WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:31 AM GMT

PDB ID	:	2R1A
Title	:	Crystal structure of the periplasmic lipopolysaccharide transport protein LptA (YhbN), trigonal form
Authors	:	Suits, M.D.L.; Polissi, A.; Jia, Z.; Montreal-Kingston Bacterial Structural Genomics Initiative (BSGI)
Deposited on	:	2007-08-22
Resolution	:	3.26 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

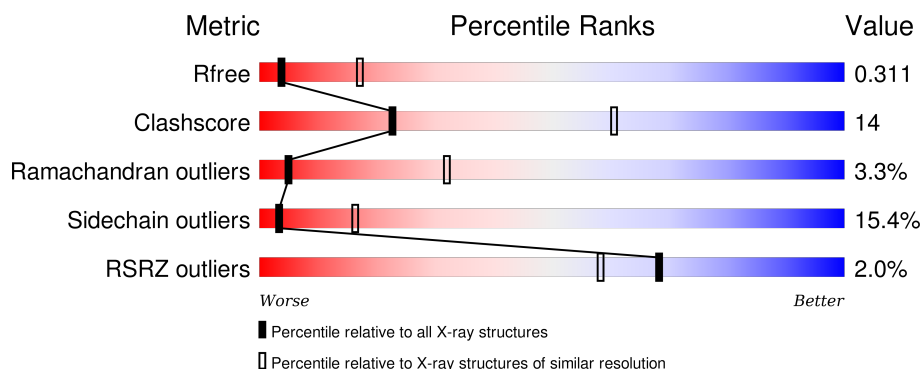|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.26 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	Percentile Ranks	Value
Rfree		0.311
Clashscore		14
Ramachandran outliers		3.3%
Sidechain outliers		15.4%
RSRZ outliers		2.0%

Worse *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1624 (3.32-3.20)
Clashscore	102246	1806 (3.32-3.20)
Ramachandran outliers	100387	1773 (3.32-3.20)
Sidechain outliers	100360	1771 (3.32-3.20)
RSRZ outliers	91569	1632 (3.32-3.20)

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 <=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	159	58% 22% •• 18%
1	B	159	47% 34% • • 16%
1	C	159	60% 20% 5% 15%
1	D	159	61% 21% • 14%
1	E	159	55% 24% • • 17%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	159	4% ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ 45% 30% 8% 17%
1	G	159	▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ 67% 13% 5% 15%
1	H	159	3% ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ 60% 18% • 18%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7620 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 Protein yhbN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	131	Total 927	C 578	N 154	O 191	S 4	0	0	0
1	B	134	Total 956	C 594	N 160	O 198	S 4	0	0	0
1	C	135	Total 989	C 617	N 163	O 206	S 3	0	0	0
1	D	136	Total 923	C 572	N 156	O 193	S 2	0	0	0
1	E	132	Total 918	C 569	N 155	O 192	S 2	0	0	0
1	F	132	Total 938	C 584	N 155	O 196	S 3	0	0	0
1	G	135	Total 980	C 609	N 165	O 203	S 3	0	0	0
1	H	131	Total 921	C 572	N 152	O 195	S 2	0	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	9	Total 9	O 9	0	0
2	B	8	Total 8	O 8	0	0
2	C	2	Total 2	O 2	0	0
2	D	8	Total 8	O 8	0	0
2	E	14	Total 14	O 14	0	0
2	F	9	Total 9	O 9	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	14	Total 14	O 14	0	0
2	H	4	Total 4	O 4	0	0

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: Protein yhbN

• Molecule 1: Protein yhbN

• Molecule 1: Protein yhbN

• Molecule 1: Protein yhbN

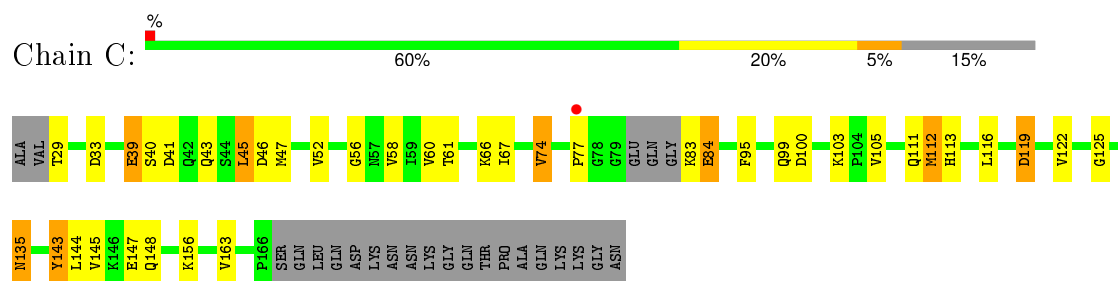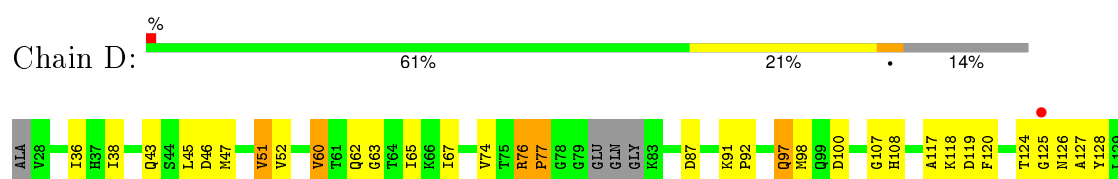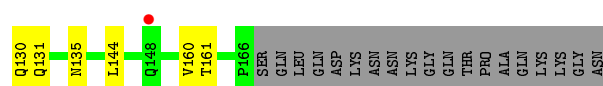Q130 Q131 N135 L144 Q148 V160 T161 P166 SER GLN LEU GLN ASP LYS ASN ASN LYS GLY GLN THR PRO ALA GLN LYS LYS GLY ASN

