Enforcing local DNA kinks by sequence-selective trisintercalating oligopeptides of a tricationic porphyrin. A polarizable Molecular Dynamics study.

Nohad Gresh*[a], Krystel El Hage[b], Louis Lagardère[a], Frédérique Brégier[c], Jérémy Godard[c], Jean-Philip Piquemal[a], Martine Perrée-Fauvet[d], Vincent Sol*[c]

[a]Laboratoire de Chimie Théorique, UMR 7616 CNRS, Sorbonne Université, F-75005 Paris

[b]Qubit Pharmaceuticals, Faculté de Médecine, 24, rue du Faubourg Saint-Jacques, 75014 Paris

[c]LABCis UR 22722, Université de Limoges, Faculté des Sciences et Techniques, 87060 Limoges, France

[d]Laboratoire de Chimie Biorganique et Bioinorganique, Université Paris-Orsay, 912405 Orsay

Summary. Bisacridinyl-bisarginyl porphyrin (BABAP) is a trisintercalating derivative of a tricationic porphyrin, formerly designed and synthesized in order to selectively target and photosensitize the ten-base pair palindromic sequence d(CGGGCGCCCG)$_2$. We resorted to the previously derived (Far et al., 2004) lowest energy-minimized (EM) structure of the BABAP complex with this sequence as a starting point. We performed polarizable molecular dynamics (MD) on this complex. It showed, over a 150 ns duration, the persistent binding of the Arg side-chain on each BABAP arm to the two G bases upstream from the central porphyrin intercalation site. We subsequently performed progressive shortenings of the connector chain linking the Arg-Gly backbone to the acridine, from n=6 methylenes to 4, followed by removal of the Gly backbone and further connector shortenings, from n=4 to n=1. These resulted into progressive deformations ('kinks') of the DNA backbone. In its most accented kinked structure, the DNA backbone was found to have a close overlap with that of DNA bound to *Cre* recombinase, with, at the level of one acridine intercalation site, negative roll and positive tilt values consistent with those experimentally found for this DNA at its own kinked dinucleotide sequence. Thus, in addition to their photosensitizing properties, some BABAP derivatives could induce sequence-selective, controlled DNA deformations, which are targets for cleavage by endonucleases or for repair enzymes.

Introduction.

The d(GGCGCC)$_2$ palindrome is encountered in oncogenic (1) and retroviral sequences (2), and was recently found at three insertion sites of *Alu* repeats (3). *Alu* repeats are transposon polynucleotides which can detach from the genome and integrate into other DNA sequences, as well as into extrachromosomal circulating DNA (ec-DNA). Ec-DNA is identified as a major cause in tumor progression, and is a much more accessible target than nuclear DNA (4) (5). Such occurrences are a strong incentive to design DNA-binding ligands which could bind with augmented selectivity to this palindrome compared to all possible hexameric sequences, of which there is a total of 2080 (6).

Toward this aim, we have previously designed oligopeptide derivatives of two intercalators. The first is the anthraquinone ring of the antitumor drugs ametantrone (AMT) and mitoxantrone (MTX) (7) . The second is a cationic porphyrin, which acts as a photosensitizer in dynamic phototherapy (8) (9). We previously reported the results of joint computational and experimental studies regarding both anthraquinone (10) and porphyrin (11) oligopeptides. A short-duration molecular dynamics (MD) study was reported in a 2010 paper (10), but most computational studies had resorted so far to energy-minimization. Advances in the Tinker-HP software (12) recently operational on GPU supercomputers (13) have paved the way to long-duration MD simulations of the complexes of oligonucleotides with newly-designed mono- and trisintercalating MTX derivatives (14). Such simulations are now extended to bisacridinyl-bisarginyl porphyrin (BABAP) and derivatives.

The molecular structure of BABAP is represented in Figure 1. In its design, while the porphyrin ring intercalates at the central d(CpG)$_2$ step of the palindrome, the Arg side-chain of each arm binds along one DNA strand in the major groove to the two successive G bases upstream. The Arg-Gly backbone is connected to an amino-acridine (AA) by a C6-paraffinic chain, and each AA ring intercalates at a d(CpG)$_2$ site upstream. Thus the overall targeted sequence is d(C GGGC GCCC G)$_2$, a blank denoting an intercalation site. Both circular dichroism experiments and topoisomerase I unwinding experiments lent support to trisintercalation in the BABAP complex with an oligonucleotide with this decanucleotide as a central sequence, while mono- and bisintercalation took place, respectively, with BAP (devoid of the two AA extension) and MABAP having one single AA intercalator (11) . The present study starts from the lowest energy-minimum structure of the complex reported in (11), in which intercalation of both AA arms takes place from the major groove. We first evaluate the extent to which long-duration MD with a multipolar polarizable potential in a water bath with counterions retains the most important feature from EM. Polarizable potentials are known to display high accuracy for polar and charged systems, including proteins and nucleic acids (15-18). Because AA intercalation is found to take place without significant distortion of the sugar-phosphate backbone between the central and both AA intercalation sites, namely at steps d(GGGC), this led us to next address the question: as long as AA intercalation remains enabled, what could be the impact of progressive shortenings of the connector on

the conformation of this step, and could local 'kinks' be sequence-selectively enforced by some shorter-chain BABAP derivatives?

Kinked DNA conformations were reported earlier in complexes with proteins (19-25). This list is not exhaustive. It encompasses proteins such as CAP (21), (22), *Cre* recombinase (23), the CENP centromere protein (24) , the TATA-binding box (25). The binding of intercalators and steroidal diamines was also reported to stabilize kinks. This was experimentally reported for ethidium bromide and actinomycin (26), irehdiamine (26), (27), (28) , and, in the context of a JUMNA (29) energy-minimization study, for dipyrandium (30).

In the present study, the C6 chain connecting the Arg-Gly backbone is first shortened by one and two methylene groups, yielding the C5 and C4 analogs, respectively, of BABAP. This is followed by removal of the Gly backbone, and by up to three methylene groups, yielding the –C4 to -C1 Arg analogs of BABAP. The structural formulas of BABAP and its shorter-chain analogs are given in Figure 1. For each analog, we repeat an identical computational protocol, as detailed below. It consists of restrained energy-minimization then NVT equilibration, followed by limited-duration restrained MD and long-duration, unrestrained MD production. At its outcome, the most important structural features of the complexes and the complexed DNA conformation are analyzed.

Computational procedure.

We used the same procedure as in (14). MD simulations use the AMOEBA polarizable multipole potential (31) (and Refs. therein) coded in the Tinker-HP software (12) in its recent GPU implementation (13). We use the 2018 DNA parameters (31). The distributed multipoles of all derivatives compounds were derived by a Stone analysis (32) on their molecular wave functions computed with the B3LYP DFT functional (33) (34) and a cc-pVDZ basis set (35) (36) with the G09 (37) software.

All simulations were done on the 18-base palindromic sequence d(CGTA**C G**GG**C G**CCC **G**TACG)$_2$, the same sequence as in the complex of a trisintercalating ligand, denoted **III** in (14). **III** is an oligopeptide derivative of an anthraquinone drug, mitoxantrone, the two arms of which were extended with an oligopeptide motif with an alkydiammonium side-chain connected by an amide-imidazole motif to an AA ring (14).

