



# Full wwPDB X-ray Structure Validation Report ⓘ

May 23, 2020 – 10:08 pm BST

PDB ID : 4ZN4  
Title : Crystal structure of Sqt1 from Chaetomium thermophilum solved by MR  
Authors : Ahmed, Y.L.; Sinning, I.  
Deposited on : 2015-05-04  
Resolution : 1.94 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

<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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

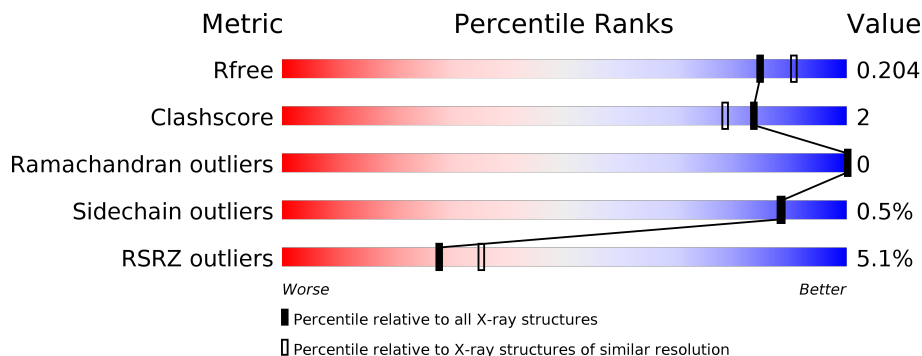
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.94 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	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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 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	541	
1	B	541	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6362 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 sqt1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	409	3016	1901	514	593	8	0	2	0
1	B	411	3024	1907	516	593	8	0	1	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	534	GLY	-	expression tag	UNP G0S0R0
A	535	SER	-	expression tag	UNP G0S0R0
A	536	HIS	-	expression tag	UNP G0S0R0
A	537	HIS	-	expression tag	UNP G0S0R0
A	538	HIS	-	expression tag	UNP G0S0R0
A	539	HIS	-	expression tag	UNP G0S0R0
A	540	HIS	-	expression tag	UNP G0S0R0
A	541	HIS	-	expression tag	UNP G0S0R0
B	534	GLY	-	expression tag	UNP G0S0R0
B	535	SER	-	expression tag	UNP G0S0R0
B	536	HIS	-	expression tag	UNP G0S0R0
B	537	HIS	-	expression tag	UNP G0S0R0
B	538	HIS	-	expression tag	UNP G0S0R0
B	539	HIS	-	expression tag	UNP G0S0R0
B	540	HIS	-	expression tag	UNP G0S0R0
B	541	HIS	-	expression tag	UNP G0S0R0

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	3	3		
2	A	1	Total	C	O	0	0
			6	3	3		

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	A	156	Total 156	O 156	0	0
4	B	149	Total 149	O 149	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	111.57Å 94.82Å 94.95Å 90.00° 109.90° 90.00°	Depositor
Resolution (Å)	47.41 – 1.94 47.41 – 1.94	Depositor EDS
% Data completeness (in resolution range)	98.5 (47.41-1.94) 97.6 (47.41-1.94)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.56 (at 1.94Å)	Xtrriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.171 , 0.205 0.173 , 0.204	Depositor DCC
$R_{free}$ test set	3398 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	33.6	Xtrriage
Anisotropy	0.290	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 42.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	6362	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.54% 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: GOL, SO4

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	0.43	0/3081	0.61	2/4197 (0.0%)
1	B	0.43	0/3090	0.59	0/4208
All	All	0.43	0/6171	0.60	2/8405 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	231	GLN	CA-CB-CG	5.97	126.53	113.40
1	A	53	LEU	CA-CB-CG	5.01	126.83	115.30

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	3016	0	2917	17	0
1	B	3024	0	2928	13	0
2	A	12	0	16	0	0
3	A	5	0	0	0	0
4	A	156	0	0	3	0
4	B	149	0	0	1	0
All	All	6362	0	5861	29	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 2.

