WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 10:07 am BST

PDB ID	:	1CGP
Title	:	CATABOLITE GENE ACTIVATOR PROTEIN (CAP)/DNA COMPLEX + ADENOSINE-3',5'-CYCLIC-MONOPHOSPHATE
Authors	:	Schultz, S.C.; Shields, G.C.; Steitz, T.A.
Deposited on	:	1991-08-12
Resolution	:	3.00 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

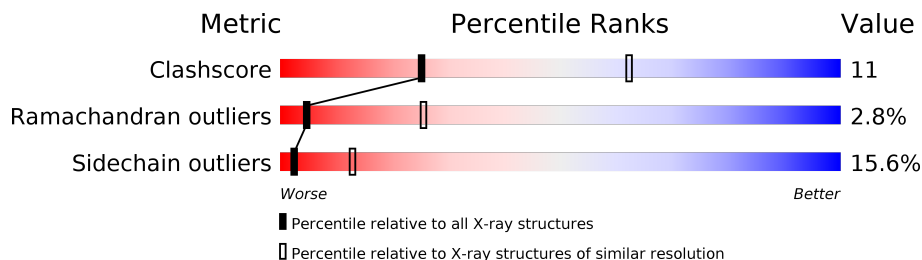|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.00 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
Clashscore		11
Ramachandran outliers		2.8%
Sidechain outliers		15.6%

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))
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)

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 respectively. 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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	18	61% · 33% · 6%
1	E	18	78% · 22%
2	D	13	54% · 46%
2	F	13	77% · 15% · 8%
3	A	205	58% · 30% · 7% · ·
3	B	205	57% · 31% · 7% · ·

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 5274 atoms, of which 856 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 (5'-D(*GP*CP*GP*AP*AP*AP*AP*GP*TP*GP* TP*GP*AP*CP*AP*TP*AP*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	C	18	Total	C	H	N	O	P	0	0	0
			411	178	39	74	103	17			
1	E	18	Total	C	H	N	O	P	0	0	0
			411	178	39	74	103	17			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	D	13	Total	C	H	N	O	P	0	0	0
			284	126	25	42	79	12			
2	F	13	Total	C	H	N	O	P	0	0	0
			284	126	25	42	79	12			

- Molecule 3 is a protein called PROTEIN (CATABOLITE GENE ACTIVATOR PROTEIN (CAP)).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	A	197	Total	C	H	N	O	S	0	0	0
			1917	986	361	273	288	9			
3	B	197	Total	C	H	N	O	S	0	0	0
			1917	986	361	273	288	9			

- Molecule 4 is ADENOSINE-3',5'-CYCLIC-MONOPHOSPHATE (three-letter code: CMP) (formula: $C_{10}H_{12}N_5O_6P$).

CMP
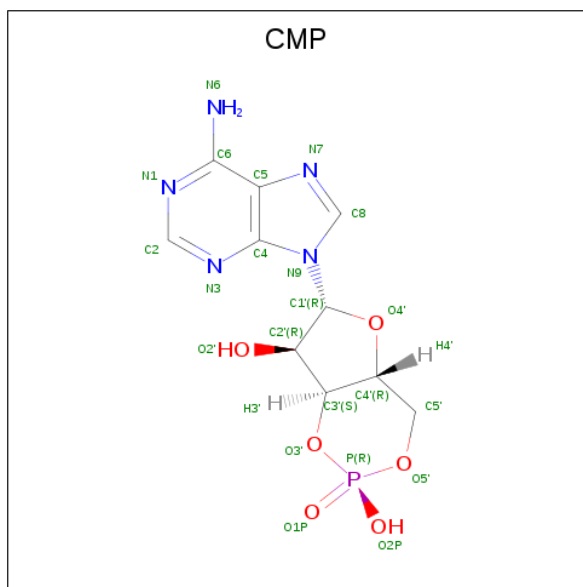
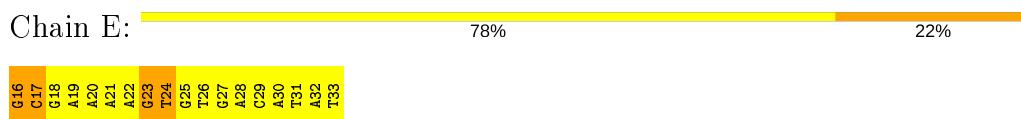| Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
4	A	1	Total	C	H	N	O	P	0	0
			25	10	3	5	6	1		
4	B	1	Total	C	H	N	O	P	0	0
			25	10	3	5	6	1		

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. 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. 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.