The starting trisintercalated DNA conformation was extracted from the previous MD simulation of its complex with ligand **III**. In a first step, manual docking was performed with the help of computer graphics with the Insight II software (Accelrys Inc., San Diego). The porphyrin ring was intercalated in the central d(CpG)$_2$ sequence, locating its three cationic N-methylpyridinium rings in the minor groove,

and the phenyl ring in the major groove. This is the ring location corresponding to the lowest energy minima of the complexes of BABAP (11) and BAP (38) with oligonucleotides having the target central palindromic sequence. On each DNA strand, the guanidinium side-chain was oriented in order to face the two successive G bases upstream from the central intercalation site, subsequently denoted as G_4-G_3/$G_{4'}$-$G_{3'}$. Appropriate torsional rotations around the C-C and C-N bonds of the connector then enabled AA intercalation in the corresponding $d(C_1pG_2).d(C_9pG_{10})$ or $d(C_{1'}pG_{2'}).d(C_{9'}pG_{10'})$ intercalation site of the target ten base-pair sequence. Manual docking was completed with a first round of constrained energy-minimization, bearing on the sole ligand in the presence of the following constraints: for each Arg side-chain, a 2.0 Å distance between the imino proton and O_6 of G_3/$G_{3'}$ and that of the *cis*-amino proton and O_6 of G_4/$G_{4'}$; for each AA ring, a distance of 3.8 Å between its N atom and the N_3 atom of the two G bases and the two C bases of the corresponding intercalation site. Intramolecular restraints were added. To retain planarity of each AA ring, two distance restraints of 7.3 Å were set between the two outermost C atoms, namely between atoms C_2 and C_7 and atoms C_3 and C_8. The planarity of all amide bonds were enforced by distance restraints of 3.15 Å between the carbonyl O and the amide H atom.

The DNA-ligand complex was then immersed in a bath of water molecules in a box of dimensions 51, 54 and 90 Å on the x, y, and z dimensions. Neutrality was ensured by placing 29 Na^+ replacing the last 29 water molecules of the water bath. Periodic boundary conditions (PBC) were applied along with Smooth Particle Mesh Ewald (PME) (39). We used cutoff values of 12 Å and 9 Å for van der Waals and Ewald interactions, respectively.

We then followed a similar procedure to that of (14).

Restrained energy-minimization was resumed, now also relaxing DNA, water and counterions. Molecular Dynamics was subsequently run. Equilibration was started by 50 K stepwise raises in temperature for a duration of 1 ns at constant volume, from 0 to 300 K. Production was then started at 300 K with a Bussi thermostat (40) and at constant pressure 1 atm using a Monte-Carlo barostat. Coordinates were saved every 100 ps.

We have retained the same inter- and intramolecular constraints as the preliminary stage. Additional restraints were added in the NVT and first 48ns NPT stages, destined to preserve the H-bonds between the complementary WC bases at, and capping, the three intercalation sites. These were set to prevent premature distortions of these base pairs. In addition, as motivated in (14), six additional restraints were introduced, which prevent the 'fraying' of the two end GC base-pairs in the course of MD, namely 2.0 Å between the corresponding O_6-HN_4, N_1H-N_3, and N_2H-O_2 atoms. This inclusion has precedents in DNA MD simulations including those with polarizable potentials (31). This was justified by the limited length of the oligonucleotide. It prevents unwanted 'danglings' of the end base-pairs.

During the first 48 ns production stage, the ligand-DNA distance restraints were maintained, as well as those between the above-mentioned complementary H-bonded bases. This was motivated in Ref. (14): large amplitude motion of DNA and the solvent were observed during the initial phases of the production

runs, which cause an excess kinetic energy transfer to the bound ligand, which could destabilize and disrupt its complex with DNA, as well as distort the H-bond patterns of the bases. Unrestrained MD was resumed past 48ns, retaining, however, the intramolecular restraints enforcing acridine, guanidinium and formamide planarity and the six restraints enforcing H-bonding between the two-end base-pairs. The length of such unrestrained production runs was in the 150-200 ns range.

Structural analyses of the DNA at the outcome of MD were done with the CURVES+ software (41).

The intermolecular interaction energies between the four bases, C_1, G_2, $C_{5'}$, $G_{6'}$ making up one acridine intercalation site, $d(C_1G_2).d(C_{5'}G_{6'})$, and the acridine ring were computed by DFT quantum chemistry and decomposed into their five contributions: Coulomb (E_C), short-range repulsion (E_X), polarization (E_{pol}), charge-transfer (E_{ct}) and dispersion (E_{disp}). This was done by the ALMOEDA procedure (42) coded in the QChem package (43). We resorted to the ωB97-XD DFT functional (44) and the cc-pVTZ(-f) basis set (36).

Results and discussion.

BABAP. A representation of the last MD pose of the complex of BABAP with the target oligonucleotide is given in Figures 2a-c. Figure 2a gives an overall representation of the complex. Figures 2b-c show the detailed H-bond interactions which take place on each strand between the imino and *cis*-amino protons of the Arg side-chain and O_6/N_7 of the two targeted bases, G_4-G_3/$G_{4'}$-$G_{3'}$ as well as the H-bond between the Arg main-chain NH and $G_6/G_{6'}$ of the central intercalation site. Figures 3a and 3b show the (normalized) distribution functions of the H-bond distances between the first and second Arg side-chain, respectively, and the two targeted G bases. Figure 3c shows the corresponding distributions for the H-bond between each Arg main-chain NH group and O_6 of bases $G_6/G_{6'}$ of the central intercalation site. Figures 4a-c show the corresponding time evolutions of these distances. The structural results are fully consistent with the energy-minimization reported earlier (11). Each Arg side-chain can align along the z axis enabling it to bind to the two successive G bases on the same strand. This mode of binding, rather than the more frequent one in which an Arg side-chain binds coplanar to a single G base in a bidentate fashion, was reported in an X-ray study of the complex between yeast RAP1 protein and telomeric DNA. Thus Arg_{404} was found to bind to N_7 of two successive G bases, $G_{8'}$ and $G_{9'}$ on the same strand (45). In the present complex, however, the interaction takes place preferentially with O_6. With both arms, the distance distributions are much better peaked around O_6 than around N_7 (figures 3a-b). The distributions around O_6 of $G_6/G_{6'}$ (figure 3c) are shallower, possibly due to the neutrality of the main-chain Arg NH group, compared to the partly positively charged imino and amino side-chain groups.

Shorter BABAP derivatives.

1) Arg-Gly with a C5 or a C4 connector.

Using the one-letter code for the peptide backbone, these derivatives will be denoted RG-C5 and RG-C4. We retain the acronym BABAP for the initial derivative rather than RG-C6.

RG-C5. An overall view is given in Supp. Info S1a, while the H-bond distances with G_3-G_4/$G_{3'}$-$G_{4'}$ are represented in Supp. Info S1b-c.

RG-C4. An overall view is given in Supp. Info S2a, while the H-bond distances with G_3-G_4/$G_{3'}$-$G_{4'}$ are represented in Supp. Info S2b-c.

