



# Full wwPDB X-ray Structure Validation Report ⓘ

May 19, 2020 – 09:25 am BST

PDB ID : 1ZUP  
Title : CRYSTAL STRUCTURE OF A PUTATIVE NUCLEASE WITH A RIBONUCLEASE H-LIKE MOTIF FOLD (TM1739) FROM THERMOTOGA MARITIMA AT 2.20 Å RESOLUTION  
Authors : Joint Center for Structural Genomics (JCSG)  
Deposited on : 2005-05-31  
Resolution : 2.20 Å (reported)

This is a Full wwPDB X-ray Structure Validation 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
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

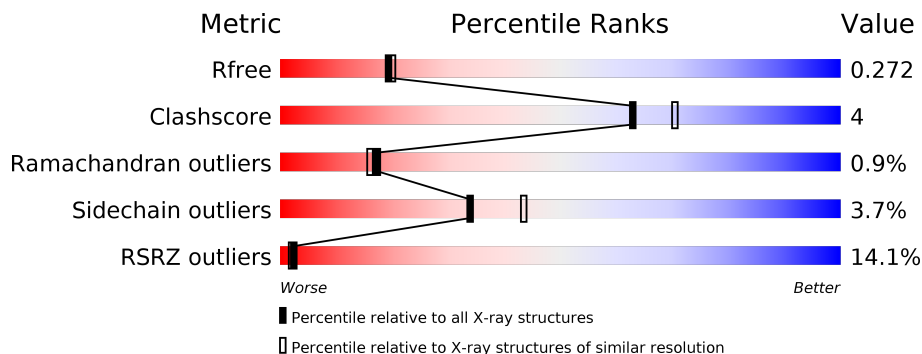
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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 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	315	
1	B	315	

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 4471 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 TM1739.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	293	2336	1500	394	435	1	6	0	5	0
1	B	287	2065	1317	347	395	1	5	0	1	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MSE	-	LEADER SEQUENCE	UNP Q9X261
A	-10	GLY	-	LEADER SEQUENCE	UNP Q9X261
A	-9	SER	-	LEADER SEQUENCE	UNP Q9X261
A	-8	ASP	-	LEADER SEQUENCE	UNP Q9X261
A	-7	LYS	-	LEADER SEQUENCE	UNP Q9X261
A	-6	ILE	-	LEADER SEQUENCE	UNP Q9X261
A	-5	HIS	-	LEADER SEQUENCE	UNP Q9X261
A	-4	HIS	-	LEADER SEQUENCE	UNP Q9X261
A	-3	HIS	-	LEADER SEQUENCE	UNP Q9X261
A	-2	HIS	-	LEADER SEQUENCE	UNP Q9X261
A	-1	HIS	-	LEADER SEQUENCE	UNP Q9X261
A	0	HIS	-	LEADER SEQUENCE	UNP Q9X261
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
A	129	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
A	136	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
A	148	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
A	203	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
A	293	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
B	-11	MSE	-	LEADER SEQUENCE	UNP Q9X261
B	-10	GLY	-	LEADER SEQUENCE	UNP Q9X261
B	-9	SER	-	LEADER SEQUENCE	UNP Q9X261
B	-8	ASP	-	LEADER SEQUENCE	UNP Q9X261
B	-7	LYS	-	LEADER SEQUENCE	UNP Q9X261
B	-6	ILE	-	LEADER SEQUENCE	UNP Q9X261
B	-5	HIS	-	LEADER SEQUENCE	UNP Q9X261

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	HIS	-	LEADER SEQUENCE	UNP Q9X261
B	-3	HIS	-	LEADER SEQUENCE	UNP Q9X261
B	-2	HIS	-	LEADER SEQUENCE	UNP Q9X261
B	-1	HIS	-	LEADER SEQUENCE	UNP Q9X261
B	0	HIS	-	LEADER SEQUENCE	UNP Q9X261
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
B	129	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
B	136	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
B	148	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
B	203	MSE	MET	MODIFIED RESIDUE	UNP Q9X261
B	293	MSE	MET	MODIFIED RESIDUE	UNP Q9X261

