Simulations of DNA coiling around a synthetic major groove binding ligand

Syma Khalid, Michael J. Hannon, Alison Rodger, P. Mark Rodger**
Department of Chemistry, University of Warwick
Coventry CV4 7AL, U.K.

Abstract
In this work we present the results of a molecular simulation study of the interaction between a tetracationic bis iron(II) supramolecular cylinder, [Fe$_2$(C$_{25}$H$_{20}$N$_4$)]$^{4+}$, and DNA. This supramolecular cylinder has been shown to bind in the major groove of DNA and to induce dramatic coiling of the DNA. The simulations have been designed to elucidate the interactions that lead the cylinder to target the major groove and that drive the subsequent DNA conformational changes. Three different simulations have been performed: one of the uncomplexed d(CCCCCTTTTTCC)$_2$ dodecamer; one of this DNA complexed with the cylinder molecule; and one of this DNA complexed with a neutralised version of the cylinder. It is shown that the DNA response is insensitive to large changes in the cylinder charge distribution, but is instead dominated by the short range interactions that define the molecular shape.

* email address: p.m.rodger@warwick.ac.uk
1 Introduction

Gene expression is one of the most fundamental processes in biology. It involves the transfer of information encoded within the gene to produce a biologically active protein. However, not all genes are expressed in every cell all the time. Genes are expressed only when they are needed for a cell to function. Regulation of gene expression is controlled by proteins that activate or repress transcription by binding to short, specific DNA sequences. Control of gene expression is currently one of the key areas of interest in molecular medicine. The ability to turn genes on or off artificially by the action of synthetic analogues of DNA-binding proteins is an important goal that would open up new possibilities for disease control and prevention as well as cure.

Proteins that bind DNA frequently achieve sequence-specific code recognition by binding in a non-covalent way in the major groove of DNA.1,2,3 In biological systems, the major groove is the preferred binding site for sequence recognition as it shows the greatest variation in size and shape with base sequence; it would therefore be the ideal target for synthetic molecules designed to recognise and bind to specific DNA sequences. However, until recently, relatively little progress had been made in producing synthetic major groove binders. Most of the compounds synthesised are either minor groove binders,4 intercalators5,6 or metal complexes that span only 2–3 base pairs.7,8,9 In each case the interaction with DNA is too limited for sequence selectivity. Synthetic molecules that target the major groove in a sequence-selective way remain a major goal in molecular medicine.

Recently a major step has been taken toward achieving this goal: Hannon et al. have developed a novel compound that binds strongly into the major groove of DNA and is large enough to span more than four base pairs. The compound is a metallo-supramolecular tetracation (1) with a triple helical framework. The tetracation is approximately cylindrical in shape with length 19 Å and diameter 11 Å; hereafter it will be referred to as the cylinder. This cylinder is too big to bind in the minor groove of DNA, but appears to have just the right shape and size to lie along the major groove.10

Experimental techniques have provided some information regarding the binding strength and preferred binding sites of 1 on DNA.11 Spectroscopic studies indicate that the cylinder has a binding constant well in excess of $10^7 \text{M}^{-1}$. Both flow linear dichroism (LD) and tapping mode AFM,11 have shown that the DNA bends or coils upon binding the cylinder. NMR data have confirmed the major groove-binding mode for the cylinder.12 From all the experimental data it is clear that the cylinder binds in the major groove and is able to induce dramatic conformational changes in the DNA; these are unprecedented effects with synthetic DNA.

1 the tetracationic cylinder: $Fe^{lll}$ is pink, $N$ is blue, $C$ is grey; $H$ atoms have been omitted for clarity
binders. However, gaining molecular level information in such a macromolecular system is challenging. Molecular dynamics (MD) simulations can provide information at the molecular level that is complementary to experiment and therefore are an ideal way to get a better understanding of this system. In this paper we present initial results of such an MD study, paying particular attention to the nature of the intermolecular forces that drive the DNA response. Docking calculations, both manual and using high temperature MD with constrained or rigid molecules, have been used to identify favourable binding sites for the cylinder on a DNA dodecamer, and then multi-nanosecond MD simulations performed with explicit solvent to monitor the DNA response. We note that some MD simulations have been reported on this system before,13 but these were restrained to NOE data and did not report the induced mismatches in the DNA duplex reported here; the present paper also presents a much more comprehensive analysis of the DNA structure and of the forces that induce the DNA response to binding.

