



# Full wwPDB X-ray Structure Validation Report ⓘ

May 16, 2020 – 07:59 am BST

PDB ID : 4DH2  
Title : Crystal structure of Coh-OlpC(Cthe\_0452)-Doc435(Cthe\_0435) complex: A novel type I Cohesin-Dockerin complex from Clostridium thermocellum ATTC 27405  
Authors : Alves, V.D.; Carvalho, A.L.; Najmudin, S.H.; Bras, J.; Prates, J.A.M.; Fontes, C.M.G.A.  
Deposited on : 2012-01-27  
Resolution : 1.75 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

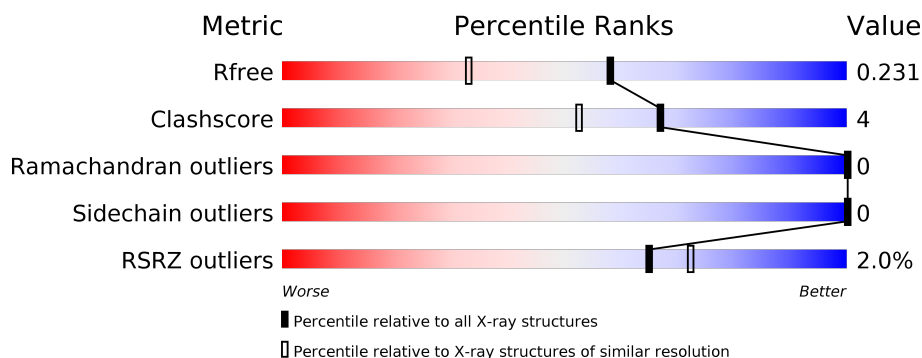
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.75 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	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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 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	170	<div> <div>2%</div> <div> <div></div> <div>80%</div> <div>13%</div> <div>7%</div> </div> </div>
1	C	170	<div> <div></div> <div> <div>86%</div> <div>6%</div> <div>8%</div> </div> </div>
2	B	82	<div> <div>%</div> <div> <div></div> <div>85%</div> <div>•</div> <div>12%</div> </div> </div>
2	D	82	<div> <div>5%</div> <div> <div></div> <div>83%</div> <div>•</div> <div>15%</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7829 atoms, of which 3648 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 Cellulosome anchoring protein cohesin region.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	158	Total	C	H	N	O	S	0	4	0
			2539	825	1268	192	253	1			
1	C	156	Total	C	H	N	O	S	0	4	0
			2522	814	1268	190	249	1			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP A3DCL1
A	2	ALA	-	EXPRESSION TAG	UNP A3DCL1
A	3	SER	-	EXPRESSION TAG	UNP A3DCL1
A	163	LEU	-	EXPRESSION TAG	UNP A3DCL1
A	164	GLU	-	EXPRESSION TAG	UNP A3DCL1
A	165	HIS	-	EXPRESSION TAG	UNP A3DCL1
A	166	HIS	-	EXPRESSION TAG	UNP A3DCL1
A	167	HIS	-	EXPRESSION TAG	UNP A3DCL1
A	168	HIS	-	EXPRESSION TAG	UNP A3DCL1
A	169	HIS	-	EXPRESSION TAG	UNP A3DCL1
A	170	HIS	-	EXPRESSION TAG	UNP A3DCL1
C	1	MET	-	INITIATING METHIONINE	UNP A3DCL1
C	2	ALA	-	EXPRESSION TAG	UNP A3DCL1
C	3	SER	-	EXPRESSION TAG	UNP A3DCL1
C	163	LEU	-	EXPRESSION TAG	UNP A3DCL1
C	164	GLU	-	EXPRESSION TAG	UNP A3DCL1
C	165	HIS	-	EXPRESSION TAG	UNP A3DCL1
C	166	HIS	-	EXPRESSION TAG	UNP A3DCL1
C	167	HIS	-	EXPRESSION TAG	UNP A3DCL1
C	168	HIS	-	EXPRESSION TAG	UNP A3DCL1
C	169	HIS	-	EXPRESSION TAG	UNP A3DCL1
C	170	HIS	-	EXPRESSION TAG	UNP A3DCL1

