



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 5, 2017 – 12:46 AM EDT

PDB ID : 4X0Q
Title : Ternary complex of human DNA polymerase theta C-terminal domain binding
ddGTP opposite dCMP
Authors : Zahn, K.E.; Doubleie, S.
Deposited on : unknown
Resolution : 3.90 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 : rb-20029824
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) : rb-20029824

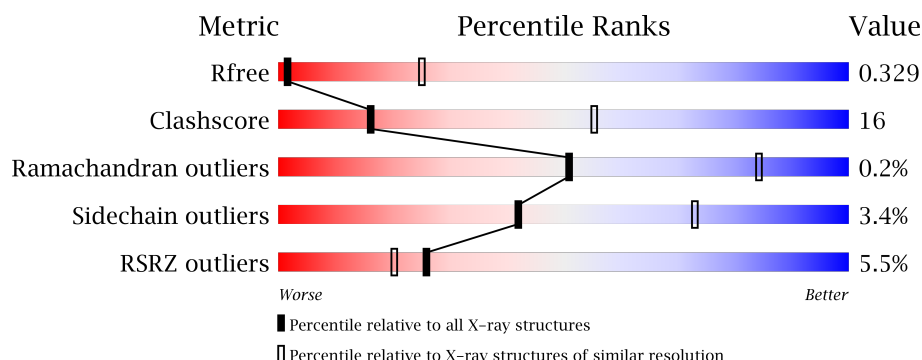
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.90 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	1007 (4.20-3.60)
Clashscore	112137	1103 (4.20-3.60)
Ramachandran outliers	110173	1062 (4.20-3.60)
Sidechain outliers	110143	1053 (4.20-3.60)
RSRZ outliers	101464	1020 (4.20-3.60)

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 $\leq 5\%$. 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	772	<div> <div>4%</div> <div>51%</div> <div>29%</div> <div>19%</div> </div>
1	B	772	<div> <div>5%</div> <div>53%</div> <div>27%</div> <div>19%</div> </div>
2	E	17	<div> <div>6%</div> <div>24%</div> <div>53%</div> <div>24%</div> </div>
2	G	17	<div> <div>6%</div> <div>6%</div> <div>65%</div> <div>6%</div> <div>24%</div> </div>
3	F	13	<div> <div>31%</div> <div>38%</div> <div>31%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	13	

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
4	GOL	A	2601	-	-	-	X
4	GOL	A	2602	-	-	-	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 10902 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 theta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	625	Total	C	N	O	S	0	0	0
			4963	3157	848	929	29			
1	B	625	Total	C	N	O	S	0	0	0
			4963	3157	848	929	29			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	13	Total	C	N	O	P	0	0	0
			263	126	51	74	12			
2	G	13	Total	C	N	O	P	0	0	0
			263	126	51	74	12			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	9	Total	C	N	O	P	0	0	0
			182	88	29	56	9			
3	H	9	Total	C	N	O	P	0	0	0
			182	88	29	56	9			

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).

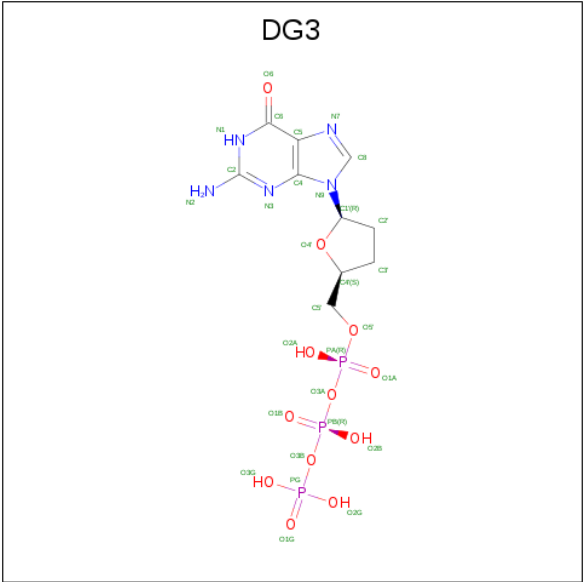


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			6	3	3		
4	A	1	Total	C	O	0	0
			6	3	3		
4	B	1	Total	C	O	0	0
			6	3	3		
4	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Mg	0	0
			1	1		
5	A	1	Total	Mg	0	0
			1	1		

- Molecule 6 is 2'-3'-DIDEOXYGUANOSINE-5'-TRIPHOSPHATE (three-letter code: DG3) (formula: C₁₀H₁₆N₅O₁₂P₃).

