



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 12:46 AM GMT

PDB ID : 2BGG  
Title : THE STRUCTURE OF A PIWI PROTEIN FROM ARCHAEOGLOBUS  
FULGIDUS COMPLEXED WITH A 16NT SIRNA DUPLEX.  
Authors : Parker, J.S.; Roe, S.M.; Barford, D.  
Deposited on : 2004-12-22  
Resolution : 2.20 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

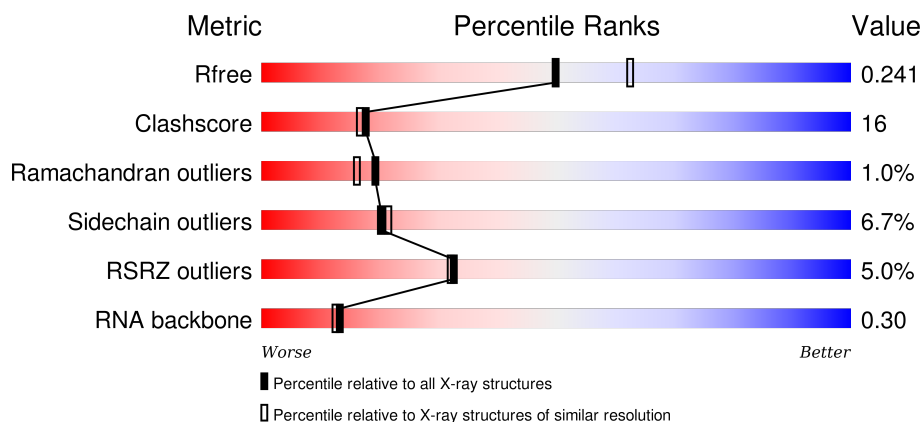
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)
RNA backbone	2183	1062 (2.80-1.60)

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	427	<div> <div>3%</div> <div>70% 18% 7%</div> </div>
1	B	427	<div> <div>5%</div> <div>65% 22% 5% 8%</div> </div>
2	P	8	<div> <div>38%</div> <div>13% 63% 25%</div> </div>
2	R	8	<div> <div>38%</div> <div>38% 50% 13%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	Q	8	<div><div></div><div></div><div></div><div></div></div> <div>13%13%25%25%38%</div>
3	S	8	<div><div></div><div></div><div></div><div></div></div> <div>25%25%13%38%</div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7557 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 PROTEIN AF1318.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	395	Total	C	N	O	S	0	2	0
			3237	2097	535	596	9			
1	B	394	Total	C	N	O	S	0	0	0
			3193	2065	530	589	9			

- Molecule 2 is a RNA chain called 5'-R(\*UP\*UP\*CP\*GP\*AP\*CP\*GP\*CP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	8	Total	C	N	O	P	0	0	0
			169	75	28	58	8			
2	R	8	Total	C	N	O	P	0	0	0
			169	75	28	58	8			

- Molecule 3 is a RNA chain called 5'-R(\*GP\*UP\*CP\*GP\*AP\*AP\*UP\*UP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	Q	8	Total	C	N	O	P	0	0	0
			170	76	29	57	8			
3	S	5	Total	C	N	O	P	0	0	0
			108	48	20	35	5			

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mn	0	0
			1	1		
4	A	1	Total	Mn	0	0
			1	1		

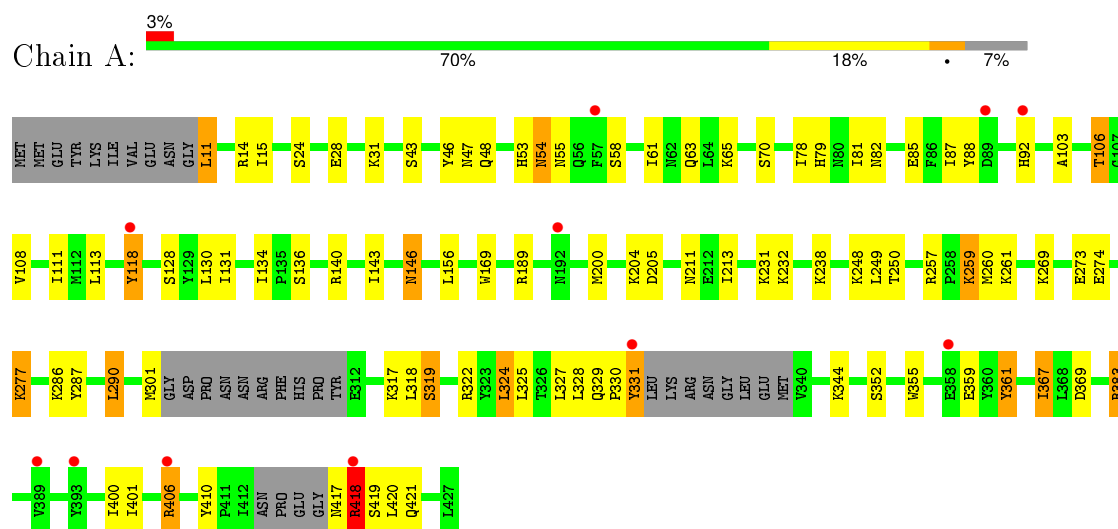
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	283	Total 283	O 283	0	0
5	B	188	Total 188	O 188	0	0
5	P	12	Total 12	O 12	0	0
5	Q	20	Total 20	O 20	0	0
5	R	3	Total 3	O 3	0	0
5	S	3	Total 3	O 3	0	0

