wwPDB X-ray Structure Validation Summary Report (i)

Aug 6, 2023 – 06:19 AM EDT

PDB ID	:	1K4S
Title	:	HUMAN DNA TOPOISOMERASE I IN COVALENT COMPLEX WITH A 22 BASE PAIR DNA DUPLEX
Authors	:	Staker, B.L.; Hjerrild, K.; Feese, M.D.; Behnke, C.A.; Burgin Jr., A.B.; Stewart, L.J.
Deposited on	:	2001-10-08
Resolution	:	3.20 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

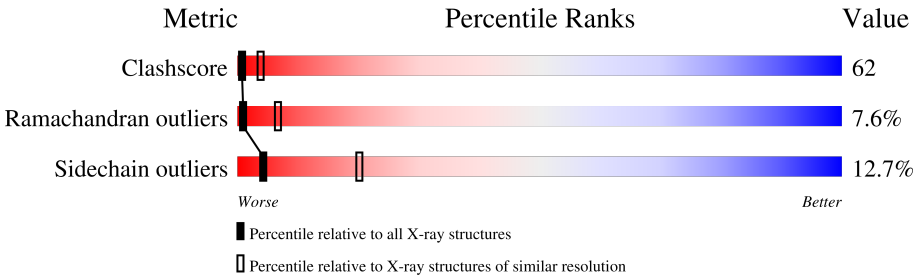
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.20 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	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)

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 <=5%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	B	10	70% 30%
2	C	12	8% 58% 33%
3	D	22	45% 55%
4	A	592	24% 47% 10% • 17%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
1	5IU	B	9	-	-	X	-
2	5IU	C	18	-	-	X	-
2	5IU	C	19	-	-	X	-
2	5IU	C	20	-	-	X	-
3	5IU	D	107	-	-	X	-
3	5IU	D	109	-	-	X	-
3	5IU	D	110	-	-	X	-
3	5IU	D	116	-	-	X	-
3	5IU	D	118	-	-	X	-
3	5IU	D	119	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4858 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(*AP*AP*AP*AP*AP*GP*AP*CP*(5IU)P*(5IU))-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	I	N	O	P			
1	B	10	203	97	2	42	53	9	0	0	0

- Molecule 2 is a DNA chain called 5'-D(*(SPT)P*GP*AP*AP*AP*AP*AP*(5IU)P*(5IU)P*(5IU)P*(5IU)P*T)-3'.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
			Total	C	I	N	O	P	S			
2	C	12	244	116	4	42	70	11	1	0	0	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	I	N	O	P			
3	D	22	445	209	9	73	133	21	0	0	0

- Molecule 4 is a protein called DNA topoisomerase I.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
4	A	490	3966	2540	689	715	1	21	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	723	PTR	TYR	modified residue	UNP P11387

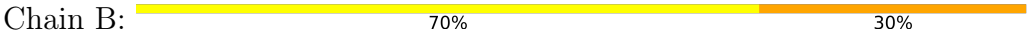
3 Residue-property plots (i)

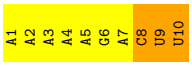
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded 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. 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(*AP*AP*AP*AP*AP*GP*AP*CP*(5IU)P*(5IU))-3'

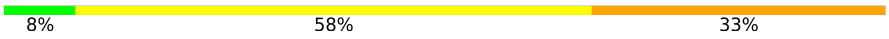
Chain B:





• Molecule 2: 5'-D(*(SPT)P*GP*AP*AP*AP*AP*AP*(5IU)P*(5IU)P*(5IU)P*(5IU)P*T)-3'

Chain C:

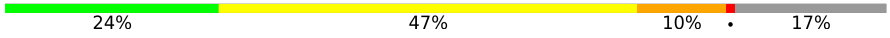


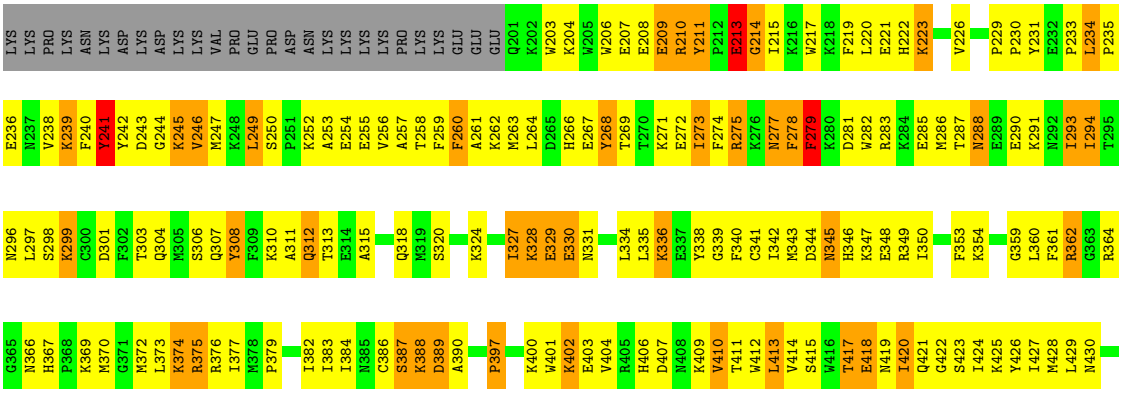
• Molecule 3: 5'-D(*AP*AP*AP*AP*AP*TP*(IDO)UP*(IDO)UP*(IDO)UP*(IDO)UP*CP*AP*AP*AP*GP*(IDO)UP*CP*(IDO)UP*(IDO)UP*(IDO)UP*(IDO)UP*T)-3'

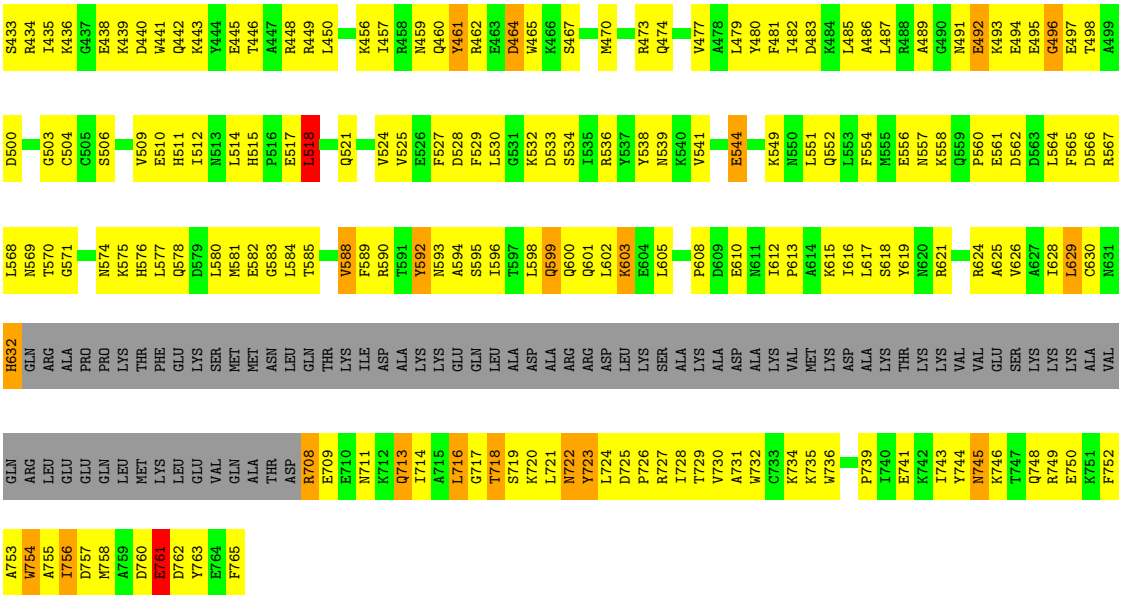
Chain D:

• Molecule 4: DNA topoisomerase I

Chain A:






S433 R434 I435 K436 G437 E438 K439 D440 W441 Q442 K443 Y444 E445 T446 A447 R448 R449 L450 K456 I457 R458 N459 Q460 Y461 R462 E463 D464 W465 K466 S467 M470 R473 Q474 V477 A478 L479 Y480 F481 I482 D483 K484 L485 A486 L487 R488 A489 G490 N491 E492 K493 E494 E495 G496 E497 T498 A499
D500 G503 C504 C505 S506 V509 E510 H511 I512 N513 L514 H515 P516 E517 L518 Q521 V524 V525 E526 F527 D528 F529 L530 G531 K532 D533 S534 I535 R536 Y537 Y538 N539 K540 V541 E544 K549 N550 L551 Q552 L553 F554 M555 E556 N557 K558 Q559 P560 E561 D562 D563 L564 F565 D566 R567
L568 N569 T570 G571 N574 K575 H576 L577 Q578 D579 L580 M581 E582 G583 L584 T585 V588 F589 R590 T591 Y592 N593 A594 S595 I596 T597 L598 Q599 Q600 Q601 L602 K603 E604 L605 P608 D609 E610 M611 I612 P613 A614 K615 I616 L617 S618 Y619 N620 R621 R624 A625 V626 A627 I628 L629 C630 N631
H632 GLN ARG ALA PRO PRO LYS THR PHE GLU LYS SER MET MET ASN LEU GLN THR LYS ILE ASP ALA LYS LYS GLU GLN LEU ALA ASP ALA ARG ARG ASP LEU LYS SER ALA LYS ALA ASP ALA LYS VAL MET LYS ASP ALA LYS THR LYS LYS VAL GLU SER LYS LYS VAL GLU
GLN ARG LEU GLU GLU GLN LEU MET LYS LEU GLU VAL GLN ALA THR ASP R708 E709 E710 N711 K712 Q713 I714 A715 L716 G717 T718 S719 K720 L721 N722 Y723 L724 D725 P726 R727 I728 T729 V730 A731 W732 G733 K734 K735 W736 P739 I740 E741 K742 I743 Y744 N745 K746 T747 Q748 R749 E750 K751 F752
A753 W754 A755 I756 D757 M758 A759 D760 E761 D762 Y763 E764 F765

4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	73.23Å 73.23Å 186.63Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 3.20	Depositor
% Data completeness (in resolution range)	96.8 (50.00-3.20)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.13	Depositor
Refinement program	CNX	Depositor
R, R_{free}	0.217 , 0.222	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4858	wwPDB-VP
Average B, all atoms (Å^2)	41.0	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: SPT, 5IU, PTR

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	0.56	0/186	0.90	1/285 (0.4%)
2	C	0.44	0/165	0.69	0/250
3	D	0.54	0/296	0.81	0/447
4	A	0.49	0/4046	0.69	0/5464
All	All	0.49	0/4693	0.71	1/6446 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
3	D	0	3

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	8	DC	C1'-O4'-C4'	-5.57	104.53	110.10

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	111	DC	Sidechain
3	D	114	DA	Sidechain
3	D	117	DC	Sidechain

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	203	0	106	45	0
2	C	244	0	127	28	0
3	D	445	0	228	95	0
4	A	3966	0	3841	434	0
All	All	4858	0	4302	559	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 62.

The worst 5 of 559 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:B:8:DC:H2"	1:B:9:5IU:I5	1.95	1.35
3:D:115:DG:H2'	3:D:116:5IU:I5	2.10	1.20
3:D:118:5IU:H2"	3:D:119:5IU:H5'	1.21	1.11
3:D:118:5IU:C2'	3:D:119:5IU:H5'	1.88	1.03
3:D:119:5IU:H2'	3:D:120:5IU:I5	2.28	1.03

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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
4	A	485/592 (82%)	367 (76%)	81 (17%)	37 (8%)	1 7

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	344	ASP
4	A	387	SER
4	A	388	LYS
4	A	389	ASP
4	A	495	GLU

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
4	A	408/535 (76%)	356 (87%)	52 (13%)	4 20

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	413	LEU
4	A	464	ASP
4	A	713	GLN
4	A	417	THR
4	A	446	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
4	A	539	ASN
4	A	632	HIS
4	A	748	GLN
4	A	722	ASN
4	A	599	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)

