



# wwPDB X-ray Structure Validation Summary Report

Apr 6, 2014 – 07:19 PM EDT

PDB ID : 4O1R  
Title : Crystal structure of NpuDnaB intein  
Authors : Aranko, A.S.; Oeemig, J.S.; Zhou, D.; Kajander, T.; Wlodawer, A.; Iwai, H.  
Deposited on : 2013-12-16  
Resolution : 1.40 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>

---

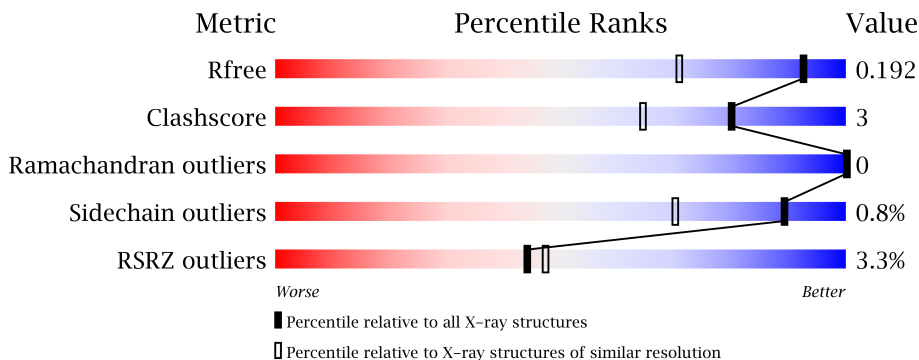
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	<b>FAILED</b>
Xtriage (Phenix)	:	dev-1439
EDS	:	stable22978
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)	:	stable22978

# 1 Overall quality at a glance

The reported resolution of this entry is 1.40 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	1097 (1.42-1.38)
Clashscore	79885	1246 (1.42-1.38)
Ramachandran outliers	78287	1206 (1.42-1.38)
Sidechain outliers	78261	1205 (1.42-1.38)
RSRZ outliers	66119	1097 (1.42-1.38)

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	142	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	NA	A	203	-	X
3	CL	A	207	-	X
4	GOL	A	208	-	X
4	GOL	A	209	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 1451 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 Replicative DNA helicase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	142	Total	C	N	O	S	0	17	0
			1154	740	197	216	1			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	EXPRESSION TAG	UNP B2IVH9
A	-1	GLY	-	EXPRESSION TAG	UNP B2IVH9
A	1	ALA	CYS	CONFLICT	UNP B2IVH9
A	94	ALA	CYS	CONFLICT	UNP B2IVH9
A	99	SER	-	EXPRESSION TAG	UNP B2IVH9
A	100	GLY	-	EXPRESSION TAG	UNP B2IVH9
A	103	ILE	VAL	CONFLICT	UNP B2IVH9
A	112	THR	GLU	CONFLICT	UNP B2IVH9

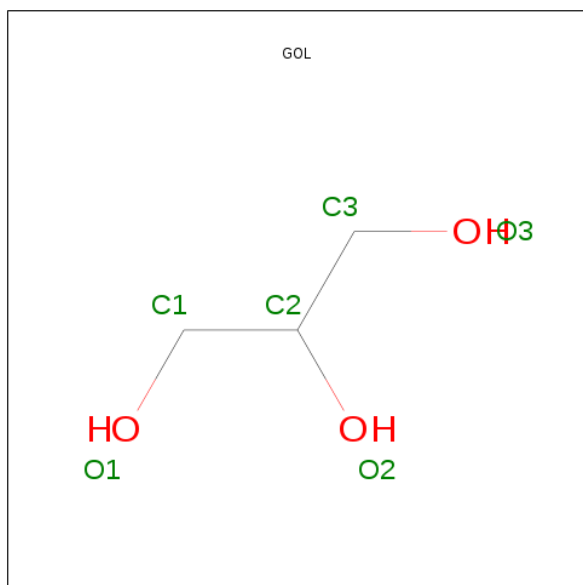
- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Na	0	0
			3	3		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	4	Total	Cl	0	0
			4	4		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			6	3	3		
4	A	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	278	Total	O	0	0
			278	278		

i

- Molecule 1: Replicative DNA helicase

Schematic representation of the protein structure of the C-terminal domain of the human protein. The protein is shown as a series of yellow rectangular blocks representing  $\alpha$ -helices and green rectangular blocks representing  $\beta$ -strands. The residues are labeled from S-2 to N139. Red dots above the residues indicate specific sites of interest: S-2, A36, N98, and S99.

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.05Å 61.05Å 122.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	61.05 – 1.40 29.62 – 1.40	Depositor EDS
% Data completeness (in resolution range)	99.8 (61.05-1.40) 94.6 (29.62-1.40)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.47 (at 1.40Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.161 , 0.189 0.161 , 0.192	Depositor DCC
$R_{free}$ test set	894 reflections (2.08%)	DCC
Wilson B-factor (Å <sup>2</sup> )	16.7	Xtriage
Anisotropy	0.450	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 48.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 46331 reflections	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	1451	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.19% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, CL, NA

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	0/1252	0.64	0/1696

There are no bond length outliers.

There are no bond angle outliers.

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	1154	0	0	4	1
2	A	3	0	0	0	0
3	A	4	0	0	0	0
4	A	12	0	0	0	0
5	A	278	0	0	4	2
All	All	1451	0	0	4	2

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 3.

All (4) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:82:LYS:NZ	5:A:418:HOH:O	2.50	0.44
1:A:54[A]:LYS:CE	5:A:445:HOH:O	2.66	0.43
1:A:82:LYS:CE	5:A:418:HOH:O	2.66	0.43
1:A:20[B]:LYS:NZ	5:A:427:HOH:O	2.52	0.41

All (2) 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)
5:A:306:HOH:O	5:A:307:HOH:O[5_555]	1.30	0.90
1:A:90:LYS:NZ	5:A:549:HOH:O[7_645]	1.97	0.23

## 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	156/142 (110%)	155 (99%)	1 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	134/119 (113%)	132 (98%)	2 (2%)	76	46

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	63[A]	LEU
1	A	63[B]	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

Mogul failed to run properly - this section will therefore be empty.

## 5.5 Carbohydrates ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.6 Ligand geometry ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.7 Other polymers ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 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	142/142 (100%)	-0.13	4 (2%) 50 54	10, 16, 30, 41	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	98	ASN	4.4
1	A	-2	SER	3.9
1	A	99	SER	3.6
1	A	36	ALA	2.2

### 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 ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	NA	A	203	1/1	0.32	12.09	32,32,32,32	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	GOL	A	208	6/6	0.22	11.19	29,39,40,41	0
3	CL	A	207	1/1	0.10	4.53	32,32,32,32	0
4	GOL	A	209	6/6	0.14	4.37	28,29,30,32	0
2	NA	A	202	1/1	0.11	1.03	23,23,23,23	0
2	NA	A	201	1/1	0.07	-0.13	14,14,14,14	0
3	CL	A	204	1/1	0.05	-1.24	15,15,15,15	0
3	CL	A	206	1/1	0.04	-2.00	21,21,21,21	0
3	CL	A	205	1/1	0.02	-20.50	14,14,14,14	1

## 6.5 Other polymers ⓘ

There are no such residues in this entry.