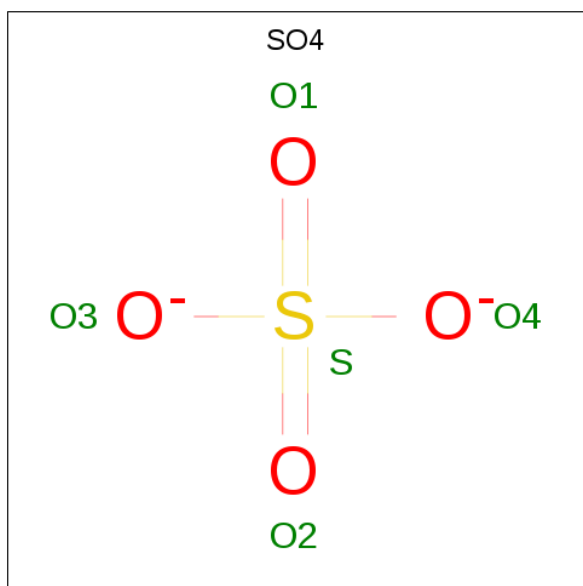
- Molecule 2 is a protein called Dockerin type 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	72	Total	C	H	N	O	S	0	0	0
			1142	371	560	97	110	4			
2	D	70	Total	C	H	N	O	S	0	0	0
			1116	359	552	95	106	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	INITIATING METHIONINE	UNP A3DCJ4
D	1	MET	-	INITIATING METHIONINE	UNP A3DCJ4

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



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

*Continued on next page...*

*Continued from previous page...*

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

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	Ca	0	0
			2	2		
4	D	2	Total	Ca	0	0
			2	2		

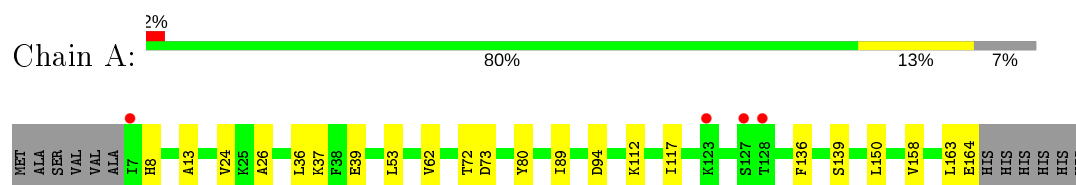
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	163	Total	O	0	0
			163	163		
5	B	65	Total	O	0	0
			65	65		
5	C	177	Total	O	0	0
			177	177		
5	D	56	Total	O	0	0
			56	56		

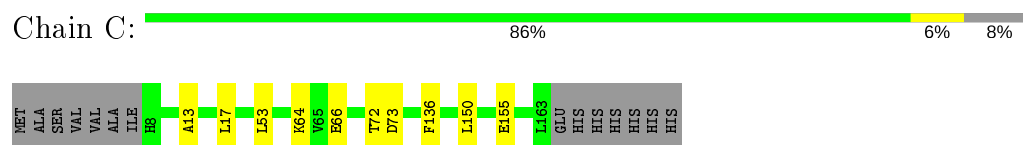
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

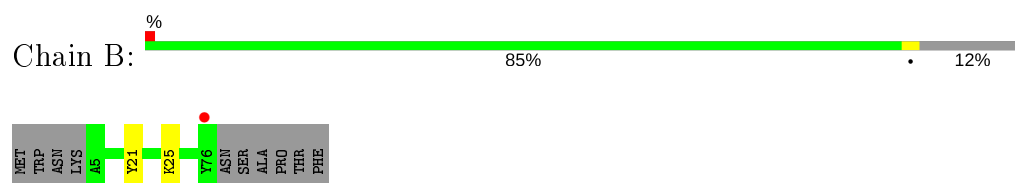
- Molecule 1: Cellulosome anchoring protein cohesin region



