



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 02:55 AM GMT

PDB ID : 2JAG  
Title : ground state and L1 intermediate of halorhodopsin  
Authors : Gmelin, W.; Zeth, K.; Efremov, R.; Heberle, J.; Tittor, J.; Oesterhelt, D.  
Deposited on : 2006-11-28  
Resolution : 1.93 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

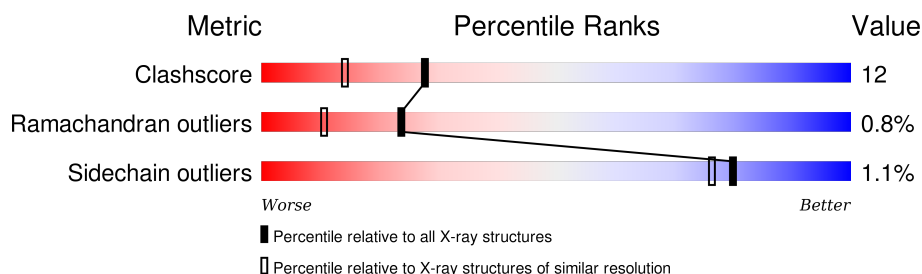
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.93 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	102246	3095 (1.96-1.92)
Ramachandran outliers	100387	3062 (1.96-1.92)
Sidechain outliers	100360	3062 (1.96-1.92)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	274	

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 2024 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 HALORHODOPSIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	242	Total	C	N	O	S	0	8	0
			1901	1259	305	327	10			

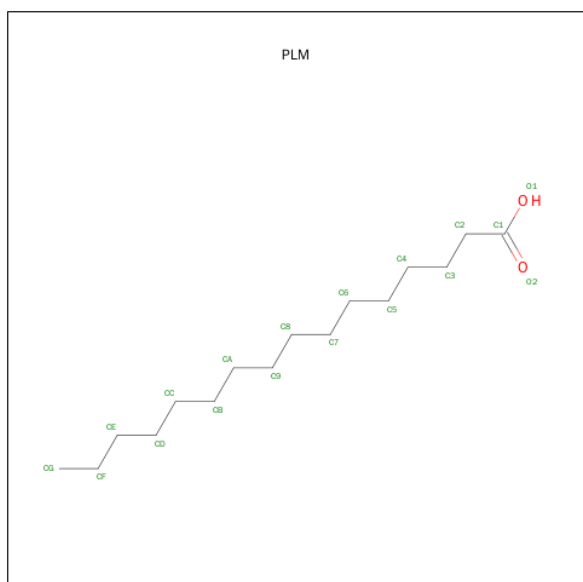
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	203	VAL	THR	ENGINEERED MUTATION	UNP P16102
A	229	ALA	VAL	ENGINEERED MUTATION	UNP P16102

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

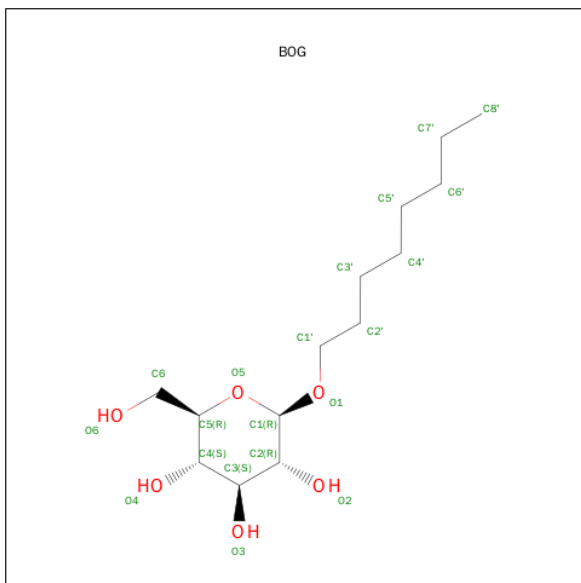
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Cl	0	2
			4	4		

- Molecule 3 is PALMITIC ACID (three-letter code: PLM) (formula: C<sub>16</sub>H<sub>32</sub>O<sub>2</sub>).