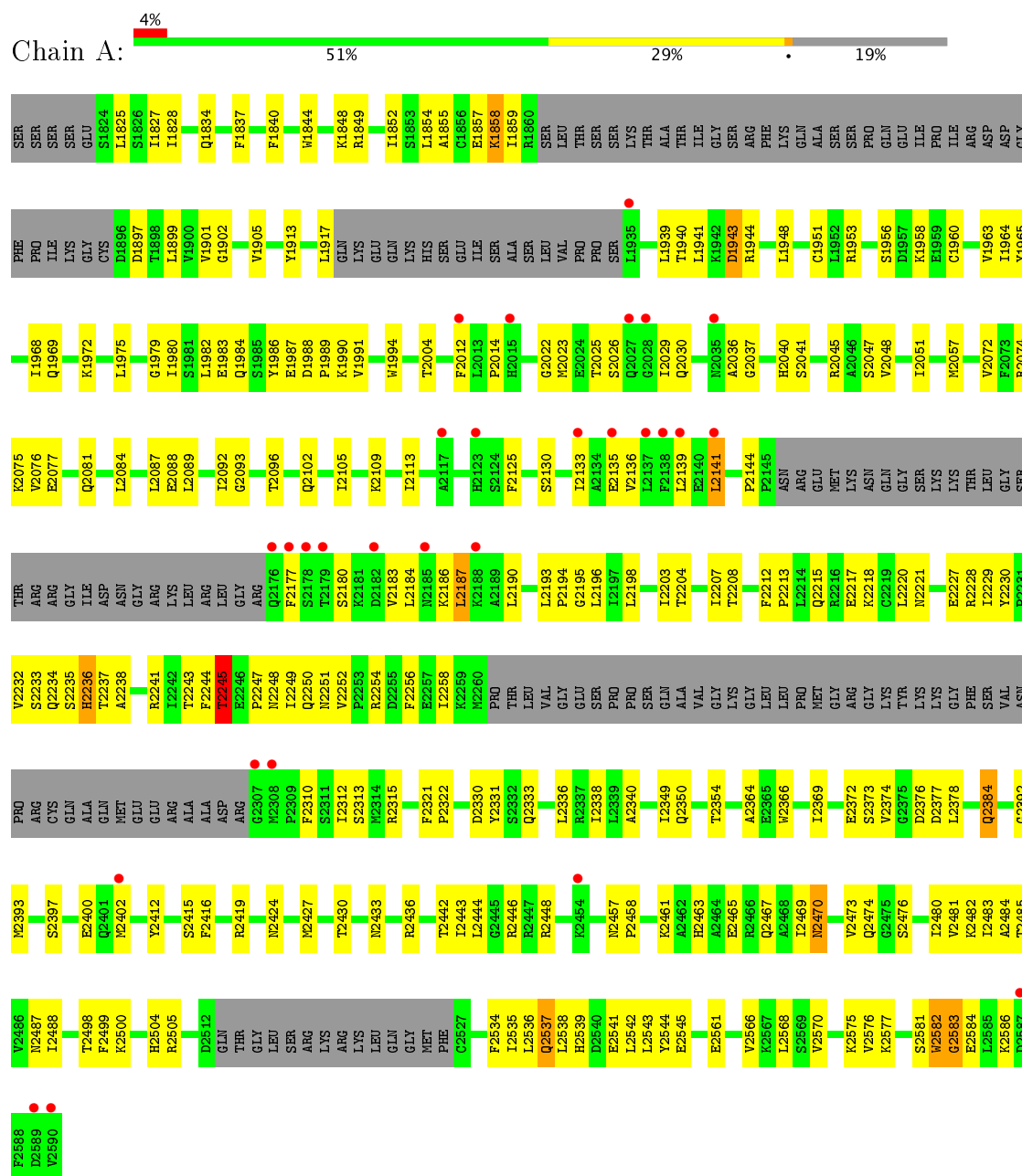


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	A	1	Total	C	N	O	P	0	0
			30	10	5	12	3		
6	B	1	Total	C	N	O	P	0	0
			30	10	5	12	3		

