



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 19, 2024 – 04:48 pm BST

PDB ID : 8ROH  
Title : Human cohesin SMC3-HD(EQ)/RAD21-N complex - Apo conformation  
Authors : Vitoria Gomes, M.; Romier, C.  
Deposited on : 2024-01-11  
Resolution : 2.60 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  
Mogul : 1.8.4, CSD as541be (2020)  
Xtrriage (Phenix) : 1.13  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.002 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.38.2

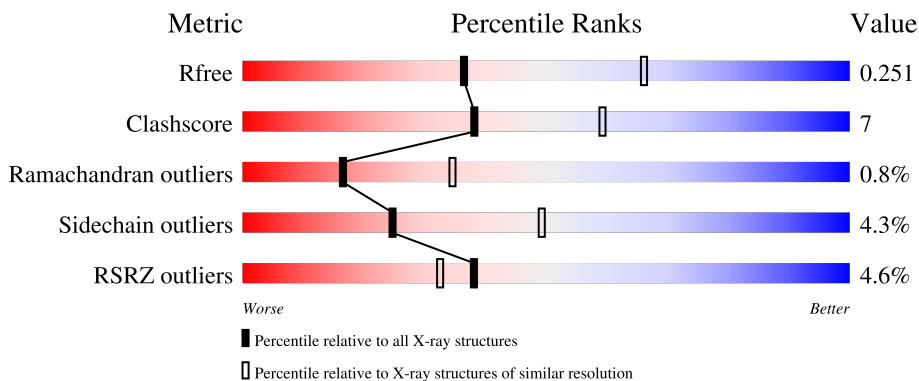
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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))
$R_{free}$	164625	3775 (2.60-2.60)
Clashscore	180529	4181 (2.60-2.60)
Ramachandran outliers	177936	4129 (2.60-2.60)
Sidechain outliers	177891	4129 (2.60-2.60)
RSRZ outliers	164620	3775 (2.60-2.60)

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	478	 3% 70% 15% 14%
2	B	110	 8% 57% 22% 21%

## 2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 4081 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 Structural maintenance of chromosomes protein 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	413	3351	2131	578	628	14	0	0	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	959	GLY	-	linker	UNP Q9UQE7
A	960	SER	-	linker	UNP Q9UQE7
A	961	GLY	-	linker	UNP Q9UQE7
A	962	SER	-	linker	UNP Q9UQE7
A	963	LEU	-	linker	UNP Q9UQE7
A	964	VAL	-	linker	UNP Q9UQE7
A	965	PRO	-	linker	UNP Q9UQE7
A	966	ARG	-	linker	UNP Q9UQE7
A	967	GLY	-	linker	UNP Q9UQE7
A	968	SER	-	linker	UNP Q9UQE7
A	969	GLY	-	linker	UNP Q9UQE7
A	970	SER	-	linker	UNP Q9UQE7
A	1144	GLN	GLU	engineered mutation	UNP Q9UQE7

- Molecule 2 is a protein called Double-strand-break repair protein rad21 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	87	700	458	127	111	4	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	103	GLY	-	expression tag	UNP O60216
B	104	SER	-	expression tag	UNP O60216
B	105	LEU	-	expression tag	UNP O60216

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	106	GLU	-	expression tag	UNP O60216
B	107	VAL	-	expression tag	UNP O60216
B	108	LEU	-	expression tag	UNP O60216
B	109	PHE	-	expression tag	UNP O60216
B	110	GLN	-	expression tag	UNP O60216

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
3	A	1	5	4	1	0	0
3	A	1	5	4	1	0	0
3	A	1	5	4	1	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 6 3 3	0	0

