



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 5, 2017 – 04:22 PM EDT

PDB ID : 5N1A
Title : Crystal structure of Utp4 from Chaetomium thermophilum
Authors : Calvino, F.R.; Ahmed, Y.L.; Wild, K.; Sinning, I.
Deposited on : 2017-02-05
Resolution : 2.15 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

A user guide is available at

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	rb-20029077
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20029077

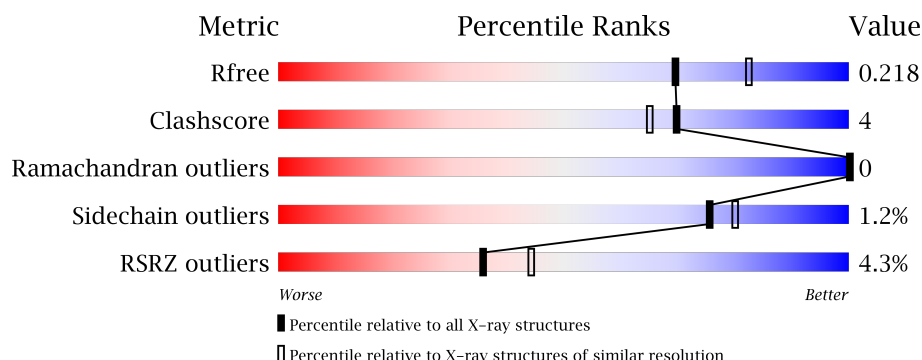
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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}	100719	1170 (2.16-2.16)
Clashscore	112137	1278 (2.16-2.16)
Ramachandran outliers	110173	1256 (2.16-2.16)
Sidechain outliers	110143	1255 (2.16-2.16)
RSRZ outliers	101464	1175 (2.16-2.16)

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 ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	908	<div> <div>3%</div> <div>68%</div> <div>9%</div> <div>23%</div> </div>
1	B	908	<div> <div>4%</div> <div>69%</div> <div>8%</div> <div>22%</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11717 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 utp4.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	703	Total	C	N	O	S	Se	0	1	0
			5527	3517	994	993	8	15			
1	B	706	Total	C	N	O	S	Se	0	1	0
			5542	3525	996	998	8	15			

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-23	MSE	-	initiating methionine	UNP G0SCT7
A	-22	GLY	-	expression tag	UNP G0SCT7
A	-21	HIS	-	expression tag	UNP G0SCT7
A	-20	HIS	-	expression tag	UNP G0SCT7
A	-19	HIS	-	expression tag	UNP G0SCT7
A	-18	HIS	-	expression tag	UNP G0SCT7
A	-17	HIS	-	expression tag	UNP G0SCT7
A	-16	HIS	-	expression tag	UNP G0SCT7
A	-15	ASP	-	expression tag	UNP G0SCT7
A	-14	TYR	-	expression tag	UNP G0SCT7
A	-13	ASP	-	expression tag	UNP G0SCT7
A	-12	ILE	-	expression tag	UNP G0SCT7
A	-11	PRO	-	expression tag	UNP G0SCT7
A	-10	THR	-	expression tag	UNP G0SCT7
A	-9	THR	-	expression tag	UNP G0SCT7
A	-8	GLU	-	expression tag	UNP G0SCT7
A	-7	ASN	-	expression tag	UNP G0SCT7
A	-6	LEU	-	expression tag	UNP G0SCT7
A	-5	TYR	-	expression tag	UNP G0SCT7
A	-4	PHE	-	expression tag	UNP G0SCT7
A	-3	GLN	-	expression tag	UNP G0SCT7
A	-2	GLY	-	expression tag	UNP G0SCT7
A	-1	ALA	-	expression tag	UNP G0SCT7
A	0	HIS	-	expression tag	UNP G0SCT7
A	1	MSE	-	expression tag	UNP G0SCT7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-23	MSE	-	initiating methionine	UNP G0SCT7
B	-22	GLY	-	expression tag	UNP G0SCT7
B	-21	HIS	-	expression tag	UNP G0SCT7
B	-20	HIS	-	expression tag	UNP G0SCT7
B	-19	HIS	-	expression tag	UNP G0SCT7
B	-18	HIS	-	expression tag	UNP G0SCT7
B	-17	HIS	-	expression tag	UNP G0SCT7
B	-16	HIS	-	expression tag	UNP G0SCT7
B	-15	ASP	-	expression tag	UNP G0SCT7
B	-14	TYR	-	expression tag	UNP G0SCT7
B	-13	ASP	-	expression tag	UNP G0SCT7
B	-12	ILE	-	expression tag	UNP G0SCT7
B	-11	PRO	-	expression tag	UNP G0SCT7
B	-10	THR	-	expression tag	UNP G0SCT7
B	-9	THR	-	expression tag	UNP G0SCT7
B	-8	GLU	-	expression tag	UNP G0SCT7
B	-7	ASN	-	expression tag	UNP G0SCT7
B	-6	LEU	-	expression tag	UNP G0SCT7
B	-5	TYR	-	expression tag	UNP G0SCT7
B	-4	PHE	-	expression tag	UNP G0SCT7
B	-3	GLN	-	expression tag	UNP G0SCT7
B	-2	GLY	-	expression tag	UNP G0SCT7
B	-1	ALA	-	expression tag	UNP G0SCT7
B	0	HIS	-	expression tag	UNP G0SCT7
B	1	MSE	-	expression tag	UNP G0SCT7

