



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 11, 2018 – 12:31 am GMT

PDB ID : 4GOP  
Title : Structure and Conformational Change of a Replication Protein A Heterotrimer Bound to ssDNA  
Authors : Pavletich, N.P.; Jie, F.  
Deposited on : 2012-08-20  
Resolution : 3.10 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  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

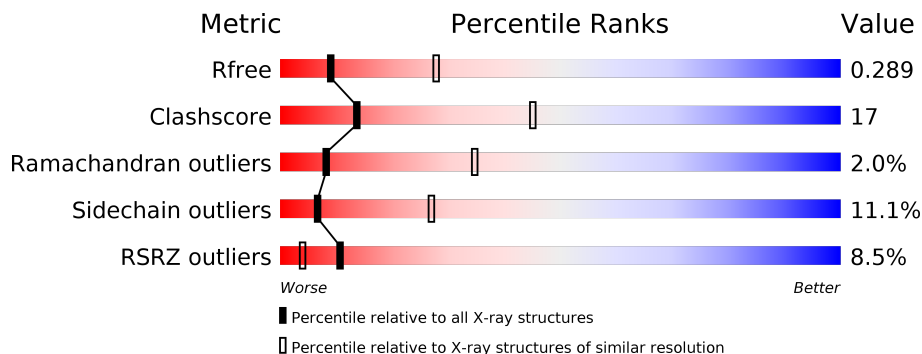
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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.



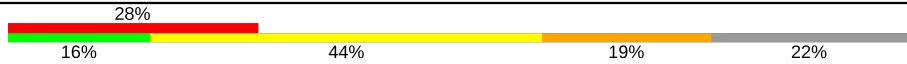

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	111664	1115 (3.12-3.08)
Clashscore	122126	1042 (3.10-3.10)
Ramachandran outliers	120053	1010 (3.10-3.10)
Sidechain outliers	120020	1010 (3.10-3.10)
RSRZ outliers	108989	1089 (3.12-3.08)

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. 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	114	 4% (poor fit), 51% (0 outliers), 35% (1 outlier), 7% (2 outliers), 5% (3+ outliers)
1	X	114	 4% (poor fit), 53% (0 outliers), 32% (1 outlier), 9% (2 outliers), 5% (3+ outliers)
2	B	136	 7% (poor fit), 55% (0 outliers), 31% (1 outlier), 10% (2+ outliers)
2	Y	136	 3% (poor fit), 49% (0 outliers), 35% (1 outlier), 5% (2 outliers), 10% (3+ outliers)
3	C	444	 8% (poor fit), 64% (0 outliers), 30% (1 outlier), 2% (2+ outliers)
3	Z	444	 10% (poor fit), 62% (0 outliers), 31% (1 outlier), 5% (2+ outliers)

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	K	32	
4	L	32	

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 11350 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 Putative uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	108	809	505	132	168	4	0	0	0
1	X	108	809	505	132	168	4	0	0	0

- Molecule 2 is a protein called Putative uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	122	978	609	186	181	2	0	0	0
2	Y	122	978	609	186	181	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	173	VAL	ALA	CONFLICT	UNP Q4PBD4
Y	173	VAL	ALA	CONFLICT	UNP Q4PBD4

- Molecule 3 is a protein called Putative uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	433	3437	2144	600	675	18	0	0	0
3	Z	433	3437	2144	600	675	18	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	314	GLN	THR	CONFLICT	UNP Q4P407
Z	314	GLN	THR	CONFLICT	UNP Q4P407

- Molecule 4 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	K	25	Total 500	250	50	175	25	0	0	0
4	L	20	Total 400	200	40	140	20	0	0	0

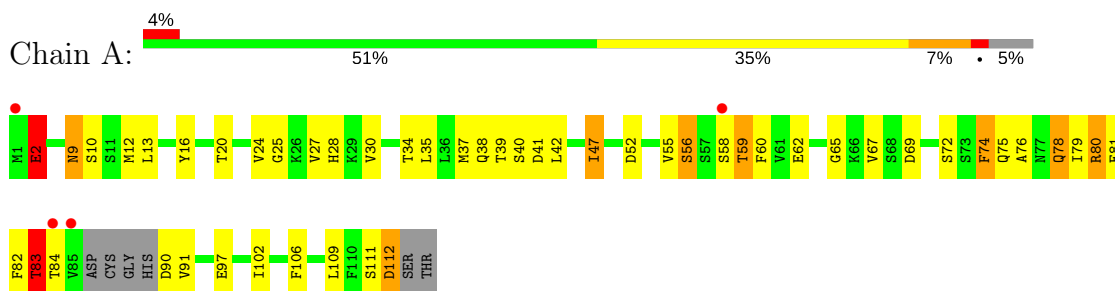
- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	Z	1	Total 1	Zn 1	0	0
5	C	1	Total 1	Zn 1	0	0

