Sequence dependence in base flipping: experimental and computational studies†

Lauren L. O'Neil and Olaf Wiest*

Received 30th August 2007, Accepted 29th November 2007
First published as an Advance Article on the web 13th December 2007
DOI: 10.1039/b713318a

Base flipping is the movement of a DNA base from an intrahelical, base-stacked position to an extrahelical, solvent-exposed position. As there are favorable interactions for an intrahelical base, both hydrogen bonding and base stacking, base flipping is expected to be energetically prohibitive for an undamaged DNA duplex. For damaged DNA bases, however, the energetic cost of base flipping may be considerably lower. Using a selective, non-covalent assay for base flipping, the sequence dependence of base flipping in DNA sequences containing an abasic site has been studied. The dissociation constants of the zinc–cyclen complex to small molecules and single strands of DNA as well as the equilibrium constants for base flipping have been determined for these sequences. Molecular dynamics simulations of the zinc-cyclen complex bound to both single- and double-stranded DNA have been performed in an attempt to rationalize the differences in the dissociation constants obtained for the two systems. The results are compared to previous studies of base flipping in DNA containing an abasic site.

Introduction

The double helical structure of DNA, first proposed by Watson and Crick, is characterized by the hydrogen bonding of purine–pyrimidine base pairs, adenine with thymine and guanine with cytosine, and the stacking of adjacent bases.[1] Even while maintaining optimal geometry, the structure of DNA is quite dynamic. In addition, there are also dynamic processes that interrupt hydrogen bonding and/or base stacking, one example being base flipping. Base flipping is the rotation of one or more bases from the intrahelical, hydrogen-bonded, base-stacked position to an extrahelical, solvent-exposed position. The increased accessibility of flipped-out bases to solvent, other small molecules and proteins makes it no surprise that most DNA base modification and repair enzymes require base flipping. Examples include the crystal structures of the DNA–enzyme complexes of M. HaeI, M. HaeIII, hOGG1, T4 endonuclease and CPD photolyase, all of which include one or more bases that have undergone base flipping.[2] Since base flipping requires the breaking of hydrogen bonds and the loss of base-stacking interactions, it is expected to be energetically costly.

Dornberger *et al.* have studied the spontaneous flipping of DNA bases using imino proton exchange and NMR detection.[3] Using these methods, the equilibrium constant for a guanine base in a GCGC tetramer was found to be 3.3×10^{-7}, which corresponds to a free energy difference of ~9 kcal mol^{-1}. Using potential of mean force calculations, base flipping of non-damaged DNA bases has been studied computationally.[4] The free energy required for base flipping of a cytosine base in a GCGC containing sequence, which is recognized by the enzyme M.HhaI, was found to be ~15 kcal mol^{-1}.[5] Other studies have estimated the energy required for a base to undergo base flipping to be between 13 and 23 kcal mol^{-1}, depending upon both the identity of the base and the base flipping pathway.[6]

The most commonly used method to detect base flipping of DNA bases is the incorporation of 2-aminopurine (2-ap) into the DNA sequence of interest. The fluorescent base 2-ap is quenched when in an intrahelical, base-stacked position, but the fluorescence increases and shifts when in an extrahelical, flipped-out position.[7] The ease of detection the base flipping of 2-ap has made this an attractive method which has been used to detect both spontaneous base flipping as well as the base flipping into DNA modification and repair enzymes. As with any method, the use of 2-ap to detect base flipping is not without limitations. As 2-ap is not a native DNA base, native DNA cannot be used and any study relies on the synthesis of the 2-ap analog of the sequence in question. This requires an effort in both time and synthesis. Furthermore, the DNA structure may be perturbed upon its inclusion in the duplex. Finally, the closest related native DNA base is adenine, but 2-ap is not able to mimic this base, nor other damaged bases which may be the subject of DNA–enzyme binding studies.

In a previous report, we demonstrated the use of a new, non-covalent assay that is simply added to native DNA to detect base flipping in free solution, as summarized in Fig. 1.[8] This assay exploits the selective binding of zinc-cyclen complexes to deprotonated, anionic pyrimidine bases.[9] When conjugated to a solvatochromic dye, in our case the widely used dansyl group as shown in Fig. 2, the change in environment from bulk solution to DNA-bound is detected as a change in fluorescence. The dansyl-cyclen reporter group was used to detect base flipping in a DNA sequence containing an abasic unit across from a thymine base, which has been previously shown to induce base flipping using time-resolved fluorescence.[10]

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana, 46556-5670, USA. E-mail: owiest@nd.edu

† Electronic supplementary information (ESI) available: Fluorescence difference plots, statistical parameters of data fit; geometries, charges and atom types used in MD simulations, plots of RMSd and selected geometric parameter *vs.* simulation times. See DOI: 10.1039/b713318a

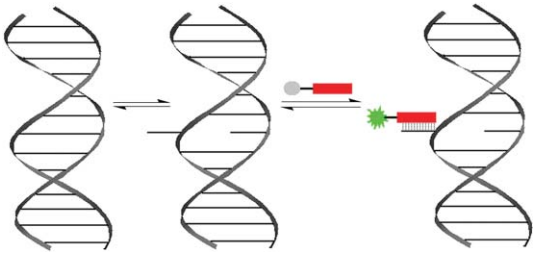

Fig. 1 Schematic representation of the base flipping assay.

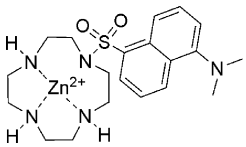

Fig. 2 Fluorescent reporter and binding unit used in base flipping assay, dansyl-cyclen.

