wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:44 PM GMT

PDB ID	:	4JC8
Title	:	Crystal Structure of HOPS component Vps33 from Chaetomium thermophilum
Authors	:	Baker, R.W.; Jeffrey, P.D.; Hughson, F.M.
Deposited on	:	2013-02-21
Resolution	:	2.60 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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:

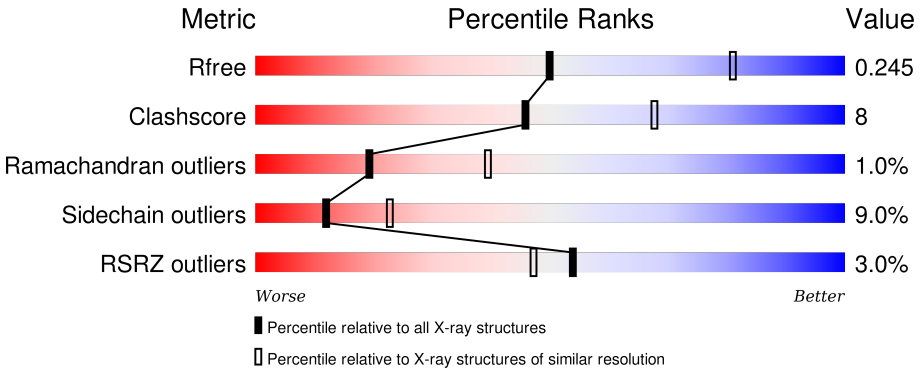
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	669	3% 68% 18% • 12%
1	B	669	2% 68% 17% • 12%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9565 atoms, of which 0 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 HOPS component Vps33.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	590	Total	C	N	O	S	Se	0	0	0
			4665	2958	822	874	4	7			
1	B	589	Total	C	N	O	S	Se	0	0	0
			4675	2965	824	875	4	7			

- Molecule 2 is water.

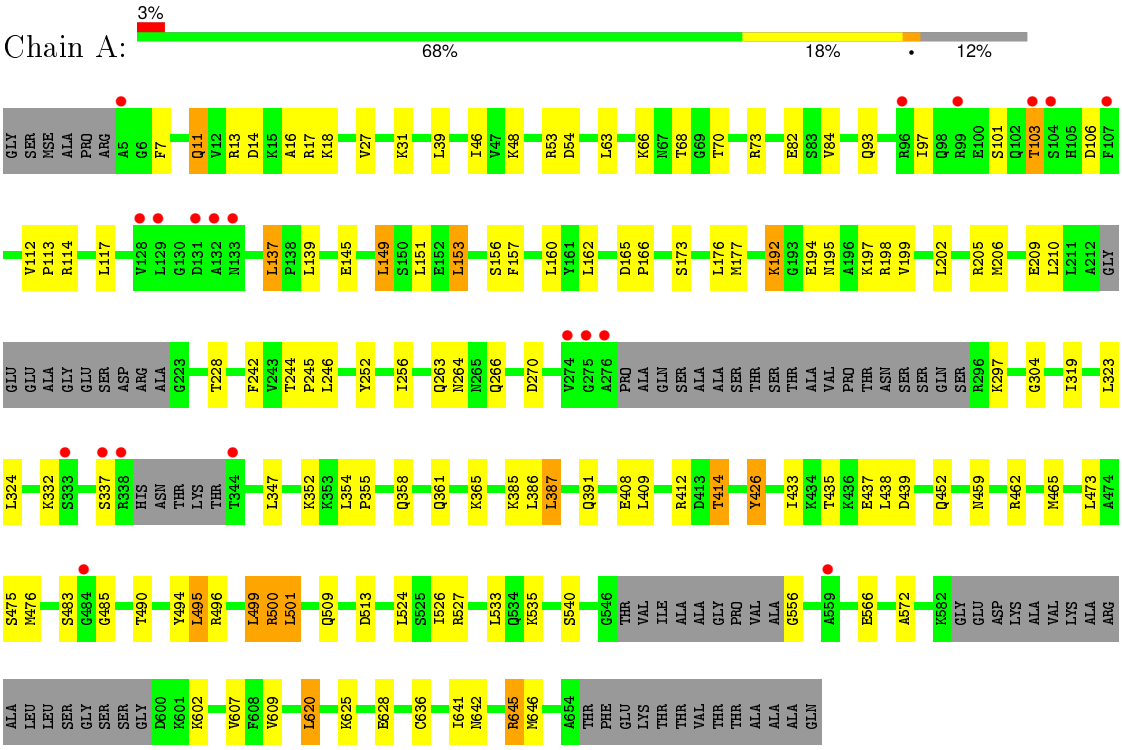
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	125	Total	O	0	0
			125	125		
2	B	100	Total	O	0	0
			100	100		

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.

