



UHASSELT

KNOWLEDGE IN ACTION

Doctoral dissertation submitted to obtain the degree of
Doctor of Sciences: Physics, to be defended by

Yannick Stulens

DOCTORAL DISSERTATION

Enhanced DNA Hybridization-based Biosensors: A Thermodynamic Approach for Single-Nucleotide Variant Detection

Promoter: Prof. Dr Jef Hooyberghs | UHasselt

Co-promoter: Prof. Dr Bart Cleuren | UHasselt



UHASSELT

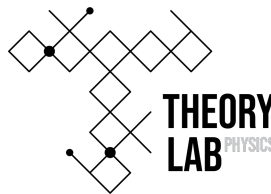
KNOWLEDGE IN ACTION

Doctoral dissertation submitted to obtain the degree of
Doctor of Sciences: Physics, to be defended by

Yannick Stulens

DOCTORAL DISSERTATION

Enhanced DNA Hybridization-based Biosensors: A Thermodynamic Approach for Single-Nucleotide Variant Detection



Promoter: Prof. Dr Jef Hooyberghs | UHasselt

Co-promoter: Prof. Dr Bart Cleuren | UHasselt

Declaration of Integrity

The PhD researcher and the UHasselt supervisor hereby formally declare that the research conducted for the purpose of this PhD thesis was executed in accordance with the principles of good scientific conduct, as stipulated in the UHasselt Integrity charter, the UHasselt charter supervisor - PhD Researcher, the UHasselt Integrity Policy and the UHasselt guidelines for the use of (generative) AI in research. The tools were used exclusively as an assistant to enhance writing clarity and supported the illustrations in the Introduction. They did not contribute to the generation, analysis, or interpretation of any scientific results.

Yannick Stulens,

Enhanced DNA Hybridization-based Biosensors:

A Thermodynamic Approach for Single-Nucleotide Variant Detection,
PhD thesis, UHasselt, Belgium, 2026.

The author asserts that this PhD thesis is made Open Access immediately upon submission. The full text is publicly available without restrictions.

Copyright © 2026 Yannick Stulens. All rights reserved.

Jury

Chairman	Prof. em. Dr Marc Gyssens UHasselt
Promoter	Prof. Dr Jef Hooyberghs UHasselt
Co-promoter	Prof. Dr Bart Cleuren UHasselt
Members of the jury	Prof. Dr Dmitry M. Kolpashchikov University of Central Florida United States Dr Wahyu Wijaya Hadiwikarta DKFZ German Cancer Research Center Germany Prof. Dr Karolien De Wael University of Antwerp Belgium Prof. Dr Patrick Wagner KU Leuven Belgium Prof. Dr Jean V. Manca UHasselt Belgium

Acknowledgements

Dankwoord

Na mijn master fysica in Gent wist ik dat ik graag een doctoraat wilde doen, en waar kan dat beter dan aan het warme UHasselt, waar ik eerder mijn bachelor behaalde. Ik kreeg te horen dat er binnenkort een nieuwe professor zou worden aangesteld, die nog op zoek was naar een doctoraatsstudent. Op aanraden van Jean Manca sprak ik af met mijn promotor, Jef, om te bekijken of er een match was. Jef had veel ervaring met DNA onderzoek en wilde deze expertise verder uitbouwen in zijn nieuwe onderzoeksgroep. Ik had altijd al een sterke interesse in biologische systemen, en deze vanuit een theoretisch perspectief beschrijven sprak mij meteen aan. Het bleek een perfecte match te zijn.

Ik wil daarom beginnen met het bedanken van mijn promotor. Jef, ik kijk enorm naar u op als onderzoeker. Onze start op de universiteit was veel zoeken, maar ze was ook heel leerrijk. Samen het vak Statistische Fysica uitwerken vond ik heel tof. Over deze vier jaren is het voor u veel drukker geworden, maar ondanks die drukke agenda maakte je steeds tijd, zelfs voor de kleinste details. Ik ben heel dankbaar dat mijn doctoraat onder uw begeleiding plaatsvond.

Ook binnen de onderzoeksgroep ben ik veel mensen dankbaar. Pieter-Jan startte niet veel later dan ik aan het doctoraatstraject, waardoor we heel wat ervaringen hebben kunnen delen. PJ, het was fijn om uw enthousiasme en doorzettingsvermogen naast mij te hebben. De toekomst van de onderzoeksgroep ziet er ook goed uit. Robbe, Thibau en Alexander, het was heel leuk om het DNA enthousiasme met jullie te delen, en ik geloof erin dat jullie dit onderzoek op een mooie manier zullen verderzetten.

Aangezien een groot deel van mijn doctoraat bestaat uit samenwerkingen, zijn er ook enkele mensen die ik in het bijzonder wil bedanken. Rebekka, tijdens onze samenwerking was jij al een senior PhD-onderzoeker, en ik heb veel geleerd van jouw ervaring, grondigheid en manier van werken. Brittany, we did so much research together, and to this day, I still feel a little bad that I had you do so many experiments ☺. Thank you for being who you are. I wish you and Cheuk all the best. Mijn laatste partner-in-crime, Thijs, ik ben blij dat ik met u dit interessante onderzoek heb kunnen uitwerken, en dat we onderweg de problemen samen hebben kunnen oplossen. Ik zal met veel plezier terugdenken aan onze tijd in Lissabon, aan de conferentie, maar zeker ook aan het toerist zijn.

My research stay in Florida was one of the most amazing experiences of the past four years. Dr. K, thank you for welcoming me into your research group at UCF. Your critical attitude always seems to find exactly the right question needed to move the research forward. I am grateful to have met you and the people in your group, and I will remember labsgiving at your house as one of the nicest moments of my PhD.

Een constante doorheen mijn traject in de fysica zijn Nico en Siebe. Bedankt om te bestaan. Jullie hebben mijn leven oprecht mooier gemaakt. We hebben samen veel meegemaakt en elkaar zien groeien als personen. Ik wens jullie, samen met Jessica en Birte, het allerbeste toe in het leven. Ook voor onze vriend Marc: bedankt voor uw gezelschap en wijze raad. Dr. Allyson, ik ben blij dat ik de eindfase samen met u heb overleefd. En ook in dit dankwoord, Siebe deed het mij voor, moet de nerdy Magic-vriendengroep aangehaald worden. Hendrik, Siebe, Martijn en Toon, jullie zijn stuk voor stuk topkerels, en ik hoop dat onze vriendschap nog heel lang mag blijven bestaan.

Ik sluit af bij mijn familie. Mama, papa, jullie hebben mij alles gegeven wat ik mij maar kon wensen. Axelle en Céderic, ik ben blij dat jullie elkaar gevonden hebben. Hoe jullie in het leven staan, daar kan ik veel van leren. 最後に、日本の家族にも感謝の気持ちを伝えたいと思います。お母さん、ボブ、俊、おじいちゃん、日本での経験は、私の世界観を大きく変えてくれました。皆さんを家族と呼べることを心から感謝しています。

Ruby, gij zijt een braaf hondje!

Marieke, ik hou van u. We houden nog wel contact ☺.

Summary

This PhD thesis investigates how thermodynamic insights into the properties of DNA hybridization can improve the selectivity of biosensors for mutation detection. By understanding, characterizing, and exploring ways to control the underlying interactions, this research contributes to the development of more accurate biotechnological applications.

Chapter 1 outlines the scientific context and describes, starting from fundamental principles, the selectivity problem within a universal model. In this framework, selectivity is reduced to a single parameter with a direct physical interpretation.

Chapters 2, 3, and 4 present three scientific articles that together form the core of this work. Each article was prepared in collaboration with a different experimental research group and approaches the selectivity problem from a different perspective. Various experimental platforms are employed, demonstrating the broad applicability of the concepts.

In addition, Chapter 5 investigates “kinetic inversion” , a phenomenon in systems with multiple reaction steps. We formulate a mathematical description of this behavior, which has the potential to improve existing biosensors.

Finally, Chapter 6 provides an overview of the main conclusions of this research and discusses possible directions for future work.

Samenvatting

Dit doctoraatsproefschrift onderzoekt hoe thermodynamische inzichten in DNA-hybridisatie de selectiviteit van biosensoren voor mutatiedetectie kunnen verbeteren. Door de onderliggende interacties beter te begrijpen, nauwkeuriger te beschrijven en gericht te beïnvloeden, draagt dit onderzoek bij aan de ontwikkeling van nauwkeurigere biotechnologische toepassingen.

Hoofdstuk 1 schetst de wetenschappelijke context en beschrijft, vertrekkend vanuit basisprincipes, het selectiviteitsprobleem binnen een universeel model. Daarbij wordt selectiviteit herleid tot één parameter met een directe fysische interpretatie.

In Hoofdstukken 2, 3 en 4 worden drie wetenschappelijke artikels gepresenteerd die samen de kern van dit proefschrift vormen. Elk artikel werd opgesteld in samenwerking met een andere experimentele onderzoeksgroep en benadert het selectiviteitsprobleem vanuit een andere invalshoek. Daarbij wordt gebruikgemaakt van verschillende experimentele platformen, wat de brede toepasbaarheid van de concepten aantoont.

Daarnaast onderzoekt Hoofdstuk 5 het fenomeen van ‘kinetic inversion’ in systemen met meerdere reactiestappen. We formuleren een wiskundige beschrijving van dit gedrag, dat potentieel bestaande biosensoren kan verbeteren.

Tot slot geeft Hoofdstuk 6 een overzicht van de belangrijkste conclusies van dit onderzoek en worden mogelijke toekomstige onderzoeksrichtingen besproken.

Contents

Acknowledgements / Dankwoord	i
Summary	iii
Samenvatting	v
List of Publications	5
List of Acronyms	9
1 Introduction	13
1.1 Research Context and Motivation	14
1.2 DNA	16
1.3 Hybridization Biosensor & Free-energy Difference	19
1.4 Readout Techniques	25
1.4.1 The Microarray	26
1.4.2 The Molecular Beacon	27
1.4.3 Polymerase Chain Reaction (PCR)	29
1.4.4 Next-generation Sequencing (NGS)	31
1.5 Nearest-Neighbor model	33
1.6 Aims of this Thesis	35
References	49
2 Hybridization-based sensor with large dynamic range for detection of circulating tumor DNA in clinical samples	51
2.1 Introduction	53
2.2 Materials and Methods	55
2.2.1 Hybridization method: the reference probe and the concept of depletion	55

2.2.2 NSCLC patients' cfDNA samples	61
2.2.3 dPCR	62
2.2.4 PCR amplicon generation	62
2.2.5 Lambda exonuclease digestion	63
2.2.6 DNA hybridization experiment	63
2.3 Results and Discussion	65
2.3.1 Application to clinical samples	65
2.3.2 Evaluation of the quantitative agreement with dPCR . .	67
2.3.3 Extension of the model to accurately determine high mutant concentrations	68
2.4 Conclusions	74
Supplementary Information	76
References	93
3 Breaking the Limit of Hybridization Selectivity by Strain- forming Probes and Wild-type Blocker	95
3.1 Introduction	98
3.2 Materials and Methods	100
3.2.1 Reagents	100
3.2.2 Sensor annealing and reaction conditions	100
3.2.3 Fluorescent assay and analysis	100
3.3 Results	101
3.3.1 Detection Signal and Limit of Detection	101
3.3.2 Strategies to enhance $\Delta\Delta G$	104
3.3.3 Optimal detection	106
3.3.4 Contribution of Strain	108
3.3.5 Contribution of Blocker	109
3.3.6 Blocker strands counteract probe-induced mutant de- pletion	112
3.3.7 Standard Mismatch Penalty	112
3.4 Discussion	113
Supplementary Information	116
References	128
4 Double Mismatch Interactions in Probe Design enable Im- proved Hybridization-based SNV Discrimination	129
4.1 Introduction	132
4.2 Materials and Methods	134

4.2.1 Probe enrichment experiment	.134
4.2.2 Hybridization conditions	.135
4.2.3 Library preparation	.135
4.2.4 Sequencing & data processing	.136
4.2.5 Free-energy Penalty Calculation	.137
4.2.6 Photoelectrochemical biosensor	.139
4.3 Results	.139
4.3.1 Mismatch Interactions	.139
4.3.2 High-throughput platform	.141
4.3.3 Poly T	.143
4.3.4 <i>KRAS</i> & <i>BRAF</i>	.145
4.3.5 Photoelectrochemical benchmark	.150
4.4 Discussion	.153
Supplementary Information	.156
References	.171
5 Kinetic Inversion in Multi-stranded DNA Biosensors	173
5.1 Introduction	.174
5.2 Kinetics of the one-step process	.176
5.3 Kinetics of the two-step process	.180
5.3.1 Optimal discrimination time	.184
5.3.2 Two-step process differentiates in on-rate	.186
5.4 Enhanced Kinetic Inversion using Wild-type Blockers	.187
5.5 Preliminary Experimental Results	.189
References	.194
6 Conclusions & Outlook	195
6.1 Thermodynamics as a design principle	.196
6.2 Selectivity improvement beyond $\Delta\Delta G$	.197
6.3 Depletion as a versatile strategy	.198
6.4 Strain and Mismatch interactions: a unified perspective?	.199
6.5 Beyond equilibrium	.201
6.6 Reflection on the past four years: theory and experiment	.202
References	.204

List of Publications

Publications in Journals

Paper A:

Y. Stulens, R. Van Hoof, K. Hollanders, I. Nelissen, M. Szymonik, P. Wagner, G. Froyen, B. Maes, and J. Hooyberghs, "Hybridization-based sensor with large dynamic range for detection of circulating tumor DNA in clinical samples," *Biosensors and Bioelectronics*, Vol. 281 (2025).

<https://doi.org/10.1016/j.bios.2025.117342>

Paper B:

Y. Stulens, B. L. Mueller, J. Hammock, A. Perez, R. Van Hoof, D. M. Kolpashchikov, and J. Hooyberghs, "Breaking the Limit of Hybridization Selectivity by Strain-forming Probes and Wild-type Blocker," (in review at *Nucleic Acids Research*)

Paper C:

T. Van der Snickt, **Y. Stulens**, L. Claudino, S. Bassini, S. Ailliet, K. De Wael, J. Hooyberghs, P. Mestdagh, "Double Mismatch Interactions in Probe Design enable Improved Hybridization-based SNV Discrimination," (submitted to *Sensors & Actuators, B: Chemical*)

Paper D:

L.-P. Jorissen, L. Jorissen, R. Van Haverbeke, J. Vandevenne, **Y. Stulens**, J. Hooyberghs, J. Hendrix, R. Thoelen, "Versatile Micro-heating Platform for Measuring Biomolecular Reaction Dynamics with Fluorescence Microscopy," (submitted to *Nature Communications*)

Patent Application

J. Hooyberghs, B. Cleuren, B. Wijns, **Y. Stulens**, "Systems and methods to quantify the transition rates of a stochastic system having a plurality of states," *International Patent Application WO2025078587* (2025).

Available at [Espacenet \(WO2025078587\)](https://Espacenet.org/WO2025078587)

International Research Stay

University of Central Florida

Kolpashchikov Lab

Orlando, FL, USA

01 September 2024 - 01 November 2024

Supported by FWO Travel Grant & European Erasmus+

Poster and oral contributions

Belgian Physical Society

Oral Presentation 17 May 2023 | Université de Namur, Belgium

"Use of DNA thermodynamics for low-abundance mutation detection by DNA hybridization"

4th European BioSensor Symposium

Poster Contribution 28-30 August 2023 | FH Aachen, Germany

"Use of DNA thermodynamics for low-abundance mutation detection by DNA hybridization"

VITO Data Science Lustrum

Poster Contribution 10 November 2023 | Mol, Belgium

"DNA Physics & Data Science: Improving hybridization-based sensors"

UHasselt Research Day

Poster Contribution 23 May 2024 | UHasselt, Belgium

"Pairing Theoretical and Experimental Science: The Physics of DNA Hybridization Sensors"

Kolpashchikov Lab

Oral Presentation 13 September 2024 | Orlando, FL, USA

"Biosensor Kinetics: One- and two-step processes"

Biosensors 2025 - 35th Anniversary World Congress on Biosensors

Poster Contribution 19-22 May 2025 | Lisbon, Portugal

"Kinetic inversion in hybridization-based biosensors with multi-step reaction mechanisms"

List of Acronyms

A	adenine
bp	base-pair
BRAF	Rapidly Accelerated Fibrosarcoma B
C	cytosine
cfDNA	circulating free DNA
ctDNA	circulating tumor DNA
ddPCR	digital droplet PCR
DNA	deoxyribonuclease acid
dPCR	digital PCR
dsDNA	double-stranded DNA
EGFR	epidermal growth factor receptor
EIS	electrochemical impedance spectroscopy
F	fluorophore
FA	fractional abundance
G	guanine
KI	kinetic inversion
KRAS	Kirsten rat sarcoma
LB	liquid biopsy
LNA	locked nucleic acid
LoD	limit of detection
MB	molecular beacon
MM	mismatch
MRD	minimal residual disease
MT	mutant
mut	mutant

NGS	next-generation sequencing
NN	nearest-neighbor
NSCLC	non-small cell lung cancer
nt	nucleotide
NTC	no-template control
ODE	ordinary differential equation
OWL	OWL sensor named after its resemblance to owl eyes
OWL2	Second-generation OWL sensor with unfolding arms
PCR	polymerase chain reaction
PEC	photoelectrochemical
PM	perfect match
PNA	Peptide nucleic acid
polyA	poly-adenine
polyT	poly-thymidine
Q	quencher
qPCR	quantitative PCR
SNV	single-nucleotide variant
SPR	surface plasmon resonance
ssDNA	single-stranded DNA
T	thymine
UMB	universal molecular beacon
VAF	variant allele frequency
WC	Watson-Crick
WT	wild-type

Chapter 1

Introduction

1.1 Research Context and Motivation

Increased life expectancy and demographic aging have led to a rising global burden of chronic and genetic diseases. Many of these diseases, particularly cancer, are driven by molecular changes that accumulate over time [1, 2, 3]. As healthcare systems shift from reactive intervention to early detection and prevention, there is a growing demand for scalable and cost-effective molecular diagnostics that enable early identification and longitudinal monitoring of disease [4, 5]. Precision medicine relies on detecting and interpreting specific genetic variants that determine disease subtype, therapeutic response, and prognosis. Consequently, the ability to accurately detect genetic variants is central to modern clinical practice.

The molecular basis of many diseases lies in DNA, which encodes proteins that perform most structural and functional tasks within the cell. A change in the DNA sequence can therefore alter protein structure and function. While most genetic variations are harmless, certain mutations can disrupt regulatory pathways, impair repair mechanisms, or activate uncontrolled cell growth. In cancer, such mutations drive tumor initiation and progression [6]. Modern therapies inhibit pathways that are known to be associated with particular mutations [7]. This has given rise to personalized medicine, where targeted therapy is guided by a patient's unique genetic variants [8]. However, tumors are not static systems. Under therapeutic pressure, new mutations can arise that lead to treatment resistance [9]. Importantly, these resistant variants often exist at extremely low abundance long before they become clinically dominant. For effective disease management, it is therefore essential not only to identify mutations at diagnosis but also to monitor their presence over time. Detecting rare emerging variants before relapse occurs is a key objective of precision medicine.

Traditional tissue biopsies provide only a snapshot of a tumor at a single location and time point. In contrast, liquid biopsy analyzes circulating tumor DNA in body fluids such as blood [10, 11]. This minimally invasive approach enables repeated sampling and improved monitoring of tumor dynamics [12, 13, 14]. The major challenge, however, lies in the extremely low relative concentration of mutant DNA in these samples, which

can fall below 0.1% [15]. When this is the case, mutant sequences are vastly outnumbered by nearly identical wild-type DNA, which easily obscures their detection. The detection task, therefore, becomes one of extreme selectivity: distinguishing a mutation within a large background of wild-type sequences. This challenge becomes most severe when the mutant and wild-type sequences differ by only a single nucleotide, creating a needle-in-a-haystack problem. At its core, this is fundamentally a physics problem. The detection probes interact with both mutant and wild-type DNA, and thermodynamic principles govern their binding behavior [16].

Although cancer monitoring provides a primary motivation for this work, highly selective DNA detection extends far beyond oncology. In infectious diseases, single-nucleotide mutations can lead to antibiotic resistance and must be identified with high specificity [17]. In transplantation medicine and non-invasive prenatal testing, rare genetic variants must be detected within a large background of closely related DNA [18, 19, 20]. Beyond healthcare, control over hybridization processes is crucial for emerging technologies such as DNA computing and DNA origami, where sequence-dependent binding governs information processing and self-assembly [21, 22]. The problem of selective mutant detection is therefore not only clinically relevant but also, more broadly, fundamental to programmable biomolecular systems.

This work takes a thermodynamic approach to study DNA hybridization and to improve biotechnological applications. We adopt a physics-based framework that reduces the complex system to its essential interacting components, leading to a consistent description of DNA hybridization behavior. This introduction briefly presents the DNA structure and relevant properties for this thesis, followed by a thermodynamic description of DNA-based biosensing. We put particular emphasis on how single mismatches affect the binding free energy and thereby the sensor selectivity. Within this framework, selectivity is quantified by the free-energy difference ($\Delta\Delta G$) between the desired and undesired binding, and this quantity serves as the fundamental constraint in low-abundance mutant detection. Subsequently, we outline the strategies explored in this thesis to overcome this intrinsic limitation and improve selectivity detection beyond conventional bounds.

1.2 DNA

The central molecule of this thesis is deoxyribonucleic acid (DNA). This polymer consists of smaller building blocks called nucleotides, which come in four versions depending on the nitrogen-containing bases: adenine (A), cytosine (C), guanine (G), and thymine (T). These can bind to form a chain of nucleotides, featuring a semi-rigid sugar-phosphate backbone and exposed bases in a sequence. The asymmetric backbone gives each DNA strand a 5' end and a 3' end, named after the fifth and third carbon atoms to which the phosphate group is attached. The sequence of the single-stranded DNA (ssDNA) presented in Figure 1.1 is denoted by 5'-GCAT-3'.

The exposed bases allow for binding to enzymes and other biomolecules, but its main property is the binding to another ssDNA. As two ssDNA bind to form double-stranded DNA (dsDNA), called hybridization, they naturally twist to form a double helix structure or duplex [24]. Hybridization is driven by hydrogen bonds between opposing nucleobases, as well as stabilizing π - π stacking interactions between adjacent base pairs. Hybridization occurs when duplex formation is thermodynamically favorable, i.e., when the Gibbs free energy change (ΔG) is negative under the given conditions. This depends on multiple factors, including the length of the binding region, the temperature, and reaction conditions such as salinity. The three and two hydrogen bonds in the G-C and A-T bindings, respectively, result in stable base pairs (bp's), such that binding occurs selectively when the A-T and G-C complementarity is fulfilled along the length of the DNA. We refer to these complementary bindings as Watson-Crick (WC) base pairs. This base-pairing rule means that each strand can essentially "read" or "check" the sequence of another strand and will serve as the basis for the DNA biosensor. This property is widely used in nature, where DNA complementarity allows each strand to serve as a precise template for copying genetic information with precision during DNA replication, gene expression, and for controlled biological function [25]. Importantly, duplex formation is not an all-or-none process but is governed by thermodynamic equilibrium between single-stranded and double-stranded states [26]. When many DNA molecules are present, hybridization must therefore be understood statistically. At any given time, only a fraction of strands is in the duplex state, while the rest remains single-stranded. At equilibrium, this fraction is determined by the duplex stability quantified by ΔG under the given

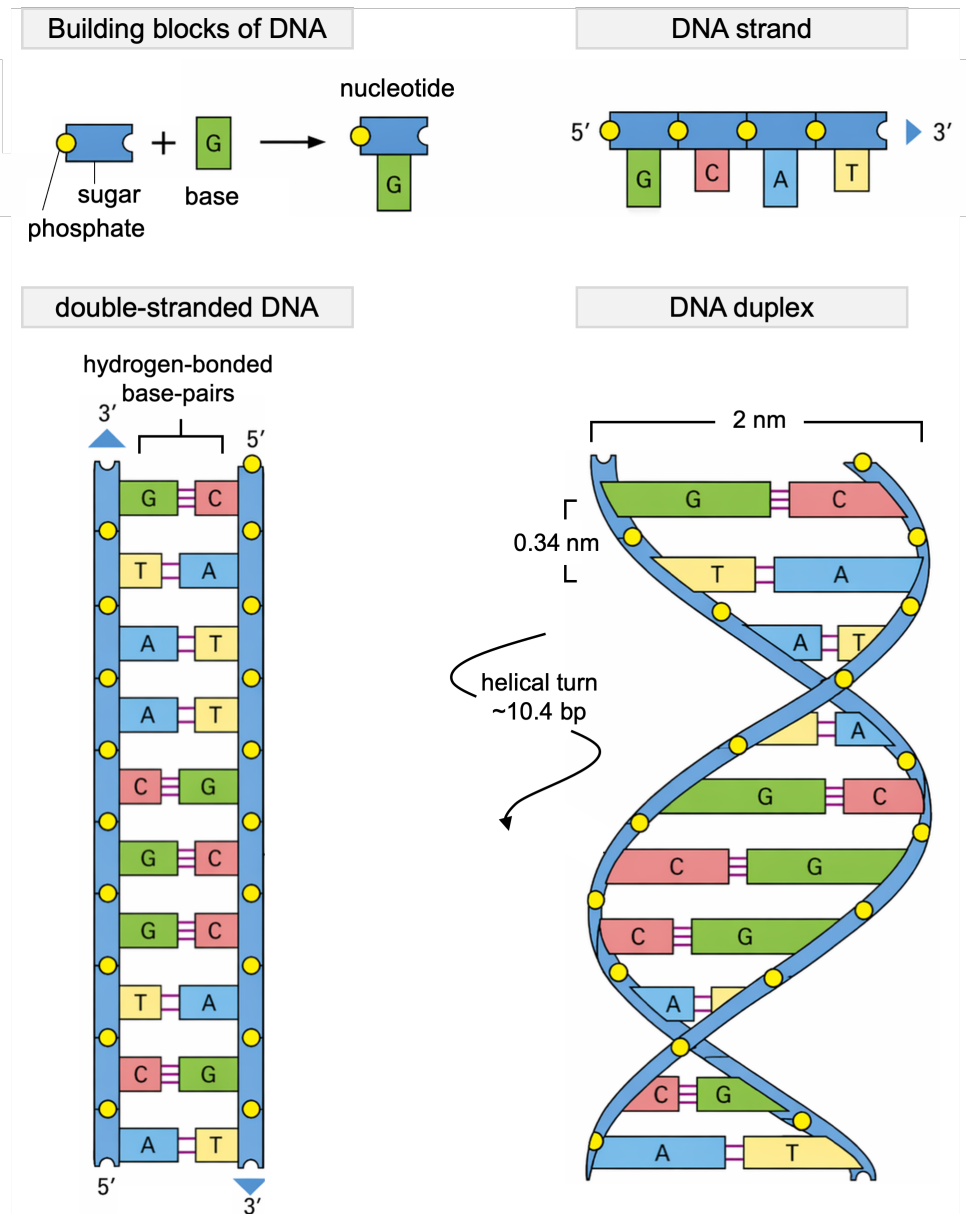


Figure 1.1: Schematic illustration of the structure of DNA. The combination of a phosphate group, a deoxyribose sugar, and a nitrogenous base forms a nucleotide. Sugar phosphates link together to form a chain of nucleotides, resulting in a single DNA strand. Two complementary strands can hybridize to form double-stranded DNA through hydrogen bonding between complementary base pairs: adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C). The double-stranded molecule naturally coils into the characteristic DNA double-helical structure, with a helical turn over ~ 10.4 base pairs. Adapted from [23].

conditions. The fraction is then maintained by a balance between ssDNA associating to form dsDNA, and dsDNA dissociating back into ssDNA.

If WC complementarity is not fulfilled at one or more sites, duplex formation becomes less favorable but not impossible. Instead of the A-T and G-C bp's, eight possible mismatched bp's can form (e.g., A-G or C-T). These mismatches (MM) locally disrupt hydrogen bonding and stacking interactions, thereby reducing overall duplex stability [27]. Although hybridization can still occur, the equilibrium shifts toward the single-stranded state. The magnitude of this destabilization depends on several factors, including the identity of the mismatch, its position within the duplex, and the surrounding sequence context. For most applications, including DNA biosensing, such cross-hybridization is undesirable because it reduces sequence selectivity. A probe designed to detect a specific target sequence should ideally hybridize only with perfectly complementary strands while strongly discriminating against mismatched ones. Understanding and controlling the differences between perfectly matched and mismatched duplexes, therefore, plays a crucial role in the design of hybridization-based biotechnology. Biological systems face a similar challenge: identifying specific genes or, during DNA replication, preventing incorrect base incorporation that leads to mutations. To maintain the high fidelity required for operation and reliable genetic information transfer, cells employ several error-correction strategies [28]. Among these are mechanisms that ensure high selectivity during base pairing, mismatch repair systems that recognize and remove incorrectly paired nucleotides after replication, and repeated proofreading steps that exploit kinetic differences between correctly matched and mismatched hybridization [29, 30, 31].

Finally, it is important to note that a single strand of DNA can interact with itself. If partial complementarity exists within its own sequence, an ssDNA molecule can fold back onto itself and form intramolecular base pairs. This self-interaction leads to the formation of secondary structures such as "hairpins", where a single-stranded loop connects a double-stranded stem. Depending on the sequence, more complex structures involving multiple loops and stems may also form. These secondary structures can influence hybridization behavior by reducing the accessibility of certain bases or by introducing additional barriers that must be overcome during duplex formation [32, 33].

1.3 Hybridization Biosensor & Free-energy Difference

In a typical hybridization-based biosensor, a ssDNA functions as a *probe* as it is designed to be complementary to a specified sequence. This is the ‘target’ sequence, as we are specifically aiming to detect the presence of these target DNA. For the oligonucleotides of interest (~ 5 -100 bp long), the hybridization process can often be approximated by a two-state model [26, 34], where probe P and target T are either in the duplex (hybridized or bound) state or remain in the unbound state.



This system is then described using a mean-field approach, not describing individual molecules but their average behavior using concentrations of probe, target and duplex, denoted by square brackets, [P], [T] and [PT] respectively. The association and dissociation rates are given by

$$\text{Association rate: } k_+[P][T] \quad (1.2)$$

$$\text{Dissociation rate: } k_-[PT] \quad (1.3)$$

The association rate constant is k_+ and the dissociation rate constant is k_- . While these rate constants are in fact constant, the association and dissociation rates depend on the currently available concentrations. Assuming that no duplex is initially present, the dissociation rate increases during the binding process, while the association rate decreases as ssDNA becomes bound. The rate constants determine the equilibrium constant K and relate to the free-energy ΔG and gas constant times temperature RT :

$$K = \frac{k_+}{k_-} = e^{-\Delta G/RT} \quad (1.4)$$

This represents the stability of the duplex, where a negative ΔG results in a forward-driving reaction whose magnitude determines the affinity of the probe and target towards each other. The free-energy ΔG and therefore also K , depends on the DNA sequences of both strands, their lengths and the reaction conditions. For now, we are interested in the binding at equilibrium. A detailed balance exists between the association of ssDNA and the dissociation of dsDNA, allowing the duplex concentration [PT] to

be determined by equating Eq. (1.2) and (1.3).

$$[PT] = [P][T]K = [P][T]e^{-\Delta G/RT} \quad (1.5)$$

If only a small amount of duplex is formed, $[P]$ and $[T]$ are close to their initial or supplied concentrations $[P]_0$ and $[T]_0$, such that the duplex concentration is linear in ssDNA concentrations and binding affinity K .

$$\text{Reservoir condition:} \quad [PT] = [P]_0 [T]_0 K \quad (1.6)$$

The ssDNA functions as an effective reservoir, which will turn out to be the ideal situation for a hybridization-based biosensor. However, at high binding affinity and concentrations, the reduction of available ssDNA limits the formation of duplexes considerably:

$$[PT] = \underbrace{([P]_0 - [PT])}_{\text{probe saturation}} \underbrace{([T]_0 - [PT])}_{\text{target depletion}} K \quad (1.7)$$

When probes run out during the binding process, *probe saturation* occurs [35], and when probes largely capture the target, *target depletion* takes place [36, 37, 38]. Usually, target is in excess compared to probes, such that $[T]_0 \gg [PT]$ and target depletion can be neglected. In this case, the equilibrium solution becomes

$$\text{Probe saturation:} \quad [PT] = \frac{[P]_0 [T]_0 K}{1 + [T]_0 K} \quad (1.8)$$

The denominator captures the saturation effect, leading to less hybridization than in the reservoir conditions. At low binding, $[T]_0 K \ll 1$, the linear relationship is recovered. The opposite effect occurs when target becomes the limiting component, i.e. when a large amount of probes depletes the available target. As a result of the symmetry between target and probe, the concentration $[PT]$ takes on the form

$$\text{Target depletion:} \quad [PT] = \frac{[P]_0 [T]_0 K}{1 + [P]_0 K} \quad (1.9)$$

Again, at low binding $[P]_0 K \ll 1$, the linear relationship is recovered. Notably, both of these effects shift the equilibrium towards the unbound state. Counterintuitively, the reservoir condition occurs at low concentrations or affinity. In general, it is possible that the effects of probe saturation and target depletion occur simultaneously, and one should solve the quadratic form of Eq. (1.7). Such an approach leads to a general description of $[PT]$, at the cost of interpretability [39, 38].

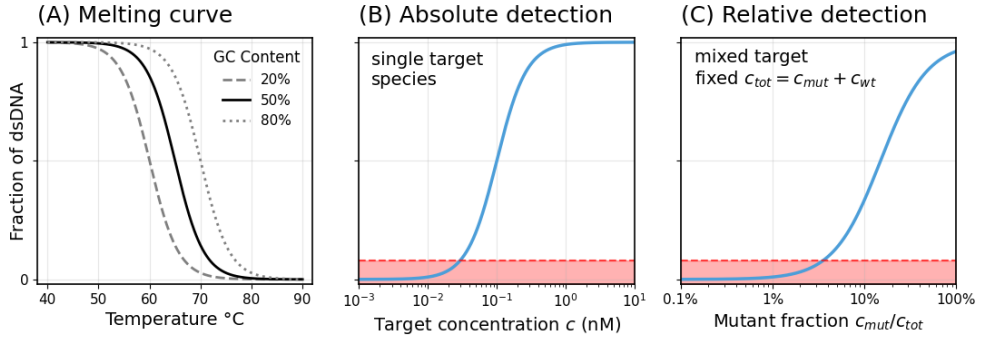


Figure 1.2: (A) DNA melting curves for sequences with different GC content, showing that higher GC content increases duplex stability and melting temperature. (B) Absolute detection: sensor response increases sigmoidally with target concentration c_t , with the dashed red line indicating the detection threshold. (C) Relative detection: response depends on mutant fraction c_{mut}/c_{tot} at fixed total concentration.

As temperature increases, the balance between stabilizing interactions, including base pairing and base stacking, and the entropic cost of duplex formation shifts towards the single-stranded state. Consequently, duplex stability decreases. This process is referred to as *DNA melting*. The temperature dependence of hybridization is typically characterized by a melting curve, obtained by measuring the amount of dsDNA as a function of temperature (Figure 1.2A). At sufficiently high temperature, hybridization becomes thermodynamically unfavorable and almost no duplexes are present. In contrast, at low temperature duplex formation is strongly favored and most ssDNA resides in the bound state, unless duplex formation is limited by probe saturation or target depletion. A useful characteristic of the melting curve is the melting temperature T_m , defined as the temperature at which half of the DNA is hybridized (i.e., half of the limiting component between $[P]_0$ or $[T]_0$). Around T_m , the duplex concentration is particularly sensitive to changes in thermodynamic parameters and concentrations, making this region well suited to study hybridization while remaining sufficiently above the background signal.

This brings us to the basic function of a biosensor: relating a change in input to a measurable change in output. A large field of DNA biosensors is concerned with absolute target detection for which our probe-target system can be used [40, 41]. Section 1.4 details some technologies that relate the amount of duplexes to a measurable output, e.g., a fluorescent

intensity, electrical readout, or even absolute number of duplex copies [42, 43, 44]. When probes are subject to an increasing amount of target, the readout results in a *dose-response curve*. For ease of notation, we denote the initial target concentration simply by $c \equiv [T]_0$. Figure 1.2B shows an example of the sensor response in function of c , taking on the familiar sigmoid curve. Above the background, the signal initially increases rapidly with increasing c , where proportional target binding occurs in the reservoir condition. As c continues to rise, probe saturation starts to limit the readout. The slope therefore decreases and the response reaches a plateau, where additional increases in c produce only small changes in signal. In this regime, the sensitivity is substantially reduced. The measurable change in output is therefore focused on the initial linear segment, defining the sensor's *dynamic range*, within which the measured signal can be inverted to quantify the target concentration c .

For low-abundance mutant detection, we need to consider two types of target, mutant T_{mut} and wild-type T_{wt} , whose initial concentrations we denote by c_{mut} and c_{wt} , respectively. Instead of absolute detection, we are concerned with the detection of mutant DNA in the presence of a large amount of wild-type DNA. Here, relative detection limits matter [45]. From an application point of view, a sample is taken at a fixed total concentration $c_{\text{tot}} = c_{\text{mut}} + c_{\text{wt}}$ and we determine if mutant DNA is present and at which relative amount. For this purpose, little emphasis is put on the total concentration, as the target DNA mixture could be amplified to the desired amount. Our approach once again exposes the mutant matching probe P to the target sample. If the sample consists of both mutant T_{mut} and wild-type sequences T_{wt} , the output reflects the total amount of hybridized complexes, as, importantly, the readout cannot distinguish between mutant or wild-type bound probes.

$$\text{Readout} \sim [PT_{\text{wt}}] + [PT_{\text{mut}}] \quad (1.10)$$

The dose-response curve now varies the mutant fraction, or equivalently the fractional abundance or variant allele frequency (VAF) given by

$$\text{Mutant fraction} = \frac{c_{\text{mut}}}{c_{\text{tot}}} \quad (0\% \rightarrow 100\%) \quad (1.11)$$

An example of a dose-response curve in relative mutant detection is illustrated on Figure 1.2C. Although this takes on a similar form as before, there are some important differences we should highlight. Instead

of adding more target going to the right, wild-type molecules are replaced by mutant ones, keeping the total concentration fixed. In contrast to absolute detection, a definite endpoint exist, where the target consists of 100% mutant DNA. At these high mutant fraction probe saturation might occur if the total concentration is high enough. For low-abundance mutant detection, we are interested in the behavior at low mutant fractions where probe saturation can be neglected. At low mutant fractions, the majority of hybridization originates from the large amount of wild-type DNA and no measurable change in output is observed between a sample with or without a small amount of mutant DNA. At a higher mutant fraction, the behavior transitions to a responsive linear region, where mutant DNA contributes significantly to the readout. We should mention the effect of target depletion: as the mutant probes have a higher affinity to probe P , they are more easily subject to the effect of depletion, which negatively impacts the dose-response curve and shifts it to the right. Interestingly, target depletion can be used to our advantage by introducing a second kind of probe that depletes wild-type over mutant [46]. This strategy of wild-type target depletion is part of Papers A and B, and will be discussed in detail in this thesis.

Under reservoir conditions, the $P + T_{\text{mut}}$ and $P + T_{\text{wt}}$ bindings are independent [39]. As above, the affinities are represented by K_{mut} and K_{wt} , or equivalently by free energies ΔG_{mut} and ΔG_{wt} . The total readout is described by

$$[P]c_{\text{wt}}K_{\text{wt}} + [P]c_{\text{mut}}K_{\text{mut}}, \quad (1.12)$$

and the relative contribution of mutant isolates an important physical parameter, $\Delta\Delta G$, that quantifies the selectivity of probe P :

$$\frac{[PT_{\text{mut}}]}{[PT_{\text{wt}}]} = \frac{c_{\text{mut}}K_{\text{mut}}}{c_{\text{wt}}K_{\text{wt}}} = \frac{c_{\text{mut}}}{c_{\text{wt}}}e^{\Delta\Delta G/RT} \quad (1.13)$$

The mutant fraction at which the system transitions from wild-type dominated to mutant dominated behavior therefore depends on the difference in free-energy between the mutant and wild-type binding, $\Delta\Delta G \equiv \Delta G_{\text{wt}} - \Delta G_{\text{mut}}$. In our case of interest, the wild-type binding involves a single-nucleotide MM while the mutant binding is perfectly matching. Therefore $\Delta\Delta G > 0$ quantifies and isolates the contribution of the MM to the duplex stability. The free-energy penalty of the MM $\Delta\Delta G$ is the central quantity in this work. Using a representative values for the penalty and

temperature, a single MM roughly reduces binding by a factor of ~ 5 -30 fold [47, 48, 49]. This might sound like a large change, but in samples with 10000 times more wild-type compared to mutant, the signal originating from mutant gets completely overshadowed by wild-type contributions.

In paper B, an elegant equation is presented that connects the penalty $\Delta\Delta G$ to the lowest detectable mutant percentage, defined as the limit of detection LoD :

$$LoD = S_{\min} e^{-\Delta\Delta G/RT}. \quad (1.14)$$

This provides a general expression that is valid for any readout mechanism, experimental conditions or DNA sequences involved. The relative detection limit is determined by two factors: the first is a technical factor, captured by $S_{\min}$, and the second is the $\Delta\Delta G$, a fundamental property of the mutation under consideration. In order to reach a low LoD , the sensor should have a small $S_{\min}$ and a large $\Delta\Delta G$. $S_{\min}$ reflects the noise characteristics of the detection system, and represents the minimum detection signal required for differentiating a pure wild-type sample from a mutant-containing mixture. In essence, the signal should be above a certain threshold to account for noise on the pure sample, which is why considerable efforts have been made to optimize this threshold. The LoD is linear in $S_{\min}$, such that improvements in measurement uncertainty proportionally reduce the LoD . In contrast, the LoD depends on $\Delta\Delta G$ in a fundamentally different way. There is an exponential dependence on $\Delta\Delta G$, such that small increments to $\Delta\Delta G$ lead to exponential improvements in the LoD .

The mismatch penalty $\Delta\Delta G$ is, however, fully determined by the physico-chemical properties of the DNA mutation involved, suggesting that $\Delta\Delta G$ is fixed and imposes a hard lower limit on the LoD . For many mismatches, the $\Delta\Delta G$ is significantly large such that no detrimental implication occurs due to this limit, but some mismatches are close to the stability of the perfectly matching situation, having a small $\Delta\Delta G$ and a limit on the LoD resulting from a large amount of cross-hybridization.

1.4 Readout Techniques

A variety of physical and biochemical techniques have been developed to detect and quantify nucleic acids. Some of the earliest methods rely on basic physical properties of DNA. For example, DNA has strong ultraviolet absorption around 260 nm, allowing concentration measurements through the Lambert-Beer law. Furthermore, structural differences between ssDNA and dsDNA lead to measurable optical effects: ssDNA absorbs approximately 30% more ultraviolet light than dsDNA, allowing the two to be distinguished. Early analytical methods also exploited differences in mobility during electrophoresis, that lead to nucleic acids of different sizes or conformations being separated and detected.

Over time, more advanced detection technologies have emerged, broadly categorized into three major types: hybridization-based assays, polymerase chain reaction (PCR), and next-generation sequencing (NGS) [50]. These methods differ in their mechanisms for recognition, readout, and operation, but all ultimately rely on the ability to identify DNA sequences specifically. They are often combined to leverage each other's strengths and achieve higher performance and versatility [51, 52]. It should be stressed that hybridization is frequently a crucial step, and improvements in selective hybridization therefore have implications across a wide range of applications.

A major development in nucleic-acid detection was the introduction of covalent labeling. In many assays, reporter molecules, most commonly fluorophores, are attached to nucleic acid strands, allowing their presence to be detected through fluorescence. The fluorescence signal serves as a proxy for strand concentration and can also provide spatial information if used on imaging-based platforms. Fluorescence-based detection also enabled multiplex analysis for the parallel detection of multiple specified sequences [53]. On the other hand, label-free approaches remain attractive due to their simplicity, reduced cost, and the absence of chemical modification of nucleic acids. Examples include techniques that detect changes in mass such as quartz crystal microbalance sensors, changes in refractive index in surface-plasmon-resonance-based platforms or even changes in heat-transfer resistance upon denaturation [54, 55, 43].

In the context of this thesis, the primary interest is not the development of new readout methods but improving the fundamental recognition process as described in the previous chapter. The hybridization processes must however be translated into measurable signals. Below, we briefly introduce several readout techniques that are particularly relevant to our work.

1.4.1 The Microarray

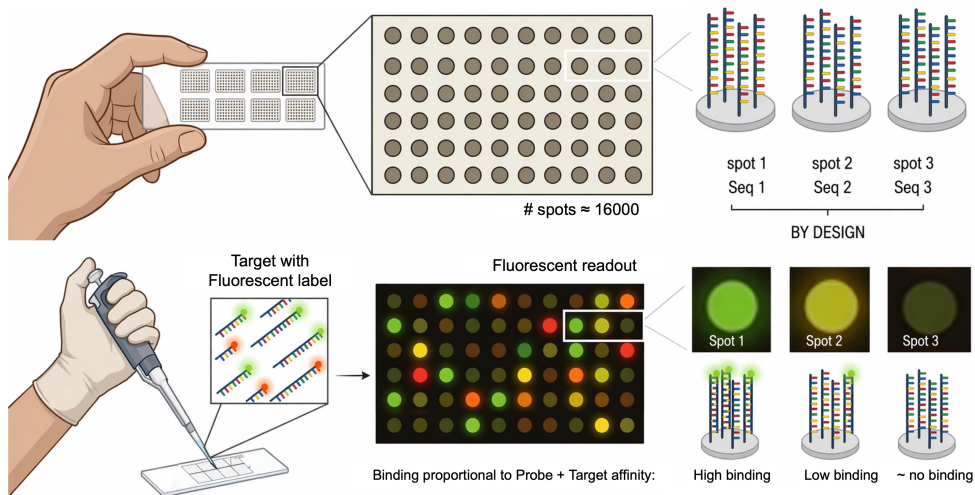


Figure 1.3: DNA microarrays rely on parallel hybridization to many probes immobilized on a solid substrate. Each spot contains millions of copies of a predefined probe sequence to which fluorescently labeled targets are allowed to hybridize. The degree of binding depends on probe-target affinity, such that after washing to remove unbound strands, fluorescence intensity provides a quantitative readout of binding strength.

The ability to spatially localize bound DNA strands forms the basis of DNA microarray technology [56]. A microarray consists of tens of thousands of spatially separated spots of immobilized ssDNA probes on a surface substrate (Figure 1.3). Each spot contains probes of a known and predefined sequence. A fluorescently labeled target sample, potentially a complex mixture, is deposited on the array under controlled hybridization conditions. After equilibration, unbound target strands are removed by washing, and the fluorescence intensity at each spot is quantified using laser scanning. As hybridization is mostly sequence-specific, the fluores-

cence intensity at each location reflects the relative abundance of complementary sequences in the sample. Microarrays have been extensively used for gene expression profiling and genotyping. Although many diagnostic applications have increasingly shifted toward sequencing-based approaches, microarrays still retain an important role in controlled studies where many predefined sequences are tested in parallel. A key limitation is that microarrays must be designed in advance: they can only detect sequences for which corresponding probes are included on the array. Sequencing, in contrast, reports the sequences present in a sample more directly and provides much richer information. As sequencing costs have decreased substantially in recent years, this has further shifted many applications toward sequencing-based technologies [57]. In addition, microarray measurements can be affected by surface effects, limited dynamic range, and cross-hybridization. But, while cross-hybridization complicates quantitative gene expression analysis and rare mutation detection, it also offers a controlled framework for studying the stability decrease associated with single-base mismatches. In particular, microarrays provide a powerful platform to systematically probe mismatch effects [48]. By designing probe sets that include the perfectly matching sequence alongside all possible single-nucleotide substitutions, one can quantitatively evaluate cross-hybridization, mismatch position dependence, and sequence-context effects.

1.4.2 The Molecular Beacon

In addition to fluorophore labeling, DNA strands can also be modified with a quencher molecule. A quencher absorbs the fluorescence emitted by a nearby fluorophore. This mechanism extends the utility of fluorophores: whereas a fluorophore alone reports the presence or location of a labeled molecule, the combination of a fluorophore and quencher enables measurement of their relative proximity. A straightforward implementation is to attach a fluorophore to the probe strand and place a quencher on the complementary base of the target strand (Figure 1.4). Hybridization then brings the fluorophore and quencher into close proximity, resulting in quenching of the fluorescence signal. Consequently, the observed fluorescence intensity becomes inversely correlated with the amount of hybridization taking place. An alternative design places both modifications

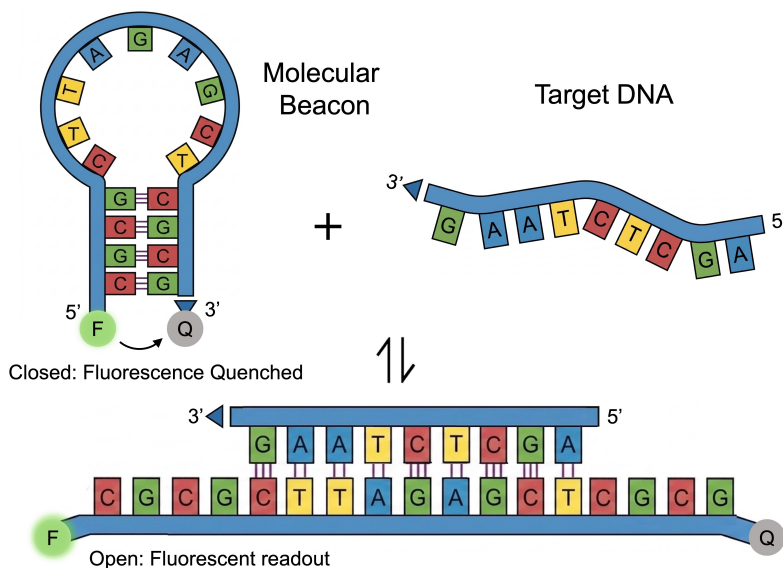


Figure 1.4: Schematic of molecular beacon-based DNA readout. In its closed formation, the beacon adopts a stem-loop hairpin structure that brings the fluorophore (F) and quencher (Q) into close proximity, suppressing fluorescence. Hybridization with a target sequence opens the hairpin, separating F and Q and resulting in a fluorescent readout.

on the same ssDNA, typically near the 5' and 3' termini of a probe strand. In this configuration, fluorescence depends on the distance between the fluorophore and quencher, which reflects the secondary structure and spatial conformation of the ssDNA. A more advanced approach uses donor-acceptor fluorophore pairs that enable Förster resonance energy transfer, where transfer efficiency is sensitive to their separation [58]. These conformational dynamics are commonly investigated using single-molecule fluorescence techniques.

A particularly important conformation is the hairpin structure, in which complementary sequences at the ends of the ssDNA form a short double-stranded stem while leaving a single-stranded loop region exposed. If the loop sequence is designed to be complementary to a target strand, the resulting probe is known as a Molecular Beacon (MB) [59]. In the absence of target, the hairpin conformation keeps the fluorophore and quencher close to each other, suppressing fluorescence. When the target is present, it hybridizes with the loop region, which opens the hairpin. This structural transition separates the fluorophore from the quencher, restor-

ing fluorescence and thereby producing a detectable signal. This design offers several advantages: the stability and responsiveness of the hair-pin can be tuned by adjusting the stem length, allowing optimization of sensitivity and specificity, and no labeling of the target strand is required, simplifying assay design and reducing costs [60, 32].

1.4.3 Polymerase Chain Reaction (PCR)

A widely used strategy is amplification via Polymerase Chain Reaction (PCR) [61]. In PCR, target DNA is amplified using synthetic DNA primers and the DNA polymerase enzyme [62]. The reaction is driven by repeated thermal cycling between high and low temperatures (Figure 1.5). During the denaturation step (typically $\sim 95^\circ\text{C}$), dsDNA gets separated into ssDNA. After being cooled to the annealing temperature, primers hybridize to their complementary target sequences, and the DNA polymerase extends the primers, synthesizing complementary strands as new dsDNA. Each cycle ideally results in a doubling of the target DNA, leading to exponential amplification. Importantly, amplification efficiency is determined primarily by primer-target complementarity, and the sequence internal to the primer region does not influence replication unless it affects primer binding or polymerase progression. Consequently, conventional PCR amplifies both wild-type and mutant targets with similar efficiency when primer binding regions are conserved.

