



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 08:31 PM GMT

PDB ID : 1KOH  
Title : THE CRYSTAL STRUCTURE AND MUTATIONAL ANALYSIS OF A NOVEL RNA-BINDING DOMAIN FOUND IN THE HUMAN TAP NUCLEAR MRNA EXPORT FACTOR  
Authors : Ho, D.N.; Coburn, G.A.; Kang, Y.; Cullen, B.R.; Georgiadis, M.M.  
Deposited on : 2001-12-20  
Resolution : 3.80 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

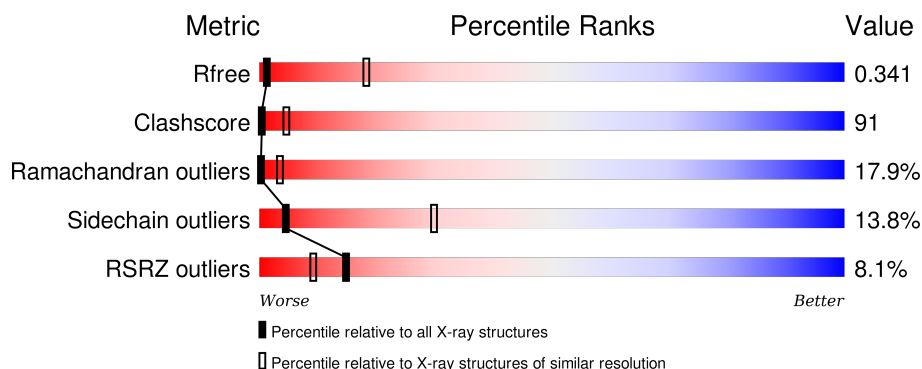
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.80 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))
$R_{free}$	91344	1317 (4.10-3.50)
Clashscore	102246	1458 (4.10-3.50)
Ramachandran outliers	100387	1397 (4.10-3.50)
Sidechain outliers	100360	1392 (4.10-3.50)
RSRZ outliers	91569	1325 (4.10-3.50)

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  $\geq 3$ , 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	277	<div> <div>4%</div> <div>9%</div> <div>58%</div> <div>21%</div> <div>5%</div> <div>7%</div> </div>
1	B	277	<div> <div>3%</div> <div>9%</div> <div>37%</div> <div>13%</div> <div>40%</div> </div>
1	C	277	<div> <div>17%</div> <div>18%</div> <div>56%</div> <div>18%</div> <div>6%</div> </div>
1	D	277	<div> <div>8%</div> <div>41%</div> <div>12%</div> <div>38%</div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 6854 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 TIP ASSOCIATING PROTEIN.

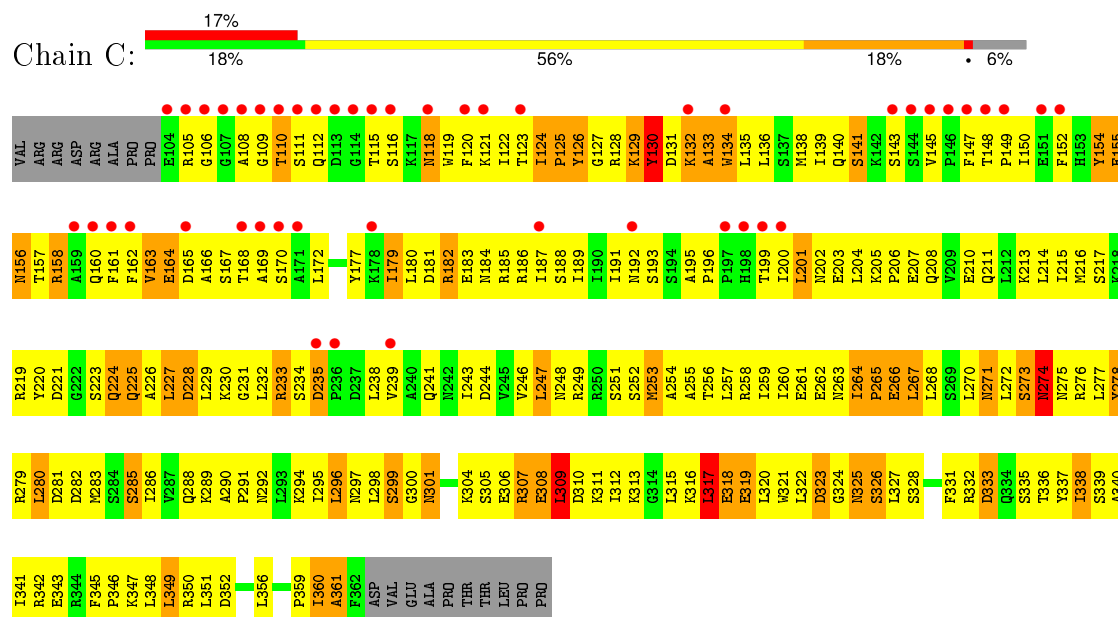
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	258	Total	C	N	O	S	0	0	0
			2067	1305	365	393	4			
1	B	167	Total	C	N	O	S	0	0	0
			1337	839	235	260	3			
1	C	259	Total	C	N	O	S	0	0	0
			2076	1310	366	396	4			
1	D	172	Total	C	N	O	S	0	0	0
			1374	863	240	268	3			

