wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 06:27 PM GMT

PDB ID	:	4LOG
Title	:	The crystal structure of the orphan nuclear receptor PNR ligand binding domain fused with MBP
Authors	:	Tan, M.E.; Zhou, X.E.; Soon, F.-F.; Li, X.; Li, J.; Yong, E.-L.; Melcher, K.; Xu, H.E.
Deposited on	:	2013-07-12
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
http://wwpdb.org/validation/2016/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:

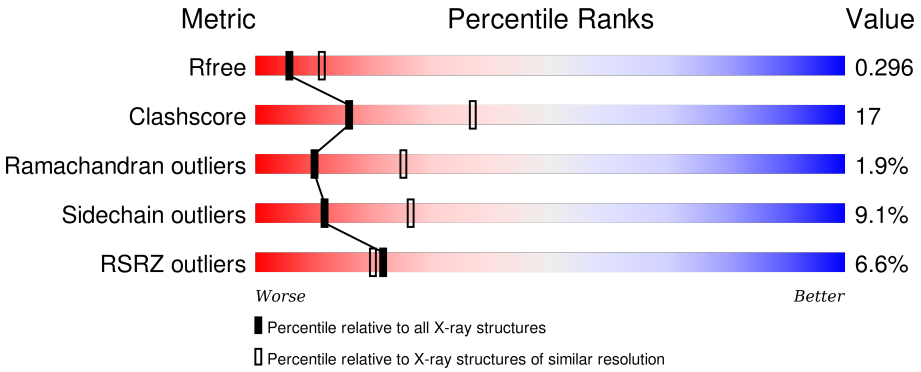
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 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}	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (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 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	574	4% 61% 31%
1	B	574	5% 24% 14% 60%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6234 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 ABC transporter periplasmic protein and NR2E3 protein chimeric construct.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	553	Total	C	N	O	S	0	0	0
			4334	2799	712	809	14			
1	B	232	Total	C	N	O	S	0	0	0
			1838	1184	314	330	10			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP K0BGG6
A	369	ASN	-	LINKER	UNP K0BGG6
A	370	ALA	-	LINKER	UNP K0BGG6
A	371	ALA	-	LINKER	UNP K0BGG6
A	372	ALA	-	LINKER	UNP K0BGG6
A	1215	GLU	-	LINKER	UNP K0BGG6
A	1216	PHE	-	LINKER	UNP K0BGG6
A	1405	SER	CYS	ENGINEERED MUTATION	UNP Q8IVZ9
A	1411	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
A	1412	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
A	1413	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
A	1414	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
A	1415	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
A	1416	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
B	1	MET	-	INITIATING METHIONINE	UNP K0BGG6
B	369	ASN	-	LINKER	UNP K0BGG6
B	370	ALA	-	LINKER	UNP K0BGG6
B	371	ALA	-	LINKER	UNP K0BGG6
B	372	ALA	-	LINKER	UNP K0BGG6
B	1215	GLU	-	LINKER	UNP K0BGG6
B	1216	PHE	-	LINKER	UNP K0BGG6
B	1405	SER	CYS	ENGINEERED MUTATION	UNP Q8IVZ9
B	1411	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
B	1412	HIS	-	EXPRESSION TAG	UNP Q8IVZ9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1413	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
B	1414	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
B	1415	HIS	-	EXPRESSION TAG	UNP Q8IVZ9
B	1416	HIS	-	EXPRESSION TAG	UNP Q8IVZ9

- Molecule 2 is water.

