



# wwPDB X-ray Structure Validation Summary Report

Sep 24, 2020 – 04:09 PM BST

PDB ID : 6Y4Y  
Title : The crystal structure of human MACROD2 in space group P41212  
Authors : Wazir, S.; Maksimainen, M.M.; Lehtio, L.  
Deposited on : 2020-02-24  
Resolution : 1.75 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.14.6  
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) : 2.14.6

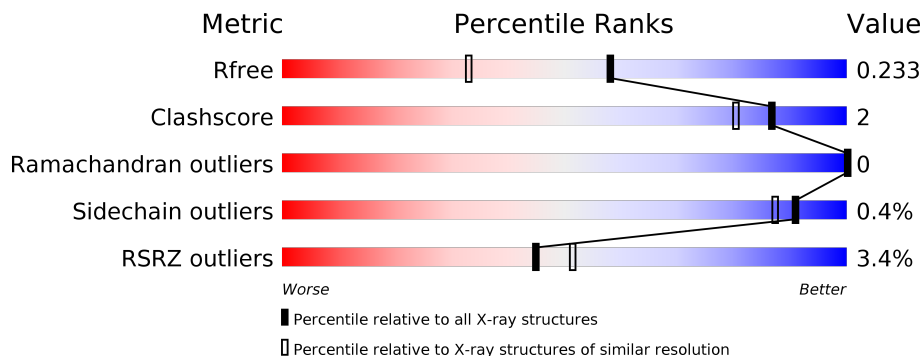
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.75 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	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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	366	<p>2% 57% 42%</p>
1	B	366	<p>2% 56% 41%</p>
1	C	366	<p>3% 54% 6% 40%</p>
1	D	366	<p>2% 55% 41%</p>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7402 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 Thioredoxin 1,ADP-ribose glycohydrolase MACROD2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	214	Total 1703	C 1084	N 299	O 309	S 11	0	1	0
1	B	215	Total 1719	C 1095	N 300	O 312	S 12	0	3	0
1	C	218	Total 1743	C 1111	N 307	O 314	S 11	0	3	0
1	D	217	Total 1716	C 1093	N 299	O 313	S 11	0	0	0

There are 80 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-122	MET	-	initiating methionine	UNP P0AA25
A	-121	HIS	-	expression tag	UNP P0AA25
A	-120	HIS	-	expression tag	UNP P0AA25
A	-119	HIS	-	expression tag	UNP P0AA25
A	-118	HIS	-	expression tag	UNP P0AA25
A	-117	HIS	-	expression tag	UNP P0AA25
A	-116	HIS	-	expression tag	UNP P0AA25
A	-115	SER	-	expression tag	UNP P0AA25
A	-114	SER	-	expression tag	UNP P0AA25
A	-113	GLY	-	expression tag	UNP P0AA25
A	-3	GLY	-	linker	UNP P0AA25
A	-2	THR	-	linker	UNP P0AA25
A	-1	GLU	-	linker	UNP P0AA25
A	0	ASN	-	linker	UNP P0AA25
A	1	LEU	-	linker	UNP P0AA25
A	2	TYR	-	linker	UNP P0AA25
A	3	PHE	-	linker	UNP P0AA25
A	4	GLN	-	linker	UNP P0AA25
A	5	SER	-	linker	UNP P0AA25
A	6	MET	-	linker	UNP P0AA25
B	-122	MET	-	initiating methionine	UNP P0AA25