Note EDS was not executed.

• Molecule 1: DNA (5'-D(*GP*CP*GP*AP*AP*AP*AP*GP*TP*GP*TP*GP*AP*CP*AP*TP*AP*T)-3')

Chain C:
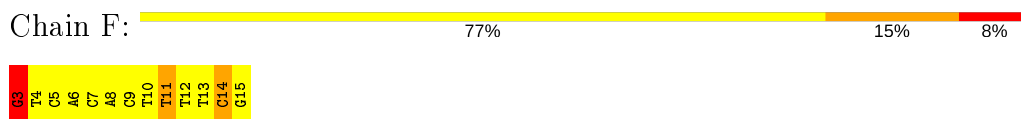

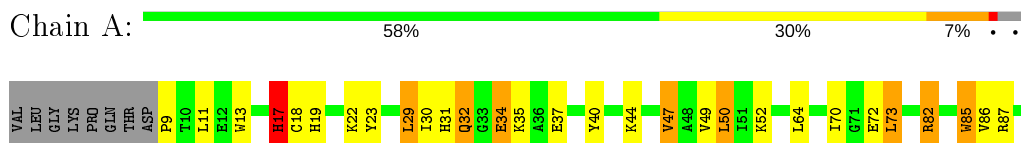• Molecule 1: DNA (5'-D(*GP*CP*GP*AP*AP*AP*AP*GP*TP*GP*TP*GP*AP*CP*AP*TP*AP*T)-3')

Chain E:

• Molecule 2: DNA (5'-D(*GP*TP*CP*AP*CP*AP*CP*TP*TP*TP*TP*CP*G)-3')

Chain D:

• Molecule 2: DNA (5'-D(*GP*TP*CP*AP*CP*AP*CP*TP*TP*TP*TP*CP*G)-3')

Chain F:

• Molecule 3: PROTEIN (CATABOLITE GENE ACTIVATOR PROTEIN (CAP))

Chain A:

L195 H199 G200 K201 T202 I203 V204 V205

● Molecule 3: PROTEIN (CATABOLITE GENE ACTIVATOR PROTEIN (CAP))

Chain B:

57%　　31%　　7%

VAL LEU GLY LYS PRO GLN THR ASP P9 T10 L11 E12 W13 C18 H19 K22 K26 L29 I30 H31 Q32 G33 E34 K35 A36 E37 T38 L39 Y40 G45 S46 V47 A48 V49 L50 E54 E55 Y63 G71 E72 L73 G74 L75 E78 G79 Q80 E81 R82 W85 V86

R87 A88 K89 T90 A91 C92 E93 V94 I97 K100 K101 F102 R103 Q104 L105 I106 N109 P110 D111 I112 L113 M114 Q119 R122 R123 L124 Q125 V139 R142 I143 L148 Q153 P154 D155 A156 M157 D161 T168 R169 Q170 E171 V176 S179 R180 R185 I186

L187 L195 I196 S197 A198 H199 I203 V204 V205

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	138.00Å 152.60Å 76.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 3.00	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-3.00)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	X-PLOR	Depositor
R, R_{free}	0.235 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	5274	wwPDB-VP
Average B, all atoms (Å²)	55.0	wwPDB-VP

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: CMP

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	1.70	5/419 (1.2%)	4.04	72/646 (11.1%)
1	E	1.53	3/419 (0.7%)	4.32	70/646 (10.8%)
2	D	1.39	1/288 (0.3%)	4.42	54/442 (12.2%)
2	F	1.61	2/288 (0.7%)	3.78	53/442 (12.0%)
3	A	0.81	1/1580 (0.1%)	1.57	21/2127 (1.0%)
3	B	0.82	0/1580	1.59	25/2127 (1.2%)
All	All	1.11	12/4574 (0.3%)	2.74	295/6430 (4.6%)

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
2	D	0	1
2	F	0	2
3	A	0	1
All	All	0	5

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	11	DT	C5-C7	7.91	1.54	1.50
1	C	23	DG	C4'-C3'	-5.93	1.46	1.52
2	F	11	DT	C4'-O4'	-5.80	1.39	1.45
1	E	20	DA	C5'-C4'	5.59	1.57	1.51
1	E	28	DA	P-O5'	5.29	1.65	1.59

