



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 22, 2022 – 08:03 AM EST

PDB ID : 7SHG  
Title : Polysaccharide ribofuranosyl transferase from *Thermobacillus composti*  
Authors : Kimber, M.S.; Kelly, S.D.  
Deposited on : 2021-10-08  
Resolution : 2.50 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  
Xtriage (Phenix) : 1.13  
EDS : 2.26  
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) : 2.26

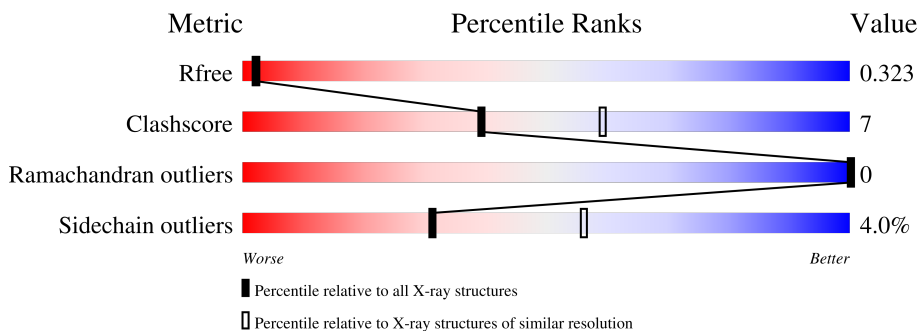
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	648	 72%      18%      • 9%
1	B	648	 70%      20%      • 9%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 10017 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 Ribofuranosyl transferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	591	4923	3191	804	911	17	0	0	0
1	B	590	4912	3182	803	910	17	0	0	0

There are 46 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-24	MET	-	initiating methionine	UNP L0EJI9
A	-23	GLY	-	expression tag	UNP L0EJI9
A	-22	SER	-	expression tag	UNP L0EJI9
A	-21	SER	-	expression tag	UNP L0EJI9
A	-20	HIS	-	expression tag	UNP L0EJI9
A	-19	HIS	-	expression tag	UNP L0EJI9
A	-18	HIS	-	expression tag	UNP L0EJI9
A	-17	HIS	-	expression tag	UNP L0EJI9
A	-16	HIS	-	expression tag	UNP L0EJI9
A	-15	HIS	-	expression tag	UNP L0EJI9
A	-14	SER	-	expression tag	UNP L0EJI9
A	-13	SER	-	expression tag	UNP L0EJI9
A	-12	GLY	-	expression tag	UNP L0EJI9
A	-11	LEU	-	expression tag	UNP L0EJI9
A	-10	VAL	-	expression tag	UNP L0EJI9
A	-9	PRO	-	expression tag	UNP L0EJI9
A	-8	ARG	-	expression tag	UNP L0EJI9
A	-7	GLY	-	expression tag	UNP L0EJI9
A	-6	SER	-	expression tag	UNP L0EJI9
A	-5	HIS	-	expression tag	UNP L0EJI9
A	-4	MET	-	expression tag	UNP L0EJI9
A	-3	ALA	-	expression tag	UNP L0EJI9
A	-2	SER	-	expression tag	UNP L0EJI9
B	-24	MET	-	initiating methionine	UNP L0EJI9
B	-23	GLY	-	expression tag	UNP L0EJI9

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-22	SER	-	expression tag	UNP L0EJI9
B	-21	SER	-	expression tag	UNP L0EJI9
B	-20	HIS	-	expression tag	UNP L0EJI9
B	-19	HIS	-	expression tag	UNP L0EJI9
B	-18	HIS	-	expression tag	UNP L0EJI9
B	-17	HIS	-	expression tag	UNP L0EJI9
B	-16	HIS	-	expression tag	UNP L0EJI9
B	-15	HIS	-	expression tag	UNP L0EJI9
B	-14	SER	-	expression tag	UNP L0EJI9
B	-13	SER	-	expression tag	UNP L0EJI9
B	-12	GLY	-	expression tag	UNP L0EJI9
B	-11	LEU	-	expression tag	UNP L0EJI9
B	-10	VAL	-	expression tag	UNP L0EJI9
B	-9	PRO	-	expression tag	UNP L0EJI9
B	-8	ARG	-	expression tag	UNP L0EJI9
B	-7	GLY	-	expression tag	UNP L0EJI9
B	-6	SER	-	expression tag	UNP L0EJI9
B	-5	HIS	-	expression tag	UNP L0EJI9
B	-4	MET	-	expression tag	UNP L0EJI9
B	-3	ALA	-	expression tag	UNP L0EJI9
B	-2	SER	-	expression tag	UNP L0EJI9

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Cl 1 1	0	0
2	B	1	Total Cl 1 1	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total Mg 2 2	0	0
3	B	1	Total Mg 1 1	0	0

- Molecule 4 is water.

