



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 10, 2018 – 01:39 PM EST

PDB ID : 5VOP  
Title : Methionine synthase folate-binding domain from *Thermus thermophilus* HB8 native  
Authors : Koutmos, M.; Yamada, K.  
Deposited on : 2017-05-03  
Resolution : 2.10 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

<http://wwpdb.org/validation/2016/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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

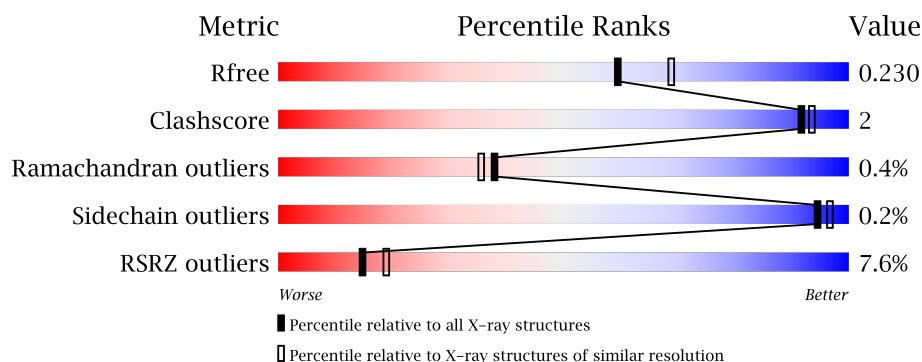
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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))
$R_{free}$	100719	4243 (2.10-2.10)
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (2.10-2.10)

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. 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	296	
1	B	296	

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
3	GOL	A	3002	-	-	-	X
5	NA	A	3005	-	-	-	X
5	NA	B	3002	-	-	-	X

## 2 Entry composition [i](#)

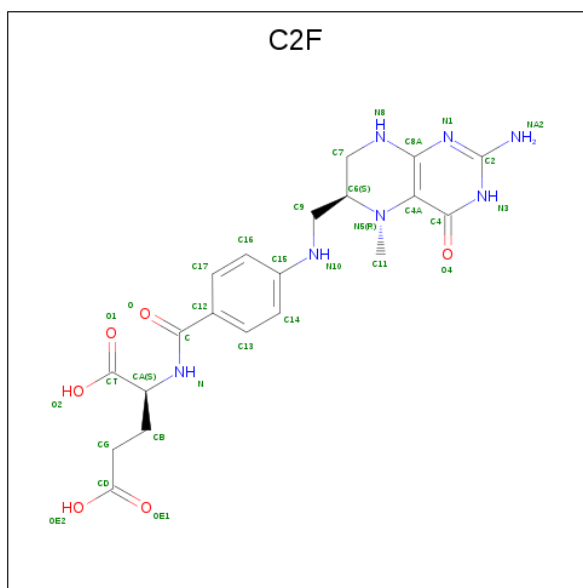
There are 6 unique types of molecules in this entry. The entry contains 9262 atoms, of which 4601 are hydrogens and 0 are deuteriums.

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 5-methyltetrahydrofolate homocysteine S-methyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	282	Total	C	H	N	O	S	0	0	0
			4480	1412	2263	390	408	7			
1	B	282	Total	C	H	N	O	S	0	0	0
			4480	1412	2263	390	408	7			

- Molecule 2 is 5-METHYL-5,6,7,8-TETRAHYDROFOLIC ACID (three-letter code: C2F) (formula:  $C_{20}H_{25}N_7O_6$ ).



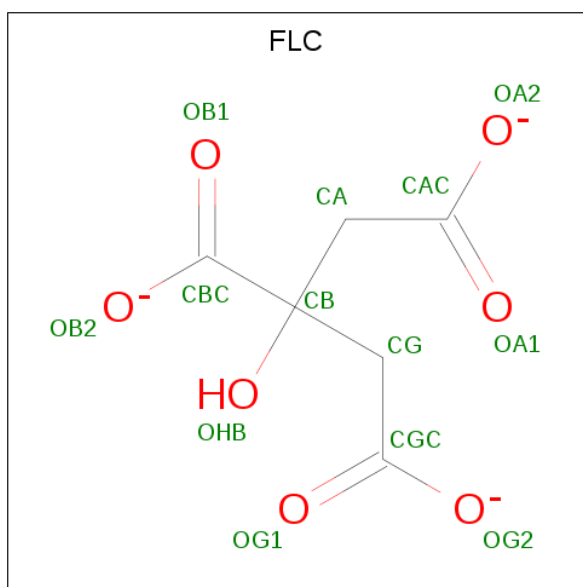
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	0	0
			56	20	23	7	6		
2	B	1	Total	C	H	N	O	0	0
			56	20	23	7	6		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			14	3	8	3		
3	A	1	Total	C	H	O	0	0
			14	3	8	3		
3	A	1	Total	C	H	O	0	0
			14	3	8	3		

