wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 08:26 PM GMT

PDB ID	:	1K82
Title	:	Crystal structure of E.coli formamidopyrimidine-DNA glycosylase (Fpg) covalently trapped with DNA
Authors	:	Gilboa, R.; Zharkov, D.O.; Golan, G.; Fernandes, A.S.; Gerchman, S.E.; Matz, E.; Kycia, J.H.; Grollman, A.P.; Shoham, G.
Deposited on	:	2001-10-22
Resolution	:	2.10 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
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

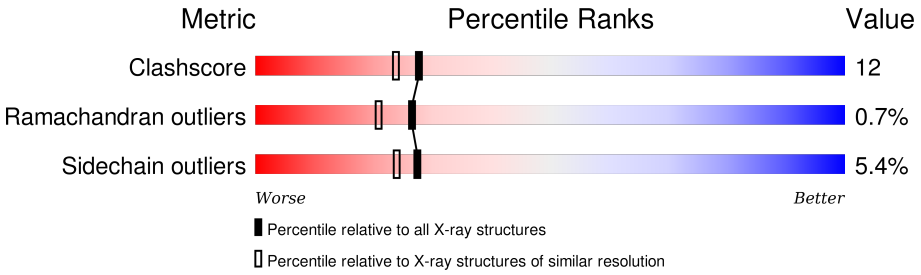
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.10 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))
Clashscore	102246	4460 (2.10-2.10)
Ramachandran outliers	100387	4413 (2.10-2.10)
Sidechain outliers	100360	4414 (2.10-2.10)

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 >=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 <=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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	E	13	62% 38%
1	F	13	69% 31%
1	G	13	69% 31%
1	H	13	69% 31%
2	I	13	8% 62% 31%
2	J	13	8% 54% 38%
2	K	13	8% 69% 23%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	L	13	8% 54% 38%
3	A	268	72% 22%
3	B	268	70% 25%
3	C	268	72% 21%
3	D	268	76% 19%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10716 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*CP*TP*TP*CP*CP*TP*CP*CP*TP*GP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	13	Total	C	N	O	P	0	0	0
			260	125	43	80	12			
1	F	13	Total	C	N	O	P	0	0	0
			260	125	43	80	12			
1	G	13	Total	C	N	O	P	0	0	0
			260	125	43	80	12			
1	H	13	Total	C	N	O	P	0	0	0
			260	125	43	80	12			

- Molecule 2 is a DNA chain called 5'-D(*CP*CP*AP*GP*GP*AP*(PED)P*GP*AP*AP*GP*CP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	13	Total	C	N	O	P	0	0	0
			256	121	52	71	12			
2	J	13	Total	C	N	O	P	0	0	0
			256	121	52	71	12			
2	K	13	Total	C	N	O	P	0	0	0
			256	121	52	71	12			
2	L	13	Total	C	N	O	P	0	0	0
			256	121	52	71	12			

- Molecule 3 is a protein called formamidopyrimidine-DNA glycosylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	260	Total	C	N	O	S	0	0	0
			2038	1297	370	362	9			
3	B	260	Total	C	N	O	S	0	0	0
			2030	1293	367	361	9			
3	C	260	Total	C	N	O	S	0	0	0
			2042	1299	368	366	9			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	260	Total	C	N	O	S	0	0	0
			2039	1298	370	362	9			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Zn	0	0
			1	1		
4	A	1	Total	Zn	0	0
			1	1		
4	D	1	Total	Zn	0	0
			1	1		
4	C	1	Total	Zn	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	93	Total	O	0	0
			93	93		
5	B	82	Total	O	0	0
			82	82		
5	C	77	Total	O	0	0
			77	77		
5	D	92	Total	O	0	0
			92	92		
5	E	15	Total	O	0	0
			15	15		
5	F	16	Total	O	0	0
			16	16		
5	G	18	Total	O	0	0
			18	18		
5	H	18	Total	O	0	0
			18	18		
5	I	23	Total	O	0	0
			23	23		
5	J	24	Total	O	0	0
			24	24		
5	K	20	Total	O	0	0
			20	20		
5	L	21	Total	O	0	0
			21	21		

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 errors 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.

