Full wwPDB X-ray Structure Validation Report (i)

May 23, 2020 – 07:53 pm BST

PDB ID	:	5CWW
Title	:	Crystal structure of the Chaetomium thermophilum heterotrimeric Nup82 NTD-Nup159 TAIL-Nup145N APD complex
Authors	:	Stuwe, T.; Bley, C.J.; Thierbach, K.; Petrovic, S.; Schilbach, S.; Mayo, D.J.; Perriches, T.; Rundlet, E.J.; Jeon, Y.E.; Collins, L.N.; Lin, D.H.; Paduch, M.; Koide, A.; Lu, V.; Fischer, J.; Hurt, E.; Koide, S.; Kossiakoff, A.A.; Hoelz, A.
Deposited on	:	2015-07-28
Resolution	:	2.20 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 (i) symbol.

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

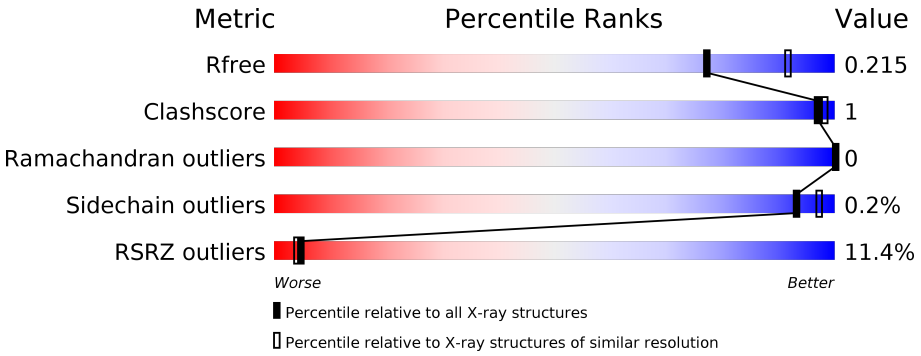
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

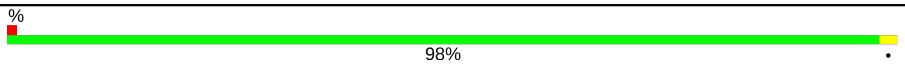
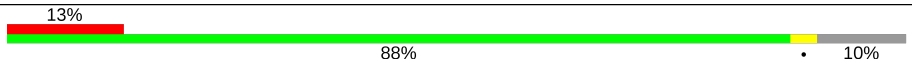
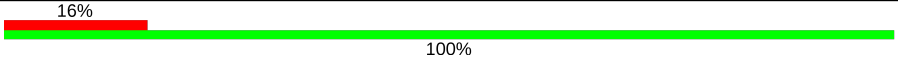
The reported resolution of this entry is 2.20 Å.

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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	140	
2	B	595	
3	C	32	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11319 atoms, of which 5534 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 Nucleoporin NUP145N.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	140	Total	C	H	N	O	S	0	1	0
			2212	705	1101	199	202	5			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	854	GLY	-	expression tag	UNP G0SAK3
A	855	PRO	-	expression tag	UNP G0SAK3
A	856	HIS	-	expression tag	UNP G0SAK3
A	857	MET	-	expression tag	UNP G0SAK3

- Molecule 2 is a protein called Nucleoporin NUP82.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	538	Total	C	H	N	O	S	0	7	0
			8387	2708	4145	705	817	12			

- Molecule 3 is a protein called Nucleoporin NUP159.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	32	Total	C	H	N	O	S	0	0	0
			549	158	288	59	43	1			

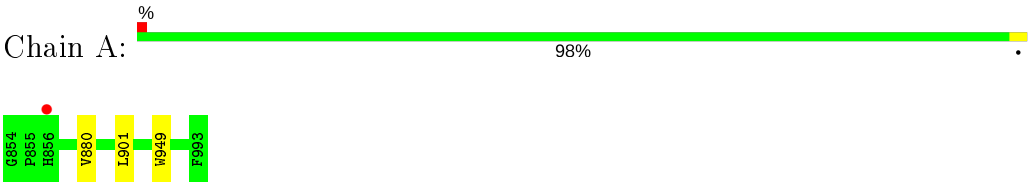
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	30	Total	O	0	0
			30	30		
4	B	131	Total	O	0	7
			138	138		
4	C	3	Total	O	0	0
			3	3		