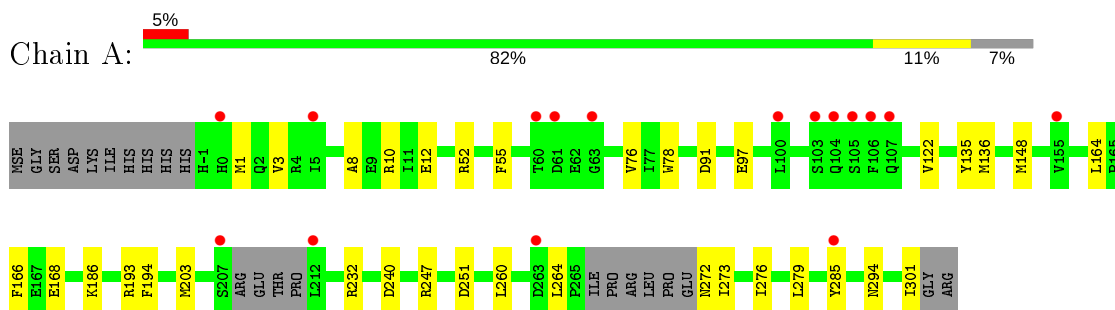
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	57	Total O 57 57	0	0
2	B	13	Total O 13 13	0	0

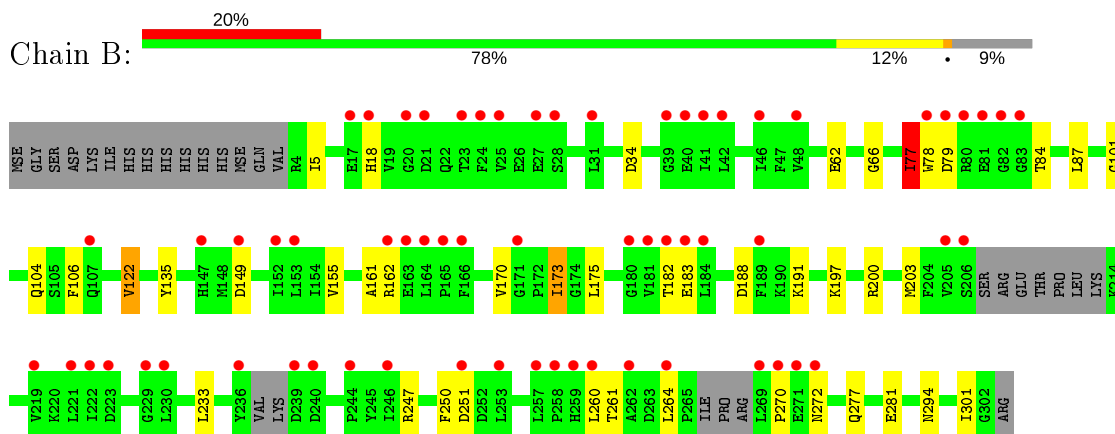
### 3 Residue-property plots [i](#)

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: hypothetical protein TM1739



- Molecule 1: hypothetical protein TM1739



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	82.43Å 82.43Å 234.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.20 41.22 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.9 (20.00-2.20) 99.9 (41.22-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.97 (at 2.10Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.228 , 0.266 0.239 , 0.272	Depositor DCC
$R_{free}$ test set	2114 reflections (4.39%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	47.3	Xtrriage
Anisotropy	0.304	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 75.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	4471	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.67% 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	0.86	1/2369 (0.0%)	0.87	3/3177 (0.1%)
1	B	0.71	2/2092 (0.1%)	0.77	2/2810 (0.1%)
All	All	0.79	3/4461 (0.1%)	0.82	5/5987 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	136	MSE	CB-CG	6.46	1.71	1.52
1	B	251	ASP	C-O	5.26	1.33	1.23
1	B	251	ASP	CA-C	5.22	1.66	1.52

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	34	ASP	O-C-N	7.17	134.18	122.70
1	A	232	ARG	NE-CZ-NH1	6.89	123.74	120.30
1	A	251	ASP	CB-CG-OD1	6.73	124.36	118.30
1	A	232	ARG	NE-CZ-NH2	-6.24	117.18	120.30
1	B	77	ILE	CB-CA-C	-5.66	100.28	111.60

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	2336	0	2302	24	0
1	B	2065	0	1915	19	0
2	A	57	0	0	0	0
2	B	13	0	0	0	0
All	All	4471	0	4217	37	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 4.