Figure 5 represents the structure of DNA bound by BABAP, along with those with RG-C5 and RG-C4. It displays the distance between one P atom, P_8, of the central intercalation site, and the two P atoms, P_{14} and P_{31}, of the $d(TpA)_2$ step six base-pairs downstream. P_8 is closer to P_{31} on the opposite strand by 4.9 Å in the BABAP complex. This difference is comparable to those found with RG-C5, namely 4.1 Å and it increases to 5.7 Å with RG-C4.

We next monitor the impact of further shortenings of the connecting chain, which follow the removal of the Gly backbone.

2) Arg with C4 to C1 connectors. With the one-letter code, these derivatives will be denoted as R-C4 till R-C1.

Representations of the overall complexes of R-C4 till R-C2, and of the Arg H-bonds with the G bases are given as Supp. Info S3-S5. Figures 6a-d represent the structures of DNA in its complexes with the four derivatives focusing on the P_8-P_{14} and P_8-P_{31} distances. Figure 7a gives a representation of the kinked DNA---R-C1 complex and figures 7b-c display the relevant H-bonds between the G bases and Arg from the last MD pose. On the unprimed strand, the interactions of the Arg side-chain involve only the amino hydrogens. Each N atom donates its H atom *trans* to the imino proton to N_7 of G_4, and both amino hydrogens belonging to the N atom *cis* to the imino proton are donated to N_7 to G_3. Owing to the 'pinching' of the major groove, the amino H *cis* to the imino proton is also donated to N_7 of G_2, one more step upstream, and on the 5'site of the AA intercalation site. The carbonyl O of the peptide bond preceding the Arg residue accepts a proton from the extracyclic amino group of cytosine C_5 on the 3'side of the porphyrin intercalation site. On the primed strand, the imino proton and its *cis*-amino proton bind to $G_{4'}$ and $G_{3'}$, respectively but with N_7 rather than with O_6 as was the case with BABAP. The main-chain NH is donated to N_7 of G_6 of the intercalation site on the unprimed strand.

Figures 8a-b give the normalized distributions of the relevant Arg-guanine H-bond distances. On the unprimed strand, three of the five H-bonds represented in Figure 7a have their peaks at 2.2-2.3 Å. The first H-bond involving the proton *cis* to the imino proton (denoted as N_cH_c) and N_7 of G_3, has a slightly shifted peak at 2.35 Å, while the second H-bond, donated to N_7 of G_2, is shifted to 2.6 Å. On the primed strand, the distributions are much more accented with marked peaks at similar distances, 2.1 and 2.2 Å

for the imino and *cis*-amino hydrogens, respectively. The corresponding structural informations of the DNA complexes with R-C4 till R-C2 are given as Supp. Infos 3-6.

A revealing indicator of the progressive ligand-induced DNA kink is the P_8-P_{31} distance. It decreases from 27.3 Å in the BABAP complex to 24.3 Å in those with RG-C5 and RG-C4, then to 20.7-20.5 Å with R-C4 and R-C3, and has its smallest values, 17.4-18.0 Å, with R-C2 and R-C1. The difference between the P_8-P_{14} and P_8-P_{31} distances has increased between the two 'extreme' DNA ligands, BABAP and R-C1, from 4.9 to 10 Å. Relevant P-P distances are reported in Table I.

Figure 9 recasts the structures of DNA in its R-C1 complex, along with its structure extracted from its complex with the *Cre* recombinase (23). The R-C1 complex bears a striking resemblance with the kinked *Cre*-bound segment of DNA. The two P-P distances corresponding to the P_8-P_{14} and P_8-P_{31} ones amount to 15.6 and 26.1 Å, closely comparable to the respective ones of 18.0 and 28.0 Å in R-C1. Such structural similarities imply that the R-C1 induced kinks are experimentally accessible. It could also be noted that the bases at the level at the kink site (denoted as 'the second and third base pairs from one end of the central crossover region' in Guo et al. 1999) have a negative roll of -49° and a positive tilt of 16°. The roll and tilt values found for the R-C1 complex at the AA intercalation site, as determined by the Curves+ software, are -72° and 33°, respectively. Such negative roll and positive tilt values could not, however, be generalized to the other complexes from the present study, nor in the other experimental DNA-protein complexes from Refs. (20-22, 24-25), which have less accented DNA kinks. For the last pose of the BABAP and of the R-C1 DNA complexes, we have computed the intermolecular interaction energies stabilizing the stacking of the acridine ring with the four bases, C_1, G_2, $C_{5'}$ and $G_{6'}$ of the $d(C_1G_2).(C_{5'}G_{6'})$ dinucleotide. The acridine ring had its amide group substituted with a terminal -CH_3 group. The four bases were disconnected from their sugars and the valence of their N atoms, N_1 or N_9, completed by an H atom replacing the $C_{1'}$ sugar. ΔE_{tot} embodies the acridine-base interactions and the base-base interactions. The results of the ALMOEDA energy analyses are reported in Table II. ΔE_{tot} has close values in the BABAP and R-C1 complexes, and the individual energy contributions have comparable magnitudes in both. This indicates that the stacking of the acridine is not impaired by the distortion of DNA that occurs in the R-C1 complex.

Conclusions and Perspectives.

The present polarizable MD simulations show the possibility of inducing sequence-selective, progressive deformations on DNA along a series of its complexes with trisintercalating oligopeptide derivatives. In this study, the starting ligand is BABAP (bisacridinyl-bisarginyl porphyrin). It has a tricationic porphyrin which intercalates at the central $d(CpG)_2$ sequence of the $d(C_1G_2G_3G_4C_5G_6C_7C_8C_9G_{10})_2$ palindrome. The phenyl ring is substituted by an Arg-Gly (RG) peptide connected by a C6 linker to an AA ring which intercalates at an end $d(CpG)_2$ sequence. Trisintercalation was previously demonstrated by circular dichroism and by Topoisomerase I unwinding experiments

(11). MD showed the persistence of the H-bonds between the imino proton of each Arg side-chain and O_6 of $G_4/G_{4'}$ and of its *cis*-amino proton with O_6 of $G_3/G_{3'}$. There were no distortions of the sugar-phosphate backbone along the non-intercalated sequences. Progressive shortenings of the linker chain to the AA ring were subsequently performed, from C6 to C4, followed by removal of the Gly backbone and further shortenings, from C4 till C1. MD simulations on the six DNA-ligand complexes showed DNA to undergo distortions ('kinks') of increasing magnitude. A sensitive indicator is the distance between one P atom of the central intercalation site, denoted as P_8, and, on the opposite strand, the P atom (P_{31}) of the $d(TpA)_2$ dinucleotide six base-pairs downstream. It underwent decreases from 27.3 Å in the BABAP complex, to 24.3 Å in the RG-C5 and RG-C4 complexes, to app. 20.5 Å in the R-C4 and R-C3 ones, to 17.4-18.0 Å in the R-C2 and R-C1 complexes.

The kinked structure of the DNA-R-C1 complex bears a striking resemblance to that of a DNA complexed with *Cre* recombinase. It is highlighted in Figure 9. This proves that such 'extreme' kinked structures could be energetically accessible to DNA. Small DNA-ligands are also able to kink DNA. Early examples are provided by steroid diamines (26) (27) (28) and the bisintercalator ditercalinium (46) (47), even though the DNA deformation (48) is less than with the extreme compounds R-C2 and R-C1. Kinked DNA structures are hypersensitive sites for cleavage by endonucleases (49). This implies that, along with their DNA photosensitizer properties, some BABAP derivatives could also, in the presence of endonucleases, act to induce sequence-selective DNA cleavage.