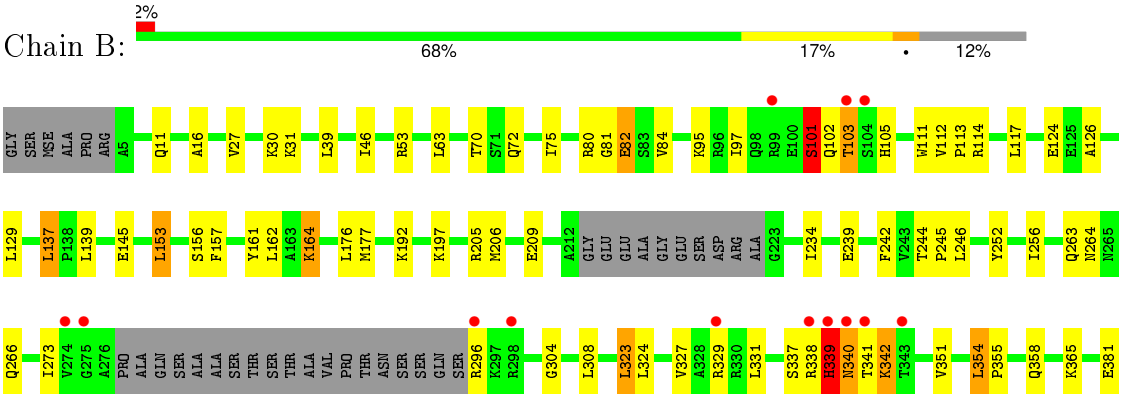
- Molecule 1: HOPS component Vps33

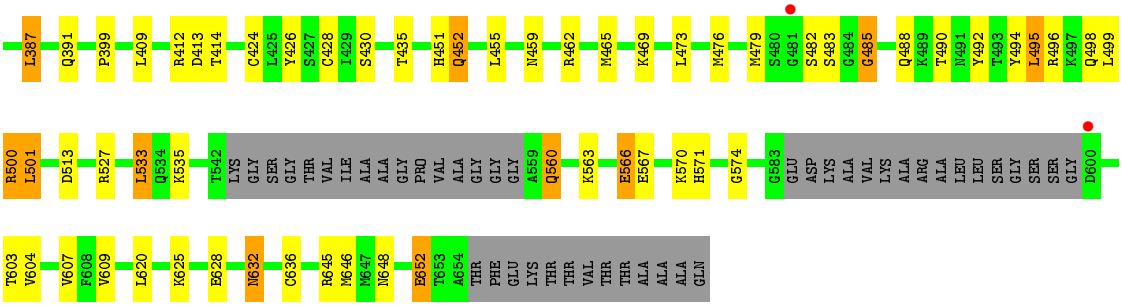
Chain A:



- Molecule 1: HOPS component Vps33

Chain B:




L387 Q391 P399 L409 R412 D413 T414 C424 L425 Y426 S427 C428 I429 S430 T435 H451 Q452 L455 N459 R462 M465 K469 L473 M476 M479 S480 G481 S482 S483 G484 G485 Q488 K489 T490 N491 Y492 T493 Y494 L495 R496 K497 Q498 L499
R500 L501 D513 R527 L533 Q534 K535 T542 LYS GLY SER GLY THR VAL ILE ALA ALA GLY PRO VAL ALA GLY GLY GLY A559 Q560 K563 E566 E567 K570 H571 G574 G583 GLU ASP LYS ALA VAL LYS ALA ARG ALA LEU LEU SER GLY SER SER GLY D600
T603 V604 V607 F608 V609 L620 K625 E628 N632 C636 R645 M646 M647 N648 E652 T653 A654 THR PHE GLU LYS THR THR VAL THR THR ALA ALA ALA GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	71.94Å 64.44Å 151.69Å 90.00° 91.77° 90.00°	Depositor
Resolution (Å)	35.95 – 2.60 49.10 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.7 (35.95-2.60) 99.7 (49.10-2.60)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.32 (at 2.61Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1144)	Depositor
R, R_{free}	0.185 , 0.247 0.181 , 0.245	Depositor DCC
R_{free} test set	2154 reflections (5.01%)	DCC
Wilson B-factor (Å^2)	46.7	Xtriage
Anisotropy	0.034	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 42.6	EDS
Estimated twinning fraction	0.108 for h,-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 43014 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9565	wwPDB-VP
Average B, all atoms (Å^2)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% 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

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	0.45	0/4729	0.63	0/6363
1	B	0.40	0/4741	0.60	0/6384
All	All	0.42	0/9470	0.62	0/12747

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	4665	0	4745	70	0
1	B	4675	0	4756	80	0
2	A	125	0	0	15	0
2	B	100	0	0	9	0
All	All	9565	0	9501	145	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 8.

The worst 5 of 145 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:361:GLN:HG2	1:B:354:LEU:HG	1.58	0.83

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:82:GLU:OE1	2:B:704:HOH:O	2.00	0.77
1:A:426:TYR:OH	1:A:437:GLU:OE2	2.03	0.75
1:A:160:LEU:O	2:A:780:HOH:O	2.05	0.75
1:A:347:LEU:HD11	1:B:329:ARG:HG3	1.71	0.72

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	578/669 (86%)	554 (96%)	20 (4%)	4 (1%)	26	51
1	B	579/669 (86%)	548 (95%)	24 (4%)	7 (1%)	16	33
All	All	1157/1338 (86%)	1102 (95%)	44 (4%)	11 (1%)	19	39

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	101	SER
1	A	483	SER
1	B	101	SER
1	A	103	THR
1	B	103	THR

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	508/556 (91%)	460 (91%)	48 (9%)	11 20
1	B	511/556 (92%)	467 (91%)	44 (9%)	13 25
All	All	1019/1112 (92%)	927 (91%)	92 (9%)	12 23

5 of 92 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	524	LEU
1	B	39	LEU
1	B	560	GLN
1	A	533	LEU
1	A	620	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	72	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	583/669 (87%)	-0.11	20 (3%)	49	41	3, 18, 60, 85	0
1	B	582/669 (86%)	-0.13	15 (2%)	59	53	7, 25, 55, 81	0
All	All	1165/1338 (87%)	-0.12	35 (3%)	54	47	3, 22, 57, 85	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	484	GLY	4.5
1	B	339	HIS	4.0
1	A	99	ARG	3.9
1	B	343	THR	3.9
1	A	132	ALA	3.9

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.