Copper carbenes alkylate guanine chemoselectively through a substrate directed reaction†

Stefanie N. Geigle,a Laura A. Wyss,b Shana J. Sturlab and Dennis G. Gillingham*a

Cu(i) carbenes derived from α-diazocarbonyl compounds lead to selective alkylation of the O6 position in guanine (O6-G) in mono- and oligonucleotides. Only purine-type lactam oxygens are targeted – other types of amides or lactams are poorly reactive under conditions that give smooth alkylation of guanine. Mechanistic studies point to N7G as a directing group that controls selectivity. Given the importance of O6-G adducts in biology and biotechnology we expect that Cu(i)-catalyzed O6-G alkylation will be a broadly used synthetic tool. While the propensity for transition metals to increase redox damage is well-appreciated, our results suggest that transition metals might also increase the vulnerability of nucleic acids to alkylation damage.

Introduction

Catalytic reactions with metal carbenes are broadly applied in synthetic chemistry.1,2 Although metal carbenes are reactive intermediates, their chemoselectivity can be steered through the right combination of metal, ligand, and carbene precursor.3 Even in water metal carbenes can still perform an impressive array of transformations.4–7 In our studies on the post-synthetic tailoring of large nucleic acids through metal catalysis we have found that donor–acceptor substituted carbenes8 derived from Rh(ii) and Cu(i) give primarily N–H insertion of exocyclic amine groups in unpaired regions of DNAs and RNAs.9–11 Here we show that unstabilized Cu(i) carbenes derived from α-diazo acetate and amide derivatives have a completely different chemoselectivity: they quickly and cleanly alkylate the O6 position in guanine and inosine (see Fig. 1B for a representative reaction with GMP and ethyl α-diazoacetate (EDA)). Cu(i) is known to bind amides (especially amidates),12 but there is no precedence which would have predicted such a pronounced selectivity for the lactam-like oxygen of guanine. While most instances of copper13 direction by heterocycles in synthetic chemistry14–16 and biology17,18 employ histidine or pyridine (see Fig. 1A), our results show that the purine ring system of guanine is also a powerful directing group in copper catalysed alkylation (see Fig. 1B) reactions.

DNA surveillance and repair pathways to maintain genomic integrity.19 Alkylation at the O6 of guanine20 is particularly dangerous because polymerases tend to misincorporate thymidines in response to O6-(alkyl)-G during replication, leading to
potentially carcinogenic G-to-A transition mutations.\textsuperscript{22-25} The colon cells of people on high red meat diets have increased carboxymethyl damage at O\textsuperscript{6}-G.\textsuperscript{26-27} Epidemiological studies have shown that the high-red meat diet demographic in Western cultures suffer as well from a higher prevalence of gastrointestinal cancers.\textsuperscript{28-29} Whether there is a causative connection between these two occurrences is currently unclear. Studying the biology of O\textsuperscript{6}-(carboxymethyl)-G damage (O\textsuperscript{6}-cmG), however, is hampered by synthetic challenges. Multi-step procedures, protecting group manipulations,\textsuperscript{30} and the need for strong electrophilic phosphorous reagents,\textsuperscript{31} combine to make both the synthesis and purification of O\textsuperscript{6}-(alkyl)-G derivatives laborious. With copper catalysed alkylation the carboxymethyl motif is introduced in a single step from naturally occurring precursors.

Results and discussion

Reaction discovery and substrate scope

Cu(i) carbene(s) derived from \(\alpha\)-diazo-carbonyl compounds are highly selective for O\textsuperscript{6} alkylation of unpaired guanines and inosines in ssDNA, monophosphates (MP) and triphosphates (TP) (see Fig. 1B and Table 1). The surprising selectivity prompted us to synthesize an O\textsuperscript{6}-alkyl-ino sine adduct by traditional routes to check the site-selectivity: identical characteri zation data confirm the proposed structures (see the ESI, page S5-6). The method is the most concise and selective approach ever reported to the O\textsuperscript{6}-(alkyl)-G motif and it is post-synthetic, obviating the need for solid phase DNA synthesis with modified phosphoramidites. The convergent assembly opens the door to a range of O\textsuperscript{6}-(alkyl)-G derivatives from simple starting materials.

