



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 12:46 am GMT

PDB ID : 3KHL
Title : Dpo4 post-extension ternary complex with misinserted A opposite the 2-aminofluorene-guanine [AF]G lesion
Authors : Rechkoblit, O.; Malinina, L.; Patel, D.J.
Deposited on : 2009-10-30
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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : trunk28620
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : recalc28949

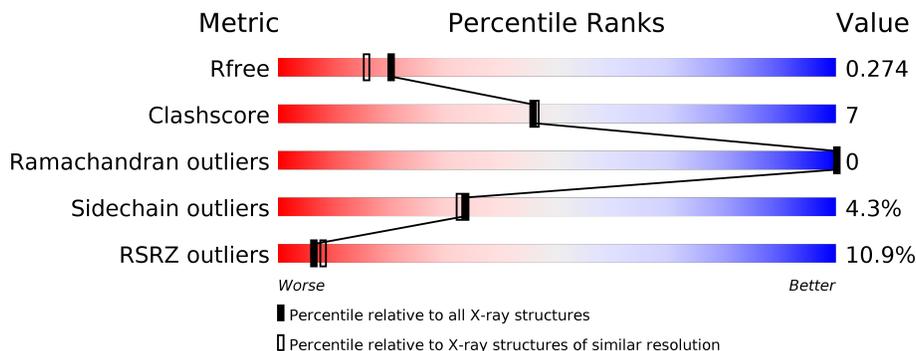
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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}	100719	4243 (2.10-2.10)
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (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 $\leq 5\%$. 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	341	
1	B	341	
2	D	13	
2	H	13	
3	E	20	
3	J	20	

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 7145 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 DNA polymerase IV.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	341	Total	C	N	O	S	0	1	0
			2749	1763	474	506	6			
1	B	341	Total	C	N	O	S	0	2	0
			2761	1769	479	507	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP Q97W02
B	1001	GLY	-	EXPRESSION TAG	UNP Q97W02

- Molecule 2 is a DNA chain called 5'-D(*TP*TP*GP*GP*AP*TP*GP*GP*TP*AP*GP*A P*(DDG))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	13	Total	C	N	O	P	0	0	0
			271	130	53	76	12			
2	H	13	Total	C	N	O	P	0	0	0
			261	125	48	76	12			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	815	DDG	G	ENGINEERED	PDB 3KHL
H	1815	DDG	G	ENGINEERED	PDB 3KHL

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

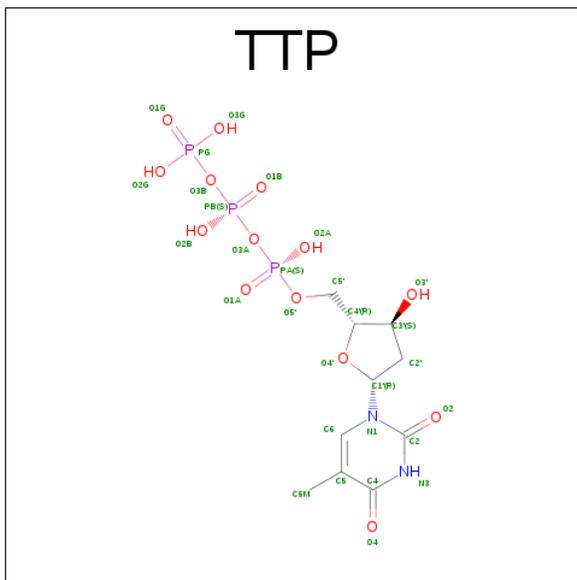
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	E	17	Total	C	N	O	P	0	0	0
			338	163	62	97	16			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	J	18	357	172	65	103	17	0	0	0

- Molecule 4 is THYMIDINE-5'-TRIPHOSPHATE (three-letter code: TTP) (formula: $C_{10}H_{17}N_2O_{14}P_3$).

