ARTICLE

Check for updates

Programmable site-selective labeling of oligonucleotides based on carbene catalysis

Yang-Ha Lee[1], Eunsoo Yu [1] & Cheol-Min Park [1✉]

Site-selective modification of oligonucleotides serves as an indispensable tool in many fields of research including research of fundamental biological processes, biotechnology, and nanotechnology. Here we report chemo- and regioselective modification of oligonucleotides based on rhodium(I)-carbene catalysis in a programmable fashion. Extensive screening identifies a rhodium(I)-catalyst that displays robust chemoselectivity toward base-unpaired guanosines in single and double-strand oligonucleotides with structurally complex secondary structures. Moreover, high regioselectivity among multiple guanosines in a substrate is achieved by introducing guanosine-bulge loops in a duplex. This approach allows the introduction of multiple unique functional handles in an iterative fashion, the utility of which is exemplified in DNA-protein cross-linking in cell lysates.

[1] Department of Chemistry, UNIST (Ulsan National Institute of Science & Technology), Ulsan, Korea. ✉email: cmpark@unist.ac.kr

Sequence-specific DNA-binding proteins such as transcription factors (TFs), polymerases, and DNA modification enzymes play a crucial role in regulating gene expression and maintaining various cellular functions. Therefore, the identification of these proteins and elucidation of their functions have drawn a great deal of attention[1–4]. The transient interactions between DNA and the corresponding binding proteins led to the development of various covalent trapping methods including diazirine-derived carbenes[5] and UV-based cross-linking[6,7]. Also, successful trapping of DNA-binding proteins has been demonstrated by employing self-ligating ligands (HaloTag, SNAP, CLIP) fused to the proteins of interest and DNA probes linked to the corresponding tags[8]. These indispensable methods for the study of DNA-binding proteins require chemical modification of natural nucleic acids to introduce functional handles, often with more than one functionalities for sophisticated manipulations including cross-linking and detection.

The current standard method for the site-selective introduction of orthogonal functionalities is the prior synthesis of pre-functionalized nucleoside phosphoramidite monomers via laborious multistep synthesis, followed by solid-phase oligonucleotide (ON) synthesis[9–12]. More importantly, the ability to introduce multiple non-redundant functional groups is limited by the availability of chemically compatible orthogonal protecting groups that can be removed selectively while preserving the labile glycosidic bonds in oligonucleotides intact.

Alternatively, post-synthetic modification of ONs may provide more efficient access to site-selective functionalization. While elegant methods based on enzyme-mediated and chemical-based modifications have been reported to date[13–15], significant challenges still remain. For example, those tailored for the modification of terminal sites include oxidative cleavage of ribose 3′-end[16,17], 3′-phosphorylation by deoxyribozyme-mediated phosphate transfer[18], 5′-phosphoramidation of 5′-phosphate via phosphorimidazolide[19], and 5′-phosphate alkylation by diazo compounds[20].

Meanwhile, internal sites-specific alkylation of bases has been developed. The SELEX (systematic evolution of ligands by exponential enrichment)-based modifications include the ribozyme-catalyzed protocol[21] and self-alkylation of RNAs by employing high affinity electrophilic ligands[22–24]. Also, proximity-driven modifications by hybridizing opposite strands bearing reactive groups have been reported[25–27]. Enzymes such as methyltransferase and tRNA-guanine transglycosylase have been employed for selective modification[28,29], which necessitates specific sequences for enzyme recognition and laborious engineering of enzymes/cofactors. Also, 2′-OH modification of RNAs has been shown by using acyl imidazole reagents[30,31] and SELEX-based ribozymes[32].

The unique reactivity of metal carbene complexes has led to the development of diverse synthetic applications including bioconjugation[33–35]. Their use for the modification of proteins has been reported by several research groups[36–39]. Recently, in their pioneering work, Gillingham and co-workers reported efficient post-synthetic modification of ONs based on Rh(II)-carbenes, which was, however, lacking base selectivity. Later, the development of Cu(I)-carbenes allowed the chemoselectivity for guanosine (G) bases, in which the limited substrate scope of sterically unhindered single strands remained to be addressed[40–42]. While the chelation-driven selectivity of Cu(I)-catalysts works smoothly on simple substrates, complex substrates/matrices bearing various multiple chelation sites pose significant challenges to this type of catalysts regarding catalyst turnover (See Supplementary Fig. 1 for the comparison of the previous and this works).

In our efforts to develop a catalytic system that allows site-selective modification among multiple Gs embedded in structurally complex oligonucleotides, we aim to address the two fundamental selectivities toward substrates; chemo- and regio-selectivity. In this regard, we reason that the discovery of a chelation-free catalyst is crucial to overcome the issue of catalyst turnover in such a highly chelating environment, in which the chemoselectivity among the different nucleotide bases needs to be driven by the intrinsic reactivity of a catalyst, instead of a chelation effect. Secondly, we envisage that the regioselectivity among multiple Gs could be realized by the introduction of G-bulge loops in duplexes, in which the difference in reactivity between base-paired and unpaired Gs in a duplex may provide the regioselectivity. The prerequisite for this approach is the identification of a catalyst, which displays robust catalytic activity toward G-bulges located in structurally complex duplexes. The successful development of this approach would allow a programmable synthesis of ONs with multiple unique functional groups in an iterative fashion by simple replacement of the corresponding opposite strands (Fig. 1a).

Here, we describe the post-synthetic modification of oligonucleotides at unpaired Gs in a chemo- and regio-selective manner based on rhodium(I)-carbene catalysis. The ability of the protocol for the selective modification of G-bulges allows the introduction of multiple unique functional handles in an iterative fashion, the utility of which is exemplified in DNA-protein cross-linking in cell lysates.

Results and discussion

Optimization and substrate scope. To identify an optimal catalyst for the selective modification of nucleobases, we commenced with screening of various metal catalysts (5 mol%)[43–45] in the presence of diazoacetone and a mixture of four nucleosides as a model substrate for base-selectivity in 50% THF-H_2O (Fig. 2, Entries 1 and 2; Supplementary Table 2, Entries 5–9), however, all of which turned out to be ineffective. Gratifyingly, rapid conversion was observed with exclusive formation of a product within 30 min when [Rh(COD)Cl]$_2$ was employed as a catalyst (45%, Fig. 2, Entry 3). The structure was unambiguously assigned as O^6-acetonyl deoxyguanosine (dG) **6** based on heteronuclear multiple bond correlation (HMBC) spectroscopic analysis (correlation between C6–H9, Fig. 3a). Since the secondary structures of ONs may be affected by the presence of an organic solvent, we examined the reaction in a reduced fraction of THF (10% THF-H_2O), and found that it afforded a comparable yield compared to that in 50% THF-H_2O (96% vs. 93%, respectively, Supplementary Table 2, Entry 1 and 2). We also confirmed by performing a reaction with **7b** in pure aqueous buffer that 10% THF does not affect the reaction (Supplementary Fig. 5). The effect of pH revealed a higher conversion at pH = 6.0 than pH = 7.4 (Supplementary Table 2, Entry 3 and 4). Eventually, the reaction conditions were optimized as 10 mol% catalyst loading in 10% THF-H_2O at pH = 6.0, MES buffer (98%; Fig. 2, Entry 4).

Encouraged by the excellent catalytic activity of [Rh(COD)Cl]$_2$, an extensive screening with various Rh(I) salts was performed, however, other Rh(I) catalysts turned out to be less effective (Fig. 2, Entries 5–8). Likewise, an examination of various diazo compounds revealed that diazoacetone possesses unique reactivity that affords excellent yield (Fig. 2, Entries 4 vs. 9; see Supplementary Table 2 for a complete list of diazo compounds and buffer conditions). Deoxyinosine (dI), a structural analog of dG, was also examined for its reactivity relative to dG and was found to be slightly less reactive (dG/dI = 1.3: 1, Supplementary Fig. 2).

