W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Apr 25, 2016 – 07:19 PM EDT

PDB ID	:	5HB3
Title	:	Crystal structure of Chaetomium thermophilum Nic96 SOL-Nup53 complex
Authors	:	Lin, D.H.; Stuwe, T.; Hoelz, A.
Deposited on	:	2015-12-31
Resolution	:	2.65 Å(reported)

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

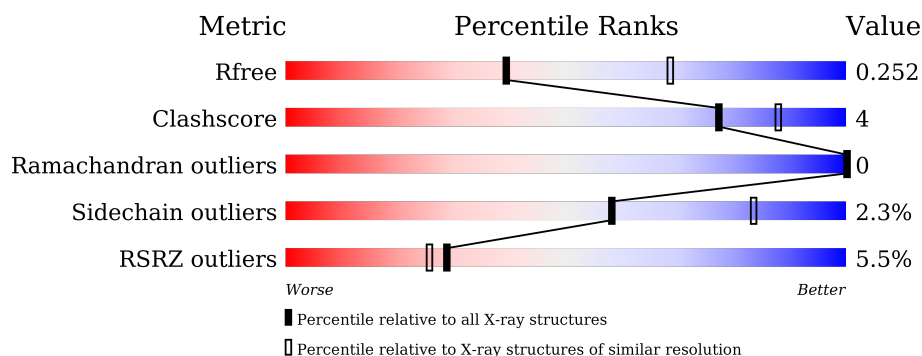|		
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027257
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (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 2.65 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	Percentile Ranks	Value
Rfree		0.252
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		2.3%
RSRZ outliers		5.5%

Worse   *Better*

∎ Percentile relative to all X-ray structures

⧠ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	3152 (2.70-2.62)
Clashscore	102246	3524 (2.70-2.62)
Ramachandran outliers	100387	3469 (2.70-2.62)
Sidechain outliers	100360	3469 (2.70-2.62)
RSRZ outliers	91569	3161 (2.70-2.62)

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	723	7% · 86% · 10% · ·
1	C	723	3% · 89% · 9% · ·
2	B	55	5% · 27% · · 67%
2	D	55	2% · 22% · 9% · 69%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 23000 atoms, of which 11431 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 NIC96.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	703	Total 11153	C 3520	H 5564	N 999	O 1036	S 34	0	1	0
1	C	709	Total 11250	C 3549	H 5610	N 1011	O 1047	S 33	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	390	SER	-	expression tag	UNP G0S024
C	390	SER	-	expression tag	UNP G0S024

- Molecule 2 is a protein called Nucleoporin NUP53.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	18	Total 267	C 88	H 130	N 22	O 27	0	0	0
2	D	17	Total 256	C 84	H 127	N 21	O 24	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	30	SER	-	expression tag	UNP G0S156
D	30	SER	-	expression tag	UNP G0S156

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Cl 1	0	0
3	C	2	Total 2	Cl 2	0	0

• Molecule 4 is water.

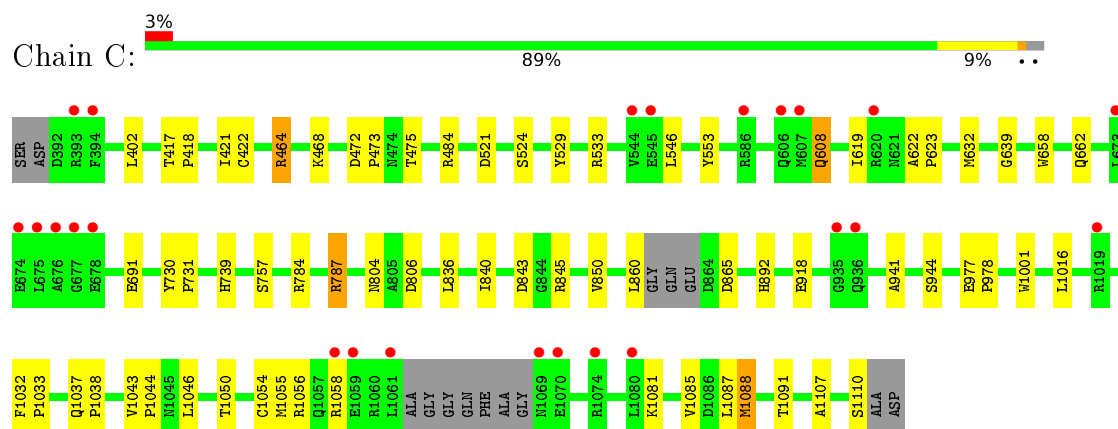| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	36	Total 36	O 36	0	0
4	C	35	Total 35	O 35	0	0

