



wwPDB X-ray Structure Validation Summary Report

Feb 1, 2016 – 08:33 PM GMT

PDB ID : 4TNX
Title : Structure basis of cellular dNTP regulation, SAMHD1-GTP-dGTP complex
Authors : Ji, X.; Tang, C.; Zhao, Q.; Wang, W.; Xiong, Y.
Deposited on : 2014-06-05
Resolution : 2.31 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 (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

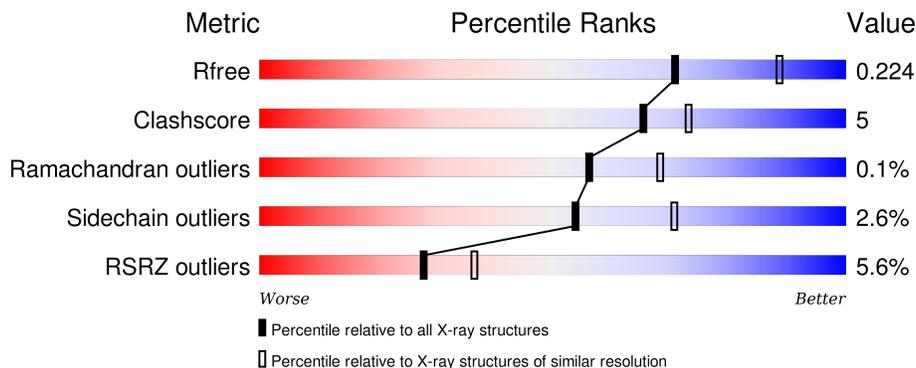
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.31 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}	91344	4425 (2.34-2.30)
Clashscore	102246	5057 (2.34-2.30)
Ramachandran outliers	100387	5008 (2.34-2.30)
Sidechain outliers	100360	5007 (2.34-2.30)
RSRZ outliers	91569	4432 (2.34-2.30)

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	514	 2% 82% 10% • 6%
1	B	514	 5% 84% 8% • 7%
1	C	514	 10% 84% 9% • 6%
1	D	514	 4% 84% 10% • 6%

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
5	3PO	A	705	-	-	-	X

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 16322 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.

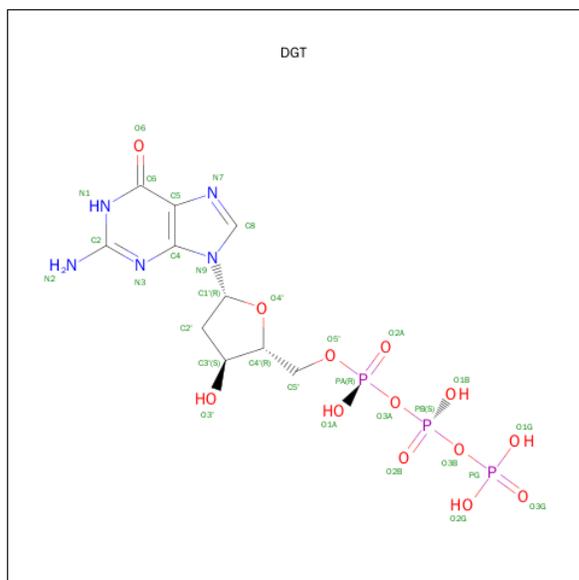
- Molecule 1 is a protein called Deoxynucleoside triphosphate triphosphohydrolase SAMHD1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	481	3956	2529	689	717	21	0	3	0
1	B	479	3940	2521	686	712	21	0	3	0
1	C	481	3947	2524	688	714	21	0	2	0
1	D	483	3961	2531	690	719	21	0	2	0

There are 8 discrepancies between the modelled and reference sequences:

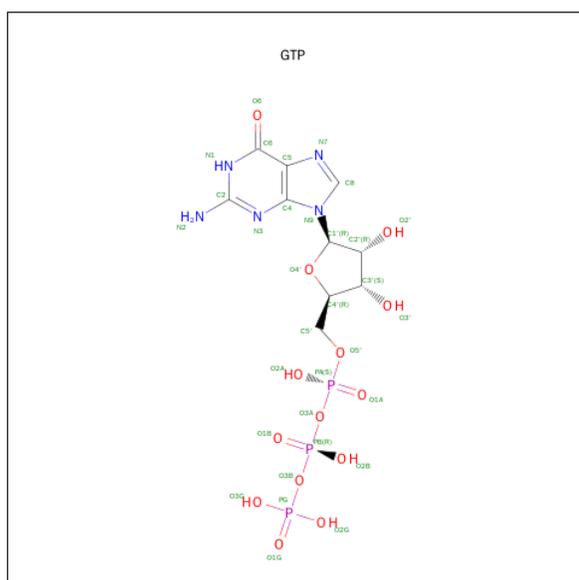
Chain	Residue	Modelled	Actual	Comment	Reference
A	206	ARG	HIS	engineered mutation	UNP Q9Y3Z3
A	207	ASN	ASP	engineered mutation	UNP Q9Y3Z3
B	206	ARG	HIS	engineered mutation	UNP Q9Y3Z3
B	207	ASN	ASP	engineered mutation	UNP Q9Y3Z3
C	206	ARG	HIS	engineered mutation	UNP Q9Y3Z3
C	207	ASN	ASP	engineered mutation	UNP Q9Y3Z3
D	206	ARG	HIS	engineered mutation	UNP Q9Y3Z3
D	207	ASN	ASP	engineered mutation	UNP Q9Y3Z3

- Molecule 2 is 2'-DEOXYGUANOSINE-5'-TRIPHOSPHATE (three-letter code: DGT) (formula: $C_{10}H_{16}N_5O_{13}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
2	A	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
2	B	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
2	C	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
2	D	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

- Molecule 3 is GUANOSINE-5'-TRIPHOSPHATE (three-letter code: GTP) (formula: $C_{10}H_{16}N_5O_{14}P_3$).

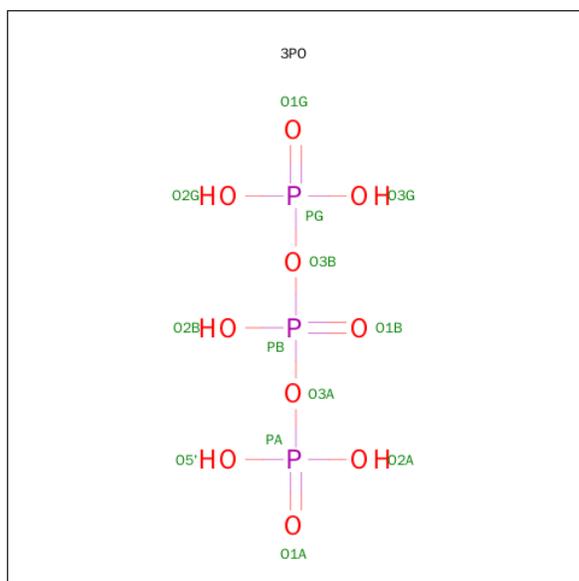


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			32	10	5	14	3		
3	A	1	Total	C	N	O	P	0	0
			32	10	5	14	3		
3	B	1	Total	C	N	O	P	0	0
			32	10	5	14	3		
3	C	1	Total	C	N	O	P	0	0
			32	10	5	14	3		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		
4	A	1	Total	Mg	0	0
			1	1		
4	C	2	Total	Mg	0	0
			2	2		

- Molecule 5 is TRIPHOSPHATE (three-letter code: 3PO) (formula: H₅O₁₀P₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	P	0	0
			13	10	3		

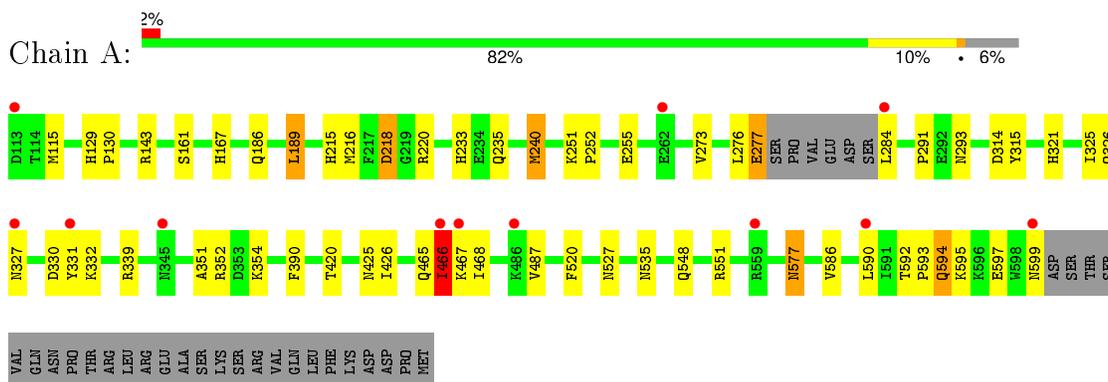
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	80	Total 80	O 80	0	0
6	B	49	Total 49	O 49	0	0
6	C	47	Total 47	O 47	0	0
6	D	73	Total 73	O 73	0	0