The optimized conditions feature an optimal balance between the reactivity/selectivity toward Gs and stability in aqueous buffer media. Also, the advantages of diazoacetone as a carbene precursor include the keto group serving as an orthogonal
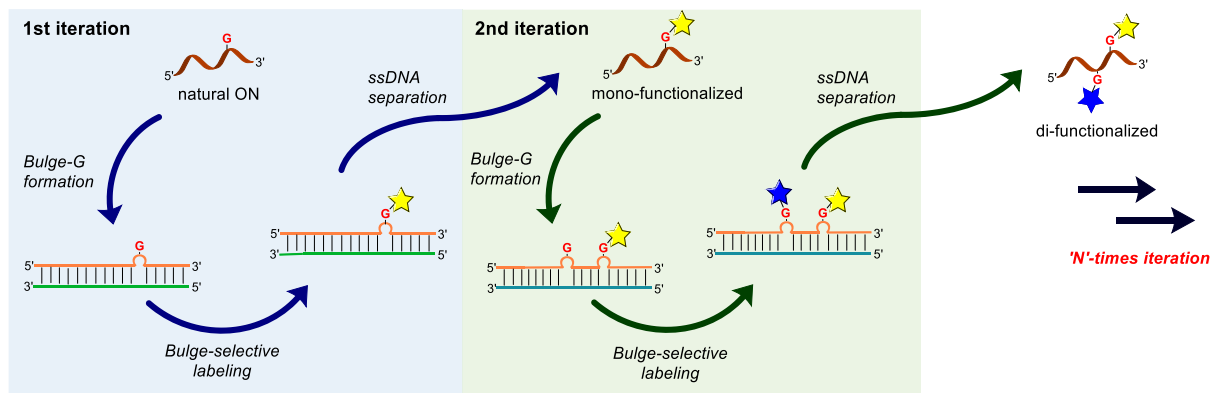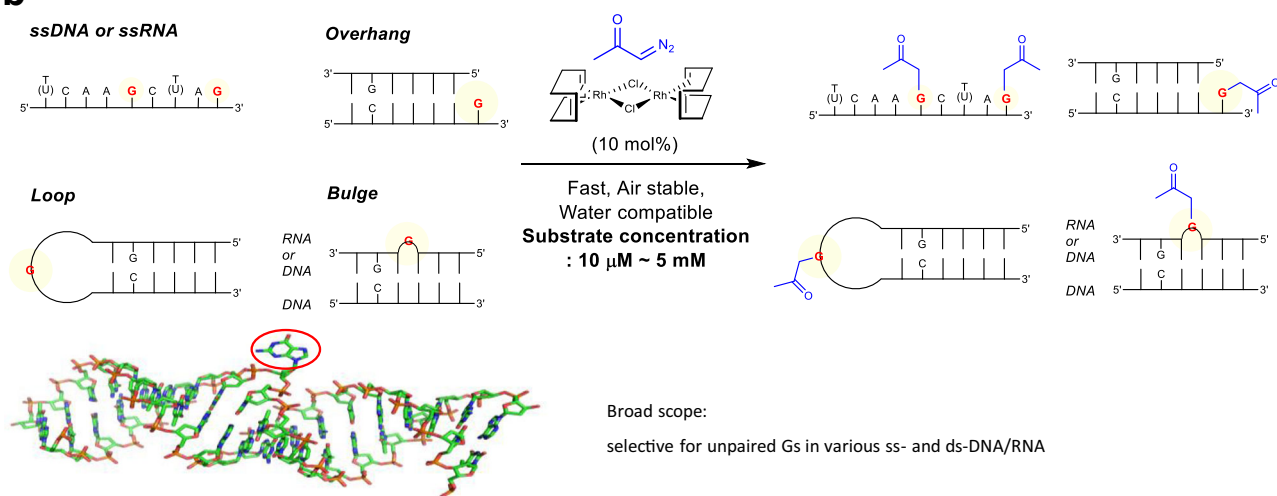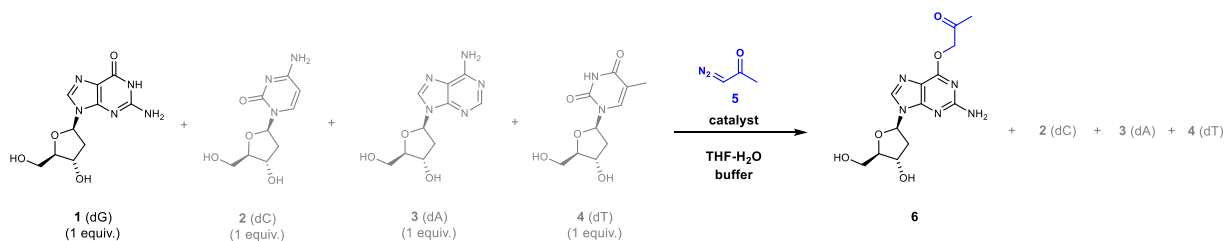
Fig. 1 Site-selective modification of oligonucleotides. a Iterative functionalization of oligonucleotides. **b** Compatibility with various secondary structures in a wide range of concentrations; 3D model of bulge-guanosine (G) dsDNA.

Entry	Catalyst	Time	Yield[a]	Entry	Catalyst	Time	Yield[a]
1[b]	Rh$_2$(OAc)$_4$	24 h	N/R	6[c]	[Rh(COD)(MeCN)$_2$]BF$_4$	o/n	76%
2[b]	Rh$_2$(TFA)$_4$	24 h	N/R	7[c]	[Rh(COD)$_2$]OTf	o/n	48%
3[b]	[Rh(COD)Cl]$_2$	30 min	45%	8[c]	Rh(CO)$_2$(acac)	o/n	N/R
4[c]	[Rh(COD)Cl]$_2$	10 min	98%	9[d]	[Rh(COD)Cl]$_2$	3 h	N/R
5[c]	Rh(COD)(acac)	60 min	83%				

Fig. 2 Optimization of Rh(I)-catalyzed O^6-G labeling. [a]Based on HPLC analysis. [b]Conditions a: each nucleoside (5 mM), catalyst (5 mol%), diazoacetone (1.2 equiv.), 50% THF-H$_2$O, r.t. [c]Conditions b: each nucleoside (5 mM), catalyst (10 mol%), diazoacetone (8 equiv.), 10% THF-H$_2$O, MES (20 mM), pH 6.0, r.t. [d]Used ethyl diazoacetate instead of diazoacetone (deviation from Conditions a). N/R = no reaction. o/n = overnight.
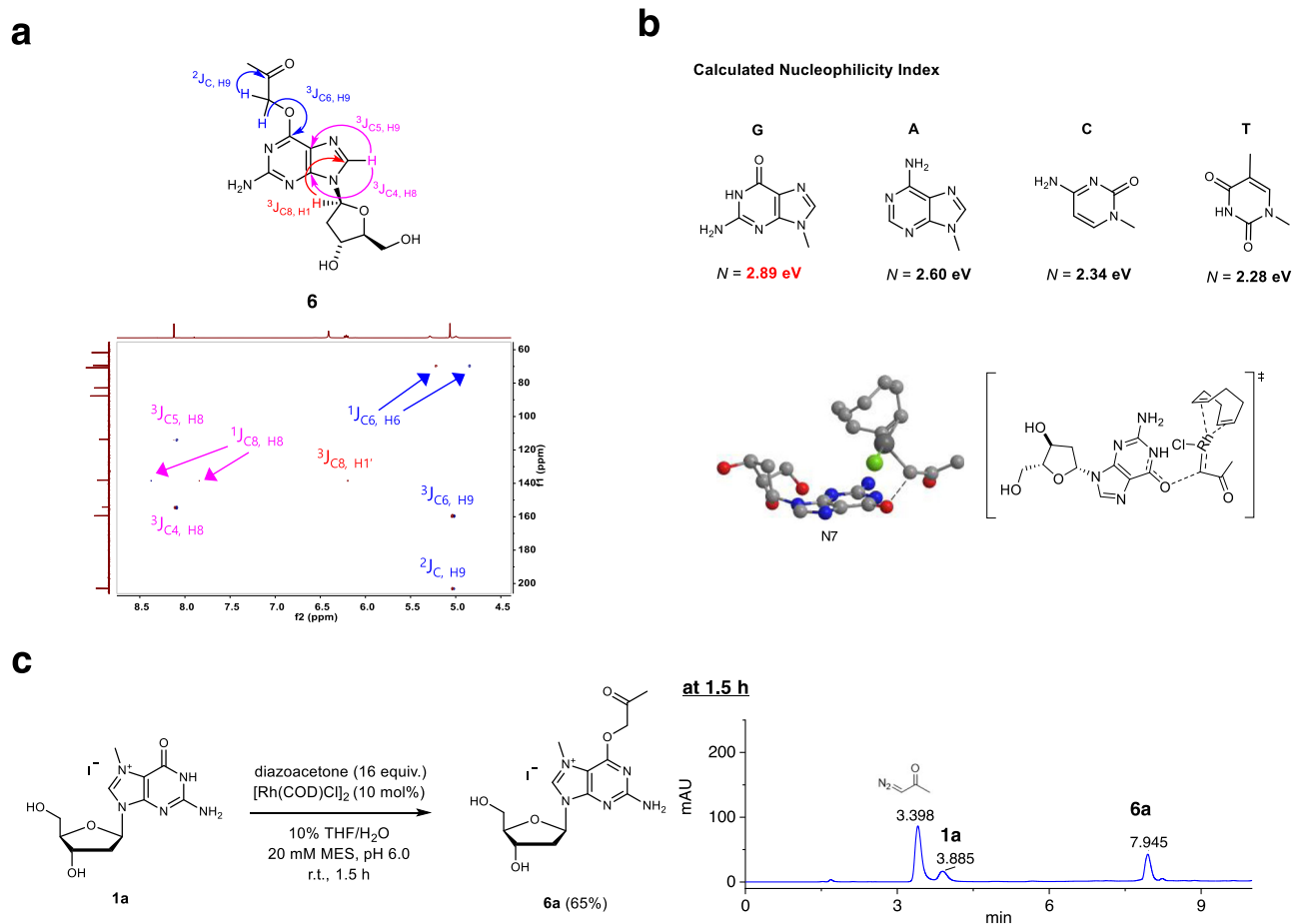
Fig. 3 Structure and mechanistic analysis of Rh(I)-catalyzed O^6-G labeling. a Expanded ^1H-^13C HMBC (DMSO-d_6) spectrum of **6**. **b** Calculated nucleophilicity index of four canonical nucleobases (A, T, G, and C); transition state geometry of the catalyst approaching guanosine obtained by DFT calculations. **c** Reaction performed with N^7-methyl-2′-deoxyguanosine and its HPLC trace.

