



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 02:58 am GMT

PDB ID : 1NQI  
Title : crystal structure of lactose synthase, a 1:1 complex between beta1,4-galactosyltransferase and alpha-lactalbumin in the presence of GlcNAc  
Authors : Ramakrishnan, B.; Qasba, P.K.  
Deposited on : 2003-01-21  
Resolution : 2.00 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

<https://www.wwpdb.org/validation/2017/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	:	1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix)	:	1.13
EDS	:	trunk30967
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30967

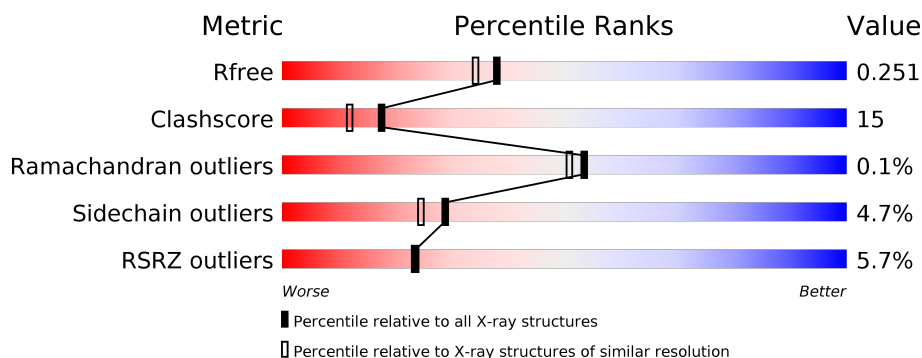
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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))
$R_{free}$	111664	7193 (2.00-2.00)
Clashscore	122126	8267 (2.00-2.00)
Ramachandran outliers	120053	8166 (2.00-2.00)
Sidechain outliers	120020	8165 (2.00-2.00)
RSRZ outliers	108989	7011 (2.00-2.00)

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	123	<div> <div>5%</div> <div> <div></div> <div>84%</div> <div>14%</div> <div>•</div> </div> </div>
1	C	123	<div> <div>5%</div> <div> <div></div> <div>84%</div> <div>14%</div> <div>•</div> </div> </div>
2	B	286	<div> <div>3%</div> <div> <div></div> <div>78%</div> <div>15%</div> <div>• 5%</div> </div> </div>
2	D	286	<div> <div>9%</div> <div> <div></div> <div>56%</div> <div>36%</div> <div>• 5%</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7198 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-LACTALBUMIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	123	Total	C	N	O	S	0	0	0
			980	620	156	195	9			
1	C	123	Total	C	N	O	S	0	0	0
			980	620	156	195	9			

- Molecule 2 is a protein called BETA-1,4-GALACTOSYLTRANSFERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	272	Total	C	N	O	S	0	0	0
			2218	1424	382	398	14			
2	D	272	Total	C	N	O	S	0	0	0
			2218	1424	382	398	14			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	117	ALA	-	SEE REMARK 999	UNP P08037
B	118	SER	-	SEE REMARK 999	UNP P08037
B	119	MET	-	SEE REMARK 999	UNP P08037
B	120	THR	-	SEE REMARK 999	UNP P08037
B	121	GLY	-	SEE REMARK 999	UNP P08037
B	122	GLY	-	SEE REMARK 999	UNP P08037
B	123	GLN	-	SEE REMARK 999	UNP P08037
B	124	GLN	-	SEE REMARK 999	UNP P08037
B	125	MET	-	SEE REMARK 999	UNP P08037
B	126	GLY	-	SEE REMARK 999	UNP P08037
B	127	ARG	-	SEE REMARK 999	UNP P08037
B	128	GLY	-	SEE REMARK 999	UNP P08037
B	129	SER	-	SEE REMARK 999	UNP P08037
D	117	ALA	-	SEE REMARK 999	UNP P08037
D	118	SER	-	SEE REMARK 999	UNP P08037
D	119	MET	-	SEE REMARK 999	UNP P08037

*Continued on next page...*

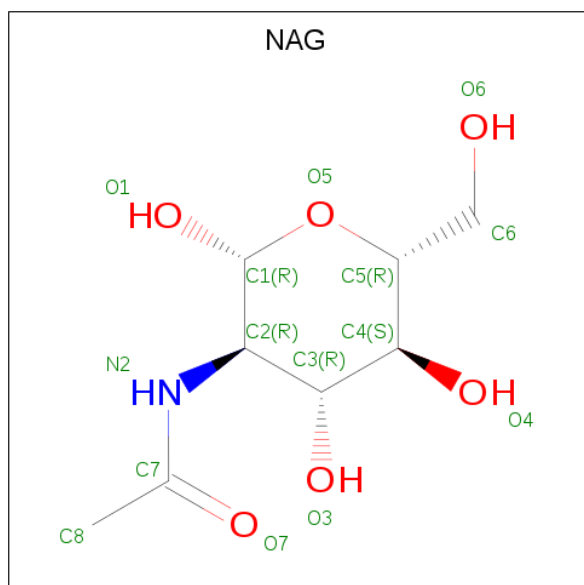
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	120	THR	-	SEE REMARK 999	UNP P08037
D	121	GLY	-	SEE REMARK 999	UNP P08037
D	122	GLY	-	SEE REMARK 999	UNP P08037
D	123	GLN	-	SEE REMARK 999	UNP P08037
D	124	GLN	-	SEE REMARK 999	UNP P08037
D	125	MET	-	SEE REMARK 999	UNP P08037
D	126	GLY	-	SEE REMARK 999	UNP P08037
D	127	ARG	-	SEE REMARK 999	UNP P08037
D	128	GLY	-	SEE REMARK 999	UNP P08037
D	129	SER	-	SEE REMARK 999	UNP P08037

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Ca 1 1	0	0
3	C	1	Total Ca 1 1	0	0

- Molecule 4 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).

