3408–3419 *Nucleic Acids Research, 2016, Vol. 44, No. 7* *Published online 20 February 2016*
doi: 10.1093/nar/gkw077

Quantitative sampling of conformational heterogeneity of a DNA hairpin using molecular dynamics simulations and ultrafast fluorescence spectroscopy

Karine Voltz[1,*], Jérémie Léonard[2,*], Patricia Tourón Touceda[2], Jamie Conyard[2], Ziyad Chaker[2], Annick Dejaegere[1], Julien Godet[3], Yves Mély[3], Stefan Haacke[2] and Roland H. Stote[1,*]

[1]Department of Integrative Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (INSERM) U964, Centre National de Recherche Scientifique (CNRS) UMR 7104, Université de Strasbourg, 67404 Illkirch, France, [2]Institut de Physique et Chimie des Matériaux de Strasbourg & Labex NIE, CNRS UMR 7504, Université de Strasbourg, Strasbourg, France and [3]Laboratoire de Biophotonique et Pharmacologie, UMR 7213 CNRS, Faculté de Pharmacie, Université de Strasbourg, Illkirch Cedex, France

Received November 12, 2015; Revised January 24, 2016; Accepted January 29, 2016

ABSTRACT

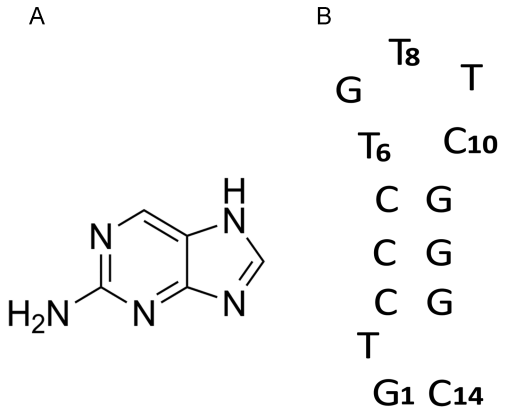
Molecular dynamics (MD) simulations and time resolved fluorescence (TRF) spectroscopy were combined to quantitatively describe the conformational landscape of the DNA primary binding sequence (PBS) of the HIV-1 genome, a short hairpin targeted by retroviral nucleocapsid proteins implicated in the viral reverse transcription. Three 2-aminopurine (2AP) labeled PBS constructs were studied. For each variant, the complete distribution of fluorescence lifetimes covering 5 orders of magnitude in timescale was measured and the populations of conformers experimentally observed to undergo static quenching were quantified. A binary quantification permitted the comparison of populations from experimental lifetime amplitudes to populations of aromatically stacked 2AP conformers obtained from simulation. Both populations agreed well, supporting the general assumption that quenching of 2AP fluorescence results from pi-stacking interactions with neighboring nucleobases and demonstrating the success of the proposed methodology for the combined analysis of TRF and MD data. Cluster analysis of the latter further identified predominant conformations that were consistent with the fluorescence decay times and amplitudes, providing a structure-based rationalization for the wide range of fluorescence lifetimes. Finally, the simulations provided evidence of local structural perturbations induced by 2AP. The approach presented is a general tool to investigate fine structural heterogeneity in nucleic acid and nucleoprotein assemblies.

INTRODUCTION

Nucleic acid hairpins can occur when two regions of a single-stranded DNA (ssDNA) or RNA form a helix that ends in a loop with unpaired or non-Watson–Crick-paired nucleotides. RNA hairpins regulate functions as varied as RNA translation, assistance of RNA folding and protection of mRNA from degradation (1), while DNA hairpins play essential roles in replication initiation and viral reverse transcription (2,3). To accomplish their functions, nucleic acid hairpins have to adapt to a wide range of interactions and often undergo multistep conformational rearrangements including global bending or twisting of helical domains and local rearrangements in base-pairing or stacking (4,5). This high flexibility, coupled to the various temporal regimes of hairpin dynamics, makes the study of their functional mechanism a challenging task.

Time-resolved fluorescence (TRF) spectroscopy has long been used to study the structural heterogeneity and dynamics of fluorescently-labeled biomolecules including DNA. In particular, 2-aminopurine (2AP), see Scheme 1A, a fluorescent analogue of adenine, has been employed to probe

*To whom correspondence should be addressed. Tel: +33 3 68 85 46 90; Fax: +33 3 88 65 32 01; Email: rstote@igbmc.fr
Correspondence may also be addressed to Karine Voltz. Tel: +33 3 68 85 47 30; Fax: +33 3 88 85 32 01; Email: voltz@igbmc.fr
Correspondence may also be addressed to Jérémie Léonard. Tel: +33 3 88 10 71 14; Fax: +33 3 88 10 72 45; Email: Jeremie.Leonard@ipcms.unistra.fr



Scheme 1. (**A**) 2-aminopurine (2AP) and (**B**) Δ(-)PBS.

DNA structure at the single nucleic acid level, to report, e.g. on nucleotide and protein binding reactions (6–10), conformations of looped-out bulge DNA duplexes (11), dynamics of DNA hairpins (12) or 'breathing' motion in ssDNA–dsDNA junctions (13). In fluorescently-labeled biopolymers, distributions of excited state (i.e. fluorescence) lifetimes are commonly observed due to interactions of the fluorophore excited state with its local environment. Different environments in distinct conformers result in different fluorescence decay times, and the relative amplitudes associated with these decay times are related to the relative populations of the corresponding conformers. In the case of 2AP, the isolated fluorophore in solution has a monoexponential excited-state decay with a lifetime of 10 ns (14,15), whereas upon incorporation in dinucleotides, ssDNA or dsDNA sequences, 2AP shows a dramatic fluorescence quenching characterized by a distribution of excited-state lifetimes spanning from the subpico- to nanosecond time scales (12,14–23). The quenching of 2AP fluorescence has long been suggested (14,15,20,21,24,25) to result from pi-stacking interactions with neighboring nucleobases and the broad range of observed times scales suggests a diversity of local geometrical arrangements of 2AP with respect to nearby nucleotides.

Retrieving more detailed structural information from TRF data requires the characterization of the structural and energetic landscape of the system under study. Molecular dynamics (MD) simulations allow extensive configurational sampling of biomolecules at the atomic level. They have been largely employed to interpret TRF anisotropy decay kinetics underlying local molecular motions of Trp in proteins (26–30). In one study, simple geometric measures were developed enabling a coarse-grained analysis of conformation-dependent fluorescence decay times of Trp/dye complexes (31). To elucidate molecular geometries involved in fluorescence quenching of these complexes, Vaiana *et al*. (31) analyzed TRF data to categorize the system into two sub-populations, either fluorescent or non-fluorescent and then used MD simulations to introduce a threshold distance 'above which no quenching occurs'. They showed that analysis at this binary level was sufficient to define a reasonably accurate 'quenching distance' for these systems.

Fewer studies, however, have combined MD and fluorescence spectroscopy for the interpretation of fluorescence decay kinetics of 2AP in nucleic acids. Early attempts to rationalize fluorescence decay measurements by MD were realized by Nordlund *et al*. on a 2AP-labeled dsDNA decamer (20). To interpret TRF data, Hall *et al*. used stochastic molecular dynamics simulations of native RNA to characterize the motions that a 2AP-substituted RNA hairpin loop might undergo (32). The detailed calculation of fluorescence decay times associated to various 2AP-labeled conformers would ideally require high-level quantum mechanical calculations of the 2AP excited state in each of the interaction geometries of the fluorophore with neighboring nucleobases in these conformers. However, the prediction of fluorophore excited state lifetimes in a given fixed biomolecular environment remains currently out-of-reach; current electronic structure calculations are limited to 2AP in dinucleotides (21,25,33,34) or small water clusters (35).

In the present work, we have developed a binary-level approach for the direct comparison of TRF and MD data in order to characterize the conformational heterogeneity of a DNA hairpin. ΔP(-)PBS is a segment of (-)PBS, the primary binding sequence of the HIV-1 genome. During the reverse transcription of the viral RNA, (-)PBS undergoes structural rearrangements chaperoned by the nucleocapsid protein 7 (NCp7) that enable its annealing to its complementary (+)PBS sequence and the processing of proviral DNA (18,36). The ΔP(-)PBS segment with 2AP incorporated at three different positions of the stem loop was previously studied by TRF spectroscopy at different time resolutions by Gelot *et al*. and Godet *et al*., respectively (17,18). Taking the results of these experiments together, time scales from <1 ps to ∼8 ns were observed, suggesting a high degree of structural heterogeneity of the ΔP(-)PBS loop.

