WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 24, 2024 – 12:35 PM EDT

PDB ID	:	6TRS
Title	:	Crystal structure of TFIIH subunit p52 in complex with p8
Authors	:	Koelmel, W.; Kuper, J.; Schoenwetter, E.; Kisker, C.
Deposited on	:	2019-12-19
Resolution	:	2.68 Å(reported)

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

|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
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)	:	2.37.1

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 2.68 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.
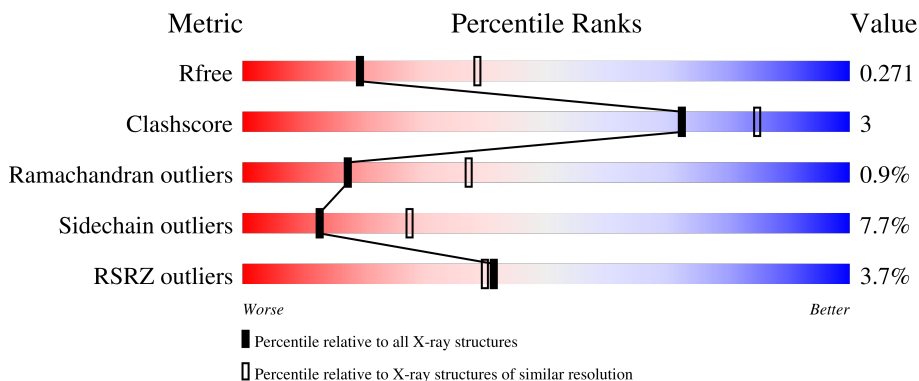
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3863 (2.70-2.66)
Clashscore	141614	4210 (2.70-2.66)
Ramachandran outliers	138981	4141 (2.70-2.66)
Sidechain outliers	138945	4141 (2.70-2.66)
RSRZ outliers	127900	3780 (2.70-2.66)

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	420	% 60% 10% 30%
1	B	420	4% 72% 11% • 17%
2	D	98	7% 43% 22% • 33%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 5606 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 RNA polymerase II transcription factor B subunit 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	292	Total	C	N	O	S	0	0	0
			2281	1461	396	418	6			
1	B	350	Total	C	N	O	S	0	0	0
			2784	1784	481	510	9			

There are 62 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	95	MET	-	initiating methionine	UNP G0S965
A	96	LYS	-	expression tag	UNP G0S965
A	97	HIS	-	expression tag	UNP G0S965
A	98	HIS	-	expression tag	UNP G0S965
A	99	HIS	-	expression tag	UNP G0S965
A	100	HIS	-	expression tag	UNP G0S965
A	101	HIS	-	expression tag	UNP G0S965
A	102	HIS	-	expression tag	UNP G0S965
A	103	PRO	-	expression tag	UNP G0S965
A	104	MET	-	expression tag	UNP G0S965
A	105	SER	-	expression tag	UNP G0S965
A	106	ASP	-	expression tag	UNP G0S965
A	107	TYR	-	expression tag	UNP G0S965
A	108	ASP	-	expression tag	UNP G0S965
A	109	ILE	-	expression tag	UNP G0S965
A	110	PRO	-	expression tag	UNP G0S965
A	111	THR	-	expression tag	UNP G0S965
A	112	THR	-	expression tag	UNP G0S965
A	113	GLU	-	expression tag	UNP G0S965
A	114	ASN	-	expression tag	UNP G0S965
A	115	LEU	-	expression tag	UNP G0S965
A	116	TYR	-	expression tag	UNP G0S965
A	117	PHE	-	expression tag	UNP G0S965
A	118	GLN	-	expression tag	UNP G0S965
A	119	GLY	-	expression tag	UNP G0S965

