



Full wwPDB X-ray Structure Validation Report ⓘ

Jun 13, 2022 – 04:11 pm BST

PDB ID : 7OUE
Title : Crystal structure of a trapped Pab-AGOG/single-standed DNA covalent intermediate
Authors : Coste, F.; Goffinont, S.; Flament, D.; Castaing, B.
Deposited on : 2021-06-11
Resolution : 2.04 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

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
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.28.1
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.28.1

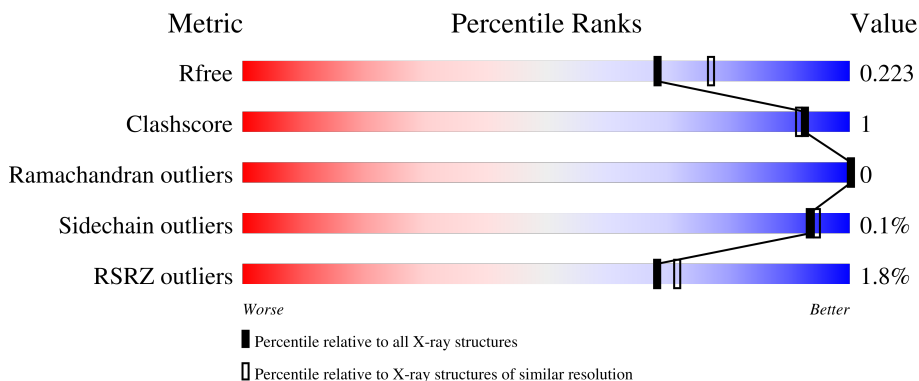
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.04 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.



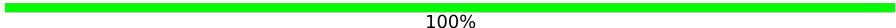


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 $\leq 5\%$. 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	242	97%
1	C	242	96%
1	E	242	5% (red), 94% (green), 6% (yellow)
1	G	242	2% (red), 94% (green), 6% (yellow)
2	B	9	100%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	9	 100%
2	F	9	 78% 22%
2	H	9	 11% 67% 11% 22%

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 8993 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 N-glycosylase/DNA lyase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	242	Total 1932	C 1259	N 320	O 348	S 5	0	0	0
1	C	242	Total 1924	C 1253	N 318	O 348	S 5	0	0	0
1	E	242	Total 1900	C 1236	N 314	O 346	S 4	0	0	0
1	G	241	Total 1876	C 1226	N 311	O 335	S 4	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP Q9UZY0
A	-1	SER	-	expression tag	UNP Q9UZY0
A	0	HIS	-	expression tag	UNP Q9UZY0
C	-2	GLY	-	expression tag	UNP Q9UZY0
C	-1	SER	-	expression tag	UNP Q9UZY0
C	0	HIS	-	expression tag	UNP Q9UZY0
E	-2	GLY	-	expression tag	UNP Q9UZY0
E	-1	SER	-	expression tag	UNP Q9UZY0
E	0	HIS	-	expression tag	UNP Q9UZY0
G	-2	GLY	-	expression tag	UNP Q9UZY0
G	-1	SER	-	expression tag	UNP Q9UZY0
G	0	HIS	-	expression tag	UNP Q9UZY0

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*TP*TP*(PED)P*TP*TP*TP*CP*T)-3').

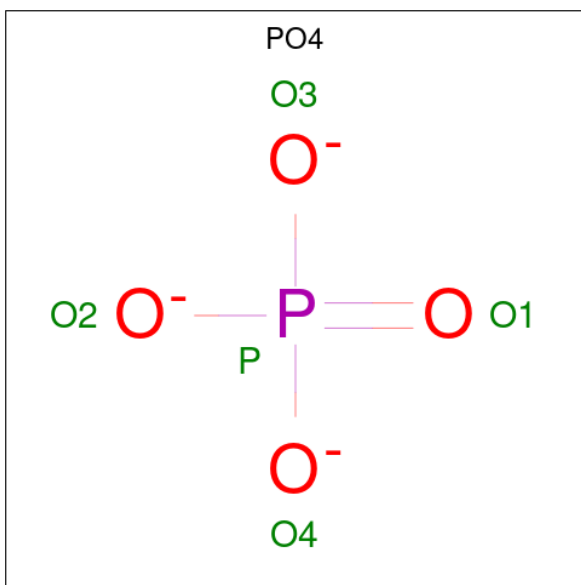
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	9	Total 161	C 78	N 15	O 59	P 9	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
2	D	9	Total	C	N	O	P	0	2	0
			161	78	15	59	9			
2	F	9	Total	C	N	O	P	0	0	0
			167	84	17	58	8			
2	H	7	Total	C	N	O	P	0	2	0
			140	68	13	51	8			

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O₄P).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	
3	C	1	Total	O	P	0	0
			5	4	1		
3	G	1	Total	O	P	0	0
			5	4	1		

- Molecule 4 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total	K	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	217	Total	O	0	0
			217	217		

Continued on next page...