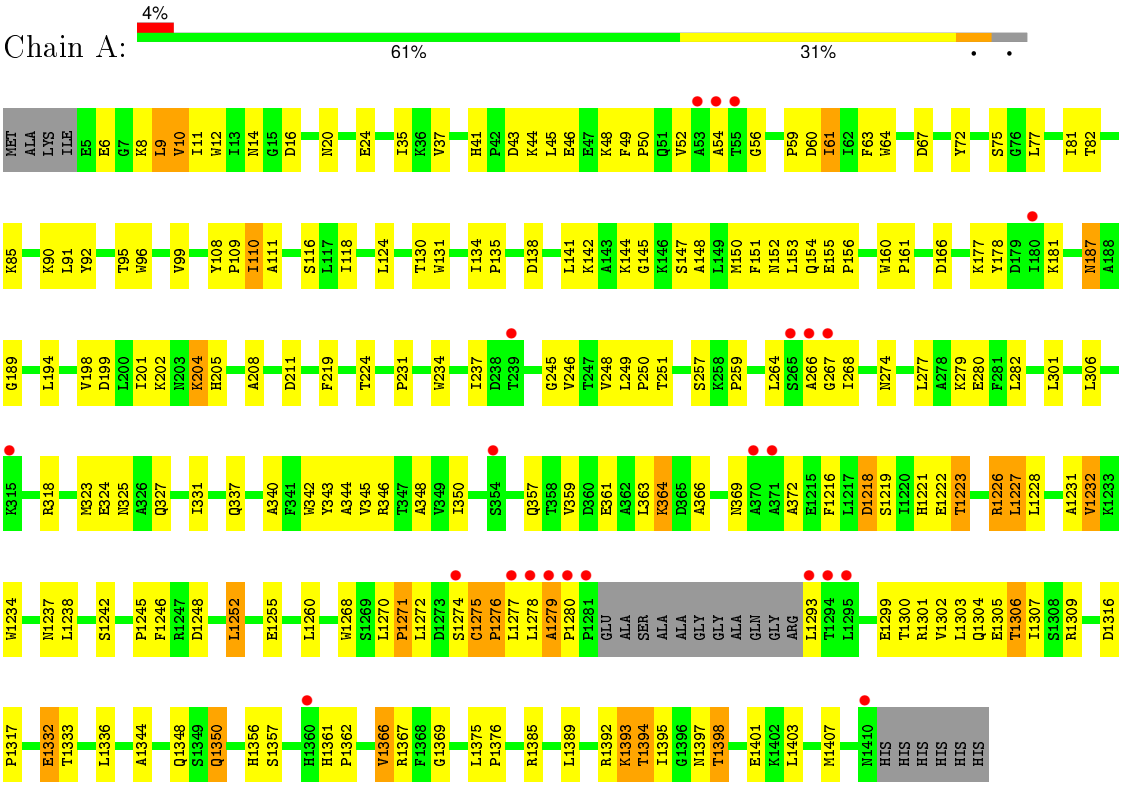
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	33	Total O 33 33	0	0
2	B	29	Total O 29 29	0	0

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 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.

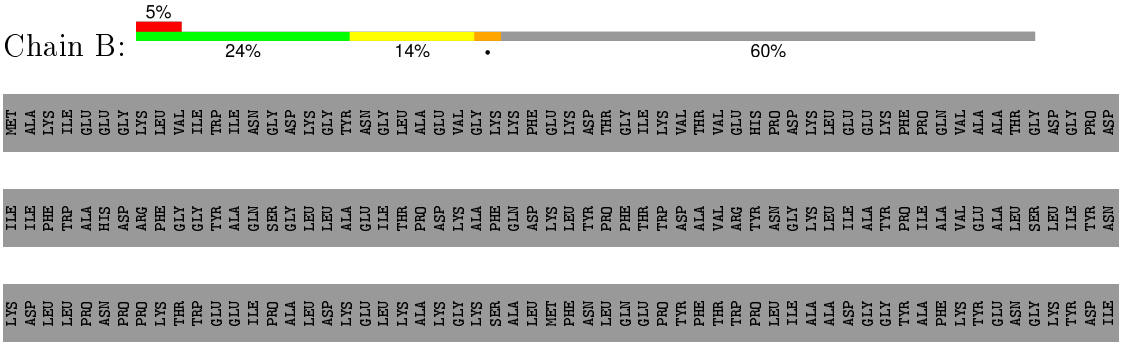
• Molecule 1: Maltose ABC transporter periplasmic protein and NR2E3 protein chimeric construct

