Conformational Energies and Equilibria of Cyclic Dinucleotides *In Vacuo* and In Solution: Computational Chemistry vs. NMR Experiments

Supplementary Information

Ondrej Gutten,[1,*] *Petr Jurečka,*[2] *Zahra Aliakbar Tehrani,*[1,#] *Miloš Budešínský,*[1] *Jan Řezáč,*[1] *Lubomír Rulíšek*[1,*]

[1] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo náměstí 2, 166 10, Praha 6, Czech Republic

[2] Department of Physical Chemistry, Faculty of Science, Palacky University Olomouc, 17. listopadu 12, 771 46, Olomouc Czech Republic

[#] Currently at Institute of Biotechnology of the Czech Academy of Sciences, BIOCEV, Průmyslová 595, 252 50 Vestec, Czech Republic

* Corresponding Authors: gutten@uochb.cas.cz; rulisek@uochb.cas.cz

Table S1: Software used for QM calculations.

Method	Software	Version	Integration grid for single-point calculations
B-LYP	TurboMole	7.2	m3 (default)
B-P	TurboMole	7.2	m3 (default)
B97-D	TurboMole	7.2	m3 (default)
OLYP	ORCA	4.0.1.2	Grid 2, FinalGrid 4, (default)
PBE	TurboMole	7.2	m3 (default)
revPBE	ORCA	4.0.1.2	Grid 2, FinalGrid 4, (default)
M06-L	TurboMole	7.2	m3 (default)
revTPSS	ORCA	4.0.1.2	Grid 2, FinalGrid 4, (default)
SCAN	ORCA	4.0.1.2	Grid 5, IntAcc 5.5
TPSS	TurboMole	7.2	m3 (default)
B3-LYP	TurboMole	7.2	m3 (default)
BH-LYP	TurboMole	7.2	m3 (default)
M06	TurboMole	7.2	m3 (default)
M06-2X	TurboMole	7.2	m3 (default)
PBE0	TurboMole	7.2	m3 (default)
PW6B95	TurboMole	7.2	m3 (default)
TPSSH	TurboMole	7.2	m3 (default)
ωB-97X	ORCA	4.0.1.2	Grid 2, FinalGrid 4, (default)
B2PLYP	ORCA	4.1.0	Grid 5, FinalGrid 6
PWPB95	ORCA	4.1.0	Grid 5, FinalGrid 6
CCSD(T)	TurboMole	7.2	
DLPNO-CCSD(T)	ORCA	4.0.1.2	
MP2	TurboMole	7.2	
MP2-F12	TurboMole	7.2	
HF-3c	ORCA	4.0.1.2	
B97-3c	ORCA	4.0.1.2	Grid 3 (default)
PBEh-3c	ORCA	4.0.1.2	Grid 3 (default)

Table S2: Averaged MUE values (see **Equation 4**) of several DFT functionals obtained obtained with different integration grids for the CDN set. Triple-ζ (def2-TZVPD) basis set was used in all cases. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol-1.

Functional	Integration Grid	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	m3	6.249	0.675	0.761	0.748
B-LYP	m5	6.241	0.676	0.759	0.745
B-P	m3	5.341	1.038	1.026	0.911
B-P	m5	5.331	1.043	1.028	0.910
B3-LYP	m3	5.254	0.480	0.592	0.585
B3-LYP	m5	5.248	0.477	0.591	0.586
SCAN	Grid 4, IntAcc 4	1.533	0.459	0.432	0.464
SCAN	Grid 5, IntAcc 5.5	1.527	0.466	0.445	0.478
SCAN	Grid 6, IntAcc 6	1.526	0.464	0.446	0.476

S1 Estimation of error of DLPNO-CCSD(T) TightPNO/MP2F12 with respect to CCSD(T)/MP2-CBS

Using CBS-extrapolated CCSD(T) energies, see Equation S1, is prohibitive for systems of CDN size (around 70 atoms).

$$E_{CBS} = \frac{4^3 * E^{MP2-F12/aug-cc-pVQZ} - 3^3 * E^{MP2-F12/aug-cc-pVTZ}}{4^3 - 3^3} \quad \text{(S1)}$$

$$E_{hoc} = E^{CCSD(T)/aug-cc-pVDZ} - E^{MP2/aug-cc-pVDZ}$$

In the main text we use an alternative reference scheme, DLPNO-CCSD(T) TightPNO/MP2-F12, accuracy of which we try to justify by comparing the error of different schemes for the 20 conformers of 3'3'-c-GAMP from the CDN dataset (see **Methods** section) and their fragments, see **Figure S1**. This allows for estimation of the error on the unfragmented original system.

