



wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 10:59 AM GMT

PDB ID : 1H2U
Title : STRUCTURE OF THE HUMAN NUCLEAR CAP-BINDING-COMPLEX
(CBC) IN COMPLEX WITH A CAP ANALOGUE M7GPPPG
Authors : Mazza, C.; Segref, A.; Mattaj, I.W.; Cusack, S.
Deposited on : 2002-08-16
Resolution : 2.40 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

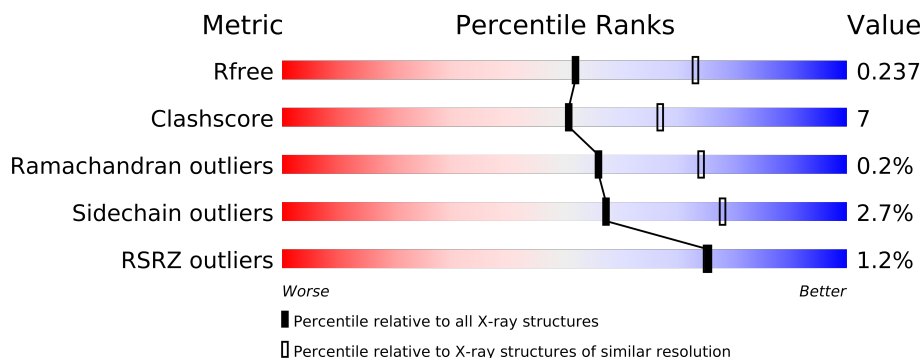
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

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}	66092	2207 (2.40-2.40)
Clashscore	79885	2789 (2.40-2.40)
Ramachandran outliers	78287	2736 (2.40-2.40)
Sidechain outliers	78261	2737 (2.40-2.40)
RSRZ outliers	66119	2210 (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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	723	
1	B	723	
2	X	156	
2	Y	156	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 15364 atoms, of which 0 are hydrogen and 0 are deuterium.

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 80 KDA NUCLEAR CAP BINDING PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	706	Total	C	N	O	S	0	0	0
			5765	3722	971	1034	38			
1	B	706	Total	C	N	O	S	0	0	0
			5762	3720	971	1033	38			

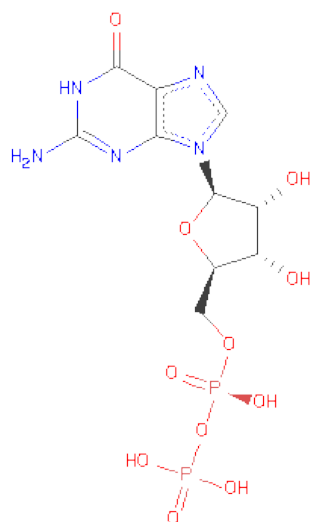
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	479	SER	ALA	ENGINEERED MUTATION	UNP Q09161
B	479	SER	ALA	ENGINEERED MUTATION	UNP Q09161

- Molecule 2 is a protein called 20 KDA NUCLEAR CAP BINDING PROTEIN.

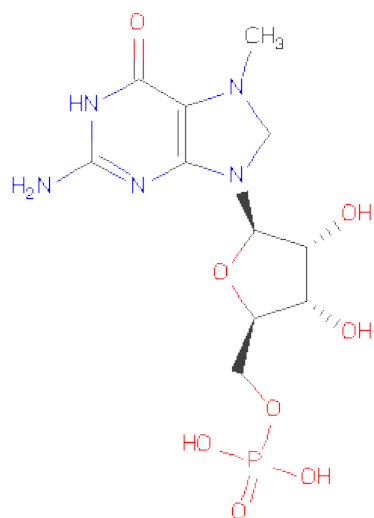
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	149	Total	C	N	O	S	0	0	0
			1214	757	217	234	6			
2	Y	147	Total	C	N	O	S	0	0	0
			1202	749	215	232	6			

- Molecule 3 is GUANOSINE-5'-DIPHOSPHATE (three-letter code: GDP) (formula: C₁₀H₁₅N₅O₁₁P₂).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	X	1	Total	C	N	O	P	0	0
			28	10	5	11	2		
3	Y	1	Total	C	N	O	P	0	0
			28	10	5	11	2		

- Molecule 4 is 7N-METHYL-8-HYDROGUANOSINE-5'-MONOPHOSPHATE (three-letter code: 7MG) (formula: C₁₁H₁₈N₅O₈P).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	X	1	Total	C	N	O	P	0	0
			24	11	5	7	1		

Continued on next page...

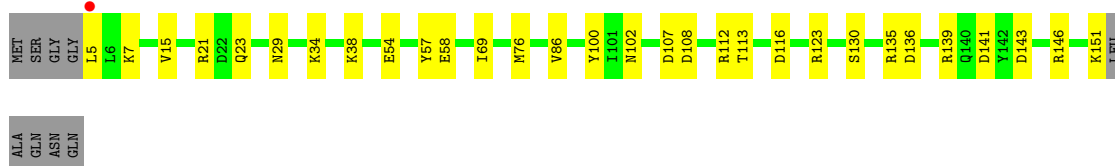
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	Y	1	Total 24	C 11	N 5	O 7	P 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	509	Total 509	O 509	0	0
5	B	538	Total 538	O 538	0	0
5	X	130	Total 130	O 130	0	0
5	Y	140	Total 140	O 140	0	0

