wwPDB X-ray Structure Validation Summary Report (i)

Aug 16, 2023 – 08:22 PM EDT

PDB ID	:	2AU0
Title	:	Unmodified preinsertion binary complex
Authors	:	Rechkoblit, O.; Malinina, L.; Cheng, Y.; Kuryavyi, V.; Broyde, S.; Geacintov, N.E.; Patel, D.J.
Deposited on	:	2005-08-26
Resolution	:	2.70 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*
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 (i)) were used in the production of this report:

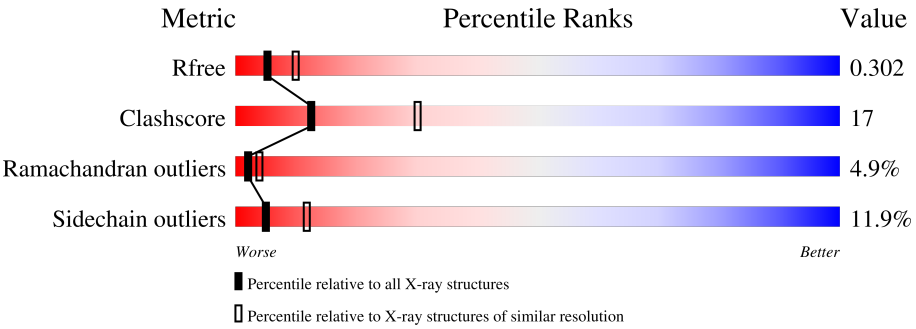
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)

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 >=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 <=5%

Mol	Chain	Length	Quality of chain
1	D	13	23% 54% 23%
1	H	13	31% 54% 15%
2	E	19	16% 42% 11% 32%
2	J	19	11% 37% 21% 32%
3	A	360	49% 39% 6% 5%
3	B	360	56% 33% 6% 5%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6651 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 DNA chain called 5'-D(*GP*GP*TP*TP*GP*GP*AP*TP*GP*GP*TP*AP*(DDG))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	13	Total	C	N	O	P	0	0	0
			272	130	53	77	12			
1	H	13	Total	C	N	O	P	0	0	0
			272	130	53	77	12			

- Molecule 2 is a DNA chain called 5'-D(*CP*TP*AP*AP*CP*G*CP*TP*AP*CP*CP*AP*TP*CP*CP*AP*AP*CP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	13	Total	C	N	O	P	0	0	0
			254	123	45	74	12			
2	J	13	Total	C	N	O	P	0	0	0
			254	123	45	74	12			

- Molecule 3 is a protein called Dpo4 polymerase IV.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	341	Total	C	N	O	S	0	0	0
			2739	1757	472	504	6			
3	B	341	Total	C	N	O	S	0	0	0
			2739	1757	472	504	6			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	GLY	-	cloning artifact	UNP Q97W02
A	-6	SER	-	cloning artifact	UNP Q97W02
A	-5	HIS	-	cloning artifact	UNP Q97W02
A	-4	MET	-	cloning artifact	UNP Q97W02
A	-3	GLY	-	cloning artifact	UNP Q97W02
A	-2	GLY	-	cloning artifact	UNP Q97W02

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	cloning artifact	UNP Q97W02
A	0	GLY	-	cloning artifact	UNP Q97W02
A	1	GLY	-	cloning artifact	UNP Q97W02
B	-7	GLY	-	cloning artifact	UNP Q97W02
B	-6	SER	-	cloning artifact	UNP Q97W02
B	-5	HIS	-	cloning artifact	UNP Q97W02
B	-4	MET	-	cloning artifact	UNP Q97W02
B	-3	GLY	-	cloning artifact	UNP Q97W02
B	-2	GLY	-	cloning artifact	UNP Q97W02
B	-1	GLY	-	cloning artifact	UNP Q97W02
B	0	GLY	-	cloning artifact	UNP Q97W02
B	1001	GLY	-	cloning artifact	UNP Q97W02

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	H	1	Total Ca 1 1	0	0
4	J	1	Total Ca 1 1	0	0
4	A	1	Total Ca 1 1	0	0

- Molecule 5 is water.

