



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 6, 2024 – 11:07 pm GMT

PDB ID : 5OW3  
Title : Crystal structure of a C-terminally truncated trimeric ectodomain of the Arabidopsis thaliana gamete fusion protein HAP2  
Authors : Fedry, J.; Legrand, P.; Rey, F.A.; Krey, T.  
Deposited on : 2017-08-30  
Resolution : 2.75 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

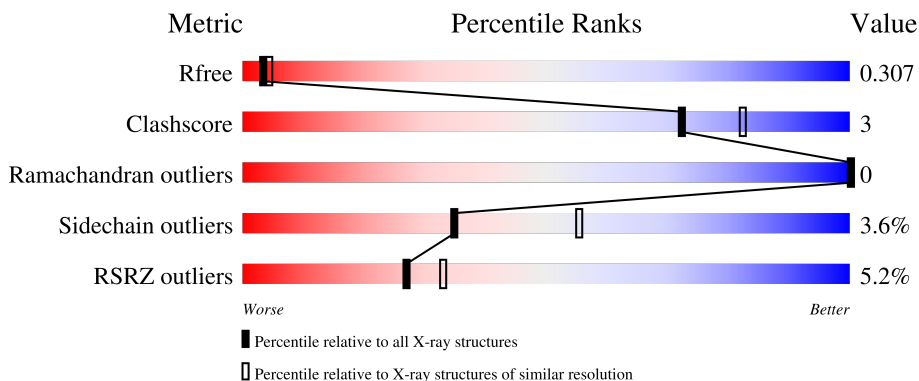
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.75 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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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	C	524	

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 3619 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 Protein HAPLESS 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	C	446	3511	2229	599	663	20	0	0	0

There are 57 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	7	MET	-	initiating methionine	UNP F4JP36
C	8	LYS	-	expression tag	UNP F4JP36
C	9	LEU	-	expression tag	UNP F4JP36
C	10	CYS	-	expression tag	UNP F4JP36
C	11	ILE	-	expression tag	UNP F4JP36
C	12	LEU	-	expression tag	UNP F4JP36
C	13	LEU	-	expression tag	UNP F4JP36
C	14	ALA	-	expression tag	UNP F4JP36
C	15	VAL	-	expression tag	UNP F4JP36
C	16	VAL	-	expression tag	UNP F4JP36
C	17	ALA	-	expression tag	UNP F4JP36
C	18	PHE	-	expression tag	UNP F4JP36
C	19	VAL	-	expression tag	UNP F4JP36
C	20	GLY	-	expression tag	UNP F4JP36
C	21	LEU	-	expression tag	UNP F4JP36
C	22	SER	-	expression tag	UNP F4JP36
C	23	LEU	-	expression tag	UNP F4JP36
C	127	PRO	HIS	variant	UNP F4JP36
C	492	THR	-	expression tag	UNP F4JP36
C	493	ALA	-	expression tag	UNP F4JP36
C	494	THR	-	expression tag	UNP F4JP36
C	495	GLY	-	expression tag	UNP F4JP36
C	496	ASP	-	expression tag	UNP F4JP36
C	497	ASP	-	expression tag	UNP F4JP36
C	498	ASP	-	expression tag	UNP F4JP36
C	499	ASP	-	expression tag	UNP F4JP36
C	500	LYS	-	expression tag	UNP F4JP36

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	501	ALA	-	expression tag	UNP F4JP36
C	502	GLY	-	expression tag	UNP F4JP36
C	503	TRP	-	expression tag	UNP F4JP36
C	504	SER	-	expression tag	UNP F4JP36
C	505	HIS	-	expression tag	UNP F4JP36
C	506	PRO	-	expression tag	UNP F4JP36
C	507	GLN	-	expression tag	UNP F4JP36
C	508	PHE	-	expression tag	UNP F4JP36
C	509	GLU	-	expression tag	UNP F4JP36
C	510	LYS	-	expression tag	UNP F4JP36
C	511	GLY	-	expression tag	UNP F4JP36
C	512	GLY	-	expression tag	UNP F4JP36
C	513	GLY	-	expression tag	UNP F4JP36
C	514	SER	-	expression tag	UNP F4JP36
C	515	GLY	-	expression tag	UNP F4JP36
C	516	GLY	-	expression tag	UNP F4JP36
C	517	GLY	-	expression tag	UNP F4JP36
C	518	SER	-	expression tag	UNP F4JP36
C	519	GLY	-	expression tag	UNP F4JP36
C	520	GLY	-	expression tag	UNP F4JP36
C	521	GLY	-	expression tag	UNP F4JP36
C	522	SER	-	expression tag	UNP F4JP36
C	523	TRP	-	expression tag	UNP F4JP36
C	524	SER	-	expression tag	UNP F4JP36
C	525	HIS	-	expression tag	UNP F4JP36
C	526	PRO	-	expression tag	UNP F4JP36
C	527	GLN	-	expression tag	UNP F4JP36
C	528	PHE	-	expression tag	UNP F4JP36
C	529	GLU	-	expression tag	UNP F4JP36
C	530	LYS	-	expression tag	UNP F4JP36