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	120	ALA	-	expression tag	UNP G0S965
A	320T	SER	ASP	conflict	UNP G0S965
A	320U	ASN	ALA	conflict	UNP G0S965
A	320V	GLY	SER	conflict	UNP G0S965
A	320W	ASN	SER	conflict	UNP G0S965
A	320X	GLY	LEU	conflict	UNP G0S965
B	95	MET	-	initiating methionine	UNP G0S965
B	96	LYS	-	expression tag	UNP G0S965
B	97	HIS	-	expression tag	UNP G0S965
B	98	HIS	-	expression tag	UNP G0S965
B	99	HIS	-	expression tag	UNP G0S965
B	100	HIS	-	expression tag	UNP G0S965
B	101	HIS	-	expression tag	UNP G0S965
B	102	HIS	-	expression tag	UNP G0S965
B	103	PRO	-	expression tag	UNP G0S965
B	104	MET	-	expression tag	UNP G0S965
B	105	SER	-	expression tag	UNP G0S965
B	106	ASP	-	expression tag	UNP G0S965
B	107	TYR	-	expression tag	UNP G0S965
B	108	ASP	-	expression tag	UNP G0S965
B	109	ILE	-	expression tag	UNP G0S965
B	110	PRO	-	expression tag	UNP G0S965
B	111	THR	-	expression tag	UNP G0S965
B	112	THR	-	expression tag	UNP G0S965
B	113	GLU	-	expression tag	UNP G0S965
B	114	ASN	-	expression tag	UNP G0S965
B	115	LEU	-	expression tag	UNP G0S965
B	116	TYR	-	expression tag	UNP G0S965
B	117	PHE	-	expression tag	UNP G0S965
B	118	GLN	-	expression tag	UNP G0S965
B	119	GLY	-	expression tag	UNP G0S965
B	120	ALA	-	expression tag	UNP G0S965
B	321S	SER	ASP	conflict	UNP G0S965
B	321T	ASN	ALA	conflict	UNP G0S965
B	321U	GLY	SER	conflict	UNP G0S965
B	321V	ASN	SER	conflict	UNP G0S965
B	321W	GLY	LEU	conflict	UNP G0S965

- Molecule 2 is a protein called Uncharacterized protein.

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	66	541	343	94	102	2	0	0	0

There are 26 discrepancies between the modelled and reference sequences:

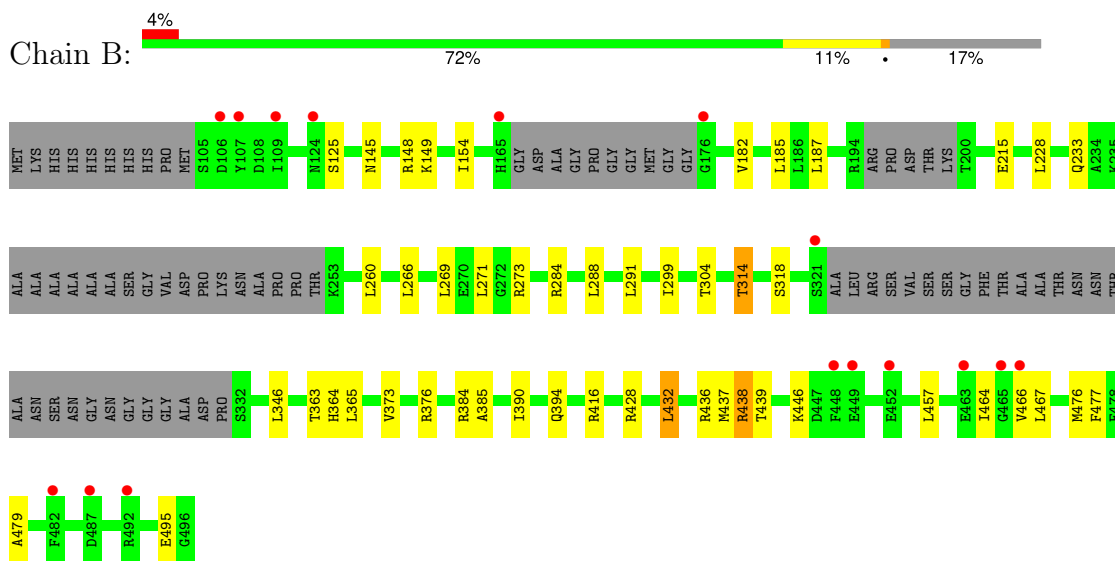| Chain	Residue	Modelled	Actual	Comment	Reference
D	-25	MET	-	initiating methionine	UNP G0SDQ1
D	-24	LYS	-	expression tag	UNP G0SDQ1
D	-23	HIS	-	expression tag	UNP G0SDQ1
D	-22	HIS	-	expression tag	UNP G0SDQ1
D	-21	HIS	-	expression tag	UNP G0SDQ1
D	-20	HIS	-	expression tag	UNP G0SDQ1
D	-19	HIS	-	expression tag	UNP G0SDQ1
D	-18	HIS	-	expression tag	UNP G0SDQ1
D	-17	PRO	-	expression tag	UNP G0SDQ1
D	-16	MET	-	expression tag	UNP G0SDQ1
D	-15	SER	-	expression tag	UNP G0SDQ1
D	-14	ASP	-	expression tag	UNP G0SDQ1
D	-13	TYR	-	expression tag	UNP G0SDQ1
D	-12	ASP	-	expression tag	UNP G0SDQ1
D	-11	ILE	-	expression tag	UNP G0SDQ1
D	-10	PRO	-	expression tag	UNP G0SDQ1
D	-9	THR	-	expression tag	UNP G0SDQ1
D	-8	THR	-	expression tag	UNP G0SDQ1
D	-7	GLU	-	expression tag	UNP G0SDQ1
D	-6	ASN	-	expression tag	UNP G0SDQ1
D	-5	LEU	-	expression tag	UNP G0SDQ1
D	-4	TYR	-	expression tag	UNP G0SDQ1
D	-3	PHE	-	expression tag	UNP G0SDQ1
D	-2	GLN	-	expression tag	UNP G0SDQ1
D	-1	GLY	-	expression tag	UNP G0SDQ1
D	0	ALA	-	expression tag	UNP G0SDQ1