*Continued on next page...*

*Continued from previous page...*

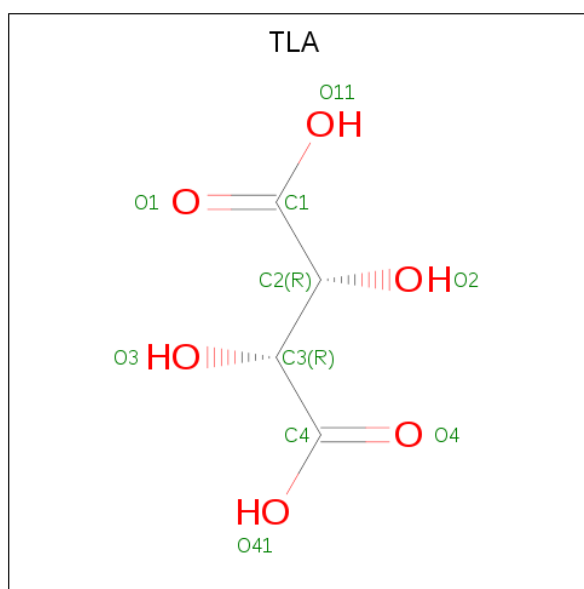
Chain	Residue	Modelled	Actual	Comment	Reference
B	-121	HIS	-	expression tag	UNP P0AA25
B	-120	HIS	-	expression tag	UNP P0AA25
B	-119	HIS	-	expression tag	UNP P0AA25
B	-118	HIS	-	expression tag	UNP P0AA25
B	-117	HIS	-	expression tag	UNP P0AA25
B	-116	HIS	-	expression tag	UNP P0AA25
B	-115	SER	-	expression tag	UNP P0AA25
B	-114	SER	-	expression tag	UNP P0AA25
B	-113	GLY	-	expression tag	UNP P0AA25
B	-3	GLY	-	linker	UNP P0AA25
B	-2	THR	-	linker	UNP P0AA25
B	-1	GLU	-	linker	UNP P0AA25
B	0	ASN	-	linker	UNP P0AA25
B	1	LEU	-	linker	UNP P0AA25
B	2	TYR	-	linker	UNP P0AA25
B	3	PHE	-	linker	UNP P0AA25
B	4	GLN	-	linker	UNP P0AA25
B	5	SER	-	linker	UNP P0AA25
B	6	MET	-	linker	UNP P0AA25
C	-122	MET	-	initiating methionine	UNP P0AA25
C	-121	HIS	-	expression tag	UNP P0AA25
C	-120	HIS	-	expression tag	UNP P0AA25
C	-119	HIS	-	expression tag	UNP P0AA25
C	-118	HIS	-	expression tag	UNP P0AA25
C	-117	HIS	-	expression tag	UNP P0AA25
C	-116	HIS	-	expression tag	UNP P0AA25
C	-115	SER	-	expression tag	UNP P0AA25
C	-114	SER	-	expression tag	UNP P0AA25
C	-113	GLY	-	expression tag	UNP P0AA25
C	-3	GLY	-	linker	UNP P0AA25
C	-2	THR	-	linker	UNP P0AA25
C	-1	GLU	-	linker	UNP P0AA25
C	0	ASN	-	linker	UNP P0AA25
C	1	LEU	-	linker	UNP P0AA25
C	2	TYR	-	linker	UNP P0AA25
C	3	PHE	-	linker	UNP P0AA25
C	4	GLN	-	linker	UNP P0AA25
C	5	SER	-	linker	UNP P0AA25
C	6	MET	-	linker	UNP P0AA25
D	-122	MET	-	initiating methionine	UNP P0AA25
D	-121	HIS	-	expression tag	UNP P0AA25
D	-120	HIS	-	expression tag	UNP P0AA25

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-119	HIS	-	expression tag	UNP P0AA25
D	-118	HIS	-	expression tag	UNP P0AA25
D	-117	HIS	-	expression tag	UNP P0AA25
D	-116	HIS	-	expression tag	UNP P0AA25
D	-115	SER	-	expression tag	UNP P0AA25
D	-114	SER	-	expression tag	UNP P0AA25
D	-113	GLY	-	expression tag	UNP P0AA25
D	-3	GLY	-	linker	UNP P0AA25
D	-2	THR	-	linker	UNP P0AA25
D	-1	GLU	-	linker	UNP P0AA25
D	0	ASN	-	linker	UNP P0AA25
D	1	LEU	-	linker	UNP P0AA25
D	2	TYR	-	linker	UNP P0AA25
D	3	PHE	-	linker	UNP P0AA25
D	4	GLN	-	linker	UNP P0AA25
D	5	SER	-	linker	UNP P0AA25
D	6	MET	-	linker	UNP P0AA25

- Molecule 2 is L(+)-TARTARIC ACID (three-letter code: TLA) (formula: C<sub>4</sub>H<sub>6</sub>O<sub>6</sub>).

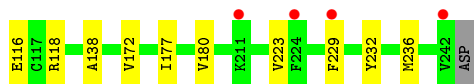


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	C	O	0	0
			10	4	6		

- Molecule 3 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
3	A	138	Total 138	O 138	0	0
3	B	135	Total 135	O 135	0	0
3	C	98	Total 98	O 98	0	0
3	D	140	Total 140	O 140	0	0





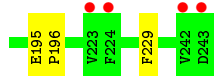
• Molecule 1: Thioredoxin 1,ADP-ribose glycohydrolase MACROD2



MET HIS HIS HIS HIS HIS HIS SER SER MET MET SER ASP ASP LYS ILE ILE HIS HIS LEU LEU THR ASP ASP PHE PHE ASP ASP THR VAL VAL ASP ASP PHE TRP TRP TRP TRP CYS CYS PRO CYS LYS MET MET ILE ALA PRO ILE LEU LEU ASP ASP GLU ILE ALA ASP ASP TYR