- Molecule 4 is CITRATE ANION (three-letter code: FLC) (formula:  $C_6H_5O_7$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	H	O	0	0
			18	6	5	7		

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Na	0	0
			1	1		
5	A	1	Total	Na	0	0
			1	1		

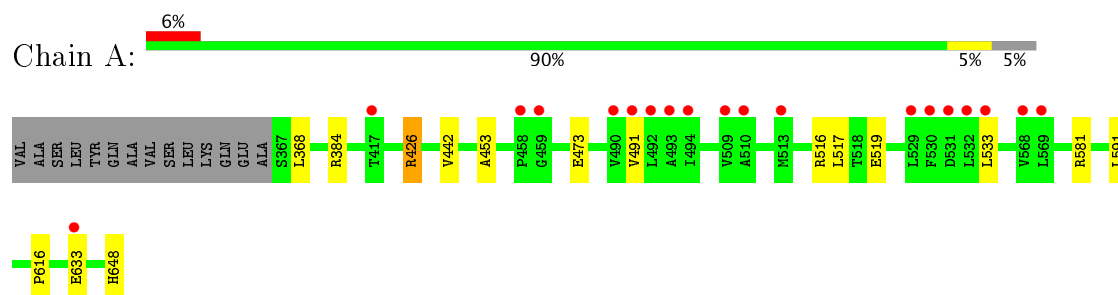
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	81	Total	O	0	0
			81	81		
6	B	47	Total	O	0	0
			47	47		

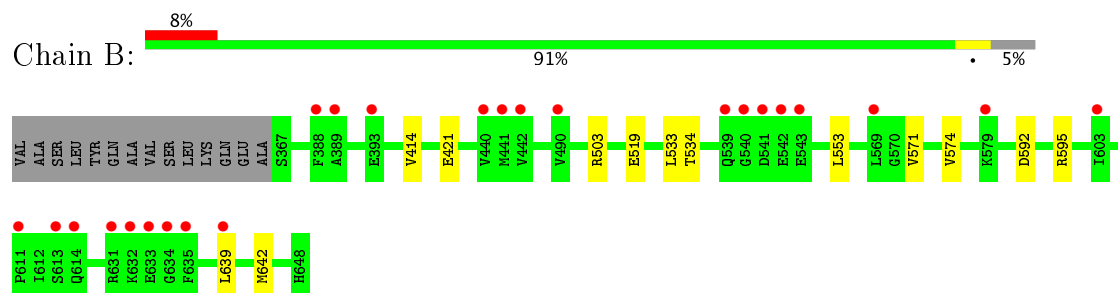
### 3 Residue-property plots [i](#)

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: 5-methyltetrahydrofolate homocysteine S-methyltransferase



- Molecule 1: 5-methyltetrahydrofolate homocysteine S-methyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	111.48 Å   111.48 Å   242.68 Å 90.00°   90.00°   120.00°	Depositor
Resolution (Å)	29.44 – 2.10 29.44 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.9 (29.44-2.10) 99.9 (29.44-2.10)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.14 (at 2.10 Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.194   ,   0.222 0.203   ,   0.230	Depositor DCC
$R_{free}$ test set	2667 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.6	Xtriage
Anisotropy	0.012	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 43.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	9262	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.73% 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, FLC, C2F, NA

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	0.50	0/2255	0.75	4/3048 (0.1%)
1	B	0.47	0/2255	0.70	0/3048
All	All	0.49	0/4510	0.73	4/6096 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	426	ARG	NE-CZ-NH2	5.86	123.23	120.30
1	A	426	ARG	NE-CZ-NH1	-5.76	117.42	120.30
1	A	616	PRO	C-N-CA	5.64	135.80	121.70
1	A	384	ARG	NE-CZ-NH1	5.23	122.92	120.30

