

Assembly of a biocompatible triazole-linked gene by one-pot click-DNA ligation.

Mikiembo Kukwikila,¹ Nittaya Gale,² Afaf H. El-Sagheer^{3,4} Tom Brown^{3*} and Ali Tavassoli^{1,5,*}

1. Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom.
2. ATDBio, Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom.
3. Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Oxford, OX1 3TA, United Kingdom.
4. Chemistry Branch, Department of Science and Mathematics, Suez University, Suez 43721, Egypt.
5. Institute for Life Science, University of Southampton, Southampton, SO17 1BJ, United Kingdom,

*e-mail: a.tavassoli@soton.ac.uk and tom.brown@chem.ox.ac.uk

The chemical synthesis of oligonucleotides and their enzyme-mediated assembly into genes and genomes has significantly advanced multiple scientific disciplines. However, these approaches are not without their shortcomings; enzymatic amplification and ligation of oligonucleotides into genes and genomes makes automation challenging, while multiple site-specific incorporation of epigenetic information and/or modified bases into large constructs is not feasible. Here, we present a fully chemical, one-pot method for the assembly of oligonucleotides into a gene by click-DNA ligation. We synthesize the 335 base-pair gene encoding the green fluorescent protein iLOV from ten functionalized oligonucleotides containing 5'-azide and 3'-alkyne units. The resulting click-linked iLOV contains eight triazoles at the sites of chemical ligation, yet is fully biocompatible; it is replicated by DNA polymerases *in vitro* and encodes a functional iLOV protein in *E. coli*. We demonstrate the power and potential of our one-pot gene assembly method by preparing an epigenetically modified variant of the iLOV gene.

The ability to design and synthesize large fragments of DNA has underpinned and revolutionized multiple fields including cell biology, biotechnology and synthetic biology.^{1, 2, 3, 4} While chemical synthesis of short oligonucleotide fragments (<100 bases) is routine, the synthesis of longer fragments is often plagued by poor yields and high error rates which occur as a function of oligonucleotide length.⁵ As a result, large DNA fragments need to be assembled from multiple short oligonucleotides using enzymes.^{3, 6, 7} Current assembly approaches typically make use of PCR amplification^{8, 9} or enzymatic ligation,^{1, 4} and although these approaches are well established and form the cornerstone of current gene and genome synthesis efforts, they have some limitations. First, chemical modifications and epigenetic information cannot be introduced into a gene or genome site-specifically by polymerase-based methods as modified bases are not differentiated by these enzymes. Second, current assembly methods do not readily lend themselves to automation, and therefore require significant effort and time. Third, the assembly reactions are often low yielding so require a final PCR amplification step to isolate the full length product from the partially assembled fragments, thus erasing epigenetic information even in ligase-based methods.^{4, 8, 9} Despite these limitations, enzymatic assembly has been used to prepare genomes of over a million base pairs¹⁰ and is routinely employed for the preparation of genes in everyday research. However, given the challenge of preparing ever increasing lengths of epigenetically modified DNA at larger scale, and the need for chemically modified DNA constructs,¹¹ alternative DNA assembly methods and ligation chemistries are needed.

We envisage that a purely chemical method for linking and assembling synthetic oligonucleotides into larger fragments would overcome the above limitations and potentially reduce the time and cost associated with gene and genome synthesis. In this strategy, a highly specific and selective chemical reaction is used to join oligonucleotides with the required functional groups at each terminus (Figure 1a). Given the self-templating properties of DNA, a one-pot assembly strategy can be envisaged whereby the sense and antisense strands of the desired gene are segmented into overlapping fragments that are chemically synthesized as functionalized oligonucleotides (Figure 1ai). These are annealed in the correct order by DNA self-templating to bring the neighbouring functional groups in sufficient proximity to enable bond formation (Figure 1aii). The chemical reaction between the functionalized termini may be initiated through the addition of a catalyst leading to covalent bonding between the oligonucleotides to give the desired gene (Figure 1aiii). However, an absolute requirement for the utility of such a process is the tolerance of the chemical ‘scar’ formed at the site of ligation (in place of the phosphodiester bond) in biological systems, so that the resulting DNA strand is correctly replicated and transcribed in cells.

There are several examples of nucleotide analogues linked with non-phosphodiester linkages, using a variety of chemical reactions,¹² including nucleophilic displacement,^{13, 14} photoligation,¹⁵ and disulfide bonding.^{16, 17} DNA self-templating has also been used to template a variety of organic reactions.^{18, 19, 20, 21} For our purpose, we selected the copper-catalyzed azide-alkyne cycloaddition (CuAAC), which has previously been used to link oligonucleotides functionalized with a 3'-propargyl and a 5'-methylene azide (Figure 1b).^{22, 23, 24, 25, 26} The resulting click-linked DNA backbone was shown by us to be accurately replicated and transcribed in bacterial^{22, 27} and human cells²⁸ when a single triazole linker was incorporated into each strand of a DNA duplex. In addition, the linker is accurately transcribed *in vitro* by T7 RNA polymerase,²⁹ while structural studies illustrated minor disturbance to the double helix structure from a single triazole-linker.³⁰ Collectively these results demonstrate the surprising biocompatibility of a single triazole-linker incorporated into DNA. However, it remains unknown whether multiple triazole-linkers incorporated into the same DNA strand are tolerated by living systems, and whether multiple click-DNA ligation reactions can be used in combination with the self-assembling properties of DNA for one-pot gene synthesis.

Here we show gene assembly via this one-pot click-ligation strategy and demonstrate the biocompatibility of the resulting triazole-containing DNA in *E. coli*. The viability of this approach is illustrated with the assembly of the 335 bp gene that encodes the fluorescent protein iLOV³¹ from ten functionalized oligonucleotides (Figure 1c). This work serves as proof-of-concept for enzyme-free gene assembly and could pave the way for the fully chemical synthesis of genes and genomes.