Base flipping of DNA sequences containing an abasic unit has been studied extensively using NMR spectroscopy and time-resolved fluorescence.[11,10] There are three possibilities for the location of the abasic site and the opposite base: i) both the abasic site and opposite base are intrahelical and the "gap" left by the absence of the base is filled by solvent, presumably water, ii) the abasic sugar moves closer to the opposite base, which "pinches" the DNA backbone to allow for interactions with the "orphaned" base and iii) both the abasic sugar and opposite base are extrahelical and the duplex collapses to fill the empty space.[11e] For apyrimidinic sites where the opposite base is adenine, there have been no structures reported in which the adenine base is extrahelical. However, in both experimental and computational studies the extrahelical conformation of the ribose at the abasic site has been observed.[11g,h] Base flipping of 2-ap across from a tetrahydrofuran analog of an abasic site has been studied using time-resolved fluorescence.[10] Open conformations of 2-ap were observed for sequences in which both adenines and thymines flank the abasic site analog. Cuniasse *et al.* studied the conformations of 5′-CGTGXGTGC-3′/3′-GCACNCACG-5′ where X is the anhydroribitol analog of an abasic site and N is either G, C or T.[11a] Only the intrahelical form was observed when guanine was opposite the abasic site, while a mixture of intra- and extrahelical conformations were observed when thymine was opposite the abasic site. Only the extrahelical form was observed when cytosine was opposite. This result was confirmed by Singh *et al.* who found a mixture of intra- and extrahelical conformations for a thymine base opposite a true abasic site, X, in the sequence 5′-CGTGXGTGC-3′/3′-GCACTCACG-5′.[11b] Interestingly, in a different sequence, 5′-CGCACXCACGC-3′/3′-GCGTGTGTGCG-5′, in which the abasic site is flanked by pyrimidine bases, Coppel *et al.* reported that the thymine base opposite the abasic site was intrahelical.[11c] The resulting hypothesis was that the flipping of the ribose at the abasic site and the opposite base are dependent upon sequence context. In sequences in which the abasic site is flanked by purine bases the drive for those bases to stack is large and, as a result, the abasic site and the opposite base are expelled from the duplex which then contracts to allow for maximum overlap between the purine bases. In sequences in which the abasic site is flanked by pyrimidine bases, the drive for those bases to stack is not as large and the abasic site and the opposite base are not expelled from the duplex. However, this hypothesis would not explain the observation of the extrahelical conformation of 2-ap in a sequence in which the bases to the 5′ and 3′ sides of 2-ap were adenine.

As previously mentioned, we have developed a non-covalent assay to detect base flipping of thymine bases in free solution. Since our assay requires both base flipping and binding of the fluorescent reporter to the flipped-out base, as shown in Fig. 1, the observed equilibrium constant obtained is a product of the individual equilibrium constants of the two events. Here, we report the dissociation constant of the complex formed between the fluorescent reporter unit and thymine and the application of that dissociation constant to the determination of equilibrium constants for base flipping. We have used both thymine and thymidine monophosphate as well as short single strands of DNA in which base flipping is not required for binding to occur. Molecular dynamics (MD) simulations were used to explore the conformations of the dansyl-cyclen bound to short single strands of DNA. As a comparison to previously reported data, the dissociation constant of the guanosine monophosphate–dansyl-cyclen complex has been determined. Finally, we will discuss the results of the base flipping assay as applied to duplex DNAs containing abasic sites in various sequence contexts in order to study the sequence dependence of base flipping in these systems using the non-covalent base flipping assay.

Results and discussion

To determine the dissociation constants, three independent titration experiments of dansyl-cyclen with different DNA sequences and building blocks were performed for each analyte, as shown in the respective figure(s). The dissociation constants reported are the averages (± standard deviation) of the dissociation constants obtained from the three trials. Each of the data points presented in the graphs of normalized relative fluorescence *versus* [analyte], is the average of ten individual measurements to increase the signal-to-noise ratio and to reduce error in each individual measurement. Fitting parameters, including the R^2 of the best fit line and maximum fluorescence, are presented in the ESI.† The errors in the dissociation constants obtained from the fitting of the data from one trial to a one-site binding model were between ~5 and 10% for thymine, thymidine monophosphate, the single strands of DNA and the duplex DNAs studied. The differences in the reported dissociation constants from the different trial runs were between 36% and 57%, corresponding to free energy differences between measurements of only 0.3 kcal mol^{-1} or less. Considering the data averaging discussed above, it can therefore be stated that the error limit for this method is expected to be sufficiently small to investigate the sequence dependence of base flipping.

The dissociation constant of the dansyl-cyclen–thymine complex was determined to be 1.04 ± 0.32 μM. The plots of normalized relative fluorescence *versus* concentration of DNA and the best-fit binding curves are shown in Fig. 3. In order to determine whether phosphate may influence the dissociation constant, thymidine monophosphate was also subjected to the same analysis, as shown in Fig. 4. The dissociation constant obtained for the thymidine monophosphate titration was 1.01 ± 0.30 μM. This indicates that the included phosphate did not influence the

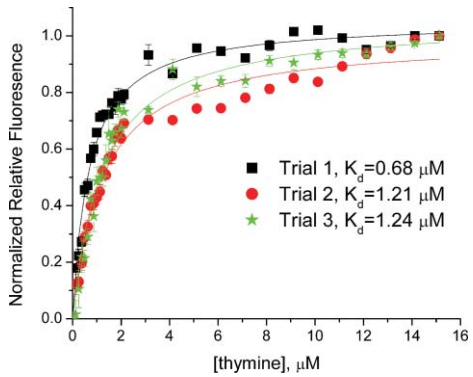

Fig. 3 Normalized relative fluorescence emission of 1 μM dansyl-cyclen in the presence of increasing concentrations of thymine in 100 mM phosphate buffer, pH 7.0.

dissociation constant of the dansyl-cyclen–thymine complex. The binding constant of thymidine monophosphate to zinc-cyclen complexes has been previously determined using potentiometric pH titrations.[9e] For unfunctionalized zinc-cyclen complexes, the reported log$K$ was 3.5 M^{-1}, which corresponds to a dissociation constant of 3.1×10^{-4} M. For a benzyl-functionalized cyclen the dissociation constants with deoxythymidine and thymidine monophosphate were reported to be 6.3×10^{-4} M and 4.0×10^{-4} M.[9c] For a zinc-cyclen complex in which one of the secondary amines has been functionalized with a (4-quinolyl)methyl group, the dissociation constant was found to be 5.0×10^{-5} M. In contrast, the dissociation constant obtained using fluorescence titrations is 1.01×10^{-6} M, which is overestimated by a factor of ~200 as compared to the previously determined values. These differences have to be due to structural features in the dansyl-cyclen system that are not present in the systems previously studied. Although the dansylsulfonamide adds sites of potential interactions with complexed ligands that will be discussed later, the fact that essentially identical binding constants are obtained for thymine and thymidine makes this explanation unlikely. We rather hypothesize that the effect is on the binder itself, presumably due to the sulfonamide linkage, which restricts the cyclen nitrogen *via* sp^2 hybridization and changes the conformational behavior the cyclen ring.

