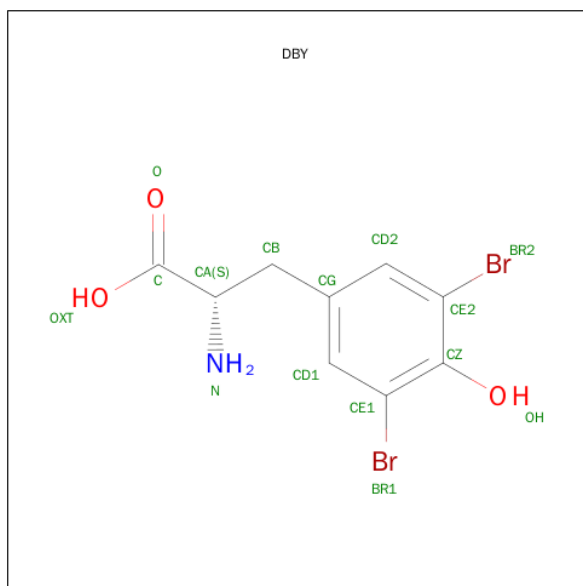
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			35	24	11		
3	B	1	Total	C	O	0	0
			35	24	11		
3	B	1	Total	C	O	0	0
			35	24	11		

- Molecule 4 is ALANINE (three-letter code: ALA) (formula: C₃H₇NO₂).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			5	3	1	1		

- Molecule 5 is 3,5 DIBROMOTYROSINE (three-letter code: DBY) (formula: $C_9H_9Br_2NO_3$).

