BIOASSEMBLER

Integrating bio-inspired assembly into semiconductor manufacturing technology for biosensors

D4.2 Optimized zipcode sequences

Version: V1.0
Status: Final
Date: 31st August 2023
Project Coordinator: VTT(FI)
Project Start Date: 1st September 2022
Project Duration: 36 months
Deliverable Nº: D4.2
Dissemination level: Public

Authors: Gisela Ibáñez-Redín, Petri Saviranta, Mark Somoza
Contact Person: Mark Somoza

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070589. This document reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein

Revision History

Version	Status	Date	Contributor / Partner	Summary of Changes
0.1	Draft 1	1.8.2023	Gisela Ibáñez-Redín /UNIVIE	First draft version
0.2	Review and comments	15.8.2023	Mark Somoza / Univie	Second draft version
1.0	Finalization of report	31.8.2023	Petri Saviranta/VTT	Finalization

List of contents

List of Tables

List of figures

1. Background and objective

The DNA-guided self-assembly of proteins represents a methodology that utilizes the precise and programmable base-pairing interactions intrinsic to DNA molecules to direct the assembly of proteins onto solid surfaces.[1] The principle is simple and involves designing specific DNA sequences that, when attached to proteins, can guide their assembly on surfaces functionalized with the complementary oligos (zipcodes). This approach is important in applications involving multiplex detection of biomarkers, as in BIOASSEMBLER, since it allows for the programmable immobilization of different antibodies in specific device positions. Moreover, it offers the possibility to precisely engineer the orientation and density of proteins on the sensor surface, which can enhance the sensor's sensitivity and specificity.

Applying this methodology in manufacturing multiplexed devices in real scenarios demands sequences with high selectivity toward their target to avoid cross-hybridization and false positives. These sequences must also be compatible with relatively low hybridization temperatures necessary to avoid denaturation of the proteins. Addressing these requirements demands designing and validating a set of candidates through studying the thermodynamic binding parameters using computational tools and testing the hybridization with fluorophore-labelled complementary oligonucleotides. The design and test of the DNA zipcode sequences is the objective of task 4.2 associated with this deliverable.

This deliverable provides information on the strategies adopted for designing DNA Zipcodes by VTT and the University of Vienna. The main results of the hybridization experiments with the candidate sequences are presented and discussed. Such experiments were done using microarrays produced by *in situ* photolithographic DNA synthesis. The primary limitations observed, which mainly related to the occurrence of cross-hybridization between most of the candidate sequences, are addressed. Some results of optimization experiments of synthesis parameters are also discussed, accompanied by a brief overview of strategies intended for implementation to enhance selectivity.

2. Results

2.1 Design of the sequences

A set of sequences was developed by considering the criteria mentioned in the previous section. To accomplish this, we began with an experimentally validated orthogonal library previously reported by Xu et al.[2] Within this library, sequences that adhered to specific criteria were chosen: (1) inclusion of a fixed section (CTTTT) at the 3' end, used for optimizing and equalizing fluorescence signal levels in hybridization experiments with complementary oligos labelled with Cy3[3]; (2) a variable region of 12 nucleotides, followed by (3) either a G or C. Various software tools were used to analyze the thermodynamic binding parameters of the candidates against their target complementary sequences, for internal secondary structures, and for cross-hybridization to the other sequences. After applying these filters, we chose a set of 18-mers with melting temperatures ranging from 55.0 to 57.6 °C. Table

1 shows the 14 sequences selected for the hybridization experiments and their respective complementary sequences.

Table 1. Set of sequences selected for the hybridization experiments and their respective complementary sequences. The hybrid melting temperatures were estimated using IDT Oligo Analyzer Tool (conditions: 0.1 µM oligo conc., 150 mM Na^{+}, 0 mM Mg^{2+}, 0 mM dNTP).

