



# wwPDB X-ray Structure Validation Summary Report

Dec 16, 2023 – 11:13 AM EST

PDB ID : 2NWI  
Title : Crystal structure of protein AF1396 from *Archaeoglobus fulgidus*, Pfam DUF98  
Authors : Bonanno, J.B.; Dickey, M.; Bain, K.T.; Adams, J.; Sridhar, V.; Wasserman, S.; Smith, D.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2006-11-14  
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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the  symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36  
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) : 2.36

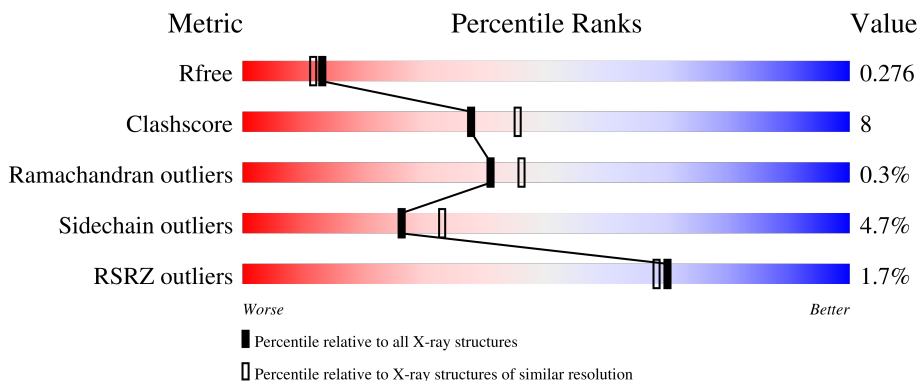
# 1 Overall quality at a glance

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))
$R_{free}$	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 of 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	A	172	 2% 80% 12% • 7%
1	B	172	 3% 73% 12% • • 11%
1	C	172	 % 81% 10% • 7%
1	D	172	 80% 12% • 7%
1	E	172	 2% 70% 21% • 7%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	172	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into four segments: a small red segment at the beginning labeled '2%', a large green segment labeled '74%', a yellow segment labeled '15%', and a small grey segment at the end labeled '8%'.</p>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 8333 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 Hypothetical protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	160	1312	828	240	240	4	0	0	
1	B	153	1274	807	231	233	5	2	0	
1	C	160	1318	832	240	242	4	1	0	
1	D	160	1324	836	240	244	0	2	0	
1	E	160	1312	828	240	240	4	0	0	
1	F	159	1304	822	239	239	4	0	0	

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	initiating methionine	UNP O28875
A	0	SER	-	cloning artifact	UNP O28875
A	1	LEU	-	cloning artifact	UNP O28875
A	163	GLU	-	cloning artifact	UNP O28875
A	164	GLY	-	cloning artifact	UNP O28875
A	165	HIS	-	expression tag	UNP O28875
A	166	HIS	-	expression tag	UNP O28875
A	167	HIS	-	expression tag	UNP O28875
A	168	HIS	-	expression tag	UNP O28875
A	169	HIS	-	expression tag	UNP O28875
A	170	HIS	-	expression tag	UNP O28875
B	-1	MET	-	initiating methionine	UNP O28875
B	0	SER	-	cloning artifact	UNP O28875
B	1	LEU	-	cloning artifact	UNP O28875
B	163	GLU	-	cloning artifact	UNP O28875
B	164	GLY	-	cloning artifact	UNP O28875
B	165	HIS	-	expression tag	UNP O28875

