What is the conformation of physiologically-active dinucleoside polyphosphates in solution? Conformational analysis of free dinucleoside polyphosphates by NMR and molecular dynamics simulations†

Noa Stern, Dan Thomas Major, Hugo Emilio Gottlieb, Daniel Weizman and Bilha Fischer*

Received 15th April 2010, Accepted 24th June 2010
DOI: 10.1039/c005122e

Dinucleoside polyphosphates, or dinucleotides (Np_nN'; N, N′ = A, U, G, C; n = 2–7), are naturally occurring ubiquitous physiologically active compounds. Despite the interest in dinucleotides, and the relevance of their conformation to their biological function, the conformation of dinucleotides has been insufficiently studied. Therefore, here we performed conformational analysis of a series of Np_nN' Na^+ salts (N = A, G, U, C; N′ = A, G, U, C; n = 2–5) by various NMR techniques. All studied dinucleotides, except for $Up_{4/5}U$, formed intramolecular base stacking interactions in aqueous solutions as indicated by NMR. The conformation around the glycosidic angle in Np_nN's was found to be *anti*/*high anti* and the preferred conformation around the C4′–C5′, C5′–O5′ bonds was found to be *gauche–gauche* (*gg*). The ribose moiety in Np_nN's showed a small preference for the *S* conformation, but when attached to cytosine the ribose ring preferred the *N* conformation. However, no predominant conformation was observed for the ribose moiety in any of the dinucleotides. Molecular dynamics simulations of Ap_2A and Ap_4A Na^+ salts supported the experimental results. In addition, three modes of base-stacking were found for $Ap_{2/4}A$: α–α, β–β and α–β, which exist in equilibrium, while none is dominant. We conclude that natural, free Np_nN's (n = 2–5) at physiological pH exist mostly in a folded (stacked), rather than extended conformation, in several interconverting stacking modes. Intramolecular base stacking of Np_nN's does not alter the conformation of each of the nucleotide moieties, which remains the same as that of the mononucleotides in solution.

Introduction

Dinucleoside polyphosphates, or dinucleotides (Np_nN'; N, N′ = A, U, G, C; n = 2–7) Fig. 1, are naturally occurring ubiquitous compounds. They consist of a polyphosphate chain linked to a nucleoside, at the 5′-position, on both edges. Many important functions, both intra- and extracellular, are suggested for Ap_nAs, Ap_nGs and Gp_nGs, including involvement in proliferation processes, in conditions of cellular stress, regulation of enzymes, neurotransmission, platelet disaggregation and modulation of vascular tone.[1–6] Np_nN's may also function as inhibitors of enzymatic reactions, acting as competitive or noncompetitive inhibitors or as alternative substrates. For example, Ap_5A is an inhibitor of adenylate kinase (K_i = 2.9 × 10^{-8} M).[7] Various dinucleotides primarily inhibit enzymes from the kinase family, but also IMP dehydrogenase, endoribonuclease, adenylsuccinate synthetase, poly(ADP-ribose) polymerase, *etc.*[7] In some cases, dinucleotides may act as positive effectors, wherein an Np_nN' binds to an enzyme (such as GMP reductase or AMP deaminase), thus increasing the activity rate of the enzyme.[7] In addition, dinucleotides, such as Ap_4A, may be used as primers for DNA synthesis by DNA polymerase or for RNA polymerase.[7] Moreover, dinucleotides were also used as probes for analyzing biochemical processes. Thus, the Ap_5A–Zn^{2+} complex was used to reveal preferred pathways that create the configuration capable of proficient chemistry in the catalysis of adenylate kinase.[8]

One of the physiological roles of both purine and pyrimidine dinucleotides is extracellular signaling mediated by the ubiquitous nucleotide P2-receptors.[4,6] Dincleotides may also bind to adenosine receptors, P1 receptors.[1,2] Furthermore, dinucleotides are attractive as therapeutic agents due to their increased chemical and metabolic stability as compared to the corresponding mononucleotide P2-receptor ligands.[9] For example, several Np_nN' analogues have been administered in human clinical trials, including Ap_4A, for lowering blood pressure during anesthesia, and Up_4U and Up_4dC (P1-(uridine 5′-)-P4-(2′-deoxycytidine 5′-) tetraphosphate) for the treatment of dry eye disease and cystic fibrosis, respectively.[5,9]

Although there is a significant volume of data on the pharmacology and biochemistry of Ap_nAs (Gp_nGs)[2,3] and despite the increasing interest in the members of the dinucleotide family as therapeutic agents, the current information on their physicochemical properties is insufficient. For instance, the actual conformation of sodium or potassium salts of dinucleotides, in a physiological medium, was scarcely determined experimentally. Such data are required for the understanding of the structure–activity relationship of either natural or synthetic dinucleotides and the design of potent therapeutic agents based on a dinucleotide scaffold.

Previous conformational studies on dinucleotides in solution, mainly on Ap_nAs, focused on the base stacking interactions using CD, UV and ^{1}H NMR.[10–14] Yet the mode of π-stacking in Ap_nAs

Department of Chemistry and the Lise Meitner-Minerva Center of Computational Quantum Chemistry, Bar Ilan University, Ramat-Gan, 52900, Israel; Fax: +972-3-6354907; Tel: +972-3-5318303. E-mail: bfischer@mail.biu.ac.il

† Electronic supplementary information (ESI) available: Experimental details. See DOI: 10.1039/c005122e

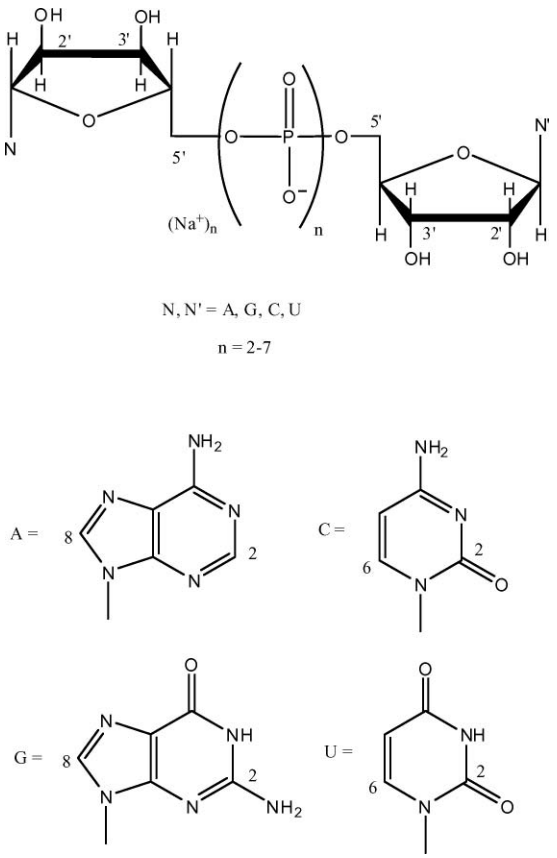


Fig. 1 Structures of dinucleoside 5′,5′- polyphosphate analogues.

(*i.e.* α–α, α–β or β–β stacking, Fig. 2) is still inconclusive.[15,16] Internal stacking interactions in other Np_nN's (N ≠ A) were studied but no conformational parameters (*e.g.* sugar puckering, α, β, γ, χ angles, Fig. 3) were provided.[17,18] Likewise, detailed data on the conformation of free dinucleotides in solution is limited mainly to Ap_2A.[18]

Crystal structures of several kinases or hydrolases-bound dinucleotides in enzymes were determined.[19–22] In protein-bound dinucleotides, the polyphosphate chain adopts various conformations: extended,[19–22] S-shaped [21] or folded.[22] In some cases the protein-bound dinucleotides coordinate with a Mg^{2+} ion. Nevertheless, crystal structures of free Np_nN's were hardly determined and a crystal structure is available only for Ap_4A sodium salt.[23]

Likewise, the conformation of Np_nN's was scarcely studied computationally. Specifically, only Ap_2A was studied and its stable conformations were calculated by energy minimizations.[24] Computational-based conformational studies were performed for related compounds, such as 3′,5′-dinucleotides,[25–28] sugar-nucleotides (UDP-glucose),[29,30] ATP-Mg^{2+},[31] and nucleosides.[32]

The absence of a large body of comprehensive data on the conformation of a large and diverse series of free dinucleotides prompted us to conduct a systematic conformational study employing NMR and computational techniques. Specifically, we analyzed the conformation of a diverse series of Np_nN' (N′ = A, U, G, C; n = 2–5) analogues **1–17** (Fig. 4) in a neutral aqueous solution by ^{1}H and ^{31}P-NMR. To augment the experimental data, we studied the conformation of Ap_2A and Ap_4A in an explicit water environment by molecular dynamics (MD) simulations combined with potential of mean force (PMF) calculations. The conformational analysis by both NMR and theoretical calculations included parameters such as ribose puckering (North↔South equilibrium), glycosidic bond angle (χ), α, β and γ angles (*e.g.*, Fig. 3) and the mode of intramolecular π-stacking interactions (*i.e.* α–α/β–β/α–β mode) between the nucleobases. *Investigating the conformations of the dinucleotides in solution is a prerequisite for understanding their activity as enzyme inhibitors and receptor ligands.*

Results

Selection of experimental and computational techniques and data analysis

To study the effect of both the nature of the nucleobase and the length of the polyphosphate linker on the conformation of dinucleotide sodium salts in aqueous solution, we selected a series of 17 dinucleotide analogues, Fig. 4. Among those 17 analogues we investigated more extensively the Ap_nAs (analogues **1–4**), since these are the most abundant dinucleotides in organisms.[9]

Our methods of choice for the investigation of the Np_nN' conformation were NMR techniques including ^{1}H-, ^{13}C-, and ^{31}P-1D NMR and COSY, HMQC, HMBC and NOESY 2D-experiments. Chemical shifts (δ_H, δ_C, δ_P) and coupling constants (J_{HH}, J_{CH}, J_{PH}, J_{CP}, J_{PP}) of Np_nN's spectra were measured. Data are tabulated in the Supplementary Material (Table S1†). The NMR data were analyzed by eqn (1)–(10) (in Experimental) and the following conformational features of Np_nN' analogues were determined: value of χ, dominant conformers around C4′–C5′ and C5′–O5′ bonds, and dominant ribose conformer, in addition to identification of intramolecular π–π stacking interactions. Furthermore, we used MD simulations combined with PMF calculations to study the conformation of two representative dinucleotides from the Np_nN' series, Ap_2A and Ap_4A, where additional information on torsion angles α, β and γ was obtained. The analysis of the MD data yields the conformational parameters of Ap_nAs as well as the occurrence and the mode of intramolecular π-stacking (*i.e.* α–α, β–β or α–β stacking). The PMF simulations provide free energy associated with the stacking process.

Synthesis of the studied dinucleotides

Briefly, nucleoside-5′-monophosphate (NMP) or nucleoside-5′-diphosphate (NDP) tributylammonium or trioctylammonium salts were first activated as the corresponding phosphoroimidazolide analogues, using CDI in DMF and then treated by NMP or NDP trialkylammonium salts.[9,33] Synthesis of Up_5U was achieved by activating UTP tributylammonium salt with DCC followed by the addition of UDP tributylammonium salt in DMF.[9,33] The dinucleotides were obtained in 10–60% yield. Dinucleotides with long phosphate linker (with 4–5 phosphates *e.g.*, Ap_4U, Up_5U) or some of the mixed dinucleotides (with 2–3 phosphates *e.g.*, Gp_2C) were obtained in yields lower than 10%.

