Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 10:16 am BST

PDB ID	:	5CWV
Title	:	Crystal structure of Chaetomium thermophilum Nup192 TAIL domain
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	:	3.15 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 ⓘ symbol.

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

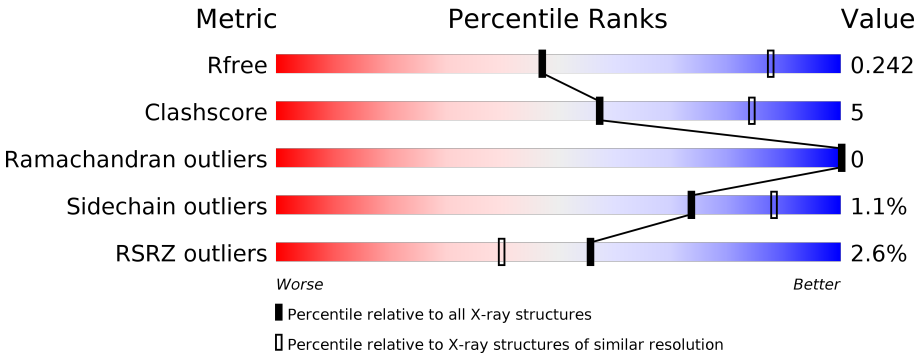
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 3.15 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}	130704	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)
RSRZ outliers	127900	1616 (3.20-3.12)

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	364	2% 68% 7% 25%
1	B	364	2% 65% 8% 27%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 8455 atoms, of which 4288 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 NUP192.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	A	273	Total	C	H	N	O	S	Se	0	0	0
			4298	1344	2181	372	390	7	4			
1	B	265	Total	C	H	N	O	S	Se	0	0	0
			4157	1303	2107	358	379	7	3			

There are 8 discrepancies between the modelled and reference sequences:

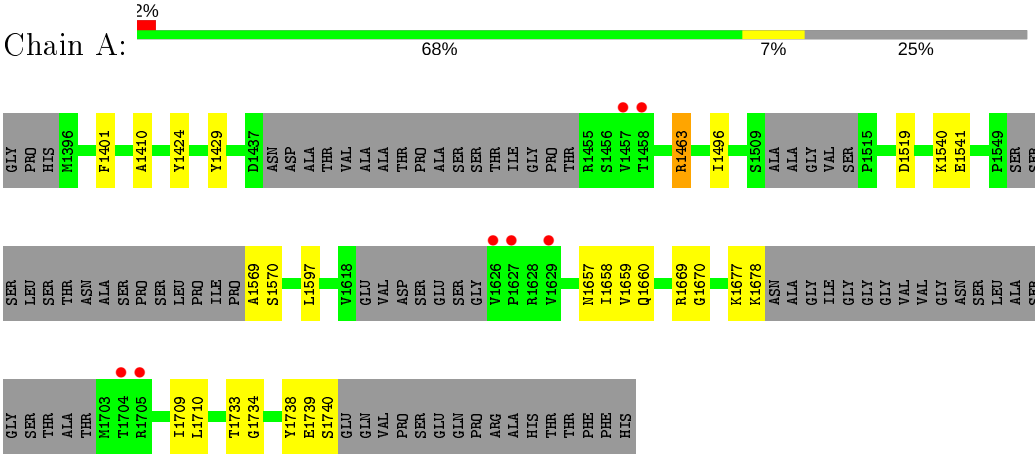
Chain	Residue	Modelled	Actual	Comment	Reference
A	1393	GLY	-	expression tag	UNP G0S4T0
A	1394	PRO	-	expression tag	UNP G0S4T0
A	1395	HIS	-	expression tag	UNP G0S4T0
A	1396	MSE	-	expression tag	UNP G0S4T0
B	1393	GLY	-	expression tag	UNP G0S4T0
B	1394	PRO	-	expression tag	UNP G0S4T0
B	1395	HIS	-	expression tag	UNP G0S4T0
B	1396	MSE	-	expression tag	UNP G0S4T0

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.

