



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 4, 2026 – 07:15 PM UTC

PDB ID : 4Y5U / pdb_00004y5u
Title : Transcription factor
Authors : Li, J.; Niu, F.; Ouyang, S.; Liu, Z.
Deposited on : 2015-02-12
Resolution : 2.71 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

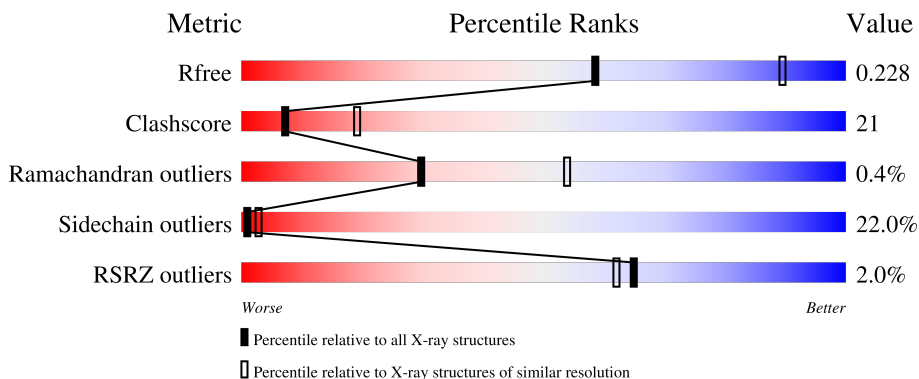
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.71 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}	180053	3538 (2.70-2.70)
Clashscore	190562	3843 (2.70-2.70)
Ramachandran outliers	187476	3778 (2.70-2.70)
Sidechain outliers	187428	3778 (2.70-2.70)
RSRZ outliers	180081	3538 (2.70-2.70)

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	549	22% 38% 17% • 19%
1	B	549	22% 39% 15% • 20%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7227 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 6.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	445	Total	C	N	O	P	S	0	0	0
			3550	2268	622	645	1	14			
1	B	438	Total	C	N	O	P	S	0	0	0
			3500	2240	611	634	1	14			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	SER	-	expression tag	UNP P42226
A	111	ASN	-	expression tag	UNP P42226
A	112	ALA	-	expression tag	UNP P42226
B	110	SER	-	expression tag	UNP P42226
B	111	ASN	-	expression tag	UNP P42226
B	112	ALA	-	expression tag	UNP P42226

- Molecule 2 is NICKEL (II) ION (CCD ID: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Ni	0	0
			1	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	98	Total	O	0	0
			98	98		
3	B	78	Total	O	0	0
			78	78		