NMR Results

Selection of NMR measurements conditions

We first established the dinucleotide concentration range suitable for the NMR experiments. The concentration range should be

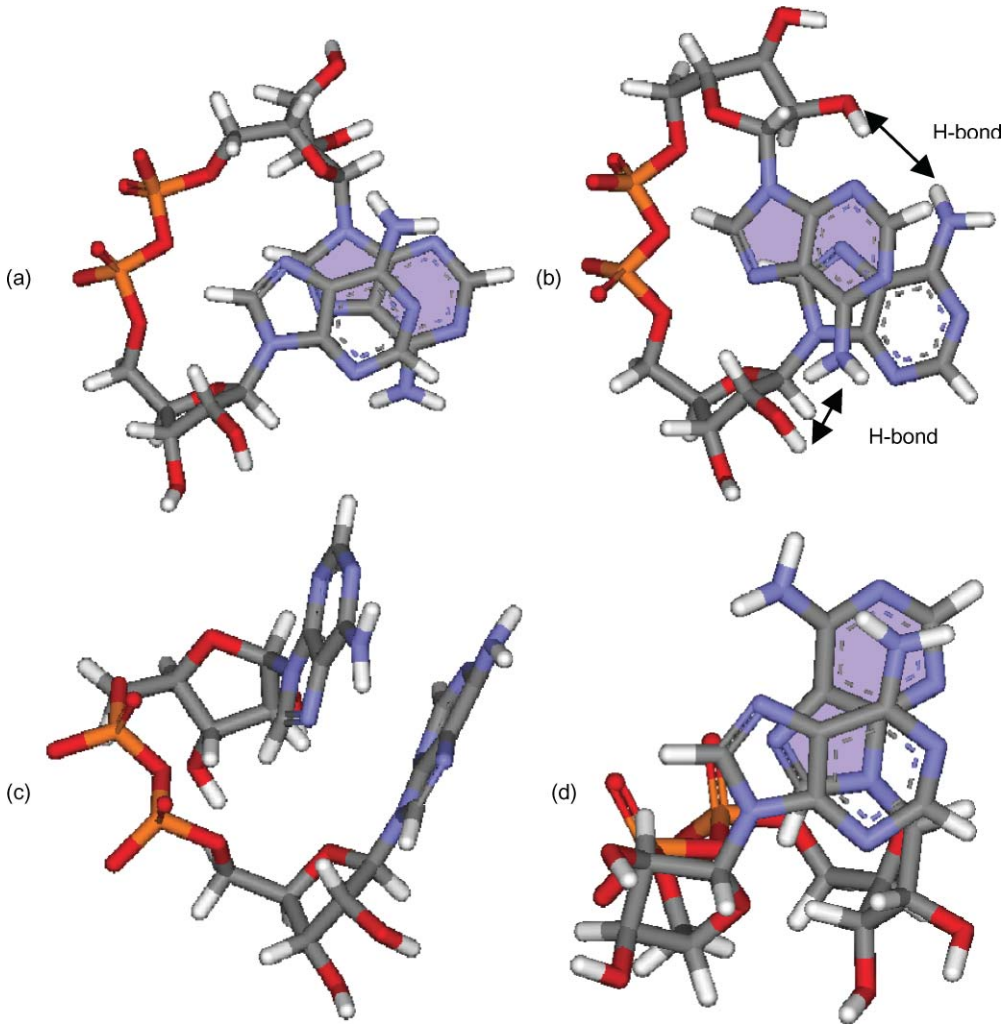


Fig. 2 Different stacking modes of Ap_2A. One base was colored purple for clarity. (a) β–β base-stacked mode (b) α–α stacked mode, stabilized by two intramolecular hydrogen bonds (black arrows) (c) α–β stacked mode. Side view. (d) α–β stacked mode. Top view.

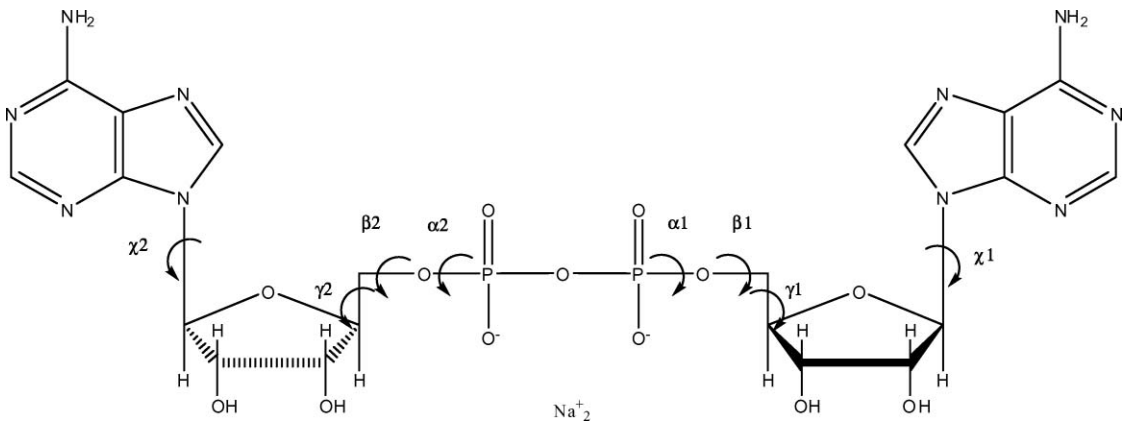
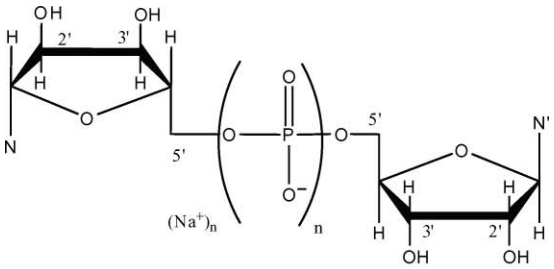


Fig. 3 Torsional angles of $Ap_{2/4}A$.

kept low enough to avoid inter-molecular base stacking (self association). Nucleotides tend to aggregate *via* π-stacking interactions in aqueous solutions.[34] However, at relatively low concentrations they remain in their monomeric form. Specifically, 97% of 9 mM ADP^{3-}, 16 mM GDP^{3-}, 22 mM CDP^{3-} and 26 mM UDP^{3-} exist in their monomeric form.[35] According to Mayo *et al.*,[18] only at dinucleotide concentrations above 20 mM are there concentration effects on the chemical shifts due to inter-molecular interactions. Therefore, we used 7.2–7.6 mM dinucleotide NMR samples. Several of our dinucleotides were measured at two concentration



Cmp. no.	Dinucleotide	N	N'	n
1	Ap_2A	A	A	2
2	Ap_3A	A	A	3
3	Ap_4A	A	A	4
4	Ap_5A	A	A	5
5	Ap_2U	A	U	2
6	Ap_3U	A	U	3
7	Ap_4U	A	U	4
8	Up_2U	U	U	2
9	Up_4U	U	U	4
10	Up_5U	U	U	5
11	Gp_2G	G	G	2
12	Gp_3G	G	G	3
13	Gp_4G	G	G	4
14	Gp_5G	G	G	5
15	Gp_2C	G	C	2
16	Gp_3C	G	C	3
17	Gp_2C'[a]	G	C'	2

[a] N^4-methyl carbamate cytosine.

Fig. 4 Dinucleoside polyphosphate (Np_nN') analogues investigated in this study.

ranges, 7.2–7.6 mM and 1.5–1.8 mM and their ^{1}H NMR spectra were compared. There were no significant differences in the coupling constants and chemical shifts ($\Delta\delta$ = 0.01–0.02 ppm) between both concentration ranges. For example, for Ap_4A $\Delta\delta_{H2} = \delta_{H2,(1.7\ mM)} - \delta_{H2,(7.4\ mM)} = 0.02$ ppm.

Assignment of the NMR signals and coupling constants of dinucleotides 1–17

Assignment of H1′, H2′, H3′, H4′, H5′, H5″ and ^{13}C NMR signals of the analogues **1–17** was based on 2D COSY, HMQC, HMBC and NOESY spectra. In several cases, assignment was achieved by comparison to mononucleotides.[36,37] For purine analogues, the ribose proton signals are always ordered as follows: δ H1′ > H2′ > H3′ > H4′ > H5′, H5″. However, for pyrimidine analogues, the ribose protons, except for H1′, appear in no specific order. The H5′ signal was assumed to be more downfield than that of the H5″ signal following Remin and Shugar's observations.[38] In most cases, parts of the spectra measured are of second order, especially for pyrimidine containing dinucleotides, and the obtained coupling constants are either not accurate or can not be obtained at all. Assignment of $^3J_{H4'H5'/H5''}$ and $^3J_{PH5'/H5''}$ was based on the J-resolved spectrum of Ap_4A and ^{31}P NMR of Ap_3A (see Supplementary Material†). The J-resolved spectrum for Ap_4A was used to analyze the spectra of other Np_nN's. The error of ^{1}H NMR coupling constants is $J \pm 0.2$ Hz. In cases where the spectra were of second order due to overlapping signals, the error increased to ±1 or ±2 Hz, as for H4′, H5′ and H5″ in Ap_2A, Up_2U and Gp_2G. Strongly coupled protons showed broadened signals and the error was *ca.* ±0.5 Hz, as for H3′ in Ap_2A and Gp_2G. The error for ^{13}C NMR and ^{31}P NMR coupling constants was $J \pm 0.5$ Hz. The error in the percentage of N/S conformers is ±2–3% for $J \pm 0.2$ Hz, which is true for most of the examined dinucleotides. However, for $J \pm 0.5$ Hz the error increases to ±5–6%, and for $J \pm 1$ Hz the error is *ca.* 11%. Thus, Up_2U has a larger error as compared to the other dinucleotides.

Analysis of NMR data of the Np_2N' series

π-stacking interactions. To study intramolecular π-stacking interactions in dinucleotides we compared the chemical shifts of the base protons of dinucleotides to those of the corresponding mononucleotides at the same concentration. An upfield shift of a dinucleotide nucleobase proton reflects that this proton is located above the other aromatic ring plane and is affected by the ring current.[18,39,40]

Base stacking was reported to be most dominant in Ap_2A, estimated as 60% of the population.[24] Here, we found that base stacking interactions are significant not only for Ap_2A, but also for all tested dinucleoside diphosphates (Np_2N') as compared to their homologues with longer phosphate linkers (Table 1). Specifically, upfield shifts with $\Delta\delta$ values of *ca.* 0.1–0.4 ppm for H2/H8 or H5/H6 for A/G or C/U, respectively, were observed as compared to the corresponding mononucleotides. We observed more significant upfield shifts of the aromatic base protons in Np_2N' as compared to dinucleotides with longer phosphate linker (for example, δ_H8: ATP>Ap_5A>Ap_4A>Ap_3A>Ap_2A). The effect of the polyphosphate chain length on base stacking was reported before for Ap_nAs and Gp_nGs, based on NMR and CD experiments.[10,13,18] Here, we show that π-stacking occurs also in Ap_nU and Gp_nC analogues (Table 1). However, for the longer Up_nU analogues, Up_4U and Up_5U, there is no clear evidence from NMR of base stacking, since the $\Delta\delta$ values are within the error range. $\Delta\delta$ values for Up_2U are larger than the error and indicate the existence of stacking interactions.

Ap_2A exhibited the largest $\Delta\delta$ values as compared to other dinucleoside diphosphates, whereas Up_2U exhibited the smallest $\Delta\delta$ values. For mixed dinucleotides (*e.g.* Ap_2U), a large effect of purine base on the pyrimidine protons was observed. However, only a small effect of the pyrimidine on purine protons was

Table 1 Differences in chemical shifts (in ppm) between dinucleoside polyphosphates and the corresponding mononucleotide. ($\Delta\delta = \delta_{mononucleotide} - \delta_{dinucleotide}$)

No.	Dinucleotide	$\Delta\delta_{H8}$	$\Delta\delta_{H2}$	$\Delta\delta_{H6}$	$\Delta\delta_{H5}$	$\Delta\delta_{H1'(pu)}$	$\Delta\delta_{H1'(py)}$
1	Ap_2A	0.37	0.24	—	—	0.20	—
2	Ap_3A	0.26	0.18	—	—	0.15	—
3	Ap_4A	0.17	0.14	—	—	0.13	—
4	Ap_5A	0.11	0.12	—	—	0.10	—
5	Ap_2U	0.10	0.06	0.34	0.31	0.07	0.19
6	Ap_3U	0.03	0.04	0.21	0.25	0.04	0.15
7	Ap_4U	0.01	0.03	0.14	0.14	0.03	0.08
8	Up_2U	—	—	0.07	0.04	—	0.05
9	Up_4U	—	—	0.02	–0.01	—	0.01
10	Up_5U	—	—	0.02	0	—	–0.01
11	Gp_2G	0.20	—	—	—	0.15	—
12	Gp_3G	0.14	—	—	—	0.11	—
13	Gp_4G	0.08	—	—	—	0.08	—
14	Gp_5G	0.06	—	—	—	0.06	—
15	Gp_2C	0.09	—	0.35	0.23	0.06	0.16
16	Gp_3C	0.03	—	0.27	0.20	0.04	0.11
17	Gp_2C'	0.15	—	–0.01	–0.81	–0.07	0.16

observed in Ap_2U and Gp_2C. The reason for those observations is the much smaller ring current of the pyrimidines *vs.* purines.[41,42] In the carbamate substituted cytosine in Gp_2C', **17**, downfield shifts were observed, especially in cytosine H5, as compared to CMP. This phenomenon is probably due to the electron withdrawing effect of the carbamate group. Nevertheless, there is an upfield shift for H8 in the guanine base which may be due to base stacking effect.

The ring current effect in dinucleotides influences not only the base protons, but also the anomeric H1′.[11] Indeed, we observed an upfield shift in the H1′ signal of most studied dinucleotides. In addition, there were also upfield shifts in the H2′ and H3′ signals (Supplementary Material, Table S2†).

Ribose puckering. We found that in most cases, for dinucleotide ribose attached to any of the bases (A,G,U,C) there was no pre-dominant conformation (*i.e.* Northern, *N*, or Southern, *S*, conformation), Table 2. For adenine and guanine dinucleotides there is a slight preference for the *S* conformer (*ca.* 55%), whereas, for the unsymmetrical dinucleotides (Ap_2U, Gp_2C) *ca.* 65% preference was observed.