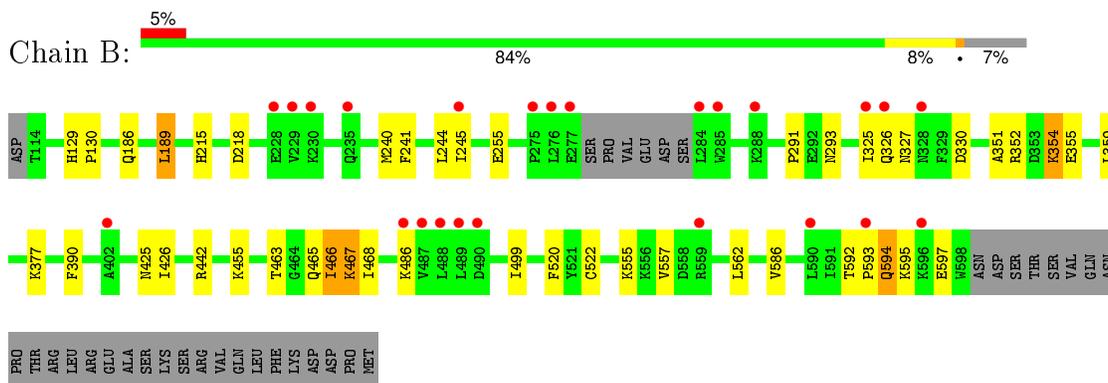
3 Residue-property plots [i](#)

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.

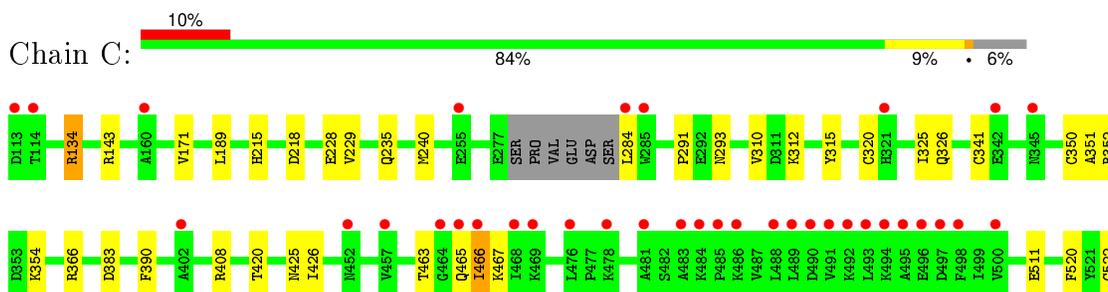
- Molecule 1: Deoxynucleoside triphosphate triphosphohydrolase SAMHD1



- Molecule 1: Deoxynucleoside triphosphate triphosphohydrolase SAMHD1

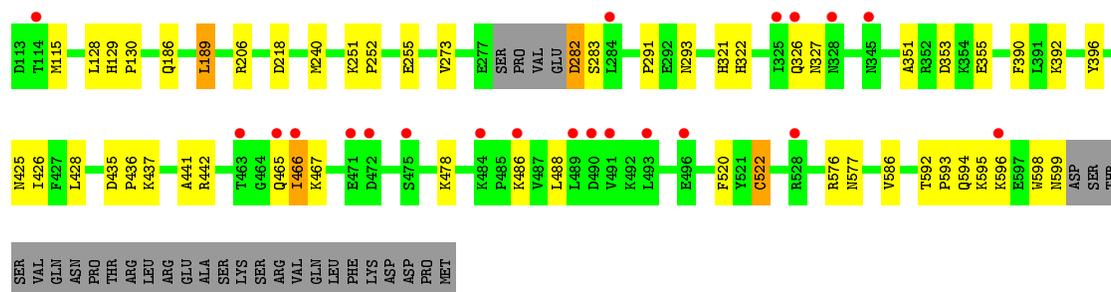
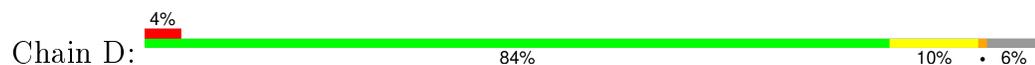


