



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 12:54 pm BST

PDB ID : 2H5G  
Title : Crystal structure of human pyrroline-5-carboxylate synthetase  
Authors : Papagrigoriou, E.; Shafqat, N.; Turnbull, A.P.; Berridge, G.; Hozjan, V.; Kavanagh, K.; Gileadi, O.; Smee, C.; Bray, J.; Gorrec, F.; Sundstrom, M.; Arrowsmith, C.; Weigelt, J.; Edwards, A.; Oppermann, U.; Structural Genomics Consortium (SGC)  
Deposited on : 2006-05-26  
Resolution : 2.25 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

<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

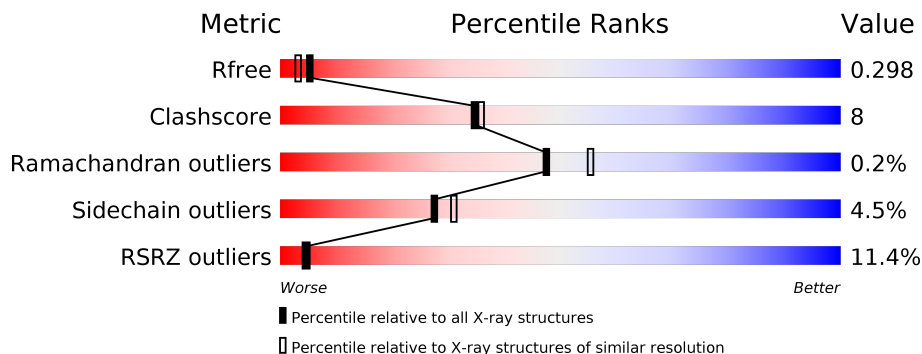
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.25 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	1377 (2.26-2.26)
Clashscore	141614	1487 (2.26-2.26)
Ramachandran outliers	138981	1449 (2.26-2.26)
Sidechain outliers	138945	1450 (2.26-2.26)
RSRZ outliers	127900	1356 (2.26-2.26)

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	463	 13% 72% 16% • 10%
1	B	463	 8% 76% 15% • 8%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	SO4	B	306	-	-	-	X

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 6695 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 Delta 1-pyrroline-5-carboxylate synthetase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	417	3144	1981	553	598	6	6	0	3	0
1	B	425	3239	2038	566	622	7	6	0	4	0

There are 70 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	333	MSE	-	INITIATING METHIONINE	UNP P54886
A	334	HIS	-	EXPRESSION TAG	UNP P54886
A	335	HIS	-	EXPRESSION TAG	UNP P54886
A	336	HIS	-	EXPRESSION TAG	UNP P54886
A	337	HIS	-	EXPRESSION TAG	UNP P54886
A	338	HIS	-	EXPRESSION TAG	UNP P54886
A	339	HIS	-	EXPRESSION TAG	UNP P54886
A	340	SER	-	CLONING ARTIFACT	UNP P54886
A	341	SER	-	CLONING ARTIFACT	UNP P54886
A	342	GLY	-	CLONING ARTIFACT	UNP P54886
A	343	VAL	-	CLONING ARTIFACT	UNP P54886
A	344	ASP	-	CLONING ARTIFACT	UNP P54886
A	345	LEU	-	CLONING ARTIFACT	UNP P54886
A	346	GLY	-	CLONING ARTIFACT	UNP P54886
A	347	THR	-	CLONING ARTIFACT	UNP P54886
A	348	GLU	-	CLONING ARTIFACT	UNP P54886
A	349	ASN	-	CLONING ARTIFACT	UNP P54886
A	350	LEU	-	CLONING ARTIFACT	UNP P54886
A	351	TYR	-	CLONING ARTIFACT	UNP P54886
A	352	PHE	-	CLONING ARTIFACT	UNP P54886
A	353	GLN	-	CLONING ARTIFACT	UNP P54886
A	354	SER	-	CLONING ARTIFACT	UNP P54886
A	355	MSE	-	CLONING ARTIFACT	UNP P54886
A	356	VAL	-	CLONING ARTIFACT	UNP P54886
A	357	LYS	-	CLONING ARTIFACT	UNP P54886