Uracil base attached to ribose induces only a slight preference for the *S* conformer in the ribose (54% *S* in Ap_2U). Up_2U exhibited almost no $N \leftrightarrow S$ preference. Ribose attached to cytosine prefers the *N* conformer (60%). N^4-Carbamate substituted C induced a greater preference for the *N* conformer (71%). This preference of the *S* conformer in purine linked ribose and *N* conformer in pyrimidine linked ribose is similar to that of the corresponding mononucleotides.[43]

Conformation around the ribose –CH_2O– exocyclic group. The percentage of *gg*, *gt* and *tg* conformer populations around the ribose –CH_2O– exocyclic group in dinucleotides **1**, **5–12**, **15–17** could not be obtained from eqn (3)–(8), since the required coupling constants could not be determined accurately from ^{1}H NMR data due to second order spectra in most cases. However, the dominant conformer could sometimes be determined by examining the $J_{4'5'/5''}$ values and comparing them to the values of the pure rotamers. The coupling constants for pure rotamers were estimated as J_g = 2.04 Hz (*gauche* rotamer) and J_t = 11.72 (*trans* rotamer) for C4′–C5′ bond and J_t = 20.9 Hz and J_g = 1.8 Hz for C5′–O5′.[44] Hence, for Ap_2A, Gp_2G and Up_2U the conformation around C4′–C5′ is *gg* ($^3J_{4'5'/5''}$ = 2.1–2.7 ± 1 Hz). For the ribose attached to U in Ap_2U, only one *J* value, $^3J_{4'5''}$ = 2.5 ± 1 Hz, was extracted from the spectrum. Therefore, we conclude that H5″ is *gauche* to H4′ (Fig. 5).

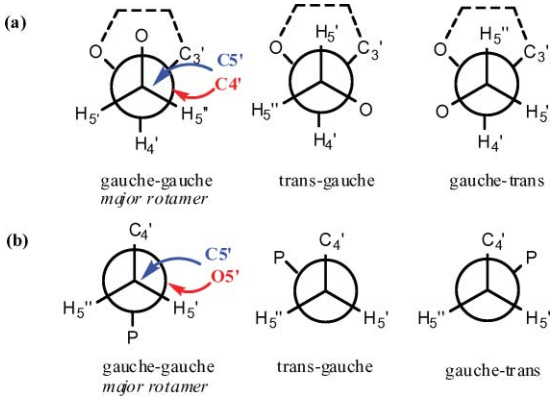


Fig. 5 Staggered rotamers around (a) C4′–C5′ and (b) C5′–O5′ bonds.

Moreover, some information may be obtained from $^3J_{C4'P}$ coupling constants about the conformation around the C5′–O5′ bond.[43] We used ATP as a reference for determining rotamers around the C5′–O5′ bond in dinucleotides. Thus, for ATP a population of 63% *gg* around C5′–O5′ was calculated applying eqn (6)–(8) using $^3J_{4'5'}$ and $^3J_{4'5''}$. In addition, a coupling constant $^3J_{C4'P}$ of 9.2 Hz was extracted for ATP from the ^{13}C NMR spectrum. For Ap_2A, Up_2U and Gp_2G we measured the following $^3J_{C4'P}$ values (±0.5 Hz): 9.0, 9.4, 9.5 Hz, respectively. Therefore, based on comparison to $^3J_{C4'P}$ in ATP, we concluded that the *gg* rotamer is the predominant one also for dinucleotides. For Ap_2U there was a slight difference in $^3J_{C4'P}$: 8.2 Hz for the adenine moiety and 7.6 Hz for the uracil moiety. This may imply that the rotamer population percentages for Ap_2U are somewhat different than those of the other mentioned Np_2N's, namely, Ap_2A, Up_2U, Gp_2G and Gp_2C.

Glycosidic angle. We calculated the glycosidic angle for Np_2N's based on $^3J_{C8/6\text{-}H1'}$ and $^3J_{C4/2\text{-}H1'}$ (eqn (9)–(10)). Each

Table 2 Conformational data obtained for the studied dinucleoside polyphosphates

No.	Compound	χ degrees ($^3J_{C8/6\text{-}H1'}$)	χ degrees ($^3J_{C4/2\text{-}H1'}$)	% Puckering	C4′–C5′			C5′–O5′		
					% *gg*	% *tg*	% *gt*	% *gg*	% *tg*	% *gt*
1	Ap_2A	213/267	213/267	55 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
2	Ap_3A	206/274	n.d.[b]	51 (N)	84	5	11	64	23	13
3	Ap_4A	218/262	n.d.	62 (S)	79	8	13	63	22	15
4	Ap_5A	219/261	213/267	65 (S)	79	8	13	64	23	13
5	$\boldsymbol{A}p_2U$	214/266	226/254	63 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
5	$Ap_2\boldsymbol{U}$	210/270	218/262	54 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
6	$\boldsymbol{A}p_3U$	214/266	226/254	64 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
6	$Ap_3\boldsymbol{U}$	211/269	213/267	52 (N)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d
7	$\boldsymbol{A}p_4U$	211/269	226/254	67 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
7	$Ap_4\boldsymbol{U}$	211/269	218/262	55 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
8	Up_2U	210/270	218/262	52 (N)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
9	Up_4U	210/270	218/262	59 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
10	Up_5U	210/270	220/260	60 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
11	Gp_2G	211/269	226/254	57 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
12	Gp_3G	211/269	230/250	57 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
13	Gp_4G	211/269	230/250	67 (S)	67	15	18	61	22	17
14	Gp_5G	214/266	[a]	69 (S)	70	15	15	60	23	17
15	$\boldsymbol{G}p_2C$	210/270	[a]	65 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
15	$Gp_2\boldsymbol{C}$	205/275	n.d.	57 (N)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
16	$\boldsymbol{G}p_3C$	211/269	240	69 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
16	$Gp_3\boldsymbol{C}$	204/276	204/276	62 (N)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
17	$\boldsymbol{G}p_2C'$	202/278	n.d.	61 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
17	$Gp_2\boldsymbol{C}'$	192/288	n.d.	71 (N)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
	ATP	210/270	198/282	64 (S)	82	8	10	63	23	14
	UTP	210/270	218/262	55 (S)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
	GTP	221/259	240	67 (S)	70	15	15	61	25	14
	CMP	210/270	213/267	61 (N)	84	7	9	73	10	17

[a] Solutions of the glycosidic angle are out of the expected range, possibly due to error of measurement (±0.5 Hz of $^3J_{C8/6\text{-}H1'}$ and $^3J_{C4/2\text{-}H1'}$). [b] n.d. not determined due to second order spectrum, The percentage of preferred conformer refers to the ribose the base of which appears in bold italics. (*e.g.* $\boldsymbol{A}p_2U$ refers to the ribose attached to adenine).

equation gives four (sometimes two) solutions. One solution in the *anti* region, one in the *high anti* region, one is in the *syn* region, and the last one is *syn* or *anti*. We selected the solutions that are in the *anti* and *high anti* regions as the most reasonable possibility, considering the NOESY and the MD results, mentioned below (Table 2). The range of glycosidic angles for analogues **1**, **5**, **8**, **11**, **15**, **17** was 192–226° (*anti*)/254–288° (*high-anti*). The error in angle values ranges from ±6° to ±22°.

Though the χ angles reported for mononucleotides were 200 ± 20° [43] and one of the solutions obtained here for χ of dinucleotides was in the range of 210 ± 20°, our MD results showed larger χ angle values, in the *high anti* region, consistent with another solution resulting from NMR data. Therefore, we concluded that the χ angle range in dinucleotides is in the *anti*/*high anti* range.

NOESY spectra provided additional evidence for the *anti* conformation of Np_2N' analogues (Fig. 6–7). Clear cross peaks between H8–H1′, H8–H3′, H8–H2′ for the purine base moiety supported the *anti* conformation in Ap_2A and Gp_2G. Likewise, one cross peak between H8–H3′ for the adenine base in Ap_2U and for the guanine base in Gp_2C, and H6–H2′/3′, H6–H1′ for the pyrimidine base moiety, supported the *anti* conformation in Up_2U, Ap_2U, Gp_2C and Gp_2C'.

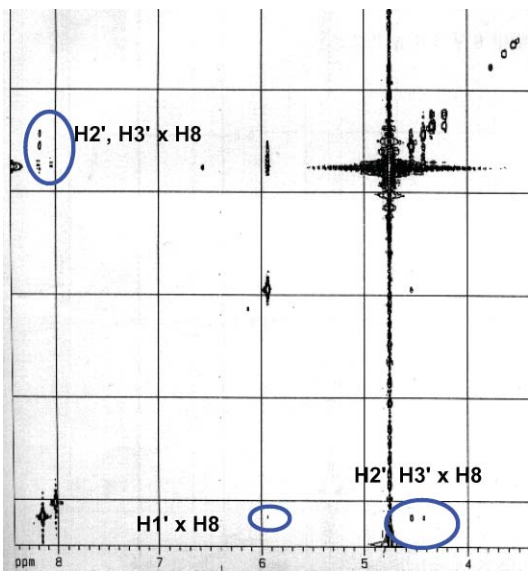


Fig. 6 NOESY spectrum of 7.4 mM Ap_2A in D_2O, pD 7.32, 300 K at 600 MHz.

Analysis of NMR data of the Np_3N' series

The upfield shifts of the nucleobase protons in Np_3N' series indicate base stacking interactions in this series, yet, to a smaller extent than those in the corresponding Np_2N' series (Table 1). In Ap_3U and Gp_3C there is a large effect of purine base on the pyrimidine protons, but only a small effect of the pyrimidine on purine protons, as observed for Ap_2U and Gp_2C.

Among the Ap_nAs (n = 2–5), Ap_3A has the largest percentage of *N* conformer. The dominant conformer around both C4′–C5′ and C5′–O5′ is *gg*, 84% and 64%, respectively (Table 2). In the

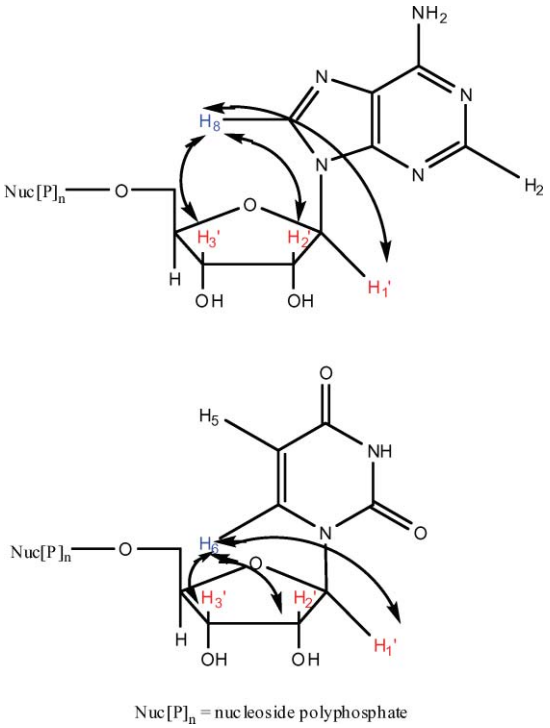


Fig. 7 NOE interactions in dinucleotides in the *anti* conformation.

Gp_nG series, Gp_3G prefers the *S* conformer (57%), as observed for Gp_2G.

The predominant conformation of the adenylated ribose in Ap_3U is *S* (64%), as observed for Ap_2U (64%). Yet, the uracilated ribose in Ap_3U slightly prefers the *N* conformer (53%). Among Ap_nUs, $n = 2$–4, Ap_3U has the highest percentage of *N* conformer in the uracilated ribose. There is a significant difference in the conformation of the adenylated ribose in Ap_3U compared to Ap_3A: 36% *S vs.* 51% *N*, respectively. Yet, the origin of the effect of the neighboring U *vs.* A on the adenylated ribose is not clear.

We used $^3J_{C4'P}$ for deducing the conformation around C5′–O5′, since $^3J_{P5'/5''}$ could not be extracted for most of the Np_3N' series. We found that the conformation around C5′–O5′ in Ap_3A, Gp_3G and Ap_3U is *gg* for both ribose moieties based on comparison with the UTP, GTP and ATP $^3J_{C4'P}$ values (Supplementary Material, Table S1†).

The conformation around the glycosidic bond in the Np_3N' series is *anti*/*high anti* according to $^3J_{C8/6\text{-}H1'}$ and $^3J_{C4/2\text{-}H1'}$ values, and NOESY data (H8–H3′, H8–H2′ and H8–H3′ cross peaks were observed for the adenine moiety in Ap_3A and Ap_3U, respectively) (Table 2).

The Gp_3C ribose conformation is similar to that of Gp_2C. The G/C bases are in *anti*/*high anti* conformation (Table 2). Comparison of $^3J_{C4'P}$ values for mononucleotides and Gp_3C (8.7 and 9.3 Hz for guanosine and cytidine moieties, respectively), shows that the conformation around C5′–O5′ in Gp_3C is *gg*.

