



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 02:38 pm GMT

PDB ID : 4S0H  
Title : TBX5 DB, NKX2.5 HD, ANF DNA Complex  
Authors : Pradhan, L.  
Deposited on : 2014-12-31  
Resolution : 2.82 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

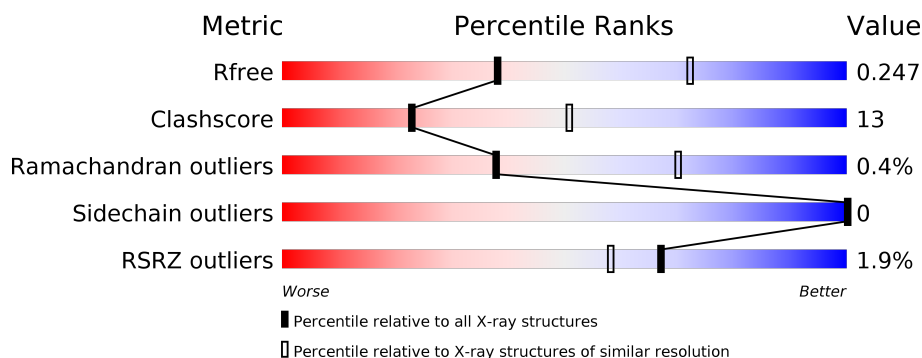
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.82 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.




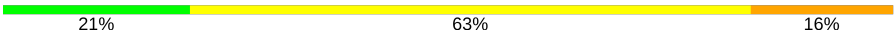
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	2917 (2.84-2.80)
Clashscore	112137	3382 (2.84-2.80)
Ramachandran outliers	110173	3324 (2.84-2.80)
Sidechain outliers	110143	3326 (2.84-2.80)
RSRZ outliers	101464	2948 (2.84-2.80)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	186	<div> <div>2%</div> <div> <div></div> <div>67%</div> <div>27%</div> <div>• 5%</div> </div> </div>
1	E	186	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>17%</div> <div>• 5%</div> </div> </div>
2	B	53	<div> <div></div> <div> <div></div> <div>85%</div> <div>15%</div> </div> </div>
2	F	53	<div> <div></div> <div> <div></div> <div>83%</div> <div>17%</div> </div> </div>
3	C	19	<div> <div></div> <div> <div></div> <div>47%</div> <div>47%</div> <div>5%</div> </div> </div>
3	G	19	<div> <div></div> <div> <div></div> <div>79%</div> <div>21%</div> </div> </div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	19	 42% 53% 5%
4	H	19	 21% 63% 16%

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 5350 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 T-box transcription factor TBX5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	176	Total	C	N	O	S	0	0	0
			1435	933	255	240	7			
1	E	177	Total	C	N	O	S	0	0	0
			1443	937	256	243	7			

- Molecule 2 is a protein called Homeobox protein Nkx-2.5.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	53	Total	C	N	O	0	0	0
			462	294	88	80			
2	F	53	Total	C	N	O	0	0	0
			462	294	88	80			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	193	SER	CYS	CONFLICT	UNP P52952
F	156	SER	CYS	CONFLICT	UNP P52952

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	19	Total	C	N	O	P	0	0	0
			384	185	67	114	18			
3	G	19	Total	C	N	O	P	0	0	0
			384	185	67	114	18			

- Molecule 4 is a DNA chain called 5'-D(\*CP\*CP\*AP\*CP\*TP\*TP\*CP\*AP\*AP\*AP\*GP\*GP\*TP\*GP\*TP\*GP\*AP\*GP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	19	Total 389	C 186	N 75	O 110	P 18	0	0	0
4	H	19	Total 389	C 186	N 75	O 110	P 18	0	0	0

