



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 08:23 AM GMT

PDB ID : 3EGX  
Title : Crystal structure of the mammalian COPII-coat protein Sec23a/24a complexed with the SNARE protein Sec22b and bound to the transport signal sequence of the SNARE protein Bet1  
Authors : Goldberg, J.; Mancias, J.D.  
Deposited on : 2008-09-11  
Resolution : 3.30 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

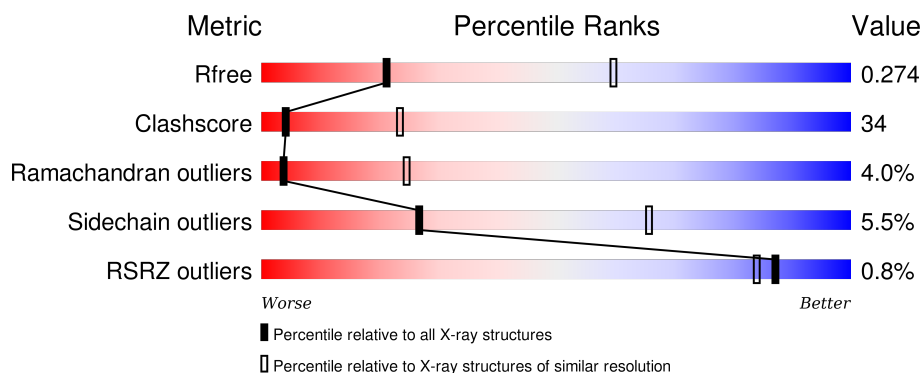
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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	2060 (3.40-3.20)
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)

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. 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	764	<div> <div></div> <div> <div></div> <div>45%</div> <div>41%</div> <div>6%</div> <div>7%</div> </div> </div>
2	B	748	<div> <div></div> <div> <div></div> <div>46%</div> <div>46%</div> <div>5%</div> <div>• •</div> </div> </div>
3	C	157	<div> <div></div> <div> <div></div> <div>41%</div> <div>40%</div> <div>•</div> <div>14%</div> </div> </div>
4	D	9	<div> <div>44%</div> <div> <div>22%</div> <div>22%</div> <div>22%</div> <div>33%</div> </div> </div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 12523 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 Protein transport protein Sec23A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	708	Total	C	N	O	S	0	0	0
			5627	3585	968	1034	40			

- Molecule 2 is a protein called Protein transport protein Sec24A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	729	Total	C	N	O	S	0	0	0
			5761	3675	981	1071	34			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1056	ALA	ARG	CONFLICT	UNP O95486

- Molecule 3 is a protein called Vesicle-trafficking protein SEC22b.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	135	Total	C	N	O	S	0	0	0
			1088	699	177	204	8			

- Molecule 4 is a protein called 9-residue synthetic peptide from SNARE protein Bet1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	6	Total	C	N	O	S	0	0	0
			45	25	6	13	1			