It is noteworthy that the chemical shifts of the α phosphates of Np_3N's are shifted upfield as compared to the other Pα chemical shifts (*e.g.* $\delta_P\alpha$ (Ap_2A) = −10.39, $\delta_P\alpha$ (Ap_4A) = −10.42, $\delta_P\alpha$ (Ap_5A) = −10.44, *vs.* $\delta_P\alpha$ (Ap_3A) = −10.75 ppm). This shift of Pα signal in Ap_3A may be due to a change in the phosphate chain conformation.[45]

Analysis of NMR data of the Np_4N' series

For the Np_4N' series base stacking interactions were still observed, although weaker than those of the corresponding Np_3N' series, except for Up_4U for which stacking interactions could not be determined by NMR due to a small ring current effect (Table 1).

Unlike the X-ray crystal structure of Ap_4A which adopts the *N* conformer,[23] Ap_4A in solution prefers the *S* conformer (62%), Table 2. Up_4U in solution prefers the *S* conformer as well (59%). However, the dominant Ap_4A conformer in solution is *gg* around both C4′–C5′ (79%) and C5′–O5′ (63%), as found in the crystal structure.[23] Gp_4G has, as Ap_4A, a larger preference for the *S* conformer (67%) as compared to $Gp_{2/3}G$, Table 2. Likewise, the preferred conformer is *gg* around both C4′–C5′ (67%) and C5′–O5′ (61%). Both ribose moieties of Ap_4U are preferentially *S* (67% in adenylated ribose, and 55% in the uracilated ribose). Uracilated ribose in Ap_4U has about the same conformation as Ap_2U (54% *S*), unlike Ap_3U (52% N).

Conformation around the exocyclic group (C5′–O5′) was not determined for most Np_4N's due to the second order ^{1}H NMR spectrum. However, $^3J_{C4'P}$ data show that for adenylated ribose in Ap_4A and Ap_4Us the C5′–O5′ conformation is *gg* as in ATP. Uracilyated ribose in Up_4U and Ap_4U conformation is also *gg* based on $^3J_{C4'P}$ (Supplementary Material, Table S1†).

The conformation around glycosidic angle in the Np_4N' series is *anti* or *high anti* around both glycosidic bonds, according to eqn (9)–(10) (Table 2) and NOESY data (showing H8–H3′ cross peak for Ap_4A and Ap_4U, and H6–H2′, H6–H3′ cross peaks for Ap_4U and Up_4U).

Analysis of NMR data of the Np_5N' series

In the Np_5N' series, the base stacking interactions are least significant of all dinucleotides investigated here. Generally, the chemical shifts of Np_5N base signals are highly similar to those of free mononucleotides (*e.g.* ATP), indicating a negligible percentage of stacked species. For Ap_5A the upfield shifts ($\Delta\delta$) are 0.12 ppm for H2, and 0.11 ppm for H8, as compared to ATP. For Gp_5G the upfield shift is negligible, *i.e.*, $\Delta\delta_H8 = 0.06$ ppm as compared to GTP H8 signal. Likewise, for Up_5U there is no upfield shift of H5, and $\Delta\delta$ of only 0.02 ppm was observed for H6 as compared to UTP H6 signal. Therefore, we assume that stacking interactions probably still occur in Ap_5A and Gp_5G, but there is no clear evidence for their occurrence in Up_5U.

In Ap_5A and Gp_5G the dominant conformer is *gg* around both C4′–C5′ (79% and 70% respectively) and C5′–O5′ (64% and 60% respectively) bonds, as in Ap_4A and Gp_4G. Population analysis for Up_5U could not be done due to missing coupling constants. However, the $^3J_{C4'P}$ constant of 9.2 Hz in Up_5U provides evidence for a *gg* conformation of C5′–O5′, as in mononucleotides.

Np_5N's are similar to the corresponding mononucleotide and the Np_4N' analogues also in the sugar puckering. Ap_5A, Gp_5G and Up_5U prefer the *S* conformer (60–69%).

The conformation around the glycosidic bonds in Gp_5G, Ap_5A, Up_5U is *anti* or *high anti*, according to the Eq. 9–10 and NOESY spectra. Thus, for instance, we observed cross peaks between H8–H3′, H8–H2′ for Ap_5A, and H6–H3′ for Up_5U.

Molecular dynamics results

We chose Ap_2A and Ap_4A for our computational investigation since they both belong to the most abundant and studied series of Ap_nAs. These two analogues helped us evaluate the dependence of the dinucleotide conformation in solution on the length of the phosphate linker.

To verify the appropriateness of the adenine nucleotide CHARMM27 force field parameters in the case of adenine dinucleotides, we performed crystal simulations, starting with the experimental structure as input.[23] According to the simulation results, there is good agreement between the experimental and simulated molecular structure and cell unit parameters of Ap_4A (Fig. 8). Ap_4A experimental crystal parameters (a, b, c, α, β, γ) are 12.75 Å, 20.27 Å, 8.56 Å, 90.00°, 90.38°, and 90.00°, and the parameters after minimization are 12.46 Å, 19.59 Å, 8.52 Å, 90.00°, 91.98°, and 90.00°. The root mean square deviation (RMSD) between the crystal structure and the structure after minimization was 0.293 Å for all the dinucleotide atoms, including hydrogens (excluding water and ions). The average unit cell values obtained during 1 ns of MD simulation are $a = 12.09 \pm 0.17$ Å, $b = 19.38 \pm 0.23$ Å, $c = 9.14 \pm 0.18$ Å, angle $\beta = 88.23 \pm 2.05°$. The graph showing the changes of a, b, c and β during the 1 ns simulation is shown in the Supplementary Material, Fig. S1–S2.† The average RMSD during simulation is 1.293 Å. The structural integrity observed during the crystal minimization and MD simulation serves as an indication of the appropriateness of the nucleotide force field parameters for dinucleotides.

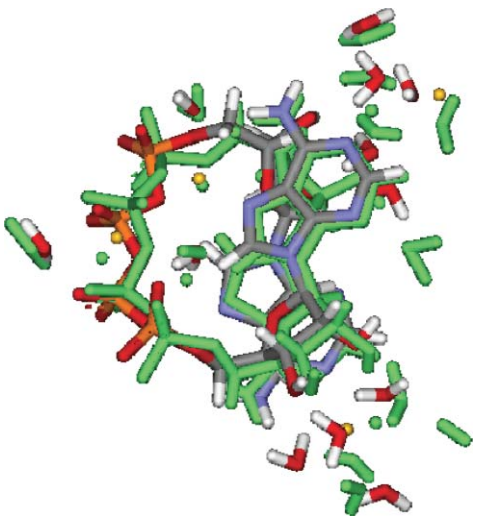


Fig. 8 Conformations of Ap_4A. A. Crystal structure (green). B. After 1 ns molecular dynamics at 293 K (other colours).

Intramolecular base stacking and hydrogen bonding in $Ap_{2/4}A$ sodium salts

Initially, to obtain qualitative information regarding the π-stacking in the adenine dinucleotides, we performed 10 ns of unbiased MD simulations (termed *free MD* below). To differentiate between open and π-stacked conformations, we defined a N9–N92 distance <6.5 Å (Fig. 9) as a stacked conformation. The somewhat large stacking limit of 6.5 Å was chosen since the base planes could be close although the N9–N92 atoms relatively far

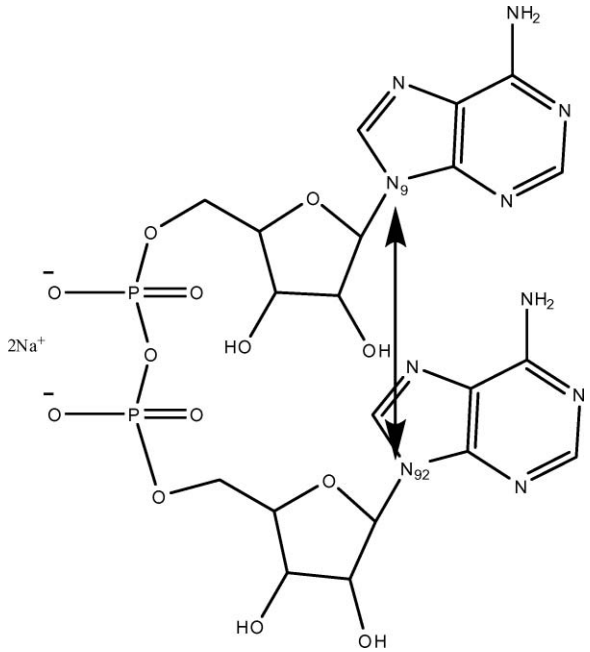


Fig. 9 The reaction coordinate, N9–N92 distance, used for examination of $Ap_{2/4}A$ base-stacking interactions.

apart. This was indeed observed in numerous snapshots. Thus, Ap_2A was found to be stacked 77% of the simulation time, while Ap_4A was stacked 70% of the simulation time (see Supplementary Material, Fig. S3†). During the course of the simulations, several occurrences of stacked conformations which opened up and subsequently re-stacked were observed.

Previous CD studies of base-stacking in dinucleotides (Ap_nA n = 2–6 as well as Gp_nGs n = 2–6 and mixed dinucleotides) proposed that there are two stereo-configurations: α–α and β–β (Fig. 2).[16] Specifically, Ap_2A was suggested to be preferentially stacked in the β–β mode, unlike other Ap_nAs which were assumed to be stacked in the α–α mode.[10,11,16] Previous reports suggested that Ap_2A can have several stacking modes, including α–α, β–β, and α–β modes.[24] It was also suggested that the α–α mode possibly enables hydrogen bonding between the amine nitrogen of each adenine ring and the C2′–OH of the opposite ribose residue for $Ap_{3/4}A$.[16]

To obtain qualitative information regarding the base stacking and its mode in the adenine dinucleotides, we performed PMF simulations for Ap_2A and Ap_4A in the α–α, β–β and α–β modes. In Fig. 2a–d we show representative stacked conformations of Ap_2A in the α–α, β–β, and α–β modes. We differentiate between α–α, β–β, and α–β stacking by visual inspection of frequent snapshots from the MD trajectory.

For Ap_2A and Ap_4A there is a preference for the stacked conformation over the extended one, regardless of the stacking mode (Fig. 10). Moreover, there is no free energy barrier separating the extended and folded forms. For Ap_2A, the α–α, β–β and α–β modes have stacking free energies of 3.1, 1.5 and 3.2 kcal mol^{-1}, respectively. This finding of several equilibrating conformers of Ap_2A is in agreement with previous studies by Thornton *et al.*[24] and Pettegrew *et al.*[15] Yet, our finding is not consistent with CD data, which suggested that Ap_2A is only β–β stacked.[10,16] This point will be addressed further in the Discussion. Ap_4A has also several stacked conformers in equilibrium. Specifically, the α–α,

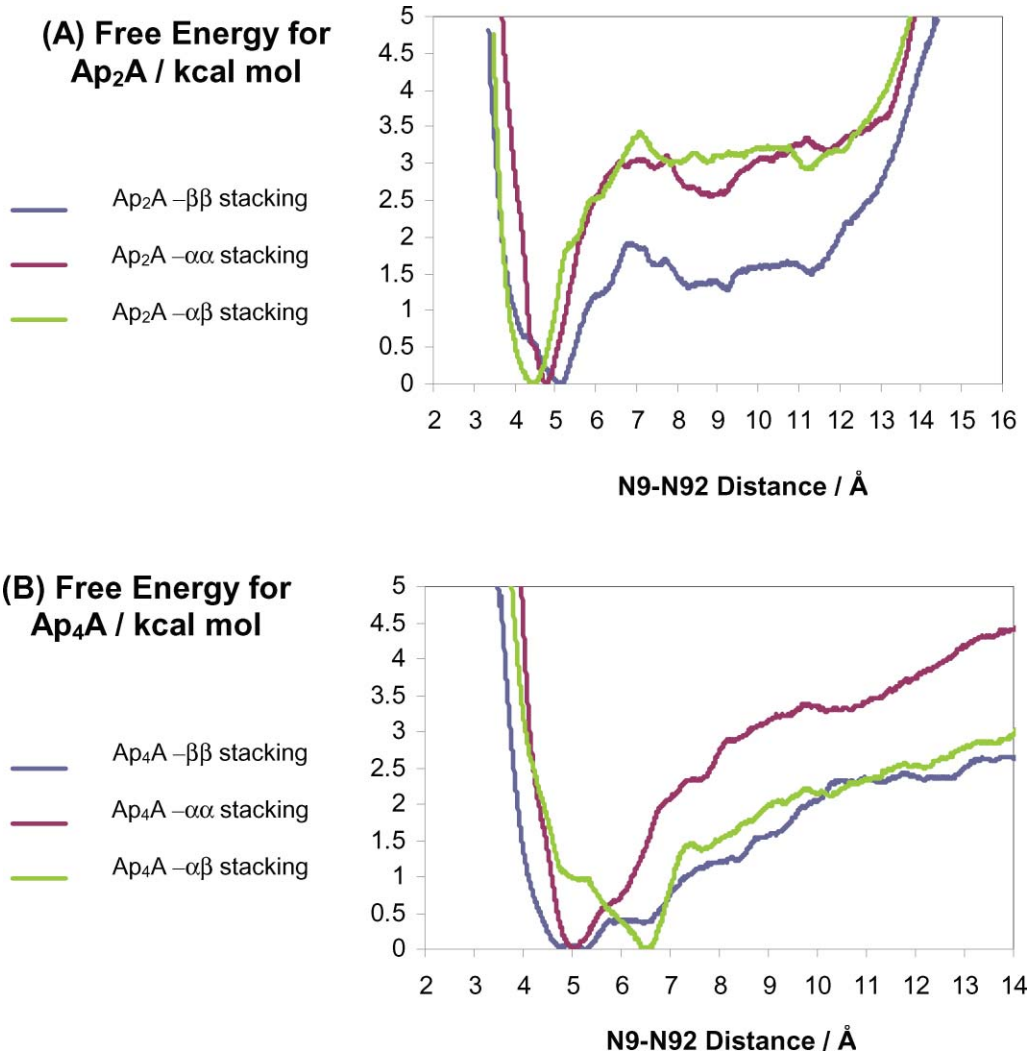


Fig. 10 Potential mean force (kcal mol^{-1}) for the base stacking in Ap$_n$A: (A) Ap$_2$A (B) Ap$_4$A in α–α mode (red line) the β–β mode (blue line) and the α–β mode (green line) as a function of the distance N9–N92 between the two adenine moieties.

