Full wwPDB X-ray Structure Validation Report (i)

May 12, 2020 – 11:57 pm BST

PDB ID	:	2YPB
Title	:	Structure of the SCL:E47 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	:	2.87 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references (1)) were used in the production of this report:

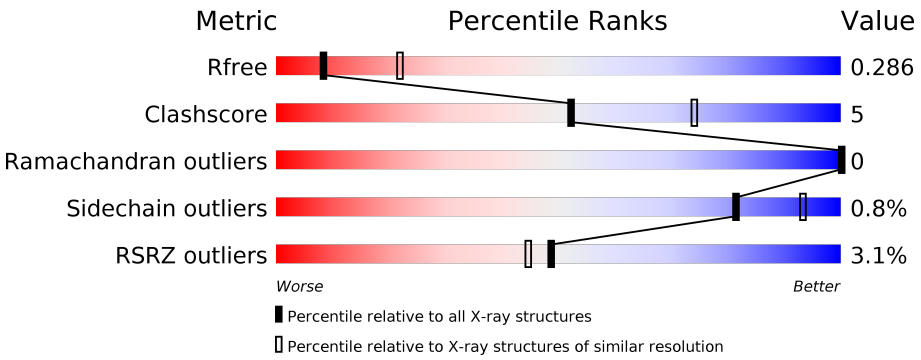
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.87 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	2691 (2.90-2.86)
Clashscore	141614	2947 (2.90-2.86)
Ramachandran outliers	138981	2868 (2.90-2.86)
Sidechain outliers	138945	2871 (2.90-2.86)
RSRZ outliers	127900	2629 (2.90-2.86)

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 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	A	91	2%, 65%, 9%, 26%
2	B	82	4%, 77%, 13%, 10%
3	E	11	55%, 36%, 9%
4	F	11	64%, 36%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1622 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	560	356	109	94	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	74	614	371	127	112	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 DNA chain called EBOX FORWARD.

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

- Molecule 4 is a DNA chain called EBOX REVERSE.

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

3 Residue-property plots (i)

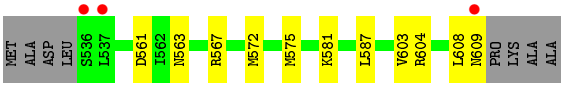
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: T-CELL ACUTE LYMPHOCYTIC LEUKEMIA PROTEIN 1

Chain A:

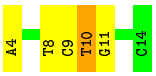
- Molecule 2: TRANSCRIPTION FACTOR E2-ALPHA

Chain B:



- Molecule 3: EBOX FORWARD

Chain E:



- Molecule 4: EBOX REVERSE

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	73.19Å 152.88Å 55.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.44 – 2.87 42.44 – 2.87	Depositor EDS
% Data completeness (in resolution range)	96.9 (42.44-2.87) 97.0 (42.44-2.87)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.254 , 0.288 0.249 , 0.286	Depositor DCC
R_{free} test set	331 reflections (4.60%)	wwPDB-VP
Wilson B-factor (Å^2)	91.5	Xtriage
Anisotropy	0.320	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 51.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	1622	wwPDB-VP
Average B, all atoms (Å^2)	117.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.97% 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 (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	0.31	0/571	0.39	0/767
2	B	0.28	0/615	0.41	0/817
3	E	0.62	0/244	1.24	1/373 (0.3%)
4	F	0.68	0/257	1.07	0/396
All	All	0.43	0/1687	0.74	1/2353 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	E	10	DT	C1'-O4'-C4'	-5.11	104.99	110.10

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	560	0	586	6	0
2	B	614	0	637	6	0
3	E	220	0	126	6	0
4	F	228	0	125	3	0
All	All	1622	0	1474	17	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 5.

All (17) 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:E:4:DA:H8	3:E:4:DA:P	2.23	0.60
2:B:572:MET:O	2:B:575:MET:HB2	2.11	0.51
3:E:8:DT:H2"	3:E:9:DC:O5'	2.10	0.51
3:E:10:DT:H2"	3:E:11:DG:C8	2.45	0.51
1:A:226:ASN:HD21	2:B:561:ASP:HB3	1.74	0.51
1:A:195:ARG:CZ	4:F:22:DG:H2'	2.41	0.51
4:F:31:DG:H2"	4:F:32:DT:OP2	2.14	0.48
1:A:197:ARG:NH1	3:E:9:DC:OP2	2.47	0.47
1:A:193:ASN:HD21	3:E:9:DC:H3'	1.80	0.46
2:B:608:LEU:HA	2:B:609:ASN:HA	1.71	0.46
1:A:235:TYR:O	1:A:239:LEU:HG	2.16	0.45
1:A:214:ILE:HA	1:A:215:PRO:HD3	1.86	0.44
2:B:581:LYS:HD2	2:B:587:LEU:HD21	2.02	0.42
3:E:4:DA:C8	3:E:4:DA:P	3.10	0.41
2:B:604:ARG:O	2:B:608:LEU:HD23	2.21	0.40
2:B:563:ASN:O	2:B:567:ARG:HG3	2.21	0.40
4:F:30:DG:H2"	4:F:31:DG:OP2	2.22	0.40

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%)	64 (98%)	1 (2%)	0	100	100
2	B	72/82 (88%)	72 (100%)	0	0	100	100
All	All	137/173 (79%)	136 (99%)	1 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	61/83 (74%)	61 (100%)	0	100	100
2	B	66/72 (92%)	65 (98%)	1 (2%)	65	86
All	All	127/155 (82%)	126 (99%)	1 (1%)	81	93

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	603	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	226	ASN
2	B	552	ASN
2	B	590	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 (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%)	0.17	2 (2%)	50	46	64, 103, 206, 259	1 (1%)
2	B	74/82 (90%)	0.28	3 (4%)	37	32	74, 125, 213, 264	0
3	E	11/11 (100%)	-0.59	0	100	100	69, 87, 121, 187	0
4	F	11/11 (100%)	-0.67	0	100	100	62, 76, 122, 179	0
All	All	163/195 (83%)	0.11	5 (3%)	49	45	62, 104, 207, 264	1 (0%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	536	SER	3.7
2	B	537	LEU	3.5
1	A	222	LYS	2.7
2	B	609	ASN	2.5
1	A	246	GLN	2.4

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.