For detection purposes, accumulation of amplified DNA can be monitored in real time using fluorescence-based methods in quantitative PCR (qPCR), where signal intensity correlates with the amount of duplex at each cycle. The initial target concentration is quantitatively described by the number of cycles C_t or C_q required for the intensity to exceed a threshold value. Alternatively, analysis can be performed after a predefined number of cycles. To enable relative quantification in mixed samples, parallel reactions are often performed using mutant-specific and wild-type-specific primers, each labeled with spectrally distinct fluorophores. The difference in quantification cycles, ΔC_q , then reflects the relative abundance of mutant versus wild-type. However, imperfect specificity due to cross-hybridization, primer dimer formation, or late-cycle nonspecific amplification can limit sensitivity, particularly when the mutant fraction is very low [50].

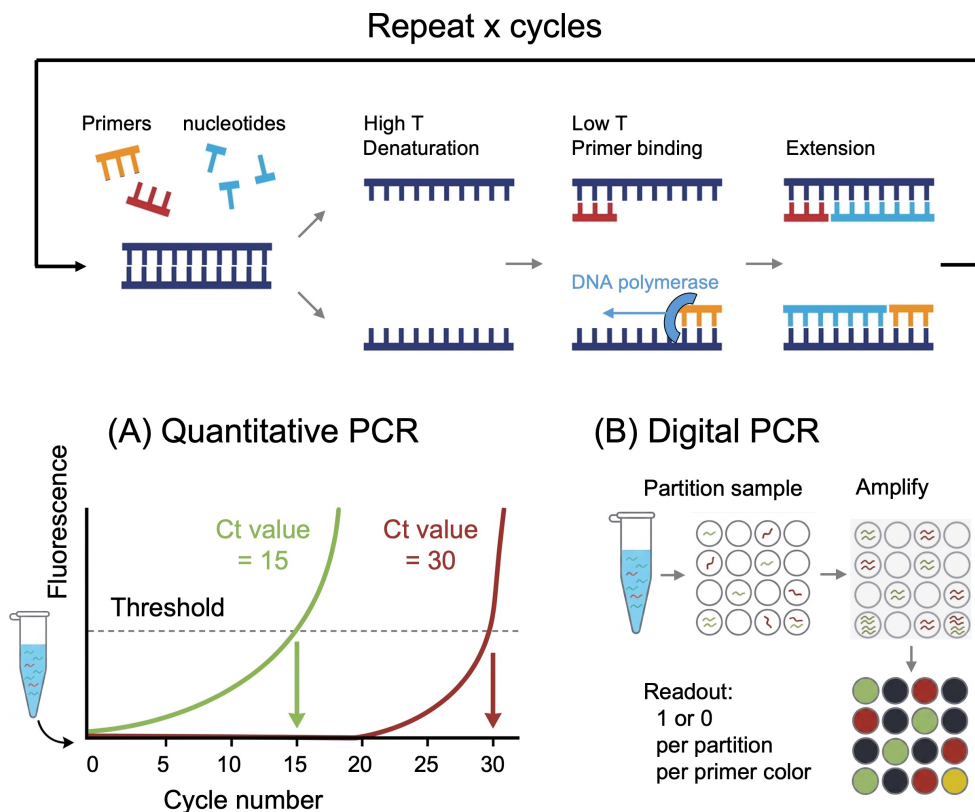


Figure 1.5: Polymerase Chain Reaction (PCR)-based amplification and detection. Repeated cycles of DNA denaturation and extension by DNA polymerase amplify an input target sequence. (A) Quantitative PCR (qPCR) monitors fluorescence during amplification, where a lower threshold cycle (Ct) value indicates a higher initial target concentration. (B) Digital PCR partitions samples into discrete partitions for quantification based on positive/negative reads in each partition.

A more precise and robust approach for rare variant detection is digital PCR (dPCR) [63, 64]. Here, the sample is partitioned into a large number (typically >10000) of nanoliter-scale volumes such that individual partitions contain only a small amount of target molecules. After amplification, each partition is classified as positive (1) or negative (0), depending on whether amplification occurred. Absolute target concentration is then inferred using Poisson statistics to account for the probability of multiple molecules per partition. As in qPCR, target-specific primers or probes labeled with distinct fluorophores enable simultaneous quantification of mutant and wild-type targets within the same assay. Because dPCR relies on

binary classification rather than amplification kinetics, it is less sensitive to variations in amplification efficiency and is widely regarded as a gold standard for low-abundance mutation detection. Detection limits on the order of $\sim 0.5\%$ mutant fraction in a wild-type background are routinely reported, making dPCR viable for applications such as minimal residual disease monitoring and liquid biopsy analysis [65].

1.4.4 Next-generation Sequencing (NGS)

Sequencing aims at determining the exact nucleotide order of DNA molecules within a sample. In contrast to hybridization-based assays, which detect the presence of specific sequences through binding, sequencing provides direct base-by-base information.

The first widely adopted method was Sanger sequencing, developed in the 1970s and ultimately used to sequence the entire Human Genome Project [66, 67]. The method is based on chain termination: during DNA synthesis by DNA polymerase, modified nucleotides are occasionally incorporated that stop further elongation. This generates DNA fragments of different lengths, each terminating at a specific base. The fragments are separated by size using electrophoresis, and fluorescence detection uniquely identifies the terminal nucleotide at each position. From this information, the DNA sequence can be reconstructed. Although highly accurate, Sanger sequencing is low-throughput, requires a relatively pure DNA sample containing essentially a single sequence, and is not well-suited for complex mixtures of different sequences.

Next-generation sequencing (NGS) technologies overcome this limitation by massively parallelizing the sequencing process [57, 68] (Figure 1.6). In widely used platforms such as those developed by Illumina, amplified DNA fragments (possibly mixtures) are immobilized on a flow cell and sequenced simultaneously. The chemistry is highly controllable, so that nucleotides are incorporated one base at a time. Each event is detected by fluorescence imaging, allowing the identity of the added base to be recorded before the next cycle begins. This cyclic imaging process generates short sequence reads for each spatially separated cluster. Alternative NGS approaches use different detection principles, such as semiconductor-

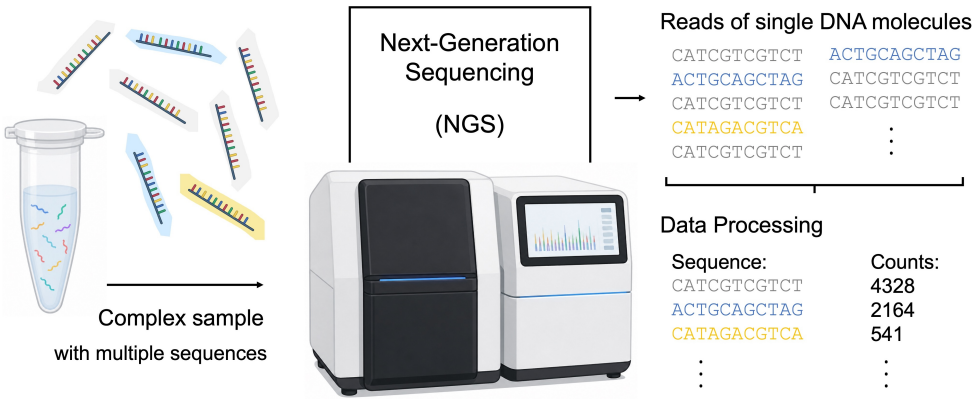


Figure 1.6: Next-generation sequencing (NGS) workflow for analysis of complex DNA mixtures. A sample containing multiple DNA sequences is processed and generates large numbers of DNA sequence “reads” from individual DNA strands. Data processing then accurately reflects relative abundances in the sample.

based sensing or single-molecule real-time sequencing [69, 70]. NGS achieves some of the most sensitive detection limits currently available, often reaching mutant fractions below 1%, and in optimized workflows (e.g., with molecular barcoding and error suppression strategies) substantially lower [44]. However, practical sensitivity is constrained by sequencing errors arising from polymerase misincorporation, incomplete fluorophore cleavage, and bioinformatic errors. Additionally, NGS requires substantial infrastructure, significant upfront investment in instrumentation, and non-negligible per-run costs. Library preparation, amplification bias, and data analysis complexity further contribute to variability. Despite these limitations, NGS has become a central tool in genomics, oncology, and rare variant detection, providing an excellent combination of multiplexing capacity, quantitative readout, and nucleotide-level resolution.

The techniques discussed above illustrate only a subset of the many available strategies for nucleic-acid detection. Despite their different readout mechanisms, they all depend, directly or indirectly, on the thermodynamic properties of DNA hybridization. To understand and predict DNA hybridization, the next section introduces the nearest-neighbor model, which provides a practical framework for estimating duplex stability from a given DNA sequence.

1.5 Nearest-Neighbor model

The Modeling of DNA can be approached in many different ways, each with distinct goals and levels of detail. At the most detailed level, all-atom molecular simulations are used to describe structural and dynamical properties of DNA, such as DNA bending, base-pair opening, major and minor groove dynamics, and DNA hydration [71, 27]. All-atom simulations are computationally demanding, which is why coarse-grained models have been developed that describe not every atom, but view DNA as simplified building blocks representing e.g. the backbone or specific nucleotides [72, 73]. This reduction allows simulations of larger systems and longer timescales. Typically, coarse-grained models rely on parameters obtained from experiments or from all-atom simulations. Some of these approaches can also provide estimates of duplex stability and melting behavior [74].

While these provide valuable insights into the mechanics and dynamics of DNA, they are often unnecessarily complex for estimating the thermodynamic stability of hybridization (ΔG). The nearest-neighbor (NN) model provides a practical method for predicting the stability of DNA based on their sequence and hybridization conditions [75, 47, 76]. It forms the basis of many widely used prediction tools in biotechnology, for calculating melting temperatures, designing primers, or determining secondary structures.

The central assumption of the NN model is that the stability of a DNA duplex is determined by hydrogen bonding between opposing bases and stacking interactions between adjacent base pairs. Importantly, the stacking interactions depend on the identity of neighboring base pairs, but not on the sequence further away, such that the NN model is a local theory. As a result, the free energy can be expressed as a sum of contributions from neighboring base-pair steps. For example, the free-energy change of duplex formation can be written as

$$\Delta G \begin{pmatrix} GCATG \\ CGTAC \end{pmatrix} = \Delta G \begin{pmatrix} GC \\ CG \end{pmatrix} + \Delta G \begin{pmatrix} CA \\ GT \end{pmatrix} + \Delta G \begin{pmatrix} AT \\ TA \end{pmatrix} \quad (1.15)$$

$$+ \Delta G \begin{pmatrix} TG \\ AC \end{pmatrix} + \Delta G_{\text{init}},$$

$$= -2.24 - 1.45 - 0.88 - 1.45 + 1.96 = -4.06 \text{ kcal/mol} \quad (1.16)$$

where the top strand is always in the 5'->3' orientation and the bottom strand 3'->5'. The neighboring base pairs $\Delta G(XY/X'Y')$ form the NN parameters [47]. Of the 16 possible combinations of two consecutive matching base pairs, some symmetry exists, reducing the variables to 10 independent parameters. On top of these, the model uses an initiation parameter and a correction for terminal AT base pairs. These parameters have been determined using multiple linear regression fluorescent measurement on ~ 100 sequences and have been refined since. Additionally, repeating such an analysis allows the separate determination of enthalpy (ΔH) and entropy (ΔS), allowing the estimation of ΔG at an arbitrary temperature.

Beyond perfectly complementary pairing, the NN model can also predict duplex stability involving mismatches. This results in 48 additional NN parameters. Analogous as before, the free energy of a mismatched duplex can be calculated as

$$\Delta G \left(\begin{smallmatrix} GC \underline{A} TG \\ CG \underline{C} AC \end{smallmatrix} \right) = \Delta G \left(\begin{smallmatrix} GC \\ CG \end{smallmatrix} \right) + \Delta G \left(\begin{smallmatrix} C \underline{A} \\ G \underline{C} \end{smallmatrix} \right) + \Delta G \left(\begin{smallmatrix} \underline{A} T \\ \underline{C} A \end{smallmatrix} \right) \quad (1.17)$$

$$+ \Delta G \left(\begin{smallmatrix} T G \\ A C \end{smallmatrix} \right) + \Delta G_{\text{init}},$$

$$= -2.24 + \underline{0.75} + \underline{0.77} - 1.45 + 1.96 = -0.21 \text{ kcal/mol} \quad (1.18)$$

where mismatched base pairs are underscored. Notice that, while this is not always the case, the mismatched NN parameters are not just lower than their matching variant, but even positive. Equation (1.15) and (1.17) may represent the binding of a detection probe to the mutant and wild-type sequences, such that a comparison leads to the quantitative measure of selectivity, $\Delta\Delta G$. Terms besides those involving the MM cancel, and we are left with

$$\Delta\Delta G \left(\begin{smallmatrix} C \underline{A} T \\ G \underline{C} A \end{smallmatrix} \right) = 3.85 \text{ kcal/mol} \quad (1.19)$$

The binding difference here isolates the impact of the mismatch. It is the mismatch penalty, such that

$$\Delta G_{WT} = \Delta G_{MT} + \Delta\Delta G \quad (1.20)$$

Notice here that the binding difference is given by the local impact of the mismatch and is not dependent on the binding length (which changes both ΔG 's with the same amount). We should note that ΔG does affect the total

amount of binding in both cases, and should usually be tuned to avoid probe saturation and target depletion.

One implication of the local nature of the NN model is that multiple MM contribute additively to the total stability. The free-energy penalty of having two MMs is estimated to be equal to the sum of their individual penalties. Previous research has studied this property and suggests that additivity breaks down when MMs are close and interact with each other [48]. On average, an exponentially decaying stabilizing interaction takes place between two MMs depending on their separation, revealing behavior beyond the NN model. In contrast, the interaction can become destabilizing under specific conditions. In such cases, the deliberate introduction of a second MM can lead to probe designs with improved selectivity (larger $\Delta\Delta G$). Paper C investigates this property in the context of mutant DNA detection.

1.6 Aims of this Thesis

This doctoral work aims to bridge the gap between thermodynamic insights into DNA hybridization and the development of new, highly selective detection strategies. The overarching objective is to improve the functionality and detection limits of hybridization-based sensors, starting from an understanding of DNA hybridization on the thermodynamic level. The work is motivated by the challenge of detecting low-abundance mutations in liquid biopsies, where rare mutant sequences must be identified against a large background of wild-type DNA. In such systems, detection is fundamentally constrained by the binding free-energy difference $\Delta\Delta G$, which captures the discrimination between the desired and undesired binding. Several strategies explored in this thesis therefore aim to effectively increase this beyond the value imposed by a single mismatch. Rather than focusing on the development of a specific technology or method, this research emphasizes general concepts that are transferable across different platforms. More broadly, this work highlights the relation between theoretical insights and the design of hybridization systems. In this sense, this thesis contributes to an improved understanding of how sequence-dependent interactions, penalties, and structural constraints determine selectivity in DNA hybridization.

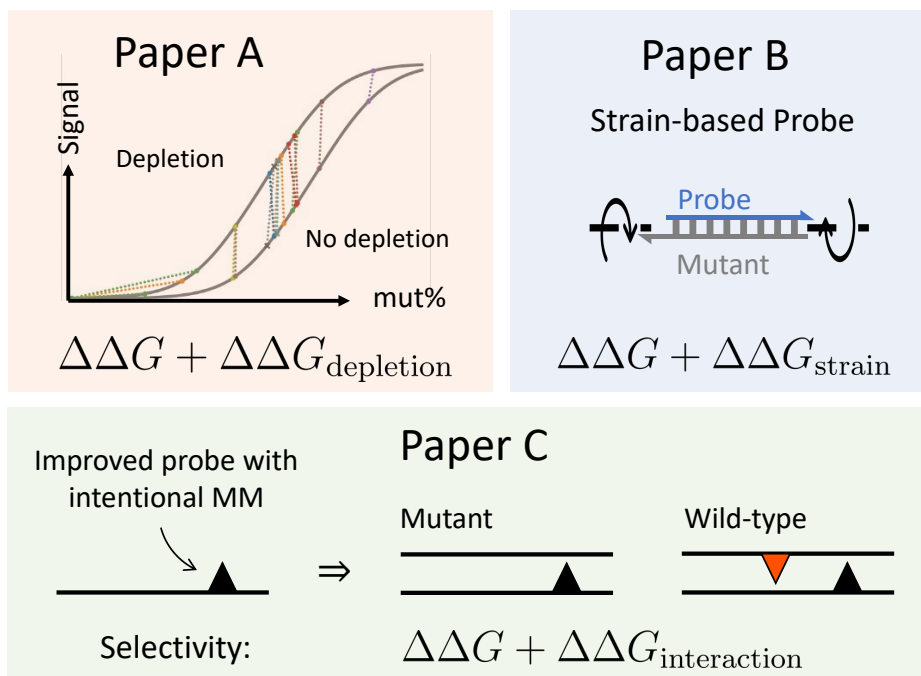


Figure 1.7: Overview of the equilibrium-based methods presented in this thesis. Papers A, B, and C introduce three strategies to enhance the detection of low-abundance mutants. Selectivity is quantified through the free-energy difference $\Delta\Delta G$ and is improved via depletion interactions, probe-induced strain, or interactions involving an intentionally introduced second mismatch (MM).

In [paper A](#), the concept of wild-type target depletion is presented on a microarray platform using clinical patient samples from the Jessa Hospital in Hasselt, Belgium. The detection method relies on a comparison to reference probes measured simultaneously in parallel, enabling the definition of a robust ratio-based detection signal [39, 46]. The thermodynamic description connects experimental readout to the $\Delta\Delta G$ of the involved MM and we investigate if the signal can be inverted to accurately determine the amount of mutant (%) present in a sample. To show the impact of wild-type target depletion, a comparison is made between a setup where a large amount of wild-type matching depletion probes are present versus where they are absent. This enables the quantification of the depletion effect, and leads to a ~ 14 -fold improvement in the smallest detectable mutant percentage. Remarkably, for samples containing the highest amount of mutant, an apparent disagreement in predicted mutant percentage was visible between the method with versus without depletion probes. We pro-

pose an extended model that includes an additional interaction between the reference probes and the targets, that correctly predicts a reduction in signal at the highest mutant fractions and recovers consistency between the two methods (Figure 1.7A). The sensor therefore has a large dynamic range. Interestingly, this extension introduces a second MM to the model, whose destabilizing effect is again present as a mismatch penalty $\Delta\Delta G$. The penalty is lower than expected from NN model calculations, consistent with the fact that it is present in close proximity to another mismatch, generally lowering its destabilizing penalty [48]. In summary, we present a robust method for low-abundant mutant detection utilizing wild-type depletion, where thermodynamic insights allowed us to describe all interactions and enable a large dynamic range.

Paper B employs an additional strategy to increase the selectivity of mutant detection. On top of wild-type depletion by 'blockers', mutant probes are modified to incorporate strain that selectively reduces the binding affinity of wild-type over mutant target (Figure 1.7B) [77]. Under optimal detection conditions, the effective $\Delta\Delta G$ is then given by the sum of three distinct contributions: the mismatch penalty, strain contribution, and the blocker interaction. The thermodynamic description allows the $\Delta\Delta G$ to be fitted, and we investigate if the individual contributions to $\Delta\Delta G$ can be quantified. We clearly present the relation between $\Delta\Delta G$ and the lowest detectable mutant percentage, framing how improvements to $\Delta\Delta G$ beyond the mismatch penalty improve the detection limit. Notably, the blocker provided a dual effect: in addition to increasing $\Delta\Delta G$, the presence of blocker suppressed mutant-probe-induced depletion that negatively impacts detection, and therefore the optimal detection conditions are more easily reached when blocker is present. On top of these insights, we operate in-solution at room temperature and reach a detection limit of 0.058% mutant in a wild-type background, illustrating the potential for cost-effective application of the involved methods [33].

The research of **Paper C** is concerned with non local effects of mismatches in the DNA duplex. In this case, the penalty of a MM is different when another MM is present somewhere in its vicinity which can be described as using an interaction taking place between the MMs (Figure 1.7C). Usually, this interaction is stabilizing, such that the presence of another MM lowers the impact of the MM relevant for distinguishing SNVs, and con-

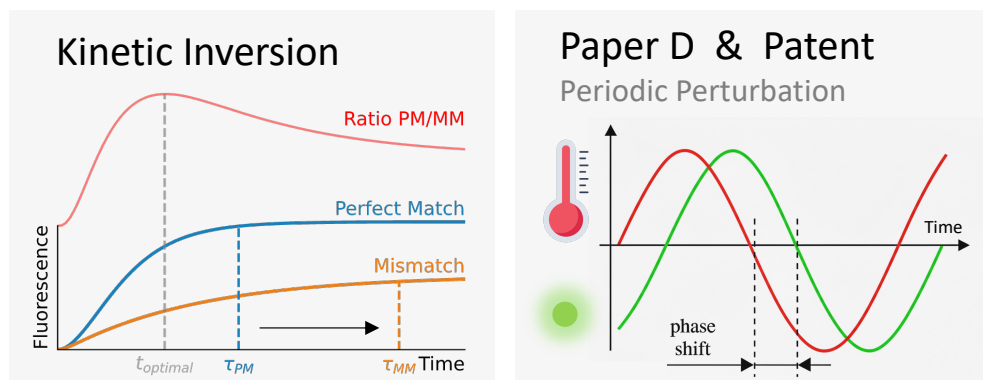


Figure 1.8: Non-equilibrium approaches for studying hybridization systems. Under specific reaction conditions, kinetic inversion can occur, where the perfect match (PM) reaches equilibrium faster than the mismatched (MM) target, leading to an optimal discrimination window during the reaction. Periodic perturbations lead to dynamic responses that reveal properties inaccessible at equilibrium.

sequently lowers the selectivity of the mutant probe [48, 78]. Our research explores under what conditions this interaction is positive, and if this leads improved discrimination and probe design. For this purpose, a highly controlled sequence is used to systematically study the interactions between MMs, depending on MM type and their (relative) positions. A high-throughput sequencing approach allows parallel measurements that are quantitatively described by thermodynamics. With the obtained results, the insights are applied to two clinically relevant sequences and a consistent probe design leads to improved discrimination: the SNV should be placed in the middle of the detection probe, and the intentional MM near one of the probe termini. To validate this design, the best probes are tested on a separate detection platform that uses a different readout mechanism [42]. The interaction strengths reach close to 20% of the size of a typical MM penalty, and the resulting probe selectivity is observed to improve ~ 3 times.

In addition to equilibrium measurements, studying how binding changes over time can provide useful information about a DNA hybridization system and be used to distinguish mutant versus wild-type samples. The speeds of binding and unbinding can differ between a perfectly matching sequence and one with a mismatch, and by tracking these kinetic differences, mutant and wild-type sequences can be distinguished [79, 80]. As the mismatched

binding dissociates more easily, its timescale is shorter and reaches equilibrium faster compared to a matching probe-target pair. An important consequence of this is that mutant and wild-type target are most easily discriminated at equilibrium. Chapters 5 challenges this, where instead of a simple probe-target system, a multi-step hybridization system is presented, based on the sensor employed in Paper B. We investigate an effect termed 'kinetic inversion' where, under certain conditions, the matching system is faster compared to the mismatched situation [81]. An important consequence is that in the case of kinetic inversion, optimal discrimination happens not at equilibrium but at a specified time during the kinetics. The optimal time can be controlled by changing concentrations and other design parameters, which allows the kinetic response and signal magnitude to be fine-tuned. As a result, kinetic inversion could enable faster detection while improving the trade-off between sensitivity and selectivity in existing biosensors.

Finally, we present here a brief overview of an out-of-equilibrium method that takes a very different approach. By continuously perturbing the DNA system, it constantly adapts and reacts to reach its ever-changing equilibrium state. This enables a continuous monitoring of kinetics that are related to the properties of the sample. Specifically, in the system under consideration, we periodically perturb the temperature with a given frequency (Figure 1.8B). If this frequency is low, the system is in quasi-equilibrium, and the response nicely follows the applied perturbation. At higher frequencies, the system fails to catch up, and a lag appears between the applied temperature oscillation and the response. A frequency response of the measured amplitude and phase lag characterizes the system. A key strength of the technology is the continuous acquisition of kinetic data, allowing the noise to be reduced and the detection limit to be improved by increasing the measurement duration. This method of characterizing properties of hybridization systems is the result of a combined effort and research topics in our research group, Theory Lab. This has led to a patent application, of which I am co-inventor, as well as a technical paper (Paper D) presenting a cost-effective experimental platform for implementing the method. Several promising research directions are currently being explored. However, further work on this topic is not discussed in this thesis in order to preserve potential intellectual property as they represent promising opportunities for future research and technology.

References

- [1] S. M. Batool *et al.*, "The Liquid Biopsy Consortium: Challenges and opportunities for early cancer detection and monitoring," *Cell Reports Medicine*, vol. 4, no. 10, 2023.
- [2] U.S. Cancer Statistics Working Group, "U.S. Cancer Statistics Data Visualizations Tool," <https://www.cdc.gov/cancer/dataviz>, 2025, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute.
- [3] C. Greenman *et al.*, "Patterns of somatic mutation in human cancer genomes," *Nature*, vol. 446, no. 7132, pp. 153–158, 2007.
- [4] J. T. Loud and J. Murphy, "Cancer Screening and Early Detection in the 21st Century," *Seminars in Oncology Nursing*, vol. 33, no. 2, pp. 121–128, 2017.
- [5] G. Alexandrou *et al.*, "The Evolution of Affordable Technologies in Liquid Biopsy Diagnostics: The Key to Clinical Implementation," *Cancers*, vol. 15, no. 22, p. 5434, 2023.
- [6] M. Peifer *et al.*, "Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer," *Nature Genetics*, vol. 44, no. 10, pp. 1104–1110, 2012.
- [7] L. A. Diaz and A. Bardelli, "Liquid biopsies: Genotyping circulating tumor dna," *Journal of Clinical Oncology*, vol. 32, no. 6, pp. 579–586, 2014.
- [8] S. Siena, A. Sartore-Bianchi, F. Di Nicolantonio, J. Balfour, and A. Bardelli, "Biomarkers Predicting Clinical Outcome of Epidermal Growth Factor Receptor-Targeted Therapy in Metastatic Colorectal

- Cancer," *JNCI: Journal of the National Cancer Institute*, vol. 101, no. 19, pp. 1308–1324, 2009.
- [9] S. Misale *et al.*, "Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer," *Nature*, vol. 486, no. 7404, pp. 532–536, 2012.
- [10] W. Xie, S. Suryaprakash, C. Wu, A. Rodriguez, and S. Fraterman, "Trends in the use of liquid biopsy in oncology," *Nature Reviews Drug Discovery*, vol. 22, no. 8, pp. 612–613, 2023.
- [11] S. Udomruk, S. Orrapin, D. Pruksakorn, and P. Chaiyawat, "Size distribution of cell-free DNA in oncology," *Critical Reviews in Oncology/Hematology*, vol. 166, p. 103455, 2021.
- [12] M. García-Pardo, M. Makarem, J. J. N. Li, D. Kelly, and N. B. Leighl, "Integrating circulating-free DNA (cfDNA) analysis into clinical practice: opportunities and challenges," *British Journal of Cancer*, vol. 127, no. 4, pp. 592–602, 2022.
- [13] J. Lu, Y. Feng, K. Guo, L. Sun, S. Ruan, and K. Zhang, "Prognostic value of preoperative circulating tumor DNA in non-small cell lung cancer: a systematic review and meta-analysis," *Journal of Cancer Research and Clinical Oncology*, vol. 150, no. 1, p. 25, 2024.
- [14] L. Sivapalan, J. C. Murray, J. V. Canzoniero, B. Landon, J. Jackson, S. Scott, V. Lam, B. P. Levy, M. Sausen, and V. Anagnostou, "Liquid biopsy approaches to capture tumor evolution and clinical outcomes during cancer immunotherapy," *Journal for Immunotherapy of Cancer*, vol. 11, no. 1, p. e005924, 2023.
- [15] O. B. Alese, N. Cook, A. Ortega-Franco, M. B. Ulanja, L. Tan, and J. Tie, "Circulating Tumor DNA: An Emerging Tool in Gastrointestinal Cancers," *American Society of Clinical Oncology Educational Book*, no. 42, pp. 279–298, 2022.
- [16] H. Willems, A. Jacobs, W. W. Hadiwikarta, T. Venken, D. Valkenburg, N. Van Roy, J. Vandesompele, and J. Hooyberghs, "Thermodynamic framework to assess low abundance DNA mutation detection by hybridization," *PLOS ONE*, vol. 12, no. 5, p. e0177384, 2017.

- [17] E. Drenkard and F. M. Ausubel, "Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation," *Nature*, vol. 416, no. 6882, pp. 740–743, 2002.
- [18] H. Yamamoto, J. Allen, G. Wilson, P. Zuzarte, J. Simpson, S. Keshavjee, and J. Yeung, "Rapid Donor-Specific Single Nucleotide Variation Detection by Nanopore Sequencing of Ex Vivo Lung Perfusate," *The Journal of Heart and Lung Transplantation*, vol. 41, no. 4, pp. S256–S257, 2022.
- [19] J. Böhmer, H. Wåhlander, K. Karason, J. Sunnegårdh, C. Wasslavik, M. Jonsson, J. Asp, A. Ricksten, and G. Dellgren, "Clinical Examples of the Additive Value of Absolute Quantification of Cell-Free DNA After Heart Transplantation," *Clinical Transplantation*, vol. 38, no. 10, p. e15477, 2024.
- [20] C. F. Wright and H. Burton, "The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis," *Human Reproduction Update*, vol. 15, no. 1, pp. 139–151, 2008.
- [21] M. N. Stojanovic, D. Stefanovic, and S. Rudchenko, "Exercises in Molecular Computing," *Accounts of Chemical Research*, vol. 47, no. 6, pp. 1845–1852, 2014.
- [22] A. C. Bardales, Q. Vo, and D. M. Kolpashchikov, "Singleton {NOT} and Doubleton {YES; NOT} Gates Act as Functionally Complete Sets in DNA-Integrated Computational Circuits," *Nanomaterials*, vol. 14, no. 7, p. 600, 2024.
- [23] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. New York: Garland Science, 2002, available from: <https://www.ncbi.nlm.nih.gov/books/NBK21054/>.
- [24] J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.
- [25] D. P. Bartel, "MicroRNAs: Target Recognition and Regulatory Functions," *Cell*, vol. 136, no. 2, pp. 215–233, 2009.

- [26] C. A. Plata, S. Marni, A. Maritan, T. Bellini, and S. Suweis, "Statistical physics of DNA hybridization," *Physical Review E*, vol. 103, no. 4, p. 042503, 2021.
- [27] G. Rossetti, P. D. Dans, I. Gomez-Pinto, I. Ivani, C. Gonzalez, and M. Orozco, "The structural impact of DNA mismatches," *Nucleic Acids Research*, vol. 43, no. 8, pp. 4309–4321, 2015.
- [28] E. Friedberg, G. Walker, W. Siede, and R. Wood, *DNA Repair and Mutagenesis*, ser. American Society Mic Series, 2005.
- [29] K. Fukui, "DNA Mismatch Repair in Eukaryotes and Bacteria," *Journal of Nucleic Acids*, vol. 2010, no. 1, p. 260512, 2010.
- [30] M. Sharma, A. V. Predeus, S. Mukherjee, and M. Feig, "DNA Bending Propensity in the Presence of Base Mismatches: Implications for DNA Repair," *The Journal of Physical Chemistry B*, vol. 117, no. 20, pp. 6194–6205, 2013.
- [31] J. J. Hopfield, "Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity," *Proceedings of the National Academy of Sciences*, vol. 71, no. 10, pp. 4135–4139, 1974.
- [32] B. L. Mueller, T. A. Molden, J. Hammock, and D. M. Kolpashchikov, "Tailed molecular beacon probes: an approach for the detection of structured DNA and RNA analytes," *Chemical Communications*, vol. 61, no. 10, pp. 2095–2098, 2025.
- [33] B. L. Mueller, M. J. Liberman, and D. M. Kolpashchikov, "OWL2: a molecular beacon-based nanostructure for highly selective detection of single-nucleotide variations in folded nucleic acids," *Nanoscale*, vol. 15, no. 12, pp. 5735–5742, 2023.
- [34] G. A. Held, G. Grinstein, and Y. Tu, "Modeling of DNA microarray data by using physical properties of hybridization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 13, pp. 7575–7580, 2003.
- [35] D. Hekstra, "Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays," *Nucleic Acids Research*, vol. 31, no. 7, pp. 1962–1968, 2003.

-
- [36] C. J. Burden and H. Binder, "Physico-chemical modelling of target depletion during hybridization on oligonucleotide microarrays," *Physical Biology*, vol. 7, no. 1, p. 016004, 2009.
- [37] L. Zhang, M. F. Miles, and K. D. Aldape, "A model of molecular interactions on short oligonucleotide microarrays," *Nature Biotechnology*, vol. 21, no. 7, pp. 818–821, 2003.
- [38] N. Ono, S. Suzuki, C. Furusawa, T. Agata, A. Kashiwagi, H. Shimizu, and T. Yomo, "An improved physico-chemical model of hybridization on high-density oligonucleotide microarrays," *Bioinformatics*, vol. 24, no. 10, pp. 1278–1285, 2008.
- [39] S. K. Nomidis, M. Szymonik, T. Venken, E. Carlon, and J. Hooyberghs, "Enhancing the Performance of DNA Surface-Hybridization Biosensors through Target Depletion," *Langmuir*, vol. 35, no. 37, pp. 12 276–12 283, 2019.
- [40] S. A. Marras, S. Tyagi, and F. R. Kramer, "Real-time assays with molecular beacons and other fluorescent nucleic acid hybridization probes," *Clinica Chimica Acta*, vol. 363, no. 1-2, pp. 48–60, 2006.
- [41] C. Lino, S. Barrias, R. Chaves, F. Adegas, P. Martins-Lopes, and J. R. Fernandes, "Biosensors as diagnostic tools in clinical applications," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1877, no. 3, p. 188726, 2022.
- [42] S. T. Shanmugam, S. Trashin, and K. De Wael, "Singlet oxygen-based photoelectrochemical detection of DNA," *Biosensors and Bioelectronics*, vol. 195, p. 113652, 2022.
- [43] B. van Grinsven *et al.*, "Heat-Transfer Resistance at Solid-Liquid Interfaces: A Tool for the Detection of Single-Nucleotide Polymorphisms in DNA," *ACS Nano*, vol. 6, no. 3, pp. 2712–2721, 2012.
- [44] V. Plagnol *et al.*, "Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling," *PLOS ONE*, vol. 13, no. 3, p. e0193802, 2018.
- [45] C. Guo, R. Ding, Z. Zhao, J. Guo, and F. Li, "Enrichment Strategies for Low-Abundant Single Nucleotide Mutations," *Chemistry -A European Journal*, vol. 31, no. 1, p. e202402872, 2025.
-

- [46] R. Van Hoof, M. Szymonik, S. K. Nomidis, K. Hollanders, A. Jacobs, I. Nelissen, P. Wagner, and J. Hooyberghs, "Depletion of wild-type target enhances the hybridization-based sensitivity of low-abundant mutation detection by reference capture probes," *Sensors and Actuators B: Chemical*, vol. 368, p. 132175, 2022.
- [47] J. SantaLucia and D. Hicks, "The Thermodynamics of DNA Structural Motifs," *Annual Review of Biophysics and Biomolecular Structure*, vol. 33, no. 1, pp. 415–440, 2004.
- [48] W. W. Hadiwikarta, J.-C. Walter, J. Hooyberghs, and E. Carlon, "Probing hybridization parameters from microarray experiments: nearest-neighbor model and beyond," *Nucleic Acids Research*, vol. 40, no. 18, p. e138, 2012.
- [49] J. S. Wang and D. Y. Zhang, "Simulation-guided DNA probe design for consistently ultraspecific hybridization," *Nature Chemistry*, vol. 7, no. 7, pp. 545–553, 2015.
- [50] D. Khodakov, C. Wang, and D. Y. Zhang, "Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches," *Advanced Drug Delivery Reviews*, vol. 105, pp. 3–19, 2016.
- [51] J. Li, L. Wang, H. Mamon, M. H. Kulke, R. Berbeco, and G. M. Makrigiorgos, "Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing," *Nature Medicine*, vol. 14, no. 5, pp. 579–584, 2008.
- [52] Y. Nagai *et al.*, "Genetic Heterogeneity of the Epidermal Growth Factor Receptor in Non-Small Cell Lung Cancer Cell Lines Revealed by a Rapid and Sensitive Detection System, the Peptide Nucleic Acid-Locked Nucleic Acid PCR Clamp," *Cancer Research*, vol. 65, no. 16, pp. 7276–7282, 2005.
- [53] R. Molenkamp, A. Van Der Ham, J. Schinkel, and M. Beld, "Simultaneous detection of five different DNA targets by real-time Taqman PCR using the Roche LightCycler480: Application in viral molecular diagnostics," *Journal of Virological Methods*, vol. 141, no. 2, pp. 205–211, 2007.

- [54] H. Yoshimine, K. Sasaki, and H. Furusawa, "Pocketable Biosensor Based on Quartz-Crystal Microbalance and Its Application to DNA Detection," *Sensors*, vol. 23, no. 1, p. 281, 2023.
- [55] A. Dillen, A. Mohrbacher, and J. Lammertyn, "A Versatile One-Step Competitive Fiber Optic Surface Plasmon Resonance Bioassay Enabled by DNA Nanotechnology," *ACS Sensors*, vol. 6, no. 10, pp. 3677–3684, 2021.
- [56] R. B. Stoughton, "Applications of DNA Microarrays in Biology," *Annual Review of Biochemistry*, vol. 74, no. 1, pp. 53–82, 2005.
- [57] G. Gydush *et al.*, "Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth," *Nature Biomedical Engineering*, vol. 6, pp. 257–266, 2022.
- [58] B. Hellenkamp *et al.*, "Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study," *Nature Methods*, vol. 15, no. 9, pp. 669–676, 2018.
- [59] W. Tan, K. Wang, and T. J. Drake, "Molecular beacons," *Current Opinion in Chemical Biology*, vol. 8, no. 5, pp. 547–553, 2004.
- [60] G. Bonnet, O. Krichevsky, and A. Libchaber, "Kinetics of conformational fluctuations in DNA hairpin-loops," *Proceedings of the National Academy of Sciences*, vol. 95, no. 15, pp. 8602–8606, 1998.
- [61] L. S. Kristensen and L. L. Hansen, "PCR-Based Methods for Detecting Single-Locus DNA Methylation Biomarkers in Cancer Diagnostics, Prognostics, and Response to Treatment," *Clinical Chemistry*, vol. 55, no. 8, pp. 1471–1483, 2009.
- [62] R. K. Saiki, D. H. Gelfand, S. Stoffel, S. Scharf, K. B. Mullis, and H. A. Erlich, "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase," *Science*, vol. 239, 1988.
- [63] B. Vogelstein and K. W. Kinzler, "Digital pcr," *Proceedings of the National Academy of Sciences*, vol. 96, no. 16, pp. 9236–9241, 1999.
- [64] V. Taly *et al.*, "Multiplex Picodroplet Digital PCR to Detect KRAS Mutations in Circulating DNA from the Plasma of Colorectal Cancer Patients," *Clinical Chemistry*, vol. 59, no. 12, pp. 1722–1731, 2013.

- [65] C. Silveira *et al.*, "Detection and quantification of EGFR T790M mutation in liquid biopsies by droplet digital PCR," *Translational Lung Cancer Research*, vol. 10, no. 3, 2021.
- [66] F. Sanger, "A rapid method for determining sequences in dna by primed synthesis with dna polymerase," *Journal of Molecular Biology*, vol. 94, no. 3, pp. 441–448, 1975.
- [67] J. C. Venter *et al.*, "The Sequence of the Human Genome," *THE HUMAN GENOME*, vol. 291, 2001.
- [68] E. L. Van Dijk, H. Auger, Y. Jaszczyszyn, and C. Thermes, "Ten years of next-generation sequencing technology," *Trends in Genetics*, vol. 30, no. 9, pp. 418–426, 2014.
- [69] B. Merriman, I. T. R&D Team, and J. M. Rothberg, "Progress in Ion Torrent semiconductor chip based sequencing," *ELECTROPHORESIS*, vol. 33, no. 23, pp. 3397–3417, 2012.
- [70] J. Eid *et al.*, "Real-Time DNA Sequencing from Single Polymerase Molecules," *Science*, 2009.
- [71] F. Lankaš, J. Šponer, J. Langowski, and T. E. Cheatham, "DNA Base-pair Step Deformability Inferred from Molecular Dynamics Simulations," *Biophysical Journal*, vol. 85, no. 5, pp. 2872–2883, 2003.
- [72] W. Laeremans, M. Segers, A. Voorspoels, E. Carlon, and J. Hooyberghs, "Insights into elastic properties of coarse-grained DNA models: q-stiffness of cgDNA vs cgDNA+," *The Journal of Chemical Physics*, vol. 160, no. 14, p. 144105, 2024.
- [73] M. Segers, A. Voorspoels, T. Sakaue, and E. Carlon, "Mechanisms of DNA-Mediated Allostery," *Physical Review Letters*, vol. 131, no. 23, p. 238402, 2023.
- [74] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis, "Sequence-dependent thermodynamics of a coarse-grained DNA model," *The Journal of Chemical Physics*, vol. 137, no. 13, p. 135101, 2012.
- [75] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," *Proceedings of the National Academy of Sciences*, vol. 95, no. 4, pp. 1460–1465, 1998.

- [76] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Walder, "Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations," *Biochemistry*, vol. 47, no. 19, pp. 5336–5353, 2008.
- [77] R. J. Karadeema, M. Stancescu, T. P. Steidl, S. C. Bertot, and D. M. Kolpashchikov, "The owl sensor: a 'fragile' DNA nanostructure for the analysis of single nucleotide variations," *Nanoscale*, vol. 10, no. 21, pp. 10 116–10 122, 2018.
- [78] Y. Stulens, R. Van Hoof, K. Hollanders, I. Nelissen, M. Szymonik, P. Wagner, G. Froyen, B. Maes, and J. Hooyberghs, "Hybridization-based sensor with large dynamic range for detection of circulating tumor DNA in clinical samples," *Biosensors and Bioelectronics*, vol. 281, p. 117342, 2025.
- [79] B. Ashwood and A. Tokmakoff, "Kinetics and dynamics of oligonucleotide hybridization," *Nature Reviews Chemistry*, vol. 9, no. 5, pp. 305–327, 2025.
- [80] M. Glazer, J. A. Fidanza, G. H. McGall, M. O. Trulson, J. E. Forman, A. Suseno, and C. W. Frank, "Kinetics of oligonucleotide hybridization to photolithographically patterned DNA arrays," *Analytical Biochemistry*, vol. 358, no. 2, pp. 225–238, 2006.
- [81] M. Stancescu, T. A. Fedotova, J. Hooyberghs, A. Balaeff, and D. M. Kolpashchikov, "Nonequilibrium Hybridization Enables Discrimination of a Point Mutation within 5-40 °C," *Journal of the American Chemical Society*, vol. 138, no. 41, pp. 13 465–13 468, 2016.

Chapter 2

Paper A

Hybridization-based sensor with large dynamic range for detection of circulating tumor DNA in clinical samples

Hybridization-based sensor with large dynamic range for detection of circulating tumor DNA in clinical samples

Yannick Stulens¹, **Rebekka Van Hoof**^{1,2,3}, Karen Hollanders², Inge Nelissen², Michal Szymonik², Patrick Wagner³, Guy Froyen^{4,5}, Brigitte Maes^{4,5}, and Jef Hooyberghs¹

Equal contribution

¹UHasselt, Data Science Institute, Theory Lab, Agoralaan, Diepenbeek, 3590, Belgium

²Flemish Institute for Technological Research (VITO), Environmental Intelligence Unit, Mol, 2400, Belgium

³KU Leuven, Faculty of Science, Department of Physics and Astronomy, Laboratory for Soft Matter and Biophysics, Celestijnenlaan 200 D, Leuven, 3001, Belgium

⁴Jessa Hospital, Laboratory for Molecular Diagnostics, Department of Clinical Biology, Hasselt, 3500, Belgium

⁵UHasselt, Faculty of Medicine and Life Sciences, Agoralaan, Diepenbeek, 3590, Belgium

PUBLISHED April 10, 2025

JOURNAL Biosensors and Bioelectronics

DOI <https://doi.org/10.1016/j.bios.2025.117342>

Contribution

An initial investigation of depletion techniques for clinical samples was carried out by Rebekka Van Hoof and Michal Szymonik before the start of my PhD. I subsequently collaborated closely with Rebekka Van Hoof to further interpret the results, where apparent discrepancies in the data motivated the development of an extended thermodynamic framework. I was responsible for theoretical modeling, formal analysis, and data curation. We prepared the manuscript together. I contributed to all sections, with focus on the methodology, theoretical description, and the design of all figures.

Abstract

In a liquid biopsy approach, targeted mutation analysis of circulating tumor DNA (ctDNA) is a valuable tool for diagnosis, monitoring and personalization of therapy. The ctDNA usually makes up only a small fraction of the total circulating free DNA (cfDNA), and ctDNA often only differs from cfDNA at a single nucleotide. This sets strong requirements on the analytical performance of hybridization-based biosensors, which is the focus of this paper. We use clinical samples and apply the concept of wild-type target depletion. Along with this, we develop an accurate thermodynamic theory for the competitive hybridization and use it for selecting optimal experimental conditions and for data analysis. The result is a biosensor with improved quantification of ctDNA mutations, both the sensitivity and dynamic range are improved by an order of magnitude. As reference techniques, we used a clinically-validated real-time PCR assay and digital PCR for absolute quantification. Our approach can be applied to a broad range of hybridization-based biosensors, providing a robust and effective method to improve the performance of existing biosensors.

2.1 Introduction

The use of biosensors for clinical applications has increased greatly in recent years [1]. Biosensors research is driven by clinical needs, and new research influences clinical practice. An important example is the need to detect somatic gene variants for cancer diagnosis, prognosis, and treatment monitoring. While tissue biopsy is the standard for genetic profiling of tumors, analyzing circulating free DNA (cfDNA) in liquid samples, such as blood, offers a less invasive alternative [2, 3]. Known as liquid biopsy (LB), this approach enables frequent testing and at the same time reduces discomfort, risk and cost. LB also overcomes the limitations of tissue biopsy, which may miss tumor heterogeneity [4, 5]. Studies highlight circulating tumor DNA (ctDNA) as a valuable biomarker for survival, mutational burden, and treatment benefit for different tumor types [6, 7, 8].

From a biosensor point-of-view, it is important to pinpoint the analytical

challenge associated with its application in oncogene mutant detection. Tumor-derived ctDNA (mutant), is present in an abundance of cfDNA from healthy cells (wild-type). Often, mutant and wild-type sequences differ in only a single nucleotide and ctDNA represents only a small fraction of the total cfDNA, less than 0.1% in early stages of tumor development [9]. By its classical definition, a biosensor is an analytical device composed of a biological sensing element in intimate contact with a physical transducer, which together relate the concentration of an analyte to a measurable signal. With the LB application in mind, however, it is not just the concentration of the analyte, but the *relative* concentration of mutant over wild-type that poses the new challenge. This increases the importance of the biological sensing element of the biosensor, because it must be able to distinguish between two very similar molecules present in a very asymmetric abundance. In many sensitive detection techniques, the biological sensing element is a DNA probe, hybridizing with the clinical target DNA sequence. This basic process is the focus of this work, showing how the quantification of the relative mutant concentration can be improved, both in its sensitivity and dynamic range.

In current clinical practice, detection of single nucleotide variants (SNVs) is typically achieved with real-time polymerase chain reaction (PCR), digital PCR (dPCR) for single gene analyses, and targeted next-generation sequencing for multigene testing, but for cfDNA analysis, sensitivity remains an issue irrespective of the used technique and body fluid type [10, 11, 12, 13, 14]. The use of direct hybridization-based biosensors has increased, both in the number of applications and the diversity of its application conditions [11, 15, 16]. Therefore, we study direct hybridization-based biosensing for SNV quantification in cfDNA samples. The approach starts from the basic biological sensing element and does not depend on the specific readout technology of the biosensor. Here, we have chosen a mature and reliable technology with optical readout, namely microarrays, but the results of our work are easily transferable to a wide variety of sensing technologies, including multiplex and cost-effective biosensor devices [17, 18, 19, 20, 21, 22, 23, 24].

Hybridization-based sensors use single-stranded oligonucleotide probes to capture and bind DNA targets of interest. The detection of SNVs relies fundamentally on the difference in binding thermodynamics between perfectly

complementary and partially mismatched DNA sequences. Here, knowledge about competitive hybridization of non-mutant target sequences to the probes is used to improve quantification of low-abundant mutations [25, 26]. The application of capture probes selected for their thermodynamic properties leads to the intentional depletion of the corresponding wild-type target. This methodology has previously been demonstrated to increase the sensitivity of hybridization-based SNV detection to a level comparable with current PCR-based benchmarking methods, and is therefore suitable for clinical applications of LB-based gene analysis [27, 28, 29, 30]. In the current work, we explore the sensitivity and the applicability of the methodology for the first time on LBs. We use clinical cfDNA samples from non-small cell lung cancer (NSCLC) patients that were tested for the p.T790M mutation of the epidermal growth factor receptor (*EGFR*). This biomarker is known to be associated with resistance to first- and second-generation tyrosine kinase inhibitor treatment [31]. The detection of the mutation is done on an assay subject to wild-type target depletion, and compared to one without depletion capture probes. Both are in turn compared to two reference techniques: a clinically-validated real-time PCR and dPCR, which is considered the golden-standard technique for low-abundant mutant detection [12, 32]. A schematic overview of the workflow is presented in Supplementary Information S1. In addition, we expand the thermodynamic theory, enabling us to estimate mutant ratios from samples with both low and high relative mutant ratios. The expansion leads to important insights for future probe designs. This crucial step increases the dynamic range of the biosensor and consequently the range of applications [33].

2.2 Materials and Methods

2.2.1 Hybridization method: the reference probe and the concept of depletion

A hybridization-based sensor detects the presence of single-stranded (ss) mutant DNA by ssDNA hybridization probes. The detection is based on the formation of a stable double helix, where the two strands obey the

Watson-Crick pairing rule. The binding of mutant target T_{mut} to the probe indicates the presence of mutant DNA. If the mutant DNA is present in a background of wild-type target T_{wt} , i.e. as a mix in the same sample, the sensor's performance is impeded by cross-hybridization with wild-type targets. This cross-hybridization is most severe when T_{mut} is a SNV of T_{wt} , where only a single nucleotide differs between T_{mut} and T_{wt} . While traditionally, cross-hybridization leads to a reduced detection sensitivity, we will show in Section 2.2.1 how cross-hybridization allows one to define a robust mutant detection signal. More specifically, we focus on the situation where target sequences T_{mut} and T_{wt} are known, and we aim to detect a minority of mutant targets in a majority of wild-type background.

An important aspect of sensor design is the selection of probe sequences needed to detect mutant DNA in a sample. The mutant probe (P_{mut}) is designed to physically cover the mutant sequence in a region around the SNV with roughly a few dozens of nucleotides (see Section 2.2.6). In Section 2.2.1 and 2.2.1, we explain how the use of additional probes improves performance of the hybridization-based sensor. These probes will either match the wild-type sequence (probe P_{wt}) or be used as a reference signal (probe P_{ref}).

When the hybridization between the sample targets and the probes has reached thermal equilibrium, the readout occurs. Many technologies and means are available for experimental readout, including fluorescence emission, changes in heat transfer resistance, or microgravimetric sensors [22, 34]. We use microarray technology, which provides a mature sensing platform and enables readout spanning up to five orders of magnitude. In microarrays, probes are spatially organized into spots and targets are fluorescently labeled. The optical readout of a spot provides a fluorescent intensity I . For example, the intensity $I_{P_{mut}}$ is proportional to the fraction of P_{mut} bound to a target, whether it is T_{wt} , T_{mut} or a combination of the two targets. Notably, no distinction can be made between T_{wt} or T_{mut} being bound to P_{mut} . This emphasizes the need for a reference probe, which is introduced in the next section.