β–β and α–β modes have stacking free energies of 3.3, 2.3 and 2.4 kcal mol^{-1}, respectively.

Based on previous suggestions[16] we examined the existence of intramolecular hydrogen bonds between the N^6 amine and the ribose O2′ groups in Ap$_2$A and Ap$_4$A. The N^6–O22′ and N^{62}–O2′ distances (Fig. 11) were measured for the adenine dinucleotides in α–α, β–β, and α–β stacking modes based on PMF simulation windows in which the bases were stacked. The distances were averaged over *ca*. 2000 structures collected during 1 ns of MD simulations. For Ap$_2$A in the α–α mode the N^6–O22′ and N^{62}–O2′ distances were 3.40 ± 0.54 Å and 3.51 ± 0.61 Å, whereas for β–β the distances were 6.90 ± 1.25 Å and 8.58 ± 1.76 Å. For the α–β mode these distances were 7.24 ± 0.65 Å and 7.37 ± 0.84 Å. This indicates that in the α–α stacked mode, Ap$_2$A forms intramolecular hydrogen bonds, while in the β–β and α–β modes no hydrogen bonds are formed. For Ap$_4$A the distances were 3.72 ± 0.89 Å and 3.16 ± 0.40 Å for α–α, whereas for β–β 5.91 ± 0.54 Å and 5.58 ± 0.61 Å, and for α–β 11.54 ± 0.98 Å and 4.74 ± 0.63 Å. As for Ap$_2$A, there was a preference for the hydrogen bonded form in the α–α stacked Ap$_4$A (Supplementary Material, Fig. S4–S9 for all Ap$_{2/4}$A species†). However, in other simulations the Ap$_4$A α–α form was not always hydrogen bonded.

From the PMF for the adenine dinucleotides we can also obtain the average minimum N9–N92 stacking distance. For Ap$_2$A the stacking distances are 4.8 Å, 5.1 Å and 4.5 Å for α–α, β–β and α–β, respectively, while for Ap$_4$A the stacking distances are 5.0 Å, 5.3 Å and 6.6 Å for α–α β–β and α–β , respectively. In the Ap$_4$A α–β structure the average N9–N92 distance is large due to the relative approach of the two bases in this case, although the base planes are close and stacking clearly occurs.

Study of conformations in stacked *vs.* extended forms of Ap$_{2/4}$A in the various stacking modes during PMF simulations

The average χ values are in the *anti* or *high anti* range for all forms (Tables 3–4). Specifically, there is a tendency towards the *high anti* for the extended conformations of Ap$_{2/4}$A. In addition, both χ angles for Ap$_4$A β–β are in the *anti* range, while the χ angles of Ap$_4$A α–α and α–β are closer to the *high anti* range . The largest χ values are for the α–β mode, with one angle in the *high anti* range.

Table 3 Torsion angles of Ap_2A during 1 ns simulation at N9–N92 distances of 4.5 (in 3 stacking modes) and 9.0 Å (extended form)

	Ap_2A α–α (°)	Ap_2A β–β (°)	Ap_2A α–β (°)	Ap_2A ex. (°)
$\alpha 1$	180.4 ± 21.9	281.7 ± 44.7	77.9 ± 16.0	210.6 ± 77.2
$\alpha 2$	184.6 ± 22.4	141.2 ± 90.0	292.2 ± 12.6	152.3 ± 110.0
$\beta 1$	190.8 ± 31.2	151.0 ± 26.3	241.1 ± 10.5	185.8 ± 26.2
$\beta 2$	189.7 ± 25.9	208.1 ± 27.2	168.0 ± 13.5	192.4 ± 25.7
$\gamma 1$	64.6 ± 13.3	54.2 ± 13.4	83.4 ± 11.9	61.0 ± 15.0
$\gamma 2$	62.3 ± 15.1	62.9 ± 23.5	58.6 ± 14.4	60.0 ± 16.2
$\chi 1$	230.7 ± 14.9	241.0 ± 30.2	267.3 ± 11.9	263.3 ± 28.7
$\chi 2$	229.9 ± 13.7	253.1 ± 22.6	304.0 ± 15.4	267.5 ± 40.5

Table 4 Torsion angles of Ap_4A during 1 ns simulation at N9–N92 distances of 4.5 (in 3 stacking modes) and 9.0 Å (extended form)

	Ap_4A α–α (°)	Ap_4A β–β (°)	Ap_4A α–β (°)	Ap_4A ex. (°)
$\alpha 1$	144.4 ± 31.4	282.0 ± 30.3	72.8 ± 18.1	294.4 ± 26.4
$\alpha 2$	108.3 ± 51.1	278.4 ± 19.0	292.4 ± 84.7	−34.4 ± 25.2
$\beta 1$	213.1 ± 31.3	149.8 ± 19.5	226.9 ± 10.7	139.8 ± 20.4
$\beta 2$	276.4 ± 53.9	112.8 ± 15.0	111.7 ± 13.2	164.2 ± 18.0
$\gamma 1$	65.5 ± 14.8	48.9 ± 15.7	69.0 ± 11.0	41.9 ± 14.6
$\gamma 2$	72.4 ± 17.8	47.0 ± 16.0	45.6 ± 15.7	62.4 ± 16.0
$\chi 1$	255.2 ± 19.7	215.3 ± 17.5	257.5 ± 28.7	256.8 ± 15.3
$\chi 2$	261.2 ± 18.3	222.4 ± 17.3	258.4 ± 22.3	275.0 ± 43.4

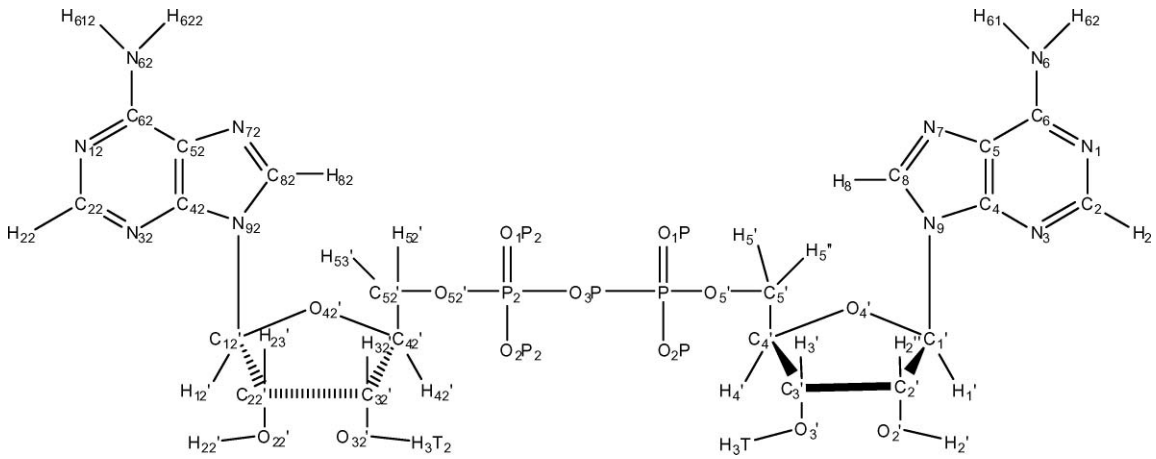


Fig. 11 Atom type definition for Ap_2A.

The average angle γ is in the *gg* range (60.0 ± 30.0°, Tables 3–4). The average angle β is in the *gg* range in most cases (180.0 ± 30.0°). Additionally, in the 10 ns simulation the average β fell mostly within the *gg* range (Supplementary Material, Table S3†). The α angle varies many times during the simulations, and it has a broad set of values with larger standard deviations. The α values may be similar for both ribose moieties (*e.g.* in the Ap_2A α–α mode the α values remain *ca.* 180°) or different (as in the Ap_2A α–β mode) (Tables 3–4).

Inspection of the results from the two 10 ns free MD simulations revealed that both *N* and *S* conformations are present for Ap_2A and Ap_4A. In agreement with the NMR data we find there is a preference for the *S* conformer over the course of 20 ns of MD simulations: The average values for the two ribose moieties in Ap_2A and Ap_4A are 84% and 92%, respectively. The preference for the *S* conformer is also found in the restrained 1 ns PMF simulations for both Ap_2A and Ap_4A (80% and 88% respectively) as in the free MD simulations. However, during 1 ns simulation in the Ap_2A α–β structure, one ribose preferred the *S* conformer and the other preferred the *N* conformer. These findings are in agreement with NMR data showing that the preference for the *S* conformer is larger in Ap_4A than in Ap_2A, although the difference is small. The dominance of the *S* conformer remained even when the simulations were extended for additional 40 ns (not shown). We note that there is a difference between the NMR and simulation *S*/*N* populations which probably results from insufficient simulation time and inaccuracies in the potential energy surface.

The effect of sodium ions was also examined by measuring the distance between the sodium ions and phosphorous atoms of $Ap_{2/4}A$ during free simulations of lasting 50 ns. The average distances were all larger than 11 Å and therefore we conclude that the sodium ion's role is neutralization of charge and it does not have an effect on the conformation of dinucleotides.

Crystal structure of Ap_4A *vs.* structure in solution

Comparison to the crystal structure of Ap_4A shows that there are differences compared to the structure in solution. The modes of stacking and the puckering are different. Specifically, in the crystal structure the stacking mode is β–β and the ribose prefers the *N* conformer while in solution the stacking mode is α–α and the preferred ribose conformer is *S*. A similar difference in puckering exists for ATP.[46] The torsion angles β and γ in the crystal structure are in the *gg* range, as was observed by NMR and in the PMF simulations. The glycosidic angle χ is in the *anti* range for the crystal structure, and is in the *high anti* (α–α)/*anti* (β–β) range according to MD simulations. The torsion angles α, β, γ, χ resulting from the PMF simulation for Ap_4A β–β have very similar values to those of the crystal structure (See Table 4, α angles: 293°, 295°, β angles: 166°, 158°, γ angles: 47°, 47°, χ angles: 210°, 208°).

Discussion

Dinucleotides in an aqueous solution exist as a mixture of equilibrating species most of which are in the folded conformation.

In this study, we addressed the question of the preferred conformations of dinucleoside polyphosphate (Np_nN'; N, N′ = A, U, G, C; n = 2–7) sodium salts in an aqueous solution. In particular, we studied whether the dinucleotides exist in the extended or folded (stacked) conformations.

Nucleic acids, nucleosides and nucleotides form π–π stacking interactions through their bases in water.[41–43,47] The base stacking that results from π–π interactions is assumed to be stabilized mainly by London dispersion forces (momentary dipole-induced dipole) and hydrophobic effects.[25] Several reports suggest that London dispersion forces are the most significant interaction,[48] while others propose that hydrophobic effects make the most significant contribution.[49] Additional minor factors that influence stacking are electrostatic interactions and the de-solvation effect.[49,50] In addition, the polyphosphate linker may influence the stacking of dinucleotide bases. The stacking interactions may occur inter- and/or intra-molecularly. While in concentrated solutions we expect intermolecular stacking, as in mononucleotides,[18,42,51] in diluted solutions, such as the physiological medium, intramolecular stacking is assumed to occur.[18]

Based on extensive NMR studies for dinucleotides **1–17** we conclude that the intramolecularly base-stacked conformation exists for all dinucleotides studied here in neutral aqueous solution, except for $Up_{4/5}U$, for which base stacking interactions could not be determined by NMR.[41] Intramolecular base stacking is supported by MD simulations for the Ap_2A and Ap_4A dinucleotides, which showed a preference of 1.5–3 kcal mol^{-1} for the π-stacked configurations (*i.e.*, equilibrating α–α, β–β, and α–β, stacked species). Indeed, intramolecular π-stacking interactions were also observed by similar PMF simulations for related dinucleoside monophosphate analogues, *e.g.* adenylyl-3′-5′-adenosine (3′-5′-ApA).[27] Specifically, dinucleotides with at least one purine base showed favorable stacking interactions with a ΔG difference of ≥2 kcal mol^{-1} between the stacked and unstacked state.[52]

A previous study on base-stacking interactions in Ap_nAs and Gp_nGs, applying CD,[10,11,16] proposed that Ap_2A adopts the β–β stacking mode, while $Ap_{3/4/5}A$ prefer the α–α mode, in contrast to our results suggesting no preferred stacking modes. However, the CD study was based on several unsettled assumptions about the transition moments in adenine.[16] In addition, in that study, the existence of only one stacking mode for each dinucleotide was assumed. Our findings are consistent with a CD study on Ap_2A[15] suggesting that Ap_2A exists in solution in a multi-conformation equilibrium, as opposed to a simple stacked-unstacked equilibrium, yet, none of the previously suggested stacked modes (α–α, β–β or α–β) was attributed to Ap_2A. On the other hand, all three stacking modes were suggested for Ap_2A by computational studies in the gas phase.[24] We found that Ap_4A equilibrates between those three stacking modes, the same as we (and others)[15,24] found for Ap_2A. The multi-conformation equilibrium found for $Ap_{2/4}A$ is likely for all other Np_nN's as well.