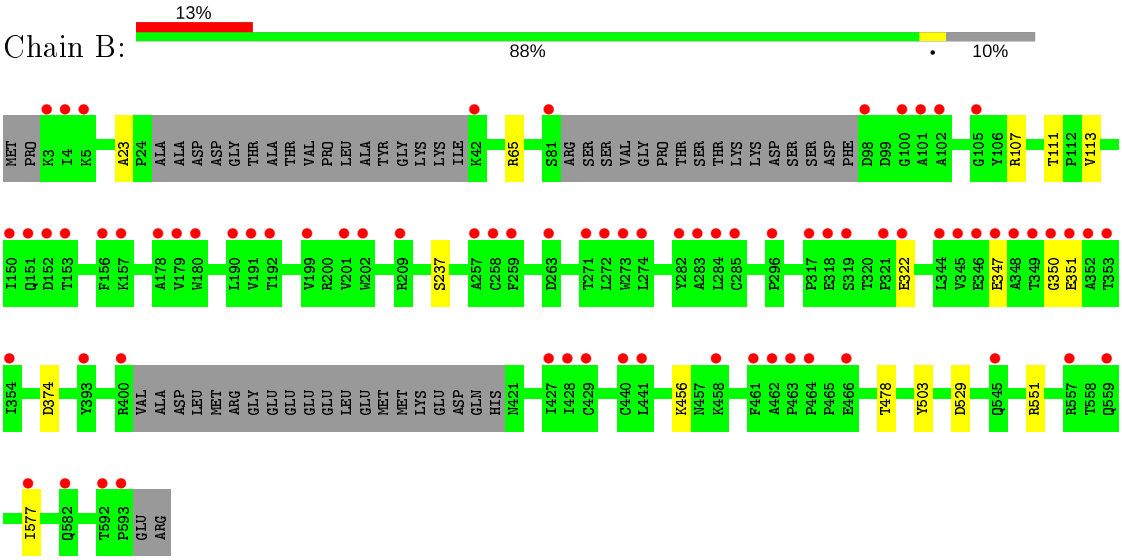
3 Residue-property plots (i)

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.

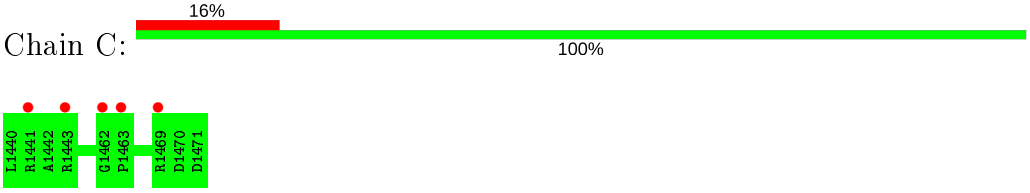
- Molecule 1: Nucleoporin NUP145N



- Molecule 2: Nucleoporin NUP82



- Molecule 3: Nucleoporin NUP159



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	122.30Å 107.96Å 69.59Å 90.00° 108.65° 90.00°	Depositor
Resolution (Å)	29.19 – 2.20 29.36 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.4 (29.19-2.20) 90.6 (29.36-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	0.79 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.185 , 0.213 0.189 , 0.215	Depositor DCC
R_{free} test set	2962 reflections (6.85%)	wwPDB-VP
Wilson B-factor (Å^2)	47.5	Xtriage
Anisotropy	0.430	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 51.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	11319	wwPDB-VP
Average B, all atoms (Å^2)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.41% 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	0/1144	0.38	0/1553
2	B	0.22	0/4366	0.41	0/5973
3	C	0.20	0/261	0.40	0/346
All	All	0.22	0/5771	0.40	0/7872

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	1111	1101	1102	1	0
2	B	4242	4145	4145	9	1
3	C	261	288	288	0	0
4	A	30	0	0	0	0
4	B	138	0	0	2	0
4	C	3	0	0	0	0
All	All	5785	5534	5535	10	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 1.