Reference probes allow for a robust sensor design

To obtain a mutant-specific signal for sample mixtures, one can introduce an additional probe P_{ref} as a reference, which is designed in such a way that T_{wt} has an equal affinity toward the P_{ref} as it does toward the P_{mut} . An appropriate reference sequence can be found by exploring, theoretically and experimentally [29], cross-hybridizations between different variants of the P_{wt} , containing one or two mismatches to T_{wt} (one or two nucleotide pairs that do not obey the Watson-Crick pairing rule). The reference probe then allows the mutant detection signal S to be defined as a robust ratio-metric quantity, given by

$$S = \ln \left(\frac{I_{P_{mut}}}{I_{P_{ref}}} \right). \quad (2.1)$$

By design, if only T_{wt} is present, cross-hybridizations $T_{wt} + P_{mut}$ and $T_{wt} + P_{ref}$ will occur equally often, i.e. $I_{P_{mut}} = I_{P_{ref}}$, and the signal S will be zero. However, when some amount of mutant is present, $I_{P_{mut}}$ will increase, and the signal will be positive, indicating the presence of mutant in the sample (Fig. 2.1A and B).

The mutant detection signal, given by equation (2.1), is described by standard Langmuir theory, which relates the occupancy of probes to target concentrations c_{mut}/c_{wt} and affinities between probes and targets [35]. At a low fractional abundance of the mutant target, the signal is described by:

Signal without depletion:

$$S_{no_depl} = \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{P_{mut}}}{RT} \right) \right) \quad (2.2)$$

where R is the universal gas constant, T is the absolute temperature and $\Delta\Delta G_{P_{mut}}$ represents the difference in hybridization free energy (ΔG) between the mutant probe and its two targets, mutant and wild-type. Notably, the sensor benefits from having a large $\Delta\Delta G_{P_{mut}}$, amplifying the mutant ratio c_{mut}/c_{wt} and increasing sensitivity. However, this free-energy difference is determined by the target sequences and it is not a design parameter. The following section will introduce a way to further improve the sensor performance, using the concept of depletion.

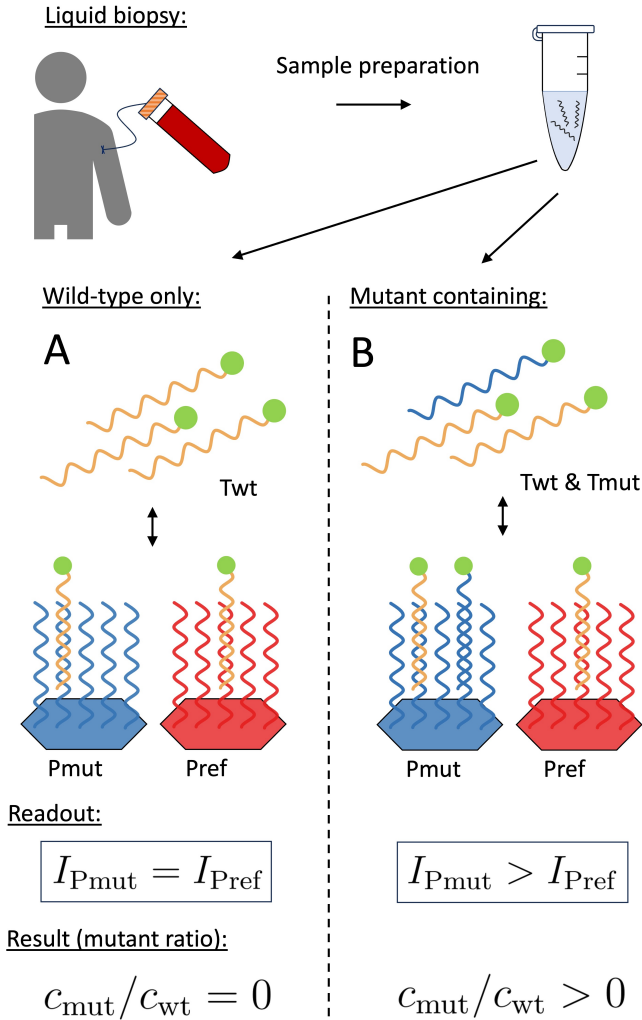


Figure 2.1: Overview of the hybridization method-based mutant detection assay [28, 29]. (A) When the assay is subject to a pure wild-type sample, the same amount of target will bind to *Pmut* and *Pref*, the intensity readout is the same and the mutant ratio is predicted to be zero. (B) When mutant DNA is present, the additional hybridization to *Pmut* will result in a higher intensity for *Pmut* compared to *Pref*. This signals the presence of mutant DNA.

Depletion probes enrich the mutant content in a sample

By introducing an abundance of wild-type matching probes *Pwt*, a large number of wild-type targets will bind to *Pwt* and become unable to contribute to the signal. The available amount of wild-type gets depleted.

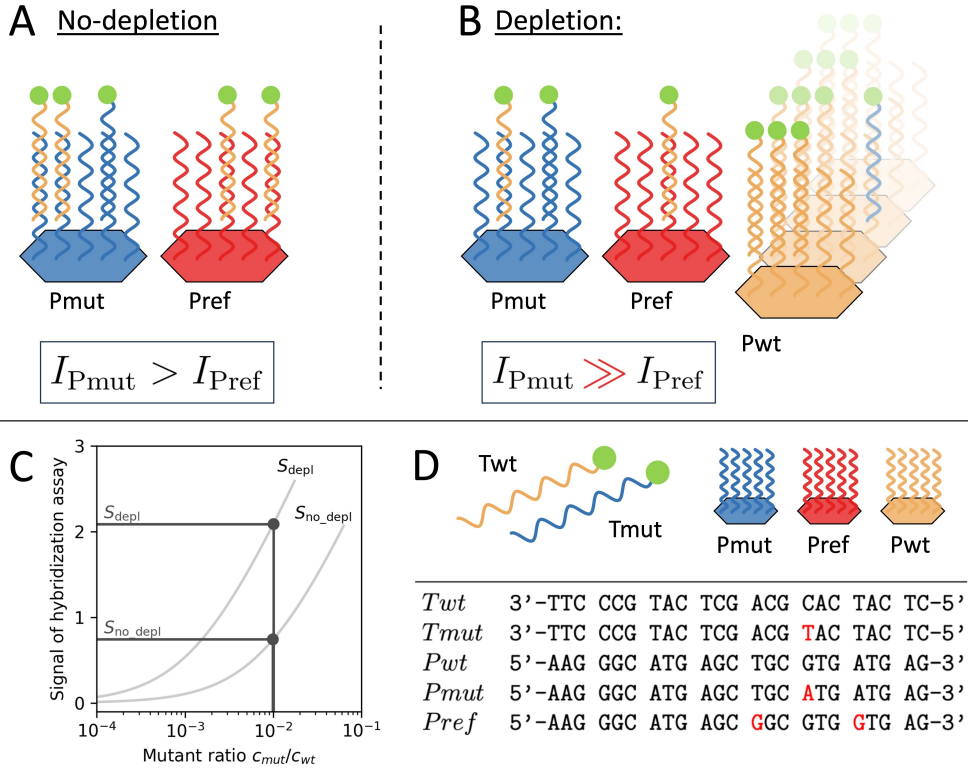


Figure 2.2: The concept of depletion in the hybridization method [28, 29]. (A) and (B) When a large amount of wild-type target depletion probes P_{wt} are present, the difference between the P_{mut} and P_{ref} intensities becomes larger. (C) Theoretical predicted signal of the hybridization assay under depletion condition (S_{depl} , equation (2.3)) and no-depletion condition (S_{no_depl} , equation (2.2)) in terms of mutant ratio c_{mut}/c_{wt} . Vertical lines demonstrate how the same mutant ratio results in a larger signal in the depletion case. Horizontal lines show how an experimental measurement of the signal can be used to obtain the mutant ratio. (D) Target and probe sequences used in the hybridization experiments. Only the 23 bp probe binding region is shown for the clinical NSCLC targets. Differences between the wild-type target sequence or the wild-type probe sequence are highlighted in red.

Of course, the available amount of mutant is also lowered, but by design there is a selective reduction of available target concentrations, such that the mutant ratio is enlarged to an effective ratio $c_{mut}^{eff}/c_{wt}^{eff}$. Because available concentrations are lowered, this usually requires an increase in total target concentration. Depletion by wild-type probes enriches the mutant content in the sample, which increases the ratio $I_{P_{mut}}/I_{P_{ref}}$ and the signal S (Fig. 2.2B). The enrichment is determined by the difference in free energy between the wild-type probe and its two targets T_{wt} and T_{mut} ,

$\Delta\Delta G_{P_{wt}}$. When the affinity to wild-type probes is high, a condition we call strong depletion [28], the enrichment of the mutant ratio reaches the theoretical maximum of a factor $\exp(\Delta\Delta G_{P_{wt}}/RT)$, and consequently the signal is maximally enhanced and given by:

Signal with depletion:

$$S_{depl} = \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{P_{mut}} + \Delta\Delta G_{P_{wt}}}{RT} \right) \right) \quad (2.3)$$

Depletion of wild-type targets further increases sensitivity and produces a higher signal when mutant targets are present in the sample compared to the no-depletion case (equation (2.2)). As in our previous work, the limit of detection (LoD) is defined as the mutant fractional abundance corresponding to a signal equal to three times the standard deviation of a pure wild-type negative control sample [29]. The higher signal in case of target depletion results in an improved LoD.

Hybridization assay allows determination of mutant content

In the hybridization experiment, the intensities of the mutant and reference probes are measured, and their ratio is used to determine the signal via equation (2.1). Figure 2.2C shows the theoretical signals of the hybridization assays as a function of the mutant ratio c_{mut}/c_{wt} , equations (2.2) and (2.3). Conversely, one can invert the equation of signal S to allow the determination of the mutant ratio c_{mut}/c_{wt} in a sample after experimental measurement. The mutant ratio of the sample is calculated using equations (2.4) and (2.5). In Section 2.3.3, we will show that these equations are only valid for samples with low signals or mutant ratios, and that accurate determination of the mutant ratio for high signals requires an extension to the model concerning the interaction between $T_{mut} + Pref$.

No-depletion:

$$\frac{c_{mut}}{c_{wt}} = (\exp(S_{no_depl}) - 1) \exp \left(-\frac{\Delta\Delta G_{P_{mut}}}{RT} \right) \quad (2.4)$$

Depletion:

$$\frac{c_{mut}}{c_{wt}} = (\exp(S_{depl}) - 1) \exp\left(-\frac{\Delta\Delta G_{Pmut} + \Delta\Delta G_{Pwt}}{RT}\right) \quad (2.5)$$

In this article, we use the same design as in our previous work [29], where we showed that adding depletion probes improves the detection sensitivity of the *EGFR* T790M mutation using synthetic DNA samples and surface-tethered probes. Synthetic DNA samples of differing mutant ratios were used to perform a least-square fit to equations (2.2) and (2.3), and free energies $\Delta\Delta G_{Pmut}/RT$ and $\Delta\Delta G_{Pwt}/RT$ were obtained. These values, determined to be 4.7 and 1.86 respectively, were in close agreement with values expected from thermodynamic nearest-neighbor parameters [36] and are used throughout this article. The LoD was determined to be $c_{mut}/c_{wt} = 0.0060$ ($c_{mut}/c_{tot} \approx 0.60\%$) under the no-depletion condition and improved to $c_{mut}/c_{wt} = 0.00048$ ($c_{mut}/c_{tot} \approx 0.048\%$) in the depletion case [29]. When c_{mut}/c_{wt} is below the LoD, samples are classified as negative. Details of the experimental implementation are given in Section 2.2.6.

2.2.2 NSCLC patients' cfDNA samples

Leftover cfDNA from 18 NSCLC patients were obtained from the Jessa Hospital (Hasselt, Belgium) with ethical approval of the Ethics Review Committee of the Jessa Hospital under code 2021/023. The QIAamp circulating nucleic acid kit (Qiagen, Hilden, Germany, #55114) was used for extraction of cfDNA and residual samples were stored for up to 3 years at -20 °C in DNA LoBind tubes (Eppendorf). A volume of 30 - 35 μ L per sample was received together with information on their *EGFR* c.2369C>T p.(T790M) mutational status based on real-time PCR (using the *EGFR* Mutation analysis kit from EntroGen, Woodland Hills, CA, #EGFR-RT52 and a Rotor-Gene X device from Qiagen) performed on freshly extracted cfDNA by the Jessa Hospital as part of their molecular diagnostic workflow. In the clinical assay, a threshold cycle (C_t) value of the housekeeping gene beta-2-microglobulin (B2M) reaction was subtracted from the C_t value for *EGFR* T790M in order to determine a delta C_t (ΔC_t) value.

$$\Delta C_t = C_{tT790M} - C_{tB2M} \quad (2.6)$$

EGFR T790M samples were classified as positive when the $\Delta C_t \leq 7.4$. According to this criterion, six samples were classified as positive and twelve as negative in the real-time PCR clinical assay. Clinical diagnostic information is provided in Supplementary Information S2.

2.2.3 dPCR

A QX200 Droplet Digital PCR System (ddPCR, Bio-Rad Laboratories, Hercules, CA) was used to perform dPCR-based mutation detection of the *EGFR* 790M mutation. The reaction mixture contained 11 μL 2 $\times$ ddPCR supermix for probes (No dUTP, Bio-Rad Laboratories, #1863023), 1.1 μL 20 $\times$ primers and probes (Bio-Rad Laboratories, assay ID: dHsaMDV2010019), 2.4 μL nuclease-free water and 7.5 μL of cfDNA sample or no-template control (NTC). Droplets were generated using the QX200™ Droplet Generator and transferred to a 96-well plate before sealing the plate using the PX1 PCR Plate Sealer according to the manufacturer's instructions. For PCR, the following amplification conditions were used: 95 °C for 10 minutes; 40 cycles of 94 °C for 30 seconds, 55 °C for 1 minute and 98 °C for 10 minutes at a temperature ramp rate of 2 °C/second and a C1000 Touch™ Thermal Cycler with 96-deep well reaction module were used. After the amplification, the plate was kept at 4°C for 30 minutes. A QX200™ Droplet Reader was used for the fluorescence intensity readout and results were analyzed using QuantaSoft™ Analysis Pro 1.0.596 Software (Bio-Rad Laboratories). For analysis, the inclusion criterion was $\geq 10,000$ droplets/well and fluorescence thresholds for the *EGFR* assay were set at 2275.10 for channel 1 (mutant) and 2051.14 for channel 2 (wild-type). Droplet counts and mutant ratios of the clinical samples with corresponding confidence intervals calculated by the software are presented in Supplementary Information S3.

2.2.4 PCR amplicon generation

For DNA hybridization experiments, amplicons of the NSCLC cfDNA samples of 107 bp comprising the region of the *EGFR* T790M mutation were generated. The PCR reaction mixture contained 25 μL 2 $\times$ Platinum™ II Hot-

Start PCR Master Mix (Invitrogen, ThermoFisher Scientific, Merelbeke, Belgium, #14000013), 2.5 μ L forward primer (10 μ M): 5' ATC TGC CTC ACC TCC AC 3', 2.5 μ L reverse primer (10 μ M) with 5'-end phosphate modification: 5' /Phos/ CTG GCG TCA TAG CTG TTT CCT GTG TGA TTG TGT TCC CGG ACA TAG TC 3', 12.5 μ L nuclease-free water and 7.5 μ L cfDNA sample. A 1.0% mixture of *EGFR* T790M mutation in a wild-type background of 500 bp gBlocks™ Gene Fragments (IDT, Leuven, Belgium) and nuclease-free water as NTC were taken along as positive and negative control, respectively. The sequence is available in Supplementary Information S4. Purification of the amplicons was done using the Zymo DNA Clean & Concentrator™-5 Kit (Zymo Research, Irvine, CA, #ZY-D4013). Gel electrophoresis was performed on a FlashGel™ System (Lonza Group AG, Basel, Switzerland #57067) to check the binding of the primers and verify the amplicon length (see Supplementary Information S4).

2.2.5 Lambda exonuclease digestion

After PCR amplicon generation, the DNA concentration of the purified PCR product was determined using a NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific), and the amplicons diluted to 10 ng/ μ L in a total reaction volume of 50 μ L containing 2 μ L of 1.25 U/ μ L Lambda exonuclease (Thermo Fisher Scientific, #EN0561) and 5 μ L of 10 $\times$ supplied reaction buffer. The reaction mixture was heated to 37 °C for 30 minutes followed by 85 °C for 10 minutes on a Veriti™ 96-Well Thermal Cycler for the selective digestion of the 5'-phosphorylated strands to generate 107 bp long single-stranded DNA. The exonuclease digestion efficiency was previously confirmed by [29]. Samples were stored at -20 °C and used without further purification.

2.2.6 DNA hybridization experiment

To perform the hybridization experiment for *EGFR* T790M detection, we used two custom 8 $\times$ 15K Agilent microarray slides (Agilent Technologies, Santa Clara, CA, USA, #G2509F) with the same design as described in [29]. Briefly, one slide contained 8 *EGFR* depletion arrays and the other

contained 8 no-depletion arrays. Depletion was induced by introducing a large number of spots containing *Pwt* (14900 spots) on one of the arrays compared to *Pmut* (20 spots) and *Pref* (20 spots). The no-depletion array contained equal numbers of *Pwt*, *Pmut* and *Pref* (20 each). The 20 replicates determine the 95% confidence interval. The readout of the slides is presented in Supplementary Information S5. Oligonucleotide sequences are presented in Fig. 2.2D. The Gene Expression Microarrays Hybridization Kit (Agilent Technologies, #5188-5242) was used and the hybridization mixture contained: 25.0 μL of $2\times$ GEx hybridization buffer, 5.0 μL $10\times$ blocking agent, 3.5 μL of Lambda exonuclease-digested *EGFR* amplicons (1 nM for the depletion array and 0.01 nM for the no-depletion array), 2.0 μL 5' Cy-3 labeled barcode oligonucleotides (1250 nM, 5'- /5Cy3/AAA AAC TGG CGT CAT AGC TGT TTC CTG TGT GA -3') and 14.5 μL nuclease-free water in a total volume of 50 μL . Hybridization and washing were done according to the manufacturer's instructions. Briefly, 40 μL of hybridization mixture was placed on an 8-well gasket slide (Agilent Technologies, #G2534-60016) that was placed inside a hybridization chamber (Agilent Technologies, #G2534A). The microarray slide was placed on top (with the active side down), the hybridization chamber was closed with the chamber cover and the assembly was clamped in place. It was checked that there were no stationary bubbles present in the assembly. The hybridization chamber was placed inside a hybridization oven (Agilent Technologies, #G2545A) and rotated at 10 rpm at 65 °C for 17 hours. Array scanning was performed on an Agilent G2565CA (G2505C) Series Microarray Scanner System and Scan Control Software A8.4.1 (Agilent Technologies) with settings for scan region: Agilent HD (61 $\times$ 21.5 mm), resolution: 5 μm , TIFF image: 20-bit, photomultiplier tube: 100% and no eXtended Dynamic Range feature. Agilent Feature Extraction Software 10.7.3.1 (Agilent Technologies) was used to analyze the array images with automatic gridding, intensity measurement, background subtraction and quality checks.

The 18 cfDNA samples were analyzed in three runs, each performed on the same slides after regeneration. Each run contained a negative control which was used for baseline subtraction, as was done by [29]. Between each run, the two slides were regenerated by washing them for 44 minutes in a 1:1 mix of wash buffer 2 from the Gene Expression Microarrays Hybridization Kit with Milli-Q water at 65 °C and drying before storage at room temperature.

2.3 Results and Discussion

2.3.1 Application to clinical samples

The sample set of this work consisted of 18 cfDNA samples from NSCLC patients. In general, these samples contained wild-type cfDNA and possibly a small fraction of tumor-originating *EGFR* T790M mutant cfDNA. All samples were tested for the presence of this point mutation in a clinically-validated real-time PCR assay. This assay relies on a ΔCt value: the difference in PCR amplification cycles before it reaches a detection threshold between the mutation, *EGFR* T790M, and a housekeeping gene (see Section 2.2.2). A higher ΔCt value corresponds to a lower mutant ratio. Both panels of Fig. 2.3 show the ΔCt values on the (reversed) x-axis. Samples that did not reach the detection threshold for *EGFR* T790M after 40 PCR cycles were classified as negative and are indicated as NaN (Not a Number) since no ΔCt value could be calculated. Samples with $\Delta Ct \geq 7.4$ were above the clinical threshold and classified as negative (12 samples) and samples with $\Delta Ct \leq 7.4$ were classified as positive (6 samples).

Next, the set of 18 samples set was evaluated using the hybridization assay of Section 2.2.1. This assay is thermodynamics-based, it uses DNA probes to capture specific target sequences in the sample and the resulting readout is determined by hybridization free-energy parameters of the sequences at hand, and by the ratio c_{mut}/c_{wt} , i.e. the mutant over the wild-type cfDNA concentration in the sample. The assay signal provides a quantitative measurement of the mutant ratio c_{mut}/c_{wt} (see Fig. 2.2C and equation (2.4) and (2.5)). The hybridization assay was developed in a no-depletion and a depletion version. The latter depletes the wild-type cfDNA in the sample leading to a lower LoD for the mutant ratio (see Section 2.2.1 for details and numerical values). When a mutant ratio is below LoD, it is classified as negative and represented by zero on Fig. 2.3. The resulting values are presented on the y-axis of Fig. 2.3.

We observed a correlation between ΔCt values and mutant ratios c_{mut}/c_{wt} of both hybridization assay variants for the clinically positive samples. For samples that were clinically assigned as negative, the majority was also negative (i.e. below LoD) in both hybridization assays. Hence, there

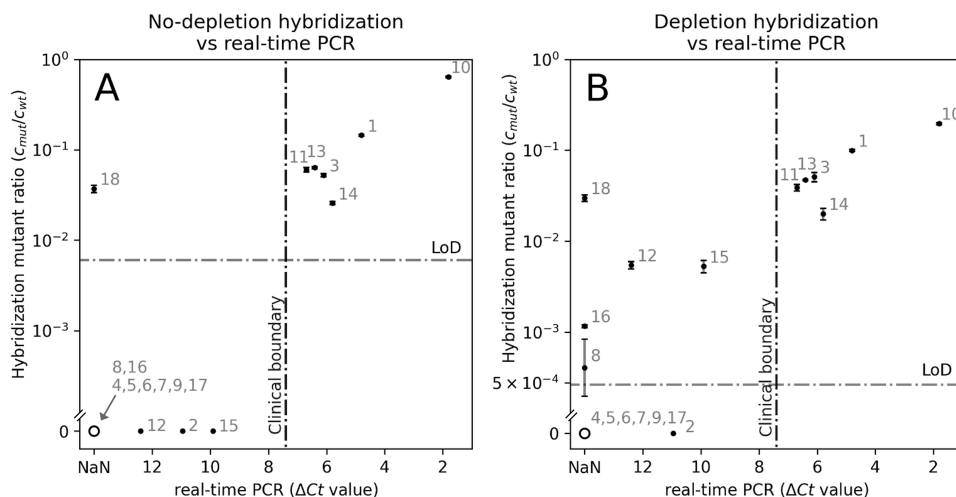


Figure 2.3: Scatter plot of mutant ratios c_{mut}/c_{wt} obtained by the hybridization assay under no-depletion condition (A) and depletion condition (B) versus ΔC_t values obtained by real-time PCR for 18 cfDNA samples from NSCLC patients. Errorbars show the 95% confidence interval. Horizontal dashed lines represent the respective LoD for each hybridization assay condition. Vertical dashed lines represent the clinical threshold between positive and negative samples for the real-time PCR assay. Samples that were both below the LoD and had no ΔC_t value (NaN) are depicted with an open circle symbol.

was a general concordance between the clinical PCR and the hybridization assays. However, there were some interesting exceptions. Both hybridization assays detected the mutation in sample 18 (further discussed in Section 2.3.2) whereas no mutation was detected using real-time PCR. Samples 12 and 15 have a $\Delta C_t \geq 7.4$, hence clinically classified negative, but received a non-zero mutant ratio c_{mut}/c_{wt} above the LoD in the depletion hybridization assay. This suggested that these samples could contain a low number of mutant targets. Further, the real-time PCR assay was not able to measure a ΔC_t value for samples 8 and 16 (assigned NaN), but again, the depletion hybridization assay detected a non-zero mutant concentration.

2.3.2 Evaluation of the quantitative agreement with dPCR

Since real-time PCR does not provide an absolute quantitative description of the mutant fraction in a sample, we performed also dPCR measurements on the clinical samples. Digital PCR is a sensitive technique for absolute quantification and regarded as the gold-standard for low-abundant mutant detection [12, 32]. All 18 clinical samples were run in a dPCR assay as described in Section 2.2.3. The assay has a technical control criterion, stating that the droplet count has to be $\geq 10,000$ droplets/well. This inclusion criterion resulted in the rejection for further analysis of samples 3, 7 and 8. For all other samples a mutant ratio c_{mut}/c_{wt} was determined by dPCR, and compared with those of the two hybridization assays (Fig. 2.4).

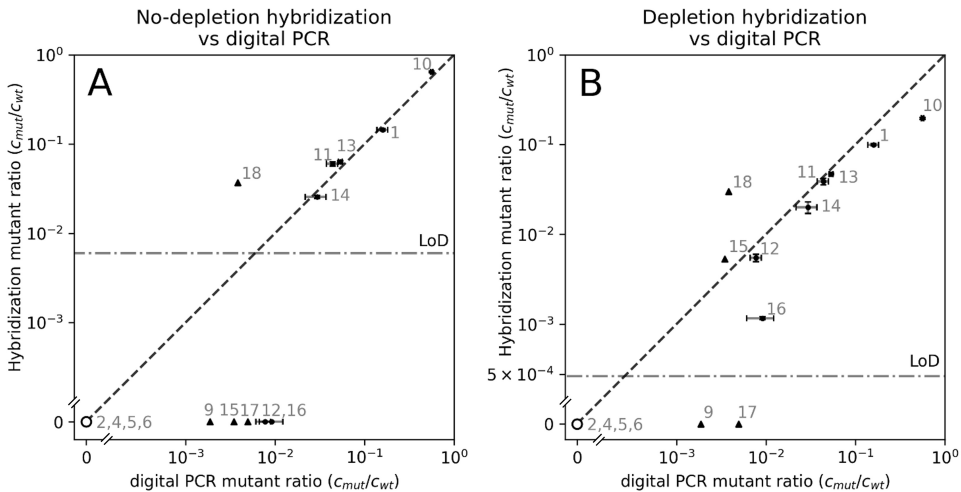


Figure 2.4: Scatter plot of mutant ratios c_{mut}/c_{wt} obtained by the hybridization assay under no-depletion condition (A) and depletion condition (B) versus dPCR for 18 cfDNA samples from NSCLC patients (1-18). Errorbars show the 95% confidence interval. A dashed diagonal is shown as a visual guide. Horizontal dashed lines represent the respective LoD for each hybridization assay condition. Samples are shown as dots, except when the number of mutant positive droplets < 3 , in which case they are denoted with a triangle. Samples 3, 7 and 8 did not meet the technical control criterion for dPCR and are not shown. The open circle symbol represents cases for which no mutant was detected in both the hybridization assays and dPCR.

Firstly, the samples which are classified as positive by the clinical assay (real-time PCR, Fig. 2.3) all show a strong quantitative correlation be-

tween the dPCR and hybridization assays with Pearson correlation coefficient of 0.89 and 0.84 for the no-depletion and depletion assays, respectively. Despite this correlation, it was found that the depletion assay underestimated the mutant ratio, which is further examined in the next section. Also, samples 2, 4, 5 and 6 were in full agreement, i.e. no mutant detected. Next, samples 12 and 16 were negatively classified by the clinical assay but showed the presence of the mutation in the depletion hybridization assay which was confirmed by the dPCR. These results showed the increased sensitivity of the depletion assay over the no-depletion assay, and confirmed that the outcome of the hybridization assays, developed on synthetic DNA, can be translated to clinical samples and lead to improved diagnostics of SNV mutations.

Furthermore, the samples in subset 9, 15, 17 and 18 (each indicated by a triangle in Fig. 2.4) were around the detection limit of the dPCR: these samples had only one or two mutant-positive droplets, which is below the three positive droplets generally adopted as a cut-off for positivity in good practice according to the manufacturer [37]. Of this subset, samples 9 and 17 were negative across all other analysis methods, indicating that either dPCR detected a very low abundance of mutants, or the mutant-positive droplets were false positives. A single mutant-positive droplet was detected in sample 18, but the measured mutant ratio by dPCR was lower than the expected ratio from the hybridization assay. Since in both hybridization assays the mutant ratios in sample 18 were equally high and well above the LoD, we suspected that the observed ratio in the hybridization assay was the result of sample contamination. Sample 15 was negative in the no-depletion hybridization assay, positive in the depletion hybridization assay, and received a ΔC_t value in the real-time PCR, which suggested a positive but very low mutant content.

2.3.3 Extension of the model to accurately determine high mutant concentrations

In the previous section we compared the difference between the two hybridization assays for positively classified samples. While both correlated well with the dPCR (Fig. 2.4), the depletion assay underestimates the mutant ratio. This issue is due to an interaction not taken into account

in the theory of section 2.2.1. This interaction will be discussed below and lead to an extension of the theory of section 2.2.1. The extension to the model will result in a more accurate description of the signal for high mutant ratios, and consequently, high mutant ratios are more accurately quantified.

The mutant ratios from the depletion and no-depletion assays are plotted in Fig. 2.5A. Again, this figure shows the concordance between the two assays for the majority of the samples and the improved LoD for the depletion assay, visible for samples 8, 12, 15, and 16. A clear correlation is visible in the upper-right quadrant. However, depletion-derived values are lower than no-depletion, and an apparent deviation from the diagonal is visible at the highest mutant ratios.

To address this issue, we re-evaluate the data analysis which leads to the mutant ratio c_{mut}/c_{wt} . Figure 2.2C illustrates how, for both the depletion and no-depletion method, a measurement leads to the determination of the mutant ratio in a sample. When theory and experiment match, the two assays agree on the amount of mutants in the sample.

Using equations (2.4) and (2.5), the same was done for the clinical samples on Fig. 2.5C. The experimental signal is plotted on the y-axis *versus* the derived mutant ratio on the x-axis for both depletion and no-depletion. Corresponding samples for depletion and no-depletion assays are connected by dashed lines, which should be vertical when theory and experiment match. The discrepancy between the two assays is apparent for samples with a high mutant fraction. This observation points to the solution of the issue: the physical model used in Fig. 2.5C (see Section 2.2.1) was developed to accurately describe samples with low mutant fractions. In this limit, it is correct to neglect the interaction between T_{mut} and the reference probe P_{ref} . Since the mutant ratio in some of the current samples is higher than those used in the work of [29], the interaction of the mutant target with the reference probe cannot be neglected for these samples. Intuitively, while for low mutant ratios, the reference probe intensity can be considered approximately constant and only dependent on the amount of T_{wt} , the intensity of the reference probe at high c_{mut}/c_{wt} is dependent on both the amount of T_{wt} and T_{mut} in the sample. The observed discrepancy between the depletion and no-depletion assay is then

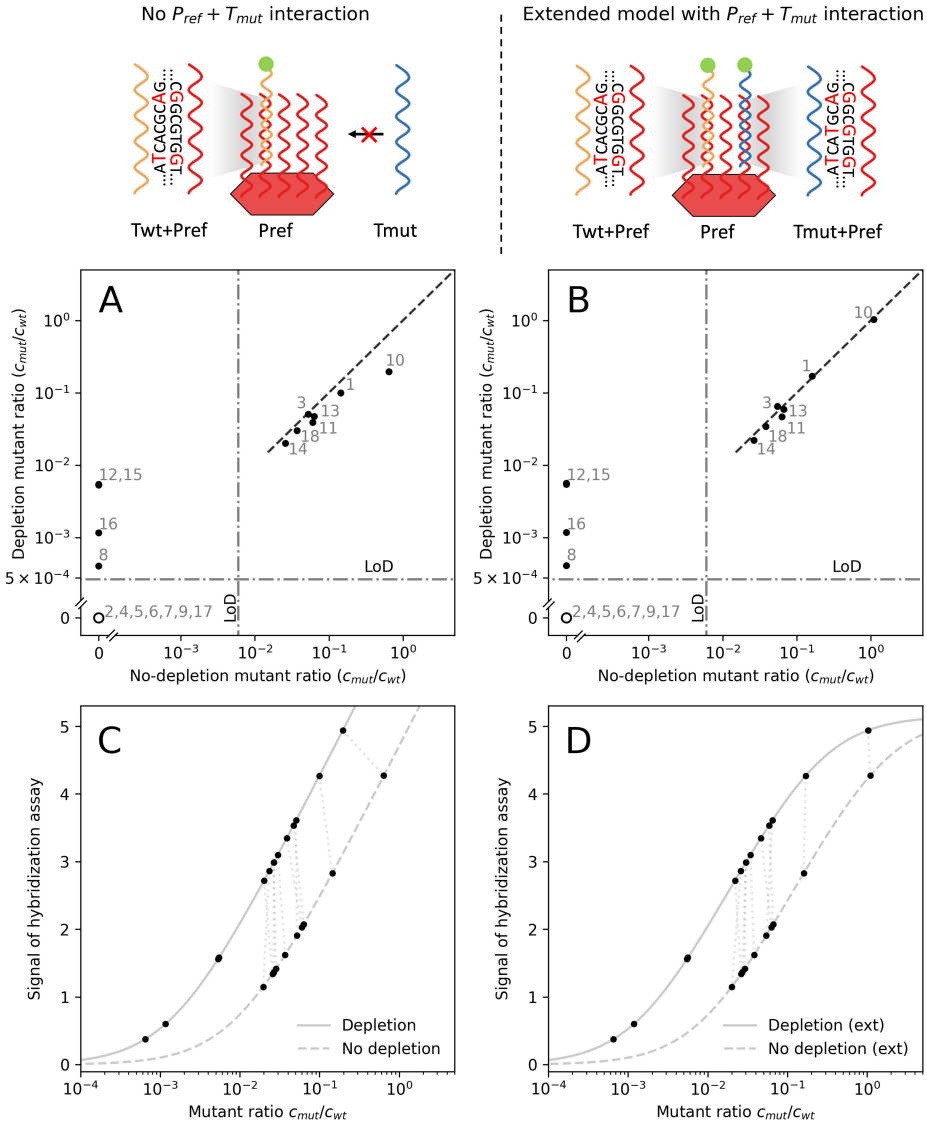


Figure 2.5: The effect of the extension of the thermodynamic hybridization model. The binding of the mutant target to the reference probe is neglected in (A) and (C) or taken into account in (B) and (D). (A) and (B) Scatter plots of mutant ratios c_{mut}/c_{wt} obtained by the two hybridization assays with and without wild-type target depletion. Horizontal and vertical lines represent the LoD for the depletion and no-depletion assay condition, respectively. The open circle symbol represents cases for which no mutant was detected using both the hybridization conditions. (C) and (D) Determination of mutant ratio c_{mut}/c_{wt} from the signal without (C) and with (D) taking into account the mutant target (T_{mut})-reference probe (P_{ref}) interaction. Solid curves represent equations (2.2) (C), and (2.7) (D). For both depletion and no-depletion hybridization conditions, the experimental signal is plotted on the y-axis against the calculated mutant ratio. The dotted lines connect corresponding samples for the depletion and no-depletion assay.

understood by the fact that the effect occurs at lower c_{mut}/c_{wt} values for the depletion assay, because this assay actively changes the ratio of T_{mut} over T_{wt} in the sample.

For a quantitative description we extended the applicability range of the physical hybridization model by incorporating all target interactions with P_{ref} . A detailed derivation of the extended theory is provided in the Supplementary Information S6. The extension adds an additional term to equations (2.2) and (2.3), leading to:

No-depletion:

$$S_{no_depl} = \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{P_{mut}}}{RT} \right) \right) - \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{P_{ref}}}{RT} \right) \right) \quad (2.7)$$

Depletion:

$$S_{depl} = \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{P_{mut}} + \Delta\Delta G_{P_{wt}}}{RT} \right) \right) - \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{P_{ref}} + \Delta\Delta G_{P_{wt}}}{RT} \right) \right) \quad (2.8)$$

where the extra parameter $\Delta\Delta G_{P_{ref}}$ represents the target interactions with P_{ref} . Specifically, $\Delta\Delta G_{P_{ref}}$ is the difference in hybridization free energy between the reference probe and the two targets, $\Delta G(P_{ref} + T_{wt}) - \Delta G(P_{ref} + T_{mut})$. Since $P_{ref} + T_{mut}$ contains an additional mismatch (see Fig. 2.5), $\Delta\Delta G_{P_{ref}}$ is negative.

The additional terms in equations (2.7) and (2.8) are not identical, which explains quantitatively the discrepancy between the two assays. More specifically, for the depletion assay, the depletion interaction $\Delta\Delta G_{P_{wt}}$ is present in the additional term, which is why the new interaction has a larger influence in the case of depletion. When all free-energy differences are known, similar to section 2.2.1, the mutant ratio of a sample is obtained by inverting equations (2.7) and (2.8). The resulting equations are provided in Supplementary Information S7.

The effect of the extended model is visible in Fig. 2.5B, where (as opposed to Fig. 2.5A) the depletion and no-depletion assays now agree on the

amount of mutants in the samples. The dynamic range is enlarged ~ 10 -fold by including the $Pref + Tmut$ interaction. The extension of the model is also illustrated by the solid curves of Fig. 2.5C with 2.5D: for high mutant ratios the signal curves flatten out, i.e. they become sigmoidal. For small mutant ratios the curves of Fig. 2.5C and 2.5D remain identical. The additional subtraction in equations (2.7) and (2.8) becomes important for high mutant ratios, where the signal was overestimated and therefore the predicted mutant ratio was underestimated. Hence, the extension allows for a more accurate determination of the amount of mutant in a sample, relevant at high mutant ratios, by introducing a single extra free-energy parameter $\Delta\Delta G_{Pref}$ with a clear physical meaning.

The parameter $\Delta\Delta G_{Pref}$ can be extracted from the data. Any sample that is measured in both the depletion and no-depletion conditions contains the same amount of mutant for both conditions. Eliminating c_{mut}/c_{wt} from equations (2.7) and (2.8) provides a mapping of the depletion signal onto the no-depletion signal for the same sample. This curve can be fitted for the values of $\Delta\Delta G_{Pwt}$ and $\Delta\Delta G_{Pref} - \Delta\Delta G_{Pmut}$. Since $\Delta\Delta G_{Pmut}$ is known from previous work by [29], this allows the value of $\Delta\Delta G_{Pref}$ to be determined. The derivation, along with the fitting procedure can be found in Supplementary Information S8. The free-energy values are summarized in Table 2.1. The fit provides a way to determine $\Delta\Delta G_{Pref}$. By adding the value of $\Delta\Delta G_{Pmut}/RT = 4.7$, we obtain $\Delta\Delta G_{Pref}/RT = -0.45$ as presented in the last column. The fitted value of the wild-type probe parameter $\Delta\Delta G_{Pwt}$ provides an independent determination of this free-energy difference. The agreement with the value reported by [29] shows the consistency of the hybridization method and the fitting procedure.

It is instructive to compare the value of $\Delta\Delta G_{Pref}$ with the theoretical value of a simple nearest-neighbor model for hybridization free energy [36]. The nearest-neighbor calculation predicts a value of -2.3 for $\Delta\Delta G_{Pref}/RT$. The measured value (Table 2.1) is noticeably smaller. This can be understood by identifying the locations of the mismatches present in the $Tmut + Pref$ binding (Fig. 2.5). Three mismatches are present in close proximity of each other, only being separated by two nucleotide pairs that obey the Watson-Crick pairing rule. It is known that, when the distance between two mismatches is closer than five nucleotides, the additivity principle, which is the basis of the nearest-neighbor model, breaks

Table 2.1: Overview of free-energy parameters relevant in the hybridization assay. [29] determined the thermodynamic parameters from synthetic samples at low mutant ratios. Adding $\Delta\Delta G_{Pmut}$ to $\Delta\Delta G_{Pref} - \Delta\Delta G_{Pmut}$ results in an estimation of the reference probe interaction $\Delta\Delta G_{Pref}$, as presented in the last column.

	Fit to clinical data		Van Hoof <i>et al.</i> (2022)		Reference probe interaction
Free-energy parameter	$\frac{\Delta\Delta G_{Pmut}}{RT}$	$\frac{\Delta\Delta G_{Pref} - \Delta\Delta G_{Pmut}}{RT}$	$\frac{\Delta\Delta G_{Pmut}}{RT}$	$\frac{\Delta\Delta G_{Pmut}}{RT}$	$\frac{\Delta\Delta G_{Pref}}{RT}$
Value	1.80 ± 0.06	-5.15 ± 0.17	1.86 ± 0.15	4.7 ± 0.1	-0.45 ± 0.27

down, and the free-energy penalty of the model is an overestimation [36]. This effect becomes even more pronounced when three mismatches are near each other, as is the case for the mutant-reference probe interaction. The overestimation (more negative) of $\Delta G(Pref + Tmut)$ implies $\Delta\Delta G_{Pref}/RT > -2.3$, which explains the observed value presented in Table 2.1. The $Tmut + Pref$ binding is therefore more stable than predicted using the nearest-neighbor model because the mismatches are in close proximity of each other. As a result, the $Tmut + Pref$ interaction, which can be seen as a flattening of the curve due to a lower increase in the signal, occurs at a lower mutant ratio than expected without accounting for the proximity of the mismatches. This thermodynamic insight leads to an important design concept for future reference probes, namely that the additional mismatch of the reference probe to $Tmut$ compared to Twt is ideally separated more than five nucleotides from the mismatches already present in $Pref + Twt$ in order to maximize the free energy penalty.

The determined value of the physical parameter $\Delta\Delta G_{Pref}$ can now be used in our extended model to accurately determine the mutant ratio in the hybridization assay. The signal as a function of c_{mut}/c_{wt} for the extended model is shown in Fig. 2.5D. The extension resolves the disagreement between depletion and no-depletion results, and agreement is observed in Fig. 2.5B. We finally note that the theoretical extension and the addition of parameter $\Delta\Delta G_{Pref}$ are important only at high mutant ratios, where it allows a more accurate determination of the mutant ratio, and the extension has a negligible effect on samples with a low mutant ratio and the analysis of the previous sections is retained.

2.4 Conclusions

We have successfully applied the thermodynamics-based concept of wild-type target depletion to clinical circulating free DNA samples using a hybridization assay for single nucleotide variation detection, demonstrating a 10-fold increase in sensitivity. We found an agreement between our data and the qualitative results of a clinically-validated assay, as well as a quantitative agreement on mutant ratios present in the samples compared to digital PCR. By incorporating all relevant molecular interactions in the thermodynamic model, we enlarged the dynamic range within which a quantitative determination of the mutant ratio can be obtained by a factor 10. The assay therefore becomes suitable for both sensitive detection of low mutant levels in early diagnosis as well as the accurate determination of high mutant abundances which are clinically relevant e.g. in the follow-up of advanced cancer. In addition, the thermodynamic analysis gave insight into future reference probe designs. The possibility for parallel detection of multiple nucleotide variants needs to be further explored since parallelization is a strength of the hybridization technology.

Our thermodynamic approach starts from physicochemical properties of DNA and is independent of the readout system of the used sensor. It can therefore be applied to a wide range of hybridization-based biosensing technologies and platforms, including cost-effective sensors operable at room temperature. This method can be utilized as a multiplex platform for the simultaneous detection of various single nucleotide variations (SNVs) using the same depletion mechanism and reference probe. Supplementary Information S9 outlines the steps required to transfer the hybridization method to alternative platforms or to modify the design for detecting different target sequences.

It is a robust and effective method to improve the performance of existing biosensors which currently lack the sensitivity required for the analysis of single nucleotide variations.

Author Contributions: Conceptualization, JH, PW, IN; methodology - investigation - data curation, RVH, YS, KH, MS, IN, JH; resources; GF, BM, RVH, IN; writing—original draft preparation, YS, RVH, JH; writing—review and editing, IN, PW, GF, BM, MS; Funding acquisition, JH, PW, IN.

Funding: This work was supported by the Research Foundation Flanders (FWO) 1S69320N/1S69322N; Special Research Fund BOF220WB11; EU H2020 M3DLoC project 760662 and VITO nv.

Ethical approval: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Review Committee of the Jessa Hospital (protocol code 2021/023, 08/09/2021). Patient consent was waived due to the fact that only residual material was used for this study.

Copyright Information: This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#). Formal publication available at:

doi:[10.1016/j.bios.2025.117342](https://doi.org/10.1016/j.bios.2025.117342).

Supplementary Information

S1 Overview and workflow of assays

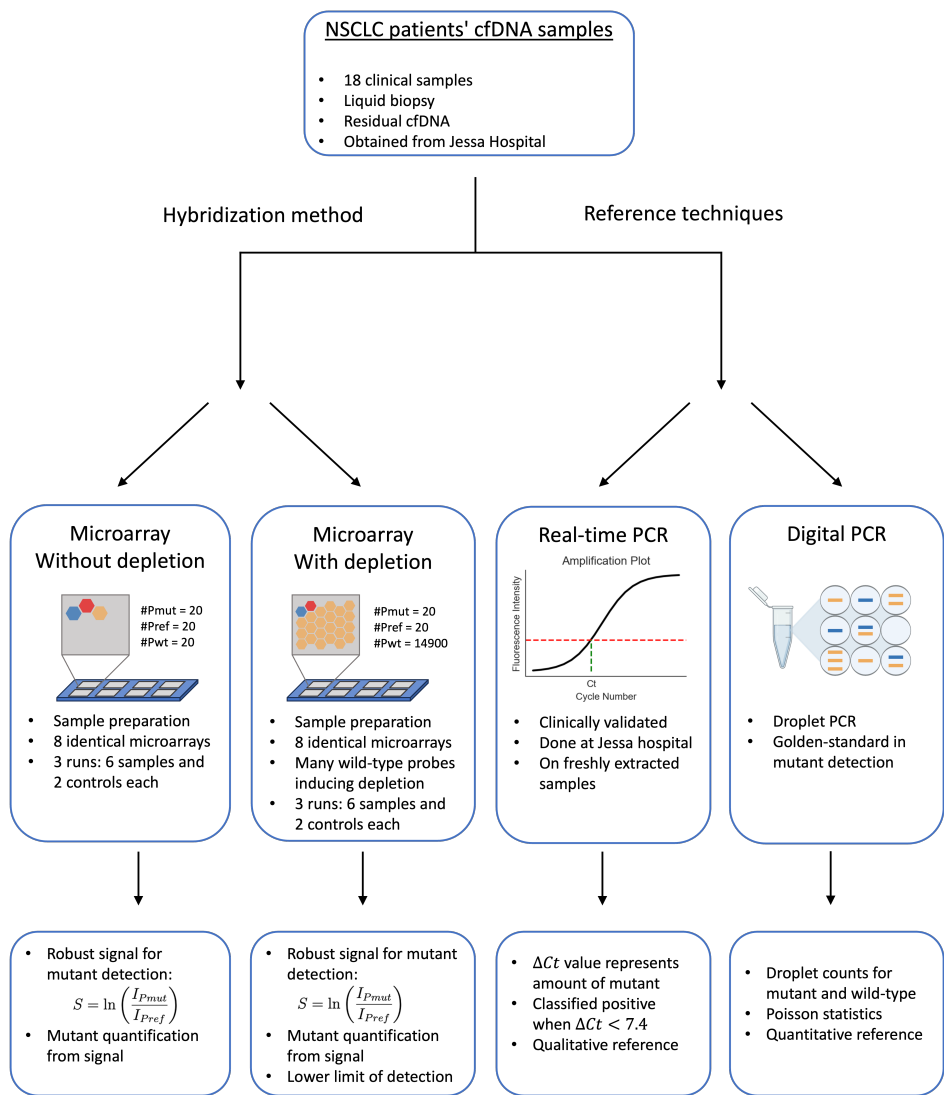


Figure S1: Overview and workflow of the used mutation detection assays. The hybridization method is tested in two variants, without and with depletion, for which separate microarray slides are used. The depletion slide, with 8 identical microarrays, has microarrays that contain a large amount of wild-type matching probes, that induce wild-type target depletion and result in a better limit of detection. The reference techniques are the clinically validated real-time PCR performed at Jessa hospital and digital PCR.

S2 Clinical diagnostic information on cfDNA samples

Table S1: Clinical diagnostic information on cfDNA samples obtained from 18 NSCLC patients. Real-time PCR threshold cycle (C_t) values are obtained for the *EGFR* T790M mutation and for the housekeeping gene beta-2-microglobulin (B2M). The C_t value for B2M was subtracted from the C_t value for *EGFR* T790M in order to determine a delta C_t (ΔC_t) value and *EGFR* T790M mutational status, which is positive when $\Delta C_t \leq 7.4$.

cfDNA sample number	DNA concentration (ng/ μ L)	Ct B2M	Ct EGFR T790M	ΔC_t	EGFR T790M mutational status
1	0.42	26.6	31.4	4.8	positive
2	0.46	26.2	37.2	11	negative
3	7.12	24.5	31.0	6.5	positive
4	0.26	26.0	>40	/	negative
5	0.21	26.3	>40	/	negative
6	0.72	26.0	>40	/	negative
7	0.89	26.6	>40	/	negative
8	0.50	26.1	>40	/	negative
9	0.37	25.9	>40	/	negative
10	2.91	27.1	28.9	1.8	positive
11	1.19	27.1	33.8	6.7	positive
12	9.02	25.6	38.0	12.4	negative
13	10.5	25.2	31.6	6.4	positive
14	0.42	25.6	31.4	5.8	positive
15	0.39	25.7	35.6	9.9	negative
16	0.71	25.4	>40	/	negative
17	0.42	N/A	>40	/	negative
18	0.33	26.2	>40	/	negative

S3 Digital PCR information

Table S2: Digital droplet PCR analysis on cfDNA samples obtained from 18 NSCLC patients. Mutant ratio and 95% confidence intervals are obtained using Poisson statistics. Two template negative controls (NTC) are included and showed the expected result.

cfDNA sam- ple num- ber	Droplets			Mutant content		
	Counts	Mutant Posi- tive	WT Posi- tive	Mutant ratio c_{mut}/c_{wt}	95% CI Min	95% CI Max
1	17391	63	391	0.160	0.117	0.202
2	17334	0	479	0		
3	2776	56	1066	/		
4	18047	0	263	0		
5	16601	0	307	0		
6	17620	0	744	0		
7	6168	11	217	/		
8	4363	0	110	/		
9	18088	1	523	0.002	0	0.006
10	15600	1847	3131	0.562	0.530	0.595
11	15848	50	1100	0.044	0.031	0.056
12	14354	49	5115	0.008	0.006	0.010
13	14569	432	6256	0.054	0.048	0.059
14	16439	15	500	0.030	0.014	0.045
15	16874	1	285	0.003	0	0.012
16	17869	9	953	0.009	0.003	0.015
17	17355	2	399	0.005	0	0.013
18	17175	1	260	0.004	0	0.013
NTC	17285	0	0			
NTC	18423	0	0			

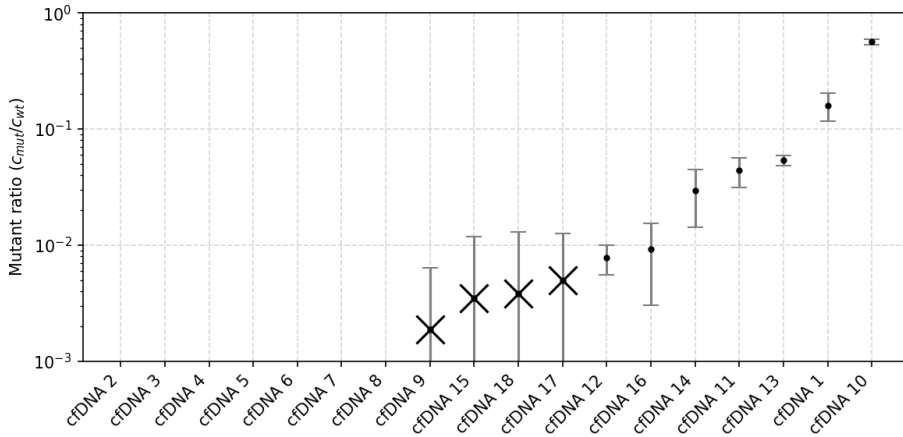


Figure S2: Mutant ratio and 95% confidence intervals resulting from digital droplet PCR analysis. cfDNA Samples on x-axis are sorted according to increasing mutant content. Samples where the number of mutant positive droplets < 3 are denoted with an 'x'.