functional handle, minimal steric disruption owing to its small size crucial for the studies involving protein-nucleic acid binding interactions, and the high aqueous solubility of diazoacetone. In addition, the by-product hydroxyacetone, which forms inevitably owing to the reaction of carbenes with water, could be readily removed by simple evaporation to facilitate the purification step.

To gain insight on the selectivity toward guanosine among the nucleosides, we performed density function theory (DFT) calculations. Based on the nucleophilicity N index, which was recently developed to describe the behavior of various nucleophiles[46,47], we calculated them for the four bases (Fig. 3b). It turned out that guanine showed the highest value on the nucleophilic scale, which is in good correlation with the observed selectivity. Also, the complete energy profile of the reaction pathway was calculated (Supplementary Fig. 62). Overall, the reaction is highly exergonic with −30.7 kcal/mol. The highest activation barrier turned out to be the initial rhodium carbene formation with 11.6 kcal/mol followed by a barrier of 6.4 kcal/mol for the O6-C9 bond formation. The transition state geometry displays the chelation-free approach of the carbene complex toward the guanosine O^6 supporting the unattenuated reactivity in the presence of numerous chelating functionalities. To further support the chelation-free model, we performed a reaction with N^7-methyl-2′-deoxyguanosine, which is unable to make chelation, and found that it provided the corresponding O^6-acetonyl product **6a** in 65% yield (Fig. 3c; Supplementary Fig. 22 for MS analysis; SI p.104 for NMR analysis after sugar was hydrolytically removed).

With the optimized conditions at the monomer level, we examined the selectivity of the carbene complex toward Gs embedded in ONs. Mg(OAc)_2 was included for subsequent applications to duplexes in mind, which stabilizes them by shielding the phosphate backbone charges[48]. The G-selectivity was confirmed by tandem mass spectrometry by employing tetramer **7a** as a model system (Fig. 4a, b).

Since the reaction site, guanosine O^6, is engaged in the base pairing, we reasoned that it might provide us another level of selectivity between paired and unpaired Gs. It turned out that exclusive modification at the unpaired G was achieved while those with base pairing remaining intact when hairpin **7b** containing multiple Gs in the stem and overhang was subjected to the reaction. The site of modification was confirmed by digestion with endonuclease EcoRI followed by identification of the acetonylated fragment by LC–MS analysis (see Fig. 4c, d). In addition, site-selectivity of the modification was further confirmed by a primer extension assay with **8h** (Supplementary Fig. 6).

