Sequence-dependence of Cy3 and Cy5 dyes in 3’ terminally-labeled single-stranded DNA

Tadija Kekić & Jory Lietard

Fluorescence is an ideal tool to see and manipulate nucleic acids, and engage in their rich and complex biophysical properties. Labeling is the preferred approach to track and quantify fluorescence with nucleic acids and cyanine dyes are emblematic in this context. The fluorescent properties of cyanine dyes are known to be sequence-dependent, with purines in the immediate vicinity increasing the fluorescence intensity of Cy3 and Cy5 dyes, and the ability of nucleobases to modulate the photophysical properties of common fluorophores may influence fluorescence measurements in critical assays such as FISH, qPCR or high-throughput sequencing. In this paper, we comprehensively map the sequence-dependence of Cy3 and Cy5 dyes in 3’-fluorescently labeled single-stranded DNA by preparing the complete permutation library of the 5 consecutive nucleotides immediately adjacent to the dye, or 1024 sequences. G-rich motifs dominate the high fluorescence range, while C-rich motifs lead to significant quenching, an observation consistent with 5’-labeled systems. We also uncover GCGC patterns in the extreme top range of fluorescence, a feature specific to 3’-Cy3 and Cy5 oligonucleotides. This study represents the final piece in linking nucleotide identity to fluorescence changes for Cy3, Cy5 and fluorescein in all 3’, 5’, single-stranded and double-stranded DNA formats.

Fluorescence labeling of nucleic acids is a ubiquitous method to detect, monitor, quantify and study DNA and RNA, in vitro and in vivo1-4. Far more convenient than radiolabeling and far more sensitive than the UV absorbance of nucleobases, fluorescence detection is commonplace for nucleic acid quantification, particularly qPCR, a staple in molecular biology that has become omnipresent ever since the outbreak of the SARS-CoV-2 virus5. Of the plethora of available fluorescent tags, cyanine dyes are perhaps the most common due to their high brightness, molar extinction coefficient and chemical stability6,7. Labeling is typically carried out at the 5’ or 3’ end of the oligonucleotide and at times on both extremities for FRET studies on molecular beacons8,9. The covalent attachment of fluorescent tags to nucleic acids brings the aromatic fluorophore in close proximity to the terminal nucleobase, < 1 nm, making it possible to foresee electronic interactions between the two10. And indeed, the fluorescence of common cyanine—but also xanthene—dyes is known to be sequence-dependent11. Nucleobases in close proximity to the dye are able to quench or enhance fluorescence through a mechanism involving the modulation of the energy barrier associated with rotational isomerization6,10-13. In particular, the loss of rotational freedom in the polyaromatic structure of fluorescent dyes is thought to be directly correlated with enhanced fluorescence properties and this chemical rigidity can be the product of π-π stacking14,15 where purines are expected to play a greater role. The ability of nucleobases to modulate the fluorescence properties of nearby dyes creates a quantitation bias where equimolar solutions of two different sequences as measured by A260 absorbance yields inconsistent fluorescence emission values. Other approaches relying on fluorescence detection such as next generation sequencing are likely to be affected by sequence-specific effects as well16.

We have extensively studied the sequence-dependence of Cy3, Cy5, fluorescein (FAM) and other cyanine derivatives by synthesizing complex DNA microarrays using photolithography, where all possible permutations of five consecutive nucleotides (45 possible combination, 1024 unique sequences) immediately adjacent to a terminal fluorophore are available17-20. Our results have showed that 5’-guanine positively affects the fluorescence intensity of Cy3 and Cy5 but quenches FAM, whereas 5’-pyrimidines yield lower Cy3 and Cy5 fluorescence but tend to enhance FAM intensity. Interestingly however, we found that in the case of FAM, 5’ and 3’ labeling yielded different motifs of sequence-dependence, with the fluorescence of 3’-FAM oligonucleotides being affected by

