



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 08:24 PM GMT

PDB ID : 1K72  
Title : The X-ray Crystal Structure Of Cel9G Complexed With cellotriose  
Authors : Mandelman, D.; Belaich, A.; Belaich, J.P.; Aghajari, N.; Driguez, H.; Haser, R.  
Deposited on : 2001-10-18  
Resolution : 1.80 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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

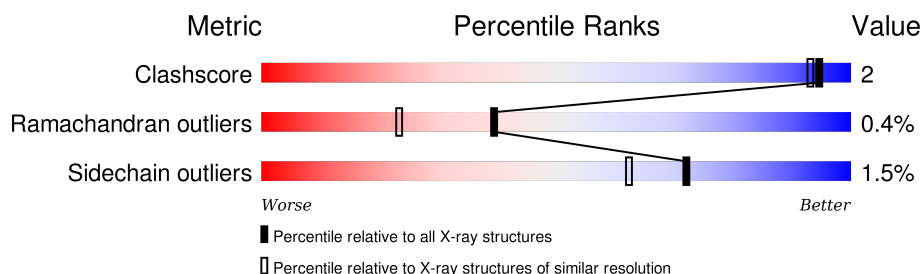
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	5383 (1.80-1.80)
Ramachandran outliers	100387	5320 (1.80-1.80)
Sidechain outliers	100360	5319 (1.80-1.80)

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	614	
1	B	614	

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	CBI	A	777[A]	X	-	-	-

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 10429 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.

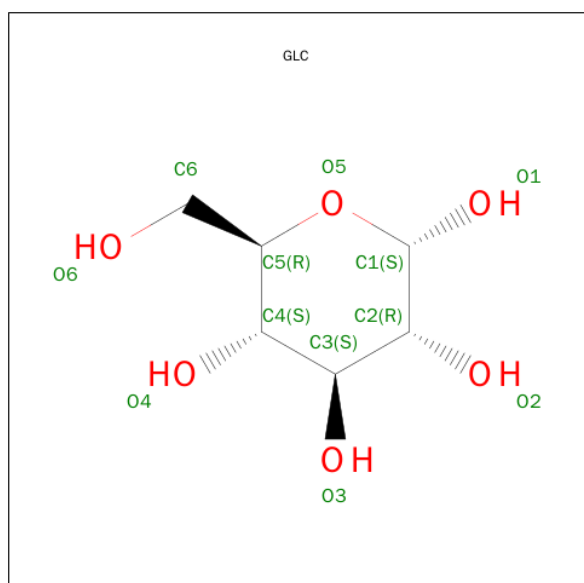
- Molecule 1 is a protein called Endoglucanase 9G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	612	Total	C	N	O	S	0	1	0
			4792	3048	785	939	20			
1	B	613	Total	C	N	O	S	0	2	0
			4794	3048	786	940	20			

There are 6 discrepancies between the modelled and reference sequences:

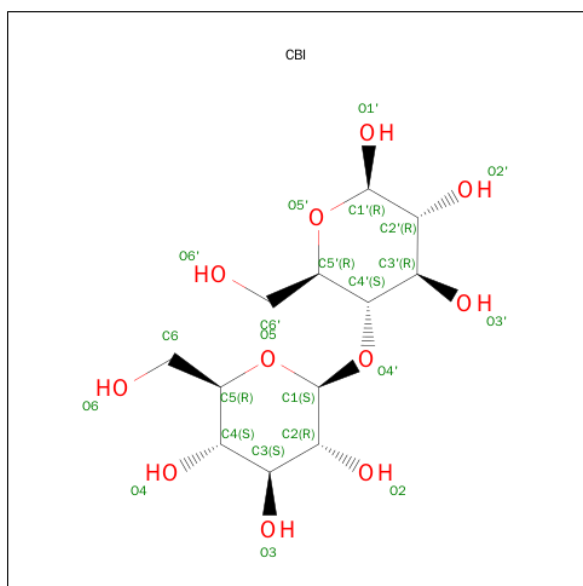
Chain	Residue	Modelled	Actual	Comment	Reference
A	257	SCH	CYS	MODIFIED RESIDUE	UNP P37700
A	574	THR	ARG	CONFLICT	UNP P37700
A	575	THR	ARG	CONFLICT	UNP P37700
B	257	SCH	CYS	MODIFIED RESIDUE	UNP P37700
B	574	THR	ARG	CONFLICT	UNP P37700
B	575	THR	ARG	CONFLICT	UNP P37700

- Molecule 2 is SUGAR (GLUCOSE) (three-letter code: GLC) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	6	6		

- Molecule 3 is SUGAR (CELLOBIOSE) (three-letter code: CBI) (formula:  $C_{12}H_{22}O_{11}$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	1
			46	24	22		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	Ca	0	0
			2	2		
4	A	2	Total	Ca	0	0
			2	2		

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Mg	0	0
			1	1		
5	A	3	Total	Mg	0	0
			3	3		

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



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

- Molecule 7 is water.

