



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 04:13 am BST

PDB ID : 1SH2  
Title : Crystal Structure of Norwalk Virus Polymerase (Metal-free, Centered Orthorhombic)  
Authors : Ng, K.K.; Pendas-Franco, N.; Rojo, J.; Boga, J.A.; Machin, A.; Alonso, J.M.; Parra, F.  
Deposited on : 2004-02-24  
Resolution : 2.30 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 ⓘ 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

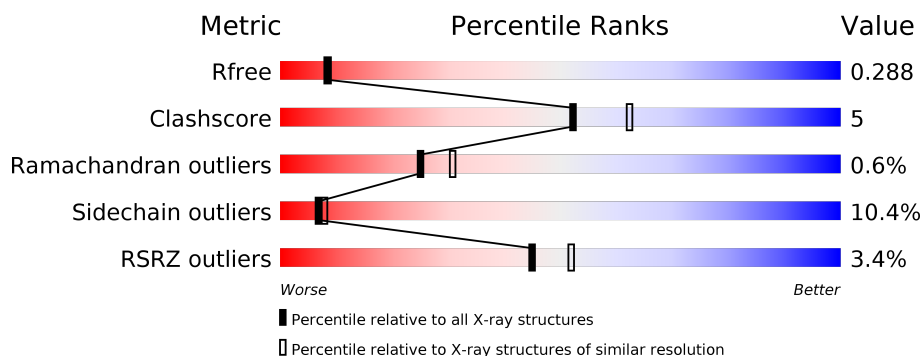
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	510	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4078 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 RNA Polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	502	Total	C	N	O	S	0	0	0
			3939	2501	675	741	22			

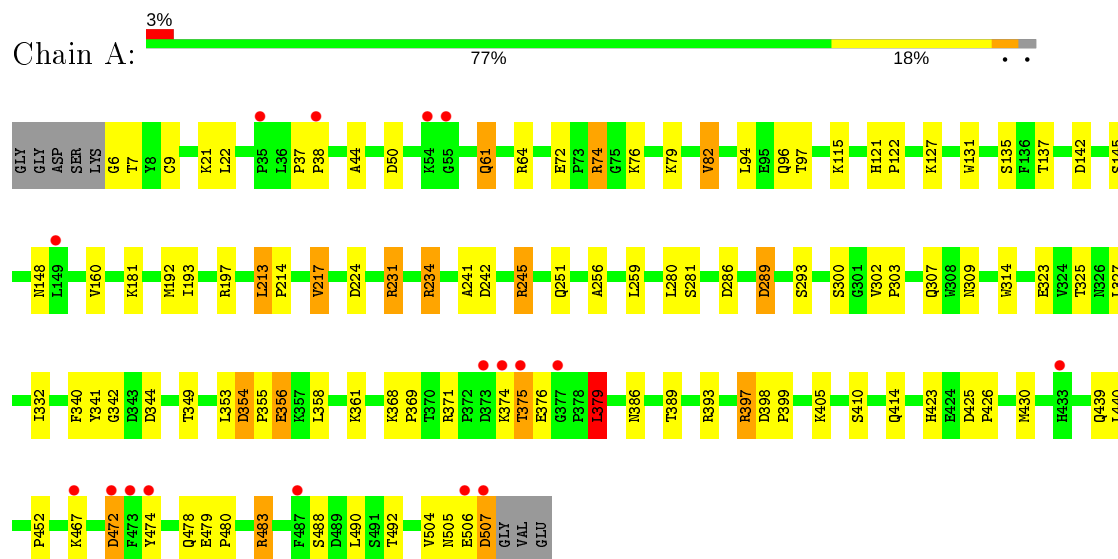
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	139	Total	O	0	0
			139	139		

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

These plots are drawn for all protein, RNA and DNA 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: RNA Polymerase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	114.27Å 115.24Å 91.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.04 – 2.30 35.93 – 2.30	Depositor EDS
% Data completeness (in resolution range)	100.0 (36.04-2.30) 98.8 (35.93-2.30)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.70 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, $R_{free}$	0.227 , 0.291 0.226 , 0.288	Depositor DCC
$R_{free}$ test set	1596 reflections (5.92%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.7	Xtriage
Anisotropy	0.199	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 30.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.032 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	4078	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.23% 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 ⓘ

### 5.1 Standard geometry ⓘ

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	0.39	1/4041 (0.0%)	0.71	10/5483 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	6	GLY	N-CA	5.31	1.54	1.46

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	379	LEU	CA-CB-CG	6.13	129.40	115.30
1	A	289	ASP	CB-CG-OD2	5.70	123.43	118.30
1	A	507	ASP	CB-CG-OD2	5.54	123.29	118.30
1	A	142	ASP	CB-CG-OD2	5.29	123.06	118.30
1	A	344	ASP	CB-CG-OD2	5.23	123.01	118.30
1	A	50	ASP	CB-CG-OD2	5.21	122.99	118.30
1	A	354	ASP	CB-CG-OD2	5.17	122.96	118.30
1	A	286	ASP	CB-CG-OD2	5.14	122.93	118.30
1	A	224	ASP	CB-CG-OD2	5.09	122.89	118.30
1	A	472	ASP	CB-CG-OD2	5.08	122.87	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts ⓘ

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	3939	0	3906	43	0
2	A	139	0	0	1	0
All	All	4078	0	3906	43	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 5.

