

# SCIENTIFIC REPORTS



OPEN

## Computational and NMR spectroscopy insights into the conformation of cyclic di-nucleotides

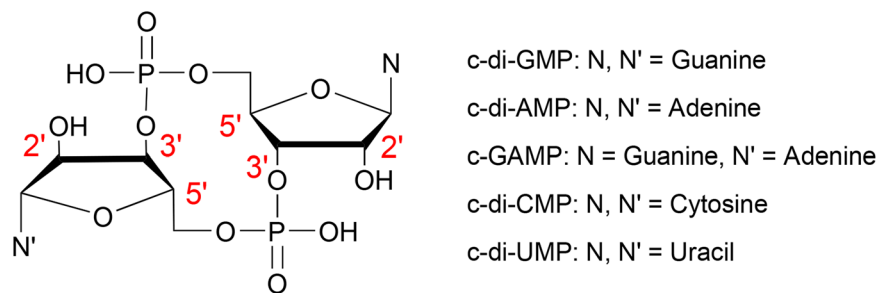
Baifan Wang<sup>1</sup>, Zhenghua Wang<sup>2</sup>, Uroš Javornik<sup>1</sup>, Zhen Xi<sup>2</sup> & Janez Plavec<sup>1,3,4</sup> 

Cyclic di-nucleotides (CDNs) are second messengers in bacteria and metazoan that are as such controlling important biological processes. Here the conformational space of CDNs was explored systematically by a combination of extensive conformational search and DFT calculations as well as NMR methods. We found that CDNs adopt pre-organized conformations in solution in which the ribose conformations are North type and glycosidic bond conformations are *anti* type. The overall flexibility of CDNs as well as the backbone torsion angles depend on the cyclization of the phosphodiester bond. Compared to di-nucleotides, CDNs display high rigidity in the macrocyclic moieties. Structural comparison studies demonstrate that the pre-organized conformations of CDNs highly resemble the biologically active conformations. These findings provide information for the design of small molecules to modulate CDNs signalling pathways in bacteria or as vaccine adjuvants. The rigidity of the backbone of CDNs enables the design of high order structures such as molecular cages based on CDNs analogues.

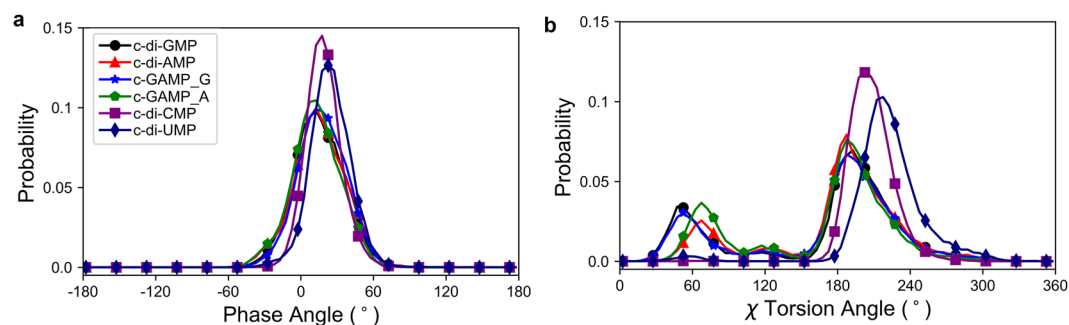
Cyclic di-nucleotides (CDNs) are composed of two nucleosides joined by two phosphate groups in a macrocycle (Fig. 1). CDNs have emerged as important second messengers in mammalian and bacteria cells<sup>1,2</sup>. In mammalian cells, cyclic GMP-AMP whose phosphate groups connect the two nucleosides from the 2'- and 5'- positions of guanosine and the 3'- and 5'- positions of adenosine (denoted as 2'3'-cGAMP), serves as a second messenger in the cell signalling pathway. 2'3'-cGAMP triggers the innate immune system by activating the adaptor protein stimulator of IFN genes (STING), which links the upstream cytosolic DNA detection and the downstream cytokine production<sup>3,4</sup>. In bacteria, bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) regulates a variety of processes such as cell motility, intercellular interactions, biofilm formation, dispersal and responses to oxidation<sup>1,2,5,6</sup>. Bis-(3'-5')-cyclic dimeric adenosine monophosphate (c-di-AMP) was identified as a crucial second messenger in the regulation of cell size, envelope stress control, fatty acid synthesis, ion transport and metabolite balance<sup>7-10</sup>. Bis-(3'-5')-cyclic guanosine monophosphate-adenosine monophosphate (c-GAMP) has been implicated in affecting bacteria intestinal colonization<sup>11</sup>. Modulating CDN signalling pathways in bacteria could represent a new way of controlling life processes in medical and industrial settings. CDNs are recognized also by mammalian immune systems as a uniquely bacterial molecule and therefore are considered promising vaccine adjuvants<sup>1,2</sup>.

There is a significant amount of data on the structural biology of CDN receptors. For example, crystal structures of c-di-GMP bound to various protein receptors which contain GGDEF I site, EAL, PilZ, BldD and VpsT domains as well as STING have been reported<sup>1,12</sup>. In the bound states c-di-GMP can adopt conformations ranging from a stacked form to a more extended form<sup>12</sup>. Moreover, c-di-GMP can bind to proteins in monomeric, dimeric or even tetrameric form<sup>13</sup>. For c-di-AMP, the structures of its complex with tricarboxylic acid cycle enzyme pyruvate carboxylase<sup>10</sup>, CTD domain<sup>14,15</sup>, BsDisA diadenylate cyclase<sup>16</sup>, phosphodiesterase PgpH HD domain<sup>17</sup> and PstA<sup>18-21</sup> have been determined. For c-GAMP, currently only two crystal structures of the molecule bound to mammalian STING are available<sup>22</sup>. Besides the protein receptors, CDNs have been shown to bind to riboswitches<sup>23-25</sup>. However, the conformation and dynamics of CDNs were rarely studied experimentally in solution.

<sup>1</sup>Slovenian NMR Center, National Institute of Chemistry, Hajdrihova 19, Ljubljana, Slovenia. <sup>2</sup>State Key Laboratory of Elemento-Organic Chemistry and Department of Chemical Biology, Nankai University, Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, 300071, P. R. China. <sup>3</sup>EN-FIST Center of Excellence, Trg OF 13, 1000, Ljubljana, Slovenia. <sup>4</sup>Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, Ljubljana, Slovenia. Baifan Wang and Zhenghua Wang contributed equally to this work. Correspondence and requests for materials should be addressed to Z.X. (email: [zhenxi@nankai.edu.cn](mailto:zhenxi@nankai.edu.cn)) or J.P. (email: [janez.plavec@ki.si](mailto:janez.plavec@ki.si))



**Figure 1.** Chemical structures of cyclic di-nucleotides.



**Figure 2.** The probability distribution of phase angle of pseudorotation (a) and  $\chi$  torsion angle (b) for CDNs. Data were obtained from REMD simulations at 300 K. The same colour scheme and style are used in both graphs.

