**Graphical Abstract**
To create your abstract, type over the instructions in the template box below. Fonts or abstract dimensions should not be changed or altered.

**The use of molecular dynamics simulations to evaluate the DNA sequence-selectivity of G-A cross-linking PBD-duocarmycin dimers**

Paul J. M. Jackson \(^{a,b}\), Khondaker M. Rahman \(^{a,b}\) and David E. Thurston \(^{a,b}\)

\(^{a}\) Femtogenix Ltd, Britannia House, 7 Trinity Street, London SE1 1DB, United Kingdom.

\(^{b}\) Institute of Pharmaceutical Science, Faculty of Life Sciences & Medicine, King’s College London, Britannia House, 7 Trinity Street, London SE1 1DB, United Kingdom.
The use of molecular dynamics simulations to evaluate the DNA sequence-selectivity of G-A cross-linking PBD-duocarmycin dimers

Paul J. M. Jackson \textsuperscript{a,b}, Khondaker M. Rahman \textsuperscript{a,b} and David E. Thurston \textsuperscript{a,b}

\textsuperscript{a} Femtogenix Ltd, Britannia House, 7 Trinity Street, London SE1 1DB, United Kingdom
\textsuperscript{b} Institute of Pharmaceutical Science, King’s College London, Britannia House, 7 Trinity Street, London SE1 1DB, United Kingdom

\textbf{ARTICLE INFO}

Article history:
Received
Revised
Accepted
Available online

Keywords:
DNA cross-linking
Pyrrolobenzodiazepine
Duocarmycin
PBD-CBI dimer
Molecular dynamics

\textbf{ABSTRACT}

The pyrrolobenzodiazepine (PBD) and duocarmycin families are DNA-interactive agents that covalently bond to guanine (G) and adenine (A) bases, respectively, and that have been joined together to create synthetic dimers capable of cross-linking G-G, A-A, and G-A bases. Three G-A alkylating dimers have been published to date, with defined DNA-binding sites proposed for two of them. In this study we have used molecular dynamics simulations to elucidate preferred DNA-binding sites for the three published molecular types. For the PBD-CPI dimer UTA-6026 (1), our simulations correctly predicted its favoured binding site (i.e., 5'-C(G\textsubscript{AATT\textsubscript{A}})-3') as identified by DNA cleavage studies. However, for the PBD-CI molecule (“Compound 11”, 3), we were unable to reconcile the results of our simulations with the reported preferred cross-linking sequence (5'-ATTTGCC(\textsubscript{G})\textsubscript{3}-3'). We found that the molecule is too short to span the five base pairs between the A and G bases as claimed, but should target instead a sequence such as 5'-ATTTTC(G\textsubscript{3})-3' with two less base pairs between the reacting G and A residues. Our simulation results for this hybrid dimer are also in accord with the very low interstrand cross-linking and \textit{in vitro} cytotoxicity activity reported for Compound 11. Although a preferred cross-linking sequence was not reported for the third hybrid dimer (“27\textsubscript{S}2”, 2), our simulations predict that it should span three base pairs between covalently reacting G and A bases (i.e., 5'-GTAT(\textsubscript{A})\textsubscript{3}-3').

2009 Elsevier Ltd. All rights reserved.
Figure 1. A. The PBD Dimers (e.g., DSB-120 and SJG-136) which can form DNA minor-groove interstrand G-G cross-links; B. A CPI Dimer (e.g., Bizelesin) which can form minor-groove interstrand A-A cross-links; C. An example of a CPI-PBD hybrid molecule that can cross-link A-G base pairs. The mode of interstrand cross-linking and likely base-pair span is shown above each family structure (covalently modified bases in red).

The incorporation of both GC-interactive and AT-interactive moieties into symmetric G-G and A-A dimers led to the concept of joining PBD and CPI units together to create hybrid molecules capable of forming cross-links to both G and A bases simultaneously (Figure 1C), which are the subject of this study. Three types of these molecules have been reported in the literature[45][46][47], and their structures are shown in Figure 2.