- Molecule 1: Deoxynucleoside triphosphate triphosphohydrolase SAMHD1





● Molecule 1: Deoxynucleoside triphosphate triphosphohydrolase SAMHD1



4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	80.89Å 140.73Å 97.60Å 90.00° 114.71° 90.00°	Depositor
Resolution (Å)	50.00 – 2.31 47.82 – 2.31	Depositor EDS
% Data completeness (in resolution range)	99.2 (50.00-2.31) 99.2 (47.82-2.31)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.56 (at 2.32Å)	Xtriage
Refinement program	REFMAC 5.8.0071	Depositor
R, R_{free}	0.188 , 0.226 0.192 , 0.224	Depositor DCC
R_{free} test set	4335 reflections (5.29%)	DCC
Wilson B-factor (Å ²)	40.8	Xtriage
Anisotropy	0.305	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 43.6	EDS
Estimated twinning fraction	0.028 for h,-k,-h-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	1 of 86352 reflections (0.001%)	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	16322	wwPDB-VP
Average B, all atoms (Å ²)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.55% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: GTP, MG, DGT, 3PO

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	0.76	1/4048 (0.0%)	0.81	5/5464 (0.1%)
1	B	0.69	1/4032 (0.0%)	0.77	4/5442 (0.1%)
1	C	0.65	0/4039	0.76	5/5452 (0.1%)
1	D	0.76	1/4053 (0.0%)	0.82	7/5471 (0.1%)
All	All	0.72	3/16172 (0.0%)	0.79	21/21829 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	355	GLU	CD-OE2	6.14	1.32	1.25
1	A	161	SER	CB-OG	5.90	1.50	1.42
1	D	355	GLU	CD-OE1	5.41	1.31	1.25

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	240	MET	CG-SD-CE	-8.38	86.80	100.20
1	A	551	ARG	NE-CZ-NH2	6.96	123.78	120.30
1	D	218	ASP	CB-CG-OD2	6.76	124.39	118.30
1	A	218	ASP	CB-CG-OD2	6.55	124.19	118.30
1	B	330	ASP	CB-CG-OD2	-6.45	112.49	118.30

There are no chirality outliers.

There are no planarity outliers.

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	A	3956	0	3935	56	0
1	B	3940	0	3925	35	0
1	C	3947	0	3930	34	0
1	D	3961	0	3939	47	0
2	A	31	0	12	4	0
2	B	31	0	12	1	0
2	C	31	0	12	1	0
2	D	31	0	12	2	0
3	A	64	0	24	0	0
3	B	32	0	12	0	0
3	C	32	0	12	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	2	0	0	0	0
5	A	13	0	0	1	0
6	A	80	0	0	4	0
6	B	49	0	0	0	0
6	C	47	0	0	2	0
6	D	73	0	0	6	0
All	All	16322	0	15825	158	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 5.

The worst 5 of 158 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:D:577:ASN:ND2	1:D:595:LYS:HZ1	1.60	1.00
1:D:596:LYS:HB3	6:D:907:HOH:O	1.63	0.98
1:A:597:GLU:OE1	1:A:597:GLU:N	1.97	0.96
1:A:216:MET:HA	6:A:869:HOH:O	1.67	0.94
1:D:577:ASN:ND2	1:D:595:LYS:NZ	2.19	0.89

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	
1	A	480/514 (93%)	474 (99%)	5 (1%)	1 (0%)	52	64
1	B	478/514 (93%)	470 (98%)	8 (2%)	0	100	100
1	C	479/514 (93%)	474 (99%)	5 (1%)	0	100	100
1	D	481/514 (94%)	474 (98%)	7 (2%)	0	100	100
All	All	1918/2056 (93%)	1892 (99%)	25 (1%)	1 (0%)	56	68

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	466	ILE

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	
1	A	430/459 (94%)	419 (97%)	11 (3%)	54	71
1	B	428/459 (93%)	419 (98%)	9 (2%)	61	78
1	C	429/459 (94%)	417 (97%)	12 (3%)	51	67
1	D	431/459 (94%)	418 (97%)	13 (3%)	48	65
All	All	1718/1836 (94%)	1673 (97%)	45 (3%)	54	71

5 of 45 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	134	ARG
1	C	315	TYR
1	D	486	LYS
1	C	235	GLN
1	C	463	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	425	ASN
1	D	577	ASN
1	C	425	ASN
1	A	425	ASN
1	C	322	HIS