• Molecule 1: Protein yhbN

Chain E:

%

55%   24%   • •   17%

ALA VAL THR GLY D31 I36 E39 S40 D41 Q42 Q43 N50 V51 V52 G56 V60 I65 K66 I67 V74 P77 GLY GLU GLN GLY K83 E84 V85 I86 D87 K91 P92 F95 Y96 Q97 D100 N101 G102 K103 P104 V105 H113 Y114 E115 L116 A117 K118

D119 V122 G125 Y128 Q131 N135 I136 K137 L144 A152 F153 S154 G157 V160 V163 L164 V165 P166 S167 GLN LEU GLN ASP LYS ASN ASN LYS GLY GLN THR PRO ALA GLN LYS LYS GLY ASN

• Molecule 1: Protein yhbN

Chain F:

4%

45%   30%   8%   17%

ALA V28 D31 H37 D41 Q42 Q43 S44 L45 G49 V52 T55 G56 N57 V58 I59 V60 T61 Q62 I65 K66 I67 N68 V72 V73 V74 T75 R76 PRO GLY GLU GLN GLY K83 Y96 Q97 M98 Q99 D100 N101 G102 E106 Q111 M112 H113 Y114 E115 L116

A117 K118 D119 F120 V121 V122 L123 T124 G125 N126 A127 Y128 L129 Q130 Q131 V132 D133 S134 N135 I136 Y143 L144 E147 M150 Q151 S154 D155 K156 G157 K158 R159 L164 V165 PRO SER GLN LEU GLN ASP LYS ASN ASN LYS GLY GLN THR PRO ALA GLN LYS LYS GLY ASN

• Molecule 1: Protein yhbN

Chain G:

67%   13%   5%   15%

ALA V28 D33 D41 Q42 Q43 D46 V52 T53 G56 N57 V58 I59 V60 I67 V72 V73 V74 T75 G79 GLU GLN GLY K83 E84 M112 H113 Y114 E115 L116 F120 L123 T124 G125 Y128 L129 Q130 N135 Y143 L144 Q148 K149 M150 V165

PRO SER GLN LEU GLN ASP LYS ASN ASN LYS GLY GLN THR PRO ALA GLN LYS LYS GLY ASN

• Molecule 1: Protein yhbN

Chain H:

3%

60%   18%   •   18%

ALA V28 I38 E39 S40 Q43 S44 L45 D46 M47 Q48 G49 N50 T53 F54 V58 I59 V60 I67 N68 V73 V74 T75 R76 PRO GLY GLU GLN GLY K83 I86 D87 Q97 M98 S110 Q111 T124 G125 Y128 Q131 V132 N135 L144 E147

Q148 S154 D155 K156 G157 V160 T161 L164 VAL PRO SER GLN LEU GLN ASP LYS ASN ASN LYS GLY GLN THR PRO ALA GLN LYS LYS GLY ASN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	146.21Å 146.21Å 186.81Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.98 – 3.26 29.98 – 3.25	Depositor EDS
% Data completeness (in resolution range)	83.6 (29.98-3.26) 97.3 (29.98-3.25)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	3.55 (at 3.24Å)	Xtriage
Refinement program	REFMAC 5.4.0065	Depositor
R, R_{free}	0.298 , 0.361 0.306 , 0.311	Depositor DCC
R_{free} test set	1788 reflections (5.29%)	DCC
Wilson B-factor (Å^2)	67.8	Xtriage
Anisotropy	0.935	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 62.6	EDS
Estimated twinning fraction	0.033 for -h,-k,l	Xtriage
L-test for twinning [2]	$< |L| > = 0.45, < L^2 > = 0.28$	Xtriage
Outliers	0 of 35752 reflections	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	7620	wwPDB-VP
Average B, all atoms (Å^2)	57.0	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 32.40 % 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 9.4591e-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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.34	0/940	0.52	0/1283
1	B	0.38	0/971	0.54	0/1327
1	C	0.36	0/1004	0.53	0/1371
1	D	0.36	0/937	0.51	0/1285
1	E	0.36	0/932	0.57	0/1280
1	F	0.39	1/951 (0.1%)	0.58	0/1301
1	G	0.34	0/994	0.52	0/1356
1	H	0.34	0/934	0.50	0/1281
All	All	0.36	1/7663 (0.0%)	0.53	0/10484

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	132	VAL	CB-CG1	-5.18	1.42	1.52

There are no bond angle outliers.