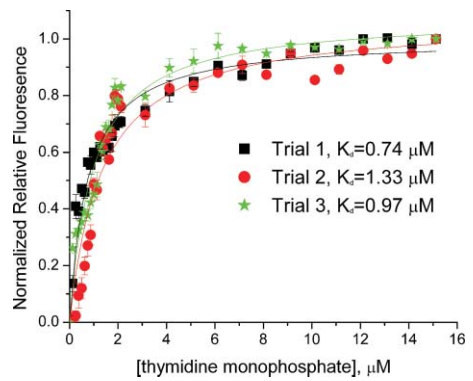

Fig. 4 Normalized relative fluorescence emission of 1 μM dansyl-cyclen in the presence of increasing concentrations of thymidine monophosphate in 100 mM phosphate buffer, pH 7.0.

Zinc-cyclen complexes have been previously shown to be selective for pyrimidine bases.[9] A titration was performed with guanosine monophosphate to determine if this selectivity exists in our system. The dissociation constant of unfunctionalized zinc-cyclen to guanosine monophosphate has been reported to be >0.01 M, while that of a (4-quinolyl)methyl-functionalized zinc-cyclen to guanosine monophosphate has been reported to be 4.0×10^{-3} M.[9e] Using our experimental setup, obtaining concentrations at which the guanosine–dansyl-cyclen system has achieved saturation would require the addition of a large volume of the guanosine monophosphate solution. The data shown in Fig. 5 clearly show that at 28 μM guanosine monophosphate, saturation has not been achieved. The fit of this data to a one-site binding model can be used to give an estimate of the dissociation constant, but because saturation has not been achieved, the reliability of the data is questionable. The experimentally obtained dissociation constant is on the order of 100 μM (1×10^{-4} M). The previously determined dissociation constant of a monofunctionalized zinc-cyclen monophosphate is an order of magnitude higher than the dissociation constant obtained using our method. This difference may be due to the difference in the systems studied. The addition of a (4-quinolyl)methyl substituent has been found to decrease the dissociation constant of the complex by an order of magnitude, consistent with the results obtained here.[9]

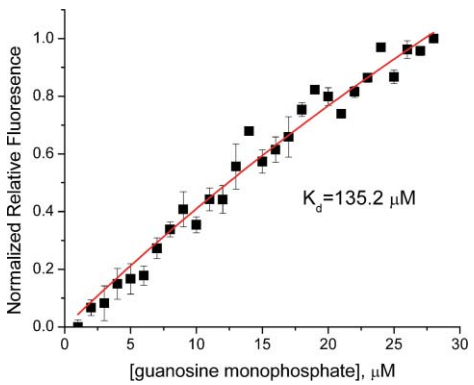

Fig. 5 Normalized relative fluorescence emission of 1 μM dansyl-cyclen in the presence of increasing concentrations of guanosine monophosphate in 100 mM phosphate buffer, pH 7.0.

Next, the dissociation constants of the dansyl-cyclen complex with short single strands of DNA were studied to determine if the full DNA backbone and the bases flanking thymine have an effect on binding. The titration curves for the sequences studied, 5′-GGTGG-3′, 5′-AATAA-3′ and 5′-CCTCC-3′, are shown in Fig. 6. The dissociation constants obtained for the dansyl-cyclen complexes with 5′-GGTGG-3′, 5′-AATAA-3′ and 5-CCTCC-3′ were found to be 0.81 ± 0.46, 1.44 ± 0.52 and 3.55 ± 1.28 μM, respectively. The dissociation constants of the dansyl-cyclen complexes with 5′-AATAA-3′ and 5′-GGTGG-3′, in which the flanking bases are purines, are ~2 and 4 times lower, respectively, than the dissociation constant obtained for 5′-CCTCC-3′, in which the flanking bases are pyrimidines. This indicates stronger binding of the dansyl-cyclen to thymines flanked by purine bases, with the strongest binding observed when those bases are guanine. Because strong preorganization can be excluded for these short

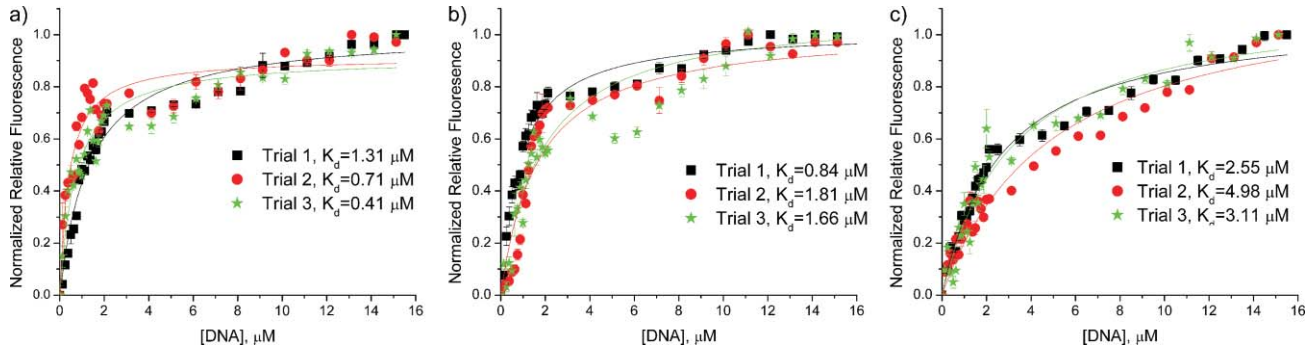

Fig. 6 Normalized relative fluorescence emission of 1 μM dansyl-cyclen in the presence of increasing concentrations of (a) 5′-GGTGG-3′, (b) 5′-AATAA-3′ and (c) 5′-CCTCC-3′, in 100 mM phosphate buffer, pH 7.0.

oligonucleotide strands, the increased binding of the adjacent guanidine is most likely due to an additional, specific interaction.