Institute of Inorganic Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria. *email: jory.lietard@univie.ac.at
guanines further away from the 3′-end, in some cases up to 4 nucleotides away from the dye. In addition, 5′ and 3′ FAM labeling yields different ranges of fluorescence intensities across all 1024 permutations. These observations prompted us to investigate the sequence-dependence of Cy3 and Cy5 dye in the context of 3′ terminally labeled single-stranded oligonucleotides, with the same 5-nt permutation scheme yielding 1024 unique combinations. To do so and to be able to perform a terminal labeling of oligonucleotide at the 3′ end, the photolithographic synthesis of DNA microarrays is carried out in the 5′→3′ direction using a set of 3′-photoprotected 5′-phosphoramidites, a method which was recently shown to be as efficient as conventional 3′→5′ nucleic acid synthesis21. We find that the sequence-dependence of 3′-Cy3 and 3′-Cy5 dyes is largely similar to their 5′ counterparts, with 3′dG dominating the high fluorescence range and 3′dC-rich sequences up to 70% less fluorescent than 3′dG sequences. But we also note unusually high fluorescence for 3′-GCGC patterns which were absent in the 5′ labeled datasets. These results should serve to finalize a comprehensive understanding of the sequence dependence of Cy3 and Cy5 dyes in all possible contexts, from 5′ to 3′ to single and double-stranded DNA.

Results and discussion

We started by studying whether Cy3 and Cy5 attached to the 3′ termini of single-stranded oligonucleotides are differently affected by sequence identity in comparison to 5′ labeled single strands. To do so, we synthesized sequence permutations over five consecutive nucleotides at the 3′ end, which amounts to 4^5 or 1024 permutations. Using reverse (5′→3′) nucleic acid synthesis by microarray photolithography and with a final coupling of Cy3 or Cy5 phosphoramidite (Fig. 1A). Reverse synthesis is very efficient (≥98% stepwise efficiency for 3′-BzNPPOC 5′-DNA phosphoramidites, Fig. 1B) and with coupling failures being systematically capped, labeling can only occur on full-length oligonucleotides. To account for the slightly lower coupling efficiency of reverse dG, all oligonucleotide sequences are designed in such a way that they all contain the same amount of G nucleotides, regardless of sequence combination. All sequences being of same length, same distance to the surface and simultaneously synthesized on the same array, the relative fluorescence of each library element can be compared to one another. In addition, the microarray and oligonucleotide sequence design mirrors that of the sequence dependence study on 5′-Cy3/Cy5-labeled DNA, allowing for the two datasets to be compared.

We indeed find that the fluorescence intensity of Cy3 and Cy5 is heavily affected by the nature of adjacent nucleotides in the context of 3′ terminal labeling (Fig. 1C). We ranked the fluorescence intensity of Cy3 and
Cy5-labeled libraries and the distribution adopts a sigmoidal shape in both cases with very large disparities in signal intensity, with a loss of up to 65 and 75% fluorescence relative to the brightest sequence for Cy3 and Cy5, respectively (Fig. 2B). This intensity span is larger than for 5′-Cy3 and Cy5 labeled single-stranded oligonucleotides, where losses of only 50 and 65% were recorded. This larger span of fluorescence intensity in 3′-labeled ssDNA may be explained by the presence of a small number of permutations displaying unusually high brightness in both cyanine cases, as illustrated by the asymmetrical shape of the sigmoidal distribution (Fig. 2). Indeed, the extreme top range of fluorescence intensity is here populated with 5–6 sequence combinations that far outperform the rest of the permutation library.