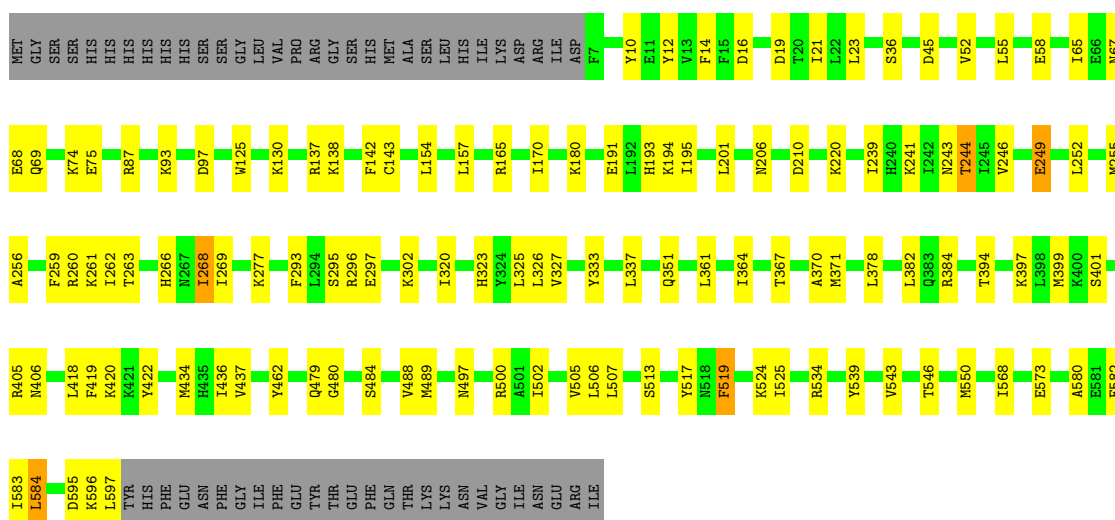
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	91	Total O 91 91	0	0
4	B	86	Total O 86 86	0	0

### 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. 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. 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.

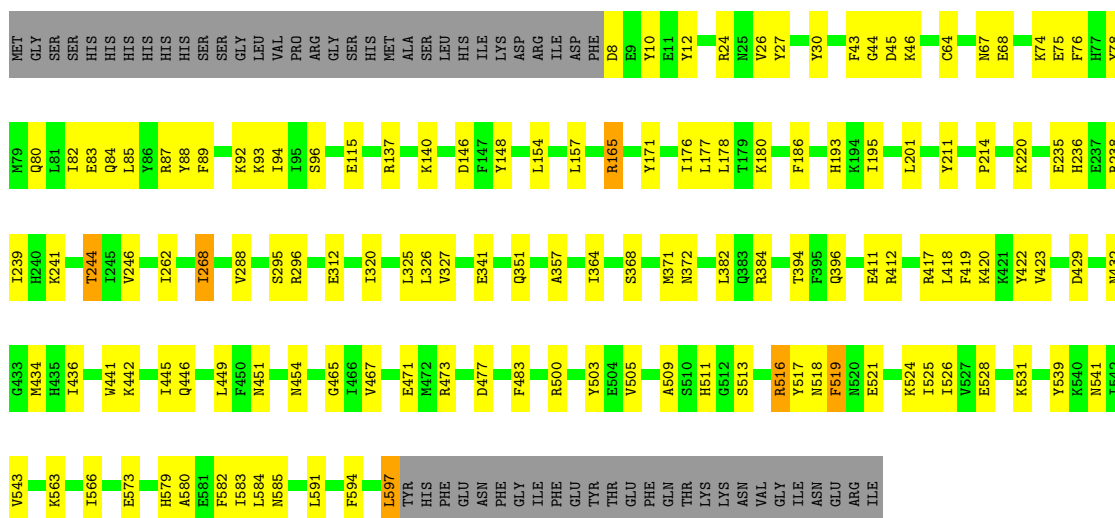
- Molecule 1: Ribofuranosyl transferase

Chain A: 



- Molecule 1: Ribofuranosyl transferase

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.74Å 82.50Å 234.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.16 – 2.50 49.16 – 2.10	Depositor EDS
% Data completeness (in resolution range)	89.6 (49.16-2.50) 64.2 (49.16-2.10)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.21 (at 2.10Å)	Xtrriage
Refinement program	PHENIX 1.19_4092	Depositor
R, $R_{free}$	0.251 , 0.295 0.249 , 0.323	Depositor DCC
$R_{free}$ test set	1775 reflections (2.58%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.7	Xtrriage
Anisotropy	0.686	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 56.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	10017	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	62.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 50.16 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.7334e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: CL, MG

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	0.26	0/5034	0.45	0/6783
1	B	0.26	0/5022	0.45	0/6767
All	All	0.26	0/10056	0.45	0/13550

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	4923	0	4883	70	1
1	B	4912	0	4874	76	1
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	2	0	0	0	0
3	B	1	0	0	0	0
4	A	91	0	0	8	0
4	B	86	0	0	14	0
All	All	10017	0	9757	144	1

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 144 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:371:MET:HB3	1:A:382:LEU:HD21	1.56	0.84
1:B:64:CYS:HB3	1:B:74:LYS:HD3	1.65	0.77
1:A:323:HIS:NE2	4:A:802:HOH:O	2.21	0.72
1:A:437:VAL:HG12	1:A:462:TYR:HB2	1.73	0.70
1:A:268:ILE:HG23	1:A:582:PHE:HZ	1.55	0.69

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:249:GLU:OE2	1:B:140:LYS:NZ[4_664]	2.07	0.13

## 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	589/648 (91%)	562 (95%)	27 (5%)	0	100	100
1	B	588/648 (91%)	560 (95%)	28 (5%)	0	100	100
All	All	1177/1296 (91%)	1122 (95%)	55 (5%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

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	547/598 (92%)	527 (96%)	20 (4%)	34	60
1	B	546/598 (91%)	522 (96%)	24 (4%)	28	52
All	All	1093/1196 (91%)	1049 (96%)	44 (4%)	31	56

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	326	LEU
1	B	454	ASN
1	B	341	GLU
1	B	412	ARG
1	B	519	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	406	ASN
1	B	236	HIS
1	B	351	GLN
1	B	518	ASN

### 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 5 ligands modelled in this entry, 5 are monoatomic - 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

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

### 6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.