WORLDWIDE wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 08:54 pm GMT

PDB ID	:	5JH5
Title	:	Structural Basis for the Hierarchical Assembly of the Core of PRC1.1
Authors	:	Wong, S.J.; Taylor, A.B.; Hart, P.J.; Kim, C.A.
Deposited on	:	2016-04-20
Resolution	:	2.55 Å(reported)

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

|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

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 2.55 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.
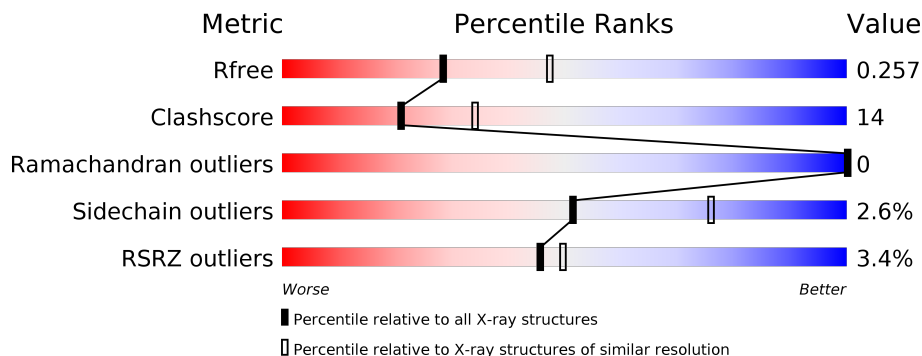
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	4993 (2.58-2.50)
Clashscore	112137	5755 (2.58-2.50)
Ramachandran outliers	110173	5652 (2.58-2.50)
Sidechain outliers	110143	5654 (2.58-2.50)
RSRZ outliers	101464	5026 (2.58-2.50)

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	281	4% / 60% / 33% / · 5%
2	B	162	6% / 65% / 23% / · 8%
3	C	109	% / 55% / 26% / 19%
4	D	122	71% / 14% / · 14%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 5015 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 Lysine-specific demethylase 2B.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	267	Total	C	N	O	S	Se	0	0	0
			2140	1344	389	384	16	7			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1056	GLY	-	expression tag	UNP Q8NHM5
A	1057	THR	-	expression tag	UNP Q8NHM5
A	1058	ARG	-	expression tag	UNP Q8NHM5

- Molecule 2 is a protein called S-phase kinase-associated protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	149	Total	C	N	O	S	Se	0	0	0
			1196	760	196	235	2	3			

- Molecule 3 is a protein called Polycomb group RING finger protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	88	Total	C	N	O	S	Se	0	0	0
			755	483	143	123	3	3			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	147	GLY	-	expression tag	UNP Q9BSM1
C	148	THR	-	expression tag	UNP Q9BSM1
C	149	ARG	-	expression tag	UNP Q9BSM1

- Molecule 4 is a protein called BCL-6 corepressor-like protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	D	105	Total 860	C 555	N 143	O 157	S 3	Se 2	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1590	MSE	-	initiating methionine	UNP Q5H9F3
D	1591	GLU	-	expression tag	UNP Q5H9F3
D	1592	THR	-	expression tag	UNP Q5H9F3
D	1593	ARG	-	expression tag	UNP Q5H9F3

- Molecule 5 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	22	Total 22	O 22	0	0
5	B	10	Total 10	O 10	0	0
5	C	16	Total 16	O 16	0	0
5	D	16	Total 16	O 16	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 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.
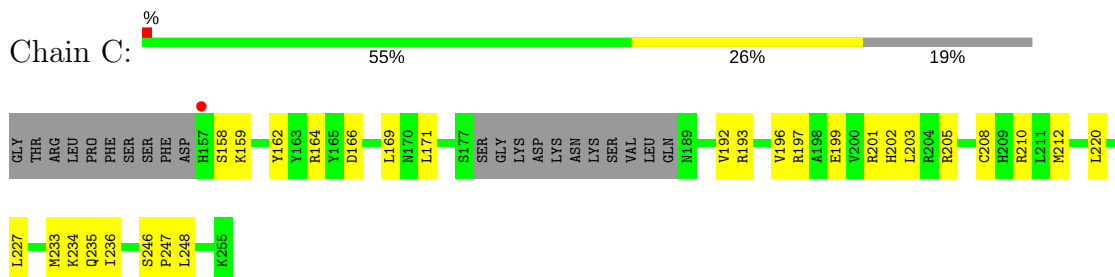
• Molecule 1: Lysine-specific demethylase 2B
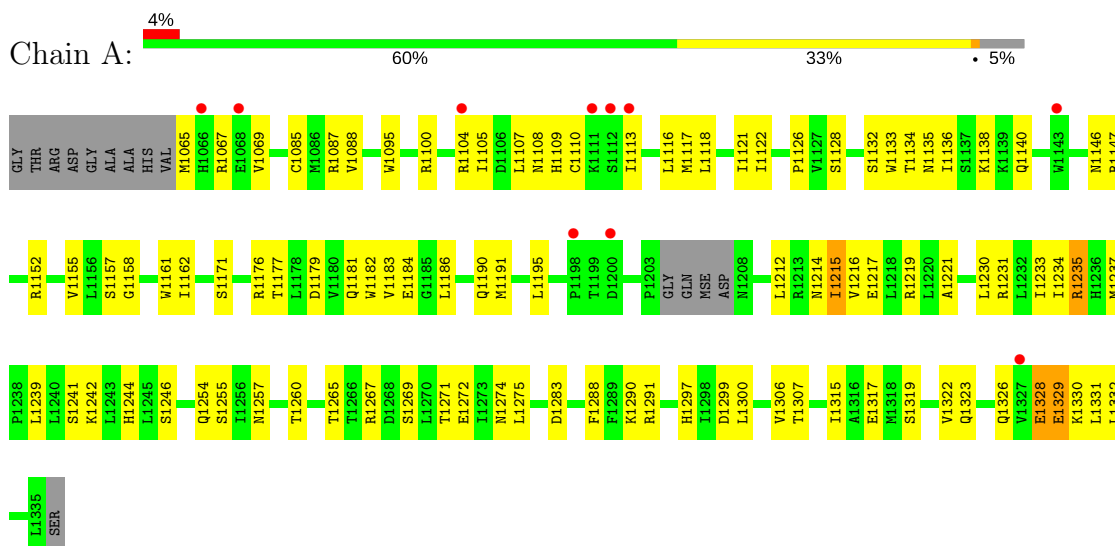
• Molecule 2: S-phase kinase-associated protein 1
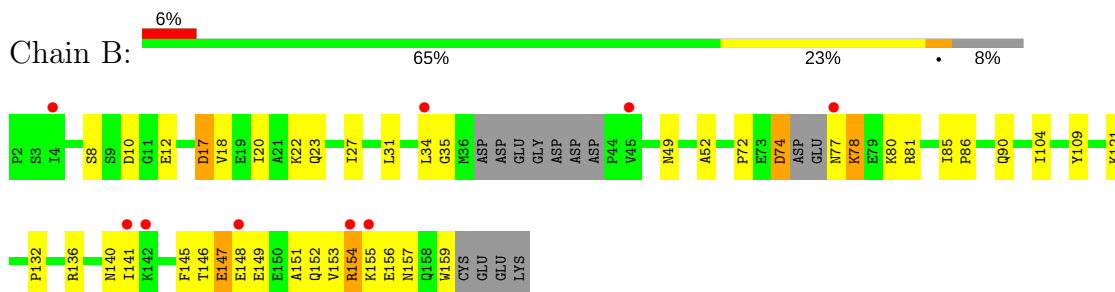
• Molecule 3: Polycomb group RING finger protein 1
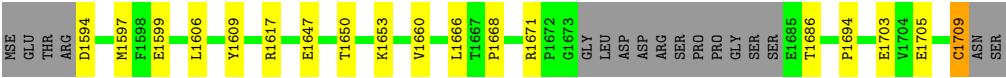
• Molecule 4: BCL-6 corepressor-like protein 1