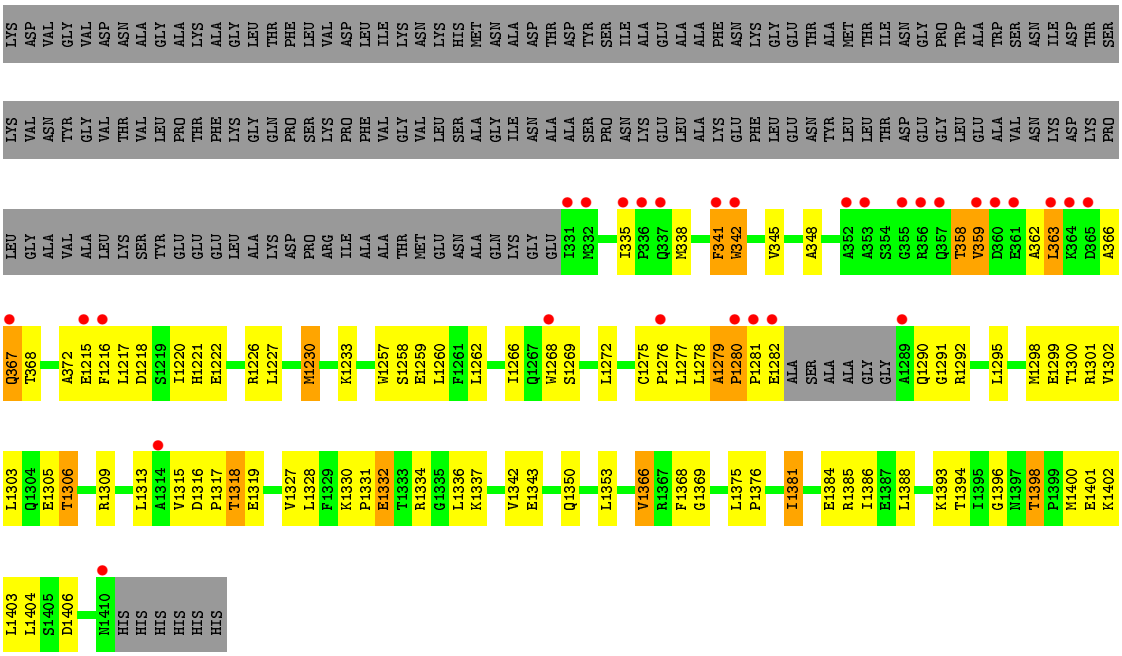
Chain A:



• Molecule 1: Maltose ABC transporter periplasmic protein and NR2E3 protein chimeric construct

Chain B:




LYS ASP VAL GLY VAL ASP ASN ALA GLY ALA LYS ALA GLY LEU THR PHE LEU VAL ASP LEU ILE LYS ASN LYS HIS MET ASN ALA ASP THR ASP TYR SER ILE ALA GLU ALA ALA PHE ASN LYS GLY GLU THR ALA MET THR ILE ASN GLY PRO TRP ALA TRP SER ASN ILE ASP THR SER
LYS VAL ASN TYR GLY VAL THR VAL LEU PRO THR PHE LYS GLY GLN PRO SER LYS PRO PHE VAL GLY VAL LEU SER ALA GLY ILE ASN ALA ALA SER PRO ASN LYS GLU LEU ALA LYS GLU PHE LEU GLU ASN TYR LEU LEU THR ASP GLU GLY LEU GLU ALA VAL ASN LYS ASP LYS PRO
LEU GLY ALA VAL ALA LEU LYS SER TYR GLU GLU GLU LEU ALA LYS ASP PRO ARG ILE ALA ALA THR MET GLU ASN ALA GLN LYS GLY GLU I331 M332 I335 P336 Q337 M338 F341 W342 V345 A348 A352 A353 S354 G355 R356 Q357 T358 V359 D360 E361 A362 L363 K364 D365 A366
Q367 T368 A372 E1215 F1216 L1217 D1218 S1219 I1220 H1221 E1222 R1226 L1227 M1230 K1233 W1257 S1258 E1259 L1260 F1261 L1262 I1266 Q1267 W1268 S1269 L1272 C1275 P1276 L1277 L1278 A1279 P1280 P1281 E1282 ALA SER ALA ALA GLY GLY A1289 Q1290 G1291 R1292 L1295 M1298 E1299 T1300 R1301 V1302
L1303 Q1304 E1305 T1306 R1309 L1313 A1314 V1315 D1316 P1317 T1318 E1319 V1327 L1328 F1329 K1330 P1331 E1332 T1333 R1334 G1335 L1336 K1337 V1342 E1343 Q1350 L1353 V1366 R1367 F1368 G1369 L1375 P1376 I1381 E1384 R1385 I1386 E1387 L1388 K1393 T1394 I1395 G1396 N1397 T1398 P1399 M1400 E1401 K1402
L1403 L1404 S1405 D1406 N1410 HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	89.44Å 184.94Å 85.97Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.70 29.70 – 2.70	Depositor EDS
% Data completeness (in resolution range)	96.5 (30.00-2.70) 96.4 (29.70-2.70)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.73 (at 2.68Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.285 , 0.307 0.278 , 0.296	Depositor DCC
R_{free} test set	2958 reflections (8.30%)	DCC
Wilson B-factor (Å^2)	62.2	Xtriage
Anisotropy	0.176	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 40.8	EDS
Estimated twinning fraction	0.017 for l,-k,h	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 38646 reflections	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	6234	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.42% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.57	0/4439	0.59	0/6025
1	B	0.62	2/1879 (0.1%)	0.64	0/2548
All	All	0.58	2/6318 (0.0%)	0.61	0/8573

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	342	TRP	CD2-CE2	5.45	1.47	1.41
1	B	1268	TRP	CD2-CE2	5.04	1.47	1.41

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	4334	0	4340	155	0
1	B	1838	0	1875	69	0
2	A	33	0	0	6	0
2	B	29	0	0	2	0
All	All	6234	0	6215	210	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 210 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:1302:VAL:O	1:B:1306:THR:HG22	1.50	1.12
1:A:61:ILE:HD11	1:A:266:ALA:HB1	1.46	0.96
1:B:1400:MET:HE2	1:B:1404:LEU:HG	1.67	0.76
1:A:1275:CYS:HB2	1:A:1276:PRO:HD3	1.68	0.76
1:A:152:ASN:HD21	1:A:155:GLU:HG2	1.53	0.74

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	549/574 (96%)	488 (89%)	53 (10%)	8 (2%)	13 32
1	B	228/574 (40%)	199 (87%)	22 (10%)	7 (3%)	5 12
All	All	777/1148 (68%)	687 (88%)	75 (10%)	15 (2%)	10 25

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1276	PRO
1	B	362	ALA
1	B	1281	PRO
1	A	1275	CYS
1	B	1277	LEU

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	459/472 (97%)	422 (92%)	37 (8%)	15 33
1	B	201/472 (43%)	178 (89%)	23 (11%)	7 16
All	All	660/944 (70%)	600 (91%)	60 (9%)	12 26

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1293	LEU
1	A	1393	LYS
1	B	1384	GLU
1	A	1332	GLU
1	A	1397	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1267	GLN
1	A	1350	GLN
1	B	1290	GLN
1	A	357	GLN
1	B	1341	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	553/574 (96%)	0.13	23 (4%) 40 39	21, 59, 99, 161	0
1	B	232/574 (40%)	0.46	29 (12%) 5 4	21, 50, 112, 148	7 (3%)
All	All	785/1148 (68%)	0.23	52 (6%) 22 20	21, 57, 104, 161	7 (0%)

The worst 5 of 52 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1281	PRO	12.1
1	A	180	ILE	8.4
1	B	360	ASP	6.8
1	A	1279	ALA	6.4
1	B	337	GLN	6.4

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 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.