



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 03:33 PM GMT

PDB ID : 4CN4  
Title : GlgE isoform 1 from Streptomyces coelicolor E423A mutant with 2-deoxy- 2-fluoro-beta-maltosyl modification  
Authors : Syson, K.; Stevenson, C.E.M.; Rashid, A.M.; Saalbach, G.; Tang, M.; Tuukainen, A.; Svergun, D.I.; Withers, S.G.; Lawson, D.M.; Bornemann, S.  
Deposited on : 2014-01-21  
Resolution : 2.40 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) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

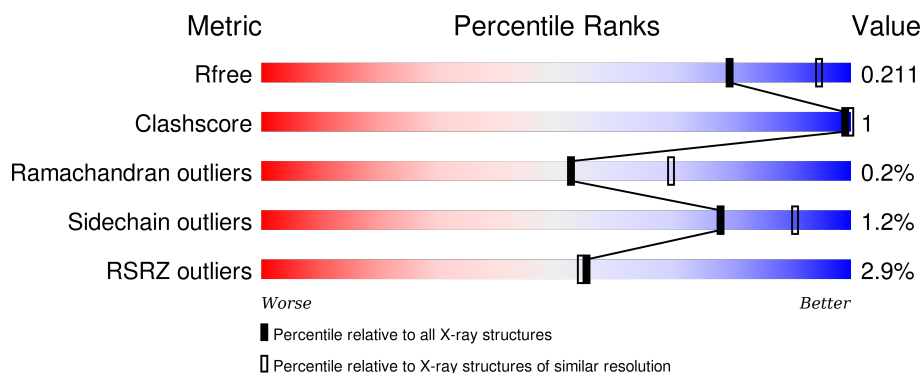
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

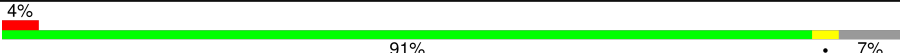

The reported resolution of this entry is 2.40 Å.

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))
$R_{free}$	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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	695	
1	B	695	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 10679 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 ALPHA-1,4-GLUCAN\;MALTOSE-1-PHOSPHATE MALTOSYLTRANSFERASE 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	649	Total	C	N	O	S	0	5	0
			5137	3253	932	942	10			
1	B	649	Total	C	N	O	S	0	2	0
			5080	3220	911	939	10			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	EXPRESSION TAG	UNP Q9L1K2
A	-18	GLY	-	EXPRESSION TAG	UNP Q9L1K2
A	-17	SER	-	EXPRESSION TAG	UNP Q9L1K2
A	-16	SER	-	EXPRESSION TAG	UNP Q9L1K2
A	-15	HIS	-	EXPRESSION TAG	UNP Q9L1K2
A	-14	HIS	-	EXPRESSION TAG	UNP Q9L1K2
A	-13	HIS	-	EXPRESSION TAG	UNP Q9L1K2
A	-12	HIS	-	EXPRESSION TAG	UNP Q9L1K2
A	-11	HIS	-	EXPRESSION TAG	UNP Q9L1K2
A	-10	HIS	-	EXPRESSION TAG	UNP Q9L1K2
A	-9	SER	-	EXPRESSION TAG	UNP Q9L1K2
A	-8	SER	-	EXPRESSION TAG	UNP Q9L1K2
A	-7	GLY	-	EXPRESSION TAG	UNP Q9L1K2
A	-6	LEU	-	EXPRESSION TAG	UNP Q9L1K2
A	-5	VAL	-	EXPRESSION TAG	UNP Q9L1K2
A	-4	PRO	-	EXPRESSION TAG	UNP Q9L1K2
A	-3	ARG	-	EXPRESSION TAG	UNP Q9L1K2
A	-2	GLY	-	EXPRESSION TAG	UNP Q9L1K2
A	-1	SER	-	EXPRESSION TAG	UNP Q9L1K2
A	0	HIS	-	EXPRESSION TAG	UNP Q9L1K2
A	423	ALA	GLU	ENGINEERED MUTATION	UNP Q9L1K2
B	-19	MET	-	EXPRESSION TAG	UNP Q9L1K2
B	-18	GLY	-	EXPRESSION TAG	UNP Q9L1K2
B	-17	SER	-	EXPRESSION TAG	UNP Q9L1K2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	SER	-	EXPRESSION TAG	UNP Q9L1K2
B	-15	HIS	-	EXPRESSION TAG	UNP Q9L1K2
B	-14	HIS	-	EXPRESSION TAG	UNP Q9L1K2
B	-13	HIS	-	EXPRESSION TAG	UNP Q9L1K2
B	-12	HIS	-	EXPRESSION TAG	UNP Q9L1K2
B	-11	HIS	-	EXPRESSION TAG	UNP Q9L1K2
B	-10	HIS	-	EXPRESSION TAG	UNP Q9L1K2
B	-9	SER	-	EXPRESSION TAG	UNP Q9L1K2
B	-8	SER	-	EXPRESSION TAG	UNP Q9L1K2
B	-7	GLY	-	EXPRESSION TAG	UNP Q9L1K2
B	-6	LEU	-	EXPRESSION TAG	UNP Q9L1K2
B	-5	VAL	-	EXPRESSION TAG	UNP Q9L1K2
B	-4	PRO	-	EXPRESSION TAG	UNP Q9L1K2
B	-3	ARG	-	EXPRESSION TAG	UNP Q9L1K2
B	-2	GLY	-	EXPRESSION TAG	UNP Q9L1K2
B	-1	SER	-	EXPRESSION TAG	UNP Q9L1K2
B	0	HIS	-	EXPRESSION TAG	UNP Q9L1K2
B	423	ALA	GLU	ENGINEERED MUTATION	UNP Q9L1K2