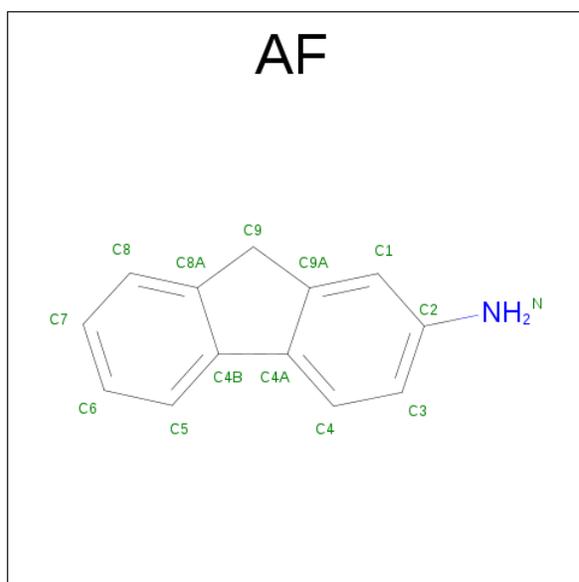


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
4	A	1	29	10	2	14	3	0	0
4	B	1	29	10	2	14	3	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	3	Total Ca 3 3	0	0
5	A	3	Total Ca 3 3	0	0

- Molecule 6 is 2-AMINOFLUORENE (three-letter code: AF) (formula: $C_{13}H_{11}N$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	E	1	Total	C	N	0	0
			14	13	1		
6	J	1	Total	C	N	0	0
			14	13	1		

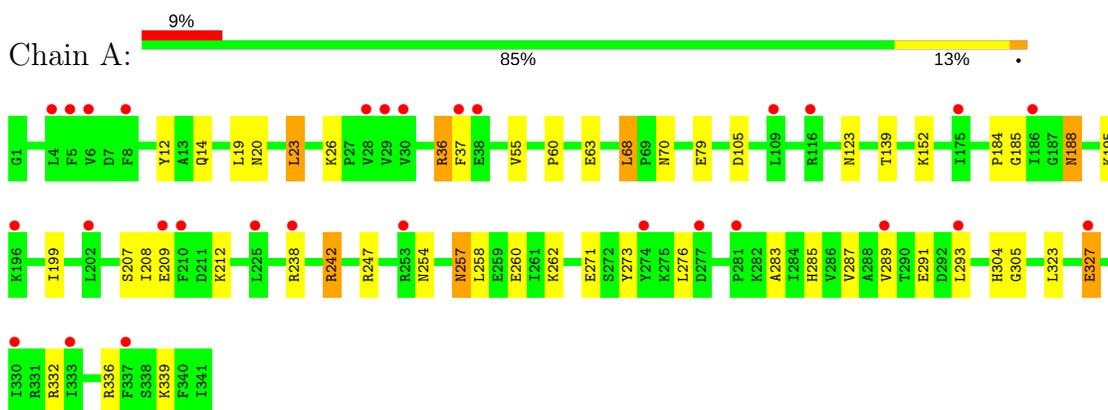
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	135	Total	O	0	0
			135	135		
7	D	15	Total	O	0	0
			15	15		
7	E	6	Total	O	0	0
			6	6		
7	B	126	Total	O	0	0
			126	126		
7	H	14	Total	O	0	0
			14	14		
7	J	20	Total	O	0	0
			20	20		

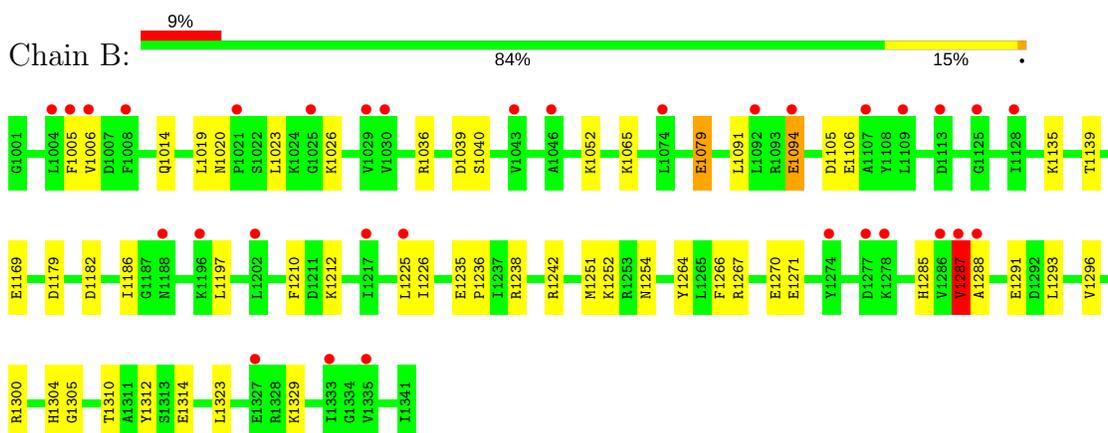
3 Residue-property plots [i](#)