Target (solution)	Sequence (5'->3')	Probe (array)	Sequence (3'->5')	Hybrid Tm (°C)
1	Cy3-GAAAACTTAGCACGTCGG	P01	CTTTTGAATCGTGCAGCC	57.3
2	Cy3-GAAAATCCTGGGTATCGG	P02	CTTTTAGGACCCATAGCC	56.0
3	Cy3-GAAAATAGGTCAGGGTCC	P03	CTTTTATCCAGTCCCAGG	55.5
4	Cy3-GAAAATGACCGAACACAC	P04	CTTTTACTGGCTTGTGTG	55.3
5	Cy3-GAAAATCTATGCCGCACC	P05	CTTTTAGATACGGCGTGG	57.6
6	Cy3-GAAAACCTCATATCGCTG	P06	CTTTTGGAGTATAGCGAC	54.2
7	Cy3-GAAAACGGTTATTGCTCC	P07	CTTTTGCCAATAACGAGG	55.1
8	Cy3-GAAAAGTGTCGCTCTCAG	P08	CTTTTCACAGCGAGAGTC	56.7
9	Cy3-GAAAACTCGACAGGCATC	P09	CTTTTGAGCTGTCCGTAG	57.1
10	Cy3-GAAAAGAATACCACCGGC	P10	CTTTTCTTATGGTGGCCG	57.2
11	Cy3-GAAAATAACTGTTCGCGC	P11	CTTTTATTGACAAGCGCG	56.0
12	Cy3-GAAAACAATCCTCTGCTG	P12	CTTTTGTTAGGAGACGAC	54.7
13	Cy3-GAAAAGTAGTTGCGTGTC	P13	CTTTTCATCAACGCACAG	54.9
14	Cy3-GAAAACAGTCAATTCCGG	P14	CTTTTGTCAGTTAAGGCC	55.0

2.2 Assessment of cross-hybridization

To explore the potential occurrence of cross-hybridization among the sequences, we conducted hybridization experiments employing microarrays containing numerous replicates of the 14 probe sequences on glass substrates. These arrays were produced using *in situ* photolithographic DNA synthesis, as detailed in prior studies.[4] Additionally, we incorporated a truncated version of each sequence, from which the CTTTT section at the 3' end was removed. This modification was included to determine whether the presence of this specific section across all sequences contributed to any observed cross-hybridization. A 25mer (QC25: 5'-GTCATCATCATGAACCACCCTGGTC-3') was also synthesized in the fiducial features used for alignment and normalization. All microarrays used in this study are based on the layout depicted in Figure 1. Four distinct subarrays are present within each microarray, which can be used in independent hybridization experiments. Each subarray adopts a 1:4 layout, as Figure 1(B) illustrates, enabling the photolithographic synthesis of 21476 sequences.

The microarrays were subjected to individual hybridization experiments using the 5'-Cy3-labelled target sequences (Biomers.net). These experiments were carried out at 37 °C, with constant rotation, for 2 hours. The hybridization solution employed consists of 90 nM of Cy3-labelled target oligo, 90 nM

of Cy3-labelled QC25 complementary oligo, 1 M Na+, 100 mM MES, 20 mM EDTA, 0.01% Tween-20, and 0.05% BSA.