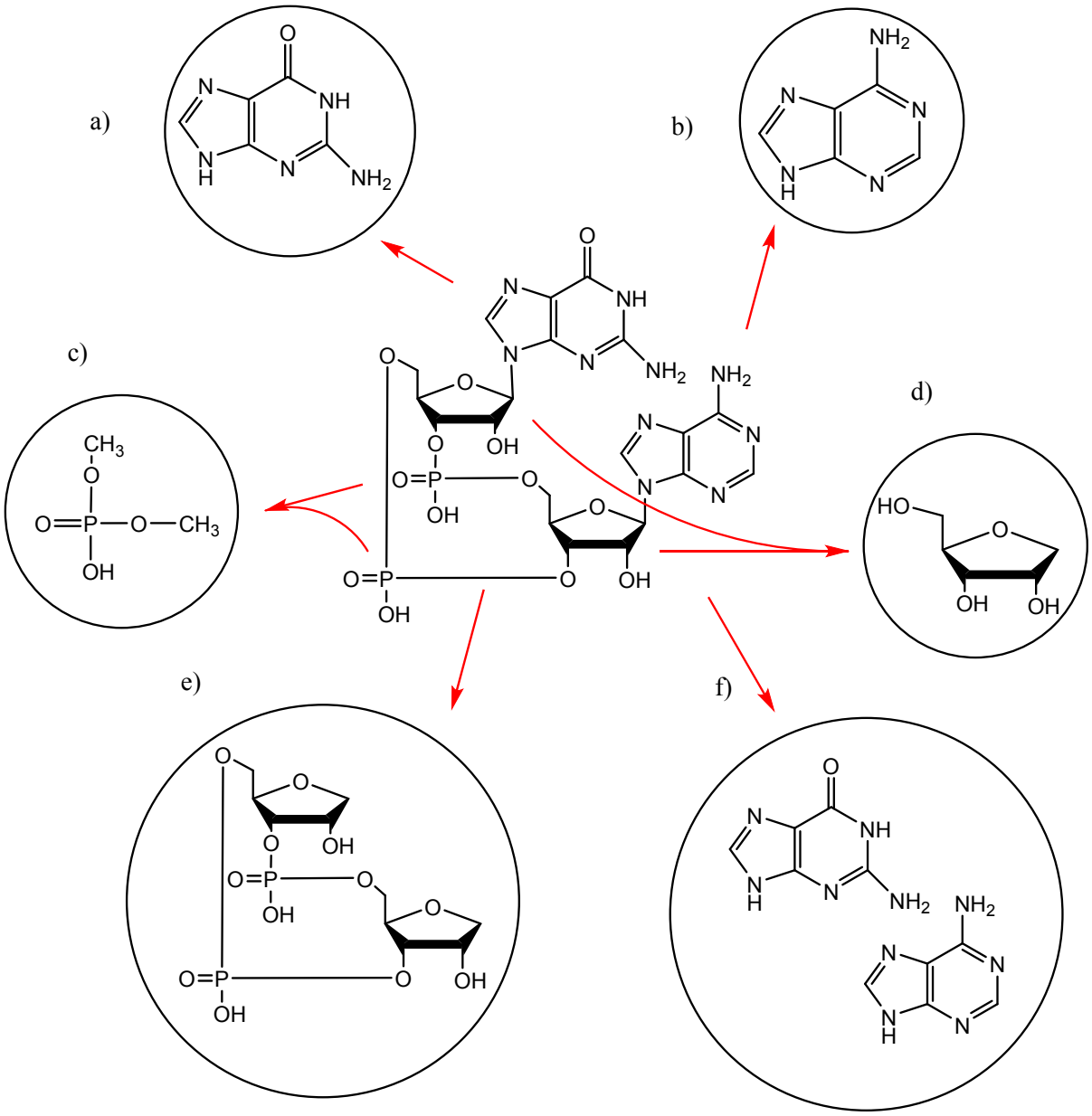


Figure S1: 16 model systems are created from each of the 20 conformers of 3'3'-c-GAMP (in the center). There are two base fragments, i.e. guanine (a) and adenine (b), and four macrocycle fragments, i.e. two dimethylphosphates (c) and two sugar moieties (d). Each of the base fragments (a) and (b) is individually combined with each of the macrocycle fragments (c) and (d), yielding 2 x 4 = 8 additional model systems. The whole macrocycle without bases (e) is considered as a separate model system, as well as both bases (f).

Using the fragments shown in **Figure S1**, the estimate may be obtained in a spirit of pairwise energy decomposition, i.e. by summing up energies of model systems made of two basic

fragments, such as **Figure S1**/f, while subtracting energies of the individual basic fragments (**Figure S1**/a-e) in a way that avoids double-counting of any part of the CDN. Specifically, the estimate is calculated as:

$$E_{estimate} = E(macrocycle) + E(bases) + E(macrocycle + bases)$$

Where E(macrocycle) is energy of a macrocycle fragment (**Figure S1**/e), E(bases) is energy of the two purine bases (**Figure S1**/f) and E(macrocycle+bases) is a sum of pairwise interaction energies between a combined model system of a base and a macrocycle fragment (e.g. guanine plus one of the phosphate fragments) and its individual basic fragments (i.e. guanine **Figure S1**/a and the phosphate **Figure S1**/c).

The errors for different schemes obtained using this estimate and errors for the whole molecule are shown in **Table S3**. Note that DLPNO-CCSD(T) TightPNO/MP2-F12, for which conformational energies of whole molecules can be calculated, is used as a reference in this case. The errors on the fragmented dataset and whole molecules in **Table S3** show correlation of ca. 0.8. We deem this correlation to be sufficiently high for the following qualitative argument. On the fragmented dataset, DLPNO-CCSD(T) TightPNO/MP2-F12 scheme exhibits similar errors with respect to CCSD(T)/MP2-CBS as DLPNO-CCSD(T) NormalPNO/MP2-CBS exhibits with respect to DLPNO-CCSD(T) TightPNO/MP2-F12. Since the mutual MUE of the latter pair for the whole systems is around 0.2 kcal.mol^{-1}, we speculate that error of comparable magnitude is to be expected for DLPNO-CCSD(T) TightPNO/MP2-F12 with respect to CCSD(T)/MP2-CBS.

Table S3: Estimates of mean unsigned error based on fragmentation scheme (left column) and true mean unsigned errors obtained for whole molecules (right column). DLPNO-CCSD(T) TightPNO/MP2-F12 is taken as reference. The missing values are not available due to high computational cost of CCSD(T) on whole CDN systems, but can be estimated.

Method	MUE Estimate [kcal/mol]	True MUE [kcal/mol]
CCSD(T)/MP2-F12	0.3	-
DLPNO-CCSD(T) NormalPNO/MP2-CBS	0.4	0.2
CCSD(T)/MP2-CBS	0.4	-
MP2/CBS	1.2	1.4
SCS-MP2/CBS	1.1	0.6
SOS-MP2/CBS	2.1	1.4
MP2F12	1.3	1.4
MP2-SCS-F12	1.2	0.6
MP2-SOS-F12	2.2	1.4
BP+D3(0)	1.2	1.3
B3LYP+D3(0)	0.4	0.4

Table S4: The ^{1}H and ^{31}P NMR data of 3‘,3‘-c-diAMP, 3‘,3‘-c-GAMP, 3‘,3‘-c-diGMP in D_2O.

Compound	Res.	H-1’	H-2’	H-3”	H-4’	H-5’a	H-5’b	Base protons	^{31}P
3‘3‘-c-di-AMP	A	6.06 s 1’,2’ < 1	4.87 bd 2’,1’ < 1 2’,3’ = 4.4	5.02 ddd 3’,2’ = 4.4 3’,4’ = 9.0 3’,P = 7.4	4.52 dm 4’,3’ = 9.0 4’,5’a = 2.3 4,5’b = 1.4 4’,P = 3.8	4.49 bdd 5a’,4’ = 2.3 5’a,5’b = 11.6 5’a,P < 1	4.14 ddd 5’b,4 = 1.4 5’b,5’a = 11.6 5’b,P = 4.0	H-2: 7.94 bs H-8: 8.32 bs	-0.56
3‘3‘-c-GAMP	G	5.83 bs 1’,2’ < 1	4.85 bd 2’,1’ < 1 2’,3’ = 4.7	5.055 ddd 3’,2’ = 4.7 3’,4’ = 8.9 3’,P’ = 7.3	4.43 ddt 4’,3’ = 8.9 4’,5’a = 2.3 4’,5’b = 1.4 4’,P = 3.7	4.40 dt 5’a,4’ = 2.3 5’a,5’b = 12.0 5’a,P = 1.8	4.08 ddd 5’b,4’ = 1.4 5’b,5’a = 12.0 5’b,P = 3.9	H-8: 7.84 s	-0.33
	A	6.01 d 1’,2’ < 1	4.865 bd 2’,1’ < 1 2’,3’ = 4.9	5.04 td 3’,2’ = 4.9 3’,4’ = 8.9 3’,P = 7.2	4.49 dm 4’,3’ = 8.9 4’,5’a = 2.2 4’,5’b = 1.7 4’,P = 3.7	4.44 dt 5’a,4’ = 2.2 5’a,5’b = 12.0 5’a,P = 2.2	4.115 ddd 5’b,4’ = 1.7 5’b,5’a = 12.0 5’b,P = 3.9	H-2: 7.99 s H-8: 8.28 s	-0.49
3‘3‘-c-di-GMP	G	5.97 s 1’,2’ < 1	4.74 bd 2’,1’ < 1 2’,3’ = 4.9	4.90 td 3’,2’ = 4.9 3’,4’ =8.5 3’,P = 8.5	4.40 dm 4’,3’ = 8.5 4’,5’a = 2.1 4’,5’b = 1.8 4’,P = 3.3	4.35 dt 5’a,4’ = 2.1 5’a,5’b = 12.1 5’a,P = 2.1	4.10 ddd 5’b,4 = 1.8 5’b,5’a = 12.1 5’b,P = 3.9	H-8: 8.01 s	-0.18

