



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 2, 2021 – 10:07 PM EDT

PDB ID : 3IAE  
Title : Structure of benzaldehyde lyase A28S mutant with benzoylphosphonate  
Authors : Brandt, G.S.; Petsko, G.A.; Ringe, D.; McLeish, M.J.  
Deposited on : 2009-07-13  
Resolution : 2.30 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	:	<b>FAILED</b>
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	<b>FAILED</b>
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)	:	2.23.2

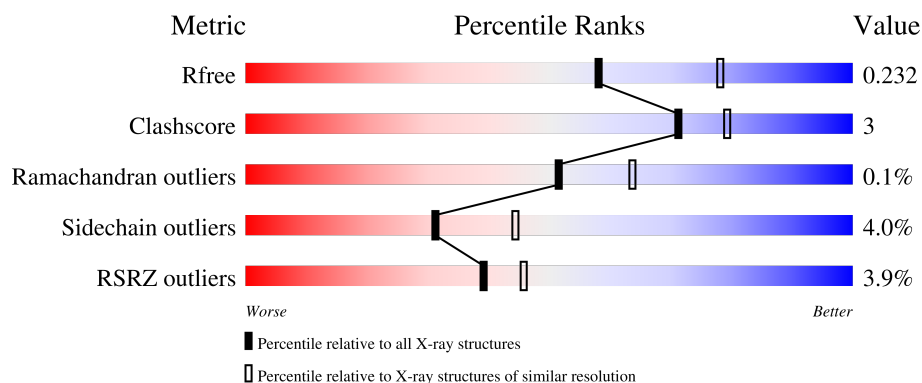
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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))
$R_{free}$	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	570	<div> <div>6%</div> <div> <div></div> <div>88%</div> <div>8%</div> <div>..</div> </div> </div>
1	B	570	<div> <div>2%</div> <div> <div></div> <div>87%</div> <div>9%</div> <div>..</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8450 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.

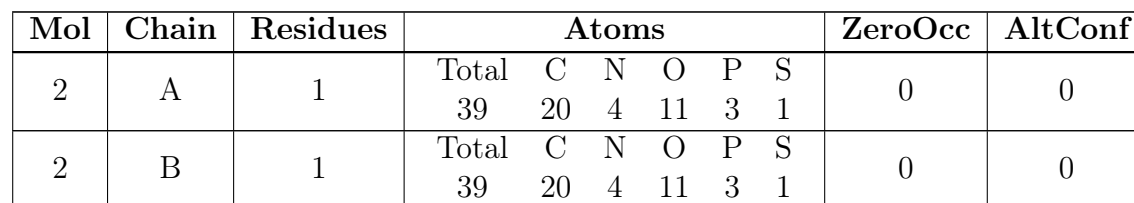
- Molecule 1 is a protein called Benzaldehyde lyase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	554	Total	C	N	O	S	0	0	0
			4080	2576	723	765	16			
1	B	554	Total	C	N	O	S	0	0	0
			4080	2576	723	765	16			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	28	SER	ALA	engineered mutation	UNP Q9F4L3
A	563	GLY	-	expression tag	UNP Q9F4L3
A	564	SER	-	expression tag	UNP Q9F4L3
A	565	HIS	-	expression tag	UNP Q9F4L3
A	566	HIS	-	expression tag	UNP Q9F4L3
A	567	HIS	-	expression tag	UNP Q9F4L3
A	568	HIS	-	expression tag	UNP Q9F4L3
A	569	HIS	-	expression tag	UNP Q9F4L3
A	570	HIS	-	expression tag	UNP Q9F4L3
B	28	SER	ALA	engineered mutation	UNP Q9F4L3
B	563	GLY	-	expression tag	UNP Q9F4L3
B	564	SER	-	expression tag	UNP Q9F4L3
B	565	HIS	-	expression tag	UNP Q9F4L3
B	566	HIS	-	expression tag	UNP Q9F4L3
B	567	HIS	-	expression tag	UNP Q9F4L3
B	568	HIS	-	expression tag	UNP Q9F4L3
B	569	HIS	-	expression tag	UNP Q9F4L3
B	570	HIS	-	expression tag	UNP Q9F4L3

- Molecule 2 is 3-[(4-amino-2-methylpyrimidin-5-yl)methyl]-2-[(S)-hydroxy[(R)-hydroxy(methoxy)phosphoryl]phenylmethyl]-5-(2-[(R)-hydroxy(phosphonooxy)phosphoryl]oxy)ethyl)-4-methyl-1,3-thiazol-3-ium (three-letter code: D7K) (formula: C<sub>20</sub>H<sub>28</sub>N<sub>4</sub>O<sub>11</sub>P<sub>3</sub>S).