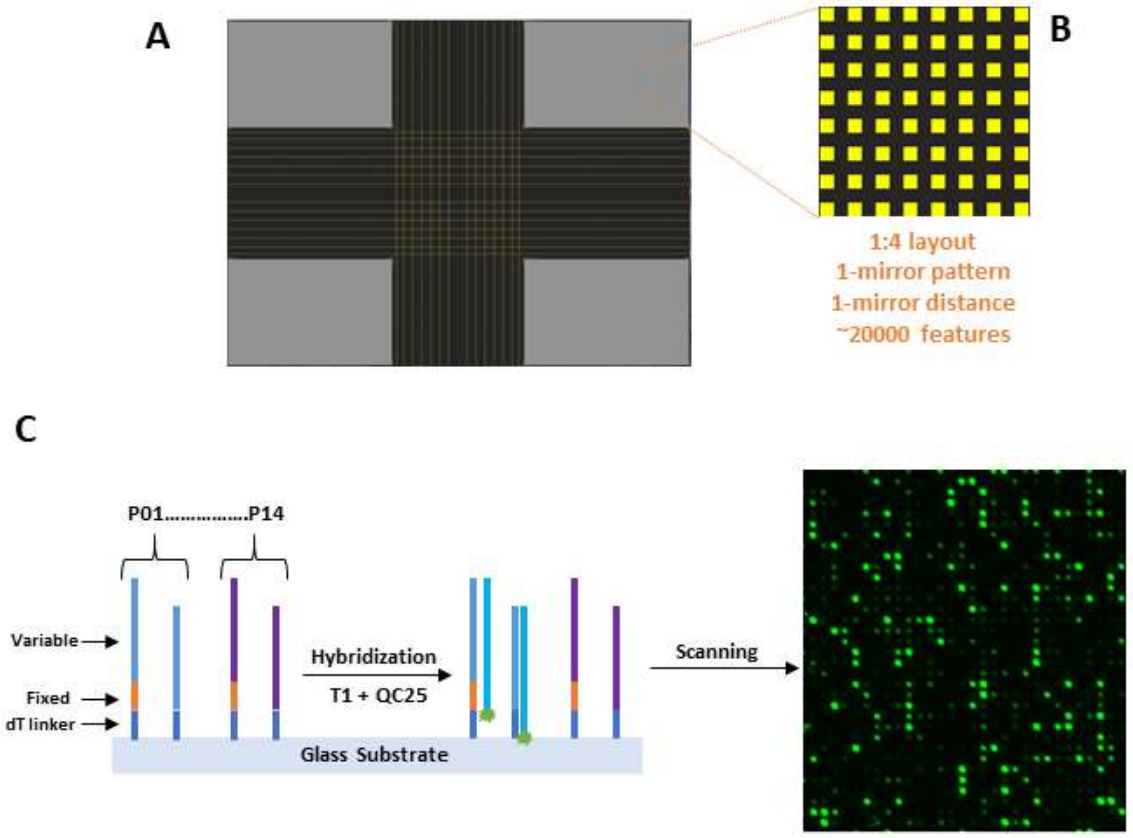

Figure 1 – Schematic representation outlining the experimental approach employed for evaluating cross-hybridization. The layout on (A-B) was adopted for the photolithographic synthesis of microarrays containing the initial 14 probe sequences and their respective truncated versions. (C) In individual experiments, these microarrays were hybridized with Cy3-labelled target sequences. The hybridization solution also contains the Cy3-labelled QC25 complementary oligo used for fiducials. Subsequent to scanning, analysis of the Cy3 fluorescent signal intensities served as an assessment of cross-hybridization.

To assess the presence of cross-hybridization, we analyzed fluorescent intensity data from replicated features for both the original and truncated sequences from experiments with each target (Figure 1 (C)). Figures 2 and 3 display the normalized fluorescence intensities. Except for targets 4 and 12, all the targets interact with other non-complementary probes on the array. As a result of these interactions, the signal intensities exceed 5% of the signal produced by hybridization with their perfectly matching probes. Notably, non-specific signals extend up to 60% of the signal observed during hybridizations with the perfect matches, as shown in Figures 2 (A) and 3 (A), which depict hybridizations involving targets 1 and 8, respectively.

Truncation of the common CTTTT sequence from the probes did (expectedly) decrease the signal intensity, but did not significantly affect the relative cross-hybridization profile. This finding suggests that cross-hybridization is not caused solely by the presence of a short shared segment in all probes. We found no correlations between sequence characteristics, such as GG content or specific nucleotide

positions, and the observed cross-hybridization. Considering the size of the sequences, we hypothesized that they might be susceptible to errors introduced during synthesis, which are inherent in chemical and photochemical reactions. For instance, incomplete photodeprotection can result in errors during subsequent coupling events, leading to deletions. Therefore, we investigated the effect of the radiant exposure used for photodeprotection, as well as other synthesis and hybridization parameters, on cross-hybridization. The results of these experiments are detailed in the following section.