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	2217	2263	2263	9	0
1	B	2217	2263	2263	8	0
2	A	33	23	23	0	0
2	B	33	23	23	0	0
3	A	18	24	24	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	13	5	5	1	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
6	A	81	0	0	1	0
6	B	47	0	0	0	0
All	All	4661	4601	4601	17	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 (17) 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:581:ARG:NH2	1:A:648:HIS:O	2.18	0.77
1:A:442:VAL:HG21	1:A:453:ALA:CB	2.35	0.56
1:A:426:ARG:NH1	6:A:3101:HOH:O	2.39	0.54
1:A:442:VAL:HG21	1:A:453:ALA:HB1	1.90	0.53
1:A:368:LEU:HD13	1:A:591:LEU:HD11	1.93	0.50
1:B:534:THR:HG23	1:B:574:VAL:HG13	1.95	0.48
1:B:592:ASP:OD1	1:B:595:ARG:NH1	2.47	0.48
1:A:491:VAL:HG11	1:A:517:LEU:HD11	1.96	0.47
1:B:571:VAL:O	1:B:574:VAL:HG22	2.15	0.47
1:B:414:VAL:HG23	1:B:421:GLU:HG3	1.97	0.46
3:A:3003:GOL:H12	1:B:519:GLU:HB2	1.97	0.46
1:A:516:ARG:O	1:A:519:GLU:O	2.34	0.45
1:A:368:LEU:CD1	1:A:591:LEU:HD11	2.47	0.45
1:B:503:ARG:NH1	1:B:553:LEU:HD13	2.33	0.44
1:A:473:GLU:HB3	4:A:3004:FLC:HA2	2.00	0.43
1:B:534:THR:CG2	1:B:574:VAL:HG13	2.50	0.42
1:B:639:LEU:HD23	1:B:642:MET:CE	2.50	0.42

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	280/296 (95%)	274 (98%)	5 (2%)	1 (0%)	38	35
1	B	280/296 (95%)	274 (98%)	5 (2%)	1 (0%)	38	35
All	All	560/592 (95%)	548 (98%)	10 (2%)	2 (0%)	38	35

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	533	LEU
1	B	533	LEU

### 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	231/242 (96%)	230 (100%)	1 (0%)	93	95
1	B	231/242 (96%)	231 (100%)	0	100	100
All	All	462/484 (96%)	461 (100%)	1 (0%)	94	97

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	633	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

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 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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	C2F	A	3001	-	25,35,35	1.74	4 (16%)	28,49,49	2.60	9 (32%)
3	GOL	A	3002	-	5,5,5	0.34	0	5,5,5	0.41	0
3	GOL	A	3003	-	5,5,5	0.32	0	5,5,5	0.38	0
4	FLC	A	3004	-	3,12,12	1.44	1 (33%)	3,17,17	3.78	2 (66%)
3	GOL	A	3006	-	5,5,5	0.75	0	5,5,5	0.86	0
2	C2F	B	3001	-	25,35,35	1.73	4 (16%)	28,49,49	1.96	7 (25%)

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	C2F	A	3001	-	-	0/16/35/35	0/3/3/3
3	GOL	A	3002	-	-	0/4/4/4	0/0/0/0
3	GOL	A	3003	-	-	0/4/4/4	0/0/0/0
4	FLC	A	3004	-	-	0/6/16/16	0/0/0/0
3	GOL	A	3006	-	-	0/4/4/4	0/0/0/0
2	C2F	B	3001	-	-	0/16/35/35	0/3/3/3

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	3001	C2F	C4-N3	-2.69	1.33	1.36
2	B	3001	C2F	C6-N5	-2.13	1.45	1.47
4	A	3004	FLC	CG-CB	-2.00	1.51	1.54
2	A	3001	C2F	C2-NA2	2.11	1.35	1.32
2	A	3001	C2F	C2-N1	2.27	1.36	1.33
2	B	3001	C2F	C4A-C8A	3.16	1.47	1.41
2	A	3001	C2F	C4A-C8A	3.76	1.48	1.41
2	A	3001	C2F	C4-C4A	6.24	1.48	1.41
2	B	3001	C2F	C4-C4A	6.38	1.48	1.41