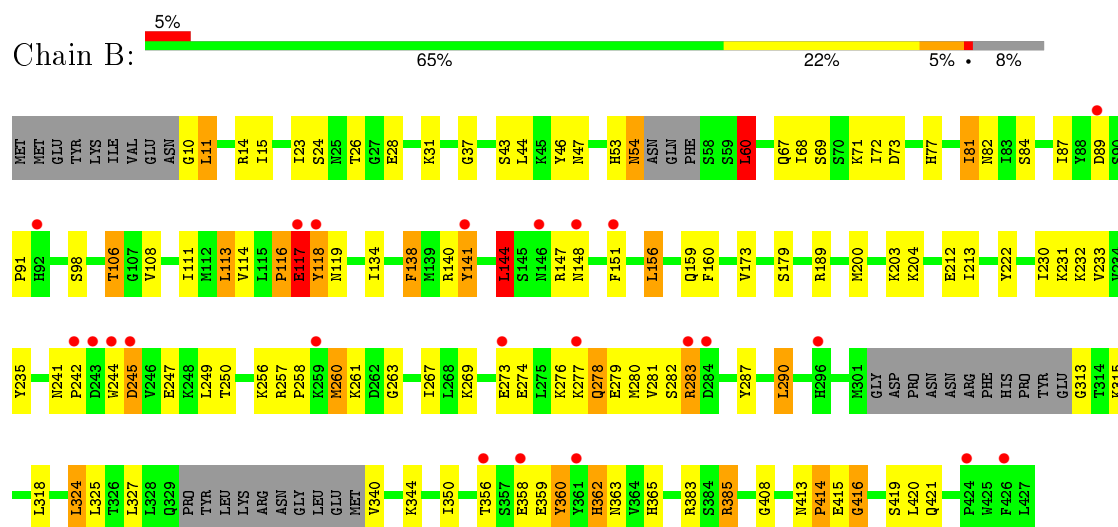
### 3 Residue-property plots

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.

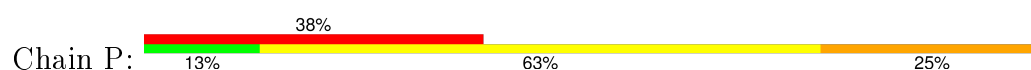
#### • Molecule 1: PROTEIN AF1318



#### • Molecule 1: PROTEIN AF1318

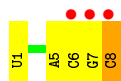


#### • Molecule 2: 5'-R(\*UP\*UP\*CP\*GP\*AP\*CP\*GP\*CP)-3'

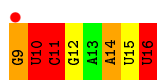




- Molecule 2: 5'-R(\*UP\*UP\*CP\*GP\*AP\*CP\*GP\*CP)-3'



- Molecule 3: 5'-R(\*GP\*UP\*CP\*GP\*AP\*AP\*UP\*UP)-3'



- Molecule 3: 5'-R(\*GP\*UP\*CP\*GP\*AP\*AP\*UP\*UP)-3'



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	51.92Å 61.21Å 104.06Å 76.54° 76.14° 79.35°	Depositor
Resolution (Å)	99.01 – 2.20 37.74 – 2.20	Depositor EDS
% Data completeness (in resolution range)	96.7 (99.01-2.20) 90.6 (37.74-2.20)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.10 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.180 , 0.243 0.180 , 0.241	Depositor DCC
$R_{free}$ test set	2972 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	27.3	Xtriage
Anisotropy	0.360	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 59.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 58833 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7557	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.03% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MN