3 Residue-property plots ⓘ

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.
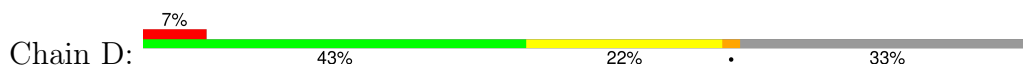
• Molecule 1: RNA polymerase II transcription factor B subunit 2
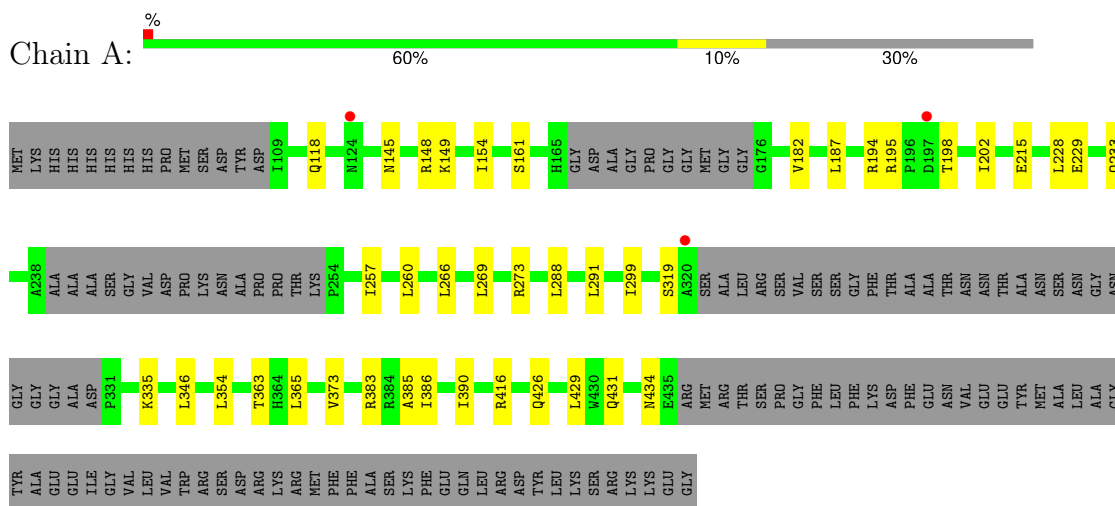
• Molecule 1: RNA polymerase II transcription factor B subunit 2

• Molecule 2: Uncharacterized protein

MET LYS HIS HIS HIS HIS HIS HIS PRO MET SER ASP TYR ASP ILE PRO THR THR GLU ASN LEU TYR PHE GLN GLY ALA MET VAL ARG A4 I5 R6 E11 C12 E13 K17 I20 N27 H28 D29 F30 I31 I32 L40 E44 N45 L50 K51 Q52 K53 L54 E55 D56