The objective of this study was to use, for the first time, molecular dynamics simulations to predict the preferred DNA recognition sequences for the three published types of G-A cross-linking agents[45][46][47] shown in Figure 2, and then compare these predictions with the cross-linking sites proposed by the authors. Hurley and co-workers published the first example of a CPI-PBD dimer (UTA-6026, 1, Figure 2) in 2001[45], reporting that it cross-links the sequence 5'-C(GAATTA)-3' (covalently-modified bases underlined; brackets denote base on opposite strand), has a CL_{50} value (a measure of cross-linking potency) of between 1-10 µM in calf thymus DNA, and an IC_{50} as low as 0.047 nM in some tumour cell lines (i.e., colon SW480). Interestingly, in their gel studies, cross-linking was not complete even at 100 µM, suggesting that the molecule is a relatively inefficient cross-linking agent. In 2003, Denny and co-workers reported[46] an extensive study of PBD-CPI hybrids which differed from the Hurley dimer in containing simple methylene linkers between the CPI and PBD units, and these molecules are now being developed as payloads for antibody-drug conjugates through Genentech. Their most potent molecule 27esS (2, Figure 2) had a CL_{50} value of ~0.013µM toward pcDNA3 plasmid DNA, and an IC_{50} as low as 0.0078 nM in some tumour cell lines (i.e., EMT6). The molecule also had limited antitumour activity in human tumour xenograft mouse models. The authors provided some evidence that G-A cross-linking was occurring, but did not suggest a specific sequence to which 2 should bind. Finally, in 2006 Lee and co-workers synthesized the seco-amino-CI-PBD hybrid ‘Compound 11’ (3, Figure 2) which they reported to provide a very low degree of cross-linking toward pBR-322 DNA at 10 µM, and to have an IC_{50} of 0.56 µM in P815 cells after 72 hours exposure. Based on the results of a thermally-induced cleavage assay, they reported that 3 forms cross-links at the 5'-AAATTGTTCC(G)-3' sequence.
gel electrophoresis was used to investigate the cross-linking orientation to alkylate DNA cooperatively[45]. Non-denaturing n = 3 to allow the PBD and CPI units to lie in the correct amide functionality (connecting to the indole ring) should be methylene linker length between the C8-position of the PBD and the amide functionality (connecting to the indole ring) should be n = 3 to allow the PBD and CPI units to lie in the correct orientation to alkylate DNA co-operatively[45]. Non-denaturing gel electrophoresis was used to investigate the cross-linking ability of 1, and a thermal cleavage gel-based assay was performed to identify adenine cleavage points. Together, these data indicated cross-link formation at two different 5’-C(G)AATTA-3’ sites (alkylated bases underlined).

In the present study, molecular dynamics simulations were carried out over a 10ns time-scale to confirm the proposed six base-pair span of 1, and the interstrand cross-link at 5’-GCC(G)AATTA-3’ as derived from the published experimental data. The simulation results were in full agreement with the experimental data. When mono-covalently bound through the CPI moiety alone to position A8 of this sequence, the PBD was located close enough to G18 for nucleophilic attack and covalent attachment to occur. Furthermore, when both the PBD and CPI units were covalently bound to G18 and A8, respectively, in the same sequence, negligible distortion of the DNA was observed (i.e., RMSD = 0.78Å, Table 1), with the molecule well accommodated in the minor groove (Figure 3).

Simulations with 1 covalently bound to sequences one base shorter and one or two bases longer (i.e., 5’-CAATA-3’, 5’-CAAAATTA-3’ and 5’-CAAAATTATTA-3’) gave higher RMSD values of 1.96, 1.00 and 3.04, respectively, thus supporting the preferred 5’-CAATA-3’ sequence (Table 1). It is noteworthy that the difference in RMSD values between the preferred and shorter sequence (i.e., 0.78 vs 1.96) was much greater than between the preferred and one base-pair longer sequences (i.e., 0.78 vs 1.00), suggesting that if the preferred sequence were to be extended by one base-pair, the ligand may still be able to form adducts. It is also interesting that the differences between all of these RMSD values for 1 (i.e., 0.78, 1.96, 1.00 and 3.04Å) are far smaller than those observed for 2 (i.e., 0.93, 2.42 and 6.45Å). 2 contains a methylene chain between alkylating moieties, whereas 1 contains an indole group, which may stabilize the adduct to a greater degree than the flexible methylene linker.

