Full wwPDB X-ray Structure Validation Report

Apr 25, 2016 – 05:45 PM EDT

PDB ID	:	5HAY
Title	:	Crystal structure of Chaetomium thermophilum Nup170 CTD Y905M L1007M L1183M V1292M mutant
Authors	:	Lin, D.H.; Fan, Y.; Hoelz, A.
Deposited on	:	2015-12-31
Resolution	:	2.80 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
http://wwpdb.org/validation/2016/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:

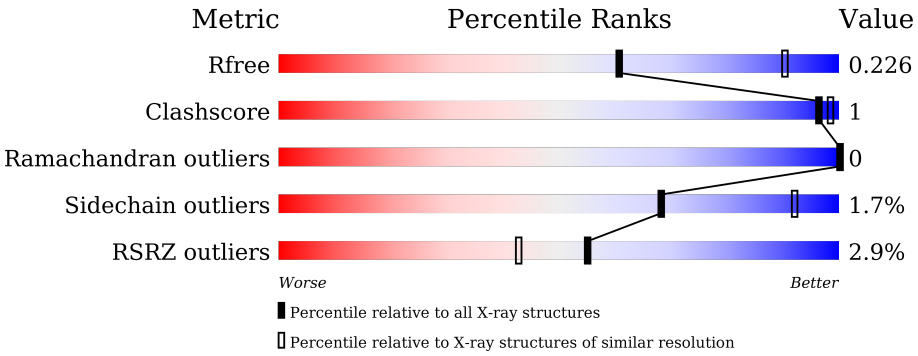
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027257
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027257

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 2.80 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}	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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 <=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	586	3%, 92%, 5%
1	B	586	3%, 82%, 6%, 12%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 17083 atoms, of which 8470 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 NUP170.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	A	559	Total	C	H	N	O	S	Se	0	0	0
			8796	2788	4370	780	839	7	12			
1	B	517	Total	C	H	N	O	S	Se	0	0	0
			8238	2615	4100	726	781	6	10			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	831	SER	-	expression tag	UNP G0S7B6
A	905	MSE	TYR	ENGINEERED MUTATION	UNP G0S7B6
A	1007	MSE	LEU	ENGINEERED MUTATION	UNP G0S7B6
A	1183	MSE	LEU	ENGINEERED MUTATION	UNP G0S7B6
A	1292	MSE	VAL	ENGINEERED MUTATION	UNP G0S7B6
B	831	SER	-	expression tag	UNP G0S7B6
B	905	MSE	TYR	ENGINEERED MUTATION	UNP G0S7B6
B	1007	MSE	LEU	ENGINEERED MUTATION	UNP G0S7B6
B	1183	MSE	LEU	ENGINEERED MUTATION	UNP G0S7B6
B	1292	MSE	VAL	ENGINEERED MUTATION	UNP G0S7B6

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Na	0	0
			1	1		

- Molecule 3 is water.

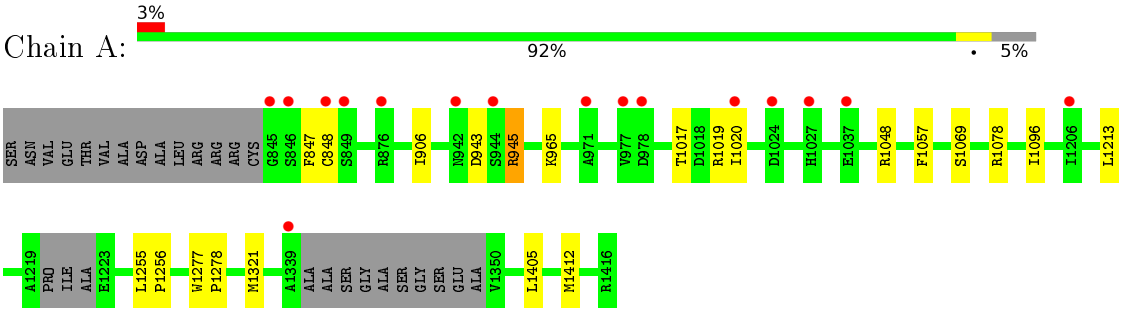
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	38	Total	O	0	0
			38	38		
3	B	10	Total	O	0	0
			10	10		

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 errors 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.

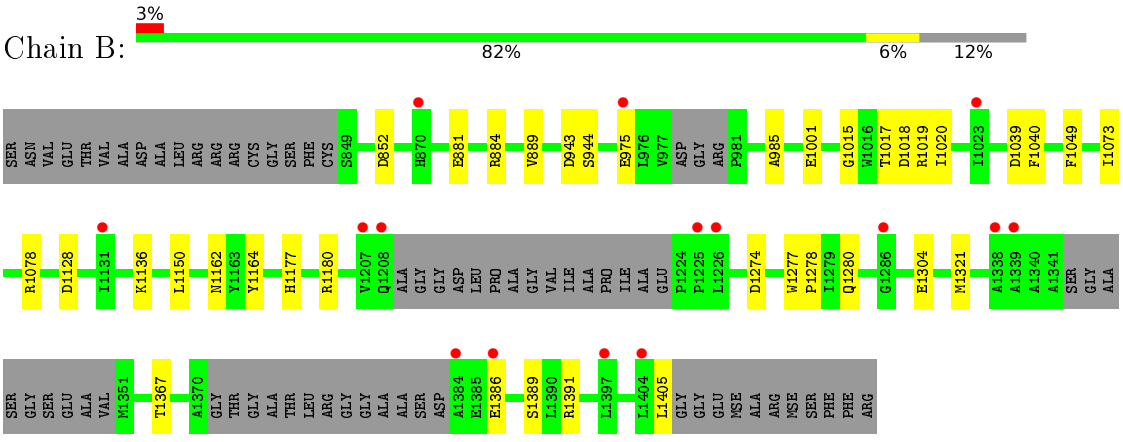
- Molecule 1: Nucleoporin NUP170

Chain A:



- Molecule 1: Nucleoporin NUP170

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	61.79Å 101.70Å 193.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.19 – 2.80 49.19 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.7 (49.19-2.80) 99.7 (49.19-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.68 (at 2.81Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.209 , 0.251 0.211 , 0.226	Depositor DCC
R_{free} test set	1540 reflections (5.00%)	DCC
Wilson B-factor (Å^2)	65.7	Xtriage
Anisotropy	0.224	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 43.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	17083	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: NA

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	0.26	0/4498	0.42	0/6081
1	B	0.25	0/4207	0.41	0/5694
All	All	0.25	0/8705	0.41	0/11775

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	B	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	1304	GLU	Peptide

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	4426	4370	4368	10	0
1	B	4138	4100	4097	14	0
2	A	1	0	0	0	0
3	A	38	0	0	0	0
3	B	10	0	0	0	0
All	All	8613	8470	8465	23	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 1.

All (23) 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:1069:SER:O	1:A:1078:ARG:NH1	2.16	0.78
1:B:881:GLU:OE1	1:B:884:ARG:NH1	2.31	0.64
1:B:1039:ASP:OD1	1:B:1040:PHE:N	2.38	0.56
1:B:943:ASP:OD1	1:B:944:SER:N	2.39	0.56
1:B:1386:GLU:O	1:B:1389:SER:N	2.43	0.52
1:A:1057:PHE:CD2	1:A:1096:ILE:HD13	2.48	0.49
1:B:1367:THR:HG22	1:B:1391:ARG:NH1	2.28	0.48
1:A:1048:ARG:NH1	1:B:1015:GLY:O	2.40	0.48
1:B:1128:ASP:O	1:B:1136:LYS:NZ	2.46	0.48
1:B:1073:ILE:O	1:B:1078:ARG:NH2	2.48	0.47
1:B:975:GLU:HA	1:B:985:ALA:HB2	1.96	0.46
1:A:847:PHE:HB2	1:A:848:CYS:SG	2.56	0.46
1:A:847:PHE:HA	1:A:848:CYS:HA	1.74	0.45
1:A:943:ASP:OD2	1:A:945:ARG:NE	2.43	0.45
1:B:1162:ASN:HB2	1:B:1164:TYR:CZ	2.54	0.42
1:A:1255:LEU:HB3	1:A:1256:PRO:HD3	2.00	0.42
1:B:1274:ASP:OD2	1:B:1280:GLN:NE2	2.53	0.42
1:A:1017:THR:HA	1:A:1020:ILE:HG22	2.02	0.41
1:B:889:VAL:HG12	1:B:889:VAL:O	2.20	0.41
1:A:1277:TRP:CG	1:A:1278:PRO:HD3	2.55	0.41
1:B:1277:TRP:N	1:B:1278:PRO:CD	2.84	0.41
1:A:906:ILE:CG2	1:A:965:LYS:HG2	2.51	0.40
1:B:1017:THR:HA	1:B:1020:ILE:HG22	2.03	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	553/586 (94%)	537 (97%)	16 (3%)	0	100	100
1	B	507/586 (86%)	493 (97%)	14 (3%)	0	100	100
All	All	1060/1172 (90%)	1030 (97%)	30 (3%)	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	473/479 (99%)	467 (99%)	6 (1%)	76	94
1	B	447/479 (93%)	437 (98%)	10 (2%)	60	89
All	All	920/958 (96%)	904 (98%)	16 (2%)	68	92

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	945	ARG
1	A	1019	ARG
1	A	1213	LEU
1	A	1321	MSE
1	A	1405	LEU
1	A	1412	MSE
1	B	852	ASP
1	B	1001	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	1018	ASP
1	B	1019	ARG
1	B	1049	PHE
1	B	1150	LEU
1	B	1177	HIS
1	B	1180	ARG
1	B	1321	MSE
1	B	1405	LEU

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

Mol	Chain	Res	Type
1	B	1233	GLN

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

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 (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	547/586 (93%)	0.31	16 (2%) 55 43	32, 56, 96, 121	0
1	B	507/586 (86%)	0.37	15 (2%) 54 41	33, 68, 98, 120	0
All	All	1054/1172 (89%)	0.34	31 (2%) 55 43	32, 62, 97, 121	0

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1024	ASP	6.8
1	B	870	HIS	5.9
1	A	848	CYS	5.5
1	A	849	SER	4.8
1	A	942	ASN	4.6
1	A	1037	GLU	4.5
1	A	978	ASP	4.4
1	A	977	VAL	4.1
1	B	1339	ALA	3.7
1	B	1397	LEU	3.6
1	A	846	SER	3.6
1	A	1027	HIS	3.4
1	A	1020	ILE	3.3
1	A	971	ALA	3.2
1	A	1339	ALA	3.2
1	B	1404	LEU	2.8
1	B	1338	ALA	2.8
1	B	1207	VAL	2.6
1	A	1206	ILE	2.6
1	B	1286	GLY	2.6
1	B	1131	ILE	2.4
1	B	1208	GLN	2.4
1	A	944	SER	2.4
1	A	876	ARG	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	845	GLY	2.3
1	B	1023	ILE	2.3
1	B	1225	PRO	2.3
1	B	1226	LEU	2.2
1	B	975	GLU	2.2
1	B	1386	GLU	2.1
1	B	1384	ALA	2.1

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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
2	NA	A	1501	1/1	0.94	0.23	0.41	62,62,62,62	0

6.5 Other polymers (i)

There are no such residues in this entry.