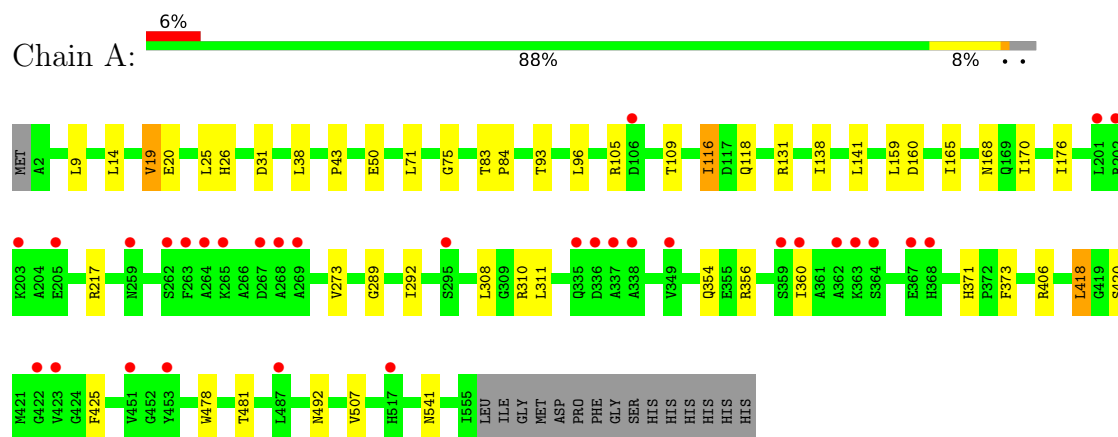
- | Mol | Chain | Residues | Atoms           | ZeroOcc | AltConf |
|-----|-------|----------|-----------------|---------|---------|
| 3   | A     | 1        | Total Ca<br>1 1 | 0       | 0       |
| 3   | B     | 1        | Total Ca<br>1 1 | 0       | 0       |

- | Mol | Chain | Residues | Atoms              | ZeroOcc | AltConf |
|-----|-------|----------|--------------------|---------|---------|
| 4   | A     | 107      | Total O<br>107 107 | 0       | 0       |
| 4   | B     | 103      | Total O<br>103 103 | 0       | 0       |

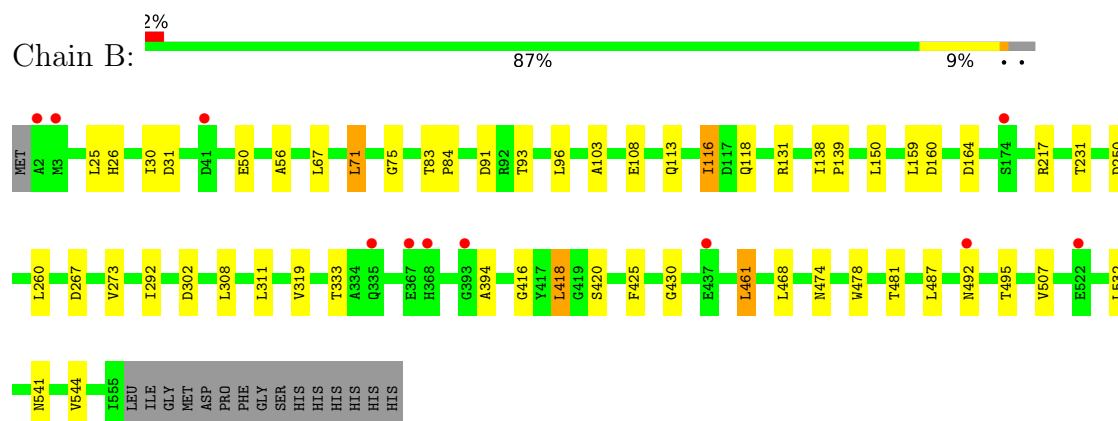
### 3 Residue-property plots [i](#)

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: Benzaldehyde lyase



#### • Molecule 1: Benzaldehyde lyase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	150.86Å 150.86Å 97.24Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.55 – 2.30 40.87 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.6 (45.55-2.30) 99.6 (40.87-2.30)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.09 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.204 , 0.234 0.203 , 0.232	Depositor DCC
$R_{free}$ test set	5711 reflections (10.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.1	Xtriage
Anisotropy	0.005	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 35.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.022 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	8450	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.48% 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: D7K, CA

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	0.31	0/4163	0.50	0/5678
1	B	0.31	0/4163	0.50	0/5678
All	All	0.31	0/8326	0.50	0/11356

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

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	4080	0	4069	30	0
1	B	4080	0	4069	32	0
2	A	39	0	24	3	0
2	B	39	0	24	4	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	107	0	0	0	0
4	B	103	0	0	0	0
All	All	8450	0	8186	57	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 3.