Table S5: The ^{13}C NMR data of 3‘,3‘-c-diAMP, 3‘,3‘-c-GAMP, 3‘,3‘-c-diGMP in D_2O. Values of observed *J*(C,P) are given in parentheses.

Compound	Res.	C-1’	C-2’	C-3’	C-4’	C-5’	C-2	C-4	C-5	C-6	C-8
3‘3‘-c-di-AMP	A	93.29	76.22 (<1)	72.99 (4.8)	83.02 (11.0; 11.0)	65.15 (4.7)	155.10	149.44	120.92	157.44	141.39
3‘3‘-c-GAMP	G	93.09	75.92 (<1)	73.07 (4.7)	83.04 (10.8; 10.8)	65.06 (4.7)	156.25	152.33	118.90	160.63	139.33
	A	93.65	75.91 (<1)	73.14 (4.9)	83.48 (10.6; 10.6)	65.21 (4.6)	155.23	149.84	121.06	157.49	141.89
3‘3‘-c-di-GMP	G	92.13	76.08 (<1)	73.23 (4.8)	82.50 (10.7; 10.1)	65.04 (5.2)	156.52	153.37	119.11	161.38	139.79

Table S6: MAD values (see **Equation 5**) of several DFT functionals obtained for the CDN set with triple-ζ (def2-TZVPD) basis set and several dispersion corrections. Entries marked as 'n.a.' indicate combinations that are not available. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol^{-1}.

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	GGA	15.7	2.1	2.2	2.5
B-P	GGA	13.2	3.0	3.1	2.9
B97-D	GGA	17.8	2.8	2.3	n.a.
OLYP	GGA	23.3	3.8	3.5	3.1
PBE	GGA	10.9	3.6	3.6	3.2
revPBE	GGA	17.2	2.7	2.9	2.7
M06-L	meta-GGA	2.7	3.7	n.a.	n.a.
revTPSS	meta-GGA	9.0	n.a.	n.a.	2.4
SCAN	meta-GGA	4.0	2.1	2.0	2.1
TPSS	meta-GGA	13.1	2.6	2.8	2.7
B3-LYP	hybrid	13.9	1.6	1.9	2.2
BH-LYP	hybrid	12.0	2.3	2.6	2.6
M06	hybrid	2.0	4.2	n.a.	n.a.
M06-2X	hybrid	2.3	2.8	n.a.	n.a.
PBE0	hybrid	10.7	2.9	2.9	2.6
PW6B95	hybrid	6.6	2.3	2.1	2.5
TPSSH	hybrid	12.8	2.4	2.3	2.5
ωB-97X	hybrid	4.5	5.0	6.0	n.a.
B2PLYP	double-hybrid	4.9	5.7	5.6	5.7
PWPB95	double-hybrid	5.1	6.4	6.3	6.1

Table S7: RMSEmax values (see **Equation 6**) of several DFT functionals obtained for the CDN set with triple-ζ (def2-TZVPD) basis set and several dispersion corrections. Entries marked as 'n.a.' indicate combinations that are not available. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol^{-1}.

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	GGA	8.6	1.0	1.1	1.2
B-P	GGA	7.3	1.6	1.6	1.4
B97-D	GGA	9.5	1.4	1.2	n.a.
OLYP	GGA	12.8	1.8	1.6	1.5
PBE	GGA	5.9	1.4	1.6	1.5
revPBE	GGA	9.5	1.4	1.4	1.3
M06-L	meta-GGA	0.9	1.7	n.a.	n.a.
revTPSS	meta-GGA	5.2	n.a.	n.a.	1.2
SCAN	meta-GGA	2.1	0.9	0.8	0.8
TPSS	meta-GGA	7.1	1.2	1.4	1.2
B3-LYP	hybrid	7.3	0.7	1.0	1.1
BH-LYP	hybrid	5.9	0.9	1.0	1.1
M06	hybrid	0.9	2.3	n.a.	n.a.
M06-2X	hybrid	0.9	1.0	n.a.	n.a.
PBE0	hybrid	5.5	1.0	1.1	1.2
PW6B95	hybrid	3.9	0.9	0.8	1.0
TPSSH	hybrid	6.9	1.1	1.1	1.1
ωB-97X	hybrid	2.3	1.6	1.7	n.a.
B2PLYP	double-hybrid	2.3	2.5	2.5	2.9
PWPB95	double-hybrid	1.7	2.7	2.3	2.1

Table S8: MSE values (see **Equation 7**) of several DFT functionals obtained for the subset of "closed" conformers (|s| = 74 conformers) of the **CDN** set with triple-ζ (def2-TZVPD) basis set and several dispersion corrections. Entries marked as 'n.a.' indicate combinations that are not available. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol^{-1}.

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	GGA	7.0	-0.2	0.0	0.1
B-P	GGA	6.0	-0.8	-0.5	-0.3
B97-D	GGA	7.7	0.8	0.3	n.a.
OLYP	GGA	10.1	1.2	0.8	1.2
PBE	GGA	4.9	0.8	0.9	0.9
revPBE	GGA	7.7	0.6	0.4	0.7
M06-L	meta-GGA	-0.3	-1.0	n.a.	n.a.
revTPSS	meta-GGA	4.3	n.a.	n.a.	-0.3
SCAN	meta-GGA	1.7	-0.2	-0.1	0.2
TPSS	meta-GGA	5.9	0.6	0.7	0.5
B3-LYP	hybrid	5.8	0.0	-0.1	0.0
BH-LYP	hybrid	4.6	0.0	0.0	-0.1
M06	hybrid	0.1	-1.8	n.a.	n.a.
M06-2X	hybrid	0.1	-0.5	n.a.	n.a.
PBE0	hybrid	4.4	0.4	0.4	0.5
PW6B95	hybrid	3.0	0.3	0.2	0.5
TPSSH	hybrid	5.6	0.5	0.5	0.5
ωB-97X	hybrid	1.1	-0.6	-0.2	n.a.
B2PLYP	double-hybrid	1.2	-1.7	-1.8	-2.1
PWPB95	double-hybrid	0.6	-1.7	-1.5	-1.2