These combinations are always of the form 3′-GCGCX and all possible X (A, C, G and T) make up the four brightest Cy3 and Cy5-labeled oligonucleotides (relative fluorescence from 0.87 to 1) and occupying the entire set of sequences in the first octile of Cy3 and Cy5 fluorescence (Fig. 3A,B), hence the very uniform GCGC sequence.
motif. This observation is particularly surprising in many respects. First, the GCGC pattern is a rare example of a multi-nucleotide motif that influences the fluorescence properties of a nearby dye, when most other systems were found to be primarily affected by the nature of the neighboring base only. Second, this motif introduces two cytidines in the top fluorescence range, when the presence of C is usually synonymous with much lower fluorescence. This can be seen in the low section of Cy3 and Cy5 fluorescence which is associated with C-rich sequences. Third, such a motif was entirely absent from the top-ranking fluorescence in 5ʹ-labeled systems. Finally, this ranking holds true even before deprotection of the DNA microarray, where the nucleobases still carry a base protecting group preventing correct Watson–Crick pairing. The formation of intermolecular G•C base pairs between neighboring strands on the same microarray feature should not occur based on the mean distance between hydroxyl groups at the surface of the glass slide. Similarly, intramolecular hairpin formation with G•C base pairs is unlikely to take place considering the fact that the formation of a loop would be impractical at best. Indeed, with the pattern 3ʹ-GCGCG for instance, the oligonucleotide synthesized on the microarray has the sequence 3ʹ-GCGCGTTTTTTTTTACTACTACTACT. Only a single G•C base pair with a 2-nt long G-C loop can occur within the pentamer, and there is no complementarity at all with the adjustment section ACT ACTACTACTACT. The presence of C nucleotides in high fluorescence range is not unheard of and was observed in 5ʹ-FAM labeled double-stranded DNA, which suggests that loss of rotational motion at the nucleobase level, like after duplex formation, makes it possible to observe fluorescence enhancement with terminal cytidines. Remarkably, the GCGC motif also yields bright fluorescence signals when in the form of 3ʹ-XGCGC, even with X as a terminal pyrimidine. For instance, the 3ʹ-GCGGC combination yields a normalized fluorescence intensity

Figure 3. Sequence-dependent variations in the fluorescence intensity of 3ʹ Cy3- and Cy5-labeled single-stranded oligonucleotides. The intensity range in Figure 2 was divided into 8 equal parts from which consensus sequences were generated and the sequence logos for each octile arranged in descending order of fluorescence intensity (left to right), for 3ʹ-Cy3 (A) and 3ʹ-Cy5 labeling (B).
of 0.76 (Cy3) and 0.66 (Cy5), making it the brightest possible combination with a terminal 3'-C, the next instance of any combination in the form of 3'-CXXXX being more than 10% darker than CGCGC.

These XGC(CG) motifs suggest that the identity of the final nucleobase may not be critical for high fluorescence. Indeed, in the 2nd–4th octiles of Cy3/Cy5 fluorescence (Fig. 3A,B), the identity of the second nucleobase appears to be as important as the terminal base and G nucleotides at the penultimate position occupy a large part of the high Cy3 and Cy5 signal intensity. In the second octile of Cy5 fluorescence (Fig. 3B), the 4th position still shows strong G preference, as strong as the 2nd position, which is likely a consequence of the presence of the GGC motif, a trimmed down version of the GC(G)C that can still enhance the fluorescence properties of Cy5, though less so for Cy3.

We then looked for patterns further away from the two terminal nucleotide positions. To do so, we studied how, in a system that can be represented in the form 3'-X1X2X3X4X5, any given X1X2X3 trinucleotide combination affects the fluorescence intensity of each possible X1X2. We performed a standardization of the fluorescence signals, with a mean value of 0 representing little to no effect. Data above the zero line indicates a positive effect –increase– on fluorescence intensity, and data below the line a negative effect on fluorescence intensity. The results, shown in Fig. 4, further mark the positive influence of GCG systems on fluorescence brightness. Indeed, all 3'-GCGCX combinations tower very significantly above any other XXGCX system. Visibly too, the CGC trinucleotide as X1X2X3 drives the fluorescence signals up only for 3'-XG dinucleotides, a pattern which concatenates into 3'-XGCGC. While this effect could be seen in pure fluorescence rankings (Fig. 2), this particular trinucleotide contrasts with all other in the form CXX which show an overall negative effect on fluorescence intensity. This observation is all the more apparent when comparing to GXX trinucleotides which, in almost all cases, have at least a slight positive effect on fluorescence intensity. Taken together, the data suggests that nucleotide identity three units away from the dye can still be relevant in the context of modulating the fluorescence properties of terminal dyes. Whether the influence of nucleotide X3 is due to electronic effects propagating through X1 and X2, or the result of a direct interaction between 3'-Cy3 and X3 is unclear at this point.