All (57) 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:9:LEU:HD11	1:A:170:ILE:HD11	1.74	0.69
1:A:371:HIS:HD2	1:A:373:PHE:H	1.43	0.64
1:B:418:LEU:HD13	1:B:420:SER:HB2	1.82	0.62
1:B:474:ASN:HD22	1:B:541:ASN:HD21	1.45	0.62
1:B:273:VAL:HG11	1:B:292:ILE:HG23	1.84	0.60
1:B:231:THR:HG23	1:B:333:THR:HG21	1.84	0.60
1:A:116:ILE:HD13	1:A:118:GLN:HE21	1.67	0.59
1:B:418:LEU:CD1	1:B:420:SER:HB2	2.33	0.57
1:B:93:THR:HG22	1:B:217:ARG:NH2	2.20	0.57
1:B:93:THR:HG22	1:B:217:ARG:HH21	1.71	0.55
1:A:93:THR:HG22	1:A:217:ARG:HH21	1.72	0.54
2:A:571:D7K:N1,	1:B:50:GLU:OE2	2.41	0.53
1:B:103:ALA:HB1	1:B:164:ASP:HB2	1.89	0.52
1:A:481:THR:HB	1:B:26:HIS:CD2	2.45	0.52
1:B:478:TRP:HB3	2:B:571:D7K:H05A	1.92	0.52
1:B:75:GLY:H	1:B:118:GLN:HE22	1.58	0.51
1:A:26:HIS:CD2	1:B:481:THR:HB	2.45	0.51
1:A:354:GLN:HE22	1:A:406:ARG:HH22	1.58	0.51
1:A:93:THR:HG22	1:A:217:ARG:NH2	2.26	0.51
1:A:14:LEU:O	1:A:19:VAL:HG13	2.12	0.50
1:A:507:VAL:HG13	1:B:507:VAL:HG13	1.94	0.50
1:A:109:THR:HA	1:B:311:LEU:HD21	1.94	0.50
1:A:26:HIS:HE1	1:A:31:ASP:OD1	1.94	0.49
1:B:116:ILE:H	1:B:116:ILE:HD13	1.78	0.48
1:A:273:VAL:HG11	1:A:292:ILE:HG23	1.96	0.48
1:B:67:LEU:HB2	1:B:150:LEU:HD21	1.96	0.47
1:B:75:GLY:H	1:B:118:GLN:NE2	2.12	0.47
1:B:30:ILE:HD12	1:B:71:LEU:HD13	1.97	0.47
1:A:50:GLU:OE2	2:B:571:D7K:N1,	2.48	0.46
1:A:83:THR:HB	1:A:84:PRO:HD3	1.96	0.46
2:B:571:D7K:O11	2:B:571:D7K:H3	2.16	0.46
1:B:302:ASP:O	1:B:319:VAL:HA	2.16	0.46
1:B:91:ASP:OD1	1:B:416:GLY:HA3	2.16	0.45
1:A:75:GLY:H	1:A:118:GLN:NE2	2.14	0.45
1:A:478:TRP:HB3	2:A:571:D7K:H05A	1.97	0.45
1:B:26:HIS:HE1	1:B:31:ASP:OD1	2.01	0.44
1:B:131:ARG:HA	1:B:160:ASP:HB3	1.99	0.44
1:A:310:ARG:NH1	1:B:108:GLU:HG3	2.33	0.44
1:A:138:ILE:HD11	1:A:165:ILE:HG22	2.00	0.43
1:A:310:ARG:HG3	1:A:311:LEU:HG	2.01	0.43

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:418:LEU:HD13	1:A:420:SER:HB2	2.00	0.43
1:A:131:ARG:HA	1:A:160:ASP:HB3	2.01	0.43
1:B:138:ILE:HB	1:B:139:PRO:HD3	2.01	0.42
1:A:26:HIS:HD2	1:B:481:THR:HB	1.84	0.42
1:A:116:ILE:HD13	1:A:118:GLN:NE2	2.34	0.42
1:A:273:VAL:HG21	1:A:292:ILE:HD12	2.02	0.41
1:A:371:HIS:CD2	1:A:373:PHE:H	2.30	0.41
2:A:571:D7K:O7	1:B:113:GLN:NE2	2.54	0.41
1:A:75:GLY:H	1:A:118:GLN:HE22	1.68	0.41
1:B:461:LEU:HD21	1:B:468:LEU:HD23	2.02	0.41
1:B:478:TRP:CE3	2:B:571:D7K:HM4B	2.56	0.41
1:A:289:GLY:HA2	1:A:292:ILE:O	2.21	0.40
1:B:273:VAL:HG21	1:B:292:ILE:HD12	2.03	0.40
1:A:20:GLU:O	1:A:43:PRO:HD2	2.22	0.40
1:B:83:THR:HB	1:B:84:PRO:HD3	2.04	0.40
1:B:56:ALA:O	1:B:430:GLY:HA3	2.21	0.40
1:A:356:ARG:O	1:A:360:ILE:HG12	2.21	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	552/570 (97%)	535 (97%)	17 (3%)	0	100	100
1	B	552/570 (97%)	537 (97%)	14 (2%)	1 (0%)	47	58
All	All	1104/1140 (97%)	1072 (97%)	31 (3%)	1 (0%)	51	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	394	ALA