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 polymerase theta



- Chain B:**

Category	Percentage
ARG	5%
LYS	53%
LEU	27%
GLN	19%
MET	5%

- Chain E:
-
- | Amino Acid | Percentage |
|------------|------------|
| C1 | 6% |
| C5 | 24% |
| A6 | 53% |
| A7 | 24% |
| T8 | 0% |
| G9 | 0% |
| A10 | 0% |
| C11 | 0% |
| A12 | 0% |
| G13 | 0% |
| DC | 0% |
| DC | 0% |
| DC | 0% |
| DC | 0% |

- Chain G: 

- 

Chain F:

31%

38%

31%



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

Chain H:

8%

31%

38%

31%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 2 21	Depositor
Cell constants a, b, c, α , β , γ	99.70Å 135.39Å 159.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.93 – 3.90 39.82 – 3.85	Depositor EDS
% Data completeness (in resolution range)	82.5 (29.93-3.90) 82.6 (39.82-3.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.05 (at 3.87Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.258 , 0.316 0.265 , 0.329	Depositor DCC
R_{free} test set	873 reflections (4.99%)	DCC
Wilson B-factor (Å ²)	102.3	Xtriage
Anisotropy	0.997	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 68.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	10902	wwPDB-VP
Average B, all atoms (Å ²)	145.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 42.70 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.9642e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical 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: GOL, MG, DG3

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	0.31	0/5056	0.57	0/6818
1	B	0.30	0/5056	0.57	0/6818
2	E	0.57	0/295	0.87	0/453
2	G	0.57	0/295	0.88	1/453 (0.2%)
3	F	0.60	0/202	1.07	0/309
3	H	0.63	0/202	1.04	0/309
All	All	0.34	0/11106	0.62	1/15160 (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	A	0	3
1	B	0	4
All	All	0	7

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	5	DC	O4'-C1'-N1	5.12	111.58	108.00

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1953	ARG	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	2245	THR	Peptide
1	A	2583	GLY	Peptide
1	B	1953	ARG	Peptide
1	B	2245	THR	Peptide

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	A	4963	0	4980	156	0
1	B	4963	0	4980	157	0
2	E	263	0	147	12	0
2	G	263	0	147	14	0
3	F	182	0	102	9	0
3	H	182	0	102	6	0
4	A	12	0	16	3	0
4	B	12	0	16	1	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
6	A	30	0	12	4	0
6	B	30	0	12	4	0
All	All	10902	0	10514	345	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 16.

The worst 5 of 345 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:2030:GLN:HB3	1:A:2045:ARG:HB3	1.53	0.91
1:A:1857:GLU:HB3	1:A:1858:LYS:HD3	1.55	0.88
1:B:2030:GLN:HB3	1:B:2045:ARG:HB3	1.56	0.87
1:A:1968:ILE:HD13	1:A:2233:SER:HB3	1.57	0.86
1:B:1968:ILE:HD13	1:B:2233:SER:HB3	1.61	0.83

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	
1	A	613/772 (79%)	545 (89%)	67 (11%)	1 (0%)	51	84
1	B	613/772 (79%)	541 (88%)	70 (11%)	2 (0%)	44	80
All	All	1226/1544 (79%)	1086 (89%)	137 (11%)	3 (0%)	51	84

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2144	PRO
1	B	2144	PRO
1	B	1859	ILE

5.3.2 Protein sidechains

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	551/675 (82%)	533 (97%)	18 (3%)	43	73
1	B	551/675 (82%)	532 (97%)	19 (3%)	42	73
All	All	1102/1350 (82%)	1065 (97%)	37 (3%)	42	73