GLN GLY LYS LEU THR VAL ALA LYS LEU ASN ILE ASP GLN ASN PRO GLY THR ALA PRO LYS TYR GLY ILE ARG ILE PRO THR LEU LEU LEU LEU PHE LYS ASN GLY VAL ALA TRP LYS VAL GLY ALA LEU SER LYS LYS GLY GLN LEU LYS GLU PHE LEU LYS PHE LEU LEU

THR GLU ASN LEU TYR PHE GLN SER MET LYS LYS K9 R26 E29 Y30 L31 I40 M47 K48 GLY LYS GLY GLN ASN ASP GLU GLU ASN THR GLU THR SER GLN VAL LYS S67 R76 L82 E83 V84 L120 C132 G133 Y134 A138 V145 R150





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	95.65Å 95.65Å 258.72Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.83 – 1.75 47.83 – 1.75	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.83-1.75) 100.0 (47.83-1.75)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.79 (at 1.75Å)	Xtrriage
Refinement program	REFMAC 5.8.0257	Depositor
R, $R_{free}$	0.204 , 0.227 0.211 , 0.233	Depositor DCC
$R_{free}$ test set	6079 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.0	Xtrriage
Anisotropy	0.396	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 34.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7402	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.00% 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](#)

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

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	0.64	0/1736	0.77	0/2340
1	B	0.67	0/1755	0.79	1/2364 (0.0%)
1	C	0.66	0/1782	0.77	0/2399
1	D	0.66	0/1749	0.77	0/2358
All	All	0.66	0/7022	0.77	1/9461 (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	B	30	TYR	CB-CG-CD1	6.21	124.73	121.00

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	1703	0	1721	5	0
1	B	1719	0	1743	4	0
1	C	1743	0	1778	13	0
1	D	1716	0	1735	9	0
2	B	10	0	4	0	0
3	A	138	0	0	1	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	135	0	0	0	0
3	C	98	0	0	0	0
3	D	140	0	0	1	0
All	All	7402	0	6981	29	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 2.

The worst 5 of 29 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:C:223:VAL:HG21	1:C:229:PHE:HB2	1.82	0.62
1:A:232:TYR:O	1:A:236:MET:HG2	2.08	0.54
1:C:21:MET:HE3	1:C:26:ARG:HA	1.89	0.53
1:A:118[B]:ARG:O	1:D:150:ARG:HD3	2.09	0.52
1:D:76:ARG:HD2	1:D:229:PHE:CD1	2.44	0.52

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	211/366 (58%)	210 (100%)	1 (0%)	0	100	100
1	B	214/366 (58%)	211 (99%)	3 (1%)	0	100	100
1	C	217/366 (59%)	215 (99%)	2 (1%)	0	100	100
1	D	213/366 (58%)	210 (99%)	3 (1%)	0	100	100
All	All	855/1464 (58%)	846 (99%)	9 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	184/310 (59%)	184 (100%)	0	100	100
1	B	187/310 (60%)	185 (99%)	2 (1%)	73	60
1	C	189/310 (61%)	189 (100%)	0	100	100
1	D	186/310 (60%)	184 (99%)	2 (1%)	73	60
All	All	746/1240 (60%)	742 (100%)	4 (0%)	91	83

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

Mol	Chain	Res	Type
1	B	222[A]	CYS
1	B	222[B]	CYS
1	D	29	GLU
1	D	47	MET

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 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

1 ligand is 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
2	TLA	B	301	-	3,9,9	1.09	0	6,12,12	1.20	1 (16%)

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
2	TLA	B	301	-	-	0/4/12/12	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	B	301	TLA	C4-C3-C2	-2.18	108.41	113.11

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	214/366 (58%)	0.05	5 (2%) 60 67	22, 32, 58, 73	0
1	B	215/366 (58%)	0.15	6 (2%) 53 58	21, 30, 53, 76	0
1	C	218/366 (59%)	0.16	11 (5%) 28 34	24, 36, 67, 89	0
1	D	217/366 (59%)	0.20	7 (3%) 47 54	20, 30, 56, 85	0
All	All	864/1464 (59%)	0.14	29 (3%) 45 51	20, 32, 59, 89	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	242	VAL	8.9
1	C	31	LEU	4.3
1	B	31	LEU	4.3
1	A	31	LEU	3.9
1	D	224	PHE	3.7

### 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 monosaccharides 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
2	TLA	B	301	10/10	0.95	0.08	27,30,31,33	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.