Induction of kinks of modulable amplitudes could, alternatively, be enabled by replacing one, or both, AA intercalator(s), by groove binders having varying lengths and appropriate shape with polar groups destined to bind O_6/N_7 of $G_2/G_{1'}$, and/or bind to the extracyclic $-NH_2$ of $C_1/C_{5'}$ (Gresh et al., work in progress). Conversely, intercalation into the $d(C_1-G_2).d(C_{5'}G_{6'})$ sequence could be enhanced with conjugated rings with augmented stacking affinities for it. This could be achieved with intercalators having enlarged rings, the kinks being enforced, as with AA, by controlled connector shortening. One possibility currently investigated is AA replacement by a Ru(II) polypyridyl complex of Ru(II) (50). Induction of kinked DNA was also found by us in a novel series of anthraquinone derivatives related to compound **III** of (14) and will be reported in due course.

Along with sequence-selectivity and binding affinities, novel DNA-binding oligopeptide derivatives of intercalators could thus be tailored in order to *de novo* modulate and control the amplitude of the kink they induce in the course of dynamics. An initial docking at a predefined kinked DNA conformation of related compounds is first done by energy-minimization. The structures of the complexes are then evolved in the course of MD, leading to the selection of those derivatives with connectors enforcing and maintaining the kinked conformation over the duration of MD. Accordingly, the structures of the DNA complexes with BABAP and R-C1 are provided as Supp. Info. These should enable such simulations on related derivatives having a central intercalating ring from other families, such as anthraquinone, anthracycline, or aminoacridine. Additional data are available upon request.

Table I. Values (Å) of the across-strand P8-P31 and along-strand P8-P14 distances and values of their differences in the DNA complexes with BABAP and derivatives with progressive linker shortenings (see text for definition).

Ligand	P8-P31	P8-P14	Difference (P8-P14)-(P8-P31
BABAP (RG-C6)	27.3	32.2	4.9
RG-C5	24.3	28.4	4.1
RG-C4	24.2	29.9	5.7
R-C4	20.2	27.6	7.4
R-C3	20.7	28.1	7.4
R-C2	17.4	28.6	11.2
R-C1	18.0	28.0	10.0

Table II. Intermolecular interaction energies in the stacked complex of the acridine ring with the four bases, $C_1, G_2, C_{5'}, G_{6'}$, of the $d(C_1G_2).d(C_{5'}G_{6'})$ sequence extracted from the complex of BABAP and R-C1 ligands with the target 18-mer in the last MD pose. See text for definition. Energies in kcal/mol.

	BABAP-complex	R-C1 complex
E_C	-68.6	-69.8
E_X	79.3	82.3
E_1	10.7	12.5
E_{pol}	-17.4	-18.9
E_{ct}	-20.3	-20.6
E_{disp}	-48.1	-47.1
ΔE_{tot}	**-75.2**	**-74.0**

Figure captions.

Figures 1a-d. Representation of the molecular structures of BABAP, and its shorter derivatives RG-C5, RG-C4, and R-C4 till R-C1.

Figures 2a-c. a) Overall representation of the DNA complex of BABAP. b-c) detailed H-bond interactions on the unprimed (b) and primed (c) strands between the imino and *cis*-amino protons of the Arg side-chain and O_6/N_7 of G3-G4/G3'-G4'.

Figures 3a-c. Normalized distribution functions of the H-bond distances between each Arg side-chain and the bases it targets in the unprimed (2a) and primed (2b) strands. 3c. Normalized distributions for the H-bond between each Arg main-chain NH group and O_6 of bases $G_6/G_{6'}$ of the central intercalation site.

Figures 4a-d. Time evolutions of the H-bond distances represented in 3a-c.

Figure 5. Representation of the structure of DNA bound by BABAP, alongside with those in its complexes with RG-C5 and RG-C4.

Figures 6a-d. Representation of the structures of DNA in its complexes with derivatives R-C4(a), R-C3(b), R-C2(c), and R-C1(d) displaying the P_8-P_{14} and P_8-P_{31} distances (see text for definition).

Figure 7a-c. a) Representation of the kinked DNA---R-C1 complex. b-c) Representation of the H-bonds between the G bases and Arg from the last MD pose, on the unprimed (b) and primed (c) strands.

Figures 8a-b. DNA--R-C1 complex. Normalized distributions of the Arg-guanine H-bond distances on the unprimed (a) and primed (b) strands.

Figure 9. Structure of DNA in its R-C1 complex, alongside with its structure extracted from its complex with the *Cre* recombinase.

Supporting Information.

Figures S1. Complex of DNA with RG-C5. Overall view (S1a) and H-bond with G_3-G_4 (S1b) and $G_{3'}-G_{4'}$ (S1c).

Figures S2. Complex of DNA with RG-C4. Overall view (S2a) and H-bond with G_3-G_4 (S2b) and $G_{3'}-G_{4'}$ (S2c).

Figures S3. Complex of DNA with R-C4. Overall view (S3a) and H-bonds in the unprimed (S3b) and primed (S3c) strands.

Figures S4. Complex of DNA with R-C3. Overall view (S4a) and H-bonds in the unprimed (S4b) and primed (S4c) strands.

Figures S5. Complex of DNA with R-C2. Overall view (S5a) and H-bonds in the unprimed (S5b) and primed (S5c) strands.

Data and Software Availability.

Four AMOEBA MD input files are provided for BABAP and ligand R-C1, with extensions *xyz (coordinates), *key and *prm (parameters) and *dyn (for MD restarts) and regrouped for each ligand as one file (lig-*-MD.pdf). They are provided with the paper as data files. The calculations were performed with the publicly available free open-source software Tinker-HP (https://github.com/TinkerTools/tinker-hp).

References.

1.   Dvorák, M., Urbánek, P., Bartůněk, P., Paces, V., Vlach, J., Pecenka, V., Arnold, L., Trávnicek, M. and Ríman, J. (1989) Transcription of the chicken myb proto-oncogene starts within a CpG island. *Nucleic Acids Res*, **17**, 5651-5664.

2.   Ding, J., Das, K., Hsiou, Y., Sarafianos, S.G., Clark, A.D., Jacobo-Molina, A., Tantillo, C., Hughes, S.H. and Arnold, E. (1998) Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody fab fragment at 2.8 Å resolution11Edited by J. Karn. *Journal of Molecular Biology*, **284**, 1095-1111.

3.   Wildschutte, J.H., Baron, A., Diroff, N.M. and Kidd, J.M. (2015) Discovery and characterization of Alu repeat sequences via precise local read assembly. *Nucleic Acids Research*, **43**, 10292-10307.

4.   Wu, S., Bafna, V. and Mischel, P.S. (2021) Extrachromosomal DNA (ecDNA) in cancer pathogenesis. *Current Opinion in Genetics & Development*, **66**, 78-82.

5.   Kim, H., Nguyen, N.-P., Turner, K., Wu, S., Gujar, A.D., Luebeck, J., Liu, J., Deshpande, V., Rajkumar, U., Namburi, S. *et al.* (2020) Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nature Genetics*, **52**, 891-897.