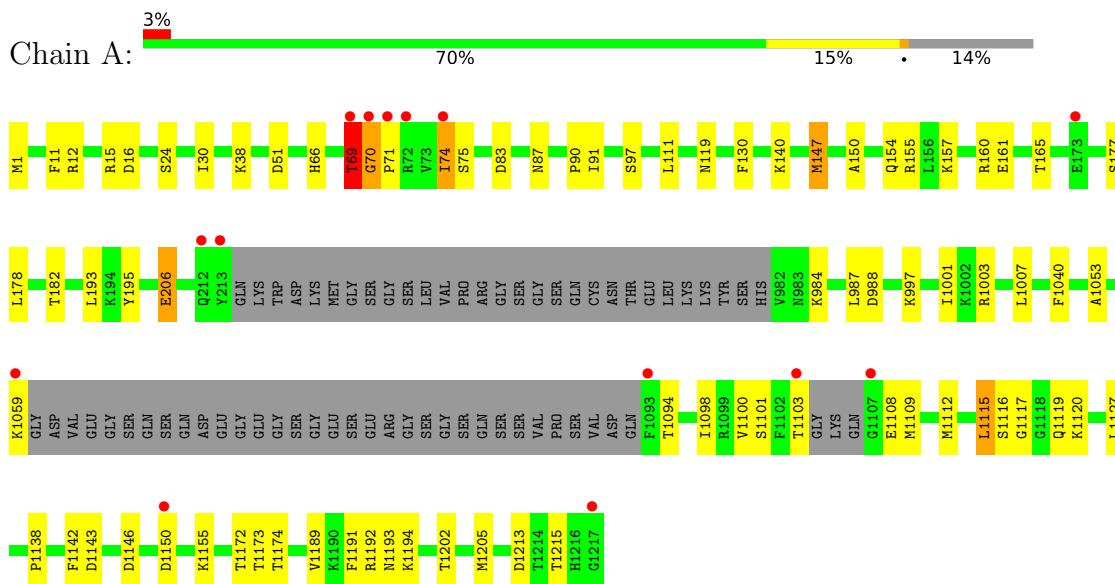
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	9	Total O 9 9	0	0

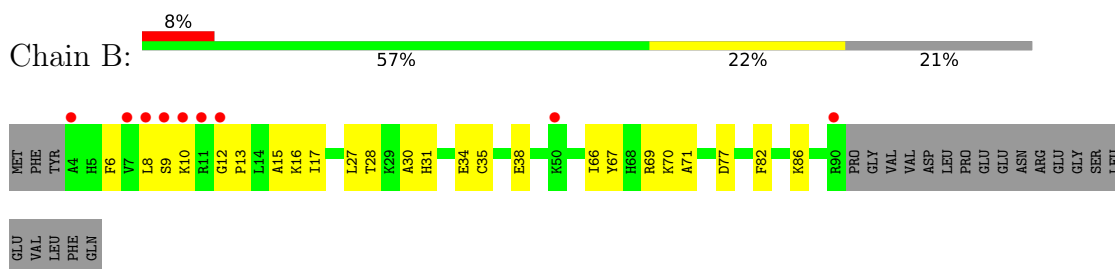
### 3 Residue-property plots [i](#)

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: Structural maintenance of chromosomes protein 3



- Molecule 2: Double-strand-break repair protein rad21 homolog



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.79Å 90.79Å 236.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.60 – 2.60 49.60 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.60-2.60) 100.0 (49.60-2.60)	Depositor EDS
$R_{merge}$	0.28	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.46 (at 2.61Å)	Xtrriage
Refinement program	PHENIX 1.19rc4_4035	Depositor
R, $R_{free}$	0.216 , 0.255 0.215 , 0.251	Depositor DCC
$R_{free}$ test set	1484 reflections (4.73%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	61.3	Xtrriage
Anisotropy	0.338	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 51.8	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	4081	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	71.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.54% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, SO4

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	0.43	0/3403	0.59	0/4559
2	B	0.39	0/716	0.58	0/961
All	All	0.43	0/4119	0.59	0/5520

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	3351	0	3397	46	0
2	B	700	0	746	14	0
3	A	15	0	0	1	0
4	A	6	0	8	0	0
5	A	9	0	0	0	0
All	All	4081	0	4151	56	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 7.