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Zn	0	0
			1	1		

*Continued on next page...*

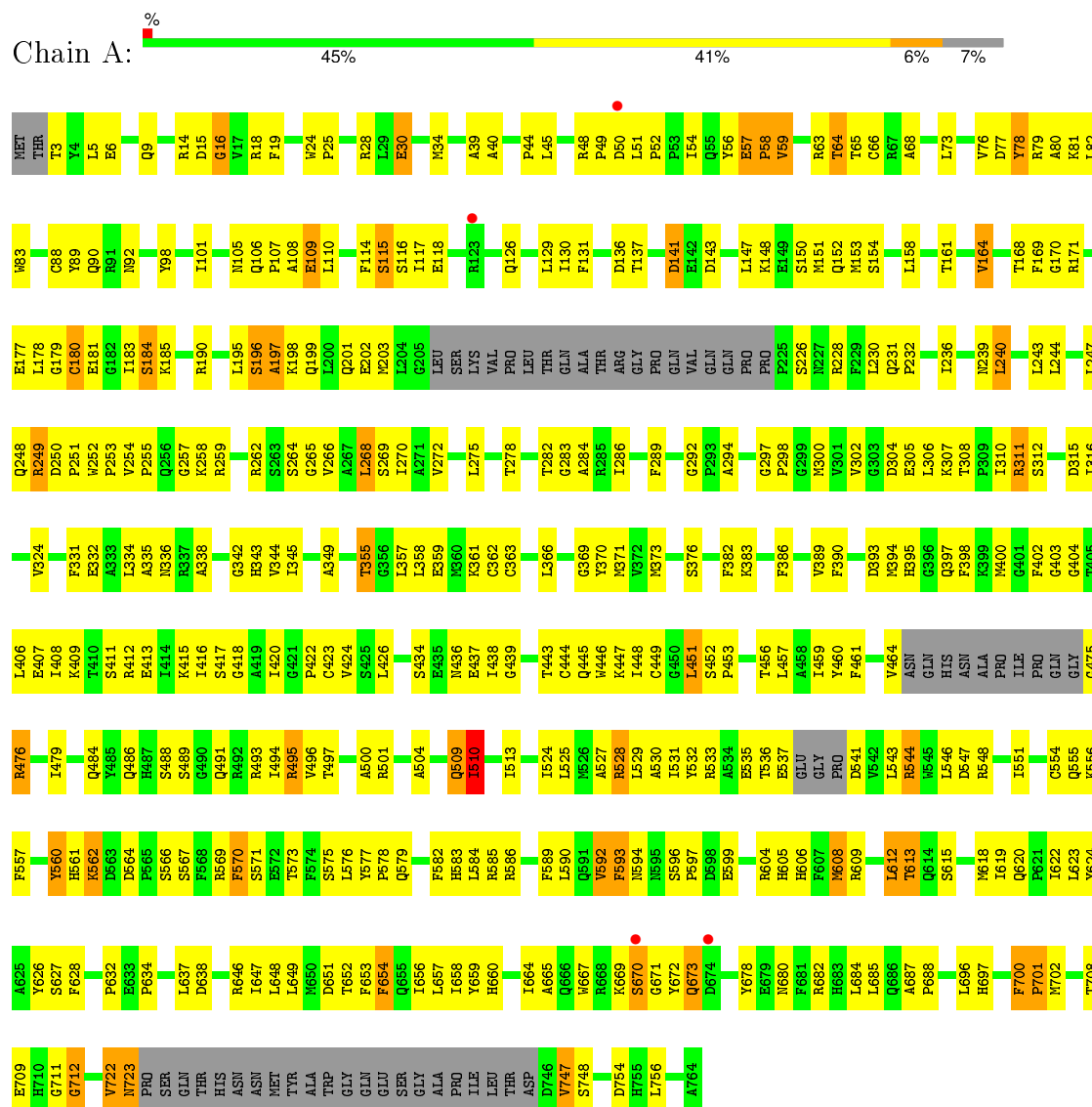
*Continued from previous page...*

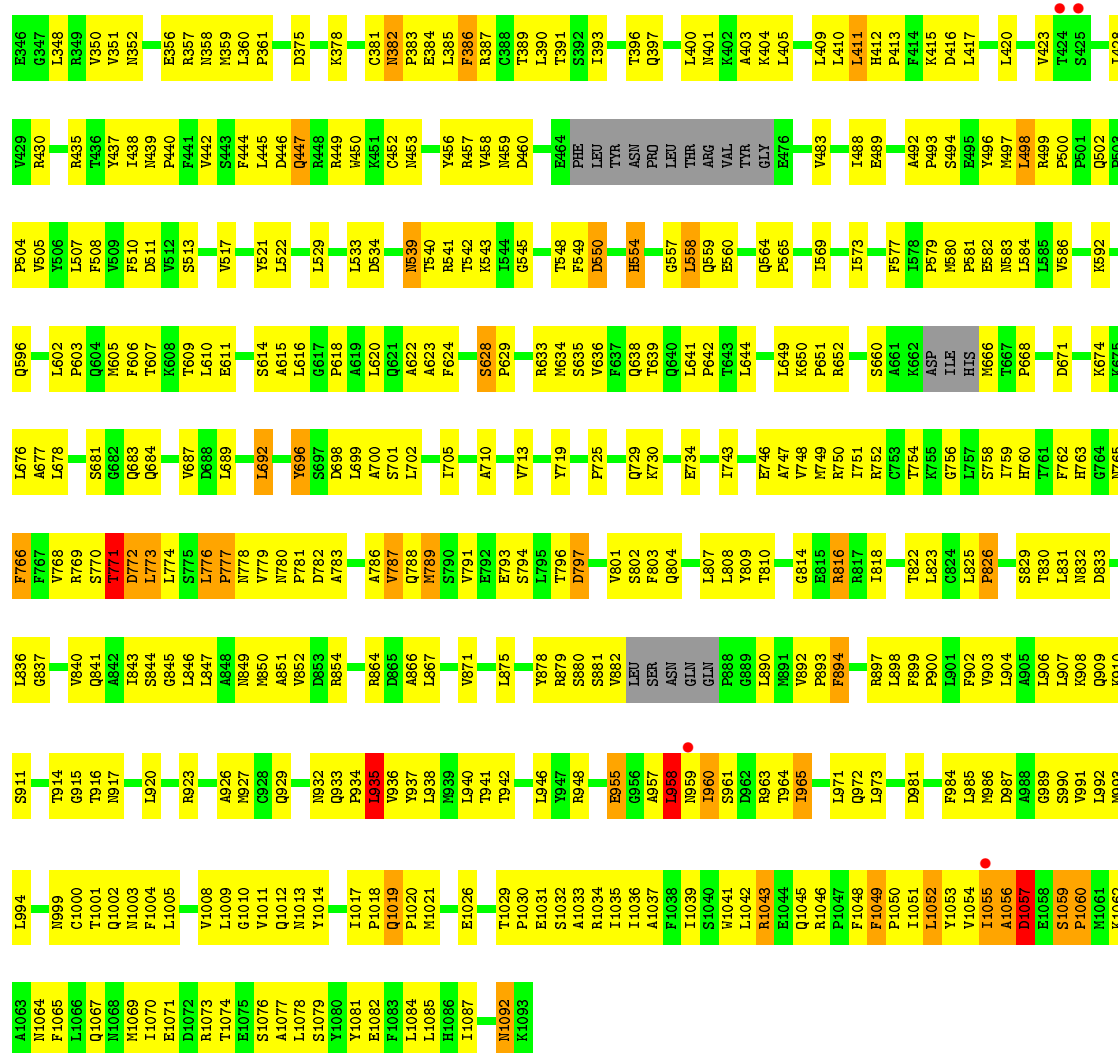
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Zn	0	0
			1	1		

