



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 04:27 am GMT

PDB ID : 1YA6
Title : alpha-glucosyltransferase in complex with UDP and a 13-mer DNA containing a central A:G mismatch
Authors : Lariviere, L.; Sommer, N.; Morera, S.
Deposited on : 2004-12-17
Resolution : 2.40 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 : trunk28620
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) : recalc28949

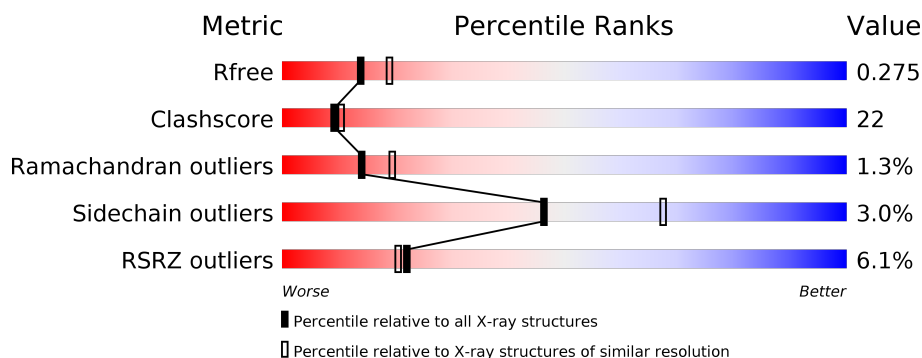
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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	3166 (2.40-2.40)
Clashscore	112137	3674 (2.40-2.40)
Ramachandran outliers	110173	3616 (2.40-2.40)
Sidechain outliers	110143	3617 (2.40-2.40)
RSRZ outliers	101464	3195 (2.40-2.40)

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	C	12	<div> <div>50%</div> <div>42%</div> <div>8%</div> </div>
2	D	12	<div> <div>75%</div> <div>25%</div> </div>
3	A	403	<div> <div>6%</div> <div>60%</div> <div>32%</div> <div>6%</div> </div>
3	B	403	<div> <div>5%</div> <div>51%</div> <div>40%</div> <div>6%</div> </div>

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	NCO	B	3002	-	-	-	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6977 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*TP*AP*CP*TP*AP*AP*GP*AP*TP*AP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	12	Total	C	N	O	P	0	0	0
			246	119	49	67	11			

- Molecule 2 is a DNA chain called 5'-D(*CP*TP*AP*TP*CP*TP*GP*AP*GP*TP*AP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	12	Total	C	N	O	P	0	0	0
			242	118	41	72	11			

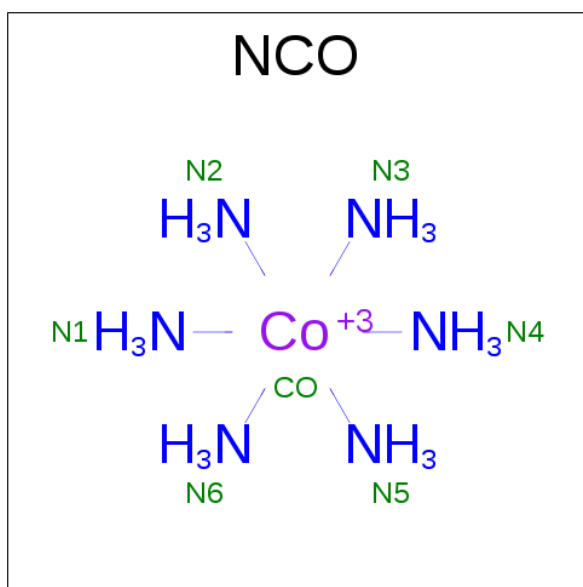
- Molecule 3 is a protein called DNA alpha-glucosyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	380	Total	C	N	O	S	0	0	0
			3118	1986	532	584	16			
3	B	377	Total	C	N	O	S	0	0	0
			3091	1970	527	577	17			

There are 6 discrepancies between the modelled and reference sequences:

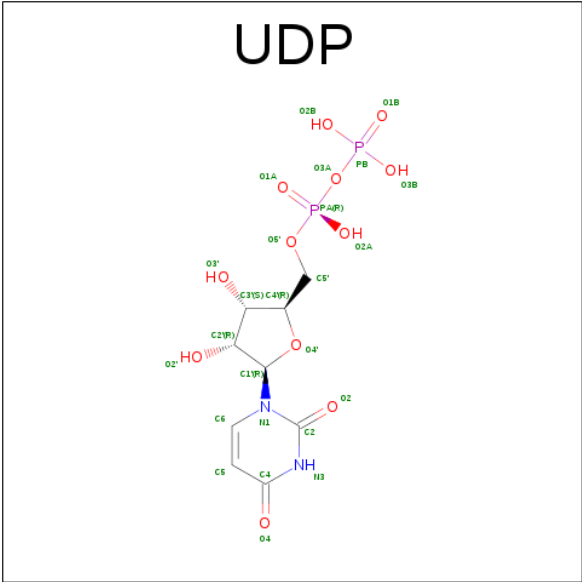
Chain	Residue	Modelled	Actual	Comment	Reference
A	998	MET	-	CLONING ARTIFACT	UNP P04519
A	999	GLY	-	CLONING ARTIFACT	UNP P04519
A	1000	SER	-	CLONING ARTIFACT	UNP P04519
B	998	MET	-	CLONING ARTIFACT	UNP P04519
B	999	GLY	-	CLONING ARTIFACT	UNP P04519
B	1000	SER	-	CLONING ARTIFACT	UNP P04519

- Molecule 4 is COBALT HEXAMMINE(III) (three-letter code: NCO) (formula: CoH₁₈N₆).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	Co	N	0	0
			7	1	6		
4	B	1	Total	Co	N	0	0
			7	1	6		
4	A	1	Total	Co	N	0	0
			7	1	6		
4	B	1	Total	Co	N	0	0
			7	1	6		
4	D	1	Total	Co	N	0	0
			7	1	6		
4	B	1	Total	Co	N	0	0
			7	1	6		
4	A	1	Total	Co	N	0	0
			7	1	6		

- Molecule 5 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula: C₉H₁₄N₂O₁₂P₂).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
5	B	1	Total	C	N	O	P	0	0
			25	9	2	12	2		

- Molecule 6 is water.

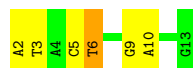
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	89	Total	O	0	0
			89	89		
6	B	81	Total	O	0	0
			81	81		
6	C	5	Total	O	0	0
			5	5		
6	D	6	Total	O	0	0
			6	6		

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.

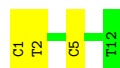
- Molecule 1: 5'-D(*AP*TP*AP*CP*TP*AP*AP*GP*AP*TP*AP*G)-3'

Chain C: 



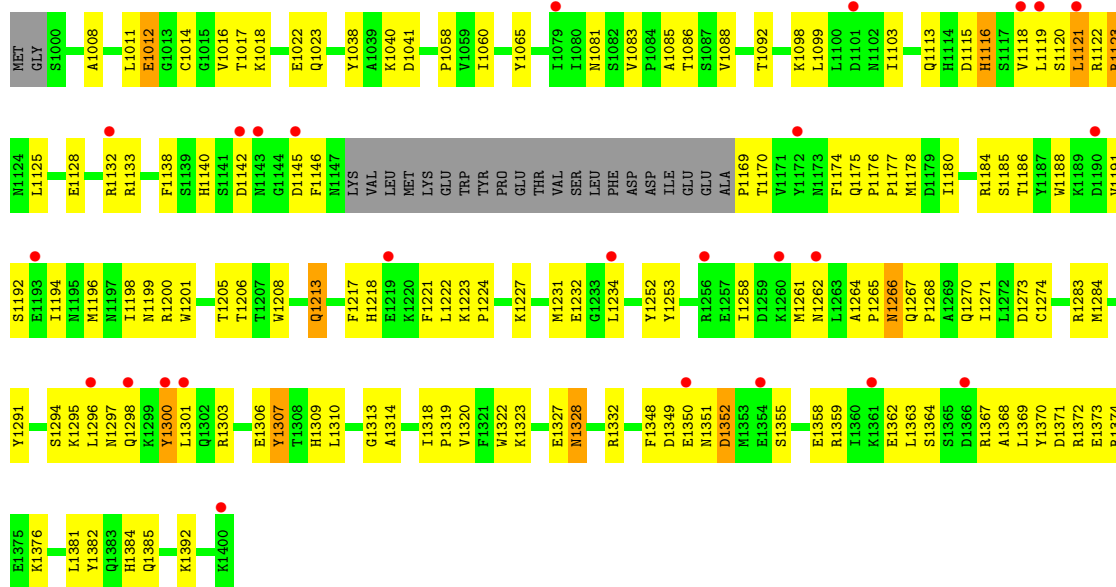
- Molecule 2: 5'-D(*CP*TP*AP*TP*CP*TP*GP*AP*GP*TP*AP*T)-3'

Chain D: 



- Molecule 3: DNA alpha-glucosyltransferase

Chain A: 



- Molecule 3: DNA alpha-glucosyltransferase

Chain B: 

