wwPDB X-ray Structure Validation Summary Report (i)

May 18, 2020 – 01:08 pm BST

PDB ID	:	4WP5
Title	:	Chaetomium thermophilum Mex67 NTF2-like domain complexed with Mtr2
Authors	:	Aibara, S.; Valkov, E.; Stewart, M.
Deposited on	:	2014-10-17
Resolution	:	2.90 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

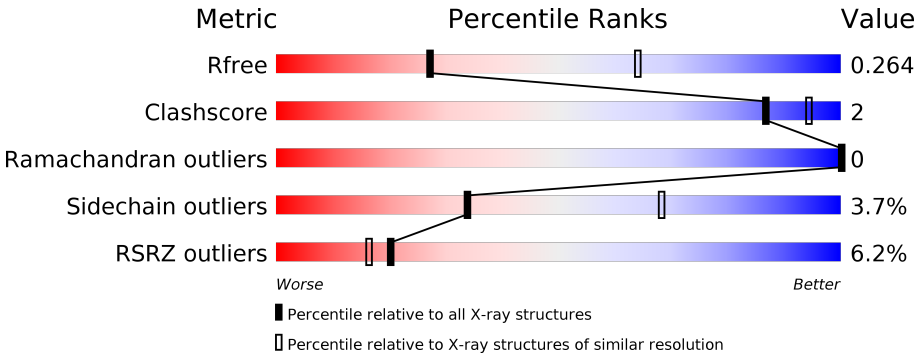
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 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	213	5% 86% 6% • 7%
2	B	183	7% 85% 8% • 7%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5822 atoms, of which 2858 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 mRNA export protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	199	3067	992	1509	265	293	8	0	0	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	352	GLY	-	expression tag	UNP G0SET4
A	353	SER	-	expression tag	UNP G0SET4
A	354	HIS	-	expression tag	UNP G0SET4
A	355	HIS	-	expression tag	UNP G0SET4
A	356	HIS	-	expression tag	UNP G0SET4
A	357	HIS	-	expression tag	UNP G0SET4
A	358	HIS	-	expression tag	UNP G0SET4
A	359	HIS	-	expression tag	UNP G0SET4
A	360	SER	-	expression tag	UNP G0SET4
A	361	GLN	-	expression tag	UNP G0SET4
A	362	ASP	-	expression tag	UNP G0SET4
A	363	PRO	-	expression tag	UNP G0SET4
A	364	MET	-	expression tag	UNP G0SET4

- Molecule 2 is a protein called Mtr2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	171	2735	879	1349	249	251	7	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	initiating methionine	UNP G0SG92
B	2	LEU	-	expression tag	UNP G0SG92
B	3	SER	-	expression tag	UNP G0SG92

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	4	ARG	-	expression tag	UNP G0SG92
B	5	ARG	-	expression tag	UNP G0SG92
B	6	TYR	-	expression tag	UNP G0SG92
B	7	ALA	-	expression tag	UNP G0SG92
B	170	ALA	GLY	conflict	UNP G0SG92

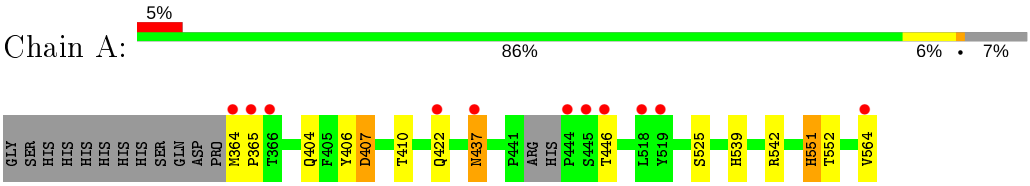
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	11	Total O 11 11	0	0
3	B	9	Total O 9 9	0	0