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	166	HIS	-	expression tag	UNP O28875
B	167	HIS	-	expression tag	UNP O28875
B	168	HIS	-	expression tag	UNP O28875
B	169	HIS	-	expression tag	UNP O28875
B	170	HIS	-	expression tag	UNP O28875
C	-1	MET	-	initiating methionine	UNP O28875
C	0	SER	-	cloning artifact	UNP O28875
C	1	LEU	-	cloning artifact	UNP O28875
C	163	GLU	-	cloning artifact	UNP O28875
C	164	GLY	-	cloning artifact	UNP O28875
C	165	HIS	-	expression tag	UNP O28875
C	166	HIS	-	expression tag	UNP O28875
C	167	HIS	-	expression tag	UNP O28875
C	168	HIS	-	expression tag	UNP O28875
C	169	HIS	-	expression tag	UNP O28875
C	170	HIS	-	expression tag	UNP O28875
D	-1	MET	-	initiating methionine	UNP O28875
D	0	SER	-	cloning artifact	UNP O28875
D	1	LEU	-	cloning artifact	UNP O28875
D	163	GLU	-	cloning artifact	UNP O28875
D	164	GLY	-	cloning artifact	UNP O28875
D	165	HIS	-	expression tag	UNP O28875
D	166	HIS	-	expression tag	UNP O28875
D	167	HIS	-	expression tag	UNP O28875
D	168	HIS	-	expression tag	UNP O28875
D	169	HIS	-	expression tag	UNP O28875
D	170	HIS	-	expression tag	UNP O28875
E	-1	MET	-	initiating methionine	UNP O28875
E	0	SER	-	cloning artifact	UNP O28875
E	1	LEU	-	cloning artifact	UNP O28875
E	163	GLU	-	cloning artifact	UNP O28875
E	164	GLY	-	cloning artifact	UNP O28875
E	165	HIS	-	expression tag	UNP O28875
E	166	HIS	-	expression tag	UNP O28875
E	167	HIS	-	expression tag	UNP O28875
E	168	HIS	-	expression tag	UNP O28875
E	169	HIS	-	expression tag	UNP O28875
E	170	HIS	-	expression tag	UNP O28875
F	-1	MET	-	initiating methionine	UNP O28875
F	0	SER	-	cloning artifact	UNP O28875
F	1	LEU	-	cloning artifact	UNP O28875
F	163	GLU	-	cloning artifact	UNP O28875

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	164	GLY	-	cloning artifact	UNP O28875
F	165	HIS	-	expression tag	UNP O28875
F	166	HIS	-	expression tag	UNP O28875
F	167	HIS	-	expression tag	UNP O28875
F	168	HIS	-	expression tag	UNP O28875
F	169	HIS	-	expression tag	UNP O28875
F	170	HIS	-	expression tag	UNP O28875

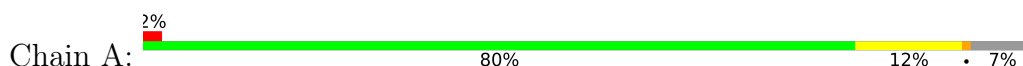
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	75	Total 75	O 75	0	0
2	B	78	Total 78	O 78	0	0
2	C	91	Total 91	O 91	0	0
2	D	101	Total 101	O 101	0	0
2	E	67	Total 67	O 67	0	0
2	F	77	Total 77	O 77	0	0

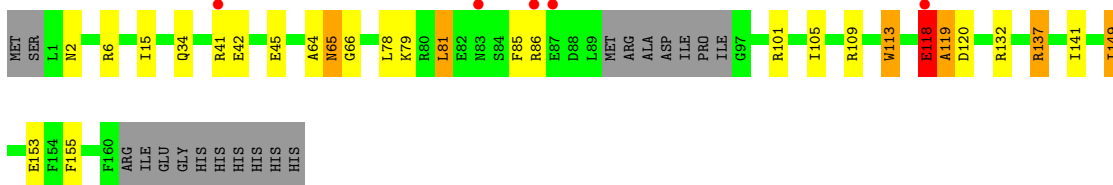
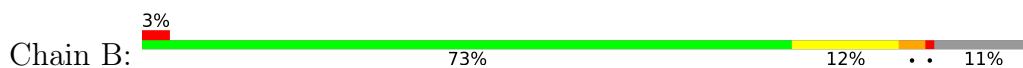
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