All (10) 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:23:ALA:O	2:B:65:ARG:NH1	2.22	0.72
2:B:107:ARG:NH2	4:B:607:HOH:O	2.35	0.58
2:B:478:THR:O	2:B:551:ARG:NH1	2.37	0.57
2:B:503:TYR:CE1	2:B:577:ILE:HD11	2.47	0.50
2:B:322[A]:GLU:OE1	2:B:456:LYS:NZ	2.38	0.49
2:B:111:THR:OG1	2:B:113:VAL:HG22	2.14	0.48
2:B:351:GLU:OE1	2:B:351:GLU:N	2.49	0.46
1:A:880:VAL:HG21	1:A:901:LEU:HD12	2.00	0.44
2:B:347:GLU:HB3	2:B:350:GLY:HA3	2.00	0.43
2:B:374[B]:ASP:OD1	4:B:601:HOH:O	2.22	0.41

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)
2:B:237:SER:OG	2:B:529:ASP:OD2[4_545]	2.18	0.02

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	139/140 (99%)	135 (97%)	4 (3%)	0	100	100
2	B	537/595 (90%)	523 (97%)	14 (3%)	0	100	100
3	C	30/32 (94%)	27 (90%)	3 (10%)	0	100	100
All	All	706/767 (92%)	685 (97%)	21 (3%)	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	121/120 (101%)	120 (99%)	1 (1%)	81 90
2	B	467/509 (92%)	467 (100%)	0	100 100
3	C	27/27 (100%)	27 (100%)	0	100 100
All	All	615/656 (94%)	614 (100%)	1 (0%)	93 97

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	949	TRP

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

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	140/140 (100%)	-0.08	1 (0%) 87 86	48, 70, 105, 146	0
2	B	538/595 (90%)	0.67	75 (13%) 2 2	39, 66, 133, 173	0
3	C	32/32 (100%)	0.82	5 (15%) 2 1	58, 81, 140, 151	0
All	All	710/767 (92%)	0.53	81 (11%) 5 4	39, 68, 128, 173	0

All (81) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	348	ALA	7.5
2	B	349	THR	6.7
2	B	352	ALA	5.9
2	B	347	GLU	5.1
2	B	345	VAL	4.8
2	B	353	THR	4.5
2	B	400	ARG	4.4
2	B	102	ALA	4.4
3	C	1462	GLY	4.4
2	B	344	LEU	4.3
2	B	545	GLN	4.2
2	B	5	LYS	4.1
2	B	81	SER	4.1
2	B	354	ILE	4.1
2	B	593	PRO	4.1
2	B	351	GLU	4.1
2	B	461	PHE	4.0
2	B	151	GLN	3.9
2	B	318	GLU	3.9
2	B	150	ILE	3.8
2	B	466	GLU	3.7
2	B	4	ILE	3.6
2	B	559	GLN	3.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	462	ALA	3.6
2	B	273	TRP	3.4
2	B	440	CYS	3.4
3	C	1441	ARG	3.4
2	B	577	ILE	3.4
2	B	557	ARG	3.3
2	B	153	THR	3.3
2	B	3	LYS	3.3
2	B	191	VAL	3.1
2	B	350	GLY	3.0
2	B	209	ARG	2.9
2	B	263	ASP	2.9
2	B	42	LYS	2.9
2	B	258	CYS	2.9
2	B	283	ALA	2.9
2	B	202	TRP	2.8
2	B	157	LYS	2.8
2	B	201	VAL	2.8
2	B	317	PRO	2.8
2	B	428	ILE	2.8
1	A	856	HIS	2.8
2	B	429	CYS	2.8
2	B	592	THR	2.8
2	B	296	PRO	2.8
2	B	272	LEU	2.7
2	B	156	PHE	2.7
3	C	1469	ARG	2.7
2	B	463	PRO	2.6
3	C	1463	PRO	2.6
2	B	190	LEU	2.6
2	B	199	VAL	2.6
2	B	321	PRO	2.6
2	B	346	GLU	2.6
2	B	284	LEU	2.5
2	B	98	ASP	2.5
3	C	1443	ARG	2.5
2	B	271	THR	2.5
2	B	582	GLN	2.4
2	B	464	PRO	2.4
2	B	285	CYS	2.3
2	B	152	ASP	2.3
2	B	322[A]	GLU	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	179	VAL	2.3
2	B	101	ALA	2.3
2	B	180	TRP	2.2
2	B	458	LYS	2.2
2	B	259	PHE	2.2
2	B	105	GLY	2.2
2	B	274	LEU	2.2
2	B	282	TYR	2.2
2	B	319	SER	2.2
2	B	393	TYR	2.2
2	B	427	ILE	2.2
2	B	192	THR	2.1
2	B	441	LEU	2.1
2	B	257	ALA	2.1
2	B	178	ALA	2.1
2	B	100	GLY	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 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.