N1335	T1336	P1337	L1338	T1339	S1340	H1341	D1342	I1346	W1347	D1352	H1353	E1354	S1355	E1358	R1359	I1360	K1361	E1362	L1363	D1366	R1367	A1368	L1369	Y1370	D1371	R1372	E1373	K1376	A1377	Y1378	Y1382	Q1383	H1384	Q1385	D1386	S1387	C1390	F1391	K1392	E1393	D1396	I1397	T1398	K1400										
PI268	A1269	Q1270	I1271	L1272	D1273	C1274	Y1275	I1276	E1279	M1280	R1283	M1284	F1289	G1290	R1291	Q1292	K1295	L1296	N1297	Q1298	K1299	Y1300	L1301	Q1302	R1303	S1304	L1305	E1306	Y1307	T1308	H1309	L1310	E1311	L1312	C1315	G1316	T1317	I1318	P1319	V1320	F1321	W1322	K1323	S1324	T1325	G1326	E1327	N1328	L1329	K1330	F1331	R1332	V1333	D1334
P1177	I1180	Y1187	W1188	K1189	D1190	V1191	S1192	E1193	I1194	N1199	R1200	W1201	I1202	G1203	R1204	T1205	T1206	T1207	W1208	K1209	Q1213	H1218	L1222	K1223	P1224	K1227	S1228	T1229	W1230	M1231	E1232	G1233	L1234	E1235	I1241	A1242	I1243	K1244	E1245	K1246	Y1252	R1256	E1257	I1258	M1261	N1262	Q1267							
K1098	L1099	L1100	D1101	R1108	H1114	D1115	H1116	L1119	R1122	R1123	N1124	L1125	G1126	L1127	E1128	E1129	T1130	W1131	R1132	R1133	F1138	D1142	ASN	GLY	ASP	PHE	ASN	LYS	VAL	LEU	MET	LYS	GLU	TRP	TYR	PRO	GLU	THR	VAL	SER	LEU	PHE	ASP	ASP	ILE	GLU	GLU	ALA	P1169	F1174	Q1175	P1176		
M998	G999	I1003	A1008	R1009	E1012	G1013	C1014	K1018	Q1023	R1024	D1025	W1026	K1029	V1034	V1037	Y1038	A1039	A1040	D1041	R1046	S1054	V1059	I1060	L1061	A1062	K1063	E1064	Y1065	D1066	K1067	A1068	L1069	N1081	S1082	V1083	T1086	Q1089	E1090	A1091	T1092	I1093	N1094	N1095	Y1096	K1097									

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	45.85Å 119.32Å 86.81Å 90.00° 92.87° 90.00°	Depositor
Resolution (Å)	20.00 – 2.40 19.98 – 2.40	Depositor EDS
% Data completeness (in resolution range)	(Not available) (20.00-2.40) 99.9 (19.98-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.30 (at 2.41Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.230 , 0.280 0.227 , 0.275	Depositor DCC
R_{free} test set	1795 reflections (5.21%)	DCC
Wilson B-factor (Å ²)	50.4	Xtriage
Anisotropy	0.718	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 50.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.030 for h,-k,-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6977	wwPDB-VP
Average B, all atoms (Å ²)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.49% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: UDP, NCO

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	C	0.48	0/277	0.76	0/426
2	D	0.44	0/270	0.71	0/415
3	A	0.41	0/3190	0.63	0/4301
3	B	0.39	0/3162	0.62	0/4262
All	All	0.41	0/6899	0.64	0/9404

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	C	0	1
3	A	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	1307	TYR	Sidechain
1	C	6	DT	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	C	246	0	137	4	0
2	D	242	0	139	6	0
3	A	3118	0	3060	143	0
3	B	3091	0	3044	141	0
4	A	21	0	0	3	0
4	B	21	0	0	4	0
4	D	7	0	0	0	0
5	A	25	0	11	0	0
5	B	25	0	11	1	0
6	A	89	0	0	3	0
6	B	81	0	0	3	0
6	C	5	0	0	0	0
6	D	6	0	0	0	0
All	All	6977	0	6402	290	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 22.

The worst 5 of 290 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:1213:GLN:H	3:A:1213:GLN:NE2	1.68	0.92
3:A:1118:VAL:HG12	3:A:1145:ASP:HB2	1.55	0.87
3:A:1113:GLN:HE22	3:A:1115:ASP:HB2	1.40	0.85
3:A:1116:HIS:CE1	4:A:2002:NCO:N3	2.45	0.85
3:A:1318:ILE:HG12	3:A:1370:TYR:CE1	2.15	0.82

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	
3	A	376/403 (93%)	336 (89%)	36 (10%)	4 (1%)	17	23
3	B	373/403 (93%)	336 (90%)	31 (8%)	6 (2%)	11	15
All	All	749/806 (93%)	672 (90%)	67 (9%)	10 (1%)	14	19

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	1352	ASP
3	B	1191	VAL
3	B	1335	ASN
3	A	1185	SER
3	B	1273	ASP