In order to more fully understand the possible interactions of the dansyl-cyclen and thymine and the experimentally observed differences in the binding of the oligonucleotides, unrestrained MD simulations were performed on dansyl-cyclen bound to deprotonated thymines in both 5′-GGTGG-3′, where the flanking bases are purines, and 5′-CCTCC-3′, where the flanking bases are pyrimidines. The total simulation time for each sequence was 6 ns. In the simulation of the dansyl-cyclen bound to 5′-GGTGG-3′, interactions between the sulfonamide oxygens and hydrogens (H1, cyclic amino hydrogen, and H22, exocyclic 1° amino hydrogen) of a guanine base were observed after ~1.2 ns, as shown in Fig. 7. The distance between the sulfonamide oxygens and the guanine hydrogens decreases from ~9 Å at the start of the simulation to ~2 Å from 1.2 to 3.7 ns and from 3.9 to 5.8 ns. The results of the simulation show that the interaction is rapidly formed, but also dissociates at longer simulation times, and indicates that this interaction is relatively weak, in agreement with the experimentally observed weak binding differences. Shown in Fig. 8 is a snapshot of the MD simulation at 4000 ps showing the interaction of the sulfonamide oxygens of the dansyl-cyclen with a guanine base. The simulation of dansyl-cyclen bound to the deprotonated thymine of 5′-CCTCC-3′ did not show any analogous interactions of the sulfonamide oxygens with the flanking DNA bases. This may explain the ~4 fold difference in the experimentally determined dissociation constants of the dansyl-cyclen to the sequences

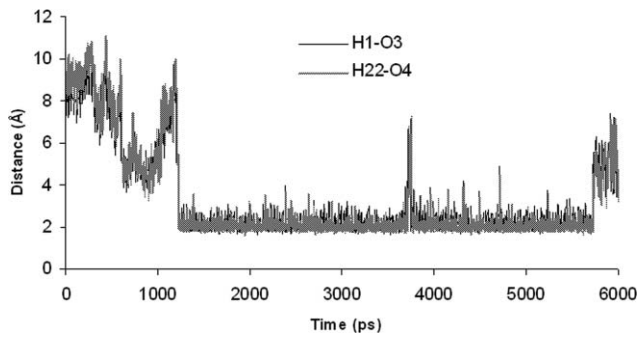

Fig. 7 Distances between sulfonamide oxygens of dansyl-cyclen and guanine hydrogens H1 (solid) and H22 (dashed) of 5′-GGTGG-3′ *vs.* time (ps).

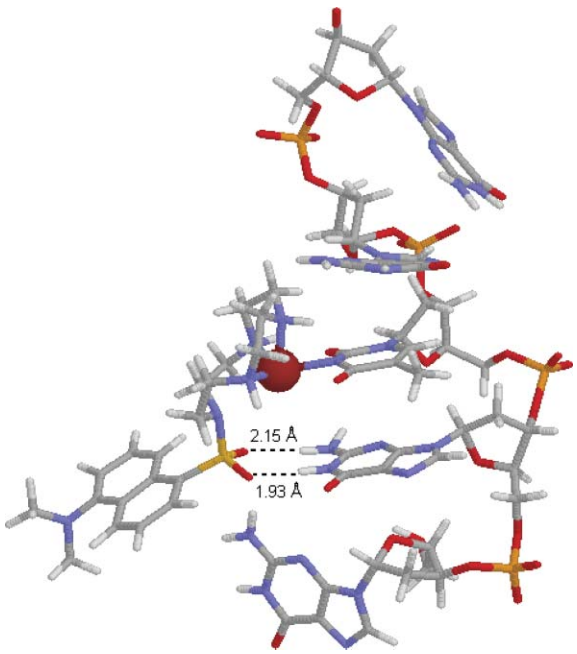

Fig. 8 Snapshot at 4000 ps of MD simulation of dansyl-cyclen bound to deprotonated thymine in 5′-GGTGG-3′ showing the interaction of sulfonamide oxygens with guanine hydrogens (O3–H1 distance 1.93 Å, O3–H1–N1 angle 177.1°, O4–H22 distance 2.15 Å, O4–H22–N2 angle 177.5°).

in which purine (5′-GGTGG-3′) and pyrimidine (5′-CCTCC-3′) bases flank the thymine.

Detection of base flipping in a duplex DNA in which an abasic unit is located across from a thymine base using dansyl-cyclen has been demonstrated previously.[5] The sequence dependence of base flipping in duplex DNAs containing an abasic site was studied by varying the two bases to both the 5′ and 3′ sides of the abasic site. The sequences studied are listed in Table 1, with the X = tetrahydrofuran analog of an abasic site, and the corresponding thymine marked in bold. The titration curves for duplex G, duplex C and duplex T are shown in Fig. 9. The average dissociation constants obtained for the dansyl-cyclen complexes with duplex G, duplex C and duplex T from the three experiments were found to be 1.45 ± 0.48, 2.14 ± 0.84 and 1.63 ± 0.26 μM, respectively. The average dissociation constant for the dansyl-cyclen complex with duplex A was previously reported to be 0.72 ± 0.25 μM.[8]

Table 1 Sequences studied (X = abasic site)

Duplex	Sequence
A[a]	5′-GCACGAA**T**AAGCAGC-3′ 3′-CGTGCTTXTTCGTGC-5′
G	5′-GCACGGG**T**GGGCAGC-3′ 3′-CGTGCCCXCCCGTGC-5′
C	5′-GCACGCC**T**CCGCAGC-3′ 3′-CGTGCGGXGGCGTCG-5′
T	5′-GCACGTT**T**TTGCAGC-3′ 3′-CGTGCAAXAACGTCG-5′

[a] Previously reported in ref. 8.