2 Computational methods

2.1 System

The DNA sequence adopted was a dodecamer: d(CCCCCCTTTTTCC)2. Preliminary simulations were actually performed with a decamer, d(CCCCCCTTTTT)2, and was chosen to give some chance of identifying sequence selectivity within a single series of simulations; this was subsequently extended by adding two C-G pairs to the T-A end so as to minimise any tendency for the DNA ends to fray in the absence of the cylinder.

The cylinder, 1 (denoted C4+), was modelled using the CHARMM 22 force field, but with the FeN6 sub-unit treated as a rigid body with geometry taken from the crystal structure. For comparison, a “thought experiment” was also performed in which the cylinder was made electrically neutral, denoted C0, via a linear scaling of the charges on all C and non-polar H atoms. From a comparison of the C4+ and C0 systems it is then possible to identify the relative importance of molecular shape (as defined by the van der Waals interactions) and electrostatics in inducing the DNA response.

2.2 Technical Details.

Compounds were modelled using the CHARMM22 all-atom force-field,14 and simulations performed using DL_POLY.15 The conversion from CHARMM to DL_POLY force-field formats was achieved using a purpose-built program that interprets the CHARMM prm and crd files. As in previous work, {san Miguel, 2003} checks on the force-field were performed on numerous configurations to ensure the energies calculated with DL_POLY and CHARMM agreed exactly. Water was modelled with the TIP3P16 potential, and kept rigid using the SHAKE algorithm, implemented in DL_POLY with a tolerance of 0.0001. All hydrogen atoms were assigned a mass of 2 u. This gave good energy conservation in NVE simulations with a 2 fs time step. An Ewald sum17 was used to evaluate electrostatic interactions, with \( k_{\text{max}} = (5,5,6) \) and \( \alpha = 0.12604 \, \text{Å}^{-1} \). Simulations were performed in the NVT ensemble using the Hoover thermostat with a time constant of 0.1 ps and a time step of 2 fs. Orthorhombic periodic boundary conditions were used to mimic an infinite system. Input files for each system are provided as supplementary material.

2.3 Protocol

Docking: a dodecamer of B–DNA and one cylinder molecule (M–enantiomer) were embedded in a neutralising atmosphere of Na+ ions. A further 58 Na+/Cl- ion pairs were added in a 45×45×60 Å box, with DNA aligned along the z (long) axis; this gave [NaCl] = 0.8 M, which is similar to the concentration used in other MD studies.18,19 The DNA was
immobilized and an NVT MD simulation performed for 1 ns at 900 K. The conformation with the lowest configurational energy was extracted and used to start fully solvated MD simulations. Manual docking calculations were also performed, but did not give more favourable docking sites than those identified using the high temperature MD. Repeat simulations were also performed with other low energy binding sites identified from the high temperature MD docking calculations; all such simulations exhibited a very similar DNA response to that reported here.20

Molecular dynamics simulations: water molecules, taken from an equilibrated liquid water simulation, were added to the configurations extracted from docking calculations (including ion positions). Any water molecule that overlapped with DNA, cylinder, Na+ or Cl− was removed, which left a total of 3720 water molecules in the final system. To equilibrate the system the DNA atoms were tethered to their original positions with a force constant of 100 kcal / mol / Å2 and an NVT MD simulation performed for 10 ps at 310 K. A further five simulations were then performed successively in which the tethering force constant was 50, 25, 10, 5 and 1 kcal / mol / Å2, respectively. The tethering potentials were then removed completely and multi-nanosecond simulations performed. We report here an analysis of the first 2 ns, which was found to be sufficient to elicit the response of the DNA to the presence of the cylinder; extensions of these simulations to 5 ns have been performed,20 but did not alter the findings presented here. Configurations from the trajectory were saved every 1 ps for subsequent analysis.