Fluorescence enhancement is thought to be directly related to loss of rotational isomerization in cyanine dyes, an effect that would be promoted by stronger stacking systems such as purine rich elements. On the other hand, pyrimidines tracts stack less efficiently, and cytidine less so than thymidine according to free energy measurements24,25, which explains why terminal Cs are associated with lower overall fluorescence. Our study focuses on single-stranded DNA, but complementarity between the permutation and "adjustment" sections (Fig. 1A) cannot be excluded. However, we expect hairpin folding to occur for only a few random candidates (Supplementary Table S1) and none of the brightest or darkest sequences can do so. The formation of secondary structures for a small number of sequence permutations should not impact the analysis as it largely stems from pattern search in aggregated data. Indeed, the intensity rankings for deprotected and non-deprotected DNA, where no Watson–Crick structures are possible, overlap fairly well. Our results are consistent with previous observations on the modulation of Cy3/Cy5 fluorescence intensity by neighboring nucleobases, where nucleotide dependency was found to follow the order dG > dA > dT > dC17,29,25. The fact that guanines do not quench Cy3 and Cy5 fluorescence is in stark contrast to its effect on other common fluorophores like fluorescein where G-mediated quenching is a known phenomenon20,26,27. There, the redox potential of the nucleobase is directly correlated with its ability to quench the fluorescence of the conjugated dye via photoinduced electron transfer, an effect that has recently been described in Cy3-labeled double-stranded DNA28. In particular, the fluorescence of Cy3 was found to be dependent on the oxidation potential of the stacking base-pair, with stacking and base-pair formation able to affect (and indeed lower) the oxidation potential of G26–31. These effects have not been identified in ssDNA but transient base-stacking interactions in ssDNA cannot be formally excluded and the resulting shifts in redox potential may then contribute to the observed sequence dependence.
The additive effect of multiple Gs or Cs in a row is not immediately apparent, with GGGGG ranking #31 and #21 (0.71 and 0.64 relative intensity) and CCCCC ranking #1009 and #1021 (0.37 and 0.25 relative intensity) for Cy3 and Cy5, respectively. The top-ranking GAAAA motif identified in 5ʹ-Cy3 oligonucleotides here ranks #334 (0.58 relative fluorescence), further illustrating the significant differences between 5ʹ and 3ʹ labeling.

Overall, the dominance of G and C in high and low fluorescence regime is very apparent in both Cy3 and Cy5 cases, as can be seen in Table 1. In the bottom section of both Cy3 and Cy5 fluorescence, the last 83 and 89 pentamers are all in the form 3ʹ-CXXXX and 75% and 98% (Cy3 and Cy5, respectively) of such motifs are under 0.5 relative fluorescence intensity. In contrast, 89% of all 3ʹ-GXXXX are above 0.5, but only 25% in the case of 3ʹ-CXXXX. G-rich sequences (> 60% of G content per pentamer) are above 0.5 relative fluorescence 99% and 60% of the time.

**Conclusion**

In this study, we looked at the sequence-dependence of 3ʹ-Cy3 and Cy5-labeled single strands by designing a library of oligonucleotides with all possible sequence permutations of the five terminal nucleotides at the 3ʹ end. The sequence dependence itself is not unlike 5ʹ-labeled Cy3/Cy5 DNA, with G-rich sequences populating high fluorescence signals and, conversely, C-rich sequences quenching Cy3 and Cy5. There are however noticeable differences, with the presence of the 3ʹ-GCGCX motif dominating all permutations (Table 2). We also found that the range of fluorescence change is larger than for the corresponding 5ʹ labeled sequences, mostly due to the very high fluorescence intensities recorded for all 3ʹ-GCGCX and all 3ʹ-XGCGC combinations. While nucleotide identity at the terminal position is a decisive factor for Cy3 and Cy5 fluorescence intensity, nucleobases further away from the dye can affect fluorescence signals. This is particularly true with dG which can still positively influence Cy3 fluorescence, regardless of the nature of the first two nucleotides at the 3ʹ end. Overall, this study completes the investigation on the sequence-dependence of common nucleic acid fluorophores (Cy3, Cy5 and

