Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 04:57 am BST

PDB ID	:	4GND
Title	:	Crystal Structure of NSD3 tandem PHD5-C5HCH domains
Authors	:	Li, F.; He, C.; Wu, J.; Shi, Y.
Deposited on	:	2012-08-17
Resolution	:	2.27 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 ①) were used in the production of this report:

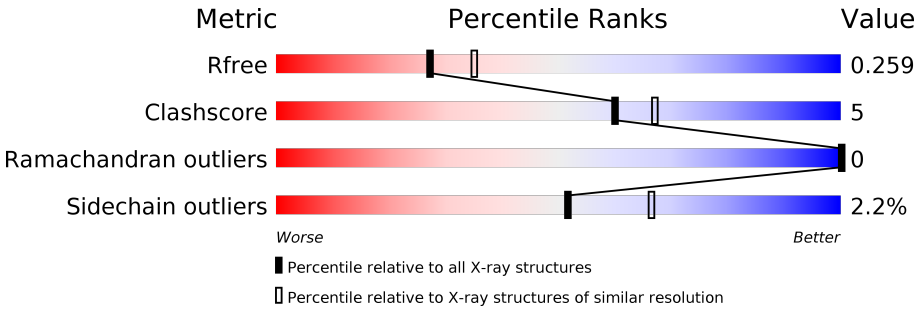
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.27 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	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)

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%

Mol	Chain	Length	Quality of chain
1	A	107	77% 15% 8%
1	C	107	92% 7% .

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 1629 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 Histone-lysine N-methyltransferase NSD3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	98	Total	C	N	O	S	0	0	0
			762	475	127	144	16			
1	C	107	Total	C	N	O	S	0	0	0
			834	518	142	158	16			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1307	SER	-	EXPRESSION TAG	UNP Q9BZ95
A	1308	ASN	-	EXPRESSION TAG	UNP Q9BZ95
A	1309	ALA	-	EXPRESSION TAG	UNP Q9BZ95
C	1307	SER	-	EXPRESSION TAG	UNP Q9BZ95
C	1308	ASN	-	EXPRESSION TAG	UNP Q9BZ95
C	1309	ALA	-	EXPRESSION TAG	UNP Q9BZ95

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	4	Total	Zn	0	0
			4	4		
2	C	4	Total	Zn	0	0
			4	4		

- Molecule 3 is water.

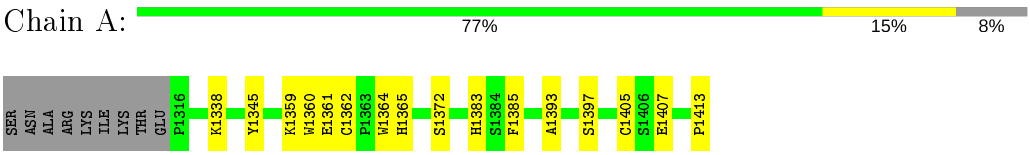
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	6	Total	O	0	0
			6	6		
3	C	19	Total	O	0	0
			19	19		

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. 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. 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.

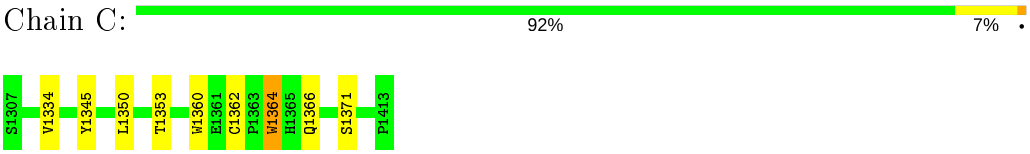
- Molecule 1: Histone-lysine N-methyltransferase NSD3

Chain A:



- Molecule 1: Histone-lysine N-methyltransferase NSD3

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α, β, γ	58.51Å 58.51Å 108.58Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.67 – 2.27 50.67 – 2.27	Depositor EDS
% Data completeness (in resolution range)	100.0 (50.67-2.27) 99.8 (50.67-2.27)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	4.92 (at 2.27Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.172 , 0.219 0.220 , 0.259	Depositor DCC
R_{free} test set	706 reflections (7.24%)	wwPDB-VP
Wilson B-factor (Å^2)	22.6	Xtriage
Anisotropy	0.048	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 36.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.088 for h,-h-k,-l	Xtriage
Reported twinning fraction	0.557 for H, K, L 0.443 for K, H, -L	Depositor
Outliers	0 of 9756 reflections	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	1629	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.80% 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

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.58	1/787 (0.1%)	0.58	0/1063
1	C	0.59	2/859 (0.2%)	0.63	0/1159
All	All	0.59	3/1646 (0.2%)	0.60	0/2222

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	1360	TRP	CD2-CE2	5.14	1.47	1.41
1	C	1364	TRP	CD2-CE2	5.13	1.47	1.41
1	A	1360	TRP	CD2-CE2	5.02	1.47	1.41

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	762	0	686	7	0
1	C	834	0	764	7	0
2	A	4	0	0	0	0
2	C	4	0	0	0	0
3	A	6	0	0	0	0
3	C	19	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	1629	0	1450	14	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 (14) 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:1334:VAL:HG11	1:C:1350:LEU:CD1	2.27	0.65
1:C:1366:GLN:HG3	1:C:1371:SER:HA	1.80	0.63
1:C:1334:VAL:HG11	1:C:1350:LEU:HD11	1.91	0.53
1:C:1345:TYR:CE2	1:C:1362:CYS:HB2	2.45	0.51
1:A:1383:HIS:CD2	1:A:1407:GLU:HG3	2.46	0.51
1:C:1334:VAL:HG11	1:C:1350:LEU:HD12	1.92	0.50
1:C:1334:VAL:HG12	1:C:1345:TYR:O	2.11	0.50
1:A:1345:TYR:CE2	1:A:1362:CYS:HB2	2.49	0.48
1:C:1334:VAL:CG1	1:C:1350:LEU:HD11	2.43	0.48
1:A:1338:LYS:HG2	1:A:1365:HIS:CE1	2.52	0.45
1:A:1385:PHE:CE2	1:A:1393:ALA:HB1	2.53	0.44
1:A:1397:SER:HB2	1:A:1413:PRO:HG3	2.00	0.44
1:A:1359:LYS:HE2	1:A:1361:GLU:HG3	2.02	0.41
1:A:1405:CYS:SG	1:A:1407:GLU:HB2	2.61	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	96/107 (90%)	95 (99%)	1 (1%)	0	100	100
1	C	105/107 (98%)	102 (97%)	3 (3%)	0	100	100
All	All	201/214 (94%)	197 (98%)	4 (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	86/94 (92%)	84 (98%)	2 (2%)	50	65
1	C	94/94 (100%)	92 (98%)	2 (2%)	53	68
All	All	180/188 (96%)	176 (98%)	4 (2%)	52	66

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1364	TRP
1	A	1372	SER
1	C	1353	THR
1	C	1364	TRP

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

Mol	Chain	Res	Type
1	C	1351	ASN

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

Of 8 ligands modelled in this entry, 8 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

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.