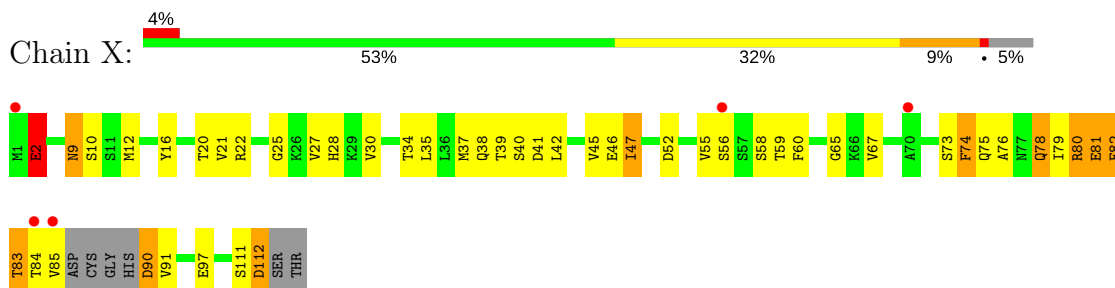
### 3 Residue-property plots [i](#)

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.

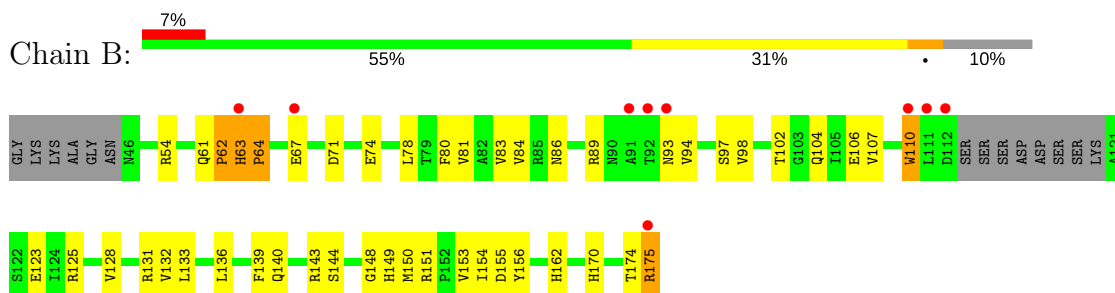
- Molecule 1: Putative uncharacterized protein



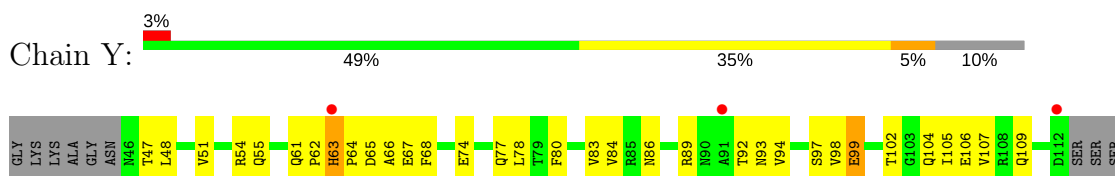
- Molecule 1: Putative uncharacterized protein



- Molecule 2: Putative uncharacterized protein



- Molecule 2: Putative uncharacterized protein









## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	184.10Å 93.80Å 120.50Å 90.00° 110.90° 90.00°	Depositor
Resolution (Å)	20.00 – 3.10 34.23 – 3.07	Depositor EDS
% Data completeness (in resolution range)	85.5 (20.00-3.10) 83.8 (34.23-3.07)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.82 (at 3.06Å)	Xtrriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.242 , 0.286 0.244 , 0.289	Depositor DCC
$R_{free}$ test set	1644 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	56.5	Xtrriage
Anisotropy	0.667	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 54.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	11350	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	88.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.56% 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:  
ZN

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	0.58	0/818	0.78	1/1111 (0.1%)
1	X	0.58	0/818	0.75	0/1111
2	B	0.54	1/994 (0.1%)	0.70	0/1348
2	Y	0.58	0/994	0.72	0/1348
3	C	0.51	4/3502 (0.1%)	0.67	1/4732 (0.0%)
3	Z	0.53	3/3502 (0.1%)	0.67	1/4732 (0.0%)
4	K	0.31	0/549	1.06	8/846 (0.9%)
4	L	0.31	0/438	0.96	4/672 (0.6%)
All	All	0.52	8/11615 (0.1%)	0.73	15/15900 (0.1%)