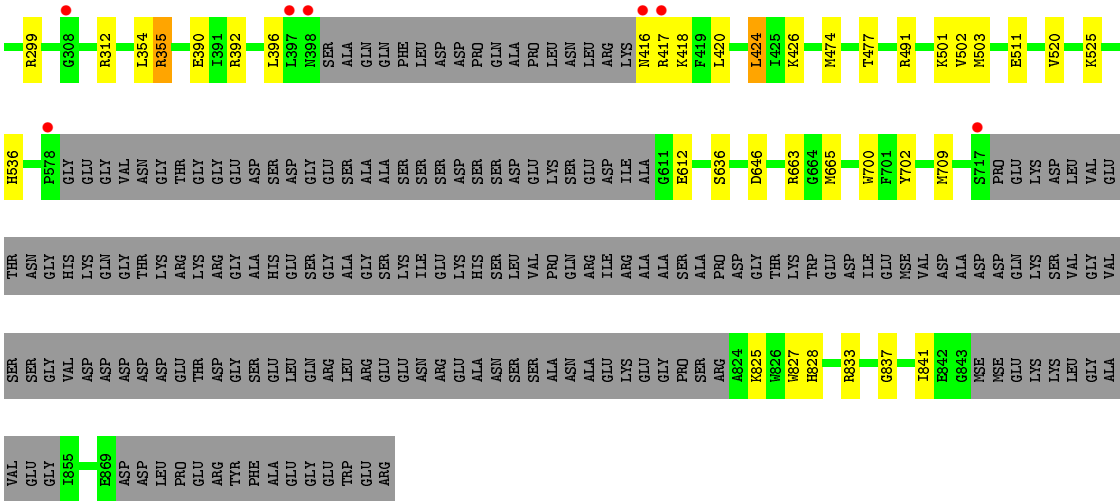
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	347	Total O 347 347	0	0
2	B	301	Total O 301 301	0	0

These plots are drawn for all protein, RNA and DNA 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 ($\text{RSRZ} > 2$). 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.

- [illegible]

- Chain B:
-
- 69% 8% 22%
- 4%
- 0.10
0.08
0.06
0.04
0.02
0.00
- bits
- 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
- MSE GLY HIS HIS HIS HIS ASP TYR ASP ILE P-11 T-10 N-7 F-4 D2 V19 V28 V29 S30 S31 S32 LYS LYS L36 Q37 K38 Q41 L44 L59 L62 W82 E89 M90 A91 D92 G93 F94 L102 Y107 K118 G134 F136
- L140 L141 P141 K143 A144 M145 A146 A147 A148 A149 Q150 M151 R152 K153 N162 Q176 R181 T182 S184 R185 I192 T198 V202 R215 M223 G226 ASP ASP LEU LEU GLY GLY S233 K234 L244 D248 I249 Q257 V261 R271 S274 P278



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	203.03Å 81.59Å 112.30Å 90.00° 110.55° 90.00°	Depositor
Resolution (Å)	47.53 – 2.15 47.53 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.53-2.15) 99.9 (47.53-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.44 (at 2.16Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.183 , 0.220 0.178 , 0.218	Depositor DCC
R_{free} test set	4680 reflections (5.01%)	DCC
Wilson B-factor (Å ²)	35.2	Xtriage
Anisotropy	0.220	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 46.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	11717	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.22% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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	0/5644	0.58	6/7626 (0.1%)
1	B	0.29	0/5663	0.66	12/7653 (0.2%)
All	All	0.28	0/11307	0.62	18/15279 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	B	38	LYS	CA-CB-CG	17.05	150.91	113.40
1	B	234	LYS	CA-CB-CG	11.19	138.02	113.40
1	B	312	ARG	CB-CG-CD	-9.77	86.20	111.60
1	B	38	LYS	CB-CA-C	-9.39	91.61	110.40
1	B	424	LEU	CB-CG-CD1	-9.33	95.13	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	417	ARG	Sidechain
1	B	511	GLU	Sidechain

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	5527	0	5553	51	1
1	B	5542	0	5550	42	2
2	A	347	0	0	7	1
2	B	301	0	0	8	0
All	All	11717	0	11103	92	2

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 4.

The worst 5 of 92 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:HIS:CD2	1:A:483:ARG:NH1	1.74	1.53
1:A:371:HIS:CD2	1:A:483:ARG:HH11	1.51	1.12
1:A:371:HIS:CD2	1:A:483:ARG:HH12	1.70	0.93
1:A:371:HIS:HD2	1:A:483:ARG:NH1	1.53	0.91
1:B:424:LEU:HD12	1:B:426:LYS:HE2	1.56	0.86

All (2) 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:128:GLN:NE2	1:B:151:ASN:OD1[4_546]	2.11	0.09
1:B:143:LYS:NZ	2:A:911:HOH:O[4_556]	2.16	0.04

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	690/908 (76%)	680 (99%)	10 (1%)	0	100	100
1	B	691/908 (76%)	676 (98%)	15 (2%)	0	100	100
All	All	1381/1816 (76%)	1356 (98%)	25 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	595/740 (80%)	588 (99%)	7 (1%)	75	80
1	B	596/740 (80%)	589 (99%)	7 (1%)	75	80
All	All	1191/1480 (80%)	1177 (99%)	14 (1%)	75	80

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	833	ARG
1	B	135	PHE
1	B	355	ARG
1	A	532	GLN
1	B	278	ASP

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

5.3.3 RNA ⓘ

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](#)

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 [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	688/908 (75%)	0.05	26 (3%)	41 48	22, 37, 68, 115	1 (0%)
1	B	691/908 (76%)	0.19	33 (4%)	31 39	26, 40, 75, 126	0
All	All	1379/1816 (75%)	0.12	59 (4%)	36 44	22, 38, 72, 126	1 (0%)

The worst 5 of 59 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	32	SER	6.8
1	B	146	ALA	6.3
1	A	32	SER	5.6
1	B	36	LEU	5.6
1	B	147	ALA	5.2

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 carbohydrates 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.