Full wwPDB X-ray Structure Validation Report (i)

Jun 24, 2024 – 02:14 PM EDT

PDB ID	:	6XWU
Title	:	Crystal structure of drosophila melanogaster CENP-C cumin domain
Authors	:	Jeyaprakash, A.A.; Medina-Pritchard, B.; Lazou, V.; Zou, J.; Byron, O.; Abad, M.A.; Rappsilber, J.; Heun, P.
Deposited on	:	2020-01-24
Resolution	:	1.82 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 (i) 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 (1)) were used in the production of this report:

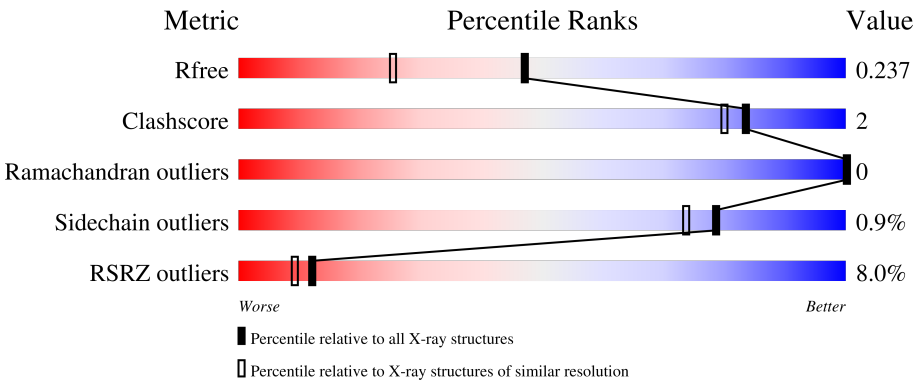
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

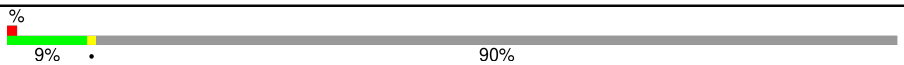
The reported resolution of this entry is 1.82 Å.

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	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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 of 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	1411	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1107 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 RE68959p.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	138	1065	670	184	203	8	0	0	0

- Molecule 2 is water.

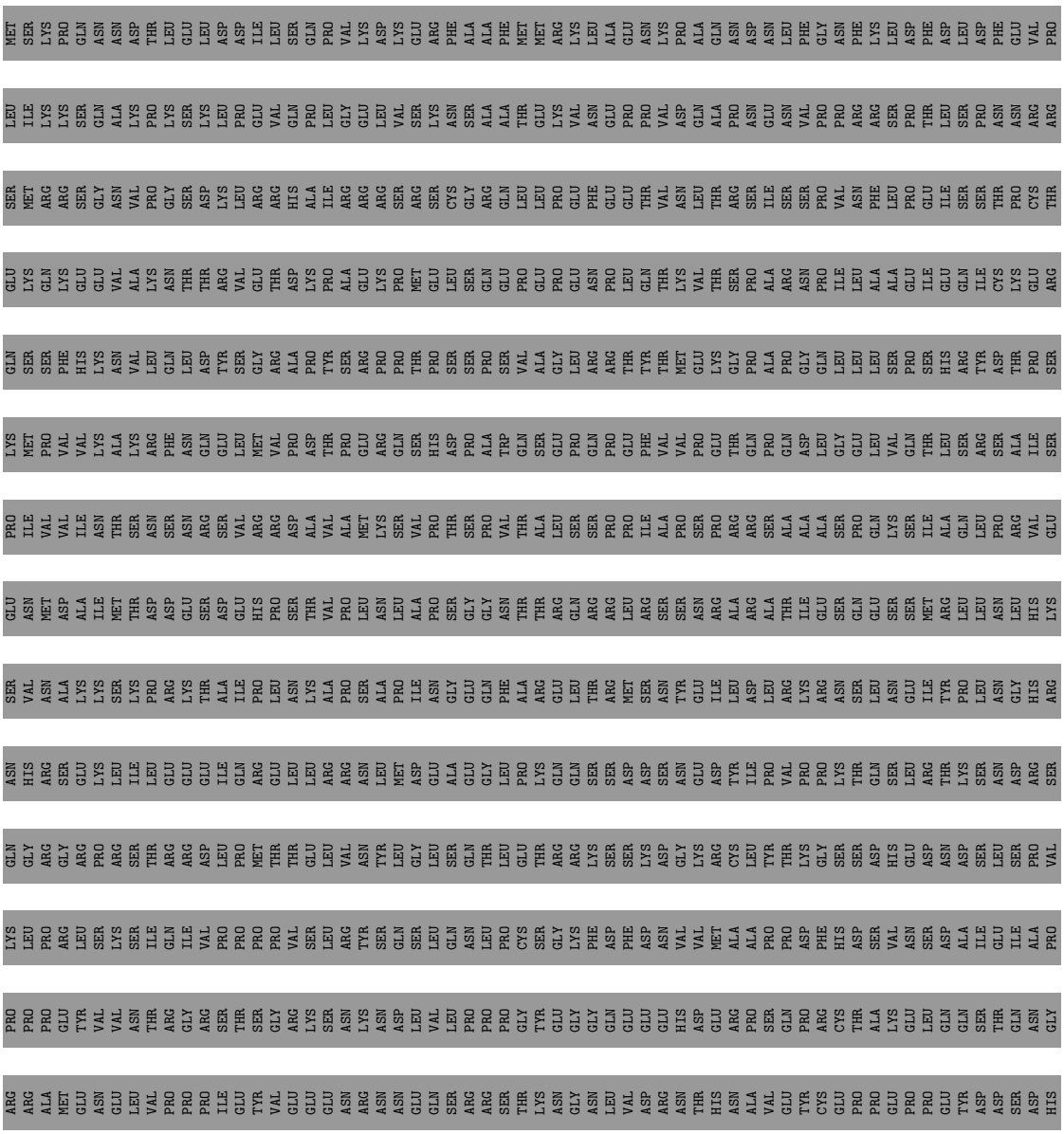
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	42	42	42	0	0