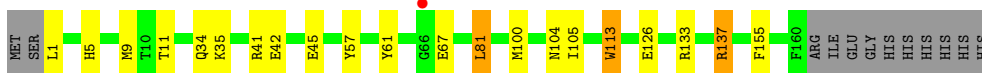
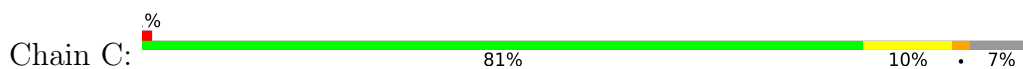
- Molecule 1: Hypothetical protein



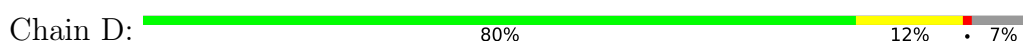
- Molecule 1: Hypothetical protein



- Molecule 1: Hypothetical protein

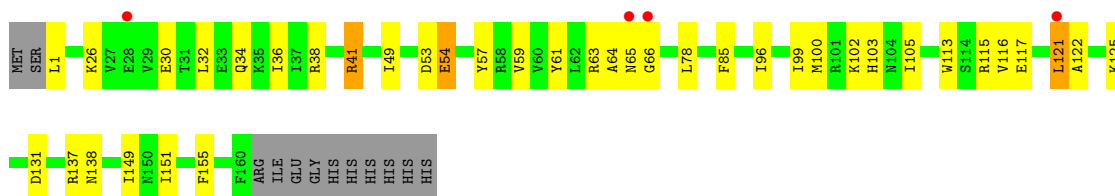


- Molecule 1: Hypothetical protein

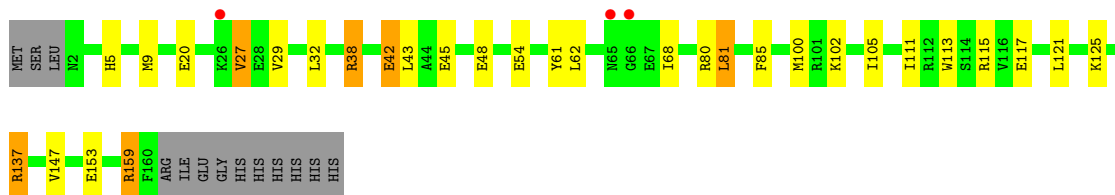
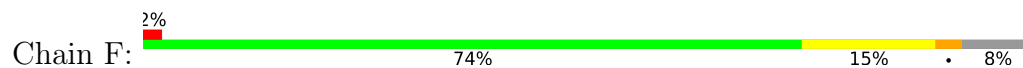


- Molecule 1: Hypothetical protein





- Molecule 1: Hypothetical protein





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.13Å 128.65Å 129.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.20 19.88 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.4 (20.00-2.20) 99.4 (19.88-2.20)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	0.17	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.77 (at 2.19Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.208 , 0.276 0.210 , 0.276	Depositor DCC
$R_{free}$ test set	2874 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	24.6	Xtrriage
Anisotropy	0.440	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 47.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.016 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	8333	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.68% 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 [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	0.64	1/1328 (0.1%)	0.72	0/1785
1	B	0.67	1/1296 (0.1%)	0.71	1/1741 (0.1%)
1	C	0.64	1/1337 (0.1%)	0.78	2/1797 (0.1%)
1	D	0.61	0/1346	0.77	1/1809 (0.1%)
1	E	0.56	0/1328	0.74	0/1785
1	F	0.59	0/1320	0.75	2/1774 (0.1%)
All	All	0.62	3/7955 (0.0%)	0.75	6/10691 (0.1%)

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	B	0	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	101	ARG	CG-CD	11.30	1.80	1.51
1	A	42	GLU	CB-CG	7.78	1.67	1.52
1	C	42	GLU	CG-CD	5.92	1.60	1.51