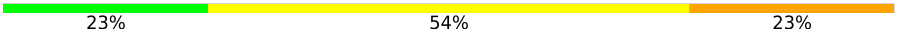
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	12	Total O 12 12	0	0
5	E	5	Total O 5 5	0	0
5	H	16	Total O 16 16	0	0
5	J	6	Total O 6 6	0	0
5	A	41	Total O 41 41	0	0
5	B	38	Total O 38 38	0	0

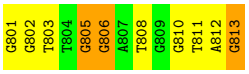
3 Residue-property plots (i)

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. 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. 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: 5'-D(*GP*GP*TP*TP*GP*GP*AP*TP*GP*GP*TP*AP*(DDG))-3'

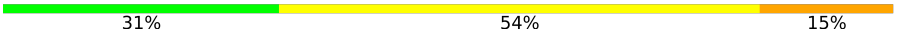
Chain D:

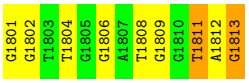




• Molecule 1: 5'-D(*GP*GP*TP*TP*GP*GP*AP*TP*GP*GP*TP*AP*(DDG))-3'

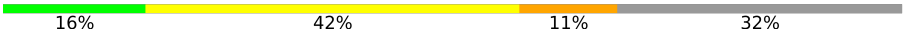
Chain H:





• Molecule 2: 5'-D(*CP*TP*AP*AP*CP*G*CP*TP*AP*CP*CP*AP*TP*CP*CP*AP*AP*CP*C)-3'

Chain E:

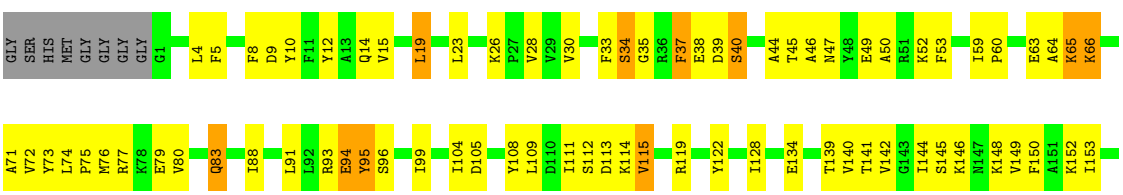


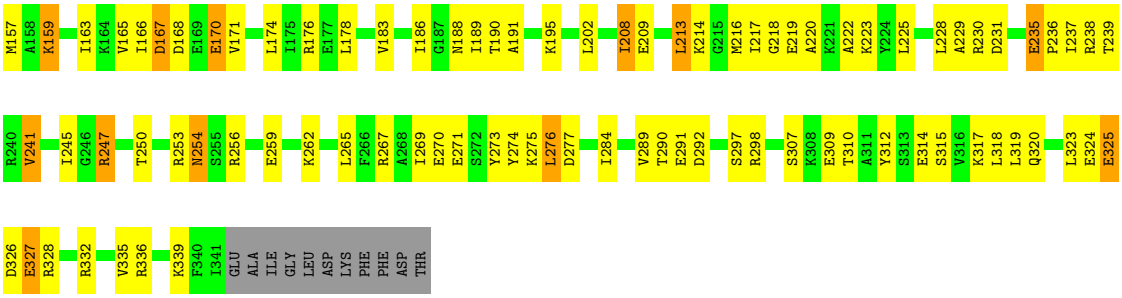
• Molecule 2: 5'-D(*CP*TP*AP*AP*CP*G*CP*TP*AP*CP*CP*AP*TP*CP*CP*AP*AP*CP*C)-3'

Chain J:

• Molecule 3: Dpo4 polymerase IV

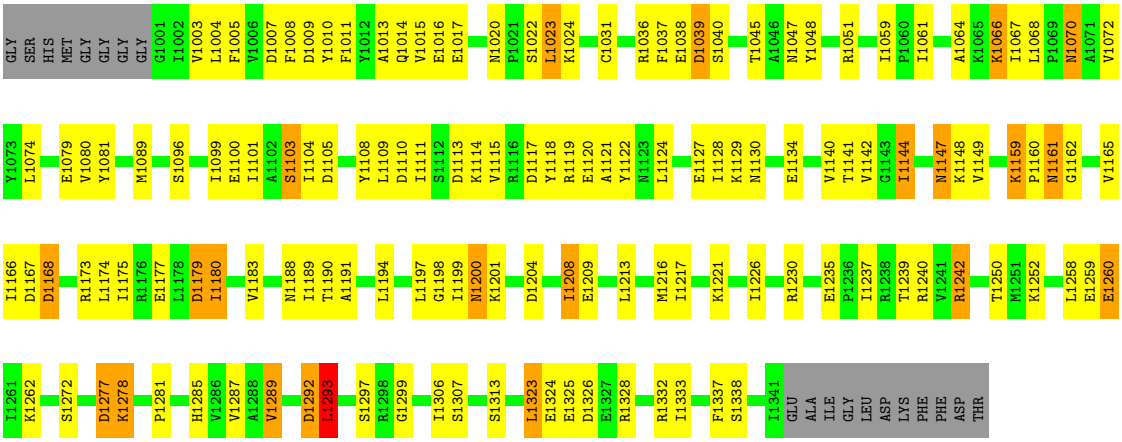
Chain A:





• Molecule 3: Dpo4 polymerase IV

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.64Å 180.96Å 51.57Å 90.00° 107.00° 90.00°	Depositor
Resolution (Å)	15.00 – 2.70 19.88 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.7 (15.00-2.70) 99.1 (19.88-2.60)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.96 (at 2.59Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.269 , 0.310 0.270 , 0.302	Depositor DCC
R_{free} test set	1305 reflections (4.76%)	wwPDB-VP
Wilson B-factor (Å^2)	61.2	Xtriage
Anisotropy	0.084	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 20.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.400 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6651	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.19% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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

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	D	0.79	0/282	1.83	7/436 (1.6%)
1	H	0.87	0/282	1.78	10/436 (2.3%)
2	E	0.93	0/283	1.96	11/432 (2.5%)
2	J	0.93	1/283 (0.4%)	1.98	9/432 (2.1%)
3	A	0.54	0/2778	0.70	0/3731
3	B	0.55	0/2778	0.70	0/3731
All	All	0.62	1/6686 (0.0%)	1.03	37/9198 (0.4%)

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
3	A	0	1
3	B	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	J	1910	DC	C3'-O3'	-6.38	1.35	1.44

The worst 5 of 37 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	918	DC	O4'-C1'-N1	15.70	118.99	108.00
2	J	1916	DA	O4'-C1'-N9	13.64	117.55	108.00
2	J	1912	DA	O4'-C1'-N9	11.47	116.03	108.00
2	J	1915	DC	O4'-C1'-N1	11.28	115.90	108.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	806	DG	O4'-C1'-N9	10.16	115.11	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	235	GLU	Peptide
3	B	1099	ILE	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	D	272	0	149	9	0
1	H	272	0	149	10	0
2	E	254	0	147	6	0
2	J	254	0	147	17	0
3	A	2739	0	2883	93	0
3	B	2739	0	2880	91	0
4	A	1	0	0	0	0
4	H	1	0	0	0	0
4	J	1	0	0	0	0
5	A	41	0	0	4	0
5	B	38	0	0	7	0
5	D	12	0	0	2	0
5	E	5	0	0	0	0
5	H	16	0	0	2	0
5	J	6	0	0	1	0
All	All	6651	0	6355	211	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 17.

The worst 5 of 211 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:H:1813:DDG:H3'1	5:B:2:HOH:O	1.56	1.05
1:D:802:DG:H1	2:E:918:DC:H42	1.11	0.93
1:H:1802:DG:N2	2:J:1918:DC:N3	2.17	0.93
3:A:141:THR:OG1	3:A:159:LYS:O	1.99	0.80
3:B:1017:GLU:OE1	3:B:1024:LYS:NZ	2.13	0.79

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
3	A	339/360 (94%)	258 (76%)	63 (19%)	18 (5%)	2 3
3	B	339/360 (94%)	279 (82%)	45 (13%)	15 (4%)	2 5
All	All	678/720 (94%)	537 (79%)	108 (16%)	33 (5%)	2 4

5 of 33 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	40	SER
3	A	276	LEU
3	A	277	ASP
3	B	1023	LEU
3	B	1036	ARG

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	
3	A	299/311 (96%)	261 (87%)	38 (13%)	4	10
3	B	299/311 (96%)	266 (89%)	33 (11%)	6	14
All	All	598/622 (96%)	527 (88%)	71 (12%)	5	12

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	1197	LEU
3	B	1208	ILE
3	B	1292	ASP
3	A	241	VAL
3	A	239	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	1070	ASN
3	B	1130	ASN
3	B	1200	ASN
3	B	1188	ASN
3	B	1014	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are 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	$\#\|Z\| > 2$
1	DDG	H	1813	1,4,2	17,23,24	1.23	3 (17%)	15,33,36	1.26	2 (13%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
1	DDG	D	813	1,4,2	17,23,24	1.31	3 (17%)	15,33,36	1.09	1 (6%)

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
1	DDG	H	1813	1,4,2	-	0/3/18/19	0/3/3/3
1	DDG	D	813	1,4,2	-	0/3/18/19	0/3/3/3

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	813	DDG	C8-N7	-3.35	1.29	1.35
1	H	1813	DDG	C8-N7	-2.92	1.30	1.35
1	H	1813	DDG	C5-C6	-2.60	1.42	1.47
1	D	813	DDG	C5-C6	-2.50	1.42	1.47
1	D	813	DDG	C6-N1	2.08	1.41	1.37

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	1813	DDG	O4'-C4'-C5'	-2.45	105.48	109.52
1	D	813	DDG	C3'-C2'-C1'	2.20	105.33	102.78
1	H	1813	DDG	C4'-O4'-C1'	2.12	111.81	109.81

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	H	1813	DDG	3	0
1	D	813	DDG	1	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 3 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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.