



# Full wwPDB X-ray Structure Validation Report i

May 29, 2014 – 12:05 AM EDT

PDB ID : 4O42  
Title : Tandem chromodomains of human CHD1 in complex with influenza NS1 C-terminal tail dimethylated at K229  
Authors : Qin, S.; Xu, C.; Tempel, W.; Arrowsmith, C.H.; Bountra, C.; Edwards, A.M.; Min, J.; Structural Genomics Consortium (SGC)  
Deposited on : 2013-12-18  
Resolution : 1.87 Å(reported)

This is a full wwPDB validation 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/ValidationPDFNotes.html>

---

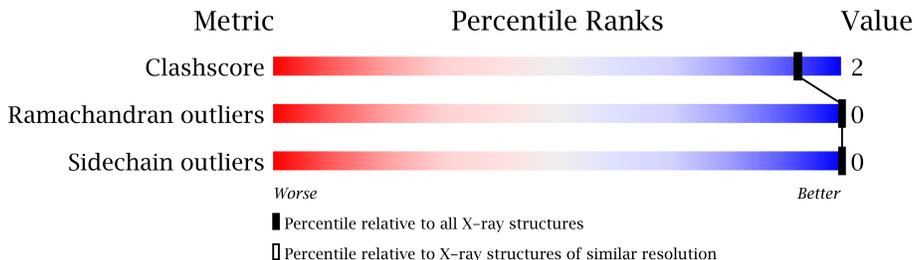
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : **FAILED**  
Xtriage (Phenix) : dev-1439  
EDS : **FAILED**  
Percentile statistics : 21963  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable23161

# 1 Overall quality at a glance

The reported resolution of this entry is 1.87 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))
Clashscore	79885	6268 (1.90-1.86)
Ramachandran outliers	78287	6195 (1.90-1.86)
Sidechain outliers	78261	6196 (1.90-1.86)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	194	
2	B	15	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 1563 atoms, of which 0 are hydrogen and 0 are deuterium.

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-bindingprotein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	173	1389	890	228	263	8	0	8	0

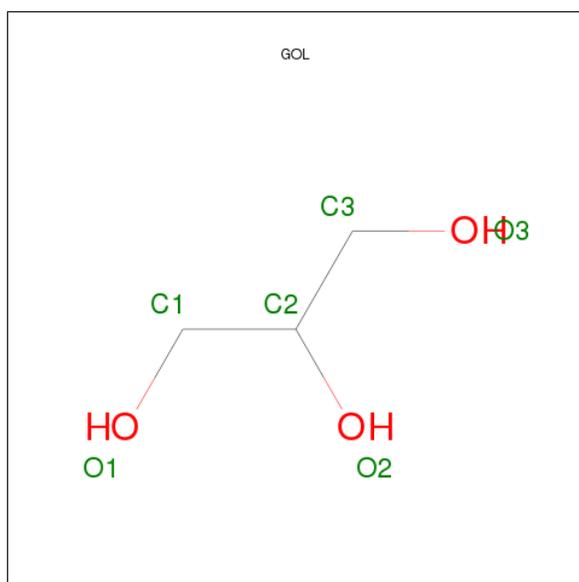
- Molecule 2 is a protein called Nonstructural protein 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	7	61	37	14	10	0	1	0

- Molecule 3 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	X	0	0
			1	1		
3	A	10	Total	X	0	0
			10	10		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C O 6 3 3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	92	Total O 92 92	0	0
5	B	4	Total O 4 4	0	0

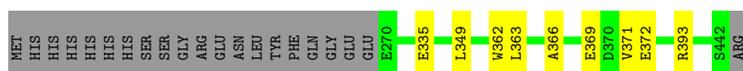
### 3 Residue-property plots i

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.

Note EDS failed to run properly.

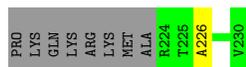
- Molecule 1: Chromodomain-helicase-DNA-bindingprotein 1

Chain A: 



- Molecule 2: Nonstructural protein 1

Chain B: 



## 4 Data and refinement statistics (i)

EDS failed to run properly - this section will therefore be incomplete.

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	110.07Å 46.06Å 46.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.06 – 1.87	Depositor
% Data completeness (in resolution range)	100.0 (46.06-1.87)	Depositor
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.71 (at 1.87Å)	Xtrriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.189 , 0.221	Depositor
Wilson B-factor (Å <sup>2</sup> )	30.6	Xtrriage
Anisotropy	0.303	Xtrriage
Estimated twinning fraction	0.033 for -h,l,k	Xtrriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Outliers	0 of 20189 reflections	Xtrriage
Total number of atoms	1563	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.66% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 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: UNX, GOL, MLY

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	0.72	0/1460	0.79	1/1981 (0.1%)
2	B	0.65	0/51	1.02	0/65
All	All	0.72	0/1511	0.80	1/2046 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	393	ARG	NE-CZ-NH2	-5.08	117.76	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1389	0	1251	4	0
2	B	61	0	72	1	0
3	A	10	0	0	0	0
3	B	1	0	0	0	0
4	B	6	0	8	1	0
5	A	92	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	4	0	0	0	0
All	All	1563	0	1331	5	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 2.

All (5) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:362:TRP:O	1:A:366:ALA:HB2	2.16	0.45
2:B:226:ALA:HA	4:B:302:GOL:O1	2.19	0.42
1:A:335:GLU:HG3	1:A:349:LEU:HD21	2.03	0.41
1:A:363:LEU:HD23	1:A:371:VAL:HG22	2.03	0.41
1:A:369:GLU:O	1:A:372:GLU:HB3	2.22	0.40

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	179/194 (92%)	178 (99%)	1 (1%)	0	100	100
2	B	5/15 (33%)	5 (100%)	0	0	100	100
All	All	184/209 (88%)	183 (100%)	1 (0%)	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	135/170 (79%)	135 (100%)	0	100	100
2	B	6/12 (50%)	6 (100%)	0	100	100
All	All	141/182 (78%)	141 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

### 5.3.3 RNA [i](#)

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

Mogul failed to run properly - this section will therefore be empty.

## 5.5 Carbohydrates [i](#)

Mogul failed to run properly - this section will therefore be empty.

## 5.6 Ligand geometry [i](#)

Mogul failed to run properly - this section will therefore be empty.

## 5.7 Other polymers [i](#)

Mogul failed to run properly - this section will therefore be empty.

## 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

EDS failed to run properly - this section will therefore be empty.

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

EDS failed to run properly - this section will therefore be empty.

### 6.3 Carbohydrates

EDS failed to run properly - this section will therefore be empty.

### 6.4 Ligands

EDS failed to run properly - this section will therefore be empty.

### 6.5 Other polymers

EDS failed to run properly - this section will therefore be empty.