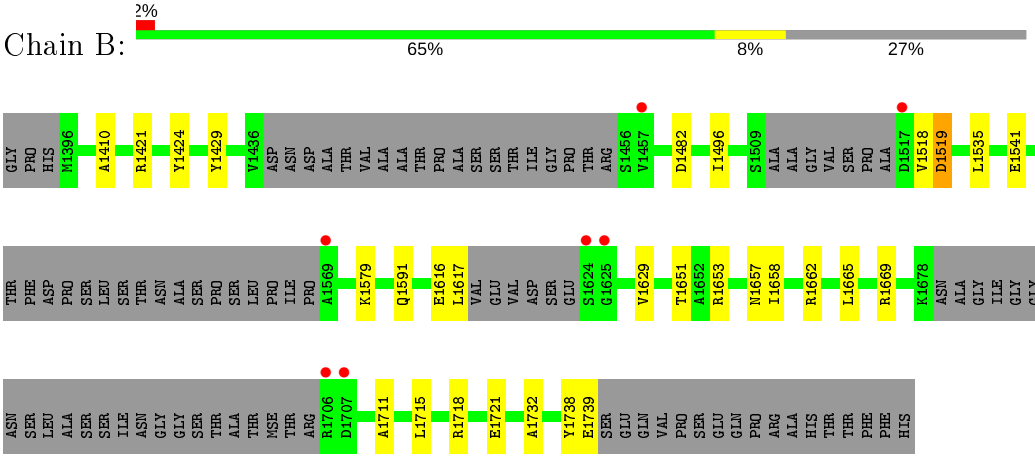
- Molecule 1: Nucleoporin NUP192

Chain A:



- Molecule 1: Nucleoporin NUP192

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	64.75Å 114.47Å 114.69Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.89 – 3.15 19.89 – 3.15	Depositor EDS
% Data completeness (in resolution range)	99.4 (19.89-3.15) 94.1 (19.89-3.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 3.15Å)	Xtriage
Refinement program	PHENIX 1.8_1069	Depositor
R, R_{free}	0.194 , 0.237 0.200 , 0.242	Depositor DCC
R_{free} test set	1518 reflections (10.05%)	wwPDB-VP
Wilson B-factor (Å^2)	107.7	Xtriage
Anisotropy	0.229	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 53.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.447 for -h,l,k	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8455	wwPDB-VP
Average B, all atoms (Å^2)	129.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.60% 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	0.24	0/2141	0.40	0/2894
1	B	0.25	0/2072	0.42	0/2800
All	All	0.24	0/4213	0.41	0/5694

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	2117	2181	2174	21	0
1	B	2050	2107	2102	21	0
All	All	4167	4288	4276	41	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 5.

All (41) 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:1669:ARG:NH1	1:A:1739:GLU:OE1	1.63	1.29
1:B:1519:ASP:HB3	1:B:1591:GLN:NE2	1.74	1.02