All (37) 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:203:MSE:HE1	1:A:247:ARG:HG3	1.67	0.76
1:A:1:MSE:HE2	1:A:3:VAL:CG2	2.14	0.76
1:A:203:MSE:HE1	1:A:247:ARG:CG	2.30	0.62
1:A:1:MSE:HE3	1:B:106:PHE:CZ	2.37	0.60
1:A:1:MSE:HE3	1:B:106:PHE:HZ	1.68	0.58
1:B:203:MSE:HE3	1:B:250:PHE:CD2	2.40	0.57
1:A:1:MSE:HE2	1:A:3:VAL:HG21	1.86	0.56
1:A:148:MSE:CE	1:A:164:LEU:HD22	2.37	0.55
1:B:277:GLN:O	1:B:281:GLU:HG3	2.10	0.52
1:A:193:ARG:NH1	1:A:194:PHE:CZ	2.79	0.51
1:A:76:VAL:HG13	1:A:279:LEU:HD13	1.93	0.50
1:A:203:MSE:HE1	1:A:247:ARG:CB	2.43	0.49
1:B:122:VAL:HG21	1:B:135:TYR:CG	2.47	0.49
1:A:273:ILE:HG12	1:A:276:ILE:HD12	1.94	0.48
1:A:301:ILE:HG23	1:B:5:ILE:HD13	1.96	0.47
1:A:10:ARG:HG3	1:B:294:ASN:ND2	2.30	0.47
1:A:148:MSE:HE1	1:A:164:LEU:HD22	1.96	0.47
1:B:161:ALA:O	1:B:162:ARG:C	2.54	0.46
1:A:10:ARG:HD3	1:A:12:GLU:OE2	2.16	0.45
1:A:78:TRP:CZ2	1:A:260:LEU:HD21	2.52	0.45
1:A:55:PHE:HB3	1:A:294:ASN:OD1	2.15	0.45
1:B:188:ASP:O	1:B:191:LYS:HB2	2.17	0.45
1:A:8:ALA:HB2	1:B:301:ILE:HG21	1.99	0.44
1:B:272[A]:ASN:N	1:B:272[A]:ASN:HD22	2.16	0.43
1:B:78:TRP:CZ2	1:B:260:LEU:HD21	2.52	0.43
1:B:155:VAL:HG22	1:B:173:ILE:HG23	1.99	0.43
1:A:285[B]:TYR:CG	1:A:285[B]:TYR:O	2.71	0.43
1:B:182:THR:O	1:B:183:GLU:HB2	2.19	0.43
1:B:66:GLY:HA2	1:B:101:GLY:O	2.19	0.43
1:A:166:PHE:CE2	1:A:168:GLU:HA	2.54	0.42

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1:MSE:HE2	1:A:3:VAL:HG23	2.00	0.42
1:B:277:GLN:O	1:B:281:GLU:CG	2.67	0.42
1:A:203:MSE:CE	1:A:247:ARG:HA	2.49	0.42
1:B:77:ILE:CD1	1:B:87:LEU:HG	2.49	0.42
1:A:186:LYS:NZ	1:B:62:GLU:OE1	2.41	0.41
1:B:200:ARG:NH2	1:B:247:ARG:HD2	2.35	0.41
1:A:97:GLU:OE1	1:A:135:TYR:OH	2.32	0.40

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	292/315 (93%)	282 (97%)	9 (3%)	1 (0%)	41 46
1	B	280/315 (89%)	252 (90%)	24 (9%)	4 (1%)	11 8
All	All	572/630 (91%)	534 (93%)	33 (6%)	5 (1%)	17 16

