



# Full wwPDB X-ray Structure Validation Report i

Dec 8, 2025 – 12:34 PM EST

PDB ID : 9MHJ / pdb\_00009mhj  
Title : The structure of native Zcp  
Authors : Bera, A.K.; Liyayi, I.K.; Criss, A.K.; Noinaj, N.  
Deposited on : 2024-12-12  
Resolution : 1.94 Å (reported)

This is a Full wwPDB X-ray Structure 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  
<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>  
with specific help available everywhere you see the i symbol.

The types of validation reports are described at  
<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references](#) ①) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.47

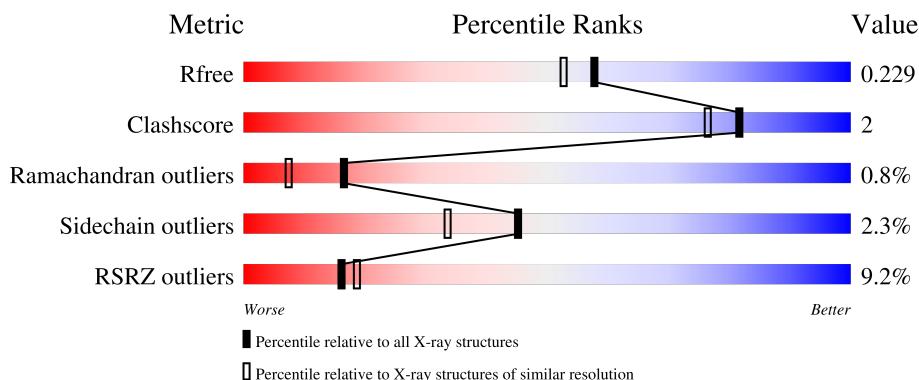
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## X-RAY DIFFRACTION

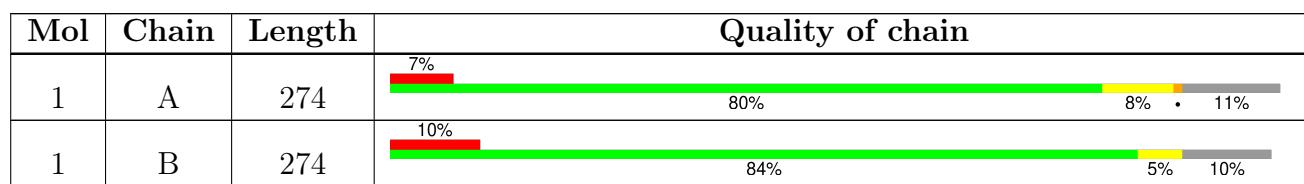
The reported resolution of this entry is 1.94 Å.

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))
$R_{free}$	164625	1306 (1.94-1.94)
Clashscore	180529	1400 (1.94-1.94)
Ramachandran outliers	177936	1387 (1.94-1.94)
Sidechain outliers	177891	1387 (1.94-1.94)
RSRZ outliers	164620	1306 (1.94-1.94)

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 of 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.



## 2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3942 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 Phosphoribosylglycinamide formyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	245	Total	C	N	O	S	0	0	0
			1914	1204	336	369	5			
1	B	246	Total	C	N	O	S	0	0	0
			1907	1202	334	366	5			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	SER	-	expression tag	UNP Q5F7W6
A	-4	TYR	-	expression tag	UNP Q5F7W6
A	-3	TYR	-	expression tag	UNP Q5F7W6
A	-2	HIS	-	expression tag	UNP Q5F7W6
A	-1	HIS	-	expression tag	UNP Q5F7W6
A	0	HIS	-	expression tag	UNP Q5F7W6
A	1	HIS	-	expression tag	UNP Q5F7W6
A	2	HIS	-	expression tag	UNP Q5F7W6
A	3	HIS	-	expression tag	UNP Q5F7W6
A	4	ASP	-	expression tag	UNP Q5F7W6
A	5	TYR	-	expression tag	UNP Q5F7W6
A	6	ASP	-	expression tag	UNP Q5F7W6
A	7	ILE	-	expression tag	UNP Q5F7W6
A	8	PRO	-	expression tag	UNP Q5F7W6
A	9	THR	-	expression tag	UNP Q5F7W6
A	10	THR	-	expression tag	UNP Q5F7W6
A	11	GLU	-	expression tag	UNP Q5F7W6
A	12	ASN	-	expression tag	UNP Q5F7W6
A	13	LEU	-	expression tag	UNP Q5F7W6
A	14	TYR	-	expression tag	UNP Q5F7W6
A	15	PHE	-	expression tag	UNP Q5F7W6
A	16	GLN	-	expression tag	UNP Q5F7W6
A	17	GLY	-	expression tag	UNP Q5F7W6
A	18	ALA	-	expression tag	UNP Q5F7W6
A	19	MET	-	expression tag	UNP Q5F7W6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	VAL	-	expression tag	UNP Q5F7W6
B	-5	SER	-	expression tag	UNP Q5F7W6
B	-4	TYR	-	expression tag	UNP Q5F7W6
B	-3	TYR	-	expression tag	UNP Q5F7W6
B	-2	HIS	-	expression tag	UNP Q5F7W6
B	-1	HIS	-	expression tag	UNP Q5F7W6
B	0	HIS	-	expression tag	UNP Q5F7W6
B	1	HIS	-	expression tag	UNP Q5F7W6
B	2	HIS	-	expression tag	UNP Q5F7W6
B	3	HIS	-	expression tag	UNP Q5F7W6
B	4	ASP	-	expression tag	UNP Q5F7W6
B	5	TYR	-	expression tag	UNP Q5F7W6
B	6	ASP	-	expression tag	UNP Q5F7W6
B	7	ILE	-	expression tag	UNP Q5F7W6
B	8	PRO	-	expression tag	UNP Q5F7W6
B	9	THR	-	expression tag	UNP Q5F7W6
B	10	THR	-	expression tag	UNP Q5F7W6
B	11	GLU	-	expression tag	UNP Q5F7W6
B	12	ASN	-	expression tag	UNP Q5F7W6
B	13	LEU	-	expression tag	UNP Q5F7W6
B	14	TYR	-	expression tag	UNP Q5F7W6
B	15	PHE	-	expression tag	UNP Q5F7W6
B	16	GLN	-	expression tag	UNP Q5F7W6
B	17	GLY	-	expression tag	UNP Q5F7W6
B	18	ALA	-	expression tag	UNP Q5F7W6
B	19	MET	-	expression tag	UNP Q5F7W6
B	20	VAL	-	expression tag	UNP Q5F7W6