*Continued on next page...*

*Continued from previous page...*

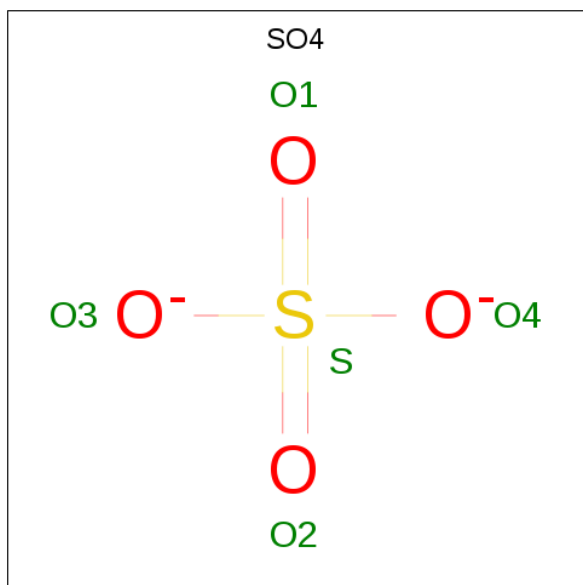
Chain	Residue	Modelled	Actual	Comment	Reference
A	358	PRO	-	CLONING ARTIFACT	UNP P54886
A	359	ALA	-	CLONING ARTIFACT	UNP P54886
A	360	GLY	-	CLONING ARTIFACT	UNP P54886
A	361	PRO	-	CLONING ARTIFACT	UNP P54886
A	369	MSE	MET	MODIFIED RESIDUE	UNP P54886
A	376	MSE	MET	MODIFIED RESIDUE	UNP P54886
A	551	MSE	MET	MODIFIED RESIDUE	UNP P54886
A	577	MSE	MET	MODIFIED RESIDUE	UNP P54886
A	586	MSE	MET	MODIFIED RESIDUE	UNP P54886
A	636	MSE	MET	MODIFIED RESIDUE	UNP P54886
B	333	MSE	-	INITIATING METHIONINE	UNP P54886
B	334	HIS	-	EXPRESSION TAG	UNP P54886
B	335	HIS	-	EXPRESSION TAG	UNP P54886
B	336	HIS	-	EXPRESSION TAG	UNP P54886
B	337	HIS	-	EXPRESSION TAG	UNP P54886
B	338	HIS	-	EXPRESSION TAG	UNP P54886
B	339	HIS	-	EXPRESSION TAG	UNP P54886
B	340	SER	-	CLONING ARTIFACT	UNP P54886
B	341	SER	-	CLONING ARTIFACT	UNP P54886
B	342	GLY	-	CLONING ARTIFACT	UNP P54886
B	343	VAL	-	CLONING ARTIFACT	UNP P54886
B	344	ASP	-	CLONING ARTIFACT	UNP P54886
B	345	LEU	-	CLONING ARTIFACT	UNP P54886
B	346	GLY	-	CLONING ARTIFACT	UNP P54886
B	347	THR	-	CLONING ARTIFACT	UNP P54886
B	348	GLU	-	CLONING ARTIFACT	UNP P54886
B	349	ASN	-	CLONING ARTIFACT	UNP P54886
B	350	LEU	-	CLONING ARTIFACT	UNP P54886
B	351	TYR	-	CLONING ARTIFACT	UNP P54886
B	352	PHE	-	CLONING ARTIFACT	UNP P54886
B	353	GLN	-	CLONING ARTIFACT	UNP P54886
B	354	SER	-	CLONING ARTIFACT	UNP P54886
B	355	MSE	-	CLONING ARTIFACT	UNP P54886
B	356	VAL	-	CLONING ARTIFACT	UNP P54886
B	357	LYS	-	CLONING ARTIFACT	UNP P54886
B	358	PRO	-	CLONING ARTIFACT	UNP P54886
B	359	ALA	-	CLONING ARTIFACT	UNP P54886
B	360	GLY	-	CLONING ARTIFACT	UNP P54886
B	361	PRO	-	CLONING ARTIFACT	UNP P54886
B	369	MSE	MET	MODIFIED RESIDUE	UNP P54886
B	376	MSE	MET	MODIFIED RESIDUE	UNP P54886
B	551	MSE	MET	MODIFIED RESIDUE	UNP P54886

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	577	MSE	MET	MODIFIED RESIDUE	UNP P54886
B	586	MSE	MET	MODIFIED RESIDUE	UNP P54886
B	636	MSE	MET	MODIFIED RESIDUE	UNP P54886

