



# wwPDB X-ray Structure Validation Summary Report

May 24, 2020 – 01:41 am BST

PDB ID : 6H02  
Title : Crystal structure of human Mediator subunit MED23  
Authors : Clantin, B.; Monte, D.; Villeret, V.  
Deposited on : 2018-07-06  
Resolution : 2.80 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

<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](#) ) 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

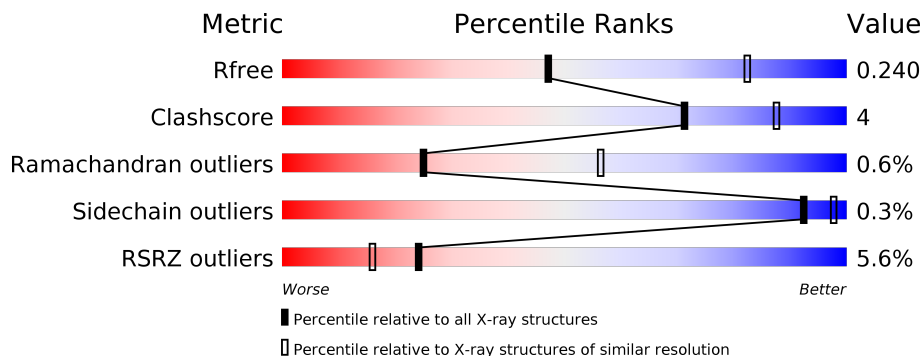
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*



The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 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  $\leq 5\%$ . 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	1382	
2	B	134	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 11714 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 Mediator of RNA polymerase II transcription subunit 23.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1334	10774	6967	1827	1909	71	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1369	ASP	-	expression tag	UNP Q9ULK4
A	1370	VAL	-	expression tag	UNP Q9ULK4
A	1371	LEU	-	expression tag	UNP Q9ULK4
A	1372	PHE	-	expression tag	UNP Q9ULK4
A	1373	GLN	-	expression tag	UNP Q9ULK4
A	1374	GLY	-	expression tag	UNP Q9ULK4
A	1375	PRO	-	expression tag	UNP Q9ULK4
A	1376	GLY	-	expression tag	UNP Q9ULK4
A	1377	HIS	-	expression tag	UNP Q9ULK4
A	1378	HIS	-	expression tag	UNP Q9ULK4
A	1379	HIS	-	expression tag	UNP Q9ULK4
A	1380	HIS	-	expression tag	UNP Q9ULK4
A	1381	HIS	-	expression tag	UNP Q9ULK4
A	1382	HIS	-	expression tag	UNP Q9ULK4

- Molecule 2 is a protein called LAMA nanobody NB106.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	123	940	588	165	181	6	0	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	75.65Å 137.18Å 90.11Å 90.00° 110.98° 90.00°	Depositor
Resolution (Å)	15.00 – 2.80 15.00 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.9 (15.00-2.80) 100.0 (15.00-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.30 (at 2.81Å)	Xtrriage
Refinement program	PHENIX 1.13_2998, PHENIX 1.13_2998	Depositor
R, $R_{free}$	0.199 , 0.240 0.199 , 0.240	Depositor DCC
$R_{free}$ test set	2010 reflections (4.79%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	75.8	Xtrriage
Anisotropy	0.590	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 64.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	11714	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	91.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.22% 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, 0.333 respectively for untwinned datasets, and 0.375, 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	0.24	0/11056	0.38	0/15023
2	B	0.24	0/959	0.46	0/1301
All	All	0.24	0/12015	0.39	0/16324

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 [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	10774	0	10838	71	0
2	B	940	0	913	12	0
All	All	11714	0	11751	83	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 4.

The worst 5 of 83 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)
2:B:47:VAL:HG21	2:B:50:ARG:HH11	1.55	0.72
1:A:45:ARG:HH22	1:A:89:THR:HG21	1.55	0.71

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:623:LEU:HD23	1:A:652:LEU:HD11	1.72	0.71
2:B:91:THR:HG22	2:B:122:VAL:H	1.54	0.70
1:A:517:SER:HB2	1:A:560:ALA:HB2	1.78	0.66

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	1332/1382 (96%)	1257 (94%)	67 (5%)	8 (1%)	25	56
2	B	121/134 (90%)	117 (97%)	4 (3%)	0	100	100
All	All	1453/1516 (96%)	1374 (95%)	71 (5%)	8 (1%)	25	56

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	327	SER
1	A	325	GLY
1	A	294	LYS
1	A	442	LYS
1	A	943	PRO

### 5.3.2 Protein sidechains [i](#)

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	1203/1244 (97%)	1200 (100%)	3 (0%)	93	98
2	B	96/106 (91%)	95 (99%)	1 (1%)	76	93
All	All	1299/1350 (96%)	1295 (100%)	4 (0%)	92	98

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

Mol	Chain	Res	Type
1	A	429	TRP
1	A	783	GLN
1	A	967	PHE
2	B	72	ARG

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

Mol	Chain	Res	Type
1	A	387	ASN

### 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

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	1334/1382 (96%)	0.09	78 (5%) 23 15	55, 84, 145, 273	0
2	B	123/134 (91%)	0.09	3 (2%) 59 49	57, 82, 111, 128	0
All	All	1457/1516 (96%)	0.09	81 (5%) 24 16	55, 84, 143, 273	0

The worst 5 of 81 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	28	ASP	9.3
1	A	29	THR	8.1
1	A	942	ASP	6.8
1	A	235	GLY	6.6
1	A	322	ASP	6.5

### 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.