To obtain better insight into the structural heterogeneity of the ΔP(-)PBS loop suggested by the TRF studies, we implemented a general protocol to directly confront results from MD simulations to TRF data. The 2AP-labeled ΔP(-)PBS segments were re-examined experimentally by combining new streak camera TRF spectroscopy data with the previously determined femtosecond data. This permitted the reconstruction of the fluorescence decay kinetics over 5 and 4 decades of time scales and signal amplitude, respectively. Femtosecond fluorescence anisotropy decay measurements were performed to determine the shortest time scale of local diffusional motion, which was subsequently used as an objective criterion to distinguish between static or dynamic quenching. The interpretation of these experimental results relied on explicit-solvent MD simulations of the same 2AP variant-ΔP(-)PBS constructs and the development of a quantitative approach for geometry characterization. Working under the general assumption that 2AP undergoes fluorescence quenching when stacked with other nucleobases, we defined two simple geometric measures that characterize the relative position of 2AP with respect to other nucleotides and used them to analyze the results of MD simulations. Threshold values of these metrics, which defined whether 2AP was stacked or not, were calculated from a set of reference geometries obtained from a MD simulation of double-stranded DNA at equilibrium. Despite the binary nature of the above criteria for analyz-

ing experimental and computational data, we demonstrate a nearly quantitative mapping between the population of conformers undergoing static quenching (TRF results) and the populations of conformers in which 2AP is stacked with at least one nearby nucleotide (MD results). This approach is designed to be a general method that enables the quantitative comparison between MD simulations and experimental TRF studies of 2AP-substituted DNA, and has several implications. First, strong support is given to the generally accepted, although largely unsubstantiated, assumption that static 2AP quenching in DNA occurs essentially upon stacking interactions with neighboring nucleotides. Then, the unprecedented level of comparison and agreement between TRF data and MD simulations achieved here strongly supports the general validity and accuracy of the MD simulations, which were in turn exploited for a more detailed structural analysis. Atomic-level structural information pertaining to the conformational heterogeneity of ΔP(-)PBS was obtained from the MD simulations through a detailed cluster analysis. Finally, comparing the MD simulations of native and labeled hairpins allowed us to extrapolate the conclusions supported by the TRF data to the native ΔP(-)PBS.

MATERIALS AND METHODS

Time-resolved fluorescence spectroscopy

The 2AP chromophore was substituted at positions 6, 8 or 10 (Scheme 1B) of ΔP(-)PBS. The three DNA hairpin variants were purchased from IBA GmbH and dissolved as received in water buffered at pH = 7.5 with 50 mM HEPES. The approach described here relies on results from two time-resolved fluorescence spectroscopy experiments performed using two different experimental set-ups (see experimental details in the Supplementary Data). The first set-up is a home-made femtosecond fluorescence spectrometer described elsewhere (17,37) which achieves a time-resolution as low as 300 fs by implementing a weakly efficient, non-linear, optical frequency conversion of the fluorescence emission (37). The second setup is a streak camera (C10627 streak tube from Hamamatsu Photonics) which affords a time resolution of 6 ps at best, but is instead much more sensitive and is therefore better suited to record the weak, ns-lived tail of the fluorescence decay.

For each labeled oligonucleotide, the population decay kinetics were measured with the streak camera and appended to the femtosecond decay kinetics previously reported (17) according to the procedure described in Supplementary Figure S1. Accurate fluorescence decay kinetics can thus be reconstructed over 5 decades of time scales and 4 decades of signal amplitude. The quantitative analysis of the reconstructed traces was performed with the Maximum Entropy Method (MEM) (38) using a home-made Matlab routine based on the work of Esposito *et al.* (39), with very good agreement with previous independent analyses of femtosecond (17) and picosecond (18) data. Here, the MEM yielded a least-biased Laplace transform of the entire experimental decay curves, that is the fully-resolved distribution of 2AP fluorescence lifetimes characterizing the distribution of oligonucleotide structures in solution and their dynamics on the sub-10 ns time-scale.

In addition, for each labeled oligonucleotide, as well as for free 2AP in solution, we used the femtosecond set-up to measure the fluorescence anisotropy decay kinetics, which revealed the time scales over which thermal motion of the 2AP fluorophores occur in the solvated oligonucleotides. The anisotropy decay kinetics were analyzed by performing the simultaneous (global), multiexponential fitting of the three traces recorded at the magic angle, perpendicular and parallel polarization directions, as described in the Supplementary Data.

Molecular dynamics simulations

Six different NMR-based structures of the ΔP(-)PBS were used for the computational study. Two structures were constructed from the 18-nucleotide (-)PBS DNA structure determined by Johnson *et al.* (40) (PDBID: 1EN1, structures 1 and 16) by deleting the 3′-GCCA overhang, and four structures were taken from the ΔP(-)PBS structures determined by Bourbigot *et al.* (41) (structures 5, 6, 9 and 10). These six models differ noticeably by the position of the base T8, which is stacked with G7 and/or with T9 and points toward the stem in structures of Bourbigot *et al.*, while in the two structures taken from the work of Johnson *et al.*, T8 is not involved in any stacking. Also, base C10 has various orientations in the different structures; it is stacked with G11 in models 6 and 10 from Bourbigot *et al.*, while it is non-stacked and less buried inside the loop for the other structures. These multiple structures were chosen as initial starting points in order to force an increased coverage of conformational space. Although NMR structure families themselves are unlikely to define a complete ensemble of structures, they are likely to represent more probable structures.

To position the 2AP in the initial structures, common heavy atom positions were retained from the experimental structures and the remaining heavy atoms were placed using the tools present in the CHARMM program (42). The CHARMM all-atom nucleic acid force field CHARMM27 (43) was used for all calculations. The topology and parameters for the 2AP moiety were constructed based on similarity to groups and parameters existing in the CHARMM27 all-atom nucleic acid force field (43) and were optimized by quantum mechanical calculations. The parameters are given in Supplementary Data. An initial energy minimization consisting of 1000 steps using the Steepest Descent method followed by 1000 steps of the Adapted Basis Newton–Raphson minimization method was realized in order to eliminate strong steric contacts prior to system solvation. The details of the simulations are given in the Supplementary Data.

For each of the above-noted NMR structures, three constructs having 2AP in positions 6, 8 and 10, were built, see Scheme 1B. The sequences of the native and substituted ΔP(-)PBS studied here are given in Table 1. From here on we refer to these variants as PBS6, PBS8 and PBS10, respectively. For each of the three ΔP(-)PBS variants and the native system, a set of eight 50ns simulations were done. Each individual simulation was started from one of the six above-noted NMR structures. For two of the simulations in each set, an experimental structure was used twice, but with different initial conditions. The correspondence be-

Table 1. Sequences of the native and substituted ΔP(-)PBS (positions 6, 8 and 10)

native PBS	ss(G T C C C T G T T C G G G C)
PBS6	ss(G T C C C 2AP G T T C G G G C)
PBS8	ss(G T C C C T G 2AP T C G G G C)
PBS10	ss(G T C C C T G T T 2AP G G G C)

tween the starting NMR structure and the placement of the trajectory segment is shown in Table S2 in Supplementary Data. The eight trajectories thus obtained for a given variant were sequentially appended to yield a single 400ns cumulated trajectory that was used for the analysis of each native and ΔP(-)PBS variant. We choose to carry out multiple 50 ns simulations using multiple starting structures and initial conditions rather than a single longer trajectory because it has been shown in numerous cases that such an approach provides a more effective sampling of conformational space (44–46). To further validate this, we compared the conformational distribution obtained by the procedure of running multiple short simulations to that obtained from a single long 400 ns simulation of native ΔP(-)PBS. The results, given in Supplementary Data, indeed show that the approach of multiple short simulations as done here provides a more extensive exploration of conformational space than that of a single long trajectory of ΔP(-)PBS.

Analysis of MD runs

The analyses were performed on all substituted and native ΔP(-)PBS constructs. Given the expected role of stacking interactions in 2AP fluorescence quenching, we characterized the conformations sampled during the MD simulations in terms of aromatic stacking of 2AP with nearby bases. Spatial arrangement of base steps in a nucleic acid helix are commonly described using parameters such as Tilt, Roll, Twist angles and Shift, Slide and Rise translations (47). In other studies that analyzed base stacking, different metrics have been used including distance criteria between atoms or between geometric center of bases (48–50) sometimes combined with angles and overlap criteria (51) or energy criteria (52). Here, we have defined two simple geometric criteria, (i) the distance $d^{com}_{i,j}$ between the centers of mass of the aromatic rings of two bases i and j (for purines, both the pyrimidine and imidazole rings are considered) and (ii) the angle between vectors normal to the aromatic ring planes $\theta_{i,j}$. We computed the distribution of distances $d^{com}_{i,j}$ and angles $\theta_{i,j}$ between 2AP (native base) and nearby nucleobases in each labeled (and native) ΔP(-)PBS construct.

To establish threshold values of these metrics that distinguish between stacked and unstacked geometries, we computed the distribution of these distances and angles occurring between adjacent, stacked bases in an ensemble of 5000 structures originating from a 10 ns MD simulation of 13-bp double-stranded DNA ds(GCTGCAAACGTCG) (PDBID 2KUZ, (53)). The same MD protocol as described above was used for this simulation. In the calculation of the distance and angle metrics, the nucleobases at the strand extremities were disregarded in order to avoid skewing the results due to fraying. This distribution (see Results) was used as a criterion to define the subset of conformations in which 2AP is stacked with neighboring bases.