| Rank | 5ʹ-Cy3 sequence | Normalized fluorescence | 3ʹ-Cy3 sequence | Normalized fluorescence | Rank | 5ʹ-Cy5 sequence | Normalized fluorescence | 3ʹ-Cy5 sequence | Normalized fluorescence |
|------|----------------|------------------------|----------------|------------------------|------|----------------|------------------------|----------------|------------------------|
| 1    | GGGGG          | 1.909 ± 0.024          | GGGGG          | 1.090 ± 0.035          | 1005 | CGCC          | 0.373 ± 0.022          | CGCC           | 0.272 ± 0.016          |
| 2    | GGGCA          | 0.973 ± 0.017          | GGGCT          | 0.964 ± 0.064          | 1006 | CTTT          | 0.373 ± 0.032          | CTTT           | 0.272 ± 0.035          |
| 3    | GGGCC          | 0.961 ± 0.026          | GGGCC          | 0.944 ± 0.034          | 1007 | CCCC          | 0.370 ± 0.028          | CCCC           | 0.272 ± 0.012          |
| 4    | GGGCT          | 0.930 ± 0.023          | GGGCA          | 0.876 ± 0.027          | 1008 | CCAA          | 0.369 ± 0.026          | CCAA           | 0.272 ± 0.020          |
| 5    | AGGGG          | 0.823 ± 0.027          | GGGCC          | 0.793 ± 0.047          | 1009 | CCCC          | 0.369 ± 0.018          | CCCC           | 0.271 ± 0.016          |
| 6    | GGGGG          | 0.814 ± 0.019          | AGGGG          | 0.792 ± 0.034          | 1010 | CGAC          | 0.369 ± 0.023          | CGAC           | 0.271 ± 0.019          |
| 7    | AGGGG          | 0.766 ± 0.033          | AGGGG          | 0.757 ± 0.038          | 1011 | CTTT          | 0.363 ± 0.017          | CTTT           | 0.271 ± 0.027          |
| 8    | TGCCG          | 0.783 ± 0.029          | AGGGT          | 0.725 ± 0.054          | 1012 | CTTT          | 0.362 ± 0.030          | CTTT           | 0.269 ± 0.018          |
| 9    | AGGAA          | 0.779 ± 0.020          | AGGGG          | 0.715 ± 0.052          | 1013 | CCCC          | 0.358 ± 0.016          | CCCC           | 0.268 ± 0.019          |
| 10   | AGGGG          | 0.773 ± 0.024          | AGGGC          | 0.714 ± 0.045          | 1014 | CTTT          | 0.357 ± 0.036          | CTTT           | 0.266 ± 0.016          |
| 11   | GGGAG          | 0.759 ± 0.020          | TGCCG          | 0.697 ± 0.048          | 1015 | CTTT          | 0.354 ± 0.016          | CTTT           | 0.266 ± 0.019          |
| 12   | GGGCC          | 0.759 ± 0.032          | AGGGG          | 0.689 ± 0.033          | 1016 | CTTT          | 0.351 ± 0.020          | CTTT           | 0.265 ± 0.016          |
| 13   | AGGGG          | 0.757 ± 0.010          | AGGGG          | 0.669 ± 0.059          | 1017 | CTTT          | 0.351 ± 0.018          | CTTT           | 0.265 ± 0.023          |
| 14   | AGGGG          | 0.753 ± 0.014          | AGGGG          | 0.660 ± 0.052          | 1018 | CTTT          | 0.350 ± 0.010          | CTTT           | 0.260 ± 0.019          |
| 15   | AGAGG          | 0.747 ± 0.023          | AGGGG          | 0.657 ± 0.043          | 1019 | CTTT          | 0.346 ± 0.012          | CTTT           | 0.257 ± 0.015          |
| 16   | AGATG          | 0.747 ± 0.018          | AGGGG          | 0.655 ± 0.076          | 1020 | CTTT          | 0.345 ± 0.016          | CTTT           | 0.253 ± 0.015          |
| 17   | TGCCG          | 0.743 ± 0.026          | AGGGG          | 0.650 ± 0.045          | 1021 | CTTT          | 0.342 ± 0.030          | CTTT           | 0.251 ± 0.018          |
| 18   | TGCCG          | 0.741 ± 0.023          | AGGGT          | 0.648 ± 0.024          | 1022 | CTTT          | 0.335 ± 0.042          | CTTT           | 0.247 ± 0.016          |
| 19   | AGGGG          | 0.734 ± 0.076          | AGGGG          | 0.642 ± 0.037          | 1023 | CTTT          | 0.330 ± 0.022          | CTTT           | 0.243 ± 0.015          |
| 20   | GGGGG          | 0.731 ± 0.012          | AGGGG          | 0.630 ± 0.054          | 1024 | CCCC          | 0.320 ± 0.030          | CCCC           | 0.229 ± 0.015          |

