



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 12:32 am BST

PDB ID : 5NT7  
Title : Structure of the LOTUS domain of Oskar in complex with the C-terminal RecA-like domain of Vasa  
Authors : Jeske, M.; Ephrussi, A.; Mueller, C.W.  
Deposited on : 2017-04-27  
Resolution : 1.40 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

<https://www.wwpdb.org/validation/2017/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) : 1.13  
EDS : 2.11  
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) : 2.11

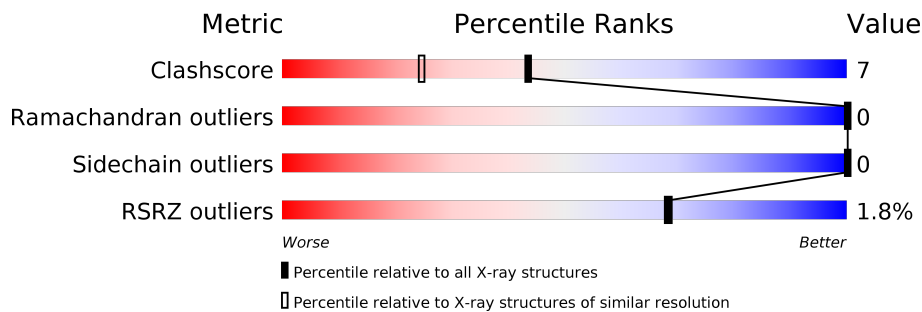
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 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))
Clashscore	141614	1812 (1.40-1.40)
Ramachandran outliers	138981	1763 (1.40-1.40)
Sidechain outliers	138945	1762 (1.40-1.40)
RSRZ outliers	127900	1674 (1.40-1.40)

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 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  $\leq 5\%$ . 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	B	163	 2% 86% 8% 6%
1	D	163	 2% 81% 13% 6%
2	A	107	 % 80% 6% 14%
2	C	107	 % 82% • 14%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4466 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 ATP-dependent RNA helicase vasa, isoform A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	153	1204	757	213	231	3	0	1	0
1	D	153	1207	759	215	230	3	0	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	461	GLY	-	expression tag	UNP P09052
B	462	HIS	-	expression tag	UNP P09052
D	461	GLY	-	expression tag	UNP P09052
D	462	HIS	-	expression tag	UNP P09052

- Molecule 2 is a protein called Maternal effect protein oskar.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	92	719	454	123	140	2	0	0	0
2	C	92	723	456	123	142	2	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	134	GLY	-	expression tag	UNP P25158
A	135	PRO	-	expression tag	UNP P25158
A	136	LEU	-	expression tag	UNP P25158
A	137	GLY	-	expression tag	UNP P25158
A	138	SER	-	expression tag	UNP P25158
C	134	GLY	-	expression tag	UNP P25158
C	135	PRO	-	expression tag	UNP P25158
C	136	LEU	-	expression tag	UNP P25158

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	137	GLY	-	expression tag	UNP P25158
C	138	SER	-	expression tag	UNP P25158

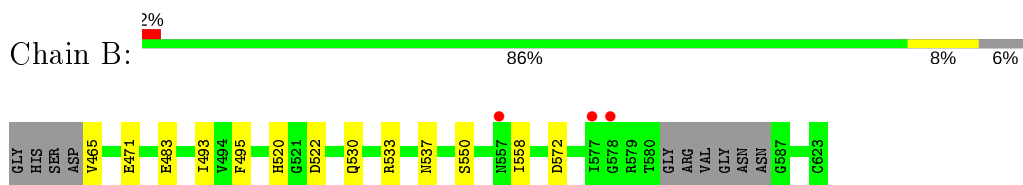
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	192	Total O 192 192	0	0
3	A	119	Total O 119 119	0	0
3	C	116	Total O 116 116	0	0
3	D	186	Total O 186 186	0	0

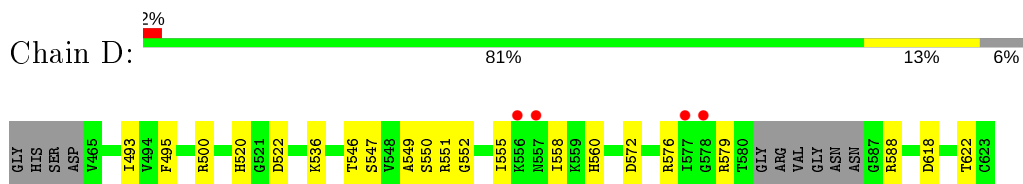
### 3 Residue-property plots [i](#)

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.