6.   Dervan, P.B. (2001) Molecular recognition of DNA by small molecules. *Bioorganic & medicinal chemistry*, **9**, 2215-2235.

7.   Krapcho, A.P., Getahun, Z., Avery, K.L., Jr., Vargas, K.J., Hacker, M.P., Spinelli, S., Pezzoni, G. and Manzotti, C. (1991) Synthesis and antitumor evaluations of symmetrically and unsymmetrically substituted 1,4-bis[(aminoalkyl)amino]anthracene-9,10-diones and 1,4-bis[(aminoalkyl)amino]-5,8-dihydroxyanthracene-9,10-diones. *Journal of Medicinal Chemistry*, **34**, 2373-2380.

8.   Praseuth, D., Gaudemer, A., Verlhac, J.B., Kraljic, I., Sissoëff, I. and Guillé, E. (1986) Photocleavage of DNA in the presence of synthetic water-soluble porphyrins. *Photochemistry and photobiology*, **44**, 717-724.

9.   Villanueva, A., Juarranz, A., Diaz, V., Gomez, J. and Cañete, M. (1992) Photodynamic effects of a cationic mesosubstituted porphyrin in cell cultures. *Anticancer Drug Des*, **7**, 297-303.

10. Gianoncelli, A., Basili, S., Scalabrin, M., Sosic, A., Moro, S., Zagotto, G., Palumbo, M., Gresh, N. and Gatto, B. (2010) Rational design, synthesis, and DNA binding properties of novel sequence-selective peptidyl congeners of ametantrone. *ChemMedChem*, **5**, 1080-1091.

11. Far, S., Kossanyi, A., Verchère-Béaur, C., Gresh, N., Taillandier, E. and Perrée-Fauvet, M. (2004) Bis- and Tris-DNA Intercalating Porphyrins Designed to Target the Major Groove: Synthesis of Acridylbis-arginyl-porphyrins, Molecular Modelling of Their DNA Complexes, and Experimental Tests. *European Journal of Organic Chemistry*, **2004**, 1781-1797.

12. Lagardère, L., Jolly, L.-H., Lipparini, F., Aviat, F., Stamm, B., Jing, Z.F., Harger, M., Torabifard, H., Cisneros, G.A., Schnieders, M.J. *et al.* (2018) Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical Science*, **9**, 956-972.

13. Adjoua, O., Lagardère, L., Jolly, L.-H., Durocher, A., Very, T., Dupays, I., Wang, Z., Inizan, T.J., Célerse, F., Ren, P. *et al.* (2021) Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields Using GPUs and Multi-GPU Systems. *Journal of Chemical Theory and Computation*, **17**, 2034-2053.

14. El Hage, K., Ribaudo, G., Lagardère, L., Ongaro, A., Kahn, P.H., Demange, L., Piquemal, J.-P., Zagotto, G. and Gresh, N. (2022) Targeting the Major Groove of the Palindromic d(GGCGCC)2 Sequence by Oligopeptide Derivatives of Anthraquinone Intercalators. *Journal of Chemical Information and Modeling*, **62**, 6649-6666.

15. Gresh, N., Cisneros, G. A., Darden, T. A., Piquemal, J.-P.(2007) Anisotropic, polarizable molecular mechanics studies of inter-, intra-molecular interactions, and ligand-macromolecule complexes. A bottom-up strategy. J. Chem. theory Comput, **3**, 1960-1986.

16. Gresh, N., Naseem-Khan, S., Lagardère, L, Piquemal, J.-P., Sponer, J. E., Sponer, J. (2017) Channeling Through Two Stacked Guanine Quartets of One and Two Alkali Cations in the Li+, Na+ , K+ and Rb+ Series. Assessment of the Accuracy of the SIBFA Anisotropic Polarizable Molecular Mechanics Potential. J. Phys. Chem. B, **121**, 3997-4014

17. Shi, Y., Schnieders, M., Ren, P. Y., Piquemal, J.-P. (2015) Polarizable Force Fields for Biomolecular Modeling. Reviews in Computational Chemistry, **28**, 51-86, Wiley, DOI: 10.1002/9781118889886.ch2

18. Melcr, J., Piquemal, J.-P. (2019) Accurate biomolecular simulations account for electronic polarization. Front. Mol. Biosci., **6**, 143

19. Dickerson, R.E. (1998) DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res*, **26**, 1906-1926.

20. Harteis, S. and Schneider, S. (2014) Making the Bend: DNA Tertiary Structure and Protein-DNA Interactions. *International Journal of Molecular Sciences*, **15**, 12335-12363.

21. McKay, D.B. and Steitz, T.A. (1981) Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature*, **290**, 744-749.

22. Lawson, C.L., Swigon, D., Murakami, K.S., Darst, S.A., Berman, H.M. and Ebright, R.H. (2004) Catabolite activator protein: DNA binding and transcription activation. *Current opinion in structural biology*, **14**, 10-20.

23. Guo, F., Gopaul, D.N. and Van Duyne, G.D. (1999) Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proceedings of the National Academy of Sciences*, **96**, 7143-7148.

24. Tanaka, Y., Nureki, O., Kurumizaka, H., Fukai, S., Kawaguchi, S., Ikuta, M., Iwahara, J., Okazaki, T. and Yokoyama, S. (2001) Crystal structure of the CENP-B protein–DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *The EMBO journal*, **20**, 6612-6618.

25. Kim, J.L. and Burley, S.K. (1994) 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature structural biology*, **1**, 638-653.

26. Sobell, H.M., Tsai, C.-C., Jain, S.C. and Gilbert, S.G. (1977) Visualization of drug-nucleic acid interactions at atomic resolution: III. Unifying structural concepts in understanding drug-DNA interactions and their broader implications in understanding protein-DNA interactions. *Journal of molecular biology*, **114**, 333-365.

27. Dattagupta, N., Hogan, M. and Crothers, D. (1978) Does irehdiamine kink DNA? *Proceedings of the National Academy of Sciences*, **75**, 4286-4290.

28. Waring, M. and Henley, S. (1975) Stereochemical aspects of the interaction between steroidal diamines and DNA. *Nucleic Acids Research*, **2**, 567-586.

29. Lavery, R., Zakrzewska, K. and Sklenar, H. (1995) JUMNA (junction minimisation of nucleic acids). *Computer Physics Communications*, **91**, 135-158.

30. Hui, X., Gresh, N. and Pullman, B. (1989) Modelling basic features of specificity in the binding of a dicationic steroid diamine to double-stranded oligonucleotides. *Nucleic acids research*, **17**, 4177-4187.

31. Zhang, C., Lu, C., Jing, Z., Wu, C., Piquemal, J.-P., Ponder, J.W. and Ren, P. (2018) AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *Journal of Chemical Theory and Computation*, **14**, 2084-2108.

32. Stone, A.J. (2005) Distributed Multipole Analysis: Stability for Large Basis Sets. *Journal of Chemical Theory and Computation*, **1**, 1128-1132.

33. Lee, C., Yang, W. and Parr, R.G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, **37**, 785-789.

34. Becke, A.D. (1993) A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics*, **98**, 1372-1377.

35. Dunning, T. (1989) Gaussian-Basis Sets for Use in Correlated Molecular Calculations .1. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.*, **90**, 1007-1023.