S4 PCR amplicon information

S4.1 EGFR template sequences

Sequence of EGFR template gBlocks™ Gene Fragments with deviating nucleotides C/T for the wild-type and the EGFR T790M mutant, respectively:

```

5' TAA ACG TCC CTG TGC TAG GTC TTT TGC AGG CAC AGC
TTT TCC TCC ATG AGT ACG TAT TTT GAA ACT CAA GAT CGC
ATT CAT GCG TCT TCA CCT GGA AGG GGT CCA TGT GCC CCT
CCT TCT GGC CAC CAT GCG AAG CCA CAC TGA CGT GCC TCT
CCC TCC CTC CAG GAA GCC TAC GTG ATG GCC AGC GTG GAC
AAC CCC CAC GTG TGC CGC CTG CTG GGC ATC TGC CTC ACC
TCC ACC GTG CAG CTC ATC AC/TG CAG CTC ATG CCC TTC
GGC TGC CTC CTG GAC TAT GTC CGG GAA CAC AAA GAC AAT
ATT GGC TCC CAG TAC CTG CTC AAC TGG TGT GTG CAG ATC
GCA AAG GTA ATC AGG GAA GGG AGA TAC GGG GAG GGG
AGA TAA GGA GCC AGG ATC CTC ACA TGC GGT CTG CGC TCC
TGG GAT AGC AAG AGT TTG CCA TGG GGA TAT GTG TGT GCG
TGC ATG CAG CAC ACA CAC ATT CCT TTA TTT TGG ATT CAA
TC 3'

```

S4.2 Gel electrophoresis

For gel electrophoresis, 4 μL of 5 ng/ μL PCR product from wild-type EGFR gBlocks™ Gene Fragments, NTC PCR product or nuclease- free water (as NTC for the gel electrophoresis) was mixed with 1 μL of 5x loading dye. The mixtures or 5 μL of 5 times diluted FlashGel™ DNA marker 50 bp - 1500 bp (Lonza Group AG, #57033) were loaded in a well of a 2.2% Agarose FlashGel™ DNA Cassette (Lonza Group AG, #57031). The gel was run at 275 V, and images were taken using the FlashGel™ camera and capture software (Lonza Group AG).



Figure S3: Evaluation of PCR amplification products. 1: FlashGel™ DNA marker 50 bp - 1500 bp (M), 2-4: PCR product from wild-type EGFR gBlocks™ Gene Fragments (EGFR PCR product), 5-7: no template control PCR product (NTC PCR), 8-9: no template control gel electrophoresis (NTC gel), and 10: FlashGel™ DNA marker (M).

S5 Hybridization method: microarray results

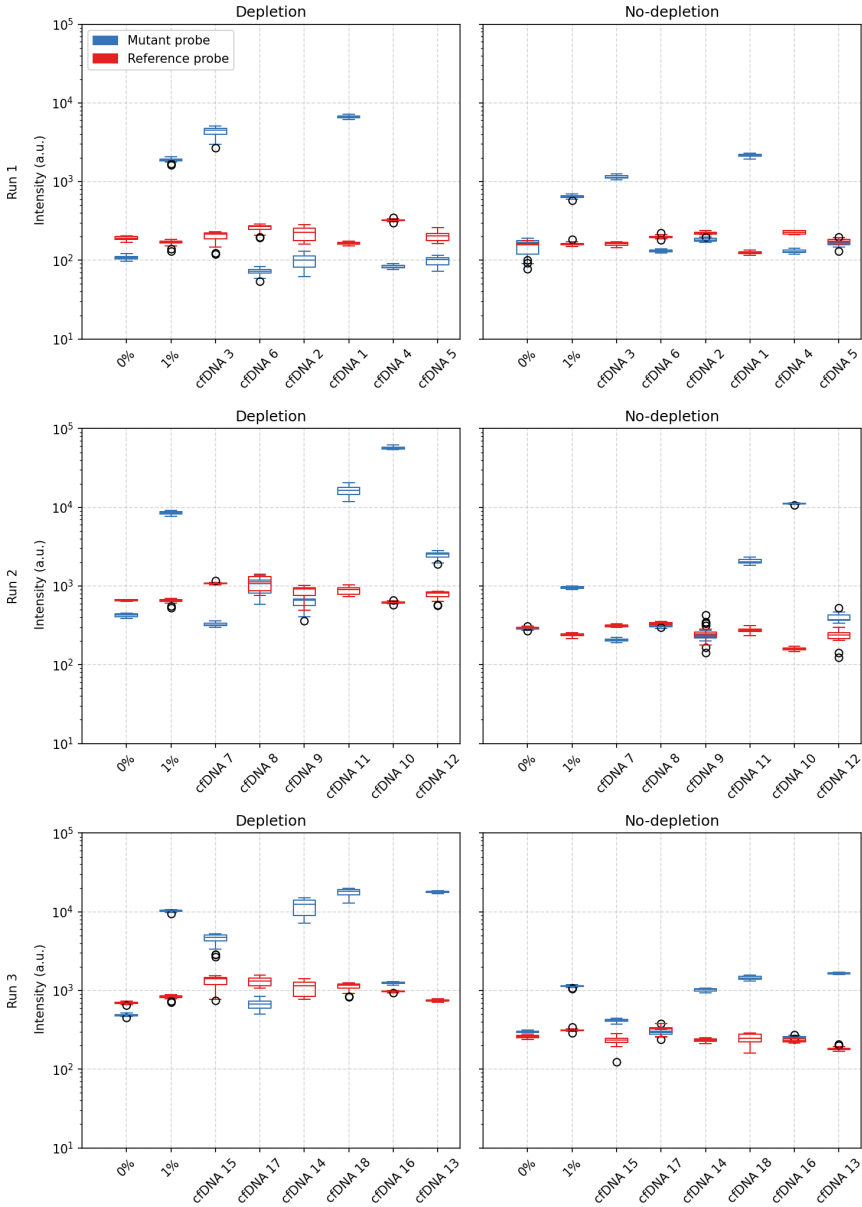


Figure S4: Microarray results for the hybridization method. cDNA samples from 18 NSCLC patients were analyzed in three runs. For each run, depletion and no-depletion conditions were evaluated using a mutant-matching probe and a reference probe for boxplots are shown. The mutant detection signal is calculated from these data using Eq. (2.1). Whiskers indicate the furthest point in the interquartile range, and outliers beyond these are plotted as open circles. In addition, 0% and 1% control samples were included for baseline subtraction.

S6 Derivation of the extension of the model for larger mutant ratios

Let θ_{ij} denote the fraction of probes of type i hybridized to target of type j . When targets are present at concentrations c_j and probes at concentrations a_i , dynamics are governed by

$$\frac{d\theta_{ij}}{dt} = k_{ij}^+ \underbrace{\left(1 - \sum_{m=1}^{n_t} \theta_{im}\right)}_{\text{saturation of probe } i} \underbrace{\left(c_j - \sum_{n=1}^{n_p} a_n \theta_{nj}\right)}_{\text{depletion of target } j} - k_{ij}^- \theta_{ij}, \quad (\text{S1})$$

where k_{ij}^+ and k_{ij}^- are the association and dissociation constants. In the case where we have three probes (P_{wt} , P_{mut} and P_{ref} , $i = 1, 2$ and 3 respectively), two targets (T_{wt} and T_{mut} , $j = 1$ and 2 respectively), and avoid saturation of the probes, this becomes

$$\frac{d\theta_{ij}}{dt} = k_{ij}^+ \underbrace{(c_j - a_1 \theta_{1j} - a_2 \theta_{2j} - a_3 \theta_{3j})}_{c_j^{\text{eff}}} - k_{ij}^- \theta_{ij}, \quad (\text{S2})$$

with $\theta_i = \theta_{i1} + \theta_{i2}$ representing the fraction of probes that are occupied by any target, such that $I_i \sim \theta_i$. We also introduced an effective target j concentration c_j^{eff} . At equilibrium

$$\theta_{ij} = c_j^{\text{eff}} K_{ij}, \quad (\text{S3})$$

where the $K_{ij} \equiv k_{ij}^+ / k_{ij}^- = e^{-\Delta G_{ij} / RT}$ are equilibrium constants.

Without depletion probes, the reservoir condition is satisfied and effective target concentrations are practically equal to c_j .

$$\text{No-depletion:} \quad c_j^{\text{eff}} = c_j. \quad (\text{S4})$$

$$(\text{S5})$$

In case of a large amount of wild-type matching depletion probes, effective wild-type and mutant concentrations are lowered significantly. When a_1 is large, we can assume depletion is only due to wild-type matching probes,

$$c_j^{\text{eff}} = c_j - a_1 \theta_{1j} - a_2 \theta_{2j} - a_3 \theta_{3j} \quad (\text{S6})$$

$$\approx c_j - a_1 \theta_{1j}. \quad (\text{S7})$$

When the effective target concentration is low, θ_{1j} is linear to effective target concentration c_j^{eff} and therefore

$$c_j^{\text{eff}} \approx c_j - a_1 c_j^{\text{eff}} K_{1j} \quad (\text{S8})$$

$$c_j^{\text{eff}} \approx \frac{c_j}{1 + a_1 K_{1j}}. \quad (\text{S9})$$

Finally, in case of strong depletion $a_1 K_{1j} \gg 1$ and the effective concentration in case of depletion become

$$\text{Depletion:} \quad c_j^{\text{eff}} = \frac{c_j}{a_1 K_{1j}}, \quad (\text{S10})$$

resulting in an enhanced effective mutant ratio

$$\frac{c_2^{\text{eff}}}{c_1^{\text{eff}}} = \frac{c_2 K_{11}}{c_1 K_{12}} = \frac{c_2}{c_1} e^{\Delta \Delta G_{P_{wt}}/RT} > \frac{c_2}{c_1}. \quad (\text{S11})$$

The detection signal is given by

$$S = \ln \frac{\theta_{\text{Pmut}}}{\theta_{\text{Pref}}} \quad (\text{S12})$$

$$= \ln \frac{\theta_{21} + \theta_{22}}{\theta_{31} + \theta_{32}} \quad (\text{S13})$$

$$= \ln \left(\frac{c_1^{\text{eff}} K_{21} + c_2^{\text{eff}} K_{22}}{c_1^{\text{eff}} K_{31} + c_2^{\text{eff}} K_{32}} \right) \quad (\text{S14})$$

$$= \ln \left(\frac{K_{21}}{K_{31}} \right) + \ln \left(\frac{1 + \frac{c_2^{\text{eff}} K_{22}}{c_1^{\text{eff}} K_{21}}}{1 + \frac{c_2^{\text{eff}} K_{32}}{c_1^{\text{eff}} K_{31}}} \right) \quad (\text{S15})$$

$$= \ln \left(\frac{K_{21}}{K_{31}} \right) + \ln \left(1 + \frac{c_2^{\text{eff}} K_{22}}{c_1^{\text{eff}} K_{21}} \right) - \ln \left(1 + \frac{c_2^{\text{eff}} K_{32}}{c_1^{\text{eff}} K_{31}} \right), \quad (\text{S16})$$

The first term is a baseline S_0 , which is close to zero due to the probe sequence selection of P_{ref} . The second term represents the mutant probe interaction, and the last term represents the reference probe interaction. The baseline-subtracted signal (which is described in the main text as just the signal S) is now given by

$$\text{No-depletion:} \quad S - S_0 = \ln \left(1 + \frac{c_2 K_{22}}{c_1 K_{21}} \right) - \ln \left(1 + \frac{c_2 K_{32}}{c_1 K_{31}} \right), \quad (\text{S17})$$

$$\text{Depletion:} \quad S - S_0 = \ln \left(1 + \frac{c_2 K_{22} K_{21}}{c_1 K_{21} K_{11}} \right) - \ln \left(1 + \frac{c_2 K_{32} K_{21}}{c_1 K_{31} K_{11}} \right). \quad (\text{S18})$$

Finally, one can write the signal in terms of free-energy differences by substituting the ratios of equilibrium constants.

S7 Mutant ratio determination in the extended model

Inverting the equations from the signal in terms of the mutant ratio results in the equations to obtain the mutant ratio from the signals.

No-depletion signal:

$$S_{no_depl} = \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{Pmut}}{RT} \right) \right) - \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{Pref}}{RT} \right) \right) \quad (S19)$$

Depletion signal:

$$S_{depl} = \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{Pmut} + \Delta\Delta G_{Pwt}}{RT} \right) \right) - \ln \left(1 + \frac{c_{mut}}{c_{wt}} \exp \left(\frac{\Delta\Delta G_{Pref} + \Delta\Delta G_{Pwt}}{RT} \right) \right) \quad (S20)$$

No-depletion mutant ratio:

$$\frac{c_{mut}}{c_{wt}} = \frac{\exp(S_{no_depl}) - 1}{\exp\left(\frac{\Delta\Delta G_{Pmut}}{RT}\right) - \exp\left(S_{no_depl} + \frac{\Delta\Delta G_{Pref}}{RT}\right)} \quad (S21)$$

Depletion mutant ratio:

$$\frac{c_{mut}}{c_{wt}} = \frac{\exp(S_{depl}) - 1}{\exp\left(\frac{\Delta\Delta G_{Pmut}}{RT}\right) - \exp\left(S_{depl} + \frac{\Delta\Delta G_{Pref}}{RT}\right)} \exp\left(-\frac{\Delta\Delta G_{Pwt}}{RT}\right) \quad (S22)$$

S8 Determination of $\Delta\Delta G_{Pref}$ and fitting procedure

For ease of notation, let us introduce parameters

$$\Delta x_1 \equiv \Delta\Delta G_{Pmut}/RT, \quad (S23)$$

$$\Delta x_2 \equiv \Delta\Delta G_{Pref}/RT, \quad (S24)$$

$$\Delta x_d \equiv \Delta\Delta G_{Pwt}/RT, \quad (S25)$$

and abbreviate $S_{nd} \equiv S_{no_depl}$ and $S_d \equiv S_{depl}$. Using this notation, the mutant ratio calculated from the hybridization experiment without depletion is expressed as

$$\frac{c_{mut}}{c_{wt}} = \frac{e^{S_{nd}} - 1}{e^{\Delta x_1} - e^{\Delta x_2 + S_{nd}}} \quad (S26)$$

The hybridization experiment with depletion for the same sample, results in a signal S_d that is given by

$$S_d = \ln \left(1 + \frac{c_{mut}}{c_{wt}} e^{\Delta x_1 + \Delta x_d} \right) - \ln \left(1 + \frac{c_{mut}}{c_{wt}} e^{\Delta x_2 + \Delta x_d} \right) \quad (S27)$$

Inserting equation (S26) into equation (S27) we find:

$$S_d = \ln \left(1 + \frac{e^{S_{nd}} - 1}{e^{\Delta x_1} - e^{\Delta x_2 + S_{nd}}} e^{\Delta x_1 + \Delta x_d} \right) \quad (S28)$$

$$- \ln \left(1 + \frac{e^{S_{nd}} - 1}{e^{\Delta x_1} - e^{\Delta x_2 + S_{nd}}} e^{\Delta x_2 + \Delta x_d} \right) \quad (S29)$$

$$= \ln (e^{\Delta x_1} - e^{\Delta x_2 + S_{nd}} + e^{\Delta x_1 + S_{nd} + \Delta x_d} - e^{\Delta x_1 + \Delta x_d}) \quad (S30)$$

$$- \ln (e^{\Delta x_1} - e^{\Delta x_2 + S_{nd}} + e^{\Delta x_2 + S_{nd} + \Delta x_d} - e^{\Delta x_2 + \Delta x_d}) \quad (S31)$$

$$= \ln \left(\frac{1 - e^{\Delta x_2 - \Delta x_1 + S_{nd}} + e^{\Delta S_{nd} + \Delta x_d} - e^{\Delta x_d}}{1 - e^{\Delta x_2 - \Delta x_1 + S_{nd}} + e^{\Delta x_2 - \Delta x_1 + S_{nd} + \Delta x_d} - e^{\Delta x_2 - \Delta x_1 + \Delta x_d}} \right) \quad (S32)$$

$$S_d = \ln \left(\frac{1 - e^{\Delta x_2 - \Delta x_1 + S_{nd}} + e^{\Delta S_{nd} + \Delta x_d} - e^{\Delta x_d}}{1 - e^{\Delta x_2 - \Delta x_1} (e^{S_{nd}} + e^{S_{nd} + \Delta x_d} - e^{\Delta x_d})} \right) \quad (S33)$$

Equation (S33) represents the depletion signal S_d in function of the no-depletion signal S_{nd} on the same sample, given the values of Δx_d and the difference $\Delta x_2 - \Delta x_1$. This allows Δx_d and $\Delta x_2 - \Delta x_1$ to be fitted when both depletion and no-depletion measurements are performed on multiple samples. Figure S5A shows S_d in function of S_{nd} for the clinical data, along

with a fit to equation (S33).

Fit parameters were determined to be

$$\Delta x_d = \Delta\Delta G_{P_{wt}}/RT = 1.87 \pm 0.08 \quad (\text{S34})$$

$$\Delta x_2 - \Delta x_1 = \Delta\Delta G_{P_{ref}}/RT - \Delta\Delta G_{P_{mut}}/RT = -5.15 \pm 0.17 \quad (\text{S35})$$

Adding $\Delta\Delta G_{P_{mut}}/RT = 4.7$ to equation (S35), we obtain $\Delta\Delta G_{P_{ref}}/RT = -0.45$ as presented in the main text.

Additionally, Fig. S5B shows two lines, one showing equation (S33) and one line where the mutant-reference probe interaction is omitted from the theory i.e. where the second term is removed from equation (S27). Figure S5 shows once again that the extension of the model only affects samples with large signals, and shows that sample 1 and sample 10 do indeed follow the suggested model, which now includes the mutant-reference probe interaction.

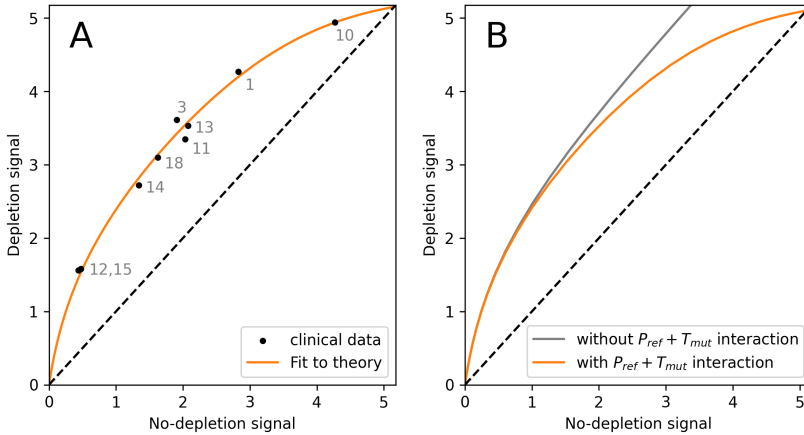


Figure S5: Depletion signal in function of the no-depletion signal of the same sample. (A) Clinical NSCLC cfDNA sample measurements. The orange line is a fit of parameters Δx_d and $\Delta x_2 - \Delta x_1$ to equation (S33). (B) The effect of the addition of the $P_{ref} + T_{mut}$ interaction. The addition of the $P_{ref} + T_{mut}$ interaction is only noticeable at larger signals and mutant ratios and does not significantly affect smaller signal. As a visual guide a dashed diagonal is plotted in black.

S9 Transfer of the hybridization method to other platforms

Here we provide an overview of the general steps involved in adapting the hybridization method to a new target sequence or transferring it to a different readout instrument. The goal remains the same: one would like to detect single-stranded mutant DNA in a majority of SNV wild-type background. The method makes use of three different probes: the mutant matching probe P_{mut} , the wild-type matching probe P_{wt} and a reference probe P_{ref} . The ratio of the readouts of P_{mut} and P_{ref} determine the signal, and the wild-type matching probes are used to induce depletion, so therefore no readout mechanism is required for P_{wt} . Readout of P_{mut} and P_{ref} can occur on many platforms and through various mechanisms where the readout is linear in the bound concentration to the probes, examples include Surface Plasmon Resonance (SPR) [18], Electrochemical impedance spectroscopy (EIS) [19] or in-solution fluorescence [25].

1. Select the length of the binding region of mutant matching probe P_{mut} depending on the used instrument and temperature of operation. Preferably center the position of the mismatch in the $P_{mut}+T_{wt}$ binding to maximize the free energy penalty due to the mismatch.
2. To maximize the effect of depletion (strong depletion), either the number of wild-type matching probes in the system can be made large or the binding affinity of the wild-type matching probes can be increased (by e.g. increasing the length of the binding region). We recommend increasing the amount of probes over increasing the length because the latter will increase the time required for the system to reach equilibrium. Depletion of wild-type target can be achieved using different mechanisms, possibly in solution by introducing matching strands attached to magnetic beads or other capture mechanisms resulting in selective capturing of the wild-type target. Depletion can take place at the same time or as a separate step before the sample is allowed to bind to the mutant and reference probes.
3. The reference probe should fulfill the following two conditions:
 - The affinity of the reference probe toward wild-type target should

be the same as the wild-type target towards the mutant matching probe to ascertain equal binding to P_{mut} and P_{ref} in case of a wild-type only target.

- The affinity of the reference probe to the mutant target should be kept as low as possible to avoid a decrease in the signal at higher mutant ratios and to keep a large dynamic range.

For the first condition, nearest-neighbor predictions can be used to explore sequences with one, two or more mismatches and determine their affinity towards wild-type targets. Additionally, equal binding can be experimentally tested for candidate reference probes [29]. For the second condition, we recommend selecting a reference probe sequence where the additional mismatch occurs at a location ideally separated more than five nucleotides from the other mismatch locations [36]. This will minimize the (stabilizing) interaction between the mismatches, which is not predicted using nearest-neighbor calculations.

4. An appropriate target concentration should not saturate the probes. We recommend determining the concentration by testing a dilution series of the total concentration, for each of the 0%, 10% and 50% mutant containing sample and ensuring the P_{mut} readout is above background and not close to saturation.

After probe selection, the free-energy penalties are to be checked experimentally by constructing a dose-response curve where the percentage of mutant in control sample is varied. The exponents of equations (S20) can be fitted and used in equation (S22) for quantification of the mutant ratio c_{mut}/c_{wt} .

On top of this, the method can be extended to design a multiplex platform for the detection of different SNVs at the same time, using the same depletion mechanism and reference probe. Multiple mutant probes are used and they should match the SNV sequences, i.e. $P_{mut}^1, P_{mut}^2, \dots$. Each SNV has a corresponding signal $S_i = \log(I_{mut}^i/I_{ref})$ that can be used to determine the relative mutant concentration in the sample.

References

- [1] C. Lino, S. Barrias, R. Chaves, F. Adegas, P. Martins-Lopes, and J. R. Fernandes, "Biosensors as diagnostic tools in clinical applications," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1877, no. 3, p. 188726, 2022.
- [2] C. Abbosh *et al.*, "Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution," *Nature*, vol. 545, no. 7655, pp. 446–451, 2017.
- [3] W. Xie, S. Suryaprakash, C. Wu, A. Rodriguez, and S. Fraterman, "Trends in the use of liquid biopsy in oncology," *Nature Reviews Drug Discovery*, vol. 22, no. 8, pp. 612–613, 2023.
- [4] M. García-Pardo, M. Makarem, J. J. N. Li, D. Kelly, and N. B. Leighl, "Integrating circulating-free DNA (cfDNA) analysis into clinical practice: opportunities and challenges," *British Journal of Cancer*, vol. 127, no. 4, pp. 592–602, 2022.
- [5] S. Udomruk, S. Orrapin, D. Pruksakorn, and P. Chaiyawat, "Size distribution of cell-free DNA in oncology," *Critical Reviews in Oncology/Hematology*, vol. 166, p. 103455, 2021.
- [6] J. Lu, Y. Feng, K. Guo, L. Sun, S. Ruan, and K. Zhang, "Prognostic value of preoperative circulating tumor DNA in non-small cell lung cancer: a systematic review and meta-analysis," *Journal of Cancer Research and Clinical Oncology*, vol. 150, no. 1, p. 25, 2024.
- [7] L. Sivapalan, J. C. Murray, J. V. Canzoniero, B. Landon, J. Jackson, S. Scott, V. Lam, B. P. Levy, M. Sausen, and V. Anagnostou, "Liquid biopsy approaches to capture tumor evolution and clinical outcomes

- during cancer immunotherapy," *Journal for Immunotherapy of Cancer*, vol. 11, no. 1, p. e005924, 2023.
- [8] J. Zhou, H. Mo, D. Hu, X. Zhao, H. Zhou, and J. Pan, "Association of ctDNA detection and recurrence assessment in patients with neoadjuvant treatment," *Cancer Medicine*, vol. 12, no. 19, pp. 19794–19806, 2023.
- [9] O. B. Alese, N. Cook, A. Ortega-Franco, M. B. Ulanja, L. Tan, and J. Tie, "Circulating Tumor DNA: An Emerging Tool in Gastrointestinal Cancers," *American Society of Clinical Oncology Educational Book*, no. 42, pp. 279–298, 2022.
- [10] G. Alexandrou *et al.*, "The Evolution of Affordable Technologies in Liquid Biopsy Diagnostics: The Key to Clinical Implementation," *Cancers*, vol. 15, no. 22, p. 5434, 2023.
- [11] F. Ferrara, S. Zoupanou, E. Primiceri, Z. Ali, and M. S. Chiriaco, "Beyond liquid biopsy: Toward non-invasive assays for distanced cancer diagnostics in pandemics," *Biosensors and Bioelectronics*, vol. 196, p. 113698, 2022.
- [12] C. Silveira *et al.*, "Detection and quantification of EGFR T790M mutation in liquid biopsies by droplet digital PCR," *Translational Lung Cancer Research*, vol. 10, no. 3, pp. 1200–1208, 2021.
- [13] H. Takahashi *et al.*, "Mutation detection of urinary cell-free DNA via catch-and-release isolation on nanowires for liquid biopsy," *Biosensors and Bioelectronics*, vol. 234, p. 115318, 2023.
- [14] N. Wang, X. Zhang, F. Wang, M. Zhang, B. Sun, W. Yin, S. Deng, Y. Wan, and W. Lu, "The Diagnostic Accuracy of Liquid Biopsy in EGFR-Mutated NSCLC: A Systematic Review and Meta-Analysis of 40 Studies," *SLAS Technology*, vol. 26, no. 1, pp. 42–54, 2021.
- [15] X. Li, M. Ye, W. Zhang, D. Tan, N. Jaffrezic-Renault, X. Yang, and Z. Guo, "Liquid biopsy of circulating tumor dna and biosensor applications," *Biosensors and Bioelectronics*, vol. 126, pp. 596–607, 2019.
- [16] J. Zhuang, L. Xia, Z. Zou, J. Yin, N. Lin, and Y. Mu, "Recent advances in integrated microfluidics for liquid biopsies and future directions," *Biosensors and Bioelectronics*, vol. 217, p. 114715, 2022.

- [17] M. Alipour, S. Jalili, H. Shirzad, E. Ansari Dezfouli, M. H. Fouani, A. A. Sadeghan, H. Bardania, and S. Hosseinkhani, "Development of dual-emission cluster of Ag atoms for genetically modified organisms detection," *Microchimica Acta*, vol. 187, no. 11, p. 628, 2020.
- [18] A. Dillen, A. Mohrbacher, and J. Lammertyn, "A Versatile One-Step Competitive Fiber Optic Surface Plasmon Resonance Bioassay Enabled by DNA Nanotechnology," *ACS Sensors*, vol. 6, no. 10, pp. 3677–3684, 2021.
- [19] H. S. Magar, R. Y. A. Hassan, and A. Mulchandani, "Electrochemical Impedance Spectroscopy (EIS): Principles, Construction, and Biosensing Applications," *Sensors*, vol. 21, no. 19, p. 6578, 2021.
- [20] B. L. Mueller, M. J. Liberman, and D. M. Kolpashchikov, "OWL2: a molecular beacon-based nanostructure for highly selective detection of single-nucleotide variations in folded nucleic acids," *Nanoscale*, vol. 15, no. 12, pp. 5735–5742, 2023.
- [21] B. van Grinsven *et al.*, "Rapid assessment of the stability of DNA duplexes by impedimetric real-time monitoring of chemically induced denaturation," *Lab on a Chip*, vol. 11, no. 9, pp. 1656–1663, 2011.
- [22] B. van Grinsven *et al.*, "Heat-Transfer Resistance at Solid-Liquid Interfaces: A Tool for the Detection of Single-Nucleotide Polymorphisms in DNA," *ACS Nano*, vol. 6, no. 3, pp. 2712–2721, 2012.
- [23] F. Wei, P. Yu, J. Cheng, F. Li, D. Chia, and D. T. W. Wong, "Single-Droplet Microsensor for Ultra-Short Circulating EGFR Mutation Detection in Lung Cancer Based on Multiplex EFIRM Liquid Biopsy," *International Journal of Molecular Sciences*, vol. 24, no. 12, p. 10387, 2023.
- [24] H. N. Wood, T. Venken, H. Willems, A. Jacobs, A. J. Reis, P. E. Almeida da Silva, S. Homolka, S. Niemann, K. H. Rohde, and J. Hooyberghs, "Molecular drug susceptibility testing and strain typing of tuberculosis by DNA hybridization," *PloS One*, vol. 14, no. 2, p. e0212064, 2019.
- [25] R. J. Karadeema, M. Stancescu, T. P. Steidl, S. C. Bertot, and D. M. Kolpashchikov, "The owl sensor: a 'fragile' DNA nanostructure for the

- analysis of single nucleotide variations," *Nanoscale*, vol. 10, no. 21, pp. 10 116–10 122, 2018.
- [26] K. Knez, D. Spasic, K. P. F. Janssen, and J. Lammertyn, "Emerging technologies for hybridization based single nucleotide polymorphism detection," *Analyst*, vol. 139, no. 2, pp. 353–370, 2013.
- [27] J. Hooyberghs, M. Baiesi, A. Ferrantini, and E. Carlon, "Breakdown of thermodynamic equilibrium for DNA hybridization in microarrays," *Physical Review E*, vol. 81, no. 1, p. 012901, 2010.
- [28] S. K. Nomidis, M. Szymonik, T. Venken, E. Carlon, and J. Hooyberghs, "Enhancing the Performance of DNA Surface-Hybridization Biosensors through Target Depletion," *Langmuir*, vol. 35, no. 37, pp. 12 276–12 283, 2019.
- [29] R. Van Hoof, M. Szymonik, S. K. Nomidis, K. Hollanders, A. Jacobs, I. Nelissen, P. Wagner, and J. Hooyberghs, "Depletion of wild-type target enhances the hybridization-based sensitivity of low-abundant mutation detection by reference capture probes," *Sensors and Actuators B: Chemical*, vol. 368, p. 132175, 2022.
- [30] H. Willems, A. Jacobs, W. W. Hadiwikarta, T. Venken, D. Valkenburg, N. Van Roy, J. Vandesompele, and J. Hooyberghs, "Thermodynamic framework to assess low abundance DNA mutation detection by hybridization," *PLOS ONE*, vol. 12, no. 5, p. e0177384, 2017.
- [31] S. Wang, S. Cang, and D. Liu, "Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer," *Journal of Hematology & Oncology*, vol. 9, no. 1, p. 34, 2016.
- [32] B. Zhang *et al.*, "Comparison of droplet digital PCR and conventional quantitative PCR for measuring EGFR gene mutation," *Experimental and Therapeutic Medicine*, vol. 9, no. 4, pp. 1383–1388, 2015.
- [33] S. A. Cohen, M. C. Liu, and A. Aleshin, "Practical recommendations for using ctDNA in clinical decision making," *Nature*, vol. 619, no. 7969, pp. 259–268, 2023.
- [34] H. Yoshimine, K. Sasaki, and H. Furusawa, "Pocketable Biosensor Based on Quartz-Crystal Microbalance and Its Application to DNA Detection," *Sensors*, vol. 23, no. 1, p. 281, 2023.

- [35] A. Harrison *et al.*, "Physico-chemical foundations underpinning microarray and next-generation sequencing experiments," *Nucleic Acids Research*, vol. 41, no. 5, pp. 2779–2796, 2013.
- [36] W. W. Hadiwikarta, J.-C. Walter, J. Hooyberghs, and E. Carlon, "Probing hybridization parameters from microarray experiments: nearest-neighbor model and beyond," *Nucleic Acids Research*, vol. 40, no. 18, p. e138, 2012.
- [37] Bio-Rad Laboratories Inc., "Bulletin_6407.pdf (accessed in june 2024)," 2015.

Chapter 3

Paper B

Breaking the Limit of Hybridization Selectivity by Strain-forming Probes and Wild-type Blocker

Breaking the Limit of Hybridization Selectivity by Strain-forming Probes and Wild-type Blocker

Yannick Stulens¹, **Brittany L. Mueller**², **Jordan Hammock**², **Antonio Perez**², **Rebekka Van Hoof**³, **Dmitry M. Kolpashchikov**², and **Jef Hooyberghs**¹

Equal contribution

¹ *UHasselt, Data Science Institute, Theory Lab, Agoralaan, Diepenbeek, 3590, Belgium*

² *Chemistry Department, University of Central Florida, Orlando, FL 32816, United States*

³ *Flemish Institute for Technological Research (VITO), Environmental Intelligence Unit, Mol, 2400, Belgium*

SUBMITTED March 9, 2026

IN REVIEW

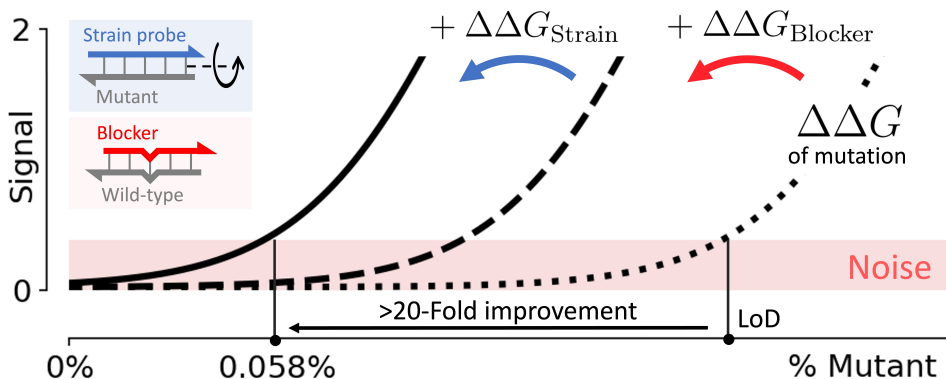
JOURNAL [Nucleic Acids Research](#)

Contribution

This work is a collaboration between the University of Central Florida and UHasselt. Brittany L. Mueller spent a research period in Belgium, while I conducted a research stay in Florida, USA. Brittany L. Mueller carried out most of the experiments, whereas I focused on formal analysis, theoretical modeling, and simulation-based optimization of the experimental design. During my stay in Florida, I also contributed to wet-lab optimization and initiated a preliminary study of the kinetic properties of the OWL sensor presented in Chapter 5. I obtained funding through an FWO travel grant and Erasmus+ support to enable the exchange. I wrote the original draft and designed the figures. All authors reviewed the manuscript.

Abstract

DNA hybridization is a fundamental molecular recognition process in nature and biotechnology. However, the finite free energy difference between matched and mismatched duplexes limits detection of rare variants. This challenge is particularly acute in liquid biopsies, where single-nucleotide variants in circulating tumor DNA must be distinguished from a vast excess of wild-type DNA. Here, we present a quantitative framework to analyze and improve the hybridization selectivity. We provide an expression for the Limit of Detection which nicely separates all parameters of the detection system from the intrinsic DNA properties. At first sight, it sets a fundamental physical constraint on low-abundant mutant detection. To overcome this limitation, we implement two complementary strategies: introducing conformational strain into the probe design and employing blocker strands to selectively deplete wild-type DNA. Using a clinically relevant EGFR mutation, we demonstrate a 21-fold improvement in detection performance, enabling identification of mutant DNA at a fractional abundance as low as 0.058%. Within our framework, both strategies contribute additively to the effective free energy penalty, and their individual effects are independently quantified. These concepts establish a general thermodynamic basis for understanding and improving hybridization selectivity and provide transferable principles for enhanced single-nucleotide variant detection in future technologies.



3.1 Introduction

Hybridization of single-stranded nucleic acids to form a duplex is a reversible chemical reaction that serves as the foundation for many biological processes and biotechnological techniques. Its utility arises from the inherent sequence selectivity of hybridization, making the quantitative understanding and exploitation of this property an active area of research. The importance of this field continues to grow with the increasing integration of sequence information into diagnostic applications. In recent years, liquid biopsies have emerged as a minimally invasive diagnostic tool in clinical diagnostics [1]. A particularly critical application is the detection of circulating tumor DNA (ctDNA) in patient samples, where single-nucleotide specificity is required. Since ctDNA is typically present in extremely low fractions of total circulating free DNA [2, 3], a major challenge remains to lower the detection limit. Many technologies, including PCR and next-generation sequencing, rely on hybridization performance [4, 5, 6]. Here, we develop a framework to understand, quantify, and improve the sequence specificity of DNA hybridization and demonstrate its impact on improving the detection limit in a relevant experimental setting. The concepts and conclusions are independent of the specific experimental implementation and are transferable to other hybridization-based technologies. We address the most challenging scenario of a minor fraction of a single-nucleotide variant (mutDNA) within a vast majority of wild-type DNA (wtDNA). The concentrations of these single-stranded targets are denoted as c_{mut} and c_{wt} respectively. Given our focus on hybridization, the primary detection tool consists of a single-stranded DNA oligonucleotide probe complementary to the mutDNA. Although the probe preferentially binds mutDNA, cross-hybridization with wtDNA is inevitable and limits the detection of mutDNA at low $c_{\text{mut}}/c_{\text{wt}}$ ratios [7, 8]. The Results section commences with a rigorous analysis of the underlying principles of the Limit of Detection (LoD, the smallest detectable ratio $c_{\text{mut}}/c_{\text{wt}}$) and we provide a suitable definition for the signal S derived from the probe-readout. We demonstrated that within the concentration regime relevant for optimal detection, an accurate description is captured by the following elegant expression:

$$LoD = S_{\min} e^{-\Delta\Delta G/RT}. \quad (3.1)$$

In equation (3.1), $S_{\min}$ is the minimal experimental signal, corresponding to the noise measured for a negative sample ($c_{\text{mut}} = 0$). Further, R is the universal gas constant, T is the temperature and $\Delta\Delta G$ is the free energy penalty of the point mutation being studied. The merit of this expression lies in its ability to encapsulate all parameters of the detection system into a single factor, $S_{\min}$, while isolating the physicochemical properties of the DNA within the other factor. When the detection system is defined, the LoD is fully determined by $\Delta\Delta G$ revealing a fundamental physical constraint that leaves no room for further improvement at first sight. The situation is analogous to optical microscopy, where the resolving power is fixed by the numerical aperture and the wavelength of the light used. The numerical aperture captures the total optical configuration, and once set, the wavelength of light represents a fundamental ‘hard limit’ on the resolving power imposed by the laws of physics. However, much as super-resolution techniques were developed for optical microscopy, it is nonetheless possible to circumvent the physical constraints on the LoD for low-abundant mutDNA.

The central hypothesis is that the effective free-energy difference $\Delta\Delta G$ between matched and mismatched probe-target complexes can be increased well beyond conventional limits through rational probe design. To demonstrate and quantify this, we apply two complementary strategies. The first strategy increases the binding selectivity through the introduction of conformational strain in the DNA probe. The second strategy does not alter the probe-target interaction, but depletes the wtDNA by complementary blocker strands. We show that, in the optimal regime, equation (3.1) remains fully valid, while the effective free energy penalty $\Delta\Delta G_{\text{eff}}$ is enhanced:

$$\Delta\Delta G_{\text{eff}} = \Delta\Delta G + \Delta\Delta G_{\text{Strain}} + \Delta\Delta G_{\text{Blocker}}. \quad (3.2)$$

Since the free energy penalty enters equation (3.1) via the exponential, small adjustments of it have a big impact on the LoD . This work introduces a thermodynamic framework for the elegant and quantitative interpretation of various LoD -improvement strategies as additive free energy terms. We provide a method for determining these terms and demonstrate their experimental application for a clinically relevant point mutation of the epidermal growth factor receptor (EGFR) [9, 10, 11].

3.2 Materials and Methods

3.2.1 Reagents

All solutions were made using DNase- and protease-free water purchased from Fisher Scientific. Synthesized oligonucleotides were obtained from Integrated DNA Technologies, Inc. (Coralville, IA). Concentrations of oligonucleotides stock solutions were determined via absorbance at 260 nm on a ThermoScientific NanoDrop One spectrophotometer (Waltham, MA). Mixed target samples were prepared by combining mutant and wild-type EGFR at the desired ratios. EGFR target sequences and sensor component sequences are listed in Supplementary Information S1.

3.2.2 Sensor annealing and reaction conditions

The OWL2 sensor was assembled by combining oligonucleotide components into a 2.0 mL microcentrifuge tube containing hybridization buffer (50 mM Tris-HCl, 50 mM MgCl₂, 0.1% (w/v) Tween-20, pH 7.4) to a final volume of 1.0 mL. Samples were heated to 95 °C for 5 min in a 2 L water bath and left to cool overnight (~8-10 h). Components T1, T2, T3 and T4 were kept constant and equal to the target concentration for each experiment. Following sensor annealing, the UMB probe was added at a fixed concentration of 50 nM. Target, blocker, and P-strand concentrations were varied as indicated for each experiment.

3.2.3 Fluorescent assay and analysis

The fluorescence of each sample was measured at 25 °C in a quartz cuvette on a Shimadzu RF-6000 spectrometer with excitation and emission wavelengths of 485 nm and 517 nm, respectively. Milli-Q water was added to samples with all oligonucleotide sequences but without target to allow for background subtraction. Three independent repeat measurements result in a mean estimation of the signal S and standard deviation error bars. Determination of $\Delta\Delta G_{\text{eff}}$ and its standard error was done by fitting

the dose-response curves using the *lmfit* module in Python on equation (3.11). The errors in the differences of $\Delta\Delta G_{\text{eff}}$ are propagated using absolute error propagation.

3.3 Results

3.3.1 Detection Signal and Limit of Detection

To establish the relation between the *LoD* and thermodynamics, we first introduce a general scenario for the detection of mutDNA in a mixture using the concept of a "basic sensor". The basic sensor utilizes a single-stranded DNA *probe* designed to match the mutDNA sequence, but partial complementarity still leads to some hybridization with wtDNA at reduced affinity. A sample contains single-stranded *target* DNA at a fixed total concentration $c_{\text{tot}} = c_{\text{wt}} + c_{\text{mut}}$ (Figure 3.1). The sensor's readout, e.g. a fluorescent intensity F , is measured at equilibrium and is proportional to the concentration of probe-target complexes. We restrict our analysis to the linear-response regime, i.e. where neither probes nor target run out during binding: the unbound probe and target concentrations effectively remain at their total concentrations. One can prove that in this regime the sensor has the best performance in distinguishing mutDNA from wtDNA (Supplementary Information S2). Under these conditions, wtDNA and mutDNA contribute independently to F , each proportional to their respective concentrations and association constants. The background-subtracted fluorescence can therefore be written as

$$F = A (c_{\text{wt}}K_{\text{P:wt}} + c_{\text{mut}}K_{\text{P:mut}}), \quad (3.3)$$

where A is a sensor-specific proportionality factor. The association constants follow the standard thermodynamic relation $K_{\text{P:X}} = \exp(-\Delta G_{\text{P:X}}/RT)$, where $\Delta G_{\text{P:X}}$ is the binding free energy of probe (P) to target X (wtDNA or mutDNA).

To detect mutDNA in a sample, a comparison is made to a reference sample containing only wtDNA, see Figure 3.1, for which the readout is $F_{\text{ref}} = A c_{\text{tot}}K_{\text{P:wt}}$. To quantify the comparison, we define the ratio-based

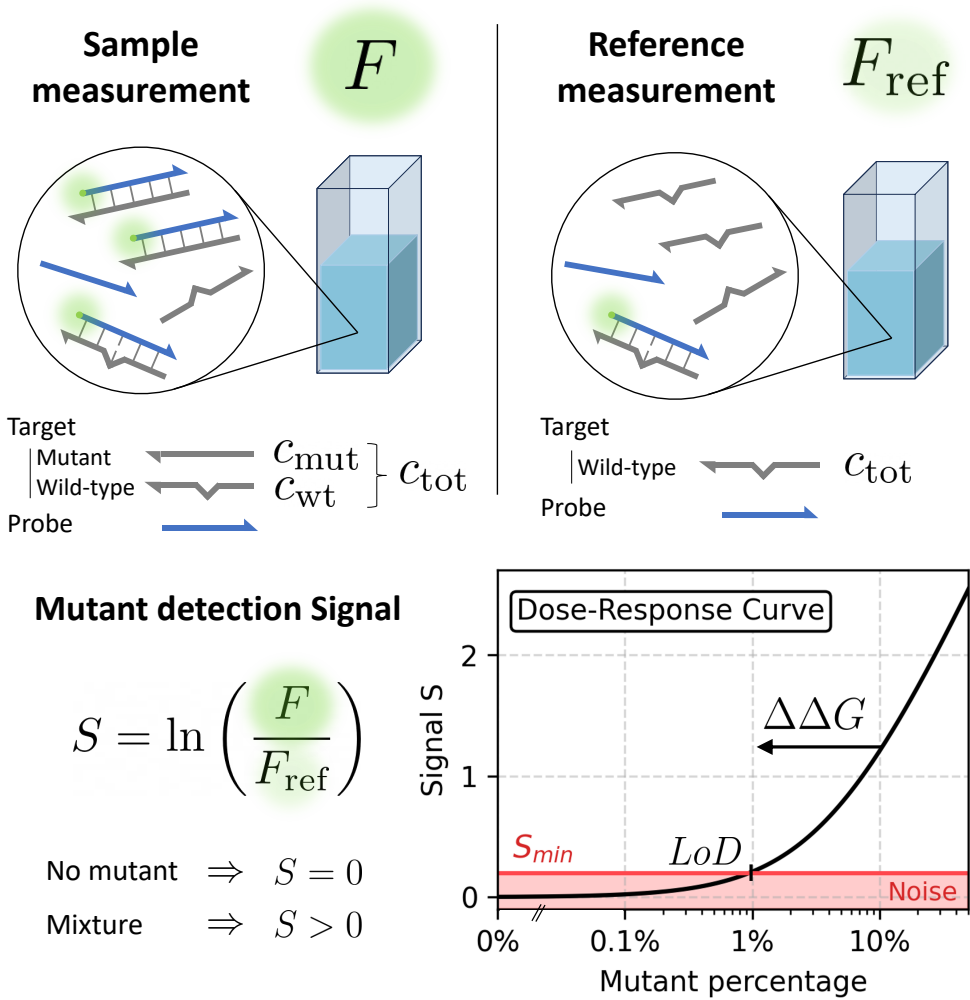


Figure 3.1: Schematic for low-abundant mutant detection. The fluorescent readout of a (possibly) mutant-containing sample is compared to a wild-type only reference sample at the same total concentration. The ratio-based mutant detection signal S changes with the mutant percentage $c_{\text{mut}}/c_{\text{tot}}$, leading to a dose-response curve which is characterized by the binding free-energy difference $\Delta\Delta G$ of the probe towards its two targets. The relative detection limit LoD is identified as the mutant percentage at which the signal exceeds the minimum detection threshold S_{min} .

mutant detection signal $S = \ln(F/F_{\text{ref}})$, leading to a robust unitless quantity [12, 13]. When no mutDNA is present $S_0 = 0$, but if present, S is positive. Expressing the mutant detection signal S in terms of the relative concentration $c_{\text{mut}}/c_{\text{tot}}$ and association constants provides insight into the determining physical parameters.

$$S = \ln\left(\frac{F}{F_{\text{ref}}}\right) = \ln\left(1 + \frac{c_{\text{mut}}K_{\text{P:mut}}}{c_{\text{tot}}K_{\text{P:wt}}}\right) \quad (3.4)$$

$$= \ln\left(1 + \frac{c_{\text{mut}}}{c_{\text{tot}}}e^{\Delta\Delta G/RT}\right) \quad (3.5)$$

Signal S increases with $c_{\text{mut}}/c_{\text{tot}}$, resulting in a dose-response curve that is characterized by the free-energy penalty of the mutation $\Delta\Delta G \equiv \Delta G_{\text{P:wt}} - \Delta G_{\text{P:mut}} > 0$. Notably, the sensor benefits from having large $\Delta\Delta G$, as the mutant percentage $c_{\text{mut}}/c_{\text{tot}}$ is multiplied with the exponent of $\Delta\Delta G$, leading to a large signal (Figure 3.1).

The LoD is the mutant percentage at which the signal exceeds a minimum detection threshold S_{min} , required to reliably distinguish a mutDNA-containing mixture from a pure wtDNA sample. In practice, this is often set at three times the standard deviation of the wtDNA reference signal, $S_{\text{min}} = 3 \times \text{Std}(S_0)$ [14]. In Supplementary Information S3, we show that S_{min} is proportional to the multiplicative noise of the measuring system. The connection between the free energy penalty $\Delta\Delta G$ and the $LoD \equiv (c_{\text{mut}}/c_{\text{tot}})_{\text{min}}$ becomes evident from Eq. (3.4):

$$S_{\text{min}} = \ln\left(1 + LoD e^{\Delta\Delta G/RT}\right) \quad (3.6)$$

$$LoD = (e^{S_{\text{min}}} - 1) e^{-\Delta\Delta G/RT} \quad (3.7)$$

$$\approx S_{\text{min}} e^{-\Delta\Delta G/RT} \quad (3.8)$$

Accordingly, the relative detection limit is determined by two key factors: (1) a technical factor, the measurement uncertainty, captured by S_{min} , which reflects the noise characteristics of the detection system, and (2) a fundamental property of the sensor, its selectivity, quantified by $\Delta\Delta G$. Ideally, to detect low fractions of mutDNA in wtDNA background, a sensor therefore has small S_{min} and large $\Delta\Delta G$. A crucial observation in Eq. (3.8) is the fundamentally different way in which the LoD depends on the technical parameter compared to the free-energy penalty. The LoD is linear in S_{min} , such that improvements in measurement uncertainty proportionally reduce the LoD . In contrast, there is an exponential dependence on

$\Delta\Delta G$, such that small increments to $\Delta\Delta G$ lead to exponential improvements in the LoD . It is important to notice that, for the basic sensor, $\Delta\Delta G$ is fully determined by the mismatch penalty of the DNA mutation involved. At first sight, this suggests that $\Delta\Delta G$ is physico-chemically fixed and imposes a hard lower limit on the LoD . Despite this apparent limit, we demonstrate in this study how two complementary strategies improve the LoD by increasing the effective $\Delta\Delta G$.

3.3.2 Strategies to enhance $\Delta\Delta G$

The first strategy increases the binding selectivity through the introduction of structural strain in the sensor design. For this, the solution-based OWL2 hybridization sensor is used, whose improved selectivity is attributed to the structural conformation adopted by the OWL structure shown in Figure 3.2A [15, 16]. The multi-stranded sensor makes use of intermediate adapter strands R and P, and a molecular beacon reporter that forms a stem loop structure in the unbound state [17]. In the presence of target, the adapter strands can hybridize adjacent to each other and bind to the molecular beacon in its elongated form, thus producing a high fluorescent readout at room temperature [15]. In order to do so, the adapter strands fold and bring the 3' and 5' ends in close proximity, where stacking interactions are present. The 'locked ends' of strand P introduce a structural strain in the complex, which is essential for the high selectivity of the OWL sensor [16]. Strain in the OWL sensor destabilizes the mismatched complex with wtDNA more than it does the matched complex with mutDNA, increasing the system's $\Delta\Delta G$ by an additional free-energy contribution $\Delta\Delta G_{\text{Strain}}$.