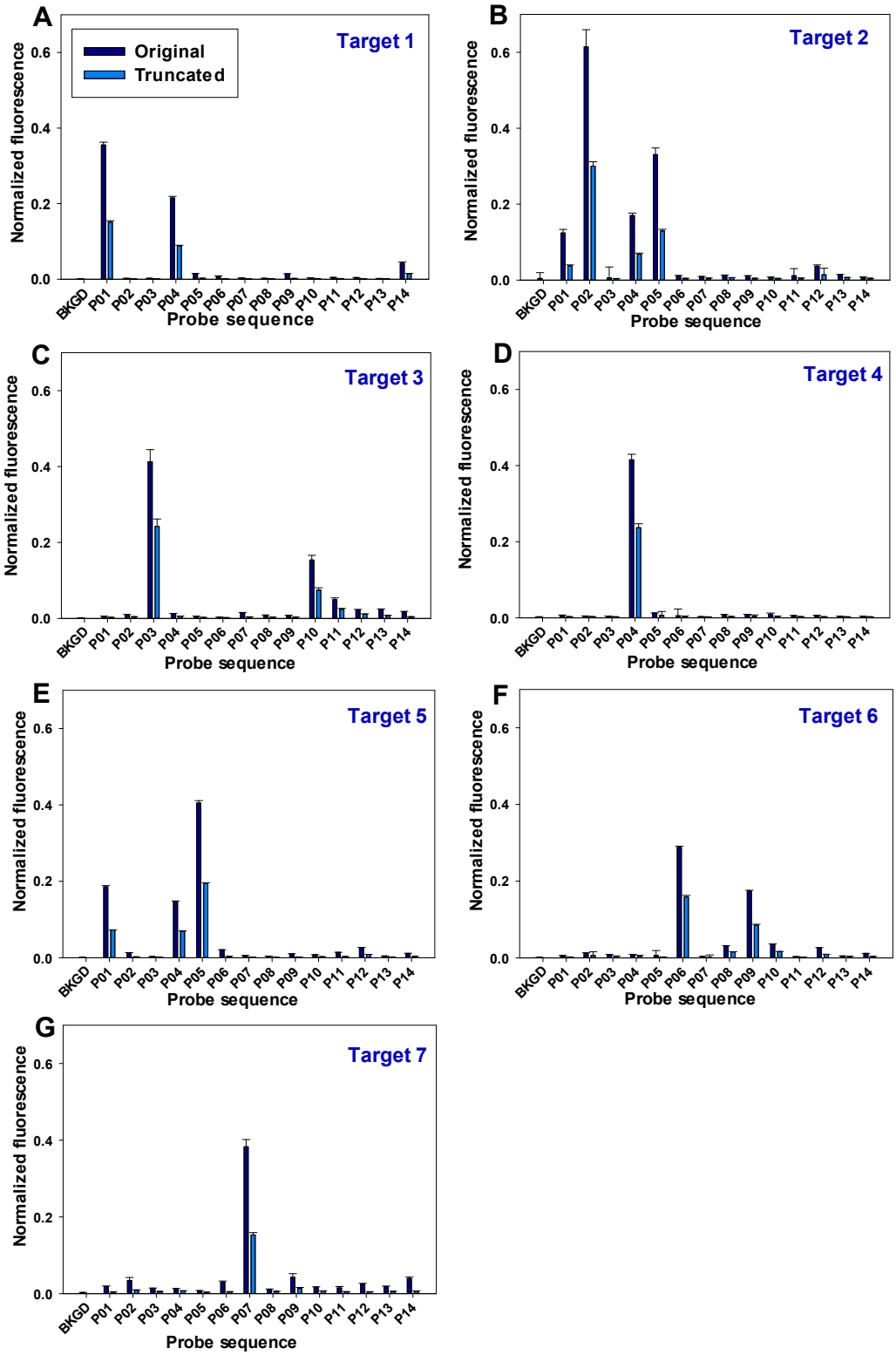

Figure 2 - Normalized fluorescence signal intensities acquired from microarrays containing the initial 14 probe sequences (Original) along with their corresponding truncated versions (Truncated) in hybridizations with targets 1 (A), 2 (B), 3 (C), 4 (D), 5 (E), 6 (F), and 7 (G). The normalization was done based on the signal obtained from the QC25 features used for fiducials.

