



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 12:36 AM GMT

PDB ID : 2B2U  
Title : Tandem chromodomains of human CHD1 complexed with Histone H3 Tail containing trimethyllysine 4 and dimethylarginine 2  
Authors : Flanagan IV, J.F.; Mi, L.-Z.; Chruszcz, M.; Cymborowski, M.; Clines, K.L.; Kim, Y.; Minor, W.; Rastinejad, F.; Khorasanizadeh, S.  
Deposited on : 2005-09-19  
Resolution : 2.95 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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

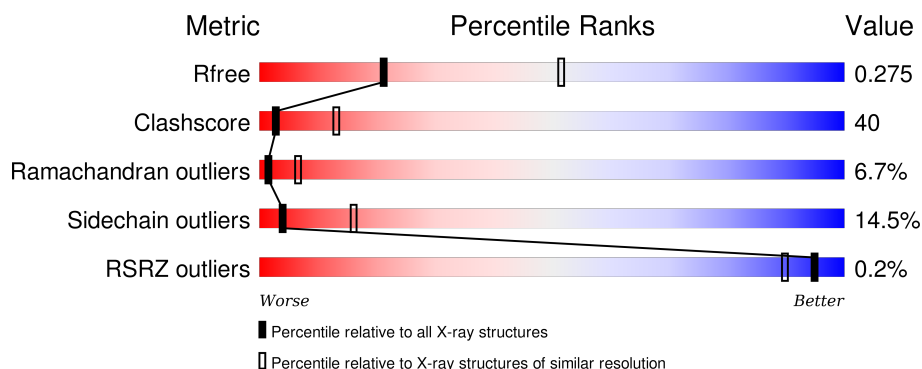
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.95 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	2184 (3.00-2.92)
Clashscore	102246	2552 (3.00-2.92)
Ramachandran outliers	100387	2468 (3.00-2.92)
Sidechain outliers	100360	2471 (3.00-2.92)
RSRZ outliers	91569	2201 (3.00-2.92)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	187	<div> <div>44%</div> <div>40%</div> <div>9%</div> <div>7%</div> </div>
1	B	187	<div> <div>39%</div> <div>36%</div> <div>15%</div> <div>8%</div> </div>
2	C	115	<div> <div>33%</div> <div>27%</div> <div>15%</div> <div>24%</div> </div>
3	D	16	<div> <div>6%</div> <div>19%</div> <div>6%</div> <div>69%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3568 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 Chromodomain-helicase-DNA-binding protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	173	Total	C	N	O	S	0	0	0
			1429	902	245	275	7			
1	B	172	Total	C	N	O	S	0	0	0
			1402	884	238	273	7			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	CLONING ARTIFACT	UNP O14646
A	2	LYS	-	CLONING ARTIFACT	UNP O14646
A	3	LYS	-	CLONING ARTIFACT	UNP O14646
A	4	HIS	-	EXPRESSION TAG	UNP O14646
A	5	HIS	-	EXPRESSION TAG	UNP O14646
A	6	HIS	-	EXPRESSION TAG	UNP O14646
A	7	HIS	-	EXPRESSION TAG	UNP O14646
A	8	HIS	-	EXPRESSION TAG	UNP O14646
A	9	HIS	-	EXPRESSION TAG	UNP O14646
A	186	LYS	-	CLONING ARTIFACT	UNP O14646
A	187	LYS	-	CLONING ARTIFACT	UNP O14646
B	1	MET	-	CLONING ARTIFACT	UNP O14646
B	2	LYS	-	CLONING ARTIFACT	UNP O14646
B	3	LYS	-	CLONING ARTIFACT	UNP O14646
B	4	HIS	-	EXPRESSION TAG	UNP O14646
B	5	HIS	-	EXPRESSION TAG	UNP O14646
B	6	HIS	-	EXPRESSION TAG	UNP O14646
B	7	HIS	-	EXPRESSION TAG	UNP O14646
B	8	HIS	-	EXPRESSION TAG	UNP O14646
B	9	HIS	-	EXPRESSION TAG	UNP O14646
B	186	LYS	-	CLONING ARTIFACT	UNP O14646
B	187	LYS	-	CLONING ARTIFACT	UNP O14646

- Molecule 2 is a protein called Chromodomain-helicase-DNA-binding protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	87	Total	C	N	O	S	0	0	0
			691	432	123	133	3			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MET	-	CLONING ARTIFACT	UNP O14646
C	2	LYS	-	CLONING ARTIFACT	UNP O14646
C	3	LYS	-	CLONING ARTIFACT	UNP O14646
C	4	HIS	-	EXPRESSION TAG	UNP O14646
C	5	HIS	-	EXPRESSION TAG	UNP O14646
C	6	HIS	-	EXPRESSION TAG	UNP O14646
C	7	HIS	-	EXPRESSION TAG	UNP O14646
C	8	HIS	-	EXPRESSION TAG	UNP O14646
C	9	HIS	-	EXPRESSION TAG	UNP O14646

- Molecule 3 is a protein called Histone H3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	D	5	Total	C	N	O	0	0	0
			46	29	10	7			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	2	DA2	ARG	MODIFIED RESIDUE	UNP P68431
D	4	M3L	LYS	MODIFIED RESIDUE	UNP P68431
D	16	TYR	-	SEE REMARK 999	UNP P68431



