W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 12:28 AM GMT

PDB ID	:	2AJQ
Title	:	Structure of replicative DNA polymerase provides insigts into the mechanisms for processivity, frameshifting and editing
Authors	:	Brieba, L.; Ellenberger, T.
Deposited on	:	2005-08-02
Resolution	:	2.60 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)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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)	:	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 2.60 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		0.289
Clashscore		21
Ramachandran outliers		1.8%
Sidechain outliers		5.3%
RSRZ outliers		3.3%

Worse ... *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
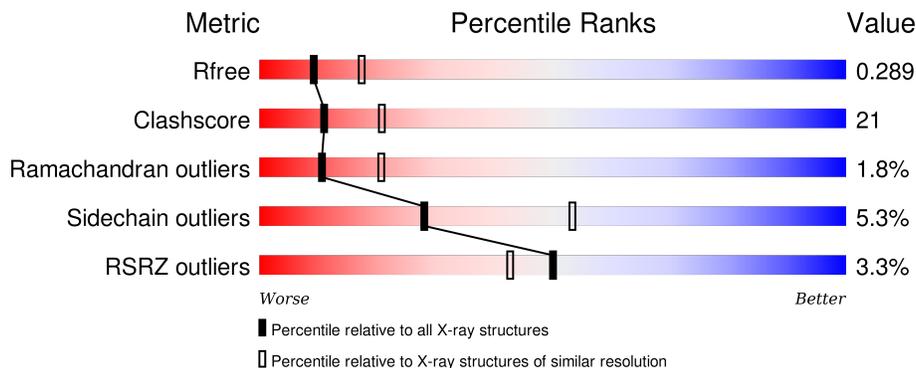
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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.

Mol	Chain	Length	Quality of chain
1	P	22	18% 64% 5% 5% 9%
1	X	22	32% 55% 9% 5%
2	T	26	27% 42% 15% 15%
2	Z	26	4% / 42% 54% •
3	A	704	% / 67% 31% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	704	66% 29%
4	B	108	6% 54% 40%
4	I	108	29% 36% 52% 9%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 14791 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 Primer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	P	20	Total 390	C 184	N 77	O 110	P 19	0	0	1
1	X	21	Total 432	C 204	N 87	O 120	P 21	0	0	0

- Molecule 2 is a DNA chain called DNA Template.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	22	Total 453	C 215	N 79	O 137	P 22	0	0	0
2	Z	25	Total 514	C 244	N 89	O 156	P 25	0	0	0

- Molecule 3 is a protein called T7 DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	704	Total 5541	C 3527	N 971	O 1019	S 24	0	0	0
3	F	689	Total 5459	C 3480	N 953	O 1003	S 23	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	5	ALA	ASP	ENGINEERED	UNP P00581
A	7	ALA	GLU	ENGINEERED	UNP P00581
F	5	ALA	ASP	ENGINEERED	UNP P00581
F	7	ALA	GLU	ENGINEERED	UNP P00581

- Molecule 4 is a protein called thioredoxin 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	105	Total 799	C 517	N 129	O 150	S 3	0	0	0
4	I	105	Total 789	C 512	N 126	O 148	S 3	0	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	P	16	Total 16	O 16	0	0
5	T	21	Total 21	O 21	0	0
5	X	21	Total 21	O 21	0	0
5	Z	27	Total 27	O 27	0	0
5	A	158	Total 158	O 158	0	0
5	B	13	Total 13	O 13	0	0
5	F	142	Total 142	O 142	0	0
5	I	16	Total 16	O 16	0	0

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.

- Molecule 1: DNA Primer

Chain P:
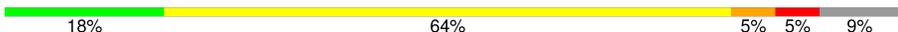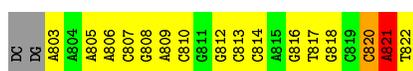
- Molecule 1: DNA Primer

Chain X:
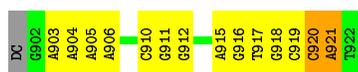
- Molecule 2: DNA Template

Chain T:
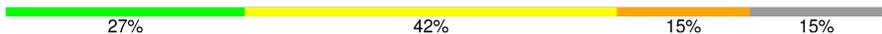
- Molecule 2: DNA Template

Chain Z:
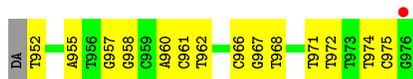
- Molecule 3: T7 DNA polymerase

