



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 24, 2022 – 12:34 pm GMT

PDB ID : 5O9I
Title : Crystal structure of transcription factor IIB Mvu mini-intein
Authors : Mikula, K.M.; Iwai, H.; Li, M.; Wlodawer, A.
Deposited on : 2017-06-19
Resolution : 2.50 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.27
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.27

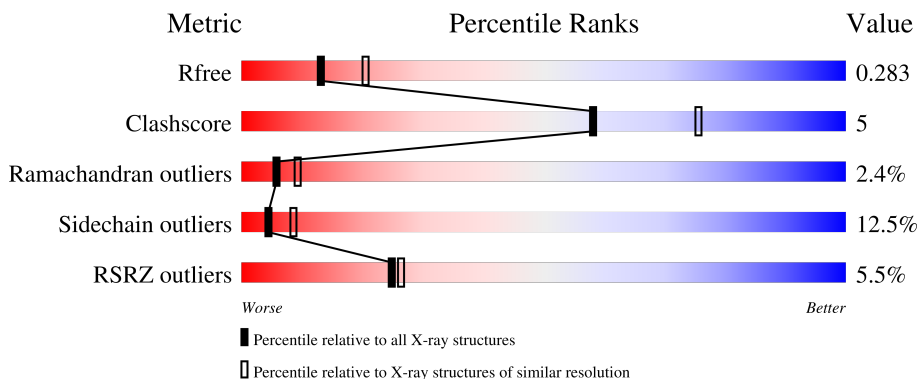
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 $\leq 5\%$. 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	195	 8% 70% 15% 2% 5%
1	B	195	 2% 69% 17% 2% 10%

2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 2773 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 Transcription initiation factor IIB, Transcription initiation factor IIB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	171	1359	878	225	253	3	0	0	0
1	B	176	1398	904	231	260	3	0	0	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	SER	-	expression tag	UNP C9REA0
A	-2	GLY	-	expression tag	UNP C9REA0
A	-1	GLY	-	expression tag	UNP C9REA0
A	0	TYR	-	expression tag	UNP C9REA0
A	1	ALA	-	expression tag	UNP C9REA0
A	129	ASN	-	linker	UNP C9REA0
A	130	ARG	-	linker	UNP C9REA0
A	131	LYS	-	linker	UNP C9REA0
A	132	LEU	-	linker	UNP C9REA0
A	133	GLU	-	linker	UNP C9REA0
A	137	GLU	TYR	conflict	UNP C9REA0
A	138	GLU	ASN	conflict	UNP C9REA0
A	191	ALA	-	expression tag	UNP C9REA0
B	-3	SER	-	expression tag	UNP C9REA0
B	-2	GLY	-	expression tag	UNP C9REA0
B	-1	GLY	-	expression tag	UNP C9REA0
B	0	TYR	-	expression tag	UNP C9REA0
B	1	ALA	-	expression tag	UNP C9REA0
B	129	ASN	-	linker	UNP C9REA0
B	130	ARG	-	linker	UNP C9REA0
B	131	LYS	-	linker	UNP C9REA0
B	132	LEU	-	linker	UNP C9REA0
B	133	GLU	-	linker	UNP C9REA0
B	137	GLU	TYR	conflict	UNP C9REA0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	138	GLU	ASN	conflict	UNP C9REA0
B	191	ALA	-	expression tag	UNP C9REA0

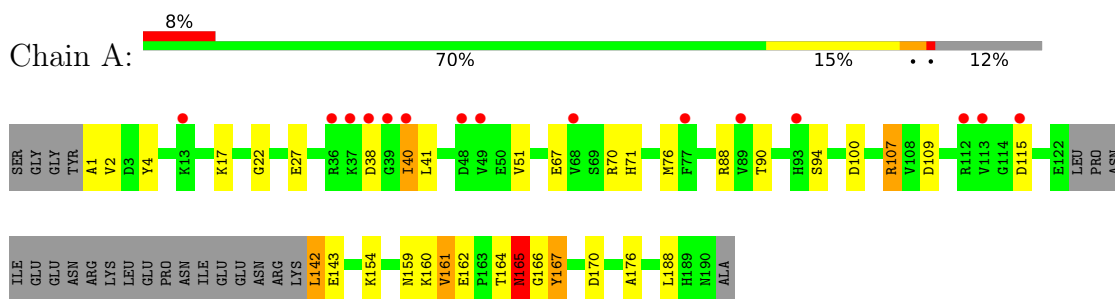
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	4	Total O 4 4	0	0
2	B	12	Total O 12 12	0	0

