



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2023 – 06:05 PM EDT

PDB ID : 7UVI  
Title : Pfs230 domain 1 bound by RUPA-55 Fab  
Authors : Ivanochko, D.; Newton, J.; Julien, J.P.  
Deposited on : 2022-05-02  
Resolution : 2.92 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 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  
Xtrriage (Phenix) : 1.13  
EDS : 2.35  
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) : 2.35

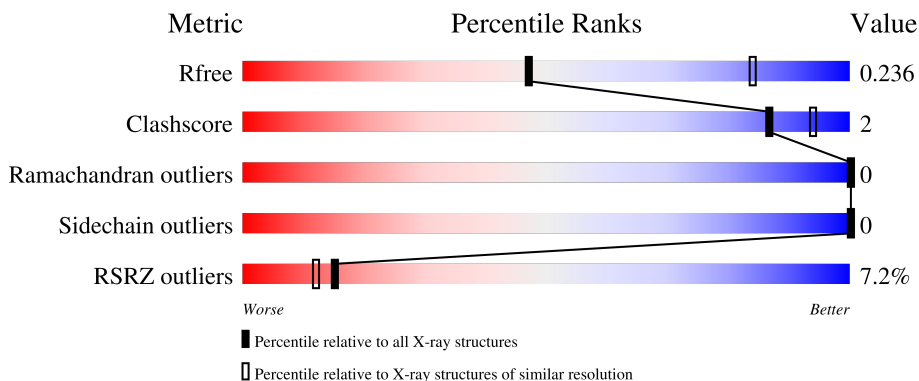
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.92 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.



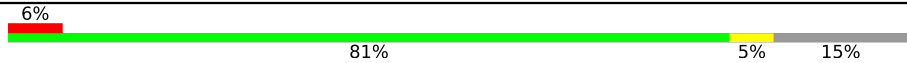
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2307 (2.94-2.90)
Clashscore	141614	2531 (2.94-2.90)
Ramachandran outliers	138981	2462 (2.94-2.90)
Sidechain outliers	138945	2464 (2.94-2.90)
RSRZ outliers	127900	2248 (2.94-2.90)

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 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	227	 10% 90% 8%
1	D	227	 3% 94% 5%
2	B	215	 9% 93% 6%
2	E	215	 4% 93% 7%
3	C	199	 10% 88% 5% 7%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	199	 <p>6% 81% 5% 15%</p>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9476 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 RUPA-55 Fab heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	224	Total 1672	C 1062	N 277	O 327	S 6	0	0	0
1	D	225	Total 1681	C 1068	N 279	O 328	S 6	0	0	0

- Molecule 2 is a protein called RUPA-55 Fab light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	212	Total 1627	C 1017	N 278	O 327	S 5	0	0	0
2	E	214	Total 1640	C 1024	N 280	O 331	S 5	0	0	0

- Molecule 3 is a protein called Gametocyte surface protein P230.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	185	Total 1483	C 953	N 232	O 294	S 4	0	0	0
3	F	170	Total 1373	C 888	N 217	O 264	S 4	0	0	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	585	GLN	ASN	conflict	UNP P68874
C	732	GLY	-	expression tag	UNP P68874
C	733	SER	-	expression tag	UNP P68874
C	734	LEU	-	expression tag	UNP P68874
C	735	LYS	-	expression tag	UNP P68874
C	736	GLU	-	expression tag	UNP P68874
C	737	ASN	-	expression tag	UNP P68874
C	738	LEU	-	expression tag	UNP P68874

