



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 10:24 PM GMT

PDB ID : 5BUZ  
Title : Crystal Structure of a Complex Between the SNARE Vam3 and the HOPS Vps33-Vps16 subcomplex from *Chaetomium thermophilum*  
Authors : Baker, R.W.; Jeffrey, P.D.; Hughson, F.M.  
Deposited on : 2015-06-04  
Resolution : 3.10 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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026982  
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) : rb-20026982

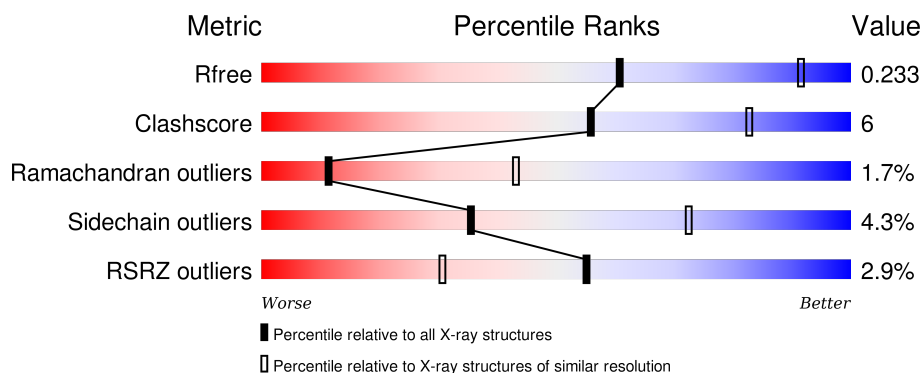
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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	669	<div> <div>73%</div> <div>16%</div> <div>10%</div> </div>
1	D	669	<div> <div>3%</div> <div>74%</div> <div>14%</div> <div>10%</div> </div>
2	B	333	<div> <div>4%</div> <div>80%</div> <div>17%</div> <div>• •</div> </div>
2	E	333	<div> <div>3%</div> <div>78%</div> <div>17%</div> <div>• •</div> </div>
3	C	67	<div> <div>7%</div> <div>42%</div> <div>15%</div> <div>43%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	67	<p>3% 40% 18% 40%</p>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 15276 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 SM (Sec1/Munc18-like) protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	605	Total	C	N	O	S	0	0	0
			4770	3019	841	899	11			
1	D	605	Total	C	N	O	S	0	0	0
			4770	3019	841	899	11			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP G0SCM5
D	-1	GLY	-	expression tag	UNP G0SCM5

- Molecule 2 is a protein called Putative vacuolar protein sorting-associated protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	324	Total	C	N	O	S	0	0	0
			2583	1622	465	487	9			
2	E	324	Total	C	N	O	S	0	0	0
			2583	1622	465	487	9			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	502	MET	-	initiating methionine	UNP G0S6M7
B	503	GLY	-	expression tag	UNP G0S6M7
B	504	SER	-	expression tag	UNP G0S6M7
B	672	ARG	-	insertion	UNP G0S6M7
B	673	MET	-	insertion	UNP G0S6M7
B	674	GLN	-	insertion	UNP G0S6M7
B	675	GLU	-	insertion	UNP G0S6M7
B	676	THR	-	insertion	UNP G0S6M7
B	677	PHE	-	insertion	UNP G0S6M7
B	678	GLU	-	insertion	UNP G0S6M7