36. Feller, D. (1996) The role of databases in support of computational chemistry calculations. *Journal of Computational Chemistry*, **17**, 1571-1586.

37. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H. *et al.* (2016), Wallingford, CT.

38. Gresh, N. and Perrée-fauvet, M. (1999) Major versus minor groove DNA binding of a bisarginylporphyrin hybrid molecule: A molecular mechanics investigation. *Journal of Computer-Aided Molecular Design*, **13**, 123-137.

39. Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, **98**, 10089-10092.

40. Bussi, G., Donadio, D. and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, **126**, 014101.

41. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic acids research*, **37**, 5917-5929.

42. Mao, Y., Ge, Q., Horn, P.R. and Head-Gordon, M. (2018) On the Computational Characterization of Charge-Transfer Effects in Noncovalently Bound Molecular Complexes. *Journal of Chemical Theory and Computation*, **14**, 2401-2417.

43. Epifanovsky, E., Gilbert, A.T.B., Feng, X., Lee, J., Mao, Y., Mardirossian, N., Pokhilko, P., White, A.F., Coons, M.P., Dempwolff, A.L. *et al.* (2021) Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *The Journal of Chemical Physics*, **155**, 084801.

44. Mardirossian, N. and Head-Gordon, M. (2014) ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Physical Chemistry Chemical Physics*, **16**, 9904-9924.

45. König, P., Giraldo, R., Chapman, L. and Rhodes, D. (1996) The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell*, **85**, 125-136.

46. Esnault, C., Roques, B.P., Jacquemin-Sablon, A. and Pecq, J.B.L. (1984) Effects of New Antitumor Bifunctional Intercalators Derived from 7 H-Pyridocarbazole on Sensitive and Resistant L1210 Cells. *Cancer research*, **44**, 4355-4360.

47. Delbarre, A., Delepierre, M., Garbay, C., Igolen, J., Le Pecq, J.-B. and Roques, B.P. (1987) Geometry of the antitumor drug ditercalinium bisintercalated into d (CpGpCpG) 2 by 1H NMR. *Proceedings of the National Academy of Sciences*, **84**, 2155-2159.

48. Gao, Q., Williams, L.D., Egli, M., Rabinovich, D., Chen, S.-L., Quigley, G.J. and Rich, A. (1991) Drug-induced DNA repair: X-ray structure of a DNA-ditercalinium complex. *Proceedings of the National Academy of Sciences*, **88**, 2422-2426.