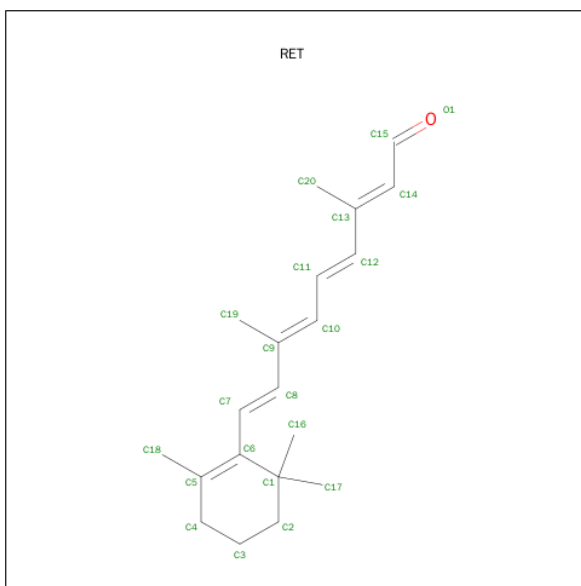
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			18	16	2		

- Molecule 4 is SUGAR (B-OCTYLGLUCOSIDE) (three-letter code: BOG) (formula:  $C_{14}H_{28}O_6$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			20	14	6		

- Molecule 5 is RETINAL (three-letter code: RET) (formula:  $C_{20}H_{28}O$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C 40 40	0	1

- Molecule 6 is water.

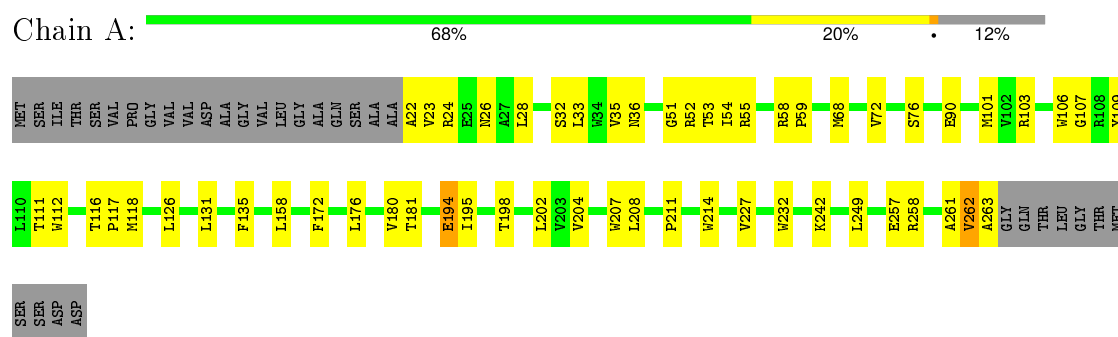
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	41	Total O 41 41	0	0

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

Note EDS was not executed.

#### • Molecule 1: HALORHODOPSIN



## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	95.04Å 95.04Å 157.67Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 1.93	Depositor
% Data completeness (in resolution range)	91.9 (20.00-1.93)	Depositor
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.237 , 0.261	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2024	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PLM, CL, RET, BOG

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	0.37	0/1949	0.54	0/2672

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	1901	0	1944	48	1
2	A	4	0	0	2	0
3	A	18	0	31	0	0
4	A	20	0	28	1	0
5	A	40	0	54	7	0
6	A	41	0	0	2	0
All	All	2024	0	2057	48	1

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 12.