Table S9: MUE values (see **Equation 4**) of several DFT functionals obtained for the **CDN** set with triple-ζ (def2-TZVP) basis set and several dispersion corrections. DLPNO-CCSD(T)/MP2-F12 is used as reference, see Equation 2. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	GGA	6.0	0.8	0.8	0.8
B-P	GGA	5.2	1.2	1.1	1.0
SCAN	meta-GGA	1.4	0.6	0.6	0.5
B2PLYP	Double-hybrid	2.3	0.5	0.6	0.8
PWPB95	Double-hybrid	1.6	0.9	0.6	0.5

Table S10: MAD values (see **Equation 5**) of several DFT functionals obtained for the **CDN** set with triple-ζ (def2-TZVP) basis set and several dispersion corrections. DLPNO-CCSD(T)/MP2-F12 is used as reference, see Equation 2. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	GGA	15.4	2.5	3.0	3.1
B-P	GGA	13.0	3.4	3.7	3.4
SCAN	meta-GGA	3.9	2.3	2.2	2.4
B2PLYP	Double-hybrid	6.2	1.8	2.1	2.7
PWPB95	Double-hybrid	4.0	3.0	2.0	2.0

Table S11: RMSEmax values (see **Equation 6**) of several DFT functionals obtained for the **CDN** set with triple-ζ (def2-TZVP) basis set and several dispersion corrections. DLPNO-CCSD(T)/MP2-F12 is used as reference, see Equation 2. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	GGA	8.4	1.1	1.2	1.3
B-P	GGA	7.1	1.7	1.7	1.4
SCAN	meta-GGA	2.1	1.2	1.1	0.9
B2PLYP	Double-hybrid	3.1	0.8	1.0	1.4
PWPB95	Double-hybrid	2.4	1.4	0.9	0.8

Table S12: MSE values (see **Equation 7**) of several DFT functionals obtained for the subset of "closed" conformers (|s| = 74 conformers) of the **CDN** set with triple-ζ (def2-TZVP) basis set and several dispersion corrections. DLPNO-CCSD(T)/MP2-F12 is used as reference, see Equation 2. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	GGA	6.7	-0.5	-0.3	-0.1
B-P	GGA	5.8	-1.0	-0.7	-0.5
SCAN	meta-GGA	1.5	-0.3	-0.3	0.0
B2PLYP	Double-hybrid	2.5	-0.5	-0.5	-0.9
PWPB95	Double-hybrid	1.7	-0.6	-0.4	-0.1

Table S13: MUE values (see **Equation 4**) of several DFT functionals obtained for the **CDN** set with quadruple-ζ (def2-QZVPD) basis set and several dispersion corrections. DLPNO-CCSD(T)/MP2-F12 is used as reference, see Equation 2. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4
B-LYP	GGA	6.3	0.6	0.7	0.7
B-P	GGA	5.5	1.0	1.0	0.9
B97-D	GGA	7.0	0.9	0.7	n.a.
PBE	GGA	4.5	1.0	1.1	1.0
TPSS	meta-GGA	5.3	0.8	0.9	0.8

Table S14: MUE values (see **Equation 4**) of several DFT functionals obtained for the **CDN** set with double-ζ (DZVP-DFT) basis set and several dispersion corrections. Entries marked as 'n.a.' indicate combinations that are not available. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4	DZVP-DFT-D3(0)	DZVP-DFT-D3(BJ)
B-LYP	GGA	5.4	1.3	1.2	1.1	0.7	0.8
B-P	GGA	4.4	1.8	1.7	1.4	0.9	0.8
B97-D	GGA	6.0	0.8	0.9	n.a.	0.7	0.7
PBE	GGA	3.5	0.8	0.8	0.7	0.7	0.7
M06-L	meta-GGA	1.8	2.4	n.a.	n.a.	n.a.	n.a.
TPSS	meta-GGA	4.2	1.0	0.9	1.0	0.8	0.7
B3-LYP	hybrid	4.3	1.2	1.2	1.2	0.8	0.7
BH-LYP	hybrid	3.0	1.4	1.4	1.5	n.a.	n.a.
M06	hybrid	1.2	2.7	n.a.	n.a.	n.a.	n.a.
M06-2X	hybrid	1.6	2.1	n.a.	n.a.	n.a.	n.a.
PBE0	hybrid	2.8	1.1	1.1	1.0	0.8	0.7
TPSSH	hybrid	1.6	1.3	1.3	1.1	n.a.	n.a.
PW6B95	hybrid	4.2	1.0	0.9	1.0	0.8	0.7

Table S15: MAD values (see **Equation 5**) of several DFT functionals obtained for the **CDN** set with double-ζ (DZVP-DFT) basis set and several dispersion corrections. Entries marked as 'n.a.' indicate combinations that are not available. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4	DZVP-DFT-D3(0)	DZVP-DFT-D3(BJ)
B-LYP	GGA	14.5	4.0	4.3	4.1	2.8	3.0
B-P	GGA	12.0	4.9	5.0	4.8	3.3	3.6
B97-D	GGA	16.5	2.9	3.1	n.a.	2.9	2.7
PBE	GGA	9.7	3.0	3.3	2.9	3.4	3.4
M06-L	meta-GGA	5.0	6.6	n.a.	n.a.	n.a.	n.a.
TPSS	meta-GGA	11.4	3.4	3.3	3.6	3.1	2.8
B3-LYP	hybrid	12.5	3.6	4.0	3.8	2.6	2.5
BH-LYP	hybrid	9.9	3.8	3.9	4.2	n.a.	n.a.
M06	hybrid	3.4	6.9	n.a.	n.a.	n.a.	n.a.
M06-2X	hybrid	5.2	5.7	n.a.	n.a.	n.a.	n.a.
PBE0	hybrid	8.6	3.3	3.3	3.3	2.5	2.4
TPSSH	hybrid	4.6	4.0	3.9	3.8	n.a.	n.a.
PW6B95	hybrid	10.9	3.5	3.5	3.4	n.a.	n.a.

Table S16: RMSEmax values (see **Equation 6**) of several DFT functionals obtained for the **CDN** set with double-ζ (DZVP-DFT) basis set and several dispersion corrections. Entries marked as 'n.a.' indicate combinations that are not available. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4	DZVP-DFT-D3(0)	DZVP-DFT-D3(BJ)
B-LYP	GGA	7.6	1.8	1.7	1.7	1.1	1.3
B-P	GGA	6.2	2.5	2.4	2.2	1.4	1.3
B97-D	GGA	8.4	1.2	1.3	n.a.	1.1	1.2
PBE	GGA	4.8	1.2	1.2	1.2	1.2	1.3
M06-L	meta-GGA	2.5	3.4	n.a.	n.a.	n.a.	n.a.
TPSS	meta-GGA	5.8	1.6	1.6	1.6	1.4	1.4
B3-LYP	hybrid	6.2	1.5	1.6	1.6	1.1	1.2
BH-LYP	hybrid	4.5	1.8	1.9	2.0	n.a.	n.a.
M06	hybrid	1.6	3.6	n.a.	n.a.	n.a.	n.a.
M06-2X	hybrid	2.3	2.7	n.a.	n.a.	n.a.	n.a.
PBE0	hybrid	4.0	1.3	1.4	1.4	1.1	1.1
TPSSH	hybrid	2.4	1.8	1.8	1.6	n.a.	n.a.
PW6B95	hybrid	5.4	1.6	1.7	1.6	n.a.	n.a.

