



# Full wwPDB X-ray Structure Validation Report ⓘ

May 16, 2020 – 11:20 pm BST

PDB ID : 1OIK  
Title : Crystal structure of the alkylsulfatase AtsK, a non-heme Fe(II)  $\alpha$ -ketoglutarate dependent Dioxygenase in complex with Fe,  $\alpha$ -ketoglutarate and 2-ethyl-1-hexanesulfuric acid  
Authors : Mueller, I.; Kahnert, A.; Pape, T.; Dierks, T.; Meyer-Klauke, W.; Kertesz, M.A.; Uson, I.  
Deposited on : 2003-06-18  
Resolution : 2.06 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

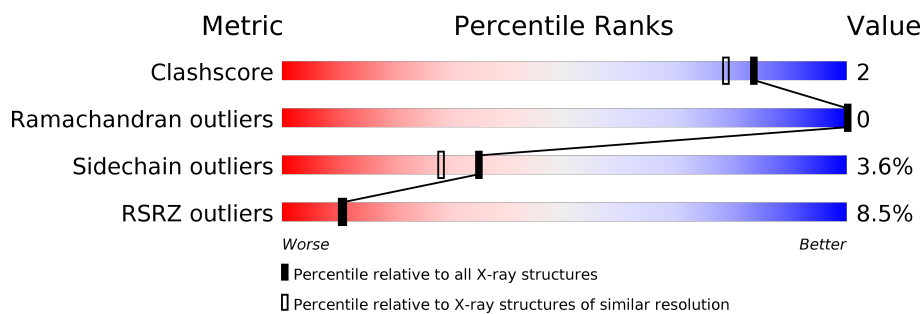
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.06 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))
Clashscore	141614	2801 (2.08-2.04)
Ramachandran outliers	138981	2768 (2.08-2.04)
Sidechain outliers	138945	2768 (2.08-2.04)
RSRZ outliers	127900	2646 (2.08-2.04)

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	301	<div> <div>5%</div> <div>71%</div> <div>8%</div> <div>20%</div> </div>
1	D	301	<div> <div>9%</div> <div>77%</div> <div>7%</div> <div>16%</div> </div>



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.

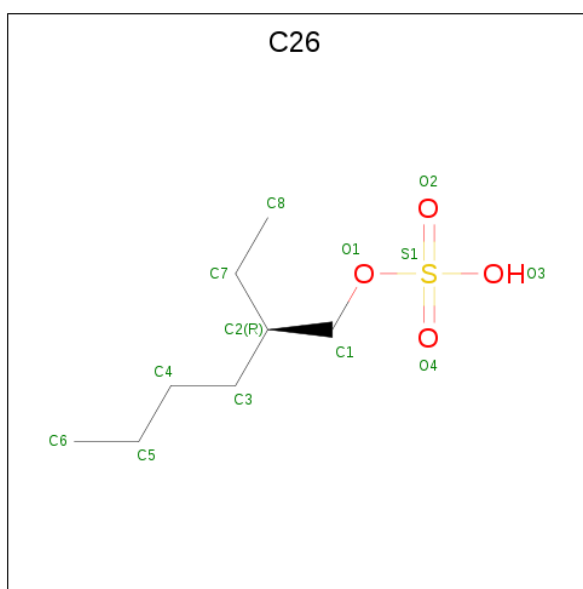
- Molecule 1 is a protein called PUTATIVE ALKYL SULFATASE ATSK.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	240	Total 1865	C 1176	N 347	O 342	0	1	0
1	D	254	Total 1958	C 1236	N 361	O 361	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	165	ARG	GLU	conflict	UNP Q9WWU5
D	165	ARG	GLU	conflict	UNP Q9WWU5

- Molecule 2 is (2R)-2-ETHYL-1-HEXANESULFONIC ACID (three-letter code: C26) (formula:  $\text{C}_8\text{H}_{18}\text{O}_4\text{S}$ ).