We explored the efficiency and selectivity of O\textsuperscript{6}-G targeting by characterizing reactions with nucleotide monophosphates. Treatment of dGMP or GMP with 20 mol\% Cu(s)SO\textsubscript{4} (in situ reduced by ascorbate) and ethyl \(\alpha\)-diazoacetate (EDA) in aqueous buffer resulted in high conversion and isolated yields of ethylcarboxymethyl product (see Table 1, entry 1 and 2) within 30 min. Under the same conditions, the structural analogues IMP, GTP, and dGTP were efficiently alkylated at their O\textsuperscript{6} positions (see Table 1, entry 6, 11 and 12). In contrast, reactions without catalyst (entry 5) or with different MPs like UMP, AMP, CMP, or TMP (entries 7–10) resulted in little or no detectable alkylation. Rather, in cases where substrates lacked a guanine, the starting material was partially converted through unselective pathways that we could not completely characterize – in each case some alkylation was usually observed in the MS but these did not give substantial HPLC peaks that we could isolate. The same was also seen in short DNA fragments such as d(TAT) (see entry 14); here again starting material was consumed over time, but no single alkylation product could be isolated. In contrast, when guanine was present post-synthetic modifications of a DNA 3-mer d(TTG) and 9-mer d(TTTTGGTTTT) with the EDA-derived Cu(i) carbene gave a single large alkylation product (only O\textsuperscript{6}-G alkylation observed, see Table 1 entries 13 and 17). In the substrates that lack guanine we cannot account for the mass balance: we believe in these cases that oxidation is a side-reaction that consumes starting material when no guanine is available for the carbene-based reaction.

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Entry & Substrate & Time (min) & Cosolvent & Conv. (%) & Yield. (%) \\
\hline
1\textsuperscript{a} & dGMP & 30 & None & 98\textsuperscript{f} & 86 \\
2\textsuperscript{a} & GMP & 30 & None & 96\textsuperscript{f} & 92 \\
3\textsuperscript{a} & GMP & 30 & DMSO & 93\textsuperscript{f} & — \\
4\textsuperscript{a} & GMP & 30 & None & 87 (ref. 32) & — \\
5\textsuperscript{a} & GMP & 150 & None & 0 & — \\
6\textsuperscript{a} & IMP & 30 & None & 92\textsuperscript{g} & 81 \\
7\textsuperscript{a} & AMP & 30 & None & 40 & — \\
8\textsuperscript{a} & CMP & 30 & None & 38 & — \\
9\textsuperscript{a} & UMP & 30 & None & 21 & — \\
10\textsuperscript{a} & TMP & 30 & None & 13 & — \\
11\textsuperscript{b} & GTP & 30 & tBuOH & 98\textsuperscript{g} & 73 \\
12\textsuperscript{b} & dGTP & 30 & tBuOH & 97\textsuperscript{g} & 65 \\
13\textsuperscript{b} & d(TGT) & 15/45 & None & 62/81\textsuperscript{f} & — \\
14\textsuperscript{b} & d(TAT) & 120 & None & 35\textsuperscript{f} & — \\
15\textsuperscript{b} & d(ATGC) & 150 & tBuOH & 38\textsuperscript{g} & — \\
16\textsuperscript{b} & d(TTTTGGTTTT) & 30 & DMSO & 82\textsuperscript{g} & — \\
\hline
\end{tabular}
\caption{Reactions of ethyl \(\alpha\)-diazoacetate (EDA) with mono- and oligonucleotides}
\end{table}

\textsuperscript{a} Complete conditions: 5 mM substrate, 1 mM CuSO\textsubscript{4}, 5 mM ascorbate, 50 mM EDA (ethyl diazoacetate), 100 mM MES (2-(N-morpholino)ethanesulfonic acid) pH 6, H\textsubscript{2}O\textsubscript{2}, 25 °C. \textsuperscript{b} Entry was assigned on the basis of tandem MS fragmentation ions. \textsuperscript{c} Six minor mono-alkylation products were detected by ESI, but none in an amount that could be isolated. \textsuperscript{d} No CuSO\textsubscript{4} or ascorbate added. \textsuperscript{e} No CuSO\textsubscript{4} or ascorbate added. \textsuperscript{f} Alkylation site was assigned on the basis of tandem MS fragmentation ions. \textsuperscript{g} Six minor mono-alkylation products were detected by ESI, but none in an amount that could be isolated.