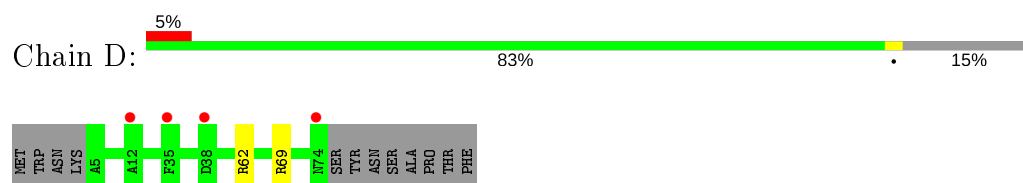
- Molecule 1: Cellulosome anchoring protein cohesin region



- Molecule 2: Dockerin type 1



- Molecule 2: Dockerin type 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.76 Å 90.76 Å 135.07 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.30 – 1.75 43.02 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.9 (39.30-1.75) 96.6 (43.02-1.75)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.02 (at 1.75 Å)	Xtriage
Refinement program	PHENIX 1.7.2_865	Depositor
R, $R_{free}$	0.181 , 0.215 0.197 , 0.231	Depositor DCC
$R_{free}$ test set	3345 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.6	Xtriage
Anisotropy	0.152	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.43 , 51.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.000 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	7829	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 45.53 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.3039e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: CA, 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	0.54	0/1306	0.69	0/1763
1	C	0.56	0/1289	0.69	0/1740
2	B	0.52	0/593	0.62	0/805
2	D	0.46	0/574	0.56	0/779
All	All	0.53	0/3762	0.66	0/5087

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	1271	1268	1283	18	0
1	C	1254	1268	1266	8	0
2	B	582	560	565	2	0
2	D	564	552	551	2	0
3	A	15	0	0	1	0
3	B	10	0	0	0	0
3	C	20	0	0	0	0
4	B	2	0	0	0	0
4	D	2	0	0	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	163	0	0	3	0
5	B	65	0	0	0	0
5	C	177	0	0	0	0
5	D	56	0	0	2	0
All	All	4181	3648	3665	29	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 4.

All (29) 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:A:201:SO4:O4	5:A:455:HOH:O	1.99	0.80
2:D:62:ARG:NH2	5:D:256:HOH:O	2.19	0.74
1:A:80:TYR:OH	5:A:392:HOH:O	2.06	0.72
1:A:112:LYS:NZ	5:A:317:HOH:O	2.36	0.57
1:A:163:LEU:O	1:A:164:GLU:HB2	2.03	0.57
1:C:53:LEU:HD21	1:C:136:PHE:CE1	2.42	0.55
1:A:53:LEU:HD21	1:A:136:PHE:CE1	2.42	0.54
2:D:69:ARG:NH2	5:D:251:HOH:O	2.39	0.53
1:A:37:LYS:NZ	1:A:39[B]:GLU:OE2	2.41	0.53
1:A:53:LEU:CD2	1:A:136:PHE:CE1	2.93	0.52
1:A:8:HIS:HD2	1:C:155:GLU:OE2	1.95	0.49
1:A:53:LEU:HD23	1:A:136:PHE:CD1	2.49	0.48
1:A:36:LEU:HD13	1:A:53:LEU:HD11	1.97	0.47
1:A:26:ALA:HB3	1:A:117:ILE:CD1	2.45	0.46
2:B:21:TYR:CZ	2:B:25:LYS:HE3	2.51	0.46
1:A:24:VAL:HG13	1:A:158:VAL:HG13	1.97	0.45
1:A:37:LYS:CE	1:A:39[B]:GLU:OE2	2.66	0.43
1:C:53:LEU:CD2	1:C:136:PHE:CE1	3.02	0.43
1:A:8:HIS:CE1	1:A:139:SER:HB3	2.52	0.43
1:A:13:ALA:HB3	1:A:150:LEU:HG	2.00	0.43
1:A:94:ASP:OD2	2:B:25:LYS:HD2	2.19	0.43
1:C:72:THR:O	1:C:73:ASP:C	2.57	0.42
1:C:13:ALA:HB3	1:C:150:LEU:HG	2.01	0.42
1:A:62:VAL:HG21	1:A:89:ILE:HD11	2.02	0.41
1:C:53:LEU:CD2	1:C:136:PHE:CD1	3.03	0.41
1:C:17:LEU:HD12	1:C:17:LEU:C	2.41	0.41
1:A:37:LYS:HE2	1:A:39[B]:GLU:OE2	2.20	0.41
1:C:64:LYS:HE3	1:C:66[B]:GLU:CG	2.51	0.41
1:A:72:THR:O	1:A:73:ASP:C	2.59	0.41