The worst 5 of 295 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	12	DT	O3'-P-O5'	30.54	162.02	104.00
1	E	24	DT	OP2-P-O3'	30.30	171.86	105.20
1	E	19	DA	O3'-P-O5'	28.85	158.82	104.00
2	D	9	DC	O3'-P-O5'	28.29	157.74	104.00
2	D	4	DT	OP2-P-O3'	27.14	164.91	105.20

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	159	HIS	Peptide
1	C	32	DA	Sidechain
2	D	3	DG	Sidechain
2	F	11	DT	Sidechain
2	F	3	DG	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	C	372	39	204	6	1
1	E	372	39	204	3	0
2	D	259	25	150	3	0
2	F	259	25	150	2	1
3	A	1556	361	1600	41	0
3	B	1556	361	1600	42	0
4	A	22	3	11	3	0
4	B	22	3	11	3	0
All	All	4418	856	3930	88	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 11.

The worst 5 of 88 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:A:206:CMP:H2	4:A:206:CMP:C2	0.97	1.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:206:CMP:C2	4:B:206:CMP:H2	0.97	1.48
3:A:17:HIS:HB2	3:A:101:LYS:HZ3	1.40	0.86
3:B:30:ILE:HG23	3:B:82:ARG:HD3	1.63	0.79
3:A:30:ILE:HG23	3:A:82:ARG:HD3	1.69	0.75

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)
1:C:16:DG:H1	2:F:15:DG:O6[3_554]	1.50	0.10

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	195/205 (95%)	176 (90%)	15 (8%)	4 (2%)	7	33
3	B	195/205 (95%)	184 (94%)	4 (2%)	7 (4%)	3	19
All	All	390/410 (95%)	360 (92%)	19 (5%)	11 (3%)	5	25

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	155	ASP
3	A	160	PRO
3	B	55	GLU
3	B	75	LEU
3	B	72	GLU

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	170/177 (96%)	144 (85%)	26 (15%)	2	13
3	B	170/177 (96%)	143 (84%)	27 (16%)	2	12
All	All	340/354 (96%)	287 (84%)	53 (16%)	2	13

5 of 53 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	153	GLN
3	B	29	LEU
3	B	155	ASP
3	A	189	MET
3	B	10	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	65	ASN
3	B	153	GLN
3	B	80	GLN
3	A	174	GLN
3	B	125	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)

2 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	#$|Z| > 2$
4	CMP	A	206	-	22,25,25	1.32	3 (13%)	24,39,39	1.73	4 (16%)
4	CMP	B	206	-	22,25,25	1.51	4 (18%)	24,39,39	1.94	8 (33%)

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	CMP	A	206	-	-	0/0/31/31	0/4/4/4
4	CMP	B	206	-	-	0/0/31/31	0/4/4/4

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	206	CMP	P-O5'	3.71	1.61	1.57
4	A	206	CMP	C5'-C4'	-2.77	1.47	1.51
4	A	206	CMP	O5'-C5'	-2.73	1.42	1.46
4	B	206	CMP	C8-N7	-2.62	1.30	1.34
4	B	206	CMP	O3'-C3'	-2.42	1.40	1.44

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	206	CMP	O5'-P-O3'	5.20	112.84	105.68
4	A	206	CMP	O3'-C3'-C4'	-4.30	107.46	110.71

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	206	CMP	O2P-P-O1P	3.60	119.99	108.73
4	A	206	CMP	O5'-P-O3'	-3.59	100.73	105.68
4	B	206	CMP	O2P-P-O1P	3.30	119.05	108.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	206	CMP	3	0
4	B	206	CMP	3	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand CMP A 206