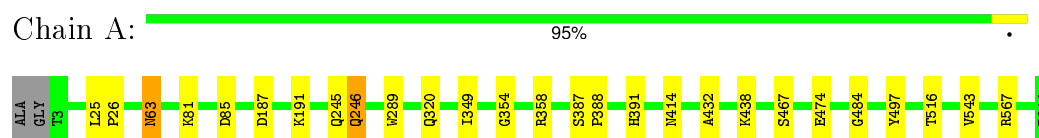
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	385	Total	O	0	0
			385	385		
7	B	380	Total	O	0	0
			380	380		

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

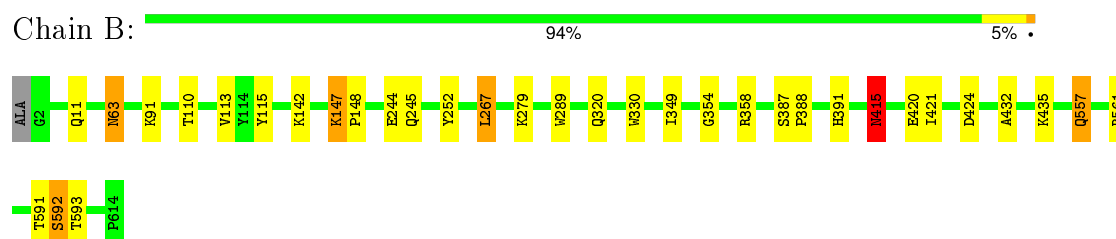
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Endoglucanase 9G



- Molecule 1: Endoglucanase 9G



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	57.00 Å   57.60 Å   86.80 Å 93.60°   100.80°   99.50°	Depositor
Resolution (Å)	36.31 – 1.80	Depositor
% Data completeness (in resolution range)	94.0 (36.31-1.80)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.05	Depositor
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.166 , 0.198	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	10429	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	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: GOL, MG, CBI, SCH, CA, GLC

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	0.53	0/4924	0.69	0/6710
1	B	0.52	0/4926	0.70	1/6714 (0.0%)
All	All	0.52	0/9850	0.70	1/13424 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	267	LEU	CA-CB-CG	6.07	129.26	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	148	PRO	Mainchain
1	B	415	ASN	Mainchain

### 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	4792	0	4488	13	0
1	B	4794	0	4480	19	0
2	A	12	0	12	0	0
3	A	46	0	42	0	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
5	A	3	0	0	0	0
5	B	1	0	0	0	0
6	A	6	0	8	0	0
6	B	6	0	8	0	0
7	A	385	0	0	0	0
7	B	380	0	0	1	0
All	All	10429	0	9038	32	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.

The worst 5 of 32 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:81[B]:LYS:NZ	1:A:85:ASP:OD2	2.14	0.79
1:B:591:THR:OG1	1:B:593:THR:HG22	1.86	0.74
1:B:91:LYS:HE2	7:B:824:HOH:O	1.91	0.71
1:A:497:TYR:HB3	1:A:543:VAL:HG22	1.74	0.68
1:B:244:GLU:HG3	1:B:252:TYR:CE1	2.33	0.62

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	A	610/614 (99%)	591 (97%)	17 (3%)	2 (0%)	46	29
1	B	612/614 (100%)	590 (96%)	19 (3%)	3 (0%)	34	17
All	All	1222/1228 (100%)	1181 (97%)	36 (3%)	5 (0%)	39	23

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	592	SER
1	B	245	GLN
1	A	354	GLY
1	B	354	GLY
1	A	484	GLY

### 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	503/506 (99%)	498 (99%)	5 (1%)	82	77
1	B	502/506 (99%)	492 (98%)	10 (2%)	63	49
All	All	1005/1012 (99%)	990 (98%)	15 (2%)	72	62

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	63	ASN
1	B	147	LYS
1	B	424	ASP
1	B	11	GLN
1	B	415	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	59	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	63	ASN
1	B	415	ASN
1	A	551	ASN
1	B	519	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

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	$\# Z  > 2$
1	SCH	A	257	1	6,7,8	0.85	0	3,7,9	1.17	0
1	SCH	B	257	1	6,7,8	0.55	0	3,7,9	1.14	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
1	SCH	A	257	1	-	0/2/6/8	0/0/0/0
1	SCH	B	257	1	-	0/2/6/8	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 13 ligands modelled in this entry, 8 are monoatomic - leaving 5 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	GLC	A	615	-	12,12,12	0.35	0	17,17,17	0.42	0
3	CBI	A	777[A]	5	24,24,24	0.95	1 (4%)	35,35,35	1.23	5 (14%)
3	CBI	A	777[B]	5	24,24,24	1.01	1 (4%)	35,35,35	1.26	6 (17%)
6	GOL	A	783	-	5,5,5	0.91	0	5,5,5	0.84	0
6	GOL	B	618	-	5,5,5	0.98	0	5,5,5	0.91	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	GLC	A	615	-	-	0/2/22/22	0/1/1/1
3	CBI	A	777[A]	5	1/1/10/10	0/8/48/48	0/2/2/2
3	CBI	A	777[B]	5	-	0/8/48/48	0/2/2/2
6	GOL	A	783	-	-	0/4/4/4	0/0/0/0
6	GOL	B	618	-	-	0/4/4/4	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	777[A]	CBI	O5'-C1'	2.19	1.47	1.43
3	A	777[B]	CBI	O5'-C1'	2.19	1.47	1.43

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	A	777[A]	CBI	C1'-C2'-C3'	-2.62	106.53	110.43
3	A	777[B]	CBI	C1'-C2'-C3'	-2.62	106.53	110.43
3	A	777[B]	CBI	O1'-C1'-O5'	-2.36	103.79	110.25
3	A	777[A]	CBI	C2'-C3'-C4'	-2.30	104.55	109.60
3	A	777[B]	CBI	C2'-C3'-C4'	-2.30	104.55	109.60

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	777[A]	CBI	C1'

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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.