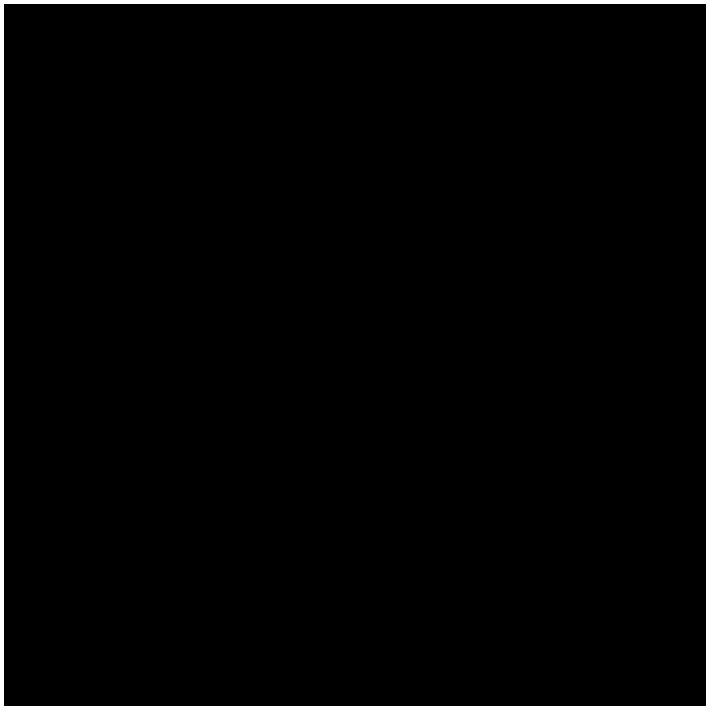
Table S17: MSE values (see **Equation 7**) of several DFT functionals obtained for the subset of "closed" conformers (|s| = 74 conformers) of the the **CDN** set with double-ζ (DZVP-DFT) basis set and several dispersion corrections. Entries marked as 'n.a.' indicate combinations that are not available. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol^{-1}

Functional	Jacob's ladder class	No dispersion correction	D3(0)	D3(BJ)	D4	DZVP-DFT-D3(0)	DZVP-DFT-D3(BJ)
B-LYP	GGA	6.1	-1.1	-0.9	-0.8	-0.2	0.2
B-P	GGA	5.0	-1.8	-1.5	-1.3	-0.4	-0.1
B97-D	GGA	6.7	-0.2	-0.7	n.a.	0.1	0.2
PBE	GGA	3.9	-0.2	-0.1	-0.2	0.0	0.2
M06-L	meta-GGA	-1.8	-2.5	n.a.	n.a.	n.a.	n.a.
TPSS	meta-GGA	4.7	-0.6	-0.5	-0.7	-0.2	0.1
B3-LYP	hybrid	4.8	-1.0	-1.1	-1.0	-0.4	-0.1
BH-LYP	hybrid	3.3	-1.3	-1.3	-1.4	n.a.	n.a.
M06	hybrid	-1.2	-3.0	n.a.	n.a.	n.a.	n.a.
M06-2X	hybrid	-1.5	-2.1	n.a.	n.a.	n.a.	n.a.
PBE0	hybrid	3.1	-0.9	-0.9	-0.9	-0.4	-0.2
TPSSH	hybrid	1.7	-1.1	-1.1	-0.9	n.a.	n.a.
PW6B95	hybrid	4.3	-0.8	-0.8	-0.8	n.a.	n.a.

Table S18: MUE values (see **Equation 4**) of several wave-function methods for the **CDN** set. DLPNO-CCSD(T)/MP2-F12 is used as reference, see **Equation 2**. All values in kcal.mol^{-1}

Functional	Basis set	Value
DLPNO-CCSD(T) NormalPNO/MP2-F12	aug-cc-pVDZ/cc-pVFZ-F12	0.24
HF	aug-cc-pVDZ	5.16
MP2	aug-cc-pVDZ	4.30
MP2-SCS	aug-cc-pVDZ	2.75
MP2-SOS	aug-cc-pVDZ	1.99
HF	aug-cc-pVTZ	5.85
MP2	aug-cc-pVTZ	2.57
MP2-SCS	aug-cc-pVTZ	1.07
MP2-SOS	aug-cc-pVTZ	0.45
DLPNO-CCSD(T) TightPNO	aug-cc-pVDZ	3.07
DLPNO-CCSD(T) NormalPNO	aug-cc-pVDZ	3.10
HF	cc-pVDZ-F12	5.97
MP2F12	cc-pVDZ-F12	1.26
SCS-MP2-F12	cc-pVDZ-F12	0.54
SOS-MP2-F12	cc-pVDZ-F12	1.29

Figure S2: Relative magnitude of coupling constants ^3J(H1′/C8) and ^3J(H1′,C4) between ribose hydrogen H-1’ and base carbon atoms C-4 and C-8 are related to glycoside torsion angle χ, see ref [1]. The relative magnitude of coupling constants ^3J(H-1’,C8) > ^3J(H-1’,C4) indicates *anti*-orientation; ^3J(H-1’,C8) < ^3J(H-1’,C4) indicates *syn*-orientation. Our observed values of ^3J(H-1’,C8) = 2.1 to 3.0 Hz and ^3J(H-1’,C4) < 1 Hz therefore indicate *anti*-orientation.



