wwPDB X-ray Structure Validation Summary Report (i)

Oct 8, 2023 – 04:59 PM EDT

PDB ID	:	6MBZ
Title	:	Structure of Transcription Factor
Authors	:	Seo, H.-S.; Dhe-Paganon, S.
Deposited on	:	2018-08-30
Resolution	:	3.21 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 (1)) were used in the production of this report:

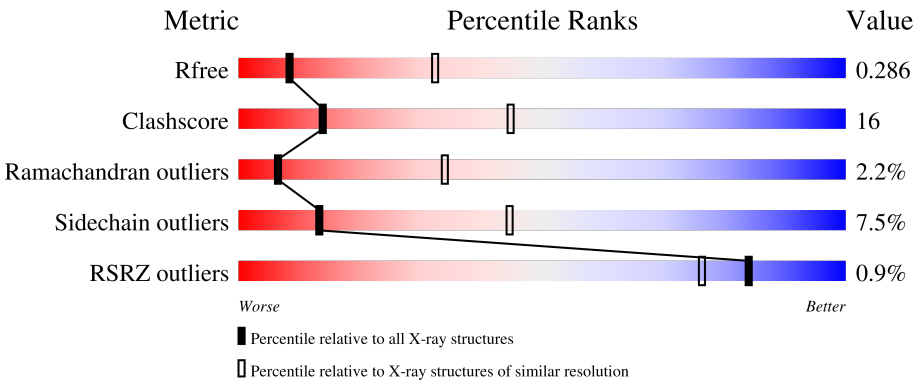
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.21 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	1335 (3.24-3.20)
Clashscore	141614	1460 (3.24-3.20)
Ramachandran outliers	138981	1437 (3.24-3.20)
Sidechain outliers	138945	1436 (3.24-3.20)
RSRZ outliers	127900	1291 (3.24-3.20)

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% 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	572	% 61% 29% • 6%
1	B	572	59% 30% • 9%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7933 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 Signal transducer and activator of transcription 5B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	536	4104	2620	709	763	12	0	0	0
1	B	522	3793	2422	657	702	12	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	132	SER	-	expression tag	UNP P51692
A	133	THR	-	expression tag	UNP P51692
A	134	GLY	-	expression tag	UNP P51692
A	135	SER	-	expression tag	UNP P51692
A	642	HIS	ASN	engineered mutation	UNP P51692
B	132	SER	-	expression tag	UNP P51692
B	133	THR	-	expression tag	UNP P51692
B	134	GLY	-	expression tag	UNP P51692
B	135	SER	-	expression tag	UNP P51692
B	642	HIS	ASN	engineered mutation	UNP P51692

- Molecule 2 is water.

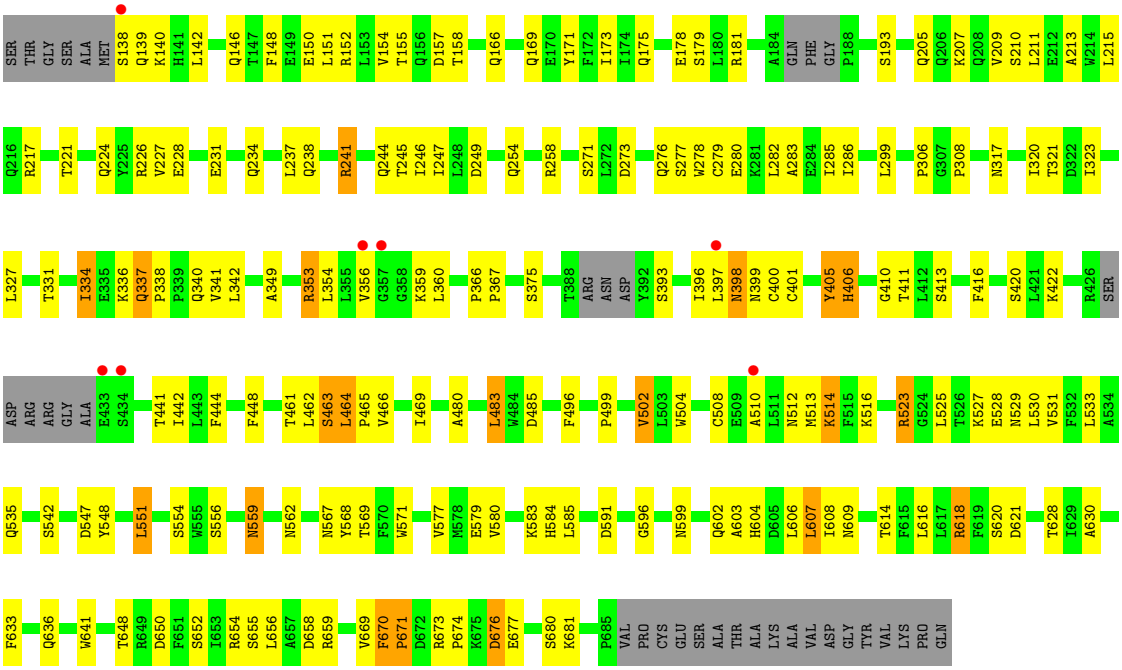
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	20	20	20	0	0
2	B	16	16	16	0	0