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	679	ARG	-	insertion	UNP G0S6M7
B	680	ASP	-	insertion	UNP G0S6M7
B	681	LEU	-	insertion	UNP G0S6M7
B	682	THR	-	insertion	UNP G0S6M7
B	683	ASP	-	insertion	UNP G0S6M7
B	684	SER	-	insertion	UNP G0S6M7
B	685	PHE	-	insertion	UNP G0S6M7
B	686	VAL	-	insertion	UNP G0S6M7
B	687	GLY	-	insertion	UNP G0S6M7
B	688	LEU	-	insertion	UNP G0S6M7
B	689	SER	-	insertion	UNP G0S6M7
B	690	VAL	LEU	engineered mutation	UNP G0S6M7
E	502	MET	-	initiating methionine	UNP G0S6M7
E	503	GLY	-	expression tag	UNP G0S6M7
E	504	SER	-	expression tag	UNP G0S6M7
E	672	ARG	-	insertion	UNP G0S6M7
E	673	MET	-	insertion	UNP G0S6M7
E	674	GLN	-	insertion	UNP G0S6M7
E	675	GLU	-	insertion	UNP G0S6M7
E	676	THR	-	insertion	UNP G0S6M7
E	677	PHE	-	insertion	UNP G0S6M7
E	678	GLU	-	insertion	UNP G0S6M7
E	679	ARG	-	insertion	UNP G0S6M7
E	680	ASP	-	insertion	UNP G0S6M7
E	681	LEU	-	insertion	UNP G0S6M7
E	682	THR	-	insertion	UNP G0S6M7
E	683	ASP	-	insertion	UNP G0S6M7
E	684	SER	-	insertion	UNP G0S6M7
E	685	PHE	-	insertion	UNP G0S6M7
E	686	VAL	-	insertion	UNP G0S6M7
E	687	GLY	-	insertion	UNP G0S6M7
E	688	LEU	-	insertion	UNP G0S6M7
E	689	SER	-	insertion	UNP G0S6M7
E	690	VAL	LEU	engineered mutation	UNP G0S6M7

- Molecule 3 is a protein called SNAP receptor-like protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	38	Total	C	N	O	0	0	0
			275	171	46	58			
3	F	40	Total	C	N	O	0	0	0
			295	182	51	62			

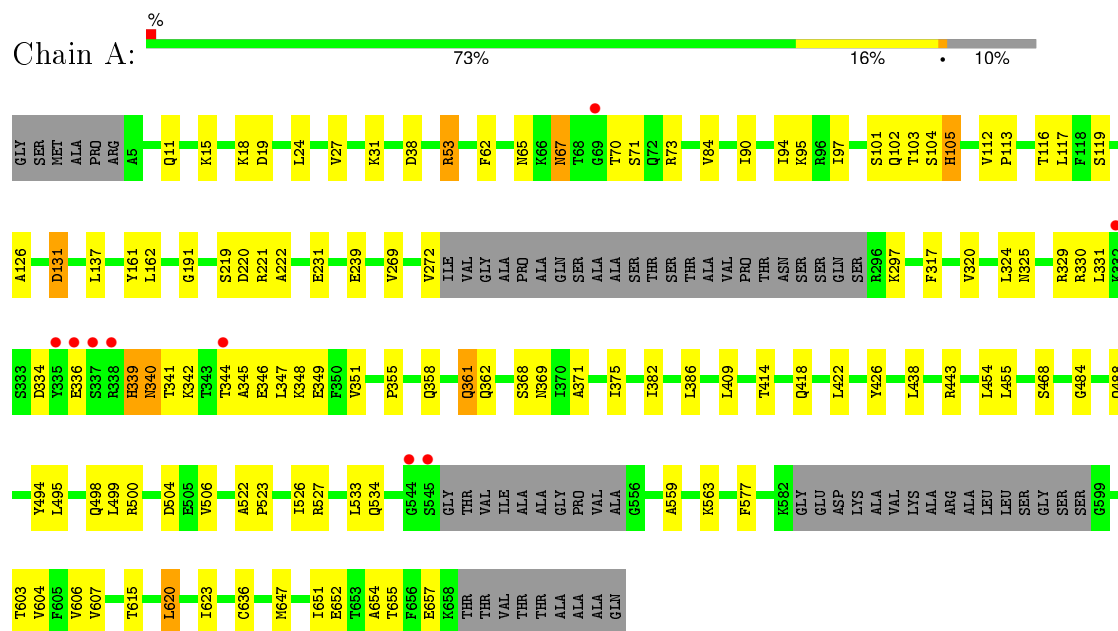
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	179	GLY	-	expression tag	UNP G0S236
C	180	SER	-	expression tag	UNP G0S236
F	179	GLY	-	expression tag	UNP G0S236
F	180	SER	-	expression tag	UNP G0S236