49. Pingoud, A. and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic acids research*, **29**, 3705-3727.

50. Heinemann, F., Karges, J. and Gasser, G. (2017) Critical overview of the use of Ru (II) polypyridyl complexes as photosensitizers in one-photon and two-photon photodynamic therapy. *Accounts of Chemical Research*, **50**, 2727-2736.
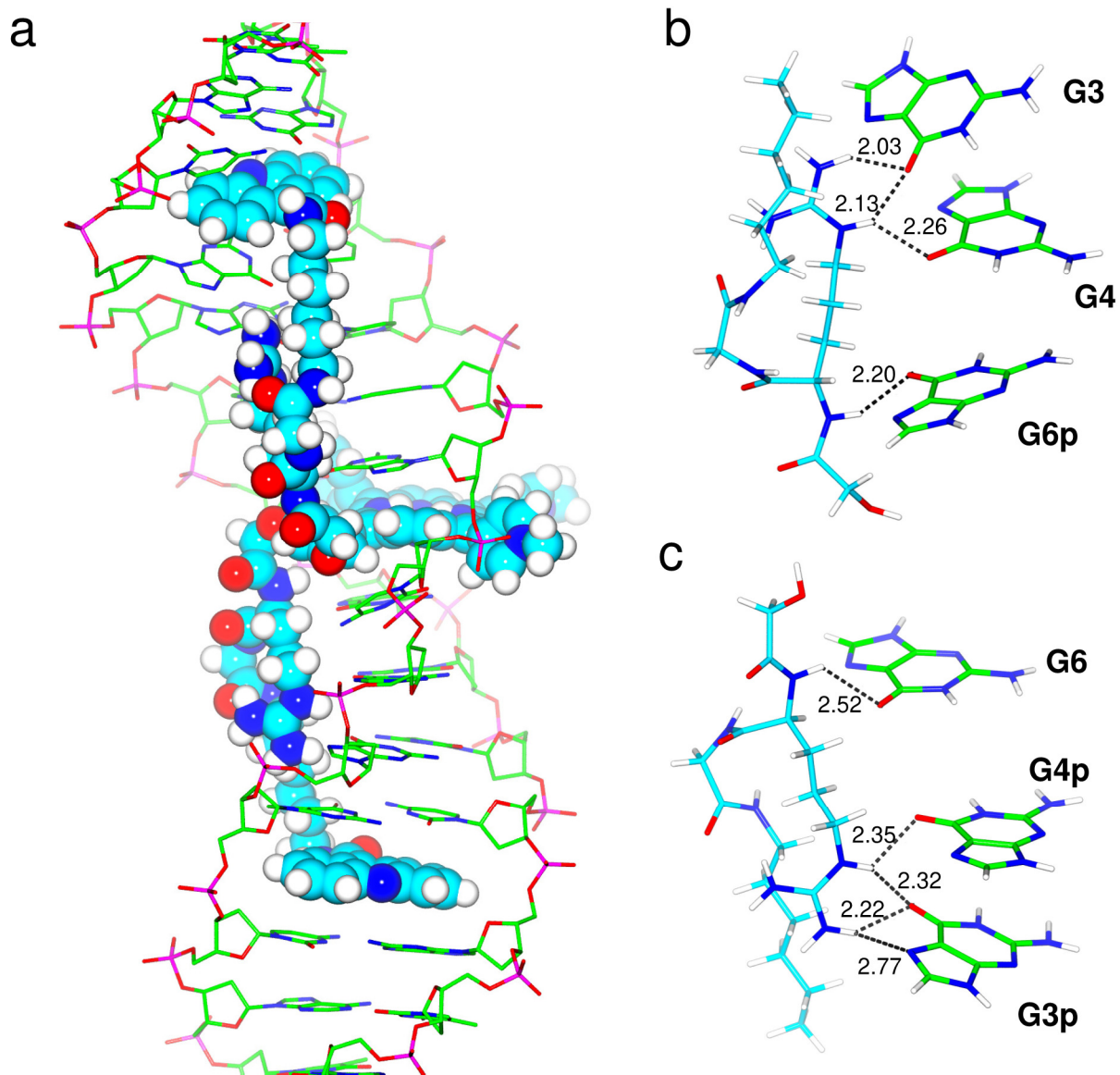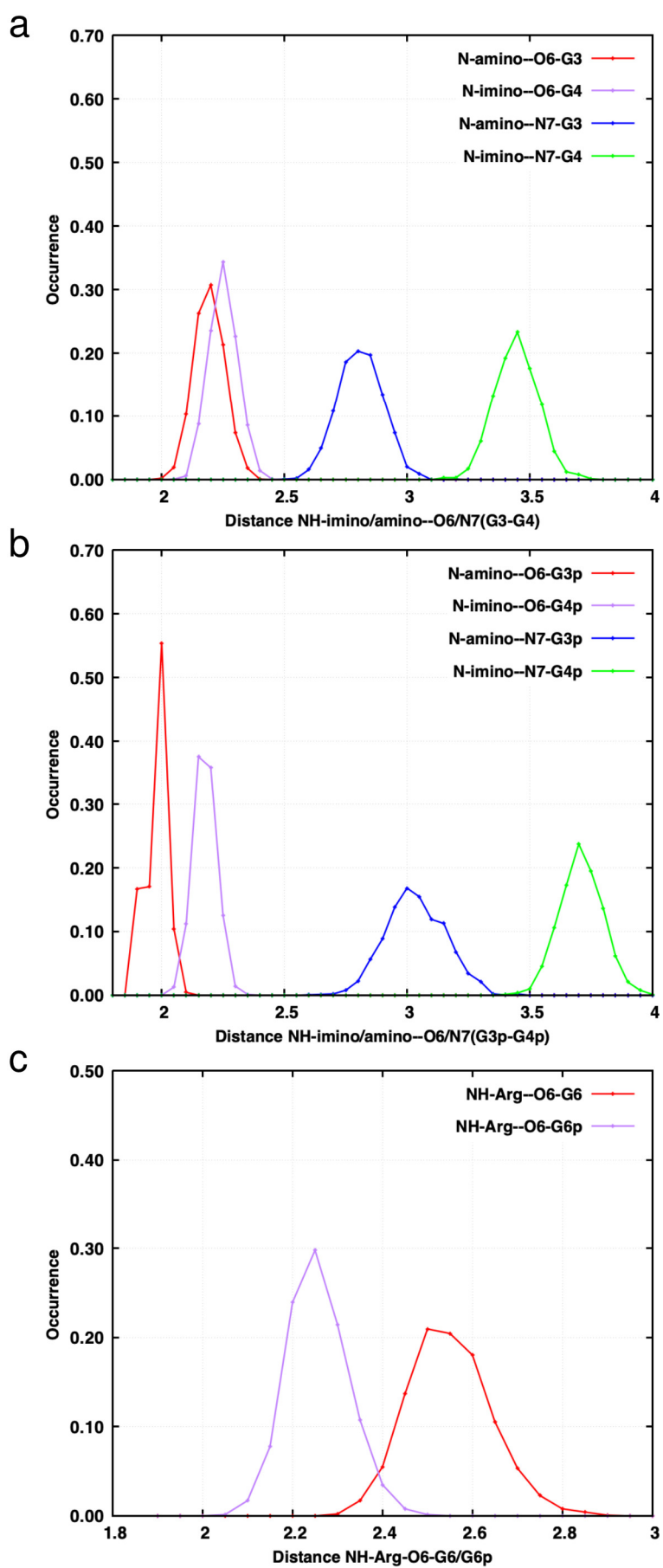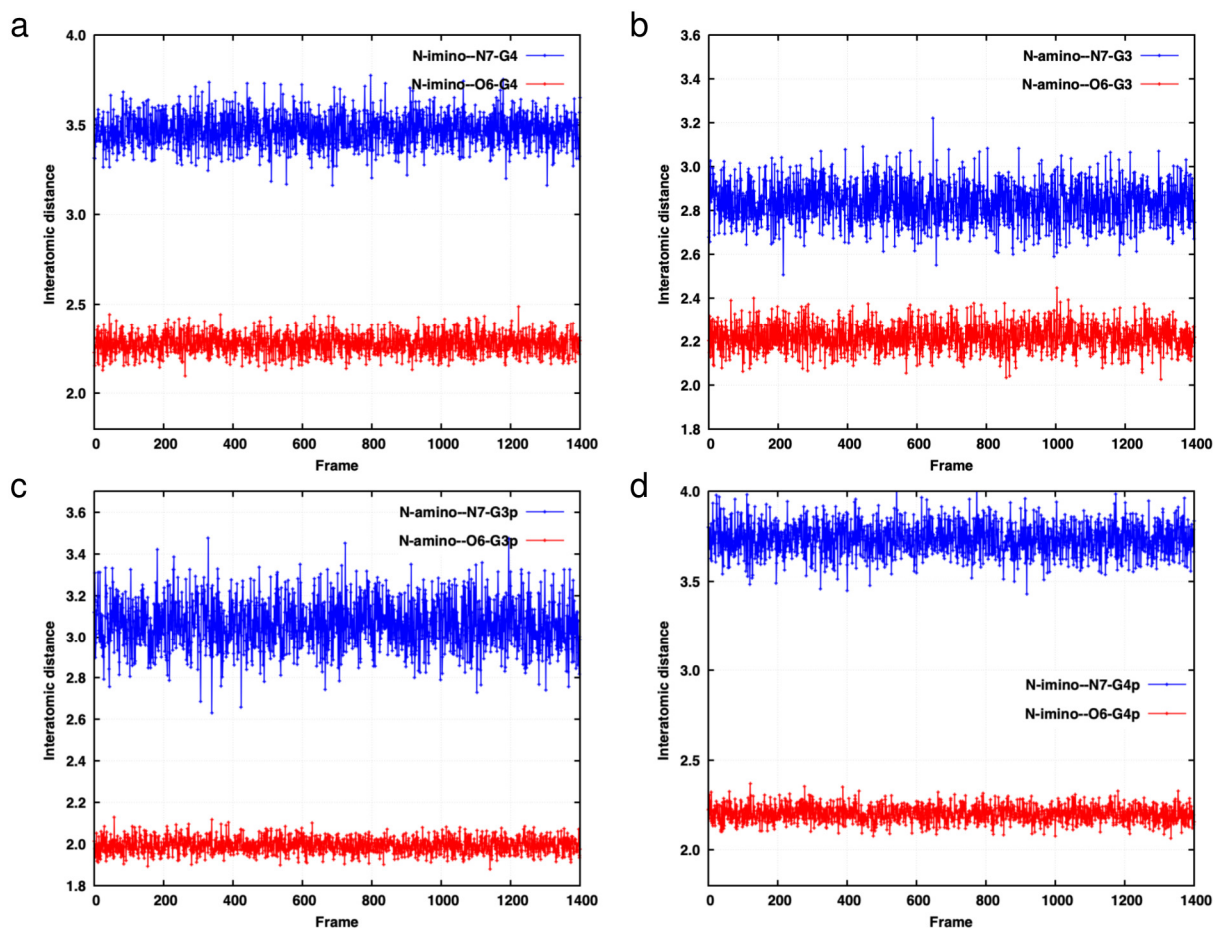
P14

roll −72°

tilt 33°

AA intercalation site

P31

28.0 Å

18.0 Å

Rationally induced DNA kinks by trisintercalators with shortened connectors

P8

Porphyrin intercalation site

TOC

a

b

With R =   BABAP : n = 6
            RG -C5 : n= 5
            RG-C4 : n= 4

With R =   R-C4 : n = 4
            R-C3 : n= 3
            R-C2 : n= 2
            R-C1 : n=1

Figure 1 (a, b)