The second, complementary strategy does not alter the probe-target interaction but instead changes the relative availability of the targets in the solution [6, 18]. This is achieved by the concept of wild-type depletion, by introducing a high concentration of blocker strands that are perfectly complementary to the wtDNA sequence [13]. These blockers selectively hybridize to and deplete the abundant wtDNA from the solution, reducing interference with mutant detection. Of course, the blockers will also cross-hybridize to mutDNA, but the larger affinity of blocker to wtDNA will result in a selective depletion of wtDNA leading to an enrichment in the relative availability of mutant target as illustrated in Figure 3.2B. In

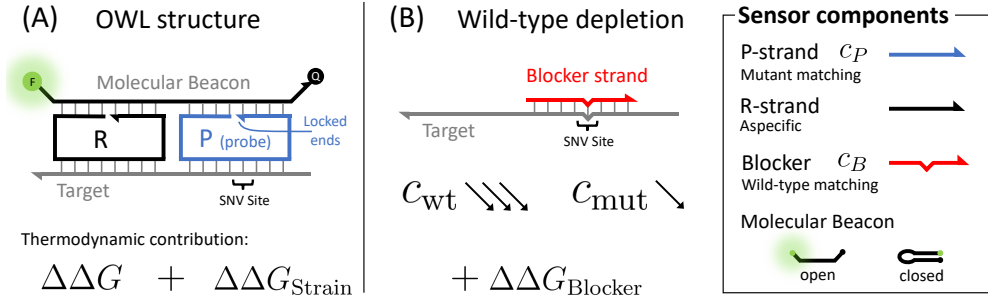


Figure 3.2: Two strategies to enhance mutant detection in DNA mixtures. (A) OWL structure introduces a strain penalty, leading to increased selectivity captured by $\Delta\Delta G$. This multi-stranded hybridization sensor uses intermediate strands R and P that fold and whose 3' and 5'-ends interact to form locked ends and provide strain in the complex, improving the selectivity towards mutant target. Sequences for all strands used in this study are listed in Table S1. (B) Blocker strands lower effective concentrations, and as a result of selective binding to wtDNA lead to an increased relative availability $c_{\text{mut}}/c_{\text{tot}}$.

previous studies, the strategy of wild-type depletion has demonstrated a ~ 10 -fold improvement in LoD and its clinical utility has been validated using low-abundance mutant samples [19, 20]. The amount of depletion is dependent on the blocker concentration, but in the regime of strong depletion, this dependence disappears as maximal enrichment is achieved. In this optimal regime, wild-type depletion manifests as an additional free energy contribution to $\Delta\Delta G$, determined by the free energy difference of the blocker towards the two targets, $\Delta\Delta G_{\text{Blocker}}$.

The two strategies are integrated into a thermodynamic framework that provides a quantitative description of their impact on sensor performance. The mutant detection signal in terms of concentrations and binding affinities is given by Eq. (3.9) (see derivation in Supplementary Information S4).

$$S = \ln \left(1 + \frac{c_{\text{mut}}}{c_{\text{tot}}} \frac{K_{\text{P:mut}}}{K_{\text{P:wt}}} \left[\frac{1 + c_P K_{\text{P:wt}} + c_B K_{\text{B:wt}}}{1 + c_P K_{\text{P:mut}} + c_B K_{\text{B:mut}}} \right] \right) \quad (3.9)$$

The detection signal is determined by $c_{\text{mut}}/c_{\text{tot}}$, and the factor by which it is multiplied. The factor outside the square brackets describes the strain-enhanced binding affinity of the probing strand to matching mutDNA and mismatched wtDNA. As before, the ratio $K_{\text{P:mut}}/K_{\text{P:wt}}$ resolves to the mismatch penalty of the probe $\Delta\Delta G_{\text{Probe}}$ increased with the strain contribution $\Delta\Delta G_{\text{Strain}}$. The factor in the square brackets is concerned with target depletion and introduces probe and blocker concentrations, c_P and c_B , and affinities of the blocker strands to matching wild-type $K_{\text{B:wt}}$ and mis-

matched mutant $K_{B:\text{mut}}$. Since $K_{B:\text{wt}} > K_{B:\text{mut}}$, the presence of blocker results in a selective depletion of wtDNA over mutDNA, leading to higher detection signals. The opposite is true for the mutant matching probes, as $K_{P:\text{mut}} > K_{P:\text{wt}}$, a large probe concentration leads to probe-induced mutant depletion and a reduced detection signal. The signal is maximal in the optimal detection regime at low c_P and high c_B , where it is given by

$$S = \ln \left(1 + \frac{c_{\text{mut}}}{c_{\text{tot}}} \frac{K_{P:\text{mut}}}{K_{P:\text{wt}}} \frac{K_{B:\text{wt}}}{K_{B:\text{mut}}} \right) \quad (3.10)$$

$$= \ln \left(1 + \frac{c_{\text{mut}}}{c_{\text{tot}}} e^{\Delta\Delta G_{\text{eff}}/RT} \right), \quad (3.11)$$

and the system's effective $\Delta\Delta G_{\text{eff}}$ is given by

$$\Delta\Delta G_{\text{eff}} = \Delta\Delta G_{\text{Probe}} + \Delta\Delta G_{\text{Strain}} + \Delta\Delta G_{\text{Blocker}}. \quad (3.12)$$

Optimal performance coincides with the situation where the thermodynamic description simplifies considerably and takes on the form of Eq. (3.4). The signal, for a given $c_{\text{mut}}/c_{\text{tot}}$ and temperature, is determined by a single physical parameter related to three contributions: the standard mismatch penalty $\Delta\Delta G_{\text{Probe}}$, the strain penalty $\Delta\Delta G_{\text{Strain}}$, and the depletion interaction of the blocker $\Delta\Delta G_{\text{Blocker}}$. This elegant thermodynamic framework, captured by Eq. (3.12), provides a basis for understanding how individual components contribute to overall performance and provides insight into the free-energy contributions and relative importance of each component. In what follows, we first describe the sensor's performance in the optimal detection regime and construct a dose-response curve to determine $\Delta\Delta G_{\text{eff}}$. Following this, either the strain component or the blocker strands are removed to separately quantify the individual contributions of the mismatch penalty, the structural strain and the blocker to the sensor's selectivity.

3.3.3 Optimal detection

To evaluate the selectivity enhancing strategies under realistic conditions, we applied the thermodynamic framework to experimental measurements using synthetic EGFR DNA. Specifically, we targeted the clinically relevant p.T790M mutation, which arises from a C→T transition in the EGFR wild-type sequence and is known to confer resistance to first- and second-generation tyrosine kinase inhibitors [21, 11].

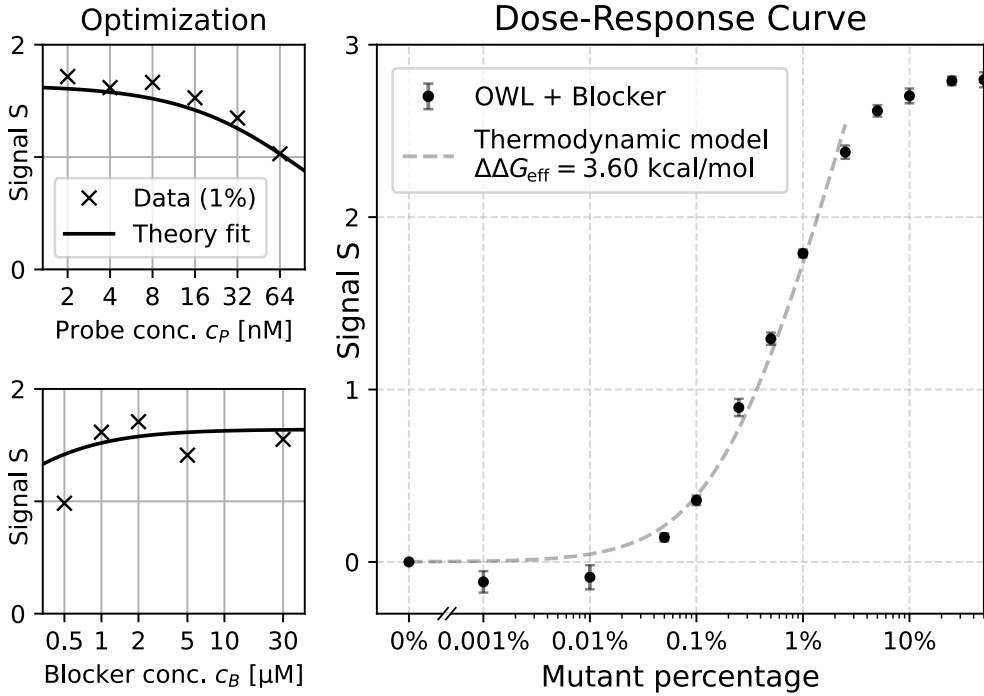


Figure 3.3: Left: probe and blocker concentration optimization. Signal S becomes maximal and enters the optimal detection regime when $c_P \leq 8$ nM and $c_B \geq 1000$ nM, where a plateau is reached and the concentration dependence disappears. Right: the dose-response curve in the optimal detection regime determines $\Delta\Delta G_{\text{eff}}$, reflecting the contribution of probe selectivity, strain contribution and the blocker interaction.

To maximize the mutant detection signal of the OWL sensor with strain and in the presence of blocker strands, we first varied probe concentrations to reach the optimal detection regime. We used a fixed target concentration of 500 nM that contains 1% mutant in excess of wild-type target. Outside of the optimal detection regime, the mutant detection signal Eq. (3.9) is dependent on probe (P-strand) concentration c_P and blocker concentration c_B . Various combinations of c_P and c_B were tested to determine the signal of this 1% sample. As predicted by the theory, small c_P in combination with large c_B resulted in the highest detection signal. We experimentally determined that optimal concentrations of 8 nM of probe and 1 μ M of blocker resulted in accurate readout with high detection signals. To explicitly verify optimality, either c_P or c_B was fixed while the other concentration was varied (Figure 3.3). The blocker optimization indeed shows that beyond 1 μ M (twice the amount of target) the concentration dependence disappears,

indicative of the optimal detection regime. Using higher c_B will further deplete targets from the solution and thus lower the fluorescent readout, but the relative availability of unbound mutant targets no longer increases. Similarly, c_P optimization shows that the signal S is maximal and constant up to 8 nM of probe and only begins to decrease above this value. This is the undesired probe-induced mutant depletion: a large amount of mutant matching probes deplete mutDNA over wtDNA and lower the signal. This is absent at concentrations 8 nM and below, where the optimal detection regime is reached.

The dose-response curve shown in Figure 3.3 shows the signal for varying c_{mut}/c_{tot} and provides data to fit the selectivity as captured by $\Delta\Delta G_{eff}$. The curve follows a sigmoidal shape and flattens at high mutant fractions, where extensive binding leads to probe saturation. For these data points the fluorescent readout is greater than 50% of the maximal readout, resulting in a reduction to the sensor's sensitivity. Probe saturation may safely be neglected at lower mutant fractions, where the signal is fitted by the thermodynamic model of Eq. 3.10. The total $\Delta\Delta G_{eff}$ resulting from the penalty for mismatch, strain and blocker is found to be 3.60 ± 0.04 kcal/mol. From this, the system's LoD can be determined using Eq. (3.7). Using 20 repeats, the standard deviation of the wild-type reference signal was quantified as $Std(S_0) = 0.08$, leading to a minimum detection threshold $S_{min} = 0.24$. This translates into $LoD = 0.058\%$ mutDNA in the presence of excess wtDNA.

3.3.4 Contribution of Strain

In order to quantify the contribution of structural strain in the OWL structure, we compared the sensor performance with and without strain while keeping all other conditions identical. The strain, caused by stacking interactions at the locked ends (Figure 3.2A), is relieved by the introduction of a polyethylene glycol (PEG) linker into the P-strand of the OWL structure [15]. As was shown earlier, the flexible PEG linker removes the structural strain of the OWL sensor complex [15]. This was followed by the comparison of the OWL sensors differing by the P stands as shown in Figure 3.4. Both sensors were subject to excess blocker strands to ensure optimal dose-response curves. Since the effect of the blocker is equivalent in

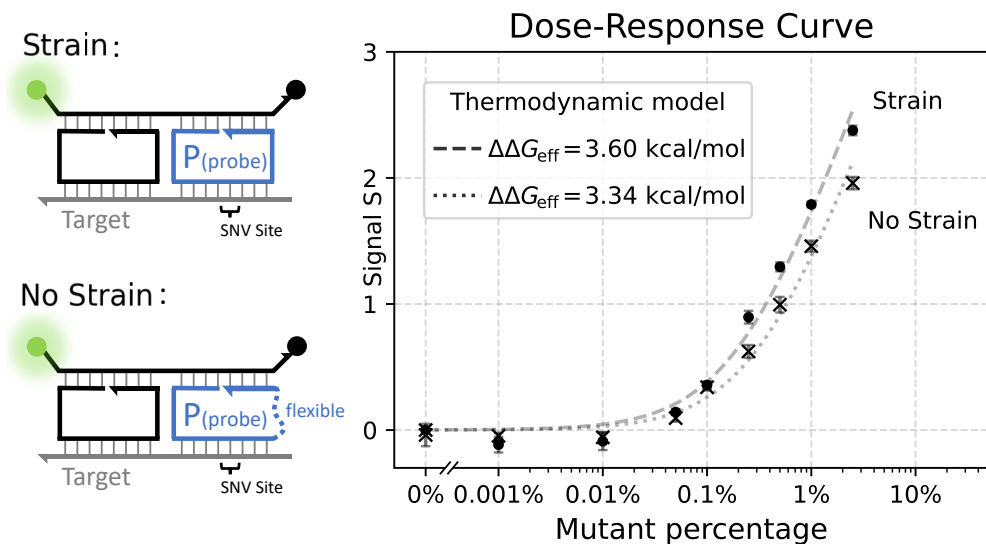


Figure 3.4: Contribution of strain to the sensor selectivity. Both dose-response curves are subject to a large amount of blocker strands. For the case without strain, the P-strand is adapted to have a flexible PEG linker that removes the strain in the sensor structure. The increased performance and the contribution of strain $\Delta\Delta G_{\text{Strain}}$ is quantified by the difference in fitted $\Delta\Delta G_{\text{eff}}$ values and leads to a 1.5-fold improvement to selectivity over the sensor without strain.

both sensors, the observed differences reflect the strain effect in the OWL structure. Introducing strain produced higher signals and improved selectivity compared to the sensor with a flexible P-strand and no strain (Figure 3.4). Fitting the dose-response curves revealed that strain increases $\Delta\Delta G_{\text{eff}}$ from 3.34 ± 0.04 kcal/mol to 3.60 ± 0.04 kcal/mol. The corresponding difference represents the strain-based free-energy penalty $\Delta\Delta G_{\text{strain}}$ determined to be 0.26 ± 0.08 kcal/mol. For comparison, this contribution equates to approximately 20% of a single mismatch penalty, which translates into a $\exp(0.26/RT) \approx 1.5$ -fold improvement in selectivity.

3.3.5 Contribution of Blocker

To determine the contribution of blocker strands to the overall sensor performance, we similarly performed experiments with and without the use of wild-type-matching blocker. Strain is maintained in both conditions to provide high signals, and because it is present equally in both cases, any

differences are attributed to the blocker interaction. Without blocker, the system is once more optimized for probe concentration c_p . The generic form of Eq. (3.9) reduces to

$$S = \ln \left(1 + \frac{c_{\text{mut}}}{c_{\text{tot}}} \frac{K_{P:\text{mut}}}{K_{P:\text{wt}}} \left[\frac{1 + c_p K_{P:\text{wt}}}{1 + c_p K_{P:\text{mut}}} \right] \right). \quad (3.13)$$

Analogous to the selective depletion of wtDNA by wild-type-matching blocker strands, the mutant-matching probes will selectively deplete mutDNA. As $K_{P:\text{mut}} > K_{P:\text{wt}}$, the signal decreases with c_p as a result of probe-induced mutant depletion. This undesired effect vanishes at sufficiently low probe concentrations, where the signal reaches a maximum. In this optimal regime, the signal is characterized by an effective $\Delta\Delta G_{\text{eff}}$ representing the standard mismatch penalty increased by the strain contribution. The left side of Figure 3.5 shows experimental measurements of the mutant detection signal for varying probe concentration c_p . We used a 5% mutant sample at a reduced total target concentration of 20 nM to avoid probe saturation since without blocker the probes are exposed to all targets. Experimentally, the optimal detection regime was not reached, resulting in the underestimation of the fitted $\Delta\Delta G_{\text{eff}}$ values. While ideally the probe concentration should be lower, the accessible range is limited to 1 nM, below which fluorescent intensities become indistinguishable from background noise [22]. At these low probe concentrations, the minimal detection threshold S_{min} increases, leading to worse LoD following Eq. (3.7). Supplementary Information S3 contains a detailed description of the signal uncertainty in terms of probe concentration. A probe concentration of 2 nM was selected as an optimal compromise between signal quality and measurement uncertainty, and was used to construct the dose-response curve.

Figure 3.5 shows a comparison of the dose-response curves with and without the blocker. The use of blocker strands leads to significantly higher signals for all mutant fractions. This is particularly visible in the range of 0.01-1%, where the sensor without blocker produces near-zero signals and the presence of blocker leads to clearly distinguishable, positive signals. This shows the significant contribution of the blocker interaction to the LoD . Fitting the experimental dose-response curves to the thermodynamic model quantifies the importance of blockers. Parameter $\Delta\Delta G_{\text{eff}}$ was found

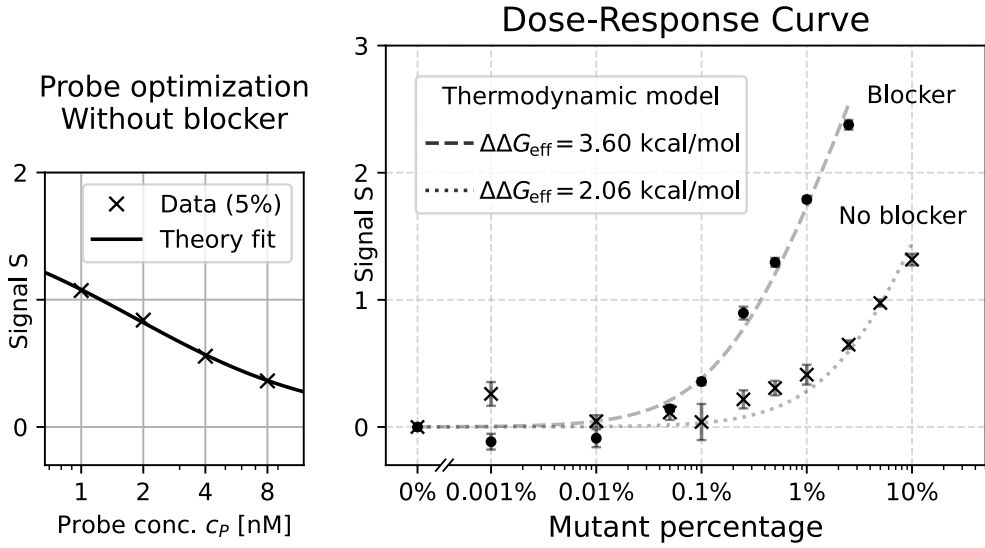


Figure 3.5: Contribution of blocker to the sensor selectivity. Left: The signal without blocker does not reach the optimal detection regime when lowering c_P . Right: at $c_P = 2$ nM, the dose-response curve for the system without blocker is compared to the system with blocker strands. Lower signals are observed without the selective depletion of the blocker, whose contribution $\Delta\Delta G_{\text{Blocker}}$ is quantified by comparing the fitted $\Delta\Delta G_{\text{eff}}$ parameters, and leads to a 14-fold improvement to selectivity.

to change from 3.60 ± 0.04 kcal/mol to 2.06 ± 0.07 kcal/mol when the blocker interaction is removed. The selective wild-type depletion, induced by the blockers, therefore leads to an increase of $\Delta\Delta G_{\text{Blocker}} = 1.54 \pm 0.11$ kcal/mol. As expected, this value is approximately a single mismatch penalty, since it corresponds to the free-energy difference of the blocker binding to its two targets. The blocker interaction results in a $\exp(1.54/RT) \approx 14$ -fold improvement in selectivity compared to the system without blockers. These results are consistent with prior studies, where surface-bound blocker strands led to a 12.5-fold improvement for the same target sequences [19, 20].

3.3.6 Blocker strands counteract probe-induced mutant depletion

Beyond their direct role in enhancing selectivity, blocker strands provide an important secondary benefit: they counteract the undesired depletion of mutant targets by the mutant probes. From the previous results, we observe that the probe concentration c_p behaved differently depending on the presence of blockers. With blockers, concentrations as high as 8 nM could be used while remaining in the optimal detection regime (Figure 3.3). In contrast, in the absence of blockers, even the lower c_p of 1 nM results in an impacted signal due to undesired depletion by mutant probes and produces larger statistical errors (Figure 3.5). Equation (3.9) provides insight into this additional benefit. Probe concentration c_p starts to impede the signal when $c_p K_{P:mut} \gg 1 + c_B K_{B:mut}$, leading to an upper bound on the probe concentration of $c_p \approx c_B K_{B:mut} / K_{P:mut}$ to be in the optimal detection regime. The presence of blocker strands effectively increases the limit by a factor $c_B K_{B:mut}$. Consequently, blocker strands allow the use of higher probe concentrations, resulting in improved signal-to-noise ratios, which are especially important for samples with low mutant fractions.

3.3.7 Standard Mismatch Penalty

Lastly, the discrimination of the sensor was assessed in the absence of both blocker strands and structural strain. In this configuration, the P-strand was again adapted to have a flexible linker and analyzed under the no-blocker conditions. This setup isolates the binding free-energy difference of the probe, $\Delta\Delta G_{Probe}$, representing the standard mismatch penalty. Supplementary Information S5 contains the corresponding dose-response curve, and fitting the data quantifies $\Delta\Delta G_{Probe} = 1.40 \pm 0.08$ kcal/mol. As encountered before without blocker, the system operates close to, but not fully within, the optimal detection regime, leading to an underestimation of the mismatch penalty.

This measurement provides an independent means of assessing the strain contribution by comparing dose-response curves obtained without the use of blocker strands. A larger free-energy difference of 0.66 ± 0.15 kcal/mol is observed in this comparison, resulting from the fact that the systems are operating outside the optimal detection regime. Under such condi-

tions, the difference between the fitted $\Delta\Delta G_{\text{eff}}$ values no longer isolates the strain contribution, leading to an inaccurate estimate. The addition of blocker strands therefore allowed for reliable quantification of the strain-based contribution, while simultaneously contributing to the sensor performance. Although strain accounts for only $\sim 20\%$ of a single-mismatch penalty under optimal conditions, the larger apparent penalty observed here indicates that it can become a major contributor in situations where blockers cannot be used, for example at low sample concentrations [20].

3.4 Discussion

Motivated by the exponential dependence of the relative detection limit LoD on the binding free energy difference $\Delta\Delta G$, in this study, we employed two strategies to improve the detection of low-abundant mutant DNA in wild-type background: structural strain in the probing complex, and wild-type depletion by blocker strands. Our thermodynamic description allows the quantification of $\Delta\Delta G$ from dose-response curves. The individual effects of blocker and strain were subsequently quantified from experimental measurements where each effect was included or excluded, as summarized in Table 3.1. The structural strain led to a 1.5-fold improvement to sensor performance, and the addition of blocker strands showed a 14-fold improvement, leading to a remarkable ~ 21 -fold increase in selectivity compared to the sensor without strain and blocker strands. The combination of these strategies resulted in a remarkably low LoD of 0.058%, purely through thermodynamic optimization of selectivity governed by $\Delta\Delta G$. This must be emphasized, because many reported approaches reach an LoD in the range of 0.01-1%, but only at the cost of increased assay complexity or enzymatic processing [18, 23, 24]. Hence, our sensor is highly competitive with state-of-the-art methods.

Beyond improving selectivity, the strain-based OWL sensor uses adapter strands and a universal molecular beacon reporter, making it operable at room temperature and in a broad temperature range of 5-38°C [15, 25]. Thus, while this study focused on EGFR as a case example for detection of low-abundance single-nucleotide variants, the sensor design is readily adaptable to alternative sequences in a cost-effective manner. Additionally, it permits the use of shorter strands, that exhibit faster binding

Table 3.1: Free-energy contributions in the strain-based OWL2 sensor with wild-type depletion. Contributions are presented relative to a typical mismatch (MM) penalty or as the fold change resulting from the improvement. The optimal sensor’s limit of detection (*LoD*) is the relative amount of mutDNA detectable in a background of wtDNA.

	$\Delta\Delta G_{\text{eff}} =$	$\Delta\Delta G_{\text{Probe}}$	$+ \Delta\Delta G_{\text{Strain}}$	$+ \Delta\Delta G_{\text{Blocker}}$
Value	3.60 ± 0.04	1.40 ± 0.08	0.26 ± 0.08	1.54 ± 0.11
(kcal/mol)				
Contribution		$\sim 1 \text{ MM}$	$\sim 20\% \text{ of MM}$	$\sim 1 \text{ MM}$
Fold change	$\downarrow$		1.5-fold	14-fold
	<i>LoD</i> 0.058%			

kinetics suitable for a homogeneous biosensor for point-of-care applications. The OWL2 sensor also introduced unfolding arms (Supplementary Information S1), which allows for selective detection even for long target sequences with stable secondary structures [16].

Wild-type blocker strands provided an additional benefit by allowing the use of higher mutant probe concentrations without compromising performance. This led to more robust dose-response curves, which allowed an accurate study of the strain-based interaction. From a broader perspective, this dual role of blockers, improving selectivity through wild-type depletion and enabling detection at higher probe concentrations, illustrates their versatility in the design of hybridization-based biosensors.

These findings suggest several promising directions for future research. The integration of strain and blocker strands into existing hybridization-based biosensors could offer immediate increases in performance without requiring major redesigns, as the concepts are easily transferable to other detection platforms and could facilitate parallel detection of multiple variants of interest. Further investigation into the mechanisms and sequence dependence of strain will help understand design principles and improve modeling. Importantly, the effectiveness of strain relies on the presence of an asymmetric destabilization that favors the mutant, thereby increasing $\Delta\Delta G$. This opens an interesting opportunity to explore diverse mechanisms by which mechanical stress introduces selective destabilization, leading to improved detection. The selective destabilization may originate from torsional stress in the OWL structure, as used here, but could also

be introduced by allosteric effects [26, 27], local perturbations [28, 29], or the formation of secondary structures like stem-loops [30]. In ongoing work, we are exploring the use of a second intentional mismatch [5] to introduce additional strain, with the aim of extending this approach to new sensor designs. It is also important to note that the blocker contribution, $\Delta\Delta G_{\text{Blocker}}$, analogous to $\Delta\Delta G_{\text{Probe}}$, is determined by the difference in binding free energy of the blocker towards its two possible targets. In the same way that the selectivity of the mutant-matching probe was enhanced through the introduction of structural strain, one could envision designing blocker strands that incorporate a strain component. Such an approach would increase the effectiveness of wild-type depletion and further enrich the mutant content of a sample. This would improve overall selectivity, resulting into an even larger $\Delta\Delta G_{\text{eff}}$ and improved LoD .

In summary, we addressed the intrinsic selectivity limitations of conventional hybridization assays by employing two synergistic strategies: structural strain within the DNA probe and wild-type blocker strands, achieving a limit of detection of 0.058% mutDNA in a majority wtDNA background. Importantly, we quantified the individual contribution of strain and blocker, providing insights for future sensor design. Together, these results establish both a theoretical and experimental validation for the development of more selective and robust nucleic acid biosensors. Given that DNA hybridization forms the basis of numerous biotechnological applications, the principles established here extend well beyond this specific implementation, with broad relevance to nucleic acid sensing and molecular diagnostics.

Author contributions: J.Ho, D.M.K., Y.S., B.L.M. and R.V.H. were responsible for conceptualization and methodology. Y.S., B.L.M., J.Ha., and A.P. conducted the investigation. Y.S. and B.L.M. performed formal analysis, data curation, and visualization. D.M.K. and J.Ho. were responsible for supervision and project administration. Y.S., B.L.M., D.M.K., and J.Ho. acquired funding. Y.S. and B.L.M. wrote the original draft. All authors reviewed and edited the manuscript.

Funding: This work was supported by Special Research Fund: [BOF220WB11], [BOF23BL05] and [BOF25GP02]; Research Foundation Flanders [V452324N]; European Erasmus+ and National Science Foundation through the CCF: Software and Hardware Foundations [SHF-1907824].

Supplementary Information

S1 OWL2 Structure with Unfolding Arms

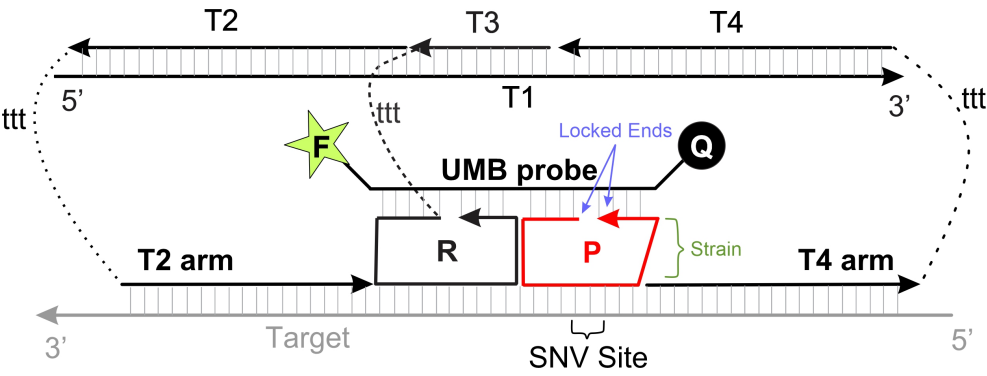


Figure S1: Diagram of OWL2 sensor in complex with complementary DNA. Sequences T1, T2, T3 and T4 function as scaffolding for active components of the unfolding arms (T2 arm and T4 arm) and bring adapter strand R in binding range to the target sequence. Binding with mutant specific P strand subsequently facilitates the opening of universal molecular beacon (UMB) probe and allows for a fluorescent readout. The sequences for all strands are listed in Table S1.

S1.1 Oligonucleotide Sequences

Table S1: Oligonucleotide sequences used in experiments for EGFR p.T790M detection using the OWL2 Sensor. Target and probe sequences used in the hybridization experiments. The 80 nt long targets *EGFR* WT and *EGFR* Mut differ at the highlighted single nucleotide C→T. Binding regions to R strand, (flexible) P strand and WT Blocker are colored, and binding regions of the UMB are underlined. Highlighted yellow are the nucleotides at the mutation site and the nucleotide of P strand complementary to *EGFR* Mut at the mutation site.

Oligo	Binding region sequence
<i>EGFR</i> WT	5'- ATC TGC CTC ACC TCC ACC GTG CAG CTCA TCAC GCAG CTCAT GCCCT TCGG CTG CCT CCT GGA CTA TGT CCG GGA ACA CAA -3'
<i>EGFR</i> Mut	5'- ATC TGC CTC ACC TCC ACC GTG CAG CTCA TCAT GCAG CTCAT GCCCT TCGG CTG CCT CCT GGA CTA TGT CCG GGA ACA CAA -3'
UMB-15	5'- FAM-CG C GTTC CCATA <u>CAAC CAATC</u> GCG-BQ1 -3'
T1	5'- GTA TCA GTC ATT ACC AGT AGT CGGAC CTAGG CTCTCGGT CTA G CCAC TTAAC -3'
T2	5'- ACT ACT GGT AAT GAC TGA TAC ttt ACA TAG TCC AGG AGG CAG CCGA -3'
T3 (R)	5'- CCTAG GTCCG ttt <u>GAACG AGGGC ATGAG TATGG</u> -3'
T4	5'- TGA GCT GCA CGG TGG AGG TGA GGC ttt GTT AA GTGG CTAG ACCGAGAG -3'
P	5'- <u>GTTG CTGC ATGA GATTG</u> -3'
Flexible P	5'- <u>GTTG CTGC ATGA</u> /iSp9/ <u>GATTG</u> -3'
WT Blocker	5'- CTGC GTGA -3'

S2 Optimal Discrimination for the Basic Sensor

We consider a probe-target system with a simple reaction model given below.



with association and dissociation rate constants k^+ and k^- , which determine association constant $K \equiv k^+/k^-$. The dynamics of this system is governed by the differential equation

$$\frac{d[PT]}{dt} = k^+ \underbrace{([P]_0 - [PT])}_{\text{probes saturation}} \underbrace{([T]_0 - [PT])}_{\text{target depletion}} - k^- [PT] \quad (S2)$$

where one identifies the origin of probe saturation, i.e. probes running out during the binding process, and target depletion, where target runs out. At equilibrium, $d[PT]_{\text{eq}}/dt = 0$ and one gets an expression for the equilibrium concentration of the probe-target complex.

$$[PT]_{\text{eq}} = K ([P]_0 - [PT]_{\text{eq}}) ([T]_0 - [PT]_{\text{eq}}) \quad (S3)$$

This expression allows the comparison of the probe subject to mutant target versus wild-type target under the same conditions, we get

$$\frac{[PT_{\text{mut}}]_{\text{eq}}}{[PT_{\text{wt}}]_{\text{eq}}} = \frac{K_{\text{mut}}}{K_{\text{wt}}} \underbrace{\frac{([P]_0 - [PT_{\text{mut}}]_{\text{eq}})}{([P]_0 - [PT_{\text{wt}}]_{\text{eq}})}}_{<1} \underbrace{\frac{([T_{\text{mut}}]_0 - [PT_{\text{mut}}]_{\text{eq}})}{([T_{\text{wt}}]_0 - [PT_{\text{wt}}]_{\text{eq}})}}_{<1} \quad (S4)$$

Given that $[PT_{\text{mut}}]_{\text{eq}} > [PT_{\text{wt}}]_{\text{eq}}$, both probe saturation and target depletion effects negatively impact discrimination between mutant and wild-type target. Finally, maximal discrimination is achieved when these effects can be neglected where $[PT]_{\text{eq}} = K[P]_0[T]_0$ and $[PT_{\text{mut}}]_{\text{eq}}/[PT_{\text{wt}}]_{\text{eq}} = K_{\text{mut}}/K_{\text{wt}}$ which resolves to the free energy difference between the two bindings, $\Delta\Delta G$.

S3 Signal Uncertainty $Std(S_0)$

Suppose $\tilde{F}$ is the fluorescent measurement of a sample of interest (which might contain some amount of mutant target), $\tilde{F}_0$ the reference wild-type only measurement and $\tilde{B}$ a background measurement without target. The (measured) signal is defined as

$$\tilde{S} = \log\left(\frac{\tilde{F} - \tilde{B}}{\tilde{F}_0 - \tilde{B}}\right) = \log\left(\frac{\tilde{X}}{\tilde{X}_0}\right). \quad (\text{S5})$$

The measurement results deviate from their true values F and B , which we assume are drawn from a normal distribution with multiplicative noise σ_m (the uncertainty is e.g. 5% of the fluorescence). The background subtracted intensities are then

$$\tilde{X} = \tilde{F} - \tilde{B} \quad (\text{S6})$$

$$= F(1 + \varepsilon_1) - B(1 + \varepsilon_2) \quad (\text{S7})$$

$$= X + F\varepsilon_1 - B\varepsilon_2 \quad (\text{S8})$$

$$\tilde{X} = X + \delta \quad (\text{S9})$$

with $\varepsilon_1, \varepsilon_2 \sim \mathcal{N}(0, \sigma_m^2)$ and δ the resulting error drawn from the new distribution with additive noise σ_X . The variance of the background subtracted measurement is subsequently given by

$$\text{Var}(\tilde{X}) = F^2 \text{Var}(\varepsilon_1) + B^2 \text{Var}(\varepsilon_2) \quad (\text{S10})$$

or equivalently

$$\sigma_X = \sqrt{F^2 \sigma_m^2 + B^2 \sigma_m^2} \quad (\text{S11})$$

$$= \sigma_m \sqrt{F^2 + B^2} \quad (\text{S12})$$

For small errors, a linear approximation is used:

$$\log \tilde{X} \approx \log X(1 + \delta/X) \quad (\text{S13})$$

$$\approx \log X + \frac{\delta}{X} \quad (\text{S14})$$

Which results into the signal

$$\tilde{S} = \log \tilde{X} - \log \tilde{X}_0 \quad (\text{S15})$$

$$= \log X + \frac{\delta}{X} - \log X_0 - \frac{\delta_0}{X_0} \quad (\text{S16})$$

$$= S + \frac{\delta}{X} - \frac{\delta_0}{X_0} \quad (\text{S17})$$

where $\text{Var}(\delta) = \sigma_X^2$ and $\text{Var}(\delta_0) = \sigma_{X_0}^2$.

Combined, this leads to an error on the measurement of the signal

$$\sigma_S^2 = \frac{\sigma_X^2}{X^2} + \frac{\sigma_{X_0}^2}{X_0^2} = \sigma_m^2 \left(\frac{F^2 + B^2}{(F - B)^2} + \frac{F_0^2 + B^2}{(F_0 - B)^2} \right) \quad (\text{S18})$$

that is dependent on the single multiplicative error σ_m but also on the (true) intensities and background. The contribution from the sample measurement is present in the first term, and the reference measurement in the second.

In particular, one is interested in the standard deviation of the reference signal, $\text{Std}(S_0) \equiv \sigma_{S_0}$, obtained by setting $F = F_0$, resulting into a closed expression for the signal uncertainty.

$$\text{Std}(S_0) = \sqrt{2} \sigma_m \sqrt{\frac{F_0^2 + B^2}{(F_0 - B)^2}} \quad (\text{S19})$$

The lowest value of $\text{Std}(S_0)$ is obtained when $F_0 \gg B$, where the background can be neglected and $\text{Std}(S_0) = \sqrt{2} \sigma_m$. The relative detection limit LoD is dependent on the minimal detection threshold $S_{\min} = 3 \times \text{Std}(S_0)$, such that

$$LoD = (e^{S_{\min}} - 1) e^{-\Delta\Delta G/RT} \quad (\text{S20})$$

$$\approx S_{\min} e^{-\Delta\Delta G/RT} \quad (\text{if } S_{\min} \ll 1) \quad (\text{S21})$$

$$\approx 3\sqrt{2} \sigma_m e^{-\Delta\Delta G/RT} \quad (\text{S22})$$

The LoD is therefore linear in the multiplicative error σ_m of fluorescent readout.

On the other hand, $\text{Std}(S_0)$ increases when fluorescence F_0 becomes close to background levels, which increases $S_{\min}$ and leads to a worse LoD . This

is important in the optimization for probe concentration c_P . The optimal detection regime (see main text) is expected at low c_P , but this lowers the fluorescence F_0 , leading to higher $Std(S_0)$. To quantify the probe concentration dependence, realize that the background subtracted intensities are linear in probe concentration, $F_0 - B = m c_P$, such that

$$Std(S_0) = \sqrt{2}\sigma_m \sqrt{\frac{(B + c_P m)^2 + B^2}{(c_P m)^2}} \quad (S23)$$

For given experimental parameters σ_m , B and m , this expression allows one to determine the lowest probe concentration that keeps the signal uncertainty $Std(S_0)$ below a defined value. Experimentally, the above parameters are determined to be (see Figure S2)

$$\sigma_m \approx 0.02 \quad B \approx 500 \text{ a.u.} \quad m \approx 80 \text{ a.u./nM} \quad (S24)$$

Setting a threshold of $Std(S_0) = 0.2$ results in probe concentrations c_P no lower than 2nM.

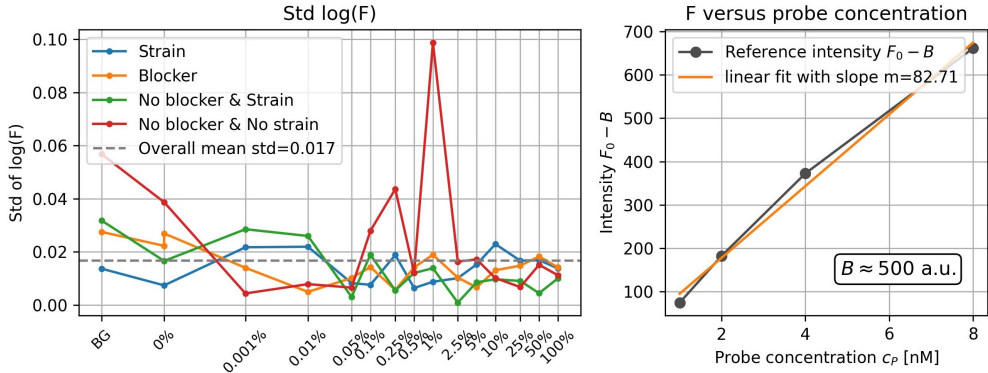


Figure S2: Top: Standard deviation on the logarithm of fluorescent measurements. Constant standard deviation indicates multiplicative error across the measurement range. Bottom: Probe concentration dependence of reference intensity. The background subtracted fluorescence $F_0 - B$ indicates a linear relationship with probe concentration c_P and results in a fit of coefficient m .

S4 Derivation of the Signal under Depletion Conditions

We consider the regime where probe saturation does not occur, i.e. in the dose-response curve for low mutant percentages. In this case, the mutant and wild-type target can be considered independent. The target can bind to the mutant matching probe P , and the wild-type matching blocker B , leading to the system below.



with their respective association and dissociation rate constants, which determine association constants $K_P \equiv k_{P+}/k_{P-}$ and $K_B \equiv k_{B+}/k_{B-}$, that are both different dependent on the target type. Without probe saturation but with target depletion the concentrations will be given by

$$[P] \approx [P]_0 = c_P, \quad [B] \approx [B]_0 = c_B, \quad \text{and} \quad [T] = c - [PT] - [BT]. \quad (\text{S27})$$

Therefore at equilibrium

$$\begin{cases} [PT] = K_P [P] [T] \\ [BT] = K_B [B] [T] \end{cases} \quad (\text{S28})$$

$$\begin{cases} [PT] = c_P K_P (c - [PT] - [BT]) \\ [BT] = c_B K_B (c - [PT] - [BT]) \end{cases} \quad (\text{S29})$$

Solving for $[PT]$ results into

$$[PT] = \frac{c_P c K_P}{1 + c_P K_P + c_B K_B}. \quad (\text{S30})$$

In a mixture, the signal is consequently described by

$$S(\%) = \ln \left(\frac{[PT]_{\text{mix}}}{[PT]_{\text{pure}}} \right) \quad (\text{S31})$$

$$= \ln \left(\frac{[PT]_{\text{wt}} + [PT]_{\text{mut}}}{[PT]_{\text{pure}}} \right) \quad (\text{S32})$$

$$= \ln \left(\frac{\frac{c_p c_{\text{wt}} K_{\text{wt}}}{1 + c_p K_{\text{wt}} + c_B K_{B:\text{wt}}} + \frac{c_p c_{\text{mut}} K_{\text{mut}}}{1 + c_p K_{\text{mut}} + c_B K_{B:\text{mut}}}}{\frac{c_p c_{\text{tot}} K_{\text{wt}}}{1 + c_p K_{\text{wt}} + c_B K_{B:\text{wt}}}} \right) \quad (\text{S33})$$

$$= \ln \left(1 + \frac{c_{\text{mut}}}{c_{\text{tot}}} \frac{K_{\text{mut}}}{K_{\text{wt}}} \left[\frac{1 + c_p K_{\text{wt}} + c_B K_{B:\text{wt}}}{1 + c_p K_{\text{mut}} + c_B K_{B:\text{mut}}} \right] \right) \quad (\text{S34})$$

$$(\text{S35})$$

since $c_{\text{wt}} \approx c_{\text{tot}}$ for low mutant percentages, resulting in the expression as presented in the main article.

S5 Full Dose-Response Curves

Figure S3 shows the dose-response curves for the three systems considered in the main article up to 100% mutant (a pure mutant sample). The theoretical fit considers only the region where an exponential increase is observed. For these low-abundant mutant mixtures, the theory from the main article is valid since probe saturation can be neglected. Beyond these percentages, saturation effects do take place, which lowers the signal response of the sensor. Because the systems with blocker use a larger total target concentration, the percentage at which saturation becomes important is reached earlier since saturation is largely dependent on the total concentration of mutant target.

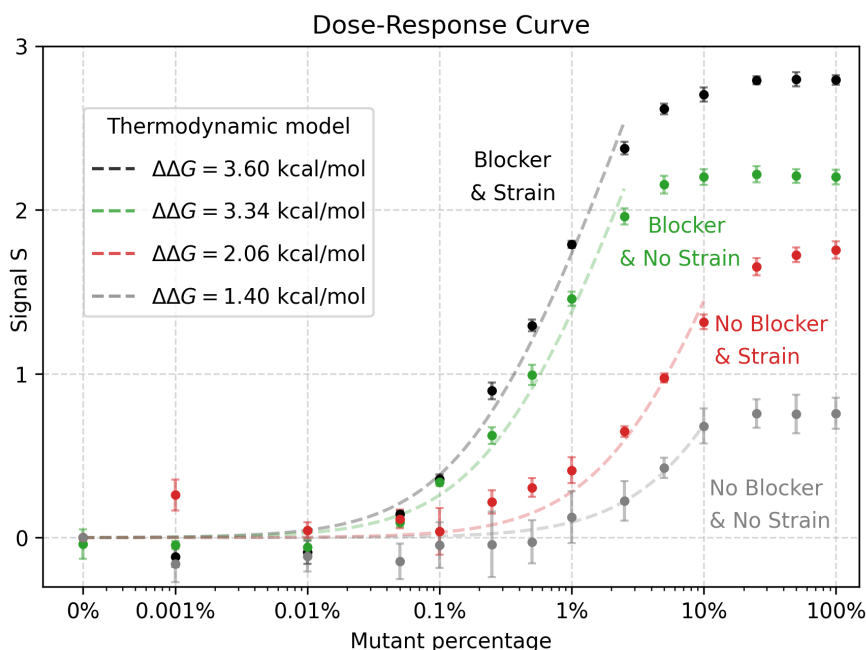


Figure S3: Dose-response curves in the full range of pure wild-type to pure mutant samples for the systems with and without blocker, and with and without strain. Error bars represent the standard deviation over three independent repeats.

References

- [1] W. Xie, S. Suryaprakash, C. Wu, A. Rodriguez, and S. Fraterman, "Trends in the use of liquid biopsy in oncology," *Nature Reviews Drug Discovery*, vol. 22, no. 8, pp. 612–613, 2023.
- [2] J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, and N. Rosenfeld, "Liquid biopsies come of age: towards implementation of circulating tumour DNA," *Nature Reviews Cancer*, vol. 17, no. 4, pp. 223–238, 2017.
- [3] K. Knez, D. Spasic, K. P. F. Janssen, and J. Lammertyn, "Emerging technologies for hybridization based single nucleotide polymorphism detection," *Analyst*, vol. 139, no. 2, pp. 353–370, 2013.
- [4] J. Stadler, J. Eder, B. Pratscher, S. Brandt, D. Schneller, R. Müllegger, C. Vogl, F. Trautinger, G. Brem, and J. P. Burgstaller, "SNPase-ARMS qPCR: Ultrasensitive Mutation-Based Detection of Cell-Free Tumor DNA in Melanoma Patients," *PLOS ONE*, vol. 10, no. 11, p. e0142273, 2015.
- [5] S. Lefever, A. Rihani, J. Van Der Meulen, F. Pattyn, T. Van Maerken, J. Van Dorpe, J. Hellemans, and J. Vandesompele, "Cost-effective and robust genotyping using double-mismatch allele-specific quantitative PCR," *Scientific Reports*, vol. 9, no. 1, p. 2150, 2019.
- [6] P. Song, L. R. Wu, Y. H. Yan, J. X. Zhang, T. Chu, L. N. Kwong, A. A. Patel, and D. Y. Zhang, "Limitations and opportunities of technologies for the analysis of cell-free DNA in cancer diagnostics," *Nature Biomedical Engineering*, vol. 6, no. 3, pp. 232–245, 2022.
- [7] V. V. Demidov and M. D. Frank-Kamenetskii, "Two sides of the coin:

- affinity and specificity of nucleic acid interactions," *Trends in Biochemical Sciences*, vol. 29, no. 2, pp. 62–71, 2004.
- [8] D. Y. Zhang, S. X. Chen, and P. Yin, "Optimizing the specificity of nucleic acid hybridization," *Nature Chemistry*, vol. 4, no. 3, pp. 208–214, 2012.
- [9] J. Lu, Y. Feng, K. Guo, L. Sun, S. Ruan, and K. Zhang, "Prognostic value of preoperative circulating tumor DNA in non-small cell lung cancer: a systematic review and meta-analysis," *Journal of Cancer Research and Clinical Oncology*, vol. 150, no. 1, p. 25, 2024.
- [10] M. A. Gouda *et al.*, "Longitudinal Monitoring of Circulating Tumor DNA to Predict Treatment Outcomes in Advanced Cancers," *JCO Precision Oncology*, vol. 6, no. 6, p. e2100512, 2022.
- [11] S. Wang, S. Cang, and D. Liu, "Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer," *Journal of Hematology & Oncology*, vol. 9, no. 1, p. 34, 2016.
- [12] H. Willems, A. Jacobs, W. W. Hadiwikarta, T. Venken, D. Valkenburg, N. V. Roy, J. Vandesompele, and J. Hooyberghs, "Thermodynamic framework to assess low abundance DNA mutation detection by hybridization," *PLOS ONE*, vol. 12, no. 5, p. e0177384, 2017.
- [13] S. K. Nomidis, M. Szymonik, T. Venken, E. Carlon, and J. Hooyberghs, "Enhancing the Performance of DNA Surface-Hybridization Biosensors through Target Depletion," *Langmuir*, vol. 35, no. 37, pp. 12 276–12 283, 2019.
- [14] D. MacDougall, W. B. Crummett, and . Et Al., "Guidelines for data acquisition and data quality evaluation in environmental chemistry," *Analytical Chemistry*, vol. 52, no. 14, pp. 2242–2249, 1980.
- [15] R. J. Karadeema, M. Stancescu, T. P. Steidl, S. C. Bertot, and D. M. Kolpashchikov, "The owl sensor: a 'fragile' DNA nanostructure for the analysis of single nucleotide variations," *Nanoscale*, vol. 10, no. 21, pp. 10 116–10 122, 2018.
- [16] B. L. Mueller, M. J. Liberman, and D. M. Kolpashchikov, "OWL2: a molecular beacon-based nanostructure for highly selective detection of single-nucleotide variations in folded nucleic acids," *Nanoscale*, vol. 15, no. 12, pp. 5735–5742, 2023.

- [17] S. Tyagi, U. Landegren, M. Tazi, P. M. Lizardi, and F. R. Kramer, "Extremely sensitive, background-free gene detection using binary probes and beta replicase," *Proceedings of the National Academy of Sciences*, vol. 93, no. 11, pp. 5395–5400, 1996.
- [18] C. Guo, R. Ding, Z. Zhao, J. Guo, and F. Li, "Enrichment Strategies for Low-Abundant Single Nucleotide Mutations," *Chemistry - A European Journal*, vol. 31, no. 1, p. e202402872, 2025.
- [19] R. Van Hoof, M. Szymonik, S. K. Nomidis, K. Hollanders, A. Jacobs, I. Nelissen, P. Wagner, and J. Hooyberghs, "Depletion of wild-type target enhances the hybridization-based sensitivity of low-abundant mutation detection by reference capture probes," *Sensors and Actuators B: Chemical*, vol. 368, p. 132175, 2022.
- [20] Y. Stulens, R. Van Hoof, K. Hollanders, I. Nelissen, M. Szymonik, P. Wagner, G. Froyen, B. Maes, and J. Hooyberghs, "Hybridization-based sensor with large dynamic range for detection of circulating tumor DNA in clinical samples," *Biosensors and Bioelectronics*, vol. 281, p. 117342, 2025.
- [21] N. Wang, X. Zhang, F. Wang, M. Zhang, B. Sun, W. Yin, S. Deng, Y. Wan, and W. Lu, "The Diagnostic Accuracy of Liquid Biopsy in EGFR-Mutated NSCLC: A Systematic Review and Meta-Analysis of 40 Studies," *SLAS Technology*, vol. 26, no. 1, pp. 42–54, 2021.
- [22] K. Huang and A. A. Martí, "Recent trends in molecular beacon design and applications," *Analytical and Bioanalytical Chemistry*, vol. 402, no. 10, pp. 3091–3102, 2012.
- [23] V. Plagnol *et al.*, "Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling," *PLOS ONE*, vol. 13, no. 3, p. e0193802, 2018.
- [24] C. Silveira *et al.*, "Detection and quantification of EGFR T790M mutation in liquid biopsies by droplet digital PCR," *Translational Lung Cancer Research*, vol. 10, no. 3, 2021.
- [25] C. Guo, H. Deng, Q. Yang, D. Huang, C. Shen, G. A. Wang, and F. Li, "Coding Intrinsic Disorder into DNA Hybridization Probes Enables Discrimination of Single Nucleotide Variants over Wide and Tunable Temperature Ranges," *Angewandte Chemie International Edition*, vol. 62, no. 49, p. e202314386, 2023.