3 Residue-property plots ⓘ

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.
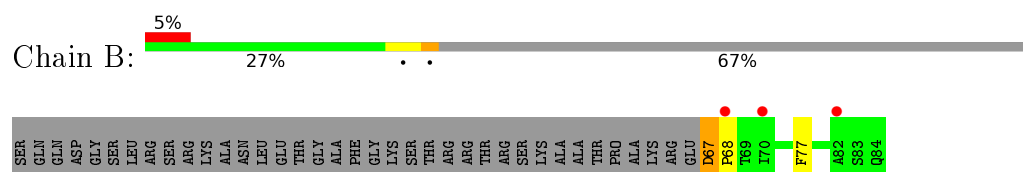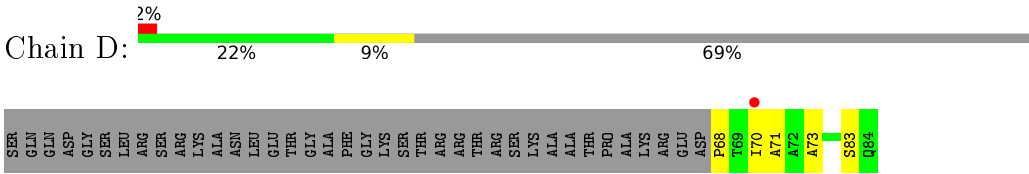
- Molecule 1: Nucleoporin NIC96
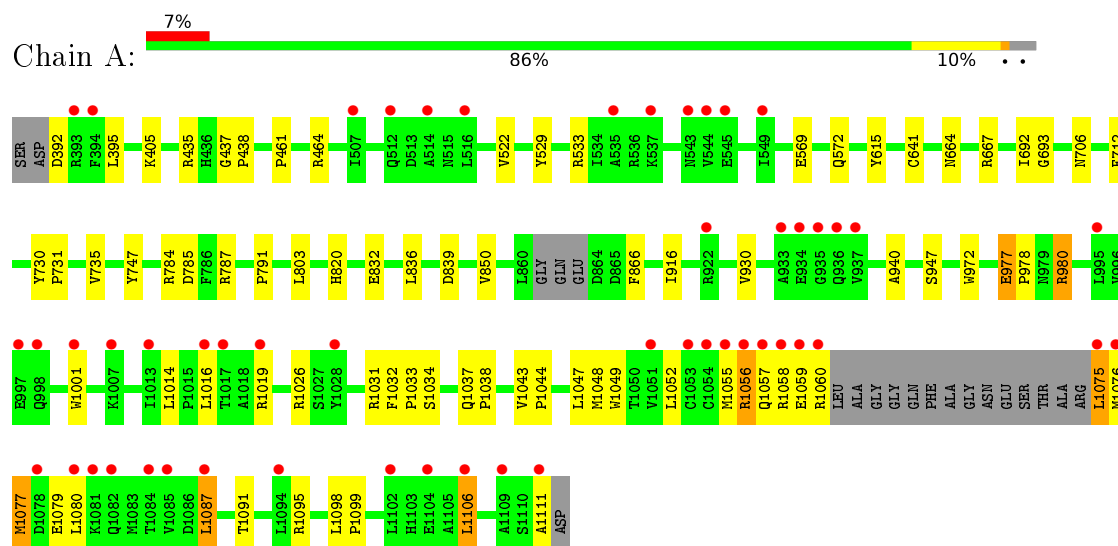
Chain A:

- Molecule 1: Nucleoporin NIC96

Chain C:

- Molecule 2: Nucleoporin NUP53

Chain B:

• Molecule 2: Nucleoporin NUP53

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	59.52Å 86.95Å 98.01Å 100.69° 99.59° 95.66°	Depositor
Resolution (Å)	47.28 – 2.65 47.29 – 2.65	Depositor EDS
% Data completeness (in resolution range)	88.5 (47.28-2.65) 83.0 (47.29-2.65)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.31 (at 2.65Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.212 , 0.249 0.215 , 0.252	Depositor DCC
R_{free} test set	2368 reflections (4.88%)	DCC
Wilson B-factor (Å^2)	57.6	Xtriage
Anisotropy	0.142	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 38.5	EDS
L-test for twinning [2]	$<\lvert L\rvert> = 0.49, <L^2> = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	23000	wwPDB-VP
Average B, all atoms (Å^2)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.86% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $<\lvert L\rvert>$, $<L^2>$ 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: CL

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	0.22	0/5696	0.38	0/7702
1	C	0.22	0/5748	0.37	0/7773
2	B	0.20	0/140	0.32	0/191
2	D	0.20	0/132	0.34	0/179
All	All	0.22	0/11716	0.38	0/15845

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	5589	5564	5549	44	0
1	C	5640	5610	5595	37	0
2	B	137	130	130	2	0
2	D	129	127	127	3	0
3	A	1	0	0	0	0
3	C	2	0	0	0	0
4	A	36	0	0	0	0
4	C	35	0	0	0	0
All	All	11569	11431	11401	85	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 4.

The worst 5 of 85 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:C:1081:LYS:O	1:C:1085:VAL:HG23	2.03	0.58
1:C:622:ALA:HB1	1:C:623:PRO:CD	2.36	0.55
1:C:622:ALA:HB1	1:C:623:PRO:HD2	1.90	0.54
2:D:68:PRO:HA	2:D:71:ALA:HB3	1.90	0.53
1:C:1037:GLN:HB3	1:C:1038:PRO:HD3	1.90	0.53

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	698/723 (96%)	668 (96%)	30 (4%)	0	100	100
1	C	704/723 (97%)	678 (96%)	26 (4%)	0	100	100
2	B	16/55 (29%)	16 (100%)	0	0	100	100
2	D	15/55 (27%)	14 (93%)	1 (7%)	0	100	100
All	All	1433/1556 (92%)	1376 (96%)	57 (4%)	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	584/596 (98%)	568 (97%)	16 (3%)	52	80
1	C	590/596 (99%)	580 (98%)	10 (2%)	68	88
2	B	14/43 (33%)	13 (93%)	1 (7%)	18	38
2	D	13/43 (30%)	13 (100%)	0	100	100
All	All	1201/1278 (94%)	1174 (98%)	27 (2%)	58	84

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1077	MET
1	A	1106	LEU
1	C	865	ASP
1	A	1087	LEU
1	A	785	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	486	GLN
1	C	932	ASN
1	C	578	GLN
1	A	760	ASN
1	C	572	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 (i)

Of 3 ligands modelled in this entry, 3 are 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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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	703/723 (97%)	0.49	52 (7%)	17 14	32, 60, 109, 124	0
1	C	709/723 (98%)	0.32	24 (3%)	49 47	35, 56, 105, 130	0
2	B	18/55 (32%)	0.53	3 (16%)	2 1	66, 79, 91, 97	0
2	D	17/55 (30%)	0.59	1 (5%)	26 23	63, 79, 87, 87	0
All	All	1447/1556 (92%)	0.41	80 (5%)	29 26	32, 58, 106, 130	0

The worst 5 of 80 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	935	GLY	8.5
1	A	1060	ARG	8.3
1	A	1057	GLN	8.2
1	A	1087	LEU	7.2
1	C	676	ALA	7.1

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, 95th 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(Å²)	Q<0.9
3	CL	A	1201	1/1	0.91	0.14	-1.48	62,62,62,62	0
3	CL	C	1202	1/1	0.97	0.14	-2.80	51,51,51,51	0
3	CL	C	1201	1/1	0.94	0.13	-2.93	57,57,57,57	0

6.5 Other polymers (i)

There are no such residues in this entry.