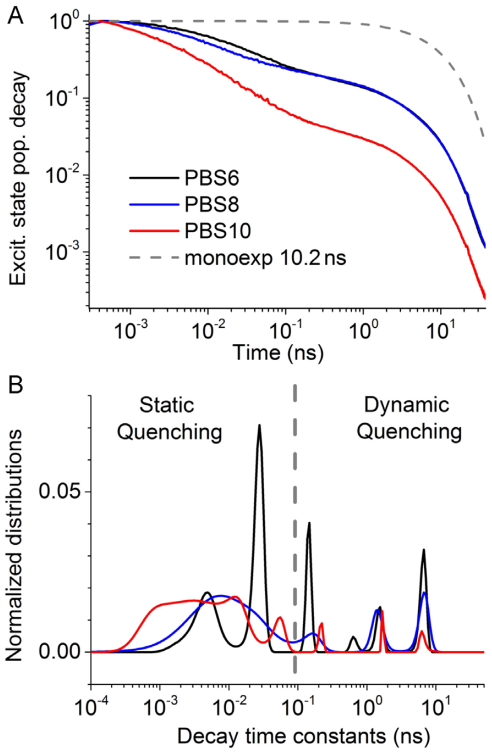


Figure 1. Excited state population decay (**A**) and the corresponding distributions of fluorescence decay times obtained by maximum entropy fitting method (MEM) analysis (**B**) for PBS6 (black), PBS8 (blue), and PBS10 (red). The dashed curve in (A) is a monoexponential decay curve with time constant 10.2 ns to simulate the fluorescence kinetics of free 2AP. The vertical dashed line in (B) represents the frontier of ∼90 ps in time scales between dynamic and static quenching, as determined by fluorescence anisotropy decay measurements (see text).

To further quantify the structural heterogeneity within the DNA hairpin loop, a cluster analysis was done focusing on three distinct segments of trinucleotides that included 2AP and its first adjacent neighbors. For PBS6 and PBS8, the trinucleotide segments were, respectively, T5–2AP6(T6)-G7 and G7–2AP8(T8)-T9, with the native base indicated in parentheses. For PBS10, the trinucleotide segment 2AP10(C10)-G11-G12 was used as it gave a more descriptive clustering of 2AP10 conformers. For each of the 2AP-labeled and native constructs, the cluster analysis was performed on 40 000 structures extracted at a time interval of 10 ps from the entire 400 ns cumulated trajectory using the MMTSB tool set (54). More details about the cluster analysis are given in Supplementary Data.

RESULTS AND DISCUSSION

Fluorescence decay measurements

For the three 2AP labeled compounds, the fluorescence population decay kinetics resolved from the 300-fs time scale to the 50 ns time scale and the corresponding distributions of fluorescence lifetimes are displayed in Figure 1A and B, respectively (see Materials and Methods and Supplementary Data). These distributions span over as many as 5

orders of magnitudes from sub-ps to several ns. We note that over all time scales, the distributions are significantly different depending on whether ΔP(-)PBS is labeled in position 6, 8 or 10, indicating significantly distinct, site-specific, 2AP local environments. This is particularly true over the sub 50 ps time range, which remains essentially inaccessible by conventional time-correlated single photon counting (TCSPC) experiments, thus justifying the need for femtosecond time resolution.

Extracting detailed, quantitative structural information from the distributions of fluorescence lifetime displayed in Figure 1B is a major challenge. Even at a qualitative level, structural interpretation is uncertain because the quenching mechanism itself is not clear. While a photo-induced charge transfer has been proposed to occur mostly to nearest neighbor guanines or next neighbor guanines via efficient charge transport along stacked nucleobases (22,55,56), other investigations point to a quenching of 2AP due to internal conversion to a long-lived dark state (25,34) induced by aromatic stacking interaction and possibly mediated by hydrogen bonding with the quencher. The mechanism of charge transfer to guanines alone appears to be insufficient to rationalize the TRF investigations of 2AP-labeled hairpins (12). In addition, multiexponential fluorescence decay kinetics are observed having significant or major contributions from sub-40 ps fluorescence lifetimes (21) in dinucleotides incorporating 2AP with any nucleobase, indicating that interaction with any base can yield static quenching.

The quantitative structural interpretation of the observed fluorescence (i.e. excited state) lifetimes is beyond the scope of this work. Focus will instead be on the corresponding amplitudes, which are related to the relative populations of distinct conformers. In order to compare with the results of the MD simulations, we analyzed the distributions displayed in Figure 1B to discriminate between static versus dynamic quenching, irrespective of the nature of the quenching mechanism. In order to define a precise criterion for this discrimination, it is most relevant to measure the fluorescence anisotropy decay kinetics, which give access to the typical time scales for 2AP motions inside the DNA hairpins (14). We therefore performed complementary anisotropy decay experiments with femtosecond time resolution. They revealed that the shortest anisotropy decay time scale for all three oligonucleotides is 90 ± 30 ps (see details in Supplementary Data). The relatively poor accuracy of this result is because, on this time scale, the fluorescence signal has already dropped strongly, as does the signal-to-noise ratio. Compared to the time scale for rotational diffusion of free 2AP or free 2AP-riboside in water, which are respectively 27 ± 2 ps (see Supplementary Data) or 80 ps (15), we conclude that the shortest time scale for fluorescence anisotropy decay must be associated to local, confined, diffusional motion of 2AP within the DNA hairpins.

Each of the fluorescence lifetimes appearing in the distributions displayed in Figure 1B correspond to an excited state decay rate $k = k_0 + k_Q$, where k_0 is the decay rate observed for isolated 2AP in water ($1/k_0 = 10$ ns), and k_Q is a quenching decay rate controlled by the interactions of the chromophore with its environment. Fluorescence lifetimes (k^{-1}) that are longer than the time scales of local intramolecular motions are to be assigned to dynamic quenching. Dynamic quenching is a process in which local motions of 2AP and neighboring nucleobases gate their interactions and hence the quenching (k_Q) of the 2AP excited state (12). In particular, the longest decay times, in the range of 6 to 8 ns, likely correspond to conformers where such gating motions occur at a typical rate of $k_Q = k - k_0 =$ 0.025 to 0.07 ns^{-1}, meaning that 2AP can remain exposed to water and distant from neighboring nucleobases on time scales k_Q^{-1} longer than 10 ns, possibly approaching 50 ns. Interestingly, no structures showing solvent exposed bases are present in the available families of NMR structure of (-)PBS. Conversely, all fluorescence lifetimes that are significantly shorter than local nucleobase motions must characterize conformers where the ground state structure (i.e. before photon absorption) is already favorable for excited state quenching. In these cases, quenching occurs soon after light excitation and before any significant motion of either interacting partners. We attribute these shorter fluorescence lifetimes to static quenching. We note here that the usual definition of static quenching refers to such ground state conformers but with the additional property that they remain 'unobserved' in TRF experiments (15,57). However, the proportion of 'unobserved' conformers obviously depends on the experimental time resolution. In fact, large proportions of labeled-PBS conformers appeared dark in a 100-ps time-resolved experiment (18), while essentially all are observed at femtosecond resolution (17). Therefore, we adopt here a definition of static quenching that does not depend on the experimental time resolution, but discriminates conformers by comparing their fluorescence lifetimes to the time scale of local motions. In order to provide quantitative proportions of conformers undergoing static versus dynamic quenching, we further propose to use the fastest fluorescence anisotropy decay time (90 fs in the present case) as a natural cut-off between the corresponding fluorescence lifetimes.

With the above definition, and irrespective of the nature of the quenching mechanism, we can deduce the relative population of conformers in which 2AP undergoes static quenching by integrating the lifetime distributions from 0 to the cut-off time scale of 90 ps (dashed line in Figure 1B). These structures represent 68%, 73% and 94% of the entire population of PBS6, PBS8 and PBS10, respectively. In such a binary analysis, although the criterion to discriminate static versus dynamic quenching is conceptually precisely defined, the exact cut-off value that determines the above percentages remains somewhat arbitrary. Reducing the cut-off value by a factor of 2 modifies these proportions by less than 5%, which is indicative of the uncertainty level of these proportions.

Molecular dynamics simulations

Binary analysis of aromatic stacking. Of interest was the definition of threshold values of the geometric metrics defined above in order to distinguish structures with 2AP in stacked versus non-stacked geometries, in line with the 'binary' interpretation of short and long experimental fluorescence lifetimes in terms of static versus dynamic quenching. To define these threshold values, the distribution of dis-

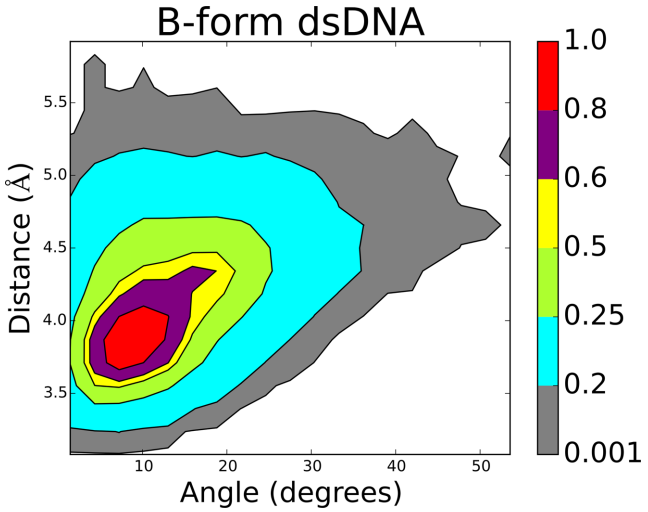


Figure 2. 2D density map of the relative distance $d^{com}_{i,j}$, and angle $\theta_{i,j}$ between adjacent bases in the reference simulation for dsDNA (see text). Level 1.0 of the density scale corresponds to the maximal density.