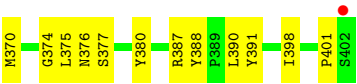


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C N O 15 8 1 6	0	0
4	D	1	Total C N O 15 8 1 6	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	155	Total 155	O 155	0	0
5	B	244	Total 244	O 244	0	0
5	C	157	Total 157	O 157	0	0
5	D	214	Total 214	O 214	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	57.16Å 96.38Å 99.48Å 90.00° 100.98° 90.00°	Depositor
Resolution (Å)	24.50 – 2.00 24.49 – 1.99	Depositor EDS
% Data completeness (in resolution range)	95.1 (24.50-2.00) 94.1 (24.49-1.99)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.25 (at 1.99Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.210 , 0.258 0.204 , 0.251	Depositor DCC
$R_{free}$ test set	6887 reflections (10.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	23.0	Xtriage
Anisotropy	0.313	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 58.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7198	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.09% 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: CA, 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	A	0.57	0/1001	0.76	1/1350 (0.1%)
1	C	0.60	0/1001	0.87	3/1350 (0.2%)
2	B	0.57	0/2278	0.75	0/3085
2	D	0.55	0/2278	0.77	0/3085
All	All	0.57	0/6558	0.78	4/8870 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	67	PRO	N-CA-C	7.56	131.75	112.10
1	C	68	GLU	N-CA-C	6.62	128.87	111.00
1	C	57	ASP	CB-CG-OD1	5.85	123.57	118.30
1	A	60	TRP	N-CA-C	5.15	124.89	111.00

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	980	0	936	14	0
1	C	980	0	936	15	0
2	B	2218	0	2185	44	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	2218	0	2185	114	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
4	B	15	0	15	0	0
4	D	15	0	15	1	0
5	A	155	0	0	2	0
5	B	244	0	0	4	0
5	C	157	0	0	3	0
5	D	214	0	0	11	0
All	All	7198	0	6272	186	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 15.

The worst 5 of 186 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)
2:D:270:PRO:HD3	2:D:330:MET:HE2	1.31	1.10
2:D:325:LEU:HD13	2:D:326:ALA:N	1.73	1.02
2:D:270:PRO:HD3	2:D:330:MET:CE	1.89	1.01
2:B:150:ASN:HD22	2:B:151:ILE:H	1.09	0.97
2:B:155:LEU:H	2:B:155:LEU:HD12	1.27	0.96

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	121/123 (98%)	117 (97%)	4 (3%)	0	100	100
1	C	121/123 (98%)	117 (97%)	4 (3%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	270/286 (94%)	260 (96%)	10 (4%)	0	100	100
2	D	270/286 (94%)	255 (94%)	14 (5%)	1 (0%)	36	31
All	All	782/818 (96%)	749 (96%)	32 (4%)	1 (0%)	53	51

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	152	PRO

### 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	109/109 (100%)	106 (97%)	3 (3%)	47	47
1	C	109/109 (100%)	105 (96%)	4 (4%)	37	35
2	B	245/254 (96%)	232 (95%)	13 (5%)	25	20
2	D	245/254 (96%)	232 (95%)	13 (5%)	25	20
All	All	708/726 (98%)	675 (95%)	33 (5%)	29	24

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	336	ASN
1	C	85	LEU
2	D	331	SER
2	B	375	LEU
1	C	65	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	336	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	39	GLN
2	D	310	ASN
2	B	347	HIS
1	C	32	HIS

### 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 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	$\# Z  > 2$
4	NAG	B	403	-	15,15,15	1.12	1 (6%)	21,21,21	0.81	0
4	NAG	D	527	-	15,15,15	0.92	0	21,21,21	0.84	1 (4%)

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
4	NAG	B	403	-	-	0/6/26/26	0/1/1/1
4	NAG	D	527	-	-	0/6/26/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	403	NAG	C2-N2	3.65	1.51	1.45

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	527	NAG	C8-C7-N2	-2.13	112.36	116.10

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	527	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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	123/123 (100%)	-0.29	6 (4%) 29 29	13, 21, 45, 64	0
1	C	123/123 (100%)	-0.17	6 (4%) 29 29	11, 21, 46, 72	0
2	B	272/286 (95%)	-0.23	8 (2%) 51 51	11, 24, 46, 60	0
2	D	272/286 (95%)	0.55	25 (9%) 9 8	16, 33, 57, 66	0
All	All	790/818 (96%)	0.04	45 (5%) 24 23	11, 26, 52, 72	0

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	123	PRO	6.7
2	D	132	THR	6.4
2	D	402	SER	5.1
2	D	131	LEU	4.8
1	C	122	LYS	4.7

### 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. 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
4	NAG	D	527	15/15	0.97	0.12	14,15,18,20	0
4	NAG	B	403	15/15	0.98	0.10	12,15,17,19	0
3	CA	C	526	1/1	0.99	0.09	18,18,18,18	0
3	CA	A	124	1/1	1.00	0.05	18,18,18,18	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.