WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 06:10 PM GMT

PDB ID : 4KTQ
Title : BINARY COMPLEX OF THE LARGE FRAGMENT OF DNA POLY-
MERASE I FROM T. AQUATICUS BOUND TO A PRIMER/TEMPLATE
DNA
Authors : Li, Y.; Waksman, G.
Deposited on : 1998-09-09
Resolution : 2.50 Å(reported)

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

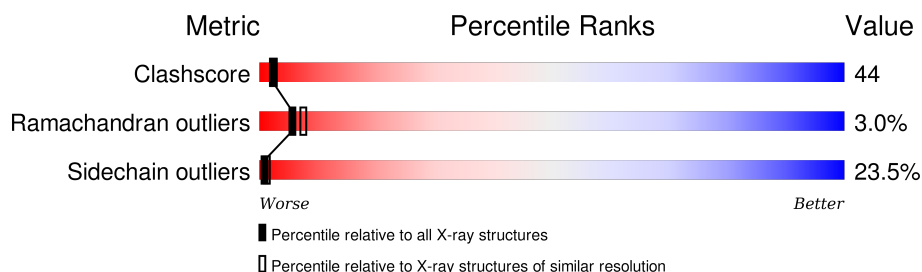|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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 2.50 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
Clashscore		44
Ramachandran outliers		3.0%
Sidechain outliers		23.5%

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))
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)

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	B	12	33% 67%
2	C	13	38% 62%
3	A	539	30% 56% 14% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4769 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 DNA (5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*GP* CP*(DOC))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	12	Total	C	N	O	P	0	0	0
			240	114	48	67	11			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	13	Total	C	N	O	P	0	0	0
			267	126	51	78	12			

- Molecule 3 is a protein called PROTEIN (LARGE FRAGMENT OF DNA POLYMERASE I).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	539	Total	C	N	O	S	0	0	0
			4159	2652	740	754	13			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	84	Total	O	0	0
			84	84		
4	B	8	Total	O	0	0
			8	8		
4	C	11	Total	O	0	0
			11	11		

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 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.
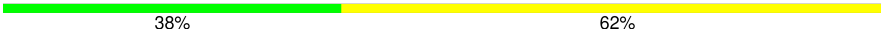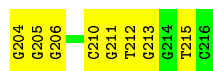
Note EDS was not executed.

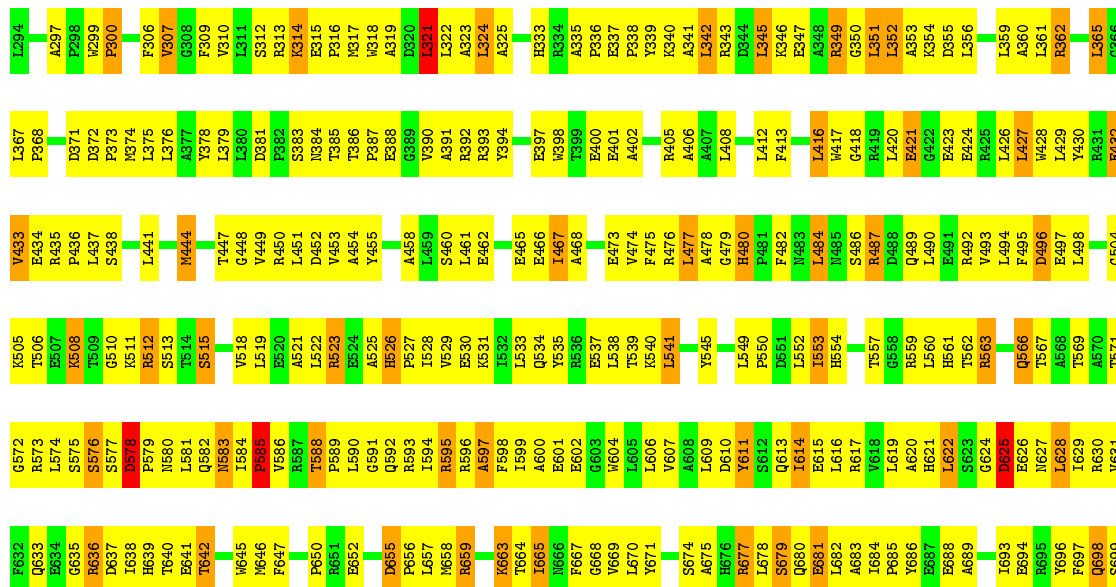- Molecule 1: DNA (5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*(DOC))-3')

Chain B:

33%	67%

G101 A102 C103 C104 A105 C106 G107 G108 C111 C112

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

Chain C:

38%	62%

G204 G205 G206 C210 G211 T212 G213 C214 T215 C216

- Molecule 3: PROTEIN (LARGE FRAGMENT OF DNA POLYMERASE I)

Chain A:

30%	56%	14%

L294 A297 P298 W299 P300 F306 V307 G308 F309 V310 L311 S312 R313 K314 E315 P316 M317 W318 A319 D320 L321 L322 A323 L324 A325 H333 R334 A335 E336 E337 P338 Y339 K340 A341 L342 D344 L345 K346 E347 A348 R349 G350 L351 L352 A353 K354 D355 L356 L359 A360 L361 R362 L365 G366