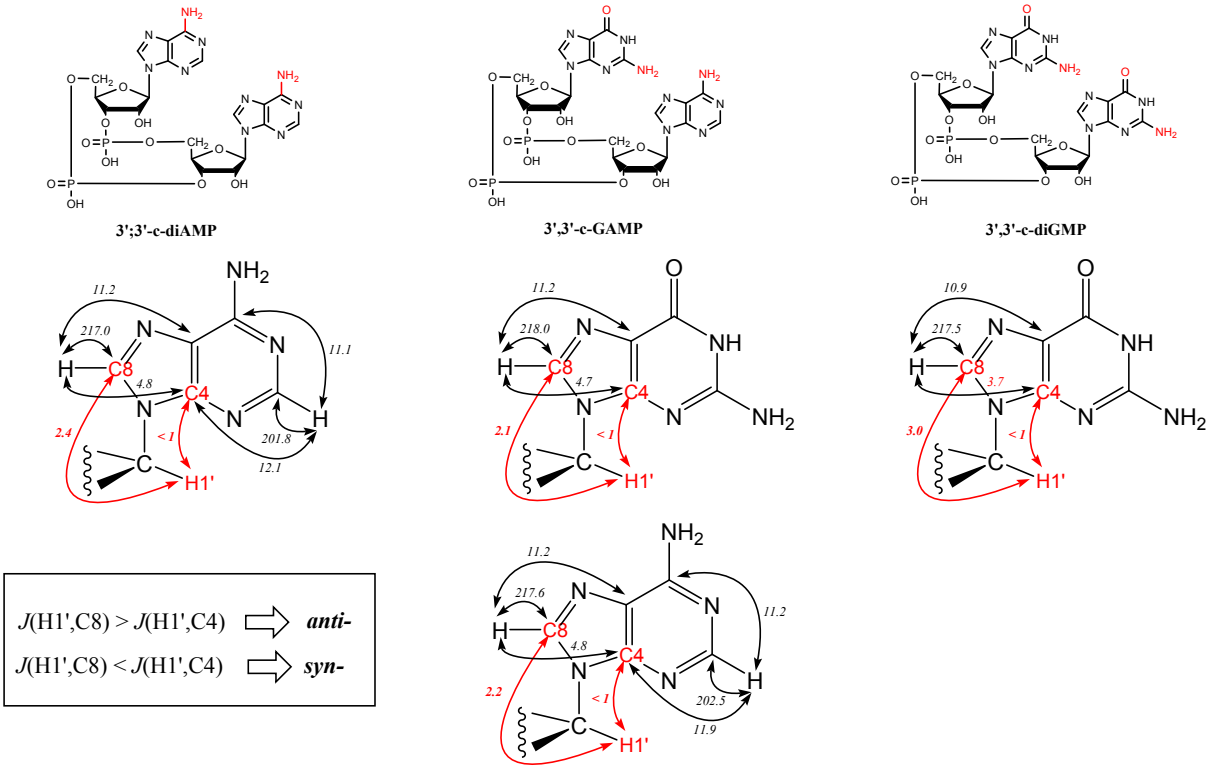


Table S19: Correlation of experimental and calculated chemical shifts. The shifts were calculated using B3-LYP/PCM model with 6-31Gd basis set for the lowest-energy conformers of the ***Syn-/Anti-*** set.

CDN	X-conformer Class	^{13}C RMSD	^{1}H RMSD
3'3'-c-di-AMP	*anti- / anti-*	0.9979	0.9912
	syn- / anti-	0.9972	0.9353
	syn- / syn-	0.9943	0.8606
	anti-anti (57 %) + syn-anti (40 %) + syn-syn (3 %) [a]	0.9980	0.9844
3'3'-c-GAMP	*anti- / anti-*	0.9971	0.9834
	anti- / syn-	0.9963	0.9552
	syn- / anti-	0.9958	0.8658
	syn- / syn-	0.9957	0.8958
	anti-anti (18 %) + syn-anti (61 %) + anti-syn (9 %) + syn-syn (12%) [a]	0.9970	0.9811
3'3'-c-di-GMP	*anti- / anti-*	0.9980	0.9706
	syn- / anti-	0.9955	0.8553
	syn- / syn-	0.9938	0.7666
	anti-anti (8 %) + syn-anti (31 %) + syn-syn (61 %) [a]	0.9968	0.8922

[a] Calculated average values of chemical shifts were obtained for conformation mixture with populations estimated by molecular dynamics.

Table S20: Experimental transient NOEs. Intensities are normalized with respect to the intensity of irradiated inverted peak = -100).

CDN	signal inverted	observed NOEs			
		H8	H1‘	H2‘	H3‘
3’3’-c-di-AMP	H8	-	0.71	0.60	2.43
	H1’	0.41	-	2.44	-0.66
3’3’-c-GAMP	H8 (G)	-	0.79	0.49	0.56
	H1‘ (G)	0.36	-	1.59	0.76
	H8 (A)	-	0.62	0.58	0.70
	H1‘ (A)	0.39	-	1.59	0.76
3’3’-c-diGMP	H8	-	0.83	0.60	0.88
	H1‘	0.58	-	1.24	0.72

Table S21: Experimental 2D ROESY signals. n.d. stands for ‘not determined’. Absolute (abs.) intensities and intensities relative (rel.) to H1’/H2’ signal strength are listed.

Proton pair	3’3’-c-di-AMP		3’3’-c-GAMP				3’3’-c-diGMP	
	A		G		A		G	
	I (abs.)	*I (rel.)*	*I (abs.)*	*I (rel.)*	*I (abs.)*	*I (rel.)*	*I (abs.)*	*I (rel.)*
H8 / H1‘	3.38	0.19	1.82	0.43	1.53	0.30	0.77	0.52
H8 / H2’	2.97	0.17	1.24	0.29	1.50	0.29	0.54	0.36
H8 / H3’	17.39	0.97	2.76	0.65	3.20	0.62	1.46	0.99
H2 / H1’	0.63	0.04	-	-	~0	~0	-	-
H2 / H2’	7.05	0.40	-	-	2.97	0.57	-	-
H2 / H3’	4.09	0.23	-	-	2.36	0.46	-	-
H1’ / H2’	17.84	1.00	4.23	1.00	5.19	1.00	1.47	1.00

