



# Full wwPDB X-ray Structure Validation Report ⓘ

May 21, 2020 – 06:56 am BST

PDB ID : 6QGM  
Title : VirX1 apo structure  
Authors : Gkotsi, D.S.; Ludewig, H.; Sharma, S.V.; Unsworth, W.P.; Taylor, R.J.K.;  
McLachlan, M.M.W.; Shanahan, S.; Naismith, J.H.; Goss, R.J.M.  
Deposited on : 2019-01-11  
Resolution : 2.75 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  
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

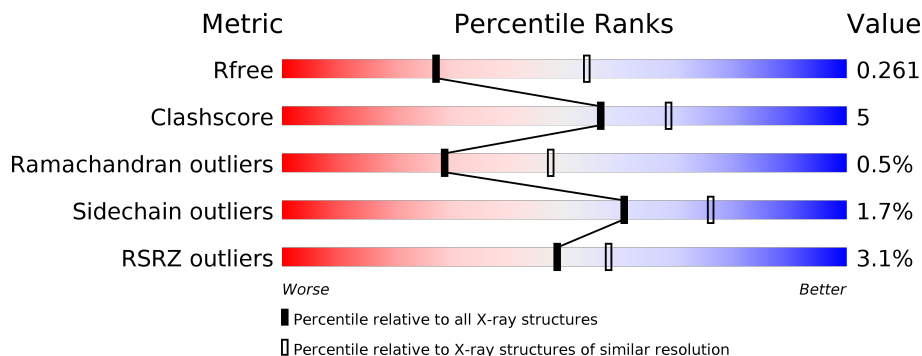
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.75 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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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	531	
1	b	531	
1	c	531	
1	d	531	
1	e	531	
1	f	531	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 25688 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 VirX1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	a	524	Total 4273	C 2724	N 722	O 806	S 21	0	0	0
1	b	525	Total 4278	C 2727	N 723	O 807	S 21	0	0	0
1	c	522	Total 4265	C 2720	N 721	O 803	S 21	0	0	0
1	d	526	Total 4285	C 2731	N 724	O 808	S 22	0	0	0
1	e	523	Total 4264	C 2718	N 721	O 804	S 21	0	0	0
1	f	525	Total 4275	C 2724	N 723	O 807	S 21	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
a	5	VAL	-	insertion	UNP M4SKV1
b	5	VAL	-	insertion	UNP M4SKV1
c	5	VAL	-	insertion	UNP M4SKV1
d	5	VAL	-	insertion	UNP M4SKV1
e	5	VAL	-	insertion	UNP M4SKV1
f	5	VAL	-	insertion	UNP M4SKV1

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	a	9	Total 9	O 9	0	0
2	b	6	Total 6	O 6	0	0
2	c	10	Total 10	O 10	0	0

*Continued on next page...*

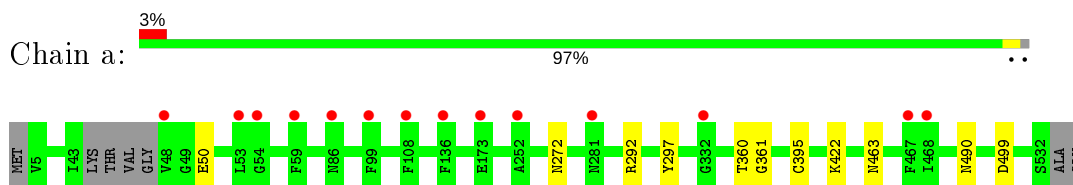
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
2	d	9	Total O 9 9	0	0
2	e	7	Total O 7 7	0	0
2	f	7	Total O 7 7	0	0

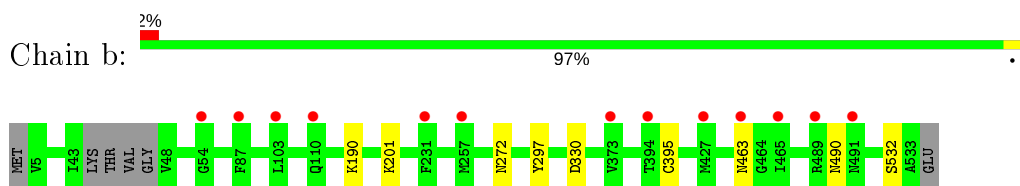
### 3 Residue-property plots [i](#)

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.

