Full wwPDB X-ray Structure Validation Report ⓘ

Jun 3, 2024 – 03:13 PM EDT

PDB ID	:	8V6Q
Title	:	Crystal structure of EcThsA in ligand-free state
Authors	:	Shi, Y.; Masic, V.; Mosaiab, T.; Goulart, C.C.; Hartley-Tassell, L.; Sorbello, M.; Vasquez, V.; Mishra, B.P.; Holt, S.; Gu, W.; Kobe, B.; Ve, T.
Deposited on	:	2023-12-02
Resolution	:	2.62 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*
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:

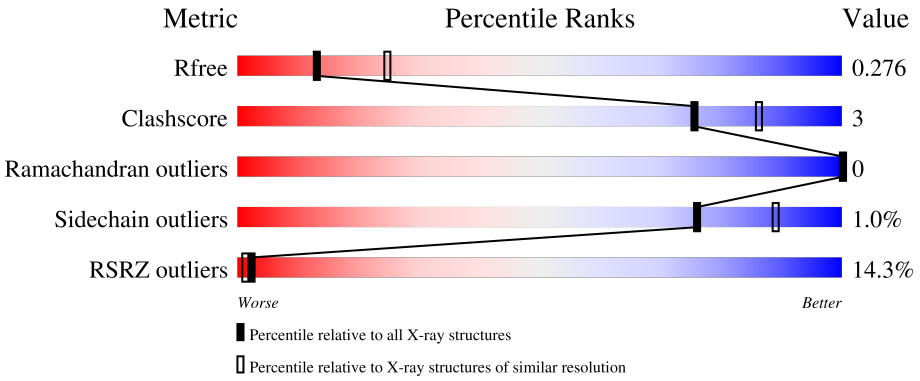
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.62 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	3797 (2.64-2.60)
Clashscore	141614	4168 (2.64-2.60)
Ramachandran outliers	138981	4093 (2.64-2.60)
Sidechain outliers	138945	4093 (2.64-2.60)
RSRZ outliers	127900	3731 (2.64-2.60)

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	241	10% 81% 7% 11%
1	B	241	15% 79% 7% 14%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6898 atoms, of which 3452 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 Thoeris protein ThsA Macro domain-containing protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	214	Total	C	H	N	O	S	0	0	0
			3502	1125	1754	294	326	3			
1	B	207	Total	C	H	N	O	S	0	0	0
			3394	1095	1698	282	316	3			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	38	MET	-	initiating methionine	UNP A0A178STJ6
A	39	HIS	-	expression tag	UNP A0A178STJ6
A	40	HIS	-	expression tag	UNP A0A178STJ6
A	41	HIS	-	expression tag	UNP A0A178STJ6
A	42	HIS	-	expression tag	UNP A0A178STJ6
A	43	HIS	-	expression tag	UNP A0A178STJ6
A	44	HIS	-	expression tag	UNP A0A178STJ6
A	45	SER	-	expression tag	UNP A0A178STJ6
A	46	SER	-	expression tag	UNP A0A178STJ6
A	47	GLY	-	expression tag	UNP A0A178STJ6
A	48	VAL	-	expression tag	UNP A0A178STJ6
A	49	ASP	-	expression tag	UNP A0A178STJ6
A	50	LEU	-	expression tag	UNP A0A178STJ6
A	51	GLY	-	expression tag	UNP A0A178STJ6
A	52	THR	-	expression tag	UNP A0A178STJ6
A	53	GLU	-	expression tag	UNP A0A178STJ6
A	54	ASN	-	expression tag	UNP A0A178STJ6
A	55	LEU	-	expression tag	UNP A0A178STJ6
A	56	TYR	-	expression tag	UNP A0A178STJ6
A	57	PHE	-	expression tag	UNP A0A178STJ6
A	58	GLN	-	expression tag	UNP A0A178STJ6
A	59	SER	-	expression tag	UNP A0A178STJ6
A	60	ASN	-	expression tag	UNP A0A178STJ6
A	61	ALA	-	expression tag	UNP A0A178STJ6
B	38	MET	-	initiating methionine	UNP A0A178STJ6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	39	HIS	-	expression tag	UNP A0A178STJ6
B	40	HIS	-	expression tag	UNP A0A178STJ6
B	41	HIS	-	expression tag	UNP A0A178STJ6
B	42	HIS	-	expression tag	UNP A0A178STJ6
B	43	HIS	-	expression tag	UNP A0A178STJ6
B	44	HIS	-	expression tag	UNP A0A178STJ6
B	45	SER	-	expression tag	UNP A0A178STJ6
B	46	SER	-	expression tag	UNP A0A178STJ6
B	47	GLY	-	expression tag	UNP A0A178STJ6
B	48	VAL	-	expression tag	UNP A0A178STJ6
B	49	ASP	-	expression tag	UNP A0A178STJ6
B	50	LEU	-	expression tag	UNP A0A178STJ6
B	51	GLY	-	expression tag	UNP A0A178STJ6
B	52	THR	-	expression tag	UNP A0A178STJ6
B	53	GLU	-	expression tag	UNP A0A178STJ6
B	54	ASN	-	expression tag	UNP A0A178STJ6
B	55	LEU	-	expression tag	UNP A0A178STJ6
B	56	TYR	-	expression tag	UNP A0A178STJ6
B	57	PHE	-	expression tag	UNP A0A178STJ6
B	58	GLN	-	expression tag	UNP A0A178STJ6
B	59	SER	-	expression tag	UNP A0A178STJ6
B	60	ASN	-	expression tag	UNP A0A178STJ6
B	61	ALA	-	expression tag	UNP A0A178STJ6