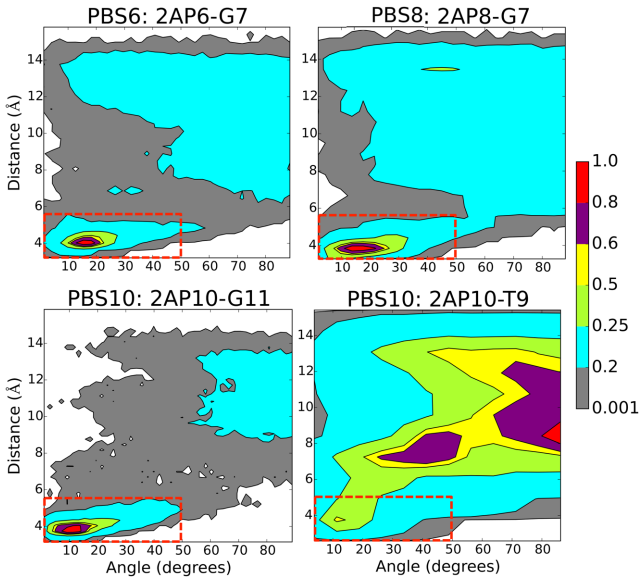


Figure 3. 2D density map of the relative distance $d^{com}_{i,j}$ and angle $\theta_{i,j}$ between 2AP6 and G7 in PBS6, 2AP10 and G11 in PBS10, 2AP8 and G7 in PBS8 and 2AP10 and T9 in PBS10. The density scale is normalized using the maximal density of the 2D plot in Figure 2. The red rectangles enclose the region $d^{com}_{i,j} < 5.5$ Å and $\theta_{i,j} < 50°$, defining B-form type stacking, (see Figure 2). For the dyads 2AP6-G7, 2AP8-G7, 2AP10-G11 and 2AP10-T9, these regions correspond to 28%, 27%, 61% and 7% of the entire population of PBS6, PBS8 and PBS10, respectively (see Table 2).

tances $d^{com}_{i,j}$ and angles $\theta_{i,j}$ between adjacent nucleobases, were computed from the MD simulation of a 13 bp dsDNA fragment and displayed as a probability density map, see Figure 2. From the contours of the distribution, we concluded that in dsDNA, a structure substantially less flexible than, e.g., loops in DNA hairpins, 100% of the pairwise stacking geometries are characterized by distances $d^{com}_{i,j} <$ 5.5 Å and relative angles $\theta_{i,j} < 50°$. These values were the same whether the dsDNA was B-form (PBDID: 2KUZ) or A-form (model calculations not presented here).

The same probability density plots were calculated from the MD simulations of the 2AP substituted PBS constructs and shown for the dyads 2AP6-G7, 2AP8-G7, 2AP10-G11 and 2AP10-T9 in Figure 3. In each of these cases, the structural heterogeneity of the dyads is much broader than that of a typical stacked base pair in dsDNA, with $d^{com}_{i,j}$ distances reaching values up to 14 Å. If we use the combined values $d^{com}_{i,j} < 5.5$ Å and $\theta_{i,j} < 50°$ derived from the simulations of the dsDNA as criteria for aromatic base stacking, subsets of structures where 2AP is aromatically stacked with nearby nucleobases are significantly populated, and likely correspond to structures where static quenching is expected. These subsets are highlighted in red dotted rectangles in Figure 3 and reveal, for instance, that dyads formed by 2AP and guanine only, i.e. 2AP6-G7, G7–2AP8 and 2AP10-G11, adopt pairwise stacking in 28%, 27% and 61% of the entire populations of PBS6, PBS8 and PBS10, respectively. More generally, when quantifying the percentage of conformations exhibiting such pairwise stacking of 2AP with *any* nucleobase, aromatic stacking of 2AP is predicted to occur in ∼60%, 50% or 70% of the structures when incorporated in positions 6, 8 or 10, respectively, see Table 2.

Given the binary level of description of (i) the 2AP fluorescence lifetimes in terms of 'static' versus 'dynamic' quenching, and of (ii) the MD simulation results in terms of 'stacking' or 'non stacking' using a simple description of stacking geometries, the computational results correlate remarkably well with the proportion of conformers determined experimentally to undergo static quenching. This high level of correlation provides strong evidence for the occurrence of static quenching of 2AP in a range of aromatically stacked geometries in flexible hairpins. In an attempt to discuss possible refinements in the comparison between experimental and computational results, we raised the question as to why the absolute values calculated from the simulations systematically appear 10 to ∼20% lower than those estimated by the TRF studies in the three hairpins. In addition to the simplified nature of our analysis, experimental relative uncertainties on fast versus slow fluorescence components arise from the procedure of appending fs and streak camera data, which is estimated to be around 3% (discussed in the Supplementary Data) and from the choice for the cut-off value, which is estimated to be of about 5% (discussed in Experimental Results). Other factors to consider include the general difficulty in achieving a full sampling of the conformational space by MD simulations, which are by necessity, finite and may lead to a biased estimate of certain populations. To check the influence of the trajectory time length on the populations of stacked conformers, we have extended each of the eight individual trajectories of PBS8 from 50 to 100 ns and calculated the percentages of pairwise stacking on this 8 × 100 ns concatenated trajectory. The percentages obtained are in the same range as the ones obtained for the 8 × 50 ns cumulated trajectory (see Supplementary Table S4).

In addition, we tested whether the criteria for base-stacking geometries were too strict and whether modulation of these parameters could allow a better match with the experimental percentages. Using expanded criteria for

Table 2. Percentage of conformers among the 40 000 MD structures exhibiting various pairwise stacking in PBS6, PBS8, PBS10 and native ΔP(-)PBS systems involving T6 or 2AP6 (left column), T8 or 2AP8 (middle column) and C10 or 2AP10 (right column)

	PBS6	Native		PBS8	Native		PBS10	Native
T6(2AP6)/C5	51%	64%	T8(2AP8)/T6	4%	<1%	C10(2-AP10)/T6	12%	6%
T6(2AP6)/G7	28%	50%	T8(2AP8)/G7	27%	39%	C10(2AP10)/G7	6%	1%
T6(2AP6)/T8	3%	<1%	T8(2AP8)/T9	12%	30%	C10(2AP10)/T8	4%	8%
T6(2AP6)/T9	6%	<1%	T8(2AP8)/C10	12%	8%	C10(2-AP10)/T9	7%	6%
T6(2AP6)/C10	6%	6%	T8(2AP8)/G11	2%	0%	C10(2AP10)/G11	61%	64%
Total	58%	79%	Total	47%	48%	Total	69%	68%

2AP10, e.g. pairwise stacking distance and angle of $d^{com}{}_{i,j} <$ 6 Å and $\theta_{i,j} < 60°$, respectively, increased the 2AP10 stacked population by only 5%, keeping it still at least 15% lower than the population of static quenching estimated by TRF for 2AP10. Also, for the stacking pairs 2AP6-G7, 2AP8-G7 and 2AP10-G11, the respective pairwise density maps show a relatively clear cut-off between pairwise stacking and non-stacked geometries (Figure 3). On the contrary, there is no such clear cut-off in the stacking probability maps for 2AP10 with T9 or G7, or for 2AP8 with T6 (Figure 3 and Supplementary Figure S5), thus suggesting that the binary discrimination analysis may be less appropriate for these two. Finally, we focused here on quantifying aromatic stacking only, but interactions of 2AP other than aromatic stacking, e.g. hydrogen bonding, may also contribute to static quenching (25,35,58).

The accurate reproduction of experimentally derived populations from MD simulations remains a challenge since it requires that the empirical force fields be able to account for subtle differences in absolute free energies. For example, the difference between the experiment (94%) and the calculated populations (70%) of 2AP stacked conformers in PBS10 in which 2AP may undergo static quenching corresponds to a difference of only 0.3 kT. While a more complete fundamental understanding of 2AP quenching mechanisms and its geometry dependence would help to refine the assignment of static quenching to more precisely defined geometries, this is not within the scope of the present work. The significance, however, of the correlation between the TRF and MD percentages is strongly reinforced by the fact that it is observed across the three hairpin systems, even though they exhibit different structural and dynamical properties in the vicinity of 2AP, as will be documented below. This validates the use of the above base stacking criteria for the identification of the 2AP conformers amenable to static quenching in substituted ΔP(-)PBS. It also gives strong confidence in the validity of the conformational sampling achieved in the MD simulations, thus prompting the more refined conformational analysis shown hereafter.

Cluster analysis. To further investigate the structural heterogeneity of both substituted and native ΔP(-)PBS sampled in the MD simulations, an RMSD-based cluster analysis (see Materials and Methods) was carried out for all PBS constructs. Time series for cluster occupation, 2AP solvent accessible surface area (SASA), the pairwise stacking interactions for the cumulated 400 ns simulations, and a representation of conformation with the smallest RMSD to the cluster centroid are presented in Figures 4–6 for PBS6, PBS8 and PBS10, respectively, and in Supplementary Figures S8, S11 and S13 for native ΔP(-)PBS. In this analysis, we will make reference to a particular cluster by the color of frame surrounding its representative conformation; this color specification is also used in the cluster time series shown in Figures 4–6, and in Supplementary Data.