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: DNA polymerase IV



- Molecule 1: DNA polymerase IV



- Molecule 2: 5'-D(*TP*TP*GP*GP*AP*TP*GP*GP*TP*AP*GP*AP*(DDG))-3'

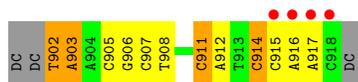
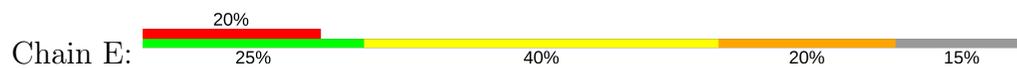


- Molecule 2: 5'-D(*TP*TP*GP*GP*AP*TP*GP*GP*TP*AP*GP*AP*(DDG))-3'





- Molecule 3: 5'-D(*CP*C*TP*AP*AP*CP*GP*CP*TP*AP*CP*CP*AP*TP*CP*CP*AP*AP*CP*C)-3'



- Molecule 3: 5'-D(*CP*C*TP*AP*AP*CP*GP*CP*TP*AP*CP*CP*AP*TP*CP*CP*AP*AP*CP*C)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	52.39Å 52.67Å 100.10Å 82.19° 76.40° 70.07°	Depositor
Resolution (Å)	20.00 – 2.10 19.76 – 2.10	Depositor EDS
% Data completeness (in resolution range)	97.2 (20.00-2.10) 95.3 (19.76-2.10)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.43 (at 2.09Å)	Xtrriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.205 , 0.263 0.213 , 0.274	Depositor DCC
R_{free} test set	2797 reflections (5.33%)	DCC
Wilson B-factor (Å ²)	41.3	Xtrriage
Anisotropy	0.145	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 51.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7145	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 6.38% 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: CA, TTP, DDG, AF

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	0/2788	0.66	0/3742
1	B	0.54	0/2801	0.65	1/3760 (0.0%)
2	D	0.93	0/281	1.53	4/434 (0.9%)
2	H	0.85	0/269	1.44	3/415 (0.7%)
3	E	0.87	0/378	1.78	14/579 (2.4%)
3	J	0.93	0/399	1.56	7/611 (1.1%)
All	All	0.64	0/6916	0.94	29/9541 (0.3%)

There are no bond length outliers.

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	E	902	DT	O3'-P-O5'	-11.03	83.04	104.00
3	E	916	DA	O4'-C1'-N9	7.94	113.56	108.00
3	E	902	DT	OP1-P-O3'	-7.85	87.92	105.20
3	J	1901	DC	O4'-C1'-N1	7.74	113.42	108.00
3	E	902	DT	P-O3'-C3'	7.67	128.91	119.70

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	2749	0	2895	38	0
1	B	2761	0	2898	38	0
2	D	271	0	149	9	0
2	H	261	0	144	4	0
3	E	338	0	192	11	0
3	J	357	0	203	6	0
4	A	29	0	13	2	0
4	B	29	0	13	0	0
5	A	3	0	0	0	0
5	B	3	0	0	0	0
6	E	14	0	10	3	0
6	J	14	0	10	4	0
7	A	135	0	0	5	0
7	B	126	0	0	5	0
7	D	15	0	0	1	0
7	E	6	0	0	1	0
7	H	14	0	0	0	0
7	J	20	0	0	1	0
All	All	7145	0	6527	97	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 7.

The worst 5 of 97 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:14:GLN:HE22	1:A:139:THR:H	1.04	0.95
1:B:1091:LEU:O	1:B:1094:GLU:HG2	1.69	0.93
3:E:902:DT:H2''	3:E:903:DA:OP2	1.66	0.92
1:B:1014:GLN:HE22	1:B:1139:THR:H	1.14	0.88
1:B:1296:VAL:HB	7:B:202:HOH:O	1.75	0.86

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	340/341 (100%)	331 (97%)	9 (3%)	0	100	100
1	B	341/341 (100%)	331 (97%)	10 (3%)	0	100	100
All	All	681/682 (100%)	662 (97%)	19 (3%)	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	300/299 (100%)	284 (95%)	16 (5%)	26	24
1	B	301/299 (101%)	291 (97%)	10 (3%)	43	45
All	All	601/598 (100%)	575 (96%)	26 (4%)	33	32

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	242	ARG
1	A	323	LEU
1	B	1287	VAL
1	A	257	ASN
1	A	291	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	285	HIS
1	A	304	HIS
1	B	1070	ASN
1	A	257	ASN
1	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
2	DDG	D	815	3,2	16,23,24	1.28	1 (6%)	15,33,36	3.12	7 (46%)
2	DDG	H	1815	3,2,5	16,23,24	1.17	1 (6%)	15,33,36	3.16	6 (40%)

