wwPDB X-ray Structure Validation Summary Report (i)

Apr 11, 2016 – 05:10 PM EDT

PDB ID	:	5J3E
Title	:	Crystal Structure of Human THYN1 protein in complex with 5-methylcytosine containing DNA
Authors	:	Halabelian, L.; Tempel, W.; Li, Y.; Bountra, C.; Edwards, A.M.; Arrowsmith, C.H.; Structural Genomics Consortium (SGC)
Deposited on	:	2016-03-30
Resolution	:	2.60 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:

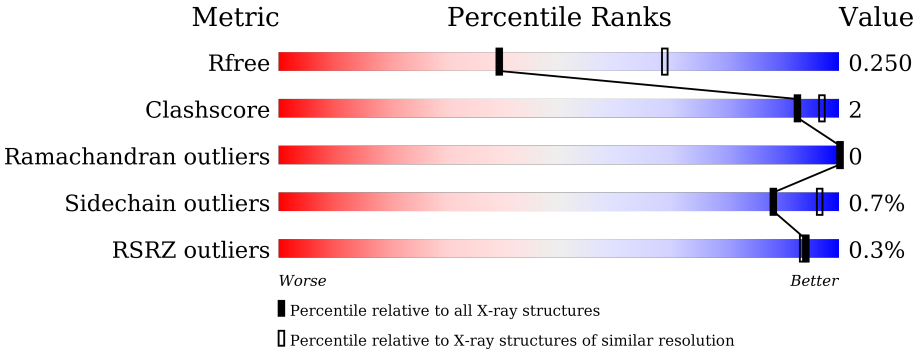
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027107
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027107

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 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	225	76% 24%
1	B	225	74% . 24%
2	C	12	83% 17%
2	D	12	67% 25% 8%
2	E	12	100%
2	F	12	92% 8%

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	UNX	A	303	-	-	-	X

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3764 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 Thymocyte nuclear protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	172	1372	876	234	253	9	0	0	0
1	B	171	1356	866	232	249	9	0	0	0

- Molecule 2 is a DNA chain called 5-methylcytosine containing DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	12	244	117	46	70	11	0	0	0
2	D	12	286	137	53	83	13	0	2	0
2	E	12	244	117	46	70	11	0	0	0
2	F	12	244	117	46	70	11	0	0	0

- Molecule 3 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	X		
3	B	5	5	5	0	0
3	A	3	3	3	0	0

- Molecule 4 is water.

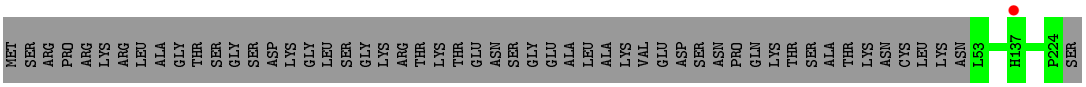
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	6	6	6	0	0
4	B	4	4	4	0	0

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.

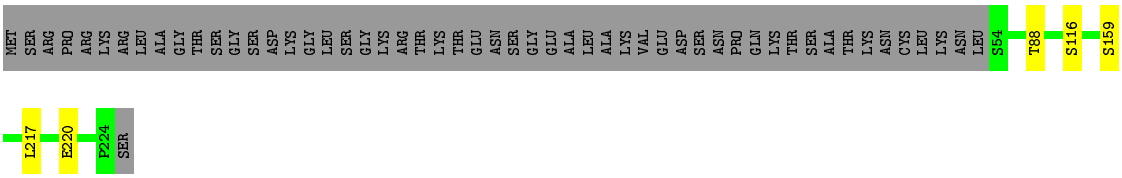
- Molecule 1: Thymocyte nuclear protein 1

Chain A:



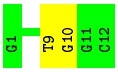
- Molecule 1: Thymocyte nuclear protein 1

Chain B:



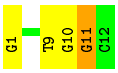
- Molecule 2: 5-methylcytosine containing DNA

Chain C:



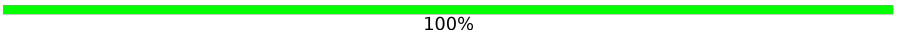
- Molecule 2: 5-methylcytosine containing DNA

Chain D:



- Molecule 2: 5-methylcytosine containing DNA

Chain E:



There are no outlier residues recorded for this chain.

- Molecule 2: 5-methylcytosine containing DNA

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	40.33Å 73.23Å 108.20Å 90.00° 91.18° 90.00°	Depositor
Resolution (Å)	43.51 – 2.60 43.51 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.8 (43.51-2.60) 99.8 (43.51-2.60)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.25 (at 2.61Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.209 , 0.249 0.214 , 0.250	Depositor DCC
R_{free} test set	941 reflections (5.07%)	DCC
Wilson B-factor (Å^2)	44.7	Xtriage
Anisotropy	0.552	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 24.1	EDS
Estimated twinning fraction	0.048 for h,-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	7 of 19500 reflections (0.036%)	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3764	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 19.21% 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: UNX, 5CM