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 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.

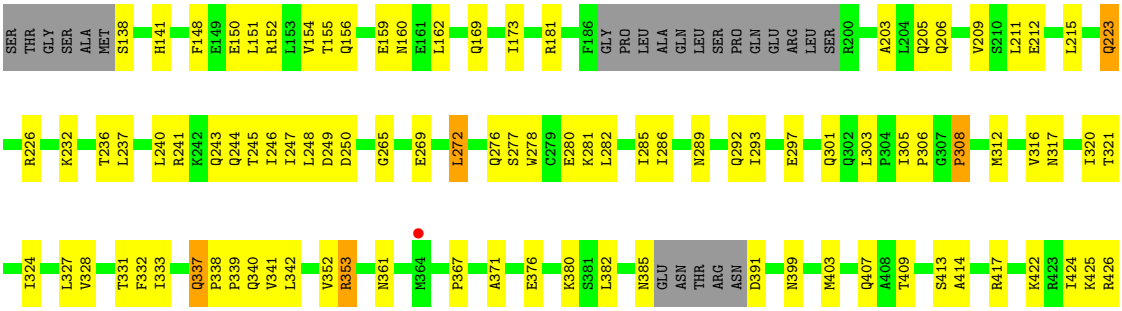
• Molecule 1: Signal transducer and activator of transcription 5B

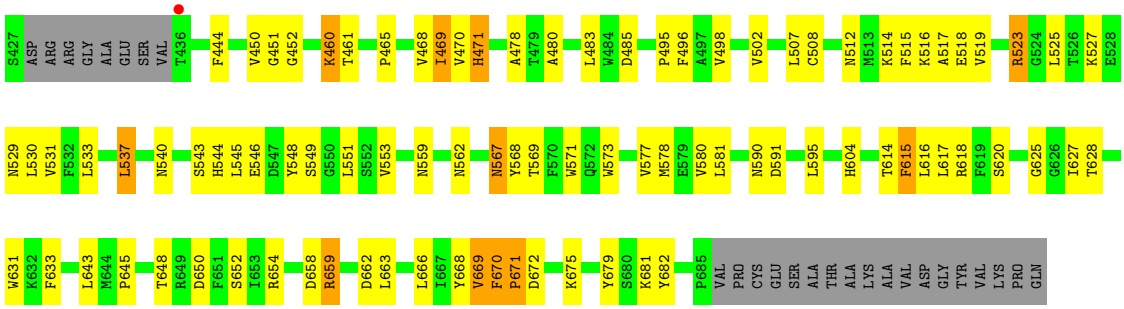
Chain A:



• Molecule 1: Signal transducer and activator of transcription 5B

Chain B:




S427 ASP ARG ARG GLY ALA GLU SER VAL T436 F444 V450 G451 G452 K460 T461 P465 V468 I469 V470 H471 A478 T479 A480 L483 W484 D485 P495 F496 A497 V498 V502 L507 C508 N512 M513 K514 F515 K516 A517 E518 V519 R523 G524 L525 T526 K527 E528
N529 L530 V531 F532 L533 L537 N540 S543 H544 L545 E546 D547 Y548 S549 G550 L551 S552 V553 N559 N562 N567 Y568 T569 F570 W571 Q572 W573 V577 M578 E579 V580 L581 N590 D591 L595 H604 T614 F615 L616 L617 R618 F619 S620 G625 G626 I627 T628
W631 K632 F633 L643 M644 P645 T648 R649 D650 F651 S652 I653 R654 D658 R659 D662 L663 L666 I667 Y668 V669 F670 P671 D672 K675 Y679 S680 K681 Y682 P685 VAL PRO CYS GLU SER ALA THR ALA LYS ALA VAL ASP GLY TYR VAL LYS PRO GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	132.38Å 141.40Å 72.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	70.70 – 3.21 70.70 – 3.21	Depositor EDS
% Data completeness (in resolution range)	99.9 (70.70-3.21) 99.9 (70.70-3.21)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.83 (at 3.19Å)	Xtriage
Refinement program	PHENIX (1.14_3260: ???)	Depositor
R, R_{free}	0.265 , 0.287 0.265 , 0.286	Depositor DCC
R_{free} test set	1143 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	95.4	Xtriage
Anisotropy	0.260	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 57.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	7933	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.68% 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	0.33	1/4194 (0.0%)	0.53	2/5715 (0.0%)
1	B	0.30	0/3874	0.53	0/5299
All	All	0.31	1/8068 (0.0%)	0.53	2/11014 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	464	LEU	C-N	8.09	1.49	1.34

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	337	GLN	CB-CA-C	-5.96	98.48	110.40
1	A	607	LEU	CA-CB-CG	5.87	128.81	115.30

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	4104	0	3863	123	0
1	B	3793	0	3400	117	0
2	A	20	0	0	5	0
2	B	16	0	0	5	0
All	All	7933	0	7263	240	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 16.

The worst 5 of 240 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:663:LEU:O	1:B:666:LEU:HD12	1.48	1.11
1:A:334:ILE:HD12	1:A:334:ILE:H	1.22	1.04
1:B:666:LEU:O	1:B:675:LYS:NZ	1.97	0.96
1:B:663:LEU:O	1:B:666:LEU:CD1	2.17	0.92
1:A:210:SER:O	2:A:801:HOH:O	1.88	0.92

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	528/572 (92%)	488 (92%)	30 (6%)	10 (2%)	8	38
1	B	514/572 (90%)	468 (91%)	33 (6%)	13 (2%)	5	31
All	All	1042/1144 (91%)	956 (92%)	63 (6%)	23 (2%)	6	34

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	306	PRO
1	A	502	VAL
1	A	568	TYR
1	A	636	GLN
1	A	670	PHE

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	416/510 (82%)	386 (93%)	30 (7%)	14	46
1	B	349/510 (68%)	322 (92%)	27 (8%)	13	43
All	All	765/1020 (75%)	708 (92%)	57 (8%)	13	44

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	652	SER
1	B	643	LEU
1	B	272	LEU
1	B	633	PHE
1	B	591	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	562	ASN

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 (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	536/572 (93%)	-0.09	7 (1%)	77	66	59, 86, 119, 139	0
1	B	522/572 (91%)	-0.26	2 (0%)	92	89	53, 85, 113, 129	0
All	All	1058/1144 (92%)	-0.18	9 (0%)	84	76	53, 86, 116, 139	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	138	SER	4.5
1	A	433	GLU	3.8
1	A	357	GLY	2.8
1	B	436	THR	2.4
1	A	356	VAL	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 monosaccharides 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.