Non-covalently bound simulations provided a similar pattern to dimers 2 and 3 in that the molecule appeared to be restrained over a guanine for the duration of the simulation, but clearly showed that six base-pairs were being spanned (Figure 3). Also, because this sequence span is independent of the length of the DNA sequences being investigated, the Free Energy of Binding calculations are similar in each case, and do not reflect the more meaningful observations based on RMSD calculations from covalently bound simulations.

The CPI-PBD hybrid UTA-6026 (1) was designed by Hurley and co-workers using a molecular modeling approach (details not published) to span six base-pairs, covalently binding to a guanine (through the PBD) and an adenine (through the CPI moiety)[45]. An indole group was included as part of the central linking moiety to increase interactivity with functional groups in the minor groove through van der Waals interactions. Hurley and co-workers predicted that, for their chosen sequence, the optimal methylene linker length between the C8-position of the PBD and the amide functionality (connecting to the indole ring) should be n = 3 to allow the PBD and CPI units to lie in the correct orientation to alkylate DNA co-operatively[45]. Non-denaturing gel electrophoresis was used to investigate the cross-linking ability of 1, and a thermal cleavage gel-based assay was used to identify adenine cleavage points. Together, these data indicated cross-link formation at two different 5’-C(G)AATTA-3’ sites (alkylated bases underlined).

In the present study, molecular dynamics simulations were carried out over a 10ns time-scale to confirm the proposed six base-pair span of 1, and the interstrand cross-link at 5’-GCC(G)AATTA-3’ as derived from the published experimental data. The simulation results were in full agreement with the experimental data. When mono-covalently bound through the CPI moiety alone to position A8 of this sequence, the PBD was located close enough to G18 for nucleophilic attack and covalent attachment to occur. Furthermore, when both the PBD and CPI units were covalently bound to G18 and A8, respectively, in the same sequence, negligible distortion of the DNA was observed (i.e., RMSD = 0.78Å, Table 1), with the molecule well accommodated in the minor groove (Figure 3).

Simulations with 1 covalently bound to sequences one base shorter and one or two bases longer (i.e., 5’-CAATA-3’, 5’-CAAAATTA-3’ and 5’-CAAAATTATTA-3’) gave higher RMSD values of 1.96, 1.00 and 3.04, respectively, thus supporting the preferred 5’-CAATA-3’ sequence (Table 1). It is noteworthy that the difference in RMSD values between the preferred and shorter sequence (i.e., 0.78 vs 1.96) was much greater than between the preferred and one base-pair longer sequences (i.e., 0.78 vs 1.00), suggesting that if the preferred sequence were to be extended by one base-pair, the ligand may still be able to form adducts. It is also interesting that the differences between all of these RMSD values for 1 (i.e., 0.78, 1.96, 1.00 and 3.04Å) are far smaller than those observed for 2 (i.e., 0.93, 2.42 and 6.45Å). 2 contains a methylene chain between alkylating moieties, whereas 1 contains an indole group, which may stabilize the adduct to a greater degree than the flexible methylene linker.

Non-covalently bound simulations provided a similar pattern to dimers 2 and 3 in that the molecule appeared to be restrained over a guanine for the duration of the simulation, but clearly showed that six base-pairs were being spanned (Figure 3). Also, because this sequence span is independent of the length of the DNA sequences being investigated, the Free Energy of Binding calculations are similar in each case, and do not reflect the more meaningful observations based on RMSD calculations from covalently bound simulations.

The CPI-PBD hybrid UTA-6026 (1) was designed by Hurley and co-workers using a molecular modeling approach (details not published) to span six base-pairs, covalently binding to a guanine (through the PBD) and an adenine (through the CPI moiety)[45]. An indole group was included as part of the central linking moiety to increase interactivity with functional groups in the minor groove through van der Waals interactions. Hurley and co-workers predicted that, for their chosen sequence, the optimal methylene linker length between the C8-position of the PBD and the amide functionality (connecting to the indole ring) should be n = 3 to allow the PBD and CPI units to lie in the correct orientation to alkylate DNA co-operatively[45]. Non-denaturing gel electrophoresis was used to investigate the cross-linking ability of 1, and a thermal cleavage gel-based assay was used to identify adenine cleavage points. Together, these data indicated cross-link formation at two different 5’-C(G)AATTA-3’ sites (alkylated bases underlined).