Chain A:
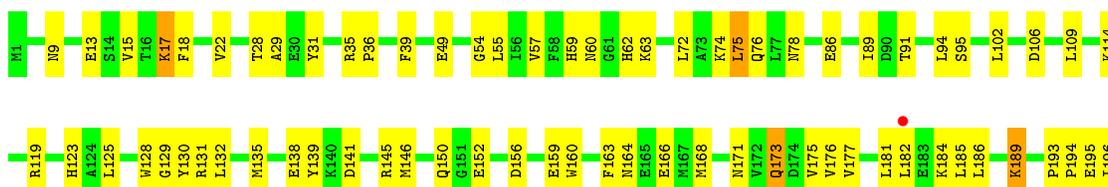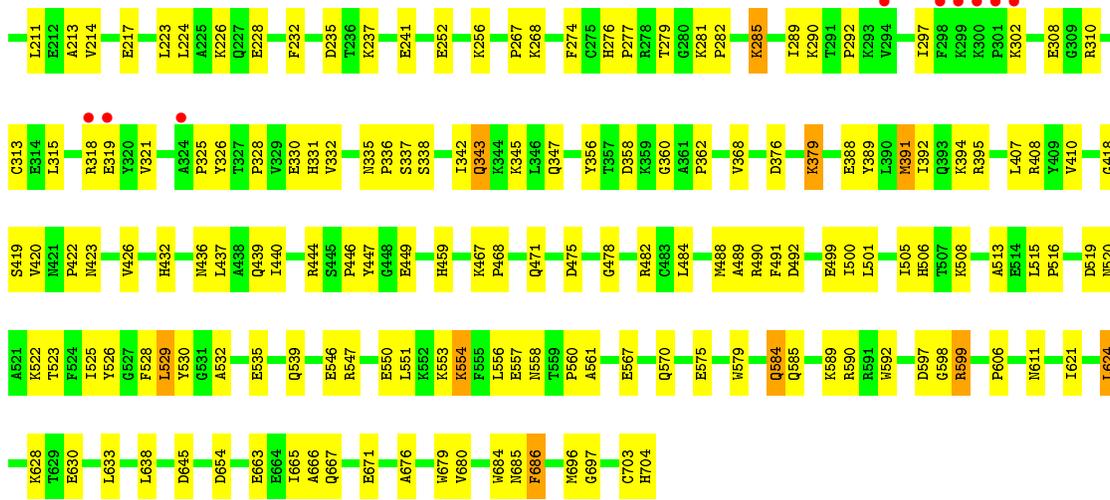
• **Molecule 3: T7 DNA polymerase**

• **Molecule 4: thioredoxin 1**

T89 G92 A93 L94 S95 K96 G97 Q98 L99 F102 L103 L107 ALA

- Molecule 4: thioredoxin 1

Chain I:

29%

36%   52%   9%   •

SER ASP K3 I4 I5 H6 L7 T8 D9 D10 S11 F12 D13 V16 L17 K18 A19 D20 G21 A22 I23 L24 V25 D26 F27 C32 C35 K36 M37 I38 A39 P40 I41 L42 D43 E44 I45 A46 Y49 Q50 G51 K52 L53 T54 V55 A56 K57 L58 N63 A67 P68 P76 T77