*Continued on next page...*

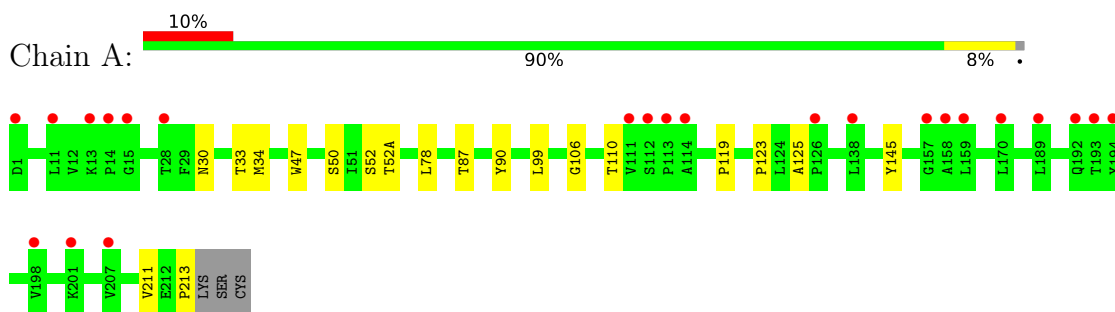
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	739	TYR	-	expression tag	UNP P68874
C	740	PHE	-	expression tag	UNP P68874
C	741	GLN	-	expression tag	UNP P68874
C	742	GLY	-	expression tag	UNP P68874
C	743	TRP	-	expression tag	UNP P68874
C	744	SER	-	expression tag	UNP P68874
C	745	HIS	-	expression tag	UNP P68874
C	746	PRO	-	expression tag	UNP P68874
C	747	GLN	-	expression tag	UNP P68874
C	748	PHE	-	expression tag	UNP P68874
C	749	GLU	-	expression tag	UNP P68874
C	750	LYS	-	expression tag	UNP P68874
F	585	GLN	ASN	conflict	UNP P68874
F	732	GLY	-	expression tag	UNP P68874
F	733	SER	-	expression tag	UNP P68874
F	734	LEU	-	expression tag	UNP P68874
F	735	LYS	-	expression tag	UNP P68874
F	736	GLU	-	expression tag	UNP P68874
F	737	ASN	-	expression tag	UNP P68874
F	738	LEU	-	expression tag	UNP P68874
F	739	TYR	-	expression tag	UNP P68874
F	740	PHE	-	expression tag	UNP P68874
F	741	GLN	-	expression tag	UNP P68874
F	742	GLY	-	expression tag	UNP P68874
F	743	TRP	-	expression tag	UNP P68874
F	744	SER	-	expression tag	UNP P68874
F	745	HIS	-	expression tag	UNP P68874
F	746	PRO	-	expression tag	UNP P68874
F	747	GLN	-	expression tag	UNP P68874
F	748	PHE	-	expression tag	UNP P68874
F	749	GLU	-	expression tag	UNP P68874
F	750	LYS	-	expression tag	UNP P68874

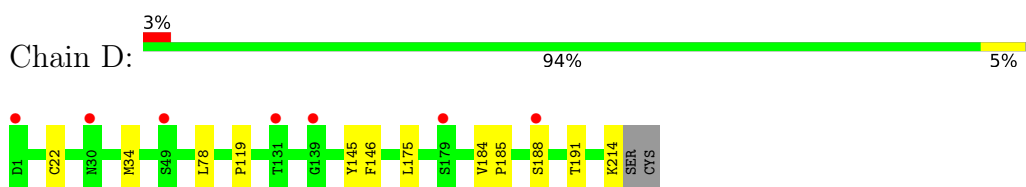
### 3 Residue-property plots [i](#)

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.

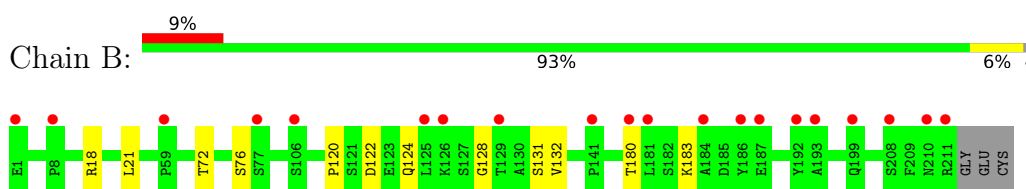
- Molecule 1: RUPA-55 Fab heavy chain



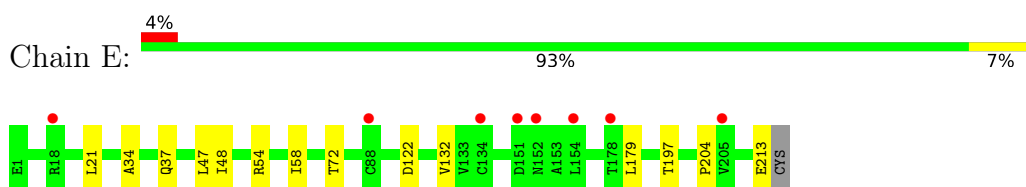
- Molecule 1: RUPA-55 Fab heavy chain



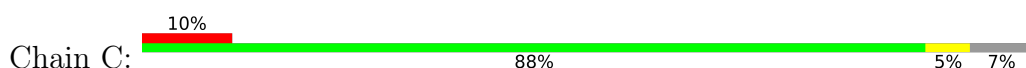
- Molecule 2: RUPA-55 Fab light chain



- Molecule 2: RUPA-55 Fab light chain



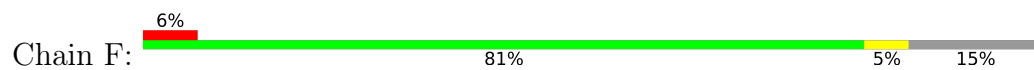
- Molecule 3: Gametocyte surface protein P230