- Molecule 2 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	2	Total	C	F	O	0	0
			22	12	1	9		
2	B	2	Total	C	F	O	0	0
			22	12	1	9		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	248	Total	O	0	0
			248	248		
3	B	170	Total	O	0	0
			170	170		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	113.00Å 113.00Å 312.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.54 – 2.40 50.54 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.5 (50.54-2.40) 99.5 (50.54-2.40)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.43 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, $R_{free}$	0.184 , 0.208 0.187 , 0.211	Depositor DCC
$R_{free}$ test set	3985 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	47.8	Xtriage
Anisotropy	0.265	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 26.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtriage
Outliers	0 of 79410 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	10679	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.21% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: GLC, SHG

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	0.58	0/5305	0.71	2/7256 (0.0%)
1	B	0.54	0/5238	0.69	4/7173 (0.1%)
All	All	0.56	0/10543	0.70	6/14429 (0.0%)

There are no bond length outliers.

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	67	ASP	CB-CG-OD1	5.99	123.69	118.30
1	A	394	ASP	CB-CG-OD2	5.93	123.64	118.30
1	A	67	ASP	CB-CG-OD1	5.67	123.40	118.30
1	B	394	ASP	CB-CG-OD2	5.46	123.22	118.30
1	B	67	ASP	CB-CG-OD2	-5.20	113.62	118.30
1	B	32	ARG	NE-CZ-NH1	5.17	122.89	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	5137	0	4972	6	0
1	B	5080	0	4872	8	0
2	A	22	0	18	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	22	0	18	0	0
3	A	248	0	0	1	0
3	B	170	0	0	2	0
All	All	10679	0	9880	14	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 1.

All (14) 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:103:THR:HG21	3:A:2059:HOH:O	1.96	0.66
1:B:26:VAL:HG22	1:B:200:LEU:HD12	1.84	0.58
1:A:423:ALA:HA	1:A:443:TYR:CD1	2.43	0.53
1:B:103:THR:HG21	3:B:2058:HOH:O	2.08	0.51
1:B:423:ALA:HA	1:B:443:TYR:CD1	2.47	0.49
1:B:53:ARG:NH1	3:B:2033:HOH:O	2.41	0.49
1:B:421:LEU:HD13	1:B:441:GLN:HB3	1.95	0.47
1:A:630:VAL:HB	1:A:659:CYS:HB3	1.98	0.46
1:A:421:LEU:HD13	1:A:441:GLN:HB3	2.00	0.44
1:B:630:VAL:HB	1:B:659:CYS:HB3	1.99	0.43
1:A:489:GLY:HA3	1:A:493:ALA:HB2	2.01	0.42
1:B:212:GLY:O	1:B:507:THR:HA	2.20	0.42
1:B:489:GLY:HA3	1:B:493:ALA:HB2	2.02	0.41
1:A:212:GLY:O	1:A:507:THR:HA	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	652/695 (94%)	637 (98%)	14 (2%)	1 (0%)	52	69
1	B	649/695 (93%)	638 (98%)	10 (2%)	1 (0%)	52	69
All	All	1301/1390 (94%)	1275 (98%)	24 (2%)	2 (0%)	52	69

