wwPDB X-ray Structure Validation Summary Report (i)

Dec 8, 2023 – 12:57 am GMT

PDB ID	:	2UZK
Title	:	Crystal structure of the human FOXO3a-DBD bound to DNA
Authors	:	Tsai, K.-L.; Sun, Y.-J.; Huang, C.-Y.; Yang, J.-Y.; Hung, M.-C.; Hsiao, C.-D.
Deposited on	:	2007-04-30
Resolution	:	2.70 Å(reported)

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

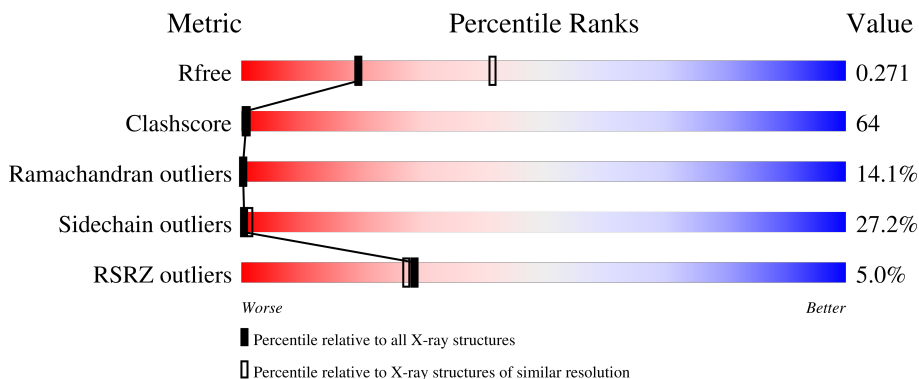|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.70 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.
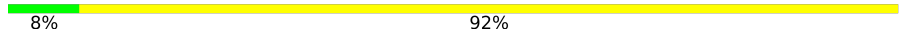
Metric	Percentile Ranks	Value
Rfree		0.271
Clashscore		64
Ramachandran outliers		14.1%
Sidechain outliers		27.2%
RSRZ outliers		5.0%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (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 <=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	97	9% 28% 44% 19% 9%
1	C	97	3% 15% 52% 24% · 5%
2	B	13	8% 85% 8%
2	D	13	15% 85%
3	E	13	8% 85% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	13	8% 92%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 2741 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 FORKHEAD BOX PROTEIN O3A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	97	Total	C	N	O	S	0	0	0
			774	481	147	142	4			
1	C	92	Total	C	N	O	S	0	0	1
			704	437	131	133	3			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	13	Total	C	N	O	P	0	0	0
			262	127	50	73	12			
2	D	13	Total	C	N	O	P	0	0	0
			262	127	50	73	12			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	13	Total	C	N	O	P	0	0	0
			265	129	45	79	12			
3	F	13	Total	C	N	O	P	0	0	0
			265	129	45	79	12			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	46	Total	O	0	0
			46	46		
4	B	15	Total	O	0	0
			15	15		
4	C	72	Total	O	0	0
			72	72		

Continued on next page...

Continued from previous page...

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	D	33	Total 33	O 33	0	0
4	E	22	Total 22	O 22	0	0
4	F	21	Total 21	O 21	0	0

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: FORKHEAD BOX PROTEIN O3A
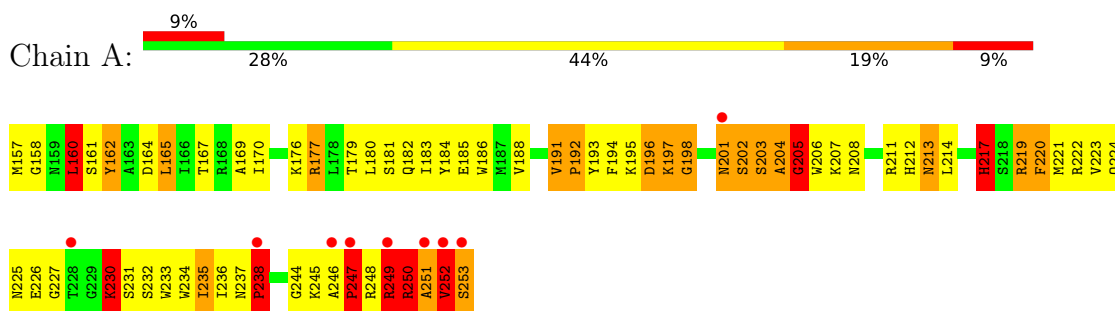
Chain A:

• Molecule 1: FORKHEAD BOX PROTEIN O3A
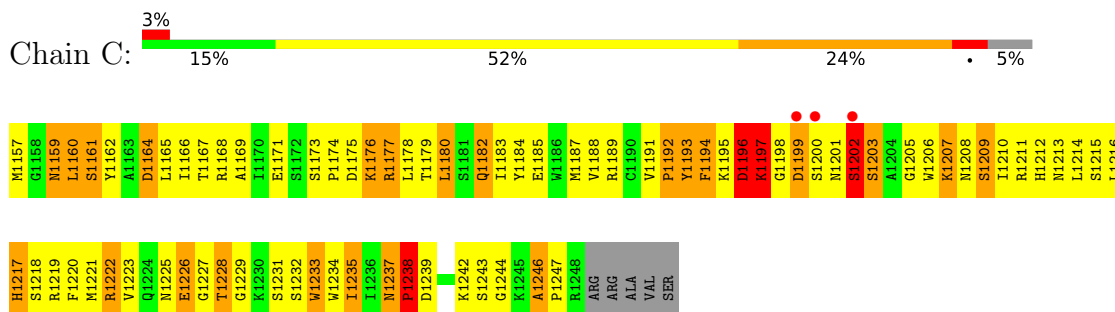
Chain C:

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

Chain B:

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

Chain D:

• Molecule 3: 5'-D(*GP*TP*TP*GP*TP*TP*TP*AP*CP*AP*TP*AP*G)-3'

Chain E:

G25 T26 T27 G28 T29 T30 T31 A32 C33 A34 T35 A36 G37

● Molecule 3: 5'-D(*GP*TP*TP*GP*TP*TP*TP*AP*CP*AP*TP*AP*G)-3'

Chain F: 8%   92%

G1025 T1026 T1027 G1028 T1029 T1030 T1031 A1032 C1033 A1034 T1035 A1036 G1037

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	41.96Å 41.96Å 354.82Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	24.13 – 2.70 24.12 – 2.61	Depositor EDS
% Data completeness (in resolution range)	94.6 (24.13-2.70) 92.2 (24.12-2.61)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.84 (at 2.60Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.242 , 0.266 0.240 , 0.271	Depositor DCC
R_{free} test set	495 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	24.2	Xtriage
Anisotropy	0.193	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.13 , 106.7	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.26, < L^2 > = 0.11$	Xtriage
Estimated twinning fraction	0.399 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	2741	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.04% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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	0.75	2/792 (0.3%)	1.22	15/1064 (1.4%)
1	C	0.63	1/721 (0.1%)	1.50	10/972 (1.0%)
2	B	0.98	3/294 (1.0%)	1.20	6/451 (1.3%)
2	D	0.59	0/294	0.79	0/451
3	E	0.78	1/296 (0.3%)	0.87	0/456
3	F	0.49	0/296	0.83	0/456
All	All	0.71	7/2693 (0.3%)	1.18	31/3850 (0.8%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1	DC	O3'-P	10.65	1.74	1.61
3	E	37	DG	P-O5'	-8.27	1.51	1.59
1	C	1203	SER	CA-CB	7.60	1.64	1.52
1	A	247	PRO	C-N	-7.59	1.16	1.34
1	A	245	LYS	CB-CG	6.80	1.71	1.52

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	1237	ASN	C-N-CD	-21.96	72.29	120.60
1	C	1237	ASN	C-N-CA	12.43	174.22	122.00
1	C	1202	SER	C-N-CA	11.24	149.79	121.70
2	B	2	DT	O5'-P-OP1	-10.15	96.57	105.70
1	C	1196	ASP	N-CA-C	9.49	136.63	111.00

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	A	774	0	764	101	0
1	C	704	0	659	100	0
2	B	262	0	148	24	0
2	D	262	0	148	29	0
3	E	265	0	151	26	0
3	F	265	0	151	37	0
4	A	46	0	0	3	0
4	B	15	0	0	0	0
4	C	72	0	0	1	0
4	D	33	0	0	0	0
4	E	22	0	0	0	0
4	F	21	0	0	0	0
All	All	2741	0	2021	285	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 64.

The worst 5 of 285 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)
2:D:1007:DA:H2"	2:D:1008:DA:H5'	1.26	1.14
1:C:1237:ASN:ND2	1:C:1238:PRO:HD3	1.68	1.07
1:A:237:ASN:HB2	1:A:238:PRO:HD2	1.11	1.05
1:A:160:LEU:HD12	1:A:165:LEU:HD22	1.39	1.05
2:B:5:DG:H2"	2:B:6:DT:H5"	1.37	1.03

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	
1	A	95/97 (98%)	64 (67%)	17 (18%)	14 (15%)	0	0
1	C	90/97 (93%)	65 (72%)	13 (14%)	12 (13%)	0	0
All	All	185/194 (95%)	129 (70%)	30 (16%)	26 (14%)	0	0

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	198	GLY
1	A	204	ALA
1	A	217	HIS
1	A	238	PRO
1	A	247	PRO

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	84/84 (100%)	64 (76%)	20 (24%)	0	2
1	C	74/84 (88%)	51 (69%)	23 (31%)	0	0
All	All	158/168 (94%)	115 (73%)	43 (27%)	0	1

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1194	PHE
1	C	1213	ASN
1	C	1195	LYS
1	C	1202	SER
1	C	1217	HIS

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	C	1201	ASN
1	C	1208	ASN
1	C	1237	ASN
1	C	1224	GLN
1	A	217	HIS

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 ⓘ

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	247:PRO	C	248:ARG	N	1.16

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	97/97 (100%)	0.40	9 (9%)	8	6	0, 19, 60, 70	0
1	C	92/97 (94%)	0.30	3 (3%)	46	46	1, 18, 48, 58	0
2	B	13/13 (100%)	-0.60	0	100	100	3, 8, 18, 20	0
2	D	13/13 (100%)	-0.41	0	100	100	9, 14, 24, 27	0
3	E	13/13 (100%)	-0.58	0	100	100	3, 8, 23, 23	0
3	F	13/13 (100%)	-0.32	0	100	100	4, 17, 26, 26	0
All	All	241/246 (97%)	0.17	12 (4%)	28	27	0, 17, 56, 70	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1200	SER	4.9
1	C	1199	ASP	4.7
1	A	228	THR	4.1
1	A	247	PRO	3.7
1	A	246	ALA	3.1

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 ⓘ

There are no such residues in this entry.