- Molecule 2 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).

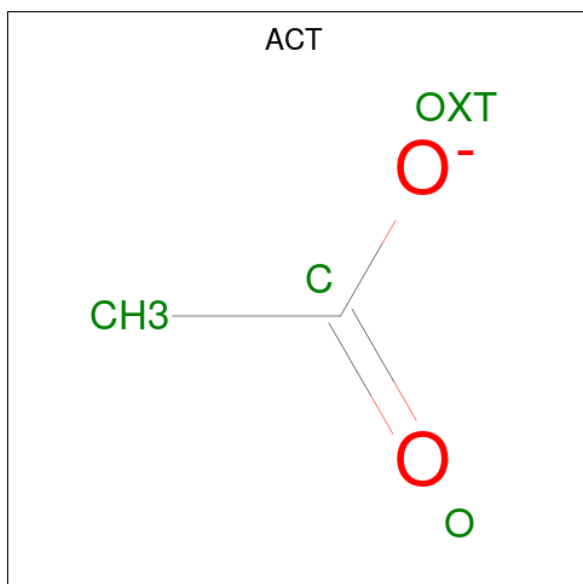


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	C	1	Total	C	N	O	0	0
			14	8	1	5		
2	C	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

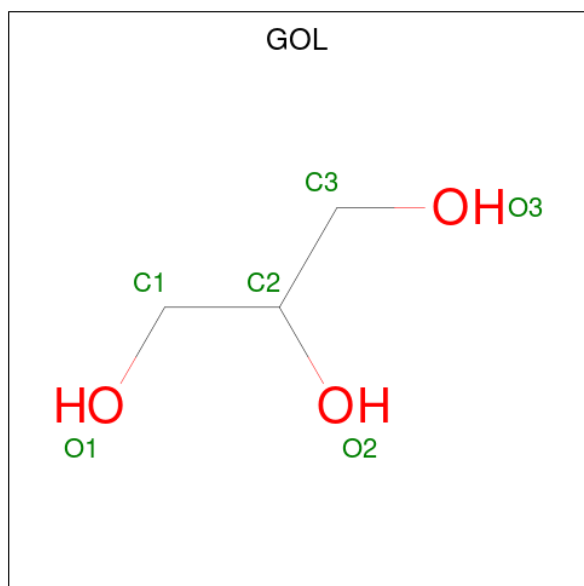
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	3	Total	Zn	0	0
			3	3		

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	1	Total C O 4 2 2	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	1	Total C O 6 3 3	0	0
5	C	1	Total C O 6 3 3	0	0
5	C	1	Total C O 6 3 3	0	0
5	C	1	Total C O 6 3 3	0	0
5	C	1	Total C O 6 3 3	0	0