Based on these results, we investigated the substrate scope with hairpins, duplexes, and hybrid duplexes (Fig. 5 and Supplementary Table 4). Beforehand, a brief examination of single-stranded ONs confirmed the selective modification of Gs with good efficiency (Supplementary Table 3). Next, we expanded our investigation on various duplex ONs (Fig. 5). The influence of base pairing on the reactivity of the catalytic system in several hairpin and duplex ONs bearing Gs at various positions including overhang, loop, base-paired, and mismatched was examined.
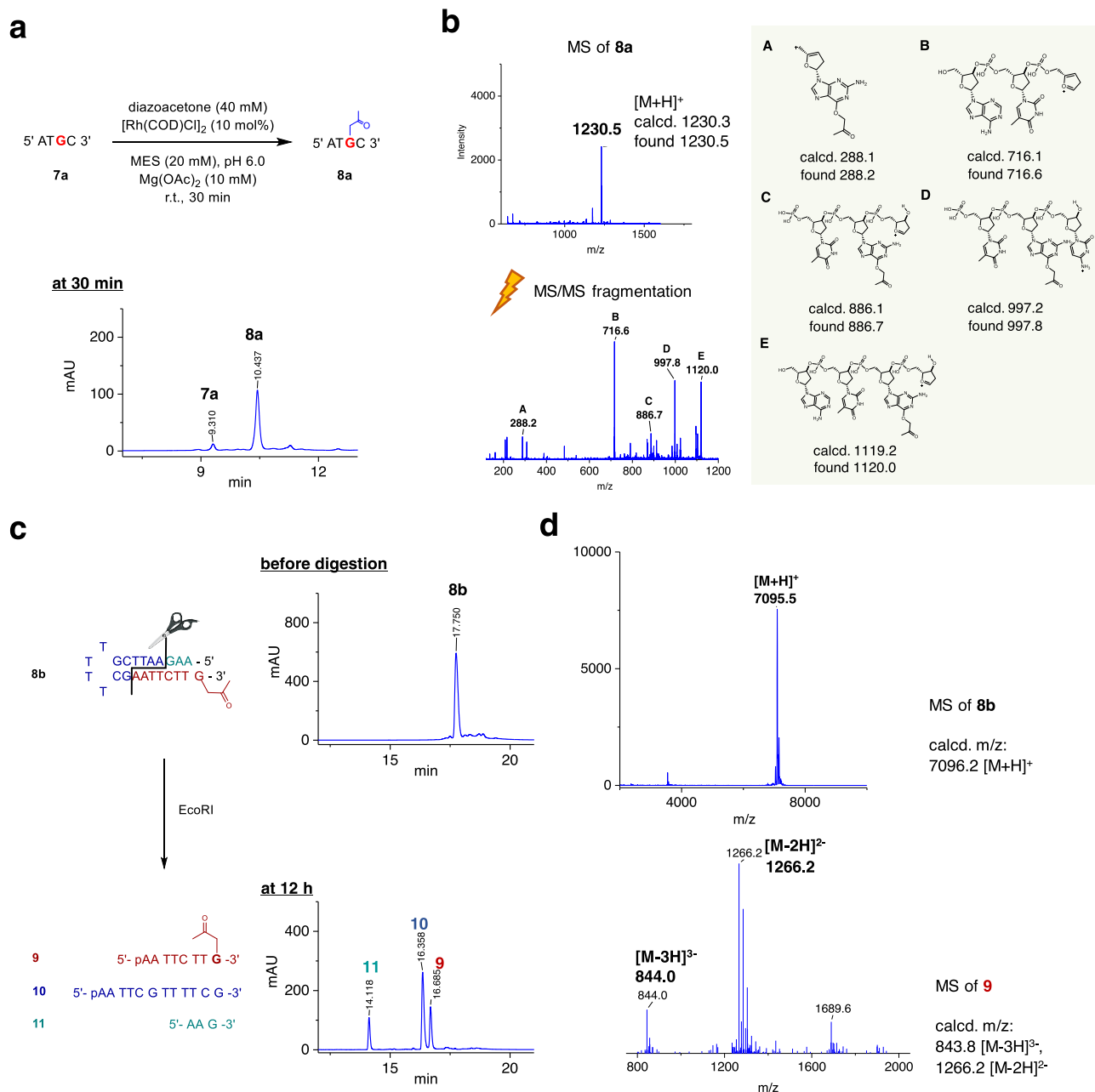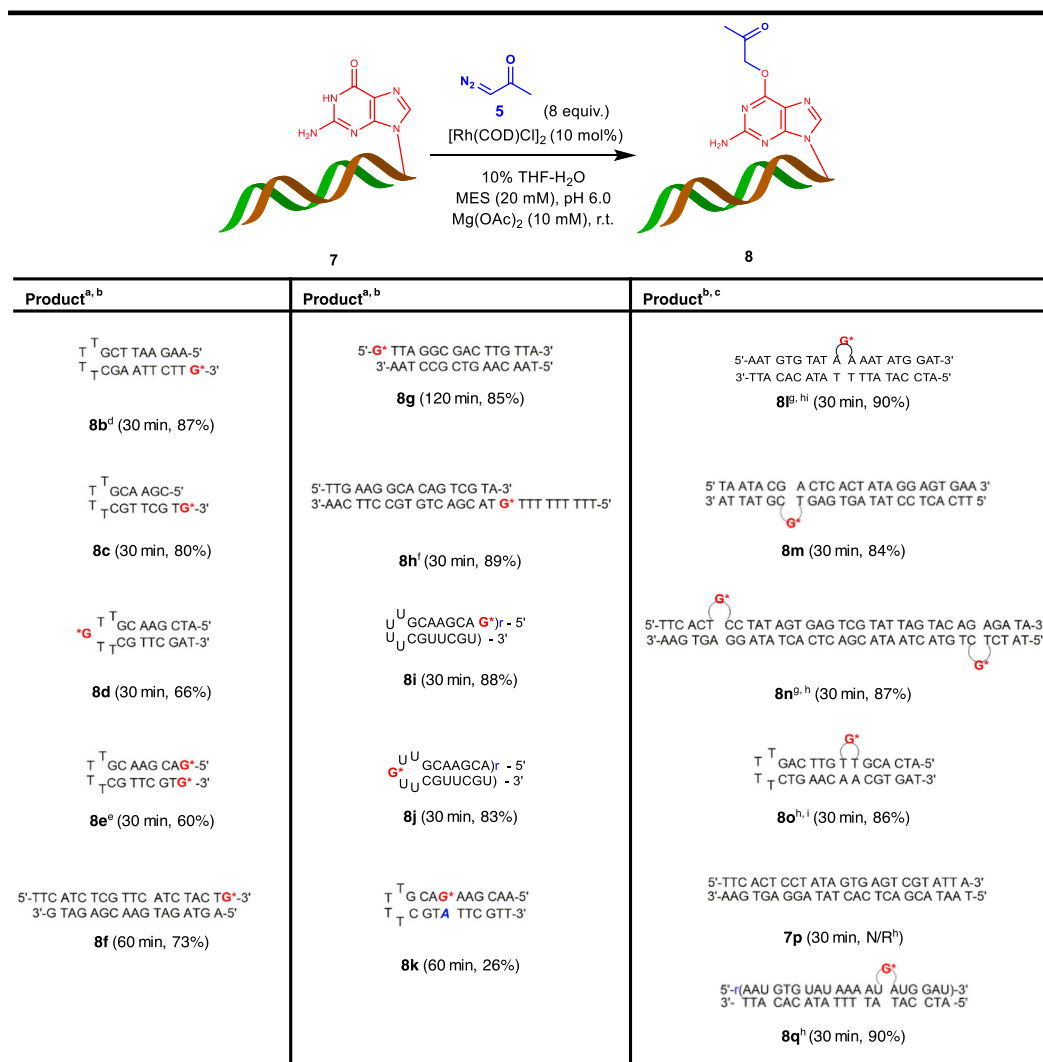
Fig. 4 Structure analysis of Rh(I)-catalyzed O⁶-G labeling of oligonucleotides. a HPLC analysis of the acetonylation of **7a**. **b** Identification of G modification by MALDI-TOF/TOF analysis of **8a**. Mass spectrum (MS) of **8a** and its MS/MS fragmentation with labeled fragments. **c** Endonuclease digestion of acetonylated hairpin oligonucleotide **8b** and the corresponding HPLC spectra for the analysis of the modification site. **d** MS of **8b** and **9**.

Overall, the reaction turned out to be highly selective toward unpaired Gs regardless of hairpins or duplexes **7b–7j** (Fig. 5). Hairpin **7e** bearing two Gs at both 5′- and 3′-ends reacted to give the corresponding dual labeling product in 60% yield. Efficient conversions were observed with an overhang G located at either 3′- or 5′-end of dsODN **7f** and **7g**. In addition, dsODN (double-stranded oligodeoxynucleotide) **7h** containing a G at an internal position of a long overhang participated well to give an excellent yield (89%) showing highly general reactivity. Moreover, RNA hairpin **7i** and **7j**, which have unpaired G at the overhang and loop positions, gave excellent yields (88 and 83%, respectively).

Although the metal-carbene complex displays a high base-selectivity toward Gs, its lack of site-selectivity on ONs bearing multiple unpaired G residues poses severe limitation. To address the problem, we initially envisaged that the introduction of

mismatched Gs would provide selectivity over those engaged in base-pairing. It turned out that the reaction on hairpin **7k** containing a mismatch at an internal position was sluggish (26%, Fig. 5). Based on the speculation that the stack-in conformation of mismatched Gs may interfere the catalyst accessibility, we reasoned that the introduction of Gs as a bulge loop may provide a better exposure toward the catalyst. We were gratified to find that the reaction proceeded with an exclusive modification at the G-bulge in excellent yield in 30 min when tested on the 20-mer and 25-mer bulge duplexes **7l** and **7m**. Dual labeling of duplex **7n**, in which both strands contain G-bulges, proceeded efficiently. Also, a comparable yield was obtained with hairpin **7o**. A control experiment with dsODN **7p** lacking a G-bulge showed no reaction, which further confirms the selectivity. We also investigated oligoribonucleotides (ORNs) as substrates. G-bulge