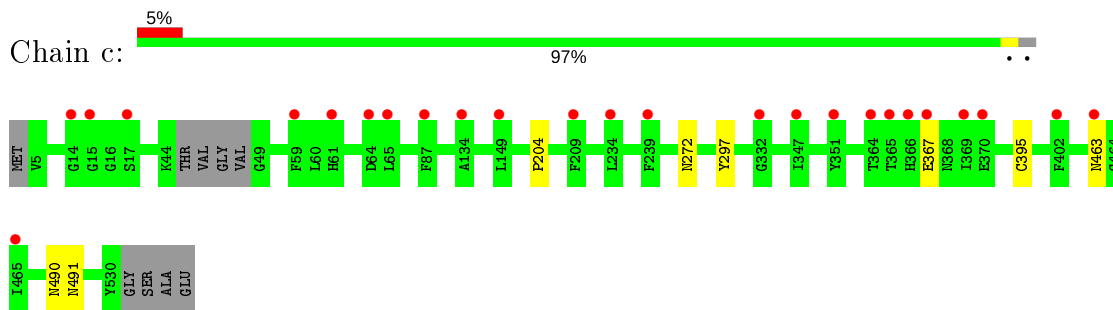
- Molecule 1: VirX1



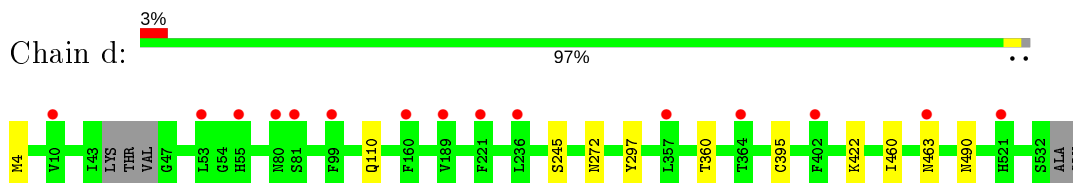
- Molecule 1: VirX1



- Molecule 1: VirX1

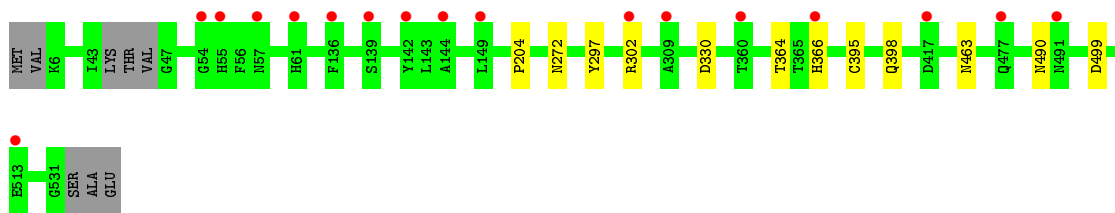


- Molecule 1: VirX1

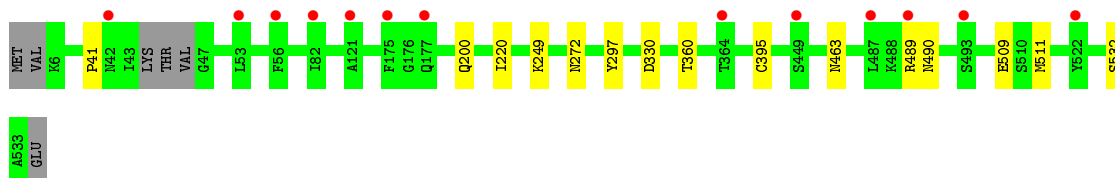


- Molecule 1: VirX1





- Molecule 1: VirX1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.13Å 172.98Å 110.42Å 90.00° 112.89° 90.00°	Depositor
Resolution (Å)	87.94 – 2.75 87.79 – 2.75	Depositor EDS
% Data completeness (in resolution range)	99.0 (87.94-2.75) 99.0 (87.79-2.75)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.12 (at 2.77Å)	Xtrriage
Refinement program	REFMAC 5.8.0232	Depositor
R, $R_{free}$	0.216 , 0.259 0.221 , 0.261	Depositor DCC
$R_{free}$ test set	4356 reflections (4.87%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	52.1	Xtrriage
Anisotropy	0.161	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 40.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	25688	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	66.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.34% 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	0.66	0/4394	0.81	2/5953 (0.0%)
1	b	0.68	0/4399	0.79	0/5960
1	c	0.66	0/4386	0.80	0/5941
1	d	0.67	0/4406	0.79	0/5968
1	e	0.65	0/4385	0.79	1/5940 (0.0%)
1	f	0.66	0/4396	0.79	0/5955
All	All	0.66	0/26366	0.79	3/35717 (0.0%)

