



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 10, 2024 – 12:09 AM EST

PDB ID : 4N9N  
Title : Crystal Structure of Saccharomyces cerevisiae Upc2 Transcription Factor fused with T4 Lysozyme  
Authors : Yang, H.; Im, Y.J.  
Deposited on : 2013-10-21  
Resolution : 2.90 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

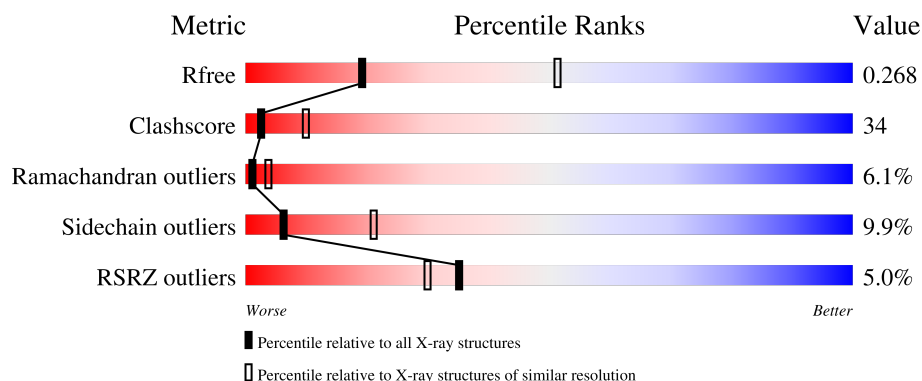
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	2335 (2.90-2.90)
Clashscore	180529	2564 (2.90-2.90)
Ramachandran outliers	177936	2514 (2.90-2.90)
Sidechain outliers	177891	2516 (2.90-2.90)
RSRZ outliers	164620	2337 (2.90-2.90)

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.

Mol	Chain	Length	Quality of chain
1	A	438	
1	B	438	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6829 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 Sterol uptake control protein 2, Lysozyme.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	428	Total	C	N	O	S	Se	0	0	0
			3400	2182	577	628	2	11			
1	B	428	Total	C	N	O	S	Se	0	0	0
			3400	2182	577	628	2	11			

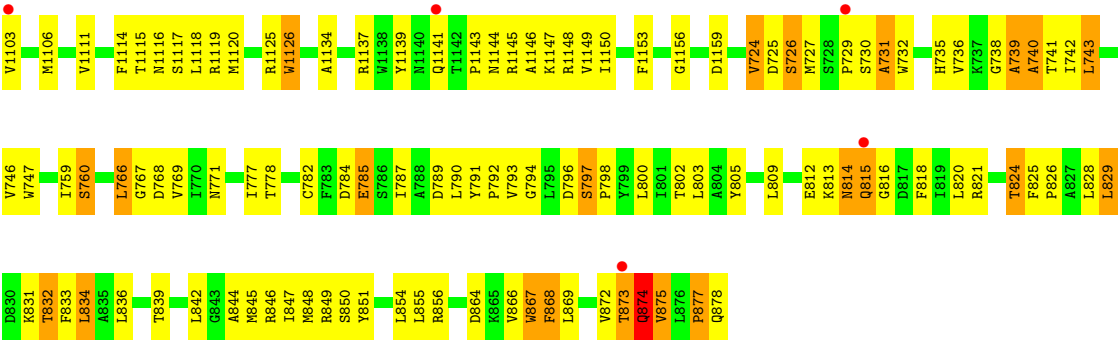
There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	594	GLY	-	expression tag	UNP Q12151
A	595	SER	-	expression tag	UNP Q12151
A	596	HIS	-	expression tag	UNP Q12151
A	597	MSE	-	expression tag	UNP Q12151
A	715	VAL	-	linker	UNP Q12151
A	716	ASP	-	linker	UNP Q12151
A	1054	THR	CYS	engineered mutation	UNP P00720
A	1097	ALA	CYS	engineered mutation	UNP P00720
A	724	VAL	-	linker	UNP Q12151
A	725	ASP	-	linker	UNP Q12151
B	594	GLY	-	expression tag	UNP Q12151
B	595	SER	-	expression tag	UNP Q12151
B	596	HIS	-	expression tag	UNP Q12151
B	597	MSE	-	expression tag	UNP Q12151
B	715	VAL	-	linker	UNP Q12151
B	716	ASP	-	linker	UNP Q12151
B	1054	THR	CYS	engineered mutation	UNP P00720
B	1097	ALA	CYS	engineered mutation	UNP P00720
B	724	VAL	-	linker	UNP Q12151
B	725	ASP	-	linker	UNP Q12151

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	21	Total 21	O 21	0	0
2	B	8	Total 8	O 8	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.17Å 67.17Å 257.51Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	58.17 – 2.90 58.17 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.7 (58.17-2.90) 99.8 (58.17-2.90)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.22 (at 2.91Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.251 , 0.272 0.249 , 0.268	Depositor DCC
$R_{free}$ test set	1442 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	69.4	Xtriage
Anisotropy	0.089	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 56.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.014 for -h,-k,l 0.062 for h,-h-k,-l 0.042 for -k,-h,-l	Xtriage
$F_o$ , $F_c$ correlation	0.92	EDS
Total number of atoms	6829	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.83% 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	0.35	0/3460	0.63	1/4678 (0.0%)
1	B	0.34	0/3460	0.64	2/4678 (0.0%)
All	All	0.35	0/6920	0.63	3/9356 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	869	LEU	CA-CB-CG	6.72	130.75	115.30
1	B	766	LEU	N-CA-C	-5.63	95.79	111.00
1	B	631	GLY	N-CA-C	-5.44	99.51	113.10

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	3400	0	3432	230	0
1	B	3400	0	3432	228	0
2	A	21	0	0	0	0
2	B	8	0	0	0	0
All	All	6829	0	6864	458	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 34.



The worst 5 of 458 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:715:VAL:HG12	1:A:716:ASP:OD2	1.40	1.18
1:A:846:ARG:HG3	1:A:846:ARG:HH11	1.04	1.18
1:B:797:SER:HB2	1:B:798:PRO:HA	1.11	1.09
1:B:825:PHE:HE2	1:B:848:MSE:HE1	1.24	1.00
1:B:1024:TYR:HB3	1:B:1035:LYS:HA	1.44	0.97

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	426/438 (97%)	355 (83%)	48 (11%)	23 (5%)	1	5
1	B	426/438 (97%)	338 (79%)	59 (14%)	29 (7%)	1	3
All	All	852/876 (97%)	693 (81%)	107 (13%)	52 (6%)	1	4

5 of 52 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	668	LEU
1	A	669	GLU
1	A	724	VAL
1	A	761	VAL
1	A	785	GLU

### 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	367/363 (101%)	329 (90%)	38 (10%)	5	18
1	B	367/363 (101%)	332 (90%)	35 (10%)	7	22
All	All	734/726 (101%)	661 (90%)	73 (10%)	6	21

5 of 73 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1094	VAL
1	B	864	ASP
1	B	725	ASP
1	B	814	ASN
1	A	752	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	771	ASN
1	B	874	GLN
1	B	651	HIS
1	B	675	HIS
1	B	694	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

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	417/438 (95%)	0.09	16 (3%) 44 38	28, 55, 91, 112	0
1	B	417/438 (95%)	0.40	26 (6%) 28 23	36, 70, 124, 131	0
All	All	834/876 (95%)	0.25	42 (5%) 35 30	28, 61, 114, 131	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	690	ILE	4.7
1	B	1042	ALA	4.4
1	A	795	LEU	4.3
1	A	877	PRO	4.2
1	B	1041	ALA	3.8

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.