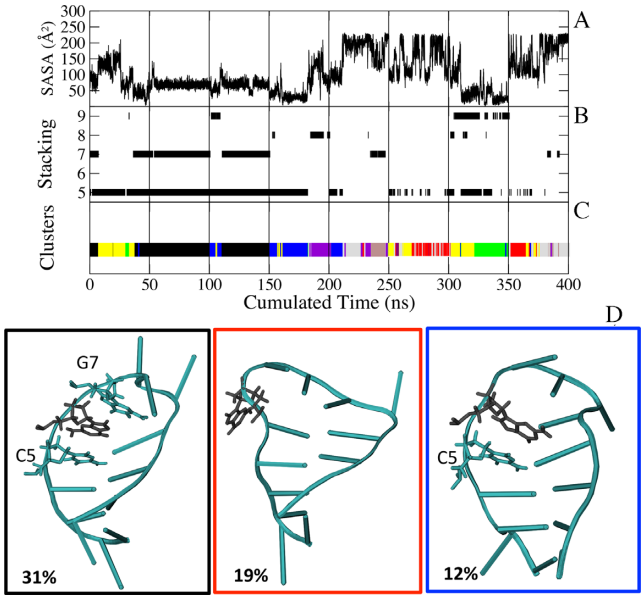


Figure 4. Distribution along the PBS6 cumulated 400-ns trajectory of (**A**) the 2AP6 SASA (**B**) the 2AP6 pairwise stacking occurrences and (**C**) the eight clusters obtained for PBS6 from the analysis of the C5–2AP6-G7 region. The vertical lines separate the eight 50-ns individual trajectories. (B) Each line represents the time-series of 2AP6 pairwise stacking occurrence with one base (5 = C5, 7 = G7, 8 = T8, 9 = T9) according to the criteria $d^{com}{}_{i,j} < 5.5$ Å and $\theta_{i,j} < 50°$C) Each color represents a particular cluster. (**D**) Schematic representation of PBS6 conformations with the smallest RMSD to the centroid for three of the eight clusters (see other clusters in Supplementary Data, Supplementary Figure S7). The color of the frames corresponds to the color of each cluster in (**C**). The 2AP6 base is represented in black.

We also use this information to discuss, for each labeled site, whether the substitution of 2AP affects the prevalent conformers by systematically comparing the clusters obtained for the substituted and native ΔP(-)PBS. We compare the clustered conformers to the motifs present in the NMR structures (see Supplementary Table S2 and Supplementary Figure S6). Finally, we address convergence issues of the simulation and the influence of the NMR structures used to start the simulations on the obtained clusters.

The cluster analysis of the PBS6 trajectories focused on the C5–2AP6-G7 sequence region and identified eight clusters, see Figure 4. Representative structures of the three

most populated clusters are shown in Figure 4D. The time series for cluster occupation shows that not all clusters are present in all trajectories, indicating that the trajectories do not all cover completely the same conformational space. This was expected as most trajectories are started from different structures, but nevertheless, there is a significant amount of overlap suggesting that the trajectories converge toward predominant motifs. The most populated cluster represents 31% of the entire population and can be populated over long times (e.g. 50 ns in the 50–100 ns segment of the trajectory, Figure 4C). This cluster represents structures that exhibit a C5–2AP6-G7 stacking triad oriented toward the interior of the DNA loop with low water exposure (SASA < 100 Å^2, Figure 4A). The average C5–2AP6-G7 stacking geometry values are $d^{com}{}_{2AP6\text{-}C5} = 4.0 \pm 0.5$ Å, $\theta_{2AP6\text{-}C5} = 16 \pm 11°$ and $d^{com}{}_{2AP6\text{-}G7} = 4.7 \pm 1.2$ Å, $\theta_{2AP6\text{-}G7} = 24 \pm 17°$. Two other clusters (blue in Figure 4 and yellow in Supplementary Figure S7) account for 21% of the population and display 2AP6-C5 stacking dyads inside the loop. Given the well-formed stacking geometries, 2AP likely undergoes static quenching in these structures. Interestingly, C5–2AP6-G7 stacking can occur within a simulation, even though such conformations are not present in the starting structure used for the simulation. Such geometries are, however, present in other members of the NMR structure family not used in this study. In about 23% of the conformers, 2AP6 can also form short-living pairwise stacking with G7, T8 or T9, that are moderately exposed to the solvent (<150 Å^2) and that can be oriented toward the exterior of the loop (see clusters brown, violet and green, Supplementary Figure S7). Finally, 25% of the population corresponds to conformers in which 2AP is non-stacked and highly exposed to the solvent (SASA > 150 Å^2), such as in the red and grey clusters in Figures 4D and Supplementary Figure S7, respectively. These latter water-exposed conformers can be populated over tens of nanoseconds, as it is the case between 380 and 400 ns (Figure 4A) and likely account for the fluorescent species undergoing dynamic quenching (Figure 1B) estimated to occur in ∼20% of the conformations (see Supplementary Table S1).

Cluster analysis of the C5-T6-G7 region of the native ΔP(-)PBS gives a distribution of six clusters (Supplementary Figure S8) similar to that observed for the PBS6, with prevailing C5-T6-G7 stacking triad and C5-T6 stacking dyad. However, the percentage of C5-T6 and T6-G7 stacking dyads is predicted to be higher for the native ΔP(-)PBS than for PBS6 (see Table 2). Referring to Supplementary Figure S9 panel A, which shows stacking time series from the PBS6 trajectories, we see that the C5–2AP6 stacking dyad is more stable when T8 is unstacked or stacked outside the loop as in trajectories 0–180 ns. In native ΔP(-)PBS (Supplementary Figure S9 panel B) instead, the T6-C5 stacking is more independent of the T8 pairwise stacking. Together, these results suggest that the replacement of T6 by 2AP creates a steric conflict within the loop affecting 2AP6 stacking. This is likely due to the larger size of the 2AP (purine) as compared to the native thymine (pyrimidine) in position 6.

For PBS8, the cluster analysis focused on the G7–2AP8-T9 region and yielded seven clusters with no particular prevailing cluster. 2AP8 predominantly stacks with G7 (see Figure 5B and Table 2) and the 2AP8-G7 dyad is either oriented inside the loop with additional stacking with T9 (17%), or outside the loop (14%), see green and yellow clusters, respectively in Figure 5D. The 2AP8-G7 stacking interaction can be very stable, as it can remain over several tens of ns as in the segment 300–350 ns (Figure 5C), but it can also disrupt spontaneously, as seen in segments 150–300 ns. In addition, 48% of the structures display shorter-lived stacking dyads of 2AP8 with T6 or C10 (red, blue, brown and grey clusters in Supplementary Figure S10). These are mostly oriented outside the loop and prone to disruption. Altogether, we suggest that these dyad conformations are representative of those where 2AP undergoes static quenching. Even though the fluorescence decay kinetics and proportions of conformers undergoing fluorescence static quenching are nearly the same for PBS6 and PBS8, the MD simulations reveal significant differences in the structural characteristics between these two systems. For PBS6, in a majority of conformers 2AP is stacked with C5 and additional stacking with G7 is often observed yielding the C5–2AP6-G7 triad. Instead, in PBS8, 2AP is located in a more flexible region of the loop and tends to favor dyad stacking mostly with G7, but also with second neighbors T6 or C10. The larger spread of structural parameters for 2AP8 may rationalize the different decay time histograms observed in Figure 1B for PBS 6 and 8 in the sub-100 ps range. As observed for PBS6, 21% of the structures are clustered around an orientation where 2AP is highly solvent exposed and sufficiently long-lived to correspond to the population of structures with long decay times (see Figure 5, black cluster).

Cluster analysis of the G7-T8-T9 region in the native ΔP(-)PBS (Supplementary Figure S11) yields similar results, highlighting the T8-G7 stacked dyad and the G7-T8-T9 triad, but with the difference that the G7-T8-T9 triad is mostly oriented outside the loop (e.g. black cluster in Sup-

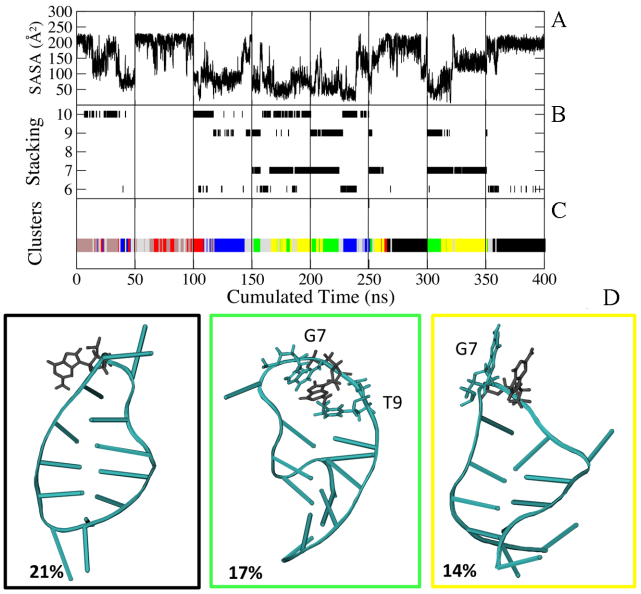


Figure 5. Distribution along the PBS8 cumulated 400-ns trajectory. See the legend of Figure 4 for details. Three out of seven clusters are displayed here (see the others in Supplementary Data, Supplementary Figure S10).

