



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:15 PM GMT

PDB ID : 4HPQ  
Title : Crystal Structure of the Atg17-Atg31-Atg29 Complex  
Authors : Stanley, R.E; Ragusa, M.J; Hurley, J.H  
Deposited on : 2012-10-24  
Resolution : 3.06 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  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

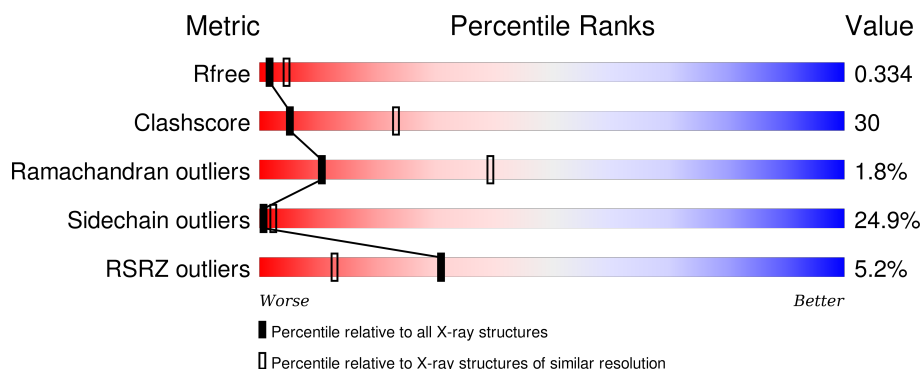
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.06 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))
$R_{free}$	91344	1191 (3.10-3.02)
Clashscore	102246	1303 (3.10-3.02)
Ramachandran outliers	100387	1254 (3.10-3.02)
Sidechain outliers	100360	1254 (3.10-3.02)
RSRZ outliers	91569	1197 (3.10-3.02)

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. 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	69	<div> <div>4%</div> <div>65% 28% 6%</div> </div>
1	D	69	<div> <div>3%</div> <div>62% 29% 7%</div> </div>
2	B	159	<div> <div>%</div> <div>39% 26% 8% 28%</div> </div>
2	E	159	<div> <div></div> <div>37% 26% 9% 28%</div> </div>
3	C	413	<div> <div>8%</div> <div>37% 44% 14%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	413	<div><div></div><div>4%</div><div>37%</div><div>46%</div><div>13%</div><div></div></div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 9220 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 Atg31.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	69	Total	C	N	O	S	0	0	0
			429	266	82	80	1			
1	D	69	Total	C	N	O	S	0	0	0
			429	266	82	80	1			

- Molecule 2 is a protein called Atg29.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	115	Total	C	N	O	S	0	0	0
			932	587	156	185	4			
2	E	115	Total	C	N	O	S	0	0	0
			932	587	156	185	4			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	87	MET	LEU	ENGINEERED MUTATION	UNP C5DEB9
B	110	MET	LEU	ENGINEERED MUTATION	UNP C5DEB9
B	146	ALA	-	EXPRESSION TAG	UNP C5DEB9
B	147	GLY	-	EXPRESSION TAG	UNP C5DEB9
B	148	GLN	-	EXPRESSION TAG	UNP C5DEB9
B	149	PHE	-	EXPRESSION TAG	UNP C5DEB9
B	150	TYR	-	EXPRESSION TAG	UNP C5DEB9
B	151	LEU	-	EXPRESSION TAG	UNP C5DEB9
B	152	ASN	-	EXPRESSION TAG	UNP C5DEB9
B	153	ALA	-	EXPRESSION TAG	UNP C5DEB9
B	154	HIS	-	EXPRESSION TAG	UNP C5DEB9
B	155	HIS	-	EXPRESSION TAG	UNP C5DEB9
B	156	HIS	-	EXPRESSION TAG	UNP C5DEB9
B	157	HIS	-	EXPRESSION TAG	UNP C5DEB9
B	158	HIS	-	EXPRESSION TAG	UNP C5DEB9
B	159	HIS	-	EXPRESSION TAG	UNP C5DEB9