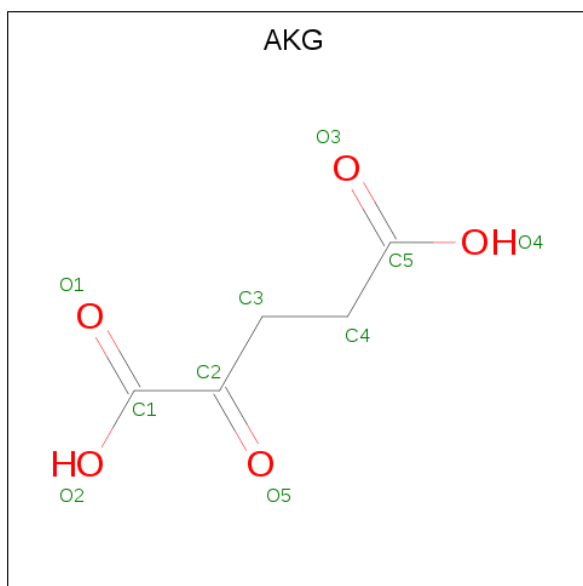
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	O	S	0	0
			13	8	4	1		

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	D	1	Total	C	O	S	0	0
			13	8	4	1		

- Molecule 3 is 2-OXOGLUTARIC ACID (three-letter code: AKG) (formula:  $C_5H_6O_5$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			10	5	5		
3	D	1	Total	C	O	0	0
			10	5	5		

- Molecule 4 is FE (II) ION (three-letter code: FE2) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Fe	0	0
			1	1		
4	D	1	Total	Fe	0	0
			1	1		

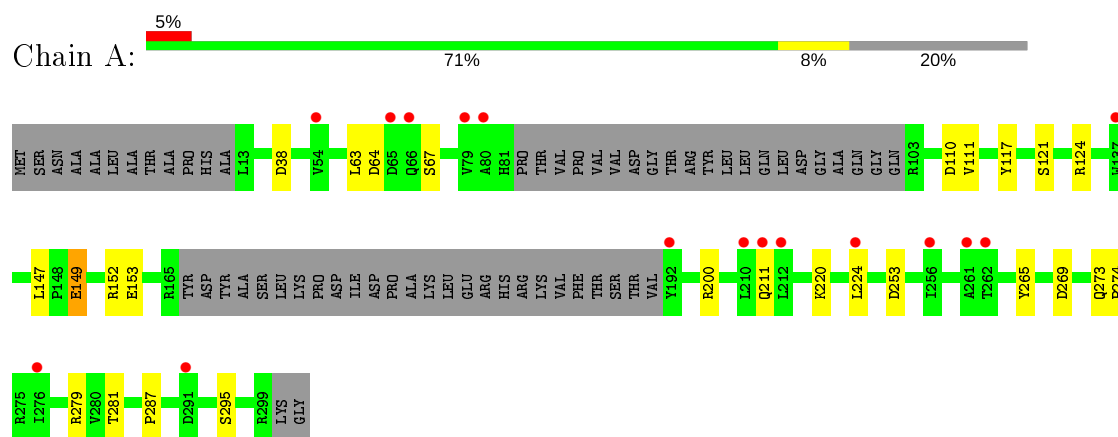
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	119	Total	O	0	0
			119	119		
5	D	126	Total	O	0	0
			126	126		

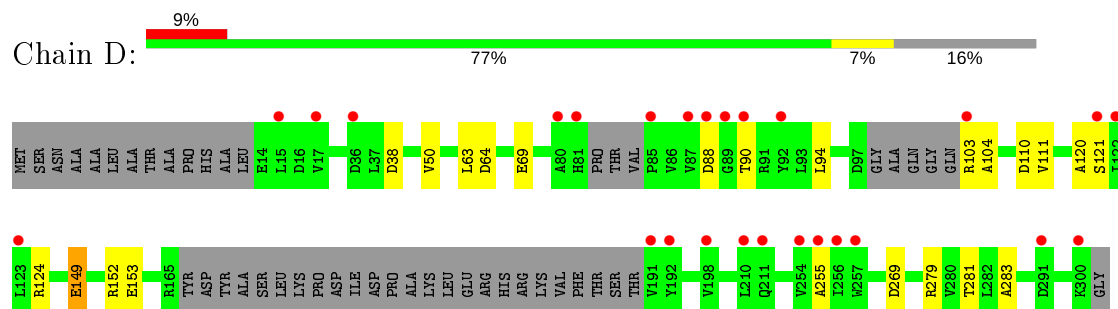
### 3 Residue-property plots [i](#)

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: PUTATIVE ALKYL SULFATASE ATSK



#### • Molecule 1: PUTATIVE ALKYL SULFATASE ATSK



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.03Å 141.89Å 161.72Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 2.06 25.15 – 2.06	Depositor EDS
% Data completeness (in resolution range)	99.9 (25.00-2.06) 99.2 (25.15-2.06)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.53 (at 2.06Å)	Xtriage
Refinement program	REFMAC 5.1.19	Depositor
R, $R_{free}$	0.199 , 0.221 0.200 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.9	Xtriage
Anisotropy	0.561	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 64.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4116	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.29% 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 ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: C26, AKG, FE2