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	643	ARG
1	B	643	ARG

### 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	532/574 (93%)	526 (99%)	6 (1%)	80	92
1	B	521/574 (91%)	514 (99%)	7 (1%)	76	89
All	All	1053/1148 (92%)	1040 (99%)	13 (1%)	78	90

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	31	GLN
1	A	196	SER
1	A	511	TYR
1	A	538	LYS
1	A	598	LEU
1	A	606	ARG
1	B	26	VAL
1	B	31	GLN
1	B	167	ARG
1	B	180	GLN
1	B	196	SER
1	B	511	TYR
1	B	538	LYS

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 ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

4 carbohydrates 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$
2	SHG	A	1394	1,2	11,11,12	2.76	4 (36%)	11,15,17	1.54	2 (18%)
2	GLC	A	1395	2	11,11,12	1.28	1 (9%)	14,15,17	1.00	1 (7%)
2	SHG	B	1394	1,2	11,11,12	2.07	3 (27%)	11,15,17	1.58	3 (27%)
2	GLC	B	1395	2	11,11,12	1.22	1 (9%)	14,15,17	1.01	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	SHG	A	1394	1,2	-	0/2/19/22	0/1/1/1
2	GLC	A	1395	2	-	0/2/19/22	0/1/1/1
2	SHG	B	1394	1,2	-	0/2/19/22	0/1/1/1
2	GLC	B	1395	2	-	0/2/19/22	0/1/1/1

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1395	GLC	O2-C2	2.28	1.48	1.43
2	A	1394	SHG	C4-C3	2.61	1.59	1.52
2	B	1394	SHG	O5-C1	3.32	1.49	1.43
2	A	1395	GLC	O2-C2	3.64	1.51	1.43
2	B	1394	SHG	C1-C2	3.66	1.57	1.52
2	B	1394	SHG	C2-C3	3.85	1.55	1.51
2	A	1394	SHG	O5-C1	4.56	1.51	1.43
2	A	1394	SHG	C2-C3	4.88	1.56	1.51
2	A	1394	SHG	C1-C2	4.98	1.59	1.52

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1394	SHG	O6-C6-C5	-3.40	100.10	111.33
2	A	1394	SHG	O6-C6-C5	-3.26	100.56	111.33
2	A	1394	SHG	C1-O5-C5	-2.53	109.03	112.25
2	B	1394	SHG	C3-C4-C5	-2.41	106.00	110.20
2	B	1394	SHG	C1-O5-C5	-2.39	109.22	112.25
2	A	1395	GLC	O2-C2-C3	2.52	115.19	110.12

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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	649/695 (93%)	-0.11	26 (4%) 42 43	31, 45, 85, 106	0
1	B	649/695 (93%)	-0.19	12 (1%) 71 71	32, 53, 91, 130	0
All	All	1298/1390 (93%)	-0.15	38 (2%) 55 54	31, 49, 87, 130	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	176	ALA	5.4
1	A	147	PRO	4.2
1	A	181	VAL	4.2
1	B	124	ALA	4.1
1	A	144	ALA	3.8
1	A	184	VAL	3.7
1	A	145	ASP	3.6
1	B	644	ALA	3.5
1	B	625	HIS	3.4
1	A	122	ILE	3.4
1	B	624	TRP	3.4
1	B	15	PRO	3.3
1	B	612	THR	3.2
1	A	180	GLN	3.1
1	A	124	ALA	3.1
1	A	123	PRO	3.0
1	A	119	ARG	2.9
1	A	178	THR	2.9
1	A	125	GLY	2.8
1	A	185	LEU	2.7
1	A	152	ARG	2.6
1	A	625	HIS	2.5
1	B	641	TRP	2.4
1	A	177	LEU	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	169	ALA	2.4
1	A	118	ALA	2.4
1	B	627	SER	2.3
1	B	618	PRO	2.3
1	B	226	HIS	2.3
1	A	183	ALA	2.3
1	A	15	PRO	2.3
1	A	71	ARG	2.2
1	A	179	PRO	2.2
1	A	173	LEU	2.2
1	B	524	GLU	2.1
1	B	225	PRO	2.1
1	A	524	GLU	2.0
1	A	148	GLY	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](#)

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( $\text{\AA}^2$ )	Q<0.9
2	GLC	B	1395	11/12	0.97	0.11	-0.30	47,48,48,49	0
2	SHG	B	1394	11/12	0.93	0.10	-1.59	50,52,54,54	0
2	GLC	A	1395	11/12	0.97	0.09	-1.89	39,41,42,42	0
2	SHG	A	1394	11/12	0.94	0.09	-2.26	42,45,49,51	0

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.