17 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	SPT	C	11	3,2	18,18,22	0.37	0	26,26,33	0.46	0
3	5IU	D	119	1,3	18,21,22	1.03	1 (5%)	26,30,33	0.45	0
3	5IU	D	109	2,3	18,21,22	0.84	1 (5%)	26,30,33	0.40	0
3	5IU	D	116	1,3	18,21,22	1.11	2 (11%)	26,30,33	0.42	0
2	5IU	C	20	3,2	18,21,22	1.21	1 (5%)	26,30,33	0.62	0
2	5IU	C	21	3,2	18,21,22	0.85	1 (5%)	26,30,33	0.38	0
4	PTR	A	723	4,1	15,16,17	0.80	0	19,22,24	0.84	1 (5%)
3	5IU	D	107	2,3	18,21,22	0.80	1 (5%)	26,30,33	0.39	0
3	5IU	D	118	1,3	18,21,22	0.95	1 (5%)	26,30,33	0.42	0
3	5IU	D	121	1,3	18,21,22	0.82	1 (5%)	26,30,33	0.41	0
3	5IU	D	110	2,3	18,21,22	0.82	1 (5%)	26,30,33	0.38	0
2	5IU	C	19	3,2	18,21,22	0.77	1 (5%)	26,30,33	0.31	0
1	5IU	B	9	3,1	18,21,22	1.00	1 (5%)	26,30,33	0.37	0
2	5IU	C	18	3,2	18,21,22	0.77	1 (5%)	26,30,33	0.50	0
3	5IU	D	108	2,3	18,21,22	0.83	1 (5%)	26,30,33	0.41	0
3	5IU	D	120	1,3	18,21,22	0.68	1 (5%)	26,30,33	0.42	0
1	5IU	B	10	3,4,1	18,20,22	1.01	1 (5%)	23,28,33	0.93	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
2	SPT	C	11	3,2	-	2/6/18/22	0/2/2/2
3	5IU	D	119	1,3	-	3/7/21/22	0/2/2/2
3	5IU	D	109	2,3	-	1/7/21/22	0/2/2/2

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	5IU	D	116	1,3	-	0/7/21/22	0/2/2/2
2	5IU	C	20	3,2	-	1/7/21/22	0/2/2/2
2	5IU	C	21	3,2	-	0/7/21/22	0/2/2/2
4	PTR	A	723	4,1	-	2/10/11/13	0/1/1/1
3	5IU	D	107	2,3	-	1/7/21/22	0/2/2/2
3	5IU	D	118	1,3	-	0/7/21/22	0/2/2/2
3	5IU	D	121	1,3	-	0/7/21/22	0/2/2/2
3	5IU	D	110	2,3	-	1/7/21/22	0/2/2/2
2	5IU	C	19	3,2	-	0/7/21/22	0/2/2/2
1	5IU	B	9	3,1	-	0/7/21/22	0/2/2/2
2	5IU	C	18	3,2	-	2/7/21/22	0/2/2/2
3	5IU	D	108	2,3	-	0/7/21/22	0/2/2/2
3	5IU	D	120	1,3	-	0/7/21/22	0/2/2/2
1	5IU	B	10	3,4,1	-	0/7/18/22	0/2/2/2

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	20	5IU	C5-I5	-4.05	1.96	2.08
3	D	116	5IU	C5-I5	-3.39	1.98	2.08
3	D	119	5IU	C5-I5	-3.38	1.98	2.08
1	B	10	5IU	C5-I5	-3.29	1.98	2.08
1	B	9	5IU	C5-I5	-3.19	1.98	2.08

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	10	5IU	C2'-C1'-N1	3.37	118.79	112.40
4	A	723	PTR	O2P-P-O1P	2.02	118.58	110.68

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	11	SPT	O4'-C4'-C5'-S5'
2	C	11	SPT	C3'-C4'-C5'-S5'
3	D	119	5IU	O4'-C4'-C5'-O5'
4	A	723	PTR	C-CA-CB-CG
3	D	119	5IU	C3'-C4'-C5'-O5'

There are no ring outliers.

17 monomers are involved in 94 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	11	SPT	5	0
3	D	119	5IU	13	0
3	D	109	5IU	9	0
3	D	116	5IU	16	0
2	C	20	5IU	10	0
2	C	21	5IU	5	0
4	A	723	PTR	3	0
3	D	107	5IU	13	0
3	D	118	5IU	8	0
3	D	121	5IU	1	0
3	D	110	5IU	8	0
2	C	19	5IU	7	0
1	B	9	5IU	9	0
2	C	18	5IU	8	0
3	D	108	5IU	4	0
3	D	120	5IU	3	0
1	B	10	5IU	4	0

5.5 Carbohydrates (i)

There are no monosaccharides 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

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.