2.4 ANALYSIS METHODS

A number of different methods have been used to monitor DNA structure. Conformational parameters have been used as defined by the Curves algorithm21 and implemented in MDToolchest.22 These use a seven torsion angles (α–ζ, ϕ, χ) to describe the DNA backbone, while the intra-base-pair geometry is described by six helicoidal parameters: three displacements (shear, stretch and stagger) and three angles (buckle, propeller and opening). This parameter set is now well established as a means of describing DNA conformation;23,24 a full definition is supplied with the supplementary material, though it is useful to note here that of the backbone angles, α and ζ refer to torsions about P–O bonds, δ and ϕ refer to the ribose ring, while χ is for the bond that links the backbone to the base.

In understanding the DNA conformation and flexibility, it is also useful to examine some of the parameters that give a more correlated view of the overall behaviour. In particular, it is useful to study the relationship between neighbouring base pairs. In this work these have been monitored using a distance (step) and three angles (tilt, roll and twist) as implemented in 3DNA25 (see supplementary information). In this case, the analysis has been applied to the average structure determined from a continuous 50 ps portion of the MD trajectory; this length of trajectory was found to be long enough to smooth out the instantaneous fluctuations in shape, but was still short with respect to the systematic relaxation induced by the presence of the cylinder, 1. The average structures have also been used to generate normal vector plots,26 which are useful for identifying linear and bent regions within the DNA.

3 Response of the DNA to a bound cylinder

It is instructive to begin by considering snapshot images of the DNA conformation in the three different systems: uncomplexed, bound to C4+, and bound to C0. Although such snapshots provide only limited information about the range of DNA conformations, for the present study they prove to give a very useful overview of the more quantitative data presented later in this article.

Images of the initial and final configurations for DNA and cylinder are given in Figure 1. Initially, the DNA adopts a nearly linear B-form. When present, the cylinder is seen to occupy the major groove, lying symmetrically between the two strands of the DNA, and spanning base pairs 6–11 (T6–A19 to C11–G14). This is consistent with the structure obtained from NMR (NOE) data.\textsuperscript{10,11,12} We note that this binding geometry was retained throughout equilibration phases, and was not disrupted until the tethering potentials that restrained the DNA conformation were removed.

After 2 ns of the production simulations, the uncomplexed DNA retains its near linear B-form, albeit with some fraying at the first base pair. In contrast, the final configurations of the DNA/cylinder simulations reveal substantial curvature of the DNA. The parent compound (C\textsuperscript{4+}) still occupies the major groove, but the location is no longer symmetrical with respect to
the two DNA strands. There is close association between the cylinder and base-pairs 6–9 (closest interatomic distances between the cylinder and each of these nucleic acids are 2.4–3.5 Å). It is particularly interesting to note that some breakdown of the Watson-Crick structure is apparent in the lower 3–4 base pairs of the DNA (Figure 1b). As will be shown below, this actually due to a mispairing between T9 and A15, which then leaves T10 and A16 unpaired.

The behaviour of the neutralised cylinder, C₀, provides a fascinating contrast. It might originally be thought that the coiling of the DNA was, in large part, due to the attraction between the large positive charge on C⁴⁺ and the negatively charge phosphates in the DNA backbone. However, it is clear from Figure 1 that neutral cylinder, C₀, induces a much tighter DNA coil than does C⁴⁺. The final configuration shows C₀ to lie symmetrically within the major groove, spanning base pairs 5–10, but with the DNA having bent substantially to create a much deeper pocket for the cylinder than was found with C⁴⁺. At the same time, C₀ does not appear to induce any disruption of the Watson-Crick base pairs. This contrast between response induced into DNA by C⁴⁺ and C₀ suggests that the DNA coiling is largely a response to the shape-dependent van der Waals forces. We note further that while the net positive charge of the parent compound undoubtedly leads to very strong binding, it also appears to stress the interbase pair hydrogen bonding of the double helix.