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	70.05Å 73.80Å 123.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	63.28 – 2.55 63.28 – 2.55	Depositor EDS
% Data completeness (in resolution range)	99.5 (63.28-2.55) 99.7 (63.28-2.55)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.46 (at 2.55Å)	Xtriage
Refinement program	PHENIX (1.10_2155)	Depositor
R, R_{free}	0.206 , 0.257 0.204 , 0.257	Depositor DCC
R_{free} test set	3724 reflections (9.32%)	DCC
Wilson B-factor (Å^2)	45.6	Xtriage
Anisotropy	0.219	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 40.9	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.022 for k,h,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5015	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.45% 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	0.27	0/2171	0.45	0/2927
2	B	0.30	0/1214	0.44	0/1636
3	C	0.26	0/771	0.49	0/1034
4	D	0.27	0/879	0.43	0/1185
All	All	0.28	0/5035	0.45	0/6782

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	2140	0	2192	76	0
2	B	1196	0	1200	38	0
3	C	755	0	764	23	0
4	D	860	0	851	16	0
5	A	22	0	0	0	0
5	B	10	0	0	0	0
5	C	16	0	0	0	0
5	D	16	0	0	0	0
All	All	5015	0	5007	136	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 14.

The worst 5 of 136 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:1306:VAL:O	1:A:1330:LYS:NZ	1.77	1.18
2:B:152:GLN:O	2:B:156:GLU:HB2	1.62	0.98
3:C:193:ARG:NH2	4:D:1606:LEU:O	2.05	0.87
2:B:152:GLN:O	2:B:156:GLU:CB	2.25	0.84
2:B:152:GLN:O	2:B:156:GLU:N	2.13	0.82

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	263/281 (94%)	254 (97%)	9 (3%)	0	100	100
2	B	143/162 (88%)	139 (97%)	4 (3%)	0	100	100
3	C	84/109 (77%)	84 (100%)	0	0	100	100
4	D	101/122 (83%)	101 (100%)	0	0	100	100
All	All	591/674 (88%)	578 (98%)	13 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	250/251 (100%)	245 (98%)	5 (2%)	60	82
2	B	137/146 (94%)	128 (93%)	9 (7%)	19	34
3	C	85/101 (84%)	85 (100%)	0	100	100
4	D	98/110 (89%)	97 (99%)	1 (1%)	80	92
All	All	570/608 (94%)	555 (97%)	15 (3%)	51	76

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	23	GLN
2	B	74	ASP
2	B	157	ASN
2	B	17	ASP
2	B	154	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 (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, 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	260/281 (92%)	0.13	10 (3%)	41	44	27, 47, 69, 79	0
2	B	146/162 (90%)	0.23	9 (6%)	21	23	31, 55, 84, 91	0
3	C	85/109 (77%)	-0.22	1 (1%)	79	81	26, 37, 54, 72	0
4	D	103/122 (84%)	-0.17	0	100	100	26, 37, 54, 68	0
All	All	594/674 (88%)	0.05	20 (3%)	46	50	26, 45, 72, 91	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1111	LYS	5.0
1	A	1198	PRO	4.0
2	B	34	LEU	3.5
1	A	1113	ILE	3.1
2	B	148	GLU	2.8

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.

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

6.5 Other polymers ⓘ

There are no such residues in this entry.