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

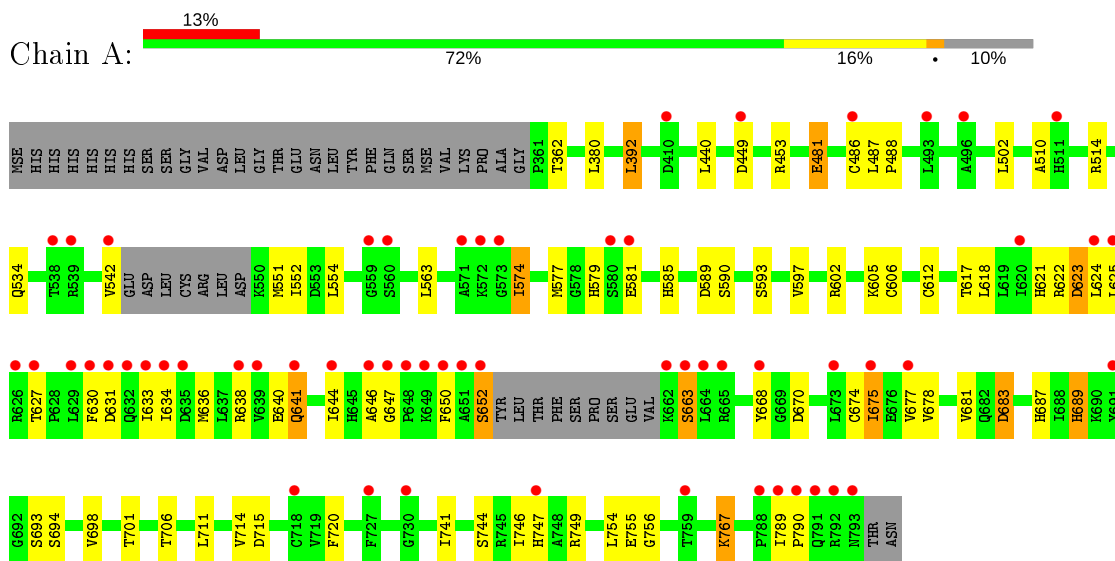
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	127	Total O 130 130	0	3
3	B	150	Total O 152 152	0	3

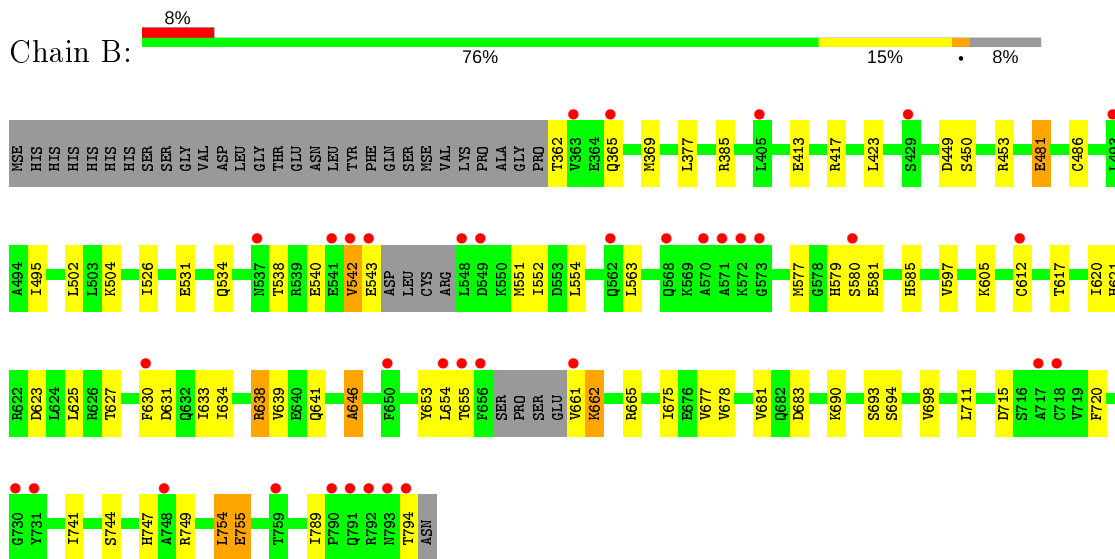
### 3 Residue-property plots [i](#)

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: Delta 1-pyrroline-5-carboxylate synthetase