### 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	412/427 (96%)	396 (96%)	16 (4%)	32	46
1	B	412/427 (96%)	395 (96%)	17 (4%)	30	43
All	All	824/854 (96%)	791 (96%)	33 (4%)	31	44

All (33) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	19	VAL
1	A	25	LEU
1	A	38	LEU
1	A	71	LEU
1	A	96	LEU
1	A	105	ARG
1	A	116	ILE
1	A	141	LEU
1	A	159	LEU
1	A	168	ASN
1	A	176	ILE
1	A	308	LEU
1	A	418	LEU
1	A	425	PHE
1	A	492	ASN
1	A	541	ASN
1	B	25	LEU
1	B	71	LEU
1	B	96	LEU
1	B	116	ILE
1	B	159	LEU
1	B	250	ASP
1	B	260	LEU
1	B	267	ASP
1	B	308	LEU
1	B	418	LEU
1	B	425	PHE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	461	LEU
1	B	487	LEU
1	B	492	ASN
1	B	495	THR
1	B	532	LEU
1	B	544	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (28) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	26	HIS
1	A	87	ASN
1	A	113	GLN
1	A	118	GLN
1	A	144	GLN
1	A	335	GLN
1	A	354	GLN
1	A	371	HIS
1	A	374	HIS
1	A	415	HIS
1	A	476	GLN
1	A	492	ASN
1	A	530	GLN
1	A	534	HIS
1	A	541	ASN
1	B	26	HIS
1	B	87	ASN
1	B	113	GLN
1	B	118	GLN
1	B	168	ASN
1	B	258	GLN
1	B	259	ASN
1	B	335	GLN
1	B	354	GLN
1	B	374	HIS
1	B	415	HIS
1	B	530	GLN
1	B	541	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

Mogul failed to run properly - this section is therefore empty.

## 5.5 Carbohydrates ⓘ

Mogul failed to run properly - this section is therefore empty.

## 5.6 Ligand geometry ⓘ

Mogul failed to run properly - this section is therefore empty.

## 5.7 Other polymers ⓘ

Mogul failed to run properly - this section is therefore empty.

## 5.8 Polymer linkage issues ⓘ

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	554/570 (97%)	0.19	32 (5%)	23 29	29, 47, 79, 101	0
1	B	554/570 (97%)	0.01	11 (1%)	65 71	31, 46, 63, 80	0
All	All	1108/1140 (97%)	0.10	43 (3%)	39 46	29, 46, 73, 101	0

All (43) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	368	HIS	7.1
1	A	337	ALA	5.7
1	A	367	GLU	5.5
1	A	265	LYS	4.5
1	B	335	GLN	4.0
1	A	267	ASP	3.9
1	B	368	HIS	3.7
1	A	268	ALA	3.5
1	A	359	SER	3.5
1	A	269	ALA	3.4
1	A	262	SER	3.4
1	B	393	GLY	3.2
1	A	335	GLN	3.2
1	B	174	SER	3.2
1	A	338	ALA	3.1
1	A	203	LYS	3.1
1	A	259	ASN	3.0
1	A	363	LYS	2.9
1	B	3	MET	2.8
1	A	451	VAL	2.7
1	A	336	ASP	2.7
1	A	349	VAL	2.6
1	A	364	SER	2.6
1	A	360	ILE	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	487	LEU	2.5
1	B	522	GLU	2.4
1	A	517	HIS	2.4
1	A	201	LEU	2.4
1	A	423	VAL	2.4
1	A	205	GLU	2.3
1	A	453	TYR	2.3
1	B	367	GLU	2.3
1	A	422	GLY	2.3
1	B	2	ALA	2.3
1	A	264	ALA	2.2
1	A	202	ARG	2.1
1	A	295	SER	2.1
1	A	362	ALA	2.1
1	B	492	ASN	2.1
1	B	41	ASP	2.1
1	B	437	GLU	2.1
1	A	263	PHE	2.0
1	A	106	ASP	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	CA	A	572	1/1	0.75	0.18	78,78,78,78	0
3	CA	B	572	1/1	0.80	0.14	75,75,75,75	0
2	D7K	A	571	39/39	0.96	0.16	35,43,57,58	1
2	D7K	B	571	39/39	0.96	0.14	23,34,44,44	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.