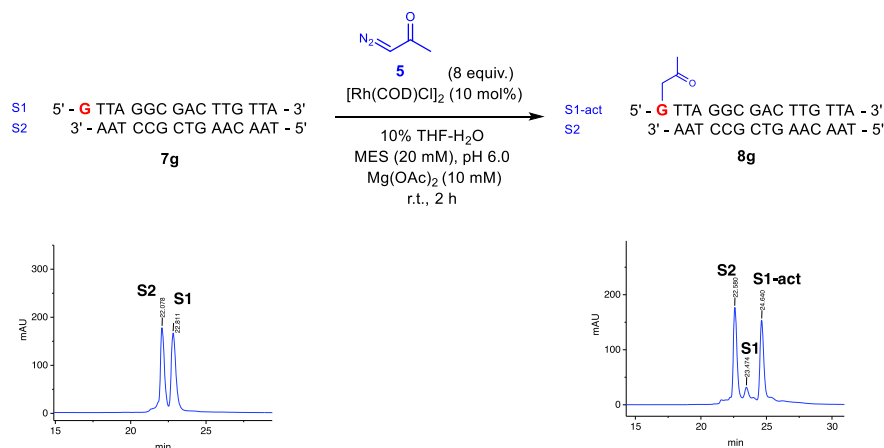| Product[a, b]	Product[a, b]	Product[b, c]
ᵀ ᵀGCT TAA GAA-5' ᵀ ᵀCGA ATT CTT **G***-3' **8b**[d] (30 min, 87%)	5'-**G*** TTA GGC GAC TTG TTA-3' 3'-AAT CCG CTG AAC AAT-5' **8g** (120 min, 85%)	5'-AAT GTG TAT A A AAT ATG GAT-3' 3'-TTA CAC ATA T T TTA TAC CTA-5' **8l**[g, hi] (30 min, 90%)
ᵀ ᵀGCA AGC-5' ᵀ ᵀCGT TCG **TG***-3' **8c** (30 min, 80%)	5'-TTG AAG GCA CAG TCG TA-3' 3'-AAC TTC CGT GTC AGC AT **G*** TTT TTT TTT-5' **8h**[f] (30 min, 89%)	5' TA ATA CG A CTC ACT ATA GG AGT GAA 3' 3' AT TAT GC T GAG TGA TAT CC TCA CTT 5' G* **8m** (30 min, 84%)
*G ᵀ ᵀGC AAG CTA-5' ᵀ ᵀCG TTC GAT-3' **8d** (30 min, 66%)	ᵁ ᵁGCAAGCA **G***)r - 5' ᵁ ᵁCGUUCGU) - 3' **8i** (30 min, 88%)	5'-TTC ACT CC TAT AGT GAG TCG TAT TAG TAC AG AGA TA-3' 3'-AAG TGA GG ATA TCA CTC AGC ATA ATC ATG TC TCT AT-5' G* **8n**[g, h] (30 min, 87%)
ᵀ ᵀGC AAG CA**G***-5' ᵀ ᵀCG TTC G**TG*** -3' **8e**[e] (30 min, 60%)	**G*** ᵁ ᵁGCAAGCA)r - 5' ᵁ ᵁCGUUCGU) - 3' **8j** (30 min, 83%)	ᵀ ᵀGAC TTG T T GCA CTA-5' ᵀ ᵀCTG AAC A A CGT GAT-3' G* **8o**[h, i] (30 min, 86%)
5'-TTC ATC TCG TTC ATC TAC T**G***-3' 3'-G TAG AGC AAG TAG ATG A-5' **8f** (60 min, 73%)	ᵀ ᵀG CA**G*** AAG CAA-5' ᵀ ᵀC GT**A** TTC GTT-3' **8k** (60 min, 26%)	5'-TTC ACT CCT ATA GTG AGT CGT ATT A-3' 3'-AAG TGA GGA TAT CAC TCA GCA TAA T-5' **7p** (30 min, N/R[h]) 5'-r(AAU GTG UAU AAA AU AUG GAU)-3' 3'- TTA CAC ATA TTT TA TAC CTA -5' G* **8q**[h] (30 min, 90%)

Example of **8g**

Fig. 5 Scope for hairpins and duplexes. All oligonucleotide (ON)s are oligodeoxynucleotides (ODN)s unless noted as r(sequence). All the reactions were analyzed by HPLC, as exemplified at the bottom of the figure. In **8k**, mismatched A was marked in blue. [a]Conditions a: ON (5 mM), diazoacetone (8 equiv.), [Rh(COD)Cl]₂ (10 mol%), 10% THF-H₂O, MES (20 mM), pH 6.0, Mg(OAc)₂ (10 mM), r.t. [b]Yield was calculated by HPLC analysis. [c]Conditions b: ON (5 mM), diazoacetone (100 mM), [Rh(COD)Cl]₂ (10 mol%), 10% THF-H₂O, MES (50 mM), pH 6.0, Mg(OAc)₂ (50 mM), r.t. [d]Modification site was determined by endonuclease digestion (See Fig. 4c, d). [e]16 equiv. of diazoacetone, 15 mol% [Rh(COD)Cl]₂ (deviation from Conditions a). [f]16 equiv. of diazoacetone (deviation from Conditions b). [g]40 equiv. of diazoacetone (deviation from Conditions b). [h]Yield was calculated after oxime ether conjugation with excess amount of benzyloxyamine. [i]12 equiv. of diazoacetone (deviation from Conditions b). N/R = no reaction.
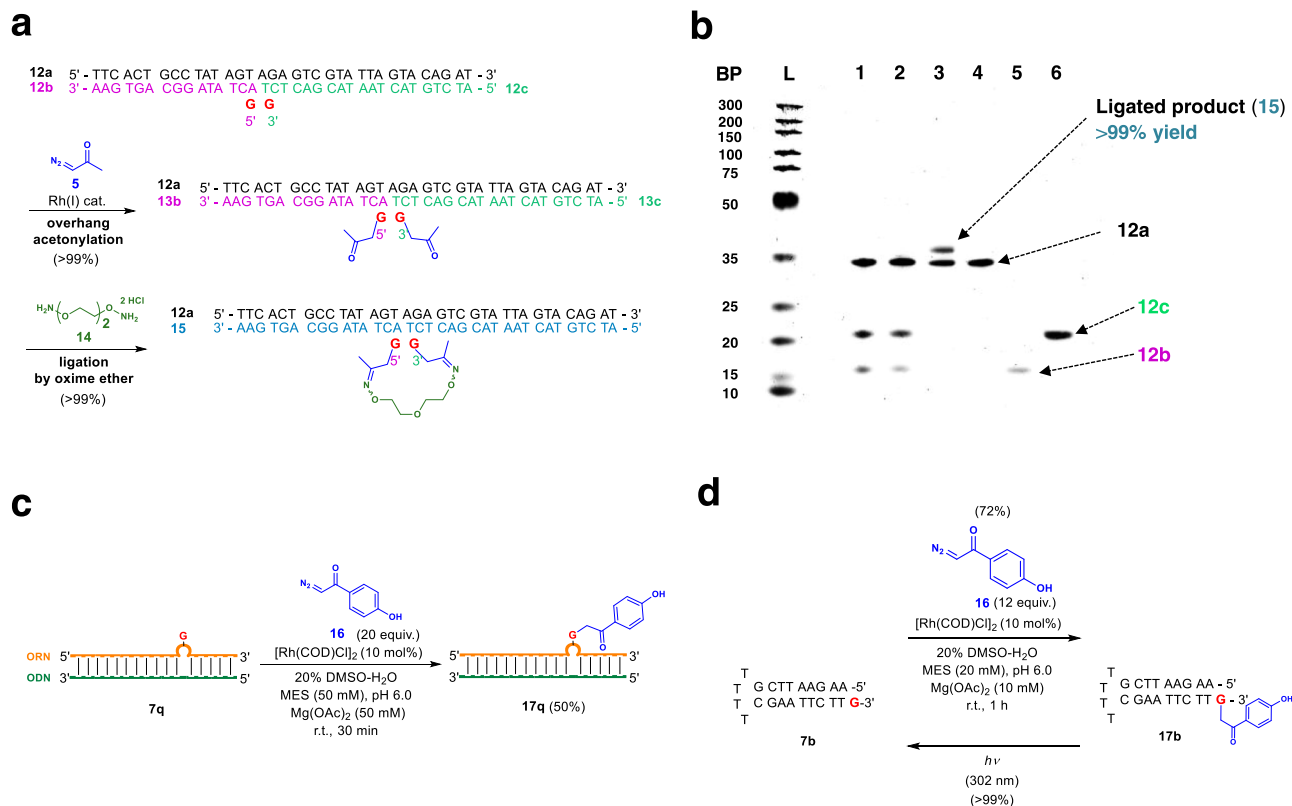
Fig. 6 Chemical ligation of oligodeoxynucleotides and single-step synthesis of photocaged oligonucleotides. a Synthetic scheme for the ligation of oligonucleotides with an oxime ether linker. **b** Denaturing PAGE analysis of the ligation visualized by SYBR gold. Lanes; L: DNA ladder, 1: before acetonylation, 2: after acetonylation, 3: after ligation, 4: **12a**, 5: **12b**, 6: **12c**. **c** Synthesis of photocaged single-stranded oligoribonucleotide based on the G-bulge protocol. **d** Caging and decaging of **17b**. Source data are provided as a Source Data file.