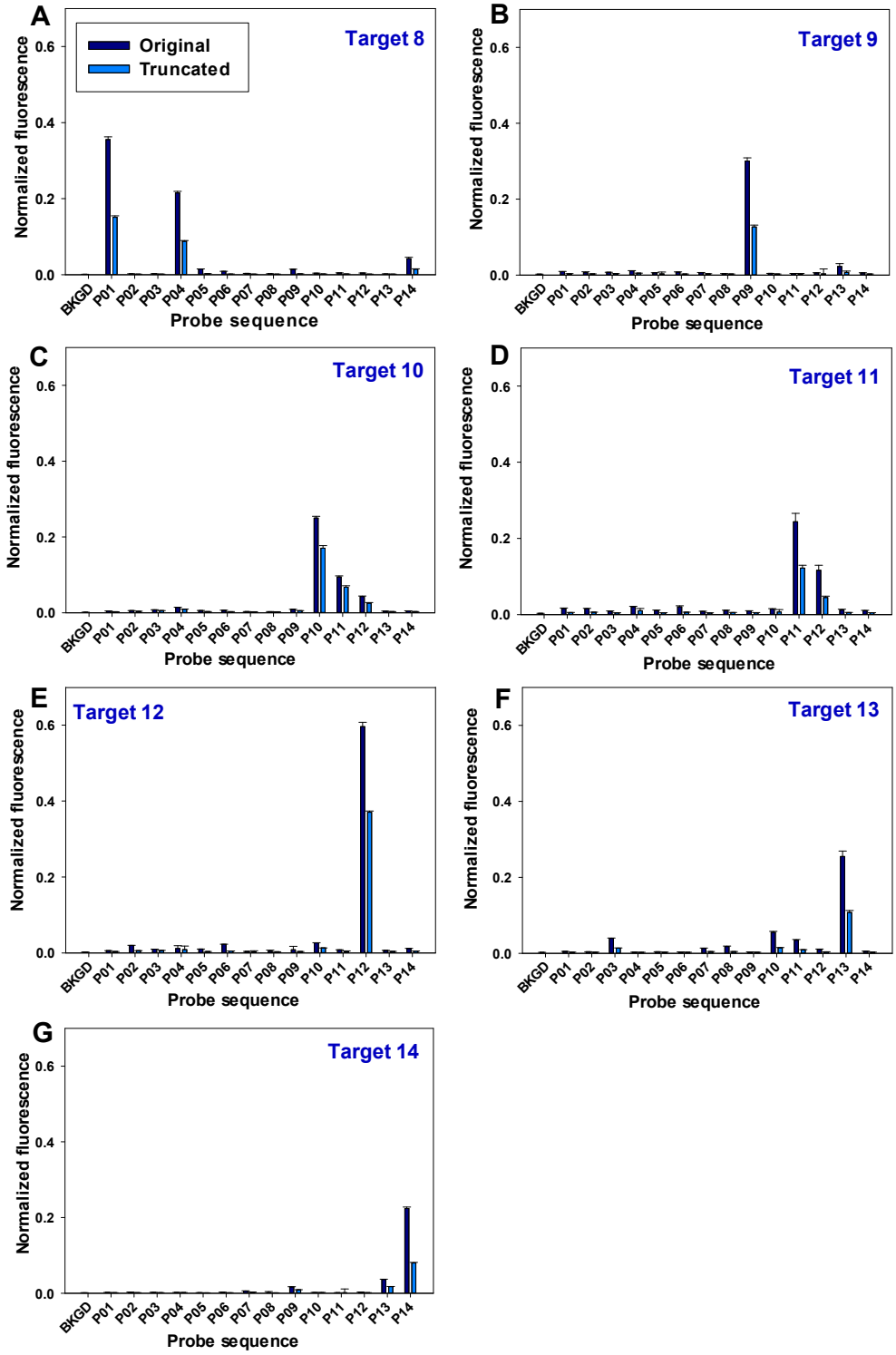

Figure 3 -Normalized fluorescence signal intensities acquired from microarrays containing the initial 14 probe sequences (Original) along with their corresponding truncated versions (Truncated) in hybridizations with targets 8 (A), 9 (B), 10 (C), 11 (D), 12 (E), 13 (F), and 14 (G). The normalization was done based on the signal obtained from the QC25 features used for fiducials.