Results

Synthesis of oligonucleotides comprising the iLOV gene. The 335-bp gene encoding iLOV was codon-optimized for expression in *E. coli*. The sites of ligation were chosen to be between adjacent deoxycytidine (dC) and deoxythymidine (dT) residues (CpT steps) due to the simplicity of preparing the appropriate modified nucleosides, and positioned throughout the gene to give an overlap of at least 10 base pairs between the sense and antisense strands (Figure 1c and Supplementary Table 1). Efforts were also made to keep the length of each oligonucleotide between 50-80 bases to enable high-fidelity synthesis and facilitate purification. The sequences of the oligonucleotides were not optimized to give favourable melting temperatures or avoid secondary structures; in fact the melting temperatures

of the individual fragments differed by as much as 6.6 °C due to varying GC-contents (Supplementary Table 2). These oligonucleotides were synthesized by phosphoramidite chemistry using a resin functionalized with 3'-propargyl dC. The monomer was isolated from a one-pot reaction in which N⁴-acetyl-2'-deoxy-5'-O-DMT cytidine was alkylated using propargyl bromide in the presence of NaH in THF. Sonication for 2 h at room temperature drove the reaction to completion without the formation of by-products. Methanol was subsequently added to quench the residual propargyl bromide and simultaneously remove the acetyl protecting group to give the propargyl modified nucleoside in 68% yield; an improvement on the previously published 2-step reaction to synthesize 3'-propargyl methyl-dC monomer.³² The monomer was then loaded onto an amino functionalized resin in preparation for DNA synthesis (Figure 2).³² The 5'-azide functional group was installed by incorporating 5'-O-dimethoxytrityl-5-iodo-2'-deoxyuridine-3'-phosphoramidite as the final monomer during oligonucleotide synthesis and treating the resin-bound oligonucleotide with sodium azide to displace the iodine.²² The terminal fragments of the sense and antisense strands (F5 and R5, Supplementary Table 1) were synthesized to contain 5'-phosphate groups rather than azides. To enable ligation of the final synthetic gene into a plasmid, the sequences of the four terminal fragments (F1, F5, R1 and R5, Supplementary Table 1) were designed to include overhangs compatible with digestion with *NdeI* and *EcoRI* restriction endonucleases (Figure 1c).

Two different methods of purifying the oligonucleotide were evaluated; the crude solutions of each oligonucleotide were divided into two fractions; one fraction purified by semi-preparative HPLC using a hexylammonium acetate/acetonitrile buffer system and the other *via* polyacrylamide gel electrophoresis (PAGE). The purity of the oligonucleotides was then quantified *via* analytical HPLC and capillary electrophoresis. We observed that HPLC purification yielded purer oligonucleotides than PAGE purification (Supplementary Table 3). Furthermore, we found that capillary electrophoresis tended to overestimate the purity of the oligonucleotides compared to analytical HPLC (Supplementary Table 3). Consequently, only the HPLC-purified oligonucleotides were used to assemble the iLOV gene.

Click-mediated assembly of the gene encoding iLOV. The ten oligonucleotide fragments synthesized above were combined in salt solution, heated at 95°C for 15 min, then cooled to room temperature over 2 h to enable annealing. We hypothesized that the self-templating properties of DNA would enable annealing of the oligonucleotides in the correct order to give the unligated iLOV gene. The alkyne- and azide-functionalized termini of these thermally assembled oligonucleotides were simultaneously reacted to form triazoles via addition of Cu(II) sulphate and sodium ascorbate. A control assembly reaction containing modified oligonucleotides, but no copper was also carried out to assess the importance of click-linking the annealed DNA. The crude click reactions were purified by PAGE under denaturing conditions to ensure that any unreacted fragments or by-products migrated separately from the assembled gene. As expected, in the absence of the copper catalyst, the individual oligonucleotide fragments did not assemble to form the iLOV gene, but rather migrated individually towards the bottom of the gel (Figure 3a, lane 1). In the copper-containing reaction however, a distinct high molecular weight band was observed (Figure 3a, lane 2), indicating correct assembly and click-DNA ligation of the functionalized oligonucleotides. The band corresponding to the assembled iLOV gene was excised from the gel and the DNA was extracted to give 95 µg (4.5% isolated yield) of pure click-linked iLOV gene as a single band at ~350 bp (Figure 3b lane 2 and insert in Figure 3b).

The effect of multiple click-linkers on the secondary structure of DNA. A single triazole incorporated into the backbone of DNA has been shown to cause minor distortions that result in displacement of the deoxyribose sugar and increased mobility of the base pairs flanking the triazole.³⁰ Given this, we were concerned that the presence of multiple triazoles might drastically perturb the secondary structure of the click-linked gene and affect its biocompatibility. We therefore conducted circular-dichroism (CD) spectroscopy analysis to probe this. CD analysis of the click-linked iLOV gene gave bisignate signals with maxima at $\lambda = +276/-248$ nm which are characteristic of the B-type DNA helix,³³ identical to that observed for the canonical iLOV gene (Figure 3b). The lack of perturbation in the CD signal for click-linked iLOV when compared to the canonical equivalent demonstrates that the multiple backbone triazoles do not significantly alter the conformation of the DNA double helix. However, melting temperature analysis of the two constructs suggests that the triazoles do have a minor destabilizing effect. The canonical iLOV gave a melting temperature of 84.3 °C while that of the click-linked gene was 81.3 °C, indicating that the eight triazoles destabilize the duplex by 3 °C (Supplementary Figure 1).

Replication of click-linked iLOV *in vitro*. The ability of DNA polymerases to replicate click-linked iLOV *in vitro* was next assessed. We used click-linked iLOV as a template for PCR amplification by either *Taq* or *Pfu* polymerase, and the unligated, modified oligonucleotides were used as a negative control. Click-linked iLOV was amplified by both DNA polymerases, giving a single band that appeared at ~350 bp markers in the ladder (Figure 3d, lanes 3 and 5). In contrast, PCR of the unligated oligonucleotides did not give rise to a band corresponding to the full length iLOV product. Instead a strong band was observed at ~150bp, with weaker bands visible at ~275 and ~300 bp (Figure 3d, lanes 2 and 4). The absence of truncation products when amplifying the click-linked iLOV, suggests that the triazole backbone linkers do not cause the polymerase to detach from the template or stall at the triazole. We also assessed the ability of DNA polymerases to read through click-linked iLOV using a primer extension assay, whereby a single strand of the click-linked DNA is copied in linear fashion by the Klenow fragment (exo⁻) of DNA polymerase I (Klenow). This assay uses the sense strand of click-linked iLOV as template, with the polymerase processing through this strand. The single-strand click-linked template required for the primer extension assay was generated using a template-mediated click-ligation strategy (Supplementary Figure 2) using the modified oligonucleotide fragments that comprised the sense strand of iLOV. These oligonucleotides were assembled in the correct order using four shorter complementary splints which spanned 20-25 bases either side of the ligation points. The assembled sense strand was observed as a high molecular weight band at the top of the gel; as with the one-pot gene assembly, cyclization and truncation products were observed in addition to the residual unreacted oligonucleotides (Supplementary Figure 2). The click-linked sense strand of iLOV was extended successfully by Klenow, depleting all the primer added to the reaction, with only a single band corresponding to the full length iLOV gene being observed (Figure 3e, lane 2). The absence of any truncated species or unreacted primer indicates that Klenow successfully reads through the four triazole backbone-linkers, extending the primer to the full length complementary strand. As expected, the primer remained unextended in the negative control reaction lacking Klenow (Figure 3e, lane 1).