Using this data, the equilibrium constants for base flipping of the thymine base can be calculated. The dissociation constants reported are a product of the dissociation constant of the dansyl-cyclen to an extrahelical thymine base and inverse of the equilibrium constant for the flipping of the thymine base to an extrahelical position. The choice of the dissociation constant of the dansyl-cyclen–thymine complex that is used in the calculation of the equilibrium constant for base flipping becomes therefore critical for the absolute value of the equilibrium constant for base flipping. While the relative equilibrium constants (and thus the sequence dependence of base flipping) should not depend on the binding constant as long as the interactions remain the same, the observation of interaction of the dansyl-cyclen with bases adjacent to thymine in the short single-strands in both the experimental studies and the MD simulations indicate that this assumption might not be valid. Although the use of short single-strands of DNA appeared to be the optimal choice for the determination of the dissociation constant of the dansyl-cyclen–thymine complex due to the similarity in environment to duplex DNA, it is not clear that the results for 5′-GGTGG-3′ discussed above are also applicable to a duplex where the guanine might stay inside the duplex. Thus, we checked whether the additional interactions to a guanine obtained during the 6 ns of MD simulations of dansyl-cyclen bound to duplex G or other interactions with the DNA backbone occurred in the larger systems.

The total time of the simulation of the dansyl-cyclen bound to an extrahelical deprotonated thymine in an undamaged DNA duplex was 8 ns. A snapshot of the MD simulation at 6 ns, which is a structure representative of the simulation, is shown in Fig. 10. Flipping of the extrahelical thymine base back into the helix was not observed. In the present case, this is due to the definition of

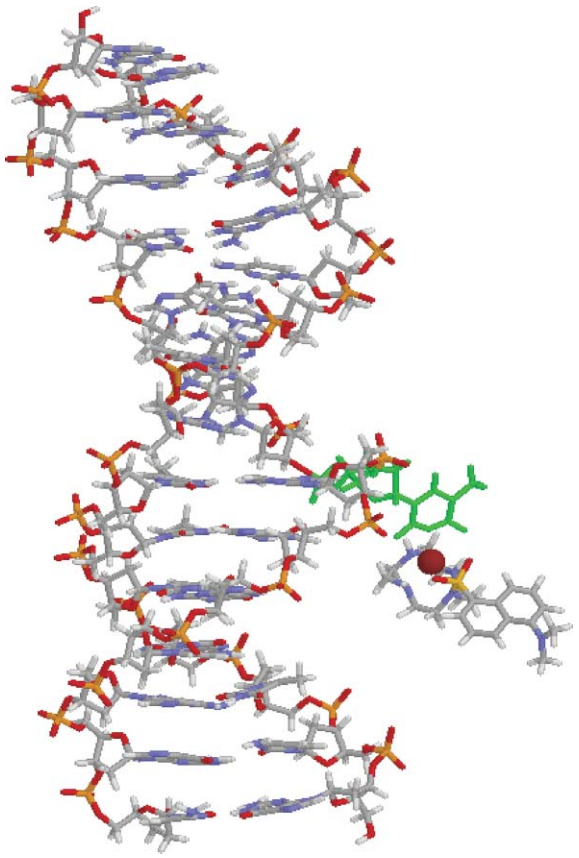

Fig. 10 Snapshot (6 ns) of MD simulation of dansyl-cyclen bound to 5′-GCACGGGT$_x$GGGCAGC-3′/3′-CGTGCCCACCCGTGC-5′ where T$_x$ represents anionic thymine.

a bond between the deprotonated thymine and the zinc ion which would prevent base flipping because the bulky dansyl-cyclen would be impossible to incorporate into the DNA structure. This can be seen by considering the strength of the zinc–nitrogen bond and the fact that base flipping is not observed in similar duplexes in MD simulations on the 10 ns timescale.[9,18] Waters included in the simulation are observed to fill the "hole" in the duplex left by the extrahelical thymine base, again in analogy to earlier simulations.

During the 8 ns of simulation, no interactions of the dansyl-cyclen with the backbone of the DNA or guanines were observed. Although it cannot be excluded that the simulation timescale

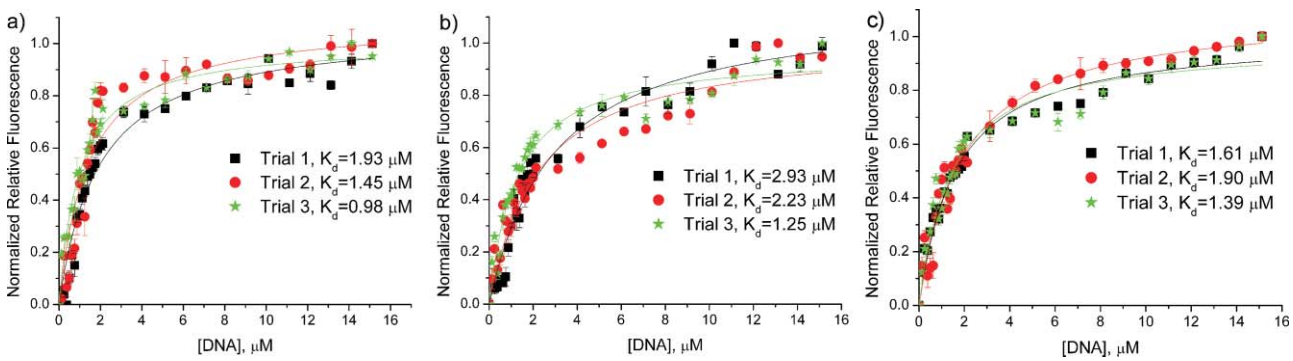

Fig. 9 Normalized relative fluorescence emission of 1 μM dansyl-cyclen in the presence of increasing concentrations of (a) duplex G, (b) duplex C and (c) duplex T, in 100 mM phosphate buffer, pH 7.0.