All (56) 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:1003:ARG:NH2	2:B:77:ASP:OD2	2.14	0.80
2:B:35:CYS:O	2:B:70:LYS:NZ	2.14	0.80
1:A:140:LYS:HG2	1:A:1143:ASP:HB3	1.74	0.69
1:A:1053:ALA:HB2	1:A:1100:VAL:HG23	1.83	0.61
1:A:1191:PHE:CZ	1:A:1194:LYS:HA	2.35	0.60
1:A:1116:SER:HB3	1:A:1119:GLN:OE1	2.06	0.55
2:B:8:LEU:HD21	2:B:12:GLY:HA3	1.87	0.54
1:A:147:MET:HE1	1:A:155:ARG:HG3	1.88	0.54
1:A:1142:PHE:HB2	1:A:1172:THR:HG22	1.89	0.54
2:B:10:LYS:HG3	2:B:15:ALA:HA	1.89	0.54
1:A:1:MET:HA	1:A:83:ASP:O	2.08	0.53
1:A:15:ARG:HD2	1:A:16:ASP:OD1	2.08	0.53
1:A:150:ALA:O	1:A:155:ARG:NH2	2.41	0.53
1:A:69:THR:OG1	1:A:70:GLY:N	2.41	0.53
1:A:119:ASN:HB2	3:A:2002:SO4:O4	2.08	0.53
1:A:1116:SER:OG	1:A:1117:GLY:N	2.43	0.52
1:A:997:LYS:O	1:A:1001:ILE:HG12	2.10	0.52
2:B:6:PHE:CE2	2:B:9:SER:HA	2.45	0.52
1:A:206:GLU:HG2	1:A:987:LEU:HD21	1.91	0.51
2:B:17:ILE:HD13	2:B:66:ILE:HD12	1.90	0.51
1:A:984:LYS:HA	1:A:987:LEU:HD12	1.93	0.49
1:A:1202:THR:OG1	1:A:1205:MET:HG3	2.13	0.49
2:B:28:THR:HG23	2:B:31:HIS:CE1	2.48	0.49
1:A:91:ILE:HD11	1:A:111:LEU:HD21	1.94	0.48
1:A:30:ILE:HB	1:A:1173:THR:HG22	1.94	0.48
1:A:38:LYS:HG2	1:A:1189:VAL:HG21	1.96	0.48
1:A:1115:LEU:HB3	1:A:1120:LYS:HG3	1.96	0.48
1:A:11:PHE:CD2	1:A:12:ARG:HG2	2.49	0.47
2:B:82:PHE:CE2	2:B:86:LYS:HD2	2.49	0.47
1:A:90:PRO:O	1:A:91:ILE:HG13	2.15	0.47
1:A:182:THR:HG21	2:B:67:TYR:CE2	2.50	0.46
1:A:38:LYS:HB2	1:A:38:LYS:HE2	1.82	0.45
1:A:12:ARG:HH11	1:A:12:ARG:HG3	1.81	0.45
1:A:147:MET:SD	1:A:1127:LEU:HD23	2.57	0.45
1:A:130:PHE:CZ	1:A:1138:PRO:HD2	2.52	0.45
1:A:160:ARG:O	1:A:165:THR:HG23	2.17	0.45
1:A:15:ARG:HB2	1:A:66:HIS:CE1	2.52	0.44
1:A:195:TYR:HB3	2:B:82:PHE:CE1	2.53	0.44
1:A:1112:MET:HE2	1:A:1112:MET:HB2	1.52	0.44
2:B:30:ALA:O	2:B:34:GLU:HG2	2.18	0.43
1:A:1059:LYS:HB2	1:A:1094:THR:OG1	2.18	0.43
1:A:157:LYS:O	1:A:161:GLU:HG3	2.19	0.43