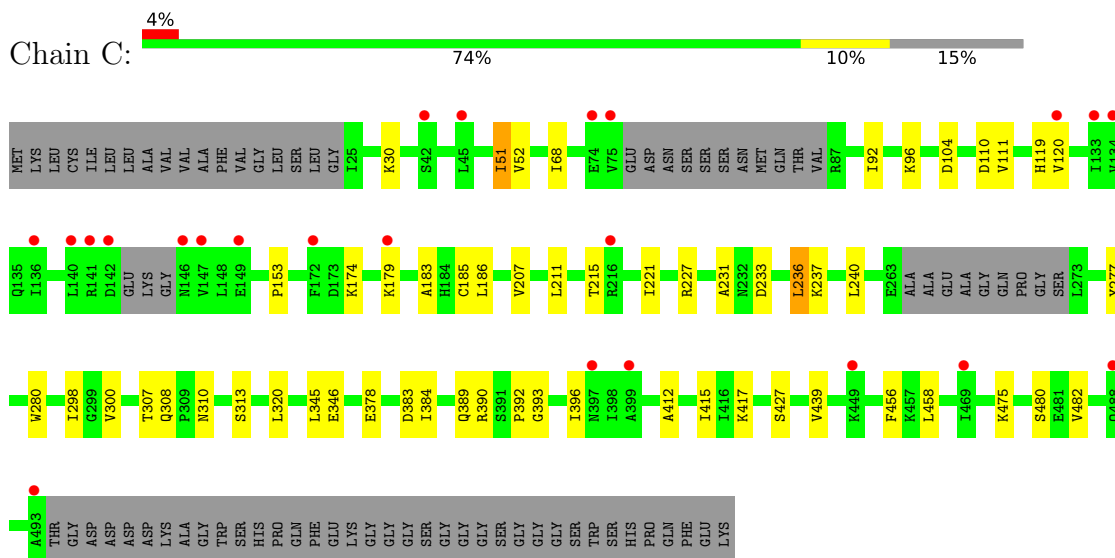
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	C	43	Total O 43 43	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: Protein HAPLESS 2



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	77.25Å 77.25Å 219.53Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.43 – 2.75 49.38 – 2.50	Depositor EDS
% Data completeness (in resolution range)	55.4 (42.43-2.75) 44.2 (49.38-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.49 (at 2.51Å)	Xtrriage
Refinement program	BUSTER 2.10.2	Depositor
R, $R_{free}$	0.231 , 0.277 0.247 , 0.307	Depositor DCC
$R_{free}$ test set	664 reflections (5.86%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.7	Xtrriage
Anisotropy	0.056	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 55.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	0.103 for h,-h-k,-l	Xtrriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	3619	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.70% 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, ACT, ZN, NAG

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	C	0.38	0/3586	0.56	0/4858

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	C	3511	0	3453	24	0
2	C	28	0	26	1	0
3	C	3	0	0	0	0
4	C	4	0	3	0	0
5	C	30	0	40	1	0
6	C	43	0	0	0	0
All	All	3619	0	3522	24	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 (24) 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:C:227:ARG:HB2	1:C:240:LEU:HB3	1.78	0.63
1:C:96:LYS:HG3	1:C:207:VAL:HG22	1.80	0.63
1:C:237:LYS:HB2	1:C:378:GLU:HB3	1.81	0.62
1:C:280:TRP:HB3	1:C:345:LEU:HD12	1.80	0.62
1:C:68:ILE:HG23	1:C:92:ILE:HB	1.89	0.54
1:C:231:ALA:HA	5:C:1008:GOL:H32	1.91	0.52
1:C:396:ILE:HB	1:C:415:ILE:HG23	1.93	0.49
1:C:396:ILE:HD11	1:C:417:LYS:HB2	1.93	0.49
1:C:52:VAL:HG22	1:C:378:GLU:HG3	1.96	0.48
1:C:174:LYS:HE3	1:C:179:LYS:HG3	1.97	0.47
1:C:427:SER:HB3	1:C:475:LYS:HB2	1.97	0.47
1:C:390:ARG:HH22	1:C:392:PRO:HG3	1.79	0.47
1:C:277:TYR:HB2	1:C:346:GLU:HG3	1.98	0.45
1:C:120:VAL:HB	1:C:183:ALA:HB3	1.99	0.45
1:C:186:LEU:HD13	1:C:320:LEU:HD11	1.98	0.45
1:C:221:ILE:H	2:C:1001:NAG:H83	1.83	0.44
1:C:153:PRO:HB3	1:C:185:CYS:HB3	1.99	0.43
1:C:310:ASN:HB3	1:C:313:SER:HB2	2.00	0.43
1:C:412:ALA:HB2	1:C:458:LEU:HD11	2.01	0.43
1:C:30:LYS:HE2	1:C:51:ILE:HD11	2.01	0.42
1:C:393:GLY:H	1:C:482:VAL:HG11	1.84	0.42
1:C:111:VAL:HG11	1:C:298:ILE:HG22	2.01	0.42
1:C:211:LEU:HD22	1:C:236:LEU:HD12	2.01	0.42
1:C:439:VAL:HG11	1:C:456:PHE:HB3	2.03	0.40

