



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 02:58 am BST

PDB ID : 1SHS  
Title : SMALL HEAT SHOCK PROTEIN FROM METHANOCOCCUS JAN-NASCHII  
Authors : Kim, K.K.; Kim, R.; Kim, S.H.; Berkeley Structural Genomics Center (BSGC)  
Deposited on : 1998-07-30  
Resolution : 2.90 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

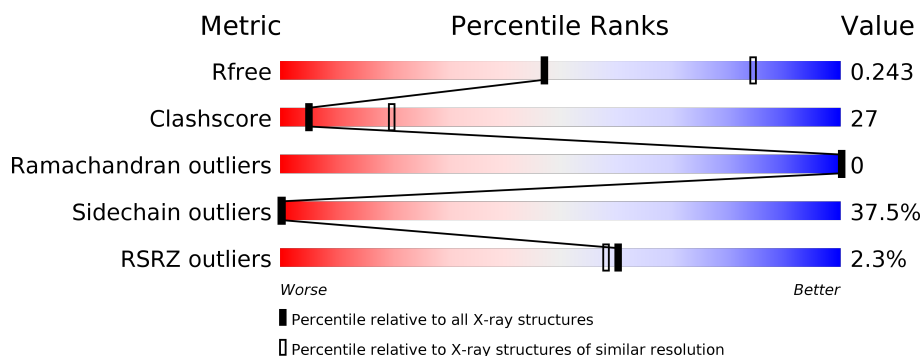
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	A	147	<div> <div>2%</div> <div> <div></div> <div>29%</div> <div>33%</div> <div>17%</div> <div>22%</div> </div> </div>
1	B	147	<div> <div></div> <div> <div>33%</div> <div>30%</div> <div>15%</div> <div>22%</div> </div> </div>
1	C	147	<div> <div>2%</div> <div> <div></div> <div>33%</div> <div>29%</div> <div>16%</div> <div>22%</div> </div> </div>
1	D	147	<div> <div>5%</div> <div> <div></div> <div>34%</div> <div>29%</div> <div>15%</div> <div>22%</div> </div> </div>
1	E	147	<div> <div>2%</div> <div> <div></div> <div>37%</div> <div>28%</div> <div>14%</div> <div>22%</div> </div> </div>
1	F	147	<div> <div>%</div> <div> <div></div> <div>38%</div> <div>27%</div> <div>13%</div> <div>22%</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	147	<div><div><div>%</div><div><div></div><div>36%</div><div>27%</div><div>15%</div><div>22%</div></div></div></div>
1	H	147	<div><div><div>2%</div><div><div></div><div>37%</div><div>28%</div><div>14%</div><div>22%</div></div></div></div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 7144 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 SMALL HEAT SHOCK PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	115	Total	C	N	O	S	0	0	0
			893	570	147	174	2			
1	B	115	Total	C	N	O	S	0	0	0
			893	570	147	174	2			
1	C	115	Total	C	N	O	S	0	0	0
			893	570	147	174	2			
1	D	115	Total	C	N	O	S	0	0	0
			893	570	147	174	2			
1	E	115	Total	C	N	O	S	0	0	0
			893	570	147	174	2			
1	F	115	Total	C	N	O	S	0	0	0
			893	570	147	174	2			
1	G	115	Total	C	N	O	S	0	0	0
			893	570	147	174	2			
1	H	115	Total	C	N	O	S	0	0	0
			893	570	147	174	2			

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 ( $\text{RSRZ} > 2$ ). 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.

Chain A:

29% 33% 17% 22%

Legend:

Amino Acid	Category
MET	Grey
PHE	Green
GLY	Green
ARG	Green
ASP	Green
PRO	Green
PHE	Green
ASP	Green
LEU	Green
SER	Green
LEU	Green
PHE	Green
GLU	Green
ARG	Green
MET	Green
PHE	Green
LYS	Green
GLU	Green
PHE	Green
PHE	Green
THR	Green
PRO	Green
MET	Green
THR	Green
GLY	Green
THR	Green
MET	Green
ILE	Green
GLN	Green
SER	Green
SER	Green
SER	Green
T33	Green
T35	Green
Q36	Yellow
I37	Yellow
S38	Yellow
G39	Yellow
K40	Yellow
M43	Orange
P44	Orange
I45	Orange
S46	Orange
I47	Orange
I48	Orange
E49	Orange
Q52	Grey
H53	Grey
I54	Grey
K55	Grey
W59	Grey
L60	Grey
V63	Orange
M64	Orange
K65	Orange