In the present study, molecular dynamics simulations were carried out over a 10ns time-scale to confirm the proposed six base-pair span of 1, and the interstrand cross-link at 5’-GCC(G)AATTA-3’ as derived from the published experimental data. The simulation results were in full agreement with the experimental data. When mono-covalently bound through the CPI moiety alone to position A8 of this sequence, the PBD was located close enough to G18 for nucleophilic attack and covalent attachment to occur. Furthermore, when both the PBD and CPI units were covalently bound to G18 and A8, respectively, in the same sequence, negligible distortion of the DNA was observed (i.e., RMSD = 0.78Å, Table 1), with the molecule well accommodated in the minor groove (Figure 3).

Simulations with 1 covalently bound to sequences one base shorter and one or two bases longer (i.e., 5’-CAATA-3’, 5’-CAAAATTA-3’ and 5’-CAAAATTATTA-3’) gave higher RMSD values of 1.96, 1.00 and 3.04, respectively, thus supporting the preferred 5’-CAATA-3’ sequence (Table 1). It is noteworthy that the difference in RMSD values between the preferred and shorter sequence (i.e., 0.78 vs 1.96) was much greater than between the preferred and one base-pair longer sequences (i.e., 0.78 vs 1.00), suggesting that if the preferred sequence were to be extended by one base-pair, the ligand may still be able to form adducts. It is also interesting that the differences between all of these RMSD values for 1 (i.e., 0.78, 1.96, 1.00 and 3.04Å) are far smaller than those observed for 2 (i.e., 0.93, 2.42 and 6.45Å). 2 contains a methylene chain between alkylating moieties, whereas 1 contains an indole group, which may stabilize the adduct to a greater degree than the flexible methylene linker.

Non-covalently bound simulations provided a similar pattern to dimers 2 and 3 in that the molecule appeared to be restrained over a guanine for the duration of the simulation, but clearly showed that six base-pairs were being spanned (Figure 3). Also, because this sequence span is independent of the length of the DNA sequences being investigated, the Free Energy of Binding calculations are similar in each case, and do not reflect the more meaningful observations based on RMSD calculations from covalently bound simulations.
within linearized pBR322 plasmid DNA. Based on this, we initially conducted an MD simulation of the published adduct (i.e. 3) covalently cross-linked to the seven base-pair duplex sequence 5'-GAATTTCC(G)C-3'), with the introduction of a ‘GC-lock’ (i.e., terminal guanine bases) to ensure that the DNA duplex remained intact throughout the simulation. As CPI/CBI-type molecules are known to be more reactive toward DNA than PBD molecules[46], the simulations involved initial covalent attachment of just the seco-amino-CI unit, leaving the PBD unreacted. This allowed us to investigate the span of the entire hybrid molecule. Next, the DNA sequence was gradually reduced in length to investigate the shortest possible adduct which was determined to be 5'-ATTC(G)-3' (Table 1). An identical set of simulations was then undertaken with both PBD and CI covalently bound to guanine and adenine residues, respectively, within the same set of sequences. Finally, simulations were undertaken with 3 bound non-covalently in the DNA minor groove to enable calculation of free energy of binding of the ligand to DNA (Table 1).

It was immediately evident from the simulations that 3 was unable to span the seven base-pair cross-link proposed by Lee and co-workers (i.e., 5'-ATTTTC(G)-3'). In particular, simulations with the molecule bis-covalently bonded to this sequence indicated that severe distortion of the DNA helix would occur with disruption of base-pairing and extensive minor groove widening (Figure 4A and 4C). This result was also supported by RMSD plots of the adduct over the duration of the simulation (Supplementary Material), and RMSD calculations between the lowest energy conformation of the molecule derived from MD simulations and the lowest energy conformation after energy minimization (pre-MD simulation). RMSD values were calculated to be 2.45Å, suggesting a high degree of distortion of the ligand:DNA complex. Crucially, simulations with only the

### Table 1. Free Energy of Binding and RMSD calculations for interaction of the PBD-duocarmycin dimer hybrids 1-3 with the DNA sequences shown, as calculated over the duration of the MD simulations.