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	C	438/524 (84%)	412 (94%)	26 (6%)	0	<b>100</b> <b>100</b>

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	C	390/446 (87%)	376 (96%)	14 (4%)	35 55

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

Mol	Chain	Res	Type
1	C	51	ILE
1	C	104	ASP
1	C	110	ASP
1	C	119	HIS
1	C	215	THR
1	C	233	ASP
1	C	236	LEU
1	C	300	VAL
1	C	307	THR
1	C	308	GLN
1	C	383	ASP
1	C	384	ILE
1	C	389	GLN
1	C	480	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	308	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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 11 ligands modelled in this entry, 3 are monoatomic - leaving 8 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	NAG	C	1002	1	14,14,15	0.31	0	17,19,21	0.67	1 (5%)
5	GOL	C	1008	-	5,5,5	0.05	0	5,5,5	0.21	0
5	GOL	C	1009	-	5,5,5	0.04	0	5,5,5	0.18	0
5	GOL	C	1011	-	5,5,5	0.04	0	5,5,5	0.13	0
4	ACT	C	1006	-	3,3,3	1.04	0	3,3,3	0.96	0
5	GOL	C	1010	-	5,5,5	0.04	0	5,5,5	0.20	0
2	NAG	C	1001	1	14,14,15	0.29	0	17,19,21	0.69	0
5	GOL	C	1007	-	5,5,5	0.05	0	5,5,5	0.13	0

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	NAG	C	1002	1	-	0/6/23/26	0/1/1/1
5	GOL	C	1008	-	-	0/4/4/4	-
5	GOL	C	1009	-	-	0/4/4/4	-
5	GOL	C	1011	-	-	0/4/4/4	-
5	GOL	C	1010	-	-	0/4/4/4	-
2	NAG	C	1001	1	-	0/6/23/26	0/1/1/1
5	GOL	C	1007	-	-	0/4/4/4	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1002	NAG	C1-O5-C5	2.46	115.52	112.19

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
5	C	1008	GOL	1	0
2	C	1001	NAG	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	C	446/524 (85%)	0.28	23 (5%) <span style="border: 1px solid red; padding: 2px;">27</span> <span style="border: 1px solid red; padding: 2px;">33</span>	21, 70, 96, 119	0

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	493	ALA	4.9
1	C	147	VAL	4.3
1	C	134	VAL	3.8
1	C	469	ILE	3.5
1	C	449	LYS	3.4
1	C	399	ALA	3.4
1	C	133	ILE	2.9
1	C	488	GLN	2.9
1	C	149	GLU	2.9
1	C	142	ASP	2.9
1	C	136	ILE	2.8
1	C	120	VAL	2.7
1	C	216	ARG	2.6
1	C	141	ARG	2.4
1	C	45	LEU	2.3
1	C	179	LYS	2.3
1	C	140	LEU	2.2
1	C	397	ASN	2.2
1	C	172	PHE	2.2
1	C	74	GLU	2.2
1	C	146	ASN	2.2
1	C	75	VAL	2.1
1	C	42	SER	2.1

## 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
5	GOL	C	1008	6/6	0.86	0.14	73,74,74,74	0
2	NAG	C	1001	14/15	0.87	0.13	86,88,89,89	0
5	GOL	C	1010	6/6	0.87	0.19	66,66,67,67	0
5	GOL	C	1009	6/6	0.90	0.20	83,83,84,84	0
5	GOL	C	1007	6/6	0.93	0.24	57,57,57,58	0
2	NAG	C	1002	14/15	0.93	0.12	78,80,81,81	0
5	GOL	C	1011	6/6	0.93	0.08	69,70,71,71	0
4	ACT	C	1006	4/4	0.95	0.36	51,51,51,51	0
3	ZN	C	1003	1/1	0.98	0.11	65,65,65,65	0
3	ZN	C	1004	1/1	0.98	0.19	81,81,81,81	0
3	ZN	C	1005	1/1	0.99	0.07	119,119,119,119	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.