



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 17, 2024 – 06:42 AM EDT

PDB ID : 5HAZ  
Title : Structure of Chaetomium thermophilum Nup170 CTD  
Authors : Lin, D.H.; Hoelz, A.  
Deposited on : 2015-12-31  
Resolution : 2.10 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](mailto: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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
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)	:	2.37.1

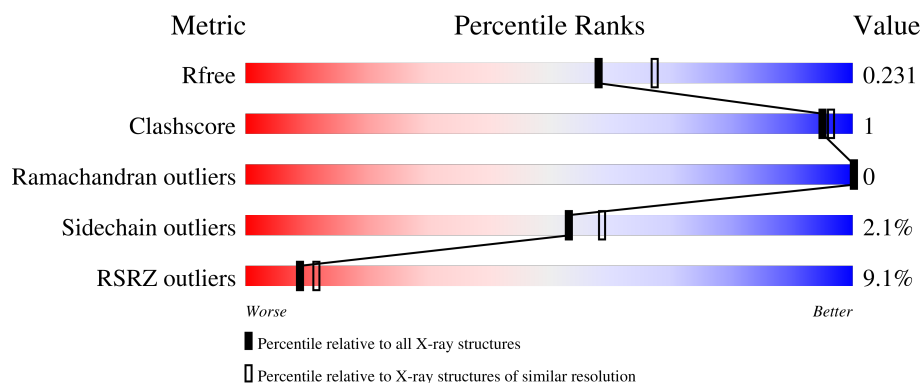
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

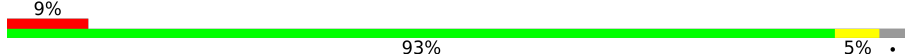
The reported resolution of this entry is 2.10 Å.

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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	567	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	EPE	A	1503[A]	-	-	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	NA	A	1504	-	-	-	X

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 9405 atoms, of which 4486 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.

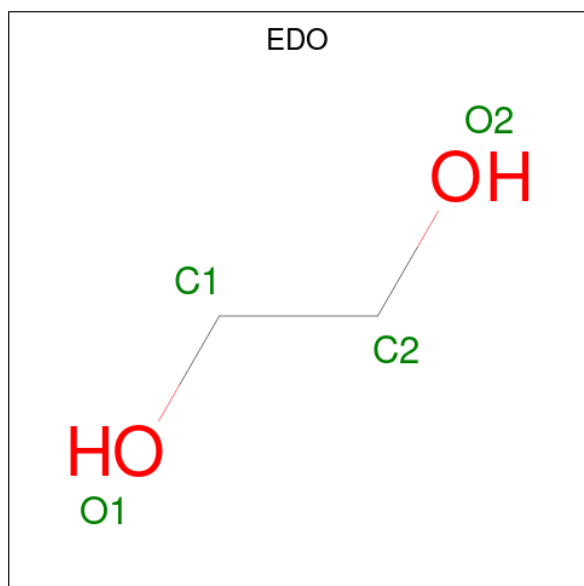
- Molecule 1 is a protein called Nucleoporin NUP170.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	551	Total	C	H	N	O	S	0	23	0
			8974	2844	4457	791	866	16			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	850	SER	-	expression tag	UNP G0S7B6

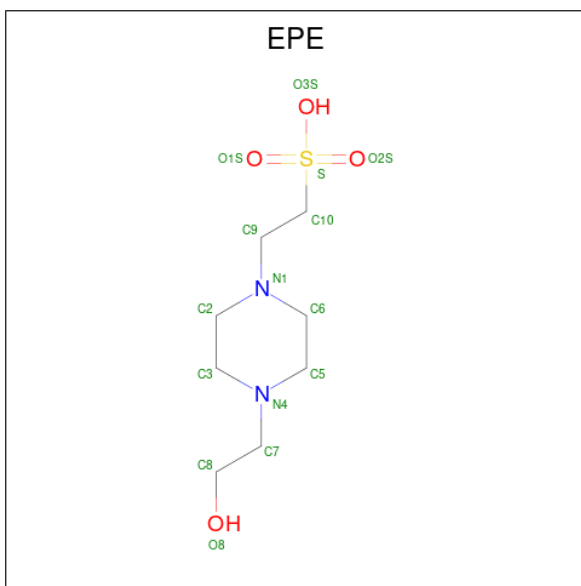
- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			10	2	6	2		
2	A	1	Total	C	H	O	0	0
			10	2	6	2		