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	87	MET	LEU	ENGINEERED MUTATION	UNP C5DEB9
E	110	MET	LEU	ENGINEERED MUTATION	UNP C5DEB9
E	146	ALA	-	EXPRESSION TAG	UNP C5DEB9
E	147	GLY	-	EXPRESSION TAG	UNP C5DEB9
E	148	GLN	-	EXPRESSION TAG	UNP C5DEB9
E	149	PHE	-	EXPRESSION TAG	UNP C5DEB9
E	150	TYR	-	EXPRESSION TAG	UNP C5DEB9
E	151	LEU	-	EXPRESSION TAG	UNP C5DEB9
E	152	ASN	-	EXPRESSION TAG	UNP C5DEB9
E	153	ALA	-	EXPRESSION TAG	UNP C5DEB9
E	154	HIS	-	EXPRESSION TAG	UNP C5DEB9
E	155	HIS	-	EXPRESSION TAG	UNP C5DEB9
E	156	HIS	-	EXPRESSION TAG	UNP C5DEB9
E	157	HIS	-	EXPRESSION TAG	UNP C5DEB9
E	158	HIS	-	EXPRESSION TAG	UNP C5DEB9
E	159	HIS	-	EXPRESSION TAG	UNP C5DEB9

- Molecule 3 is a protein called Atg17.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	396	Total	C	N	O	S	0	0	0
			3249	2036	565	635	13			
3	F	396	Total	C	N	O	S	0	0	0
			3249	2036	565	635	13			

### 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 errors 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.

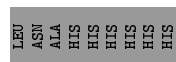
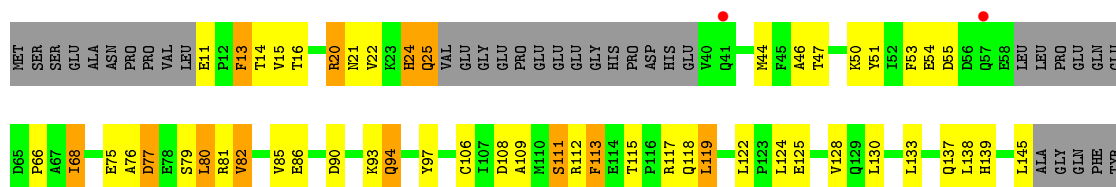
#### • Molecule 1: Atg31



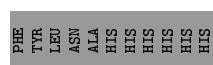
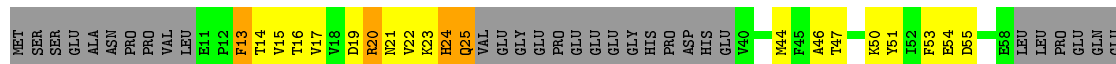
#### • Molecule 1: Atg31