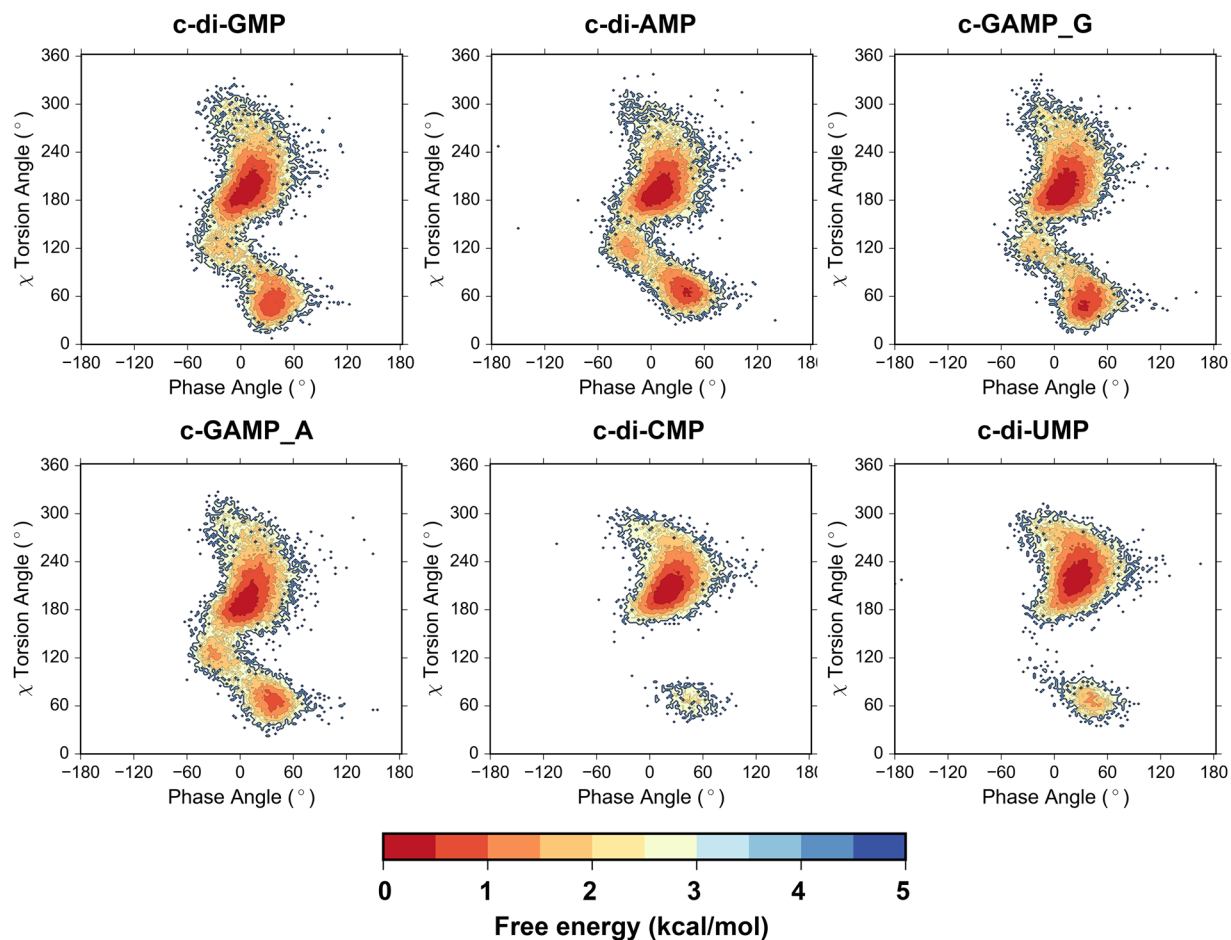
Despite the increasing interest in using CDNs to study signalling pathways in cells or as therapeutic agents, the current information on CDNs physicochemical properties is insufficient. It has been shown that the conformation of CDNs in solution is crucial for the evaluation of the binding constants to their receptors<sup>26</sup>. Such data are helpful for the understanding of the structure-activity relationship of either CDNs or their analogues and the design of potent therapeutic agents based on CDNs scaffold. The absence of comprehensive understanding on the conformation of free CDNs prompted us to perform thorough computational and NMR studies of the effect of nucleobase and cyclization through the phosphate linker. Specifically, we studied the conformation of CDNs composed of four major types of nucleobases (Fig. 1) by replica exchange molecular dynamics (REMD) simulations in implicit and explicit water environment and DFT calculation. To supplement the computational results, we synthesised five CDNs and analyzed their conformations in aqueous solution by NMR. The computational and NMR analysis provide parameters such as ribose puckering, conformational preference across glycosidic bond and backbone torsion angles.

## Results and Discussion

**Computational study of CDNs.** Firstly Replica exchange molecular dynamics (REMD)<sup>27</sup> simulations were performed on CDNs with 24 temperature states ranging from 273.0 to 583.5 K with simulation time up to 60 ns in implicit solvent. For temperature state at 300 K, a total of 30 000 conformations of CDNs were sampled. REMD was proved to effectively search the conformational space of CDNs as reflected in the transformation of  $\chi$  torsion angle from *anti* to *syn* and conversions between North (N-) and South (S-) type ribose conformations (see Supplementary Fig. S1). At 300 K the phase angles of pseudorotation<sup>28,29</sup> for all five CDNs are in the range from  $-60^\circ$  to  $60^\circ$  corresponding to the N-type conformation (Fig. 2a).

As for the conformation across glycosidic bonds, CDNs which are composed of pyrimidine bases mostly adopt *anti* conformation, while their counterparts with purines in addition display minor, but perceptible population of *syn* conformers (Fig. 2b). It is interesting to note that guanine and adenine moieties display slight differences in populations of *syn* conformers. As a consequence, free energy landscapes for  $\chi$  torsion angles and pseudorotational equilibria of sugar moieties show that CDNs composed of purine bases possess two stable states, of which the glycosidic bond in the *anti* region exhibits lower energy (about 1 kcal/mol) compared to the *syn* region; while the CDNs composed of pyrimidine bases exhibit a single state (Fig. 3).

The influence of cyclization of the phosphodiester groups on CDNs backbone torsion angles was analyzed by REMD simulations. The distribution of the six backbone torsion angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  and  $\zeta$  for individual CDN is shown in Fig. 4. Compared to the allowed ranges of backbone torsion angles in nucleosides, nucleotides, oligo- and poly-nucleotides<sup>30</sup>, it can be seen that torsion angles of CDNs show very narrow distributions. The dominant conformations of  $\alpha$  ( $g^+$ ),  $\beta$  ( $t$ ),  $\gamma$  ( $g^+$ ),  $\delta$  ( $g^+$ ),  $\epsilon$  ( $t$ ) and  $\zeta$  ( $g^+$ ) torsion angles are in accord with spatial restraints imposed by a 12-member macrocyclic moiety.

