

Full wwPDB X-ray Structure Validation Report (i)

Jun 23, 2024 – 04:36 AM EDT

PDB ID	:	4RE1
Title	:	Crystal structure of human TEAD1 and disulfide-engineered YAP
Authors	:	Xu, Z.; Zhou, Z.
Deposited on		
Resolution	:	2.20 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

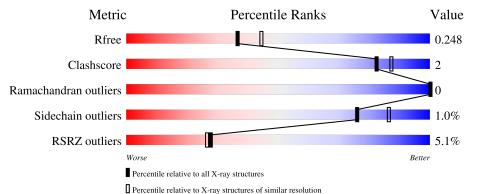
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: $X\text{-}RAY \, DIFFRACTION$

The reported resolution of this entry is 2.20 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R_{free}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	А	220	5%			89%	5% 6%	
1	В	220	2%			88%	• 8%	
2	С	149	5%	28%	5%	66%		
2	D	149	.% <mark>-</mark>	30%	•	68%		



2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4054 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 Transcriptional enhancer factor TEF-1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
1	Δ	206	Total	С	Ν	0	\mathbf{S}	0	4	0
	A	200	1646	1054	275	304	13	0	4	0
1	В	202	Total	С	Ν	0	S	0	1	0
	D	202	1621	1034	271	303	13	0	4	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	192	MET	-	expression tag	UNP P28347
А	193	ALA	-	expression tag	UNP P28347
В	192	MET	-	expression tag	UNP P28347
В	193	ALA	-	expression tag	UNP P28347

• Molecule 2 is a protein called Yorkie homolog.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
9	С	50	Total	С	Ν	Ο	S	0	0	0
	U	50	355	225	64	62	4	0	0	
0	Л	47	Total	С	Ν	Ο	S	0	0	0
	D	47	336	214	59	59	4	0	0	0

There are 62 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
С	23	MET	-	expression tag	UNP P46937
С	24	GLY	-	expression tag	UNP P46937
С	25	SER	-	expression tag	UNP P46937
С	26	SER	-	expression tag	UNP P46937
С	27	HIS	-	expression tag	UNP P46937
С	28	HIS	-	expression tag	UNP P46937
С	29	HIS	-	expression tag	UNP P46937
С	30	HIS	-	expression tag	UNP P46937



Chain	Residue	Modelled	Actual	Comment	Reference
С	31	HIS	-	expression tag	UNP P46937
С	32	HIS	-	expression tag	UNP P46937
С	33	HIS	-	expression tag	UNP P46937
С	34	HIS	-	expression tag	UNP P46937
С	35	HIS	-	expression tag	UNP P46937
С	36	HIS	-	expression tag	UNP P46937
С	37	SER	_	expression tag	UNP P46937
С	38	SER	-	expression tag	UNP P46937
С	39	GLY	-	expression tag	UNP P46937
С	40	ARG	-	expression tag	UNP P46937
С	41	GLU	-	expression tag	UNP P46937
С	42	ASN	-	expression tag	UNP P46937
С	43	LEU	-	expression tag	UNP P46937
С	44	TYR	-	expression tag	UNP P46937
С	45	PHE	-	expression tag	UNP P46937
С	46	GLN	-	expression tag	UNP P46937
С	47	SER	-	expression tag	UNP P46937
С	48	HIS	-	expression tag	UNP P46937
С	49	MET	-	expression tag	UNP P46937
С	87	CYS	ARG	engineered mutation	UNP P46937
С	93	ALA	ASP	engineered mutation	UNP P46937
С	96	CYS	PHE	engineered mutation	UNP P46937
С	100	ARG	GLU	engineered mutation	UNP P46937
D	23	MET	-	expression tag	UNP P46937
D	24	GLY	-	expression tag	UNP P46937
D	25	SER	-	expression tag	UNP P46937
D	26	SER	-	expression tag	UNP P46937
D	27	HIS	-	expression tag	UNP P46937
D	28	HIS	-	expression tag	UNP P46937
D	29	HIS	-	expression tag	UNP P46937
D	30	HIS	-	expression tag	UNP P46937
D	31	HIS	-	expression tag	UNP P46937
D	32	HIS	-	expression tag	UNP P46937
D	33	HIS	-	expression tag	UNP P46937
D	34	HIS	-	expression tag	UNP P46937
D	35	HIS	-	expression tag	UNP P46937
D	36	HIS	-	expression tag	UNP P46937
D	37	SER	-	expression tag	UNP P46937
D	38	SER	-	expression tag	UNP P46937
D	39	GLY	-	expression tag	UNP P46937
D	40	ARG	-	expression tag	UNP P46937
D	41	GLU	-	expression tag	UNP P46937

Continued from previous page...



