



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:35 AM GMT

PDB ID : 2RC4
Title : Crystal Structure of the HAT domain of the human MOZ protein
Authors : Holbert, M.A.; Sikorski, T.; Snowflack, D.; Marmorstein, R.
Deposited on : 2007-09-19
Resolution : 3.00 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
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

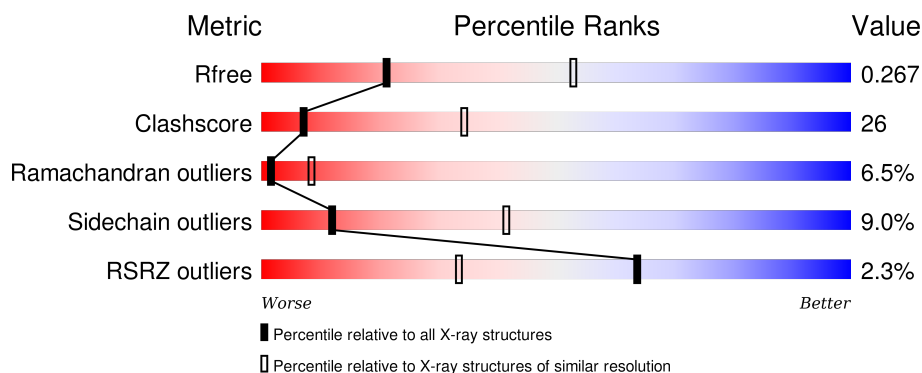
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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. 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	287	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	ACO	A	900	-	-	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2157 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 acetyltransferase MYST3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	264	Total	C	N	O	S	0	0	0
			2093	1350	353	373	17			

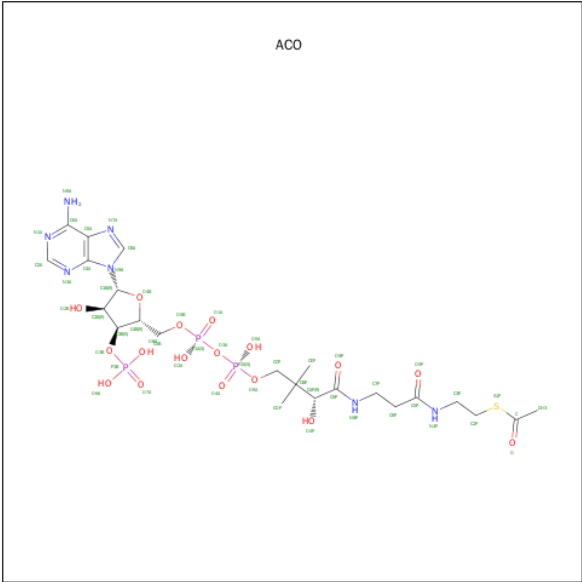
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	785	LYS	-	EXPRESSION TAG	UNP Q92794
A	786	LYS	-	EXPRESSION TAG	UNP Q92794
A	787	LYS	-	EXPRESSION TAG	UNP Q92794

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is ACETYL COENZYME *A (three-letter code: ACO) (formula: C₂₃H₃₈N₇O₁₇P₃S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
3	A	1	51	23	7	17	3	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	12	Total	0	0
			12 O		

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.

Chain A:

ARG	1669	PRO	581
SER	5670	ASP	583
ASP	6671	GLN	586
GLN	6672	THR	587
PHE	6673	LEU	588
VAL	6674	ILE	589
ARG	6675	GLY	590
ARG	6676	CYS	591
THR	6677	PRO	592
LEU	6678	ASP	593
ILE	6679	GLN	594
GLY	6680	THR	595
CYS	6681	LEU	596
PRO	6682	ASP	597
ASP	6683	GLN	598
GLN	6684	THR	599
THR	6685	LEU	600
LEU	6686	ASP	601
ASP	6687	GLN	602
GLN	6688	THR	603
THR	6689	LEU	604
LEU	6690	ASP	605
ASP	6691	GLN	606
GLN	6692	THR	607
THR	6693	LEU	608
LEU	6694	ASP	609
ASP	6695	GLN	610
GLN	6696	THR	611
THR	6697	LEU	612
LEU	6698	ASP	613
ASP	6699	GLN	614
GLN	6700	THR	615
THR	6701	LEU	616
LEU	6702	ASP	617
ASP	6703	GLN	618
GLN	6704	THR	619
THR	6705	LEU	620
LEU	6706	ASP	621
ASP	6707	GLN	622
GLN	6708	THR	623
THR	6709	LEU	624
LEU	6710	ASP	625
ASP	6711	GLN	626
GLN	6712	THR	627
THR	6713	LEU	628
LEU	6714	ASP	629
ASP	6715	GLN	630
GLN	6716	THR	631
THR	6717	LEU	632
LEU	6718	ASP	633
ASP	6719	GLN	634
GLN	6720	THR	635
THR	6721	LEU	636
LEU	6722	ASP	637
ASP	6723	GLN	638
GLN	6724	THR	639
THR	6725	LEU	640
LEU	6726	ASP	641
ASP	6727	GLN	642
GLN	6728	THR	643
THR	6729	LEU	644
LEU	6730	ASP	645
ASP	6731	GLN	646
GLN	6732	THR	647
THR	6733	LEU	648
LEU	6734	ASP	649
ASP	6735	GLN	650
GLN	6736	THR	651
THR	6737	LEU	652
LEU	6738	ASP	653
ASP	6739	GLN	654
GLN	6740	THR	655
THR	6741	LEU	656
LEU	6742	ASP	657
ASP	6743	GLN	658
GLN	6744	THR	659
THR	6745	LEU	660
LEU	6746	ASP	661
ASP	6747	GLN	662
GLN	6748	THR	663
THR	6749	LEU	664
LEU	6750	ASP	665
ASP	6751	GLN	666
GLN	6752	THR	667
THR	6753	LEU	668
LEU	6754	ASP	669
ASP	6755	GLN	670
GLN	6756	THR	671
THR	6757	LEU	672
LEU	6758	ASP	673
ASP	6759	GLN	674
GLN	6760	THR	675
THR	6761	LEU	676
LEU	6762	ASP	677
ASP	6763	GLN	678
GLN	6764	THR	679
THR	6765	LEU	680
LEU	6766	ASP	681
ASP	6767	GLN	682
GLN	6768	THR	683
THR	6769	LEU	684
LEU	6770	ASP	685
ASP	6771	GLN	686
GLN	6772	THR	687
THR	6773	LEU	688
LEU	6774	ASP	689
ASP	6775	GLN	690
GLN	6776	THR	691
THR	6777	LEU	692
LEU	6778	ASP	693
ASP	6779	GLN	694
GLN	6780	THR	695
THR	6781	LEU	696
LEU	6782	ASP	697