Chain B:

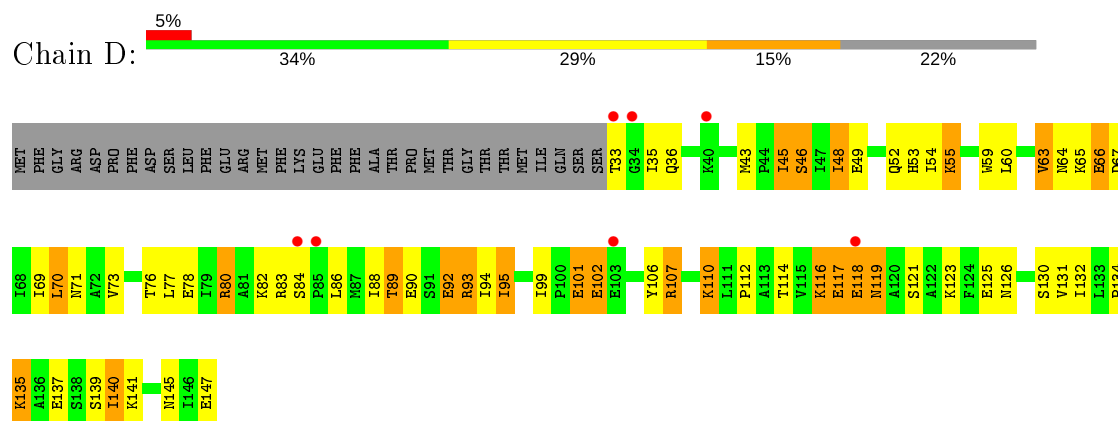
Category	Percentage
Green	33%
Yellow	30%
Orange	15%
Grey	22%

**Chain C:**

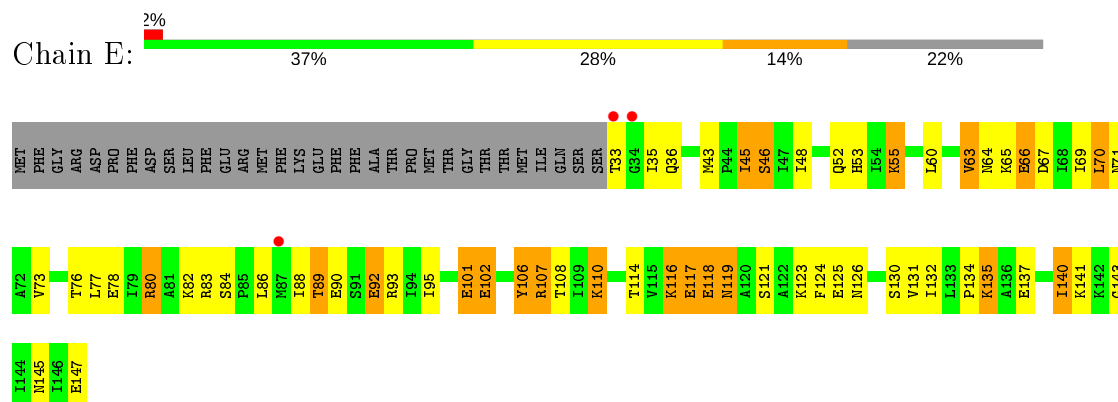
Amino Acid	Percentage
MET	2%
PHE	33%
GLY	29%
L70	16%
L71	22%
A72	
V73	
T76	
L77	
E78	
I79	
R80	
A81	
K82	
R83	
S84	
P85	
L86	
P87	
I88	
T89	
E90	
S91	
E92	
R93	
I94	
I95	
Y96	
I99	
F100	
E101	
E102	
Y106	
R107	
K110	
L111	
T114	
V115	
K116	
E117	
E118	
N119	
A120	
S121	
A122	
K123	
F124	
E125	
N126	
S130	
V131	
I132	
L133	
P134	
K135	
D136	



