wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 06:53 AM GMT

PDB ID	:	2YPA
Title	:	Structure of the SCL:E47:LMO2:LDB1 complex bound to DNA
Authors	:	El Omari, K.; Hoosdally, S.J.; Tuladhar, K.; Karia, D.; Ponsele, E.; Platonova, O.; Vyas, P.; Patient, R.; Porcher, C.; Mancini, E.J.
Deposited on	:	2012-10-30
Resolution	:	2.80 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:

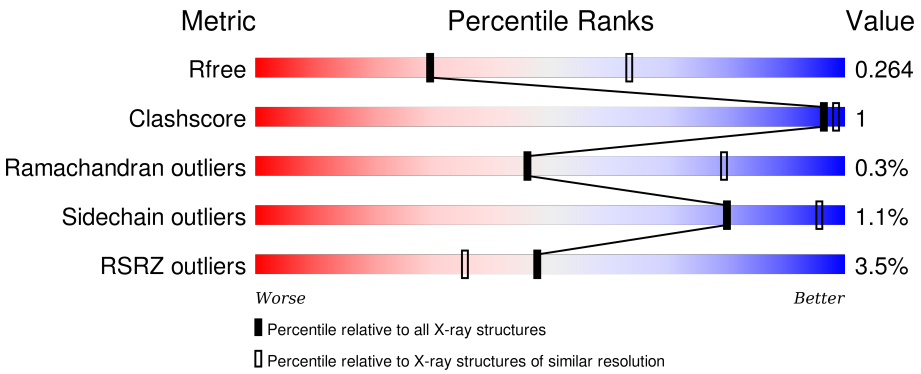
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.80 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	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	91	% 69% • • 26%
2	B	82	88% • 11%
3	C	145	% 81% 6% 13%
4	D	51	14% 57% • 41%
5	E	11	9% 82% 18%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	F	11	82% 18%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 2894 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 T-CELL ACUTE LYMPHOCYTIC LEUKEMIA PROTEIN 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	67	565	359	109	96	1	0	0	0

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	163	MET	-	EXPRESSION TAG	UNP P17542
A	164	GLY	-	EXPRESSION TAG	UNP P17542
A	165	SER	-	EXPRESSION TAG	UNP P17542
A	166	SER	-	EXPRESSION TAG	UNP P17542
A	167	HIS	-	EXPRESSION TAG	UNP P17542
A	168	HIS	-	EXPRESSION TAG	UNP P17542
A	169	HIS	-	EXPRESSION TAG	UNP P17542
A	170	HIS	-	EXPRESSION TAG	UNP P17542
A	171	HIS	-	EXPRESSION TAG	UNP P17542
A	172	HIS	-	EXPRESSION TAG	UNP P17542
A	173	SER	-	EXPRESSION TAG	UNP P17542
A	174	GLN	-	EXPRESSION TAG	UNP P17542
A	175	ASP	-	EXPRESSION TAG	UNP P17542
A	176	PRO	-	EXPRESSION TAG	UNP P17542
A	177	GLU	-	EXPRESSION TAG	UNP P17542
A	178	ILE	-	EXPRESSION TAG	UNP P17542
A	179	SER	-	EXPRESSION TAG	UNP P17542

- Molecule 2 is a protein called TRANSCRIPTION FACTOR E2-ALPHA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	73	610	369	128	109	4	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	532	MET	-	EXPRESSION TAG	UNP P15923
B	533	ALA	-	EXPRESSION TAG	UNP P15923
B	534	ASP	-	EXPRESSION TAG	UNP P15923

- Molecule 3 is a protein called RHOMBOTIN-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	126	1035	648	188	183	16	0	0	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	12	MET	-	EXPRESSION TAG	UNP P25791
C	13	GLY	-	EXPRESSION TAG	UNP P25791
C	14	SER	-	EXPRESSION TAG	UNP P25791
C	15	SER	-	EXPRESSION TAG	UNP P25791
C	16	HIS	-	EXPRESSION TAG	UNP P25791
C	17	HIS	-	EXPRESSION TAG	UNP P25791
C	18	HIS	-	EXPRESSION TAG	UNP P25791
C	19	HIS	-	EXPRESSION TAG	UNP P25791
C	20	HIS	-	EXPRESSION TAG	UNP P25791
C	21	HIS	-	EXPRESSION TAG	UNP P25791
C	22	SER	-	EXPRESSION TAG	UNP P25791
C	23	GLN	-	EXPRESSION TAG	UNP P25791
C	24	ASP	-	EXPRESSION TAG	UNP P25791

- Molecule 4 is a protein called LIM DOMAIN-BINDING PROTEIN 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	30	232	141	38	51	2	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	289	GLY	-	EXPRESSION TAG	UNP Q86U70
D	290	GLY	-	EXPRESSION TAG	UNP Q86U70
D	291	SER	-	EXPRESSION TAG	UNP Q86U70
D	292	GLY	-	EXPRESSION TAG	UNP Q86U70
D	293	GLY	-	EXPRESSION TAG	UNP Q86U70
D	294	HIS	-	EXPRESSION TAG	UNP Q86U70
D	295	MET	-	EXPRESSION TAG	UNP Q86U70

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	296	GLY	-	EXPRESSION TAG	UNP Q86U70
D	297	SER	-	EXPRESSION TAG	UNP Q86U70
D	298	GLY	-	EXPRESSION TAG	UNP Q86U70
D	299	GLY	-	EXPRESSION TAG	UNP Q86U70

- Molecule 5 is a DNA chain called EBOX FORWARD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	E	11	220	106	35	68	11	0	0	0

- Molecule 6 is a DNA chain called EBOX REVERSE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	F	11	228	109	47	62	10	0	0	0

- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

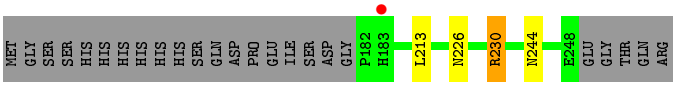
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
7	C	4	4	4	0	0

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.