- [26] G. Rosenblum, N. Elad, H. Rozenberg, F. Wiggers, J. Jungwirth, and H. Hofmann, "Allostery through DNA drives phenotype switching," *Nature Communications*, vol. 12, no. 1, p. 2967, 2021.
- [27] M. Segers, A. Voorspoels, T. Sakaue, and E. Carlon, "Mechanisms of DNA-Mediated Allostery," *Physical Review Letters*, vol. 131, no. 23, p. 238402, 2023.
- [28] G. Rossetti, P. D. Dans, I. Gomez-Pinto, I. Ivani, C. Gonzalez, and M. Orozco, "The structural impact of DNA mismatches," *Nucleic Acids Research*, vol. 43, no. 8, pp. 4309–4321, 2015.
- [29] M. Sharma, A. V. Predeus, S. Mukherjee, and M. Feig, "DNA Bending Propensity in the Presence of Base Mismatches: Implications for DNA Repair," *The Journal of Physical Chemistry B*, vol. 117, no. 20, pp. 6194–6205, 2013.
- [30] S. Kim *et al.*, "Probing Allostery Through DNA," *Science*, vol. 339, no. 6121, pp. 816–819, 2013.

Chapter 4

Paper C

Double Mismatch Interactions in Probe Design enable Improved Hybridization-based SNV Discrimination

Double Mismatch Interactions in Probe Design enable Improved Hybridization-based SNV Discrimination

Thijs Van der Snickt¹, **Yannick Stulens**², Laura Claudino¹, Simone Bassini^{3,4}, Scott Ailliet¹, Karolien De Wael^{3,4}, Jef Hooyberghs², and Pieter Mestdagh¹

Equal contribution

¹OncoRNALab, Department of Biomolecular Medicine, Ghent University, 9000, Ghent, Belgium

²UHasselt, Data Science Institute, Theory Lab, Agoralaan, Diepenbeek, 3590, Belgium

³Photoelectrochemistry and sensing (A-PECS), Department of Bioscience Engineering, University of Antwerp, 2020, Antwerp, Belgium

⁴NANOLight Center of Excellence, University of Antwerp, 2020, Antwerp, Belgium

SUBMITTED Jun 1, 2026

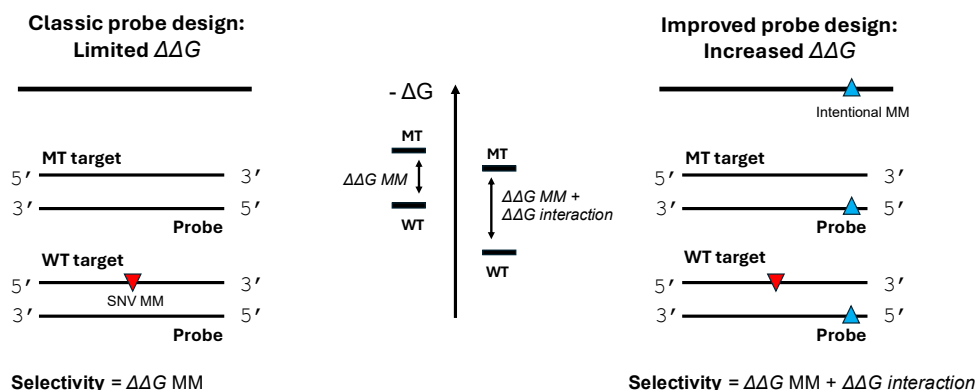
JOURNAL [Sensors & Actuators B: Chemical](#)

Contribution

Early on in my PhD, I briefly investigated interactions between two mismatches, but later shifted to other research directions as new opportunities arose. Half-way during my PhD, Ghent University contacted our research group for possible theoretical support for their data, aligning closely with my earlier interests. This led to a collaborative effort on new sensing strategies and their underlying thermodynamics, resulting in the present manuscript and collaboration which is still ongoing. Thijs Van der Snickt took the experimental lead, while I focused on the theoretical description and on guiding experimental design. We jointly contributed to conceptualization, methodology, and formal analysis. The manuscript was initially drafted by Thijs Van der Snickt, and then extensively re-written together.

Abstract

Achieving high selectivity between mutant (MT) and wild-type (WT) sequences at low variant allele frequencies is challenging because the impact of sequence context and mismatch (MM) positioning on hybridization energetics is poorly understood. To establish design principles for single-nucleotide variant discrimination, we used a newly developed high-throughput platform to systematically quantify the energetics of single and double MMs within a controlled synthetic architecture. This unbiased analysis demonstrated that strategically introducing an intentional MM significantly improves MT/WT selectivity, while further characterizing the distance-dependent interactions that govern duplex stability. By applying this framework to the clinically actionable mutations KRAS p.G12C and BRAF p.V600E, we evaluated libraries of 884 probe variants per target, where each probe carried one intentional MM relative to the MT and a dual-MM state relative to the WT. Through over 100,000 measurements, we identified pronounced positional effects, particularly at probe termini, that optimize discrimination. Benchmarking on a photoelectrochemical biosensor confirmed that these insights translate into improved MT/WT target selectivity. This study provides a data-driven framework for the predictive design of highly selective hybridization assays for clinical mutation detection.



4.1 Introduction

Nucleic acid hybridization serves as the foundational mechanism for modern molecular diagnostics, yet the reliable discrimination of single-nucleotide variants (SNVs) remains a persistent challenge. This difficulty is most acute in clinical samples characterized by low variant allele frequencies (VAFs), where a minute quantity of mutant (MT) DNA is obscured by a vast excess of the corresponding wild-type (WT) sequence [1, 2]. Because an SNV represents only a single-base substitution within an otherwise identical sequence, the MT and WT sequences exhibit near-perfect homology, resulting in nearly identical binding affinities for a MT complementary probe. Consequently, the thermodynamic selectivity, defined as the difference in Gibbs free energy ($\Delta\Delta G$) between the probe-MT and probe-WT complexes, is small compared to the binding affinities [3]. If the $\Delta\Delta G$ does not provide enough selectivity, the probe binds non-specifically to the WT, leading to a cross-hybridized DNA. Because the WT is present in large excess, the combined signal from WT-probe binding becomes indistinguishable from that of the MT. This overwhelming background signal effectively drowns out the rare MT signal, resulting in a loss of sensitivity and potential false negatives [4].

Conventional methods that mitigate cross-hybridization primarily focus on increasing the yield of the intended duplex. Strategies include adjusting probe lengths, incorporating artificial nucleobases (e.g. LNAs, PNAs), or capturing the WT target using WT complementary probes [5, 6, 7, 1, 8]. A considerable limitation of these “equilibrium-shifting” approaches is that the binding energies of both the intended and unintended targets tend to evolve in the same direction. Consequently, the $\Delta\Delta G$ remains largely unaffected, which means that intrinsic thermodynamic differentiation does not improve. This forces a compromise where any enhancement in sequence selectivity must come from increased stringency, which requires a corresponding deterioration in yield [9].

To break this inherent trade-off, we utilize a strategy of deliberate destabilization. Our method intentionally introduces an additional mismatch (MM) into a probe designed for the MT sequence, forcing the WT target into a highly unstable dual-MM state while the MT target carries only a single en-

gineered MM. Even though previous studies like double-MM allele-specific (DMAS)-qPCR or MM-Introduced Probes (MIPs), have explored this, they have primarily treated the extra MM as a static affinity tuner, much like LNAs, to find an experimental window where the WT is less stable [10, 11]. Our approach, however, moves beyond mere affinity tuning; by optimizing the MM configuration, we actively increase the $\Delta\Delta G$, fundamentally improving the thermodynamic difference between two targets.

The intrinsic, hybridization-driven effects of multiple MMs remain poorly understood, and standard nearest-neighbor (NN) models do not include these non-additive effects, thus providing limited guidance for truly predictive probe design [12, 13]. Foundational *in silico* analyses have established general heuristics for single-nucleotide MMs, such as the relative stability of terminal versus internal substitutions. However, empirical insights to increase the $\Delta\Delta G$ by the placement of secondary MMs are still lacking [14, 15]. To address these challenges, we utilized a newly developed high-throughput Next Generation Sequencing (NGS)-based hybridization platform that enables a more systematic analysis of MM interactions. Using a controlled synthetic poly-thymidine (polyT) sequence, we evaluated 430 unique probe designs (in triplicate) to quantify the intrinsic thermodynamics of single- and double-MM states. This large-scale thermodynamic dataset leads to insights beyond the predictive limitations of current models, and allows the identification of specific locations where the intentional placement of an additional MM can disproportionately penalize the WT target.

Building on this foundation, we applied this platform to two clinically actionable mutations, *KRAS* p.G12C and *BRAF* p.V600E. We screened 884 probe designs per target across diverse MM identities, positions, and intentional MM configurations. All probes were evaluated simultaneously in pooled hybridization assays and quantified by NGS, with all experiments performed in duplicate. Screening *KRAS* across 12 assay conditions and four MT/WT mixtures to simulate competitive binding produced 84,864 hybridization measurements, while a focused screen of *BRAF* generated an additional 15,912 measurements. This comprehensive dataset reveals pronounced sequence- and position-dependent effects in MM behavior, demonstrating that intentional MMs can disproportionately destabilize WT binding. Finally, selected probes were benchmarked on a photoelectro-

chemical biosensor, confirming that insights derived from high-throughput hybridization translate directly into improved MT/WT discrimination [16]. Beyond the specific application illustrated in this work, a better understanding of MM interactions directly translates to improved detection limits in high-stakes clinical scenarios, such as monitoring minimal residual disease (MRD), where the ability to resolve rare mutant alleles against a vast WT background is the critical bottleneck for precision oncology [17].

4.2 Materials and Methods

4.2.1 Probe enrichment experiment

Biotin-labeled polyA (homopolymer adenosine), MT and WT target (*KRAS* or *BRAF*) oligonucleotides (Integrated DNA technologies - IDT) were used at 0.1 μ M (Supplementary table 1). For immobilization, 10 μ L of streptavidin-coated Dynabeads M-280 (ThermoFisher Scientific) were washed with 1 mL TE buffer, resuspended in 500 μ L TE, and incubated with 500 μ L of biotinylated target on a rotary shaker (10 rpm, 15 min). Beads were then captured on a magnetic rack and the supernatant removed. Beads were resuspended with 500 μ L TE buffer, followed by 500 μ L of probe pool (Annex 1 for polyT, Annex 2 for *KRAS*, Annex 3 for *BRAF*), adjusted to maintain a 1:25 molar ratio of each individual probe to biotinylated target. For MT-to-WT competition experiments, the TE buffer contained the appropriate amount of non-biotinylated WT target. Tubes were vortexed and incubated on a thermal mixer (600 rpm) for hybridization, then returned to the magnetic rack. Beads were washed three times with 800 μ L TE buffer, transferring the solution to fresh tubes after the second wash. Hybridization was performed under varying conditions of salt, temperature, and MT/WT target ratios (Hybridization conditions). After washing, beads were resuspended in 20 μ L 95% formamide buffer (10 mM EDTA, pH 8.2) and incubated at 65 $^{\circ}$ C for 5 min to release bound target-probe complexes. Beads were separated on a magnetic rack, and the supernatant transferred to fresh tubes. To prevent PCR inhibition from residual formamide, 10 μ L of supernatant was immediately diluted in 190 μ L nuclease-free water. Enrichment yield was assessed by qPCR. For each

primer pair, 2.5 μ L SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) and 0.25 μ L of each primer (5 μ M) were mixed with 2 μ L of eluted DNA in a final 5 μ L reaction. qPCR was performed on a LightCycler 480 (Roche) using the program in (Supplementary table 2). In addition to serving as quality control, the qPCR step determined the optimal number of PCR cycles for Illumina library preparation: the fluorescence amplification curve identified the cycle corresponding to 50% of maximum signal, from which 2 cycles were subtracted to account for higher input volume.

4.2.2 Hybridization conditions

Hybridization for all targets was conducted under reservoir conditions of probe and target to ensure that binding kinetics were not limited by reagent depletion. The polyT sequence melting temperature (T_m) was established at 60°C (IDT OligoAnalyzer™). Because T_m values varied across the *KRAS* and *BRAF* probe pools (Supplementary table 3 and Supplementary table 4), different hybridization temperatures were tested. For *KRAS*, six temperatures (58°C, 60°C, 62°C, 64°C, 66°C, and 68°C) were tested in TE buffer (pH 7.5; 5 mM Tris-HCl, 0.5 mM EDTA, 1% Tween-20) at two ionic strengths (600 mM and 1000 mM NaCl). Probe performance was evaluated against MT and WT targets. Additionally, competitive assays were performed using mixtures of MT and WT, where only the target of interest was biotinylated. *KRAS* mixtures included MT/WT ratios of 100%/0%, 50%/50%, 25%/75%, and 0%/100%. The presence of competitive unbiotinylated targets did not influence hybridization of the target of interest, which confirms the reservoir conditions. A streamlined subset was used for *BRAF*, testing three temperatures (56°C, 60°C, 64°C), a single salt concentration (1000 mM NaCl), and three MT/WT ratios (100%/0%, 25%/75%, and 0%/100%).

4.2.3 Library preparation

For library preparation, 1 μ L forward and 1 μ L reverse primer were mixed with 10 μ L enriched DNA and 8 μ L 2X KapaGrobust HotStart ReadyMix (Kapa Biosystems) (Supplementary table 1). PCR was performed on a

2720 Thermal Cycler (Thermo Fisher Scientific) using the program detailed in Supplementary table 5, which includes cycles of denaturation, annealing, and extension to incorporate Illumina adapters. PCR products were purified using AMPure XP beads (Beckman Coulter) at a 1.8:1 bead-to-sample ratio. Beads were mixed thoroughly, incubated 10 min at room temperature, and captured on a magnetic rack. Beads were washed three times with 80% ethanol (200 μ L each), air-dried for 3 min, and DNA eluted in 14 μ L nuclease-free water. After a final magnetic separation, 12 μ L of eluate was transferred to fresh tubes. DNA fragment size and integrity were assessed using the Fragment Analyzer 96-capillary system (Advanced Analytical Technologies) with PROSize 2.0 software, scaling against an internal standard (Supplementary Fig. S1). This confirmed the presence of the correctly sized Illumina library fragments, including sequencing adapters and inserts.

4.2.4 Sequencing & data processing

DNA libraries from bead-cleaned samples were pooled equimolarly based on Qubit dsDNA HS quantification (Thermo Fisher Scientific) and sequenced paired-end on a NextSeq 2000 (Illumina) using the NextSeq 1000/2000 P1 kit. Libraries were loaded at 650 pM with 15% PhiX. Raw BCL files were converted and demultiplexed into FASTQ files using Illumina Bcl2fastq v2.20. Read quality was assessed with FastQC v0.11.9. For target identification, probe motifs for *KRAS* and *BRAF* (Annex 2 and 3) were mapped directly using SeqKit v2.10.0. However, for the polyT sequences, unique barcode motifs were utilized for mapping (Annex 4). This barcoding strategy was implemented to circumvent the inherent non-homogeneity of the polyT tracts; by replacing the low-diversity polyT signals with high-diversity synthetic barcodes, we prevented potential base-calling errors and registration failures that often occur when the sequencer encounters repetitive, monochromatic sequences. Ultimately, the mapped reads were converted into sequence counts, providing a quantitative readout of the relative binding affinity for each probe configuration. Annex 1-4 and more information can be found on:

<https://github.com/OncoRNALab/Probe-enrichment-platform>.

4.2.5 Free-energy Penalty Calculation

Affinity from sequence-based quantification

Sequencing-based quantification per probe, represented by counts_P , was used to infer binding affinities and relative free energy penalties. In this experimental setup hybridization occurs with a single target sequence, present in excess, and thermodynamic equilibrium is reached without saturating the probes. Under these assumptions, the number of counts for any specific probe is linear in probe concentration (c_P), in target concentration (c_T), and in the binding affinity ($K_{P:T}$). The counts of probe P can be expressed as:

$$\text{counts}_P = A \times c_P \times c_T \times K_{P:T} \quad (4.1)$$

where A is a technology-dependent proportionality factor, including the PCR amplification. The binding affinity $K_{P:T}$ relates to the free energy by the standard thermodynamic relation $K_{P:T} = \exp \Delta G_{P:T}/RT$, such that a stable complex has a negative $\Delta G < 0$. Experimental access to differences of ΔG 's are needed, which corresponds to ratio's of K 's. As a first step, counts are expressed as a ratio relative to the perfectly matching (PM) probe and denoted by an overline, this leads to:

$$\overline{\text{counts}}_P \equiv \frac{\text{counts}_P}{\text{counts}_{PM}} = \frac{c_P}{c_{PM}} \times \frac{K_P}{K_{PM}} \quad (4.2)$$

The factors A and c_T are, by design, completely identical for all probes and drop out of the expression. In principle also $c_P/c_{PM} \approx 1$, but here technical variability cannot be excluded. To measure c_P/c_{PM} , a reference pool (R) is used, which does not undergo pulldown. For the reference pool one gets

$$\overline{\text{counts}}_P^R \equiv \frac{\text{counts}_P^R}{\text{counts}_{PM}^R} = \frac{c_P}{c_{PM}} \quad (4.3)$$

This value is used to produce the fully normalized counts of probe P (double overline):

$$\overline{\overline{\text{counts}}}_P^R \equiv \frac{\overline{\text{counts}}_P}{\overline{\text{counts}}_P^R} = \frac{K_P}{K_{PM}} \quad (4.4)$$

All technical variability is removed, and experimental access is provided to the affinity ratio of the probe+target over perfectly-matching-probe+target.

Equivalently, using $\Delta G = RT \log(K)$, the free energy penalty of probe P relative to the PM probe is given by:

$$\Delta\Delta G_P = RT \log\left(\frac{K_P}{K_{PM}}\right) = RT \log(\overline{\overline{\text{counts}_P}}) \quad (4.5)$$

In our setup, a single experiment provides the normalized probe count for many probes to a single target. Note that this value depends on both probe- and target sequence. When different experiments are combined, it can be important to mention the target T explicitly, and we use the notation $\text{counts}_{P:T}$.

Selectivity from sequence-based quantification against two targets

Repeating the above for a specific probe P against both MT and WT target reveals the binding selectivity of the probe. This can either be expressed as a binding fraction between MT and WT, from the ratio of the normalized counts:

$$\begin{aligned} \text{MT/WT Binding} \\ \text{fraction of probe P} \end{aligned} = \frac{\overline{\overline{\text{counts}_{P:MT}}}}{\overline{\overline{\text{counts}_{P:WT}}}} = \frac{K_{P:MT}}{K_{PM(MT)}} \frac{K_{PM(WT)}}{K_{P:WT}} \quad (4.6)$$

Or as the binding difference towards MT versus WT:

$$\Delta\Delta G_{P:WT} - \Delta\Delta G_{P:MT} = RT \log\left(\frac{\overline{\overline{\text{counts}_P}}}{\overline{\overline{\text{counts}_P}}}\right) \quad (4.7)$$

$$= \Delta\Delta G + \text{Constant} \quad (4.8)$$

Here $\Delta\Delta G$ is the quantity of interest: for a given probe it is the difference in ΔG between a hybridization towards the MT or the WT target. It is a direct measure of how well the probe can distinguish between MT and WT. The constant is probe-independent (the small free energy difference between perfectly matching MT binding versus perfectly matching WT binding). For this purpose, the perfectly matching reference counts of the centrally located binding (scaffold 9) were used for both MT and WT.

4.2.6 Photoelectrochemical biosensor

This biosensor employs a sandwich assay with a capture probe (3' linked with Biotin on magnetic beads) and a photosensitizer-labelled (Chlorin e6) detection probe to detect the target of interest. Functionalized magnetic beads were deposited onto the working electrodes of a 96-well carbon electrode array (DRP-96X110), where an underlying magnet ensured bead attraction to the electrode surface [18]. Upon target recognition and illumination, $^1\text{O}_2$ oxidizes hydroquinone (HQ) to benzoquinone (BQ), which is subsequently reduced back to HQ at the electrode surface (-0.2 V). This electrocatalytic cycle generates the amperometric response [16].

4.3 Results

4.3.1 Mismatch Interactions

For short oligonucleotide sequences, hybridization is accurately described by a two-state model, where a single stranded DNA probe is either bound or unbound [17]. In equilibrium, the amount of DNA in the duplex state is determined by ssDNA concentrations and the binding free energy ΔG , quantifying the strength of the binding. Classic hybridization-based assays make use of a probing strand whose sequence is designed to match that of a target sequence, such that binding of the probe signals its presence. For many purposes, however, it is of critical importance to be highly selective for the MT that differs from the WT sequence by a single nucleotide (in case of a SNV). Despite not being perfectly complementary, the probe can still bind the WT, forming a single base-pair MM. The MM destabilizes the duplex, introducing a free-energy penalty $\Delta\Delta G_{\text{MM}}$ compared to the perfectly matching SNV binding (Fig. 4.1). The selectivity of this classic probe is described by the difference in free-energy towards MT versus WT,

$$\Delta\Delta G = \Delta G_{\text{WT}} - \Delta G_{\text{MT}} = \Delta\Delta G_{\text{MM}}, \quad (4.9)$$

which for the classic probe is equal to the MM penalty. The $\Delta\Delta G$ is of high importance, since it quantifies and provides a measure of selectivity and performance. From a measurement point of view, the fold-change

between MT and WT binding (i.e. the ratio of their experimental readouts) is determined by $\exp(\Delta\Delta G/RT)$. For mixtures of MT and WT, the $\Delta\Delta G$ directly relates to the lowest VAF that can be detected. Importantly, it has been shown that increases in $\Delta\Delta G$ lead to exponential improvements in performance [Paper in submission]. However, $\Delta\Delta G_{MM}$ is determined by the MM under consideration and therefore the selectivity in distinguishing MT versus WT is supposedly fixed.

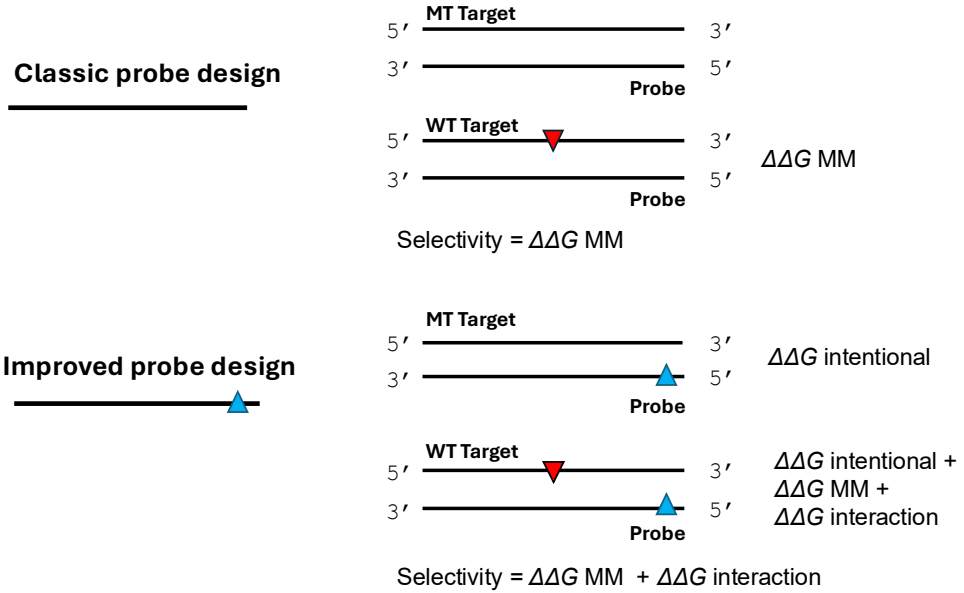


Figure 4.1: Comparison of classic and improved probe designs. Classic probe design: Selectivity is fixed by the thermodynamic penalty of a single mismatch ($\Delta\Delta G_{MM}$) at the SNV site; red triangle. Improved probe design: Introduction of an intentional mismatch (blue triangle) targets both MT and WT. Selectivity is broadened by the interaction term ($\Delta\Delta G_{interaction}$). Optimal discrimination occurs when the interaction is destabilizing ($\Delta\Delta G_{interaction} > 0$), widening the thermodynamic window beyond classic limits.

Here, we explore whether introducing an additional intentional MM can enhance selectivity as quantified by $\Delta\Delta G$. The binding with MT now contains a MM, introducing the penalty $\Delta\Delta G_{intentional}$ over the probe that perfectly matches the SNV sequence (Fig. 4.1). The binding to WT has two MMs that impact the stability of the duplex. The intentional MM contributes $\Delta\Delta G_{intentional}$, the variant position $\Delta\Delta G_{MM}$, and on top of this, the MM's might interact and contribute an interaction $\Delta\Delta G_{interaction}$. The difference in binding free-energy between this new probe towards MT versus WT is

now given by

$$\Delta\Delta G = \Delta G_{\text{WT}} - \Delta G_{\text{MT}} \quad (4.10)$$

$$= \Delta\Delta G_{\text{MM}} + \Delta\Delta G_{\text{interaction}} \quad (4.11)$$

The selectivity is therefore additionally depends on the interaction strength $\Delta\Delta G_{\text{interaction}}$, which could lead to better performance if the interaction is destabilizing and therefore $\Delta\Delta G_{\text{interaction}}$ is positive.

In previous work, the relative interaction strength between MMs was investigated using a clinically relevant sequence [19, 20]. Averaging over multiple MM types and positions, the interaction was found to be predominantly stabilizing and strongly dependent on the separation between MMs. Specifically, the interaction decayed exponentially over a distance of approximately 5 nucleotides (nts), beyond which the interaction vanished and MM penalties became effectively additive. However, in specific cases, a large interaction between MMs takes place, that is possibly even destabilizing. Notably, some of these strong interactions took place when the MM separation was as large as 10 nts, possibly hinting at the periodicity of a helical turn in duplex DNA. However, a systematic correlation between interaction strength and MM separation, position, or identity could not be established. This limitation arose from the complexity of the sequence context, including neighboring base-pairs and the intervening sequence.

A better understanding of MM interactions, specifically the conditions in which the binding is further destabilized due to a positive $\Delta\Delta G_{\text{interaction}}$, could therefore lead to higher selectivity and better performance. Thus, we decided to investigate this using a controlled sequence consisting of 21 consecutive AT base-pairs, eliminating sequence complexity.

4.3.2 High-throughput platform

A controlled hybridization system was established using a PM polyT probe complementary to a 21-nt polyA target. This simple setup eliminated sequence-context effects and enabled systematic investigation of MM stabilities and their interactions. PM probe binding, quantified by qPCR, showed that overnight incubation at 60 °C in 1000 mM NaCl reaches equilibrium without probe saturation, ensuring that observed differences in

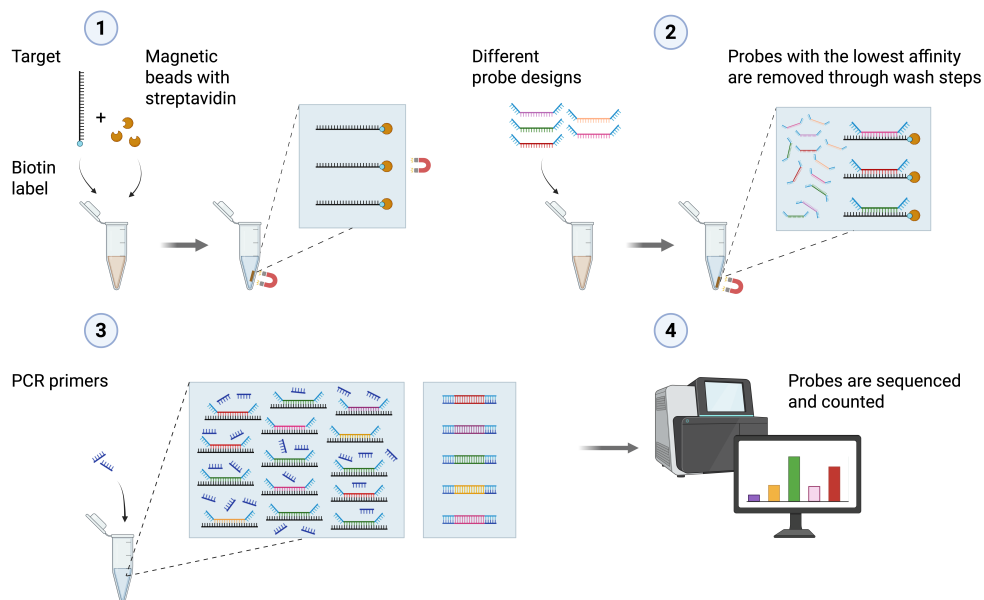


Figure 4.2: 1) Different pooled probe designs that are flanked by PCR handles are hybridized to a biotinylated target. 2) Following streptavidin pull-down, 3) probes are converted into a probe library 4) which is quantified by next generation sequencing (NGS).

binding efficiency reflect intrinsic free-energy penalties (Supplementary Fig. S2).

To enable high-throughput quantification of probe-target interactions under these conditions, we developed a novel and scalable hybridization and sequencing workflow (Fig. 4.2). Biotinylated targets were immobilized on streptavidin-coated beads and hybridized with a probe pool under defined temperature and salt conditions. All probes contained Nextera adapter sequences to allow Illumina library preparation and sequencing (Supplementary table 1). Following stringent washes, probes were eluted, amplified, sequenced, and quantified. Sequencing read counts provided a direct, quantitative measure of probe-target binding, enabling systematic evaluation of affinity and specificity across probes (Sequencing & data processing).

Using this platform, we first evaluated a pool of 20 probes containing 0-4 MMs to the target sequence. The measured binding strengths showed strong correlation with theoretical predictions, validating the ability of the

platform to capture probe-target affinities (Supplementary Fig. S3).

4.3.3 Poly T

With the platform validated, we designed a probe set to systematically study MM effects of probe binding to a polyA target sequence. Probes were designed to be 21 nt in length, and single-nucleotide substitutions were introduced to generate one or two MMs. In the first set (MM1), probes contained a single MM at each position (1-21), with T replaced by A, G, or C. In the second set (MM2), probes carried two MMs: one (MM1) fixed at positions 21, 19, 17, 15, 13, or 11 and a second (MM2) systematically varied at every 5' position, substituting T with either C or G (Fig. 4.3A). These substitutions lead to a highly destabilizing A·C type MM or a less impactful A·G type MM.

Following hybridization, pull-down, and sequencing, binding affinities are quantified as described in (Affinity from sequence-based quantification). Briefly, sequencing counts are compared to those of the PM probe while a non-hybridized probe pool is used to normalize for differences in probe abundance. These parallel measurements eliminate all technical variability and result in a robust measure for binding affinity. The highest affinity is observed for the PM, to which those of other probes are quantified using $\Delta\Delta G$ values. Probes designed to have a single MM showed the expected behavior, A·G MMs are the least destabilizing, followed by A·A and AC, having the highest $\Delta\Delta G$ penalty in this sequence context. Consistent with prior research, the free-energy penalty of a single MM near the termini was observed to be slightly lower than more central MMs (Supplementary Fig. S4) [20]. As expected, the $\Delta\Delta G$ penalties further increased with the addition of a second MM (Supplementary Fig. S5). When two MMs are present, the interaction free-energy penalty $\Delta\Delta G_{\text{interaction}}$ can be determined by comparing the $\Delta\Delta G$ penalty of the individual MMs with the total (for two MMs) penalty.

$$\Delta\Delta G_{\text{interaction}} = \Delta\Delta G_{\text{AB}} - \Delta\Delta G_{\text{A}} - \Delta\Delta G_{\text{B}} \quad (4.12)$$

Here, $\Delta\Delta G_{\text{A}}$ and $\Delta\Delta G_{\text{B}}$ represent the free-energy penalties of the individual MMs, and $\Delta\Delta G_{\text{AB}}$ is the measured penalty when both MMs are present in the same probe. To put the interaction strength relative to the present

MMs and their types, we normalize $\Delta\Delta G_{\text{interaction}}$ and define the relative interaction $\Delta\Delta G_{\text{interaction}}/(\Delta\Delta G_{\text{A}} + \Delta\Delta G_{\text{B}})$. A negative $\Delta\Delta G_{\text{interaction}}$ indicates a stabilizing interaction between two MMs, whereas a positive value reflects cooperative destabilization beyond the sum of the individual MM penalties (Fig. 4.3B). At short distances, values were predominantly negative, indicating strong stabilizing interactions between nearby MMs. With increasing MM distance, the interaction weakens and decays toward zero with a characteristic decay length of $\sim$ three nucleotides. Interestingly, a large variance around zero is present at distance 10, 11 and 12, possibly indicating that distortions can propagate more efficiently over the length of one helical turn ($\sim$ 10.5 bp) of the B-DNA duplex. Importantly, positive values correspond to additional duplex destabilization relative to the additive expectation, which directly enhances discrimination between SNVs and WT targets. Consequently, the intentional placement of a MM at positions yielding positive $\Delta\Delta G_{\text{interaction}}$ provides a design strategy for improving probe selectivity.

To focus on the most informative cases, we restricted our analysis to duplexes in which the position of the first MM (MM A) was fixed at positions 11, 13, 15, or 17 (dashed vertical lines in Fig. 4.3C). The complete dataset is provided in Supplementary Fig. S6. Each panel shows $\Delta\Delta G_{\text{interaction}}$ as a function of the position of the second MM (MM B). When MM B was positioned close to MM A, we consistently observed negative $\Delta\Delta G_{\text{interaction}}$ values, indicating stabilizing interactions between nearby MMs. By contrast, positive interactions occurred predominantly when MM B was located near either probe termini. Internal MMs produced significantly lower $\Delta\Delta G_{\text{interaction}}$ than terminal MMs ($p < 0.001$) (Fig. 4.3D). Notably, a strong positive interaction occurred when two T>G transversions were present, leading to two highly destabilizing A·G-type MMs. In contrast, the T>C transition resulted almost exclusively in stabilizing interactions. This suggests a possible mechanical origin of these interaction free-energies: purine-purine MMs are known to cause high structural distortion, whereas A·C MMs result in a more flexible helix that may be less susceptible to long-range distortions [21]. In short, these findings demonstrate that MM interactions are governed by a combination of spatial separation and base identity, providing a predictable framework for leveraging cooperative destabilization to engineer selective hybridization probes.

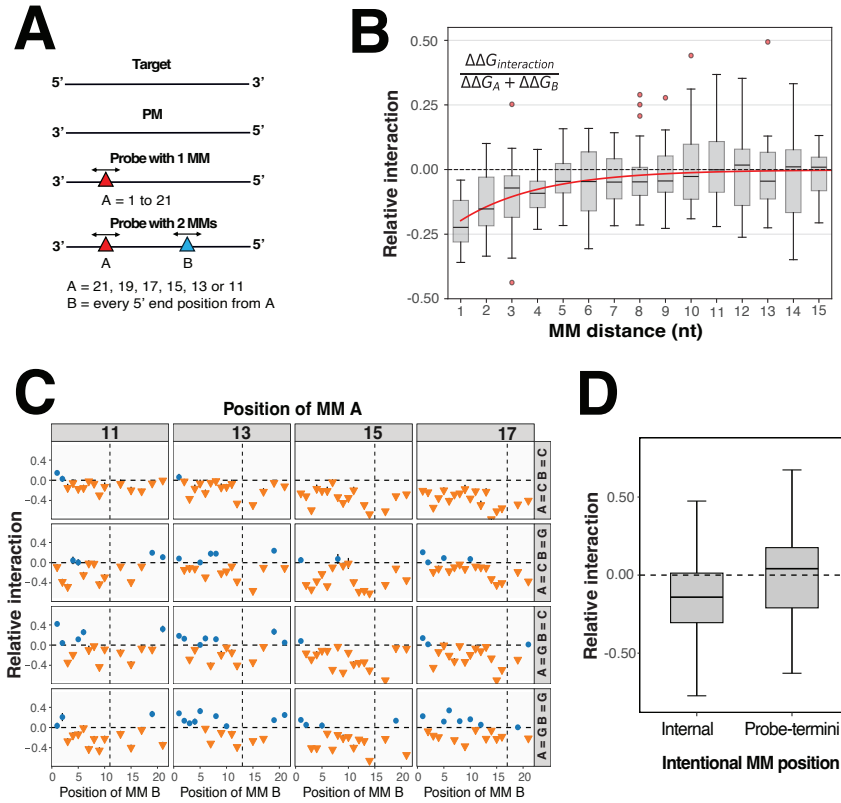


Figure 4.3: Systematic characterization of MM interactions in a polyT:polyA model system. (A) Experimental design: Probes contain a fixed MM (T>C, T>A or T>G) at positions 11-21 (MM A) and a second (T>C or T>G), systematically varied MM at all 5' positions (MM B). (B) Relative interaction strength $\Delta\Delta G_{\text{interaction}}/\Delta\Delta G_A + \Delta\Delta G_B$ as a function of MM separation Δx . (C) Interaction energy as a function of the position of MM B relative to fixed MM A (dashed lines); blue dots indicate a positive interaction energy, while orange triangles indicate a negative interaction energy. (D) Interaction energy in function of the additional MM.

4.3.4 KRAS & BRAF

Based on the systematic characterization of MM effects in a controlled sequence context, we applied this framework to clinically actionable SNVs: *KRAS* p.G12C and *BRAF* p.V600E. We designed 17 probe scaffolds (17-nt sequences shifted along the target; Fig. 4.4) that perfectly matched the MT target. To enhance the selectivity of binding to the MT over WT target, each scaffold was modified to generate 52 designs by introducing a single intentional MM at each position (three substitutions per position). Thus,

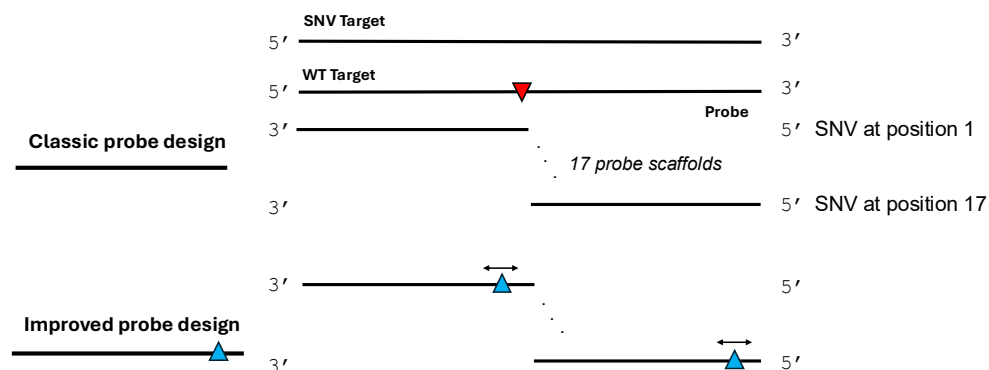


Figure 4.4: A 17-nt capture probe was systematically shifted along the target sequence (resulting in 17 probe-target binding scaffolds) to cover all possible SNV sites (top panel, classic probe design). Each probe was fully complementary to the MT target and contained a single MM with the WT sequence. To improve MT over WT selectivity, an additional intentional MM was introduced at every position by substituting each nucleotide with the three non-complementary bases (bottom panel, improved probe design). This yielded, for each of the 17 probe positions (scaffolds), one perfectly matched MT probe and 51 single-nucleotide variants ($17 \times 52 = 884$ probes per target), resulting in probes with one MM to MT and two to WT.

$17 \times 52 = 884$ probes were designed for both *KRAS* and *BRAF* (Annex 2 and Annex 3).

We first evaluated binding affinity for the *KRAS* p.G12C probe set by screening all 884 probes in two replicates across 12 hybridization reaction conditions against MT, WT, and two MT/WT mixtures to simulate competitive binding. This resulted in a total of 84,864 data points. As before, the non-hybridized probe pool was sequenced to normalize for differences in probe abundance induced during probe synthesis.

Among the 52 probes with the *KRAS* p.G12C variant positioned centrally (position 9) and hybridized to the MT target (68 °C, 1000 mM NaCl), the PM (classic probe) exhibited one of the highest binding affinities, with normalized counts averaging 4.9-fold higher than those of MM probes (Fig. 4.5A). Notably, three probes carrying substitutions at position 1 (probe 417: T>A, probe 418: T>C, and probe 419: T>G) slightly outperformed the PM probe, with a 72%, 67%, and 5% higher binding strength, respectively. This trend was consistent across hybridization conditions (Supplementary Fig. S7). Within each set of 52 probes, one probe perfectly

matches the WT sequence due to an intentional MM at the variant position. When hybridized to the WT target, this WT-matching probe displayed normalized counts 14.2-fold higher than the MT probe, further validating assay performance (Fig. 4.5A). Importantly, binding of the MT probe to the WT target introduces an A·G MM (relative binding strength: 0.07), whereas binding of the WT probe to the MT target introduces a C·T MM (relative binding strength: 0.05). These MMs differ in stability, explaining the observed asymmetry in off-target binding between MT and WT probes [22]. Together, these results demonstrate that normalized counts provide a reliable and interpretable measure of probe affinity across probe variants and hybridization conditions (Supplementary Fig. S7).

We then evaluated the MT/WT binding fraction as a measure for probe selectivity (Fig. 4.5A). Among the 52 probes with the *KRAS* p.G12C variant positioned in the middle of the probe, the probe perfectly matching the MT target showed a ratio of 15.23 ($\Delta\Delta G = 1.85$ kcal/mol) while the probe perfectly matching the WT target showed a count ratio of 0.04 ($\Delta\Delta G = -2.09$ kcal/mol). Consistent with expectations on MM positioning, probes with the additional MM near the edges outperform those with central MMs. In particular, probe 419 (T>G at position 1) and probe 421 (A>G at position 2) exhibited MT/WT ratios of 17.05 and 16.53 ($\Delta\Delta G = 1.92$ and 1.90 kcal/mol) respectively, slightly exceeding the selectivity of the perfectly matched MT probe ($\Delta\Delta G = 1.85$ kcal/mol). The same trend was observed at 600mM NaCl and different hybridization temperatures (Supplementary Fig. S7). Generally, these results further validate that off-center MMs can enhance discrimination between MT and WT targets.

We expanded our analysis to all 884 probes, covering 17 variant position scaffolds (position 1 at the 5' end to position 17 at the 3' end) (Fig. 4.4). Within each scaffold, the position of an intentional MM was systematically shifted across the sequence (Fig. 4.5B). This design revealed an emergent selectivity pattern: as the SNV moves (scaffold number) from the 5' end toward the 3' end, the $\Delta\Delta G$ of the WT-matching probes increases when the SNV is located near either duplex terminus. This is consistent with prior studies showing reduced MM penalties at duplex ends [21]. For each scaffold, the varying intentional MM position results in a characteristic "V-shape" pattern centered at the SNV location. This represents a stabilizing

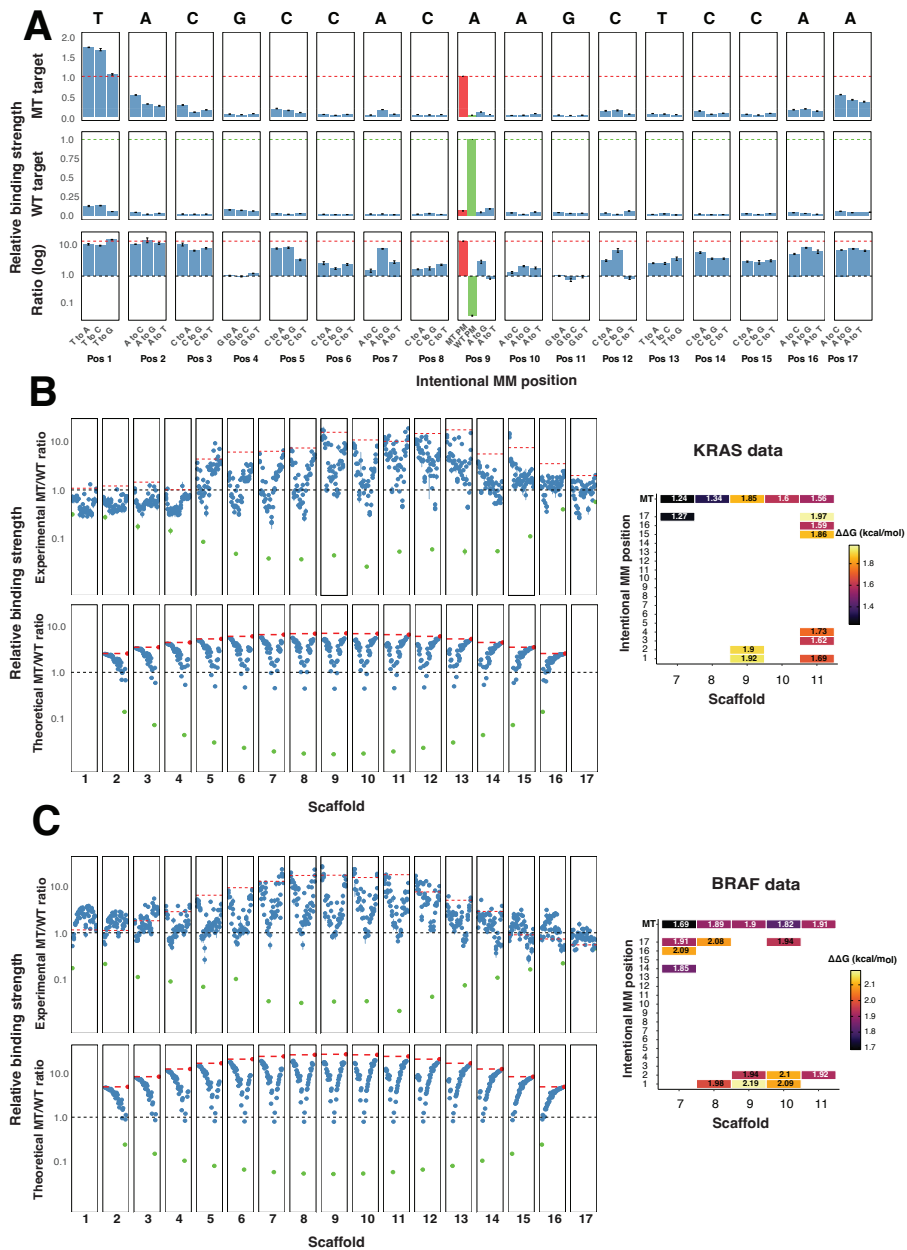


Figure 4.5: (A) Average results of two independent pulldown experiments at 68 °C in 1000 mM NaCl. Data for 52 probes containing the *KRAS* G12C variant positioned centrally (position 9). Binding profiles are shown against the MT and WT targets, in addition to the MT/WT count ratios on a log scale. (B,C) Hybridization results for all 884 probes targeting *KRAS* and *BRAF* against MT and WT targets. Red indicates MT-perfect matches, green indicates WT-perfect matches, and blue denotes probes containing an additional substitution relative to the MT target. $\Delta\Delta G$ values for probes that outperformed their MT-perfect-match counterparts within scaffolds 7-11 are shown on the heatmaps.

effect when two MMs are in close proximity. To validate these observations, we integrated these proximity effects into a NN model, which successfully replicated the experimental free-energy pattern (Fig. 4.5B). While this model correctly identifies central variant positioning (positions 5-13) as a primary driver of selectivity (mean $\Delta\Delta G$ for the MT-matching probes = 1.49 kcal/mol), the data reveals that absolute central alignment is not a strict requirement for peak performance. Specifically, MT-matching probes in scaffolds 10, 12, and 13 yielded $\Delta\Delta G$ values of 1.60, 1.80, and 1.92 kcal/mol respectively, aligning with or exceeding the 1.85 kcal/mol of the canonical central probe. More importantly, we identified 16 probes in this central region where intentional MMs considerably enhanced selectivity compared to perfectly matched MT counterparts (Fig. 4.5C). This "MM-enhancement" is most notable in probes 419, 421 and 362 which achieved $\Delta\Delta G$ gains of 0.07, 0.05, and 0.41 kcal/mol, respectively. These findings suggest that a strategic terminal intentional MM (positions 1-2) combined with a central variant location (7-11) can lead to the increased free-energy penalty of an SNV, providing an design rule for highly-selective variant detection (Fig. 4.5B). These trends were confirmed across a temperature range (58-68°C) and at another salt concentration (600mM) (Supplementary Fig. S8).

To further substantiate these findings, we screened an independent target, *BRAF* V600E, using 884 probes in duplicate across three temperatures (56-64 °C), one salt concentration (1000 mM NaCl), MT, WT and one MT/WT mixture to simulate competitive binding, yielding 15,912 additional data points. The *BRAF* results strikingly mirrored the *KRAS* data: in the centrally positioned scaffold (position 9), the MT probe exhibited high affinity (5-fold above the average across all probes), yet was again outperformed by terminal intentional MMs (Supplementary Fig. S9). Specifically, probes 419 and 421 achieved 39% and 93% higher binding strength, respectively, than the perfectly matched MT probe. When hybridized to the WT target, the WT-matching probe showed 16.6-fold higher binding than the MT probe (relative binding strength: 0.06), reflecting the higher stability of the resulting T·T MM compared to the A·A MM formed in the reciprocal condition (relative binding strength: 0.03). These trends were reproduced at different hybridization conditions (Supplementary Fig. S10). As with *KRAS*, our NN model accurately predicted the global selectivity trends, identifying the central region (scaffolds 5-13) as the primary window for high discrimi-

nation (Fig. 4.5C). Similarly, while the NN-model suggests the best selectivity for the MT-matching probes, we identified 40 mismatched probes within these central scaffolds that outperformed the theoretical optimum (Fig. 4.5C). Specifically, probes with MMs at the 3' end (position 17: mean $\Delta\Delta G = 1.98$ kcal/mol; position 16: mean $\Delta\Delta G = 2.09$ kcal/mol) and the 5' end (position 2: mean $\Delta\Delta G = 1.99$ kcal/mol; position 1: mean $\Delta\Delta G = 2.09$ kcal/mol) exceeded the performance of the canonical MT-matching probes. The highest average selectivity was recorded for position 1 variants, yielding a $\Delta\Delta G$ gain of 0.29 kcal/mol over the perfectly matched sequence (Fig. 4.5C). These trends remained robust across a range of temperatures (Supplementary Fig. S10).

To further evaluate the robustness of these design motifs on longer probes, we extended our screening to 20- and 25-nucleotide *KRAS* sequences. Consistent with our observations on 17-nt probes, mismatched probes with substitutions at positions 1 or 2 from either terminus maintained superior selectivity, demonstrating that terminal substitutions enhance MT/WT discrimination even as the total duplex length increases (Supplementary Fig. S11).

Together, the results of the polyT, *KRAS* and *BRAF* libraries identified a consistent motif that maximizes thermodynamic discrimination. Probes that outperformed the perfectly matched MT probes shared a common design: a centrally located SNV (positions 7-11) paired with an intentional MM at a duplex terminus (positions 1 or 2). Interestingly, peak selectivity consistently occurs at a separation of 8-9 nucleotides, a distance that tracks closely with one full turn of the B-DNA helix. Future research is needed, however, to pinpoint if this effect is driven by the terminal placement of the MM or by the distance between the MMs.

4.3.5 Photoelectrochemical benchmark

To benchmark the utility of intentional MM introduction as a functional strategy for enhancing MT/WT selectivity, we transitioned our findings to a photoelectrochemical (PEC) detection platform. In this system, assay selectivity is intrinsically governed by the sequence-specific hybridization between a surface-immobilized capture probe and the target nucleic acid.

For the *KRAS* p.G12C mutation, a maximum MT/WT ratio of 6.25 had been previously reported using this platform [16]. We first evaluated this benchmark probe (probe 416) and confirmed its reported performance (Supplementary Fig. S12).

Based on this baseline, we selected three probes from scaffold 9. We chose the perfectly matched MT probe (468) as a control, alongside probes 419 and 421. These specific variants were selected because they consistently emerged as the most selective probes in our previous screens, where the placement of an intentional MM at the probe termini (T>G at position 1 and A>G at position 2, respectively) maximized the $\Delta\Delta G$ and outperformed the canonical MT-matching probe ($\Delta\Delta G$ gains of 0.07 and 0.05, respectively). As expected, the absolute photocurrent generated upon MT hybridization was highest for the perfectly matched probe 468, indicating that the introduction of a secondary MM results in a modest reduction in sensitivity. However, analysis of MT/WT discrimination revealed a decisive advantage for the mismatched designs (Fig. 4.6). Specifically, probe 421 achieved the highest MT/WT ratio of 29.4-fold, outperforming the perfectly matched probe 468 (7.87-fold) by nearly 4-fold, representing a substantial reduction in non-specific WT signal. Compared to the previously reported benchmark (probe 416), probe 421 exhibited a ~ 3.3 -fold improvement in MT/WT discrimination, underscoring the superior performance of our rationally designed configurations (Fig. 4.6).