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	Z	195	TRP	CD2-CE2	5.72	1.48	1.41
2	B	110	TRP	CD2-CE2	5.30	1.47	1.41
3	Z	537	TRP	CD2-CE2	5.27	1.47	1.41
3	Z	507	TRP	CD2-CE2	5.19	1.47	1.41
3	C	359	TRP	CD2-CE2	5.17	1.47	1.41

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	K	11	DT	P-O3'-C3'	9.31	130.88	119.70
4	K	24	DT	P-O3'-C3'	7.71	128.95	119.70
4	L	4	DT	P-O3'-C3'	7.29	128.45	119.70
4	K	10	DT	P-O3'-C3'	7.10	128.22	119.70
4	K	12	DT	P-O3'-C3'	6.83	127.90	119.70

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	809	0	807	51	0
1	X	809	0	807	47	0
2	B	978	0	970	31	0
2	Y	978	0	970	35	0
3	C	3437	0	3315	106	1
3	Z	3437	0	3315	109	1
4	K	500	0	301	23	0
4	L	400	0	242	10	0
5	C	1	0	0	0	0
5	Z	1	0	0	0	0
All	All	11350	0	10727	377	1

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 17.

The worst 5 of 377 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)
3:Z:495:CYS:SG	3:Z:511:LYS:HD3	1.80	1.22
1:A:76:ALA:HB1	1:A:79:ILE:HD11	1.37	1.06
1:X:76:ALA:HB1	1:X:79:ILE:HD11	1.35	1.06
1:A:62:GLU:HB2	1:A:82:PHE:HZ	1.31	0.93
3:C:492:SER:OG	3:C:512:CYS:SG	2.26	0.93

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:596:VAL:N	3:Z:466:GLU:OE2[4_444]	2.05	0.15

## 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	104/114 (91%)	89 (86%)	12 (12%)	3 (3%)	5	27
1	X	104/114 (91%)	88 (85%)	15 (14%)	1 (1%)	17	53
2	B	118/136 (87%)	103 (87%)	12 (10%)	3 (2%)	6	30
2	Y	118/136 (87%)	104 (88%)	11 (9%)	3 (2%)	6	30
3	C	429/444 (97%)	384 (90%)	39 (9%)	6 (1%)	12	44
3	Z	429/444 (97%)	379 (88%)	40 (9%)	10 (2%)	7	32
All	All	1302/1388 (94%)	1147 (88%)	129 (10%)	26 (2%)	8	35

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2	GLU
1	A	83	THR
2	B	63	HIS
2	B	64	PRO
3	C	449	ALA

### 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	96/101 (95%)	79 (82%)	17 (18%)	2	9
1	X	96/101 (95%)	77 (80%)	19 (20%)	1	6
2	B	107/118 (91%)	96 (90%)	11 (10%)	8	30

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	Y	107/118 (91%)	90 (84%)	17 (16%)	3 12
3	C	370/374 (99%)	343 (93%)	27 (7%)	15 47
3	Z	370/374 (99%)	334 (90%)	36 (10%)	9 33
All	All	1146/1186 (97%)	1019 (89%)	127 (11%)	7 27

5 of 127 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	X	12	MET
1	X	83	THR
3	Z	485	LEU
1	X	34	THR
1	X	73	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	547	GLN
1	X	15	GLN
3	Z	500	ASN
3	C	557	HIS
3	C	582	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, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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	108/114 (94%)	0.10	4 (3%) 41 20	31, 61, 99, 121	0
1	X	108/114 (94%)	0.11	5 (4%) 32 15	29, 55, 93, 114	0
2	B	122/136 (89%)	0.18	9 (7%) 14 5	34, 62, 126, 165	0
2	Y	122/136 (89%)	0.15	4 (3%) 46 24	33, 57, 109, 135	0
3	C	433/444 (97%)	0.40	36 (8%) 11 4	36, 86, 152, 175	0
3	Z	433/444 (97%)	0.41	43 (9%) 7 2	32, 84, 146, 186	0
4	K	25/32 (78%)	1.76	9 (36%) 0 0	119, 166, 208, 221	0
4	L	20/32 (62%)	1.65	7 (35%) 0 0	107, 175, 239, 245	0
All	All	1371/1452 (94%)	0.36	117 (8%) 11 4	29, 76, 154, 245	0

The worst 5 of 117 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	Z	518	THR	9.4
3	C	517	ALA	8.7
2	Y	112	ASP	7.6
3	C	441	GLY	6.3
3	C	515	SER	6.2

### 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. 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(Å <sup>2</sup> )	Q<0.9
5	ZN	Z	701	1/1	0.80	0.15	130,130,130,130	0
5	ZN	C	701	1/1	0.97	0.18	136,136,136,136	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.