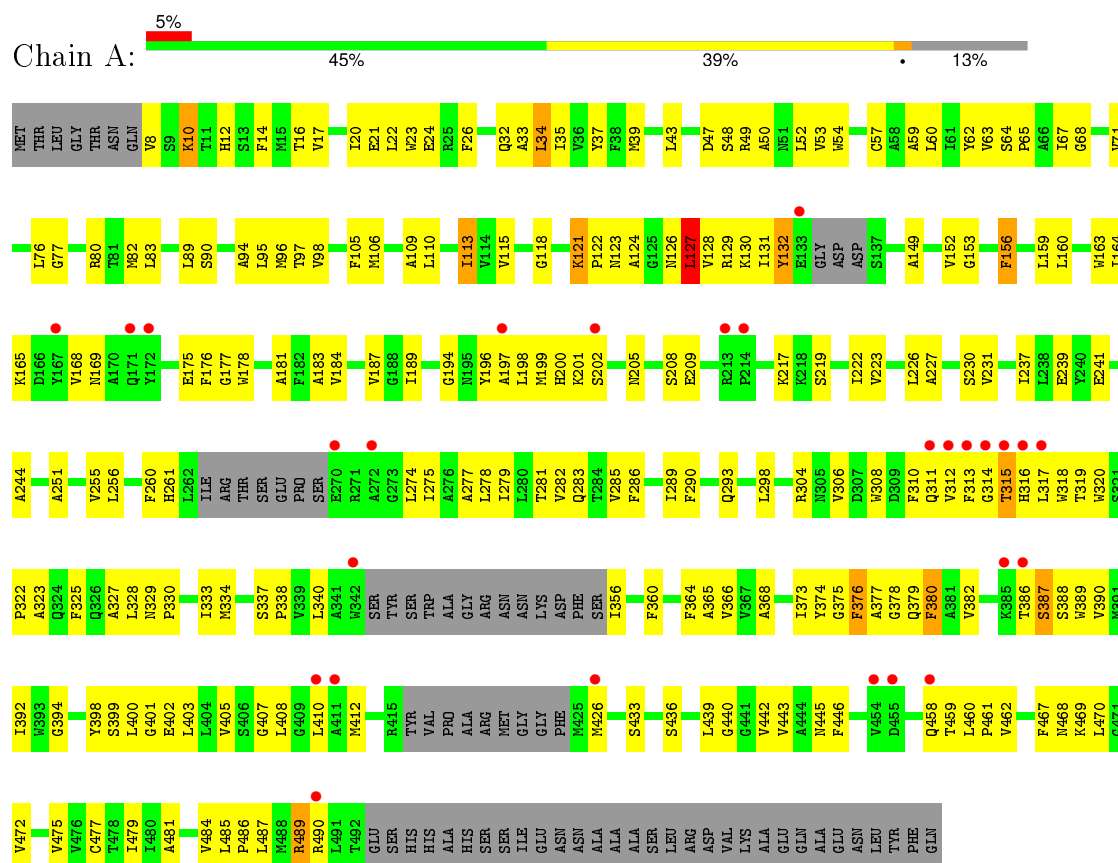


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	Br	C	N	O	0	0
			15	2	9	1	3		

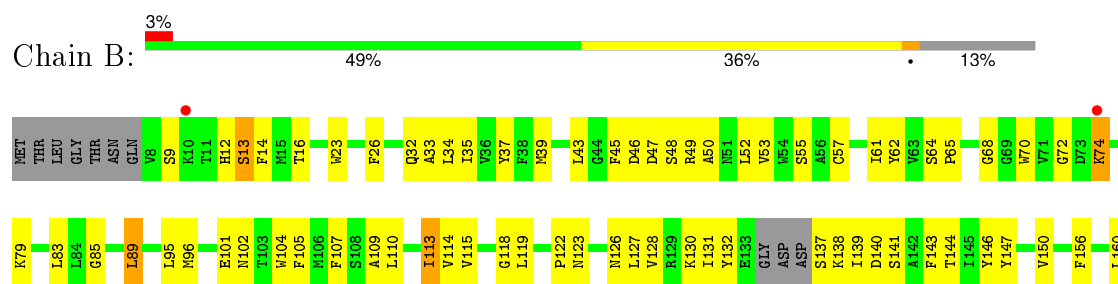
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Proton:oligopeptide symporter POT family



• Molecule 1: Proton:oligopeptide symporter POT family



SER	HIS	HIS	HIS	ALA	ALA	SER	GLY	GLY	PHE	VAL	W342	SER	TYR	TRP	ALA	GLY	GLY	ARG	ASN	ASN	LYS	ASP	PHE	SER	S356	F360	F364	A365	V366	I369	G370	F371	F372	I373	I374	A377	G378	V382	T386	S387	S388	H389	V390	M391	I392	H393	G394	S397	Y398	S399	L400	L403	L408	G409	L410	A411	M412	I413	A414	R415	TYR
VAL	PRO	ALA	ARG	MET	GLY	GLY	PHE	VAL	M425	M426	Y429	S433	S436	Q437	Y438	L439	G440	G441	V442	V443	A444	N445	F446	A447	S448	V449	P450	Q451	D452	L453	Q458	P461	L466	F467	L470	V475	V476	C477	I480	A481	V484	L485	P486	L487	M488	R489	R490	L491	T492	GLU											
R271	A272	W273	L274	A277	L278	I279	L280	T281	V282	Q283	T284	V285	F286	I289	F290	Y291	Q292	Q293	M294	L298	A299	R304	N305	V306	W307	W308	F309	F310	Q311	V312	F313	G314	T315	H316	L317	W318	T319	W320	S321	F322	A327	L328	N329	P330	I331	W332	I333	N334	V335	P338	V339	L340	A341								
T161	P162	W163	I164	K165	D166	E175	A183	V187	L190	G194	N195	Y196	A197	L198	M199	H200	K201	S202	L203	A204	N205	S208	E209	N216	K217	K218	S219	I222	S230	V231	V232	A233	I237	L238	E239	Y240	E241	A244	L262	I1E	ARG	THR	SER	GLU	PRO	SER	E270														

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	80.65Å 107.68Å 204.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.54 – 3.15 29.54 – 3.16	Depositor EDS
% Data completeness (in resolution range)	78.2 (29.54-3.15) 78.4 (29.54-3.16)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.23 (at 3.18Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, R_{free}	0.251 , 0.307 0.251 , 0.308	Depositor DCC
R_{free} test set	1255 reflections (5.12%)	DCC
Wilson B-factor (Å ²)	101.4	Xtriage
Anisotropy	0.059	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 70.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	2 of 24566 reflections (0.008%)	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	6970	wwPDB-VP
Average B, all atoms (Å ²)	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.73% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: DBY, ZN, LMT

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	0.64	1/3509 (0.0%)	0.85	4/4785 (0.1%)
1	B	0.63	1/3509 (0.0%)	0.82	3/4785 (0.1%)
All	All	0.64	2/7018 (0.0%)	0.83	7/9570 (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
1	A	0	1
1	B	0	1
All	All	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	33	ALA	C-O	5.64	1.34	1.23
1	B	33	ALA	C-O	5.27	1.33	1.23

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	127	LEU	CA-CB-CG	7.75	133.12	115.30
1	A	408	LEU	CA-CB-CG	6.61	130.50	115.30
1	B	408	LEU	CA-CB-CG	6.37	129.96	115.30
1	A	127	LEU	CB-CG-CD2	-5.89	100.99	111.00
1	A	34	LEU	CA-CB-CG	-5.38	102.94	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	132	TYR	Peptide
1	B	274	LEU	Peptide

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	3422	0	3412	172	0
1	B	3422	0	3412	143	0
2	A	1	0	0	0	0
3	A	35	0	46	9	0
3	B	70	0	92	6	0
4	A	5	0	4	0	0
5	A	15	0	6	1	0
All	All	6970	0	6972	317	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 23.