We further probed the effect of DNA backbone triazoles on DNA polymerases using real-time PCR (qPCR). We designed four primers to bind upstream of each triazole on the sense strand of click-linked iLOV to monitor the rate of

replication through the increasing number of triazoles by *Taq* polymerase. For comparison, the experiment was repeated using canonical iLOV as template. We observed an inverse relationship between the number of triazoles in the DNA backbone and rate of PCR product formation (Figure 3f). We compared the threshold cycle numbers (Ct value) of the templates (the point at which the fluorescence is first detected as statistically significant above the threshold) for click-linked iLOV and the canonical equivalent. When reading through one triazole linker, the Ct values for the click-linked iLOV and canonical iLOV were comparable at 9.4 ± 0.1 and 7.5 ± 0.1 respectively (Figure 3fi). The difference in Ct values increased when amplifying through two triazoles to 11.3 ± 0.1 for click-linked iLOV and 7.0 ± 0.1 for canonical iLOV (Figure 3fii). When reading through three triazoles, the difference in Ct values further increased to 18.5 ± 0.8 for click-linked iLOV and 6.6 ± 0.2 for canonical iLOV (Figure 3fiii), while reading through four triazoles gave Ct values of 20.7 ± 0.5 for click-linked iLOV and 6.3 ± 0.1 for canonical iLOV (Figure 3fiv). This data suggested that increasing the number of triazoles in the backbone of the template DNA slows down DNA replication by *Taq* polymerase. It should be noted that although the number cycles required to detect a significant amount of PCR product above the baseline fluorescence increases with more triazoles, the flat line at zero during these early cycles (e.g. in Figure 3fiv) does not equate to the lack of DNA replication. Rather, the amount of DNA produced in the early cycle is below the minimum fluorescence threshold and so is reported as zero by the instrument. Furthermore, total product levels for the PCR reactions with click-linked iLOV approach that of the canonical equivalent by the end of the PCR reaction in all cases. To further probe this effect, we repeated each amplification reaction (through increasing number of triazoles) with *Taq* DNA polymerase, visualized the DNA produced on an agarose gel (Supplementary Figure 3) and quantified the concentration of DNA produced (Supplementary Table 3). Interestingly, no significant difference was observed between using click-linked iLOV or the canonical iLOV gene as template, even when reading through 4 triazoles (Supplementary Figure 3 and Supplementary Table 4). This suggests that the difference in replication rates observed by qPCR is normalized by the end of the PCR reaction. Any effect from the triazole-linkages will only be relevant during the first few PCR cycles; after this, there will be a large excess of unmodified template product, so amplification will proceed at the same rate in all reactions. It is unclear whether the observed triazole-dependent slowdown in PCR replication would also affect replication of click-linked iLOV in cells.

Probing the biocompatibility of the click-linked iLOV gene in *E. coli*. The biocompatibility of our clicked-linked gene was next probed in *E. coli*. For comparison, we assembled the canonical iLOV gene with T4-DNA ligase (ligase-assembled iLOV), using 5'-phosphorylated equivalents of the oligonucleotides used for click-ligation. Both the click-linked iLOV and ligase-assembled iLOV were designed to contain sticky ends for ligation into the pRSET-mCherry plasmid³⁴ cleaved by *NdeI* and *EcoRI* restriction endonucleases (Figure 4). These restriction sites were chosen so that the click-linked gene, which encodes the green fluorescent protein iLOV, directly replaces the gene encoding for the red fluorescent protein mCherry in the backbone. Thus surviving colonies containing click-linked iLOV would be readily distinguished from false positives containing the parent plasmid by the colour of their fluorescence. The pRSET-mCherry plasmid was sequentially digested with *NdeI* and *EcoRI* restriction endonucleases, which removes the gene encoding mCherry, followed by treatment with shrimp alkaline phosphatase to remove the terminal phosphate groups to minimize backbone-only ligation. Click-linked iLOV was ligated into the purified digested vector

with T4 DNA ligase. The ligation mixture was transformed into KRX *E. coli* cells, chosen as they contain a chromosomal copy of T7 RNA polymerase and therefore allow rhamnose-dependent overexpression of iLOV from the pRSET plasmid used. Transformed cells were plated on LB agar, incubated overnight and the resulting colonies were assessed for the green fluorescent phenotype associated with transcription and translation of click-linked iLOV (Figure 5a). Of the 551 assessed colonies ($n = 15$) $56.9 \pm 9.6\%$ displayed the green fluorescent phenotype associated with iLOV, while $29.3 \pm 8.6\%$ were white and $13.8 \pm 10.4\%$ were red (Figure 5b). When using the ligase-assembled iLOV, of the 459 assessed colonies ($n = 15$) $40.1 \pm 12.1\%$ were green, $44.5 \pm 9.7\%$ were white and $15.4 \pm 12.1\%$ were red (Figure 5B). It is interesting to note that in both cases $\sim 15\%$ of the colonies still contained undigested or singly digested and religated pRSET-mCherry, suggesting poor performance by one of the restriction endonucleases despite gel purification of the double digested backbone and treatment with shrimp alkaline phosphatase (to prevent backbone-only ligation). There are two possibilities that would lead to white colonies; either the gene for iLOV in those colonies is mutated, or the plasmid does not contain an insert. To further probe this, we isolated and sequenced the plasmids from 25 white colonies; 92% of these colonies contained point mutations that led to frameshifts or deletions in the iLOV gene, while the remaining 8% of colonies contained religated vector without an insert. Plasmids from 50 colonies expressing the green phenotype of the click-linked iLOV gene were isolated and their sequences were analyzed. Of these colonies, 49 (98%) possessed the correct sequence, meaning that the whole iLOV gene, including the regions flanking the triazole linkers, is correctly replicated in *E. coli* (Figure 5c). The lone mutant contained an adenine to cytosine point mutation at thirteen bases after the second triazole (position 157 on the sense strand) which did not affect iLOV function. In comparison, of the 50 green colonies sequenced from the cohort transformed with the ligase-assembled iLOV gene, 40 (80%) were error-free while the remaining 10 (20%) contained point mutations at a variety of different positions. It is interesting that the error rate is higher for ligase-assembly than click DNA ligation, both in terms of number of colonies displaying green fluorescence ($40.1 \pm 12.1\%$ for ligase versus $59.6 \pm 9.6\%$ for click) and mutations in the iLOV gene from green colonies (20% for ligase versus 2% for click). This is likely a reflection of errors incurred during oligonucleotide synthesis, or more stringent purification of the modified oligonucleotides, rather than a consequence of the chosen method of assembly.