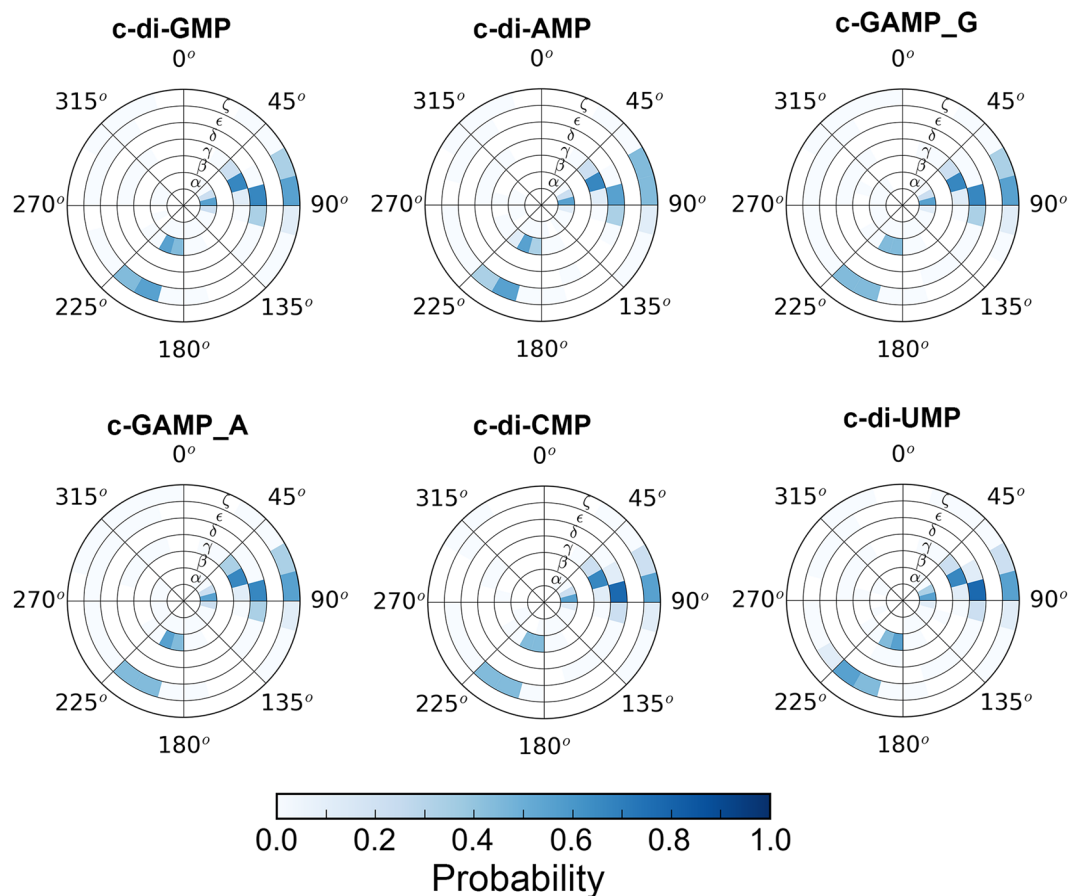


**Figure 3.** Population based free energy plot on  $\chi$ -phase angles plane for the CDNs based on REMD simulations at 300 K. Free energy estimates were calculated using the following equation:  $G_i = -k_B T \ln(N_i/N_{\max})$ , where  $k_B$  is Boltzmann constant,  $T$  is temperature,  $N_i$  is the population of bin  $i$  and  $N_{\max}$  is the population of the most populated bin.

To evaluate the effect of solvent on the conformational space of CDNs, REMD simulations were also performed in explicit water solvent. In the presence of explicit water molecules phase angles of pseudorotation are N-type for all five CDNs (see Supplementary Fig. S2). Noteworthy, similar observations can be made based on implicit solvent simulations. For the conformation across glycosidic bonds, the population of *syn* conformers of the guanine and uracil moieties (50% for guanine and 25% for uracil moiety) were significantly higher compared to REMD simulations in implicit solvent. CDNs composed of adenine and cytosine moieties displayed similar populations as in implicit solvent simulations. Examination of the CDN conformations showed that the amino group of guanine base forms a hydrogen bond with the phosphate group when the guanine moiety is in *syn* glycosidic conformation (see Supplementary Fig. S3). Comparison of the backbone torsion angles of CDNs obtained by simulations in implicit and explicit solvent showed no significant differences (see Supplementary Fig. S4).

**NMR study of CDNs.** All five CDNs examined by REMD simulations were synthesized (see Supplementary text) and subjected to NMR study. NMR spectra were measured at five temperatures ranging from 273 to 353 K to investigate the influence of temperature on the conformation of CDNs. The coupling constants between sugar protons, carbon and phosphorus ( $J_{H-H}$ ,  $J_{H-P}$  and  $J_{C-P}$ ) were determined for CDNs by various NMR spectra (see Supplementary Fig. S5–S9). The analysis of the  $^3J_{H-H}$  coupling constants (Supplementary Table S1) defining the sugar pucker and N-S populations showed that for *c*-di-CMP, *c*-di-UMP and *c*-di-GMP, the N-type ribose conformers dominate at 298 K (Table 1). The  $^3J_{H-H}$  coupling constants for the guanine moiety of *c*-GAMP cannot be determined at 298 K, but at 333 K the population of N-type conformers is about 95% (Table 1 and Supplementary Table S1). At 333 and 353 K, there is a slight tendency for CDNs composed of purine moieties to exhibit minor populations of S-type conformations, while CDNs composed of pyrimidines display 100% of N-type conformation (Supplementary Table S1).

The conformation of  $\chi$  torsion angles of the CDNs in solution was estimated from NOESY and ROESY spectra (see Supplementary Fig. S10–S14). The relative intensities of cross-peaks between aromatic H6 or H8 protons and ribose ring protons indicate a general preference for *anti* orientation of  $\chi$  torsion angle. The presence of



**Figure 4.** Conformational wheels for CDNs from REMD simulations in implicit solvent at 300 K.

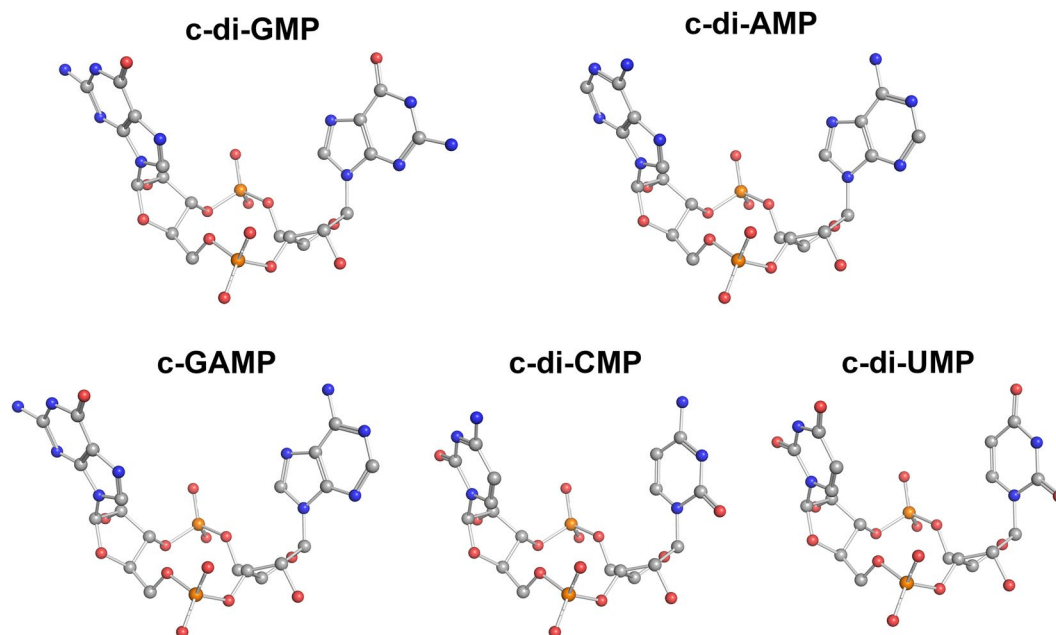
CDNs	$\chi$ [type]	$P_N$ [deg]	$\psi_N$ [deg]	ribose conformer
c-di-GMP	<i>anti</i> (211) <sup>1</sup>	0 (15)	41 (37)	100% N-type (N-type)
c-di-AMP	<i>anti</i> (210)	4 (13)	42 (35)	100% N-type (N-type)
c-GAMP (guanine moiety)	<i>anti</i> (210)	4 (13)	40 (34)	95% N-type <sup>2</sup> (N-type)
c-GAMP (adenine moiety)	<i>anti</i> (212)	1 (13)	40 (34)	100% N-type (N-type)
c-di-CMP	<i>anti</i> (201)	19 (12)	43 (37)	100% N-type (N-type)
c-di-UMP	<i>anti</i> (204)	18 (12)	40 (36)	100% N-type (N-type)

**Table 1.** The predominant glycosidic bond and ribose conformations of CDNs calculated from NMR measurements and DFT calculations. <sup>1</sup>Values in parenthesis are from DFT calculation. <sup>2</sup>The value of phase angle of pseudorotation for ribose of guanine moiety of c-GAMP was determined at 333 K.