A1	X2	T3	K4	Q5	THR	ALA	ARG	LYS	SER	THR	GLY	GLY	LYS	ALA	TYR
----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	111.98Å 54.75Å 100.27Å 90.00° 112.32° 90.00°	Depositor
Resolution (Å)	20.00 – 2.95 32.54 – 2.95	Depositor EDS
% Data completeness (in resolution range)	92.4 (20.00-2.95) 94.2 (32.54-2.95)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.72 (at 2.95Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.238 , 0.290 0.238 , 0.275	Depositor DCC
$R_{free}$ test set	813 reflections (7.74%)	DCC
Wilson B-factor (Å <sup>2</sup> )	47.2	Xtriage
Anisotropy	0.377	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 40.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 11356 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	3568	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.34% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: M3L, DA2

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	0.51	0/1464	0.80	3/1969 (0.2%)
1	B	0.50	0/1435	0.89	6/1934 (0.3%)
2	C	0.55	0/705	0.98	4/949 (0.4%)
3	D	0.51	0/19	0.83	0/23
All	All	0.51	0/3623	0.87	13/4875 (0.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1
3	D	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	84	ASN	CB-CA-C	-12.35	85.71	110.40
1	B	47	ALA	CB-CA-C	-8.58	97.23	110.10
2	C	14	GLU	N-CA-C	8.51	133.97	111.00
1	A	47	ALA	CB-CA-C	-8.36	97.56	110.10
1	A	47	ALA	N-CA-C	6.99	129.86	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	B	182	TYR	Peptide
3	D	4	M3L	Peptide

## 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	1429	0	1362	108	0
1	B	1402	0	1311	105	0
2	C	691	0	655	65	0
3	D	46	0	55	8	0
All	All	3568	0	3383	278	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 40.

The worst 5 of 278 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)
3:D:2:DA2:HCD1	3:D:2:DA2:C1	1.58	1.34
3:D:2:DA2:HC11	3:D:2:DA2:CD	1.47	1.33
1:B:181:GLU:O	1:B:183:PHE:HB3	1.19	1.25
1:B:181:GLU:O	1:B:183:PHE:CB	1.89	1.21
1:A:136:ILE:HD13	1:A:136:ILE:H	1.12	1.09

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	169/187 (90%)	145 (86%)	18 (11%)	6 (4%)	4	21
1	B	166/187 (89%)	130 (78%)	21 (13%)	15 (9%)	1	3
2	C	85/115 (74%)	70 (82%)	8 (9%)	7 (8%)	1	4
3	D	1/16 (6%)	0	1 (100%)	0	100	100
All	All	421/505 (83%)	345 (82%)	48 (11%)	28 (7%)	1	6

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	13	PHE
1	B	49	PHE
1	B	170	LEU
1	B	171	ILE
1	B	181	GLU

### 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	150/164 (92%)	137 (91%)	13 (9%)	13	40
1	B	145/164 (88%)	118 (81%)	27 (19%)	2	9
2	C	69/101 (68%)	56 (81%)	13 (19%)	2	8
3	D	2/9 (22%)	2 (100%)	0	100	100
All	All	366/438 (84%)	313 (86%)	53 (14%)	4	16

5 of 53 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	89	LYS
1	B	129	GLN
2	C	81	LYS
1	B	101	THR
1	B	106	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	117	ASN
1	B	119	GLN
1	B	176	GLN
1	B	83	GLN
1	B	156	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	DA2	D	2	3	10,12,13	1.23	1 (10%)	9,14,16	1.51	2 (22%)
3	M3L	D	4	3	10,11,12	1.15	0	12,14,16	0.80	1 (8%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	DA2	D	2	3	-	0/11/13/15	0/0/0/0
3	M3L	D	4	3	-	0/8/10/12	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	2	DA2	CB-CA	-3.17	1.50	1.53

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	2	DA2	CB-CG-CD	-2.93	103.11	112.13
3	D	2	DA2	O-C-CA	-2.72	118.40	125.49
3	D	4	M3L	O-C-CA	-2.09	120.05	125.49

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	2	DA2	6	0
3	D	4	M3L	2	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	173/187 (92%)	-0.37	1 (0%) 90 78	11, 35, 61, 73	0
1	B	172/187 (91%)	-0.14	0 100 100	14, 47, 67, 78	0
2	C	87/115 (75%)	-0.12	0 100 100	22, 42, 68, 80	0
3	D	3/16 (18%)	0.51	0 100 100	8, 8, 25, 33	0
All	All	435/505 (86%)	-0.22	1 (0%) 95 90	8, 42, 66, 80	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	187	LYS	2.2

### 6.2 Non-standard residues in protein, DNA, RNA chains [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, 95<sup>th</sup> 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(Å <sup>2</sup> )	Q<0.9
3	M3L	D	4	12/13	0.94	0.16	-	14,25,32,38	0
3	DA2	D	2	13/14	0.90	0.19	-	2,13,22,31	0

### 6.3 Carbohydrates [i](#)

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.