Figure 2 (a,b,c)
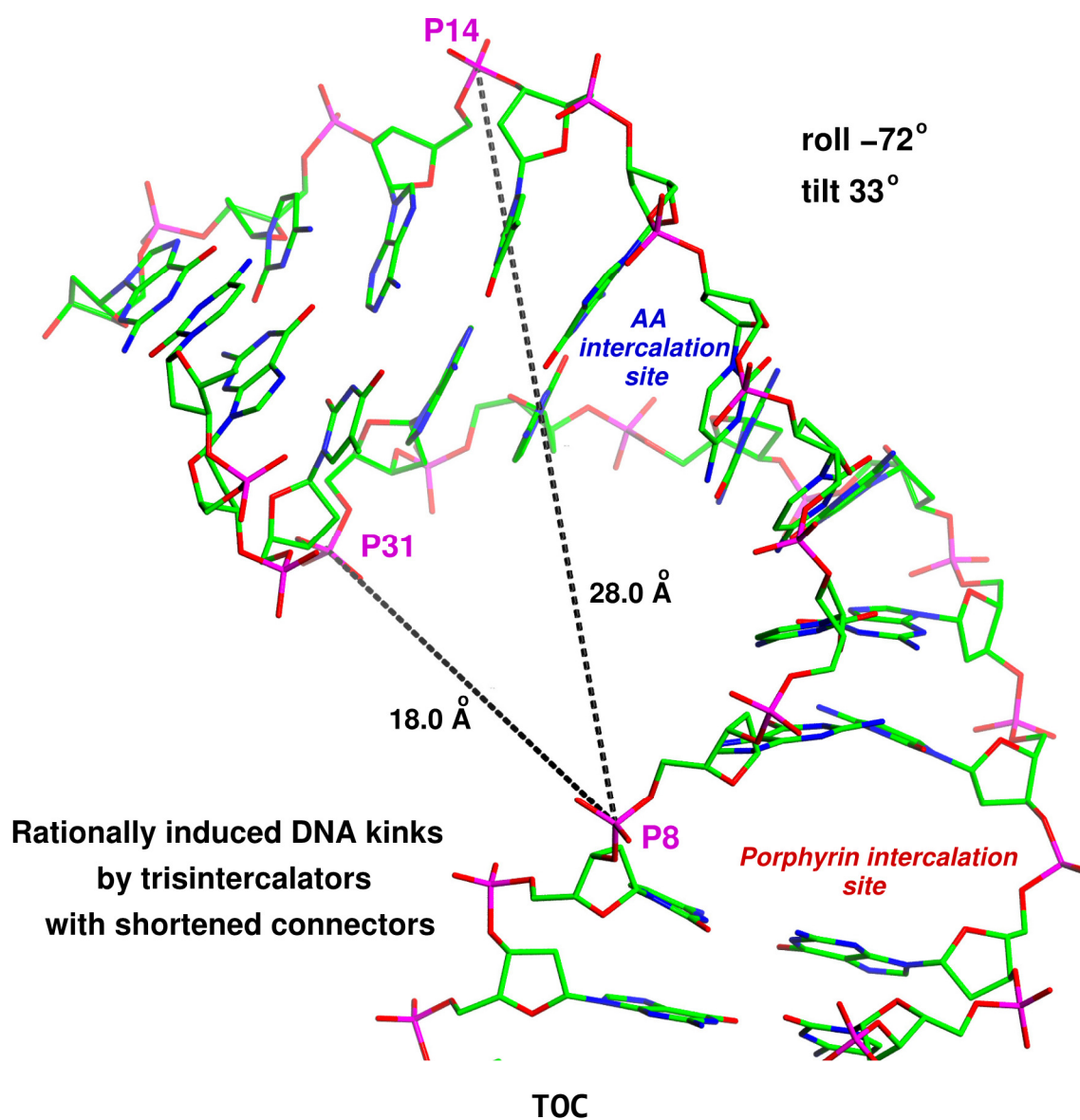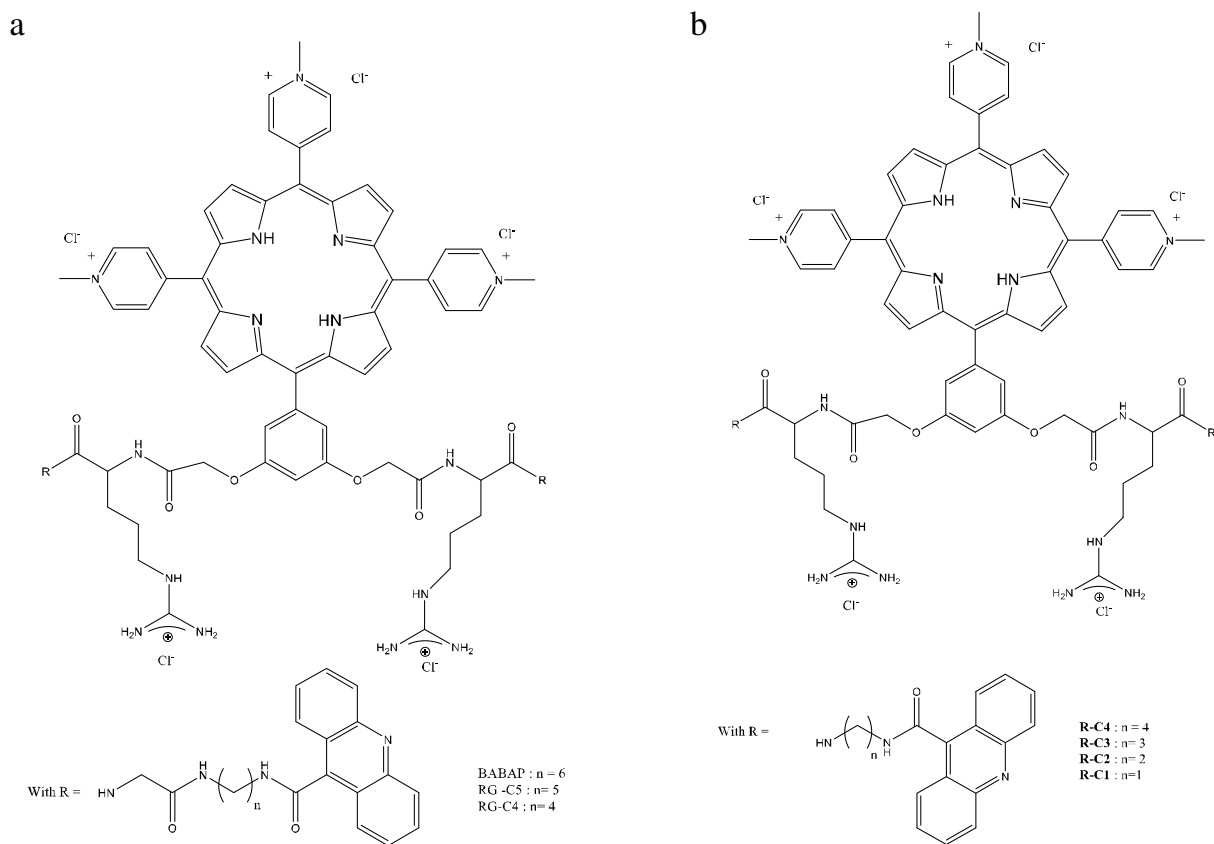
Figure 3 (a,b,c)

Figure 4 (a,b,c,d)

Figure 5

Figure 6

Figure 7 (a,b,c)

Figure 8 (a,b)

RC1–bound DNA

Cre–bound DNA at kinked sequence

Figure 9