Table 1. Normalized relative fluorescence intensity of the 20 brightest and darkest pentamers 3ʹ-Cy3 and Cy5 labeled. SEM is standard error of the mean.

| Observations | 3ʹ-Cy3 | 3ʹ-Cy5 |
|--------------|--------|--------|
| Fluorescent range (norm.) | 1–0.35 (65% variation) | 1–0.25 (75% variation) |
| Most fluorescent sequence (#1) | 3ʹ-GCCCG | 3ʹ-GCCCG |
| 3ʹ-GCGCX score | Top 4 most fluorescent | Top 4 most fluorescent |
| 3ʹ-XGCGC score | #5, 6, 8, 12 | #5, 6, 11, 14 |
| 20-most fluorescent sequences | G-rich | G-rich |
| 20-least fluorescent sequences | C-rich | C-rich |
| 3ʹ-CCCCX score | #1009, 1013, 1023, 1024 | #1007, 1011, 1021 |
| Least fluorescent sequence (#1024) | 3ʹ-CCCCC | 3ʹ-CCTTC |

Table 2. Summary of the most relevant findings and observations in the sequence-dependence of 3ʹ-Cy3 and Cy5 labeled single-stranded DNA.
Methods

Sequence design. The design of DNA oligonucleotides is based on a permutation scheme allowing for all possible combinations of 5 consecutive nucleotides immediately adjacent to a fluorescent dye to be synthesized in parallel as microarrays (P→P→ Fig. 1). The total amount of unique pentanucleotides is 4^5, or 1024 oligonucleotides to be synthesized. To account for potential variability in synthesis efficiency which would produce lower fluorescence signals for poorly-synthesized sequences, the nucleotide content in all DNA sequences is kept constant by adding a subsection composed of five "N" trinucleotides (N1 to N5) where each N4 corresponds to AGCT minus the nucleotide in P. Between the P and N sections, a T15 spacer is introduced and the entire oligotary material S1.

Microarray synthesis. The process of microarray synthesis by MAS has been extensively described elsewhere32-38. Briefly, DNA synthesis takes place on a silane-functionalized glass slide (Schott Nexterion D) and proceeds according to phosphoramidite chemistry. The major change from conventional solid-phase synthesis is the use of a photosensitive BzNPPOC protecting group instead of the acid-labile DMT, which is removed with 365 nm UV light. To control oligonucleotide elongation and spatial distribution across the surface, patterned UV light is generated by a Digital Micromirror Device (DMD, Texas Instruments) and imaged onto the surface of the slide. DMD mirrors turned OFF leave the corresponding area of the slide which in turns removes the BzNPPOC protecting group, allowing for nucleoside coupling to take place on photodeprotected areas only. DMD mirrors turned OFF leave the corresponding area on the surface of the array unaffected by UV. Photodeprotection time and efficiency is accelerated by submerging the synthesis area with a 1% solution of imidazole in DMSO. Oligonucleotide synthesis proceeds in the 3’→5’ direction using 3'-BzNPPOC protected 5’-DNA phosphoramidites (Orgentis) which were prepared as 30 mM solutions in dry acetonitrile (ACN) and coupled for 60 s in 0.25 M dicyanoimidazole in ACN. Cy3 and Cy5 phosphoramidites (LinkTech) are prepared as 50 mM solutions in ACN and coupled for 2 × 5 min. Before each coupling cycle, the ON areas are exposed for 6 J/cm^2 in order to reach ~99% photodeprotection efficiency. After coupling, unreacted hydroxyl groups are capped by coupling a DMTr-dT phosphoramidite (Sigma-Aldrich) for 60 s. This ensures that the final dye can only be incorporated onto full-length oligonucleotides. Together with high coupling and high photolysis efficiency, terminal 3’ labeling yields a highly uniform oligonucleotide surface density. The delivery of reagents and solvents is accomplished with an Expedite automated DNA synthesizer (PerSeptive Biosystems) which is attached to the reaction chamber encasing the microscope slides.