Chain	Residue	Modelled	Actual	Comment	Reference
D	42	ASN	-	expression tag	UNP P46937
D	43	LEU	-	expression tag	UNP P46937
D	44	TYR	-	expression tag	UNP P46937
D	45	PHE	-	expression tag	UNP P46937
D	46	GLN	-	expression tag	UNP P46937
D	47	SER	-	expression tag	UNP P46937
D	48	HIS	-	expression tag	UNP P46937
D	49	MET	-	expression tag	UNP P46937
D	87	CYS	ARG	engineered mutation	UNP P46937
D	93	ALA	ASP	engineered mutation	UNP P46937
D	96	CYS	PHE	engineered mutation	UNP P46937
D	100	ARG	GLU	engineered mutation	UNP P46937

Continued from previous page...

• Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	А	1	Total Cl 1 1	0	0

• Molecule 4 is water.

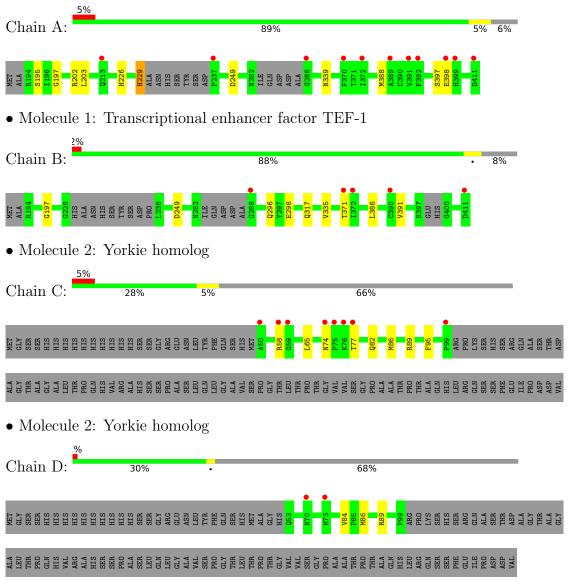
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	40	Total O 40 40	0	0
4	В	54	$\begin{array}{cc} \text{Total} & \text{O} \\ 54 & 54 \end{array}$	0	0
4	D	1	Total O 1 1	0	0



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: Transcriptional enhancer factor TEF-1





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	44.25Å 105.03Å 164.78Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	26.57 - 2.20	Depositor
Resolution (A)	26.57 - 2.20	EDS
% Data completeness	95.8 (26.57-2.20)	Depositor
(in resolution range)	95.7 (26.57-2.20)	EDS
R _{merge}	0.08	Depositor
R _{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.50 (at 2.20Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.3_1479)	Depositor
D D	0.212 , 0.248	Depositor
R, R_{free}	0.214 , 0.248	DCC
R_{free} test set	1901 reflections (4.96%)	wwPDB-VP
Wilson B-factor $(Å^2)$	44.6	Xtriage
Anisotropy	0.679	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.34, 41.3	EDS
L-test for twinning ²	$< L > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	4054	wwPDB-VP
Average B, all atoms $(Å^2)$	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 10.67% of the height of the origin peak. No significant pseudotranslation is detected.

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.



¹Intensities estimated from amplitudes.

5 Model quality (i)

5.1 Standard geometry (i)

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

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	А	0.43	0/1682	0.57	0/2272	
1	В	0.45	0/1653	0.55	0/2232	
2	С	0.41	0/364	0.55	0/500	
2	D	0.39	0/344	0.58	0/473	
All	All	0.43	0/4043	0.56	0/5477	

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 (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	А	1646	0	1556	8	0
1	В	1621	0	1529	7	0
2	С	355	0	340	5	0
2	D	336	0	325	3	0
3	А	1	0	0	0	0
4	А	40	0	0	1	0
4	В	54	0	0	0	0
4	D	1	0	0	0	0
All	All	4054	0	3750	19	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 2.