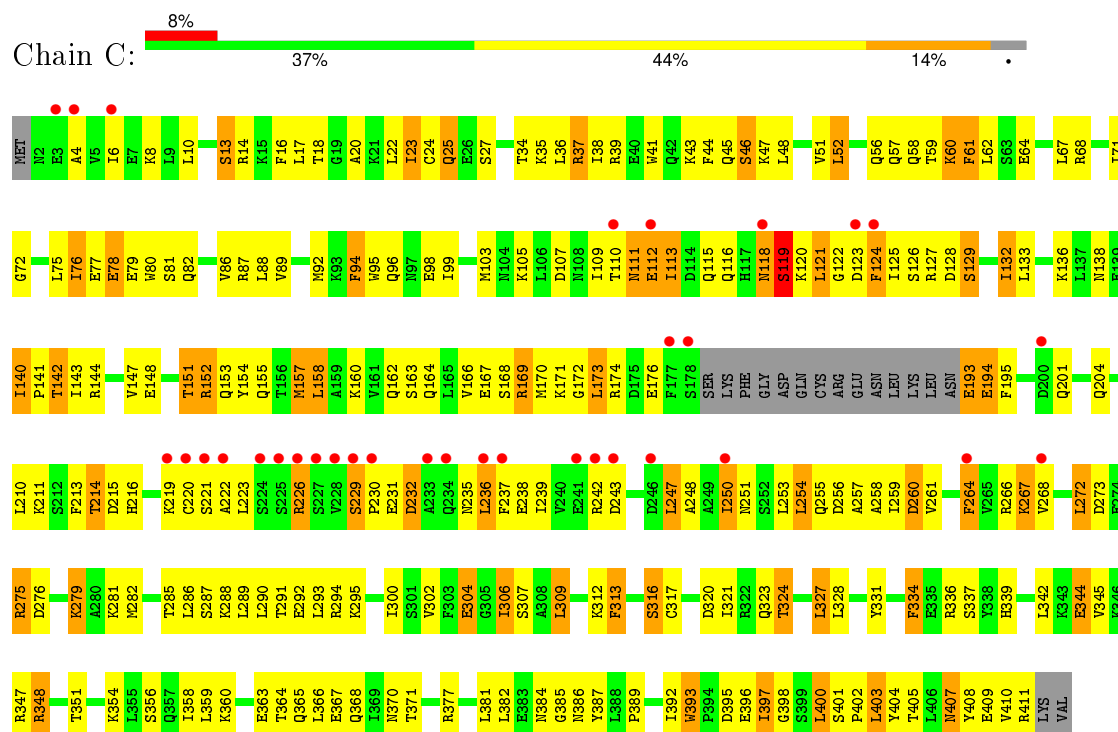
#### • Molecule 2: Atg29



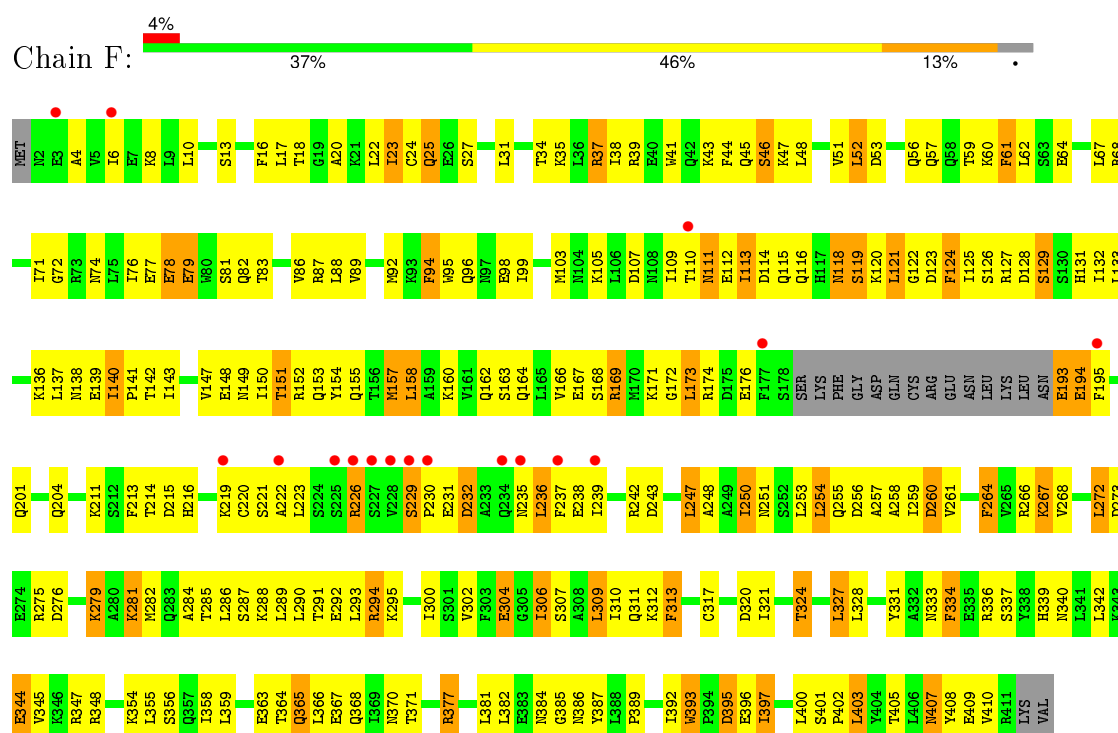
#### • Molecule 2: Atg29



• Molecule 3: Atg17



• Molecule 3: Atg17



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	144.37Å 64.20Å 184.21Å 90.00° 110.79° 90.00°	Depositor
Resolution (Å)	46.56 – 3.06 46.51 – 3.06	Depositor EDS
% Data completeness (in resolution range)	75.4 (46.56-3.06) 75.6 (46.51-3.06)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.37 (at 3.06Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.303 , 0.336 0.302 , 0.334	Depositor DCC
$R_{free}$ test set	3058 reflections (7.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	64.1	Xtriage
Anisotropy	0.108	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 43.5	EDS
Estimated twinning fraction	0.003 for h,-k,-h-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	2 of 45810 reflections (0.004%)	Xtriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	9220	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.82% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.