Cosolvent, oligonucleotide size, and diazo substrate all modulate reactivity

While oligonucleotides also reacted selectively at O\textsuperscript{6}-G, the reactions were slower. For example, d(ATGC) was alkylated at O\textsuperscript{6}-G (determined by MS-MS sequencing, see ESI page S18\textsuperscript{t}) but required longer reaction times and conversion was lower than for the mononucleotides (\textit{cf.} Table 1 entries 11 & 15). This reduced reactivity can likely be attributed to the unproductive sequestration of copper to other binding sites (phosphate backbone, other nucleobases) in the nucleic acid. The lower reactivity with longer oligonucleotides was mitigated by using 1,4-dioxane as a cosolvent. For example, changing the stock solution of EDA from 10% tBuOH in water to dioxane nearly doubled the conversion of d(ATGC) (\textit{cf.} Table 1 entries 15 & 16). Nevertheless, these observations point to a limitation of the method: the larger the oligonucleotide the slower it is likely to react.

Diazo substrate variation is another approach to improve the reaction with challenging nucleic acid substrates. Diazoacetamides\textsuperscript{33} (DAAs) led to higher conversions than EDA and high O\textsuperscript{6} selectivity was maintained (see Table 2). For example,
the alkylation of d(ATGC) with DAAs led to conversions of starting material above 60% (see entries 3 & 5, Table 2), whereas the same substrate with EDA led to only 38% conversion (see entry 15, Table 1). While DAAs are better substrates in nucleic acid alkylation, they are also less stable; it was, for example, impossible to prepare a stock solution in 10% tBuOH and water due to rapid decomposition. DMSO as co-solvent stabilizes the DAAs, with the drawback that their reactivity in nucleic acid alkylation is reduced. Changing to a less coordinating co-solvent like dioxane seems to be the best compromise between these competing effects. For example comparing entries 3 & 4 in Table 2 we see that in 20% dioxane higher conversion was obtained in a fifth of the time (62% after 150 minutes versus 74% after 30 minutes). With the more reactive DAAs, even longer DNAs such as 16-mers were alkylated efficiently (see entries 8 & 9, Table 2). The amide substrates offer the additional advantages of stability and modularity. Customization of substrates for pull-downs or with fluorescent tags would simply require the addition of the appropriate amine in the DAA synthesis.

Mechanistic hypothesis
The selectivity observed with copper carbenes does not reflect the intrinsic nucleophilicity of DNA nucleophiles. As an example, modification of guanine at the O6 position of d(ATGC) with a terminal alkyne-bearing diazo substrate (see entry 10, Table 2) followed by a click reaction with azide-conjugated biotin (see the ESI page S31†), installed a complex pull-down tag in only two steps.