2.3 Optimization of experimental parameters

All arrays employed in this study were synthesized using commercially available Bz-NPPOC DNA phosphoramidites (Orgentis). Previous research has indicated that achieving approximately 95% photodeprotection efficiency in Bz-NPPOC requires a radiant exposure of 3 J/cm².[5] Therefore, this specific light exposure was used to obtain the data in Figures 2 and 3. We also performed experiments involving a gradient of light exposure to assess its potential impact on cross-hybridization. Following synthesizing P05 sequences on microarrays with varying levels of UV light exposure, these microarrays were subjected to hybridization with the Cy3-labelled complementary oligo (Target 5) or with Target 2. Target 2 was selected for this test since it interacts with P05, as observed in Figure 2 (B). The fluorescence intensities normalized to the maximum signal for hybridization with Target 5 are shown in Figure 4 (A). As expected, there was an increase in hybridization signals by increasing UV light exposure and, consequently, sequence fidelity. However, increasing radiant exposure did not significantly reduce the signal derived from hybridizations with the unintended target. It should be noted that this behavior was also observed in hybridization experiments at 48 °C (Figure 4 (B)).

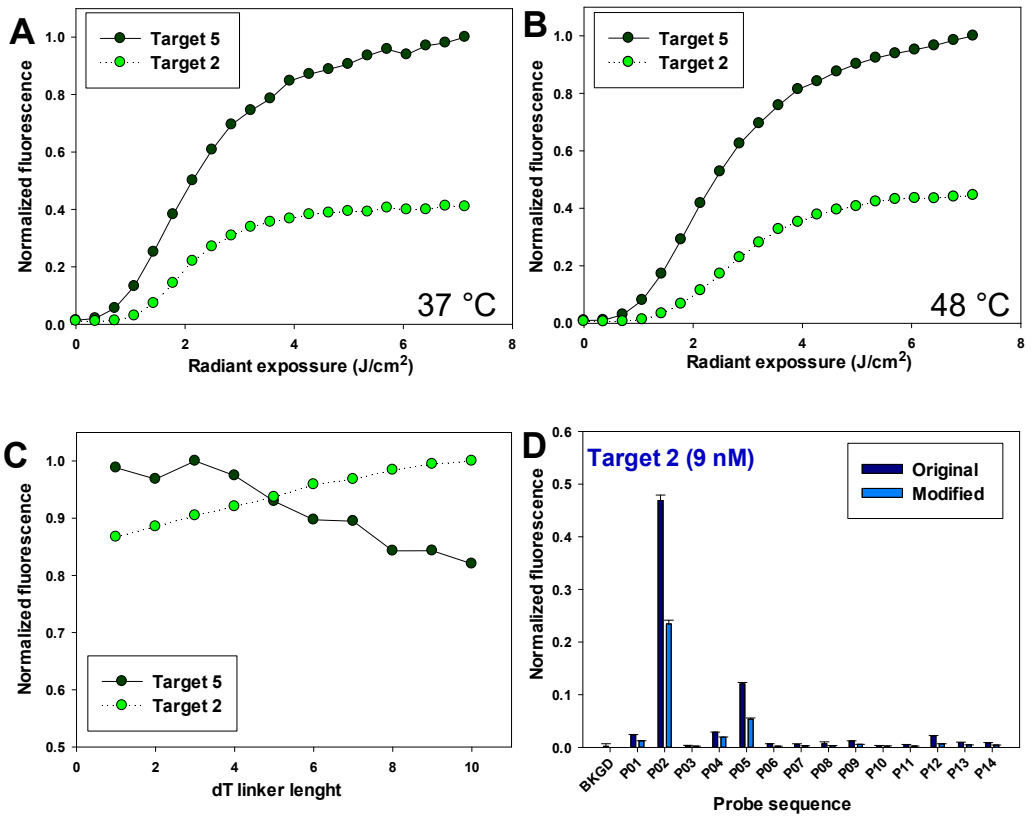

Figure 4 – Optimization of experimental parameters for DNA synthesis and hybridization. Hybridization-based fluorescence intensity for arrays of P05 sequence synthesized using a gradient of exposure with UV light. Hybridization was done using the 5'-Cy3 labelled complementary sequence (target 5) or a non-complementary sequence (target 2) at 37 °C (A) or 48 °C (B). Normalized fluorescence signal from hybridization vs linker length for arrays of P05 hybridized to target 5 or 2 (C). Normalized fluorescence intensities for arrays containing the Original and Modified (truncated) sequences in hybridization experiments with targets 2 (D).

