W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 24, 2021 – 12:09 PM EST

PDB ID	:	2P9B
Title	:	Crystal structure of putative prolidase from Bifidobacterium longum
Authors	:	Kumaran, D.; Burley, S.K.; Swaminathan, S.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2007-03-24
Resolution	:	1.70 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

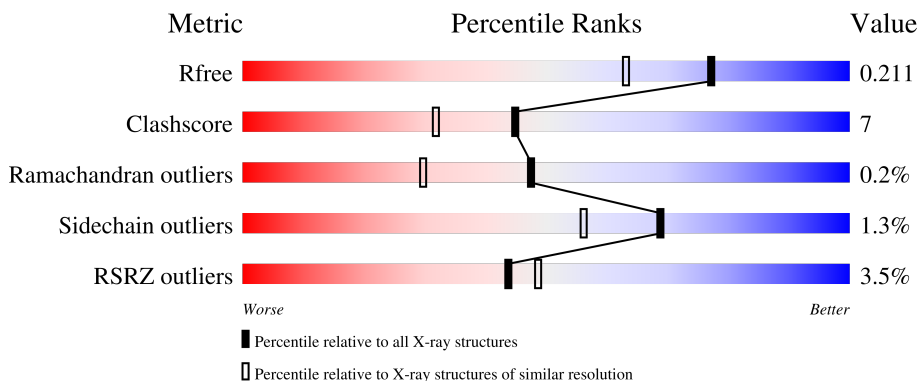|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

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 1.70 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	Percentile Ranks	Value
Rfree		0.211
Clashscore		7
Ramachandran outliers		0.2%
Sidechain outliers		1.3%
RSRZ outliers		3.5%

Worse — *Better*

▮ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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 >=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 <=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	458	3% / 78% 10% · 11%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 3277 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 Possible prolidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	407	2988	1880	518	579	1	10	0	0	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	cloning artifact	UNP Q8G4D5
A	2	SER	-	cloning artifact	UNP Q8G4D5
A	3	LEU	-	cloning artifact	UNP Q8G4D5
A	4	MSE	MET	modified residue	UNP Q8G4D5
A	34	MSE	MET	modified residue	UNP Q8G4D5
A	70	MSE	MET	modified residue	UNP Q8G4D5
A	99	MSE	MET	modified residue	UNP Q8G4D5
A	113	MSE	MET	modified residue	UNP Q8G4D5
A	166	MSE	MET	modified residue	UNP Q8G4D5
A	224	MSE	MET	modified residue	UNP Q8G4D5
A	229	MSE	MET	modified residue	UNP Q8G4D5
A	277	MSE	MET	modified residue	UNP Q8G4D5
A	326	MSE	MET	modified residue	UNP Q8G4D5
A	339	MSE	MET	modified residue	UNP Q8G4D5
A	348	MSE	MET	modified residue	UNP Q8G4D5
A	451	GLU	-	cloning artifact	UNP Q8G4D5
A	452	GLY	-	cloning artifact	UNP Q8G4D5
A	453	HIS	-	cloning artifact	UNP Q8G4D5
A	454	HIS	-	cloning artifact	UNP Q8G4D5
A	455	HIS	-	cloning artifact	UNP Q8G4D5
A	456	HIS	-	cloning artifact	UNP Q8G4D5
A	457	HIS	-	cloning artifact	UNP Q8G4D5
A	458	HIS	-	cloning artifact	UNP Q8G4D5

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	289	289	289	0	0

3 Residue-property plots ⓘ

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: Possible prolidase

Chain A:

4   Data and refinement statistics   ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	150.65Å 63.05Å 56.25Å 90.00° 101.97° 90.00°	Depositor
Resolution (Å)	40.27 – 1.70 40.27 – 1.70	Depositor EDS
% Data completeness (in resolution range)	92.9 (40.27-1.70) 92.6 (40.27-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$<I/\sigma(I)>$ [1]	3.60 (at 1.70Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.193 , 0.212 0.193 , 0.211	Depositor DCC
R_{free} test set	2666 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	16.5	Xtriage
Anisotropy	0.418	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 43.2	EDS
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3277	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.70% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $<|L|>$, $<L^2>$ 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 ⓘ

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	0.29	0/3029	0.61	0/4114

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	2988	0	2999	39	0
2	A	289	0	0	2	0
All	All	3277	0	2999	39	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 7.

The worst 5 of 39 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:258:LEU:HG	1:A:277:MSE:HE1	1.60	0.84
1:A:26:LYS:HD3	1:A:26:LYS:H	1.50	0.76
1:A:26:LYS:CD	1:A:26:LYS:H	2.02	0.72
1:A:166:MSE:HE3	1:A:190:ALA:HB3	1.72	0.70
1:A:166:MSE:HE3	1:A:190:ALA:CB	2.25	0.67

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	401/458 (88%)	391 (98%)	9 (2%)	1 (0%)	47	30

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	266	HIS

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	310/338 (92%)	306 (99%)	4 (1%)	69	56

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	26	LYS
1	A	246	HIS
1	A	248	GLN
1	A	382	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	248	GLN
1	A	407	ASN
1	A	334	HIS
1	A	246	HIS
1	A	306	GLN

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 ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 95th 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^2)	Q<0.9
1	A	397/458 (86%)	0.02	14 (3%)	44	49	9, 17, 35, 47	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	10	ILE	7.6
1	A	450	ALA	5.3
1	A	112	TYR	4.6
1	A	57	ALA	4.6
1	A	58	GLU	4.2

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.