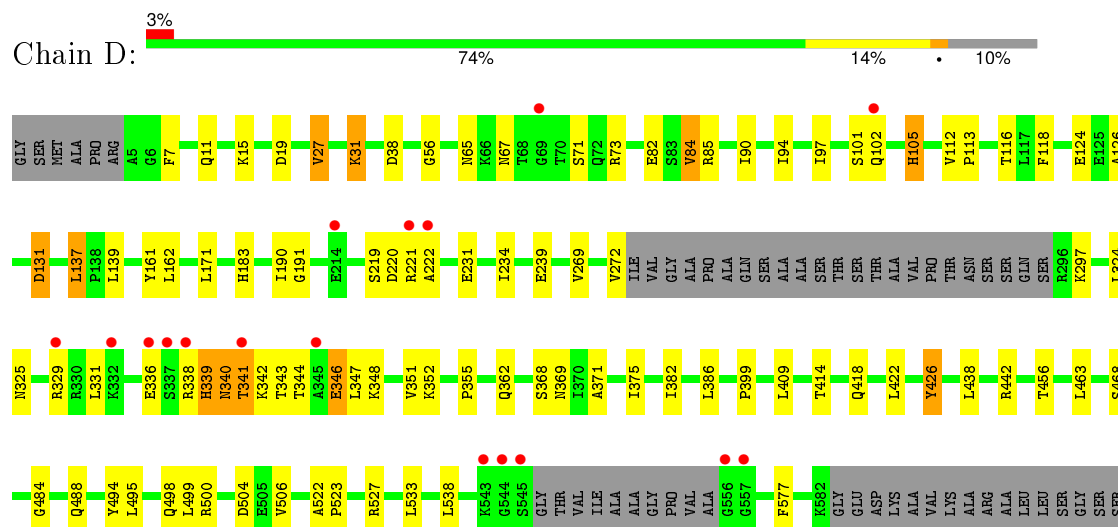
### 3 Residue-property plots [i](#)

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: SM (Sec1/Munc18-like) protein

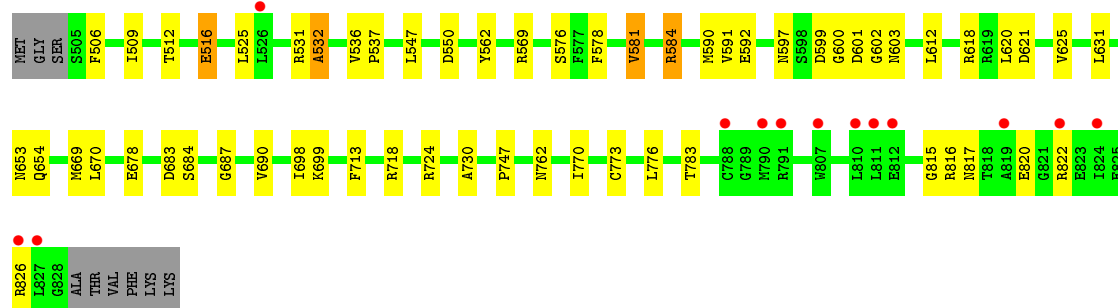
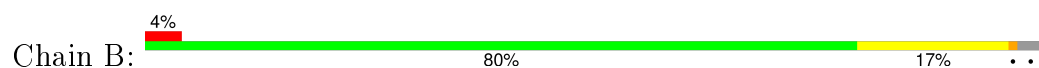


#### • Molecule 1: SM (Sec1/Munc18-like) protein

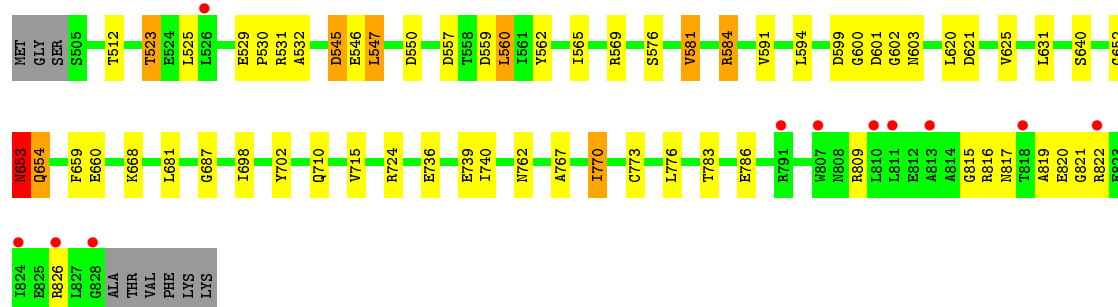
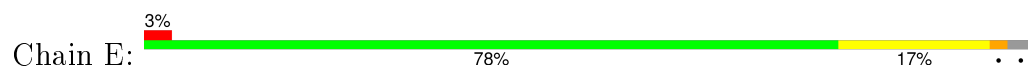