Continued from previous page...

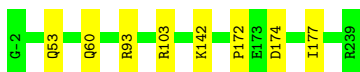
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	21	Total 21	O 21	0	0
5	C	206	Total 206	O 206	0	0
5	D	16	Total 16	O 16	0	0
5	E	112	Total 112	O 112	0	0
5	F	13	Total 13	O 13	0	0
5	G	130	Total 130	O 130	0	0
5	H	6	Total 6	O 6	0	0

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: N-glycosylase/DNA lyase

Chain A:  97%



- Molecule 1: N-glycosylase/DNA lyase

Chain C:  96%



- Molecule 1: N-glycosylase/DNA lyase

Chain E:  5% 94% 6%



- Molecule 1: N-glycosylase/DNA lyase

Chain G:  2% 94% 6%



- Molecule 2: DNA (5'-D(*TP*TP*TP*(PED)P*TP*TP*TP*CP*T)-3')

Chain B:  100%


There are no outlier residues recorded for this chain.

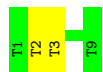
- Molecule 2: DNA (5'-D(*TP*TP*TP*(PED)P*TP*TP*TP*CP*T)-3')

Chain D:  100%

There are no outlier residues recorded for this chain.

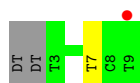
- Molecule 2: DNA (5'-D(*TP*TP*TP*(PED)P*TP*TP*TP*CP*T)-3')

Chain F:  78% 22%



- Molecule 2: DNA (5'-D(*TP*TP*TP*(PED)P*TP*TP*TP*CP*T)-3')

Chain H:  11% 67% 11% 22%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	39.74Å 74.26Å 101.73Å 92.48° 100.74° 105.46°	Depositor
Resolution (Å)	71.23 – 2.04 99.46 – 2.04	Depositor EDS
% Data completeness (in resolution range)	98.1 (71.23-2.04) 93.4 (99.46-2.04)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.23 (at 2.03Å)	Xtrriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.175 , 0.223 0.175 , 0.223	Depositor DCC
R_{free} test set	3345 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å ²)	25.3	Xtrriage
Anisotropy	0.465	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.097 for h,-h-k,-h-l	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	8993	wwPDB-VP
Average B, all atoms (Å ²)	33.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.55% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: PED, K, PO4

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	0.38	0/1979	0.51	0/2679
1	C	0.37	0/1971	0.50	0/2671
1	E	0.35	0/1947	0.50	2/2643 (0.1%)
1	G	0.36	0/1923	0.47	0/2614
2	B	0.89	0/173	1.31	0/261
2	D	0.84	0/173	1.26	0/261
2	F	0.74	0/170	1.22	0/258
2	H	0.65	0/150	1.24	0/224
All	All	0.42	0/8486	0.60	2/11611 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	237	ILE	CG1-CB-CG2	-5.15	100.07	111.40
1	E	114	GLU	CA-CB-CG	5.01	124.42	113.40

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	1932	0	1932	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1924	0	1909	5	0
1	E	1900	0	1855	6	0
1	G	1876	0	1830	7	0
2	B	161	0	90	0	0
2	D	161	0	90	0	0
2	F	167	0	106	1	0
2	H	140	0	78	1	0
3	C	5	0	0	0	0
3	G	5	0	0	0	0
4	E	1	0	0	0	0
5	A	217	0	0	2	0
5	B	21	0	0	0	0
5	C	206	0	0	0	0
5	D	16	0	0	0	0
5	E	112	0	0	0	0
5	F	13	0	0	0	0
5	G	130	0	0	0	0
5	H	6	0	0	0	0
All	All	8993	0	7890	24	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 1.