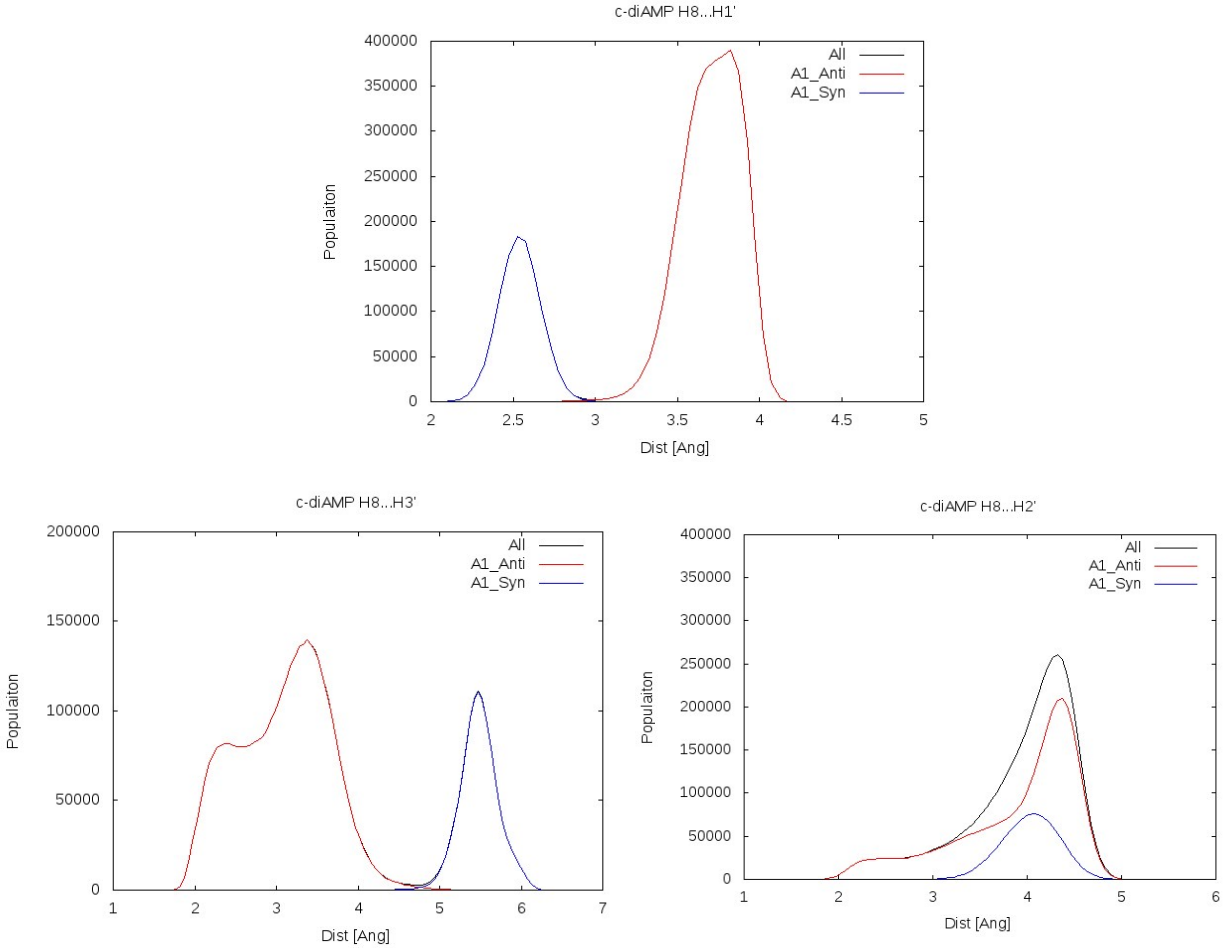


Figure S3: Distance distributions of purine H8 and ribose protons H1' (top), H2' (bottom left), and H3' (bottom right) as obtained in MD simulations of 3'3'-c-diAMP. H8/H1' distribution displays the sharpest and most separated peaks for *syn*- and *anti*- conformations, which makes it preferable for interpretation of ROESY spectra as done in section 3.2 of the main text.

Table S22: Conformational energies of lowest free-energy χ-conformers of the ***Syn-/Anti-*** set of 3'3'-c-di-AMP, 3'3'-c-GAMP, and 3'3'-c-di-GMP. Each column represents a result obtained with a unique combination of optimization method (indicated by the top row – "Opt.") and a single-point energy method (indicated by the second row – "SPE"). The original structures of the ***Syn-/Anti-*** set were optimized with B-P/COSMO protocol as described in the Computational Details section; these structures were then reoptimized by the respective methods listed in the table below. Lowest free-energy conformer for each of the molecules is highlighted in bold. All values are in kcal.mol^{-1}.

CDN	**X-conformer Class**	**Opt.**	**B-P/ COSMO**	**B-P/ COSMO**	**B-P/ COSMO**	**B-P/ COSMO**	**B3-LYP/ SMD**	**B3-LYP/ SMD**	**B3-LYP/ SMD**
		SPE	B-P/ COSMO	B-LYP/ CPCM	B-P/ COSMO-RS	DLPNO/ COSMO-RS	B3-LYP/ SMD	B3-LYP/ COSMO-RS	DLPNO/ COSMO-RS
3'3'-c-di-AMP	*anti- / anti-*		**0.0**	**0.0**	1.2	0.9	**0.0**	1.9	2.1
	syn- / anti-		1.2	2.1	**0.0**	**0.0**	2.0	**0.0**	**0.0**
	syn- / syn-		5.1	5.4	3.1	2.9	7.8	4.5	5.1
3'3'-c-GAMP	*anti- / anti-*		2.1	2.4	3.0	2.3	2.6	0.6	**0.0**
	anti- / syn-		4.5	4.7	0.4	**0.0**	4.8	0.5	1.4
	syn- / anti-		**0.0**	**0.0**	**0.0**	0.3	**0.0**	**0.0**	1.2
	syn- / syn-		2.3	1.9	3.3	2.5	1.8	0.3	1.5
3'3'-c-di-GMP	*anti- / anti-*		1.9	1.8	3.5	2.1	**0.0**	0.1	**0.0**
	syn- / anti-		**0.0**	0.6	**0.0**	**0.0**	**0.0**	**0.0**	0.7
	syn- / syn-		0.8	**0.0**	2.7	3.9	0.8	1.6	2.4

Table S22 (continued) : Conformational energies of lowest free-energy χ-conformers of the ***Syn-/Anti-*** set of 3'3'-c-di-AMP, 3'3'-c-GAMP, and 3'3'-c-di-GMP. Each column represents a result obtained with a unique combination of optimization method (indicated by the top row – "Opt.") and a single-point energy method (indicated by the second row – "SPE"). The original structures of the ***Syn-/Anti-*** set were optimized with B-P/COSMO protocol as described in the Computational Details section; these structures were then reoptimized by the respective methods listed in the table below. Lowest free-energy conformer for each of the molecules is highlighted in bold. All values are in kcal.mol^{-1}.

CDN	**X-conformer Class**	**Opt.**	**B3-LYP/ CPCM**	**B3-LYP/ CPCM**	**TPSS/ SMD**	**TPSS/ SMD**	**TPSS/ CPCM**	**TPSS/ CPCM**
		SPE	B3-LYP/ CPCM	B-P/ COSMO-RS	DLPNO/ COSMO-RS	B3-LYP/ SMD	TPSS/ CPCM	DLPNO/ COSMO-RS
3'3'-c-di-AMP	*anti- / anti-*		**0.0**	1.3	**0.0**	1.1	**0.0**	2.9
	syn- / anti-		1.9	**0.0**	2.2	**0.0**	1.2	**0.0**
	syn- / syn-		6.7	3.4	7.4	3.4	6.6	4.6
3'3'-c-GAMP	*anti- / anti-*		3.4	1.7	1.6	1.7	2.1	1.5
	anti- / syn-		3.8	0.9	3.6	1.0	5.3	2.9
	syn- / anti-		**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
	syn- / syn-		1.2	0.8	2.3	2.1	2.8	1.4
3'3'-c-di-GMP	*anti- / anti-*		1.0	0.4	**0.0**	2.3	0.5	1.6
	syn- / anti-		**0.0**	**0.0**	1.0	**0.0**	**0.0**	**0.0**
	syn- / syn-		0.3	2.0	1.9	3.0	1.4	1.2

Table S23: Conformational energies of lowest-energy χ-conformers of the **Syn/Anti** set of 3'3'-c-di-AMP, 3'3'-c-GAMP, and 3'3'-c-di-GMP . Lowest conformer for each of the molecules is highlighted in bold. Addition of zero-point vibrational energies (middle column), or ZPVE + thermal contributions (right column) introduces only minor changes into relative ranking of conformers