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	
3	A	347/368 (94%)	338 (97%)	9 (3%)	51	72
3	B	344/368 (94%)	332 (96%)	12 (4%)	41	61
All	All	691/736 (94%)	670 (97%)	21 (3%)	46	67

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	1012	GLU
3	B	1119	LEU
3	B	1291	TYR
3	A	1392	LYS
3	B	1292	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such

sidechains are listed below:

Mol	Chain	Res	Type
3	A	1384	HIS
3	B	1094	ASN
3	B	1383	GLN
3	A	1394	GLN
3	B	1023	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

9 ligands 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	$\# Z > 2$
5	UDP	A	2001	-	21,26,26	1.11	2 (9%)	22,40,40	3.40	3 (13%)
4	NCO	A	2002	-	6,6,6	1.02	0	0,15,15	0.00	-
4	NCO	A	2003	-	6,6,6	1.21	0	0,15,15	0.00	-
4	NCO	A	2004	-	6,6,6	1.15	0	0,15,15	0.00	-
5	UDP	B	3001	-	21,26,26	1.11	3 (14%)	22,40,40	3.31	3 (13%)
4	NCO	B	3002	-	6,6,6	1.09	0	0,15,15	0.00	-
4	NCO	B	3003	-	6,6,6	1.15	0	0,15,15	0.00	-
4	NCO	B	3004	-	6,6,6	1.16	0	0,15,15	0.00	-

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
4	NCO	D	1003	-	6,6,6	1.25	0	0,15,15	0.00	-

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
5	UDP	A	2001	-	-	0/12/32/32	0/2/2/2
4	NCO	A	2002	-	-	0/0/0/0	0/0/0/0
4	NCO	A	2003	-	-	0/0/0/0	0/0/0/0
4	NCO	A	2004	-	-	0/0/0/0	0/0/0/0
5	UDP	B	3001	-	-	0/12/32/32	0/2/2/2
4	NCO	B	3002	-	-	0/0/0/0	0/0/0/0
4	NCO	B	3003	-	-	0/0/0/0	0/0/0/0
4	NCO	B	3004	-	-	0/0/0/0	0/0/0/0
4	NCO	D	1003	-	-	0/0/0/0	0/0/0/0

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	3001	UDP	C6-C5	-2.03	1.33	1.38
5	B	3001	UDP	PB-O3A	2.38	1.63	1.60
5	A	2001	UDP	PB-O3A	2.47	1.64	1.60
5	A	2001	UDP	C4-N3	2.77	1.38	1.33
5	B	3001	UDP	C4-N3	2.77	1.38	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	2001	UDP	C5-C4-N3	-3.66	114.39	123.12
5	B	3001	UDP	C5-C4-N3	-3.53	114.68	123.12
5	B	3001	UDP	O2B-PB-O1B	2.23	119.22	110.50
5	A	2001	UDP	O2B-PB-O1B	2.29	119.46	110.50
5	B	3001	UDP	C4-N3-C2	14.72	126.78	114.13

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	2002	NCO	3	0
5	B	3001	UDP	1	0
4	B	3002	NCO	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	C	12/12 (100%)	0.14	0 100 100	38, 62, 71, 72	0
2	D	12/12 (100%)	-0.24	0 100 100	50, 60, 74, 74	0
3	A	380/403 (94%)	0.28	26 (6%) 18 16	34, 60, 87, 100	0
3	B	377/403 (93%)	0.33	22 (5%) 24 22	39, 65, 84, 97	0
All	All	781/830 (94%)	0.29	48 (6%) 22 20	34, 62, 86, 100	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	1143	ASN	7.1
3	A	1145	ASP	5.3
3	B	1300	TYR	4.4
3	A	1296	LEU	3.9
3	B	1132	ARG	3.4

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(\AA^2)	Q<0.9
4	NCO	B	3002	7/7	0.93	0.22	3.63	78,79,79,79	0
4	NCO	A	2002	7/7	0.96	0.19	0.32	63,63,65,66	0
4	NCO	A	2004	7/7	0.93	0.19	0.19	92,92,93,93	7
4	NCO	B	3003	7/7	0.94	0.17	-0.47	98,99,99,99	0
4	NCO	D	1003	7/7	0.97	0.14	-0.49	88,89,89,91	0
4	NCO	A	2003	7/7	0.96	0.15	-0.51	81,82,84,84	0
5	UDP	B	3001	25/25	0.98	0.10	-1.24	37,43,47,49	0
5	UDP	A	2001	25/25	0.98	0.08	-1.40	38,44,46,48	0
4	NCO	B	3004	7/7	0.92	0.20	-	71,71,71,73	7

6.5 Other polymers [i](#)

There are no such residues in this entry.