The worst 5 of 317 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:B:364:PHE:CD2	1:B:477:CYS:HB3	2.03	0.93
1:A:152:VAL:HG22	1:A:156:PHE:CE1	2.08	0.89
1:A:94:ALA:O	1:A:97:THR:HG22	1.76	0.86
1:A:443:VAL:HA	1:A:446:PHE:HD2	1.42	0.84
1:B:200:HIS:CG	1:B:201:LYS:N	2.51	0.77

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	443/523 (85%)	403 (91%)	35 (8%)	5 (1%)	17	61
1	B	443/523 (85%)	403 (91%)	35 (8%)	5 (1%)	17	61
All	All	886/1046 (85%)	806 (91%)	70 (8%)	10 (1%)	17	61

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	205	ASN
1	A	208	SER
1	A	313	PHE
1	B	205	ASN
1	B	208	SER

5.3.2 Protein sidechains

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	340/417 (82%)	323 (95%)	17 (5%)	30	70
1	B	340/417 (82%)	331 (97%)	9 (3%)	54	85
All	All	680/834 (82%)	654 (96%)	26 (4%)	40	77

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	337	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	380	PHE
1	B	369	ILE
1	A	360	PHE
1	A	376	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	329	ASN
1	B	451	GLN
1	B	102	ASN
1	A	311	GLN
1	A	383	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	LMT	A	602	-	36,36,36	1.25	3 (8%)	47,47,47	1.73	9 (19%)
4	ALA	A	603	5	3,4,5	0.95	0	0,4,6	0.00	-
5	DBY	A	604	4	12,15,15	0.95	1 (8%)	15,21,21	1.24	2 (13%)
3	LMT	B	601	-	36,36,36	1.15	2 (5%)	47,47,47	1.53	7 (14%)
3	LMT	B	602	-	36,36,36	1.22	5 (13%)	47,47,47	1.44	7 (14%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	LMT	A	602	-	-	0/21/61/61	0/2/2/2
4	ALA	A	603	5	-	0/0/2/4	0/0/0/0
5	DBY	A	604	4	-	0/4/8/8	0/1/1/1
3	LMT	B	601	-	-	0/21/61/61	0/2/2/2
3	LMT	B	602	-	-	0/21/61/61	0/2/2/2

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	602	LMT	O3'-C3'	2.00	1.47	1.43
3	B	602	LMT	O5'-C5'	2.14	1.49	1.44
3	A	602	LMT	O5'-C5'	2.19	1.49	1.44
5	A	604	DBY	BR1-CE1	2.21	1.95	1.89
3	B	602	LMT	O1B-C4'	2.23	1.49	1.43

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	602	LMT	C1B-O1B-C4'	-3.08	109.95	118.01
3	B	601	LMT	C1B-O1B-C4'	-2.60	111.20	118.01
3	A	602	LMT	O3'-C3'-C2'	-2.15	105.49	110.34
5	A	604	DBY	CD1-CE1-CZ	-2.09	119.36	121.89
3	B	602	LMT	O5'-C5'-C4'	2.03	114.03	109.75

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 16 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	602	LMT	9	0
5	A	604	DBY	1	0
3	B	601	LMT	3	0
3	B	602	LMT	3	0

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	453/523 (86%)	-0.11	27 (5%) 25 13	61, 93, 141, 159	0
1	B	453/523 (86%)	-0.16	18 (3%) 42 26	56, 91, 143, 165	0
All	All	906/1046 (86%)	-0.14	45 (4%) 32 17	56, 93, 142, 165	0

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	315	THR	6.0
1	B	315	THR	5.8
1	A	312	VAL	5.3
1	B	202	SER	5.0
1	A	411	ALA	4.3

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
4	ALA	A	603	5/6	0.72	0.43	7.25	115,116,136,137	0
3	LMT	B	602	35/35	0.70	0.61	4.83	64,117,125,128	0
5	DBY	A	604	15/15	0.78	0.37	2.41	140,168,246,276	0
3	LMT	A	602	35/35	0.83	0.43	1.85	57,114,125,135	0
3	LMT	B	601	35/35	0.84	0.40	1.43	85,115,126,134	0
2	ZN	A	601	1/1	0.73	0.35	-	150,150,150,150	0

6.5 Other polymers [i](#)

There are no such residues in this entry.