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	1.12	3/3316 (0.1%)	0.97	7/4498 (0.2%)
1	B	0.92	3/3269 (0.1%)	0.93	11/4433 (0.2%)
2	P	1.74	3/187 (1.6%)	2.22	12/287 (4.2%)
2	R	1.35	1/187 (0.5%)	1.52	0/287
3	Q	1.51	1/189 (0.5%)	2.16	8/292 (2.7%)
3	S	1.07	0/120	2.84	17/185 (9.2%)
All	All	1.08	11/7268 (0.2%)	1.14	55/9982 (0.6%)

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
1	B	0	3
All	All	0	4

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	R	1	U	OP3-P	-11.42	1.47	1.61
2	P	1	U	OP3-P	-8.91	1.50	1.61
1	B	159	GLN	CB-CG	-7.34	1.32	1.52
1	A	169	TRP	CG-CD1	6.62	1.46	1.36
2	P	4	G	N7-C5	5.94	1.42	1.39

The worst 5 of 55 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	S	9	G	O3'-P-O5'	-11.93	81.34	104.00
1	A	325	LEU	CB-CG-CD2	-11.25	91.88	111.00
3	S	10	U	OP2-P-O3'	-11.05	80.89	105.20
3	S	9	G	OP2-P-O3'	-10.14	82.88	105.20
1	B	14	ARG	NE-CZ-NH2	-9.86	115.37	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	259	LYS	Peptide
1	B	116	PRO	Peptide
1	B	117	GLU	Peptide
1	B	414	PRO	Peptide

## 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	A	3237	0	3255	88	0
1	B	3193	0	3222	118	0
2	P	169	0	87	8	0
2	R	169	0	87	1	0
3	Q	170	0	86	12	0
3	S	108	0	55	7	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	283	0	0	23	0
5	B	188	0	0	37	0
5	P	12	0	0	2	0
5	Q	20	0	0	3	0
5	R	3	0	0	0	0
5	S	3	0	0	0	0
All	All	7557	0	6792	219	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 16.

The worst 5 of 219 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:A:331:TYR:CE1	2:P:5:A:H5"	1.77	1.18
1:B:385:ARG:HG3	1:B:385:ARG:HH11	1.03	1.17
1:B:282:SER:HB2	5:B:2121:HOH:O	1.44	1.13
1:B:421:GLN:HG2	5:B:2094:HOH:O	1.48	1.12
1:A:63:GLN:HE22	1:A:146:ASN:ND2	1.47	1.12

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	389/427 (91%)	373 (96%)	12 (3%)	4 (1%)	19	16
1	B	386/427 (90%)	368 (95%)	14 (4%)	4 (1%)	19	16
All	All	775/854 (91%)	741 (96%)	26 (3%)	8 (1%)	19	16

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	118	TYR
1	A	260	MET
1	A	418	ARG
1	B	118	TYR
1	B	360	TYR

### 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	368/394 (93%)	347 (94%)	21 (6%)	25	29
1	B	363/394 (92%)	334 (92%)	29 (8%)	15	15
All	All	731/788 (93%)	681 (93%)	50 (7%)	20	21

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	31	LYS
1	B	81	ILE
1	B	344	LYS
1	B	43	SER
1	B	60	LEU

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	54	ASN
1	A	146	ASN
1	B	278	GLN
1	B	329	GLN
1	B	362	HIS

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	P	7/8 (87%)	0	0
2	R	7/8 (87%)	3 (42%)	0
3	Q	8/8 (100%)	3 (37%)	1 (12%)
3	S	4/8 (50%)	1 (25%)	0
All	All	26/32 (81%)	7 (26%)	1 (3%)

5 of 7 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	Q	10	U
3	Q	11	C
3	Q	16	U

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	R	5	A
2	R	6	C

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
3	Q	9	G

## 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](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	395/427 (92%)	-0.08	11 (2%) 56 55	11, 22, 40, 63	0
1	B	394/427 (92%)	0.22	23 (5%) 26 26	18, 32, 55, 65	0
2	P	8/8 (100%)	1.29	3 (37%) 0 0	20, 35, 78, 86	0
2	R	8/8 (100%)	1.53	3 (37%) 0 0	34, 66, 93, 94	0
3	Q	8/8 (100%)	0.93	1 (12%) 5 5	42, 47, 78, 87	0
3	S	5/8 (62%)	1.51	0 100 100	78, 80, 86, 89	0
All	All	818/886 (92%)	0.11	41 (5%) 32 32	11, 27, 54, 94	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	118	TYR	5.7
2	P	8	C	5.4
1	B	141	TYR	5.3
1	A	331	TYR	5.2
2	R	6	C	4.9

### 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](#)

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	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	MN	A	1428	1/1	0.99	0.12	-1.09	16,16,16,16	0
4	MN	B	1428	1/1	1.00	0.10	-1.50	23,23,23,23	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.