W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 01:23 pm BST

PDB ID	:	6DKL
Title	:	Crystal Structure of a Rationally Designed Six-Fold Symmetric DNA Scaffold
Authors	:	Simmons, C.R.; Zhang, F.; Yan, H.
Deposited on	:	2018-05-29
Resolution	:	3.03 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

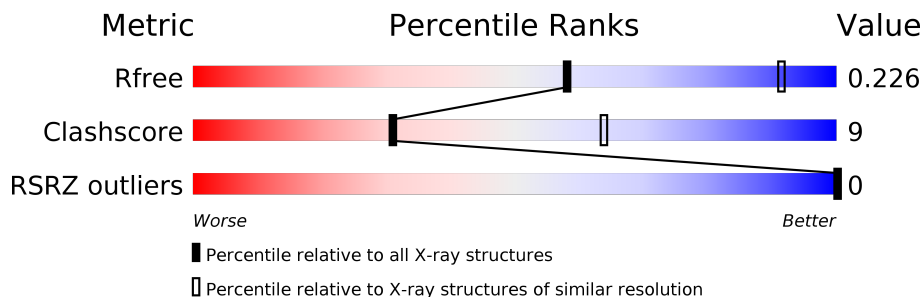|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.03 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	Percentile Ranks	Value
Rfree		0.226
Clashscore		9
RSRZ outliers		0

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2752 (3.08-3.00)
Clashscore	141614	3096 (3.08-3.00)
RSRZ outliers	127900	2636 (3.08-3.00)

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 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% 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	9	78% 22%
2	B	10	50% 50%
3	C	11	64% 27% 9%
4	D	12	50% 50%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1322 atoms, of which 471 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 DNA (5'-D(P*GP*GP*AP*TP*GP*CP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	9	Total	C	H	N	O	P	0	0	0
			281	87	99	38	49	8			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	10	Total	C	H	N	O	P	0	0	0
			305	94	110	37	55	9			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	11	Total	C	H	N	O	P	0	0	0
			353	108	127	36	71	11			

- Molecule 4 is a DNA chain called DNA (5'-D(P*GP*TP*AP*CP*GP*GP*AP*TP*CP*CP*AP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	D	12	Total	C	H	N	O	P	0	0	0
			383	117	135	48	71	12			

3 Residue-property plots ⓘ

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 (5'-D(P*GP*GP*AP*TP*GP*CP*AP*CP*A)-3')

Chain A:

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

Chain B:

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

Chain C:

- Molecule 4: DNA (5'-D(P*GP*TP*AP*CP*GP*GP*AP*TP*CP*CP*AP*G)-3')

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	68.41Å 68.41Å 143.22Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.18 – 3.03 37.17 – 3.03	Depositor EDS
% Data completeness (in resolution range)	89.5 (37.18-3.03) 89.5 (37.17-3.03)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.80 (at 3.06Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.228 , 0.234 0.219 , 0.226	Depositor DCC
R_{free} test set	660 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å^2)	102.5	Xtriage
Anisotropy	1.030	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.15 , 76.8	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.139 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	1322	wwPDB-VP
Average B, all atoms (Å^2)	128.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.25% 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 (i)

5.1 Standard geometry (i)

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	0.71	0/205	0.92	0/315
2	B	0.79	0/218	0.92	0/333
3	C	0.73	0/251	1.19	1/386 (0.3%)
4	D	0.72	0/278	1.01	0/427
All	All	0.74	0/952	1.02	1/1461 (0.1%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	6	DT	O4'-C4'-C3'	-5.94	102.12	104.50

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	182	99	98	1	0
2	B	195	110	109	4	1
3	C	226	127	127	3	1
4	D	248	135	135	7	0
All	All	851	471	469	12	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 9.

The worst 5 of 12 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:8:DT:H2'	3:C:9:DG:C8	2.35	0.61
3:C:6:DT:O4'	4:D:6:DG:H2"	2.09	0.53
2:B:6:DC:H2"	2:B:7:DG:C8	2.44	0.53
2:B:1:DC:H2'	2:B:2:DA:C8	2.48	0.49
1:A:8:DC:H2"	1:A:9:DA:O4'	2.16	0.46

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)
2:B:10:DC:HO3'	3:C:1:DT:OP1[5_554]	1.55	0.05

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

There are no protein molecules in this entry.

5.3.2 Protein sidechains (i)

There are no protein molecules in this entry.

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)

There are no ligands in this entry.

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	9/9 (100%)	-1.34	0	100	100	106, 120, 132, 133	0
2	B	10/10 (100%)	-1.36	0	100	100	112, 123, 131, 135	0
3	C	11/11 (100%)	-1.21	0	100	100	104, 111, 135, 140	0
4	D	12/12 (100%)	-1.23	0	100	100	99, 112, 148, 148	0
All	All	42/42 (100%)	-1.28	0	100	100	99, 117, 138, 148	0

There are no RSRZ outliers to report.

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.