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	270	PRO
1	B	264	LEU
1	A	264	LEU
1	B	197	LYS
1	B	170	VAL

### 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	241/274 (88%)	236 (98%)	5 (2%)	53	67
1	B	195/274 (71%)	184 (94%)	11 (6%)	21	25
All	All	436/548 (80%)	420 (96%)	16 (4%)	34	43

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	52	ARG
1	A	91	ASP
1	A	122	VAL
1	A	240	ASP
1	A	272	ASN
1	B	18	HIS
1	B	77	ILE
1	B	79	ASP
1	B	84	THR
1	B	104	GLN
1	B	122	VAL
1	B	149	ASP
1	B	173	ILE
1	B	175	LEU
1	B	233	LEU
1	B	261	THR

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 [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	287/315 (91%)	0.60	16 (5%) 24 23	53, 65, 82, 121	0
1	B	282/315 (89%)	1.19	64 (22%) 0 0	50, 63, 80, 119	0
All	All	569/630 (90%)	0.89	80 (14%) 2 2	50, 64, 81, 121	0

All (80) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	244	PRO	6.5
1	B	269	LEU	5.6
1	A	107	GLN	5.3
1	B	82	GLY	5.1
1	B	264	LEU	4.8
1	B	236	TYR	4.5
1	B	262	ALA	4.4
1	A	106	PHE	4.4
1	B	222	ILE	4.3
1	B	42	LEU	4.1
1	B	260	LEU	4.1
1	B	219	VAL	4.0
1	B	246	ILE	3.9
1	B	46	ILE	3.9
1	B	24	PHE	3.9
1	B	206	SER	3.9
1	B	162	ARG	3.7
1	B	223	ASP	3.7
1	B	259	HIS	3.7
1	B	184	LEU	3.6
1	B	41	ILE	3.5
1	B	152	ILE	3.4
1	B	257	LEU	3.4
1	A	263	ASP	3.3

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	B	205	VAL	3.3
1	B	180	GLY	3.2
1	B	149	ASP	3.1
1	B	80	ARG	3.1
1	B	23	THR	3.0
1	A	63	GLY	3.0
1	A	60	THR	3.0
1	A	207	SER	3.0
1	B	230	LEU	2.9
1	B	253	LEU	2.9
1	B	221	LEU	2.9
1	A	105	SER	2.9
1	B	258	PRO	2.9
1	B	182	THR	2.9
1	B	189	PHE	2.9
1	B	165	PRO	2.9
1	B	31	LEU	2.8
1	B	164	LEU	2.8
1	A	285[A]	TYR	2.8
1	B	239	ASP	2.8
1	B	240	ASP	2.8
1	B	153	LEU	2.8
1	B	83	GLY	2.8
1	B	28	SER	2.7
1	B	27	GLU	2.7
1	A	104[A]	GLN	2.7
1	A	61	ASP	2.7
1	B	78	TRP	2.7
1	B	251	ASP	2.7
1	B	271	GLU	2.6
1	B	270	PRO	2.6
1	B	39	GLY	2.6
1	B	20	GLY	2.6
1	B	18	HIS	2.6
1	B	229	GLY	2.5
1	B	17	GLU	2.5
1	A	103	SER	2.5
1	B	166	PHE	2.5
1	B	147	HIS	2.5
1	B	79	ASP	2.4
1	B	40	GLU	2.4
1	B	183	GLU	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	48	VAL	2.4
1	B	81	GLU	2.4
1	A	0	HIS	2.4
1	B	21	ASP	2.3
1	B	272[A]	ASN	2.2
1	B	25	VAL	2.2
1	B	163	GLU	2.2
1	B	171	GLY	2.1
1	B	107	GLN	2.1
1	A	5	ILE	2.1
1	B	181	VAL	2.1
1	A	100	LEU	2.0
1	A	212	LEU	2.0
1	A	155	VAL	2.0

## 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 carbohydrates in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.