We also explored the impact of the linker length between the functionalized glass surface and the probe sequence, as this factor can affect both the synthesis yield and the fluorescence signal. We synthesized P05 sequences on dT linkers with lengths ranging from one to 10 nucleotides. The normalized fluorescence signal for arrays hybridized with target 5 and target 2 (separate hybridizations) is shown in Figure 4 (C). In the case of hybridization with the matching target 5, the highest signal occurs with linker sizes between 1 and 3 nucleotides, followed by a decline. Conversely, for hybridizations involving the cross-hybridizing target 2, the signal increases with larger linker sizes. Similar results were observed in experiments using microarrays containing P04 sequences in hybridizations with target 5 and target 4 (data not shown). Notably, even when using a one nucleotide linker, high intensities were obtained for non-specific interactions, indicating that linker size is not a determining factor in cross-hybridization.

Overall, no significant improvements in hybridization selectivity were observed when optimizing synthesis parameters. This suggests that the observed cross-hybridization is more likely associated with sequence characteristics than with errors introduced during synthesis. Changing other parameters, such as Na^+ concentration used for hybridization, also resulted in insignificant improvements in selectivity and a decrease in intensity (data not shown). An improvement in selectivity was observed when the target concentration was decreased from 90 to 9 nM, as seen in Figures 2 (B) and 4 (D), showing hybridization data with target 2.

3. Discussion and outlook

The investigation of 14 target sequences through hybridization experiments yielded two target sequences (target 4 and target 12) displaying high selectivity, with cross-hybridization signals below 5%. When considering the aggregate signals of all probes and targets, only three probe sequences (P02, P03 and P06) demonstrated compatibility, suggesting potential utilization together. However, the development of the "zipcode" requires further refinement of hybridization conditions, encompassing variables such as target concentration and washing time. A prospective approach involves introduction of unlabeled competitors within hybridization solutions, aiming to enhance specificity and mitigate cross-reactivity.

Our findings indicate that the observed low selectivity is not attributable to the parameters employed in the synthesis process, but rather to inherent properties of the sequences themselves, potentially arising from the formation of surface structures. Consequently, the impact of introducing mismatches and truncations on selectivity will also be investigated.

4. References

(1) Niemeyer, C. M.; Sano, T.; Smith, C. L.; Cantor, C. R. Oligonucleotide-Directed Self-Assembly of Proteins: Semisynthetic DNA - Streptavidin Hybrid Molecules as Connectors for the Generation of Macroscopic Arrays and the Construction of Supramolecular Bioconjugates. *Nucleic Acids*

Res. **1994**, *22* (25), 5530–5539. https://doi.org/10.1093/nar/22.25.5530.

(2) Xu, Q.; Schlabach, M. R.; Hannon, G. J.; Elledge, S. J. Design of 240,000 Orthogonal 25mer DNA Barcode Probes. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (7), 2289–2294.

(3) Agbavwe, C.; Somoza, M. M. Sequence-Dependent Fluorescence of Cyanine Dyes on Microarrays. *PLoS One* **2011**, *6* (7), 1–12. https://doi.org/10.1371/journal.pone.0022177.

(4) Sack, M.; Hölz, K.; Holik, A. K.; Kretschy, N.; Somoza, V.; Stengele, K. P.; Somoza, M. M. Express Photolithographic DNA Microarray Synthesis with Optimized Chemistry and High-Efficiency Photolabile Groups. *J. Nanobiotechnology* **2016**, *14* (1), 1–13. https://doi.org/10.1186/s12951-016-0166-0.

(5) Kretschy, N.; Holik, A. K.; Somoza, V.; Stengele, K. P.; Somoza, M. M. Next-Generation o-Nitrobenzyl Photolabile Groups for Light-Directed Chemistry and Microarray Synthesis. *Angew. Chemie - Int. Ed.* **2015**, *54* (29), 8555–8559. https://doi.org/10.1002/anie.201502125.