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	0.90	0/1913	0.92	5/2609 (0.2%)
1	D	0.95	1/2008 (0.0%)	0.94	4/2740 (0.1%)
All	All	0.93	1/3921 (0.0%)	0.93	9/5349 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	255	ALA	CA-CB	7.27	1.67	1.52

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	269	ASP	CB-CG-OD2	7.18	124.76	118.30
1	A	110	ASP	CB-CG-OD2	6.87	124.48	118.30
1	D	38	ASP	CB-CG-OD2	6.71	124.34	118.30
1	A	38	ASP	CB-CG-OD2	6.57	124.21	118.30
1	A	269	ASP	CB-CG-OD2	6.16	123.84	118.30
1	A	200	ARG	NE-CZ-NH2	-6.00	117.30	120.30
1	D	88	ASP	CB-CG-OD2	5.61	123.34	118.30
1	D	103	ARG	NE-CZ-NH1	-5.39	117.61	120.30
1	A	253	ASP	CB-CG-OD2	5.20	122.98	118.30

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	1865	0	1805	7	0
1	D	1958	0	1882	11	0
2	A	13	0	18	0	0
2	D	13	0	18	3	0
3	A	10	0	4	0	0
3	D	10	0	4	0	0
4	A	1	0	0	0	0
4	D	1	0	0	0	0
5	A	119	0	0	0	0
5	D	126	0	0	4	0
All	All	4116	0	3731	18	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 2.

All (18) 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:D:283:ALA:HA	5:D:2119:HOH:O	1.96	0.64
1:A:63:LEU:HD23	1:A:124:ARG:HD3	1.78	0.64
1:A:147:LEU:HB2	1:A:152:ARG:HG2	1.91	0.53
1:A:149:GLU:O	1:A:153:GLU:HG3	2.10	0.52
1:D:149:GLU:HG3	1:D:152:ARG:NH2	2.27	0.50
1:D:63:LEU:HD23	1:D:124:ARG:HD3	1.94	0.49
1:D:104:ALA:HB2	2:D:1301:C26:H7C2	1.95	0.48
1:D:69:GLU:OE2	1:D:90:THR:HB	2.15	0.47
1:D:104:ALA:HA	2:D:1301:C26:H6C3	1.97	0.46
1:A:287:PRO:HB2	1:A:295:SER:OG	2.15	0.46
1:D:110:ASP:HA	2:D:1301:C26:O4	2.17	0.45
1:D:50:VAL:HG23	5:D:2014:HOH:O	2.16	0.45
1:A:117:TYR:OH	1:A:211[A]:GLN:NE2	2.50	0.45
1:A:64:ASP:H	1:A:67:SER:HB3	1.82	0.44
1:D:149:GLU:O	1:D:153:GLU:HG3	2.21	0.41
1:D:64:ASP:HB2	5:D:2022:HOH:O	2.19	0.41
1:A:273:GLN:HA	1:A:274:PRO:HD3	1.97	0.40
1:D:120:ALA:HB2	5:D:2035:HOH:O	2.20	0.40

There are no symmetry-related clashes.



## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	235/301 (78%)	228 (97%)	7 (3%)	0	100	100
1	D	247/301 (82%)	236 (96%)	11 (4%)	0	100	100
All	All	482/602 (80%)	464 (96%)	18 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	189/244 (78%)	181 (96%)	8 (4%)	30	23
1	D	197/244 (81%)	191 (97%)	6 (3%)	41	35
All	All	386/488 (79%)	372 (96%)	14 (4%)	35	28

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	111	VAL
1	A	121	SER
1	A	149	GLU
1	A	220	LYS
1	A	224	LEU
1	A	265	TYR
1	A	279	ARG
1	A	281	THR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	94	LEU
1	D	111	VAL
1	D	121	SER
1	D	149	GLU
1	D	279	ARG
1	D	281	THR

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	293	GLN
1	D	293	GLN

### 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 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 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	$\# Z  > 2$
2	C26	A	1300	-	12,12,12	1.16	1 (8%)	13,15,15	0.93	1 (7%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	AKG	D	1300	4	3,9,9	0.34	0	4,11,11	2.88	3 (75%)
2	C26	D	1301	-	12,12,12	0.77	1 (8%)	13,15,15	1.46	2 (15%)
3	AKG	A	1301	4	3,9,9	0.66	0	4,11,11	3.11	2 (50%)