Bond lengths   Bond angles

Torsions   Rings

Ligand CMP B 206

Bond lengths

Bond angles

Torsions

Rings
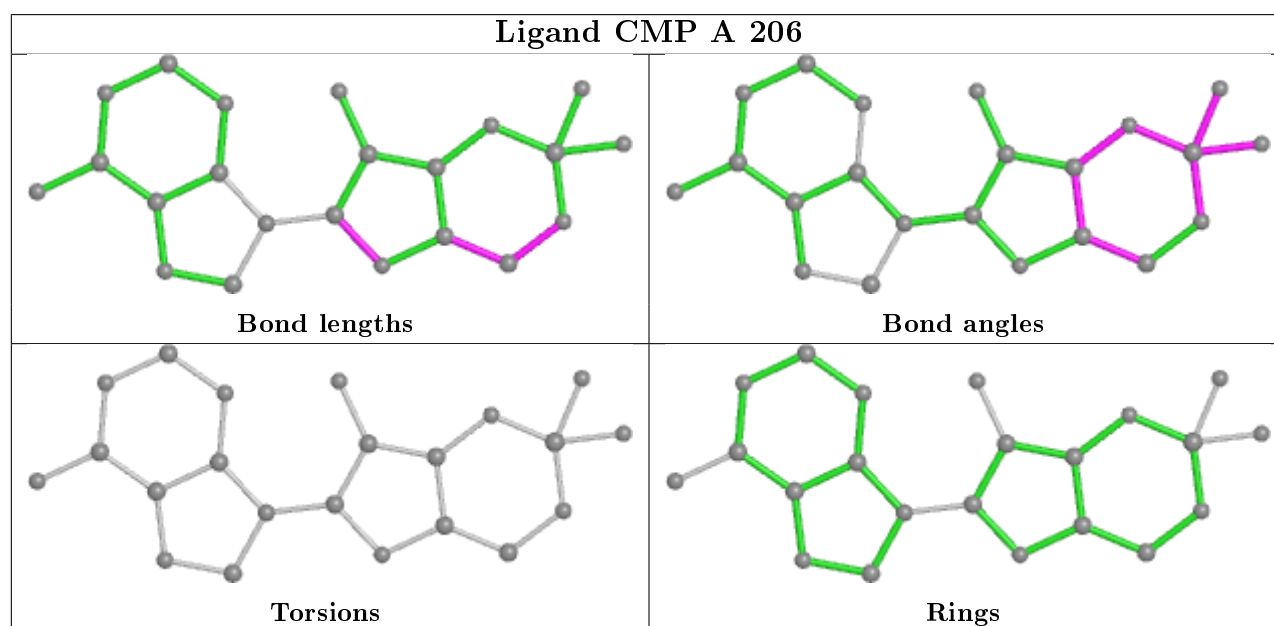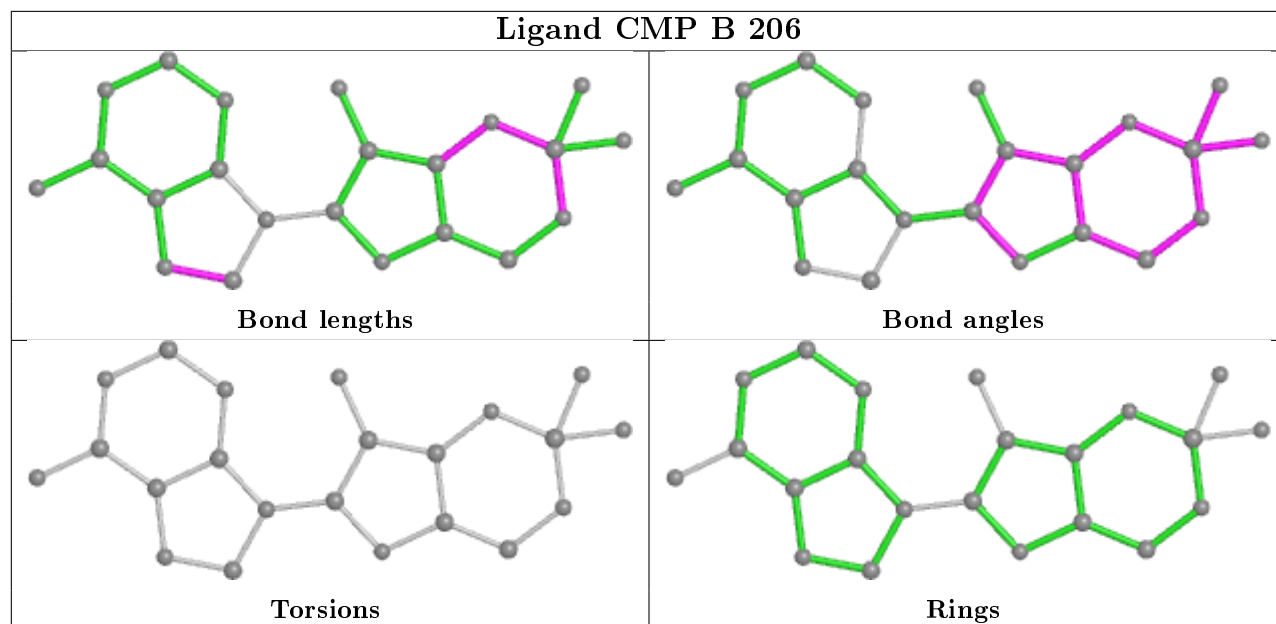
5.7 Other polymers (i)

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)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.