LYS

• Molecule 3: Gametocyte surface protein P230



GLN  
PHE  
GLU  
LYS

## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	91.02Å 100.78Å 425.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.67 – 2.92 29.67 – 2.92	Depositor EDS
% Data completeness (in resolution range)	99.7 (29.67-2.92) 99.8 (29.67-2.92)	Depositor EDS
$R_{merge}$	0.27	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.16 (at 2.90Å)	Xtrriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.218 , 0.237 0.217 , 0.236	Depositor DCC
$R_{free}$ test set	2035 reflections (4.73%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	81.7	Xtrriage
Anisotropy	0.802	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 52.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9476	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	106.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.47% 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	0.28	0/1713	0.59	0/2334
1	D	0.29	0/1722	0.58	0/2345
2	B	0.30	0/1665	0.55	0/2265
2	E	0.29	0/1678	0.56	0/2282
3	C	0.31	0/1515	0.57	0/2050
3	F	0.30	0/1404	0.56	0/1898
All	All	0.30	0/9697	0.57	0/13174

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
3	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	744	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	1672	0	1647	10	0
1	D	1681	0	1660	7	0
2	B	1627	0	1578	7	0
2	E	1640	0	1587	8	0
3	C	1483	0	1482	4	0
3	F	1373	0	1387	5	0
All	All	9476	0	9341	40	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 2.

All (40) 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:D:146:PHE:HB2	1:D:175:LEU:HD23	1.87	0.57
1:D:34:MET:HB3	1:D:78:LEU:HD22	1.89	0.55
2:B:21:LEU:O	2:B:72:THR:HA	2.09	0.53
1:D:214:LYS:HE2	2:E:213:GLU:HB2	1.90	0.53
1:A:34:MET:HB3	1:A:78:LEU:HD22	1.91	0.52
1:D:119:PRO:HB3	1:D:145:TYR:HB3	1.91	0.51
1:A:90:TYR:O	1:A:106:GLY:HA2	2.10	0.51
2:E:197:THR:HG22	2:E:204:PRO:HB3	1.91	0.51
1:A:87:THR:HG23	1:A:110:THR:HA	1.92	0.51
1:A:125:ALA:HB1	1:A:213:PRO:HA	1.92	0.51
3:C:590:GLU:HG3	3:C:621:LYS:HD2	1.92	0.50
1:A:119:PRO:HB3	1:A:145:TYR:HB3	1.94	0.49
3:C:609:LYS:O	3:C:721:GLY:HA2	2.15	0.47
3:F:589:LYS:HA	3:F:589:LYS:HD3	1.80	0.46
2:E:21:LEU:O	2:E:72:THR:HA	2.15	0.46
2:B:124:GLN:OE1	2:B:131:SER:N	2.43	0.45
3:F:633:GLU:HG3	3:F:634:LYS:HG3	2.00	0.44
1:A:30:ASN:OD1	1:A:30:ASN:N	2.46	0.44
2:B:131:SER:HB3	2:B:180:THR:HG23	2.00	0.44
2:E:122:ASP:OD1	2:E:122:ASP:N	2.44	0.44
2:B:128:GLY:HA2	2:B:183:LYS:HB2	1.98	0.44
2:B:122:ASP:N	2:B:122:ASP:OD1	2.51	0.44
1:A:47:TRP:NE1	1:A:50:SER:OG	2.44	0.43
1:A:123:PRO:HB3	1:A:211:VAL:HG12	1.99	0.43
3:C:628:LEU:HB2	3:C:631:SER:HB2	1.99	0.43
1:D:188:SER:HA	1:D:191:THR:HG22	2.00	0.43
2:B:18:ARG:HG3	2:B:76:SER:HA	2.01	0.43
2:E:54:ARG:HG2	2:E:58:ILE:HB	2.00	0.42

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:33:THR:HG21	1:A:99:LEU:HD13	2.00	0.42
1:D:22:CYS:HB3	1:D:78:LEU:HB3	2.02	0.42
2:E:34:ALA:HA	2:E:48:ILE:O	2.19	0.42
3:F:609:LYS:O	3:F:721:GLY:HA2	2.19	0.42
3:F:643:PRO:HG3	3:F:651:LEU:HG	2.02	0.42
1:A:52:SER:OG	1:A:52(A):THR:N	2.53	0.41
2:B:120:PRO:HD3	2:B:132:VAL:HG22	2.01	0.41
2:E:132:VAL:HG13	2:E:179:LEU:HB3	2.02	0.41
1:D:184:VAL:HG22	1:D:185:PRO:HD2	2.02	0.41
3:F:699:ALA:HA	3:F:727:VAL:O	2.20	0.41
2:E:37:GLN:HB2	2:E:47:LEU:HD11	2.02	0.41
3:C:742:GLY:O	3:C:745:HIS:ND1	2.54	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	222/227 (98%)	214 (96%)	8 (4%)	0	100	100
1	D	223/227 (98%)	214 (96%)	9 (4%)	0	100	100
2	B	210/215 (98%)	204 (97%)	6 (3%)	0	100	100
2	E	212/215 (99%)	208 (98%)	4 (2%)	0	100	100
3	C	183/199 (92%)	169 (92%)	14 (8%)	0	100	100
3	F	168/199 (84%)	158 (94%)	10 (6%)	0	100	100
All	All	1218/1282 (95%)	1167 (96%)	51 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	188/191 (98%)	188 (100%)	0	100	100
1	D	189/191 (99%)	189 (100%)	0	100	100
2	B	184/186 (99%)	184 (100%)	0	100	100
2	E	185/186 (100%)	185 (100%)	0	100	100
3	C	172/185 (93%)	172 (100%)	0	100	100
3	F	159/185 (86%)	159 (100%)	0	100	100
All	All	1077/1124 (96%)	1077 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes 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 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