Table 2 Reactions of diazo acetamides (DAAs) with oligonucleotides

<table>
<thead>
<tr>
<th>Entry</th>
<th>Substrate</th>
<th>Diazo compound</th>
<th>Time (min)</th>
<th>Conv. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>d(TGT)</td>
<td>N2=N=O</td>
<td>15</td>
<td>78(^a)</td>
</tr>
<tr>
<td>2</td>
<td>d(TGT)</td>
<td>N2=N=O</td>
<td>15</td>
<td>78(^b)</td>
</tr>
<tr>
<td>3</td>
<td>d(ATGC)</td>
<td>N2=N=O</td>
<td>150</td>
<td>62(^a)</td>
</tr>
<tr>
<td>4</td>
<td>d(ATGC)(^c)</td>
<td>N2=N=O</td>
<td>30</td>
<td>74(^a)</td>
</tr>
<tr>
<td>5</td>
<td>d(ATGC)</td>
<td>N2=N=O</td>
<td>150</td>
<td>61(^a)</td>
</tr>
<tr>
<td>6</td>
<td>d(TTTTGT)</td>
<td>N2=N=O</td>
<td>30</td>
<td>92(^a)</td>
</tr>
<tr>
<td>7</td>
<td>d(TTTTGT)</td>
<td>N2=N=O</td>
<td>30</td>
<td>81(^a)</td>
</tr>
<tr>
<td>8</td>
<td>d(AACAGTCATCCTTA)</td>
<td>N2=N=O</td>
<td>180</td>
<td>48</td>
</tr>
<tr>
<td>9</td>
<td>d(AACAGTCATCCTTA)</td>
<td>N2=N=O</td>
<td>180</td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>d(ATGC)(^c)</td>
<td>N2=N=O</td>
<td>150</td>
<td>66(^b)</td>
</tr>
</tbody>
</table>

\(^a\) Complete conditions: 5 mM substrate, 1 mM CuSO\(_4\), 5 mM ascorbate, 50 mM diazo compound, 100 mM MES pH 6, DMSO 20% (v/v), H\(_2\)O, 25 °C.
\(^b\) G-alkylation was assigned on the basis of tandem MS fragmentation ions. \(^c\) Same conditions as \(^a\) but with 20% (v/v) dioxane instead of DMSO.
with powerful alkylating agents, it has been established that N7G is the most potent nucleophile in both dsDNA and ssDNA. O6-G alkylation is found five to tenfold less, while the remaining heteroatoms account for approximately five percent of alkylation products. Since unstabilized Cu(I) carbenes should be highly reactive with water, we expected that the selectivity might stem from pre-association of copper with the guanine nucleobase. Guanine is the most electron rich of the nucleobases and readily chelates with divalent metals, including copper(II). DFT calculations of molecular dipoles in DNA bases indicate that guanine has the largest molecular dipole (6.9 D) and its negative pole is located between O6 and N7. On the other hand, little is known about the interaction of Cu(I) with DNA, likely because the Cu(I) oxidation state has limited stability in aqueous buffer at neutral pH (it tends to oxidize or disproportionate) and is therefore difficult to study. Nevertheless, based on analogy to Cu(II), which readily binds guanine N7s, we hypothesize that the selectivity derives from carbene formation with an N7 bound copper atom. Crystal structure data for Cu(II) and other divalent metals show a strong preference for N7G coordination, with a consequent positioning of the metal approximately 3.5 Å away from O6. The bound carbene would be thus poised to form a metal ylide with the O6 position (see lower mechanistic scheme in Fig. 2), and an irreversible protonation would complete the alkylation.

The selectivity that we observe mirrors the preference in sites of Cu-catalysed oxidative (peroxide or superoxide) damage in DNA, which show increased oxidation at and around guanine residues. The research on copper-mediated redox damage supports two main pathways. In the first an unbound copper(I) ion releases reactive oxygen species upon oxygen reduction or hydrogen peroxide cleavage, the resulting hydroxyl radical diffuses to the nearest site on DNA and typically leads to DNA fragmentation. This mechanism is unselective and presents a constant danger to DNA and RNA – it is, however, inefficient. In the second mechanism a guanine-bound copper(I) forms a copper(II) oxo complex by reacting with peroxide, leading to oxidative damage of the guanine itself as well as neighbouring bases. These reports on oxidative damage, however, provided no insight on why guanine was the target or how the copper bound. We propose that both the specificity of oxidative damage observed in earlier reports and the alkylation specificity we observe here is a result of the same thing: guanine’s superior ability to coordinate copper.

The hypothesis that guanine pre-coordination controls chemoselectivity would predict the following:

1. Other amide or lactone substrates that lack the strong coordination ability of guanine should be poorly reactive.
2. Strongly chelating multidentate ligands should be deleterious to reactivity, since the combination of the guanine and the multidentate ligand would not leave space in copper’s coordination environment for carbene formation.