plementary Figure S11) and appears more stable than in PBS8 (as seen by comparing the simultaneous occurrence of stacking with G7 and T9 in Figure 5B and Supplementary Figure S11B). In addition, more transitions between stacking partners and more stacking with second neighbors are observed for 2AP8 than T8. Together, this indicates that the introduction of 2AP in position 8 may slightly perturb the organization of the loop. Moreover, in both the native ΔP(-)PBS and PBS8, the clusters obtained from the simulations seem to depend strongly on the initial structure. Nearly all structures from Bourbigot *et al*. display stacking of T8 with T9 and/or G7, in contrast to the structures from Johnson *et al*., which show no such stacking (see Supplementary Figure S6). In the present simulations, the G7-T8(2AP8)-T9 triad is only observed in simulations using structures from Bourbigot *et al*. (segments between 150–400 ns trajectories, Figure 5b and Supplementary Figure S11B). Importantly, in trajectories started with structures from Johnson *et al*. (segments 0–150 ns), T8(2AP8) does not maintain any simultaneous stacking with G7 and T9. The transition between stacked and unstacked conformations of T8(2AP8) would require the reorganization of three highly dynamic bases located in the middle of the loop (see RMSF, Supplementary Figure S4). Even though the spontaneous formation of the G7-T8 stacked dyad is seen in the simulation of native ΔP(-)PBS (segment 100–150 ns trajectory, Supplementary Figure S11B), the 50 ns trajectory length is probably too short to allow the complete reorganization of all three bases. This is especially true considering that the simulations were started from experimental structures of flexible loops with large conformational uncertainty, as is the case here given the significant differences between the structures of Bourbigot *et al*. and Johnson *et al*. for the T8 region. This could also contribute to the lower percentage of stacking obtained for this region compared to the proportion of static quenching observed in TRF experiments. Indeed, the difficulty arising from high flexibility is reinforced by the lack of the NMR experimental structures that fully represent the conformational ensemble of viable structures. As these structures are used as starting geometries, a conformational bias is imposed on the MD simulations and the populated clusters may correlate with the initial structure. This suggests that the two NMR studies describe structures that belong to two different energy basins separated by a sufficiently high energy barrier.

For PBS10, the TRF studies estimated that 94% of the species undergo static quenching, the highest percentage of all three variants. A very significant proportion of these conformers undergo a particularly efficient fluorescence quenching on a sub-ps time scale that is not observed in other variants. This very fast component is indicative of an enhanced electronic coupling between 2AP in position 10 and its nearby quencher. Concordantly, the simulations predict that a majority (∼70%) of the conformations implicate 2AP in a stacking geometry with another nucleobase, and most particularly with G11. Interestingly, however, in the NMR structures of Johnson *et al*., C10 rarely forms any stacked geometries with G11, while in the structures of Bourbigot *et al*. the stacking of C10 with G11 occurs in about half the structures (5 out of 11, see Supplementary Figure S6).

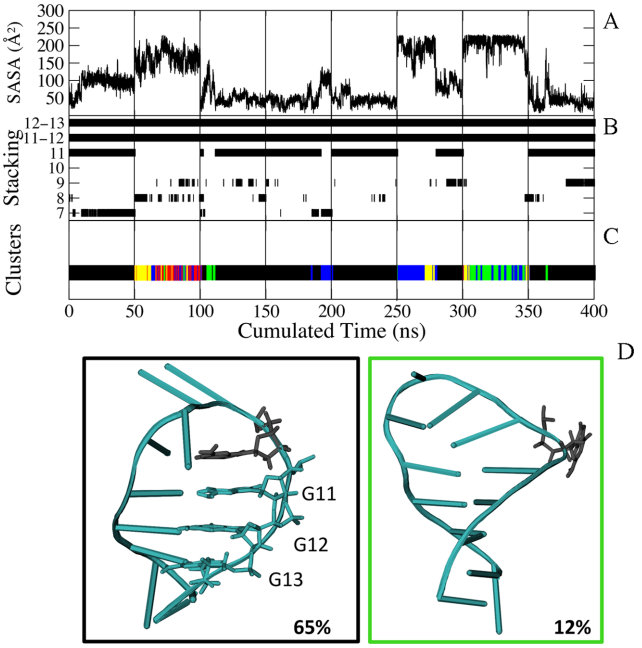


Figure 6. Distribution along the PBS10 cumulated 400-ns trajectory. See the legend of Figure 4 for details. Two out of five clusters are displayed here (see the others in the Supplementary Data, Supplementary Figure S12).

As done above, cluster analysis of the 400-ns cumulated trajectory of PBS10 focused on the 2AP10-G11-G12 region and the cumulated results are displayed in Figure 6. In total, five clusters were identified. Up to 65% of the entire population is represented by a single cluster in which 2AP10 stacks with the G11-G12-G13 stacked triad of the stem. This particular triad remains structured as such throughout all simulations, see Figure 6. The 2AP10-G11-G12-G13 stacked tetrad is oriented toward the interior of the helix-loop with 2AP essentially unexposed to solvent (SASA < 60 Å^2, Figure 6A). Within the tetrad, 2AP10 stacks with G11 with mean values of $d^{com}{}_{2AP10\text{-}G11} = 4.2 \pm 0.7$ Å and $\theta_{2AP10\text{-}G11} = 17 \pm 13°$. Importantly, the stacking of 2AP10 with the G11-G12-G13 triad forms spontaneously in trajectories that were started from NMR structures where it was initially absent, e.g. in segments 0–50, 100–150 and 200–250 ns, which were started from structures taken from both the Johnson *et al*. and Bourbigot *et al*. families (see Supplementary Table S2). Moreover, PBS10 explored conformations of the stacked tetrad (Cluster 1) over very long periods of time during which only rare excursions to other clusters were observed, (e.g. in the 100–150 ns segment). This indicates that the stacked tetrad formed by nucleobases 2AP10-G11-G12-G13 is a motif that is significantly more stable and unique when compared to all other clusters, including those formed in the PBS6 and PBS8 constructs. We therefore propose that this tetrad motif could provide a structural explanation for the ultra-short lifetime observed in the TRF experiments. The effects of multiple stacked guanines on the excited-state lifetimes of 2AP would be of interest to explore, but it does not fall within the scope of the present work.

Other clusters from the PBS10 simulations accounting for ∼23% of the population include conformations in which

2AP10 is oriented toward the exterior of the loop, but forms shorter-lived stacking motifs with T8 or G7 (clusters red, blue and yellow, Supplementary Figure S12). In these clusters, 2AP is pairwise stacked, but with geometric measures that slightly surpass those of the aromatic-stacking criterion (see, e.g. 2-AP10-T9 stacking geometries, Figure 3). It is likely that in such conformations, 2AP also undergoes static quenching. Finally, conformations where 2AP10 is not pairwise stacked and is highly exposed to the solvent with a mean SASA near the value of isolated 2AP in solution (205 ± 31 Å^2) account for 12% of the population, see Figure 6A and D. Such conformations are populated over nanoseconds and are assigned to the longest fluorescence lifetimes.

The clustering analysis of the C10-G11-G12 region for the native ΔP(-)PBS gives similar results as for PBS10 (Supplementary Figure S13). The C10-G11-G12-G13 stacked tetrad motif is found in 75% of the structures, while in the remaining structures, C10 is found outside the loop and alternatively fully exposed to solvent or stacked with neighboring bases of the loop, mainly T8. Interestingly, the larger number of clusters identified for PBS10 (5 clusters) as compared to native ΔP(-)PBS (2 clusters for position 10) indicates a larger structural heterogeneity for the 2AP10-G11-G12 segment, thus again suggesting a possible destabilizing effect of 2AP in position 10, as also observed for positions 6 and 8 and as previously reported for position 7 (59).

CONCLUSIONS

We have developed a combined MD/TRF approach to investigate the conformational heterogeneity of a DNA hairpin. The methodology introduced precise, objective criteria to enable the direct quantitative comparison of conformer populations obtained from MD simulations and lifetime amplitudes resulting from TRF spectroscopy experiments. In both experiments and MD simulations, the ΔP(-)PBS hairpin from the HIV-1 genome was studied, although the approach developed is not limited to this particular system. Three variants of this hairpin were examined, where 2AP was substituted in positions 6, 8 or 10 of the loop sequence. TRF spectroscopy studies were done covering a wide range of time scales and MD analysis was used to provide an atomistic representation in order to rationalize the distributions of excited-state lifetimes observed in the three variants.

By analyzing the relative distances and angles between stacked residues in a molecular dynamics simulation of double-stranded DNA, we defined geometric criteria to quantify the population of structures in which 2AP is in aromatic stacking interaction with other bases. Without any further hypothesis on the physical mechanisms responsible for the 2AP excited state quenching, this coarse-grained structural criterion allowed us to demonstrate that the proportion of conformers in which 2AP is in a pairwise stacked geometry correlates very well with the proportion of conformers undergoing static fluorescence quenching. The latter proportion was accurately determined from TRF spectroscopy by introducing a quantitative criterion to discriminate static versus dynamic quenching, namely the shortest time scale for local motions as determined by femtosecond fluorescence anisotropy decay measurements.