Data extraction and analysis. After synthesis, the microarrays are immediately washed in 30 ml ACN for 2 h at r.t. to remove unbound fluorescent phosphoramidite. After drying, the arrays are scanned on a GenePix 4400A or 4100A scanner at 2.5 or 5 µm resolution using the appropriate laser for Cy3 or Cy5 excitation (532 and 635 nm, respectively). The arrays are then transferred into a 1:1 solution of ethylenediamine in ethanol and left to react overnight at r.t. to effect oligonucleotide deprotection. After washing with deionized water (2 × 20 ml), the arrays are dried and scanned again. Data is then extracted from the corresponding scans using NimbleScan software (NimbleGen) and analyzed using Excel. The fluorescence intensity values are calculated as an average of all 10 replicates of each permutation per array to which background fluorescence was subtracted. Background fluorescence corresponds to the fluorescence of non-labeled versions of each oligonucleotide, where the dye binds in a non-specific manner. The reported final fluorescence is an average across at least 3 independent syntheses. Standardization (z) of fluorescence signals was performed on background-corrected values (x) by applying \( z = (x - \mu) / \sigma \) where \( \mu \) is the average fluorescence intensity across all permutations and \( \sigma \) the standard deviation of the fluorescence dataset. The consensus sequences were obtained by dividing the range of fluorescence into bins spanning equal intensity, the content of each bin was fed to a sequence logo generator (WebLogo, http://weblogo.berkeley.edu/) and the corresponding consensus sequences arranged per bin in decreasing order of fluorescence intensity. Fluorescence intensity data for all permutations is available as part of the supplementary material S1.

Data availability

Supplementary information with the relative fluorescence intensity data for all experimental sequences in spreadsheet format is available for this paper. Microarray scans and raw fluorescence signals are available from the corresponding author upon reasonable request.