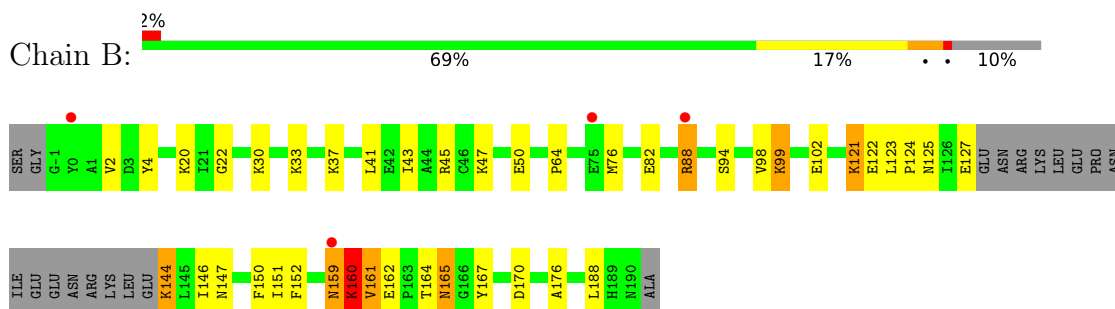
3 Residue-property plots [i](#)

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: Transcription initiation factor IIB, Transcription initiation factor IIB



- Molecule 1: Transcription initiation factor IIB, Transcription initiation factor IIB



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	108.18Å 67.25Å 51.23Å 90.00° 102.43° 90.00°	Depositor
Resolution (Å)	56.70 – 2.50 56.73 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (56.70-2.50) 99.5 (56.73-2.50)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.73 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.222 , 0.278 0.224 , 0.283	Depositor DCC
R_{free} test set	580 reflections (4.67%)	wwPDB-VP
Wilson B-factor (Å ²)	65.0	Xtriage
Anisotropy	0.224	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	2773	wwPDB-VP
Average B, all atoms (Å ²)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.92% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.88	3/1380 (0.2%)	0.98	0/1858
1	B	0.90	1/1421 (0.1%)	0.99	2/1915 (0.1%)
All	All	0.89	4/2801 (0.1%)	0.99	2/3773 (0.1%)

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	A	0	1
1	B	0	1
All	All	0	2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	165	ASN	CG-OD1	7.15	1.39	1.24
1	A	143	GLU	CG-CD	5.39	1.60	1.51
1	B	94	SER	CB-OG	5.35	1.49	1.42
1	A	143	GLU	CB-CG	5.07	1.61	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	B	170	ASP	CB-CG-OD1	5.33	123.09	118.30
1	B	88	ARG	NE-CZ-NH1	5.00	122.80	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	142	LEU	Peptide
1	B	144	LYS	Peptide

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	1359	0	1408	13	1
1	B	1398	0	1444	17	1
2	A	4	0	0	0	0
2	B	12	0	0	0	0
All	All	2773	0	2852	27	1

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.

The worst 5 of 27 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:A:1:ALA:HB1	1:A:90:THR:HB	1.72	0.71
1:A:167:TYR:CZ	1:B:45:ARG:HD3	2.27	0.69
1:B:165:ASN:O	1:B:165:ASN:ND2	2.27	0.67
1:A:165:ASN:O	1:A:167:TYR:N	2.28	0.66
1:B:98:VAL:O	1:B:99:LYS:HD2	2.03	0.59

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:100:ASP:OD1	1:B:125:ASN:ND2[4_547]	2.04	0.16

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	167/195 (86%)	158 (95%)	5 (3%)	4 (2%)	6	9
1	B	172/195 (88%)	159 (92%)	9 (5%)	4 (2%)	6	10
All	All	339/390 (87%)	317 (94%)	14 (4%)	8 (2%)	6	9

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	166	GLY
1	A	165	ASN
1	B	160	LYS
1	A	176	ALA
1	B	176	ALA

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	154/175 (88%)	134 (87%)	20 (13%)	4	7
1	B	158/175 (90%)	139 (88%)	19 (12%)	5	9
All	All	312/350 (89%)	273 (88%)	39 (12%)	4	8

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	99	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	164	THR
1	B	102	GLU
1	B	127	GLU
1	B	167	TYR

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

Mol	Chain	Res	Type
1	A	101	ASN
1	B	159	ASN
1	B	165	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	171/195 (87%)	0.59	15 (8%) 10 10	46, 91, 154, 168	0
1	B	176/195 (90%)	0.02	4 (2%) 60 63	43, 68, 100, 125	0
All	All	347/390 (88%)	0.30	19 (5%) 25 26	43, 77, 136, 168	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	39	GLY	8.4
1	A	113	VAL	5.9
1	A	38	ASP	5.6
1	A	40	ILE	4.4
1	B	0	TYR	4.4

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.