Table 2 Summary of experimental data

System	K_d/μM
Thymine	1.04 ± 0.32
Thymidine monophosphate	1.01 ± 0.30
Guanosine monophosphate	135.2 ± 61
5′-AATAA-3′	1.44 ± 0.52
5′-GGTGG-3′	0.81 ± 0.46
5′-CCTCC-3′	3.55 ± 1.28
Duplex A[a]	0.72 ± 0.25
Duplex G	1.45 ± 0.48
Duplex C	2.14 ± 0.84
Duplex T	1.63 ± 0.26

[a] Previously reported in ref. 8.

is too short to position a guanine for interaction with the sulfonamide, we consider this possibility as unlikely because the interaction energy is small compared to the energy needed to deform the backbone and disrupt base stacking of the guanine. Since interactions between the dansyl-cyclen and guanine bases were only observed in the single-stranded oligonucleotides, the dissociation constants obtained using single-stranded DNAs may not be appropriate for use in the determination of base flipping equilibrium constants. The dissociation constant of the thymine–dansyl-cyclen complex is larger than the dissociation constant obtained for the 5′-GGTGG-3′–dansyl-cyclen complex, indicating that there are interactions present in the latter system not present in the former. Use of the smaller dissociation constant may lead to underestimation of the equilibrium constants for base flipping in abasic DNAs. Therefore, the dissociation constant obtained for the dansyl-cyclen–thymine complex is best suited to be used to calculate the equilibrium constants for base flipping in abasic DNAs.

A summary of the dissociation constants obtained using the base flipping assay is presented in Table 2. The dissociation constants for the dansyl-cyclen complexes with thymine and thymidine monophosphate are essentially identical, indicating that there are no significant differences in the binding in these two systems. The large dissociation constant estimated from the titration of guanosine monophosphate demonstrates the selectivity of the dansyl-cyclen for pyrimidine bases, namely thymine. The trend in the dissociation constants of the dansyl-cyclen complexes with the single-stranded DNAs was found to be 5′-CCTCC-3′ > 5′-AATAA-3′ > 5′-GGTGG-3′. This trend follows the observation that interactions of the sulfonamide oxygens were observed in the MD simulation of the dansyl-cyclen bound to 5′-GGTGG-3′ but not in the simulation of 5′-CCTCC-3′. These interactions increase the affinity of the dansyl-cyclen for the single-stranded DNA which, in turn, decreases the dissociation constant. Based upon the differences in the dissociation constants obtained for 5′-GGTGG-3′ and 5′-CCTCC-3′, the difference in energy of binding of dansyl-cyclen has been calculated to be ~0.9 kcal.

The base flipping equilibrium constant for duplex A calculated using the binding constant of the dansyl-cyclen–thymine complex was found to be 1.44, which indicates that there is a 50% higher concentration of the flipped-out species in solution. The flipped-out conformation of duplex A is favored by 0.2 kcal, making it the only one of the four sequence studies here that was found to favor the flipped-out state. The base flipping equilibrium constants, K_{flip}, for the three other duplexes G, C and T were calculated to be 0.70, 0.49 and 0.64. This indicates that the flipped-in conformations of duplexes G, C and T are favored compared to the flipped-out conformations by ~0.2 to 0.4 kcal. The sequence dependence of base flipping in DNA containing an abasic site analog using the base flipping assay is therefore duplex A > duplex G ≥ duplex T > duplex C.

Comparison of the results obtained using the base flipping assay and previously reported structures of DNA duplexes containing an abasic site reveals some interesting differences. The results presented herein indicate that the flipped-out state is more likely to be detected in sequences in which purines flank the flipping base, *i.e.* thymine, leaving pyrimidines to flank the abasic site. This result is consistent with the observations of Rachovsky *et al.*, who reported observing the extrahelical conformation of 2-ap using time-resolved fluorescence when the abasic site was flanked by adenine and thymine bases.[10] However, the population of the extrahelical conformation was greater when the abasic site was flanked by adenines than by thymines. This is the reverse sequence dependence than we observe, opening the intriguing possibility that the reversal of the base opposite the abasic site (from the adenine-like 2-ap in ref. 11 to thymine in the present study) leads to an inversion of the sequence dependence of flipping.

It has been previously reported that in a sequence in which an abasic site is flanked by guanine bases, the extrahelical conformation of both thymine and cytosine bases opposite the abasic site was observed using NMR spectroscopy.[11a,11b] When the abasic site is flanked by cytosine bases, only the intrahelical conformation of a thymine base opposite the abasic site was observed. These observations led to the interesting hypothesis that the favorable base-stacking interactions of the purine bases flanking the abasic site drive the conformational equilibrium toward the extrahelical conformation. The base opposite the abasic site is forced out of the helix and the duplex then collapses to fill the unoccupied space, allowing the flanking purine bases to stack. Using the base flipping assay described in this report, however, the opposite is observed. The largest values for K_{flip} were observed for duplexes A and G in which pyrimidine bases flank the abasic site. The K_{flip} for duplex T was found to be approximately equal to that of duplex G. In addition, the MD simulations do not provide evidence for such a collapse of the duplex. Rather, the hole created by flipping of the thymine is filled with water molecules and the overall structure of the duplex is maintained. However, the differences in the model systems studied here and in ref. 11 are large enough that other explanations are possible.

Conclusions

The sequence dependence of base flipping in DNA containing an abasic site has been studied using a previously developed, non-covalent assay with a fluorescent readout. The dissociation constant of the dansyl-cyclen complex with thymine was determined to be approximately 1 μM. The dissociation constants of the dansyl-cyclen complexes with thymine and thymidine monophosphate are essentially identical, indicating that the added phosphate backbone did not affect binding. The dissociation constant of the dansyl-cyclen–thymine complex can be used to calculate the base flipping equilibrium constants, K_{flip}, for the four different duplex DNAs used in this study. The dissociation constants of

dansyl-cyclen complexes with single strands of DNA, which were originally intended for this purpose, are a less reliable model. This is due to additional interactions of the dansyl-cyclen with DNA bases adjacent to thymine that were observed during MD simulations of these systems as well as in the experimentally observed binding constants, which may affect the dissociation constants of these systems.