- Molecule 1: Delta 1-pyrroline-5-carboxylate synthetase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	122.02Å 137.40Å 72.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.56 – 2.25 46.56 – 2.25	Depositor EDS
% Data completeness (in resolution range)	98.1 (46.56-2.25) 98.1 (46.56-2.25)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.03	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	8.26 (at 2.24Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.230 , 0.272 0.262 , 0.298	Depositor DCC
$R_{free}$ test set	2890 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	36.8	Xtrriage
Anisotropy	0.083	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 57.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6695	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 52.64 % 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 4.7277e-05. 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: 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	0.94	14/3196 (0.4%)	0.71	1/4327 (0.0%)
1	B	1.04	15/3299 (0.5%)	0.75	6/4464 (0.1%)
All	All	1.00	29/6495 (0.4%)	0.73	7/8791 (0.1%)

The worst 5 of 29 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	663	SER	CB-OG	20.46	1.68	1.42
1	B	653	TYR	CE2-CZ	19.32	1.63	1.38
1	B	653	TYR	CG-CD2	16.46	1.60	1.39
1	B	540	GLU	CD-OE1	16.13	1.43	1.25
1	A	542	VAL	C-O	13.78	1.49	1.23

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	653	TYR	CB-CG-CD2	-9.39	115.37	121.00
1	B	653	TYR	CD1-CG-CD2	6.64	125.20	117.90
1	A	641	GLN	CG-CD-OE1	-5.76	110.08	121.60
1	B	653	TYR	CG-CD2-CE2	-5.75	116.70	121.30
1	B	653	TYR	CG-CD1-CE1	-5.52	116.88	121.30

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	3144	0	3084	61	0
1	B	3239	0	3202	44	0
2	A	15	0	0	0	0
2	B	15	0	0	0	0
3	A	130	0	0	4	0
3	B	152	0	0	1	0
All	All	6695	0	6286	105	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 8.

The worst 5 of 105 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:641:GLN:CB	1:A:641:GLN:CG	1.74	1.61
1:A:663:SER:CB	1:A:663:SER:OG	1.68	1.41
1:A:577:MSE:HE2	1:A:747[B]:HIS:CD2	1.71	1.26
1:A:577:MSE:HE1	1:A:747[A]:HIS:ND1	1.62	1.14
1:A:502:LEU:HD23	1:A:534:GLN:HB2	1.33	1.05

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	414/463 (89%)	393 (95%)	21 (5%)	0	100	100
1	B	423/463 (91%)	406 (96%)	15 (4%)	2 (0%)	29	29
All	All	837/926 (90%)	799 (96%)	36 (4%)	2 (0%)	47	55

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	662	LYS
1	B	542	VAL

### 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	328/389 (84%)	312 (95%)	16 (5%)	25	27
1	B	346/389 (89%)	331 (96%)	15 (4%)	29	33
All	All	674/778 (87%)	643 (95%)	31 (5%)	27	30

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	754	LEU
1	B	481	GLU
1	B	755	GLU
1	A	767	LYS
1	B	486	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	399	GLN
1	B	621	HIS
1	B	534	GLN
1	B	366	GLN
1	B	579	HIS

### 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](#)

6 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	$\# Z  > 2$
2	SO4	A	302	-	4,4,4	0.48	0	6,6,6	0.33	0
2	SO4	B	304	-	4,4,4	0.28	0	6,6,6	0.43	0
2	SO4	B	303	-	4,4,4	0.18	0	6,6,6	0.47	0
2	SO4	A	301	-	4,4,4	0.14	0	6,6,6	0.38	0
2	SO4	A	305	-	4,4,4	0.15	0	6,6,6	0.19	0
2	SO4	B	306	-	4,4,4	0.14	0	6,6,6	0.40	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.

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

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	411/463 (88%)	0.88	59 (14%) <b>2</b> <b>2</b>	30, 59, 83, 100	0
1	B	419/463 (90%)	0.69	36 (8%) <b>10</b> <b>11</b>	31, 58, 83, 103	0
All	All	830/926 (89%)	0.78	95 (11%) <b>5</b> <b>4</b>	30, 58, 83, 103	0

The worst 5 of 95 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	655	THR	9.2
1	A	626	ARG	7.7
1	B	549	ASP	6.7
1	A	650	PHE	6.1
1	A	625	LEU	6.0

### 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
2	SO4	B	306	5/5	0.79	0.46	75,83,90,91	0
2	SO4	A	302	5/5	0.90	0.20	52,58,79,85	0
2	SO4	A	305	5/5	0.91	0.21	94,95,99,100	0
2	SO4	B	304	5/5	0.93	0.13	52,59,78,79	0
2	SO4	A	301	5/5	0.97	0.11	52,54,57,60	0
2	SO4	B	303	5/5	0.98	0.08	50,51,57,59	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.