- Molecule 3 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID

(three-letter code: EPE) (formula:  $C_8H_{18}N_2O_4S$ ).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	C	H	N	O	S	0	1
			32	8	17	2	4	1		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

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

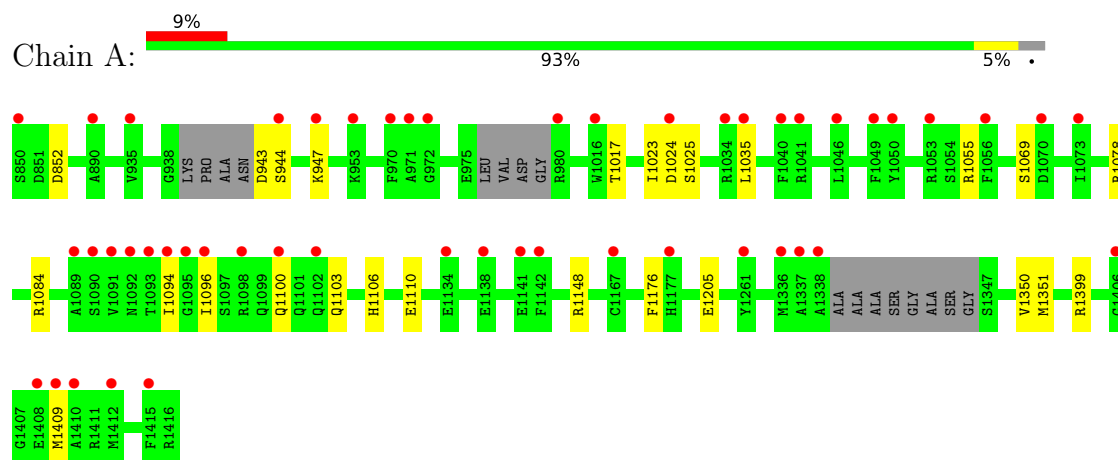
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	378	Total	O	0	2
			378	378		

### 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: Nucleoporin NUP170



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	141.26Å 54.77Å 108.20Å 90.00° 91.05° 90.00°	Depositor
Resolution (Å)	37.26 – 2.10 37.26 – 2.10	Depositor EDS
% Data completeness (in resolution range)	97.9 (37.26-2.10) 97.9 (37.26-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.70 (at 2.10Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.206 , 0.231 0.207 , 0.231	Depositor DCC
$R_{free}$ test set	2396 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	43.5	Xtriage
Anisotropy	0.327	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 53.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.014 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9405	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.03% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, EDO, EPE

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	0.29	0/4636	0.43	0/6290

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 [i](#)

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	4517	4457	4438	11	0
2	A	8	12	12	0	0
3	A	15	17	17	0	0
4	A	1	0	0	0	0
5	A	378	0	0	2	0
All	All	4919	4486	4467	11	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 (11) 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	2.27	0.67
1:A:1205[A]:GLU:OE1	5:A:1601:HOH:O	2.13	0.66
1:A:1399:ARG:NH1	5:A:1603:HOH:O	2.35	0.57
1:A:1106:HIS:NE2	1:A:1110[A]:GLU:OE2	2.39	0.56
1:A:1350:VAL:HG12	1:A:1351:MET:N	2.27	0.49
1:A:943:ASP:OD2	1:A:944:SER:N	2.45	0.49
1:A:1055:ARG:NH1	1:A:1094:ILE:O	2.48	0.45
1:A:1084[B]:ARG:HH11	1:A:1084[B]:ARG:HG2	1.82	0.44
1:A:1084[B]:ARG:HH11	1:A:1084[B]:ARG:CG	2.30	0.44
1:A:1023:ILE:HG22	1:A:1025:SER:H	1.85	0.41

