wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 11:53 PM GMT

PDB ID	:	1YTF
Title	:	YEAST TFIIA/TBP/DNA COMPLEX
Authors	:	Tan, S.; Hunziker, Y.; Sargent, D.F.; Richmond, T.J.
Deposited on	:	1996-04-05
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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

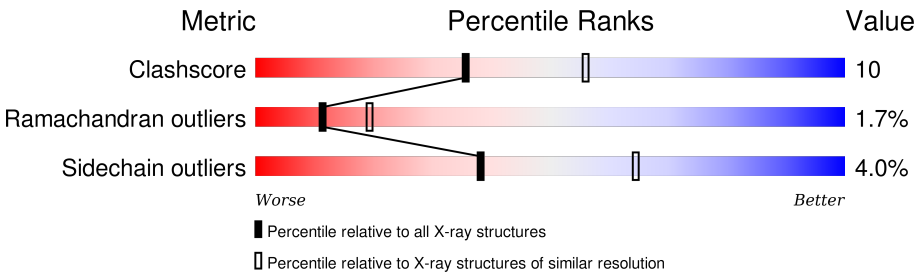
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 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))
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	E	16	19% 38% 38% 6%
2	F	16	38% 38% 25%
3	A	180	79% 20% •
4	B	53	68% 13% 6% 13%
5	C	79	37% 18% • 42%
6	D	121	65% 17% • 17%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 3605 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 DNA chain called DNA (5'-D(*TP*GP*TP*AP*TP*GP*TP*AP*TP*AP*TP*AP*AP*AP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	E	16	327	159	60	93	15	0	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*TP*TP*TP*TP*AP*TP*AP*TP*AP*CP*AP*TP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	F	16	323	158	55	95	15	0	0	0

- Molecule 3 is a protein called PROTEIN (TATA BINDING PROTEIN (TBP)).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	180	1416	921	242	247	6	0	0	0

- Molecule 4 is a protein called PROTEIN (TRANSCRIPTION FACTOR IIA - TOA1N SUBUNIT).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
4	B	46	374	232	63	79	0	0	0

- Molecule 5 is a protein called PROTEIN (TRANSCRIPTION FACTOR IIA - TOA1C SUBUNIT).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	C	46	383	242	67	71	3	0	0	0

- Molecule 6 is a protein called PROTEIN (TRANSCRIPTION FACTOR IIA - TOA2 SUB-

UNIT).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	D	100	782	492	130	156	4	0	0	0

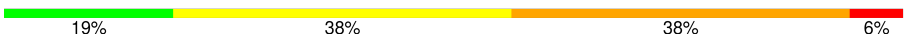
3 Residue-property plots (i)

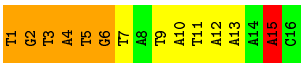
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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.

Note EDS was not executed.

• Molecule 1: DNA (5'-D(*TP*GP*TP*AP*TP*GP*TP*AP*TP*AP*TP*AP*AP*AP*AP*C)-3')

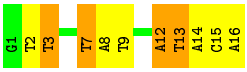
Chain E:





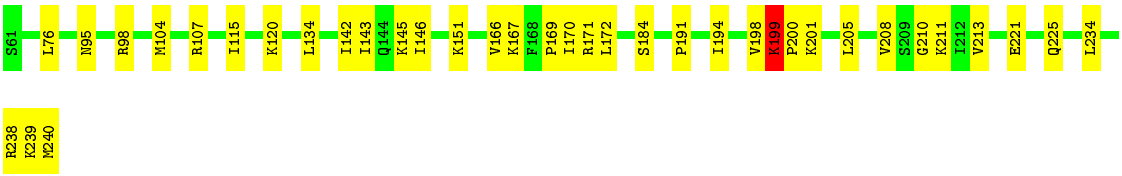
• Molecule 2: DNA (5'-D(*GP*TP*TP*TP*TP*AP*TP*AP*TP*AP*CP*AP*TP*AP*CP*A)-3')

Chain F:



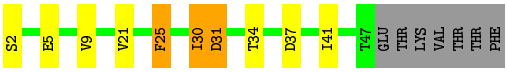
• Molecule 3: PROTEIN (TATA BINDING PROTEIN (TBP))

Chain A:



• Molecule 4: PROTEIN (TRANSCRIPTION FACTOR IIA - TOA1N SUBUNIT)

Chain B:



• Molecule 5: PROTEIN (TRANSCRIPTION FACTOR IIA - TOA1C SUBUNIT)

Chain C:

GLY SER SER ALA LEU LEU ASP THR ASP GLU VAL GLY SER GLU LEU ASP ASP SER ASP ASP ASP TYR LEU ILE SER GLU GLY GLU GLU ASP GLY PRO ASP E241 N242 L243 M244 L245 R253 A256 K259 L262 K263 D264 V267 T268 Y274 Q277 K278 A279 Q280 V281

W285 V286

- Molecule 6: PROTEIN (TRANSCRIPTION FACTOR IIA - TOA2 SUBUNIT)

Chain D:

ALA VAL PRO G5 Y6 L9 Y10 R11 T25 D29 I32 E33 A34 S35 L36 A37 M38 R39 V40 L41 E42 T51 N55 T56 Q57 L60 V87 E88 ASP SER HIS ARG ASP ALA SER GLN ASN GLY SER GLY ASP SER GLN S104 V108 K119 LYS SER GLU

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	59.15Å 93.05Å 117.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	6.00 – 2.50	Depositor
% Data completeness (in resolution range)	(Not available) (6.00-2.50)	Depositor
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	X-PLOR	Depositor
R, R_{free}	0.235 , 0.297	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3605	wwPDB-VP
Average B, all atoms (Å^2)	21.0	wwPDB-VP

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	E	1.59	2/367 (0.5%)	1.70	11/565 (1.9%)
2	F	1.40	0/361	1.69	10/555 (1.8%)
3	A	0.38	0/1443	0.64	1/1942 (0.1%)
4	B	0.36	0/377	0.46	0/509
5	C	0.39	0/389	0.72	0/523
6	D	0.36	0/789	0.62	1/1066 (0.1%)
All	All	0.74	2/3726 (0.1%)	0.96	23/5160 (0.4%)

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	E	0	3
2	F	0	3
All	All	0	6

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	15	DA	N9-C4	5.37	1.41	1.37
1	E	9	DT	O4'-C1'	5.04	1.48	1.42

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	F	2	DT	C4'-C3'-C2'	6.90	109.31	103.10
1	E	4	DA	C4'-C3'-C2'	6.76	109.19	103.10
1	E	2	DG	C4'-C3'-C2'	6.62	109.06	103.10
2	F	12	DA	C4'-C3'-C2'	6.14	108.63	103.10
1	E	10	DA	C4'-C3'-C2'	6.14	108.62	103.10

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	13	DA	Sidechain
1	E	15	DA	Sidechain
1	E	6	DG	Sidechain
2	F	12	DA	Sidechain
2	F	7	DT	Sidechain

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	E	327	0	184	6	0
2	F	323	0	185	5	0
3	A	1416	0	1493	33	0
4	B	374	0	363	11	0
5	C	383	0	384	13	0
6	D	782	0	790	12	0
All	All	3605	0	3399	67	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 10.

The worst 5 of 67 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)
4:B:2:SER:N	5:C:274:TYR:HH	1.61	0.99
3:A:199:LYS:HB2	3:A:200:PRO:HD3	1.46	0.94
3:A:199:LYS:HB2	3:A:200:PRO:CD	2.06	0.82
1:E:3:DT:H2"	1:E:4:DA:H5"	1.63	0.81
3:A:107:ARG:HD3	5:C:286:VAL:HB	1.65	0.78

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	
3	A	178/180 (99%)	171 (96%)	5 (3%)	2 (1%)	17	31
4	B	44/53 (83%)	42 (96%)	0	2 (4%)	3	3
5	C	44/79 (56%)	40 (91%)	3 (7%)	1 (2%)	8	12
6	D	96/121 (79%)	88 (92%)	7 (7%)	1 (1%)	19	34
All	All	362/433 (84%)	341 (94%)	15 (4%)	6 (2%)	11	19

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	199	LYS
3	A	201	LYS
4	B	31	ASP
5	C	264	ASP
6	D	6	TYR

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	
3	A	152/152 (100%)	149 (98%)	3 (2%)	63	86
4	B	42/49 (86%)	41 (98%)	1 (2%)	57	82
5	C	42/70 (60%)	38 (90%)	4 (10%)	11	20
6	D	90/107 (84%)	85 (94%)	5 (6%)	26	47
All	All	326/378 (86%)	313 (96%)	13 (4%)	38	64

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	C	253	ARG
5	C	267	VAL
6	D	39	ARG
5	C	241	GLU
6	D	33	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	158	GLN
6	D	57	GLN
5	C	280	GLN
3	A	95	ASN
6	D	55	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 carbohydrates 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)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.