R626	SER	G557	H494	L427	L365	L302	Q237	ALA	SER
HIS	T558	R495	W428	W429	K366	G303	L238	ASN	ASN
THR	F559	S496	D429	M430	LYS	A304	Q239	GLY	ALA
LYS	L560	V497	M431	A431	ILE	P305	Q240	THR	GLN
PRO	L561	S498	A432	F432	LYS	A306	E241	GLY	PHE
GLU	R562	W499	S433	S433	ARG	K307	V242	ARG	ARG
GLN	S563	S500	E434	E434	CYS	L310	GLY	SER	HIS
MET	S564	Q501	E435	E435	GLU	V311	ALA	GLU	LEU
GLY	F502	M435	D436	D436	ARG	A312	ALA	PRO	PRO
LYS	S566	L503	R437	R437	LYS	A313	GLY	MET	GLN
ASP	E567	K504	V438	V438	THR	D314	GLY	PRO	GLU
R638	G570	E505	P439	P439	SER	K315	LEU	PHE	HIS
R639	I571	L507	E440	E440	VAL	V316	GLU	TRP	LYS
G640	T572	L508	V441	V441	THR	K319	PRO	LYS	GLN
V642	I573	G509	V442	V442	LYS	Q220	THR	GLU	GLU
F643	A574	R310	A443	A443	E381	A321	ARG	LEU	LYS
A644	H575	G511	E444	E444	K383	R322	A255	LEU	LEU
T645	V576	F512	R445	R445	C384	E323	S256	F130	F130
T646	I577	T513	V446	V446	A385	LEU	L257	K131	K131
R647	R578	F514	W448	W448	V386	SER	R260	T132	T132
M648	P584	Q516	E449	E449	VAL	VAL	L261	G133	G133
T649	W517	W517	K450	K450	PRO	PRO	L262	L134	L134
V650	F518	D519	M451	M451	A390	GLN	E263	R135	R135
E651	E587	G520	C452	C452	S391	GLY	V264	R136	R136
ARG	N588	L522	T453	T453	F392	PRO	L265	L137	L137
ASP	I589	V521	T454	T454	L393	GLY	R266	Q138	Q138
GLN	Q590	L522	L455	L455	L394	ALA	K199	H139	H139
PRO	P591	D523	M456	M456	GLY	GLY	S271	R140	R140
LEU	F592	L524	L457	L457	PRO	ALA	C272	V141	V141
PRO	S593	T525	K458	K458	GLY	GLU	F273	G142	G142
THR	A594	K526	F459	F459	LYS	SER	L274	E143	E143
	K595	R527	M460	M460	L399	THR	V275	I144	I144
		C528	A461	A461	P400	GLY	K205	H145	H145
		L529	E462	E462	I401	E339	R206	L146	L146
	S598	R530	Q402	Q402	Q402	I340	Q207	L147	L147
	I599	S531	L403	L403	L403	T341	Q208	R148	R148
	R600	Y532	T465	T465	Q404	N342	Q209	E149	E149
	D604	W533	M466	M466	A405	N343	L210	A150	A150
	R605	S534	R467	R467	L406	T344	V282	L151	L151
	I606	D535	G468	G468	V345	V345	L283	GLN	GLN
	R607	R536	L469	L469	P409	P346	K284	LYS	LYS
	D608	L537	L470	L470	L410	L347	T285	ALA	ALA
	L609	I538	P471	P471	V412	E348	Q286	GLU	GLU
	A610	F541	E472	E472	V412	I413	T287	ALA	ALA
	Q611	I542	H473	H473	N349	S350	K288	ALA	ALA
	L612	S543	F474	F474	I414	I351	A222	GLY	GLY
	K613	S543	L475	L475	H415	P352	P223	GLY	GLY
	N614	K544	F476	F476	G416	G353	Q290	GLN	GLN
	L615	Q545	L477	L477	N417	A291	A291	VAL	VAL
	Y616	Y546	Q479	Q479	Q418	E226	E226	SER	SER
	P617	V547	K480	K480	D419	R293	R227	LEU	LEU
	K618	R618			N421	F295	E229	HIS	HIS
	K619	L551			A422	L359	S295	SER	SER
	P620	L552	S486	S486	A422	A422	L296	LEU	LEU
	K621	N553			K423	K361	L297	ILE	ILE
	E622	E554	E490	E490	A424	N362	L298	GLU	GLU
	D623	P555			T425	L363	L299	THR	THR
			Q493	Q493	I426	L364		PRO	PRO

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	66.15Å 94.32Å 179.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.36 – 2.71 46.36 – 2.71	Depositor EDS
% Data completeness (in resolution range)	90.8 (46.36-2.71) 89.1 (46.36-2.71)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.79 (at 2.73Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.185 , 0.253 0.197 , 0.228	Depositor DCC
R_{free} test set	2000 reflections (6.38%)	wwPDB-VP
Wilson B-factor (Å ²)	22.9	Xtriage
Anisotropy	0.093	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 46.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	7227	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.39% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NI, PTR

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	2.45	192/3594 (5.3%)	1.55	51/4843 (1.1%)
1	B	2.37	176/3545 (5.0%)	1.54	49/4782 (1.0%)
All	All	2.41	368/7139 (5.2%)	1.54	100/9625 (1.0%)

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	2
1	B	0	2
All	All	0	4

The worst 5 of 368 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	446	VAL	C-O	-9.76	1.16	1.24
1	A	517	TRP	C-O	-9.35	1.13	1.24
1	A	511	GLY	C-O	-9.10	1.17	1.24
1	B	446	VAL	C-O	-8.86	1.17	1.24
1	A	571	ILE	C-O	-8.83	1.14	1.24

