



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:25 AM GMT

PDB ID : 2H1E  
Title : Tandem chromodomains of budding yeast CHD1  
Authors : Flanagan IV, J.F.; Khorasanizadeh, S.  
Deposited on : 2006-05-16  
Resolution : 2.20 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



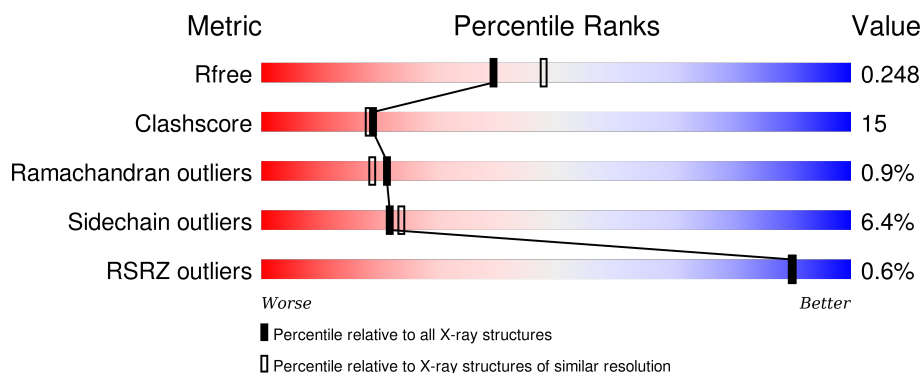
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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	177	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, green 67%, yellow 19%, orange 6%, red 6%, grey 0%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>%</span> <span>67%</span> <span>19%</span> <span>6%</span> <span>6%</span> </div> </div>
1	B	177	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, green 64%, yellow 24%, orange 8%, grey 0%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>%</span> <span>64%</span> <span>24%</span> <span>• • 8%</span> </div> </div>



## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2934 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 Chromo domain protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	166	Total	C	N	O	S	0	0	0
			1376	865	241	266	4			
1	B	163	Total	C	N	O	S	0	0	0
			1364	859	235	266	4			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP P32657
A	2	LYS	-	CLONING ARTIFACT	UNP P32657
A	3	LYS	-	CLONING ARTIFACT	UNP P32657
A	4	HIS	-	EXPRESSION TAG	UNP P32657
A	5	HIS	-	EXPRESSION TAG	UNP P32657
A	6	HIS	-	EXPRESSION TAG	UNP P32657
A	7	HIS	-	EXPRESSION TAG	UNP P32657
A	8	HIS	-	EXPRESSION TAG	UNP P32657
A	9	HIS	-	EXPRESSION TAG	UNP P32657
A	176	LYS	-	CLONING ARTIFACT	UNP P32657
A	177	LYS	-	CLONING ARTIFACT	UNP P32657
B	1	MET	-	INITIATING METHIONINE	UNP P32657
B	2	LYS	-	CLONING ARTIFACT	UNP P32657
B	3	LYS	-	CLONING ARTIFACT	UNP P32657
B	4	HIS	-	EXPRESSION TAG	UNP P32657
B	5	HIS	-	EXPRESSION TAG	UNP P32657
B	6	HIS	-	EXPRESSION TAG	UNP P32657
B	7	HIS	-	EXPRESSION TAG	UNP P32657
B	8	HIS	-	EXPRESSION TAG	UNP P32657
B	9	HIS	-	EXPRESSION TAG	UNP P32657
B	176	LYS	-	CLONING ARTIFACT	UNP P32657
B	177	LYS	-	CLONING ARTIFACT	UNP P32657

- Molecule 2 is water.



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	98	Total 98	O 98	0	0
2	B	96	Total 96	O 96	0	0





- Molecule 1: Chromo domain protein 1





## 4 Data and refinement statistics

Property	Value	Source
Space group	I 41	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	136.19Å 136.19Å 57.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.57 – 2.20 19.57 – 2.20	Depositor EDS
% Data completeness (in resolution range)	95.7 (19.57-2.20) 95.6 (19.57-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.53 (at 2.19Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.189 , 0.246 0.194 , 0.248	Depositor DCC
$R_{free}$ test set	1297 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	27.4	Xtriage
Anisotropy	0.023	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 24.9	EDS
Estimated twinning fraction	0.487 for -k,-h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 25796 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2934	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.51% 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](#)

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	1.04	0/1403	1.21	20/1898 (1.1%)
1	B	1.01	0/1392	1.03	9/1884 (0.5%)
All	All	1.02	0/2795	1.13	29/3782 (0.8%)

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	A	0	2
1	B	0	1
All	All	0	3

There are no bond length outliers.

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	151	ASN	CB-CA-C	-10.95	88.49	110.40
1	A	52	LYS	CD-CE-NZ	-10.66	87.17	111.70
1	A	96	PRO	N-CA-C	9.94	137.94	112.10
1	A	96	PRO	C-N-CA	7.74	141.05	121.70
1	A	151	ASN	N-CA-C	7.66	131.68	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	162	ASP	Peptide
1	A	87	ILE	Peptide
1	B	126	ILE	Peptide



## 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	1376	0	1316	49	0
1	B	1364	0	1306	34	0
2	A	98	0	0	5	0
2	B	96	0	0	5	0
All	All	2934	0	2622	82	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 15.

The worst 5 of 82 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:95:ASP:CB	1:A:96:PRO:HD3	1.36	1.30
1:A:95:ASP:CB	1:A:96:PRO:CD	2.08	1.29
1:B:134:LEU:HD23	1:B:138:THR:HG23	1.21	1.15
1:A:96:PRO:HB2	1:A:97:TYR:CB	1.90	1.00
1:A:146:LYS:HE2	1:A:151:ASN:CB	1.95	0.96

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	164/177 (93%)	153 (93%)	10 (6%)	1 (1%)	30	29
1	B	161/177 (91%)	154 (96%)	5 (3%)	2 (1%)	16	12

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	325/354 (92%)	307 (94%)	15 (5%)	3 (1%)	21	19

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	30	GLY
1	B	136	ASP
1	A	95	ASP

### 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	149/167 (89%)	141 (95%)	8 (5%)	27	31
1	B	150/167 (90%)	139 (93%)	11 (7%)	17	18
All	All	299/334 (90%)	280 (94%)	19 (6%)	22	24

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	24	LYS
1	B	78	LEU
1	B	134	LEU
1	A	172	LYS
1	B	135	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	170	GLN
1	B	170	GLN
1	B	60	HIS
1	A	60	HIS
1	B	13	HIS



### 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 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	166/177 (93%)	-0.47	1 (0%) 90 90	17, 29, 45, 53	0
1	B	163/177 (92%)	-0.40	1 (0%) 90 90	19, 30, 40, 44	0
All	All	329/354 (92%)	-0.43	2 (0%) 90 90	17, 29, 43, 53	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	151	ASN	2.2
1	B	97	TYR	2.1

### 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 carbohydrates 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.