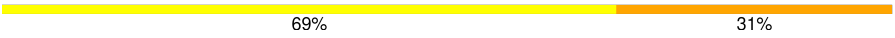
Note EDS was not executed.

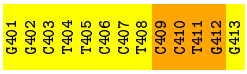
- Molecule 1: 5'-D(*GP*GP*CP*TP*TP*CP*CP*TP*CP*CP*TP*GP*G)-3'

Chain E:

- Molecule 1: 5'-D(*GP*GP*CP*TP*TP*CP*CP*TP*CP*CP*TP*GP*G)-3'

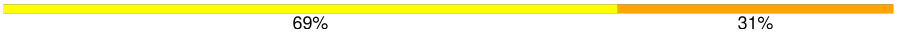
Chain F:





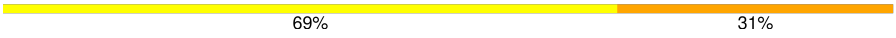
- Molecule 1: 5'-D(*GP*GP*CP*TP*TP*CP*CP*TP*CP*CP*TP*GP*G)-3'

Chain G:



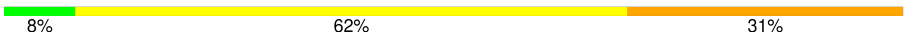
- Molecule 1: 5'-D(*GP*GP*CP*TP*TP*CP*CP*TP*CP*CP*TP*GP*G)-3'

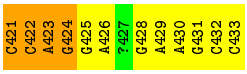
Chain H:



- Molecule 2: 5'-D(*CP*CP*AP*GP*GP*AP*(PED)P*GP*AP*AP*GP*CP*C)-3'

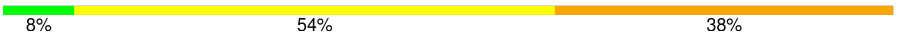
Chain I:

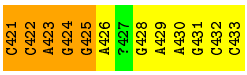




- Molecule 2: 5'-D(*CP*CP*AP*GP*GP*AP*(PED)P*GP*AP*AP*GP*CP*C)-3'

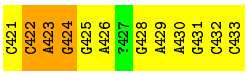
Chain J:





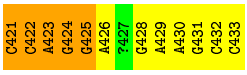
• Molecule 2: 5'-D(*CP*CP*AP*GP*GP*AP*(PED)P*GP*AP*AP*GP*CP*C)-3'

Chain K:



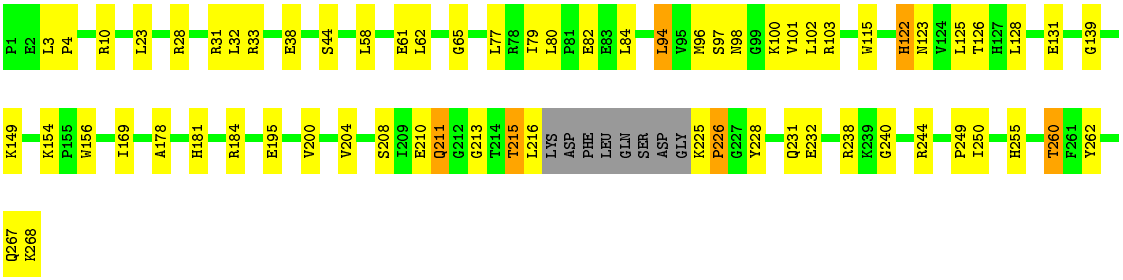
• Molecule 2: 5'-D(*CP*CP*AP*GP*GP*AP*(PED)P*GP*AP*AP*GP*CP*C)-3'

Chain L:



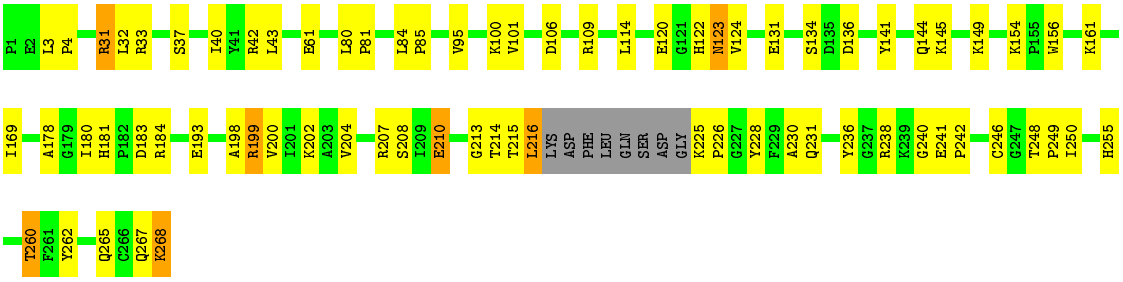
• Molecule 3: formamidopyrimidine-DNA glycosylase

Chain A:



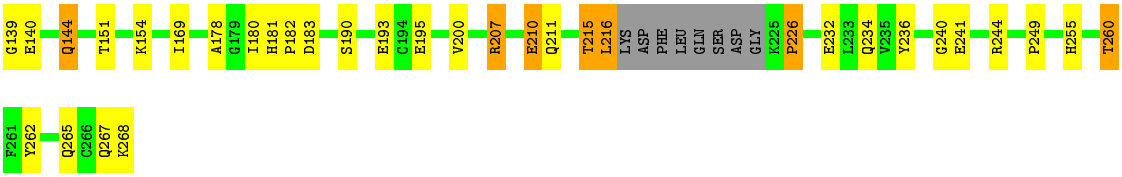
• Molecule 3: formamidopyrimidine-DNA glycosylase

Chain B:



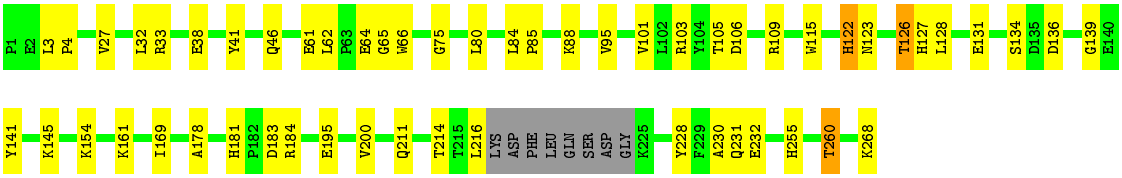
• Molecule 3: formamidopyrimidine-DNA glycosylase

Chain C:



- Molecule 3: formamidopyrimidine-DNA glycosylase

Chain D:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	80.70Å 96.03Å 96.23Å 90.00° 96.80° 90.00°	Depositor
Resolution (Å)	34.00 – 2.10	Depositor
% Data completeness (in resolution range)	(Not available) (34.00-2.10)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.214 , 0.265	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	10716	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PED, 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	E	3.03	18/289 (6.2%)	5.96	96/444 (21.6%)
1	F	2.97	19/289 (6.6%)	5.76	97/444 (21.8%)
1	G	2.99	17/289 (5.9%)	5.90	95/444 (21.4%)
1	H	2.98	18/289 (6.2%)	6.06	99/444 (22.3%)
2	I	1.76	10/275 (3.6%)	5.55	86/420 (20.5%)
2	J	1.74	9/275 (3.3%)	5.82	91/420 (21.7%)
2	K	1.73	9/275 (3.3%)	5.59	81/420 (19.3%)
2	L	1.74	10/275 (3.6%)	5.70	83/420 (19.8%)
3	A	0.33	0/2085	0.79	1/2826 (0.0%)
3	B	0.31	0/2076	0.80	3/2815 (0.1%)
3	C	0.33	0/2088	0.78	0/2830
3	D	0.32	0/2086	0.77	0/2828
All	All	1.17	110/10591 (1.0%)	2.89	732/14755 (5.0%)