- Molecule 2: Putative vacuolar protein sorting-associated protein



- Molecule 2: Putative vacuolar protein sorting-associated protein



- Molecule 3: SNAP receptor-like protein



- Molecule 3: SNAP receptor-like protein





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	95.04Å 95.95Å 164.06Å 90.00° 99.12° 90.00°	Depositor
Resolution (Å)	47.98 – 3.10 47.98 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.98-3.10) 100.0 (47.98-3.10)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.47 (at 3.12Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.194 , 0.233 0.194 , 0.233	Depositor DCC
$R_{free}$ test set	2640 reflections (4.97%)	DCC
Wilson B-factor (Å <sup>2</sup> )	89.9	Xtriage
Anisotropy	0.374	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 59.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 53146 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	15276	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	104.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.26% 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 [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	0.22	0/4845	0.43	0/6535
1	D	0.23	0/4845	0.44	0/6535
2	B	0.23	0/2626	0.44	0/3543
2	E	0.23	0/2626	0.45	0/3543
3	C	0.22	0/275	0.47	0/375
3	F	0.22	0/295	0.48	0/401
All	All	0.23	0/15512	0.44	0/20932

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	4770	0	4819	61	0
1	D	4770	0	4819	62	0
2	B	2583	0	2581	29	0
2	E	2583	0	2581	44	0
3	C	275	0	254	7	0
3	F	295	0	273	7	0
All	All	15276	0	15327	191	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 6.

The worst 5 of 191 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:500:ARG:O	1:A:527:ARG:NH1	2.20	0.74
2:B:599:ASP:H	2:B:600:GLY:HA2	1.53	0.74
1:D:500:ARG:O	1:D:527:ARG:NH1	2.24	0.70
1:A:484:GLY:O	1:A:488:GLN:NE2	2.26	0.67
1:A:65:ASN:HB3	1:A:67:ASN:OD1	1.96	0.66

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	597/669 (89%)	551 (92%)	38 (6%)	8 (1%)	15	50
1	D	597/669 (89%)	545 (91%)	44 (7%)	8 (1%)	15	50
2	B	322/333 (97%)	286 (89%)	30 (9%)	6 (2%)	10	40
2	E	322/333 (97%)	288 (89%)	26 (8%)	8 (2%)	7	32
3	C	36/67 (54%)	30 (83%)	5 (14%)	1 (3%)	6	30
3	F	38/67 (57%)	34 (90%)	3 (8%)	1 (3%)	7	32
All	All	1912/2138 (89%)	1734 (91%)	146 (8%)	32 (2%)	11	43

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	340	ASN
2	B	532	ALA
2	B	653	ASN
2	B	654	GLN
3	C	215	GLU

### 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	516/564 (92%)	494 (96%)	22 (4%)	35	72
1	D	516/564 (92%)	497 (96%)	19 (4%)	41	76
2	B	270/280 (96%)	258 (96%)	12 (4%)	35	71
2	E	270/280 (96%)	258 (96%)	12 (4%)	35	71
3	C	28/55 (51%)	27 (96%)	1 (4%)	42	77
3	F	30/55 (54%)	26 (87%)	4 (13%)	5	20
All	All	1630/1798 (91%)	1560 (96%)	70 (4%)	35	72

5 of 70 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	762	ASN
1	D	84	VAL
2	E	770	ILE
2	B	770	ILE
1	D	11	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	C	213	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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	605/669 (90%)	-0.10	9 (1%) 76 58	52, 82, 174, 256	0
1	D	605/669 (90%)	-0.10	17 (2%) 56 32	50, 83, 187, 261	0
2	B	324/333 (97%)	0.12	13 (4%) 42 20	59, 116, 210, 260	0
2	E	324/333 (97%)	0.01	11 (3%) 49 24	59, 107, 201, 293	0
3	C	38/67 (56%)	0.49	5 (13%) 4 2	101, 147, 196, 229	0
3	F	40/67 (59%)	0.31	2 (5%) 32 13	104, 141, 182, 205	0
All	All	1936/2138 (90%)	-0.03	57 (2%) 55 31	50, 93, 196, 293	0

The worst 5 of 57 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	545	SER	8.5
1	D	556	GLY	7.6
2	B	824	ILE	6.2
1	D	557	GLY	6.1
1	A	337	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.