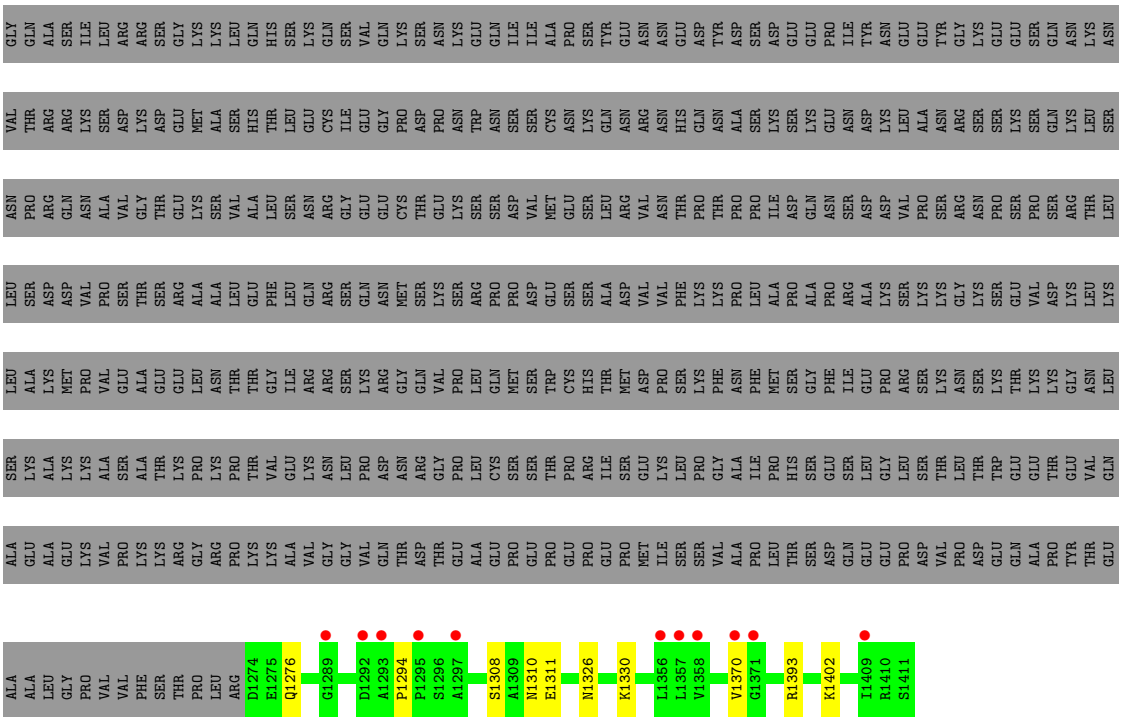
3 Residue-property plots (i)

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: RE68959p

Chain A:




GLY GLN ALA SER ILE LEU ARG ARG SER GLY LYS LYS LEU GLN HIS SER LYS GLN SER VAL GLN LYS SER ASN LYS GLU GLN ILE ILE ALA PRO SER TYR GLU ASN ASN GLU ASP TYR ASP SER ASP GLU GLU PRO ILE TYR ASN GLU GLU TYR GLY LYS GLU GLU SER GLN ASN LYS ASN
VAL THR ARG ARG LYS SER ASP LYS ASP GLU MET ALA SER HIS THR LEU GLU CYS ILE GLU GLY PRO ASP PRO ASN TRP ASN SER SER CYS ASN LYS GLN ASN ARG ASN HIS GLN ASN ALA SER LYS SER LYS GLU ASN ASP LYS LEU ALA ASN ARG SER SER LYS SER GLN LYS LEU SER
ASN PRO ARG GLN ASN ALA VAL GLY THR GLU LYS SER VAL ALA LEU SER ASN ARG GLY GLU GLU CYS THR GLU LYS SER SER ASP VAL MET GLU SER LEU ARG VAL ASN THR PRO THR PRO PRO ILE ASP GLN ASN SER ASP ASP VAL PRO SER ARG ASN PRO SER PRO ARG LEU
LEU SER ASP ASP VAL PRO SER THR SER ARG ALA ALA LEU GLU PHE LEU GLN ARG SER GLN ASN MET SER LYS SER ARG PRO PRO ASP GLU SER SER ALA ASP VAL VAL PHE LYS LYS PRO LEU ALA PRO ALA PRO ARG ALA LYS SER LYS LYS GLY LYS SER GLU VAL ASP LYS LEU LYS
LEU ALA LYS MET PRO VAL GLU ALA ALA GLU GLU LEU ASN THR THR GLY ILE ARG ARG SER LYS ARG GLY GLN VAL PRO LEU GLN MET SER TRP CYS HIS THR MET ASP PRO SER LYS PHE ASN PHE MET SER GLY PHE ILE GLU PRO ARG SER LYS ASN SER LYS THR LYS LYS GLY ASN LEU
SER LYS ALA LYS LYS ALA SER ALA THR LYS PRO PRO LYS PRO THR THR VAL GLU LYS ASN LEU PRO ASP ASN ARG GLY PRO LEU LEU CYS SER SER THR PRO ARG ILE SER GLU LYS LEU PRO GLY ALA ILE PRO HIS SER GLU SER LEU GLY LEU SER THR LEU THR TRP GLU GLU GLU THR GLU VAL GLN
ALA GLU ALA GLU LYS VAL PRO LYS LYS ARG GLY ARG PRO LYS LYS ALA VAL GLY GLY VAL GLN THR ASP THR GLU ALA GLU PRO GLU PRO GLU PRO GLU PRO MET ILE SER SER VAL ALA PRO LEU THR SER ASP GLN GLU GLU PRO ASP VAL PRO ASP GLU GLN ALA PRO TYR THR GLU
ALA ALA LEU GLY PRO VAL VAL PHE SER THR PRO LEU ARG D1274 E1275 Q1276 G1289 D1292 A1293 P1294 P1295 S1296 A1297 S1308 A1309 N1310 E1311 N1326 K1330 L1356 L1357 V1358 V1370 G1371 R1393 K1402 I1409 R1410 S1411

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	57.20Å 57.20Å 92.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.60 – 1.82 28.60 – 1.82	Depositor EDS
% Data completeness (in resolution range)	98.6 (28.60-1.82) 97.9 (28.60-1.82)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.97 (at 1.82Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660, PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.194 , 0.235 0.197 , 0.237	Depositor DCC
R_{free} test set	1429 reflections (9.90%)	wwPDB-VP
Wilson B-factor (Å^2)	32.6	Xtriage
Anisotropy	0.115	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 47.5	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	0.96	EDS
Total number of atoms	1107	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.08% 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 (i)

5.1 Standard geometry (i)

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	0.36	0/1082	0.55	0/1458

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	1065	0	1043	5	1
2	A	42	0	0	2	0
All	All	1107	0	1043	5	1

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 2.

All (5) 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:1393:ARG:NH1	2:A:1501:HOH:O	2.23	0.72
1:A:1294:PRO:O	1:A:1310:ASN:ND2	2.28	0.66
1:A:1370:VAL:HG22	2:A:1502:HOH:O	2.19	0.42
1:A:1308:SER:O	1:A:1311:GLU:HB2	2.20	0.42
1:A:1326:ASN:HD22	1:A:1326:ASN:HA	1.75	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1276:GLN:OE1	1:A:1402:LYS:NZ[7_555]	2.14	0.06

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	136/1411 (10%)	133 (98%)	3 (2%)	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	113/1275 (9%)	112 (99%)	1 (1%)	78 74

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1330	LYS

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

Mol	Chain	Res	Type
1	A	1326	ASN

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)

There are no ligands in this entry.

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	138/1411 (9%)	0.32	11 (7%) 12 9	24, 38, 66, 74	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1358	VAL	3.0
1	A	1371	GLY	2.7
1	A	1292	ASP	2.7
1	A	1370	VAL	2.6
1	A	1357	LEU	2.5
1	A	1409	ILE	2.3
1	A	1295	PRO	2.3
1	A	1297	ALA	2.2
1	A	1289	GLY	2.2
1	A	1356	LEU	2.2
1	A	1293	ALA	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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.