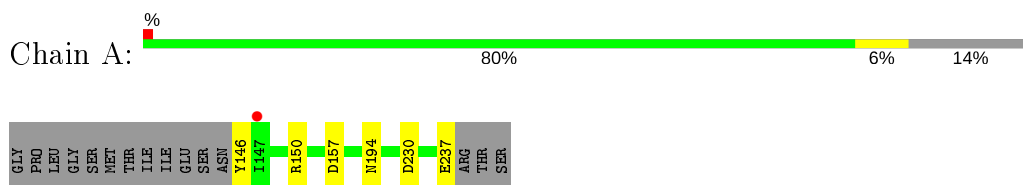
- Molecule 1: ATP-dependent RNA helicase vasa, isoform A



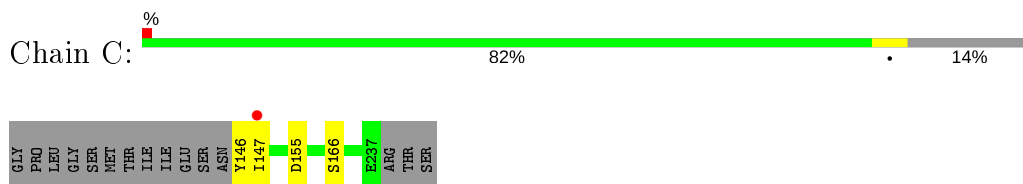
- Molecule 1: ATP-dependent RNA helicase vasa, isoform A



- Molecule 2: Maternal effect protein oskar



- Molecule 2: Maternal effect protein oskar



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	38.95Å 39.04Å 97.48Å 89.28° 91.85° 99.26°	Depositor
Resolution (Å)	48.71 – 1.40 48.71 – 1.40	Depositor EDS
% Data completeness (in resolution range)	91.4 (48.71-1.40) 91.4 (48.71-1.40)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.03 (at 1.40Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.169 , 0.199 0.171 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	21.4	Xtriage
Anisotropy	0.189	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 45.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.013 for -h,-k,l 0.015 for -k,-h,-l 0.156 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	4466	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.91% 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

### 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	B	0.33	0/1221	0.50	0/1642
1	D	0.42	0/1224	0.53	0/1645
2	A	0.31	0/730	0.51	0/988
2	C	0.32	0/734	0.49	0/993
All	All	0.35	0/3909	0.51	0/5268

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-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 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	B	1204	0	1202	10	0
1	D	1207	0	1209	23	0
2	A	719	0	723	13	0
2	C	723	0	727	7	0
3	A	119	0	0	11	2
3	B	192	0	0	6	2
3	C	116	0	0	6	2
3	D	186	0	0	12	1
All	All	4466	0	3861	52	4

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 7.

The worst 5 of 52 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:C:146:TYR:N	3:C:301:HOH:O	1.78	1.14
1:B:537[A]:ASN:ND2	3:B:702:HOH:O	1.80	1.12
1:D:549:ALA:O	3:D:701:HOH:O	1.67	1.10
1:D:546:THR:O	3:D:702:HOH:O	1.70	1.07
2:C:146:TYR:CA	3:C:301:HOH:O	2.00	1.06

All (4) 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	Interatomic distance (Å)	Clash overlap (Å)
3:B:753:HOH:O	3:B:787:HOH:O[1_455]	1.66	0.54
3:A:396:HOH:O	3:C:388:HOH:O[1_565]	1.70	0.50
3:A:401:HOH:O	3:C:410:HOH:O[1_565]	2.05	0.15
3:B:778:HOH:O	3:D:818:HOH:O[1_464]	2.19	0.01

## 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	B	150/163 (92%)	147 (98%)	3 (2%)	0	100	100
1	D	150/163 (92%)	148 (99%)	2 (1%)	0	100	100
2	A	90/107 (84%)	88 (98%)	2 (2%)	0	100	100
2	C	90/107 (84%)	88 (98%)	2 (2%)	0	100	100
All	All	480/540 (89%)	471 (98%)	9 (2%)	0	100	100

There are no Ramachandran outliers to report.



### 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	B	130/140 (93%)	130 (100%)	0	100	100
1	D	130/140 (93%)	130 (100%)	0	100	100
2	A	82/97 (84%)	82 (100%)	0	100	100
2	C	83/97 (86%)	83 (100%)	0	100	100
All	All	425/474 (90%)	425 (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

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	B	153/163 (93%)	0.17	3 (1%) 65 65	17, 25, 49, 58	0
1	D	153/163 (93%)	0.18	4 (2%) 56 55	17, 25, 49, 59	0
2	A	92/107 (85%)	0.08	1 (1%) 80 79	19, 25, 41, 48	0
2	C	92/107 (85%)	0.11	1 (1%) 80 79	20, 26, 40, 50	0
All	All	490/540 (90%)	0.15	9 (1%) 68 68	17, 25, 47, 59	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	578	GLY	3.8
2	C	147	ILE	2.8
1	D	577	ILE	2.5
1	D	556	LYS	2.4
1	B	577	ILE	2.4

### 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

There are no such residues in this entry.