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	927	0	798	25	0
1	B	956	0	831	35	0
1	C	989	0	899	28	0
1	D	923	0	762	20	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	918	0	758	25	0
1	F	938	0	817	36	0
1	G	980	0	882	16	0
1	H	921	0	783	17	0
2	A	9	0	0	1	0
2	B	8	0	0	1	0
2	C	2	0	0	0	0
2	D	8	0	0	0	0
2	E	14	0	0	1	0
2	F	9	0	0	1	0
2	G	14	0	0	1	0
2	H	4	0	0	2	0
All	All	7620	0	6530	191	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 14.

The worst 5 of 191 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:148:GLN:HG3	1:B:47:MET:HB2	1.48	0.95
1:F:98:MET:HA	1:F:99:GLN:CB	1.98	0.93
1:F:97:GLN:HG2	1:F:98:MET:H	1.43	0.83
1:B:148:GLN:HB3	1:C:47:MET:HB2	1.60	0.81
1:C:143:TYR:CE1	1:C:145:VAL:HA	2.15	0.81

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	125/159 (79%)	98 (78%)	23 (18%)	4 (3%)	5	33
1	B	130/159 (82%)	95 (73%)	31 (24%)	4 (3%)	5	34
1	C	131/159 (82%)	117 (89%)	12 (9%)	2 (2%)	13	53
1	D	132/159 (83%)	105 (80%)	19 (14%)	8 (6%)	2	15
1	E	128/159 (80%)	101 (79%)	19 (15%)	8 (6%)	2	14
1	F	128/159 (80%)	100 (78%)	24 (19%)	4 (3%)	5	34
1	G	131/159 (82%)	116 (88%)	14 (11%)	1 (1%)	24	66
1	H	127/159 (80%)	109 (86%)	15 (12%)	3 (2%)	7	41
All	All	1032/1272 (81%)	841 (82%)	157 (15%)	34 (3%)	5	32

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	156	LYS
1	D	77	PRO
1	D	119	ASP
1	E	118	LYS
1	F	99	GLN

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	88/135 (65%)	77 (88%)	11 (12%)	6	26
1	B	91/135 (67%)	73 (80%)	18 (20%)	1	8
1	C	101/135 (75%)	89 (88%)	12 (12%)	6	28
1	D	82/135 (61%)	72 (88%)	10 (12%)	6	27
1	E	82/135 (61%)	71 (87%)	11 (13%)	5	22
1	F	91/135 (67%)	65 (71%)	26 (29%)	0	1
1	G	98/135 (73%)	85 (87%)	13 (13%)	5	23
1	H	87/135 (64%)	77 (88%)	10 (12%)	7	30

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	720/1080 (67%)	609 (85%)	111 (15%)	3 16

5 of 111 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	144	LEU
1	F	31	ASP
1	H	43	GLN
1	E	39	GLU
1	E	97	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	148	GLN
1	D	130	GLN
1	H	68	ASN
1	D	97	GLN
1	E	62	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, 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(Å²)	Q<0.9
1	A	131/159 (82%)	-0.04	2 (1%)	76	67	44, 64, 75, 77	0
1	B	134/159 (84%)	0.13	3 (2%)	65	55	54, 63, 74, 80	0
1	C	135/159 (84%)	-0.19	1 (0%)	89	84	32, 41, 61, 65	0
1	D	136/159 (85%)	0.02	2 (1%)	76	67	47, 69, 76, 78	0
1	E	132/159 (83%)	0.15	2 (1%)	76	67	63, 71, 77, 80	0
1	F	132/159 (83%)	0.07	7 (5%)	30	22	46, 58, 68, 70	0
1	G	135/159 (84%)	-0.18	0	100	100	34, 46, 58, 60	0
1	H	131/159 (82%)	-0.06	4 (3%)	52	43	35, 54, 68, 71	0
All	All	1066/1272 (83%)	-0.01	21 (1%)	68	58	32, 60, 75, 80	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	77	PRO	3.4
1	H	155	ASP	3.1
1	B	77	PRO	2.8
1	F	155	ASP	2.8
1	F	154	SER	2.8

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 (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.