All (29) 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:B:470:PRO:HB2	1:B:472:MET:HB3	1.64	0.79
1:B:500:SER:OG	1:B:503:GLU:OE1	2.02	0.77
1:A:470:PRO:HB2	1:A:472:MET:HB3	1.68	0.75
1:A:415:ARG:HD2	4:A:734:HOH:O	1.99	0.61
1:B:143:ASN:ND2	1:B:160:MET:SD	2.70	0.59
1:B:366:GLN:HG2	1:B:367:ALA:N	2.18	0.57
1:A:366:GLN:N	4:A:702:HOH:O	2.43	0.51
1:A:102:ILE:HD13	1:A:520:VAL:HG21	1.93	0.51
1:A:143:ASN:ND2	1:A:160:MET:SD	2.79	0.48
1:A:425:PRO:HD2	1:A:450:MET:HB2	1.96	0.47
1:A:174:ASN:N	1:A:174:ASN:OD1	2.45	0.47
1:A:487:ALA:HB3	1:A:488:PRO:HD3	1.96	0.47
1:B:233:VAL:HG12	1:B:234[B]:GLN:HG2	1.97	0.46
1:A:470:PRO:C	1:A:472:MET:H	2.16	0.46
1:A:71:ILE:HD12	1:A:522:VAL:HG13	1.98	0.45
1:B:99:VAL:HG11	1:B:180:PHE:CZ	2.51	0.45
1:B:96:LYS:NZ	1:B:136:ASP:OD1	2.41	0.45
1:A:331:LEU:O	1:A:332:SER:HB3	2.17	0.44
1:B:326:VAL:HG22	1:B:370:ILE:HD13	2.01	0.43
1:A:77:TYR:HB2	1:A:80:ILE:HD12	2.01	0.43
1:A:470:PRO:C	1:A:472:MET:N	2.72	0.43
1:A:368:GLY:HA3	4:A:841:HOH:O	2.19	0.41
1:B:428:LYS:NZ	4:B:605:HOH:O	2.49	0.41
1:B:377:GLN:HG2	1:B:401:SER:OG	2.21	0.41
1:A:233:VAL:O	1:B:409:ARG:HD2	2.21	0.41
1:B:323:ILE:HB	1:B:374:LEU:HB2	2.02	0.41
1:A:181:LYS:O	1:A:183:LEU:HD13	2.21	0.41
1:B:425:PRO:HD2	1:B:450:MET:HB2	2.02	0.41
1:A:77:TYR:HA	1:A:78:PRO:HD2	1.96	0.40

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	401/541 (74%)	391 (98%)	10 (2%)	0	100	100
1	B	402/541 (74%)	393 (98%)	9 (2%)	0	100	100
All	All	803/1082 (74%)	784 (98%)	19 (2%)	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	319/430 (74%)	316 (99%)	3 (1%)	78	75
1	B	319/430 (74%)	318 (100%)	1 (0%)	92	93
All	All	638/860 (74%)	634 (99%)	4 (1%)	88	85

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

Mol	Chain	Res	Type
1	A	183	LEU
1	A	489[A]	SER
1	A	489[B]	SER
1	B	187	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	187	GLN

### 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](#)

3 ligands 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	GOL	A	602	-	5,5,5	0.43	0	5,5,5	0.43	0
2	GOL	A	601	-	5,5,5	0.43	0	5,5,5	0.55	0
3	SO4	A	603	-	4,4,4	0.15	0	6,6,6	0.18	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	GOL	A	602	-	-	2/4/4/4	-
2	GOL	A	601	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	601	GOL	O1-C1-C2-O2
2	A	601	GOL	O1-C1-C2-C3
2	A	602	GOL	O1-C1-C2-O2
2	A	602	GOL	O1-C1-C2-C3

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

### 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	409/541 (75%)	0.19	24 (5%) 22 28	22, 35, 65, 94	0
1	B	411/541 (75%)	0.13	18 (4%) 34 41	20, 35, 64, 87	0
All	All	820/1082 (75%)	0.16	42 (5%) 28 35	20, 35, 65, 94	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	53	LEU	9.8
1	A	127	THR	8.0
1	A	471	HIS	7.4
1	B	127	THR	7.3
1	A	177	LEU	7.1
1	A	52	ALA	6.2
1	B	488	PRO	5.6
1	A	487	ALA	5.5
1	A	104	ALA	5.2
1	A	53	LEU	4.8
1	B	534	GLY	4.7
1	B	227	SER	4.6
1	A	366	GLN	4.5
1	B	471	HIS	4.4
1	B	54	GLU	4.4
1	A	175	GLY	3.8
1	A	295	ARG	3.8
1	A	174	ASN	3.6
1	B	472	MET	3.5
1	A	176	ALA	3.5
1	B	228	ASN	3.3
1	B	502	GLN	3.2
1	A	62	ASP	3.1
1	A	281	ASP	3.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	503	GLU	3.0
1	A	294	GLN	2.9
1	A	534	GLY	2.9
1	A	91	ASP	2.8
1	B	128	SER	2.8
1	B	533	ALA	2.7
1	A	472	MET	2.7
1	A	128[A]	SER	2.7
1	B	330	ARG	2.6
1	B	499	ARG	2.5
1	B	226	PRO	2.4
1	A	228	ASN	2.4
1	A	102	ILE	2.4
1	A	367	ALA	2.3
1	A	173	GLN	2.3
1	B	506	GLY	2.2
1	A	129	LEU	2.1
1	B	91	ASP	2.1

## 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(Å <sup>2</sup> )	Q<0.9
3	SO4	A	603	5/5	0.91	0.36	55,55,56,57	0
2	GOL	A	601	6/6	0.94	0.08	26,29,33,35	0
2	GOL	A	602	6/6	0.96	0.11	28,31,35,38	0

## 6.5 Other polymers

There are no such residues in this entry.