### 3 Residue-property plots

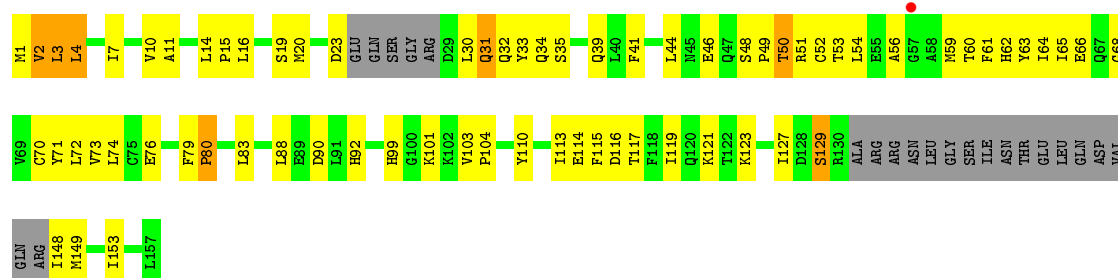
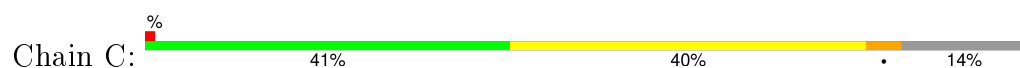
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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 ( $\text{RSRZ} > 2$ ). 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: Protein transport protein Sec23A