Received: 20 June 2022; Accepted: 24 August 2022 
Published online: 31 August 2022
References
1. Ranasinghe, R. T. & Brown, T. Ultrasensitive fluorescence-based methods for nucleic acid detection: Towards amplification-free genetic analysis. Chem. Commun. 47, 3717–3735 (2011).
2. Michel, B.Y., Dzubiha, D., Benhidja, R., Demchenko, A.P. & Burger, A. Probing of nucleic acid structures, dynamics, and interactions with environment-sensitive fluorescent labels. Front. Chem. 8 (2020).
3. Dirks, R. W. & Tanke, H. J. Advances in fluorescent tracking of nucleic acids in living cells. Biotechniques 40, 489–496 (2006).
4. Kaur, G., Lewis, J.S. & van Oijen, A.M. Shining a Spotlight on DNA: Single-Molecule Methods to Visualise DNA. Molecules 24 (2019).
5. Zhu, H.L., Fohlerova, Z., Pekarek, J., Basova, E. & Neuzil, P. Recent advances in lab-on-a-chip technologies for viral diagnosis. Biosens. Bioelectron. 153 (2020).
6. Levitus, M. & Ranjit, S. Cyanine dyes in biophysical research: the photophysics of polymethylene fluorescent dyes in biomolecular environments. Q. Rev. Biophys. 44, 123–151 (2011).
7. Mujumdar, R. B., Ernst, L. A., Mujumdar, S. R., Lewis, C. J. & Waggoner, A. S. Cyanine dye labeling reagents—sulfonocyanine Succinimidyl Esters. Bioconjugate Chem. 4, 103–111 (1993).
8. Perez-Gonzalez, C., Lafontaine, D.A. & Penedo, J.C. Fluorescence-Based Strategies to Investigate the Structure and Dynamics of Aptamer-Ligand Complexes. Front. Chem. 4 (2016).
9. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. Nat. Methods 5, 507–516 (2008).
10. Sanborn, M. E., Connolly, B. K., Gurunathan, K. & Levitus, M. Fluorescence properties and photophysics of the sulfonocyanine Cy5 linked covalently to DNA. J. Phys. Chem. B 111, 11064–11074 (2007).
11. Harvey, B. J., Perez, C. & Levitus, M. DNA sequence-dependent enhancement of Cy3 fluorescence. Photochem. Photobiol. Sci. 8, 1105–1110 (2009).
12. Huang, Z. X. et al. Spectral identification of specific photoswitches of Cy5 by means of ensemble and single molecule measurements. J. Phys. Chem. A 110, 45–50 (2006).
13. Widgren, J. & Schwille, P. Characterization of photoinduced isomerization and back-isomerization of the cyanine dye Cy5 by fluorescence correlation spectroscopy. J. Phys. Chem. A 104, 6416–6428 (2000).
14. Norman, D. G., Grainger, R. J., Uhrin, D. & Lilley, D. M. J. Location of cyanine-3 on double-stranded DNA: Importance for fluorescence resonance energy transfer studies. Biochemistry 39, 6317–6324 (2000).
15. Iqbal, A., Wang, L., Thompson, K. C., Lilley, D. M. J. & Norman, D. G. The structure of cyanine 5 terminally attached to double-stranded DNA: Implications for FRET studies. Biochemistry 47, 7857–7862 (2008).
16. Fuller, C. W. et al. The challenges of sequencing by synthesis. Nat. Biotechnol. 27, 1013–1023 (2009).
17. Agbavwe, C. & Somoza, M. M. Sequence-dependent fluorescence of cyanine dyes on microarrays. PloS ONE 6, e22177 (2011).
18. Kretschy, N. & Somoza, M.M. Comparison of the sequence-dependent fluorescence of the cyanine dyes Cy3, Cy5, DyLight DYS47 and DyLight DY464 on single-stranded DNA. PloS ONE 9 (2014).
19. Kretschy, N., Sack, M. & Somoza, M. M. Sequence-dependent fluorescence of Cy3-and Cy5-labeled double-stranded DNA. Bioconjugate Chem. 27, 840–848 (2016).
20. Lietard, J., Ameur, D. & Somoza, M. M. Sequence-dependent quenching of fluorescein fluorescence on single-stranded and double-stranded DNA. Rsc Adv. 12, 5629–5637 (2022).
21. Holz, K. et al. High-efficiency reverse (5′→3′) synthesis of complex DNA microarrays. Sci. Rep. 8, 15999 (2018).
22. Phillips, M.E. et al. In situ oligonucleotide synthesis on carbon materials: stable substrates for microarray fabrication. Nucleic Acids Res. 36 (2008).
23. Doktycz, M. J., Paner, T. M., Amaratunga, M. & Benight, A. S. Thermodynamic stability of the 5′ dangling-ended DNA hairpins formed from sequences 3′-(Xy)2ggatac(T)4gtatcc-3′, Where X, Y = a, T, G, C. Biopolymers 89–97 (2001).
24. Kretschy, N., Holik, A. K., Somoza, V. & Somoza, M. M. Next-generation o-nitrobenzyl photolabile groups for light-directed chemistry and microarray synthesis. Anal. Chem. 85, 8513–8517 (2013).
25. Lietard, J. et al. Chemical and photochemical error rates in light-directed synthesis of complex DNA libraries. ACS Sustain. Chem. Eng. 5, 828–834 (2017).
26. Schneider, T. D. & Stephens, R. M. Sequence logos—a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100 (1990).

Acknowledgements
Funding by the Austrian Science Fund (FWF grants P30596, I4923 and P34284) is gratefully acknowledged. The authors thank Mark Somoza for support and fruitful discussions and Orgentis GmbH for the synthesis of 3′-BzNPPC DNA phosphoramidites.
Author contributions
T.K. performed all experiments and analyzed the data while J.L. provided supervision, designed the experiments, acquired funding and wrote this manuscript, which all authors read and approved.

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/s41598-022-19069-9.

Correspondence and requests for materials should be addressed to J.L.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022