ORN **7q** was prepared by ORN-ODN hybrid duplex formation and was subjected to the standard conditions. Gratifyingly, the reaction proceeded to give **8q** in 90% yield.

Chemical ligation of ODNs and single-step synthesis of photocaged ONs. Chemical ligation of ODNs with unnatural linkers allows the preparation of the corresponding elongated ODNs, on which diverse functional groups could be attached. To this end, we prepared a duplex comprised of the template sequence **12a** paired with two ODN fragments **12b** and **12c** bearing a G-overhang on each fragment (Fig. 6a). The reaction with diazoacetone selectively proceeded at the two G-overhang residues to give **13b** and **13c** in quantitative yields, each of which was functionalized with an acetonylated-G (act-G) (see SI p. 33 and Supplementary Fig. 53 for details). Ligation was performed on the act-G duplex ODN with dialkoxyamine linker **14**. The ligation proceeded in quantitative yield confirmed by LC–MS and denaturing PAGE analyses (Fig. 6b; see Supplementary Fig. 54); the low mobility band slightly above the template sequence **12a** corresponds to the ligated product **15**.

Photocleavable protecting groups serve as a powerful tool in many biological studies[49–52]. The typical synthesis of photocaged ONs begins with the preparation of phosphoramidite monomers functionalized with a photocleavable group, which usually involves lengthy syntheses. For example, Woodson and coworkers[9] reported the temporally controlled Hfq-catalyzed annealing study by using photocaged ORNs. Thus, to examine the applicability of our method, G-bulge ORN **7q** prepared by ORN-ODN hybrid duplex formation was reacted with α-diazo-*p*-hydroxyacetophenone **16** (20 equiv.) in the presence of 10 mol% [Rh(COD)Cl]₂ (20 % DMSO-H₂O, MES, pH = 6.0). We found

that the reaction smoothly proceeded to afford **17q** in 50% yield (Fig. 6c, Supplementary Fig. 55). Moreover, decaging of **17b** was performed to regenerate **7b** in quantitative yield (Fig. 6d, Supplementary Fig. 59). These results demonstrate the efficiency of the single-step synthesis of photocaged ORNs and ODNs.

DNA-protein cross-linking (DPC) with a DNA sequence-specific binding protein. A programmable method enabling site-selective introduction of multiple unique functional handles in an iterative fashion on various ONs is highly desirable for the identification of numerous nucleic acid-binding proteins. As such, we sought to apply our protocol for the covalent trapping of sequence-specific DNA-binding proteins with the corresponding modified dsODNs (Fig. 7a). To examine the feasibility, we chose T7 RNA polymerase (RNAP), which possesses specificity for the T7 promoter sequence and took a two-step approach; (a) we first confirmed the cross-linking in the absence of cell lysate with a dsON bearing single acetonyl group for covalent trapping via reductive amination with an adjacent lysine residue in the binding site of RNAP. (b) Subsequently, we performed the cross-linking in the presence of cell lysate with a doubly modified dsON, in which one acetonyl group was used for covalent trapping and the other for the introduction of a fluorescent probe.

To assess the impact of the location of modification, we designed two dsDNA probes bearing a G-bulge on the antisense strands located either in the middle of or several nucleotides away from the T7 promoter sequence (Fig. 7b). The act-G-bulge dsDNAs were prepared by our standard protocol. Reductive amination was performed by incubation of T7 RNAP with the act-G-bulge dsDNAs in the presence of NaCNBH₃, and the samples were analyzed on SDS–PAGE gel visualized by silver and
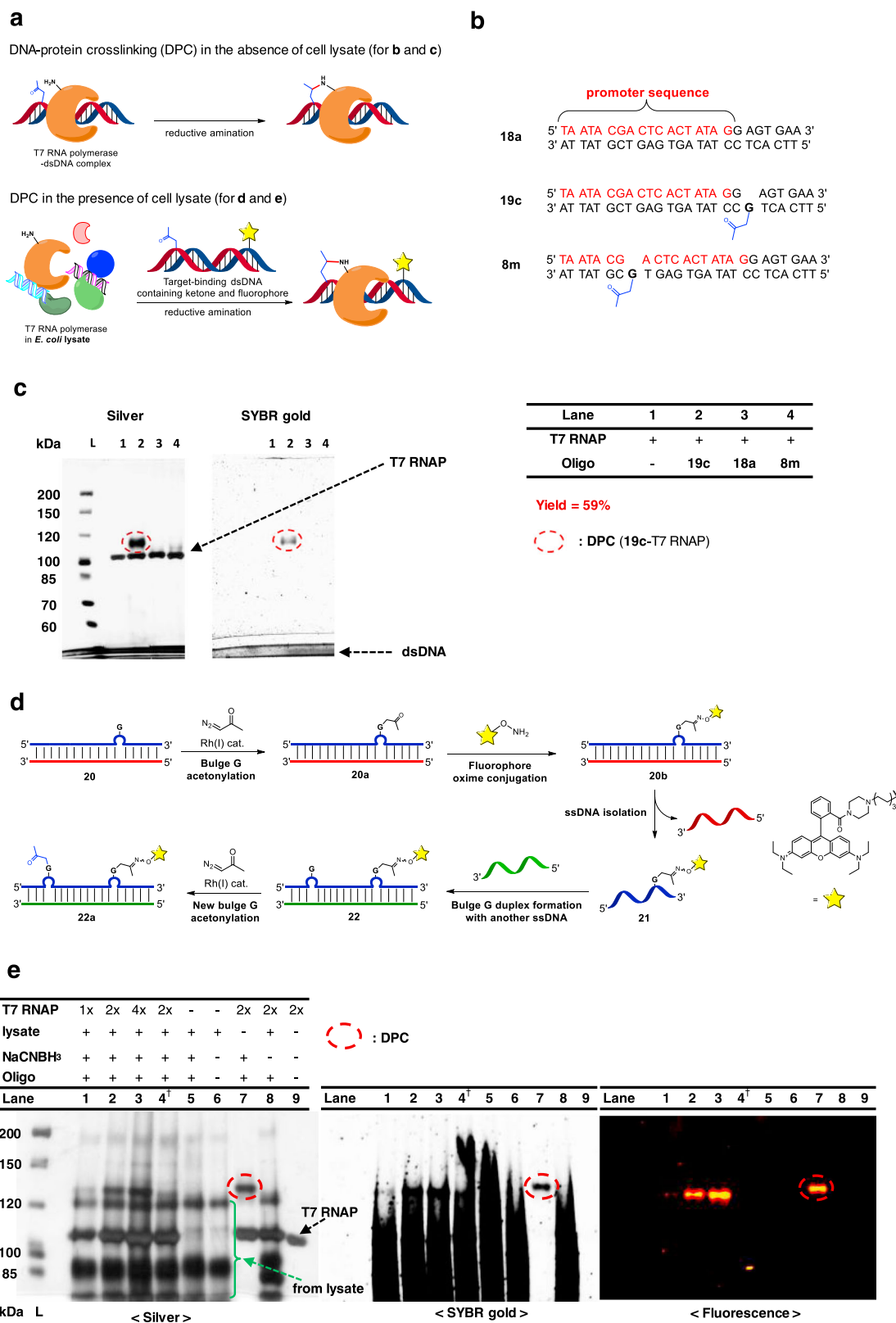
Fig. 7 Preparation of dual labeled dsDNA and T7 RNA polymerase-promoter cross-linking. a DNA-protein cross-linking of the probe and DNA-binding protein in the absence (**b** and **c**) and in the presence of *E. coli* lysate (**d** and **e**). **b** Oligonucleotide probes employed for **c**. **c** SDS–PAGE analysis of the cross-linking reaction visualized by silver and SYBR gold stains. Conditions: dsDNA (20 μM), T7 RNA polymerase (T7 RNAP) (6.8 μM), HEPES (40 mM), pH 7.4, Mg(OAc)$_2$ (5 mM), NaCNBH$_3$ (100 mM), 37 °C, 12 h. **d** Synthetic scheme for dual G-bulge probe **22a** employed for **e**. **e** SDS–PAGE analysis of DNA-protein cross-linking in *E. coli* lysate. Conditions: dsDNA (20 μM), T7 RNAP (3.4 μM, for 1x), HEPES (40 mM), pH 7.4, Mg(OAc)$_2$ (5 mM), NaCNBH$_3$ (100 mM), 30% *E. coli* lysate, 37 °C, 12 h. †unmodified oligo was used. L: protein standard ladder. Source data are provided as a Source Data file.