All (19) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic	Clash
		distance (Å)	overlap (Å)
1:A:203:LEU:HD22	1:A:388:MET:HE1	1.74	0.70
2:D:86:MET:HA	2:D:89:ARG:HD3	1.73	0.70
1:A:226:HIS:HB2	1:A:229:HIS:CD2	2.30	0.66
2:C:86:MET:HA	2:C:89:ARG:HD3	1.76	0.65
1:A:226:HIS:HB2	1:A:229:HIS:NE2	2.13	0.62
1:B:249:ASP:OD1	2:C:89:ARG:NH2	2.35	0.59
1:A:249:ASP:OD1	2:D:89:ARG:NH2	2.42	0.52
2:C:74:ASN:HB3	2:C:77:THR:HB	1.94	0.50
1:B:371:THR:HG22	1:B:391:VAL:HG12	1.95	0.49
1:B:197:GLY:HA2	1:B:386:LEU:HD21	1.96	0.48
1:B:391:VAL:HG21	2:C:95:PHE:HZ	1.78	0.47
1:A:195:SER:OG	1:A:202:ARG:HD3	2.14	0.47
1:A:397:SER:OG	1:A:398:GLU:N	2.49	0.46
2:D:84:VAL:O	2:D:89:ARG:NH1	2.39	0.46
1:B:317:GLN:OE1	2:C:58:ARG:NE	2.36	0.44
1:B:298[A]:GLU:HG2	1:B:335:VAL:HG22	1.99	0.44
1:A:202:ARG:HG3	4:A:630:HOH:O	2.18	0.43
1:A:197:GLY:HA3	1:A:202:ARG:HG2	2.02	0.42
1:B:296:GLN:NE2	1:B:298[A]:GLU:OE2	2.52	0.41

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	Perce	ntiles
1	А	204/220~(93%)	203 (100%)	1 (0%)	0	100	100
1	В	198/220~(90%)	197 (100%)	1 (0%)	0	100	100



Contre	Continueu front previous page										
Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles					
2	С	48/149~(32%)	48 (100%)	0	0	100 100					
2	D	45/149~(30%)	43~(96%)	2 (4%)	0	100 100					
All	All	495/738~(67%)	491 (99%)	4 (1%)	0	100 100					

Continued from previous page...

There are no Ramachandran outliers to report.

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	А	176/199~(88%)	174~(99%)	2(1%)	73 85
1	В	174/199~(87%)	174 (100%)	0	100 100
2	С	37/124~(30%)	35~(95%)	2~(5%)	22 26
2	D	36/124~(29%)	36 (100%)	0	100 100
All	All	423/646~(66%)	419 (99%)	4 (1%)	76 88

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	229	HIS
1	А	339	ASN
2	С	65	LEU
2	С	82	GLN

Sometimes side chains can be flipped to improve hydrogen bonding and reduce clashes. There are no such side chains 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)

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 (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, 95^{th} 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	$\langle RSRZ \rangle$	#RSRZ>2	$OWAB(Å^2)$	Q<0.9
1	А	206/220~(93%)	0.23	11 (5%) 26 25	35, 47, 68, 76	0
1	В	202/220~(91%)	0.16	5 (2%) 57 55	33, 48, 67, 78	0
2	С	50/149~(33%)	0.56	8 (16%) 1 1	40, 58, 79, 88	0
2	D	47/149 (31%)	0.19	2 (4%) 35 33	41, 53, 66, 69	0
All	All	505/738~(68%)	0.23	26 (5%) 28 26	33, 49, 69, 88	0

All (26) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	С	50	ALA	4.4
1	А	399	HIS	3.9
2	С	77	THR	3.8
2	С	76	LYS	3.7
1	В	288	GLY	3.7
2	С	99	PRO	3.1
1	А	411	ASP	3.1
2	С	75	PRO	2.9
1	А	237	PRO	2.8
2	С	58	ARG	2.8
1	А	398	GLU	2.8
1	В	411	ASP	2.6
1	А	288	GLY	2.4
1	А	392	PHE	2.4
1	В	372	ILE	2.4
1	В	371	THR	2.3
2	D	70	ASN	2.3
1	А	389	ALA	2.2
1	В	390	CYS	2.2
2	С	74	ASN	2.1
1	А	370	PHE	2.1



Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	А	372	ILE	2.0
1	А	213	GLN	2.0
1	А	391	VAL	2.0
2	D	73	MET	2.0
2	С	59	GLY	2.0

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)

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, 95^{th} 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	$\mathbf{B} ext{-factors}(\mathbf{A}^2)$	Q < 0.9
3	CL	А	501	1/1	0.92	0.08	75, 75, 75, 75, 75	0

6.5 Other polymers (i)

There are no such residues in this entry.