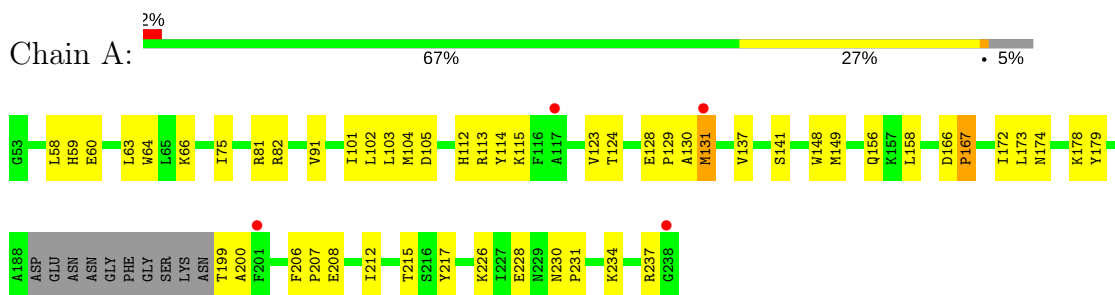
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	F	1	Total 1	O 1	0	0
5	C	1	Total 1	O 1	0	0

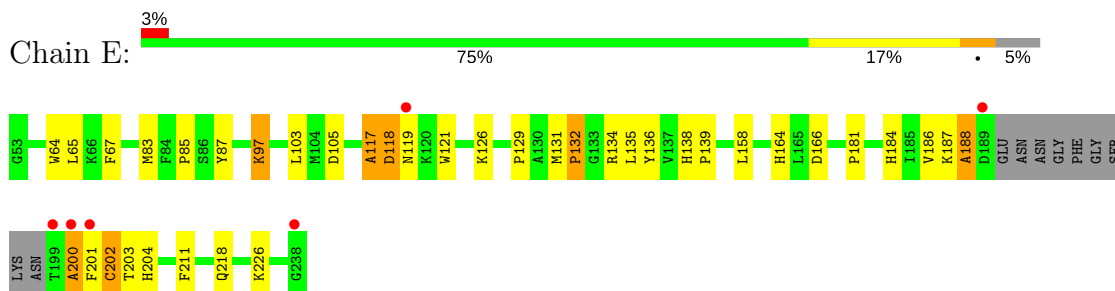
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

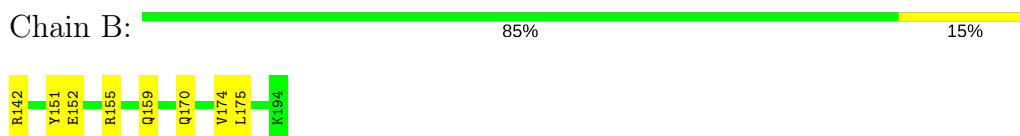
- Molecule 1: T-box transcription factor TBX5



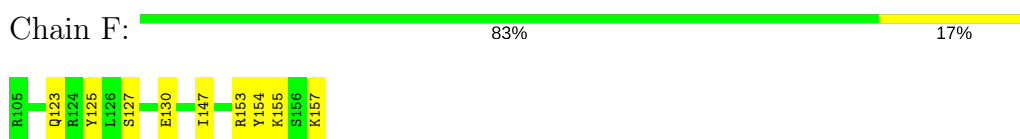
- Molecule 1: T-box transcription factor TBX5



- Molecule 2: Homeobox protein Nkx-2.5



- Molecule 2: Homeobox protein Nkx-2.5



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

T1	C2	T3	C4	A5	C6	A7	C8	C9	T10	T11	T12	G13	G19
----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----

A diagram of a 4x4 grid. The central 2x2 area is yellow and contains the labels C8, C9, T10, and T11. The top and bottom rows are green and contain the labels T1 and G19. The left and right columns are green and contain the labels C8 and C9. The grid is connected by lines.

C1	A8
C2	A9
A3	A10
C4	G11
T5	G12
	T13
	G14
	T15
	G16
	A17
	G18
	A19

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.22Å 78.46Å 78.69Å 90.00° 108.83° 90.00°	Depositor
Resolution (Å)	38.16 – 2.82 43.15 – 2.82	Depositor EDS
% Data completeness (in resolution range)	91.6 (38.16-2.82) 92.1 (43.15-2.82)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.42 (at 2.81Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.189 , 0.247 0.190 , 0.247	Depositor DCC
$R_{free}$ test set	1807 reflections (10.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.9	Xtriage
Anisotropy	1.016	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 40.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	5350	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 [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	A	0.52	0/1478	0.67	1/1996 (0.1%)
1	E	0.58	1/1486 (0.1%)	0.81	7/2007 (0.3%)
2	B	0.45	0/470	0.60	0/627
2	F	0.51	0/470	0.56	0/627
3	C	0.94	1/429 (0.2%)	1.13	0/660
3	G	0.95	0/429	1.10	0/660
4	D	0.91	0/437	1.13	3/673 (0.4%)
4	H	1.13	2/437 (0.5%)	1.23	3/673 (0.4%)
All	All	0.71	4/5636 (0.1%)	0.88	14/7923 (0.2%)