The clinical relevance of these thermodynamic refinements is most evident in the fractional abundance (FA) analysis, the ratio of MT signal to the total (MT+WT) signal. At low VAFs, even marginal WT cross-reactivity can disproportionately suppress the FA, obscuring the presence of rare mutant alleles. To illustrate the impact of our design, we compared the performance of the classical MT probe (468, 7.87-fold selectivity) against our optimized double-mismatched variant (421, 29.4-fold selectivity) across two key clinical benchmarks. At a 10% VAF, the FA increases from 46.7% to 76.6%, representing a 1.6 times increase in the relative MT signal. The impact is even more pronounced at a 1% VAF, where the FA rises from 7.4% to 22.9%; this triple increase in fractional signal effectively rescues the variant from the WT noise (Fig. 4.6B).

To confirm the generalizability of this strategy, we extended PEC valida-

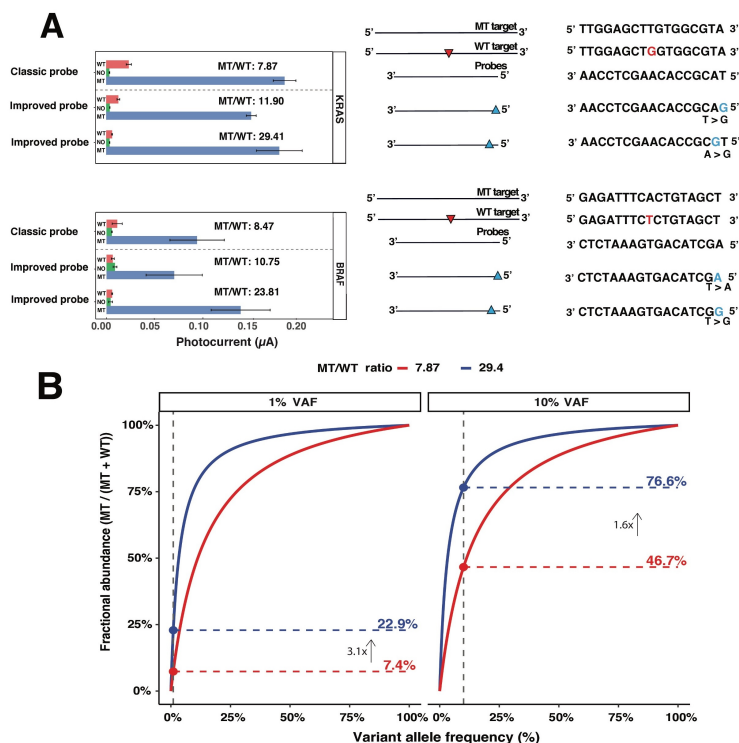


Figure 4.6: (A) Chronoamperometric responses recorded for *KRAS* p.G12C and *BRAF* V600E targets (blue), WT targets (red), and no-target controls (green) following hybridization at 68 °C for *KRAS* or 60 °C for *BRAF*. The sequences and designs of the probes used are shown right to these results. (B) Observed fractional abundance ($FA = MT\ signal / [MT\ signal + WT\ signal]$) as a function of VAF, simulated across a MT/WT ratio of 7.87 and 29.4 at a VAF of 1%, and a VAF of 10%.

tion to the *BRAF* p.V600E sequence. Consistent with the *KRAS* results, the classical perfectly matched MT probe and the improved mismatched variants followed a nearly identical hierarchical trend, with probes incorporating intentional mismatches again exhibiting the highest MT/WT fold discrimination (Fig. 4.6B). Together, these data validate the introduction of strategically placed mismatches as a robust and broadly applicable approach to improving SNV discrimination on PEC-based nucleic acid detection platforms.

4.4 Discussion

The ability to distinguish SNVs from an overwhelming background of WT DNA is a fundamental requirement for liquid biopsy and early cancer detection. While traditional hybridization-based assays rely on the free-energy penalty of a single MM ($\Delta\Delta G_{\text{MM}}$), it is often too small to prevent cross-hybridization at low VAFs [23, 13]. In this study, we demonstrate that the rational introduction of a second, intentional MM does not merely shift the binding equilibrium, but actively broadens the thermodynamic discrimination window by exploiting non-additive MM-MM interactions. While MM-MM interactions are generally stabilizing (a trend we confirmed follows an exponential decay in terms of MM separation), we identified an important spatial transition [24]. As one of the MMs moves closer to the probe termini, the interaction occasionally shifts from stabilizing (negative $\Delta\Delta G_{\text{interaction}}$) to destabilizing (positive $\Delta\Delta G_{\text{interaction}}$). Since destabilizing interactions increase the penalty associated with an SNV MM, intentional MMs provide a new design parameter to enhance probe selectivity.

Our data indicated specific design rules for the placement of the SNV and intentional MM. The high-throughput analysis revealed that in the improved probe design, the SNV placement should be central, with the intentional MM at one or two nucleotides from the probe termini. The SNV position can be moved slightly off-center while maintaining high selectivity in the improved probe design. This has significant practical importance, for example in avoiding sequence complexities, such as high GC-content clusters or unfavorable secondary structures. Although positional rules for the intentional MM placement proved to be remarkably robust across the *KRAS*, *BRAF*, and polyT sequences, the optimal substitution type varied. Our finding that T>G substitutions (forming G·A pairs) are effective for polyT, was not transferable to the clinical sequences, highlighting the influence of sequence context. We propose that while the “where” is a universal design motif, the “what” should be screened or modeled against multiple diverse sequence contexts to find the maximum $\Delta\Delta G$ gain.

The observed selectivity improvement possibly originates from structural destabilization of the duplex. The intentional MM disrupts the local base-pairing, thereby introducing mechanical strain into the duplex. This im-

pact could propagate along the duplex and affect stability of base pairing at remote locations. Another possibility is related to the partitioning of the duplex into shorter segments by the central SNV. The additional terminal MM (at position 1 or 2) further shortens one segment, approaching the minimum length required for stable hybridization [24, 25]. The selectivity gain might therefore arise from the presence of unstable fragments adjacent to the SNV, which disproportionately destabilize the double-mismatched duplex. Future work should investigate the underlying mechanism and the role of MM type in the observed selectivity enhancement, as different MM types are known to induce distinct structural distortions and thermodynamic penalties.

A central perception in molecular diagnostics is the inherent compromise between sensitivity and selectivity. Our results are valid far from saturation, and we improved the selectivity on the level of the free-energy penalty $\Delta\Delta G$. Thus, while our results indicate that adding a MM incurs a reduction of sensitivity, concentrations can be adjusted to offset this reduction while maintaining the increase in selectivity. This is supported by established thermodynamic frameworks that demonstrate that improvements in $\Delta\Delta G$ have an exponential impact on binding discrimination. Consequently, the relative detection limit improves exponentially with increasing thermodynamic selectivity [26]. These exponential gains are critical for applications such as monitoring MRD, where the primary diagnostic challenge is not the detection of the MT itself, but the suppression of the vast WT background that can obscure rare alleles [27].

In our PEC biosensor benchmarking, a modest reduction in absolute photocurrent was offset by a 3.7-fold reduction in non-specific WT signal. This decoupling allows for a non-linear amplification of the FA at ultra-low VAFs. By effectively tripling the relative MT signal at a 1% VAF, this strategy provides the resolution required to identify rare oncogenic mutations that would otherwise remain obscured, a prerequisite for high-sensitivity applications such as post-treatment monitoring.

In conclusion, we provided a data-driven framework for the rational design of MM-introduced probes. By exploiting non-additive MM-MM interactions, we move beyond traditional probe design rules. This strategy is platform-agnostic and cost-effective, offering a clear path toward the development

of highly selective SNV assays for clinical diagnostics without the need for complex chemical modifications or expensive artificial nucleotides.

Author contributions: TVS, YS, PM and JH were responsible for conceptualization and methodology. TVS and YS conducted the investigation. LC, SB, and SA contributed to the experiments and participated in the discussions. TVS and YS performed formal analysis, data curation, and visualization. PM and JH were responsible for supervision and project administration. PM, JH, and KDW acquired funding. TVS and YS wrote the original draft. All authors reviewed and edited the manuscript.

Funding: This work was supported by the Interuniversity Special Research Fund [iBOF/23/030], Special Research Fund [BOF22OWB11], and Research Foundation Flanders (FWO) [G078223N, 1S39325N]. We would like to thank the SOCan consortium for fostering a collaborative environment and engaging in fruitful discussions.

Supplementary Information

Data Availability

Sequencing data

Raw sequencing data (FASTQ files) have been deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under BioProject accession PRJNA1442078. Individual datasets are available under the following SRA submission accessions:

- polyT experiments (SUB16080050)
- KRAS G12C experiments with 17-mer probes (SUB16078753)
- KRAS G12C experiments with 20-mer probes (SUB16081042)
- KRAS G12C experiments with 25-mer probes (SUB16081095)
- BRAF V600E experiments (SUB16079470)

The code used for the probe enrichment platform, together with processed data (including metadata and count matrices), is publicly available at GitHub: <https://github.com/OncoRNALab/Probe-enrichment-platform>.

Quantitative PCR

Quantitative PCR data obtained to optimize reaction conditions (Fig. S2) will be made available online according to MIQE guidelines.

Photoelectrochemical readout

Electrochemical readout (μA current) of the photoelectrochemical detection platform (Fig. 6 and Fig. S12) will be made available online.

Supplementary Methods

Free-energy penalty computations with edge and separation correction

Given a probe sequence P and target sequence T, a MM is located and its free-energy penalty is predicted in the context of its neighbors using parameters from [30]. For example,

$$\Delta\Delta G_{MM1} = \Delta G\left(\begin{smallmatrix} \text{AAG} \\ \text{TGC} \end{smallmatrix}\right) - \Delta G\left(\begin{smallmatrix} \text{ACG} \\ \text{TGC} \end{smallmatrix}\right) \quad (\text{S1})$$

isolates the contribution of the MM. Subsequently, a first empirical correction is applied: the penalty is lowered based on its distance from either end of the duplex using a proportionality factor

$$\Delta\Delta G_{MM1}^* = \Delta\Delta G_{MM1} \left(1 - e^{-d_{3'}/L_e}\right) \left(1 - e^{-d_{5'}/L_e}\right), \quad (\text{S2})$$

where $d_{3'}$ and $d_{5'}$ are the distances (number of nucleotides) of the MM from the 3' or 5' end, respectively, and L_e is the persistence length of this edge effect. If the distance from the edges is much larger than L_e , the penalty of the MM remains unchanged. If the MM is near an edge, its penalty is lowered significantly.

A second empirical correction is applied only when two MM are present, possibly already influenced by the edge effect, e.g. $\Delta\Delta G_{MM1}^*$ and $\Delta\Delta G_{MM2}^*$. The idea is that the configuration where two MMs are present is less destabilizing than the sum of the two penalties, and the effect depends on the distance between the MMs Δx . This is captured by:

$$\Delta\Delta G_{2MM}^* = (\Delta\Delta G_{MM1}^* + \Delta\Delta G_{MM2}^*) \left(1 - \frac{1}{2} e^{\Delta x/L_a}\right), \quad (\text{S3})$$

such that when the MM separation is much larger than the persistence length of the interaction L_a , the total penalty is just the sum of the two individual MM penalties, but when they are close there is a stabilizing interaction.

The above allows for comparison of stabilities of a probe towards MT versus WT (where in the latter possibly two MMs are present), resulting in the selectivity measure $\Delta\Delta G$:

$$\Delta\Delta G = \Delta G_{WT} - \Delta G_{MT}, \quad \text{or equivalently, the MT/WT Binding Ratio} = e^{\Delta\Delta G/RT} \quad (\text{S4})$$

Supplementary Tables

Table S1: Sequences used in the probe enrichment experiments and in the subsequent PCR step. The position of the variant (KRAS G12C or BRAF V600E) is highlighted.

Oligonucleotide	5' modification	Sequence (5'→3')
Nextera Transposase Adapter -i5	-	5'- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG -3'
Nextera Transposase Adapter -i7	-	5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -3'
Index 2 (i5) Adapters	-	5'- AATGATACGGCGACCACCGAGATCTA CAC[i5]TCGTCGGCAGCGTC -3'
Index 1 (i7) Adapters	-	5'- CAAGCAGAAGACGGCATACGAG AT[i7]GTCTCGTGGGCTCGG -3'
Poly A	Biotin-TEG	5'- AATGACTGAATATAAACTTGTG GGCAAAAAAAAAAAAAAAAAAAAAACG -3'
KRAS G12C Target WT	Biotin-TEG	5'- AATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGT GGCGTAGGCAAGAGTGCCTTGACGATACAGCT -3'
KRAS G12C Target MT	Biotin-TEG	5'- AATGACTGAATATAAACTTGTGGTAGTTGGAGCTTGT GGCGTAGGCAAGAGTGCCTTGACGATACAGCT -3'
BRAF V600E Target WT	Biotin-TEG	5'- TTCAAACCTGATGGGACCCACTCCATCGAGATTTCTCT GTAGCTAGACCAAAATCACCTATTTTACTG -3'
BRAF V600E Target MT	Biotin-TEG	5'- TTCAAACCTGATGGGACCCACTCCATCGAGATTTCACT GTAGCTAGACCAAAATCACCTATTTTACTG -3'

Table S2: qPCR cycling program for probe enrichment quantification

	Temperature (°C)	Time (s)	
Pre-incubation	95	300	
Denaturation	95	10	45 cycles
Annealing	60	10	
Replication	72	10	
Melting curves	95	5	
Continuous acquisition	65	60	
Cooling	40	30	

Table S3: The position of the variant in the probe sequence, the delta T_m, and the GC% content of the 17 MT probes (perfect match MT target) developed for KRAS G12C

Position of the variant	Probe sequence (5' → 3')	$\Delta T_m(MT - WT)$	GC%
1	CTCTGCCTACGCCACA A	0.1	58.82
2	TCTTGCCTACGCCACA AA	4.9	52.94
3	CTTGCCTACGCCACA AAG	5.2	58.82
4	TTGCCTACGCCACA AAGC	4.9	58.82
5	TGCCTACGCCACA AAGCT	4.8	58.82
6	GCCTACGCCACA AAGCTC	4.9	64.71
7	CCTACGCCACA AAGCTCC	5.0	64.71
8	CTACGCCACA AAGCTCCA	5.2	58.82
9	TACGCCACA AAGCTCCAA	4.9	52.94
10	ACGCCACA AAGCTCCAAC	5.0	58.82
11	CGCCACA AAGCTCCAAC	5.1	58.82
12	GCCACA AAGCTCCAAC	5.2	52.94
13	CCACA AAGCTCCAAC	5.4	52.94
14	CACA AAGCTCCAAC	5.4	52.94
15	ACA AAGCTCCAAC	6.1	47.06
16	CA AAGCTCCAAC	5.4	52.94
17	AAGCTCCAAC	0.3	47.06

Table S4: The position of the variant in the probe sequence, the delta T_m, and the GC% content of the 17 MT probes (perfect match MT target) developed for BRAF V600E.

Position of the variant	Probe sequence (5' → 3')	$\Delta T_m (MT - WT)$	GC%
1	TTTGGTCTAGCTACAG T	2.0	41.18
2	TTGGTCTAGCTACAG TG	7.9	47.06
3	TGGTCTAGCTACAG TGA	7.9	47.06
4	GGTCTAGCTACAG TGAA	7.9	47.06
5	GTCTAGCTACAG TGAAA	8.0	41.20
6	TCTAGCTACAG TGAAAT	7.7	35.20
7	CTAGCTACAG TGAAATC	8.0	41.18
8	TAGCTACAG TGAAATCT	7.7	35.30
9	AGCTACAG TGAAATCTC	8.0	41.18
10	GCTACAG TGAAATCTCG	5.0	47.06
11	CTACAG TGAAATCTCGA	7.9	41.18
12	TACAG TGAAATCTCGAT	7.7	35.30
13	ACAG TGAAATCTCGATG	7.8	41.18
14	CAG TGAAATCTCGATGG	7.7	47.06
15	AG TGAAATCTCGATGGA	8.0	41.18
16	GT TGAAATCTCGATGGAG	7.8	47.06
17	TGAAATCTCGATGGAGT	0.9	41.18

Table S5: PCR cycling conditions for Illumina library creation

	Temperature (°C)	Time (s)	
Pre-incubation	95	15	
Denaturation	95	15	
Annealing	63	10	X cycles
Replication	72	15	
Extension	72	60	
Cooling	4	∞	

Supplementary Figures

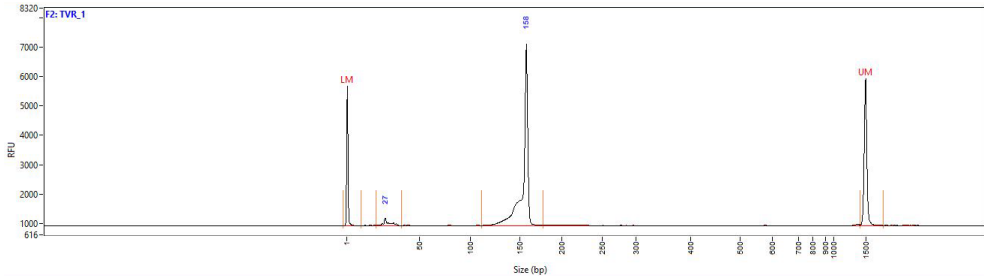


Figure S1: Fragment analyzer electropherogram showing a single, sharp peak corresponding to the 158 bp expected amplicon (KRAS target). This results confirms the specificity of the PCR conditions and the effectiveness of AMPure XP bead cleanup (1.8x ratio) in removing primer dimers and non-specific products.

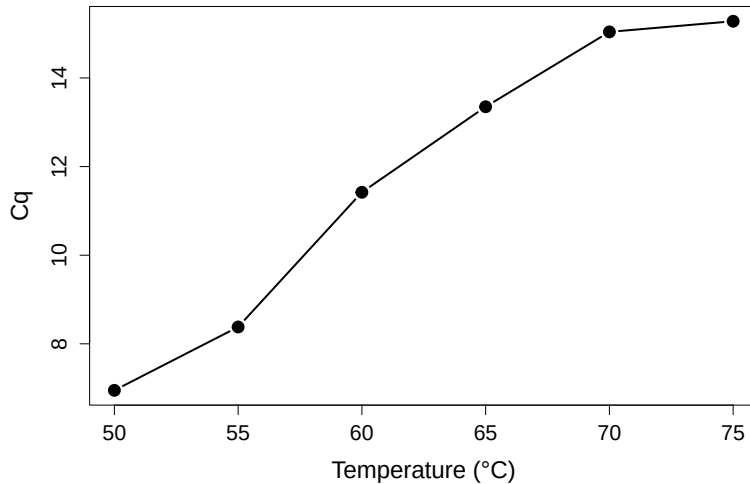


Figure S2: qPCR quantification of probe binding (perfect match) as a function of temperature (55-75 °C). 60 °C was identified as the optimal hybridization temperature for the polyT:polyA system. At 60 °C, the system reaches a stable thermodynamic equilibrium without probe saturation, ensuring that the measured binding efficiencies accurately reflect intrinsic free-energy penalties ($\Delta\Delta G$). On the Y-axis the Cq results are shown, on the X axis the corresponding hybridization temperature in the pulldown experiment.

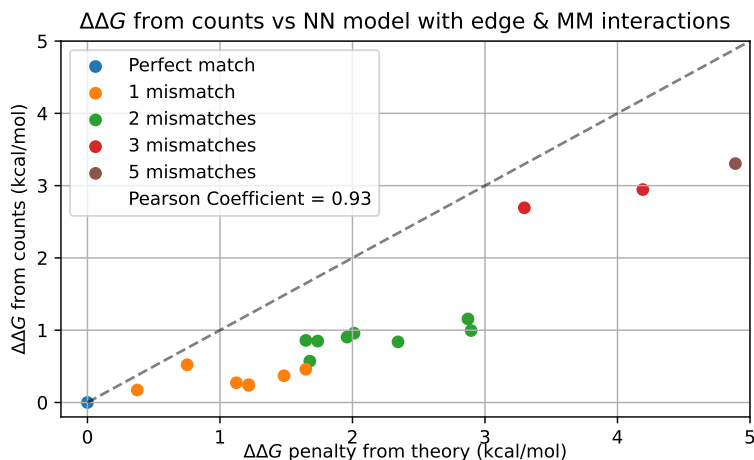


Figure S3: Correlation of measured versus calculated free-energy penalties $\Delta\Delta G$ of 20 test probes involving 0-5 mismatches (MMs). Theoretical predictions involve a nearest neighbor (NN) model estimation in addition to two empirical corrections as outlined in supplementary methods: a reduction in penalty when the MM is close to the duplex end, and a mutual stabilizing interaction between mismatches. The remaining deviation from the one-to-one (dashed line) behavior is the result of higher order stabilizing effect or slightly out-of-equilibrium measurements.

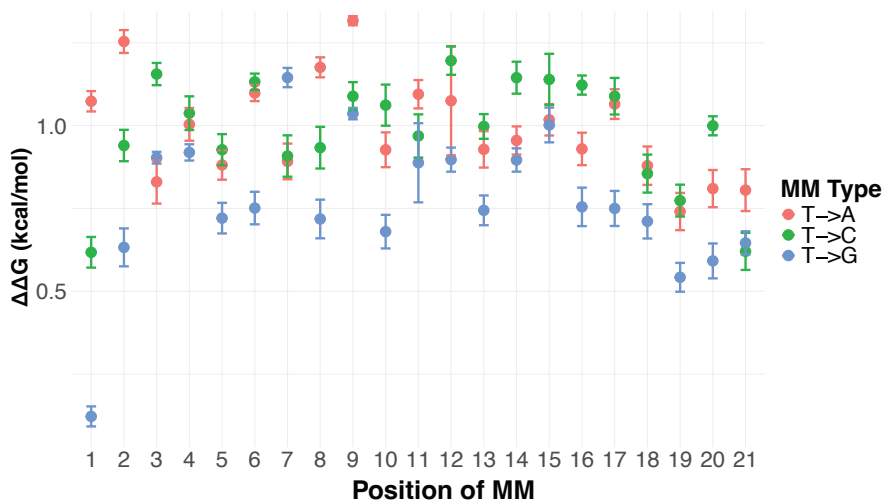


Figure S4: Binding affinities ($\Delta\Delta G$) for the single-mismatch (MM1) probe library. Position-dependent penalty profiles demonstrate that MM destabilization is maximal in the central region of the 21-nt duplex and decreases as the substitution approaches the termini. Influence of mismatch identity on duplex stability is consistent with established thermodynamic models: A·G mismatches were found to be the least destabilizing, while A·C mismatches yielded the highest free-energy penalties.

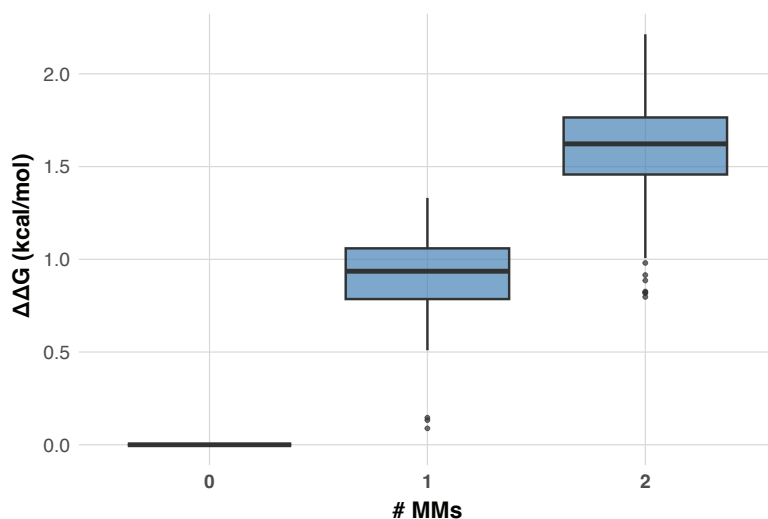


Figure S5: Distribution of $\Delta\Delta G$ penalties for the perfect match, the single-mismatch (MM1) and the double-mismatch (MM2) library.

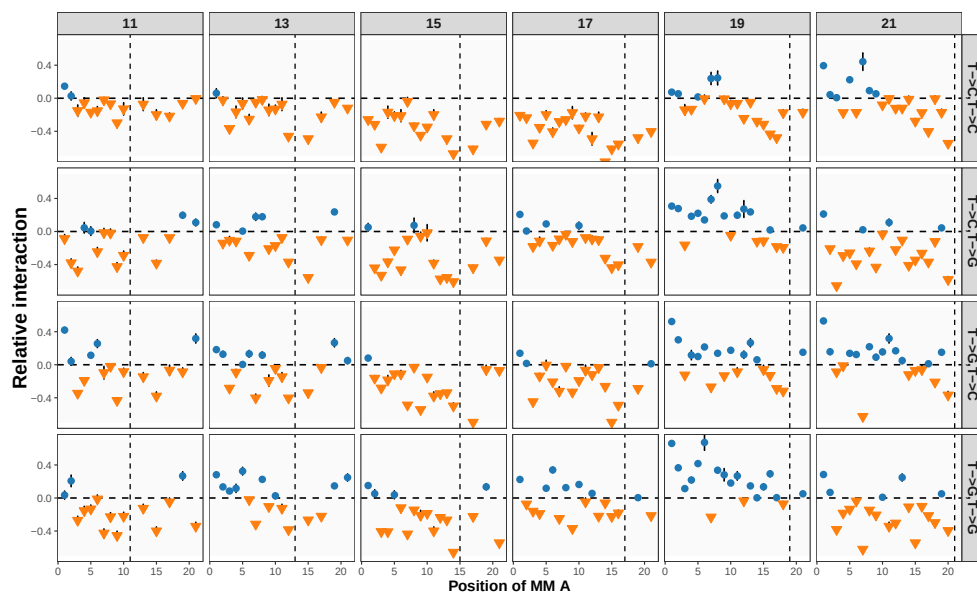


Figure S6: Interaction energy as a function of the position of MM B relative to fixed MM A (dashed lines); blue dots indicate a positive interaction energy, while orange triangles indicate a negative interaction energy.

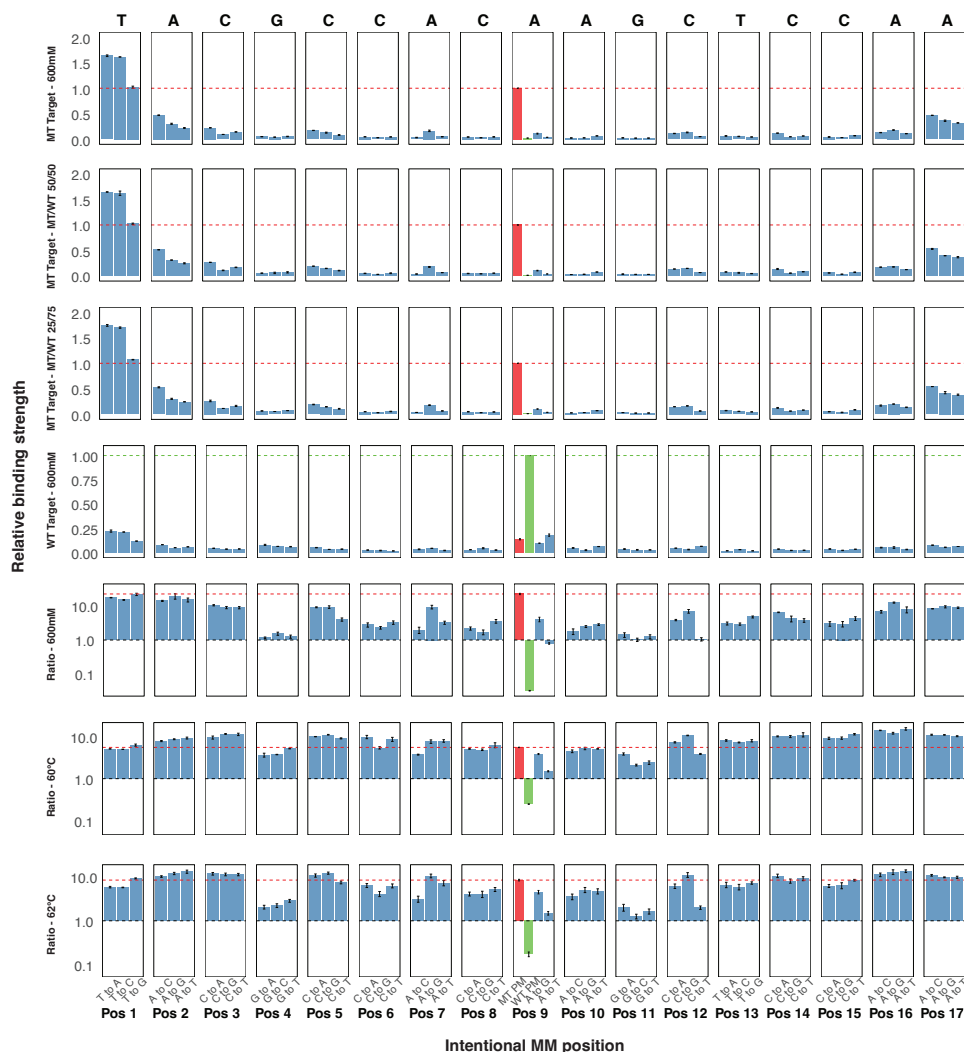


Figure S7: Average results from two independent pulldown experiments using the 884-probe KRAS G12C pool. A set of 52 probes containing the KRAS p.G12C variant centrally positioned (position 9) are shown. Results are presented for the MT target with 600mM salt, MT/WT ratio of 50/50, MT/WT ratio of 25/75, and for the WT target without MT 600mM salt. In addition, the normalized MT/WT count ratio is shown in log scale for 52 probes carrying the KRAS p.G12C variant centrally positioned (position 9). Experiments were performed under different hybridization conditions: 68 °C, 600 mM NaCl; 60 °C, 1000 mM NaCl; and 62 °C, 1000 mM NaCl. In all panels, red denotes probes perfectly matched to the MT target, green marks WT-perfect matches, and blue indicates probes carrying an additional substitution relative to the MT target. The X-axis shows the position and type of substitution.

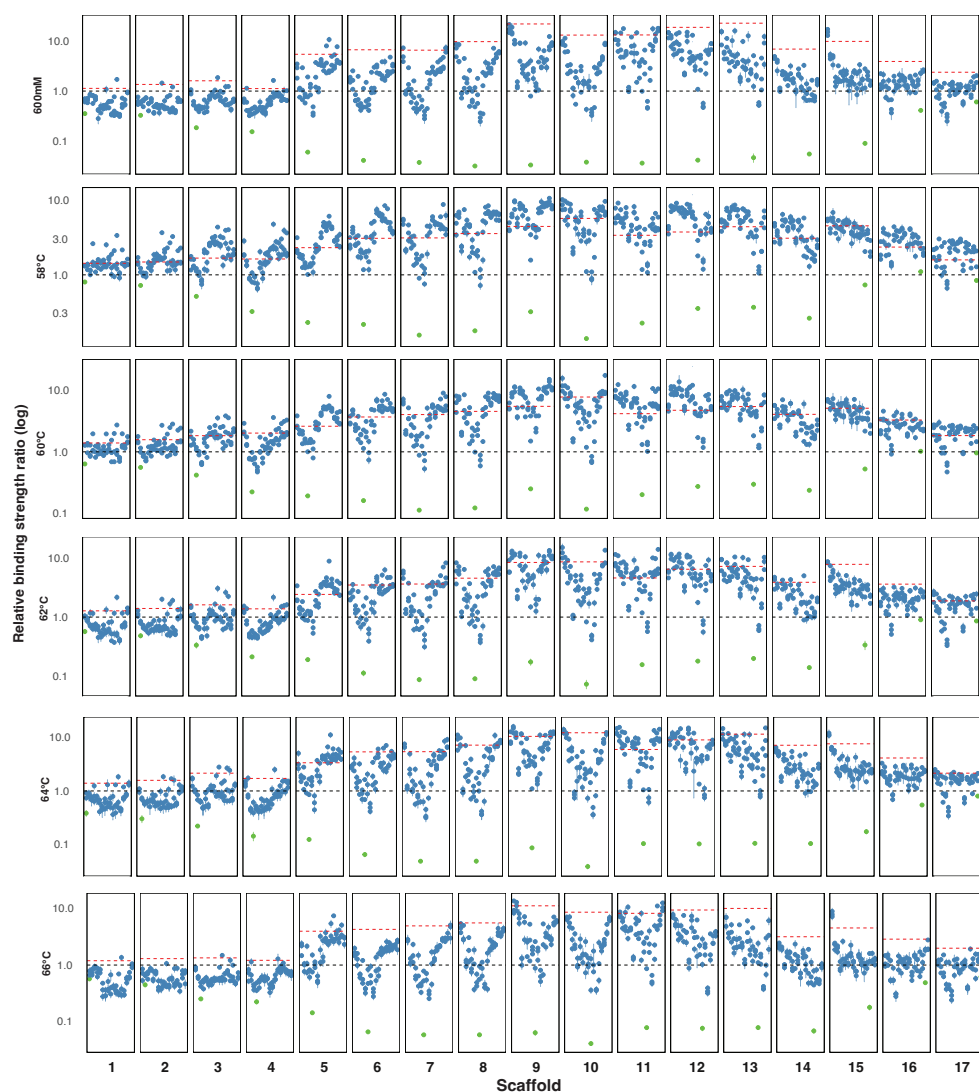


Figure S8: Average results from two independent pulldown experiments using the 884-probe KRAS G12C pool against MT and WT targets at 600mM and a temperature range 58-66°C with 1000mM NaCl. The normalized MT/WT count ratio is shown in log scale for the complete 884-probe KRAS p.G12C pool. In all panels, red denotes probes perfectly matched to the MT target, green marks WT-perfect matches, and blue indicates probes carrying an additional substitution relative to the MT target. The X-axis shows the scaffold number.

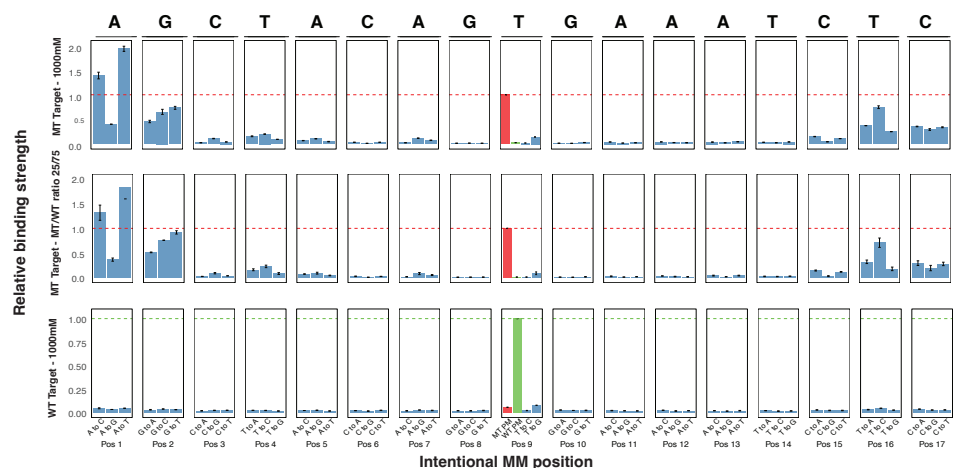


Figure S9: Average results from two independent pulldown experiments using the 884-probe BRAF V600E pool. A set of 52 probes containing the BRAF V600E variant centrally positioned (position 9) are shown. Results are presented for the MT target with 1000mM salt, MT/WT ratio of 25/75 and for the WT target. In all panels, red denotes probes perfectly matched to the MT target, green marks WT-perfect matches, and blue indicates probes carrying an additional substitution relative to the MT target. The X-axis shows the position and type of substitution.

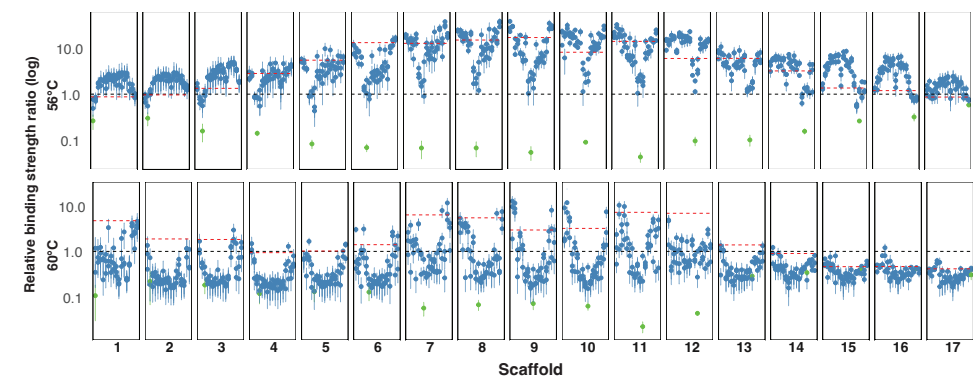


Figure S10: Average results from two independent pulldown experiments using the 884-probe BRAF V600E pool. The normalized MT/WT count ratio is shown in log scale for the complete 884-probe BRAF V600E pool at 56°C and 64°C. In all panels, red denotes probes perfectly matched to the MT target, green marks WT-perfect matches, and blue indicates probes carrying an additional substitution relative to the MT target. The X-axis shows the scaffold number.

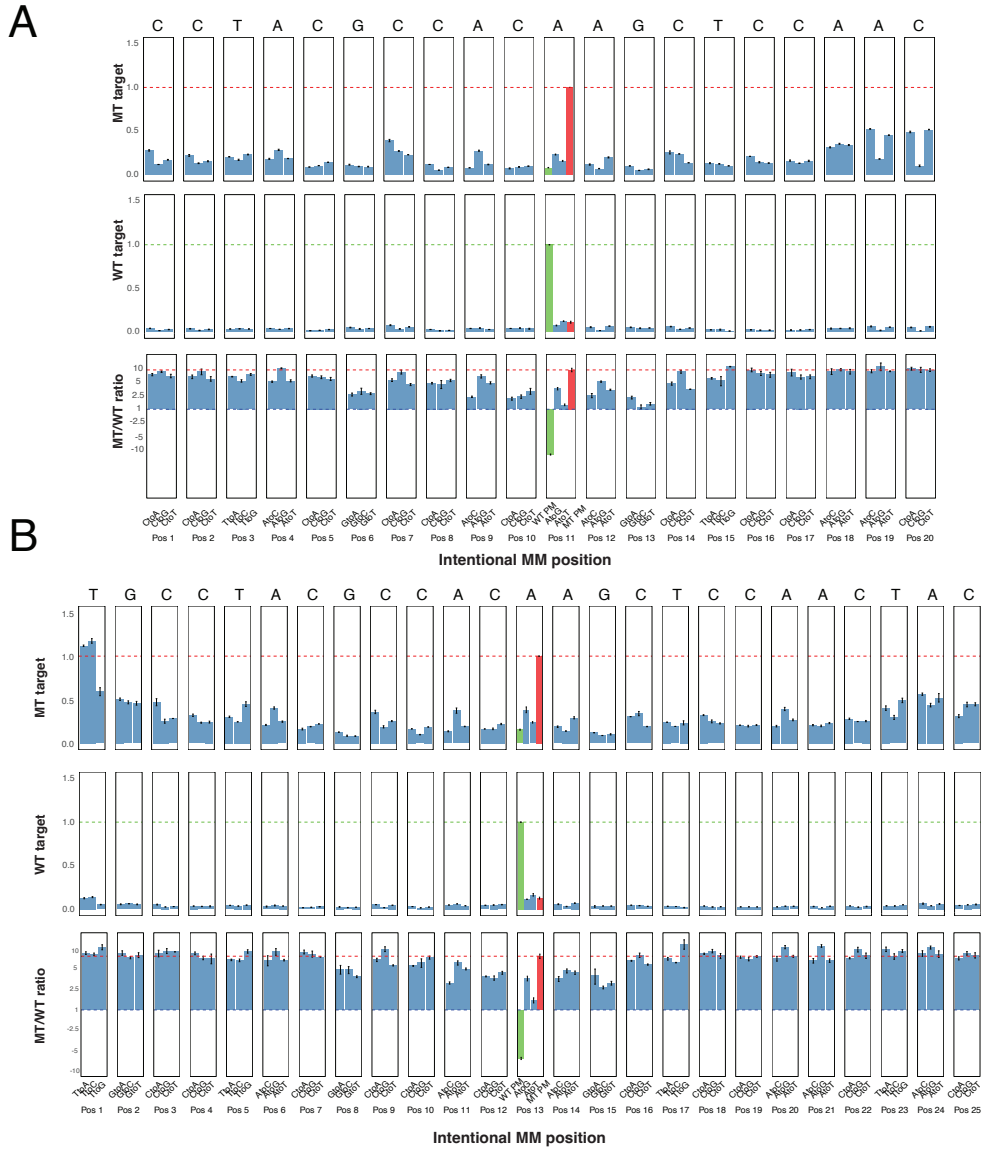


Figure S11: Average results from two independent pulldown experiments using 20 and 25nt long probe designs for the KRAS G12C variant. At In all panels, red denotes probes perfectly matched to the MT target, green marks WT-perfect matches, and blue indicates probes carrying an additional substitution relative to the MT target. The X-axis shows the scaffold number.

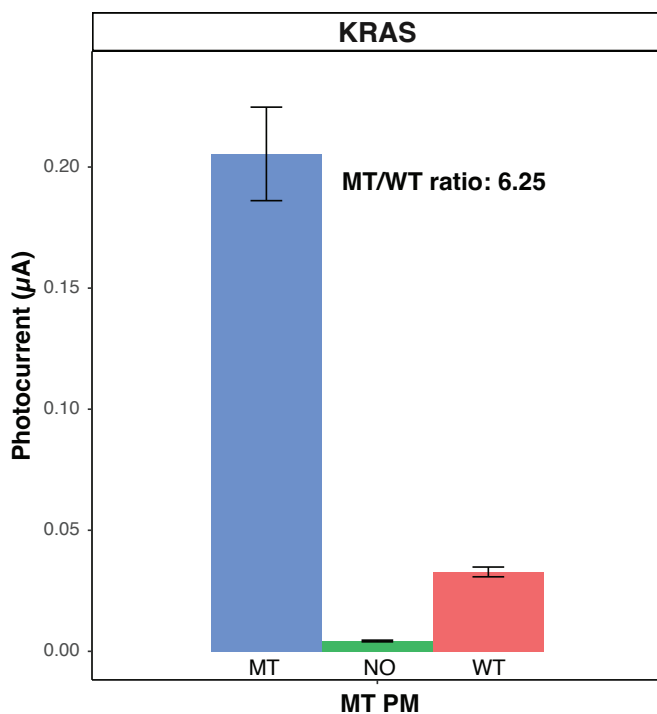


Figure S12: Chronoamperometric responses were recorded for the KRAS p.G12C targets (blue), WT target (red), and no-target control (green) following hybridization at 68°C KRAS in buffer containing 0.1% Tween-20.

References

- [1] D. Huang, Y. Liu, G. A. Wang, S. Lv, Y. Tan, and F. Li, "Chemical tools for discriminating single nucleotide variants: from design principles to clinical applications," *Chemical Society Reviews*, 2026.
- [2] S. E. Lincoln *et al.*, "One in seven pathogenic variants can be challenging to detect by ngs: an analysis of 450,000 patients with implications for clinical sensitivity and genetic test implementation," *Genetics in Medicine*, vol. 23, pp. 1673–1680, 2021.
- [3] D. Y. Zhang, S. X. Chen, and P. Yin, "Optimizing the specificity of nucleic acid hybridization," *Nature Chemistry*, vol. 4, pp. 208–214, 2012.
- [4] C. Guo, R. Ding, Z. Zhao, J. Guo, and F. Li, "Enrichment strategies for low-abundant single nucleotide mutations," 2025.
- [5] S. Suzuki, N. Ono, C. Furusawa, A. Kashiwagi, and T. Yomo, "Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays," *BMC Genomics*, vol. 8, 2007.
- [6] M. P. Johnson, L. M. Haupt, and L. R. Griffiths, "Locked nucleic acid (lna) single nucleotide polymorphism (snp) genotype analysis and validation using real-time pcr," *Nucleic acids research*, vol. 32, 2004.
- [7] B. S. Gaylord, M. R. Massie, S. C. Feinstein, and G. C. Bazan, "SNP detection using peptide nucleic acid probes and conjugated polymers: Applications in neurodegenerative disease identification," *Proceedings of the National Academy of Sciences*, vol. 102, no. 1, pp. 34–39, 2005.

- [8] R. Van Hoof, M. Szymonik, S. K. Nomidis, K. Hollanders, A. Jacobs, I. Nelissen, P. Wagner, and J. Hooyberghs, "Depletion of wild-type target enhances the hybridization-based sensitivity of low-abundant mutation detection by reference capture probes," *Sensors and Actuators B: Chemical*, vol. 368, 2022.
- [9] V. V. Demidov and M. D. Frank-Kamenetskii, "Two sides of the coin: Affinity and specificity of nucleic acid interactions," pp. 62–71, 2004.
- [10] S. W. Shin, C. Baek, and J. Min, "Mismatch-introduced dna probes constructed on the basis of thermodynamic analysis enable the discrimination of single nucleotide variants," *Analytical and Bioanalytical Chemistry*, vol. 414, pp. 5337–5345, 2022.
- [11] S. Lefever, F. Pattyn, J. Hellemans, and J. Vandesompele, "Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative pcr assays," *Clinical chemistry*, vol. 59, pp. 1470–1480, 2013.
- [12] J. SantaLucia and D. Hicks, "The thermodynamics of dna structural motifs," pp. 415–440, 2004.
- [13] J. S. Wang and D. Y. Zhang, "Simulation-guided dna probe design for consistently ultraspecific hybridization," *Nature Chemistry*, vol. 7, pp. 545–553, 2015.
- [14] I. Lee, A. A. Dombkowski, and B. D. Athey, "Guidelines for incorporating non-perfectly matched oligonucleotides into target-specific hybridization probes for a dna microarray," *Nucleic Acids Research*, vol. 32, pp. 681–690, 2004.
- [15] I. Lee, S. S. Ajay, H. Chen, A. Maruyama, N. Wang, M. G. McInnis, and B. D. Athey, "Discriminating single-base difference mirna expressions using microarray probe design guru (prodeg)," *Nucleic Acids Research*, vol. 36, 2008.
- [16] E. Daems *et al.*, "Singlet oxygen-based photoelectrochemical detection of single-point mutations in the kras oncogene," *Biosensors and Bioelectronics*, p. 115957, 2023.
- [17] G. Gydush *et al.*, "Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth," *Nature Biomedical Engineering*, vol. 6, pp. 257–266, 2022.

- [18] S. T. Shanmugam, A. Steijlen, D. Laurijssen, R. Campos, J. Steckel, W. Daems, S. Bassini, E. Daems, and K. D. Wael, "A 96-well led array for multiplexed photoelectrochemical detection of nucleic acids," *Analytical Chemistry*, 2024.
- [19] J. C. Araque and M. A. Robert, "Lattice model of oligonucleotide hybridization in solution. ii. specificity and cooperativity," *Journal of Chemical Physics*, vol. 144, 2016.
- [20] J. Hooyberghs, P. V. Hummelen, and E. Carlon, "The effects of mismatches on hybridization in dna microarrays: Determination of nearest neighbor parameters," *Nucleic Acids Research*, vol. 37, 2009.
- [21] W. W. Hadiwikarta, J. C. Walter, J. Hooyberghs, and E. Carlon, "Probing hybridization parameters from microarray experiments: Nearest-neighbor model and beyond," *Nucleic Acids Research*, vol. 40, 2012.
- [22] G. Rossetti, P. D. Dans, I. Gomez-Pinto, I. Ivani, C. Gonzalez, and M. Orozco, "The structural impact of dna mismatches," *Nucleic Acids Research*, vol. 43, pp. 4309–4321, 2015.
- [23] J. Santalucia, "A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics," *Tech. Rep.*, 1998.
- [24] N. E. Watkins, W. J. Kennelly, M. J. Tsay, A. Tuin, L. Swenson, H. R. Lee, S. Morosyuk, D. A. Hicks, and J. Santalucia, "Thermodynamic contributions of single internal rada, rcdc, rgdg and rudt mismatches in rna/dna duplexes," *Nucleic Acids Research*, vol. 39, pp. 1894–1902, 2011.
- [25] I. I. Cisse, H. Kim, and T. Ha, "A rule of seven in watson-crick base-pairing of mismatched sequences," *Nature Structural and Molecular Biology*, vol. 19, pp. 623–627, 2012.
- [26] H. Willems, A. Jacobs, W. W. Hadiwikarta, T. Venken, D. Valkenborg, N. V. Roy, J. Vandesompele, and J. Hooyberghs, "Thermodynamic framework to assess low abundance dna mutation detection by hybridization," *PLoS ONE*, vol. 12, 2017.
- [27] A. G. Singal, K. N. Kurtzman, and M. J. Thompson, "Leveraging multi-cancer blood tests to improve diagnostic efficiency for patients with nonspecific signs and symptoms," pp. 2603–2607, 2024.

Chapter 5

Kinetic Inversion in Multi-stranded DNA Biosensors

Abstract

Hybridization-based sensors are often optimized in terms of equilibrium selectivity, but the path toward equilibrium can also determine how well two targets are distinguished. Here, kinetic discrimination is analyzed by comparing one-step and two-step hybridization mechanisms. In a simple one-step process, mismatched targets equilibrate faster because their lower affinity is associated with faster dissociation. For two-step detection, however, saturation effects can reverse this behavior: the matched target reaches equilibrium first, while the mismatched signal continues to increase. This kinetic inversion creates an optimal readout time before equilibrium is reached. Wild-type blockers further strengthen the effect by selectively reducing the wild-type target available for probe binding. Experimental OWL2 measurements support these theoretical predictions.

5.1 Introduction

Previous chapters of this thesis have focused exclusively on DNA hybridization at equilibrium. However, the non-equilibrium dynamics of hybridization systems also contain valuable information that can be exploited for target discrimination [1, 2]. In contrast to equilibrium, where the final concentrations are determined solely by the ratio of reaction rates, kinetic measurements track complex formation over time, providing access to individual rates as additional information [3]. This opens the possibility of distinguishing mutant from wild-type targets beyond differences in equilibrium affinity alone.

In single-nucleotide variant (SNV) hybridization sensors, selectivity usually arises from differences in binding affinity between the matched (mutant) and mismatched (wild-type) targets. These systems are typically described by a single-step (two-state) binding process, characterized by an association rate (k^+) and a dissociation rate (k^-). Experimentally, it is well established that mismatches have little effect on the association rate, while significantly increasing the dissociation rate [4, 5]. As a result, mismatched duplexes not only exhibit lower equilibrium stability but are also expected to reach equilibrium more rapidly.

However, this expectation does not hold for more complex hybridization systems. When studying the kinetics of the multi-stranded OWL sensor presented in [Paper B](#), we observed the opposite behavior: the matched (mutant) target equilibrates faster than the mismatched (wild-type) target. This observation is counterintuitive, as higher stability is typically associated with longer residence times and slower relaxation. We attribute this inversion to the presence of multi-step reaction mechanisms operating under specific conditions of saturation or depletion. Here, the effective kinetics are no longer determined solely by the dissociation rate of a single duplex, but instead emerge from the rates of multiple reaction steps.

Interestingly, similar behavior was reported in earlier work that used multi-stranded detection under non-equilibrium conditions, where improved selectivity was achieved through faster equilibration of matched targets [6]. This phenomenon was termed kinetic inversion, but the origin of the inversion was neither formalized mathematically nor clearly understood.

In this chapter, we develop a theoretical framework to explain this effect. We begin by analyzing the kinetics of a simple one-step binding model, which serves as a reference. We then extend the analysis to a two-step process representative of the multi-stranded OWL sensor [7]. This approach reveals how, under specific conditions, selectivity shifts from being governed by differences in dissociation rates to differences in effective association rates. As a consequence, the ordering of relaxation times is reversed, with matched systems equilibrating faster than mismatched ones, resulting in kinetic inversion. This behavior naturally defines an optimal time window for mismatch discrimination, which differs from the equilibrium limit.