• Molecule 3: Vesicle-trafficking protein SEC22b



• Molecule 4: 9-residue synthetic peptide from SNARE protein Bet1



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	148.15Å 97.23Å 129.51Å 90.00° 90.01° 90.00°	Depositor
Resolution (Å)	25.00 – 3.30 24.69 – 3.30	Depositor EDS
% Data completeness (in resolution range)	(Not available) (25.00-3.30) 86.6 (24.69-3.30)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.55 (at 3.30Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.206 , 0.292 0.199 , 0.274	Depositor DCC
$R_{free}$ test set	1226 reflections (5.09%)	DCC
Wilson B-factor (Å <sup>2</sup> )	59.4	Xtriage
Anisotropy	0.122	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 50.7	EDS
Estimated twinning fraction	0.024 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 25569 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	12523	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.42	0/5758	0.71	2/7795 (0.0%)
2	B	0.45	1/5884 (0.0%)	0.74	7/7997 (0.1%)
3	C	0.39	0/1107	0.66	0/1489
4	D	0.90	0/44	1.04	0/56
All	All	0.44	1/12793 (0.0%)	0.72	9/17337 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	771	THR	CA-CB	6.55	1.70	1.53

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	57	GLU	N-CA-C	-8.09	89.17	111.00
1	A	115	SER	N-CA-C	-7.83	89.84	111.00
2	B	772	ASP	CB-CG-OD1	6.24	123.91	118.30
2	B	773	LEU	CB-CG-CD1	-6.19	100.48	111.00
2	B	773	LEU	CA-CB-CG	-5.67	102.26	115.30

There are no chirality outliers.

There are no planarity outliers.

### 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	5627	0	5575	365	0
2	B	5761	0	5815	423	0
3	C	1088	0	1091	65	0
4	D	45	0	31	17	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
All	All	12523	0	12512	845	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 34.

The worst 5 of 845 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:510:ILE:HD12	1:A:510:ILE:H	1.05	1.11
2:B:620:LEU:HD22	2:B:634:MET:HE3	1.33	1.06
1:A:48:ARG:HB2	1:A:49:PRO:HD2	1.33	1.06
2:B:1019:GLN:HB3	2:B:1020:PRO:HD3	1.38	1.01
2:B:437:TYR:HB2	2:B:804:GLN:NE2	1.77	0.99

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	698/764 (91%)	583 (84%)	84 (12%)	31 (4%)	3	22
2	B	721/748 (96%)	608 (84%)	90 (12%)	23 (3%)	5	31
3	C	129/157 (82%)	106 (82%)	16 (12%)	7 (5%)	2	17
4	D	4/9 (44%)	3 (75%)	0	1 (25%)	0	0
All	All	1552/1678 (92%)	1300 (84%)	190 (12%)	62 (4%)	4	24

5 of 62 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	64	THR
1	A	196	SER
1	A	324	VAL
1	A	509	GLN
1	A	700	PHE

### 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	619/666 (93%)	588 (95%)	31 (5%)	30	68
2	B	660/678 (97%)	622 (94%)	38 (6%)	25	64
3	C	119/138 (86%)	112 (94%)	7 (6%)	24	63
4	D	4/7 (57%)	3 (75%)	1 (25%)	1	3
All	All	1402/1489 (94%)	1325 (94%)	77 (6%)	27	66

5 of 77 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	387	ARG
2	B	635	SER
3	C	46	GLU
2	B	411	LEU
2	B	554	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	673	GLN
2	B	382	ASN
2	B	1067	GLN
1	A	714	GLN
1	A	723	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](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 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	708/764 (92%)	-0.40	4 (0%) 90 88	6, 41, 90, 131	0
2	B	729/748 (97%)	-0.55	4 (0%) 91 90	2, 33, 81, 125	0
3	C	135/157 (85%)	-0.11	1 (0%) 89 86	14, 69, 106, 135	0
4	D	6/9 (66%)	2.24	4 (66%) 0 0	89, 99, 104, 107	0
All	All	1578/1678 (94%)	-0.43	13 (0%) 87 84	2, 39, 94, 135	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	123	ARG	4.1
4	D	32	ASN	4.0
1	A	670	SER	3.6
4	D	28	CYS	3.1
2	B	1055	ILE	3.1

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	ZN	B	1094	1/1	0.99	0.06	-1.76	43,43,43,43	0
5	ZN	A	765	1/1	0.99	0.03	-2.10	34,34,34,34	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.