L78 L79 L80 F81 K82 N83 G84 E85 V86 A87 A88 T89 K90 V91 G92 A93 L94 S95 Q98 L99 K100 E101 F102 L103 D104 A105 N106 L107 ALA
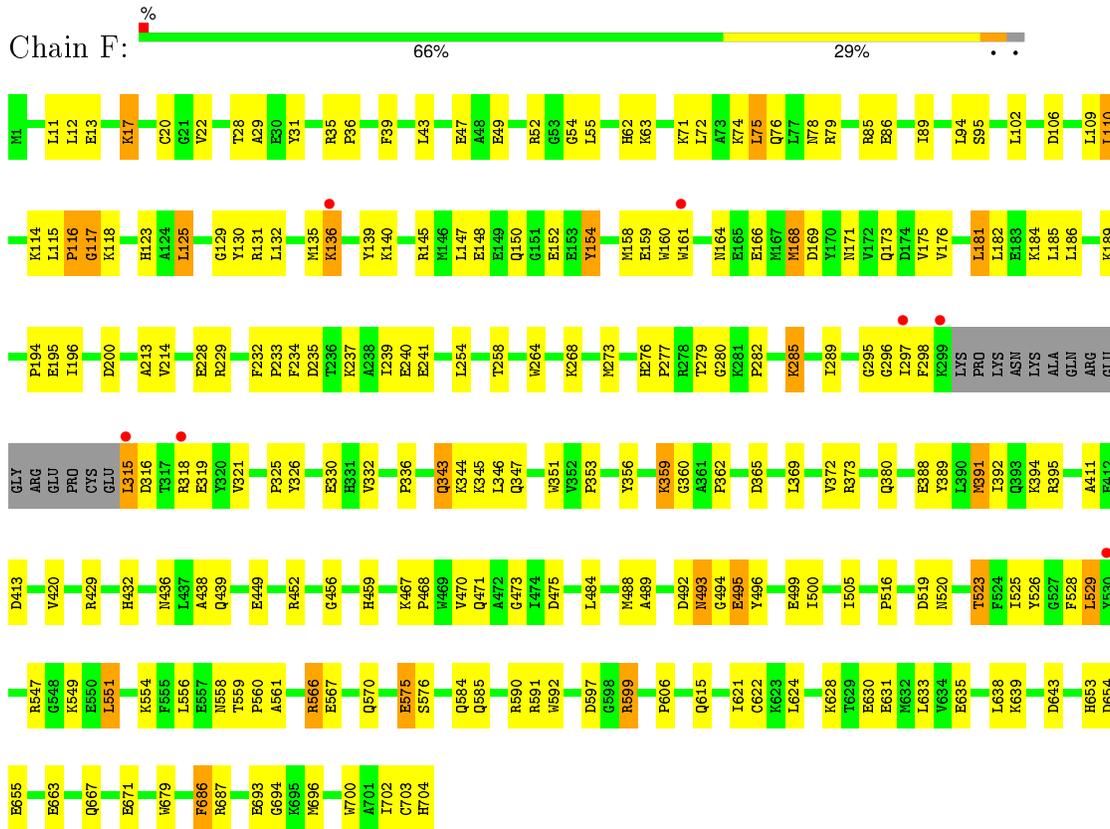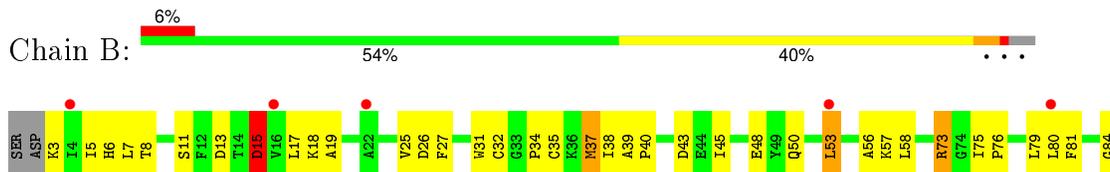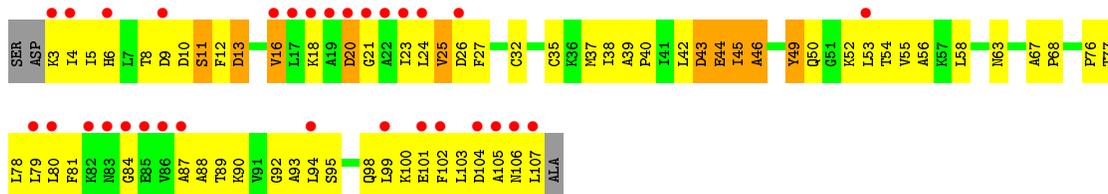
4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	168.31Å 169.24Å 179.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.97 – 2.60 45.97 – 2.38	Depositor EDS
% Data completeness (in resolution range)	99.8 (45.97-2.60) 95.7 (45.97-2.38)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	1.93 (at 2.37Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.233 , 0.284 0.245 , 0.289	Depositor DCC
R_{free} test set	3929 reflections (4.99%)	DCC
Wilson B-factor (Å^2)	60.0	Xtriage
Anisotropy	0.315	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 40.7	EDS
Estimated twinning fraction	0.449 for -k,-h,-l	Xtriage
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Outliers	1 of 98459 reflections (0.001%)	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	14791	wwPDB-VP
Average B, all atoms (Å^2)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.61% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: 2DT

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	P	0.66	1/417 (0.2%)	0.90	1/642 (0.2%)
1	X	0.41	0/465	0.81	0/715
2	T	0.55	0/506	0.83	0/780
2	Z	0.46	0/574	0.77	0/885
3	A	0.39	0/5680	0.59	0/7693
3	F	0.39	0/5594	0.60	0/7570
4	B	0.31	0/814	0.55	0/1104
4	I	0.32	0/803	0.55	0/1089
All	All	0.40	1/14853 (0.0%)	0.63	1/20478 (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
1	P	0	2
1	X	0	2
2	T	0	4
2	Z	0	2
All	All	0	10

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	P	803	DA	O3'-P	-7.13	1.52	1.61

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	P	821	DA	N9-C1'-C2'	-5.54	102.08	112.60

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	P	820	DC	Sidechain
1	P	821	DA	Sidechain
2	T	854	DG	Sidechain
2	T	855	DA	Sidechain
2	T	856	DT	Sidechain

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	P	390	0	212	37	0
1	X	432	0	234	29	0
2	T	453	0	250	21	0
2	Z	514	0	284	13	0
3	A	5541	0	5382	186	0
3	F	5459	0	5338	179	0
4	B	799	0	814	45	0
4	I	789	0	794	92	0
5	A	158	0	0	14	0
5	B	13	0	0	1	0
5	F	142	0	0	7	1
5	I	16	0	0	4	0
5	P	16	0	0	0	0
5	T	21	0	0	0	0
5	X	21	0	0	0	0
5	Z	27	0	0	0	0
All	All	14791	0	13308	577	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 21.