Over the next two sections we will examine experimental data for each of these scenarios and demonstrate that the hypothesis holds.

1. Normal amide or lactone substrates are poorly reactive.

The most compelling example for the importance of N7G coordination is the 7-deaza-2′-deoxyguanosinetriphosphate (deaza-dGTP), which shows substantially reduced reactivity and selectivity in comparison to normal guanine (see the HPLC traces in Fig. 3). The simple thymidine substrate as well as the cyclic lactam piperidin-2-one are similarly unreactive, adding further weight to the idea that the N7 coordination is the decisive factor in O6 alkylation specificity. Although no literature precedence exists for this Cu(I)-based mechanism, a recent report has shown that histidine residues in proteins can control selectivity in the Cu(II)-catalysed alkenylation reaction of the amide backbone.

2. Multidentate ligands inhibit and N-heterocyclic carbene (NHC) ligands promote guanine alkylation. The copper-catalysed azide alkyne cycloaddition (CuAAC) is a widely used ligation reaction in chemical biology. Ligands such as bathocuproine and...
tristriazoles are often used to stabilize Cu(I) and accelerate the reaction. In our case, however, the addition of strongly coordinating ligands such as tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) or bathocuproinedisulfonic acid disodium salt (BCA) lead to a suppression of guanine alkylation. Even after several hours little or no alkylation product could be detected (see Fig. 4A). In the case of THPTA the copper carbene was still formed (see disappearance of EDA in HPLC trace) but no guanine alkylation was observed. Instead the diazo compound is consumed in unproductive decomposition reactions.

A copper catalyst bearing an N-heterocyclic carbene (NHC) ligand behaved differently: mesitylimidazolinium copper chloride (MesCuCl) gave the same alkylation profile as the standard 'ligandless' conditions, with O6-G alkylation being the dominant product (see Fig. 4B). It is revealing that the NHC ligand, which is not bidentate and hence would still allow coordination with N7G, is the only type tested that maintained alkylation activity. The NHC ligands have other favourable properties for the alkylation reaction (such as reducing oxidative damage) and we are currently exploring the full potential of NHC ligands in DNA and RNA alkylation.

Biochemical studies with O6-(alkyl)-G adducts

With ready access to O6-G derivatives we next used the products to study two open questions in the biochemistry of O6-G adducts: first, their repair by alkylguanine transferases (AGTs) and second, their incorporation during DNA replication.

Human AGT repairs peptide-like carboxamide adducts

AGTs are found in all domains of life and are responsible for repairing O6-G adducts in the genome. Human AGT (hAGT) repairs O6-(alkyl)-G adducts from single stranded and duplex DNAs. Evolution has found no effective catalytic solution to the direct reversal of O6-G damage; instead AGT acts as a suicide protein. It bears a nucleophile cysteine that attacks the alkyl group of the adduct and restores the native guanine. Previous reports conflict as to whether carboxymethyl damage is repaired by hAGT. In addition, since diazotized peptides are the likely source of these alkylation adducts, it is important to understand whether adducts bearing amides are also effectively repaired. Bacteria also produce diazo peptide natural products that may alkylate DNA, adding additional impetus to understand the repair of amide adducts. Thus, we tested whether a DNA 9-mer modified with a cyclohexyl carboxymethylamide (see Table 2, entry 6), was a substrate for hAGT (see Fig. 5). Although removal of the adduct was slow (reduced rates are expected for ssDNAs), analysis of both the deoxyoligonucleotide (see panel B in Fig. 5) and the protein (see panel C in Fig. 5) indicate repair occurred. These results support the most recent claim that carboxymethyl adducts are indeed repaired, and it adds the carboxymethyl amide motif to the list of deoxyoligonucleotides repaired by hAGT.

Fig. 4  (A) Multidentate ligands like THPTA and bathocuproine disulfonic acid disodium salt (BCA) suppress dGMP alkylation. (B) dGMP alkylation catalysed by MesCuCl is highly efficient. HPLC trace was recorded after 30 minutes reaction time.