All (48) 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:242[A]:LYS:CE	5:A:900[A]:RET:C15	2.27	1.12
1:A:242[A]:LYS:NZ	5:A:900[A]:RET:C14	2.13	1.11
1:A:242[A]:LYS:HZ1	5:A:900[A]:RET:C14	1.62	1.09
1:A:242[A]:LYS:NZ	5:A:900[A]:RET:C15	0.82	0.97
1:A:24:ARG:HH12	1:A:33:LEU:HD12	1.45	0.81
1:A:111[B]:THR:OG1	6:A:2019:HOH:O	2.05	0.74
1:A:76:SER:HB2	1:A:111[A]:THR:HG21	1.71	0.71
1:A:242[A]:LYS:HZ2	5:A:900[A]:RET:C15	0.96	0.70
1:A:198:THR:O	1:A:202:LEU:HD13	1.92	0.70
1:A:194:GLU:HG3	1:A:195:ILE:H	1.58	0.67
1:A:55:ARG:HD3	1:A:257:GLU:HB3	1.76	0.67
1:A:194:GLU:HG3	1:A:195:ILE:N	2.11	0.65
1:A:58:ARG:HB2	1:A:59:PRO:HD3	1.80	0.63
1:A:107:GLY:O	1:A:111[B]:THR:HG23	2.01	0.60
1:A:103:ARG:HD2	1:A:227:VAL:HG21	1.89	0.55
1:A:131:LEU:H	1:A:131:LEU:HD22	1.72	0.54
1:A:258:ARG:O	1:A:262:VAL:HG13	2.08	0.54
1:A:242[A]:LYS:HZ1	5:A:900[A]:RET:C15	0.88	0.54
1:A:90:GLU:HB2	1:A:101:MET:SD	2.48	0.53
1:A:24:ARG:NH1	1:A:33:LEU:HD12	2.19	0.53
1:A:176:LEU:HD22	1:A:204:VAL:HG22	1.91	0.53
1:A:103:ARG:HD3	2:A:502[A]:CL:CL	2.48	0.51
1:A:116:THR:OG1	1:A:117:PRO:HD3	2.11	0.50
1:A:126:LEU:HD22	1:A:249:LEU:CD1	2.41	0.50
1:A:131:LEU:HD22	1:A:131:LEU:N	2.27	0.50
1:A:51:GLY:HA2	1:A:54:ILE:HG12	1.94	0.49
1:A:180:VAL:HG23	1:A:181:THR:HG23	1.94	0.49
1:A:242[A]:LYS:HZ3	5:A:900[A]:RET:C15	1.13	0.48
1:A:118:MET:HA	1:A:118:MET:CE	2.43	0.47
1:A:24:ARG:NH2	2:A:502[B]:CL:CL	2.85	0.47
1:A:172:PHE:HE2	1:A:208:LEU:HD23	1.80	0.47
1:A:112[B]:TRP:O	1:A:116:THR:OG1	2.34	0.46
1:A:227:VAL:HG23	6:A:2038:HOH:O	2.15	0.45
1:A:22:ALA:C	1:A:24:ARG:H	2.19	0.45
1:A:106:TRP:HA	1:A:109[A]:TYR:HD2	1.82	0.45
1:A:131:LEU:CD2	1:A:131:LEU:H	2.30	0.44
1:A:109[A]:TYR:OH	1:A:214:TRP:NE1	2.45	0.44
1:A:35:VAL:HG13	1:A:36:ASN:N	2.33	0.44
1:A:158:LEU:HD21	4:A:801:BOG:H2	2.00	0.43
1:A:51:GLY:HA2	1:A:54:ILE:CG1	2.49	0.42
1:A:68:MET:O	1:A:72:VAL:HG23	2.20	0.42
1:A:26:ASN:OD1	1:A:28:LEU:HB3	2.20	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:52:ARG:HG3	1:A:53:THR:N	2.35	0.42
1:A:207[B]:TRP:O	1:A:211:PRO:HD2	2.20	0.41
1:A:261:ALA:C	1:A:263:ALA:H	2.24	0.41
1:A:116:THR:HG1	1:A:117:PRO:HD3	1.85	0.41
1:A:109[A]:TYR:HA	1:A:112[A]:TRP:CE3	2.55	0.41
1:A:32:SER:HB3	1:A:232:TRP:CD1	2.55	0.41

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:258:ARG:NH2	1:A:258:ARG:NH2[5_555]	1.92	0.28

### 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	248/274 (90%)	239 (96%)	7 (3%)	2 (1%)	24 10

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	23	VAL
1	A	262	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	195/210 (93%)	193 (99%)	2 (1%)	82	79