The base flipping equilibrium constants for duplex A, duplex G, duplex C and duplex T were found to be 1.44, 0.70, 0.49 and 0.64, indicating that of the four sequences studied only duplex A was found to favor the flipped-out state. The energy differences between the flipped-out conformations of all four sequences studied are small, between 0.4 and 0.6 kcal, but the trends are readily discernible using the assay employed here. The differences between this work using native abasic DNA and a noncovalent probe and previous studies of base flipping using time-resolved fluorescence of covalently modified DNA may indicate a difference in the energetics of base flipping in the two types of systems that reflect the subtle effects on DNA dynamics and warrant further study. The combination of experimental and computational methods used here is a promising approach to elucidate these subtle effects.

Experimental

Thymine and thymidine monophosphate were purchased from Sigma Chemical Co., guanosine monophosphate from VWR International, all DNA strands from Integrated DNA Technologies, and were used as received. All fluorescence spectra were recorded on a Jobin-Yvon Fluoromax-3 spectrometer in a 10 × 4 mm, 1.6 mL quartz cell (Starna Cells, Atascadero, CA). The slit width was 5 nm and the integration time was 0.5 s. The temperature was regulated using an external water bath. The excitation wavelength used for all experiments was 335 nm. Duplex DNA was prepared by mixing equal concentrations of complementary single strands (500 μM in 100 mM phosphate buffer, pH 7.0 with 50 mM NaCl) and heating to 95 °C for 5 min in a water bath. The solutions were then allowed to slowly cool to room temperature. All DNA stock solutions were stored in a −20 °C freezer when not in use. The synthesis of the fluorescent reporter, dansyl-cyclen, and the development of the base flipping assay was previously reported.[8]

The maximum difference in the fluorescence emission for duplex DNAs containing an abasic site and native DNA was previously determined to be 533 nm.[8] For thymine, thymidine monophosphate, guanosine monophosphate, 5′-AATAA-3′, 5′-CCTCC-3′ and 5′-GGTGG-3′, the fluorescence emission was monitored at 500 nm (see ESI†). The assay was performed as follows: To a 1 μM solution of dansyl-cyclen was titrated 17 additions of 0.5 μL of a 250 μM solution of the compound of interest and 13 additions of 4 μL of the same solution. The fluorescence emission of the titrations was monitored at the maximum difference wavelength and each data point is the average of ten repeated measurements, each with a standard error of either <0.15% (duplex DNA, ssDNA) or <0.25% (thymine, thymidine monophosphate, guanosine monophosphate). All experiments were performed in 100 mM phosphate buffer, pH 7.0, at 25 °C. The data were fit using a one-site binding model as provided in Origin 6.1.[12] The error bars shown are the 95% confidence interval for each of the measurements. Three trials were performed for each compound, as shown in the text, and the average dissociation constants (± standard deviation) are reported.

Computational details

All molecular dynamics simulations were performed using the Amber 9 suite of programs using the Cornell *et al.* force field with the adjustments added by Wang *et al.* and the general Amber force field (gaff).[13] The DNA structures, as well as the dansyl-cyclen, were prepared using Insight II.[14] The starting structures for the single-stranded DNA simulations were identical except for the replacement of the appropriate bases. Using Gaussian03, the optimized geometry and charges for the deprotonated thymine bound to the dansyl-cyclen, both at the B3LYP/6-31G* level of theory, were calculated.[15] The *antechamber* module of Amber was then used to perform two-stage restrained electrostatic potential (RESP) fitting to obtain the charges for the new residue.[16] The anionic thymine and the zinc ion were assigned AMBER atom types; the cyclen ring and the dansyl moiety were assigned general Amber force field (gaff) atom types (see ESI† for full listings of charges and atom types). The systems were neutralized using Na^+ counterions and solvated using the TIP3P water model as provided in *xleap*. The solvent box extended 8 Å beyond the DNA structure in each direction. The final systems, 5′-GGTGG-3′, 5′-CCTCC-3′ and duplex DNA, consisted of 4000, 4189 and 20 545 atoms including 1258, 1325 and 6503 water molecules and 4, 4 and 28 sodium ions, respectively. The systems were minimized for 60 000 steps, first with restraints placed on the DNA heavy atoms to allow the water box and hydrogen atoms to equilibrate. A second round of minimization with no restraints was performed. The minimized system was then equilibrated in the constant volume, isothermal (NVT) ensemble for 20 ps with restraints of 10 kcal mol^{-1} $Å^{-2}$ on the DNA. The system was then heated to the final temperature of 300 K over 100 ps in the constant pressure, isothermal (NPT) ensemble. Isotropic position scaling with a relaxation time of 2 ps was used to maintain a pressure of 1 atm and Langevin dynamics with a collision frequency of 1.0 ps^{-1} was used to maintain the temperature at 300 K. All calculations used SHAKE to constrain covalent bonds to hydrogen, which allowed the use of a 0.002 ps time step. Long-range electrostatic interactions were treated using PME with long-range cutoffs of 10 Å applied to the Lennard-Jones interactions.[17] Periodic boundary conditions were used in all calculations. All simulations were analyzed using the *ptraj* module of Amber.

Acknowledgements

We acknowledge the generous allocation of computer resources by the Center for Research Computing at the University of Notre Dame. We are grateful to Prof. B. Smith (Notre Dame) for the use of the fluorometer. L. L. O. is the recipient of a Schmitt Fellowship and a Grace Fellowship from the University of Notre Dame.