All (43) 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:325:THR:HG21	1:A:332:ILE:HD11	1.54	0.89
1:A:148:ASN:HD21	1:A:197:ARG:HH11	1.35	0.73
1:A:386:ASN:HD21	1:A:397:ARG:HB3	1.54	0.71
1:A:309:ASN:HB3	1:A:342:GLY:HA2	1.78	0.66
1:A:452:PRO:HA	1:A:478:GLN:HE22	1.64	0.63
1:A:371:ARG:HB2	1:A:375:THR:H	1.66	0.61
1:A:74:ARG:HG2	1:A:251:GLN:HG2	1.82	0.60
1:A:398:ASP:HB2	1:A:399:PRO:HD2	1.84	0.59
1:A:97:THR:HG21	1:A:213:LEU:HD13	1.86	0.57
1:A:479:GLU:O	1:A:483:ARG:HG2	2.06	0.56
1:A:414:GLN:HE21	1:A:439:GLN:HE21	1.53	0.56
1:A:256:ALA:HB2	1:A:280:LEU:HD21	1.89	0.54
1:A:302:VAL:HG22	1:A:303:PRO:HD2	1.90	0.53
1:A:325:THR:HG23	1:A:327:LEU:H	1.76	0.49
1:A:79:LYS:HB3	1:A:82:VAL:HG13	1.94	0.48
1:A:148:ASN:ND2	1:A:197:ARG:HH11	2.08	0.48
1:A:127:LYS:HE2	1:A:192:MET:SD	2.53	0.48
1:A:231:ARG:O	1:A:234:ARG:HG2	2.14	0.47
1:A:389:THR:HA	1:A:393:ARG:O	2.14	0.47
1:A:358:LEU:HD23	1:A:379:LEU:HD21	1.97	0.47
1:A:217:VAL:HA	1:A:341:TYR:CE2	2.50	0.46
1:A:115:LYS:HE2	1:A:131:TRP:CE2	2.51	0.46
1:A:325:THR:HG23	1:A:327:LEU:N	2.31	0.46
1:A:121:HIS:HA	1:A:122:PRO:HA	1.72	0.45
1:A:214:PRO:HG2	1:A:314:TRP:CZ2	2.51	0.45
1:A:74:ARG:CG	1:A:251:GLN:HG2	2.46	0.44
1:A:44:ALA:HB1	1:A:181:LYS:HG3	2.01	0.43
1:A:242:ASP:O	1:A:369:PRO:HA	2.19	0.43
1:A:356:GLU:H	1:A:356:GLU:HG2	1.48	0.43
1:A:479:GLU:HB3	1:A:480:PRO:HD3	2.01	0.43
1:A:398:ASP:HB2	1:A:399:PRO:CD	2.47	0.42
1:A:349:THR:HG21	1:A:353:LEU:HD11	2.00	0.42

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:452:PRO:CA	1:A:478:GLN:HE22	2.30	0.42
1:A:245:ARG:HE	1:A:245:ARG:HB2	1.55	0.42
1:A:259:LEU:HD12	1:A:280:LEU:HD13	2.02	0.41
1:A:354:ASP:HA	1:A:355:PRO:HD3	1.89	0.41
1:A:425:ASP:HA	1:A:426:PRO:HD3	1.90	0.41
1:A:307:GLN:HG2	2:A:634:HOH:O	2.21	0.41
1:A:241:ALA:HB2	1:A:379:LEU:HD12	2.03	0.41
1:A:314:TRP:HA	1:A:340:PHE:CE2	2.56	0.40
1:A:61:GLN:HG2	1:A:64:ARG:NH1	2.36	0.40
1:A:323:GLU:HB3	1:A:361:LYS:HE3	2.02	0.40
1:A:37:PRO:HA	1:A:38:PRO:HD3	2.01	0.40

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	500/510 (98%)	484 (97%)	13 (3%)	3 (1%)	25 31

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	375	THR
1	A	505	ASN
1	A	9	CYS

### 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	433/438 (99%)	388 (90%)	45 (10%)	<b>7</b> <b>8</b>

All (45) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	7	THR
1	A	21	LYS
1	A	22	LEU
1	A	61	GLN
1	A	72	GLU
1	A	74	ARG
1	A	76	LYS
1	A	82	VAL
1	A	94	LEU
1	A	96	GLN
1	A	135	SER
1	A	137	THR
1	A	145	SER
1	A	160	VAL
1	A	193	ILE
1	A	213	LEU
1	A	217	VAL
1	A	231	ARG
1	A	234	ARG
1	A	245	ARG
1	A	281	SER
1	A	289	ASP
1	A	293	SER
1	A	300	SER
1	A	356	GLU
1	A	368	LYS
1	A	374	LYS
1	A	376	GLU
1	A	379	LEU
1	A	397	ARG
1	A	405	LYS
1	A	410	SER
1	A	423	HIS
1	A	430	MET
1	A	440	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	467	LYS
1	A	472	ASP
1	A	474	TYR
1	A	483	ARG
1	A	488	SER
1	A	490	LEU
1	A	492	THR
1	A	504	VAL
1	A	506	GLU
1	A	507	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	96	GLN
1	A	123	HIS
1	A	143	GLN
1	A	148	ASN
1	A	156	ASN
1	A	251	GLN
1	A	326	ASN
1	A	414	GLN
1	A	478	GLN

### 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 carbohydrates in this entry.

### 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 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, 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	502/510 (98%)	0.08	17 (3%) 45 52	8, 23, 30, 39	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	433	HIS	4.6
1	A	38	PRO	4.2
1	A	507	ASP	4.2
1	A	472	ASP	4.1
1	A	473	PHE	3.8
1	A	474	TYR	3.4
1	A	506	GLU	3.4
1	A	487	PHE	2.7
1	A	55	GLY	2.7
1	A	35	PRO	2.6
1	A	373	ASP	2.5
1	A	374	LYS	2.5
1	A	467	LYS	2.4
1	A	377	GLY	2.2
1	A	375	THR	2.2
1	A	54	LYS	2.1
1	A	149	LEU	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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.