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

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	a	292	ARG	NE-CZ-NH1	-6.21	117.19	120.30
1	e	366	HIS	CB-CA-C	-5.89	98.62	110.40
1	a	292	ARG	NE-CZ-NH2	5.22	122.91	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	b	532	SER	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	4273	0	4046	0	0
1	b	4278	0	4051	0	0
1	c	4265	0	4042	0	0
1	d	4285	0	4058	0	0
1	e	4264	0	4035	0	0
1	f	4275	0	4045	0	0
2	a	9	0	0	0	0
2	b	6	0	0	0	0
2	c	10	0	0	0	0
2	d	9	0	0	0	0
2	e	7	0	0	0	0
2	f	7	0	0	0	0
All	All	25688	0	24277	0	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 5.

There are no clashes within the asymmetric unit.

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	520/531 (98%)	488 (94%)	29 (6%)	3 (1%)	25 42
1	b	521/531 (98%)	490 (94%)	29 (6%)	2 (0%)	34 53
1	c	518/531 (98%)	488 (94%)	28 (5%)	2 (0%)	34 53

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	d	522/531 (98%)	492 (94%)	28 (5%)	2 (0%)	34	53
1	e	519/531 (98%)	489 (94%)	27 (5%)	3 (1%)	25	42
1	f	521/531 (98%)	492 (94%)	25 (5%)	4 (1%)	19	34
All	All	3121/3186 (98%)	2939 (94%)	166 (5%)	16 (0%)	29	47

All (16) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	a	272	ASN
1	a	463	ASN
1	b	272	ASN
1	b	463	ASN
1	c	272	ASN
1	c	463	ASN
1	d	272	ASN
1	d	463	ASN
1	e	272	ASN
1	f	41	PRO
1	f	272	ASN
1	f	463	ASN
1	e	463	ASN
1	e	364	THR
1	f	532	SER
1	a	361	GLY

### 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	459/464 (99%)	452 (98%)	7 (2%)	65	78
1	b	459/464 (99%)	453 (99%)	6 (1%)	69	81
1	c	458/464 (99%)	452 (99%)	6 (1%)	69	81
1	d	460/464 (99%)	451 (98%)	9 (2%)	55	72

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	e	457/464 (98%)	449 (98%)	8 (2%)	59	75
1	f	458/464 (99%)	447 (98%)	11 (2%)	49	68
All	All	2751/2784 (99%)	2704 (98%)	47 (2%)	60	76

All (47) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	a	50	GLU
1	a	297	TYR
1	a	360	THR
1	a	395	CYS
1	a	422	LYS
1	a	490	ASN
1	a	499	ASP
1	b	190	LYS
1	b	201	LYS
1	b	297	TYR
1	b	330	ASP
1	b	395	CYS
1	b	490	ASN
1	c	204	PRO
1	c	297	TYR
1	c	367	GLU
1	c	395	CYS
1	c	490	ASN
1	c	491	ASN
1	d	4	MET
1	d	110	GLN
1	d	245	SER
1	d	297	TYR
1	d	360	THR
1	d	395	CYS
1	d	422	LYS
1	d	460	ILE
1	d	490	ASN
1	e	204	PRO
1	e	297	TYR
1	e	302	ARG
1	e	330	ASP
1	e	395	CYS
1	e	398	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	e	490	ASN
1	e	499	ASP
1	f	200	GLN
1	f	220	ILE
1	f	249	LYS
1	f	297	TYR
1	f	330	ASP
1	f	360	THR
1	f	395	CYS
1	f	489	ARG
1	f	490	ASN
1	f	509	GLU
1	f	511	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	d	191	HIS
1	d	253	ASN
1	e	366	HIS
1	f	446	GLN

