



# wwPDB X-ray Structure Validation Summary Report i

Mar 31, 2014 – 03:12 PM BST

PDB ID : 4NQF  
Title : Crystal structure of HEPN domain protein  
Authors : Padavannil, A.; Jobichen, C.; Sivaraman, J.  
Deposited on : 2013-11-25  
Resolution : 2.70 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

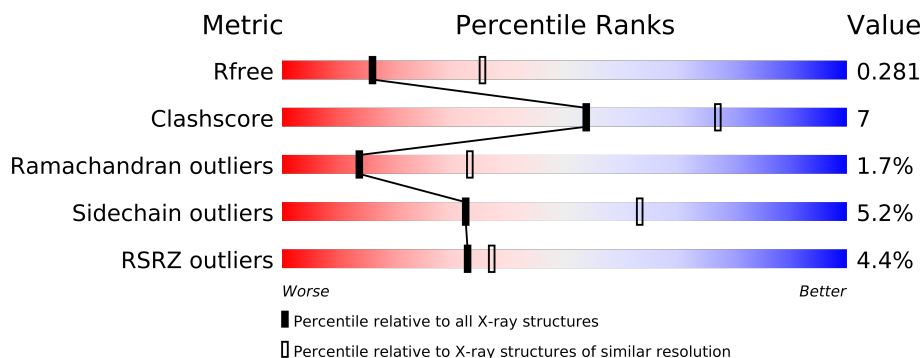
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable23004
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable23004

# 1 Overall quality at a glance



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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	66092	1557 (2.70-2.70)
Clashscore	79885	1939 (2.70-2.70)
Ramachandran outliers	78287	1905 (2.70-2.70)
Sidechain outliers	78261	1905 (2.70-2.70)
RSRZ outliers	66119	1559 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	158	
1	B	158	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2453 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Nucleotidyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	145	Total	C	N	O	S	Se	0	0	0
			1197	759	217	216	2	3			
1	B	151	Total	C	N	O	S	Se	0	0	0
			1230	779	224	222	2	3			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	168	MSE	LEU	ENGINEERED MUTATION	UNP Q7CW19
A	219	MSE	LEU	ENGINEERED MUTATION	UNP Q7CW19
A	251	MSE	LEU	ENGINEERED MUTATION	UNP Q7CW19
B	168	MSE	LEU	ENGINEERED MUTATION	UNP Q7CW19
B	219	MSE	LEU	ENGINEERED MUTATION	UNP Q7CW19
B	251	MSE	LEU	ENGINEERED MUTATION	UNP Q7CW19

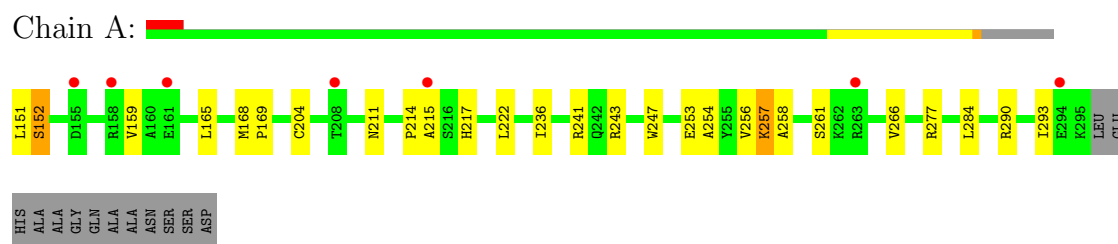
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	12	Total	O	0	0
			12	12		
2	B	14	Total	O	0	0
			14	14		

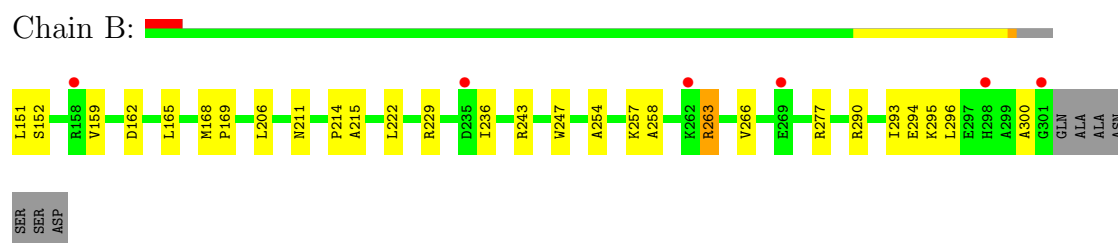
### 3 Residue-property plots

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.

- Molecule 1: Nucleotidyltransferase



- Molecule 1: Nucleotidyltransferase



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	63.10Å 71.57Å 83.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.04 – 2.70 36.04 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.0 (36.04-2.70) 99.1 (36.04-2.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.05	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	30.52 (at 2.68Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.218 , 0.280 0.223 , 0.281	Depositor DCC
$R_{free}$ test set	1075 reflections (11.12%)	DCC
Wilson B-factor (Å <sup>2</sup> )	58.7	Xtriage
Anisotropy	0.350	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 43.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 10746 reflections (0.009%)	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	2453	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 25.90 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 2.9071e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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

## 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	0.57	0/1223	0.68	0/1652
1	B	0.60	0/1257	0.74	1/1700 (0.1%)
All	All	0.58	0/2480	0.71	1/3352 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	229	ARG	NE-CZ-NH1	5.34	122.97	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Close contacts

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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1197	0	1161	19	1
1	B	1230	0	1183	14	1
2	A	12	0	0	2	0
2	B	14	0	0	0	0
All	All	2453	0	2344	33	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 7.

The worst 5 of 33 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:159:VAL:HG11	1:A:211:ASN:HD21	1.42	0.82
1:B:159:VAL:HG11	1:B:211:ASN:HD21	1.43	0.81
1:A:168:MSE:CE	1:A:204:CYS:SG	2.73	0.77
1:A:168:MSE:HE2	1:A:204:CYS:SG	2.24	0.77
1:A:151:LEU:N	2:A:406:HOH:O	2.23	0.70

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:241:ARG:NH2	1:B:162:ASP:OD1[2_445]	2.17	0.03

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	143/158 (90%)	135 (94%)	6 (4%)	2 (1%)	16	41
1	B	149/158 (94%)	137 (92%)	9 (6%)	3 (2%)	11	28
All	All	292/316 (92%)	272 (93%)	15 (5%)	5 (2%)	14	33

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	152	SER
1	B	152	SER
1	B	300	ALA
1	B	257	LYS
1	A	257	LYS

### 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	124/131 (95%)	117 (94%)	7 (6%)	30	59
1	B	125/131 (95%)	119 (95%)	6 (5%)	35	68
All	All	249/262 (95%)	236 (95%)	13 (5%)	32	63

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

Mol	Chain	Res	Type
1	A	266	VAL
1	A	284	LEU
1	B	266	VAL
1	A	261	SER
1	B	263	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	211	ASN
1	B	163	HIS
1	B	211	ASN
1	B	275	GLN
1	B	298	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates 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

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	145/158 (91%)	0.34	7 (4%) 29 33	38, 69, 109, 127	0
1	B	151/158 (95%)	0.14	6 (3%) 36 41	34, 58, 94, 111	0
All	All	296/316 (93%)	0.24	13 (4%) 33 37	34, 63, 104, 127	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	262	LYS	4.0
1	B	158	ARG	2.8
1	B	301	GLY	2.6
1	A	294	GLU	2.5
1	A	155	ASP	2.4

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

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.