



# wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 9, 2023 – 06:10 PM JST

PDB ID : 8GN6  
Title : Crystallization of Sialidase from Porphyromonas gingivalis  
Authors : Dong, W.B.  
Deposited on : 2022-08-23  
Resolution : 2.10 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 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
Xtriage (Phenix)	:	1.13
EDS	:	2.32.2
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)	:	2.32.2

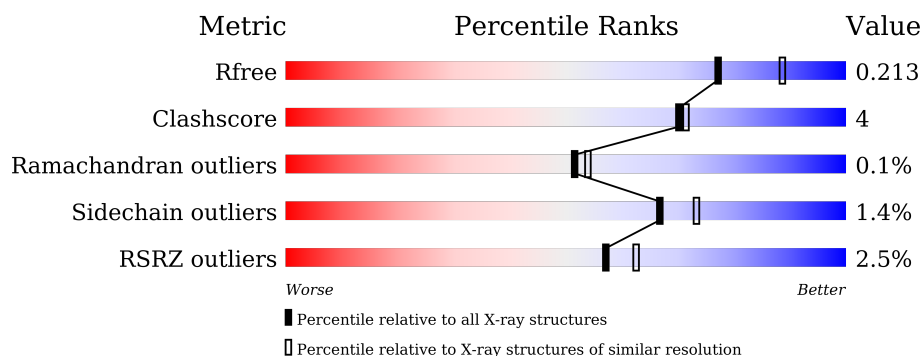
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	A	526	<div> <div>4%</div> <div> <div></div> <div>83%</div> <div>10%</div> <div>6%</div> </div> </div>
1	B	526	<div> <div>2%</div> <div> <div></div> <div>84%</div> <div>10%</div> <div>6%</div> </div> </div>
2	C	524	<div> <div>0%</div> <div> <div></div> <div>86%</div> <div>8%</div> <div>6%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 12483 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 Sialidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	493	Total	C	N	O	S	0	0	0
			3857	2416	686	741	14			
1	B	496	Total	C	N	O	S	0	0	0
			3882	2430	693	745	14			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	18	VAL	GLY	conflict	UNP A0A1R4DV85
A	21	CYS	TRP	conflict	UNP A0A1R4DV85
A	23	VAL	MET	conflict	UNP A0A1R4DV85
A	26	LEU	MET	conflict	UNP A0A1R4DV85
A	29	THR	SER	conflict	UNP A0A1R4DV85
A	30	ALA	GLY	conflict	UNP A0A1R4DV85
A	83	ALA	VAL	conflict	UNP A0A1R4DV85
A	120	LEU	PHE	conflict	UNP A0A1R4DV85
A	156	LEU	VAL	conflict	UNP A0A1R4DV85
A	178	GLU	GLY	conflict	UNP A0A1R4DV85
A	292	GLY	GLU	conflict	UNP A0A1R4DV85
A	342	MET	THR	conflict	UNP A0A1R4DV85
A	371	GLN	HIS	conflict	UNP A0A1R4DV85
A	397	ILE	VAL	conflict	UNP A0A1R4DV85
B	18	VAL	GLY	conflict	UNP A0A1R4DV85
B	21	CYS	TRP	conflict	UNP A0A1R4DV85
B	23	VAL	MET	conflict	UNP A0A1R4DV85
B	26	LEU	MET	conflict	UNP A0A1R4DV85
B	29	THR	SER	conflict	UNP A0A1R4DV85
B	30	ALA	GLY	conflict	UNP A0A1R4DV85
B	83	ALA	VAL	conflict	UNP A0A1R4DV85
B	120	LEU	PHE	conflict	UNP A0A1R4DV85
B	156	LEU	VAL	conflict	UNP A0A1R4DV85
B	178	GLU	GLY	conflict	UNP A0A1R4DV85
B	292	GLY	GLU	conflict	UNP A0A1R4DV85

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	342	MET	THR	conflict	UNP A0A1R4DV85
B	371	GLN	HIS	conflict	UNP A0A1R4DV85
B	397	ILE	VAL	conflict	UNP A0A1R4DV85

- Molecule 2 is a protein called Sialidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	494	Total	C	N	O	S	0	0	0
			3862	2419	687	742	14			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	18	VAL	GLY	conflict	UNP A0A1R4DV85
C	21	CYS	TRP	conflict	UNP A0A1R4DV85
C	23	VAL	MET	conflict	UNP A0A1R4DV85
C	26	LEU	MET	conflict	UNP A0A1R4DV85
C	29	THR	SER	conflict	UNP A0A1R4DV85
C	30	ALA	GLY	conflict	UNP A0A1R4DV85
C	83	ALA	VAL	conflict	UNP A0A1R4DV85
C	120	LEU	PHE	conflict	UNP A0A1R4DV85
C	156	LEU	VAL	conflict	UNP A0A1R4DV85
C	178	GLU	GLY	conflict	UNP A0A1R4DV85
C	292	GLY	GLU	conflict	UNP A0A1R4DV85
C	342	MET	THR	conflict	UNP A0A1R4DV85
C	371	GLN	HIS	conflict	UNP A0A1R4DV85
C	397	ILE	VAL	conflict	UNP A0A1R4DV85