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	135	PHE
1	A	194	GLU

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

### 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 8 ligands modelled in this entry, 4 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	$\# Z  > 2$
3	PLM	A	700	-	14,17,17	0.17	0	14,17,17	1.03	1 (7%)
4	BOG	A	801	-	20,20,20	1.09	1 (5%)	25,25,25	0.65	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
5	RET	A	900[A]	-	19,20,21	4.55	11 (57%)	27,27,28	1.85	6 (22%)
5	RET	A	900[B]	1	19,20,21	4.77	14 (73%)	27,27,28	1.97	9 (33%)

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
3	PLM	A	700	-	-	0/13/15/15	0/0/0/0
4	BOG	A	801	-	-	0/11/31/31	0/1/1/1
5	RET	A	900[A]	-	-	0/13/30/31	0/1/1/1
5	RET	A	900[B]	1	-	0/13/30/31	0/1/1/1

All (26) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	900[B]	RET	C20-C13	-3.78	1.42	1.50
5	A	900[A]	RET	C15-C14	-2.72	1.38	1.49
5	A	900[B]	RET	C15-C14	-2.25	1.40	1.49
5	A	900[B]	RET	C7-C6	2.21	1.53	1.45
5	A	900[A]	RET	C16-C1	2.22	1.58	1.53
5	A	900[B]	RET	C2-C3	2.37	1.59	1.52
5	A	900[A]	RET	C3-C4	2.40	1.59	1.52
5	A	900[A]	RET	C2-C1	2.45	1.60	1.54
5	A	900[A]	RET	C11-C12	2.55	1.41	1.34
5	A	900[B]	RET	C3-C4	2.70	1.60	1.52
5	A	900[B]	RET	C11-C12	2.86	1.41	1.34
5	A	900[B]	RET	C12-C13	2.87	1.52	1.45
5	A	900[A]	RET	C10-C9	2.89	1.39	1.35
4	A	801	BOG	O5-C1	2.93	1.49	1.41
5	A	900[B]	RET	C18-C5	3.02	1.56	1.51
5	A	900[A]	RET	C4-C5	3.74	1.58	1.51
5	A	900[B]	RET	C2-C1	3.84	1.63	1.54
5	A	900[A]	RET	C17-C1	4.38	1.63	1.53
5	A	900[B]	RET	C19-C9	4.52	1.60	1.50
5	A	900[B]	RET	C10-C9	4.91	1.42	1.35
5	A	900[B]	RET	C4-C5	5.72	1.63	1.51
5	A	900[A]	RET	C19-C9	5.96	1.63	1.50
5	A	900[A]	RET	C1-C6	8.92	1.66	1.53
5	A	900[B]	RET	C5-C6	10.96	1.51	1.34
5	A	900[B]	RET	C1-C6	12.39	1.71	1.53

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	900[A]	RET	C5-C6	13.83	1.56	1.34

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	900[A]	RET	C8-C9-C10	-5.51	110.10	118.98
5	A	900[B]	RET	C8-C9-C10	-3.93	112.65	118.98
5	A	900[B]	RET	C20-C13-C14	-3.80	111.70	123.36
5	A	900[B]	RET	C1-C6-C5	-2.65	118.77	122.66
5	A	900[B]	RET	C11-C10-C9	-2.49	123.60	127.20
3	A	700	PLM	CE-CD-CC	-2.27	102.79	114.53
5	A	900[B]	RET	C7-C6-C5	-2.13	116.49	121.37
5	A	900[A]	RET	C1-C6-C5	-2.07	119.63	122.66
5	A	900[A]	RET	C19-C9-C10	2.53	126.64	122.90
5	A	900[A]	RET	C1-C6-C7	2.56	122.98	115.82
5	A	900[B]	RET	C2-C1-C6	2.72	114.68	110.36
5	A	900[B]	RET	C19-C9-C8	2.75	122.68	118.10
5	A	900[A]	RET	C19-C9-C8	2.79	122.74	118.10
5	A	900[A]	RET	C20-C13-C12	2.95	123.00	118.10
5	A	900[B]	RET	C1-C6-C7	3.18	124.72	115.82
5	A	900[B]	RET	C20-C13-C12	3.35	123.66	118.10

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	801	BOG	1	0
5	A	900[A]	RET	7	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 ⓘ

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section will therefore be empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section will therefore be empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section will therefore be empty.