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1519:ASP:OD2	1:B:1519:ASP:OD1	1.99	0.81
1:B:1669:ARG:NH1	1:B:1739:GLU:OE1	2.21	0.74
1:B:1519:ASP:HB3	1:B:1591:GLN:HE22	1.52	0.73
1:A:1669:ARG:HH12	1:A:1739:GLU:CD	1.96	0.69
1:A:1677:LYS:O	1:A:1678:LYS:HE2	1.92	0.69
1:A:1401:PHE:CZ	1:A:1463:ARG:HD2	2.34	0.63
1:B:1616:GLU:HA	1:B:1617:LEU:C	2.21	0.60
1:A:1733:THR:HB	1:A:1734:GLY:HA2	1.85	0.59
1:B:1665:LEU:O	1:B:1669:ARG:N	2.37	0.58
1:B:1653:ARG:HD3	1:B:1657:ASN:OD1	2.04	0.58
1:A:1677:LYS:C	1:A:1678:LYS:HG2	2.26	0.56
1:B:1519:ASP:OD2	1:B:1519:ASP:N	2.40	0.55
1:B:1616:GLU:N	1:B:1617:LEU:HB3	2.24	0.52
1:B:1718:ARG:NH1	1:B:1721:GLU:OE1	2.43	0.52
1:A:1429:TYR:CD1	1:A:1496:ILE:HG22	2.45	0.51
1:B:1519:ASP:CB	1:B:1591:GLN:NE2	2.62	0.51
1:B:1629:VAL:HG11	1:B:1711:ALA:HB1	1.91	0.51
1:B:1429:TYR:CD1	1:B:1496:ILE:HG22	2.46	0.50
1:A:1540:LYS:HG3	1:A:1541:GLU:HG3	1.93	0.50
1:A:1709:ILE:HG23	1:A:1710:LEU:CD2	2.42	0.49
1:A:1401:PHE:CE1	1:A:1463:ARG:HD2	2.47	0.48
1:A:1669:ARG:CG	1:A:1670:GLY:N	2.79	0.45
1:B:1518:VAL:CG1	1:B:1519:ASP:N	2.80	0.45
1:A:1569:ALA:O	1:A:1570:SER:OG	2.28	0.44
1:B:1421:ARG:NH2	1:B:1482:ASP:OD1	2.51	0.43
1:B:1541:GLU:N	1:B:1541:GLU:OE1	2.51	0.43
1:B:1658:ILE:O	1:B:1662:ARG:HG3	2.17	0.42
1:A:1659:VAL:HG13	1:A:1660:GLN:N	2.34	0.42
1:A:1733:THR:H	1:A:1734:GLY:HA2	1.84	0.42
1:B:1629:VAL:HG13	1:B:1715:LEU:HD22	2.01	0.42
1:A:1733:THR:HB	1:A:1734:GLY:CA	2.49	0.41
1:A:1669:ARG:HG3	1:A:1670:GLY:H	1.85	0.41
1:B:1410:ALA:HB1	1:B:1424:TYR:CE2	2.55	0.41
1:A:1709:ILE:HG23	1:A:1710:LEU:HD22	2.03	0.41
1:A:1410:ALA:HB1	1:A:1424:TYR:CE2	2.56	0.41
1:B:1535:LEU:HD11	1:B:1579:LYS:HG3	2.03	0.40
1:A:1597:LEU:HD13	1:A:1657:ASN:ND2	2.37	0.40
1:A:1738:TYR:C	1:A:1738:TYR:CD2	2.93	0.40
1:B:1651:THR:HG22	1:B:1732:ALA:HB1	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	261/364 (72%)	246 (94%)	15 (6%)	0	100	100
1	B	253/364 (70%)	242 (96%)	11 (4%)	0	100	100
All	All	514/728 (71%)	488 (95%)	26 (5%)	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	227/292 (78%)	224 (99%)	3 (1%)	69	86
1	B	219/292 (75%)	217 (99%)	2 (1%)	78	91
All	All	446/584 (76%)	441 (99%)	5 (1%)	73	88

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1463	ARG
1	A	1658	ILE
1	A	1740	SER
1	B	1519	ASP
1	B	1738	TYR

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	1413	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

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	269/364 (73%)	0.02	7 (2%)	56	40	74, 109, 190, 254	0
1	B	262/364 (71%)	0.00	7 (2%)	54	38	72, 106, 185, 217	0
All	All	531/728 (72%)	0.01	14 (2%)	56	40	72, 107, 191, 254	0

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1626	VAL	5.1
1	A	1704	THR	5.0
1	A	1457	VAL	4.8
1	B	1624	SER	4.3
1	B	1706	ARG	4.1
1	B	1569	ALA	3.3
1	A	1627	PRO	2.9
1	B	1457	VAL	2.7
1	A	1458	THR	2.6
1	A	1705	ARG	2.5
1	B	1517	ASP	2.4
1	B	1707	ASP	2.3
1	B	1625	GLY	2.2
1	A	1629	VAL	2.1

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.