H2-H2' and H2-H3' correlations for adenine moieties, which are not possible for *anti* orientation of  $\chi$ , indicates that a fraction of the population takes up *syn* conformation.

A series of transient 1D NOESY experiments was conducted for all five CDNs to obtain further semi-quantitative information about the conformational preferences of  $\chi$  torsion angles. The results show a general preference for *anti* conformation for all CDNs, reflected by higher intensities of H8/H6 to H2'/H3' relative to H8/H6 to H1' NOE enhancements. The preference is greater for pyrimidine analogues, where  $\eta$ (H6-H1') is undetectable for c-di-CMP.  $\eta$ (H6-H1') enhancement is approximately 8 times lower than  $\eta$ (H6-H2'/H3') for c-di-UMP. The ratio of  $\eta$ (H8-H1') to  $\eta$ (H8-H2'/H3') is approximately 1:2 for c-di-GMP and 1:4 for c-di-AMP. This indicates that significant populations of  $\chi$ -*syn* conformers exist for these CDNs. The data also suggests that the prevalence of *syn* conformation of  $\chi$  torsion angle is higher in c-di-GMP than c-di-AMP (Supplementary Table S2).

The results of REMD simulations in implicit solvent indicated significant population of molecules with *syn* conformation of  $\chi$ , but did not indicate any difference in populations between c-di-GMP and c-di-AMP. On the other hand, REMD simulations with explicit solvent display higher fraction of *syn* conformers for guanine in comparison to adenine moiety. However, the fraction of *syn* conformers of  $\chi$  for guanine moiety (about 50%) was overestimated according to NMR results. We suggest that formation of a hydrogen bond between the guanine



**Figure 5.** Lowest energy structures of CDNs optimized at B3LYP/6-31 G(d,p) level.

amino and phosphate groups plays a role in the conformational equilibrium of  $\chi$  conformers. Explicit water molecules can stabilize this bonding, but with the applied force field, this effect might have been exaggerated.

The populations of  $g^+$ ,  $g^-$  and  $t$  conformers across C4'-C5' bond were determined through the analysis of  $^3J_{\text{H4}'\text{-H5}'}$  and  $^3J_{\text{H4}'\text{-H5}''}$  coupling constants (Supplementary Table S3). At 298 K, all CDNs adopt 100%  $g^+$  conformer population along  $\gamma$  torsion angle. The  $^3J_{\text{H4}'\text{-H5}'}$  and  $^3J_{\text{H4}'\text{-H5}''}$  coupling constants for c-GAMP could not be determined. The  $\beta$  and  $\varepsilon$  torsion angles which describe the rotation of O5'-C5' and C3'-O3' bonds were calculated based on  $^3J_{\text{H-P}}$ ,  $^4J_{\text{H-P}}$  and  $^3J_{\text{C-P}}$  coupling constants (Supplementary Table S3) using the parametrization of Karplus equation for CCOP fragments<sup>31,32</sup>. The  $\beta$  torsion angles for all CDNs are in the range from  $-170^\circ$  to  $-160^\circ$ , while the  $\varepsilon$  torsion angles are in the range from  $-160^\circ$  to  $-150^\circ$ .

Since the ribose and backbone conformations of CDNs predicted by both implicit and explicit REMD simulations are highly similar, and considering that the predominant glycosidic bond population observed by NMR is *anti*, we chose the lowest energy conformation observed on the free energy landscape of CDNs from implicit REMD simulation (Fig. 3) for further analysis. DFT calculations were performed on these structures. The comparison of DFT optimized structures shows that despite different orientations of nucleobases in CDNs, conformations of their macrocyclic moieties highly resemble each other (Fig. 5). The structural parameters computed from DFT optimized CDNs show nice agreement with NMR measurements (Tables 1 and 2).

**Comparison of the conformation between CDNs and linear di-nucleotides.** Earlier NMR studies have demonstrated that di-nucleotides, ApA, ApC and CpG adopt both S- and N-type ribose conformations<sup>33-35</sup>. REMD simulations showed that GpG and GpA adopt both S- and N-type ribose as well as *syn* and *anti* conformations of  $\chi$  torsion angle (Supplementary Fig. S15). As a consequence, four conformation states were observed for each moiety of di-nucleotides (Fig. 6). The backbone torsion angles of di-nucleotides were found to have much wider distributions compared to CDNs (Supplementary Fig. S16).

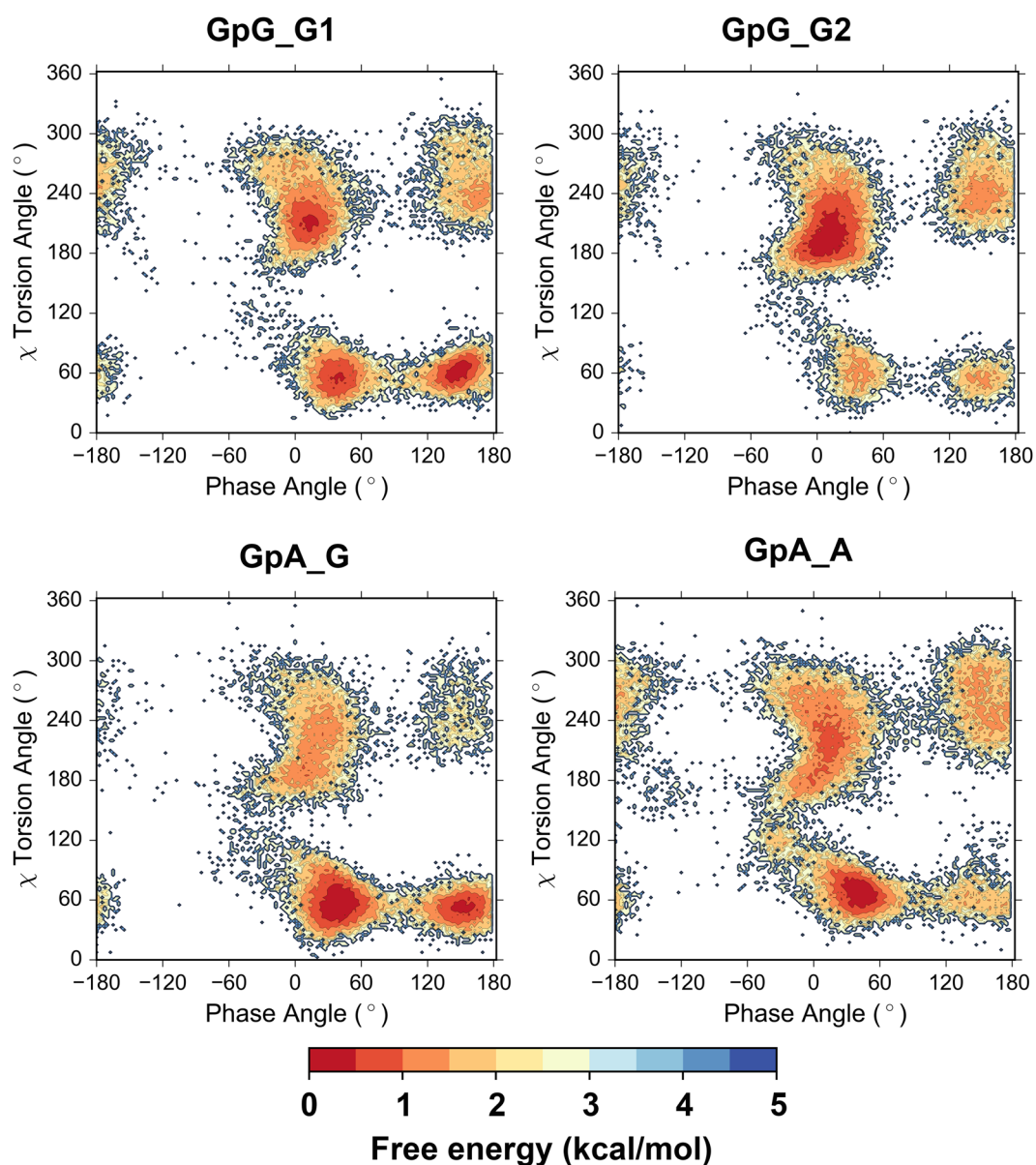
For the di-nucleoside polyphosphates, the ribose moieties showed a small preference for S-type conformation, but when attached to cytosine the ribose ring preferred N-type conformation<sup>36</sup>. The conformation around the glycosidic bond in di-nucleoside polyphosphates was found to be *anti/high anti*<sup>36</sup>. The conformations of di-nucleoside polyphosphates were found to be highly dynamic, hence no single predominant conformation was observed.