- Molecule 2: 20 KDA NUCLEAR CAP BINDING PROTEIN



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	111.84Å 125.72Å 188.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.98 – 2.40 19.98 – 2.30	Depositor EDS
% Data completeness (in resolution range)	92.3 (19.98-2.40) 89.2 (19.98-2.30)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.64 (at 2.30Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.194 , 0.245 0.188 , 0.237	Depositor DCC
R_{free} test set	4794 reflections (4.98%)	DCC
Wilson B-factor (Å ²)	29.4	Xtriage
Anisotropy	0.665	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 37.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 105502 reflections (0.001%)	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	15364	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.88% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GDP, 7MG

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	0.34	0/5919	0.52	0/8041
1	B	0.34	0/5916	0.53	0/8037
2	X	0.36	0/1234	0.61	0/1646
2	Y	0.36	0/1222	0.61	0/1630
All	All	0.34	0/14291	0.54	0/19354

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	5765	0	5737	93	0
1	B	5762	0	5736	62	0
2	X	1214	0	1174	32	0
2	Y	1202	0	1160	24	0
3	X	28	0	12	1	0
3	Y	28	0	12	1	0
4	X	24	0	16	1	0
4	Y	24	0	16	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	509	0	0	15	0
5	B	538	0	0	9	0
5	X	130	0	0	5	0
5	Y	140	0	0	4	0
All	All	15364	0	13863	205	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 7.

The worst 5 of 205 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:28:GLU:HB3	1:B:67:LYS:HE2	1.52	0.92
2:X:136:ASP:HA	2:X:139:ARG:HG3	1.58	0.85
1:A:503:CYS:HB3	5:A:2375:HOH:O	1.77	0.84
1:A:557:LYS:HD3	1:A:561:HIS:NE2	2.01	0.76
1:B:425:GLN:HG2	5:Y:2102:HOH:O	1.85	0.76

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	702/723 (97%)	679 (97%)	21 (3%)	2 (0%)	50	68
1	B	702/723 (97%)	685 (98%)	17 (2%)	0	100	100
2	X	147/156 (94%)	142 (97%)	4 (3%)	1 (1%)	30	43
2	Y	145/156 (93%)	142 (98%)	3 (2%)	0	100	100
All	All	1696/1758 (96%)	1648 (97%)	45 (3%)	3 (0%)	56	74

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	X	151	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	491	SER
1	A	363	PRO

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	647/662 (98%)	633 (98%)	14 (2%)	64	83
1	B	647/662 (98%)	630 (97%)	17 (3%)	59	79
2	X	125/130 (96%)	121 (97%)	4 (3%)	51	72
2	Y	124/130 (95%)	118 (95%)	6 (5%)	35	53
All	All	1543/1584 (97%)	1502 (97%)	41 (3%)	57	78

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	279	ASP
1	B	387	LEU
2	Y	57	TYR
1	B	299	GLU
1	B	342	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	49	ASN
1	B	425	GLN
2	X	29	ASN
1	B	157	GLN
1	A	553	HIS

5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

4 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	$\# Z > 2$
3	GDP	X	1153	4	30,30,30	0.81	0	44,47,47	4.39	4 (9%)
4	7MG	X	1154	3	24,26,27	2.55	3 (12%)	34,39,42	1.98	6 (17%)
3	GDP	Y	1152	4	30,30,30	0.80	0	44,47,47	4.38	4 (9%)
4	7MG	Y	1153	3	24,26,27	2.56	3 (12%)	34,39,42	1.97	6 (17%)

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
3	GDP	X	1153	4	-	0/16/32/32	0/1/3/3
4	7MG	X	1154	3	-	0/8/37/38	0/1/3/3
3	GDP	Y	1152	4	-	0/16/32/32	0/1/3/3
4	7MG	Y	1153	3	-	0/8/37/38	0/1/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	X	1154	7MG	C8-N9	-10.94	1.37	1.46
4	Y	1153	7MG	C8-N9	-10.92	1.37	1.46

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	X	1154	7MG	P-OP1	4.26	1.51	1.46
4	Y	1153	7MG	P-OP1	4.22	1.51	1.46
4	Y	1153	7MG	C8-N7	-3.02	1.37	1.45

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	X	1153	GDP	C6-C5-N7	-28.29	130.33	134.14
3	Y	1152	GDP	C6-C5-N7	-28.23	130.34	134.14
4	Y	1153	7MG	N7-C8-N9	7.27	112.69	103.08
4	X	1154	7MG	N7-C8-N9	7.24	112.66	103.08
4	X	1154	7MG	C4-C5-N7	4.13	111.58	106.82

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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	706/723 (97%)	-0.50	11 (1%) 68 67	18, 35, 62, 91	0
1	B	706/723 (97%)	-0.57	6 (0%) 83 82	18, 34, 60, 80	0
2	X	149/156 (95%)	-0.33	2 (1%) 74 73	20, 37, 70, 75	0
2	Y	147/156 (94%)	-0.63	1 (0%) 84 84	17, 29, 53, 68	0
All	All	1708/1758 (97%)	-0.52	20 (1%) 75 75	17, 34, 61, 91	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	527	PRO	3.6
1	A	41	LYS	3.5
1	B	262	GLU	3.4
1	A	526	ASN	3.2
1	A	490	GLU	3.0

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

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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	RSR	LLDF	B-factors(\AA^2)	Q<0.9
3	GDP	Y	1152	28/28	0.16	-0.16	30,73,79,82	0
3	GDP	X	1153	28/28	0.11	-0.46	39,50,59,65	0
4	7MG	Y	1153	24/25	0.09	-0.71	17,25,28,35	0
4	7MG	X	1154	24/25	0.07	-1.74	17,27,41,46	0

6.5 Other polymers ⓘ

There are no such residues in this entry.