All (18) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	3001	C2F	C7-C6-N5	-7.52	100.44	108.79
2	A	3001	C2F	CB-CA-CT	-6.42	102.95	112.28
4	A	3004	FLC	CB-CG-CGC	-5.48	106.39	114.95
2	B	3001	C2F	C7-C6-N5	-5.41	102.78	108.79
2	A	3001	C2F	O-C-C12	-2.69	116.16	120.94
2	B	3001	C2F	C6-C9-N10	-2.23	105.91	111.89
2	A	3001	C2F	C9-C6-N5	2.03	115.14	110.90
2	B	3001	C2F	CG-CB-CA	2.13	117.44	113.19
2	B	3001	C2F	C12-C-N	2.25	121.12	116.97
2	A	3001	C2F	C12-C-N	2.84	122.22	116.97
2	B	3001	C2F	C4-C4A-N5	2.99	127.72	119.21
2	A	3001	C2F	C4-C4A-N5	3.05	127.89	119.21
2	A	3001	C2F	C11-N5-C4A	3.23	117.78	113.13
2	A	3001	C2F	CB-CA-N	3.24	115.14	110.22
4	A	3004	FLC	CB-CA-CAC	3.37	120.22	114.95
2	B	3001	C2F	C11-N5-C4A	3.86	118.68	113.13
2	B	3001	C2F	C4-C4A-C8A	4.36	117.90	114.43
2	A	3001	C2F	C4-C4A-C8A	4.84	118.28	114.43

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	3003	GOL	1	0
4	A	3004	FLC	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	282/296 (95%)	0.18	19 (6%)	19 23	43, 53, 74, 95	0
1	B	282/296 (95%)	0.62	24 (8%)	11 15	47, 61, 102, 134	0
All	All	564/592 (95%)	0.40	43 (7%)	15 19	43, 56, 93, 134	0

All (43) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	635	PHE	5.6
1	B	633	GLU	5.2
1	B	632	LYS	4.6
1	B	634	GLY	4.5
1	B	542	GLU	4.3
1	A	509	VAL	3.8
1	B	631	ARG	3.5
1	B	543	GLU	3.5
1	A	510	ALA	3.4
1	A	493	ALA	3.4
1	A	458	PRO	3.3
1	B	540	GLY	3.2
1	A	491	VAL	3.0
1	A	530	PHE	3.0
1	B	388	PHE	2.9
1	B	539	GLN	2.9
1	A	492	LEU	2.8
1	B	393	GLU	2.8
1	A	490	VAL	2.7
1	B	639	LEU	2.7
1	B	611	PRO	2.6
1	A	459	GLY	2.6
1	B	579	LYS	2.6
1	B	541	ASP	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	440	VAL	2.5
1	A	417	THR	2.4
1	A	513	MET	2.4
1	A	633	GLU	2.4
1	A	531	ASP	2.4
1	B	441	MET	2.3
1	B	603	ILE	2.3
1	A	529	LEU	2.2
1	B	569	LEU	2.2
1	B	613	SER	2.2
1	A	532	LEU	2.2
1	B	442	VAL	2.2
1	A	494	ILE	2.2
1	B	614	GLN	2.1
1	B	389	ALA	2.1
1	A	568	VAL	2.1
1	B	490	VAL	2.1
1	A	533	LEU	2.1
1	A	569	LEU	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	NA	B	3002	1/1	0.90	0.24	6.60	82,82,82,82	0
3	GOL	A	3002	6/6	0.62	0.29	4.86	78,90,95,95	0

*Continued on next page...*



*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	NA	A	3005	1/1	0.90	0.29	2.97	68,68,68,68	0
3	GOL	A	3006	6/6	0.70	0.21	1.81	55,67,69,69	0
2	C2F	B	3001	33/33	0.82	0.23	0.71	55,64,78,82	56
2	C2F	A	3001	33/33	0.81	0.17	0.13	42,59,69,72	56
3	GOL	A	3003	6/6	0.94	0.15	-0.11	54,60,65,65	0
4	FLC	A	3004	13/13	0.95	0.15	-	40,48,56,63	18

## 6.5 Other polymers [i](#)

There are no such residues in this entry.