<table>
<thead>
<tr>
<th>Hybrid Dimer</th>
<th>Molecule Span (number of base-pairs)</th>
<th>DNA Sequence (Sequence being analysed is highlighted in bold and red)</th>
<th>Free Energy of Binding Calculations (generated from non-covalently bound simulations) (kcal/mol)</th>
<th>RMSD Calculations (generated from covalently bound simulations) (Å)</th>
<th>Base-pairing maintained during simulation? (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5-6</td>
<td>5'-GCCAAATGC-3'</td>
<td>-55.45</td>
<td>1.96</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GCCAAATGGC-3'</td>
<td>-54.15</td>
<td>0.78</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GCCAAATTGC-3'</td>
<td>-53.80</td>
<td>1.00</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GCCAAAATTCG-3'</td>
<td>-53.64</td>
<td>3.04</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>4-5</td>
<td>5'-GCGTATGC-3'</td>
<td>-41.11</td>
<td>0.93</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GCCGTAATGC-3'</td>
<td>-40.97</td>
<td>2.42</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GCCGTAATGC-3'</td>
<td>-38.93</td>
<td>6.45</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GCCGTAATGC-3'</td>
<td>-41.60</td>
<td>5.23</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>4-5</td>
<td>5'-GAATTCG-3'</td>
<td>-48.78</td>
<td>1.82</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GAATTTGC-3'</td>
<td>-51.41</td>
<td>1.15</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GAATTTTCG-3'</td>
<td>-52.84</td>
<td>5.72</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GAATTTTCG-3'</td>
<td>-52.63</td>
<td>8.45</td>
<td>N</td>
</tr>
</tbody>
</table>

seco-amino-CI unit covalently bonded to the adenine indicated that the PBD moiety is then situated directly over the T6:A15 base-pair (Figure 4B and 4C), at least two base-pairs away from the reacting guanine proposed by Lee and co-workers[49], thus rendering the proposed second alkylation event unlikely.

A similar pattern was evident in simulations involving the six base-pair sequence 5'-ATTTTC(G)-3', where mono-covalent simulations indicated difficulty in cross-link formation, and dual-covalently bound simulations suggested significant DNA distortion (indicated by a RMSD value of 5.72 (Table 1). Although mono-covalently bound simulations suggested a potential cross-link at the four base-pair sequence 5'-ATTC(G)-3', dual-covalently bound simulations (RMSD = 1.82) suggested that the five base-pair sequence 5'-ATTTTC(G)-3' should form the most favoured adduct. This was supported by the RMSD calculations (Table 1), which indicated less distortion for both the four and five base-pair adducts (i.e., RMSD = 1.82 and 1.15, respectively), suggesting that both might form.

Overall these simulations highlighted the importance of DNA breathing in the accommodation of ligands in the minor groove. Our observation that 3 is likely to cross-link the five base-pair DNA sequence 5'-ATTTTC(G)-3' rather than the extended sequence 5'-ATTTTTTC(G)-3' as suggested by the authors was further supported by the proximal location of the unbound PBD imine close to the reacting guanine in the mono-alkylated adduct in the former sequence. Although significant distortion was observed in simulations when dual-covalently bound to the idealized DNA sequence, models suggested that the molecule is likely to span five base-pairs (see Supplementary Material for molecular model). Furthermore, as 5'-ATTTTC(G)-3' is not obvious as a potential binding site from the published gel studies, it is likely that the authors were observing mono-alkylation rather than cross-linking events in their experiments.