Dinucleotides with longer phosphate chain have weaker stacking interactions

Our NMR data show chemical shift differences ($\Delta\delta$) for the dinucleotide nucleobase protons *vs.* the corresponding mononucleotide protons that may be interpreted as a different degree of base-stacking for the different Np_nN's. In each dinucleotide series, *e.g.*, Ap_nUs n = 2–4, the longer the dinucleotide phosphate chain, the smaller were the $\Delta\delta$ values, suggesting greater stacking for the shorter dinucleotides in each series.

Though according to NMR results we may suggest that Ap_2A and Ap_4A have different populations of stacked conformation, we do not obtain a significant difference between the populations of Ap_2A and Ap_4A in the PMF simulations. The stacking free energy difference between Ap_2A and Ap_4A is expected to be within the statistical error of the simulations. Moreover, the difference in $\Delta\delta$ in the NMR spectra does not provide quantitative information on the populations of stacked structures.

The preference for stacking in shorter dinucleotides may be a result of an entropic effect. The stacking in Ap_4A involves the loss of more degrees of freedom (of the phosphate linker) as compared to Ap_2A and therefore it may be less favorable. Another possible explanation is the repulsion of more negative charges in a longer phosphate chain that may be reduced in an extended conformation.

Purine dinucleotides show greater stacking interactions than pyrimidine dinucleotides

The base type, namely purine or pyrimidine, affects the interaction between the bases. The self association tendency of bases in mononucleotides was rated as: adenine > guanine > cytosine ~ uracil.[42] Similar data were obtained for oligonucleotides,.[49] Purine bases in nucleosides interact with each other more favorably than purine and pyrimidine bases, while pyrimidine–pyrimidine interactions could not be determined by NMR methods, though there was evidence for their association.[41] A similar tendency for base–base interactions was found in a computational work on dinucleoside monophosphate analogues.[27]

Here, the larger $\Delta\delta$ shifts were observed for purine dinucleotides, or mixed purine–pyrimidine dinucleotides, where the purine base affects the pyrimidine base (*i.e.* $\Delta\delta$ of H5, H6 of U or C in dinucleotides *vs.* mononucleotides UTP and CMP), while the effect of

pyrimidine on purine (*i.e.* $\Delta\delta$ of H2, H8 of A or G in dinucleotides *vs.* mononucleotides ATP and GTP), is small. Though the $\Delta\delta$ values in Up_nUs (*vs.* UTP) are small, it does not necessarily mean that there are no stacking interactions, only that they cannot be monitored by NMR. Indeed, computational data for related ribodinucleotides monophosphates (*e.g.* 3′-5′-ApA, ApG, GpA *etc.*) showed the following stacking preference: purine–purine > purine–pyrimidine > pyrimidine–pyrimidine.[27] Therefore, we suggest that Ap_nAs and Gp_nGs undergo intramolecular stacking to a greater extent as compared to Up_nUs. The stacking preference of purines over pyrimidines could be explained by the larger surface area of purines which increases dispersion interactions and also affects hydrophobic effects. Moreover, the aromaticity of purines is greater than that of pyrimidines,[53] thus possibly resulting in more significant dispersion forces. Among purines, we observed greater stacking for the adenine *vs.* guanine analogues, based on the respective $\Delta\delta$ values, consistent with literature data.[39,54]

Each nucleoside moiety in a folded dinucleotide retains a standard mononucleotide conformation

We systematically investigated the conformation of dinucleotide analogues and the corresponding mononucleotides (ATP, UTP, GTP and CMP) as reference compounds (Table 2). We found that the preferred conformation around the glycosidic-; C4′–C5′-; C5′–O5′-bonds are: *anti*/*high anti*, *gg*, and *gg*, respectively, for the investigated mononucleotides. Both purine and pyrimidine mononucleotides prefer the *S* conformer, except for CMP that prefers the *N* conformer. The reported sugar puckering for purine nucleoside monophosphates is preferentially *S* (60%), while for pyrimidine mononucleotides it is *N* (60%). Thus, in both cases there is no predominant sugar puckering, as we found here.[43] The reported conformation for mononucleotides around the glycosidic angle is *anti* and around C4′–C5′– and C5′–O5′– bonds it is *gg*.[43]

All dinucleotides studied here adopted preferentially the *high anti*/*anti* range of the χ angle according to C–H coupling constants, NOESY spectra, and MD simulation for $Ap_{2/4}A$. The β and γ torsion angles in Ap_nAs and $Gp_{4/5}Gs$ are in the *gg* conformation range, yet the *gg* preference is greater for γ- than for β-angles according to NMR (~70–80% and ~60%, respectively), and MD data. For the other dinucleotides, due to second order NMR spectra, the only information about the exocyclic group (C4′–C5′ and C5′–O5′) is derived from $J_{C4'P}$, which upon comparison to $J_{C4'P}$ of mononucleotides, provides information about the C5′–O5′ bond. The β angle in those cases adopts the *gg* conformation similar to mononucleotides.[43]

Most of the dinucleotides prefer the *S* conformer (~50–70%), except for cytosinylated ribose that prefers the *N* conformer (~60–70%). MD simulation for $Ap_{2/4}A$ also supported the preference for the *S* conformer. A slight preference for the *N* conformer was found for: Ap_3A (51%), Ap_3U (52%), Up_2U (52%) dinucleotides. These sugar puckering preferences are similar to those of mononucleotides.[43]

Here, we studied the conformation of various dinucleotides sodium salts. Yet, in a physiological medium, dinucleotides may exist either as sodium (or potassium) salts or coordinate to Mg^{2+}/Ca^{2+} ions. The preference of dinucleotides to Mg^{2+}/Ca^{2+} coordination and the possible conformations of these complexes of dinucleotides at a physiological pH will be reported in due course.

Conclusions

Based on our MD and NMR data, obtained at physiological pH, we propose that under physiological conditions dinucleotide sodium salts adopt preferentially the folded structure, which exists as an equilibrating mixture of α–α, β–β and α–β stacking modes. However, the conformation of each nucleotide moiety in a dinucleotide is similar to the standard mononucleotide conformation.

Investigating the conformations of the dinucleotides in solution is a prerequisite for understanding their activity as enzyme inhibitors and receptor ligands, since the ΔG of dinucleotide binding to proteins depends on the change of the dinucleotide conformation during the transition from solution to protein environment, in addition to de-solvation effects. Therefore, we expect our findings will be valuable to the community of biochemists and medicinal chemists.

Experimental

General. Ap_4A and Ap_5A, **3** and **4**, were purchased from Sigma (Steinheim, Germany) and Fluka (Steinheim, Germany). Dinucleotides **1–2** and **5–17** were synthesized according to literature.[9,33] The air- and moisture-sensitive reactions were performed in flame dried, nitrogen or argon flushed two necked flasks sealed with rubber septa. The reagents were introduced with a syringe. The progress of the reactions was monitored by TLC on precoated Merck silica-gel plates (60F-254) and visualization was accomplished by UV light. Primary purification of the dinucleoside polysphates was achieved on a LC (Isco UA-6) system using a Sephadex DEAE-A25 column, which was swelled in 1 M $NaHCO_3$. Final purification was achieved on a HPLC (Merck-Hitachi) system using semipreparative reverse-phase Gemini 5u C18 110A 250-10 (Phenomenex, Torrance CA, USA) or LichroCART 250-10 column (Merck, Darmstadt, Germany) in the following solvent system: MeOH/0.1 M TEAA (triethylammonium acetate, pH = 7.5). Each dinucleotide was purified using a different gradient. The gradients ranged from 1 : 99 to 13 : 87 MeOH–TEAA in 20 min, flow rate 5 mL min^{-1}. The purity of the dinucleotides was evaluated on an analytical reverse-phase column system (Gemini 5u, C-18, 110A, 150x4.60 mm, 5 µm, Phenomenex, Torrance, CA), in two solvent systems as described below. To obtain the dinucleotides as sodium salts they were passed through a column of Sephadex-CM C25 (Na^+-form) column. Some of the dinucleotides, *e.g.* $Gp_{2/3}C$, were passed instead through a column of Dowex 50WX8-200 (H^+ form that was washed with NaOH to obtain the Na^+ form of this resin), since they degraded on Sephadex- CM C25 (Na^+ form). Compounds were characterized by NMR spectroscopy using Bruker DPX-300, DMX-600, AC-200, Avance III -500 and Avance III - 700 spectrometers. Chemical shifts are reported in ppm relative to glycerol and D_2O as internal standard (^1H and ^{13}C NMR). The chemical shifts of ^{31}P NMR are reported in ppm relative to 85% H_3PO_4 as an external standard. The pD values of the NMR samples were adjusted with diluted NH_4HCO_3–TFA D_2O solutions or NaOD–DCl D_2O solutions. Measurements of pD

were performed with a Knick pH-meter (Knick, Berlin, Germany) equipped with biotrode 238140 or tiptrode 238080 Hamilton electrode (Hamilton, Bonaduz, Switzerland). Mass spectra were recorded on Micromass QToF low resolution mass spectrometer in ESI negative mode.

Typical procedure of dinucleotide synthesis: Guanosine-5′-diphospho-5′-cytidine (15) and guanosine-5′-diphospho -5′-N⁴-methyl carbamate-cytidine (17). CMP(Bu_3NH^+)$_2$ was prepared from the CMP free acid (119.7 mg, 0.37 mmol) and Bu_3N (176 μL, 2 eq) in 95% EtOH. GMP(Bu_3NH^+)$_2$ was prepared from GMP disodium salt by passing it through a column of activated Dowex 50WX-8 200 mesh, H^+ form. The column eluate was collected in an ice-cooled flask containing Bu_3N (2 eq) and EtOH. The resulting solution was freeze-dried to yield GMP(Bu_3NH^+)$_2$. Bis-tributylammonium CMP and CDI (348 mg, 2.15 mmol, 5.8 eq) were stirred in dry DMF under argon atmosphere at room temperature for 2 h. Dry MeOH (140 μL, 3.45 mmol, 9.3 eq) was added. After 20 min GMP(Bu_3NH^+)$_2$ (311 mg, 0.42 mmol) in dry DMF (6 mL) was added. The resulting solution was stirred at room temperature for 24 h. Then DMF was evaporated and a viscous oil was obtained. The resulting oil was separated on Sephadex DEAE-A25, eluting the product with a buffer gradient of 0–0.3 M NH_4HCO_3 (total volume 1000 mL). The product was obtained in *ca.* 10% yield (26 mg) after LC separation. N⁴-Me-carbamate-Gp_2C, **17**, was obtained in *ca.* 54% yield (152 mg). Final separation of **15** was achieved by applying isocratic elution of TEAA–MeOH 94 : 6 in 30 min (5 mL min^{-1}): t_R 8.07 min. Final separation of **17** was achieved by applying isocratic elution of TEAA–MeOH 93 : 7 in 20 min (5 mL min^{-1}): t_R 16.73 min. Purity data obtained on an analytical column for **15** and **17**: t_R 1.60 min; 3.73 min (99% and 98% purity, respectively) using solvent system I (99 : 1 to 97 : 3 ammonium acetate–MeOH over 23 min, 1 mL min^{-1}); t_R 2.43 min; 5.72 min (93% and 99% purity respectively) using solvent system II (99 : 1 to 97 : 3 of 0.01M KH_2PO_4 (pH=4.6)–MeOH over 23 min, 1 mL min^{-1}).