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	137	ARG	NE-CZ-NH2	-6.37	117.12	120.30
1	F	45	GLU	CA-CB-CG	5.77	126.09	113.40
1	B	101	ARG	CB-CG-CD	-5.75	96.65	111.60
1	D	137	ARG	NE-CZ-NH1	5.55	123.08	120.30
1	F	38	ARG	NE-CZ-NH1	-5.41	117.60	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	118	GLU	Peptide

## 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	1312	0	1360	23	0
1	B	1274	0	1315	19	0
1	C	1318	0	1366	10	0
1	D	1324	0	1372	17	0
1	E	1312	0	1360	31	0
1	F	1304	0	1346	23	0
2	A	75	0	0	4	0
2	B	78	0	0	1	0
2	C	91	0	0	0	0
2	D	101	0	0	1	0
2	E	67	0	0	0	0
2	F	77	0	0	2	0
All	All	8333	0	8119	122	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 8.

The worst 5 of 122 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:117:GLU:HB2	1:A:118:GLU:HB3	1.24	1.16
1:E:121:LEU:O	1:E:125:LYS:HG3	1.51	1.06
1:A:30:GLU:OE2	1:A:63:ARG:NH2	1.95	0.99
1:C:137:ARG:HD2	1:C:155:PHE:HE2	1.29	0.95
1:B:118:GLU:HA	1:B:119:ALA:HB3	1.48	0.95

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	158/172 (92%)	154 (98%)	3 (2%)	1 (1%)	25	26
1	B	151/172 (88%)	146 (97%)	3 (2%)	2 (1%)	12	9
1	C	159/172 (92%)	155 (98%)	4 (2%)	0	100	100
1	D	160/172 (93%)	156 (98%)	4 (2%)	0	100	100
1	E	158/172 (92%)	154 (98%)	4 (2%)	0	100	100
1	F	157/172 (91%)	155 (99%)	2 (1%)	0	100	100
All	All	943/1032 (91%)	920 (98%)	20 (2%)	3 (0%)	41	46

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	118	GLU
1	B	65	ASN
1	B	119	ALA

### 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	140/151 (93%)	135 (96%)	5 (4%)	35	45
1	B	136/151 (90%)	126 (93%)	10 (7%)	13	14
1	C	141/151 (93%)	135 (96%)	6 (4%)	29	36
1	D	142/151 (94%)	139 (98%)	3 (2%)	53	67

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	140/151 (93%)	134 (96%)	6 (4%)	29	36
1	F	139/151 (92%)	129 (93%)	10 (7%)	14	15
All	All	838/906 (92%)	798 (95%)	40 (5%)	26	32

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	113	TRP
1	F	81	LEU
1	E	121	LEU
1	F	42	GLU
1	F	113	TRP

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

Mol	Chain	Res	Type
1	A	104	ASN
1	B	56	ASN
1	C	65	ASN
1	D	104	ASN
1	F	104	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 monosaccharides 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	160/172 (93%)	-0.13	3 (1%) 66 65	23, 33, 50, 56	1 (0%)
1	B	153/172 (88%)	-0.04	5 (3%) 46 44	22, 34, 54, 60	1 (0%)
1	C	160/172 (93%)	-0.18	1 (0%) 89 88	23, 33, 42, 46	1 (0%)
1	D	160/172 (93%)	-0.35	0 100 100	21, 30, 37, 44	0
1	E	160/172 (93%)	-0.02	4 (2%) 57 55	29, 41, 51, 57	1 (0%)
1	F	159/172 (92%)	-0.11	3 (1%) 66 65	24, 36, 45, 49	1 (0%)
All	All	952/1032 (92%)	-0.14	16 (1%) 70 68	21, 34, 48, 60	5 (0%)

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	66	GLY	4.1
1	B	87	GLU	3.6
1	C	66	GLY	3.2
1	B	86	ARG	3.2
1	B	83	ASN	3.0

### 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 monosaccharides 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.