In comparison, CDNs display mainly N-type ribose conformation and *anti* glycosidic bond conformation. Shifting the ribose conformation from N- to S-type results in changing the O3'-C3' bond orientation from pseudo-equatorial (*e*) to pseudo-axial (*a*) position (Supplementary Fig. S17) which might be accompanied by rotation of the phosphate group. The phosphate group attached at the O3' is relatively free to rotate in di-nucleotides, while rotation is restricted in the CDNs due to the macrocycle. This effect may be the origin of the different conformational preferences for CDNs and linear di-nucleotides.

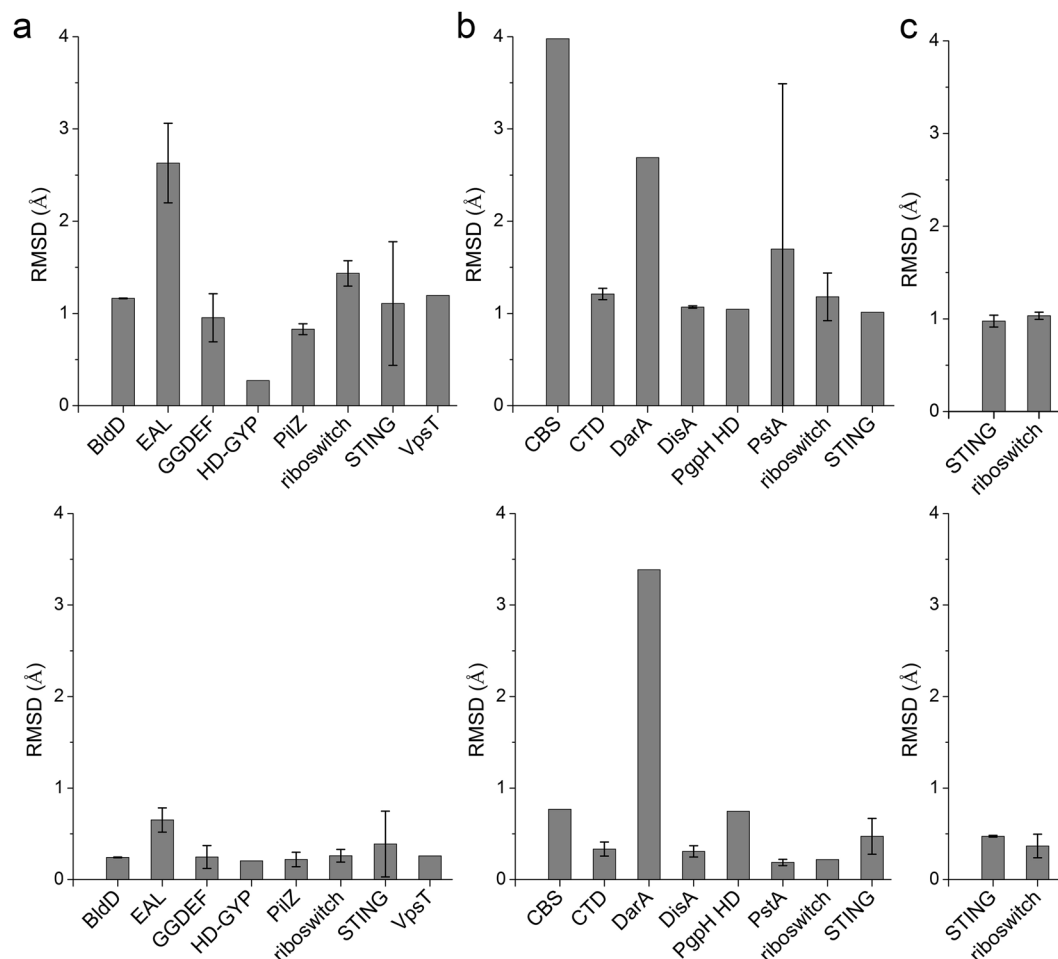
**Structural comparison of CDNs in solution and bound state.** The coordinates of CDNs bound to receptors were extracted and compared to the solution conformations. Except for the c-di-GMP bound to EAL domain based proteins, the superimpositions of the DFT optimized conformation and the receptor-bound conformations show overall and backbone (macrocyclic moieties) RMSD values less than 2.0 and 1.0 Å, respectively (Fig. 7a and Supplementary Fig. S18). Our DFT-optimized c-di-GMP conformation showed striking similarity

CDNs	$\alpha$ [deg]	$\beta$ [deg]	$\gamma$ [type]	$\delta$ [deg]	$\epsilon$ [deg]	$\zeta$ [deg]
c-di-GMP	n.d. (75) <sup>1</sup>	-170 (-173)	g <sup>+</sup> (g <sup>+</sup> )	86 (84)	-150 (-155)	n.d. (69)
c-di-AMP	n.d. (70)	-170 (-174)	g <sup>+</sup> (g <sup>+</sup> )	84 (87)	-150 (-154)	n.d. (70)
c-GAMP (guanine moiety)	n.d. (71)	-160 (-174)	n.d. (g <sup>+</sup> )	86 (87)	-150 (-155)	n.d. (69)
c-GAMP (adenine moiety)	n.d. (71)	-170 (-174)	g <sup>+</sup> (g <sup>+</sup> )	87 (88)	-160 (-155)	n.d. (69)
c-di-CMP	n.d. (73)	-170 (-170)	g <sup>+</sup> (g <sup>+</sup> )	79 (81)	-150 (-161)	n.d. (73)
c-di-UMP	n.d. (76)	-170 (-174)	g <sup>+</sup> (g <sup>+</sup> )	82 (81)	-150 (-158)	n.d. (71)

**Table 2.** Backbone parameters of CDNs calculated from NMR measurements and DFT calculations. <sup>1</sup>Values in parenthesis are from DFT calculation. n.d.: not determined.



**Figure 6.** Population based free energy plot on the  $\chi$ -phase angles plane for di-nucleotides from REMD simulation at 300 K.