L367 P368 D371 D372 P373 N374 L375 L376 A377 Y378 L379 L380 D381 P382 S383 N384 T385 T386 P387 E388 G389 V390 A391 R392 R393 Y394 E397 W398 T399 E400 E401 A402 R405 A406 A407 L408 L412 F413 L416 W417 G418 R419 L420 E421 G422 E423 A424 R425 L426 L427 W428 L429 Y430 R431 E432

V433 E434 R435 P436 L437 S438 L441 M444 T447 G448 V449 R450 L451 D452 V453 A454 Y455 A458 L459 S460 L461 E462 E465 E466 I467 A468 E473 V474 F475 R476 L477 A478 G479 H480 P481 F482 N483 L484 N485 S486 R487 D498 Q489 L490 E491 R492 V493 L494 F495 D496 E497 L498 G504

K505 T506 E507 K508 T509 G510 K511 R512 S513 T514 S515 V518 L519 E520 A521 L522 R523 E524 A525 H526 P527 I528 V529 E530 K531 I532 L533 Q534 Y535 R536 E537 L538 T539 K540 L541 Y545 P550 D551 L552 I553 H554 T557 G558 R559 L560 H561 T562 R563 Q566 T567 A568 T569 A570 T571

G572 R573 L574 S575 S576 S577 D578 P579 N580 L581 Q582 N583 I584 P585 V586 R587 T588 P589 L590 G591 Q592 R593 I594 R595 R596 A597 F598 I599 A600 E601 E602 G603 W604 L605 L606 V607 A608 L609 D610 Y611 S612 Q613 I614 E615 L616 R617 V618 L619 A620 H621 L622 S623 G624 D625 E626 N627 L628 I629 R630 V631

F632 Q633 E634 G635 R636 D637 I638 H639 T640 E641 T642 W645 M646 F647 P650 R651 E652 D655 P656 L657 M658 R659 K663 T664 I665 N666 F667 G668 V669 L670 Y671 S674 A675 H676 R677 L678 S679 Q680 N681 L682 A683 I684 P685 Y686 E687 E688 A689 I693 E694 R695 Y696 F697 Q698 S699

F700 V703 R704 I707 E708 K709 T710 L711 E712 E713 G714 R715 R716 R717 G718 Y719 V720 E721 T722 L723 F724 G725 R726 R727 R728 Y729 V730 L733 R736 V737 K738 S739 V740 R741 E742 E745 R746 M747 A748 F749 N750 M751 P752 V753 Q754 G755 T756 A757 A758 D759 L760 M761 K762 L763 A764

M765 V766 K767 L768 F769 P770 R771 L772 E773 E774 M775 R778 M779 L780 L781 Q782 V783 H784 D785 E786 L787 V788 L789 E790 A791 P792 E797 A800 R801 L802 A803 K304 E805 V806 M807 E808 G809 V810 Y811 P812 L813 A814 V815 P816 L817 E818 V819 E820 V821 G822 I823 D826 W827 L828 S829 A830

K831 E832

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 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	110.84Å　　110.84Å　　90.84Å 90.00°　　　90.00°　　120.00°	Depositor
Resolution (Å)	30.00　–　2.50	Depositor
% Data completeness (in resolution range)	91.8 (30.00-2.50)	Depositor
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
Refinement program	X-PLOR 3.851	Depositor
R, R_{free}	0.227　,　0.298	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4769	wwPDB-VP
Average B, all atoms (Å^2)	51.0	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: DOC

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	B	0.47	0/249	0.73	0/382
2	C	0.53	0/299	0.83	0/461
3	A	0.36	0/4249	0.58	0/5776
All	All	0.38	0/4797	0.61	0/6619

There are no bond length outliers.

There are no bond angle outliers.

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	B	240	0	134	12	0
2	C	267	0	147	17	0
3	A	4159	0	4117	385	0
4	A	84	0	0	8	0
4	B	8	0	0	1	0
4	C	11	0	0	1	0
All	All	4769	0	4398	397	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 44.

The worst 5 of 397 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:A:351:LEU:HD23	3:A:374:MET:HG2	1.34	1.06
3:A:780:LEU:HD11	3:A:790:GLU:HG3	1.40	1.02
3:A:614:ILE:HD11	3:A:760:LEU:HD12	1.49	0.95
3:A:779:MET:HA	3:A:789:LEU:HD12	1.50	0.93
3:A:677:ARG:HH12	3:A:746:ARG:HH12	1.17	0.91

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	537/539 (100%)	419 (78%)	102 (19%)	16 (3%)	5	7

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	525	ALA
3	A	625	ASP
3	A	681	GLU
3	A	586	VAL
3	A	597	ALA

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	412/441 (93%)	315 (76%)	97 (24%)	1	1

5 of 97 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	595	ARG
3	A	657	LEU
3	A	789	LEU
3	A	614	ILE
3	A	636	ARG

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

Mol	Chain	Res	Type
3	A	561	HIS
3	A	566	GLN
3	A	583	ASN
3	A	534	GLN
3	A	582	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	DOC	B	112	1,2	11,19,20	0.87	1 (9%)	14,26,29	0.99	1 (7%)

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	DOC	B	112	1,2	-	0/3/18/19	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	112	DOC	C6-C5	-2.09	1.33	1.38

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	112	DOC	C2-N3-C4	3.20	120.12	115.61

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	112	DOC	2	0

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)

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

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

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

6.3 Carbohydrates (i)

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

6.4 Ligands (i)

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

6.5 Other polymers (i)

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