CDN	X-conformer Class	$E_{gp} + G_{solv}$	$E_{gp} + G_{solv}$ + ZPVE	$E_{gp} + G_{solv}$ + ZPVE - RTlnQ
3'3'-c-di-AMP	*anti- / anti-*	0.9	0.6	**0.0**
	syn- / anti-	**0.0**	**0.0**	0.5
	syn- / syn-	2.9	2.6	2.8
3'3'-c-GAMP	*anti- / anti-*	2.3	2.8	3.2
	anti- / syn-	**0.0**	**0.0**	**0.0**
	syn- / anti-	0.3	1.1	1.9
	syn- / syn-	2.5	2.8	2.7
3'3'-c-di-GMP	*anti- / anti-*	2.1	2.3	2.1
	syn- / anti-	**0.0**	**0.0**	**0.0**
	syn- / syn-	3.9	5.1	5.0

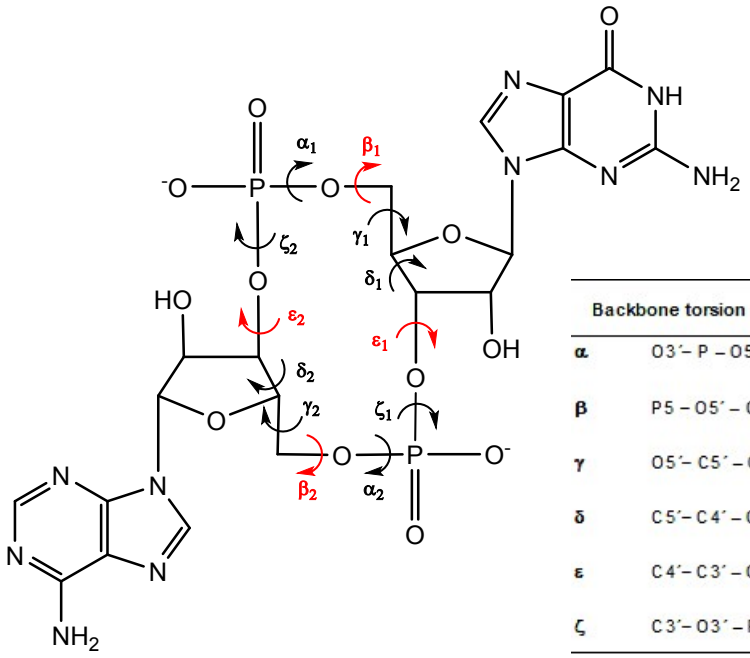
Table S24: Comparison of χ-conformer populations from MD simulations as described in Table 7 of the main text (left column) and an independent-base model (right column), assuming p(A,*anti*-) = 0.756 and p(G,*anti*-) = 0.216.

CDN	χ-conformer	MD population %	Independent-base model %
3'3'-c-di-AMP	*anti- / anti-*	57	57
	syn- / anti-	40	37
	syn- / syn-	3	6
3'3'-c-GAMP	*anti- / anti-*	18	16
	Gsyn- /A anti-	61	59
	Ganti- /A syn-	9	5
	syn- / syn-	12	19
3'3'-c-di-GMP	*anti- / anti-*	8	5
	syn- / anti-	31	34
	syn- / syn-	61	61

Table S25: Number of syn-/anti- interconversions observed in MD simulations. Total time of the simulation was 50 μs.

CDN	no. of interconversions	transition time scale [ns]
3'3'-c-di-AMP	9186 and 9432	5.4
G of 3',3'-c-GAMP	1233	40.6
A of 3',3'-c-GAMP	3808	13.1
3'3'-c-GAMP	659 and 689	74.2

Figure S4: Definitions of torsion angles of the macrocycle. Dihedrals β and ε are highlighted in red.



Backbone torsion angles		J (H,X)	J (C,X)
α	O3′– P – O5′– C5′	--	--
β	P5 – O5′ – C5′– C4′	J (H5′,P5) J (H5″,P5)	J (C4′,P5)
γ	O5′– C5′ – C4′– C3′	J (H4′,H5′) J (H4′,H5″)	J (C3′,H5′) J (C3′,H5″)
δ	C5′– C4′ – C3′– O3′	J (H3′,H4′)	J (C2′,H4′) J (C5′,H3′)
ε	C4′– C3′ – O3′– P3	J (P3,H3′)	J (C4′,P3) J (C2′,P3)
ζ	C3′– O3′ – P3′– O5′	--	--

Table S26: Values of $^{3}J(C4',P5')$ and comparison of the β-dihedral angle as calculated from NMR data, DFT and MD calculations.

CDN	^{3}J(C4',P5') [Hz]	β [degrees]		
		NMR[a]	DFT[b]	MD
3'3'-c-di-AMP	11.0	-164	-145	-171 (-164)[c]
G of 3',3'-c-GAMP	10.8	-162	-139	
A of 3',3'-c-GAMP	10.6	-160	-141	
3'3'-c-di-GMP	10.7	-161	-142; -140	

[a] Torsion angles determined from Karplus-type equation: $^{3}J(C,P) = 9.1\cos^2\Phi - 1.9\cos\Phi + 0.8$, see ref [2].
[b] Lowest-energy *Anti-/anti-* conformations from the ***Syn-/Anti-*** dataset are considered.
[c] Modal values of observed *anti-/anti-* populations are listed. Modus of the smaller population is listed in brackets.

Table S27: Values of $^{3}J(C4',P3')$ and comparison of the ε -dihedral angle as calculated from NMR data, DFT and MD calculations.

CDN	^{3}J(C4',P3') [Hz]	ε [degrees]		
		NMR[a]	DFT[b]	MD
3'3'-c-di-AMP	11.0	-164	176	-154 (-164)[c]
G of 3',3'-c-GAMP	10.8	-162	172	
A of 3',3'-c-GAMP	10.6	-160	172	
3'3'-c-di-GMP	10.1	-156	173; 172	

[a] Torsion angles determined from Karplus-type equation: $^{3}J(C,P) = 9.1\cos^2\Phi - 1.9\cos\Phi + 0.8$, see ref [2]
[b] Lowest-energy *anti-/anti-* conformations from the ***Syn-/Anti-*** dataset are considered.
[c] Modal values of observed *anti-/anti-* populations are listed. Modus of the smaller population is listed in brackets.

References:

1 J. H. Ippel, S. S. Wijmenga, R. de Jong, H. A. Heus, C. W. Hilbers, E. de Vroom, G. A. van der Marel and J. H. van Boom, Heteronuclear Scalar Couplings in the Bases and Sugar Rings of Nucleic Acids: Their Determination and Application in Assignment and Conformational Analysis, *Magnetic Resonance in Chemistry*, 1996, **34**, S156–S176.

2 J. Plavec and J. Chattopadhyaya, Reparametrization of Karplus equation relating 3JC□C□O□P to torsion angle, *Tetrahedron Letters*, 1995, **36**, 1949–1952.