- Molecule 3 is UNKNOWN LIGAND (three-letter code: UNL) (formula: ).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			10	4	6		
3	B	1	Total	C	O	0	0
			10	4	6		
3	C	1	Total	C	O	0	0
			10	4	6		

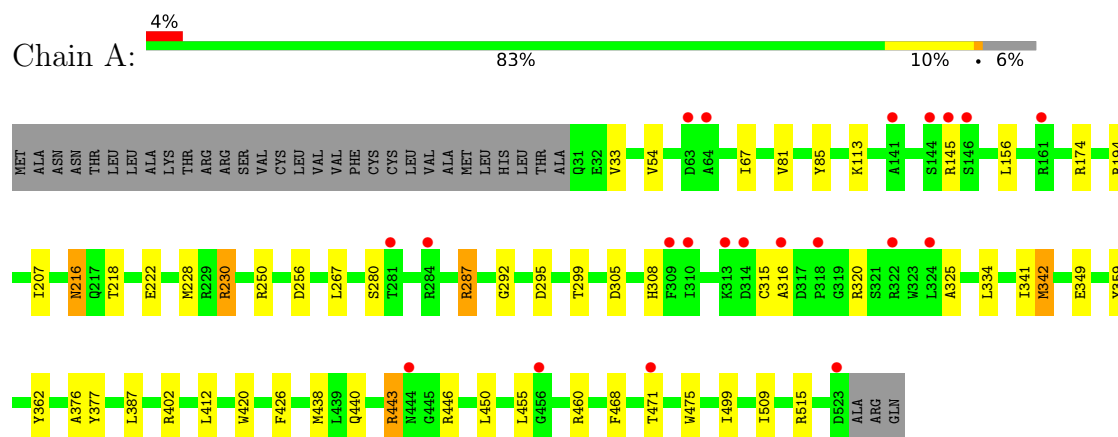
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	239	Total 239	O 239	0	0
4	B	264	Total 264	O 264	0	0
4	C	349	Total 349	O 349	0	0

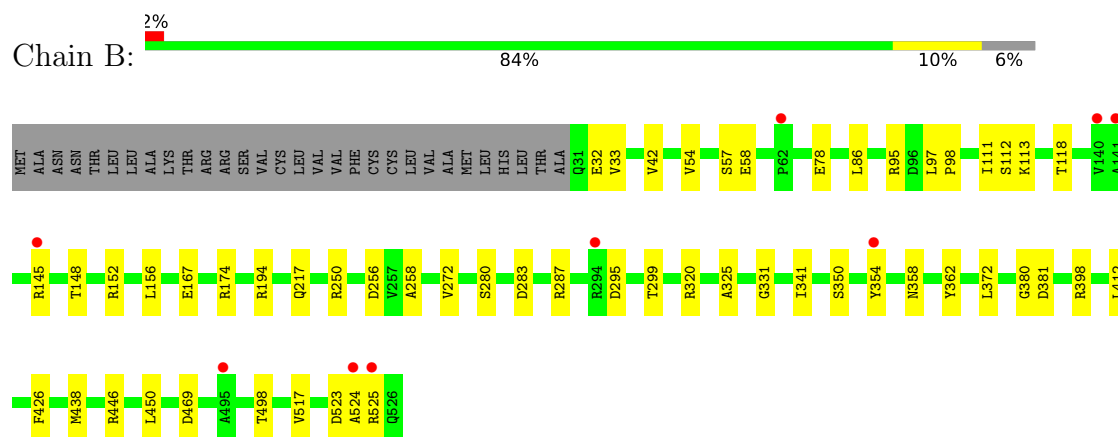
### 3 Residue-property plots [i](#)

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.