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1007:LEU:HD11	2:B:71:ALA:HA	2.02	0.42
1:A:1101:SER:HB2	1:A:1109:MET:HG2	2.01	0.42
1:A:1103:THR:OG1	1:A:1103:THR:O	2.38	0.42
1:A:1098:ILE:HD11	1:A:1127:LEU:HD22	2.02	0.42
1:A:1146:ASP:OD2	1:A:1174:THR:OG1	2.33	0.42
1:A:193:LEU:HD13	1:A:1001:ILE:HD13	2.01	0.42
1:A:38:LYS:HB3	1:A:1173:THR:HG21	2.02	0.42
1:A:74:ILE:H	1:A:74:ILE:HG13	1.76	0.41
1:A:984:LYS:O	1:A:988:ASP:N	2.44	0.41
1:A:30:ILE:HB	1:A:1173:THR:CG2	2.51	0.41
2:B:16:LYS:NZ	2:B:34:GLU:HG3	2.36	0.41
2:B:27:LEU:O	2:B:69:ARG:NH2	2.45	0.40
1:A:1192:ARG:O	1:A:1193:ASN:HB2	2.22	0.40
1:A:1040:PHE:CE2	1:A:1053:ALA:HB1	2.56	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	405/478 (85%)	386 (95%)	16 (4%)	3 (1%)	19	38
2	B	85/110 (77%)	78 (92%)	6 (7%)	1 (1%)	11	24
All	All	490/588 (83%)	464 (95%)	22 (4%)	4 (1%)	16	34

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	70	GLY
1	A	69	THR
1	A	71	PRO
2	B	13	PRO

### 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	368/421 (87%)	350 (95%)	18 (5%)	21	43
2	B	74/95 (78%)	73 (99%)	1 (1%)	62	82
All	All	442/516 (86%)	423 (96%)	19 (4%)	25	49

All (19) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	24	SER
1	A	51	ASP
1	A	69	THR
1	A	74	ILE
1	A	75	SER
1	A	87	ASN
1	A	97	SER
1	A	147	MET
1	A	154	GLN
1	A	177	SER
1	A	178	LEU
1	A	206	GLU
1	A	1108	GLU
1	A	1115	LEU
1	A	1150	ASP
1	A	1155	LYS
1	A	1213	ASP
1	A	1215	THR
2	B	38	GLU

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 oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

4 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	SO4	A	2002	-	4,4,4	0.80	0	6,6,6	1.04	1 (16%)
4	GOL	A	2004	-	5,5,5	1.05	0	5,5,5	0.96	0
3	SO4	A	2001	-	4,4,4	0.77	0	6,6,6	1.00	0
3	SO4	A	2003	-	4,4,4	0.88	0	6,6,6	0.83	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	GOL	A	2004	-	-	4/4/4/4	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2002	SO4	O4-S-O3	-2.17	99.78	109.06

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	2004	GOL	C1-C2-C3-O3
4	A	2004	GOL	O1-C1-C2-O2
4	A	2004	GOL	O1-C1-C2-C3
4	A	2004	GOL	O2-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	2002	SO4	1	0

## 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	413/478 (86%)	0.31	14 (3%) 48 42	46, 66, 111, 152	0
2	B	87/110 (79%)	0.50	9 (10%) 13 11	55, 67, 112, 132	0
All	All	500/588 (85%)	0.34	23 (4%) 38 32	46, 66, 112, 152	0

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	7	VAL	7.8
2	B	8	LEU	6.4
1	A	1107	GLY	5.8
1	A	1093	PHE	5.6
1	A	1059	LYS	5.4
1	A	213	TYR	5.0
2	B	9	SER	4.0
1	A	1217	GLY	4.0
1	A	71	PRO	3.9
1	A	72	ARG	3.9
2	B	4	ALA	3.6
1	A	1103	THR	3.4
2	B	10	LYS	3.3
1	A	69	THR	3.3
2	B	11	ARG	2.8
2	B	90	ARG	2.6
1	A	74	ILE	2.5
2	B	12	GLY	2.5
1	A	173	GLU	2.4
1	A	1150	ASP	2.2
1	A	212	GLN	2.1
1	A	70	GLY	2.1
2	B	50	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](#)

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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	SO4	A	2003	5/5	0.64	0.11	119,123,148,152	0
4	GOL	A	2004	6/6	0.81	0.18	78,91,93,95	0
3	SO4	A	2002	5/5	0.93	0.13	82,83,94,95	0
3	SO4	A	2001	5/5	0.95	0.08	53,59,60,66	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.