### 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

### 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	524/531 (98%)	0.36	14 (2%) 54 63	46, 61, 80, 100	0
1	b	525/531 (98%)	0.27	13 (2%) 57 66	43, 61, 79, 96	0
1	c	522/531 (98%)	0.47	25 (4%) 30 36	47, 66, 89, 110	0
1	d	526/531 (99%)	0.42	15 (2%) 51 61	51, 68, 90, 108	0
1	e	523/531 (98%)	0.39	17 (3%) 46 54	49, 66, 87, 103	0
1	f	525/531 (98%)	0.33	13 (2%) 57 66	45, 65, 84, 108	0
All	All	3145/3186 (98%)	0.37	97 (3%) 49 58	43, 65, 86, 110	0

All (97) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	c	364	THR	6.1
1	e	54	GLY	4.6
1	d	53	LEU	4.2
1	d	55	HIS	4.0
1	f	42	ASN	3.9
1	c	15	GLY	3.9
1	c	59	PHE	3.9
1	a	136	PHE	3.9
1	e	366	HIS	3.8
1	f	489	ARG	3.8
1	b	463	ASN	3.7
1	e	302	ARG	3.5
1	d	80	ASN	3.5
1	e	142	TYR	3.2
1	c	134	ALA	3.2
1	e	491	ASN	3.1
1	b	491	ASN	3.0
1	e	477	GLN	2.9
1	f	364	THR	2.9

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	e	360	THR	2.9
1	f	449	SER	2.8
1	c	64	ASP	2.8
1	b	54	GLY	2.8
1	c	209	PHE	2.7
1	f	175	PHE	2.7
1	c	402	PHE	2.7
1	d	364	THR	2.7
1	c	367	GLU	2.7
1	c	351	TYR	2.6
1	d	81	SER	2.6
1	d	99	PHE	2.6
1	b	103	LEU	2.6
1	c	332	GLY	2.6
1	b	373	VAL	2.6
1	a	108	PHE	2.5
1	b	87	PHE	2.5
1	d	402	PHE	2.5
1	b	489	ARG	2.5
1	c	149	LEU	2.5
1	d	221	PHE	2.5
1	e	149	LEU	2.5
1	a	59	PHE	2.5
1	a	53	LEU	2.5
1	c	234	LEU	2.5
1	c	87	PHE	2.5
1	d	236	LEU	2.5
1	e	139	SER	2.5
1	a	86	ASN	2.5
1	d	357	LEU	2.5
1	e	144	ALA	2.4
1	b	394	THR	2.4
1	b	465	ILE	2.4
1	d	160	PHE	2.4
1	c	17	SER	2.4
1	a	467	PHE	2.4
1	a	48	VAL	2.4
1	b	231	PHE	2.4
1	c	366	HIS	2.3
1	c	65	LEU	2.3
1	c	347	ILE	2.3
1	a	173	GLU	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	e	57	ASN	2.3
1	e	417	ASP	2.3
1	c	14	GLY	2.3
1	c	463	ASN	2.3
1	c	465	ILE	2.3
1	a	332	GLY	2.3
1	a	99	PHE	2.3
1	f	493	SER	2.2
1	d	521	HIS	2.2
1	f	177	GLN	2.2
1	d	189	VAL	2.2
1	b	110	GLN	2.2
1	e	136	PHE	2.2
1	d	10	VAL	2.2
1	e	55	HIS	2.2
1	c	239	PHE	2.2
1	f	487	LEU	2.2
1	b	257	MET	2.2
1	a	281	ASN	2.1
1	e	513	GLU	2.1
1	f	56	PHE	2.1
1	e	309	ALA	2.1
1	c	365	THR	2.1
1	b	427	MET	2.1
1	d	463	ASN	2.1
1	a	252	ALA	2.1
1	f	53	LEU	2.1
1	f	121	ALA	2.1
1	c	61	HIS	2.1
1	e	61	HIS	2.0
1	f	82	ILE	2.0
1	a	468	ILE	2.0
1	c	369	ILE	2.0
1	f	522	TYR	2.0
1	c	370	GLU	2.0
1	a	54	GLY	2.0

## 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.