Fig. 5  (A) De-alkylation of O6-G modified ssDNA 9 mer by hAGT at 37 °C. (B) HPLC analysis of the reaction mixture after 0 (green), 0.5 (purple) and 4 hours (red). (C) High resolution ESI-MS from hAGT after 4 h at 37 °C (the second set of peaks derive from a post-translational modification (likely methylation) of hAGT that we cannot identify, but which has no effect on its function since both masses shift after the dealkylation).
O6-G alkylated triphosphates can be misincorporated during DNA synthesis

To test the impact of O6-alkylation on nucleotide incorporation during DNA synthesis, we characterized O6-(alkyl)-G triphosphates as polymerase substrates. Thus, we synthesized the modified deoxynucleotidetriphosphates O6-(ethyl carbamoylmethyl)-dGTP and (O6-(ethylcm)-dGTP) and O6-(carboxymethyl)-dGTP (O6-(cm)-dGTP) with our Cu(I) carbene methodology. Selective carbamoylmethylation of commercial dGTP at the O6 position with EDA as carbene precursor directly delivers O6-(ethylcm)-dGTP and O6-(cm)-dGTP; hence both molecules are obtained in a total of two chemical operations (see Fig. 6A). The direct modification approach offers unprecedented efficiency in the formation of O6-G adducts (in this case from a commercial triphosphate without any protecting group or phosphorylation steps). Previous syntheses of O6-G modified nucleosides and triphosphates involved four43 to seven55–58 steps including triphosphate introduction using strong electrophilic phosphorous reagents.

To investigate the incorporation of O6-(cm)-dGTP and O6-(ethylcm)-dGTP during DNA synthesis, we performed single nucleotide primer extension experiments with KTqM747K, a mutant of KlenTaq polymerase that bypasses several DNA polymerase stalls during DNA synthesis, we characterized O6-(alkyl)-G triphosphates can be misincorporated during DNA synthesis, we performed single

Conclusions

We have discovered a new way to create O6-G adducts post-synthetically in nucleic acids of varying size, from nucleotide mono- or triphosphates to long oligonucleotides. The ability to obtain high yields of precisely defined O6-G adducts eliminates a major roadblock in studying their biochemistry, a point we have illustrated through the first synthesis of O6-(cm)-dGTP and the first demonstration of its incorporation during DNA replication.

Copper-catalysed O6-G alkylation is a dump-and-stir procedure that employs readily available materials. α-Diazo ester substrates are stable enough that several are sold commercially and a simple54 two-step (one-pot) protocol for α-diazo ester synthesis makes nearly any diazo compound in this class synthetically accessible. Research on O6-G adducts in biology67 and biotechnology68 demands ready synthetic access to these adducts and the method we have described is the simplest to date.

Nucleic acids are found in diverse contexts and can fold in myriad ways, complicating attempts at broad generalizations of reactivity. Nevertheless, the weight of evidence suggests N7G is the most powerful nucleophile34 and ligand67 in DNA and RNA. It is a primary target of organic electrophiles,34 as well as organometallic anticancer drugs.69 To the list of N7G’s functions we add the role of directing group in copper catalysed reactions. We have focussed on alkylation reactions, but the same effect is likely responsible for the well-known propensity of Fenton-type oxidative damage to be localized around guanines.40

Acknowledgements

D.G.G and S.N.G acknowledges the Swiss National Science Foundation (SNF) for financial support (grant number 200021_149805). L.A.W and S.J.S acknowledge the European Research Council (260341), the SNF (156280), and the ETH...
research commission (ETH-43 14-1) for financial support. Prof. Nathan Leudtke and Mr Aaron Johnson are gratefully acknowledged for helpful discussions.

Notes and references

21 We use the convention that exocyclic heteroatoms are labelled with a superscript and endocyclic ones with normal font: E. C. Friedberg, G. C. Walker, W. Siede and R. D. Wood, DNA repair and mutagenesis, American Society for Microbiology Press, 2005.