While it is broadly accepted that static quenching should be assigned to aromatic stacking of 2AP with other nucleobases, the only supportive factual argument that we are aware of is that equilibrium association constants of 2AP deoxyriboside with nucleosides in solution lie 'within the range' of expected values for base stacking (15). This present work provides further evidence that static quenching is mostly observed in conformers where 2AP is engaged in pairwise stacking with nearby nucleobases. Conversely, unstacked conformers likely undergo dynamic quenching with fluorescence lifetimes up to the ns time range, indicating that, on these time scales, structural dynamics gate the quenching process by limiting the transition of unstacked and largely water exposed excited-state 2AP to stacked conformations. We additionally showed that 2AP slightly alters the hairpin organization since (i) the populations of structures where the base in position 6 is stacked is larger by ∼25% in native ΔP(-)PBS as compared to PBS6 (see Table 2), and (ii) incorporation of 2AP in position 8 and 10 enhances structural heterogeneity. This is an important result that is in line with previous demonstrations of 2-AP-induced structural changes (53,59).

While the general trends in population were in excellent agreement between the TRF experiments and the MD simulations, the latter approach estimated lower populations of conformers in which 2AP is in aromatic stacking interaction as compared to the experimental estimates of populations undergoing static quenching. Despite the coarse-grained level of experimental and computational data analysis, the systematic difference is likely significant and could arise from several sources. In the absence of detailed quantum mechanical calculations, questions remain as to what degree the statically-quenched 2AP geometries can deviate from ideal pairwise stacking. Another factor could be that convergence in the conformational sampling was not achieved and longer simulations would be required to further refine the population estimates. To fully explore the configurational space of this hairpin region would require longer simulation times and/or alternative sampling methods such as replica exchange (60), accelerated dynamics (61) or metadynamics simulations (62). While these methods may improve sampling, one also loses information on time and the longevity of particular conformers, important properties for the interpretation of time-resolved TRF data.

After closely inspecting and taking into account the relevance and limitations of the TRF investigation of 2AP-labeled oligonucleotides, of the MD simulations and of the NMR structure investigation of (-)PBS, a precise and quantitative description of the structural heterogeneity of the native ΔP(-)PBS can be summarized as follows. Motifs that are significantly populated in the loop region of the ΔP(-)PBS hairpin are stacked triads such as C5-T6-G7 and G7-T8-T9. The latter triad can be oriented either inside or outside the loop. Other motifs with significant populations are dyads that include G7-T8 and T8-T9, C5-T6 and T6-G7. Interestingly, the latter dyad presents an orientation that is found in the NMR structures of the NCp7-(native)PBS complex (41). According to that work, G7 is oriented outside the loop and stacks with Trp37 of NCp7.

In this orientation, the T6-G7 dyad may correspond to a pre-organized arrangement that promotes NCp7 binding. In general, the structural heterogeneity around T8 is larger due to increased flexibility. For the C10 region, a largely dominant conformer corresponds to a unique, and very stable stacked tetrad, C10-G11-G12-G13 that may represent the structural motif responsible for the ultra-short lifetime observed for 2AP in this position, suggesting that multiple stacked guanines can contribute to ultra-fast (i.e. sub-ps, hence static) quenching. In all investigated positions, a significant percentage of conformers show the nucleobase being extensively exposed to solvent, in good agreement with the proportion of long fluorescence lifetimes observed experimentally.

From such a combined study, we have been able to extract detailed and quantitative information concerning the ensemble of structures populated by a DNA hairpin, providing new insights into the local structural dynamics of a physiologically important loop region of the (-)PBS of the HIV-1 genome that were only indirectly hinted upon in the two independent NMR studies. The method we proposed here for the combined analysis of TRF and MD simulation data is simple in its principle and it can be generalized to infer quantitative proportions of conformers of other flexible 2AP-labeled nucleic structures and their complexes. The approach can also be extended to evaluate the impact that protein binding might have on the conformational ensemble, as well as to evaluate the role of DNA sequence on protein recognition.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors thank Sarah Bourbigot and Nelly Morellet for providing us with their unreleased NMR structures of the ΔP(-)PBS.

FUNDING

Université de Strasbourg; Centre National de la Recherche Scientifique (CNRS); Institut de la Santé et de la Recherche Médicale (INSERM); Agence Nationale de la Recherche [FEMTOSTACK, number ANR-10-BLAN-1529-01]. Computing time was provided at the French national computing centers by GENCI, (Grand Equipement National de Calcul Intensif) and the Meso-center for High Performance Computing at the Université de Strasbourg and supported by the project EQUIP@MESO. Funding for open access charge were provided by the Agence Nationale de la Recherche.

Conflict of interest statement. None declared.

REFERENCES

1. Svoboda,P. and Di Cara,A. (2006) Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.*, **63**, 901–908.
2. Chou,S.H., Tseng,Y.Y. and Chu,B.Y. (1999) Stable formation of a pyrimidine-rich loop hairpin in a cruciform promoter. *J. Mol. Biol.*, **292**, 309–320.
3. Glucksmann-Kuis,M.A., Dai,X., Markiewicz,P. and Rothman-Denes,L.B. (1996) E. coli SSB activates N4 virion RNA polymerase promoters by stabilizing a DNA hairpin required for promoter recognition. *Cell*, **84**, 147–154.
4. Slama-Schwok,A., Brossalina,E., Demchenko,Y., Best-Belpomme,M. and Vlassov,V. (1998) Structural flexibility of a DNA hairpin located in the long terminal repeat of the Drosophila 1731 retrotransposon. *Nucleic Acids Res.*, **26**, 5142–5151.
5. Yingling,Y.G. and Shapiro,B.A. (2005) Dynamic behavior of the telomerase RNA hairpin structure and its relationship to dyskeratosis congenita. *J. Mol. Biol.*, **348**, 27–42.
6. Godet,J., Kenfack,C., Przybilla,F., Richert,L., Duportail,G. and Mely,Y. (2013) Site-selective probing of cTAR destabilization highlights the necessary plasticity of the HIV-1 nucleocapsid protein to chaperone the first strand transfer. *Nucleic Acids Res.*, **41**, 5036–5048.
7. Goel,T., Mukherjee,T., Rao,B.J. and Krishnamoorthy,G. (2010) Fluorescence dynamics of double- and single-stranded DNA bound to histone and micellar surfaces. *J. Phys. Chem. B*, **114**, 8986–8993.
8. Hariharan,C., Bloom,L.B., Helquist,S.A., Kool,E.T. and Reha-Krantz,L.J. (2006) Dynamics of nucleotide incorporation: snapshots revealed by 2-aminopurine fluorescence studies. *Biochemistry*, **45**, 2836–2844.
9. Lenz,T., Bonnist,E.Y., Pljevaljcic,G., Neely,R.K., Dryden,D.T., Scheidig,A.J., Jones,A.C. and Weinhold,E. (2007) 2-Aminopurine flipped into the active site of the adenine-specific DNA methyltransferase M.TaqI: crystal structures and time-resolved fluorescence. *J. Am. Chem. Soc.*, **129**, 6240–6248.
10. Jones,A.C. and Neely,R.K. (2015) 2-aminopurine as a fluorescent probe of DNA conformation and the DNA-enzyme interface. *Q. Rev. Biophys.*, **48**, 244–279.
11. Jiao,Y.G., Stringfellow,S. and Yu,H.T. (2002) Distinguishing "looped-out" and "stacked-in" DNA bulge conformation using fluorescent 2-aminopurine replacing a purine base. *J. Biomol. Struct. Dyn.*, **19**, 929–934.
12. Larsen,O.F., van Stokkum,I.H., Gobets,B., van Grondelle,R. and van Amerongen,H. (2001) Probing the structure and dynamics of a DNA hairpin by ultrafast quenching and fluorescence depolarization. *Biophys. J.*, **81**, 1115–1126.
13. Jose,D., Datta,K., Johnson,N.P. and von Hippel,P.H. (2009) Spectroscopic studies of position-specific DNA "breathing" fluctuations at replication forks and primer-template junctions. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 4231–4236.
14. Guest,C.R., Hochstrasser,R.A., Sowers,L.C. and Millar,D.P. (1991) Dynamics of mismatched base pairs in DNA. *Biochemistry*, **30**, 3271–3279.
15. Rachofsky,E.L., Osman,R. and Ross,J.B.A. (2001) Probing structure and dynamics of DNA with 2-aminopurine: effects of local environment on fluorescence. *Biochemistry*, **40**, 946–956.
16. Avilov,S.V., Piemont,E., Shvadchak,V., de Rocquigny,H. and Mély,Y. (2008) Probing dynamics of HIV-1 nucleocapsid protein/target hexanucleotide complexes by 2-aminopurine. *Nucleic Acids Res.*, **36**, 885–896.
17. Gelot,T., Tourón-Touceda,P., Crégut,O., Léonard,J. and Haacke,S. (2012) Ultrafast site-specific fluorescence quenching of 2-aminopurine in a DNA hairpin studied by femtosecond down-conversion. *J. Phys. Chem. A*, **116**, 2819–2825.
18. Godet,J., Ramalanjaona,N., Sharma,K.K., Richert,L., de Rocquigny,H., Darlix,J.L., Duportail,G. and Mely,Y. (2011) Specific implications of the HIV-1 nucleocapsid zinc fingers in the annealing of the primer binding site complementary sequences during the obligatory plus strand transfer. *Nucleic Acids Res.*, **39**, 6633–6645.
19. Manoj,P., Min,C.-K., Aravindakumar,C.T. and Joo,T. (2008) Ultrafast charge transfer dynamics in 2-aminopurine modified double helical DNA. *Chem. Phys.*, **352**, 333–338.
20. Nordlund,T.M., Andersson,S., Nilsson,L., Rigler,R., Graslund,A. and McLaughlin,L.W. (1989) Structure and dynamics of a fluorescent DNA oligomer containing the EcoRI recognition sequence: fluorescence, molecular dynamics, and NMR studies. *Biochemistry*, **28**, 9095–9103.
21. Somsen,O.J.G., Hoek,V.A. and Amerongen,V.H. (2005) Fluorescence quenching of 2-aminopurine in dinucleotides. *Chem. Phys. Lett.*, **402**, 61–65.