SYBR gold stains (Fig. 7c). Whereas strong bands corresponding to both the cross-linked product and T7 RNAP were observed for **19c** in the silver staining, **8m** bearing a bulge in the middle of the promoter sequence showed only the band for RNAP with no cross-linking (lanes 2 and 4, respectively). We speculate that the failure of **8m** in cross-linking may have resulted from the disruption of the binding of RNAP caused by the bulge or the absence of a proximal lysine residue. The presence of dsODN **19c** in the cross-linked product was further confirmed by SYBR gold staining (lane 2), in which the band matched with the upper band in the silver stain (lane 2), as opposed to the absence of the corresponding band for **8m** (lane 4, SYBR gold). Moreover, no cross-linking was observed in the reductive amination when a natural duplex **18a** (lane 3) or a duplex **19i** bearing a non-promoter sequence (Supplementary Fig. 10) was employed as negative controls. These results establish that the cross-linking is sequence-specific and modification site-specific.

Next, we examined the cross-linking by employing *E. coli* whole cell lysate as a complex matrix (Fig. 7a). To detect the cross-linked product in the cell lysate, we designed a dual G-bulge dsDNA probe, in which the second acetonyl group was used for conjugation with a fluorescence probe (Fig. 7d; see SI p. 35, Supplementary Figs. 60 and 61). For the selective manipulation of the two acetonyl groups, they were sequentially introduced in an iterative manner. Thus, the formation of a first G-bulge dsDNA at the desired position was achieved by using the corresponding opposite strand, which was subsequently reacted with diazoacetone to give **20a**. After Rhodamine B alkoxyamine was linked to the ketone via an oxime ether, the ssDNA **21** bearing the fluorophore was isolated. Introduction of a second G-bulge by duplex formation with the corresponding opposite strand followed by the reaction with diazoacetone afforded the dual act-G-bulge dsDNA probe **22a** ready for the cross-linking experiment. Notably, the reaction proceeded selectively at the bulge-G while the dye moiety remaining intact. The stability of the duplex **22** may be slightly lower than **20** owing to an additional mismatch pair, however, it does not appear to affect the reaction based on the observation that the second labeling also proceeds uneventfully.

Cross-linking was performed by reductive amination, in which **22a** was incubated with NaCNBH₃ in the presence of *E. coli* lysate pretreated with T7 RNAP (Fig. 7e, lanes 1–3). Control experiments were also performed (a) with nonmodified form of **22a** (lane 4), (b) without T7 RNAP (lane 5), (c) with lysate only (lane 6), (d) without lysate (lane 7), (e) without NaCNBH₃ (lane 8), (f) with T7 RNAP only (lane 9). The SDS–PAGE gel was visualized by fluorescence, silver stain, and SYBR gold. The visualization with fluorescence confirmed that selective cross-linking was successfully achieved even in the presence of the cell lysate (lanes 1–3); the bands with the same mobility compared to that in the positive control experiment (lane 7) was detected.

To confirm that the band corresponding to the cross-linking of **22a** with RNAP is indeed derived from the sequence-specific conjugation among the proteins in the lysate, we performed a DPC experiment in the presence of varying concentrations of RNAP (lanes 1–3). The observation that the intensity of the fluorescent bands on the SDS–PAGE gel increases with the increasing concentration of RNAP clearly supports that the bands are derived from the sequence-specific conjugation of RNAP with **22a**. Further staining with silver and SYBR gold also provided consistent results with the expected stain patterns (lanes 1–3 vs. positive control 7). Moreover, the results from several negative control experiments further support the specific cross-linking (lanes 4–6, 8, and 9).

In summary, we developed chemo- and regio-selective modification of ONs based on Rh(I)-carbene catalysis, which allows the selective introduction of an acetonyl group at G-bulges located in various ssODNs, ssORNs, hairpins, duplexes, and hybrid duplexes. Moreover, the protocol enables the introduction of multiple non-redundant functional groups in an iterative fashion. The utility of the protocol has been demonstrated with several applications.

Methods

Rh(I)-catalyzed O⁶-G acetonylation of ODNs or ORNs. All reactions were performed in PCR tubes in a total volume of 10 μL. An aqueous solution of ODN or ORN (50 nmol) in a PCR tube was concentrated, and the ON resuspended in water (4 μL) was treated with an aqueous solution of Mg(OAc)₂ (2 μL, 50 mM), MES buffer (2 μL, 100 mM, pH 6.0), and an aqueous solution of diazoacetone (1 μL, 400 mM). A solution of [Rh(COD)Cl]₂ in THF (1 μL, 5 mM) was added and mixed thoroughly. The reaction mixture was incubated at r.t. for 0.5 h unless otherwise noted and analyzed by HPLC-MS.

Derivatization of the acetonylated products by oxime ether conjugation with benzyloxyamine was performed when partial separation of HPLC peaks was observed. The crude material from the reaction was diluted with water (20 μL), extracted with ethyl acetate (30 μL × 3), and the resulting aqueous layer was concentrated. The crude material was dissolved in water, and a small portion (5 nmol of dsDNA) was diluted with water to a final volume of 4 μL. The solution was treated with benzyloxyamine (1 μL, 50 mM in DMSO), kept at r.t. for 1 h (unless otherwise noted), and analyzed by HPLC-MS.

Statistics and reproducibility. All the experiments were at least duplicated to confirm reproducibility. All attempts at replication were successful.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Source data are provided with this paper.

References

1. Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. & Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669–672 (2020).
2. Papai, G. et al. Structure of SAGA and mechanism of TBP deposition on gene promoters. *Nature* **577**, 711–716 (2020).
3. Patel, A. B. et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, eaau8872 (2018).
4. Napolitano, G. et al. A substrate-specific mTORC1 pathway underlies Birt–Hogg–Dubé syndrome. *Nature* **585**, 597–602 (2020).
5. Liu, Y. et al. Photoaffinity labeling of transcription factors by DNA-templated crosslinking. *Chem. Sci.* **6**, 745–751 (2015).
6. Zhang, L., Zhang, K., Prändl, R. & Schöffl, F. Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation. *Biochem. Biophys. Res. Comm.* **322**, 705–711 (2004).
7. Preston, G. W. & Wilson, A. J. Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem. Soc. Rev.* **42**, 3289–3301 (2013).
8. Nguyen, T. M., Nakata, E., Saimura, M., Dinh, H. & Morii, T. Design of modular protein tags for orthogonal covalent bond formation at specific DNA sequences. *J. Am. Chem. Soc.* **139**, 8487–8496 (2017).
9. Panja, S., Paul, R., Greenberg, M. M. & Woodson, S. A. Light-triggered RNA annealing by an RNA chaperone. *Angew. Chem. Int. Ed.* **54**, 7281–7284 (2015).
10. Goodchild, J. Conjugates of oligonucleotides and modified oligonucleotides: a review of their synthesis and properties. *Bioconjug. Chem.* **1**, 165–187 (1990).
11. Dai, Q. & He, C. Syntheses of 5-formyl- and 5-carboxyl-dC containing DNA oligos as potential oxidation products of 5-hydroxymethylcytosine in DNA. *Org. Lett.* **13**, 3446–3449 (2011).
12. Zewge, D. et al. Post-synthetic modification of oligonucleotides via orthogonal amidation and copper catalyzed cycloaddition reactions. *Bioconjugate Chem.* **29**, 1859–1865 (2018).
13. Tomkuvienė, M., Mickutė, M., Vilkaitis, G. & Klimašauskas, S. Repurposing enzymatic transferase reactions for targeted labeling and analysis of DNA and RNA. *Curr. Opin. Biotechnol.* **55**, 114–123 (2019).
14. Deen, J. et al. Methyltransferase-directed labeling of biomolecules and its applications. *Angew. Chem. Int. Ed.* **56**, 5182–5200 (2017).