The worst 5 of 100 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	241	GLU	N-CA-C	-11.46	95.40	110.53
1	A	502	PHE	N-CA-C	10.46	122.43	111.14
1	A	493	GLN	N-CA-C	9.61	121.75	111.28
1	B	502	PHE	N-CA-C	9.33	121.39	111.03

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	419	ASP	N-CA-C	8.99	124.48	112.88

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	303	GLY	Peptide
1	A	436	ASP	Peptide
1	B	184	GLN	Peptide
1	B	418	GLN	Peptide

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	3550	0	3614	146	1
1	B	3500	0	3568	153	0
2	A	1	0	0	0	0
3	A	98	0	0	8	0
3	B	78	0	0	12	0
All	All	7227	0	7182	295	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 21.

The worst 5 of 295 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:183:LEU:O	1:B:186:THR:OG1	1.56	1.23
1:B:144:ILE:CG2	1:B:183:LEU:HD12	1.73	1.17
1:A:153:LYS:O	1:A:155:ALA:N	1.86	1.09
1:A:233:ASP:OD1	1:A:300:ARG:NH1	1.85	1.08
1:A:153:LYS:C	1:A:155:ALA:H	1.50	1.06

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:180:ALA:CB	1:A:344:THR:O[3_544]	1.90	0.30

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	428/549 (78%)	413 (96%)	13 (3%)	2 (0%)	24	48
1	B	423/549 (77%)	416 (98%)	6 (1%)	1 (0%)	43	68
All	All	851/1098 (78%)	829 (97%)	19 (2%)	3 (0%)	30	54

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	154	GLY
1	B	415	HIS
1	A	638	GLY

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	386/469 (82%)	303 (78%)	83 (22%)	1	3
1	B	382/469 (81%)	296 (78%)	86 (22%)	1	3
All	All	768/938 (82%)	599 (78%)	169 (22%)	1	3

5 of 169 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	323	GLU
1	B	500	SER
1	B	360	PHE
1	B	426	ILE
1	B	551	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	207	GLN
1	B	286	GLN
1	B	342	ASN
1	B	320	GLN
1	A	421	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 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	$\# Z > 2$
1	PTR	A	641	1	15,16,17	2.93	7 (46%)	17,22,24	1.22	2 (11%)
1	PTR	B	641	1	15,16,17	2.46	5 (33%)	17,22,24	2.04	6 (35%)

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
1	PTR	A	641	1	-	0/10/11/13	0/1/1/1
1	PTR	B	641	1	-	0/10/11/13	0/1/1/1

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	641	PTR	P-OH	-5.43	1.49	1.59
1	A	641	PTR	P-O2P	-5.21	1.35	1.54
1	A	641	PTR	P-O3P	-5.10	1.35	1.54
1	B	641	PTR	P-O1P	-4.85	1.35	1.50
1	B	641	PTR	P-O2P	-4.53	1.38	1.54

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	641	PTR	O3P-P-OH	4.24	117.85	105.32
1	B	641	PTR	O2P-P-OH	-3.98	93.56	105.32
1	B	641	PTR	O3P-P-O2P	3.20	119.79	107.80
1	B	641	PTR	CD1-CE1-CZ	3.03	123.19	119.73
1	A	641	PTR	P-OH-CZ	2.41	132.48	123.88

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	641	PTR	1	0

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - 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

There are no such residues in this entry.

5.8 Polymer linkage issues

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	444/549 (80%)	-0.21	7 (1%) 70 68	1, 17, 50, 73	0
1	B	437/549 (79%)	-0.12	11 (2%) 58 55	2, 18, 58, 90	0
All	All	881/1098 (80%)	-0.16	18 (2%) 65 62	1, 18, 54, 90	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	182	LEU	3.5
1	B	180	ALA	3.2
1	B	414	VAL	3.2
1	B	151	LEU	3.1
1	A	638	GLY	2.8

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. 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	B-factors(Å ²)	Q<0.9
1	PTR	B	641	16/17	0.97	0.07	5,10,17,19	0
1	PTR	A	641	16/17	0.98	0.06	6,11,17,24	0

6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands

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. 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	B-factors(Å ²)	Q<0.9
2	NI	A	701	1/1	1.00	0.02	11,11,11,11	0

6.5 Other polymers

There are no such residues in this entry.