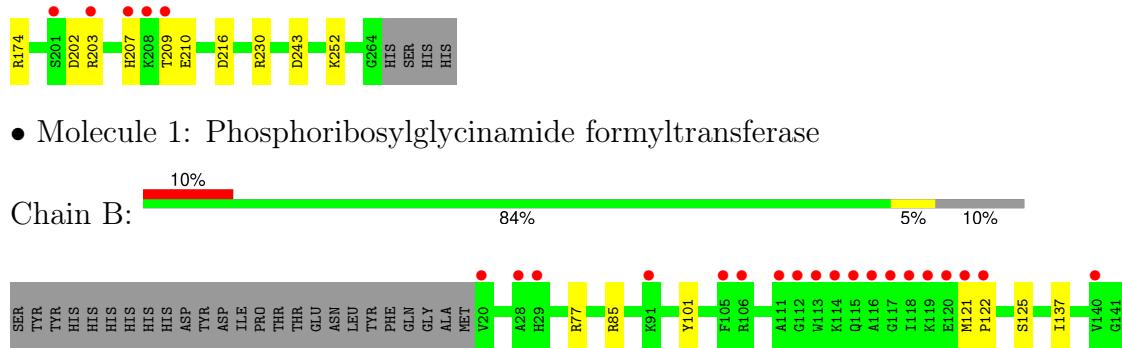
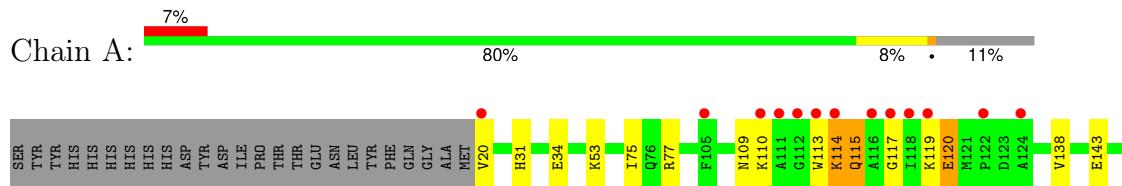
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	77	Total O 77 77	0	0
2	B	44	Total O 44 44	0	0

### 3 Residue-property plots [\(i\)](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

- Molecule 1: Phosphoribosylglycinamide formyltransferase



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	55.51 $\text{\AA}$ 92.17 $\text{\AA}$ 126.62 $\text{\AA}$ 90.00° 90.00° 90.00°	Depositor
Resolution ( $\text{\AA}$ )	27.76 – 1.94 27.76 – 1.94	Depositor EDS
% Data completeness (in resolution range)	99.2 (27.76-1.94) 99.1 (27.76-1.94)	Depositor EDS
$R_{\text{merge}}$	(Not available)	Depositor
$R_{\text{sym}}$	0.10	Depositor
$\langle I/\sigma(I) \rangle^1$	1.02 (at 1.95 $\text{\AA}$ )	Xtriage
Refinement program	PHENIX (1.18rc1_3777: ???)	Depositor
$R$ , $R_{\text{free}}$	0.194 , 0.225 0.198 , 0.229	Depositor DCC
$R_{\text{free}}$ test set	2000 reflections (4.12%)	wwPDB-VP
Wilson B-factor ( $\text{\AA}^2$ )	48.7	Xtriage
Anisotropy	0.124	Xtriage
Bulk solvent $k_{\text{sol}}$ (e/ $\text{\AA}^3$ ), $B_{\text{sol}}$ ( $\text{\AA}^2$ )	0.35 , 47.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3942	wwPDB-VP
Average B, all atoms ( $\text{\AA}^2$ )	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.60% 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	0.29	0/1958	0.57	1/2654 (0.0%)
1	B	0.24	0/1951	0.46	0/2646
All	All	0.27	0/3909	0.51	1/5300 (0.0%)

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	4
1	B	0	1
All	All	0	5

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	114	LYS	N-CA-C	5.60	122.74	110.80

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	113	TRP	Peptide
1	A	114	LYS	Peptide
1	A	115	GLN	Peptide
1	A	120	GLU	Peptide
1	B	121	MET	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	1914	0	1825	13	0
1	B	1907	0	1806	10	0
2	A	77	0	0	2	0
2	B	44	0	0	1	0
All	All	3942	0	3631	18	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.