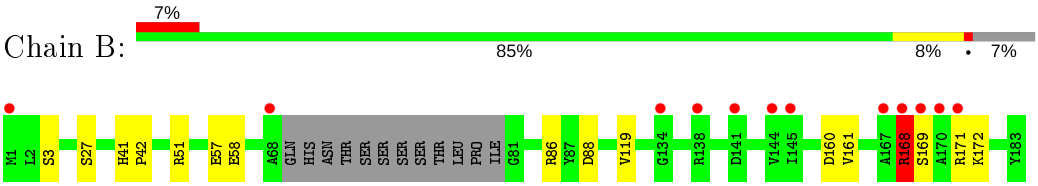
3 Residue-property plots (i)

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 (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: mRNA export protein



- Molecule 2: Mtr2



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	103.22Å 103.22Å 89.01Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	44.65 – 2.90 44.65 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.0 (44.65-2.90) 98.1 (44.65-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.48 (at 2.90Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.221 , 0.254 0.231 , 0.264	Depositor DCC
R_{free} test set	593 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	74.6	Xtriage
Anisotropy	0.388	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 52.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.034 for -h,-k,l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5822	wwPDB-VP
Average B, all atoms (Å^2)	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% 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	0.21	0/1606	0.41	0/2198
2	B	0.22	0/1423	1.05	4/1929 (0.2%)
All	All	0.22	0/3029	0.78	4/4127 (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
2	B	0	1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	168	ARG	CD-NE-CZ	27.79	162.51	123.60
2	B	168	ARG	NH1-CZ-NH2	-23.02	94.08	119.40
2	B	168	ARG	NE-CZ-NH2	-17.67	111.46	120.30
2	B	168	ARG	NE-CZ-NH1	14.50	127.55	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	168	ARG	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	1558	1509	1508	6	1
2	B	1386	1349	1349	6	9
3	A	11	0	0	0	0
3	B	9	0	0	0	0
All	All	2964	2858	2857	12	10

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.

The worst 5 of 12 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)
2:B:161:VAL:HG21	2:B:171:ARG:HB3	1.74	0.70
2:B:51:ARG:NH2	2:B:58:GLU:OE1	2.27	0.67
2:B:168:ARG:HG3	2:B:171:ARG:HB2	1.85	0.57
2:B:86:ARG:NH1	2:B:88:ASP:OD1	2.41	0.54
1:A:437:ASN:OD1	1:A:437:ASN:N	2.41	0.52

The worst 5 of 10 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)
2:B:168:ARG:NE	2:B:168:ARG:NE[4_555]	1.29	0.91
2:B:168:ARG:NE	2:B:168:ARG:CZ[4_555]	1.33	0.87
2:B:168:ARG:CZ	2:B:168:ARG:NH2[4_555]	1.34	0.86
2:B:168:ARG:CZ	2:B:168:ARG:NH1[4_555]	1.34	0.86
2:B:168:ARG:NE	2:B:168:ARG:NH1[4_555]	1.58	0.62

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	195/213 (92%)	191 (98%)	4 (2%)	0	100	100
2	B	167/183 (91%)	165 (99%)	2 (1%)	0	100	100
All	All	362/396 (91%)	356 (98%)	6 (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	178/191 (93%)	173 (97%)	5 (3%)	43	76
2	B	145/157 (92%)	138 (95%)	7 (5%)	25	58
All	All	323/348 (93%)	311 (96%)	12 (4%)	34	68

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	3	SER
2	B	27	SER
2	B	168	ARG
1	A	551	HIS
2	B	160	ASP

Some 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

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

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

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	199/213 (93%)	0.36	11 (5%)	25	21	48, 73, 136, 168	0
2	B	171/183 (93%)	0.66	12 (7%)	16	12	46, 78, 146, 318	0
All	All	370/396 (93%)	0.50	23 (6%)	20	16	46, 75, 142, 318	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	169	SER	5.4
2	B	168	ARG	5.2
2	B	144	VAL	4.6
1	A	364	MET	4.5
2	B	171	ARG	4.5

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.