There are 12 discrepancies between the modelled and reference sequences:

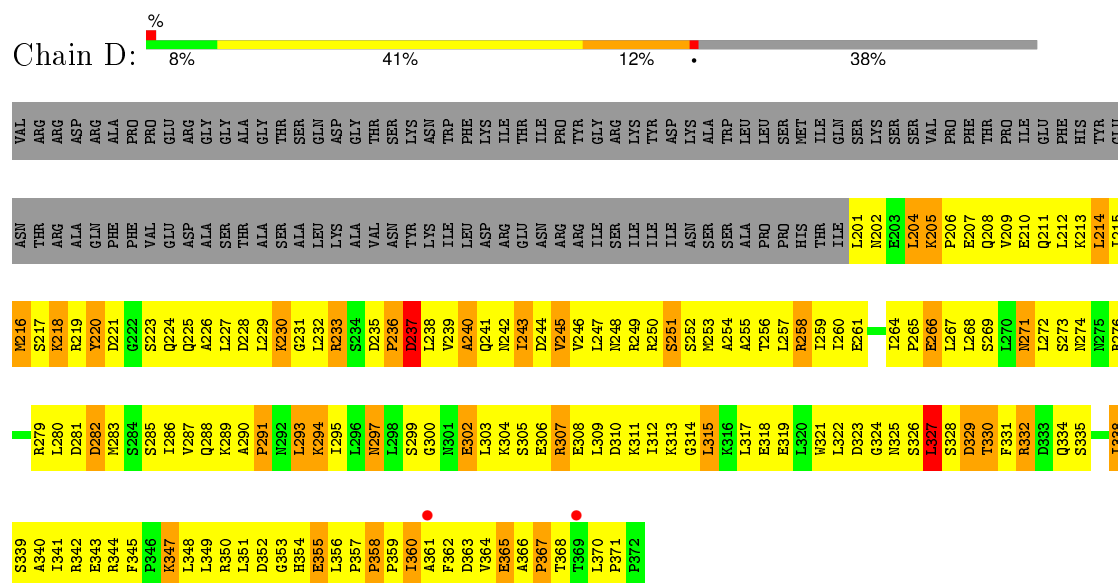
Chain	Residue	Modelled	Actual	Comment	Reference
A	143	SER	CYS	ENGINEERED	UNP Q9UBU9
A	252	SER	CYS	ENGINEERED	UNP Q9UBU9
A	328	SER	CYS	ENGINEERED	UNP Q9UBU9
B	143	SER	CYS	ENGINEERED	UNP Q9UBU9
B	252	SER	CYS	ENGINEERED	UNP Q9UBU9
B	328	SER	CYS	ENGINEERED	UNP Q9UBU9
C	143	SER	CYS	ENGINEERED	UNP Q9UBU9
C	252	SER	CYS	ENGINEERED	UNP Q9UBU9
C	328	SER	CYS	ENGINEERED	UNP Q9UBU9
D	143	SER	CYS	ENGINEERED	UNP Q9UBU9
D	252	SER	CYS	ENGINEERED	UNP Q9UBU9
D	328	SER	CYS	ENGINEERED	UNP Q9UBU9