Assessing the role of DNA repair in the biocompatibility of click-linked iLOV. The green fluorescent phenotype observed in the majority of colonies transformed with click-linked iLOV and subsequent sequencing analysis confirmed the biocompatibility and high fidelity replication of this click-linked gene in *E. coli* (Figure 5a and 5b). However, it is possible that the triazoles present on the DNA backbone of click-linked iLOV are being converted to the canonical phosphodiester linkage *via* activation of the nucleotide excision repair (NER) system rather than being directly replicated in *E. coli*. To probe this possibility, a doubly-digested pRSET plasmid ligated with click-linked iLOV as insert (as above) was transformed into a Δ UvrB strain of *E. coli* (JW0762-2);³⁵ UvrB protein is an essential component of the DNA damage recognition complex which is responsible for excising DNA lesions in the NER system.^{36, 37} If biocompatibility of click-linked iLOV was due to excision of the triazole-linkers through DNA repair, then this would be lost in the Δ UvrB strain of *E. coli*, which is incapable of NER. As JW0762-2 does not contain a copy of T7 RNA polymerase, iLOV will not be transcribed from the pRSET plasmid, therefore the green fluorescent phenotype cannot be used as an indication of biocompatibility. We therefore isolated and sequenced plasmids from 50

colonies transformed with click-linked iLOV and 50 colonies transformed with ligase-linked iLOV. Of the 50 copies of click-iLOV sequenced, 30 plasmids (60%) were error-free while the remaining 20 plasmids (40%) contained mutations; either simple point mutations (36%) or missing two or more consecutive bases (4%). Similar results were obtained for ligase-assembled iLOV; 30 plasmids (60%) were error free while the remaining 20 (40%) contained mutations; either simple point mutations (28%) or missing two or more consecutive bases (12%). The high rate of error observed is likely a consequence of a lack of NER in these cells. Nonetheless, the similarity in the ratio of error-free clones observed when using click-linked iLOV or ligase-assembled iLOV suggests that the cellular machinery of *E. coli* is able to correctly read through the multiple triazoles contained in click-linked iLOV.

Click-mediated assembly of fluorescent and epigenetically modified iLOV. While the approach detailed above can be used for the assembly of unmodified genes, the power of a purely chemical, one-pot method for gene synthesis is in the relative ease with which a variety of site-specific modifications may be incorporated into a target gene. To illustrate this, we designed and assembled an epigenetically modified iLOV gene containing 5-methylcytosine (5-meC) in the four CpGpCpG motif present (two on each strand). The four oligos encoding this region of the gene (F1, F4, R3, and R5; Supplementary Table 1) were resynthesized to contain 5-meC in the desired locations (^mcF1, ^mcF4, ^mcR3 and ^mcR5; used in place of F1, F4, R3 and R5. Supplementary Table 1). The 5-meC-containing variant of iLOV was assembled by one-pot click ligation (Figure 6a), purified, and characterized by agarose gel analysis to show a band corresponding to the size expected for iLOV (Figure 6b). The iLOV gene contains 3 AciI restriction endonuclease sites, the first of which will be blocked by the incorporation of 5-meC (Figure 6a). The click-linked 5-meC containing gene was digested with AciI and the resulting fragments analyzed by mass spectrometry. For comparison, the unmodified click-linked gene was also digested with AciI and analyzed by mass spectrometry. We observed fragments corresponding to AciI-cleavage of all three sites in the unmodified gene (Figure 6c). In contrast, the mass spectrum of the 5-meC containing gene fragments showed the expected increase in mass corresponding to two methyl groups on the expected fragments. Furthermore, only peaks associated with restriction digestion of 2 of the 3 AciI sites were observed, due to the 5-meC blocking the first AciI restriction site (Figure 6d).

To further demonstrate the ease with which genes with site-specific modifications are prepared by click-DNA ligation, we assembled a second variant of the iLOV gene that contained Cy3 on the forward strand and Oregon Green 488 on the reverse strand. Assembly of the click-linked iLOV enables FRET between these fluorophores. After synthesis of the fluorophore-containing oligos (^{fluo}F5 and ^{fluo}R1; used in place of F5 and R1. Supplementary Table 1), the gene was assembled by click-DNA ligation and purified. The product was analyzed by agarose gel, showing a band corresponding to iLOV that fluoresced at 570 nm when excited at 460 nm, indicating FRET between the fluorophores on each strand of the gene (Figure 6e). This was further analyzed using a fluorescence microplate reader, and again, the expected FRET signal was observed (Figure 6f).