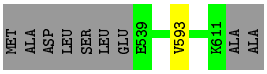
- Molecule 1: T-CELL ACUTE LYMPHOCYTIC LEUKEMIA PROTEIN 1

Chain A:



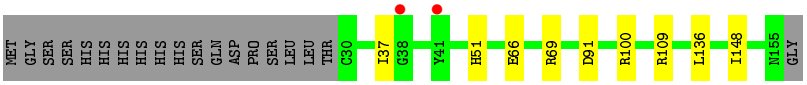
- Molecule 2: TRANSCRIPTION FACTOR E2-ALPHA

Chain B:



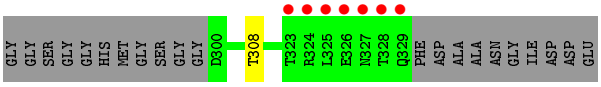
- Molecule 3: RHOMBOTIN-2

Chain C:



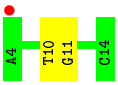
- Molecule 4: LIM DOMAIN-BINDING PROTEIN 1

Chain D:



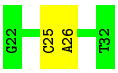
- Molecule 5: EBOX FORWARD

Chain E:



- Molecule 6: EBOX REVERSE

Chain F:


G22
C25
A26
T32

4 Data and refinement statistics (i)

Property	Value	Source
Space group	F 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	102.97Å 141.04Å 148.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.15 – 2.80 30.15 – 2.60	Depositor EDS
% Data completeness (in resolution range)	98.5 (30.15-2.80) 96.4 (30.15-2.60)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.45 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.224 , 0.269 0.223 , 0.264	Depositor DCC
R_{free} test set	640 reflections (5.04%)	DCC
Wilson B-factor (Å^2)	59.6	Xtriage
Anisotropy	0.314	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 44.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 16248 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2894	wwPDB-VP
Average B, all atoms (Å^2)	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.08% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.40	0/576	0.58	0/773
2	B	0.41	0/612	0.63	0/813
3	C	0.33	0/1054	0.56	0/1410
4	D	0.33	0/233	0.51	0/313
5	E	0.27	0/244	0.73	0/373
6	F	0.32	0/257	0.72	0/396
All	All	0.36	0/2976	0.61	0/4078

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 (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	565	0	590	2	0
2	B	610	0	639	1	0
3	C	1035	0	1002	4	0
4	D	232	0	219	1	0
5	E	220	0	126	1	0
6	F	228	0	125	1	0
7	C	4	0	0	0	0
All	All	2894	0	2701	8	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 1.

The worst 5 of 8 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)
3:C:91:ASP:OD2	3:C:100:ARG:NH1	2.36	0.58
5:E:10:DT:H2"	5:E:11:DG:C8	2.43	0.53
1:A:226:ASN:OD1	1:A:230:ARG:NH1	2.44	0.51
3:C:66:GLU:HB2	3:C:69:ARG:HG3	1.92	0.51
6:F:25:DC:H2"	6:F:26:DA:C8	2.47	0.49

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	65/91 (71%)	63 (97%)	2 (3%)	0	100	100
2	B	71/82 (87%)	69 (97%)	2 (3%)	0	100	100
3	C	124/145 (86%)	111 (90%)	12 (10%)	1 (1%)	24	58
4	D	28/51 (55%)	28 (100%)	0	0	100	100
All	All	288/369 (78%)	271 (94%)	16 (6%)	1 (0%)	46	79

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	37	ILE

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	62/83 (75%)	60 (97%)	2 (3%)	46	80
2	B	66/72 (92%)	66 (100%)	0	100	100
3	C	111/128 (87%)	110 (99%)	1 (1%)	84	96
4	D	26/37 (70%)	26 (100%)	0	100	100
All	All	265/320 (83%)	262 (99%)	3 (1%)	80	95

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	230	ARG
1	A	244	ASN
3	C	51	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	552	ASN
2	B	601	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)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	67/91 (73%)	-0.09	1 (1%) 76 68	55, 71, 121, 175	1 (1%)
2	B	73/82 (89%)	-0.12	0 100 100	46, 68, 143, 158	0
3	C	126/145 (86%)	-0.09	2 (1%) 74 66	70, 94, 144, 168	0
4	D	30/51 (58%)	0.72	7 (23%) 1 1	88, 111, 202, 205	0
5	E	11/11 (100%)	-0.44	1 (9%) 11 6	64, 69, 91, 145	0
6	F	11/11 (100%)	-0.75	0 100 100	56, 67, 104, 138	0
All	All	318/391 (81%)	-0.06	11 (3%) 48 35	46, 87, 150, 205	1 (0%)

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	325	LEU	5.2
4	D	329	GLN	4.6
4	D	328	THR	4.3
4	D	327	ASN	3.4
3	C	41	TYR	3.3

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(Å^2)	Q<0.9
7	ZN	C	1159	1/1	0.97	0.06	-1.46	107,107,107,107	0
7	ZN	C	1158	1/1	0.98	0.05	-2.32	78,78,78,78	0
7	ZN	C	1157	1/1	0.99	0.02	-2.68	79,79,79,79	0
7	ZN	C	1156	1/1	0.97	0.03	-5.54	91,91,91,91	0

6.5 Other polymers (i)

There are no such residues in this entry.