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	160/170 (94%)	158 (99%)	2 (1%)	0	100	100
1	C	158/170 (93%)	157 (99%)	1 (1%)	0	100	100
2	B	70/82 (85%)	68 (97%)	2 (3%)	0	100	100
2	D	68/82 (83%)	66 (97%)	2 (3%)	0	100	100
All	All	456/504 (90%)	449 (98%)	7 (2%)	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	144/150 (96%)	144 (100%)	0	100	100
1	C	142/150 (95%)	142 (100%)	0	100	100
2	B	64/73 (88%)	64 (100%)	0	100	100
2	D	62/73 (85%)	62 (100%)	0	100	100
All	All	412/446 (92%)	412 (100%)	0	100	100

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

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	8	HIS
2	B	58	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 13 ligands modelled in this entry, 4 are monoatomic - leaving 9 for Mogul analysis.

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	C	202	-	4,4,4	0.15	0	6,6,6	0.24	0
3	SO4	A	201	-	4,4,4	0.34	0	6,6,6	0.28	0
3	SO4	A	203	-	4,4,4	0.13	0	6,6,6	0.21	0
3	SO4	C	204	-	4,4,4	0.15	0	6,6,6	0.09	0
3	SO4	C	203	-	4,4,4	0.15	0	6,6,6	0.19	0
3	SO4	B	104	-	4,4,4	0.14	0	6,6,6	0.12	0
3	SO4	B	103	-	4,4,4	0.14	0	6,6,6	0.33	0
3	SO4	C	201	-	4,4,4	0.13	0	6,6,6	0.31	0
3	SO4	A	202	-	4,4,4	0.15	0	6,6,6	0.46	0

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	201	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 [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	158/170 (92%)	-0.07	4 (2%) 57 63	15, 27, 53, 116	0
1	C	156/170 (91%)	0.00	0 100 100	17, 26, 43, 59	0
2	B	72/82 (87%)	-0.12	1 (1%) 75 82	16, 26, 59, 85	0
2	D	70/82 (85%)	0.24	4 (5%) 23 29	21, 32, 59, 90	0
All	All	456/504 (90%)	-0.01	9 (1%) 65 72	15, 27, 54, 116	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	12	ALA	3.7
1	A	7	ILE	3.7
2	D	74	ASN	2.8
1	A	127	SER	2.7
2	B	76	TYR	2.5
1	A	128	THR	2.3
2	D	38	ASP	2.3
1	A	123	LYS	2.3
2	D	35	PHE	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 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. 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	C	202	5/5	0.91	0.15	65,84,86,91	0
3	SO4	B	103	5/5	0.92	0.18	42,49,60,69	1
3	SO4	C	204	5/5	0.94	0.18	45,82,87,89	0
3	SO4	A	203	5/5	0.94	0.12	59,74,77,82	0
3	SO4	B	104	5/5	0.95	0.14	69,72,73,74	0
3	SO4	C	203	5/5	0.95	0.23	55,78,83,87	0
3	SO4	A	201	5/5	0.98	0.07	25,34,40,48	0
4	CA	D	101	1/1	0.98	0.06	29,29,29,29	0
4	CA	D	102	1/1	0.98	0.08	27,27,27,27	0
3	SO4	C	201	5/5	0.98	0.08	34,39,55,59	0
3	SO4	A	202	5/5	0.98	0.11	52,58,72,72	0
4	CA	B	101	1/1	0.99	0.13	20,20,20,20	0
4	CA	B	102	1/1	0.99	0.06	21,21,21,21	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.