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	0.82	0/1409	0.62	0/1906
1	B	0.89	0/1393	0.68	0/1887
2	C	0.60	0/250	0.75	0/382
2	D	0.71	1/297 (0.3%)	0.90	1/455 (0.2%)
2	E	0.71	0/250	0.86	0/382
2	F	0.74	1/250 (0.4%)	0.77	0/382
All	All	0.82	2/3849 (0.1%)	0.71	1/5394 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	3	DC	O3'-P	-5.80	1.54	1.61
2	D	1	DG	O3'-P	5.19	1.67	1.61

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	11	DG	O5'-P-OP2	-5.77	100.51	105.70

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	1372	0	1299	0	0
1	B	1356	0	1272	2	0
2	C	244	0	138	1	0
2	D	286	0	161	12	0
2	E	244	0	138	0	0
2	F	244	0	138	0	0
3	A	3	0	0	0	0
3	B	5	0	0	0	0
4	A	6	0	0	0	0
4	B	4	0	0	0	0
All	All	3764	0	3146	14	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 2.

The worst 5 of 14 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)
2:D:9[A]:DT:H4'	2:D:10[A]:DG:OP1	1.38	1.10
2:D:9[A]:DT:H2"	2:D:10[A]:DG:C8	2.06	0.89
2:D:9[A]:DT:C4'	2:D:10[A]:DG:OP1	2.30	0.66
2:D:9[A]:DT:C2'	2:D:10[A]:DG:C8	2.80	0.62
2:D:9[B]:DT:H2"	2:D:10[B]:DG:C8	2.37	0.60

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	170/225 (76%)	164 (96%)	6 (4%)	0	100	100
1	B	169/225 (75%)	165 (98%)	4 (2%)	0	100	100
All	All	339/450 (75%)	329 (97%)	10 (3%)	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	144/200 (72%)	144 (100%)	0	100	100
1	B	140/200 (70%)	138 (99%)	2 (1%)	74	90
All	All	284/400 (71%)	282 (99%)	2 (1%)	88	96

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	88	THR
1	B	159	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

4 non-standard protein/DNA/RNA residues are modelled in this entry.

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	$\#\|Z\| > 2$
2	5CM	C	6	2	13,21,22	1.57	1 (7%)	17,30,33	0.92	1 (5%)
2	5CM	D	6	2	13,21,22	1.49	1 (7%)	17,30,33	1.04	1 (5%)
2	5CM	E	6	2	13,21,22	1.97	1 (7%)	17,30,33	1.64	2 (11%)
2	5CM	F	6	2	13,21,22	1.34	1 (7%)	17,30,33	1.79	2 (11%)

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
2	5CM	C	6	2	-	0/3/21/22	0/2/2/2
2	5CM	D	6	2	-	0/3/21/22	0/2/2/2
2	5CM	E	6	2	-	0/3/21/22	0/2/2/2
2	5CM	F	6	2	-	0/3/21/22	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	6	5CM	C5-C4	4.52	1.48	1.41
2	D	6	5CM	C5-C4	4.91	1.49	1.41
2	C	6	5CM	C5-C4	5.33	1.49	1.41
2	E	6	5CM	C5-C4	6.22	1.51	1.41

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	6	5CM	C2'-C1'-N1	-3.29	106.02	114.14
2	F	6	5CM	C4'-O4'-C1'	-2.50	103.01	109.42
2	C	6	5CM	O4'-C1'-N1	2.06	111.32	107.71
2	D	6	5CM	O4'-C1'-N1	2.85	112.70	107.71
2	E	6	5CM	O4'-C1'-N1	4.49	115.56	107.71

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 8 are unknown - 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	172/225 (76%)	-0.40	1 (0%)	90	88	35, 53, 79, 96	0
1	B	171/225 (76%)	-0.45	0	100	100	35, 53, 71, 85	0
2	C	11/12 (91%)	-0.59	0	100	100	61, 65, 71, 72	0
2	D	11/12 (91%)	-0.39	0	100	100	62, 67, 74, 74	0
2	E	11/12 (91%)	-0.50	0	100	100	47, 52, 69, 71	0
2	F	11/12 (91%)	-0.67	0	100	100	57, 60, 62, 64	0
All	All	387/498 (77%)	-0.44	1 (0%)	94	93	35, 55, 75, 96	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	137	HIS	2.3

6.2 Non-standard residues in protein, DNA, RNA chains (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
2	5CM	F	6	20/21	0.94	0.16	-	48,55,66,68	0
2	5CM	C	6	20/21	0.96	0.13	-	55,60,68,69	0
2	5CM	E	6	20/21	0.92	0.16	-	61,74,90,92	0
2	5CM	D	6	20/21	0.94	0.18	-	64,74,85,88	0

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
3	UNX	A	303	1/1	0.87	0.22	3.02	29,29,29,29	0
3	UNX	A	301	1/1	0.81	0.35	-	43,43,43,43	0
3	UNX	B	301	1/1	0.44	0.39	-	46,46,46,46	0
3	UNX	B	305	1/1	0.71	0.24	-	38,38,38,38	0
3	UNX	B	302	1/1	0.75	0.49	-	38,38,38,38	0
3	UNX	B	304	1/1	0.78	0.14	-	30,30,30,30	0
3	UNX	B	303	1/1	0.88	0.26	-	24,24,24,24	0
3	UNX	A	302	1/1	0.79	0.34	-	27,27,27,27	0

6.5 Other polymers (i)

There are no such residues in this entry.