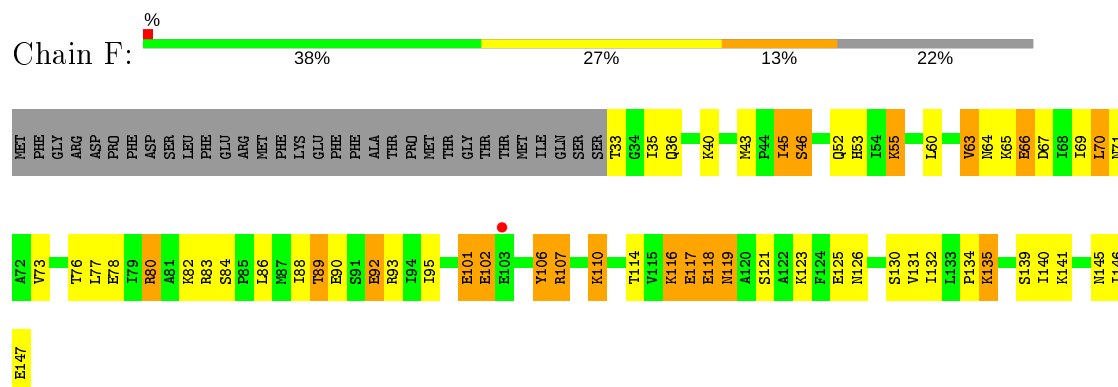
• Molecule 1: SMALL HEAT SHOCK PROTEIN



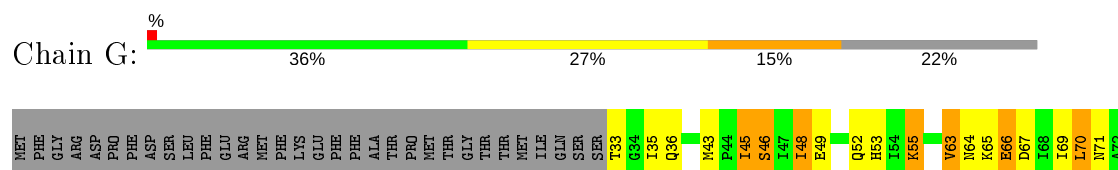
• Molecule 1: SMALL HEAT SHOCK PROTEIN

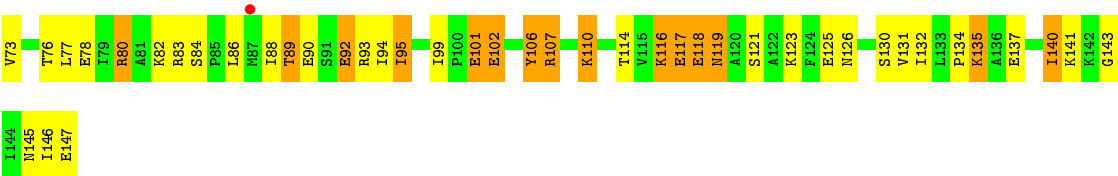


• Molecule 1: SMALL HEAT SHOCK PROTEIN

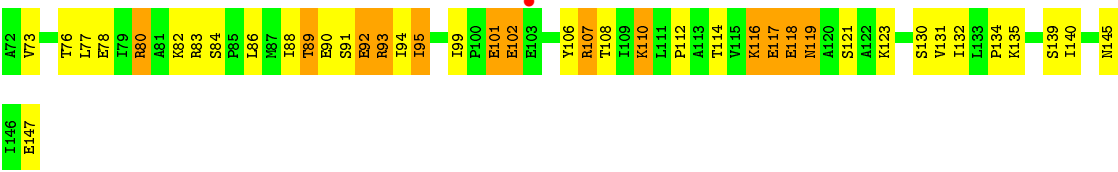


• Molecule 1: SMALL HEAT SHOCK PROTEIN





● Molecule 1: SMALL HEAT SHOCK PROTEIN



## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	171.44Å 171.44Å 101.98Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	15.00 – 2.90 28.57 – 2.90	Depositor EDS
% Data completeness (in resolution range)	89.6 (15.00-2.90) 96.7 (28.57-2.90)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	0.04	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.52 (at 2.90Å)	Xtriage
Refinement program	X-PLOR 3.85	Depositor
R, $R_{free}$	0.216 , 0.251 0.222 , 0.243	Depositor DCC
$R_{free}$ test set	2448 reflections (9.92%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	83.9	Xtriage
Anisotropy	0.011	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 62.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.000 for -2/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+4/3*l,-1/3*h+1/3*k+1/3*l 0.000 for -h,1/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+1/3*l 0.000 for -1/3*h+1/3*k+4/3*l,-k,2/3*h+1/3*k+1/3*l 0.000 for -h,2/3*h+1/3*k+4/3*l,1/3*h+2/3*k-1/3*l 0.000 for -1/3*h-2/3*k+4/3*l,-2/3*h-1/3*k-4/3*l,1/3*h-1/3*k-1/3*l 0.000 for 1/3*h+2/3*k-4/3*l,-k,-2/3*h-1/3*k-1/3*l 0.007 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7144	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.58% 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	0.59	0/905	0.73	0/1221
1	B	0.55	0/905	0.74	0/1221
1	C	0.63	0/905	0.73	0/1221
1	D	0.58	0/905	0.73	0/1221
1	E	0.58	0/905	0.72	0/1221
1	F	0.55	0/905	0.73	0/1221
1	G	0.62	0/905	0.73	0/1221
1	H	0.58	0/905	0.73	0/1221
All	All	0.59	0/7240	0.73	0/9768

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 [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	893	0	932	66	0
1	B	893	0	932	67	0
1	C	893	0	932	60	0
1	D	893	0	932	56	0
1	E	893	0	932	46	0
1	F	893	0	932	47	0
1	G	893	0	932	53	0
1	H	893	0	932	59	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	7144	0	7456	393	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 27.