R57 L58 T61 Y62 R63 P64 E65 E66 P67 L68 A69 ASP SER ASP
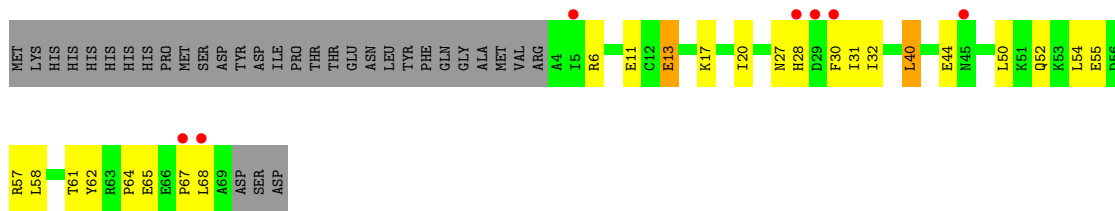
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	74.22Å 85.95Å 92.17Å 90.00° 94.93° 90.00°	Depositor
Resolution (Å)	19.92 – 2.68 19.92 – 2.68	Depositor EDS
% Data completeness (in resolution range)	52.7 (19.92-2.68) 52.7 (19.92-2.68)	Depositor EDS
R_{merge}	0.28	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.19 (at 2.67Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.216 , 0.240 0.238 , 0.271	Depositor DCC
R_{free} test set	1091 reflections (6.38%)	wwPDB-VP
Wilson B-factor (Å^2)	37.8	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 6.3	EDS
L-test for twinning[2]	$< |L| > = 0.45, < L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	5606	wwPDB-VP
Average B, all atoms (Å^2)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.39% 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	0.40	0/2330	0.58	0/3168
1	B	0.39	0/2843	0.59	0/3848
2	D	0.47	0/549	0.66	0/742
All	All	0.40	0/5722	0.59	0/7758

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	2281	0	2299	18	0
1	B	2784	0	2784	22	0
2	D	541	0	551	6	0
All	All	5606	0	5634	38	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 3.

The worst 5 of 38 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:346:LEU:HD13	1:B:363:THR:HG21	1.80	0.64

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:346:LEU:HD13	1:A:363:THR:HG21	1.79	0.63
1:B:145:ASN:HD22	1:B:148:ARG:HH21	1.47	0.62
1:A:145:ASN:HD22	1:A:148:ARG:HH21	1.47	0.61
1:A:383:ARG:HA	1:A:386:ILE:HD12	1.85	0.57

There are no symmetry-related clashes.

5.3   Torsion angles ⓘ

5.3.1   Protein backbone ⓘ

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	284/420 (68%)	277 (98%)	7 (2%)	0	100	100
1	B	340/420 (81%)	326 (96%)	14 (4%)	0	100	100
2	D	64/98 (65%)	52 (81%)	6 (9%)	6 (9%)	0	0
All	All	688/938 (73%)	655 (95%)	27 (4%)	6 (1%)	17	37

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	6	ARG
2	D	30	PHE
2	D	64	PRO
2	D	65	GLU
2	D	44	GLU

5.3.2   Protein sidechains ⓘ

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	241/339 (71%)	229 (95%)	12 (5%)	24	47
1	B	294/339 (87%)	272 (92%)	22 (8%)	13	29
2	D	62/92 (67%)	50 (81%)	12 (19%)	1	3
All	All	597/770 (78%)	551 (92%)	46 (8%)	13	27

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	457	LEU
2	D	27	ASN
1	B	464	ILE
1	B	477	PHE
2	D	32	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	394	GLN
1	B	406	GLN
1	B	426	GLN
1	A	426	GLN
1	B	145	ASN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5   Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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(Å²)	Q<0.9
1	A	292/420 (69%)	-0.06	3 (1%)	82	82	22, 41, 74, 111	0
1	B	350/420 (83%)	0.15	16 (4%)	32	30	26, 47, 109, 124	0
2	D	66/98 (67%)	0.62	7 (10%)	6	4	46, 69, 107, 115	0
All	All	708/938 (75%)	0.10	26 (3%)	41	40	22, 46, 102, 124	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	5	ILE	6.5
1	B	448	PHE	5.9
1	B	165	HIS	4.9
2	D	28	HIS	4.4
2	D	67	PRO	4.2

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.