**Figure 7.** Comparison of the overall (upper panel) and backbone (macrocyclic moieties, lower panel) structural similarity expressed in terms of RMSD between various receptor-bound and DFT optimized CDN conformations. (a): c-di-GMP; (b): c-di-AMP, (c): c-GAMP.

in the conformation of macrocyclic moieties to the bound states for most type of receptors. RMSD values for the nucleobases are higher, however the glycosidic bond conformations of the bound states are still mostly *anti*. c-di-GMP mainly displays an extended conformation (Supplementary Fig. S18) when bound to the EAL domain based protein. The EAL domain originated from c-di-GMP-specific phosphodiesterases (PDEs) whose function is to hydrolyze c-di-GMP. The fully extended monomeric conformation of c-di-GMP is apparently more susceptible to cleavage of the ribose-phosphate ring<sup>37–39</sup>. The same binding mode of the extended monomeric c-di-GMP is retained in the enzymatically inactive EAL domains that function exclusively as c-di-GMP receptors<sup>40–42</sup>, which makes the bound states display large differences to the free state.

It has been observed that c-di-GMP can bind to protein in monomeric, dimeric and tetrameric forms<sup>14</sup> (Supplementary Fig. S19). In the dimeric form, c-di-GMP molecules adopt a “U” shape like structure and intercalate with each other (Supplementary Fig. S19a). The c-di-GMP tetramer is composed of two dimers (Supplementary Fig. S19b). The driving force of the intercalation mostly comes from the stacking between guanine bases as well as hydrogen bonds between amino and imino proton of guanine with the phosphate group. Interestingly, the DFT optimized conformation of c-di-GMP shows high similarity with conformations in the dimeric and tetrameric forms, which suggests that the backbone conformation of c-di-GMP is pre-organized for intercalation. The comparison of the DFT optimized monomer and dimer conformations showed that the monomer conformation also resembles the conformations in the dimeric form (Supplementary Fig. S19c).

As for c-di-AMP, the comparison of the DFT optimized conformations to those bound to receptors such as tri-carboxylic acid cycle enzyme pyruvate carboxylase, CTD domain, BsDisA diadenylate cyclase, phosphodiesterase PgpH HD domain and PstA (Fig. 7b and Supplementary Fig. S20) reveals backbone RMSD values below 1.0 Å, indicating similar backbone conformation. Molecules of c-di-AMP bound to CBS and PstA display overall RMSD values larger than 2.0 Å relative to DFT optimized structures, which can be attributed to differences in nucleobase orientations. For c-di-AMP bound to DarA, both overall and backbone RMSD values are around 3.0 Å indicating larger conformational change upon binding of free c-di-AMP to DarA.

For c-GAMP, the comparison of the DFT optimized structures with STING and riboswitch-bound conformations (see Fig. 7c and Supplementary Fig. S21) shows high similarity in the conformation of the macrocyclic moieties and differences in the orientation of the nucleobases.

**Conformational study of 2'-OH modification of c-di-GMP.** Several 2'-OH modified analogues of c-di-GMP (Supplementary Fig. S22), which serve as potential inhibitors of c-di-GMP receptors<sup>43</sup>, have been subjected to REMD simulations. It has been shown that, in agreement with NMR measurement<sup>43</sup>, the substitution of 2'-OH by fluorine or methoxyl group does not affect the backbone conformation and phase angle of pseudorotation of c-di-GMP, while the substitution of 2'-OH by hydrogen results in unbiased sugar conformation (Supplementary Fig. S23). The glycosidic bonds of these analogues were shifted to *syn* conformations (Supplementary Fig. S23) resulting in a single major conformational state (Supplementary Fig. S24). The c-di-GMP\_2'F, in which the 2'-OH groups were replaced by fluorine, displayed 4 times higher binding affinity to GGDEF I-site based diguanylate cyclase than c-di-GMP<sup>43</sup>. Since the backbone and sugar puckering are unaffected by the modification (see Supplementary Fig. S25–S26), the enhancement of the binding affinity may result from the fluorine-hydrogen bonding interactions between 2'-F and arginine in the I-site. The native CDN displayed 10 times higher binding affinity to EAL-based phosphodiesterases than c-di-GMP\_2'F, which may be due to *syn* glycosidic bond conformations.

## Conclusions

In summary, we have systematically explored the conformational space of CDNs by computational and NMR studies. The ribose moieties of CDNs adopt predominantly N-type conformation, while the glycosidic bonds prefer *anti* conformations. The 12-membered rings formed by the backbone of CDNs are shown to be highly rigid even upon the 2'-OH modification of ribose. The conformations of macrocyclic moieties of free CDNs are very similar to conformations observed for CDNs bound to most of receptors. These findings shed light on molecular mechanisms underlying the activities of CDNs in a structural way and provide incentives for the design of small molecules to modulate CDN signalling pathways in bacteria or as vaccine adjuvants, that is, further study can focus on the modification of nucleobase or substitutions of the ribose with the aim to improve the binding constant or cell permeability of the CDN analogues. The rigidity of the backbone of CDNs also enables the design of high order structures such as molecular cages based on CDNs analogues. This study also shows that combination of REMD simulation and DFT calculations could efficiently probe the conformational space of CDNs and its analogues for further studies.

## Materials and Methods

**Replica exchange molecular dynamics simulations.** Replica exchange molecular dynamics (REMD) simulation is an advanced sampling technique that promotes efficient conformational sampling by enhancing the probability of sampling high-energy configurations at elevated temperatures. Several identical copies (replicas) of the system are run in parallel, each differing in temperature. The neighboring replicas may exchange their temperature states based on a Boltzmann-weighted probability. Replicas are allowed to communicate at regular intervals during which exchange attempts are made based on a Monte Carlo criterion. When this condition is satisfied an exchange attempt is considered successful, and the conformations in neighboring replica temperatures are swapped. The velocity of the corresponding replica is then rescaled to the new replica temperature. The process is repeated iteratively during the simulation such that each replica evolves with a wide range of temperatures, enhancing conformational sampling. At present study, a 75% exchange rate is observed for all of the CDNs.

The partial charges for CDNs were generated by geometry optimization and electrostatic potential (ESP) calculations with Gaussian 09<sup>44</sup> at the level of HF/6–31 G\* by the RED server<sup>45–47</sup>. Other force field parameters were taken from Amber ff14SB force field basic version parm99<sup>48</sup> with the bsc0<sup>49</sup> and  $\chi$ OL3 refinements<sup>50,51</sup>.

Simulations were performed with AMBER 14 molecular modeling package<sup>52,53</sup>. The structures were subjected to energy minimization for 2000 cycles, where first 500 cycles were performed by steepest descent energy minimization, and the remaining 1500 cycles were minimized by conjugate gradient minimization. The minimized structure was used to generate chirality constraints to prevent unwanted rotation around the backbone bonds, which might occur at higher temperature during the REMD simulation. The generalized Born implicit solvation model<sup>54,55</sup> and NVT ensembles were used in the REMD. The SHAKE algorithm<sup>56</sup> was used to constrain the bond stretching freedom of all bonds involving hydrogens, and the nonbonded van der Waals and electrostatic cutoffs were taken as 999 Å. For REMD, 24 replica temperatures were used (273.0, 286.5, 300.0, 313.5, 327, 340.5, 354, 367.5, 381, 394.5, 408.0, 421.5, 435, 448.5, 462.0, 475.5, 489.0, 502.5, 516.0, 529.5, 543, 556.5, 570.0 and 583.5 K). Replica temperatures were maintained by weak coupling to the Langevin thermostat with a collision frequency of 1 ps<sup>-1</sup>. Prior to REMD simulations, the system was equilibrated for 200 ps, during which the temperature of each replica was gradually increased from 0 to the target temperature of that replica. After equilibration, REMD simulation was performed using the multisander module. The integration step for the production run was 0.002 ps. Replica temperature exchange attempts were performed every 1.5 ps. The output and coordinate files were saved every 1 ps, and the total length for each of the simulations was 60 ns. All trajectories were processed with cpptraj to filter the trajectory corresponding to 300 K.

