wwPDB X-ray Structure Validation Summary Report (i)

Mar 13, 2023 – 12:52 pm GMT

PDB ID	:	7Z5K
Title	:	Transcription factor MYF5 bound to non-symmetrical site
Authors	:	Morgunova, E.; Popov, A.; Yin, Y.; Taipale, J.
Deposited on	:	2022-03-09
Resolution	:	2.28 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

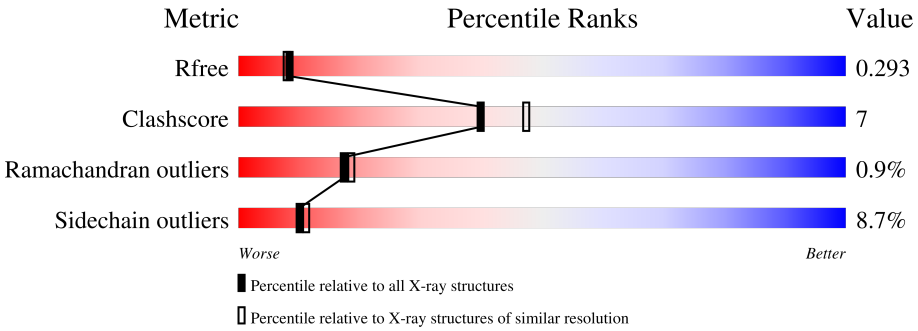
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.32.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.1

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 2.28 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	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)

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 of 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%

Mol	Chain	Length	Quality of chain
1	A	57	91% 7% .
1	B	57	65% 25% 9% .
2	E	18	50% 50%
3	F	18	72% 28%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1775 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 Myogenic factor 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	56	Total	C	N	O	S	0	1	0
			477	290	99	84	4			
1	B	56	Total	C	N	O	S	0	1	0
			484	296	102	83	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	81	SER	-	expression tag	UNP P13349
B	81	SER	-	expression tag	UNP P13349

- Molecule 2 is a DNA chain called DNA (5'-D(P*GP*CP*GP*CP*AP*AP*CP*AP*GP*CP*TP*GP*AP*CP*GP*CP*GP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	18	Total	C	N	O	P	0	0	0
			370	174	72	106	18			

- Molecule 3 is a DNA chain called DNA (5'-D(P*AP*CP*GP*CP*GP*TP*CP*AP*GP*CP*TP*GP*TP*TP*GP*CP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	18	Total	C	N	O	P	0	0	0
			368	174	66	110	18			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	26	Total	O	0	0
			26	26		
4	B	22	Total	O	0	0
			22	22		

Continued on next page...

Continued from previous page...

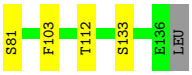
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	15	Total O 15 15	0	0
4	F	13	Total O 13 13	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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.

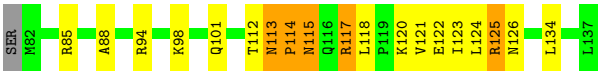
• Molecule 1: Myogenic factor 5

Chain A:



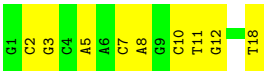
• Molecule 1: Myogenic factor 5

Chain B:



• Molecule 2: DNA (5'-D(P*GP*CP*GP*CP*AP*AP*CP*AP*GP*CP*TP*GP*AP*CP*GP*CP*GP*T)-3')

Chain E:



• Molecule 3: DNA (5'-D(P*AP*CP*GP*CP*GP*TP*CP*AP*GP*CP*TP*GP*TP*TP*GP*CP*GP*C)-3')

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	170.44Å 33.84Å 53.54Å 90.00° 91.88° 90.00°	Depositor
Resolution (Å)	46.00 – 2.28 46.00 – 1.99	Depositor EDS
% Data completeness (in resolution range)	96.4 (46.00-2.28) 88.4 (46.00-1.99)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.11 (at 1.98Å)	Xtriage
Refinement program	PHENIX 1.13_2998	Depositor
R, R_{free}	0.229 , 0.266 0.256 , 0.293	Depositor DCC
R_{free} test set	927 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å^2)	52.6	Xtriage
Anisotropy	0.238	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 49.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.005 for -h,-k,l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	1775	wwPDB-VP
Average B, all atoms (Å^2)	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.57% 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

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	0.31	0/480	0.47	0/639
1	B	0.38	0/487	0.53	0/648
2	E	0.54	0/415	0.81	0/638
3	F	0.55	0/411	0.90	0/632
All	All	0.45	0/1793	0.70	0/2557

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	125[A]	ARG	Mainchain
1	B	125[B]	ARG	Mainchain

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	477	0	506	1	0
1	B	484	0	520	12	0
2	E	370	0	201	7	0
3	F	368	0	203	3	0
4	A	26	0	0	1	0
4	B	22	0	0	1	0
4	E	15	0	0	2	0
4	F	13	0	0	1	0
All	All	1775	0	1430	23	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 7.

The worst 5 of 23 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:E:5:DA:N7	4:E:103:HOH:O	2.25	0.68
3:F:15:DG:N3	4:F:101:HOH:O	2.29	0.66
1:B:122:GLU:HA	1:B:125[A]:ARG:HD2	1.80	0.64
1:B:112:THR:O	1:B:112:THR:HG22	2.03	0.58
1:B:85:ARG:NH1	1:B:88:ALA:HB3	2.22	0.55

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	55/57 (96%)	55 (100%)	0	0	100 100
1	B	55/57 (96%)	46 (84%)	8 (14%)	1 (2%)	8 7
All	All	110/114 (96%)	101 (92%)	8 (7%)	1 (1%)	17 18

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	114	PRO

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	53/53 (100%)	50 (94%)	3 (6%)	20 26
1	B	53/53 (100%)	47 (89%)	6 (11%)	6 5
All	All	106/106 (100%)	97 (92%)	9 (8%)	10 12

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	117	ARG
1	B	134	LEU
1	B	94	ARG
1	B	101	GLN
1	B	113	ASN

Sometimes 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)

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

5.5 Carbohydrates (i)

There are no monosaccharides 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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.