4 Data and refinement statistics

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, α , β , γ	109.20 Å 109.20 Å 144.78 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.00 43.59 – 3.00	Depositor EDS
% Data completeness (in resolution range)	96.6 (50.00-3.00) 96.6 (43.59-3.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.79 (at 3.01 Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.276 , 0.283 0.278 , 0.267	Depositor DCC
R_{free} test set	909 reflections (11.51%)	DCC
Wilson B-factor (Å ²)	79.5	Xtriage
Anisotropy	0.110	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 76.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 8805 reflections	Xtriage
F_o, F_c correlation	0.88	EDS
Total number of atoms	2157	wwPDB-VP
Average B, all atoms (Å ²)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.74% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² 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: ZN, ACO

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	0.45	0/2150	0.67	1/2914 (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	767	VAL	N-CA-C	-5.26	96.81	111.00

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	2093	0	1988	109	0
2	A	1	0	0	0	0
3	A	51	0	34	2	0
4	A	12	0	0	1	0
All	All	2157	0	2022	109	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 26.

The worst 5 of 109 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:765:ARG:HG2	1:A:766:PRO:HD2	1.28	1.14
1:A:765:ARG:HG2	1:A:766:PRO:CD	1.88	1.02
1:A:716:LEU:HD23	1:A:717:SER:H	1.25	0.97
1:A:731:LEU:HA	1:A:736:MET:HB2	1.55	0.89
1:A:711:ILE:HG22	1:A:712:SER:H	1.43	0.82

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	260/287 (91%)	196 (75%)	47 (18%)	17 (6%)	1 8

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	516	LYS
1	A	557	MET
1	A	706	GLN
1	A	716	LEU
1	A	717	SER

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	221/271 (82%)	201 (91%)	20 (9%)	12	41

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	642	TYR
1	A	678	SER
1	A	715	LYS
1	A	593	LEU
1	A	629	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	603	HIS
1	A	756	HIS
1	A	654	GLN
1	A	592	ASN
1	A	707	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](#)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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	ACO	A	900	-	43,53,53	0.83	2 (4%)	55,79,79	1.60	6 (10%)

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	ACO	A	900	-	-	0/47/67/67	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	900	ACO	O4B-C1B	2.10	1.43	1.41
3	A	900	ACO	C5A-C4A	2.63	1.46	1.40

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	900	ACO	N3A-C2A-N1A	-7.05	123.50	128.89
3	A	900	ACO	C4A-C5A-N7A	-4.38	105.45	109.48
3	A	900	ACO	P2A-O3A-P1A	-3.86	121.90	132.73
3	A	900	ACO	C3P-N4P-C5P	2.06	126.85	122.79
3	A	900	ACO	O4B-C1B-N9A	2.50	113.33	108.10

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	900	ACO	2	0

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 [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, 95th 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 ²)	Q<0.9
1	A	264/287 (91%)	-0.08	6 (2%) 64 33	43, 72, 115, 133	14 (5%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	706	GLN	3.5
1	A	708	ASP	3.0
1	A	738	ASP	2.2
1	A	705	HIS	2.1
1	A	748	ARG	2.0

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](#)

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, 95th 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 ²)	Q<0.9
3	ACO	A	900	51/51	0.90	0.30	3.40	36,52,68,70	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
2	ZN	A	1	1/1	0.94	0.04	-	85,85,85,85	0

6.5 Other polymers [i](#)

There are no such residues in this entry.