## 5 Model quality

### 5.1 Standard geometry

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	0.44	0/288	0.69	0/391
1	D	0.53	0/288	0.73	0/391
2	B	0.50	0/944	0.75	0/1276
2	E	0.66	0/944	0.85	0/1276
3	C	0.50	0/3292	0.73	1/4426 (0.0%)
3	F	0.63	0/3292	0.77	0/4426
All	All	0.57	0/9048	0.76	1/12186 (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
2	B	0	1
2	E	0	1
All	All	0	2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	152	ARG	NE-CZ-NH1	5.24	122.92	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	111	SER	Peptide
2	E	111	SER	Peptide

## 5.2 Too-close contacts ⓘ

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	429	0	278	38	0
1	D	429	0	280	42	0
2	B	932	0	911	60	0
2	E	932	0	911	64	0
3	C	3249	0	3253	224	0
3	F	3249	0	3253	223	0
All	All	9220	0	8886	552	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 30.

The worst 5 of 552 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)
3:C:109:ILE:HG21	3:C:121:LEU:HD22	1.17	1.14
3:F:109:ILE:HG21	3:F:121:LEU:HD22	1.18	1.13
1:D:1:MET:HB3	2:E:44:MET:SD	2.11	0.90
3:F:110:THR:O	3:F:111:ASN:CB	2.19	0.89
3:C:393:TRP:CZ2	3:F:345:VAL:HG23	2.08	0.88

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	38/69 (55%)	31 (82%)	6 (16%)	1 (3%)	<b>7</b> 29

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	38/69 (55%)	30 (79%)	7 (18%)	1 (3%)	7	29
2	B	109/159 (69%)	94 (86%)	10 (9%)	5 (5%)	3	16
2	E	109/159 (69%)	94 (86%)	9 (8%)	6 (6%)	2	13
3	C	392/413 (95%)	339 (86%)	49 (12%)	4 (1%)	19	56
3	F	392/413 (95%)	340 (87%)	50 (13%)	2 (0%)	34	70
All	All	1078/1282 (84%)	928 (86%)	131 (12%)	19 (2%)	11	39

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	111	ASN
3	F	111	ASN
2	B	77	ASP
2	B	113	PHE
3	C	113	ILE

### 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	23/36 (64%)	17 (74%)	6 (26%)	0	1
1	D	23/36 (64%)	17 (74%)	6 (26%)	0	1
2	B	107/145 (74%)	87 (81%)	20 (19%)	2	7
2	E	107/145 (74%)	87 (81%)	20 (19%)	2	7
3	C	367/383 (96%)	266 (72%)	101 (28%)	0	1
3	F	367/383 (96%)	272 (74%)	95 (26%)	0	2
All	All	994/1128 (88%)	746 (75%)	248 (25%)	1	2

5 of 248 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	348	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	E	25	GLN
3	F	334	PHE
3	C	363	GLU
3	C	403	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	283	GLN
1	D	2	ASN
3	F	283	GLN
3	C	311	GLN
1	D	5	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 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	40/69 (57%)	0.06	3 (7%) 17 6	59, 94, 149, 158	0
1	D	40/69 (57%)	0.34	2 (5%) 32 13	53, 69, 117, 161	0
2	B	115/159 (72%)	-0.10	2 (1%) 73 49	45, 78, 111, 126	0
2	E	115/159 (72%)	-0.08	0 100 100	23, 50, 94, 115	0
3	C	396/413 (95%)	0.32	33 (8%) 14 5	31, 72, 148, 189	0
3	F	396/413 (95%)	0.16	17 (4%) 39 17	22, 52, 126, 165	0
All	All	1102/1282 (85%)	0.17	57 (5%) 31 13	22, 67, 135, 189	0

The worst 5 of 57 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	233	ALA	8.4
3	C	237	PHE	6.9
3	C	226	ARG	6.4
3	C	228	VAL	5.8
3	C	178	SER	5.0

### 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.