Discussion

All current gene synthesis methods use enzymes to assemble synthetic oligonucleotides into gene- or genome-sized fragments.^{3, 6, 38} While these methods have been pushed to their limit, achieving challenging feats such as the synthesis of whole prokaryotic genomes^{2, 10} and a eukaryotic chromosome,³⁹ the reliance on enzymatic assembly limits

scalability, as well as the ability to incorporate site-specific modifications such as epigenetic markers (which cannot be read or written by the polymerases used for gene assembly). Yet the ability to include epigenetic information in synthetic genes will be critical in meeting the challenge of the next phase of DNA synthesis, in particular the goal of synthesizing a functional human genome.¹¹ Given the extensive level of cytosine methylation and hydroxymethylation in the human genome, and the critical role it plays in gene regulation,^{40, 41} synthesizing the human genome will only be biologically relevant if it contains epigenetic information. We therefore propose a fully chemical approach to gene assembly, using chemically modified oligonucleotides that are covalently bound into genes by a suitable chemical reaction. Such an approach has the potential to overcome the limitations of current enzyme-based methods; allowing scale-up of gene and genome assembly while concurrently reducing the associated time and costs. We have demonstrated this possibility using click-chemistry and the CuAAC reaction, but the principle of chemical DNA ligation is not limited to this reaction; it can (and should) be applicable to a variety of chemical reactions. The one key requirement however, is that the functional group produced in the DNA backbone by chemical ligation is biocompatible. We have chosen the CuAAC reaction owing to its fast reaction rate, high yield, compatibility with aqueous media, orthogonality with the functional groups present in oligonucleotides, and its biocompatibility.^{22, 27, 28} However, there are other examples of DNA backbone mimics^{12, 42, 43} whose biocompatibility remains little explored (with a few exceptions).^{44, 45} There is therefore much potential for identifying other chemical reactions that can be used for DNA ligation; combining conventional oligonucleotide synthesis with chemical DNA ligation would not only allow the synthesis of genes bearing site-specific epigenetic and fluorescent modifications on the μg -mg scale, it would also enable the automation of gene assembly, a critical step for scaling-up production and reducing the time taken to make a gene or genome.

To this end, we demonstrate the one-pot synthesis of the 335 bp iLOV gene, an epigenetically modified variant, and a fluorescently labelled derivative by click-DNA ligation of ten oligonucleotides. The resulting click-iLOV construct has eight triazole moieties in its backbone, yet is fully biocompatible in *E. coli*. Using click-DNA ligation, we prepared and isolated 95 μg of the click-linked iLOV gene (after purification), a challenging feat when using conventional enzymatic methods. We initially compared the properties of click-linked iLOV to the canonical equivalent (generated by PCR) *in vitro*; CD spectroscopy showed that both had similar secondary structures, and despite the presence of eight triazole backbone linkers, the melting temperature of click iLOV was only 3° C lower than the canonical gene. We also observed that DNA polymerases can read through the click-linked iLOV, but qPCR experiments showed a considerable effect on DNA replication by PCR from each additional triazole linker that the polymerase had to read through. The data indicate that the polymerase either slows down or stalls while reading through the triazole, resulting in additional PCR cycles being required for click-linked iLOV to pass the minimum fluorescence threshold in the qPCR assay. It should be noted however, that at the end of the multiple cycles of a typical PCR reaction, the difference in total DNA concentration was found to be negligible. Although our data indicate that under PCR conditions the triazole linkages in click-linked DNA affect amplification, these conditions, and the mechanism of DNA amplification, are far removed from those that occur in cellular replication, where DNA is copied by a holoenzyme with multiple subunits and at a lower temperature.

We next assessed the biocompatibility of the click-iLOV gene in *E. coli*. To determine the source of the mutations observed in some progeny of the click-iLOV gene (i.e. whether they are caused by the triazole linkers, arising during

oligonucleotide synthesis, or introduced during cloning), we assembled a control gene using T4 DNA ligase. The gene was constructed from ten oligonucleotides which are canonical equivalents of those used for click-ligation. The ligase-assembled iLOV was transformed into *E. coli* and sequence fidelity of the progeny assessed as for click-linked iLOV. Interestingly, we found fewer mutations in iLOV genes isolated from cells transformed with click-linked iLOV than those isolated from cells transformed with ligase-linked iLOV, both in terms of ratio of functional genes produced ($56.9 \pm 9.6\%$ for click-linked iLOV versus $40.1 \pm 12.1\%$ for ligase-linked iLOV), and ratio of functional genes containing errors (2% for click-linked iLOV versus 20% for ligase-linked iLOV). The mutations observed in the click-linked gene were not located at, or adjacent to, the sites of click-ligation, and are therefore unlikely to be a consequence of the triazole-linked backbone. The most likely cause of these errors lies in the oligonucleotide synthesis and purification steps, and this may account for the higher error rates when using non-modified oligonucleotides. We next assessed the contribution of NER to the observed biocompatibility in cells using a UvrB-deficient *E. coli* strain (incapable of NER). We observed a similar rate of error in the iLOV sequence isolated from the progeny of cells transformed with click-linked iLOV as for those transformed with ligase-linked iLOV. The above indicates that the triazole-linkers are truly biocompatible and furthermore our experiments demonstrate the viability of using a chemical ligation strategy for gene synthesis.

The limits of gene and genome synthesis are continuously driven forward by new techniques and technologies. The one-pot click-mediated DNA ligation approach presented here yields biocompatible genes that contain multiple triazole linkers at each site of chemical ligation and offers an alternative, fully chemical approach to gene and genome assembly. The ability of this method to incorporate site-specific epigenetic information and without a sacrifice in scale is significant, and is a necessity for overcoming the challenge of synthesizing epigenetically modified genes and genomes.

Methods

For complete experimental methods see Supplementary Information.

One-pot assembly of iLOV gene. The oligonucleotides which comprised the sense and antisense strands of iLOV (F1-F5 and R1-R5, 4 nmol each, Supplementary Table 1) were combined and lyophilized. The oligonucleotides were suspended in 0.2 M NaCl (400 μ L) then annealed by heating at 95°C for 15 min then gradually cooled to room temperature over 2 h; the temperature was reduced by 10°C every 15 min. A Cu^I click catalyst solution was prepared from tris(3-hydroxypropyltriazolylmethyl)amine ligand (4.6 μ mol in 0.2 M NaCl, 154 μ L), sodium ascorbate (7 μ mol in 0.2 M NaCl, 14 μ L) and CuSO₄·5H₂O (0.7 μ mol in 0.2 M NaCl, 7 μ L). The pre-mixed catalyst solution (160 μ L) was added to give the following equivalents per click reaction: 140 eqv. ligand, 200 eqv. ascorbate and 20 eqv. copper. The mixture thoroughly degassed using argon then left at room temperature for 2 h. Formamide (560 μ L) was added and the samples analyzed by denaturing 8% polyacrylamide gel electrophoresis by applying 550 V for 3.5 h. The gel was visualised by placing it on top of a TLC plate and using a UV lamp at 254 nm. Bands corresponding to the assembled gene were excised and extracted from the gel using the ‘crush and soak method’. In brief, the excised polyacrylamide pieces were broken down into small pieces then suspended in distilled water (25 mL). The suspension was shaken at 37 °C for 18 h then filtered through a plug of cotton wool. The filtrate was concentrated to

approximately 2 mL then desalted using through two NAP-25 and one NAP-10 columns. The desalted eluent was lyophilized prior to use.