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	566/567 (100%)	551 (97%)	15 (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	489/475 (103%)	479 (98%)	10 (2%)	55	60

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	852	ASP
1	A	947	LYS
1	A	1017	THR
1	A	1024	ASP
1	A	1035	LEU
1	A	1096	ILE
1	A	1100	GLN
1	A	1103	GLN
1	A	1148	ARG
1	A	1409	MET

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](#)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	EPE	A	1503[A]	3	15,15,15	0.82	1 (6%)	18,20,20	1.91	6 (33%)
2	EDO	A	1501	-	3,3,3	0.46	0	2,2,2	0.31	0
2	EDO	A	1502	-	3,3,3	0.41	0	2,2,2	0.35	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	EPE	A	1503[A]	3	-	3/9/19/19	0/1/1/1
2	EDO	A	1501	-	-	0/1/1/1	-
2	EDO	A	1502	-	-	0/1/1/1	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1503[A]	EPE	C10-S	2.53	1.81	1.77

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1503[A]	EPE	C5-N4-C3	3.66	117.08	108.83
3	A	1503[A]	EPE	C7-N4-C5	3.58	120.38	111.23
3	A	1503[A]	EPE	O1S-S-C10	3.50	111.13	106.92
3	A	1503[A]	EPE	C7-N4-C3	2.45	117.49	111.23
3	A	1503[A]	EPE	C6-N1-C2	2.36	114.15	108.83
3	A	1503[A]	EPE	O2S-S-C10	2.34	109.73	106.92

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1503[A]	EPE	S-C10-C9-N1
3	A	1503[A]	EPE	C8-C7-N4-C3
3	A	1503[A]	EPE	N4-C7-C8-O8

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	551/567 (97%)	0.69	50 (9%)	9 12	34, 60, 136, 193	0

All (50) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1337	ALA	6.7
1	A	972	GLY	6.5
1	A	1409	MET	6.0
1	A	1412	MET	5.2
1	A	1093	THR	5.1
1	A	1408	GLU	4.7
1	A	1167	CYS	4.7
1	A	980	ARG	3.9
1	A	1089	ALA	3.9
1	A	1035	LEU	3.8
1	A	970	PHE	3.6
1	A	1410	ALA	3.6
1	A	1053	ARG	3.6
1	A	850	SER	3.6
1	A	953	LYS	3.5
1	A	1056	PHE	3.2
1	A	1338	ALA	3.1
1	A	1094	ILE	3.1
1	A	1070	ASP	3.1
1	A	1034	ARG	3.0
1	A	1102	GLN	2.9
1	A	1142	PHE	2.9
1	A	947	LYS	2.9
1	A	1100	GLN	2.8
1	A	1024	ASP	2.8
1	A	935	VAL	2.8
1	A	1096	ILE	2.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	890	ALA	2.7
1	A	1134	GLU	2.7
1	A	1406	GLY	2.7
1	A	944	SER	2.7
1	A	1090	SER	2.6
1	A	1050	TYR	2.5
1	A	1141	GLU	2.5
1	A	1336	MET	2.5
1	A	1016	TRP	2.5
1	A	1040	PHE	2.4
1	A	1046	LEU	2.3
1	A	1073	ILE	2.2
1	A	971	ALA	2.2
1	A	1138[A]	GLU	2.2
1	A	1177	HIS	2.1
1	A	1092	ASN	2.1
1	A	1415	PHE	2.1
1	A	1095	GLY	2.1
1	A	1049	PHE	2.1
1	A	1261	TYR	2.1
1	A	1098	ARG	2.0
1	A	1041	ARG	2.0
1	A	1091	VAL	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](#)

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. The B-factors column lists the minimum, median, 95<sup>th</sup> 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.

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
-----	------	-------	-----	-------	------	-----	-----------------------------	-------

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	NA	A	1504	1/1	-0.35	0.93	245,245,245,245	0
3	EPE	A	1503[A]	15/15	0.60	0.55	112,135,153,175	32
2	EDO	A	1501	4/4	0.75	0.19	95,114,114,114	0
2	EDO	A	1502	4/4	0.87	0.36	86,103,104,104	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.