- Molecule 2 is water.

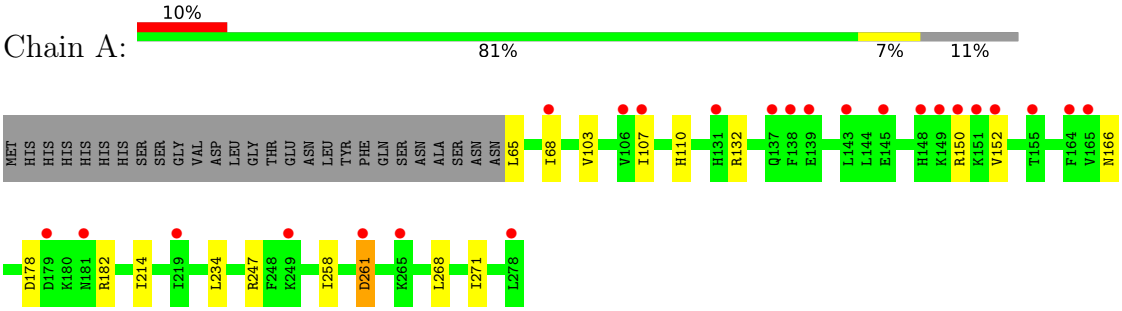
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O 1 1	0	0
2	B	1	Total O 1 1	0	0

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.

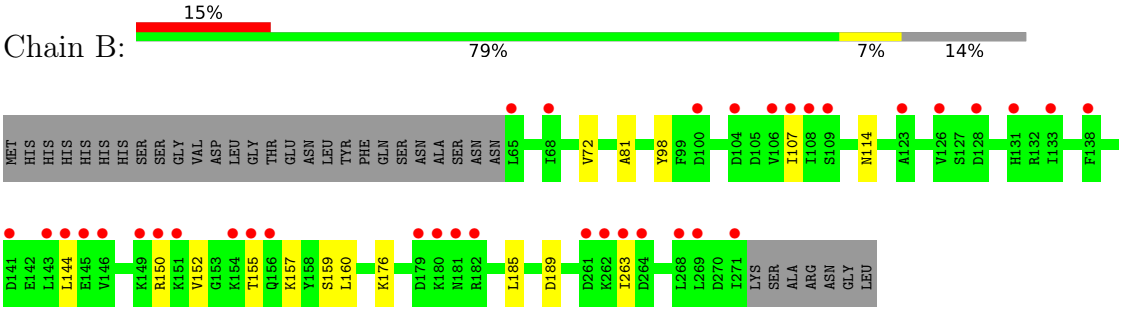
• Molecule 1: Thoeris protein ThsA Macro domain-containing protein

Chain A:



• Molecule 1: Thoeris protein ThsA Macro domain-containing protein

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	40.04Å 93.54Å 140.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.38 – 2.62 44.38 – 2.62	Depositor EDS
% Data completeness (in resolution range)	99.4 (44.38-2.62) 99.4 (44.38-2.62)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.05 (at 2.61Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.241 , 0.269 0.249 , 0.276	Depositor DCC
R_{free} test set	821 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	57.3	Xtriage
Anisotropy	0.364	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.44 , 46.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	6898	wwPDB-VP
Average B, all atoms (Å^2)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.29% 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	0.27	0/1782	0.48	0/2401
1	B	0.26	0/1730	0.48	0/2334
All	All	0.27	0/3512	0.48	0/4735

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	1748	1754	1753	10	0
1	B	1696	1698	1697	11	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
All	All	3446	3452	3450	21	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 3.