Ligation of click-assembled iLOV gene into pRSET backbone. The pRSET backbone required for ligation was prepared from the double digestion of pRSET mCherry.³⁴ The plasmid was digested sequentially between the *NdeI* and *EcoRI* restriction sites using enzymes and buffer supplied by New England Biolabs, UK. Restriction digestions were performed in a 50 μ L reaction volume with between 1- 2.5 μ g of plasmid, 10 X CutSmart® buffer (5 μ L) and the restriction enzyme (20 U/ μ g plasmid). The restriction digestion reactions were incubated at 37 °C (60 min/ μ g plasmid) then the 5'-terminus dephosphorylated by addition of shrimp alkaline phosphatase (1 U/ μ g plasmid, New England Biolab, UK). The reactions were incubated at 37 °C for a further 30 min then the enzymes inactivated by incubating at 70 °C for 10 min. The digested plasmid was analyzed by gel electrophoresis using 1% agarose gel in 1X Tris/Borate/EDTA buffer (TBE) by applying 100 V for 30 min. The band corresponding to the backbone excised and the DNA was isolated and purified using GeneJet PCR Purification Kit (Thermo Fisher Scientific). Ligation of click-ligated iLOV gene was performed in a total volume of 10 μ L using 50 ng of pRSET vector, T4 DNA Ligase (1 μ L, 3 U, Promega UK) and T4 DNA Ligase 10X reaction buffer (1 μ L, Promega). The lyophilized click-ligated gene was resuspended in ultrapure water and an aliquot diluted to a concentration of 19.7 ng/ μ L. The click-ligated gene was added to the ligation reaction to give 1:3 backbone to insert molar ratio. Negative control ligations were set up as above, using water instead of insert. The ligation reactions were incubated at 4°C for 16 h then at room temperature for 1 h. The T4 DNA ligase enzyme was subsequently deactivated by heating at 70 °C for 10 min.

Transformation of pRSET-iLOV into *E. coli*. The inactivated ligation reactions were dialysed for 1.5 h against ultrapure water using a 0.025 μ m membrane filter (Millipore, Cat No: VSWP02500). The recovered ligation mixtures (approximately 7 μ L) were added to frozen aliquots of electrocompetent KRX cells (100 μ L). Electroporation of the plasmids was achieved using MicroPulser system (Bio-Rad) using standard protocols. The transformants were immediately recovered using ice-cold SOC medium (890 μ L) then incubated at 37 °C for 1 h. An aliquot (100 μ L) of the recovered cells were spread on LB agar plates supplemented with carbenicillin (100 μ g/ μ L) and rhamnose (0.1 % w/v) then incubated at 37 °C for 18 h. Individual colonies were selected and grown in LB Broth (25 mL) supplemented with carbenicillin (100 μ g/ μ L) at 37 °C for 16 h. Plasmid DNA was extracted from these cells using a QiaPrep® Miniprep kit. The plasmids were sequenced by Eurofins MWG Operon (Ebersberg, Germany) using the T7 forward and reverse primers. Trace files were aligned against reference sequences using Clustal Omega web-based software.

Acknowledgement

We thank Roger Tsien (RIP) for his kind gift of pRSET-mCherry, and Dorcas Brown for access to ATDBio's oligonucleotide synthesis facilities. This work was supported by UK BBSRC grants BB/J001694/2 and BB/M025624/1.

Author Contribution

A.T. conceived, designed and supervised the study. M.K. synthesized the modified DNA bases, conducted click-DNA assembly, carried out all biological experiments, and analyzed data. N.G. and A.H.E.S. conducted oligonucleotide synthesis. A.H.E.S. conducted mass spectrometry and analyzed data. T.B. provided material and supervision for oligonucleotide synthesis. The manuscript was written by M.K. and A.T. with input from all authors.

Data Availability

Data underpinning this study are openly available from the University of Southampton repository at <https://doi.org/10.5258/SOTON/D0103>

Competing Financial Interests

M.K. and A.T. are co-inventors on the US patent application “One-Pot Gene Synthesis”. A.H.E.S. and T.B. are co-inventors on US patent 8,846,883 B2 “Oligonucleotide Ligation”.

References

1. de Kok S, Stanton LH, Slaby T, Durot M, Holmes VF, Patel KG, *et al.* Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. *ACS Synthetic Biology* 2014, **3**(2): 97-106.
2. Hutchison CA, 3rd, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, *et al.* Design and synthesis of a minimal bacterial genome. *Science* 2016, **351**(6280): aad6253.
3. Kosuri S, Church GM. Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods* 2014, **11**(5): 499-507.
4. Smith HO, Hutchison CA, 3rd, Pfannkoch C, Venter JC. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proceedings of the National Academy of Sciences U.S.A.* 2003, **100**(26): 15440-15445.
5. LeProust EM, Peck BJ, Spirin K, McCuen HB, Moore B, Namsaraev E, *et al.* Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Research* 2010, **38**(8): 2522-2540.
6. Hughes RA, Miklos AE, Ellington AD. Gene synthesis: methods and applications. *Methods in Enzymology* 2011, **498**: 277-309.
7. Czar MJ, Anderson JC, Bader JS, Peccoud J. Gene synthesis demystified. *Trends in Biotechnology* 2009, **27**(2): 63-72.
8. Gao X, Yo P, Keith A, Ragan TJ, Harris TK. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Research* 2003, **31**(22): e143.
9. Stemmer WP, Cramer A, Ha KD, Brennan TM, Heyneker HL. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 1995, **164**(1): 49-53.
10. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 2010, **329**(5987): 52-56.
11. Boeke JD, Church G, Hessel A, Kelley NJ, Arkin A, Cai Y, *et al.* GENOME ENGINEERING. The Genome Project-Write. *Science* 2016, **353**(6295): 126-127.