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
2	DDG	D	815	3,2	-	0/3/18/19	0/3/3/3
2	DDG	H	1815	3,2,5	-	0/3/18/19	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	1815	DDG	C6-N1	3.68	1.39	1.33
2	D	815	DDG	C6-N1	4.14	1.40	1.33

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	1815	DDG	C5-C6-N1	-8.42	111.50	123.48
2	D	815	DDG	C5-C6-N1	-8.26	111.73	123.48
2	D	815	DDG	N3-C2-N1	-2.87	123.27	127.46
2	D	815	DDG	C6-C5-C4	-2.79	118.07	120.84
2	H	1815	DDG	N3-C2-N1	-2.61	123.65	127.46

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	815	DDG	1	0
2	H	1815	DDG	1	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 6 are monoatomic - leaving 4 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
4	TTP	A	414	5	22,30,30	0.70	0	25,47,47	2.23	5 (20%)
4	TTP	B	1414	5	22,30,30	0.84	1 (4%)	25,47,47	1.87	4 (16%)
6	AF	E	926	3	16,16,16	0.53	0	23,23,23	0.79	0
6	AF	J	1926	3	16,16,16	0.62	0	23,23,23	0.88	1 (4%)

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
4	TTP	A	414	5	-	0/18/34/34	0/2/2/2
4	TTP	B	1414	5	-	0/18/34/34	0/2/2/2
6	AF	E	926	3	-	0/0/8/8	0/3/3/3
6	AF	J	1926	3	-	0/0/8/8	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	1414	TTP	PG-O3B	2.16	1.63	1.60

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	414	TTP	C5-C4-N3	-5.82	118.82	125.24
4	B	1414	TTP	C5-C4-N3	-4.71	120.05	125.24
4	A	414	TTP	C2'-C1'-N1	-3.13	106.83	114.23
4	B	1414	TTP	C2'-C1'-N1	-2.71	107.83	114.23
4	A	414	TTP	C5-C6-N1	-2.19	119.78	122.15

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	414	TTP	2	0
6	E	926	AF	3	0
6	J	1926	AF	4	0

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, 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	341/341 (100%)	0.57	29 (8%) 11 15	15, 29, 44, 52	0
1	B	341/341 (100%)	0.61	32 (9%) 9 12	18, 30, 43, 51	0
2	D	12/13 (92%)	1.53	5 (41%) 0 0	20, 29, 80, 82	0
2	H	12/13 (92%)	1.28	4 (33%) 0 0	16, 26, 86, 89	0
3	E	17/20 (85%)	1.56	4 (23%) 1 1	22, 33, 91, 95	0
3	J	18/20 (90%)	1.76	7 (38%) 0 0	19, 36, 94, 95	0
All	All	741/748 (99%)	0.67	81 (10%) 6 8	15, 30, 46, 95	0

The worst 5 of 81 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	803	DT	8.8
3	E	918	DC	7.8
2	H	1803	DT	7.4
3	J	1918	DC	7.3
3	J	1901	DC	7.1

6.2 Non-standard residues in protein, DNA, RNA chains [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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å ²)	Q<0.9
2	DDG	D	815	21/22	0.86	0.19	-	49,50,64,65	0
2	DDG	H	1815	21/22	0.90	0.15	-	44,51,61,62	0

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
5	CA	A	417	1/1	0.96	0.18	0.17	51,51,51,51	0
6	AF	E	926	14/14	0.91	0.13	-0.61	40,44,45,45	0
6	AF	J	1926	14/14	0.90	0.14	-0.71	41,44,46,47	0
5	CA	A	416	1/1	0.98	0.15	-0.77	28,28,28,28	0
4	TTP	A	414	29/29	0.97	0.12	-0.81	8,13,18,30	0
5	CA	B	1416	1/1	0.99	0.13	-0.82	27,27,27,27	0
4	TTP	B	1414	29/29	0.96	0.12	-0.93	7,13,15,18	0
5	CA	B	1417	1/1	0.88	0.11	-0.99	54,54,54,54	0
5	CA	B	1415	1/1	0.98	0.07	-2.84	54,54,54,54	0
5	CA	A	415	1/1	0.99	0.15	-	44,44,44,44	0

6.5 Other polymers [i](#)

There are no such residues in this entry.