Full wwPDB EM Validation Report ⓘ

Sep 7, 2023 – 04:24 pm BST

PDB ID	:	6ZHN
Title	:	3D electron diffraction structure of thaumatin from Thaumatococcus daniellii
Authors	:	Blum, T.; Housset, D.; Clabbers, M.T.B.; van Genderen, E.; Schoehn, G.; Ling, W.L.; Abrahams, J.P.
Deposited on	:	2020-06-23
Resolution	:	2.76 Å(reported)

This is a Full wwPDB EM Validation Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
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:

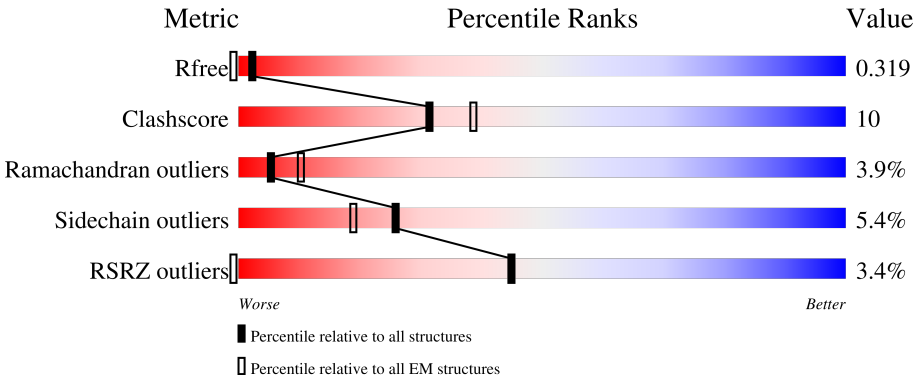
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON CRYSTALLOGRAPHY

The reported resolution of this entry is 2.76 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)	EM structures (#Entries)
R_{free}	130704	0
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826
RSRZ outliers	127900	0

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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%

Mol	Chain	Length	Quality of chain
1	A	207	3% 75% 22% .

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1551 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Thaumatin-1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	207	1550	963	271	299	17	0	0

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

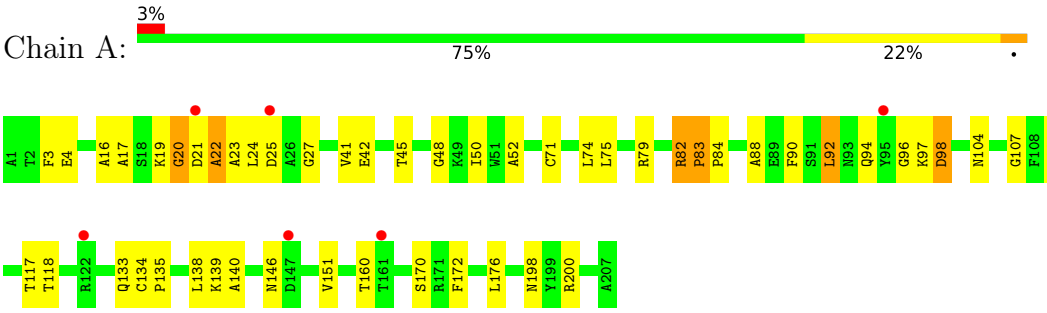
Mol	Chain	Residues	Atoms		AltConf
			Total	Cl	
2	A	1	1	1	0

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: Thaumatin-1

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	57.72Å 57.72Å 149.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.69 – 2.76 45.65 – 2.76	Depositor EDS
% Data completeness (in resolution range)	65.6 (45.69-2.76) 65.6 (45.65-2.76)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.48 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.8.0230	Depositor
R, R_{free}	0.280 , 0.321 0.285 , 0.319	Depositor DCC
R_{free} test set	231 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	45.4	Xtriage
Anisotropy	0.017	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.18 , -10.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.83	EDS
Total number of atoms	1551	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.61% 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

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

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	0.34	0/1586	0.58	0/2150

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	1550	0	1474	29	1
2	A	1	0	0	1	0
All	All	1551	0	1474	29	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 10.

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:A:82:ARG:NH1	2:A:301:CL:CL	2.61	0.70
1:A:52:ALA:O	1:A:109:ASN:ND2	2.30	0.65

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:17:ALA:HB1	1:A:24:LEU:HD12	1.87	0.57
1:A:133:GLN:O	1:A:172:PHE:HE1	1.88	0.56
1:A:17:ALA:HB2	1:A:50:ILE:HG12	1.90	0.54
1:A:88:ALA:HB1	1:A:114:PHE:CD2	2.46	0.51
1:A:19:LYS:O	1:A:20:GLY:C	2.48	0.51
1:A:21:ASP:O	1:A:22:ALA:HB2	2.11	0.50
1:A:4:GLU:OE1	1:A:200:ARG:NH1	2.44	0.50
1:A:133:GLN:O	1:A:172:PHE:CE1	2.66	0.49
1:A:42:GLU:O	1:A:45:THR:HG23	2.14	0.48
1:A:74:LEU:O	1:A:75:LEU:C	2.53	0.47
1:A:146:ASN:HB3	1:A:151:VAL:HG23	1.96	0.47
1:A:92:LEU:O	1:A:98:ASP:HA	2.16	0.46
1:A:88:ALA:HB1	1:A:114:PHE:CE2	2.51	0.46
1:A:138:LEU:O	1:A:140:ALA:N	2.50	0.45
1:A:23:ALA:HB1	1:A:27:GLY:HA2	1.97	0.45
1:A:48:GLY:O	1:A:90:PHE:N	2.50	0.45
1:A:135:PRO:HD3	1:A:172:PHE:CZ	2.51	0.44
1:A:24:LEU:HD22	1:A:41:VAL:HG13	1.98	0.44
1:A:71:CYS:SG	1:A:71:CYS:O	2.75	0.44
1:A:41:VAL:HG12	1:A:92:LEU:HD11	2.00	0.43
1:A:3:PHE:CD2	1:A:50:ILE:CD1	3.02	0.43
1:A:96:GLY:O	1:A:97:LYS:HD2	2.19	0.42
1:A:104:ASN:HA	1:A:107:GLY:O	2.20	0.42
1:A:16:ALA:O	1:A:50:ILE:HA	2.20	0.41
1:A:146:ASN:HB3	1:A:151:VAL:CG2	2.51	0.41
1:A:82:ARG:O	1:A:83:PRO:C	2.60	0.41
1:A:134:CYS:HA	1:A:172:PHE:CZ	2.57	0.40

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:79:ARG:NH1	1:A:79:ARG:NH1[7_555]	1.96	0.24

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 PDB entries followed by that with respect to all EM

entries.

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	205/207 (99%)	159 (78%)	38 (18%)	8 (4%)	3 4

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	20	GLY
1	A	139	LYS
1	A	22	ALA
1	A	25	ASP
1	A	83	PRO
1	A	92	LEU
1	A	84	PRO
1	A	94	GLN

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 PDB entries followed by that with respect to all EM entries.

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	167/167 (100%)	158 (95%)	9 (5%)	22 38

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	82	ARG
1	A	98	ASP
1	A	114	PHE
1	A	117	THR
1	A	118	THR
1	A	160	THR
1	A	170	SER
1	A	176	LEU
1	A	198	ASN

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

Mol	Chain	Res	Type
1	A	46	ASN

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

Of 1 ligands modelled in this entry, 1 is 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.