12. Micklefield J. Backbone modification of nucleic acids: Synthesis, structure and therapeutic applications. *Current Medicinal Chemistry* 2001, **8**(10): 1157-1179.
13. Kresse J, Nagpal KL, Nagyvary J, Uchic JT. The use of s-2-cyanoethyl phosphorothioate in the preparation of oligo 5'-deoxy-5'-thiothymidylates. *Nucleic Acids Research* 1975, **2**(1): 1-9.
14. Xu Y, Kool ET. A Novel 5'-Iodonucleoside Allows Efficient Nonenzymatic Ligation of Single-stranded and Duplex DNAs. *Tetrahedron Letters* 1997, **38**(32): 5595-5598.
15. Liu J, Taylor JS. Template-directed photoligation of oligodeoxyribonucleotides via 4-thiothymidine. *Nucleic Acids Research* 1998, **26**(13): 3300-3304.
16. Hansen DJ, Manuguerra I, Kjelstrup MB, Gothelf KV. Synthesis, dynamic combinatorial chemistry, and PCR amplification of 3'-5' and 3'-6' disulfide-linked oligonucleotides. *Angewandte Chemie International Edition in English* 2014, **53**(52): 14415-14418.
17. Xu Y, Kool ET. Chemical and enzymatic properties of bridging 5'-S-phosphorothioester linkages in DNA. *Nucleic Acids Research* 1998, **26**(13): 3159-3164.
18. Cai J, Li X, Yue X, Taylor JS. Nucleic acid-triggered fluorescent probe activation by the Staudinger reaction. *Journal of the American Chemical Society* 2004, **126**(50): 16324-16325.
19. Li X, Liu DR. DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angewandte Chemie International Edition in English* 2004, **43**(37): 4848-4870.
20. Roloff A, Seitz O. Bioorthogonal reactions challenged: DNA templated native chemical ligation during PCR. *Chemical Science* 2013, **4**(1): 432-436.
21. Shibata A, Uzawa T, Nakashima Y, Ito M, Nakano Y, Shuto S, *et al.* Very rapid DNA-templated reaction for efficient signal amplification and its steady-state kinetic analysis of the turnover cycle. *Journal of the American Chemical Society* 2013, **135**(38): 14172-14178.
22. El-Sagheer AH, Sanzone AP, Gao R, Tavassoli A, Brown T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli. *Proceedings of the National Academy of Sciences U.S.A.* 2011, **108**(28): 11338-11343.
23. Litovchick A, Dumelin CE, Habeshian S, Gikunju D, Guie MA, Centrella P, *et al.* Encoded Library Synthesis Using Chemical Ligation and the Discovery of sEH Inhibitors from a 334-Million Member Library. *Scientific Reports* 2015, **5**.
24. Lucas R, Elchinger PH, Faugeras PA, Zerrouki R. Pyrimidine-purine and pyrimidine heterodinucleosides synthesis containing a triazole linkage. *Nucleosides Nucleotides Nucleic Acids* 2010, **29**(3): 168-177.
25. Lucas R, Neto V, Bouazza AH, Zerrouki R, Granet R, Krausz P, *et al.* Microwave-assisted synthesis of a triazole-linked 3'-5' dithymidine using click chemistry. *Tetrahedron Letters* 2008, **49**(6): 1004-1007.
26. Lucas R, Zerrouki R, Granet R, Krausz P, Champavier Y. A rapid efficient microwave-assisted synthesis of a 3',5'-pentathymidine by copper(I)-catalyzed [3+2] cycloaddition. *Tetrahedron* 2008, **64**(23): 5467-5471.
27. Sanzone AP, El-Sagheer AH, Brown T, Tavassoli A. Assessing the biocompatibility of click-linked DNA in Escherichia coli. *Nucleic Acids Research* 2012, **40**(20): 10567-10575.
28. Birts CN, Sanzone AP, El-Sagheer AH, Blaydes JP, Brown T, Tavassoli A. Transcription of click-linked DNA in human cells. *Angewandte Chemie International Edition in English* 2014, **53**(9): 2362-2365.
29. el-Sagheer AH, Brown T. Efficient RNA synthesis by in vitro transcription of a triazole-modified DNA template. *Chemical Communications* 2011, **47**(44): 12057-12058.

30. Dallmann A, El-Sagheer AH, Dehmel L, Mugge C, Griesinger C, Ernsting NP, *et al.* Structure and dynamics of triazole-linked DNA: biocompatibility explained. *Chemistry* 2011, **17**(52): 14714-14717.
31. Chapman S, Faulkner C, Kaiserli E, Garcia-Mata C, Savenkov EI, Roberts AG, *et al.* The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proceedings of the National Academy of Sciences U.S.A.* 2008, **105**(50): 20038-20043.
32. El-Sagheer AH, Brown T. New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proceedings of the National Academy of Sciences U.S.A.* 2010, **107**(35): 15329-15334.
33. Kypr J, Kejnovská I, Renčiuk D, Vorlíčková M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research* 2009, **37**(6): 1713-1725.
34. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnology* 2004, **22**(12): 1567-1572.
35. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* 2006, **2**: 2006 0008.
36. Hsu DS, Kim ST, Sun Q, Sancar A. Structure and Function of the UvrB Protein. *Journal of Biological Chemistry* 1995, **270**(14): 8319-8327.
37. Truglio JJ, Croteau DL, Van Houten B, Kisker C. Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical Reviews* 2006, **106**(2): 233-252.
38. Merryman C, Gibson DG. Methods and applications for assembling large DNA constructs. *Metabolic Engineering* 2012, **14**(3): 196-204.
39. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* 2014, **344**(6179): 55-58.
40. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics* 2007, **39**(4): 457-466.
41. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* 2013, **500**(7463): 477-481.
42. Nielsen PE. DNA analogues with nonphosphodiester backbones. *Annual Review of Biophysics and Biomolecular Structure* 1995, **24**: 167-183.
43. Pinheiro VB, Holliger P. The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Current Opinion in Chemical Biology* 2012, **16**(3-4): 245-252.
44. Sergueev D, Hasan A, Ramaswamy M, Shaw BR. Boranophosphate oligonucleotides: New synthetic approaches. *Nucleosides and Nucleotides* 1997, **16**(7-9): 1533-1538.
45. Shivalingam A, Tyburn AES, El-Sagheer AH, Brown T. Molecular requirements of high-fidelity replication-competent DNA backbones for orthogonal chemical ligation. *Journal of the American Chemical Society* 2017.