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
2	C26	A	1300	-	-	5/12/12/12	-
3	AKG	D	1300	4	-	0/3/9/9	-
2	C26	D	1301	-	-	10/12/12/12	-
3	AKG	A	1301	4	-	0/3/9/9	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1300	C26	O1-S1	3.70	1.66	1.56
2	D	1301	C26	O1-S1	2.18	1.62	1.56

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1301	AKG	C4-C3-C2	-4.41	103.62	113.14
3	A	1301	AKG	C3-C4-C5	-4.26	105.52	112.67
3	D	1300	AKG	C4-C3-C2	-3.99	104.54	113.14
2	D	1301	C26	O1-C1-C2	-3.92	100.16	108.64
3	D	1300	AKG	O5-C2-C3	3.11	125.68	120.38
3	D	1300	AKG	C3-C4-C5	-2.70	108.14	112.67
2	D	1301	C26	C7-C2-C1	-2.60	103.69	111.23
2	A	1300	C26	O1-C1-C2	2.58	114.21	108.64

There are no chirality outliers.

All (15) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	1301	C26	O1-C1-C2-C7
2	D	1301	C26	O1-C1-C2-C3
2	A	1300	C26	C1-C2-C7-C8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms
2	D	1301	C26	C1-O1-S1-O3
2	D	1301	C26	C7-C2-C3-C4
2	D	1301	C26	C1-O1-S1-O2
2	D	1301	C26	C3-C4-C5-C6
2	A	1300	C26	C3-C4-C5-C6
2	D	1301	C26	C1-C2-C7-C8
2	D	1301	C26	C3-C2-C7-C8
2	D	1301	C26	C1-C2-C3-C4
2	A	1300	C26	C3-C2-C7-C8
2	D	1301	C26	C1-O1-S1-O4
2	A	1300	C26	C7-C2-C3-C4
2	A	1300	C26	C1-C2-C3-C4

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	1301	C26	3	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	240/301 (79%)	0.29	16 (6%) 17 18	32, 47, 68, 83	1 (0%)
1	D	254/301 (84%)	0.35	26 (10%) 6 7	33, 48, 74, 96	1 (0%)
All	All	494/602 (82%)	0.32	42 (8%) 10 11	32, 48, 71, 96	2 (0%)

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	80	ALA	9.9
1	D	192	TYR	5.0
1	D	191	VAL	3.8
1	D	122	ILE	3.5
1	D	256	ILE	3.5
1	D	123	LEU	3.4
1	D	85	PRO	3.4
1	A	210	LEU	3.3
1	A	192	TYR	3.2
1	D	88	ASP	3.1
1	D	210	LEU	3.0
1	A	224	LEU	2.8
1	D	257	TRP	2.7
1	D	90	THR	2.7
1	A	79	VAL	2.6
1	A	66	GLN	2.6
1	D	81	HIS	2.5
1	A	211[A]	GLN	2.5
1	D	291	ASP	2.5
1	D	89	GLY	2.4
1	A	256	ILE	2.3
1	D	80	ALA	2.2
1	D	36	ASP	2.2
1	A	65	ASP	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	276	ILE	2.2
1	A	54	VAL	2.2
1	A	291	ASP	2.2
1	A	137	TRP	2.1
1	D	198	VAL	2.1
1	A	262	THR	2.1
1	D	300	LYS	2.1
1	D	255	ALA	2.1
1	D	17	VAL	2.1
1	D	87	VAL	2.1
1	D	15	LEU	2.1
1	D	254	VAL	2.1
1	D	121	SER	2.0
1	A	212	LEU	2.0
1	D	103	ARG	2.0
1	D	211[A]	GLN	2.0
1	D	92	TYR	2.0
1	A	261	ALA	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
2	C26	D	1301	13/13	0.77	0.29	63,65,67,67	13
2	C26	A	1300	13/13	0.78	0.26	62,76,79,80	13
3	AKG	A	1301	10/10	0.94	0.10	47,54,57,64	0
3	AKG	D	1300	10/10	0.95	0.15	49,55,69,71	0
4	FE2	D	1302	1/1	0.95	0.15	41,41,41,41	1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	FE2	A	1302	1/1	0.99	0.05	46,46,46,46	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.