Non-covalently bound simulations of 3 were also not reflective of the reported experimental results. Previous studies have shown that adenine-binding molecules such as the CPI family direct the sequence targeting of unsymmetrical PBD-CPI hybrids[46]. However, in this case, the non-covalently bound simulations suggested that the PBD moiety directs binding due to the formation of a sequence-selective H-bond between the lone pair of the N10 of the PBD and the exocyclic amine of a guanine residue. As a result, in the case of every non-covalently bound simulation undertaken, and thus every Free Energy of Binding calculation, the PBD moiety remained restrained over the intended reacting guanine (i.e., G14 on the reverse strand in the case of 5'-GAATTTTC(G)-3'), spanning DNA in an A-ring-5' orientation in the process. As 3 spans four base-pairs, addition of an extra thymine base to the centre of the binding site does not alter the sequence with which the molecule is interacting, as the PBD is restrained over a G:C base-pair. As a result, calculated free energy of binding values are similar for all sequences analysed.
Bioinformatics analysis of the linearized DNA plasmid pBR322 suggests 20 potential binding sites for 3 (from within a total of 4361 bases) based on the proposed preferred five base-pair 5'-GWWT(W)-3' binding site (where W represents adenine or thymine, A represents guanine and C is cytosine) from our modeling studies. The relative prevalence of the sequence 5'-GWWT(W)-3' suggests that a significant band in the cross-linking gel relating to this sequence should have been observed. However, as this was not the case, it is likely that other factors may contribute to the lack of interaction of 3 with DNA. In this context, it is noteworthy that, unlike the potent PBD dimers in the literature (Figure 1A)[50], 3 possesses a linker between the C8-position of the PBD and the amnic carbonyl based on two methylene groups. Symmetrical PBD dimers in which the linker is odd-numbered (i.e. n = 3 or 5) are known to be more DNA interactive and cytotoxic (e.g. 10 more potent in some cancer cell lines) than even-numbered equivalent molecules (i.e. n = 2 or 4), as the alkylating moieties are more-appropriately oriented for the alkylation events[51]. Therefore, we considered whether this effect could have contributed to the poor cross-linking ability observed for 3 by carrying out simulations on analogues containing an alternative trimethylene linker (n=3) (data not shown). This led us to conclude that these rules derived from the PBD dimers may also apply to asymmetric hybrid dimer molecules, whereby odd-numbered methylene linkers appear to be preferential for bis-alkylation events to occur.

For the PBD-CPI hybrids produced by Tercel and co-workers[46], their thermal cleavage experiments suggested that a number of mono-alkylation events occurred between the CPI moiety of 2 and adenine bases, and results from their Comet assay analysis suggested greater interstrand cross-linking potency for some of their hybrid dimers compared to the PBD dimer DSB-120[46]. However, potential DNA cross-linking sites were not identified in their study. Therefore, we used our molecular dynamics simulations to try to identify potential cross-linking sites.

Starting with a four base-pair sequence (5'-GTAT-3'), an increasing number of adenine-thymine base-pairs were inserted between the 5'-T and 5'-A bases (i.e., 5'-GT(A),T-3'), increasing the total span of base-pairs up to a maximum of seven (i.e., 5'-GTAAAAAT-3'). A:T base-pairs were selected as the central components in an effort to maximize non-covalent interaction with functional groups in the DNA minor groove, and minimize potential steric hindrance from the C2-amino groups of guanine bases. In a similar manner to the simulations undertaken on the other hybrid dimers, the interaction of 2 with DNA was simulated in three ways; non-covalently bound, mono-alkylated with the CPI moiety covalently bound to an adenine base, and dual-alkylated. These simulations suggested that the four base-pair sequence 5'-GTAT-3' is the most favoured cross-linking site for the PBD-CPI hybrid dimer (Figure 5). This was mainly reflected in the bis-covalently bound simulations, where little distortion of the DNA structure was observed (i.e., RMSD = 0.93Å, Table 1), and base-pairing was maintained. Introduction of additional base pairs beyond the established four base-pair optimum resulted in increased DNA disorder, which was reflected in higher RMSD values (e.g., RMSD values of up to 6.45Å in the longer sequences analyzed, Table 1).