We also performed REMD simulations in explicit solvent. The CDNs were solvated with TIP3P water molecules in a truncated octahedron periodic box and the total system charge was neutralized with magnesium ions. Energy minimization was performed for 1000 steps with the steepest descent algorithm followed by 4000 steps with the conjugate gradient algorithm. After minimization, the systems were heated from 0 K to the desired temperature in 50 ps of each replica at constant volume with 10 kcal/mol Å<sup>2</sup> atomic positional restraints on CDNs. The temperature was controlled using a Langevin thermostat with a collision frequency of 2.0 ps<sup>-1</sup>. Prior to production simulations, a 200 ps equilibrium period was employed to equilibrate each replica. The parameters for the production runs were identical to the implicit solvent simulations.

**DFT calculation.** The DFT calculations were executed using Gaussian 09<sup>43</sup>. The geometries of CDNs were optimized at the B3LYP/6–31 G(d,p) level and incorporated Tomasi's Polarized Continuum Model (PCM)



corrections for water as the bulk solvent. Frequency calculations for all stationary points were carried out to confirm them as minima ( $i = 0$ ).

**NMR experiment.** Spectra were recorded on Agilent Technologies DD2 600 MHz NMR spectrometer equipped with  $^1\text{H}\{^{13}\text{C},^{15}\text{N}\}$  cold probe. Phosphorus decoupled  $^1\text{H}$  NMR spectra and spectra at  $80^\circ\text{C}$  were recorded on the same spectrometer using OneNMR probe. NOESY and ROESY spectra were recorded at  $20^\circ\text{C}$  on Agilent Technologies DD2 300 MHz NMR spectrometer equipped with ID/PFG probe with various mixing times between 60 and 400 ms for samples of each of the dinucleotides. For NMR experiments samples were dissolved in a TRIS/HCl buffer with 100 mM NaCl and 5 mM  $\text{MgCl}_2$  in  $\text{D}_2\text{O}$  at pH 7.4 to a concentration between 2 and 6 mM.  $^1\text{H}$ ,  $^{13}\text{C}$ , TOCSY, COSY,  $^1\text{H}$ - $^{13}\text{C}$  HSQC and  $^1\text{H}$ - $^{13}\text{C}$  HMBC spectra were recorded for assignment.  $J_{\text{H-H}}$  and  $J_{\text{H-P}}$  coupling constants were determined from analysis of multiplet structure in  $^1\text{H}$  spectra.  $^{31}\text{P}$  decoupled  $^1\text{H}$  NMR spectra were recorded to facilitate interpretation of multiplets involving  $J_{\text{HP}}$  couplings.  $^1\text{H}$  spectra were recorded between 0 and  $80^\circ\text{C}$  in  $20^\circ\text{C}$  steps.  $J_{\text{C-P}}$  values constants were evaluated from  $^1\text{H}$  decoupled  $^{13}\text{C}$  NMR spectra recorded at  $20^\circ\text{C}$ .

Orientations of the  $\chi$  dihedral angles were estimated from cross-peaks in NOESY spectra as well as a series of transient 1D NOESY spectra. 1D NOESY experiments were recorded at  $20^\circ\text{C}$  using the NOESY1D pulse sequence. NOE build-up curves were measured to determine the optimal mixing times in the linear regime. The spectra were recorded with mixing times of 0.6 s for all CDNs except c-di-GMP, for which a mixing time of 0.35 s was employed.

$^3J_{\text{H-H}}$  coupling constants values were fitted to a mixture of two dominant conformers and optimized using a Pseurot<sup>57</sup> based software for pseudorotation analysis of saturated five-membered ring systems implemented in Matlab<sup>58</sup>. The minor (S-type) conformation was kept fixed at  $P_s 163^\circ$  and  $\psi_s 38^\circ$  for the optimization, while  $P_N$ ,  $\psi_N$  and fractions of both populations were optimized.  $\gamma$  torsion angles were estimated from  $^3J_{\text{H}4'-\text{H}5'}$  and  $^3J_{\text{H}4'-\text{H}5''}$  coupling constants. The presence of three conformations of  $\gamma$  torsion angle at  $53^\circ$ ,  $180^\circ$  and  $-70^\circ$  was assumed. Values of scalar coupling constants were calculated for these angles and fractions were optimized for the best match to experimental values.

$\beta$  and  $\epsilon$  angles were estimated from the relevant  $J_{\text{C-P}}$  and  $J_{\text{H-P}}$  coupling constants.  $J_{\text{C-P}}$  coupling constants were analyzed according to parametrization of the Karplus equation offered by Plavec and Chattopadhyaya<sup>29</sup>.

## References

- Romling, U., Galperin, M. Y. & Gomelsky, M. Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiol. Mol. Biol. Rev.* **77**, 1–52 (2013).
- Gao, J., Tao, J., Liang, W. & Jiang, Z. Cyclic (di)nucleotides: the common language shared by microbe and host. *Curr. Opin. Microbiol.* **30**, 79–87 (2016).
- Cai, X., Chiu, Y. H. & Chen, Z. J. The cGAS-cGAMP-STING Pathway of Cytosolic DNA Sensing and Signaling. *Mol. Cell* **54**, 289–296 (2014).
- Wu, J. X. & Chen, Z. J. Innate Immune Sensing and Signaling of Cytosolic Nucleic Acids. *Annu. Rev. Immunol.* **32**, 461–488 (2014).
- Boyd, C. D. & O'Toole, G. A. Second Messenger Regulation of Biofilm Formation: Breakthroughs in Understanding c-di-GMP Effector Systems. *Annu. Rev. Cell. Dev. Biol.* **28**, 439–462 (2012).
- Ryan, R. P. Cyclic di-GMP signalling and the regulation of bacterial virulence. *Microbiology* **159**, 1286–1297 (2013).
- Corrigan, R. M. *et al.* Systematic identification of conserved bacterial c-di-AMP receptor proteins. *Proc. Natl. Acad. Sci. USA* **110**, 9084–9089 (2013).
- Zhang, L., Li, W. H. & He, Z. G. DarR, a TetR-like Transcriptional Factor, Is a Cyclic Di-AMP-responsive Repressor in *Mycobacterium smegmatis*. *J. Biol. Chem.* **288**, 3085–3096 (2013).
- Corrigan, R. M. & Grundling, A. Cyclic di-AMP: another second messenger enters the fray. *Nat. Rev. Microbiol.* **11**, 513–524 (2013).
- Sureka, K. *et al.* The Cyclic Dinucleotide c-di-AMP Is an Allosteric Regulator of Metabolic Enzyme Function. *Cell* **158**, 1389–1401 (2014).
- Davies, B. W., Bogard, R. W., Young, T. S. & Mekalanos, J. J. Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for *V. cholerae* Virulence. *Cell* **149**, 358–370 (2012).
- Chou, S. H. & Galperin, M. Y. Diversity of Cyclic Di-GMP-Binding Proteins and Mechanisms. *J. Bacteriol.* **198**, 32–46 (2016).
- Tschowri, N. *et al.* Tetrameric c-di-GMP Mediates Effective Transcription Factor Dimerization to Control *Streptomyces* Development. *Cell* **158**, 1136–1147 (2014).
- Chin, K. H. *et al.* Structural Insights into the Distinct Binding Mode of Cyclic Di-AMP with SaCpaA\_RCK. *Biochemistry* **54**, 4936–4951 (2015).
- Kim, H. *et al.* Structural Studies of Potassium Transport Protein KtrA Regulator of Conductance of K<sup>+</sup> (RCK) C Domain in Complex with Cyclic Diadenosine Monophosphate (c-di-AMP). *J. Biol. Chem.* **290**, 16393–16402 (2015).
- Witte, G., Hartung, S., Büttner, K. & Hopfner, K.-P. Structural Biochemistry of a Bacterial Checkpoint Protein Reveals Diadenylate Cyclase Activity Regulated by DNA Recombination Intermediates. *Mol. Cell* **30**, 167–178 (2008).
- Huynh, T. N. *et al.* An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. *Proc. Natl. Acad. Sci. USA* **112**, E747–E756 (2015).
- Müller, M., Hopfner, K.-P. & Witte, G. c-di-AMP recognition by *Staphylococcus aureus* PstA. *FEBS Lett.* **589**, 45–51 (2015).
- Gundlach, J. *et al.* Identification, Characterization, and Structure Analysis of the Cyclic di-AMP-binding PII-like Signal Transduction Protein DarA. *J. Biol. Chem.* **290**, 3069–3080 (2015).
- Choi, P. H., Sureka, K., Woodward, J. J. & Tong, L. Molecular basis for the recognition of cyclic-di-AMP by PstA, a PII-like signal transduction protein. *Microbiologyopen* **4**, 361–374 (2015).
- Campeotto, I., Zhang, Y., Mladenov, M. G., Freemont, P. S. & Grundling, A. Complex Structure and Biochemical Characterization of the *Staphylococcus aureus* Cyclic Diadenylate Monophosphate (c-di-AMP)-binding Protein PstA, the Founding Member of a New Signal Transduction Protein Family. *J. Biol. Chem.* **290**, 2888–2901 (2015).
- Kranzusch, P. J. *et al.* Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. *Mol. Cell* **59**, 891–903 (2015).
- Ren, A. M. *et al.* Structural Basis for Molecular Discrimination by a 3',3'-cGAMP Sensing Riboswitch. *Cell Rep.* **11**, 1–12 (2015).
- Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat. Rev. Microbiol.* **7**, 263–273 (2009).
- Smith, K. D. *et al.* Structural basis of ligand binding by a c-di-GMP riboswitch. *Nat. Struct. Mol. Biol.* **16**, 1218–U1227 (2009).
- Shi, H., Wu, J., Chen, Z. J. & Chen, C. Molecular basis for the specific recognition of the metazoan cyclic GMP-AMP by the innate immune adaptor protein STING. *Proc. Natl. Acad. Sci. USA* **112**, 8947–8952 (2015).

27. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
28. Altona, C. & Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94**, 8205–8212 (1972).
29. Plavec, J., Tong, W. & Chattopadhyaya, J. How do the gauche and anomeric effects drive the pseudorotational equilibrium of the pentofuranose moiety of nucleosides? *J. Am. Chem. Soc.* **115**, 9734–9746 (1993).
30. Neidle, S. *Principles of Nucleic Acid Structure*. (Academic Press, 2010).
31. Mooren, M. M. W., Wijmenga, S. S., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. The solution structure of the circular trinucleotide cr(GpGpGp) determined by NMR and molecular mechanics calculation. *Nucleic Acids Res.* **22**, 2658–2666 (1994).
32. Plavec, J. & Chattopadhyaya, J. Reparametrization of Karplus equation relating  $^3J_{C-C-O-P}$  to torsion angle. *Tetrahedron Lett.* **36**, 1949–1952 (1995).
33. Kondo, N. S. & Danyluk, S. S. NMR studies of nucleic acids in solution. III. Conformational properties of adenylyl-3' → 5'-adenosine in aqueous solution. *Biochemistry* **15**, 756–768 (1976).
34. Damha, M. J. & Ogilvie, K. K. Conformational properties of branched RNA fragments in aqueous solution. *Biochemistry* **27**, 6403–6416 (1988).
35. Premraj, B. J. *et al.* NMR Structure of a 2',5' RNA Favors A Type Duplex with Compact C2'endo Nucleotide Repeat. *Biochem. Biophys. Res. Commun.* **283**, 537–543 (2001).
36. Stern, N., Major, D. T., Gottlieb, H. E., Weizman, D. & Fischer, B. What is the conformation of physiologically-active dinucleoside polyphosphates in solution? Conformational analysis of free dinucleoside polyphosphates by NMR and molecular dynamics simulations. *Org. Biomol. Chem.* **8**, 4637–4652 (2010).
37. Barends, T. R. M. *et al.* Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature* **459**, 1015–1018 (2009).
38. Minasov, G. *et al.* Crystal Structures of YkuI and Its Complex with Second Messenger Cyclic Di-GMP Suggest Catalytic Mechanism of Phosphodiester Bond Cleavage by EAL Domains. *J. Biol. Chem.* **284**, 13174–13184 (2009).
39. Tchigvintsev, A. *et al.* Structural Insight into the Mechanism of c-di-GMP Hydrolysis by EAL Domain Phosphodiesterases. *J. Mol. Biol.* **402**, 524–538 (2010).
40. Guzzo, C. R., Dunger, G., Salinas, R. K. & Farah, C. S. Structure of the PilZ-FimXEAL-c-di-GMP Complex Responsible for the Regulation of Bacterial Type IV Pilus Biogenesis. *J. Mol. Biol.* **425**, 2174–2197 (2013).
41. Navarro, M. V. A. S., De, N., Bae, N., Wang, Q. & Sondermann, H. Structural Analysis of the GGDEF-EAL Domain-Containing c-di-GMP Receptor FimX. *Structure* **17**, 1104–1116 (2009).
42. Navarro, M. V. A. S. *et al.* Structural Basis for c-di-GMP-Mediated Inside-Out Signaling Controlling Periplasmic Proteolysis. *PLoS Biol.* **9**, e1000588 (2011).
43. Zhou, J. *et al.* Potent suppression of c-di-GMP synthesis via I-site allosteric inhibition of diguanylate cyclases with 2'-F-c-di-GMP. *Bioorg. Med. Chem.* **21**, 4396–4404 (2013).
44. Frisch, M. J. *et al.* Fox. *Gaussian 09 (Revision E. 01)*. Gaussian, Inc., Wallingford, CT (2009).
45. Vanqulelf, E. *et al.* RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **39**, W511–W517 (2011).
46. Dupradeau, F.-Y. *et al.* The RED. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **12**, 7821–7839 (2010).
47. Bayly, C. L., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).
48. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **118**, 2309–2309 (1996).
49. Perez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.* **92**, 3817–3829 (2007).
50. Banás, P. *et al.* Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **6**, 3836–3849 (2010).
51. Zgarbová, M. *et al.* Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).
52. Salomon-Ferrer, R., Goetz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).
53. Case, D. A. *et al.* AMBER 2015, University of California, San Francisco. (2015).
54. Onufriev, A., Bashford, D. & Case, D. A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **104**, 3712–3720 (2000).
55. Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**, 383–394 (2004).
56. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
57. De Leeuw, F. A. A. M. & Altona, C. Computer-assisted pseudorotation analysis of five-membered rings by means of proton spin-spin coupling constants: Program PSEUROT. *J. Comput. Chem.* **4**, 428–437 (1983).
58. Hendrickx, P. M. & Martins, J. C. A user-friendly Matlab program and GUI for the pseudorotation analysis of saturated five-membered ring systems based on scalar coupling constants. *Chem. Cent. J.* **2**, 20 (2008).

## Acknowledgements

This work was supported by the Slovenian Research Agency (ARRS, grants P1-0242 and J1-6733) and the National Natural Science Foundation of China (21332004, 21472101). We are grateful to National Supercomputing Center in Tianjin for providing computational resources.

## Author Contributions

The project was conceived and designed by J.P., Z.X. and B.W. B.W. performed the computational work, Z.W. prepared all the synthesis, U.J. performed the NMR work. B.W., Z.W., U.J., J.P. and Z.X. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-16794-4>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017