22. Wan,C., Xia,T., Becker,H.-C. and Zewail,A.H. (2005) Ultrafast unequilibrated charge transfer: a new channel in the quenching of fluorescent biological probes. *Chem. Phys. Lett.*, **412**, 158–163.
23. Larsen,O.F.A., van Stokkum,I.H.M., Gobets,B., van Grondelle,R. and van Amerongen,H. (2001) Probing the structure and dynamics of a DNA hairpin by ultrafast quenching and fluorescence depolarization. *Biophys. J.*, **81**, 1115–1126.
24. Jean,J.M. and Hall,K.B. (2001) 2-aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 37–41.
25. Liang,J. and Matsika,S. (2011) Pathways for fluorescence quenching in 2-aminopurine pi-stacked with pyrimidine nucleobases. *J. Am. Chem. Soc.*, **133**, 6799–6808.
26. Daura,X., Suter,R. and van Gunsteren,W.F. (1999) Validation of molecular simulation by comparison with experiment: rotational reorientation of tryptophan in water. *J. Chem. Phys.*, **110**, 3049–3055.
27. Ichiye,T. and Karplus,M. (1983) Fluorescence depolarization of tryptophan residues in proteins: a molecular dynamics study. *Biochemistry*, **22**, 2884–2893.
28. Schroder,G.F., Alexiev,U. and Grubmuller,H. (2005) Simulation of fluorescence anisotropy experiments: probing protein dynamics. *Biophys. J.*, **89**, 3757–3770.
29. Yeh,I.C. and Hummer,G. (2002) Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J. Am. Chem. Soc.*, **124**, 6563–6568.
30. Sigtryggsdottir,A.R., Papaleo,E., Thorbjarnardottir,S.H. and Kristjansson,M.M. (2014) Flexibility of cold- and heat-adapted subtilisin-like serine proteinases evaluated with fluorescence quenching and molecular dynamics. *Biochim. Biophys. Acta*, **1844**, 705–712.
31. Vaiana,A.C., Neuweiler,H., Schulz,A., Wolfrum,J., Sauer,M. and Smith,J.C. (2003) Fluorescence quenching of dyes by tryptophan: Interactions at atomic detail from combination of experiment and computer simulation. *J. Am. Chem. Soc.*, **125**, 14564–14572.
32. Hall,K.B. and Williams,J. (2004) Dynamics of the IRE RNA hairpin loop probed by 2-aminopurine fluorescence and stochastic dynamics simulations. *RNA*, **10**, 34–47.
33. Brovarets,O.O. and Hovorun,D.M. (2011) IR Vibrational spectra of H-bonded complexes of adenine, 2-aminopurine and 2-aminopurine(+) with cytosine and thymine: Quantum-chemical study. *Opt. Spectr.*, **111**, 750–757.
34. Larsen,O.F.A., van Stokkum,I.H.M., de Weerd,F.L., Vengris,M., Aravindakumar,C.T., van Grondelle,R., Geacintov,N.E. and van Amerongen,H. (2004) Ultrafast transient-absorption and steady-state fluorescence measurements on 2-aminopurine substituted dinucleotides and 2-aminopurine substituted DNA duplexes. *Phys. Chem. Chem. Phys.*, **6**, 154–160.
35. Lobsiger,S., Blaser,S., Sinha,R.K., Frey,H.M. and Leutwyler,S. (2014) Switching on the fluorescence of 2-aminopurine by site-selective microhydration. *Nat.. Chem*, **6**, 989–993.
36. Darlix,J.L., Godet,J., Ivanyi-Nagy,R., Fosse,P., Mauffret,O. and Mely,Y. (2011) Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J. Mol. Biol.*, **410**, 565–581.
37. Léonard,J., Gelot,T., Torgasin,K. and Haacke,S. (2011) Ultrafast fluorescence spectroscopy of biologically relevant chromophores using type II difference frequency generation. *J. Phys.Conf. Ser.*, **277**, 012017.
38. Livesey,A.K. and Brochon,J.C. (1987) Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method. *Biophys. J.*, **52**, 693–706.
39. Esposito,R., Altucci,C. and Velotta,R. (2013) Analysis of simulated fluorescence intensities decays by a new maximum entropy method algorithm. *J. Fluoresc.*, **23**, 203–211.
40. Johnson,P.E., Turner,R.B., Wu,Z.R., Hairston,L., Guo,J., Levin,J.G. and Summers,M.F. (2000) A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry*, **39**, 9084–9091.
41. Bourbigot,S., Ramalanjaona,N., Boudier,C., Salgado,G.F., Roques,B.P., Mely,Y., Bouaziz,S. and Morellet,N. (2008) How the HIV-1 nucleocapsid protein binds and destabilises the (-)primer binding site during reverse transcription. *J. Mol. Biol.*, **383**, 1112–1128.
42. Brooks,B.R., Brooks,C.L., Mackerell,A.D., Nilsson,L., Petrella,R.J., Roux,B., Won,Y., Archontis,G., Bartels,C., Boresch,S. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.
43. Foloppe,N. and MacKerell,A.D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, **21**, 86–104.
44. Auffinger,P. and Westhof,E. (1996) H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys. J.*, **71**, 940–954.
45. Auffinger,P. and Westhof,E. (1997) RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J. Mol. Biol.*, **269**, 326–341.
46. Caves,L.S., Evanseck,J.D. and Karplus,M. (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.*, **7**, 649–666.
47. Lu,X.J., Babcock,M.S. and Olson,W.K. (1999) Overview of nucleic acid analysis programs. *J. Biomol. Struct. Dyn.*, **16**, 833–843.
48. Jean,J.M. and Krueger,B.P. (2006) Structural fluctuations and excitation transfer between adenine and 2-aminopurine in single-stranded deoxytrinucleotides. *J. Phys. Chem. B*, **110**, 2899–2909.
49. Nagan,M.C., Beuning,P., Musier-Forsyth,K. and Cramer,C.J. (2000) Importance of discriminator base stacking interactions: molecular dynamics analysis of A73 microhelix(Ala) variants. *Nucleic Acids Res.*, **28**, 2527–2534.
50. Norberg,J. and Nilsson,L. (1995) Stacking free-energy profiles for all 16 natural ribodinucleoside monophosphates in aqueous-solution. *J. Am. Chem. Soc.*, **117**, 10832–10840.
51. Gabb,H.A., Sanghani,S.R., Robert,C.H. and Prevost,C. (1996) Finding and visualizing nucleic acid base stacking. *J. Mol. Graph.*, **14**, 6–11.
52. Gorska,A., Jasinski,M. and Trylska,J. (2015) MINT: software to identify motifs and short-range interactions in trajectories of nucleic acids. *Nucleic Acids Res.*, **43**, e114.
53. Dallmann,A., Dehmel,L., Peters,T., Mugge,C., Griesinger,C., Tuma,J. and Ernsting,N.P. (2010) 2-Aminopurine incorporation perturbs the dynamics and structure of DNA. *Angew. Chem. Int. Ed. Engl.*, **49**, 5989–5992.
54. Feig,M., Karanicolas,J. and Brooks,C.L. 3rd (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.*, **22**, 377–395.
55. Kelley,S.O. and Barton,J.K. (1999) Electron transfer between bases in double helical DNA. *Science*, **283**, 375–381.
56. Wan,C., Fiebig,T., Schiemann,O., Barton,J.K. and Zewail,A.H. (2000) Femtosecond direct observation of charge transfer between bases in DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 14052–14055.
57. Lakowicz,J.R. (2006) *Principles of fluorescence spectroscopy*. 3rd edn. Springer, NY.
58. Barman,N., Singha,D. and Sahu,K. (2013) Fluorescence quenching of hydrogen-bonded coumarin 102-phenol complex: effect of excited-state hydrogen bonding strength. *J. Phys. Chem. A*, **117**, 3945–3953.
59. Sholokh,M., Sharma,R., Shin,D., Das,R., Zaporozhets,O.A., Tor,Y. and Mely,Y. (2015) Conquering 2-aminopurine's deficiencies: highly emissive isomorphic guanosine surrogate faithfully monitors guanosine conformation and dynamics in DNA. *J. Am. Chem. Soc.*, **137**, 3185–3188.
60. Sugita,Y. and Okamoto,Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, **314**, 141–151.
61. Hamelberg,D., Mongan,J. and McCammon,J.A. (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, **120**, 11919–11929.
62. Laio,A. and Parrinello,M. (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12562–12566.