The worst 5 of 393 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:C:107:ARG:HH11	1:C:107:ARG:HG2	1.42	0.85
1:F:63:VAL:HG23	1:F:83:ARG:HB3	1.60	0.84
1:E:107:ARG:HH11	1:E:107:ARG:HG2	1.44	0.82
1:B:141:LYS:HA	1:F:73:VAL:HB	1.61	0.82
1:B:93:ARG:CZ	1:H:80:ARG:HH21	1.92	0.81

There are no symmetry-related clashes.

## 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	113/147 (77%)	107 (95%)	6 (5%)	0	100	100
1	B	113/147 (77%)	108 (96%)	5 (4%)	0	100	100
1	C	113/147 (77%)	107 (95%)	6 (5%)	0	100	100
1	D	113/147 (77%)	108 (96%)	5 (4%)	0	100	100
1	E	113/147 (77%)	106 (94%)	7 (6%)	0	100	100
1	F	113/147 (77%)	107 (95%)	6 (5%)	0	100	100
1	G	113/147 (77%)	106 (94%)	7 (6%)	0	100	100
1	H	113/147 (77%)	107 (95%)	6 (5%)	0	100	100
All	All	904/1176 (77%)	856 (95%)	48 (5%)	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	100/129 (78%)	62 (62%)	38 (38%)	0	0
1	B	100/129 (78%)	63 (63%)	37 (37%)	0	0
1	C	100/129 (78%)	62 (62%)	38 (38%)	0	0
1	D	100/129 (78%)	62 (62%)	38 (38%)	0	0
1	E	100/129 (78%)	63 (63%)	37 (37%)	0	0
1	F	100/129 (78%)	63 (63%)	37 (37%)	0	0
1	G	100/129 (78%)	63 (63%)	37 (37%)	0	0
1	H	100/129 (78%)	62 (62%)	38 (38%)	0	0
All	All	800/1032 (78%)	500 (62%)	300 (38%)	0	0

5 of 300 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	106	TYR
1	E	90	GLU
1	H	89	THR
1	D	116	LYS
1	E	46	SER

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 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 [i](#)

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	115/147 (78%)	-0.06	3 (2%) 56 52	32, 57, 86, 100	0
1	B	115/147 (78%)	-0.20	0 100 100	38, 56, 86, 95	0
1	C	115/147 (78%)	-0.02	3 (2%) 56 52	30, 52, 86, 98	0
1	D	115/147 (78%)	0.07	7 (6%) 21 17	35, 56, 87, 98	0
1	E	115/147 (78%)	-0.20	3 (2%) 56 52	36, 56, 88, 98	0
1	F	115/147 (78%)	-0.16	1 (0%) 84 84	40, 58, 87, 95	0
1	G	115/147 (78%)	-0.22	1 (0%) 84 84	28, 52, 86, 96	0
1	H	115/147 (78%)	-0.15	3 (2%) 56 52	35, 55, 86, 97	0
All	All	920/1176 (78%)	-0.12	21 (2%) 60 58	28, 56, 88, 100	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	103	GLU	4.8
1	A	33	THR	4.7
1	H	33	THR	4.2
1	H	34	GLY	3.7
1	D	85	PRO	3.5

### 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.