Finally, we introduce blocker strands and investigate their impact on the system's kinetic behavior. In previous chapters, blockers were shown to improve equilibrium selectivity by increasing effective affinity differences through selective depletion [8, 9]. Here, we extend this analysis and examine how these mechanisms influence the system's kinetics. We find that the presence of blockers not only amplifies equilibrium selectivity but also significantly enhances kinetic discrimination. In particular, the blocker interaction alters the effective reaction rates in the system, leading to a greater separation in relaxation times between matched and mismatched

targets. Consequently, the conditions under which kinetic inversion occurs are broadened, and the inversion itself becomes more pronounced. Building on this theoretical framework, we present preliminary experimental results obtained with the OWL sensor platform introduced in Chapter 3. These results establish a general framework for understanding kinetics in multi-step hybridization processes and extend naturally to enzymatic approaches used for single-nucleotide variant (SNV) discrimination [10, 11]. Moreover, they suggest new strategies that exploit kinetic inversion to enhance mismatch detection sensitivity.

5.2 Kinetics of the one-step process

We study the hybridization of a basic one-step process in which probe strands P bind to target strands T , forming a duplex PT as shown in the reaction diagram below.



As previously described in section 1.3, dynamics are described in a mean-field approach assuming mass-action kinetics, where reaction rates are proportional to concentrations of reactants, and forward/reverse rate constants are k^+ and k^- . Together, they determine the binding equilibrium constant K , which is related to the free-energy difference $\Delta G < 0$ between the bound and unbound state. We assume the system starts in an initial state in which only ssDNA is present. The duplex concentration $[PT]$ then evolves according to the general kinetic equation:

$$\frac{d[PT]}{dt} = k^+[P][T] - k^-[PT], \quad (5.2)$$

where available concentrations of probes $[P]$ and targets $[T]$ might change during the reaction and deviate from their initial values $[P]_0 = a$ and $[T]_0 = c$. At any time, they are given by

$$[P] = a - [PT], \quad [T] = c - [PT]. \quad (5.3)$$

These result in a decreasing association rate during the binding process and lead to the effects referred to as probe saturation and target depletion.

In summary:

$$\frac{d[PT]}{dt} = k^+ \underbrace{(a - [PT])}_{\text{saturation}} \underbrace{(c - [PT])}_{\text{depletion}} - k^- [PT]. \quad (5.4)$$

$$\left. \begin{array}{l} [PT] = \text{duplex conc.} \\ c = \text{initial target conc.} \\ a = \text{initial probe conc.} \end{array} \right\} \begin{array}{l} k^- = \text{off-rate} \\ k^+ = \text{on-rate} \end{array} \Rightarrow K = \frac{k^+}{k^-} = e^{-\Delta G/RT} \quad (5.5)$$

Saturation: occupied **probes** no longer contribute to duplex formation

Depletion: captured **target** no longer contributes to duplex formation

This is a nonlinear ordinary differential equation (ODE) that is solved after rewriting the ODE in terms of its fixed points $h_{\pm}$, leading to the complete solution for the kinetics [8]:

$$[PT](t) = \frac{h_+ h_- (1 - e^{-t/\tau})}{h_+ - h_- e^{-t/\tau}}, \quad \text{with } \tau \equiv [k^+ (h_+ - h_-)]^{-1}, \quad (5.6)$$

$$h_{\pm} = \frac{1}{2K} \left[1 + aK + cK \pm \sqrt{(1 + aK + cK)^2 - 4acK^2} \right] \quad (5.7)$$

Although this does not describe a simple exponential process, the timescale is given by a single characteristic value τ . At times beyond $t \gg \tau$, the solution converges to h_- , which is a stable fixed point, whereas h_+ is unstable. The equilibrium solution is therefore $[PT]_{\text{eq}} \equiv [PT](t \rightarrow \infty) = h_-$. While this description provides a general solution for the behavior of this one-step system, the expressions are complicated, and interpretation is quite difficult. We will therefore consider simplified cases in which either saturation or depletion is taken into account. To start, however, we analyze the reservoir condition, where both saturation and depletion can be neglected.

Reservoir condition: τ only depends on k^-

Under reservoir conditions, available ssDNA concentration remains at its initial values such that probe saturation and target depletion do not occur. This approximation is valid in a low-binding regime, e.g., at low concentrations or low affinities, where the reduction in available ssDNA can be neglected.

The ODE simplifies considerably,

$$\frac{d[PT]}{dt} = k^+ c a - k^- [PT], \quad (5.8)$$

leading to the solution

$$[PT](t) = \underbrace{acK}_{\text{equilibrium}} \underbrace{\left(1 - e^{-k^- t}\right)}_{\text{kinetics}} \quad (5.9)$$

In contrast to the general description from before, this represents simple exponential decay towards the equilibrium solution. The equilibrium concentration is given by acK , consistent with the linear relationship of Eq. (1.6). The time t appears only in the exponent, such that the characteristic time τ is given by the factor in the exponent, $\tau = 1/k^-$. This indicates the system's timescale and serves as a measure of the total time required to reach equilibrium. Specifically, at time $t = \tau$, the system has reached 63% of its final value, and at times $t = 2\tau, 3\tau, 4\tau$ and 5τ , the system is 86%, 95%, 98% and 99% towards equilibrium. Roughly, one can estimate that the system reaches equilibrium after $t \approx 3\tau$. Early in the kinetics, the concentration increases linearly with time by $[PT](t) = k^+ c a t$, which can be seen by approximating the exponent for small $t \ll \tau$ or by setting $[PT] = 0$ in Eq. (5.8).

For SNV detection, we analyze the difference between perfectly matched (PM) and mismatched (MM) targets by comparing their kinetics, as illustrated in Figure 5.1. At equilibrium, the matched duplex reaches a higher equilibrium concentration compared to the mismatched one, which has a lower affinity $K = k^+/k^-$ as a result of the increased dissociation rate k_- . During the kinetics, both cases exhibit the same initial behavior, which is independent of k_- . A difference is visible in the timescales of the two: the mismatched duplex reaches equilibrium faster as a result of the higher k_- , consistent with known behavior of short DNA strands [4].

Although discrimination is possible based on kinetic behavior, the difference between matched and mismatched cases strictly increases over time. Notably, in this simple model, the equilibrium time is independent of on-rate k^+ , and ssDNA concentrations a and c . One way to understand this is as follows: the initial behavior is linear in k^+ , a , and c , but the equilibrium value $k^+ c a / k^-$ is also linear in these. The two effects cancel out, making the necessary time to reach equilibrium independent of $k^+ c a$.

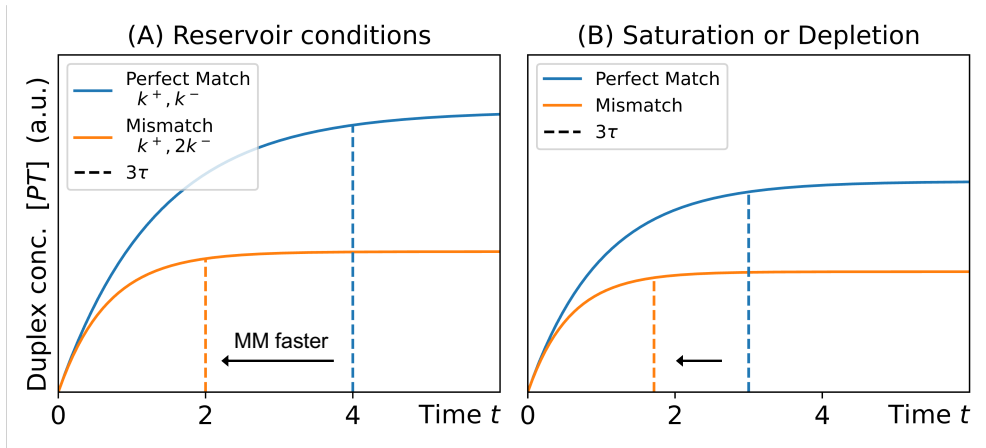


Figure 5.1: Kinetic response of perfectly matched (PM) and mismatched (MM) duplex formation in a one-step process. (A) Under reservoir conditions, the PM system reaches a higher equilibrium and takes longer to reach equilibrium compared to the MM system. (B) Saturation or depletion lead to shorter timescales, that are closer to each other, and lower equilibrium values.

Saturation introduces k^+c in τ_{eq}

The effect of probe saturation lowers the available probe concentration as probes bind to the target, leading to the following description.

$$\frac{d[PT]}{dt} = k^+c(a - [PT]) - k^-[PT], \quad (5.10)$$

$$= k^+ca - (k^- + k^+c)[PT] \quad (5.11)$$

This has the same form as before, and results in the solution

$$[PT](t) = \frac{acK}{1 + cK} \left(1 - e^{-(k^- + k^+c)t} \right) \quad (5.12)$$

Saturation therefore leads to a lower equilibrium value and a timescale that is given by $\tau = (k^- + k^+c)^{-1}$, which is slightly faster. The initial behavior is not affected by saturation, as only a small amount of duplex forms at the start of the reaction. We can now better understand the reduced equilibrium time. While the initial increase remains linear in k^+c , the equilibrium value is lowered by saturation, so it is reached faster with increasing k^+c .

Depletion introduces k^+a in τ_{eq}

Similarly, target depletion results in lower equilibrium and timescales, while leaving the initial kinetics unaffected. The results of this one-step process are summarized in Table 5.1. Kinetic discrimination of SNVs, as with equilibrium discrimination, becomes more difficult due to saturation and depletion, as a change in k^- has a lower impact on the timescale τ .

Case	Initial behavior $t \ll \tau$	Timescale (τ)	Equilibrium
<u>Reservoir</u>			
Saturation ✗	$k^+ c a t$	$1/k^-$	acK
Depletion ✗			
Saturation ✓	$k^+ c a t$	$1/(k^- + k^+ c)$	$\frac{acK}{1 + cK}$
Depletion ✗			
Saturation ✗	$k^+ c a t$	$1/(k^- + k^+ a)$	$\frac{acK}{1 + aK}$
Depletion ✓			

Table 5.1: Kinetic behavior of the one-step process with saturation or depletion.

5.3 Kinetics of the two-step process

The two-step process is inspired by the multi-stranded OWL sensor of **Paper B**. Here, binding of the probe to target can be followed by the binding of a reporter strand, resulting in a target-probe-reporter complex that contributes to the detection signal. A simplified illustration of the detection process is shown in Figure 5.2. The reaction diagram is expanded to include this reporter and two new rate constants, k_R^+ and k_R^- , which also determine the reporter's binding affinity K_R to PT .



Importantly, the reporter is insensitive to whether the probe-target is perfectly matching or contains a mismatch, i.e., k_R^+ and k_R^- do not change when the underlying target is the mutant or wild-type sequence. For our analysis, the reporter rate constants are assumed to be much lower than those of the probe-target interaction. The reporter is present at an initial concentration of $[R]_0 = r$, and when available concentrations are lowered, similar to probe saturation, the second step is subject to reporter saturation: $[R] = r - [PTR]$.



Figure 5.2: Simplified reaction diagram of the two-step process. Only after the selective binding of probe P to target T can the reporter interact with the PT complex to form the detectable, fluorescent PTR complex.

The two-step process is then described by two time-dependent variables, $[PT](t)$ and $[PTR](t)$, that evolve according to the coupled ODEs:

$$\frac{d[PT]}{dt} = k^+ \underbrace{(a - [PT])}_{\text{probe saturation}} \underbrace{(c - [PT])}_{\text{target depletion}} - k^- [PT] - k_R^+ [PT] \underbrace{(r - [PTR])}_{\text{reporter saturation}} + k_R^- [PTR], \quad (5.15)$$

$$\frac{d[PTR]}{dt} = k_R^+ [PT] \underbrace{(r - [PTR])}_{\text{reporter saturation}} - k_R^- [PTR] \quad (5.16)$$

$[PT]$ = duplex conc. $[PTR]$ = target-probe-reporter complex conc. c = initial target conc. a = initial probe conc. r = initial reporter conc.	k^- = off-rate of probe k^+ = on-rate of probe $\Rightarrow K = \frac{k^+}{k^-} = e^{-\Delta G/RT}$ k_R^- = off-rate of reporter k_R^+ = on-rate of reporter $\Rightarrow K_R = \frac{k_R^+}{k_R^-} = e^{-\Delta G_R/RT}$
---	--

(5.17)

Saturation: occupied **probes** no longer contribute to duplex formation

Depletion: captured **target** no longer contributes to duplex formation

Reporter saturation: occupied **reporter** no longer contributes to target-probe-reporter complex formation

Reservoir condition: no kinetic inversion

Once again, the analysis starts with the case where P , T and R form a reservoir and saturation/depletion effects can be neglected. Moreover, in the low-binding regime $[PTR] \ll [PT]$, the reporter does not influence the probe-target system and the ODEs decouple. In this case, substituting the one-step solution of Eq. (5.9) into Eq. (5.16) (without reporter saturation) leads to solution

$$[PTR](t) = acKrK_R \left[1 - \frac{e^{-k^-t}}{1 - k^-/k_R^-} - \frac{e^{-k_R^-t}}{1 - k_R^-/k^-} \right]. \quad (5.18)$$

At equilibrium, the exponents decay and the complex concentration is given by $acKrK_R$, now additionally linear in r and K_R . The two-step system has two timescales that can be extracted from the exponents:

$$\tau_P = 1/k^- \quad \& \quad \tau_R = 1/k_R^-, \quad (5.19)$$

that obey $\tau \ll \tau_R$ by our assumption. The exponents can be approximated, leading to a quadratic regime followed by a linear regime.

$$[PTR](t) = \frac{1}{2}acrk_R^+ k_R^+ t^2 \quad \text{for } t \ll \tau \ll \tau_R \quad (5.20)$$

$$[PTR](t) = acKrK_R^+ t \quad \text{for } \tau \ll t \ll \tau_R \quad (5.21)$$

Comparing the responses of matched and mismatched targets, differing only in the off-rate k^- of the middle step, reveals the effect on SNV detection (Figure 5.3). As k^+ is the same for both cases, the initial quadratic part is identical for the mutant and wild-type targets, but the signal transitions to the linear regime earlier for the mismatch than for the match. Interestingly, in the linear region, the signals rise with a slope that is dependent on the probe-target affinity K . Finally, equilibrium is reached on the reporter's timescale, and, as with the one-step process without depletion or saturation, the equilibrium time is fully determined by its off-rate. Because $\tau_R \gg \tau$, the difference in probe affinity does not change the total equilibrium time, and no kinetic inversion is present under the current conditions. When significant reporter binding occurs outside the low-binding regime, the probe-target system is influenced by reporter binding, and its equilibrium shifts to the right. The timescales τ_P and τ_R mix and the equilibrium time increases from $1/k_R^-$ to $1/k_R^- (1 + rk_R^+/k^-)$. Although this introduces k^- into the expression, the matching target is still slower and no kinetic inversion takes place.

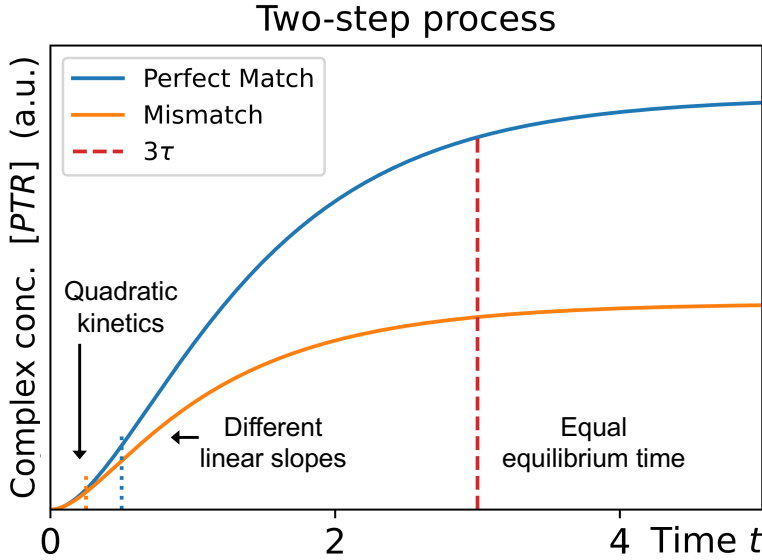


Figure 5.3: Two-step hybridization kinetics under reservoir conditions. Both perfect match and mismatch follow identical quadratic behavior, followed by a linear regime with slopes determined by probe affinity. The mismatch transitions earlier and grows more slowly. Equilibrium is reached on the reporter timescale $1/k_R^+$, and is identical for both cases, demonstrating the absence of kinetic inversion.

Reporter saturation: kinetic inversion

To facilitate our further analysis, we note that the P+T step is much faster than the R interaction. This leads to a separation of timescales (SOT) such that the reporter kinetics can be described by assuming quasi-steady-state behavior of the first step. Although this description fails to describe kinetics at the timescale of the first step, early kinetics are still accurately predicted by Eq. (5.20) since reservoir conditions hold at short timescales. In general, SOT reduces the complexity considerably, leading to more interpretable results.

Under low-binding conditions of the reporter, the first step is completely determined by the equilibrium description of section 5.2. In the simplest case, $[PT] = acK$, and inserting into Eq. (5.16) leads to the solution with reporter saturation:

$$[PTR](t) = \frac{acKrK_R}{1 + acKrK_R} \left[1 - e^{-(k_R^- + acKk_R^+)t} \right] \quad (5.22)$$

This result is straightforwardly adapted to cases where the probe-target system is subject to probe saturation or target depletion by substituting $[PT]$ with the corresponding equilibrium solution. As expected, the equilibrium is lower compared to that under reservoir conditions of Eq. (5.18). As a result of the SOT, only a single timescale is present (Figure 5.4) and the time to reach equilibrium is characterized by

$$\tau = \frac{1}{k_R^- + a c K k_R^+}. \quad (5.23)$$

Similar to the results of the one-step system, reporter saturation introduces an additional term to the characteristic time τ . Notably, the probe-target affinity K is present in this expression, such that τ changes for a PM or MM target. We observe kinetic inversion: higher affinity lowers τ and therefore decreases the total equilibrium time. In contrast to the one-step process, the PM situation equilibrates faster than the MM situation.

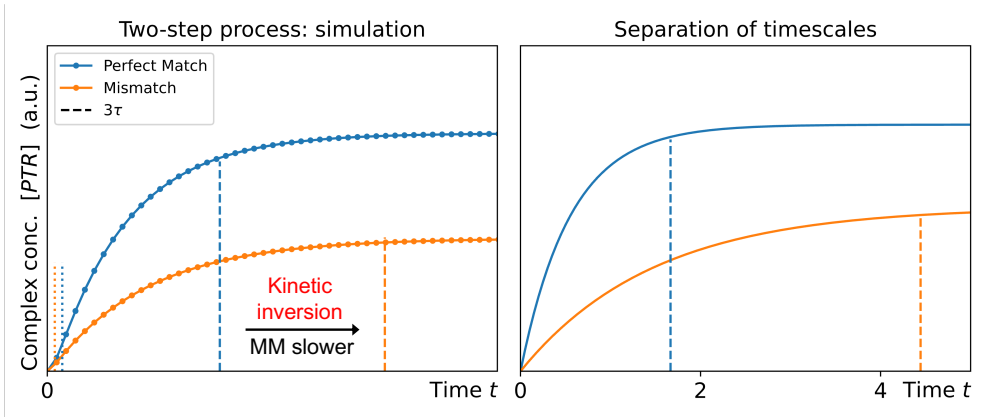


Figure 5.4: Kinetic inversion in the two-step system with reporter saturation. Left: Simulation of the system shows that the characteristic time depends on affinity and that the perfect match reaches equilibrium faster than the mismatched case. Right: Theoretical model of Eq. (5.22) employing the separation of timescales approach, where information on early timescales is lost, but a description including kinetic inversion remains.

5.3.1 Optimal discrimination time

Kinetic inversion has a direct implication concerning the discrimination of SNVs. Specifically, the optimal time to distinguish between matching and mismatched targets occurs during the kinetics rather than at equilibrium.

Once the matching target reaches equilibrium, its signal plateaus, while the mismatched target (being further from equilibrium) continues to increase (Figure 5.5). As a result, the difference between the matched and mismatched targets decreases only when the systems are allowed to fully equilibrate. In the regime before τ , the response can be approximated by

$$[PTR](t) = a c K r k_M^+ t \quad \text{for } t \ll \tau. \quad (5.24)$$

The initial behavior is linear and proportional to K , such that the ratio between PM and MM targets is constant and related to the free-energy binding difference $\Delta\Delta G$.

$$\frac{[PTR]^{PM}(t)}{[PTR]^{MM}(t)} = \frac{K^{PM}}{K^{MM}} = e^{-\Delta\Delta G/RT} \quad (5.25)$$

The ratio between PM and MM responses further reduces after $t = \tau_{PM}$, indicating an optimal discrimination time just before τ_{PM} where the sensitivity is largest. The optimal time is easily illustrated by comparing the kinetics of matching and mismatched bindings in the presence of a background contribution (Figure 5.5). This ratio is non-monotone and has a maximum value at a finite time, whereas without kinetic inversion the best time to discriminate is at $t = \infty$, i.e., equilibrium.

An appropriate trade-off between selectivity and sensitivity is provided by the fold-change parameter β , which compares the desired PM signal to the undesired background and MM contribution [12].

$$\beta(t) = \frac{([PTR]^{PM}/\text{VAF})}{\text{background} + [PTR]^{MM}}. \quad (5.26)$$

This formula was originally introduced for mixtures at a given VAF, but also works when comparing two sample responses containing only PM or MM targets. The background can represent all unavoidable sources of signal as a result of experimental artifacts, but it can also represent the minimal signal required by the detection mechanism.

The definition of β makes it a useful measure for the trade-off between selectivity and sensitivity. In the limit where the background is zero, β reduces to the usual PM/MM ratio, and therefore describes selectivity. When the background is non-zero, however, a low response will reduce β because the background term remains present in the denominator. In this way, β also accounts for sensitivity. Kinetic inversion does not necessarily

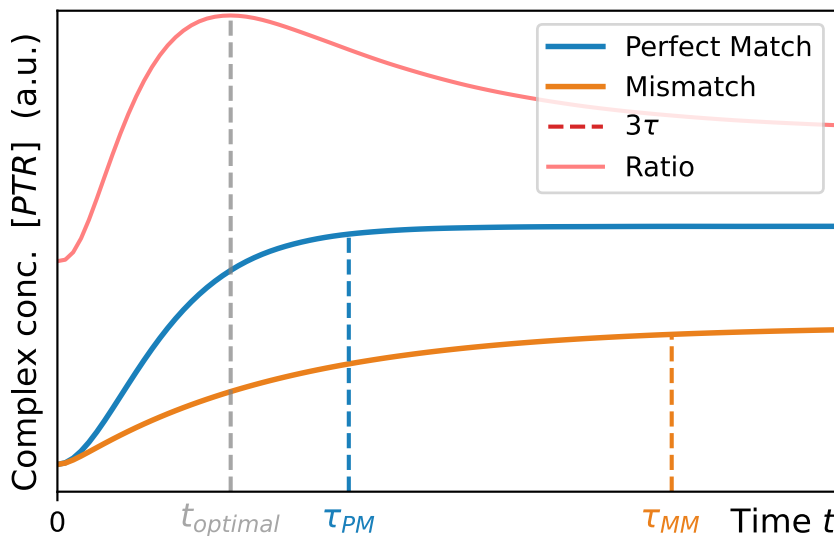


Figure 5.5: Optimal discrimination time in a system with kinetic inversion and a background contribution. The matching signal saturates earlier, while the mismatch continues to increase, leading to optimal discrimination at finite time.

improve the intrinsic selectivity at the level of $\Delta\Delta G$, which remains determined by the PM/MM affinity difference. Instead, it improves the practical trade-off between selectivity and sensitivity by allowing discrimination at a finite time, before equilibrium saturation becomes limiting. Higher probe concentrations can therefore be used to accelerate the kinetics and increase the signal, while retaining a high PM/MM ratio in the optimal time window.

5.3.2 Two-step process differentiates in on-rate

The reporter molecule, binding after the selective probe-target interaction, introduced a second, non-selective step. Under reservoir conditions, the characteristic timescale is determined solely by k_R^- . However, a k_R^+ dependence appears when reporter saturation takes place. This behavior mirrors that of the one-step process, where the characteristic timescale is determined by k^- and additionally by $c k^+$ in the presence of probe saturation. In the two-step system, the concentration in front of k_R^+ therefore

corresponds to the equilibrium concentration of the probe-target complex, which differs significantly between matched and mismatched binding. A major consequence of this is that the difference in k^- of the target is present in the on-rate of the reporter, whereas it was present in the off-rate in the one-step process. The two-step process therefore translates the difference in the off-rate of the selective probe, which differs between the targets, into a difference in the on-rate of the reporter. This effectively required saturation to enable discrimination based on the equilibrium timescale. Simultaneously, this also maximizes the magnitude of the response and allows differentiation early in the kinetics of existing detection mechanisms using a two-step process.

- One-step process: Affinity differentiation in off-rate:

$$\tau_{PM} = \frac{1}{k_{PM}^-} > \tau_{MM} = \frac{1}{k_{MM}^-} \quad (5.27)$$

- Two-step process: Affinity differentiation in on-rate:

$$\tau_{PM} = \frac{1}{k_R^- + a c K_{PM} k_R^+} < \tau_{MM} = \frac{1}{k_R^- + a c K_{MM} k_R^+} \quad (5.28)$$

5.4 Enhanced Kinetic Inversion using Wild-type Blockers

The effect of wild-type target depletion previously resulted in improved detection limits at equilibrium by increasing the effective free-energy penalty $\Delta\Delta G_{\text{eff}}$ for mutant versus wild-type binding. Wild-type target depletion was achieved by introducing a large number of short blocker strands complementary to the wild-type sequence, which prevent binding of the mutant-matching probes. The mutant target will also cross-hybridize with the blocker, so the target concentration is increased to maintain a sufficiently high concentration to interact with probes. Here, a short description is presented of the impact of blocker strands on the kinetics of the two-step process and its implications for kinetic inversion.

Blocker B introduces an additional interaction to the two-step process, the formation of BT , to which the reporter cannot bind. Although blocker

strands could run out, they are usually supplied in excess concentration $[B]_0 = b \gg c$, such that blocker saturation can be neglected. Importantly, blocker contributes to the depletion of target, and its high affinity to wild-type leads to a selective reduction of wild-type over mutant target. Under conditions of *strong depletion*, the effective target concentration is fully determined by the blocker interaction and given by $c^* = c/bK_B$. In summary:

$$\begin{aligned} B + T &\rightleftharpoons BT & \Rightarrow & \frac{d[BT]}{dt} = k_B^+ b [T] - k_B^- [BT], \\ & & & [T] = c - [PT] - [BT] \\ & & & \approx \frac{c}{bK_B} \end{aligned} \quad (5.29)$$

The blocker introduces a third timescale, τ_B , that we will assume is similar to that of the probe, τ_P , and much smaller than τ_R . We describe the initial kinetics and the changes to the characteristic time due to the blocker interaction. Initial formation of $[PTR]$ is given by

$$[PTR](t) = \frac{1}{6} a c k^+ r k_R^+ k_B^- t^3 \quad \text{for } t \ll \tau_B \approx \tau_P \ll \tau_R, \quad (5.30)$$

where, interestingly, the off-rate of the blocker k_B^- is present and the rise is cubic in time. Blocker strands perfectly match the wild-type and have a single mismatch against the mutant target, such that k_B^- is larger for the mutant target. This leads to an early difference in kinetics compared to the system without blocker, where the initial behavior was quadratic and the same for both targets.

The SOT approach leads to the kinetic solution with a similar form as in Eq. (5.22), now given by

$$[PTR](t) = \frac{a c^* K r K_R}{1 + a c^* K r K_R} \left[1 - e^{-\left(k_R^- + a \frac{c}{b K_B} K k_R^+ \right) t} \right]. \quad (5.31)$$

From this expression, the timescale is subsequently given by

$$\tau = \frac{1}{k_R^- + \frac{a c K k_R^+}{b K_B}}. \quad (5.32)$$

Here, kinetic inversion is once again present. In addition to probe affinity K , blocker affinity K_B is present reciprocally. The difference in timescale between mutant (PM to probe, MM to blocker) and wild-type (MM to probe,

PM to blocker) is therefore amplified by the addition of blockers (Figure 5.6). One might think that the blocker concentration b lowers the difference in timescales between the two, as the k_R^- term becomes more important, but this is counteracted by the required increase of target concentration c .

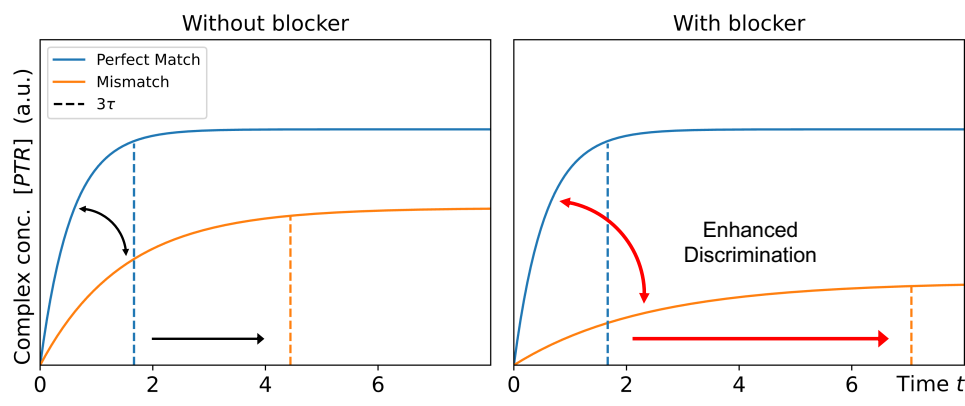


Figure 5.6: The addition of wild-type blockers enhances kinetic inversion. Within the separation-of-timescales framework, blocker-induced depletion amplifies the timescale difference between matching and mismatched cases, highlighting its benefit for both equilibrium and kinetic discrimination.

5.5 Preliminary Experimental Results

A limited number of experiments were conducted to investigate kinetic inversion using the multi-stranded OWL2 sensor involving a two-step detection mechanism. The OWL2 sensor contains long target-binding arms that unfold the target's secondary structure and allow the R-strand, P-strand, and universal molecular beacon (UMB) to interact with the target in its unfolded state [13]. The two-step process involves the binding of the selective P-strand, followed by the opening of the UMB. When the UMB opens, a fluorophore and quencher become spatially separated, resulting in a fluorescent readout. To isolate the interaction of the two-step process, the target was incubated overnight with the unfolding arms and the R-strand. The kinetic measurements are then performed by first adding UMB, and then adding P-strand to initiate the two-step process.

Figure 5.7 shows the fluorescent response of the pure mutant target and the pure wild-type target over 60 min. The initial signal is quadratic in time, followed by a linear increase with a slope that differs between mutant and wild-type. As the mutant signal approaches its equilibrium value, the wild-type signal is still linearly increasing. The time to reach equilibrium for the matching target is lower than that of the mismatched target, and kinetic inversion occurs. As a result, the ratio between the mutant and wild-type signals shows an optimal time to distinguish the SNVs.

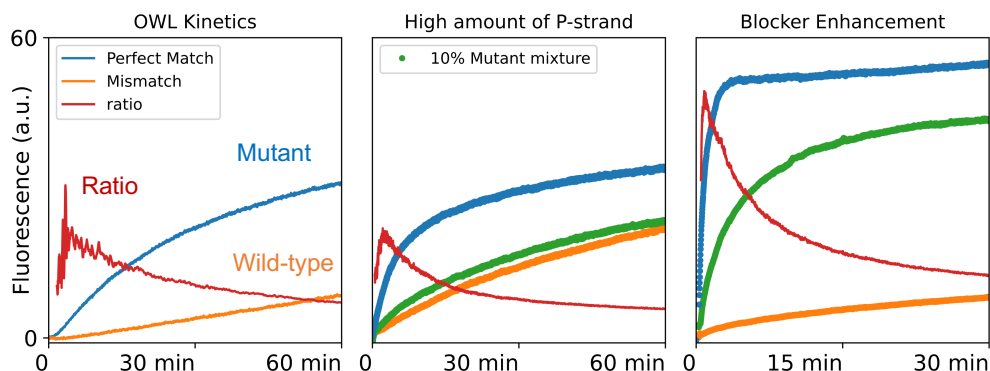


Figure 5.7: Experimental validation of kinetic inversion using the OWL2 sensor. Fluorescence kinetics show quadratic-to-linear behavior, with mutant reaching equilibrium faster than wild-type, leading to kinetic inversion and an optimal discrimination time. Increasing the P-strand concentration accelerates the timescales, while the addition of wild-type blockers enhances slope and timescale differences.

The measurement was repeated using a 10x higher amount of P-strand. This results in a much larger formation of the intermediate complex to which the UMB can bind, increasing the reporter saturation. As before, the mutant signal is observed to increase and reach equilibrium faster than the wild-type signal, and as a result of the increased P-strand concentration, the overall timescales have become lower, as predicted by Eq. (5.23). Data for a mixture containing 10% mutant is also presented, showing an intermediate slope and equilibrium time. Finally, the OWL2 sensor kinetics were measured in the presence of wild-type blockers. Target concentrations were increased, and the expected result is observed: a larger difference in the slopes in the linear part, and a larger difference in equilibrium times between matching, mismatched, and even mixed targets. The initial kinetics are very fast and the cubic behavior is difficult to detect.

Although these experiments remain preliminary, they provide useful guidance for future work. They show that kinetic inversion can emerge in a realistic multi-stranded sensor with reporter saturation. The broader insight is that multi-step hybridization systems can convert an affinity difference in an early selective step into a difference in the effective rate of a later detection step. Kinetic design may therefore improve the practical trade-off between sensitivity and selectivity.

Further development requires a more systematic exploration of the design space. Theoretically, the hybridization system can be adjusted in terms of concentrations, reaction rates and binding affinities. These make simulation-guided approaches useful for identifying conditions where kinetic inversion maximally improves performance. Experimentally, future work should focus on precise control of hybridization rates and early kinetic measurements. For this, toehold-mediated strand displacement is a promising model system, because it provides tunable and predictable multi-step kinetics [14]. Finally, the analysis should be extended to low-abundance mixtures to test whether kinetic measurements can improve SNV discrimination beyond equilibrium limits.

References

- [1] B. Ashwood and A. Tokmakoff, "Kinetics and dynamics of oligonucleotide hybridization," *Nature Reviews Chemistry*, vol. 9, no. 5, pp. 305–327, 2025.
- [2] J. K.-K. Ng, H. Feng, and W.-T. Liu, "Rapid discrimination of single-nucleotide mismatches using a microfluidic device with monolayered beads," *Analytica Chimica Acta*, vol. 582, no. 2, pp. 295–303, 2007.
- [3] K. Tawa, "Mismatching base-pair dependence of the kinetics of DNA-DNA hybridization studied by surface plasmon fluorescence spectroscopy," *Nucleic Acids Research*, vol. 32, no. 8, pp. 2372–2377, 2004.
- [4] M. Glazer, J. A. Fidanza, G. H. McGall, M. O. Trulson, J. E. Forman, A. Suseno, and C. W. Frank, "Kinetics of oligonucleotide hybridization to photolithographically patterned DNA arrays," *Analytical Biochemistry*, vol. 358, no. 2, pp. 225–238, 2006.
- [5] M. Gotoh, Y. Hasegawa, Y. Shinohara, M. Shimizu, and M. Tosu, "A New Approach to Determine the Effect of Mismatches on Kinetic Parameters in DNA Hybridization Using an Optical Biosensor," *DNA Research*, vol. 2, no. 6, pp. 285–293, 1995.
- [6] M. Stancescu, T. A. Fedotova, J. Hooyberghs, A. Balaeff, and D. M. Kolpashchikov, "Nonequilibrium Hybridization Enables Discrimination of a Point Mutation within 5–40 °C," *Journal of the American Chemical Society*, vol. 138, no. 41, pp. 13 465–13 468, 2016.
- [7] R. J. Karadeema, M. Stancescu, T. P. Steidl, S. C. Bertot, and D. M. Kolpashchikov, "The owl sensor: a 'fragile' DNA nanostructure for the

- analysis of single nucleotide variations," *Nanoscale*, vol. 10, no. 21, pp. 10 116–10 122, 2018.
- [8] S. K. Nomidis, M. Szymonik, T. Venken, E. Carlon, and J. Hooyberghs, "Enhancing the Performance of DNA Surface-Hybridization Biosensors through Target Depletion," *Langmuir*, vol. 35, no. 37, pp. 12 276–12 283, 2019.
- [9] Y. Stulens, R. Van Hoof, K. Hollanders, I. Nelissen, M. Szymonik, P. Wagner, G. Froyen, B. Maes, and J. Hooyberghs, "Hybridization-based sensor with large dynamic range for detection of circulating tumor DNA in clinical samples," *Biosensors and Bioelectronics*, vol. 281, p. 117342, 2025.
- [10] M. Cui, H. Feng, D. Guo, D. Wang, and B. Zheng, "A droplet-based microfluidic platform for kinetics-based detection of single nucleotide variation at room temperature with large discrimination factors," *Sensors and Actuators B: Chemical*, vol. 253, pp. 731–737, 2017.
- [11] S. X. Chen and G. Seelig, "An Engineered Kinetic Amplification Mechanism for Single Nucleotide Variant Discrimination by DNA Hybridization Probes," *Journal of the American Chemical Society*, vol. 138, no. 15, pp. 5076–5086, 2016.
- [12] J. S. Wang and D. Y. Zhang, "Simulation-guided DNA probe design for consistently ultraspecific hybridization," *Nature Chemistry*, vol. 7, no. 7, pp. 545–553, 2015.
- [13] B. L. Mueller, M. J. Liberman, and D. M. Kolpashchikov, "OWL2: a molecular beacon-based nanostructure for highly selective detection of single-nucleotide variations in folded nucleic acids," *Nanoscale*, vol. 15, no. 12, pp. 5735–5742, 2023.
- [14] D. Y. Zhang and E. Winfree, "Control of DNA Strand Displacement Kinetics Using Toehold Exchange," *Journal of the American Chemical Society*, vol. 131, no. 47, pp. 17 303–17 314, 2009.

Chapter 6

Conclusions & Outlook

6.1 Thermodynamics as a design principle

This thesis has explored the detection of single-nucleotide variants (SNVs) through a unified approach based on DNA thermodynamics. Rather than focusing on a specific biosensing technology or readout mechanism, the central idea has been that molecular recognition and selectivity are fundamentally determined by hybridization behavior, and that manipulating this behavior provides a general route to improved performance. A key strength of this approach is its platform independence. The thermodynamic principles that govern DNA hybridization are universal and do not depend on whether the readout is optical, electrochemical, or sequencing-based. This abstraction allowed us to develop strategies that can be transferred across a wide range of biosensing technologies. At the same time, we demonstrated these concepts using emerging, cost-effective sensing platforms, which further highlight their practical relevance for point-of-care diagnostics and scalable applications.

Improvements start from a detailed understanding of DNA hybridization. We described the thermodynamics of duplex formation and established that, under optimal conditions, selectivity is determined by the free-energy difference between the mutant and wild-type bindings, $\Delta\Delta G$. Chapters 2, 3 and 4 present three research papers, along with three approaches that increase the effective free-energy difference. These pushed detection limits to clinically relevant levels or provided valuable insight into properties that help guide future sensor design. Kinetic properties of hybridization were discussed in Chapter 5, and a kinetic method based on periodic perturbations was briefly mentioned in the introduction (Paper D and the patent application).

Across the studies presented in this thesis, a consistent theme emerged: thermodynamic optimization enables both enhanced sensitivity and improved selectivity without increasing the complexity of the detection mechanism. Moreover, hybridization-based technologies are inherently well-suited for parallelization, making these design principles readily adaptable to multiplexed platforms for the simultaneous detection of multiple SNVs. While the primary focus has been on biosensing applications, the results also provide broader insights into the fundamental properties of selec-

tive DNA hybridization. They contribute to the development of a more quantitative and predictive framework for understanding and controlling hybridization behavior in the context of SNV detection.

6.2 Selectivity improvement beyond $\Delta\Delta G$

A central result of this work is that the detection limit of mutant DNA in a wild-type background depends exponentially on the free-energy difference, $\Delta\Delta G$. Together with the noise of the readout mechanism, this directly determines the lowest detectable mutant fraction. $\Delta\Delta G$, therefore, is the key parameter that determines selectivity, and maximizing it is essential for low-abundance mutant detection. While $\Delta\Delta G$ is traditionally limited by the intrinsic penalty of a single mismatch (MM), we showed that its effective value can be increased through several thermodynamics-based strategies that act independently and can be synergistic. Three such mechanisms were investigated in detail:

- Wild-type depletion via blocker strands
- Structural strain in probe-target complexes
- Non-additive mismatch-mismatch interactions

Each mechanism contributes to an increase in $\Delta\Delta G$, but does so through distinct physical origins. Wild-type depletion selectively reduces the available concentration of targets, thereby amplifying the relative contribution of mutant binding. In contrast, structural strain and MM-MM interactions introduce asymmetric destabilization that preferentially affects wild-type duplexes.

An important insight of this thesis is that these strategies are not mutually exclusive. On the contrary, their combination leads to enhanced performance beyond what can be achieved individually. For example, the use of blocker strands allows strain-based selectivity improvements to be more effectively expressed. This can be extended: structural strain or intentional mismatches could be introduced into the blocker design to further

increase depletion efficiency. These results point toward a framework of thermodynamic design, in which multiple strategies are combined depending on the sensing context. Because DNA hybridization underlies amplification and sequencing technologies, these improvements have the potential to achieve robust discrimination at ultra-low mutant fractions [1, 2].

6.3 Depletion as a versatile strategy

Wild-type depletion emerged as a particularly versatile strategy. Beyond its primary role in enhancing selectivity, it allows the use of higher probe concentrations without compromising discrimination, as supported by both theory and experimental results.

This has several implications. First, an important experimental advantage: using higher probe concentrations improves signal robustness and reduces noise, leading to more reliable signals and more accurate determination of the amount of mutant DNA. Second, they enable more precise quantification of thermodynamic parameters, facilitating the extraction of $\Delta\Delta G$ from experimental data. In this sense, wild-type depletion is not only a practical tool for improving detection, but also a methodological tool for studying hybridization thermodynamics in more complex sensors.

Wild-type blockers, however, reduce the readout signal, requiring higher target concentrations to remain above background. While increased probe concentrations partially compensate for this effect, balancing sensitivity and selectivity within diagnostically relevant concentration ranges remains a challenge [3]. Future work should therefore focus on a detailed exploration of the concentrations involved. Optimizing probe, blocker, and target concentrations and affinities could allow the design to be tuned for specific diagnostic regimes. Examples include sensors that are maximally sensitive in a narrow range of low mutant fractions, relevant for early detection, or that improve quantitative accuracy at higher abundances for treatment monitoring.

6.4 Strain and Mismatch interactions: a unified perspective?

The improved performance of the OWL sensor was attributed to structural strain present in the selective probe-target binding. Formation of the fluorescence complex requires the probe to bend in a specific way, bringing its termini into close proximity for stacking interactions. Because such stacking depends on a well-defined rotational orientation, this suggests torsional strain within the duplex that propagates to the selective binding site and preferentially destabilizes the mismatched binding, leading to an increased $\Delta\Delta G$. Although the structural origin of this effect is reasonably well understood [4], it remains difficult to identify the precise design requirements needed to achieve improved SNV discrimination. While the OWL sensor provides a highly selective, cost-effective platform with relative detection limits as low as 0.058%, it is less suited for isolating the molecular contributions of strain. The OWL2 architecture, for instance, relies on long unfolding arms and an additional intermediate strand that must also bend. These components are essential to the sensor's operation, but they introduce additional interactions that may influence the strain at the selective binding site, complicating a more direct interpretation.

Similarly, $\Delta\Delta G$ was increased by the introduction of intentional mismatches in the probe design, revealing hybridization thermodynamics beyond the nearest-neighbor (NN) model. Rather than behaving as independent contributions, mismatch-mismatch (MM-MM) interactions were found to be strongly non-additive and highly position-dependent. In particular, we observed a transition from stabilizing to destabilizing interactions as a mismatch approaches the duplex termini. These observations indicate that mismatches not only introduce local perturbations but also influence the global structure and stability of the duplex. The dependence on both mismatch position and type further indicates that the resulting affinities are determined by a complex interplay of structural distortion, base-stacking interactions, and conformational flexibility [5]. While our results established general design principles, such as the advantage of placing an intentional mismatch near the termini, the optimal mismatch type remains unclear and lacks a unifying framework. It is not yet understood whether the optimal type mismatch depends on the specific SNV mismatch involved,

nor whether the observed effects are primarily driven by enthalpic or entropic contributions. Additionally, it remains to be determined whether these effects arise from structural perturbations or from the instability of short binding segments. These questions are currently being addressed in follow-up work, in which sequences covering all possible mismatch types are being investigated to describe the underlying physical mechanisms of these effects.

A useful perspective emerges by comparing MM-MM interactions with structural strain in the OWL sensor. Although the mechanisms underlying selective destabilization differ, the phenomenology is similar: both introduce perturbations that propagate along the duplex and modify its stability nonlocally. This raises the question of whether MM-MM interactions and structural strain are fundamentally distinct or can be understood within a common framework.

One possible unifying view is to study the DNA duplex as a mechanically coupled system, in which disruptions, whether caused by mismatches or external constraints, influence stability at remote base-pairs. Such an approach could isolate general principles and conditions required for the amplification of mismatch penalties. This perspective also explains why these effects are context-dependent, as the propagation of structural perturbations depends on sequence composition, duplex length, and boundary conditions. These systems can be studied using several approaches. Coarse-grained models, such as spring-based representations of DNA, could identify essential features and help determine whether the destabilizing interactions are primarily enthalpic or entropic [6, 7]. More detailed computational methods, such as molecular dynamics simulations, can further investigate the contributions of steric effects, torsional strain and whether perturbations propagate along the backbone or through adjacent base-stacking interactions [8, 9]. Experimentally, torque can be applied to DNA duplexes using magnetic optical tweezers, enabling direct investigation of how externally imposed mechanical strain influences mismatch stability [10]. Alternatively, strategies that are designed to introduce strain, such as the deliberate formation of secondary structures or DNA-protein interactions, provide additional routes to probe these effects and isolate their physical origin [11, 12]. Several of these approaches are currently being pursued within our research group.

In general, establishing a common framework for MM-MM interactions and structural strain has important implications for probe design. Instead of separate tools, they could be turned into generalized concepts, where local sequence modifications or structural constraints maximize selectivity and result in a physics-based optimization of hybridization probe design.

6.5 Beyond equilibrium

While this thesis has primarily focused on equilibrium thermodynamics, Chapter 5 introduced hybridization kinetics as an alternative route to selectivity. In particular, the concept of kinetic inversion, where mismatched systems take longer to equilibrate than matched ones, provided a fundamentally different perspective on mutant versus wild-type discrimination.

In simple one-step hybridization, discrimination arises from differences in off-rate: mismatched duplexes equilibrate faster because of their reduced stability. In multi-step systems such as the OWL sensor, an additional reporter binding step changes this picture. As the reporter interacts with the probe-target complexes, discrimination shifts from the probe's off-rate to the reporter's effective on-rate. When saturation effects become relevant, the overall equilibration time can depend on the probe-target affinity itself, leading to an inversion of equilibrium times.

Unlike equilibrium approaches, where selectivity is determined solely by $\Delta\Delta G$, kinetic strategies depend on rate constants and reaction mechanisms. This introduces a larger parameter space, but also additional design flexibility. An important question is therefore: under which conditions do kinetic effects provide a clear advantage over equilibrium-based optimization? In particular, kinetic inversion may enable trade-offs between sensitivity and selectivity that are not accessible at equilibrium.

Further development requires both theoretical and experimental advances. Solving kinetics with all relevant interactions is theoretically challenging, making computational approaches better suited to identify conditions where kinetic effects improve performance [3, 13]. Extending current models to include additional reaction steps or strands may further amplify these

effects. Experimentally, precise control over hybridization rates and time-resolved measurements is essential. A promising hybridization system for studying these kinetics is toehold-mediated strand displacement, as it provides tunable, predictable reaction kinetics and inherently involves multiple steps [14]. This makes it well-suited for studying and evaluating kinetic design principles. The analysis should also be extended to mixtures, particularly for low-abundance mutant detection, to evaluate whether kinetics enables improved discrimination beyond equilibrium limits.

6.6 Reflection on the past four years: theory and experiment

One of the defining aspects of my PhD has been the continuous exchange between theoretical modeling and experimental research. As a physicist working primarily on theory, many ideas begin as concepts or thought experiments. However, their feasibility and true value only become clear through experimental validation.

At the same time, experimental researchers often develop strong intuition through hands-on experience, and theoretical modeling helps translate these insights into quantitative and predictive models. In this way, theory and experiment continuously strengthen one another.

This project also involved multiple disciplines, including physics, chemistry, and medicine, which highlighted the importance of communication. Bridging the differences in scientific language and perspective between these fields was sometimes challenging but, in the end, highly rewarding. The exchange of ideas across disciplines improved the quality of our work and opened new directions for future research.

Overall, the combination of theory and experiment, physics and other disciplines, proved more powerful than either approach alone: the whole was truly greater than the sum of its parts. I am particularly pleased that these collaborations will continue beyond this project, allowing future PhD students in our research group to benefit from the same close interaction with other research teams.

References

- [1] C. Guo, R. Ding, Z. Zhao, J. Guo, and F. Li, "Enrichment Strategies for Low-Abundant Single Nucleotide Mutations," *Chemistry - A European Journal*, vol. 31, no. 1, p. e202402872, 2025.
- [2] D. Khodakov, C. Wang, and D. Y. Zhang, "Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches," *Advanced Drug Delivery Reviews*, vol. 105, pp. 3–19, 2016.
- [3] J. S. Wang and D. Y. Zhang, "Simulation-guided DNA probe design for consistently ultraspecific hybridization," *Nature Chemistry*, vol. 7, no. 7, pp. 545–553, 2015.
- [4] R. J. Karadeema, M. Stancescu, T. P. Steidl, S. C. Bertot, and D. M. Kolpashchikov, "The owl sensor: a 'fragile' DNA nanostructure for the analysis of single nucleotide variations," *Nanoscale*, vol. 10, no. 21, pp. 10 116–10 122, 2018.
- [5] G. Rossetti, P. D. Dans, I. Gomez-Pinto, I. Ivani, C. Gonzalez, and M. Orozco, "The structural impact of DNA mismatches," *Nucleic Acids Research*, vol. 43, no. 8, pp. 4309–4321, 2015.
- [6] B. E. K. Snodin *et al.*, "Introducing improved structural properties and salt dependence into a coarse-grained model of DNA," *The Journal of Chemical Physics*, vol. 142, no. 23, p. 234901, 2015.
- [7] D. Petkevičiūtė, M. Pasi, O. Gonzalez, and J. Maddocks, "cgDNA: a software package for the prediction of sequence-dependent coarse-grain free energies of B-form DNA," *Nucleic Acids Research*, vol. 42, no. 20, pp. e153–e153, 2014.

- [8] R. Lavery *et al.*, "A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA," *Nucleic Acids Research*, vol. 38, no. 1, pp. 299–313, 2010.
- [9] R. J. Isaacs and H. P. Spielmann, "Insight into G–T Mismatch Recognition Using Molecular Dynamics with Time-Averaged Restraints Derived from NMR Spectroscopy," *Journal of the American Chemical Society*, vol. 126, no. 2, pp. 583–590, 2004.
- [10] I. Heller, T. P. Hoekstra, G. A. King, E. J. G. Peterman, and G. J. L. Wuite, "Optical Tweezers Analysis of DNA-Protein Complexes," *Chemical Reviews*, vol. 114, no. 6, pp. 3087–3119, 2014.
- [11] G. Rosenblum, N. Elad, H. Rozenberg, F. Wiggers, J. Jungwirth, and H. Hofmann, "Allostery through DNA drives phenotype switching," *Nature Communications*, vol. 12, no. 1, p. 2967, 2021.
- [12] S. Kim *et al.*, "Probing Allostery Through DNA," *Science*, vol. 339, no. 6121, pp. 816–819, 2013.
- [13] S. X. Chen and G. Seelig, "An Engineered Kinetic Amplification Mechanism for Single Nucleotide Variant Discrimination by DNA Hybridization Probes," *Journal of the American Chemical Society*, vol. 138, no. 15, pp. 5076–5086, 2016.
- [14] D. Y. Zhang and E. Winfree, "Control of DNA Strand Displacement Kinetics Using Toehold Exchange," *Journal of the American Chemical Society*, vol. 131, no. 47, pp. 17 303–17 314, 2009.



www.uhasselt.be
Hasselt University
Martelarenlaan 42 | BE-3500 Hasselt