References

1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
2 (*a*) S. Klimasauskas, S. Kumar, R. J. Roberts, R. and X. Cheng, *Cell*, 1994, **76**, 357–369; (*b*) K. M. Reinisch, L. Chen, G. L. Verdine and W. N. Lipscomb, *Cell*, 1995, **82**, 143–153; (*c*) A. Banerjee, W. Yang, M. Karplus and G. L. Verdine, *Nature*, 2005, **434**, 612–618; (*d*) D. G. Vassylyev, T. Kashiwagi, Y. Mikami, M. Ariyoshi, S. Iwai, E. Ohtsuka

and K. Morikawa, *Cell*, 1995, **83**, 773–782; (*e*) A. Mees, T. Klar, P. Gnau, U. Hennecke, A. P. M. Eker, T. Carell and L.-O. Essen, *Science*, 2004, **306**, 1789–1793.
3 U. Dornberger, M. Leijon and Fritzsche, *J. Biol. Chem.*, 1999, **274**, 6957–6962.
4 U. D. Priyakumar and A. D. MacKerrel, Jr., *Chem. Rev.*, 2006, **106**, 489–505.
5 N. K. Banavali and A. D. MacKerrel, Jr., *J. Mol. Biol.*, 2002, **319**, 141–160.
6 (*a*) E. Guidice, P. Várnai and R. Lavery, *ChemPhysChem*, 2001, **11**, 673–677; (*b*) E. Guidice, P. Várnai and R. Lavery, *Nucleic Acids Res.*, 2003, **31**, 1434–1443.
7 (*a*) D. C. Ward and E. Reich, *J. Biol. Chem.*, 1969, **244**, 1228–1237; (*b*) C. R. Guest, R. A. Hochstrasser, L. C. Sowers and D. P. Millar, *Biochemistry*, 1991, **30**, 3271–3279; (*c*) D. Xu, K. O. Evans and T. M. Nordlund, *Biochemistry*, 1994, **33**, 9592–9599; (*d*) B. Holz, S. Klimasauskas, S. Serva and E. Weinhold, *Nucleic Acids Res.*, 1998, **26**, 1076–1083; (*e*) K. P. Bjornson, K. J. M. Moore and T. M. Lohman, *Biochemistry*, 1996, **35**, 2268–2282; (*f*) K. S. Christine, A. W. MacFarlane, IV, K. Yang and R. J. Stanley, *J. Biol. Chem.*, 2002, **277**, 38339–38344.
8 (*a*) L. L. O'Neil and O. Wiest, *J. Am. Chem. Soc.*, 2005, **127**, 16800–16801; (*b*) For other application of non-covalent fluorescent assays to the detection of conformational changes in biomolecules, see for example: R. A. Vishwakarma, S. Vehring, A. Mehta, A. Sinha, T. Pomorski, A. Herrmann and A. K. Menon, *Org. Biomol. Chem.*, 2005, **3**, 1275–1283.
9 (*a*) M. Shionoya, E. Kimura and M. Shiro, *J. Am. Chem. Soc.*, 1993, **115**, 6730–6737; (*b*) M. Shionoya, T. Ikeda, E. Kimura and M. J. Shiro, *J. Am. Chem. Soc.*, 1994, **116**, 3848–3859; (*c*) S. Aoki, C. Sigimura and E. Kimura, *J. Am. Chem. Soc.*, 2000, **122**, 4542–4548; (*d*) S. Aoki, C. Sigimura and E. Kimura, *J. Am. Chem. Soc.*, 1998, **120**, 10094–10102; (*e*) E. Kikuta, M. Murata, N. Katsube, T. Koike and E. Kimura, *J. Am. Chem. Soc.*, 1999, **121**, 5426–5436; (*f*) E. Kimura, M. Kikuchi, H. Kitamura and T. Koike, *Chem. Eur. J.*, 1999, **5**, 3113–3213; (*g*) E. Kimura, H. Kitamura, K. Ohtani and T. Koike, *J. Am. Chem. Soc.*, 2000, **122**, 4668–4677; (*h*) E. Kimura, S. Aoki and M. Shiro, *J. Am. Chem. Soc.*, 1997, **119**, 3068–3076; (*i*) E. Kimura, T. Ikeda, S. Aoki and M. Shionoya, *J. Biol. Inorg. Chem.*, 1998, **3**, 259–267.
10 E. L. Rachovsky, E. Seibert, J. T. Stivers, R. Osman and J. B. A. Ross, *Biochemistry*, 2001, **40**, 957–967.
11 Ph. Cuniasse (*a*) G. V. Fazakerley, W. Guschlbauer, B. E. Kaplan and L. C. Sowers, *J. Mol. Biol.*, 1990, **213**, 303–314; (*b*) M. P. Singh, G. C. Hill, D. Péoc'h, B. Rayner, J.-L. Imbach and J. W. Lown, *Biochemistry*, 1994, **33**, 10271–10285; (*c*) Y. Coppel, N. Berthet, C. Colombeau, C. Colombeau, J. Garcia and J. Lhomme, *Biochemistry*, 1997, **36**, 4817–4830; (*d*) K. Y. Wang, S. A. Parker, I. Goljer and P. H. Bolton, *Biochemistry*, 1997, **36**, 11629–11639; (*e*) C. A. Gelfand, G. E. Plum, A. P. Grollman, F. Johnson and K. J. Breslauer, *Biochemistry*, 1998, **37**, 7321–7327; (*f*) R. D. Berger and P. H. Bolton, *J. Biol. Chem.*, 1998, **273**, 15565–15573; (*g*) D. Barsky, N. Foloppe, S. Ahmadia, D. M. Wilson, III and A. D. MacKerell, Jr., *Nucleic Acids Res.*, 2000, **28**, 2613–2626; (*h*) S. T. Hoehn, C. J. Turner and J. Stubbe, *Nucleic Acids Res.*, 2001, **29**, 3413–3423; (*i*) J. Chen, F.-Y. Dupradeau, D. A. Case, C. J. Turner and J. Stubbe, *Biochemistry*, 2007, **46**, 3096–3107.
12 *Origin 6.1*, OriginLab Corp, Northampton, MA, 2000.
13 (*a*) D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross and P. A. Kollman, *AMBER 9*, University of California, San Francisco, 2006; (*b*) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, T. J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197; (*c*) J. Wang, P. Cieplak, P. and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074; (*d*) J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollamn and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
14 *Insight II*, Accelrys Software Inc., San Diego, CA, 2005.
15 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03 (Revision C.02)*, Gaussian, Inc., Wallingford, CT, 2004.
16 (*a*) C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280; (*b*) W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollman, *J. Am. Chem. Soc.*, 1993, **115**, 9620–9631.
17 T. Darden, D. York and L. G. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
18 L. L. O'Neil, A. Grossfield and O. Wiest, *J. Phys. Chem. B*, 2007, **111**, 11843–11849.