The helicoidal parameters developed by Lavery and Sklenar²¹ are a good source of quantitative data on how these ligands affect the integrity of the base pairs within the DNA double helix. These parameters describe how two nucleic acids move and rotate relative to each other and are defined such that zero displacement/rotation corresponds to the ideal base pair geometry. Helicoidal parameters have been calculated from all three simulations and are presented in Figure 2. For the uncomplexed DNA, nearly all the base pairs can be seen to be stable, with typically small fluctuations about values of zero (the ideal base-pair geometry). The only exception is the first pair, C₁–G₂₄, which, from about 1 ns, showed large amplitude motions characteristic of fraying at the end of the double helix; end-fraying is a not uncommon event for DNA, both in simulations and in vivo. Some large-amplitude fluctuations were also seen for T₈–A₁₇ and T₉–A₁₆ pair at about 1 ns, but these were transient and rapidly returned to stable values; such behaviour is indicative of the flexibility inherent within the DNA duplex rather than of irreversible conformational changes.

Given the amount of curvature evident in Figure 1c, the C₀ ligand gives rise to helicoidal parameters that are remarkably similar to those of the uncomplexed DNA. Indeed, the only substantive difference between these two sets of data is that the transient large amplitude oscillations seen in the uncomplexed DNA are completely absent for DNA in the presence of C₀. As with the uncomplexed DNA, fraying is again found at the first base pair. Thus, the presence of C₀ actually appears to enhance the stability of the Watson-Crick base pairs, despite the overall curvature of the DNA evident in Figure 1c.

For C⁴⁺ bound to DNA, the helicoidal parameters clearly reveal a deformation of the base-pair stack. There are two distinct zones in the double helix for this system. The first seven base pairs show stable behaviour as seen in the uncomplexed DNA, with just small fluctuations around zero. Similar behaviour is seen for T₈–A₁₇, albeit with larger amplitude vibrations for SHR, BKL and OPN. However, the plots for the last four base pairs show extremely large variations that are simply inconsistent with a stable Watson-Crick base pair: shear deformations vary by as much as 8 Å during a simulation, while the base pair “stretches” are, at times, 3–4 Å smaller than their equilibrium value.
Visual inspection of the DNA/C$^{4+}$ trajectory showed that this apparent disruption of the double helix was actually a complex rearrangement amongst the base pairs, resulting in the formation of a new pairing between T9 and A15, so that T10 and A16 were left unpaired (see Figure 3). The sequence of events for the formation of this defect was as follows:

1. some stress became evident in base pairs 10–12 from about 170 ps
2. T9–A16 broke apart at 390 ps
3. a new base pair, T9–A15, formed within the following 10 ps.
T10 was then observed to swing out of the helix, kinking the backbone, and subsequently disrupting the last two base pairs (C11–G14 and C12–G13). It is interesting to note that all this activity occurred adjacent to the cylinder, which remained bound to base pairs 5–9 throughout the simulation.

\[ \text{Figure 3: Snapshot of DNA showing the base-pair defect induced by } C^{4+} \]

This change in base-pair stacking should also be revealed in the pattern of hydrogen bonding between the nucleic acids at the end of the simulation. Hydrogen bond distances were obtained from the average DNA structure, as calculated by the 3DNA package, using the final 50 ps of each trajectory. This time window was found to be long enough to smooth out the instantaneous vibrations, but still short enough to avoid artefacts due to long timescale conformational motions of the DNA. The results for the two DNA/cylinder simulations are listed in Table 1 and confirm the formation of a mismatch link between T9 and A15 in the presence of C\(^{4+}\), with the two hydrogen bonds expected for a T–A Watson-Crick base pair.

<table>
<thead>
<tr>
<th>Base pairings</th>
<th>Length Å</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA / C(^{4+})</td>
</tr>
<tr>
<td></td>
<td>O2-N2</td>
</tr>
<tr>
<td>[1–24] C–G</td>
<td>2.75</td>
</tr>
<tr>
<td>[2–23] C–G</td>
<td>2.96</td>
</tr>
<tr>
<td>[3–22] C–G</td>
<td>2.86</td>
</tr>
<tr>
<td>[4–21] C–G</td>
<td>2.80</td>
</tr>
<tr>
<td>[5–20] C–G</td>
<td>2.91</td>
</tr>
<tr>
<td>[6–19] T–A</td>
<td>2.98</td>
</tr>
<tr>
<td>[7–18] T–A</td>
<td>3.05</td>
</tr>
<tr>
<td>[8–17] T–A</td>
<td>3.03</td>
</tr>
</tbody>
</table>

\[ \text{Table 1: Hydrogen bond distances for DNA/cylinder simulations} \]
Table 1: Intra-base pair hydrogen bond distance, taken from the average DNA structure observed during the final 50 ps of the simulation; blank values indicate no H-bond was found. The mismatch induced by C4+ is indicated in bold.