15. Shu, X. et al. N6-Allyladenosine: a new small molecule for RNA labeling identified by mutation assay. *J. Am. Chem. Soc.* **139**, 17213–17216 (2017).
16. Proudnikov, D. & Mirzabekov, A. Chemical methods of DNA and RNA fluorescent labeling. *Nucleic Acids Res.* **24**, 4535–4542 (1996).
17. Defrancq, E., Singh, Y. & Spinelli, N. Chemical strategies for oligonucleotide-conjugates synthesis. *Curr. Org. Chem.* **12**, 263–290 (2008).
18. Camden, A. J., Walsh, S. M., Suk, S. H. & Silverman, S. K. DNA oligonucleotide 3′-phosphorylation by a DNA enzyme. *Biochemistry* **55**, 2671–2676 (2016).
19. Qin, P. Z. & Pyle, A. M. Site-specific labeling of RNA with fluorophores and other structural probes. *Methods* **18**, 60–70 (1999).
20. Gampe, C. M., Hollis-Symynkywicz, M. & Zécri, F. Covalent chemical 5′-functionalization of RNA with diazo reagents. *Angew. Chem. Int. Ed.* **55**, 10283–10286 (2016).
21. Poudyal, R. R. et al. Nucleobase modification by an RNA enzyme. *Nucleic Acids Res.* **45**, 1345–1354 (2016).
22. Wilson, C. & Szostak, J. W. In vitro evolution of a self-alkylating ribozyme. *Nature* **374**, 777–782 (1995).
23. McDonald, R. I. et al. Electrophilic activity-based RNA probes reveal a self-alkylating RNA for RNA labeling. *Nat. Chem. Biol.* **10**, 1049–1054 (2014).
24. Sharma, A. K. et al. Fluorescent RNA labeling using self-alkylating ribozymes. *ACS Chem. Biol.* **9**, 1680–1684 (2014).
25. Sasaki, S., Onizuka, K. & Taniguchi, Y. The oligodeoxynucleotide probes for the site-specific modification of RNA. *Chem. Soc. Rev.* **40**, 5698–5706 (2011).
26. Egloff, D., Oleinich, I. A. & Freisinger, E. Sequence-specific generation of 1,N6-ethenoadenine and 3,N4-ethenocytosine in single-stranded unmodified DNA. *ACS Chem. Biol.* **10**, 547–553 (2015).
27. Jitsuzaki, D. et al. Remarkable acceleration of a DNA/RNA inter-strand functionality transfer reaction to modify a cytosine residue: the proximity effect via complexation with a metal cation. *Nucleic Acids Res.* **42**, 8808–8815 (2014).
28. Zhou, C. Y., Alexander, S. C. & Devaraj, N. K. Fluorescent turn-on probes for wash-free mRNA imaging via covalent site-specific enzymatic labeling. *Chem. Sci.* **8**, 7169–7173 (2017).
29. Alexander, S. C., Busby, K. N., Cole, C. M., Zhou, C. Y. & Devaraj, N. K. Site-specific covalent labeling of RNA by enzymatic transglycosylation. *J. Am. Chem. Soc.* **137**, 12756–12759 (2015).
30. Park, H. S., Kietrys, A. M. & Kool, E. T. Simple alkanoyl acylating agents for reversible RNA functionalization and control. *Chem. Commun.* **55**, 5135–5138 (2019).
31. Habibian, M. et al. Reversible RNA acylation for control of CRISPR–Cas9 gene editing. *Chem. Sci.* **11**, 1011–1016 (2020).
32. Ghaem Maghami, M., Scheitl, C. P. M. & Höbartner, C. Direct in vitro selection of trans-acting ribozymes for posttranscriptional, site-specific, and covalent fluorescent labeling of RNA. *J. Am. Chem. Soc.* **141**, 19546–19549 (2019).
33. Ford, A. et al. Modern organic synthesis with α-diazocarbonyl compounds. *Chem. Rev.* **115**, 9981–10080 (2015).
34. Ohata, J., Martin, S. C. & Ball, Z. T. Metal-mediated functionalization of natural peptides and proteins: panning for bioconjugation gold. *Angew. Chem. Int. Ed.* **58**, 6176–6199 (2019).
35. Zhou, C.-Y., Huang, J.-S. & Che, C.-M. Ruthenium-porphyrin-catalyzed carbenoid transfer reactions. *Synlett* **2010**, 2681–2700 (2010).
36. Antos, J. M., McFarland, J. M., Ivarone, A. T. & Francis, M. B. Chemoselective tryptophan labeling with rhodium carbenoids at mild pH. *J. Am. Chem. Soc.* **131**, 6301–6308 (2009).
37. Ball, Z. T. Designing enzyme-like catalysts: a rhodium(II) metallopeptide case study. *Acc. Chem. Res.* **46**, 560–570 (2013).
38. Ho, C.-M. et al. A water-soluble ruthenium glycosylated porphyrin catalyst for carbenoid transfer reactions in aqueous media with applications in bioconjugation reactions. *J. Am. Chem. Soc.* **132**, 1886–1894 (2010).
39. Antos, J. M. & Francis, M. B. Selective tryptophan modification with rhodium carbenoids in aqueous solution. *J. Am. Chem. Soc.* **126**, 10256–10257 (2004).
40. Tishinov, K., Schmidt, K., Häussinger, D. & Gillingham, D. G. Structure-selective catalytic alkylation of DNA and RNA. *Angew. Chem. Int. Ed.* **51**, 12000–12004 (2012).
41. Geigle, S. N., Wyss, L. A., Sturla, S. J. & Gillingham, D. G. Copper carbenes alkylate guanine chemoselectively through a substrate directed reaction. *Chem. Sci.* **8**, 499–506 (2017).
42. Rasale, D. et al. A new water soluble copper N-heterocyclic carbene complex delivers mild O6G-selective RNA alkylation. *Chem. Commun.* **54**, 9174–9177 (2018).
43. Xia, Y., Qiu, D. & Wang, J. Transition-metal-catalyzed cross-couplings through carbene migratory insertion. *Chem. Rev.* **117**, 13810–13889 (2017).
44. Doyle, M. P., Duffy, R., Ratnikov, M. & Zhou, L. Catalytic carbene Insertion into C−H Bonds. *Chem. Rev.* **110**, 704–724 (2010).
45. Davies, H. M. L. & Hedley, S. J. Intermolecular reactions of electron-rich heterocycles with copper and rhodium carbenoids. *Chem. Soc. Rev.* **36**, 1109–1119 (2007).
46. Domingo, L. R. & Pérez, P. The nucleophilicity N index in organic chemistry. *Org. Biomol. Chem.* **9**, 7168–7175 (2011).
47. Domingo, L. R., Chamorro, E. & Pérez, P. Understanding the reactivity of captodative ethylenes in polar cycloaddition reactions. A theoretical study. *J. Org. Chem.* **73**, 4615–4624 (2008).
48. Serra, M. J. et al. Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA* **8**, 307–323 (2002).
49. Liu, Q. & Deiters, A. Optochemical control of deoxyoligonucleotide function via a nucleobase-caging approach. *Acc. Chem. Res.* **47**, 45–55 (2014).
50. Zhang, L. et al. Caged circular siRNAs for photomodulation of gene expression in cells and mice. *Chem. Sci.* **9**, 44–51 (2018).
51. Debart, F., Dupouy, C. & Vasseur, J.-J. Stimuli-responsive oligonucleotides in prodrug-based approaches for gene silencing. *Beilstein J. Org. Chem.* **14**, 436–469 (2018).
52. Seyfried, P., Eiden, L., Grebenovsky, N., Mayer, G. & Heckel, A. Photo-tethers for the (Multi-)cyclic, conformational caging of long oligonucleotides. *Angew. Chem. Int. Ed.* **56**, 359–363 (2017).

Acknowledgements

Author contributions

C.-M.P. conceived, designed, and directed the project; Y.-H.L. performed the experiments and analyzed the data; E.Y. performed the experiments, DFT calculations, and analyzed the data; C.-M.P., Y.-H.L., and E.Y. prepared the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21839-4.

Correspondence and requests for materials should be addressed to C.-M.P.

Peer review information *Nature Communications* thanks Dennis Gillingham, Takashi Morii and Zachary Ball for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.