All (24) 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:174:ASP:H	1:A:177:ILE:HG22	1.62	0.63
1:C:174:ASP:H	1:C:177:ILE:HG22	1.65	0.60
1:E:12:GLU:HG3	1:E:13:GLY:N	2.17	0.59
1:C:158:VAL:HG23	1:C:159:LEU:HG	1.85	0.57
1:E:208:ASP:OD1	1:E:212:TRP:NE1	2.34	0.57
1:E:181:THR:HG21	1:E:189:PRO:HA	1.88	0.55
1:A:53:GLN:HE21	1:A:93:ARG:HH11	1.55	0.55
1:G:76:SER:HB3	1:G:79:LEU:HB2	1.93	0.51
1:C:79:LEU:O	1:C:83:GLU:HG3	2.13	0.49
1:E:214:LEU:HD21	1:E:231:LEU:HB3	1.94	0.49
1:E:8:GLU:HB3	1:E:230:LYS:HD3	1.95	0.47
2:F:2:DT:H5''	2:F:3:DT:H5'	1.96	0.47
1:G:9:ILE:HG23	1:G:227:LEU:HD11	1.96	0.47
1:G:116:ARG:O	1:G:119:GLU:HG3	2.15	0.47
1:A:60:GLN:HG3	5:A:466:HOH:O	2.15	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:103:ARG:NH2	5:A:309:HOH:O	2.48	0.46
1:G:-1:SER:HA	1:G:2:ILE:HD12	1.97	0.46
1:A:142:LYS:HA	1:A:172:PRO:HG3	2.00	0.43
1:C:63:TRP:O	1:C:67:LYS:HE3	2.18	0.43
1:C:142:LYS:HA	1:C:172:PRO:HG3	1.99	0.43
1:E:93:ARG:HD2	2:H:7:DT:H2''	2.01	0.42
1:G:171:ILE:HG13	1:G:208:ASP:HB2	2.03	0.41
1:G:226:GLU:O	1:G:230:LYS:HG3	2.20	0.41
1:G:181:THR:HG21	1:G:189:PRO:HA	2.03	0.41

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	240/242 (99%)	235 (98%)	5 (2%)	0	100	100
1	C	240/242 (99%)	236 (98%)	4 (2%)	0	100	100
1	E	240/242 (99%)	232 (97%)	8 (3%)	0	100	100
1	G	239/242 (99%)	234 (98%)	5 (2%)	0	100	100
All	All	959/968 (99%)	937 (98%)	22 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	202/213 (95%)	202 (100%)	0	100	100
1	C	200/213 (94%)	200 (100%)	0	100	100
1	E	194/213 (91%)	193 (100%)	1 (0%)	88	89
1	G	188/213 (88%)	188 (100%)	0	100	100
All	All	784/852 (92%)	783 (100%)	1 (0%)	93	94

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	24	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	97	GLN
1	G	97	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](#)

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

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 for Mogul analysis.

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	$\# Z > 2$
3	PO4	G	301	-	4,4,4	0.84	0	6,6,6	0.43	0
3	PO4	C	301	-	4,4,4	0.87	0	6,6,6	0.66	0

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

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, 95th 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 ²)	Q<0.9
1	A	242/242 (100%)	-0.58	0 100 100	14, 24, 39, 50	0
1	C	242/242 (100%)	-0.53	0 100 100	15, 26, 41, 49	0
1	E	242/242 (100%)	-0.26	11 (4%) 33 35	18, 32, 74, 103	0
1	G	241/242 (99%)	-0.20	6 (2%) 57 61	19, 32, 76, 97	0
2	B	8/9 (88%)	-0.17	0 100 100	28, 46, 59, 78	0
2	D	8/9 (88%)	-0.33	0 100 100	30, 46, 62, 69	0
2	F	8/9 (88%)	0.15	0 100 100	38, 54, 92, 95	0
2	H	6/9 (66%)	0.34	1 (16%) 1 1	37, 40, 81, 123	0
All	All	997/1004 (99%)	-0.38	18 (1%) 68 71	14, 29, 67, 123	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	212	TRP	4.3
1	G	214	LEU	3.6
1	E	212	TRP	3.5
1	G	225	PRO	3.4
1	E	235	ILE	3.4
1	E	224	PRO	3.1
1	G	175	SER	3.1
1	E	214	LEU	3.0
1	G	172	PRO	2.9
1	E	219	SER	2.9
2	H	9	DT	2.8
1	E	225	PRO	2.6
1	E	238	ILE	2.3
1	E	239	ARG	2.3
1	E	172	PRO	2.2
1	E	213	PRO	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	E	215	LEU	2.1
1	G	215	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 monosaccharides 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, 95th 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 ²)	Q<0.9
3	PO4	G	301	5/5	0.93	0.17	60,67,74,80	0
3	PO4	C	301	5/5	0.97	0.09	35,40,47,49	0
4	K	E	301	1/1	0.99	0.10	44,44,44,44	0

6.5 Other polymers [i](#)

There are no such residues in this entry.