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 13 ligands modelled in this entry, 4 are monoatomic - leaving 9 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
2	DGT	A	701	4	25,33,33	1.38	3 (12%)	35,52,52	2.03	9 (25%)
3	GTP	A	702	4	25,34,34	1.28	3 (12%)	34,54,54	1.94	8 (23%)
3	GTP	A	704	4	25,34,34	1.53	3 (12%)	34,54,54	2.08	8 (23%)
5	3PO	A	705	-	8,12,12	0.63	0	15,20,20	1.55	4 (26%)
3	GTP	B	701	4	25,34,34	1.40	3 (12%)	34,54,54	1.99	10 (29%)
2	DGT	B	703	4	25,33,33	1.21	3 (12%)	35,52,52	2.19	11 (31%)
2	DGT	C	703	4	25,33,33	1.30	3 (12%)	35,52,52	2.02	10 (28%)
3	GTP	C	704	4	25,34,34	1.23	2 (8%)	34,54,54	2.09	11 (32%)
2	DGT	D	800	4	25,33,33	1.15	3 (12%)	35,52,52	2.00	11 (31%)

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	DGT	A	701	4	-	0/18/34/34	0/3/3/3
3	GTP	A	702	4	-	0/18/38/38	0/3/3/3
3	GTP	A	704	4	-	0/18/38/38	0/3/3/3
5	3PO	A	705	-	-	0/12/12/12	0/0/0/0
3	GTP	B	701	4	-	0/18/38/38	0/3/3/3
2	DGT	B	703	4	-	0/18/34/34	0/3/3/3
2	DGT	C	703	4	-	0/18/34/34	0/3/3/3
3	GTP	C	704	4	-	0/18/38/38	0/3/3/3
2	DGT	D	800	4	-	0/18/34/34	0/3/3/3

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	701	DGT	O4'-C4'	-2.33	1.39	1.45
2	B	703	DGT	C2-N1	-2.26	1.31	1.35
3	A	702	GTP	PG-O2G	-2.13	1.47	1.54
2	D	800	DGT	PB-O1B	-2.02	1.46	1.54
3	A	704	GTP	O4'-C1'	2.33	1.44	1.41

The worst 5 of 82 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	704	GTP	C5-C6-N1	-5.77	115.70	123.59
2	C	703	DGT	C6-C5-C4	-5.21	114.67	120.90

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	703	DGT	C6-C5-C4	-4.97	114.96	120.90
3	A	702	GTP	C5-C6-N1	-4.92	116.86	123.59
2	B	703	DGT	C5-C6-N1	-4.81	117.02	123.59

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	701	DGT	4	0
5	A	705	3PO	1	0
2	B	703	DGT	1	0
2	C	703	DGT	1	0
2	D	800	DGT	2	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	481/514 (93%)	0.05	12 (2%) 61 69	27, 44, 71, 98	0
1	B	479/514 (93%)	0.24	24 (5%) 32 41	32, 54, 86, 113	0
1	C	481/514 (93%)	0.49	51 (10%) 8 12	32, 55, 101, 129	0
1	D	483/514 (93%)	0.11	21 (4%) 39 48	28, 45, 86, 112	0
All	All	1924/2056 (93%)	0.22	108 (5%) 28 36	27, 50, 88, 129	0

The worst 5 of 108 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	488	LEU	7.0
1	C	490	ASP	6.8
1	A	466	ILE	6.6
1	C	493	LEU	6.4
1	C	485	PRO	6.2

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
5	3PO	A	705	13/13	0.88	0.21	2.52	62,89,154,155	0
2	DGT	B	703	31/31	0.98	0.14	-0.41	29,30,34,36	0
2	DGT	D	800	31/31	0.98	0.13	-0.42	34,36,38,38	0
2	DGT	A	701	31/31	0.98	0.13	-0.82	31,34,38,39	0
2	DGT	C	703	31/31	0.98	0.13	-1.10	37,40,48,49	0
3	GTP	C	704	32/32	0.98	0.11	-1.17	36,43,49,50	0
3	GTP	A	702	32/32	0.99	0.10	-1.37	30,31,34,35	0
3	GTP	A	704	32/32	0.98	0.10	-1.43	29,30,32,32	0
3	GTP	B	701	32/32	0.98	0.09	-1.58	37,40,44,45	0
4	MG	C	702	1/1	0.93	0.06	-	49,49,49,49	0
4	MG	C	701	1/1	0.98	0.07	-	43,43,43,43	0
4	MG	A	703	1/1	0.96	0.09	-	34,34,34,34	0
4	MG	B	702	1/1	0.93	0.10	-	49,49,49,49	0

6.5 Other polymers [i](#)

There are no such residues in this entry.