The worst 5 of 577 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)
4:B:73:ARG:HH11	4:B:73:ARG:HB3	1.11	1.09
2:T:867:DG:H2"	2:T:868:DT:H5'	1.32	1.07
4:I:16:VAL:HA	4:I:23:ILE:HG21	1.36	1.04
1:P:807:DC:H2"	1:P:808:DG:H5'	1.38	1.03
4:B:73:ARG:NH1	4:B:73:ARG:HB3	1.79	0.96

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)
5:F:5072:HOH:O	5:F:5072:HOH:O[3_555]	1.46	0.74

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	702/704 (100%)	657 (94%)	39 (6%)	6 (1%)	21	42
3	F	685/704 (97%)	635 (93%)	43 (6%)	7 (1%)	19	39
4	B	103/108 (95%)	89 (86%)	10 (10%)	4 (4%)	4	5
4	I	101/108 (94%)	71 (70%)	19 (19%)	11 (11%)	0	0
All	All	1591/1624 (98%)	1452 (91%)	111 (7%)	28 (2%)	11	21

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	302	LYS
3	A	310	ARG
4	B	18	LYS
4	B	19	ALA
3	F	114	LYS

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	563/582 (97%)	533 (95%)	30 (5%)	28	53
3	F	559/582 (96%)	528 (94%)	31 (6%)	27	51
4	B	84/87 (97%)	79 (94%)	5 (6%)	24	47
4	I	81/87 (93%)	79 (98%)	2 (2%)	55	81
All	All	1287/1338 (96%)	1219 (95%)	68 (5%)	28	53

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	15	ASP
3	F	85	ARG
3	F	599	ARG
4	B	37	MET
4	B	80	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	667	GLN
3	F	123	HIS
3	F	667	GLN
4	B	83	ASN
3	F	173	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	#$|Z| > 2$
1	2DT	P	822	1,2	11,20,21	1.30	3 (27%)	12,28,31	4.94	2 (16%)
1	2DT	X	922	1,2	11,20,21	1.33	3 (27%)	12,28,31	5.00	3 (25%)

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	2DT	P	822	1,2	-	0/3/18/19	0/2/2/2
1	2DT	X	922	1,2	-	0/3/18/19	0/2/2/2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	P	822	2DT	C6-C5	-2.40	1.33	1.40
1	X	922	2DT	C6-C5	-2.32	1.33	1.40
1	P	822	2DT	C6-N1	2.17	1.38	1.35
1	X	922	2DT	C6-N1	2.18	1.38	1.35
1	P	822	2DT	C4-N3	2.62	1.38	1.33

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	X	922	2DT	C5-C4-N3	-9.04	115.08	125.14
1	P	822	2DT	C5-C4-N3	-8.94	115.19	125.14
1	X	922	2DT	C5M-C5-C6	2.04	122.73	118.62
1	P	822	2DT	C4-N3-C2	14.25	127.57	115.25
1	X	922	2DT	C4-N3-C2	14.38	127.68	115.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	P	822	2DT	1	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)

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	P	19/22 (86%)	-0.55	0	100	100	50, 100, 170, 171	0
1	X	20/22 (90%)	-0.35	0	100	100	49, 100, 201, 201	0
2	T	22/26 (84%)	-0.62	0	100	100	49, 90, 145, 153	0
2	Z	25/26 (96%)	-0.26	1 (4%)	42	34	49, 97, 183, 192	0
3	A	704/704 (100%)	-0.06	10 (1%)	78	74	38, 61, 116, 153	0
3	F	689/704 (97%)	-0.08	7 (1%)	84	81	38, 60, 105, 143	0
4	B	105/108 (97%)	0.27	7 (6%)	21	15	50, 93, 113, 116	0
4	I	105/108 (97%)	1.86	31 (29%)	1	0	51, 108, 147, 153	0
All	All	1689/1720 (98%)	0.06	56 (3%)	50	43	38, 64, 135, 201	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	I	16	VAL	18.1
4	I	84	GLY	11.8
4	I	53	LEU	11.5
4	I	87	ALA	9.8
4	I	83	ASN	9.6

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
1	2DT	X	922	19/20	0.97	0.14	-	48,52,57,59	0
1	2DT	P	822	19/20	0.97	0.14	-	48,50,55,56	0

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.