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2566	VAL
1	B	1899	LEU
1	B	2537	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	2582	TRP
1	B	1854	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2221	ASN
1	A	2384	GLN
1	B	2215	GLN
1	A	2215	GLN
1	B	1984	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](#)

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](#)

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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	GOL	A	2601	-	5,5,5	0.31	0	5,5,5	0.31	0
4	GOL	A	2602	-	5,5,5	0.37	0	5,5,5	0.33	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
6	DG3	A	2604	5	25,32,32	1.12	2 (8%)	23,50,50	2.43	8 (34%)
4	GOL	B	2601	-	5,5,5	0.33	0	5,5,5	0.31	0
4	GOL	B	2602	-	5,5,5	0.37	0	5,5,5	0.43	0
6	DG3	B	2604	5	25,32,32	1.13	2 (8%)	23,50,50	2.36	9 (39%)

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	GOL	A	2601	-	-	0/4/4/4	0/0/0/0
4	GOL	A	2602	-	-	0/4/4/4	0/0/0/0
6	DG3	A	2604	5	-	0/18/31/31	0/3/3/3
4	GOL	B	2601	-	-	0/4/4/4	0/0/0/0
4	GOL	B	2602	-	-	0/4/4/4	0/0/0/0
6	DG3	B	2604	5	-	0/18/31/31	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	A	2604	DG3	C5-C4	2.99	1.47	1.40
6	B	2604	DG3	C5-C4	3.06	1.47	1.40
6	A	2604	DG3	C6-C5	3.87	1.48	1.41
6	B	2604	DG3	C6-C5	3.94	1.48	1.41

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	2604	DG3	C6-C5-C4	-4.15	116.72	120.84
6	B	2604	DG3	C6-C5-C4	-4.02	116.84	120.84
6	B	2604	DG3	C5-C6-N1	-3.77	118.12	123.48
6	A	2604	DG3	C5-C6-N1	-3.72	118.19	123.48
6	A	2604	DG3	N3-C2-N1	-3.30	122.65	127.46

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	2601	GOL	2	0
4	A	2602	GOL	1	0
6	A	2604	DG3	4	0
4	B	2601	GOL	1	0
6	B	2604	DG3	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	625/772 (80%)	0.18	28 (4%) 34 27	72, 137, 251, 323	0
1	B	625/772 (80%)	0.19	40 (6%) 20 14	65, 128, 234, 303	0
2	E	13/17 (76%)	0.86	1 (7%) 14 11	118, 139, 269, 363	0
2	G	13/17 (76%)	0.50	1 (7%) 14 11	98, 130, 231, 307	0
3	F	9/13 (69%)	0.23	0 100 100	115, 135, 286, 323	0
3	H	9/13 (69%)	0.24	1 (11%) 6 6	110, 134, 214, 256	0
All	All	1294/1604 (80%)	0.20	71 (5%) 26 20	65, 132, 242, 363	0

The worst 5 of 71 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	2178	SER	8.6
1	B	2307	GLY	7.8
1	A	2179	THR	7.2
1	A	2307	GLY	6.8
1	B	2135	GLU	6.3

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](#)

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
4	GOL	A	2601	6/6	0.61	0.60	3.85	120,136,145,151	0
4	GOL	A	2602	6/6	0.91	0.44	2.90	72,72,73,79	0
4	GOL	B	2602	6/6	0.88	0.36	1.16	54,55,56,56	0
6	DG3	A	2604	30/30	0.92	0.32	0.88	95,102,141,146	0
6	DG3	B	2604	30/30	0.92	0.32	0.05	72,76,113,120	0
5	MG	B	2603	1/1	0.91	0.27	-	76,76,76,76	0
5	MG	A	2603	1/1	0.93	0.31	-	95,95,95,95	0
4	GOL	B	2601	6/6	0.63	0.56	-	107,137,143,147	0

6.5 Other polymers [i](#)

There are no such residues in this entry.