Dial plots of the DNA backbone torsion angles from the three sets of simulations are given in Figure 4. These show less variation between the three systems than was seen in the helicoidal parameters. Interestingly, while the helicoidal parameters indicated that the base-pair structure was more constrained in DNA/C0 than in the uncomplexed DNA, the opposite appears to be true for the backbone, with the backbone torsions showing larger fluctuations in the presence of C0 than without. The changes in base-pair stacking induced by C4+ are also seen in the backbone, but the difference between the two cylinders is less obvious than it was with the direct measures of base-pair geometry. The most obvious indication is in the O–P–O–C5′ angle (α), which shows almost random angular variations between base pairs 10–12 in the presence of C4+, suggestive of rotation that is unhindered by base-pair formation.
Figure 4: DNA backbone torsion angles for (a) uncomplexed DNA, (b) DNA/C^{4+} and (c) DNA/C^{0}. Circular polar plot of time (radius, 0–2 ns) against torsion angle (0° top, 90° right).
Overall curvature for the DNA can be analysed by examining the geometric relationship between neighbouring base pairs. This is conventionally done using the tilt, roll and twist angles. These angles have been calculated between all stable base pairs found in the average DNA structure observed during the final 50 ps of each MD simulation (using 3DNA); plots of these angles are given in Figure 5.

\[ \text{Figure 5: tilt, roll and twist angles between adjacent base pairs. The horizontal axis corresponds to the step between base pair — step 1 compares the first two base pairs (C1–G24 and C2–G23), etc.} \]

The biggest difference between the complexed and uncomplexed DNA appears in the roll angle. For C^0 there is a consistent shift to higher (i.e. more positive) roll angle across steps 2–8 that averages 12° per step compared with the uncomplexed DNA. A positive shift in the roll
angles is also induced by the parent ligand, C\textsuperscript{4+}, although in this case the change is concentrated in the region where the cylinder binds (steps 6–8). Similar trends are seen in the tilt, although in this case the size of the shift is considerably smaller (only +3° per step for C\textsuperscript{0} compared with the uncomplexed DNA). The net effect of these shifts is clearly seen in the Calladine-Drew schematic representations depicted in Figure 6.

Figure 6: representation of the average DNA structure in the (a) uncomplexed, (b) DNA / C\textsuperscript{4+} and (c) DNA / C\textsuperscript{0} simulations. Averages were calculated from the final 50 ps of each simulation using 3DNA.\textsuperscript{25}

As a final measure of DNA helix curvature we report the normal vector plots (NVPs)\textsuperscript{27} in Figure 7. NVPs are projections of the unit vector normal to the plane of each base pair onto a plane that is normal to the helix axis. Linear segments of the DNA can then be identified as clusters of neighbouring points while coiling is seen as a steady change of orientation across a number of adjacent base pairs. Clusters of closely-spaced points are, indeed, apparent in Figure 7 for the uncomplexed DNA. In the presence of the neutralised cylinder (C\textsuperscript{0}), however, there is a large and systematic variation in orientation across base pairs 3–9, consistent with the strong curvature noted above. For the fully charged cylinder, there is strong curvature but it is confined to base pairs 6–9, i.e. to the vicinity of the bound ligand, and is suggestive of a localised kink in the double helix.
4 Discussion

The results of our simulations confirm that the tetracationic macromolecular cylinder, 1 (C\textsuperscript{4+}), binds strongly to the major groove of DNA, spanning 4–5 base pairs. Our simulations also suggest that it has a marked effect on the DNA conformation, but that the extensive coiling observed in experiments is also be associated with the introduction of defects into the base-pair sequence. It is important to stress that the coiling of the DNA is not necessarily caused by the base-pair defects. Indeed, we found that the greatest degree of coiling was actually induced by a neutralized version of the parent cylinder that also appeared to strengthen the original base pairs. Thus it appears that it is the short-range van der Waals interactions arising from the cylinder that induce the coiling in the DNA, but that when these are supplemented by a strong cationic charge, they can place excessive strain on the intra-base-pair hydrogen bonding that usually stabilizes the DNA double helix.