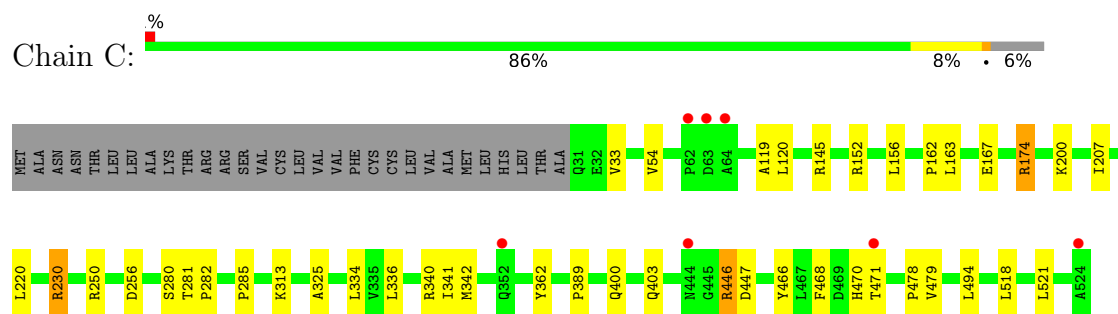
#### • Molecule 1: Sialidase



#### • Molecule 1: Sialidase



#### • Molecule 2: Sialidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	131.16Å 131.16Å 181.17Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	56.86 – 2.10 56.79 – 2.10	Depositor EDS
% Data completeness (in resolution range)	95.9 (56.86-2.10) 95.9 (56.79-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.45 (at 2.10Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, $R_{free}$	0.173 , 0.216 0.178 , 0.213	Depositor DCC
$R_{free}$ test set	5017 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	32.9	Xtriage
Anisotropy	0.030	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 43.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.019 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	12483	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.99% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: UNL

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	0.42	0/3940	0.76	0/5336
1	B	0.41	0/3965	0.78	2/5369 (0.0%)
2	C	0.44	0/3945	0.77	1/5343 (0.0%)
All	All	0.42	0/11850	0.77	3/16048 (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	A	0	5
2	C	0	4
All	All	0	9

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	194	ARG	NE-CZ-NH1	-5.95	117.33	120.30
2	C	340	ARG	NE-CZ-NH1	5.58	123.09	120.30
1	B	95	ARG	NE-CZ-NH2	-5.06	117.77	120.30

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	230	ARG	Sidechain
1	A	250	ARG	Sidechain

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	287	ARG	Sidechain
1	A	455	LEU	Peptide
1	A	85	TYR	Peptide

## 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	3857	0	3763	30	0
1	B	3882	0	3789	26	0
2	C	3862	0	3768	27	0
3	A	10	0	0	1	0
3	B	10	0	0	0	0
3	C	10	0	0	0	0
4	A	239	0	0	2	0
4	B	264	0	0	2	0
4	C	349	0	0	2	0
All	All	12483	0	11320	83	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 4.

The worst 5 of 83 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:216:ASN:HD21	1:A:222:GLU:HG3	1.40	0.84
1:A:315:CYS:SG	1:A:320:ARG:NH1	2.54	0.81
1:A:216:ASN:HD22	1:A:218:THR:H	1.33	0.74
2:C:145:ARG:O	2:C:174:ARG:NH1	2.23	0.72
1:B:354:TYR:CD1	1:B:354:TYR:N	2.64	0.66

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	491/526 (93%)	467 (95%)	23 (5%)	1 (0%)	47	49
1	B	494/526 (94%)	473 (96%)	20 (4%)	1 (0%)	47	49
2	C	492/524 (94%)	471 (96%)	21 (4%)	0	100	100
All	All	1477/1576 (94%)	1411 (96%)	64 (4%)	2 (0%)	51	54

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	524	ALA
1	A	316	ALA

### 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	413/441 (94%)	408 (99%)	5 (1%)	71	77
1	B	415/441 (94%)	407 (98%)	8 (2%)	57	63
2	C	413/439 (94%)	409 (99%)	4 (1%)	76	82
All	All	1241/1321 (94%)	1224 (99%)	17 (1%)	67	73

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	174	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	C	256	ASP
1	B	58	GLU
1	B	112	SER
1	B	145	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	192	ASN
1	A	216	ASN
1	A	308	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 monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 3 are unknown - leaving 0 for Mogul analysis.

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.

## 5.7 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	493/526 (93%)	0.01	21 (4%) 35 41	23, 38, 63, 92	0
1	B	496/526 (94%)	-0.06	9 (1%) 68 72	22, 36, 61, 93	0
2	C	494/524 (94%)	-0.13	7 (1%) 75 78	22, 31, 55, 82	0
All	All	1483/1576 (94%)	-0.06	37 (2%) 57 62	22, 35, 60, 93	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	316	ALA	4.8
1	A	523	ASP	4.3
1	A	318	PRO	4.2
2	C	524	ALA	4.1
1	B	354	TYR	3.9

### 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
3	UNL	C	601	10/-	0.94	0.15	35,41,45,50	0
3	UNL	B	601	10/-	0.96	0.09	36,43,48,49	0
3	UNL	A	601	10/-	0.96	0.09	38,43,48,50	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.