^{1}H NMR for **15** (600 MHz, D_2O) δ: 8.01 (s, 1H, H-8), 7.72 (d, J = 7.6, 1H, H-6), 5.88 (d, J = 7.6, 1H, H-5), 5.84 (d, J = 6.0, 1H, H-1′G), 5.82 (d, J = 4.0, 1H, H-1′C), 4.78 (dd, J = 6.0, 5.3,1H, H-2′G), 4.46 (dd, J = 5.3, 3.3, 1H, H-3′G), 4.31 (m, 1H, H-4′G), 4.24 (m, 3H, H-5′G), 4.23 (m, 3H, H-5′C), 4.22 (m, 3H, H-3′C), 4.18 (m, 3H, H-4′C), 4.16 (m, 3H, H-2′C), 4.15 (m, 3H, H-5″G), 4.12 (m, 1H, H-5″C) ppm. ^{31}P NMR (243 MHz, D_2O) δ: −10.72 (s, Pα) ppm. ^{13}C NMR (151 MHz, D_2O) δ :165.27 (C4C), 158.24 (C6G), 156.81 (C2C), 153.24 (C2G), 151.13 (C4G), 140.27 (C6C), 137.00 (C8G), 115.68 (C5G), 95.55 (C5C), 88.81 (C1′C), 86.40 (C1′G), 83.21 (d, J_{CP} = 9.0, C4′G), 81.74 (d, J_{CP} = 9.0, C4′C), 73.99 (C2′C), 72.87 (C2′G), 69.88 (C3′G), 68.31 (C3′C), 64.83 (d, J_{CP} = 4.0, C5′G), 63.84 (d, J_{CP} = 4.0, C5′C) ppm. MS (MALDI): m/z 667 ((M − H)$^-$, 100%).

^{1}H NMR for **17** (600 MHz, D_2O) δ: 8.08 (d, J = 7.6, 1H, H-6), 7.95 (s, 1H, H-8), 6.92 (d, J = 7.6, 1H, H-5), 5.97 (d, J = 5.7, 1H, H-1′G), 5.82 (d, J = 2.7, 1H, H-1′C), 4.71 (dd, J = 5.7, 5.3,1H, H-2′G), 4.45 (dd, J = 5.3, 3.7, 1H, H-3′G), 4.34 (m, 1H, H-5′C), 4.29 (m, 2H, H-4′G), 4.28 (m, 2H, H-5′G), 4.24 (m, 3H, H-3′C), 4.24 (m, 3H, H-4′C), 4.21 (m, 3H, H-2′C), 4.18 (m, 2H, H-5″G), 4.17 (m, 2H, H-5″C), 3.77 (s, 3H, OMe) ppm. ^{31}P NMR (243 MHz, D_2O) δ: −10.48 (s, Pα) ppm. ^{13}C NMR (151 MHz, D_2O) δ : 163.53 (C4C), 159.29 (C6G), 157.24 (C2C), 154.80 (COOMe) 154.37 (C2G), 152.22 (C4G), 145.07 (C6C), 138.14 (C8G), 116.81 (C5G), 97.37 (C5C), 91.15 (C1′C), 87.75 (C1′G), 84.21 (d, J_{CP} = 8.7, C4′G), 83.01 (d, J_{CP} = 8.7, C4′C), 75.50 (C2′C), 74.22 (C2′G), 70.97 (C3′G), 68.98 (C3′C), 66.05 (d, J_{CP} = 5.0, C5′G), 64.76 (d, J_{CP} = 3.7, C5′C), 53.99 (CH_3) ppm. MS (MALDI): m/z 725((M − H)$^-$, 69%) 515(33), 548(100), 699(51), 705(15).

Guanosine-5′-triphospho-5′-cytidine (16). CMP(Bu_3NH^+)$_2$ was prepared from the CMP free acid (139 mg, 0.43 mmol) and Bu_3N (205 μL, 2 eq) in EtOH. GDP(Bu_3NH^+)$_2$ was prepared from GDP disodium salt (199.8 mg, 0.41 mmol) by passing it through a column of activated Dowex 50WX-8 200 mesh, H^+ form. The column eluate was collected in an ice-cooled flask containing Bu_3N (2 eq) and EtOH. The resulting solution was freeze-dried to yield GDP(Bu_3NH^+)$_2$. Bis-tributylammonium CMP and CDI (350.5 mg, 2.16 mmol, 5 eq) were stirred in dry DMF under argon atmosphere at room temperature for 3.5 h. Dry MeOH (140 μL, 3.44 mmol, 8 eq) was added. After 5 min GDP(Bu_3NH^+)$_2$ in dry DMF (4 mL) was added. The resulting solution was stirred at room temperature overnight. Then DMF was evaporated and a viscous oil was obtained. The resulting oil was separated on Sephadex DEAE-A25, eluting the product with a buffer gradient of 0–0.5 M NH_4HCO_3 (total volume 1000 mL). The product was obtained in *ca.* 3% yield (12 mg). Final purification of Gp_3C was achieved by HPLC isocratic elution with MeOH–0.1 M TEAA (pH = 7.5) 5 : 95 over 30 min, 5 mL min^{-1}. Purity data obtained on an analytical column for **16**: t_R 1.81 min (90% purity) using solvent system I (100% ammonium acetate over 23 min, 1 mL min^{-1}); t_R 1.87 min (99% purity) using solvent system II (100% of 0.01 M KH_2PO_4 (pH 4.6) over 23 min, 1 mL min^{-1}).

^{1}H NMR (600 MHz, D_2O) δ: 8.07 (s, 1H, H-8), 7.80 (d, J = 7.8, 1H, H-6), 5.91 (d, J = 7.8, 1H, H-5), 5.87 (d, J = 3.5, 1H, H-1′C), 5.86 (d, J = 6.5, 1H, H-1′G), 4.49 (dd, J = 5.0, 3.0, 1H, H-3′G), 4.31 (quint, 1H, H-4′G), 4.25 (m, 6H, H-3′C), 4.24 (m, 6H, H-5′C), 4.24 (m, 6H, H-5″C), 4.21 (m, 6H, H-2′C), 4.20 (m, 6H, H-5′G), 4.20 (m, 6H, H-5″G), 4.18 (m, 6H, H-4′C) ppm. ^{31}P NMR (243 MHz, D_2O) δ: −10.78 (d, J = 18.7, Pα), −10.90 (d, J = 19.7, Pα), −22.42 (t, J = 19.1, Pβ) ppm. ^{13}C NMR (151 MHz, D_2O) δ : 166.46 (C4C), 159.42 (C6G), 158.00 (C2C), 154.52 (C2G), 152.37 (C4G), 141.51 (C6C), 138.11 (C8G), 116.78 (C5G), 96.79 (C5C), 89.91 (C1′C), 87.22 (C1′G), 84.58 (d, J_{CP} = 8.7, C4′G), 82.82 (d, J_{CP} = 9.3, C4′C), 75.08 (C2′C), 74.50 (C2′G), 71.32 (C3′G), 69.33 (C3′C), 65.95 (d, J_{CP} = 4.5, C5′G), 64.87 (d, J_{CP} = 4.5, C5′C) ppm. MS (MALDI): m/z 747((M − H)$^-$, 100%) 515(15), 548(48).

NMR Experiments

Preparation of NMR samples. dinucleotide sodium salts were disssolved in 99.9% D_2O and pD was adjusted to a physiological pH (pD = 7.4 ± 0.2, pD = pH + 0.4). NMR spectra were measured at 600 MHz, 300 K ± 0.5, using glycerol as an internal reference. The sample concentration was kept low enough to avoid intermolecular base stacking yet, high enough to enable the detection of clear and sharp NMR signals.[42,51] Therefore, the concentration of dinucleotides in the NMR samples was 7.5 mM.

Analysis of ^{1}H, ^{13}C, ^{31}P-NMR data for compounds 1–17. The ribose conformation of the dinucleotides was analyzed in terms of a dynamic equilibrium between two favored puckered conformations *North* (*N*) conformer and *South* (*S*) conformer *N*

and *S* equilibrium populations were calculated from the observed $J_{1'2'}$ and $J_{3'4'}$ couplings as reported previously (eqn (1)–(2)).[43,44,55,56]

$$J_{1'2'} = 9.3(1 - X_N) = 9.3X_S \quad (1)$$

$$J_{3'4'} = 9.3X_N \quad (2)$$

Information on the torsional angles around C4′–C5′ (γ-angle) and C5′–O5′ (β-angle), Fig. 3, was obtained by measuring $J_{4'5'}$ and $J_{4'5''}$, $J_{P5'}$ and $J_{P5''}$ values and using previously reported equations eqn (3)–(5) and (6)–(8), respectively[44]

$$\rho_{gg} = [(J_t + J_g) - (J_{4'5'} + J_{4'5''})]/(J_t - J_g) \quad (3)$$

$$\rho_{tg} = (J_{4'5'} - J_g)/(J_t - J_g) \quad (4)$$

$$\rho_{gt} = (J_{4'5''} - J_g)/(J_t - J_g) \quad (5)$$

where $J_g = 2.04$ Hz, $J_t = 11.72$ Hz.

$$\rho_{gg} = [(J_t + J_g) - (J_{P5'} + J_{P5''})]/(J_t - J_g) \quad (6)$$

$$\rho_{tg} = (J_{P5'} - J_g)/(J_t - J_g) \quad (7)$$

$$\rho_{gt} = (J_{P5''} - J_g)/(J_t - J_g) \quad (8)$$

where $J_t = 20.9$ Hz, $J_g = 1.8$ Hz.

The determination of the conformation around the glycosidic bond (χ-angle), Fig. 3., was obtained by Karplus-type equations using the vicinal coupling constants $^3J_{C8/6\text{-}H1'}$ and $^3J_{C4/2\text{-}H1'}$, which were extracted from ^{13}C NMR spectra for each dinucleotide (eqn (9)–(10)).[57]

$$^3J_{C6/8\text{-}H1'} = 4.5\cos^2(\chi - 60°) - 0.6\cos(\chi - 60°) + 0.1 \quad (9)$$

$$^3J_{C2/4\text{-}H1'} = 4.7\cos^2(\chi - 60°) + 2.3\cos(\chi - 60°) + 0.1 \quad (10)$$

π–π Stacking interactions were deduced based on $\Delta\delta$ of H2/H8 of the adenine/guanine ring and H5/H6 of the uracil/cytosine ring in each of the analogues as compared to their values in ATP and GTP, UTP, and CMP, correspondingly.

MD Simulations

System preparation. Initially, the Ap_2A molecule was built in its extended conformation with the bases in the *anti* conformation (as in mono-nucleotides[43]). The initial conformation of Ap_4A was generated from crystallographic data.[23] Each dinucleotide was placed in a 40 × 40 × 40 Å^3 pre-equilibrated box of 2044 TIP3 water molecules.[58] Sodium counter ions were added in order to neutralize the system. Water molecules that overlapped with the solute or sodium ions were deleted (<2.5 Å heavy atom distance). The initial systems were minimized by 150 cycles of the Adopted Basis Newton-Raphson (ABNR) method.[59] Periodic boundary conditions were applied. The hydrogen atom–heavy atom bond lengths were constrained by the SHAKE algorithm.[60] A relative dielectric constant of 1.0 was used. The Particle Mesh Ewald method was employed to treat electrostatics.[61] The number of grid point for the Ewald summation in each dimension was 32, while the κ parameter was set to 0.34 Å^{-1}. A spherical cutoff scheme was employed for van der Waals and real-space electrostatic interactions with a value of 12.0 Å. Heating dynamics were performed from 50 to 300 K in 5 intervals of 5 psec with the velocity Verlet algorithm for a total time of 25 psec (timestep 1 fsec).[62] Equilibration of the system at 300 K was performed for 75 psec in the constant particle–volume–temperature (NVT) ensemble using the Nose thermostat with an effective mass of 100 kcal mol^{-1}-ps^2. Production dynamics was performed at 300 K for Ap_2A and Ap_4A in the particle–pressure–temperature (NPT)[63] ensemble employing the leapfrog algorithm.[62] Specifically, the extended system pressure algorithm of Andersen with an effective mass of 500.0 amu,[64] and the Hoover thermostat with an effective mass of 1000.0 kcal mol^{-1}-ps^2.[65] For each dinucleotide a 10 ns production simulation was performed. The trajectories were analyzed and structural data were collected. All the minimizations and simulations were performed with CHARMM program[59,66] and the CHARMM27 force-field.[67,68] All visualization employed Discovery Studio 2.1 (Accelrys Inc.).

The distance between the dinucleotide bases is defined as the distance between the glycosidic nitrogen atoms of the bases (N9-N92).[26,27] The torsional angles were defined according to Saenger (Fig. 3).[46] χ: O4′–C1′–N9–C4, α: O3P–P–O5′–C5′, β: P–O5′–C5′–C4′, γ: O5′–C5′–C4′–C3′. Endocyclic torsion angles: ν0: C4′–O4′–C1′–C2′, ν1: O4′–C1′–C2′–C3′, ν2: C1′–C2′–C3′–C4′, ν3: C2′–C3′–C4′–O4′, ν4: C3′–C4′–O4′–C1′.