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

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	97	LYS	CE-NZ	8.75	1.71	1.49
3	C	6	DC	C1'-N1	6.28	1.57	1.49
4	H	13	DT	C3'-O3'	5.19	1.50	1.44
4	H	3	DA	C3'-O3'	-5.07	1.37	1.44

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	119	ASN	N-CA-CB	-14.19	85.06	110.60
4	H	13	DT	O4'-C4'-C3'	-10.67	99.60	106.00
1	E	118	ASP	CB-CA-C	10.17	130.74	110.40
1	A	131	MET	CG-SD-CE	9.86	115.97	100.20

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	H	4	DC	O4'-C4'-C3'	-7.29	101.58	104.50

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	117	ALA	Peptide
1	E	200	ALA	Mainchain
1	E	201	PHE	Peptide

## 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	1435	0	1451	47	0
1	E	1443	0	1455	35	0
2	B	462	0	477	7	0
2	F	462	0	477	15	0
3	C	384	0	217	9	0
3	G	384	0	217	5	0
4	D	389	0	215	8	0
4	H	389	0	215	17	0
5	C	1	0	0	0	0
5	F	1	0	0	0	0
All	All	5350	0	4724	128	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 13.

The worst 5 of 128 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:E:97:LYS:NZ	1:E:97:LYS:CE	1.70	1.51
1:A:131:MET:CE	1:A:179:TYR:OH	1.74	1.35
1:A:131:MET:SD	1:A:173:LEU:HA	1.79	1.23

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:131:MET:HE1	1:A:179:TYR:OH	1.28	1.20
1:A:131:MET:HE1	1:A:179:TYR:CZ	1.79	1.18

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	172/186 (92%)	157 (91%)	14 (8%)	1 (1%)	28	61
1	E	173/186 (93%)	161 (93%)	11 (6%)	1 (1%)	28	61
2	B	51/53 (96%)	46 (90%)	5 (10%)	0	100	100
2	F	51/53 (96%)	47 (92%)	4 (8%)	0	100	100
All	All	447/478 (94%)	411 (92%)	34 (8%)	2 (0%)	38	70

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	132	PRO
1	A	167	PRO

### 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	154/162 (95%)	154 (100%)	0	100	100
1	E	155/162 (96%)	155 (100%)	0	100	100
2	B	50/50 (100%)	50 (100%)	0	100	100
2	F	50/50 (100%)	50 (100%)	0	100	100
All	All	409/424 (96%)	409 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains 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 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](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	176/186 (94%)	0.10	4 (2%) 61 51	30, 52, 76, 100	0
1	E	177/186 (95%)	0.11	6 (3%) 46 34	31, 50, 81, 125	0
2	B	53/53 (100%)	0.14	0 100 100	38, 55, 77, 82	0
2	F	53/53 (100%)	0.12	0 100 100	37, 50, 71, 76	0
3	C	19/19 (100%)	0.21	0 100 100	38, 50, 68, 69	0
3	G	19/19 (100%)	-0.17	0 100 100	39, 55, 61, 65	0
4	D	19/19 (100%)	0.01	0 100 100	40, 50, 74, 84	0
4	H	19/19 (100%)	0.04	0 100 100	39, 53, 66, 73	0
All	All	535/554 (96%)	0.10	10 (1%) 67 58	30, 52, 76, 125	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	199	THR	10.5
1	E	200	ALA	4.7
1	A	201	PHE	3.7
1	E	201	PHE	3.7
1	E	238	GLY	3.7

### 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 carbohydrates in this entry.

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.