These results showed that 5'-GWWT-3' (where W represents A or T) should be the most favoured cross-linking site for 2. In a similar manner to simulations of 3, Free Energy of Binding
calculations did not reflect the predicted cross-linking sites. During simulations between 2 and each sequence, a sequence-selective hydrogen bond was formed between the C2-amino group of guanine and the lone pair of the N10 of the PBD, thereby restraining 2 over a guanine base for the duration of the simulation. Therefore, as the span of the molecule is four to five base-pairs, there was little difference between each sequence from the simulation perspective (Table 1), and so negligible differences in Free Energy of Binding values were observed.

Finally, it is known that symmetrical PBD dimers in which the linker attached to the C8/C8’-positions is odd-numbered in length (e.g., n = 3 or n = 5) have a higher DNA binding affinity and are more cytotoxic (e.g., 10^7 more potent in some cancer cell lines) than even-numbered equivalent molecules (e.g., n = 2 or 4). This is thought to be due to the alkylating moieties being in a more optimal orientation and alignment for alkylation events to occur in the case of odd-numbered linkers[51]. The hybrid dimers synthesized by Tercel and co-workers[46] appear to follow this rule. For example, they synthesized and evaluated a series based on CPI and PBD moieties separated by a methylene linker (i.e., based on 2, Figure 2), with differentiation achieved through variation of linker length (i.e., n = 1-5). Compounds with odd-numbered linker lengths (in this case n = 3 and n = 5) significantly out-performed those with even-numbered linker lengths, providing a cross-linking hierarchy of n = 2 < 1 < 4 < 5 < 3. This experimental observation is in accord with our models of 3, which suggest that the analogue with a linker of n = 3 should be more DNA-interactive than the n = 2 analogue.

It is important to acknowledge that this molecular dynamics simulation approach may have a number of limitations. For example, one problem is that the covalent binding energy of forming just one covalent bond between any ligand and DNA is so large (i.e., ~75 kcal/mol per C-N bond) that there is abundant energy available to radically distort the DNA structure, and for cross-linking agents which form two covalent bonds even more free energy is available (i.e., ~150 kcal/mol). This covalent binding free energy is not encompassed in the molecular dynamics simulation paradigm, and so this approach may not be able to predict the formation of highly distorted DNA adducts (e.g., those produced by cross-linking agents such as cisplatin)[3]. One practical problem when carrying out a molecular dynamics simulation is the need to establish which of the many conformations formed during the simulations are potentially relevant in real life. In the case of ligand-DNA interactions, if the simulations produce solutions that strongly resemble canonical B-form DNA with the ligand fitting perfectly in the minor groove and making good Van der Waals interactions in all directions, then it might be reasonable to assume that such "equilibrium" configurations represent plausible structures in real life. Although we cannot be completely confident that the conformation chosen post-minimization (i.e., before undertaking simulation) is the most biologically relevant, it is evident from our results that measuring the degree of distortion over the duration of a simulation can provide an insight into cross-linking efficiency. In the case of the molecules studied here, cross-linking potency, cytotoxicity and degree of distortion of ligand-DNA adducts are experimentally correlated, and it is gratifying that our MD simulations appear to follow this trend. Interestingly, in the case of the PBD dimers (Figure 1A), their significant cytotoxicity in cells is thought to be due to their perfect fit within the minor groove and consequent lack of distortion upon covalent cross-linking, thus avoiding the attention of DNA repair enzymes[17].

In summary, we report the first molecular dynamics simulations of the interaction of PBD-CPI hybrid dimers with DNA. In the case of the Hurley hybrid dimer UTA-6026 (1)[45], our simulations correctly predicted the cross-link shown to form experimentally in a DNA cleavage assay, and this served to validate our modeling protocols for use with the other G-A cross-linking agents, 2 and 3. However, our modeling studies illustrated that the cross-link proposed by Purnell and co-workers for 2 is unlikely to occur, and that the results of their reported gel studies most likely relate to mono-alkylation rather than cross-linking events. Also, our calculations demonstrate that RMSD values between the lowest energy derived from energy minimization studies and those obtained through molecular dynamics simulations are a valuable tool in predicting cross-link formation, and may be useful in the design of more potent next-generation PBD-(CBI/CPI/CI) dimers targeted to specific DNA sequences.

Acknowledgement

The Frederick Gardner Cottrell Foundation (Arizona, USA) is thanked for financial support of this work.

References and Notes