All (21) 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:107:ILE:HD11	1:A:152:VAL:HG12	1.49	0.91

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:81:ALA:HB2	1:B:263:ILE:HG21	1.57	0.85
1:A:132:ARG:NH1	1:A:166:ASN:OD1	2.33	0.60
1:A:214:ILE:HD11	1:A:258:ILE:HG12	1.85	0.58
1:A:234:LEU:HD23	1:A:271:ILE:CD1	2.38	0.53
1:A:68:ILE:HD12	1:A:268:LEU:HB3	1.91	0.51
1:B:160:LEU:HD11	1:B:176:LYS:HE3	1.94	0.49
1:B:107:ILE:HD12	1:B:152:VAL:HG23	1.96	0.47
1:B:150:ARG:HB3	1:B:155:THR:HG22	1.96	0.46
1:A:178:ASP:OD2	1:A:182:ARG:NE	2.49	0.46
1:B:72:VAL:O	1:B:72:VAL:HG23	2.15	0.45
1:B:98:TYR:CE1	1:B:157:LYS:HG2	2.53	0.44
1:B:81:ALA:CB	1:B:263:ILE:HD13	2.48	0.44
1:A:65:LEU:HD21	1:A:68:ILE:HD11	2.01	0.43
1:B:81:ALA:HB2	1:B:263:ILE:HD13	2.01	0.42
1:A:103:VAL:HG12	1:A:110:HIS:CD2	2.55	0.41
1:A:261:ASP:OD1	1:A:261:ASP:N	2.52	0.41
1:B:144:LEU:HD11	1:B:159:SER:HA	2.03	0.41
1:A:107:ILE:CD1	1:A:152:VAL:HG12	2.34	0.40
1:B:81:ALA:HB2	1:B:263:ILE:CG2	2.40	0.40
1:B:185:LEU:HD22	1:B:189:ASP:HB3	2.02	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	212/241 (88%)	205 (97%)	7 (3%)	0	100	100
1	B	205/241 (85%)	188 (92%)	17 (8%)	0	100	100
All	All	417/482 (86%)	393 (94%)	24 (6%)	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	195/219 (89%)	192 (98%)	3 (2%)	65 82
1	B	190/219 (87%)	189 (100%)	1 (0%)	88 95
All	All	385/438 (88%)	381 (99%)	4 (1%)	76 89

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

Mol	Chain	Res	Type
1	A	150	ARG
1	A	247	ARG
1	A	261	ASP
1	B	114	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

There are no ligands in this entry.

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^{th} 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	214/241 (88%)	0.94	24 (11%)	5	3	40, 60, 84, 98	0
1	B	207/241 (85%)	1.08	36 (17%)	1	0	40, 61, 93, 106	0
All	All	421/482 (87%)	1.01	60 (14%)	2	1	40, 61, 89, 106	0

All (60) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	181	ASN	5.4
1	A	137	GLN	5.3
1	A	138	PHE	5.2
1	A	152	VAL	5.1
1	A	181	ASN	5.1
1	A	149	LYS	4.7
1	B	179	ASP	4.6
1	B	269	LEU	4.6
1	B	149	LYS	4.4
1	B	138	PHE	4.2
1	B	264	ASP	4.2
1	A	145	GLU	4.1
1	B	65	LEU	4.0
1	B	133	ILE	4.0
1	B	68	ILE	3.9
1	A	150	ARG	3.7
1	B	262	LYS	3.6
1	A	151	LYS	3.6
1	B	107	ILE	3.4
1	A	143	LEU	3.3
1	B	150	ARG	3.3
1	B	268	LEU	2.9
1	B	126	VAL	2.9
1	B	261	ASP	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	107	ILE	2.8
1	B	271	ILE	2.8
1	A	249	LYS	2.8
1	B	143	LEU	2.8
1	B	128	ASP	2.7
1	A	278	LEU	2.7
1	B	151	LYS	2.6
1	B	123	ALA	2.6
1	B	141	ASP	2.6
1	B	145	GLU	2.5
1	A	219	ILE	2.5
1	B	100	ASP	2.5
1	B	106	VAL	2.5
1	A	155	THR	2.5
1	B	144	LEU	2.5
1	B	131	HIS	2.5
1	B	182	ARG	2.4
1	A	106	VAL	2.4
1	B	263	ILE	2.4
1	B	156	GLN	2.4
1	A	68	ILE	2.4
1	A	265	LYS	2.4
1	A	164	PHE	2.4
1	B	146	VAL	2.4
1	A	139	GLU	2.3
1	B	180	LYS	2.3
1	B	154	LYS	2.3
1	A	131	HIS	2.3
1	B	109	SER	2.2
1	A	261	ASP	2.1
1	B	104	ASP	2.1
1	A	148	HIS	2.1
1	B	108	ILE	2.1
1	B	155	THR	2.0
1	A	165	VAL	2.0
1	A	179	ASP	2.0

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.