All (18) 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:B:231:GLN:O	2:B:301:HOH:O	1.88	0.93
1:A:230:ARG:HH12	1:B:230:ARG:HH12	1.34	0.74
1:A:174:ARG:NH2	2:A:301:HOH:O	2.22	0.72
1:A:203:ARG:HB2	1:A:209:THR:HG21	1.76	0.67
1:A:230:ARG:NH1	1:B:230:ARG:HH12	1.95	0.65
1:A:230:ARG:HH12	1:B:230:ARG:NH1	1.98	0.62
1:A:31:HIS:O	1:A:34:GLU:HG2	2.10	0.50
1:A:117:GLY:O	1:B:242:ALA:HB1	2.15	0.46
1:B:205:LYS:C	1:B:207:HIS:H	2.23	0.46
1:A:243:ASP:OD1	1:A:252:LYS:NZ	2.43	0.46
1:B:137:ILE:HD11	1:B:162:PRO:HG3	1.97	0.45
1:A:75:ILE:HG22	1:A:77:ARG:HG2	1.98	0.45
1:A:53:LYS:HD3	1:A:53:LYS:HA	1.82	0.44
1:A:210:GLU:HB3	2:A:305:HOH:O	2.17	0.44
1:B:77:ARG:CZ	1:B:85:ARG:HH21	2.30	0.44
1:B:77:ARG:NE	1:B:85:ARG:HH21	2.16	0.43
1:A:216:ASP:OD2	1:B:208:LYS:HE2	2.18	0.43
1:A:202:ASP:O	1:A:207:HIS:ND1	2.52	0.43

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	243/274 (89%)	224 (92%)	16 (7%)	3 (1%)	11 3
1	B	244/274 (89%)	233 (96%)	10 (4%)	1 (0%)	30 22
All	All	487/548 (89%)	457 (94%)	26 (5%)	4 (1%)	16 7

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	115	GLN
1	A	120	GLU
1	A	119	LYS
1	B	122	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	200/236 (85%)	195 (98%)	5 (2%)	42 30
1	B	196/236 (83%)	192 (98%)	4 (2%)	50 38
All	All	396/472 (84%)	387 (98%)	9 (2%)	45 33

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	20	VAL
1	A	109	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	110	LYS
1	A	138	VAL
1	A	143	GLU
1	B	101	TYR
1	B	125	SER
1	B	208	LYS
1	B	265	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	29	HIS
1	A	212	GLN
1	B	136	ASN
1	B	142	HIS
1	B	212	GLN
1	B	265	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 oligosaccharides 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	245/274 (89%)	0.29	18 (7%) 22 25	37, 54, 138, 181	0
1	B	246/274 (89%)	0.66	27 (10%) 12 14	43, 67, 131, 160	0
All	All	491/548 (89%)	0.48	45 (9%) 16 19	37, 62, 135, 181	0

All (45) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	116	ALA	6.3
1	B	20	VAL	6.1
1	A	209	THR	5.9
1	A	116	ALA	4.8
1	A	118	ILE	4.7
1	A	208	LYS	4.7
1	A	20	VAL	4.3
1	B	206	THR	4.3
1	A	111	ALA	3.9
1	B	113	TRP	3.9
1	B	122	PRO	3.9
1	B	118	ILE	3.8
1	A	117	GLY	3.6
1	B	203	ARG	3.5
1	B	114	LYS	3.5
1	A	113	TRP	3.4
1	B	111	ALA	3.3
1	B	115	GLN	3.3
1	B	120	GLU	3.2
1	A	201	SER	3.2
1	B	204	SER	3.1
1	A	114	LYS	3.0
1	A	122	PRO	2.9
1	B	119	LYS	2.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	265	HIS	2.8
1	A	105	PHE	2.8
1	B	112	GLY	2.4
1	B	208	LYS	2.4
1	A	203	ARG	2.4
1	B	205	LYS	2.4
1	B	106	ARG	2.3
1	A	119	LYS	2.3
1	B	29	HIS	2.2
1	B	140	VAL	2.2
1	A	112	GLY	2.1
1	A	124	ALA	2.1
1	B	121	MET	2.1
1	B	117	GLY	2.1
1	B	91	LYS	2.1
1	B	105	PHE	2.1
1	A	207	HIS	2.0
1	B	142	HIS	2.0
1	A	110	LYS	2.0
1	B	264	GLY	2.0
1	B	28	ALA	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 oligosaccharides 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.