The conformation around the glycosidic bond is defined as *anti* for 180 ± 90° and *syn* for 0 ± 90°.In the *syn* region, the range between 270–330° is also referred to as *high anti.* The exocyclic group torsion angles β and γ are in the *gg* range when $\beta = 180 \pm 30°$ and $\gamma = 60 \pm 30°$. Pseudorotation was calculated according to Altona and Sundaralingam:[55]

$$\tan P = \frac{[(\nu4 + \nu1) - (\nu3 + \nu0)]}{2\nu2(\sin 36 + \sin 72)} \quad (11)$$

Where we add 180 to the calculated P if $\nu2 < 0$. Here $P = 0$ is for the *Northern* conformer and P =180 is for the *Southern* conformer.

Crystal simulation of Ap_4A

Coordinates for Ap_4A were obtained from the Cambridge Structural Database (entry HIFKEM)[23] for the tetrasodium-dodecahydrate form of Ap_4A. We tested the CHARMM27 force field for the dinucletides by crystal minimization and simulation.[69] The crystal unit cell of Ap_4A with 4 sodium ions and 12 water molecules was minimized by 120 cycles of the ABNR method. A minimization of the lattice parameters was done as well by 60 cycles of the ABNR method. The system was heated stepwise from 50 to 293 K[23] for 25 psec in 4 intervals of 50 K and one final interval of 43 K, and further equilibrated for 25 psec in the NVT ensemble. Finally, a simulation was run in the NPT ensemble for 1 ns.

PMF simulations

The PMF was obtained from umbrella sampling MD simulations.[70] The PMF was defined as:

$$W(\zeta) = -k_B T \ln\rho^*(\zeta) - U^{bias}(\zeta) \quad (12)$$

where k_B is the Boltzman constant, T is the temperature and $\rho^*(\zeta)$ is the probability distribution in the presence of a biasing potential. To enable uniform sampling along the reaction coordinate, ζ, a biasing potential is added to the potential energy of the system, $U^{bias}(\zeta) = U^{umbr}(\zeta) + U^{harm}(\zeta)$. The umbrella potential, $U^{umbr}(\zeta)$, is defined as $-W(\zeta)$, which essentially removes the free energy barrier, thereby enabling uniform sampling. Furthermore, the reaction coordinate is divided into regions (windows) centered around ζ_0, and a harmonic potential, $U^{harm}(\zeta) = k(\zeta - \zeta_0)^2/2$, is added. Here k is a force constant, ζ is the reaction coordinate defined as the distance between the N9 atoms in the two nucleotide bases (Fig. 9)[26] and ζ_0 is the distance which the system is restrained to in a specific simulation window. This allows the dinucleotide/solvent system to relax along the reaction coordinate. Practically, since $W(\zeta)$ is required for $U^{umbr}(\zeta)$ and the optimal force constant, k, in $U^{harm}(\zeta)$ are not known at the outset, the PMF is obtained by employing adaptive umbrella sampling molecular dynamics simulations whereby the biasing potential is refined in an iterative manner.

The PMF windows were connected using the weighed-histogram analysis method.[71] The reaction coordinate is expected to be representative of the principal stacking features, although there can be situations where bases are close but not parallel (*i.e.* T-stacked as opposed to parallel stacking).[26,27]

For each energy profile there were 22 windows. The reaction coordinate range was 3.5–14.0 Å for both dinucleotides. Each window was equilibrated for *ca.* 50 psec and further sampled for *ca.* 500 ps. Convergence was verified by building the PMF in stages starting from 100 ps to 500 ps with increments of 100 ps. The PMFs obtained from the different stages were identical to within ±1 kcal mol^{-1}. Based on this analysis, we estimate the statistical error to be *ca.* ±1 kcal mol^{-1}, (see Supplementary Material Fig. S12–S17†). More PMF simulations were run for 1 ns for Ap_2A and Ap_4A in all the stacking modes at 4.5 Å, except for Ap_4A α–β which was at 6.5 Å, corresponding to the stacked configuration, and at 9.0 Å, corresponding to the open configuration. The torsion angles χ, α, β and γ were monitored as well as pseudorotation angle and the distances N6–O22′, N62–O2′.

Abbreviations

A	adenine
G	guanine
U	uracil
C	cytosine
Ap_nAs	diadenosine polyphosphates
Gp_nGs	diguanosine polyphosphates
CD	circular dichroism
NMP	nucleoside monophosphate
NDP	nucleoside diphosphate
NTP	nucleoside triphosphate
CDI	carbonyl diimidazole
DCC	dicyclohexylcarbonyldiimide
COSY	correlation spectroscopy
HMQC	heteronuclear multiple quantum coherence
HMBC	heteronuclear multiple bond coherence
NOESY	nuclear Overhauser effect spectroscopy
MD	molecular dynamics
PMF	potential of mean force

References

1 A. G. McLennan, *Pharmacol. Ther.*, 2000, **87**, 73–89.
2 C. H. V. Hoyle, R. H. Hilderman, J. J. Pintor, H. Schluter and B. F. King, *Drug Dev. Res.*, 2001, **52**, 260–273.
3 V. Ralevic, J. Jankowski and H. Schluter, *Br. J. Pharmacol.*, 2001, **134**, 1073–1083.
4 E. G. Delicado, M. T. Miras-Portugal, L. M. Carrasquero, D. Leon, R. Perez-Sen and J. Gualix, *Pflugers Arch.-Eur. J. Physiol.*, 2006, 563–572.
5 A. Guzman-Aranguez, A. Crooke, A. Peral, C. H. V. Hoyle and J. Pintor, *Prog. Retinal Eye Res.*, 2007, **26**, 674–687.
6 J. G. Douglass, R. I. Patel, B. R. Yerxa, S. R. Shaver, P. S. Watson, K. Bednarski, R. Plourde, C. C. Redick, K. Brubaker, A. C. Jones and J. L. Boyer, *J. Med. Chem.*, 2008, **51**, 1007–1025.
7 A. G. McLennan, *Ap4A and other dinucleoside polyphosphates*, CRC Press, Boca Raton, Florida, 1992.
8 K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hubner and D. Kern, *Nature*, 2007, **450**, 838–844.
9 S. R. Shaver, J. L. Rideout, W. Pendergast, J. G. Douglass, E. G. Brown, J. L. Boyer, R. I. Patel, C. C. Redick, A. C. Jones, M. Picher and B. R. Yerxa, *Purinergic Signalling*, 2005, **1**, 183–191.
10 E. Holler, B. Holmquist, B. L. Vallee, K. Taneja and P. Zamecnik, *Biochemistry*, 1983, **22**, 4924–4933.
11 M. Ikehara, S. Uesugi and K. Yoshida, *Biochemistry*, 1972, **11**, 836–842.
12 N. H. Kolodny and L. J. Collins, *J. Biol. Chem.*, 1986, **261**, 14571–14575.
13 N. H. Kolodny, E. Kisteneff, C. Redfield and E. Rapaport, *FEBS Lett.*, 1979, **107**, 121–124.
14 R. B. Westkaemper, *Biochem. Biophys. Res. Commun.*, 1987, **144**, 922–929.
15 J. W. Pettegrew, D. W. Miles and H. Eyring, *Proc. Natl. Acad. Sci. U. S. A.*, 1977, **74**, 1785–1788.
16 J. F. Scott and P. C. Zamecnik, *Proc. Natl. Acad. Sci. U. S. A.*, 1969, **64**, 1308–1314.
17 J. Stepinski, M. Bretner, M. Jankowska, K. Felczak, R. Stolarski, Z. Wieczorek, A.-L. Cai, R. E. Rhoads and A. Temeriusz, *Nucleosides, Nucleotides Nucleic Acids*, 1995, **14**, 717–721.
18 K. H. Mayo, O. M. Mvele and R. N. Puri, *FEBS Lett.*, 1990, **265**, 97–100.
19 K. Scheffzek, W. Kliche, L. Wiesmueller and J. Reinstein, *Biochemistry*, 1996, **35**, 9716–9727.
20 M. B. Berry and G. N. Phillips, *Proteins: Struct., Funct., Genet.*, 1998, **32**, 276–288.
21 M. D. Baker, D. E. Holloway, G. J. Swaminathan and K. R. Acharya, *Biochemistry*, 2006, **45**, 416–426.
22 D. K. Simanshu, H. S. Savithri and M. R. N. Murthy, *Proteins: Struct., Funct., Bioinf.*, 2008, **70**, 1379–1388.
23 D. Watanabe, M. Ishikawa, M. Yamasaki, M. Ozaki, T. Katayama and H. Nakajima, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 1996, **52**, 338–340.
24 J. M. Thornton and P. M. Bayley, *Biopolymers*, 1976, **15**, 955–975.
25 J. Norberg and L. Nilsson, *Biophys. J.*, 1994, **67**, 812–824.
26 J. Norberg and L. Nilsson, *Biophys. J.*, 1995, **69**, 2277–2285.
27 J. Norberg and L. Nilsson, *J. Am. Chem. Soc.*, 1995, **117**, 10832–10840.
28 J. Norberg and L. Nilsson, *J. Phys. Chem.*, 1996, **100**, 2550–2554.
29 P. Petrova, J. Koca and A. Imberty, *Eur. J. Biochem.*, 2001, **268**, 5365–5374.
30 P. Petrova, C. Monteiro, C. Herve du Penhoat, J. Koca and A. Imberty, *Biopolymers*, 2001, **58**, 617–635.
31 J. Liao, S. Sun, D. Chandler and G. Oster, *Eur. Biophys. J.*, 2004, **33**, 29–37.
32 N. Foloppe and L. Nilsson, *J. Phys. Chem. B*, 2005, **109**, 9119–9131.
33 W. Pendergast, B. R. Yerxa, J. G. Douglass, S. R. Shaver, R. W. Dougherty, C. C. Redick, I. F. Sims and J. L. Rideout, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 157–160.
34 H. Sigel and R. Griesser, *Chem. Soc. Rev.*, 2005, **34**, 875–900.
35 K. H. Scheller and H. Sigel, *J. Am. Chem. Soc.*, 1983, **105**, 5891–5900.
36 O. K. Kalinowski, S. Berger, and S. Braun, *Carbon-13 NMR Spectroscopy*, John Wiley and Sons, Chichester, 1994.
37 T. Son and C. Chachaty, *Biochim. Biophys. Acta.*, 1977, **500**, 405–418.
38 M. Remin and D. Shugar, *Biochem. Biophys. Res. Commun.*, 1972, **48**, 636–642.
39 C. Giessner-Prettre and B. Pullman, *J. Theor. Biol.*, 1970, **27**, 87–95.
40 P. O. Ts'o, N. S. Kondo, M. P. Schweizer and D. P. Hollis, *Biochemistry*, 1969, **8**, 997–1029.

41 M. P. Schweizer, S. I. Chan and P. O. P. Ts'o, *J. Am. Chem. Soc.*, 1965, **87**, 5241–5247.
42 K. H. Scheller and H. Sigel, *J. Am. Chem. Soc.*, 1983, **105**, 5891–5900.
43 D. B. Davies, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1978, **12**, 135–225.
44 D. B. Davies and S. S. Danyluk, *Biochemistry*, 1974, **13**, 4417–4434.
45 M. Cohn, *Structural and Chemical properties of ATP and its Metal Complexes in Solution*, The New York Academy of Sciences, New York, 1990.
46 W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New-York, 1984.
47 K. H. Scheller, F. Hofstetter, P. R. Mitchell, B. Prijs and H. Sigel, *J. Am. Chem. Soc.*, 1981, **103**, 247–260.
48 J. Sponer, K. E. Riley and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2595–2610.
49 K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, D. C. Tahmassebi and E. T. Kool, *J. Am. Chem. Soc.*, 2000, **122**, 2213–2222.
50 J. Sponer, P. Jurecka, I. Marchan, F. J. Luque, M. Orozco and P. Hobza, *Chem.–Eur. J.*, 2006, **12**, 2854–2865.
51 H. Sigel, R. Tribolet, R. Malinibalakrishnan and R. B. Martin, *Inorg. Chem.*, 1987, **26**, 2149–2157.
52 J. Norberg and L. Nilsson, *Acc. Chem. Res.*, 2002, **35**, 465–472.
53 A. F. Pozharskii, *Khim. Geterotsikl. Soedin.*, 1985, 867–905.
54 B. W. K. Shum and D. M. Crothers, *Biopolymers*, 1983, **22**, 919–933.
55 C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, 1972, **94**, 8205–8212.
56 C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, 1973, **95**, 2333–2344.
57 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, *Magn. Reson. Chem.*, 1996, **34**, S156–S176.
58 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
59 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
60 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
61 K. Nam, J. Gao and D. M. York, *J. Chem. Theory Comput.*, 2005, **1**, 2–13.
62 M. P. Allen, and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, U.K., 1987.
63 S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**, 4613–4621.
64 H. C. Andersen, *J. Chem. Phys.*, 1980, **72**, 2384–2393.
65 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
66 B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–1614.
67 N. Foloppe and J. MacKerell, A.D., *J. Comput. Chem.*, 2000, **21**, 86–104.
68 J. MacKerell, A. D. and N. Banavali, *J. Comput. Chem.*, 2000, **21**, 105–120.
69 A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. G. Gao, H., S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
70 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
71 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