### 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	224/227 (98%)	0.44	23 (10%) 6 5	84, 127, 171, 200	0
1	D	225/227 (99%)	0.05	7 (3%) 49 45	60, 81, 126, 163	0
2	B	212/215 (98%)	0.40	20 (9%) 8 6	86, 116, 186, 217	0
2	E	214/215 (99%)	0.17	8 (3%) 41 38	67, 95, 130, 151	0
3	C	185/199 (92%)	0.51	19 (10%) 6 5	70, 100, 154, 170	0
3	F	170/199 (85%)	0.42	12 (7%) 16 13	62, 93, 142, 165	0
All	All	1230/1282 (95%)	0.32	89 (7%) 15 12	60, 101, 163, 217	0

All (89) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	138	LEU	4.8
1	A	194	TYR	4.2
3	F	604	SER	4.0
3	F	656	THR	4.0
1	A	207	VAL	4.0
3	C	745	HIS	4.0
3	C	736	GLU	3.8
1	A	1	ASP	3.7
3	C	746	PRO	3.6
1	A	170	LEU	3.5
3	C	713	ASP	3.5
3	C	656	THR	3.5
2	B	186	TYR	3.4
3	F	714	ASP	3.4
2	B	141	PRO	3.3
3	C	678	GLU	3.3
2	B	126	LYS	3.3
3	F	605	GLY	3.2
3	F	657	LYS	3.2

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
3	C	562	LEU	3.2
3	F	659	LYS	3.2
2	E	152	ASN	3.2
3	F	655	GLU	3.1
1	A	193	THR	3.1
1	A	15	GLY	3.1
3	F	654	GLU	3.0
3	C	564	TYR	3.0
2	B	129	THR	3.0
3	C	655	GLU	2.9
2	B	180	THR	2.9
1	A	192	GLN	2.9
2	E	134	CYS	2.9
2	E	178	THR	2.8
2	B	1	GLU	2.8
1	A	13	LYS	2.8
3	C	676	VAL	2.8
2	B	208	SER	2.7
2	B	199	GLN	2.7
1	D	188	SER	2.7
2	B	125	LEU	2.7
1	A	14	PRO	2.6
1	D	30	ASN	2.6
1	D	179	SER	2.6
1	A	111	VAL	2.6
2	B	59	PRO	2.6
1	D	139	GLY	2.6
3	C	563	SER	2.6
2	B	181	LEU	2.5
2	B	192	TYR	2.5
3	F	633	GLU	2.5
2	E	18	ARG	2.4
1	A	28	THR	2.4
3	F	713	ASP	2.4
1	A	114	ALA	2.4
2	E	88	CYS	2.4
1	A	158	ALA	2.4
3	C	635	LEU	2.4
3	C	606	PRO	2.4
1	D	131	THR	2.3
2	B	210	ASN	2.3
2	B	187	GLU	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	157	GLY	2.3
1	A	189	LEU	2.3
2	E	205	VAL	2.3
2	E	154	LEU	2.3
2	B	8	PRO	2.3
2	B	211	ARG	2.2
1	A	113	PRO	2.2
3	C	604	SER	2.2
3	C	566	THR	2.2
1	A	112	SER	2.2
3	C	611	CYS	2.2
3	F	736	GLU	2.2
1	A	126	PRO	2.1
1	A	198	VAL	2.1
1	D	1	ASP	2.1
2	B	184	ALA	2.1
1	D	49	SER	2.1
3	F	658	LEU	2.1
1	A	11	LEU	2.0
2	B	77	SER	2.0
3	C	654	GLU	2.0
2	B	106	SER	2.0
2	E	151	ASP	2.0
1	A	159	LEU	2.0
2	B	193	ALA	2.0
3	C	640	GLU	2.0
1	A	201	LYS	2.0
3	C	659	LYS	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 monosaccharides 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.