Figure Legends:

Figure1: One pot gene synthesis by click-DNA ligation. (a) Schematic representation of our one-pot click-ligation strategy. (b) CuAAC reaction between adjacent terminal cytidine and thymidine residues giving rise to a 1,4-linked 1,2,3-triazole. (c) Our approach is exemplified with the assembly of a gene encoding iLOV. The sequences of the individual oligonucleotides used for the assembly of the whole gene are shown. The positions of the triazole linkages in the sense and antisense strands are highlighted by the orange and yellow boxes respectively and restriction sites are in italics.

Figure 2. Synthesis of alkyne-modified deoxycytidine and loading onto solid support in preparation for oligonucleotide synthesis. N⁴-acetyl-2'-deoxy-5'-O-DMT cytidine was alkylated using propargyl bromide in the presence of NaH in THF; Methanol was subsequently added to quench the residual propargyl bromide and simultaneously remove the acetyl protecting group. The resulting alkyne-modified deoxycytidine was coupled to an amino functionalized CPG resin that had been treated with succinic anhydride prior to coupling.

Figure 3: Characterisation of click-ligated iLOV gene. (a) PAGE analysis of the constituent modified oligonucleotides in the absence of Cu^I (Lane 1) and the product of the click reaction (Lane 2), visualized under UV light. The position of the click-iLOV product is indicated. In Lane 2, the lower molecular weight bands have been attributed to cyclized and truncated products. The difference in band intensity between lanes 1 and 2 is likely due to DNA loss upon removal of Cu(II) using a NAP-5 column. (b) Agarose gel of the purified click-linked iLOV gene (lane 2), which shows a 350 bp band (ladder in lane 1); inserted image shows a vial of the purified click-linked iLOV gene. (c) CD spectra of canonical and click-iLOV genes (d) PCR amplification of click-iLOV gene. Lanes 2 and 4 show control experiments in which the unligated modified oligonucleotides were used as template for amplification by *Taq* and *Pfu* DNA polymerases. Lanes 3 and 5 show the results of amplification of the click-linked iLOV gene by *Taq* and *Pfu* DNA polymerases. In both lanes, a band was observed between 300-350 bp, corresponding to the 335 bp click-ligated iLOV gene. Lane 1 is a DNA Ladder. (e) Primer extension analysis of the sense strand of click-iLOV gene using Klenow fragment (exo)⁻ DNA polymerase. Lane 1: negative control in which the enzyme was omitted from the reaction. The positions of the template and primer are indicated. Lane 2: primer extension reaction; the primer was extended to the same length as the template. (f) Quantitative real-time PCR (qPCR) analysis of the effect of triazole linkers on DNA replication. Amplification of fragments of click-linked iLOV gene (green) containing increasing numbers of triazoles and their canonical equivalent (black) were assessed by qPCR. The binding sites of the forward primer have been represented schematically above each graph, with replication through (i) one, (ii) two, (iii) three and (iv) four triazoles assessed. Reverse primer is not shown, but anneals to the 3'-terminus of the upper strand. All data presented as mean \pm SD, n=3.

Figure 4: Assembly of click-ligated iLOV plasmid. The iLOV gene was assembled from functionalized oligonucleotides by click DNA ligation; the gene was designed so that it contained ‘sticky ends’ ready for ligation into a plasmid backbone cleaved by *NdeI* and *EcoRI* restriction endonucleases. The pRSET-mCherry plasmid, digested with *NdeI* and *EcoRI* to excise the mCherry gene (encoding a red fluorescent protein) was used as the backbone. The click-linked iLOV gene (encoding a green fluorescent protein) was ligated into this backbone using T4 DNA ligase.

Figure 5: Biocompatibility of click-linked iLOV in *E. coli* cells. (a) *E. coli* cells were transformed with plasmids containing click-linked iLOV or ligase-assembled iLOV, and incubated on LB-agar plates for 16 h. The top image was captured under white light. The bottom image shows iLOV (green, 488 nm) and mCherry (red, 610 nm) fluorescence on the same plate, with the green and red fluorescence images merged. Colonies containing iLOV or mCherry are readily identified (b) Percentage of colonies displaying the green, white and red phenotypes when cells were transformed with plasmids containing the click-ligated or ligase-assembled iLOV. The total number of colonies assessed is shown. Plasmids were isolated from 50 colonies displaying green fluorescence for each type of insert, and submitted for DNA sequencing. The percentage of green colonies containing the correct sequence for iLOV is shown beneath the graph. All data are presented as mean \pm SD, n=15. (c) Representative sequence electropherogram from a single colony of KRX *E. coli* transformed with click-linked iLOV. The click-linked gene has been correctly replicated in the bacteria and is error free. The red boxes highlight the position of the triazole-linked CT dinucleotides in the sense strand, while the yellow boxes highlight triazole-linked CT dinucleotides in the antisense strand.

Figure 6: click-assembly and characterization of epigenetically modified iLOV. (a) A click-linked iLOV gene containing 5-methylcytosines (5-meC) within the two CGCG motifs on each strand (5-meC represented by purple lines) was assembled. iLOV contains three *Acil* restriction sites (represented by green dotted line), however the first *Acil* site overlaps the CGCG motif and restriction digestion will be blocked by incorporation of 5-meC. (b) Agarose gel of the purified 5-meC-containing iLOV shows a 350 bp band (ladder in lane 1). (c) Mass spectrum of the *Acil*-digested click-linked iLOV; digestion fragment labels correspond to those in panel A. (d) Mass spectrum of the *Acil*-digested click-linked 5-meC containing iLOV; digestion fragment labels correspond to those in panel A. (e) Agarose gel of the fluorophore-labeled iLOV gene shows a product of the expected size, and expected FRET. Lane 1 is ladder; lane 2 is the multiplex image (overlay of UV and all fluorescent channels); lane 3 shows the gene when excited at 460 nm, and imaged with a 520 nm filter, showing little fluorescence from the Oregon Green 488 dye; lane 4 shows the gene when excited at 460 nm and imaged with a 570 nm filter, the bright fluorescent band is due to Cy3 and indicative of FRET between the fluorophores on each strand of the iLOV gene. (f) Fluorescence spectrum of the iLOV gene containing two fluorophores illustrates FRET associated with duplex formation. The ^{Fluo}R1 oligonucleotide is tagged with Cy3, while the ^{Fluo}F5 oligonucleotide is tagged with Oregon Green 488 (Supplementary Table 1).