• Molecule 1: TIP ASSOCIATING PROTEIN



• Molecule 1: TIP ASSOCIATING PROTEIN



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	136.78Å 136.78Å 202.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 3.80 37.81 – 3.70	Depositor EDS
% Data completeness (in resolution range)	94.9 (8.00-3.80) 94.0 (37.81-3.70)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.14 (at 3.66Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.275 , 0.355 0.265 , 0.341	Depositor DCC
$R_{free}$ test set	481 reflections (2.92%)	DCC
Wilson B-factor (Å <sup>2</sup> )	133.4	Xtriage
Anisotropy	0.009	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 70.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 21079 reflections	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	6854	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	89.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.05% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.46	0/2103	0.74	1/2840 (0.0%)
1	B	0.45	0/1354	0.73	0/1827
1	C	0.39	0/2112	0.67	0/2852
1	D	0.49	0/1393	0.77	0/1882
All	All	0.45	0/6962	0.73	1/9401 (0.0%)

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	A	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	229	LEU	CA-CB-CG	5.05	126.92	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	278	TYR	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	A	2067	0	2107	409	0
1	B	1337	0	1376	274	0
1	C	2076	0	2113	365	0
1	D	1374	0	1415	243	0
All	All	6854	0	7011	1257	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 91.

The worst 5 of 1257 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:D:365:GLU:HG2	1:D:366:ALA:H	1.15	1.08
1:B:338:ILE:HA	1:B:341:ILE:HD11	1.30	1.08
1:A:334:GLN:HA	1:A:334:GLN:HE21	1.18	1.07
1:C:233:ARG:HH21	1:C:246:VAL:HG12	1.20	1.06
1:A:278:TYR:HD1	1:A:279:ARG:H	1.07	0.97

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	256/277 (92%)	137 (54%)	60 (23%)	59 (23%)	0	1
1	B	165/277 (60%)	98 (59%)	38 (23%)	29 (18%)	0	3
1	C	257/277 (93%)	133 (52%)	74 (29%)	50 (20%)	0	3
1	D	170/277 (61%)	110 (65%)	46 (27%)	14 (8%)	1	18
All	All	848/1108 (76%)	478 (56%)	218 (26%)	152 (18%)	0	3



5 of 152 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	113	ASP
1	A	123	THR
1	A	139	ILE
1	A	148	THR
1	A	182	ARG

### 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	232/249 (93%)	196 (84%)	36 (16%)	3	24
1	B	153/249 (61%)	132 (86%)	21 (14%)	4	30
1	C	233/249 (94%)	211 (91%)	22 (9%)	11	47
1	D	158/249 (64%)	130 (82%)	28 (18%)	2	17
All	All	776/996 (78%)	669 (86%)	107 (14%)	4	30

5 of 107 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	306	GLU
1	C	154	TYR
1	D	330	THR
1	B	315	LEU
1	B	341	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	274	ASN
1	C	112	GLN
1	D	225	GLN
1	B	288	GLN
1	B	297	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](#)

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	258/277 (93%)	0.11	12 (4%) 35 23	47, 81, 120, 130	0
1	B	167/277 (60%)	0.24	8 (4%) 34 22	46, 83, 111, 120	0
1	C	259/277 (93%)	0.64	47 (18%) 2 2	58, 104, 142, 148	0
1	D	172/277 (62%)	0.05	2 (1%) 81 67	46, 72, 111, 122	0
All	All	856/1108 (77%)	0.28	69 (8%) 15 9	46, 84, 138, 148	0

The worst 5 of 69 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	113	ASP	6.4
1	B	366	ALA	5.2
1	B	201	LEU	4.9
1	C	160	GLN	4.9
1	A	106	GLY	4.6

### 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 [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.