The sensitivity of these results to the initial choice of ligand binding site on the DNA was examined by performing several different multi-nanosecond simulations that started from different initial configurations. Four different locations for the C\textsuperscript{4+} on the linear B-DNA were identified from the original docking calculations and used to implement the full protocol outlined in section 2.3. The results were very similar to those presented above; they are not reproduced here to avoid needless duplication, but can be found in the PhD. thesis of Khalid.\textsuperscript{20} Most significant was that in every case, defects in the alignment of the base pairs were found to arise within the A–T tract of the DNA on a timescale of about 0.5 ns.

To determine whether the defects were introduced as part of the coiling process (i.e. whether the additional strain during coiling made the DNA susceptible to defects in the base-pair sequence), a further simulation was performed in which C\textsuperscript{4+} was bound to a pre-coiled DNA.
This was achieved by taking the final configuration from the DNA/C\textsuperscript{0} simulation reported above, and changing the atomic charges back to those of C\textsuperscript{4+}; the extended simulation was also performed with C\textsuperscript{0} as a control. It was found that within just 250 ps, mismatches in the A–T region again appeared in the presence of C\textsuperscript{4+}, but not with C\textsuperscript{0}. Snapshots illustrating this are shown in Figure 8. Thus we conclude that the induction of defects is not particularly sensitive to the DNA/ligand encounter geometry and so is likely to be well represented in the equilibrium mixture.

![Figure 8: configuration of DNA 250 ps after continuing from the final configuration of the 2 ns DNA/C\textsuperscript{0} simulation with (a) C\textsuperscript{0} and (b) C\textsuperscript{4+}. The purple arrows point to the unpaired A and T bases that are displaced by the formation of the new A–T mismatch (green arrow)](image)

Finally, we note that the induction of mismatches in the DNA sequence alignment was also relatively insensitive to the overall charge of the cylinder. Some simulations were also performed with a variant of the cylinder in which the overall charge was set to +3e (C\textsuperscript{3+}), again achieved by a linear scaling of the atomic charges for all C and non-polar H atoms. Within 250 ps this system also showed disruption of the base pairing within the double helix (see Figure 9). Since the strength of the DNA cylinder interactions are reduced by 25% with C\textsuperscript{3+} compared with C\textsuperscript{4+}, we conclude that disruption of the base pairing cannot be due to any minor inadequacies in the nucleic acid potentials, but represents a real effect of the cylinder binding. Analogous calculations with a +2e variant of the cylinder (not shown) gave results that were intermediate between C\textsuperscript{3+} and C\textsuperscript{0}, with frequent bifurcated pairings in which an A was found to bridge two T bases in the opposite strand, but these did not lead to irreversible breakdown in the base pairing on a nanosecond timescale.
Thus we conclude that the tetracationic supramolecular cylinder, 1, does have two substantive effects on DNA. In the first place it can induce coiling of the DNA, with potentially very large curvature being introduced in DNA dodecamers. However, at the same time it tends to introduce defects into the alignment of the base-pair sequences in the two strands of the double helix. In our simulations this occurred consistently within the A–T rich region of the DNA and resulted in misaligned A–T cross-links between the two strands. It is not clear whether this represents a selectivity for A–T over G–C and simulations with other DNA sequences are in progress to resolve this issue.

Acknowledgements
This work was supported through an EPSRC postgraduate studentship.

References
previous MD


S. Khalid, PhD Thesis,


Normal vector plots

Normal vector plot references: (1) Dickerson; (2) Beveridge et al; (3) McDonald & Wu