The worst 5 of 110 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	404	DT	C5-C7	18.17	1.60	1.50
1	F	411	DT	C5-C7	17.36	1.60	1.50
1	G	411	DT	C5-C7	17.27	1.60	1.50
1	G	404	DT	C5-C7	17.03	1.60	1.50
1	E	405	DT	C5-C7	16.97	1.60	1.50

The worst 5 of 732 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	413	DG	N7-C8-N9	22.34	124.27	113.10
1	H	413	DG	N7-C8-N9	21.92	124.06	113.10
1	F	412	DG	N7-C8-N9	21.85	124.02	113.10
2	L	424	DG	N7-C8-N9	21.82	124.01	113.10
2	J	431	DG	N7-C8-N9	21.74	123.97	113.10

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	E	260	0	149	9	0
1	F	260	0	149	5	0
1	G	260	0	149	2	0
1	H	260	0	149	9	0
2	I	256	0	142	6	0
2	J	256	0	142	5	0
2	K	256	0	142	3	0
2	L	256	0	142	6	0
3	A	2038	0	2058	42	0
3	B	2030	0	2048	53	0
3	C	2042	0	2062	53	0
3	D	2039	0	2059	40	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	93	0	0	1	0
5	B	82	0	0	2	0
5	C	77	0	0	1	0
5	D	92	0	0	1	0
5	E	15	0	0	0	0
5	F	16	0	0	0	0
5	G	18	0	0	1	0
5	H	18	0	0	0	0
5	I	23	0	0	0	0
5	J	24	0	0	0	0
5	K	20	0	0	1	0
5	L	21	0	0	1	0
All	All	10716	0	9391	229	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 12.

The worst 5 of 229 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:C:122:HIS:ND1	3:C:123:ASN:O	1.76	1.16
3:D:211:GLN:OE1	3:D:232:GLU:OE2	1.76	1.03
3:A:80:LEU:HD13	3:A:84:LEU:HD23	1.45	0.99
3:C:122:HIS:CE1	3:C:123:ASN:O	2.17	0.98
3:D:122:HIS:O	3:D:126:THR:OG1	1.82	0.96

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	256/268 (96%)	243 (95%)	10 (4%)	3 (1%)	16	10
3	B	256/268 (96%)	244 (95%)	11 (4%)	1 (0%)	39	37
3	C	256/268 (96%)	249 (97%)	5 (2%)	2 (1%)	24	17
3	D	256/268 (96%)	240 (94%)	15 (6%)	1 (0%)	39	37
All	All	1024/1072 (96%)	976 (95%)	41 (4%)	7 (1%)	26	21

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	226	PRO
3	B	123	ASN
3	C	124	VAL
3	A	82	GLU
3	A	123	ASN

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	218/230 (95%)	206 (94%)	12 (6%)	27	23
3	B	216/230 (94%)	204 (94%)	12 (6%)	26	22
3	C	219/230 (95%)	204 (93%)	15 (7%)	20	16
3	D	218/230 (95%)	210 (96%)	8 (4%)	41	41
All	All	871/920 (95%)	824 (95%)	47 (5%)	27	24

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	216	LEU
3	C	38	GLU
3	D	122	HIS
3	B	268	LYS
3	C	82	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	144	GLN
3	D	211	GLN
3	C	211	GLN
3	B	122	HIS
3	C	181	HIS

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 4 ligands modelled in this entry, 4 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

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates

EDS was not executed - this section will therefore be empty.

6.4 Ligands

EDS was not executed - this section will therefore be empty.

6.5 Other polymers

EDS was not executed - this section will therefore be empty.