Mesoscopic Models for RNA Salt Dependence and for Oncogene Probe Design with Locked Nucleic Acids

PhD Thesis

Izabela Ferreira da Silva
Departamento de Física
Universidade Federal de Minas Gerais

Belo Horizonte, Tuesday 21st September, 2021

Mesoscopic Models for RNA Salt Dependence and for Oncogene Probe Design with Locked Nucleic Acids

Doctoral thesis presented to the Interunit Graduate Program in Bioinformatics of the Institute of Biological Sciences of the Universidade Federal de Minas Gerais as a requirement for obtaining the degree of PhD in Bioinformatics.

Advisor: Prof.Dr. Gerald Weber
Co-Advisor: Prof.Dr[a] Kira Astakhova

PhD Thesis

Izabela Ferreira da Silva
Departamento de Física
Universidade Federal de Minas Gerais

Belo Horizonte, Tuesday 21st September, 2021

043 Silva, Izabela Ferreira da.
Mesoscopic models for RNA salt dependence and for oncogene probe design with locked nucleic acids [manuscrito] / Izabela Ferreira da Silva. – 2021.
159 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Gerald Weber. Coorientadora: Prof. Dra Kira Astakhova.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação Interunidades em Bioinformática.

1. Biologia Computacional. 2. DNA. 3. RNA. 4. Modelos Teóricos. 5. Desnaturação de Ácido Nucleico. 6. Ligações de Hidrogênio. I. Weber, Gerald. II. Astakhova, Kira. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 - 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

ATA DE DEFESA DE TESE

IZABELA FERREIRA DA SILVA

Às nove horas do dia **21 de setembro de 2021**, reuniu-se, através de videoconferência, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Mesoscopic Models for RNA Salt Dependence and for Oncogene Probe Design with Locked Nucleic Acids"**, requisito para obtenção do grau de Doutora em **Bioinformática.** Abrindo a sessão, o Presidente da Comissão, **Dr. Gerald Weber**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Professor(a)/Pesquisador(a)	**Instituição**	**Indicação**
Dr. Gerald Weber - Orientador	Universidade Federal de Minas Gerais	**Aprovada**
Dra. Mariana Torquato Quezado de Magalhaes	Universidade Federal de Minas Gerais	**Aprovada**
Dra. Simone Silva Alexandre	Universidade Federal de Minas Gerais	**Aprovada**
Dra. Luciana Márcia de Oliveira	Institut national de recherche pour l'agriculture/France	**Aprovada**
Dr. Anderson de Sá Pinheiro	Universidade Federal do Rio de Janeiro	**Aprovada**

Pelas indicações, a candidata foi considerada: **Aprovada**

O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 21 de setembro de 2021.

Documento assinado eletronicamente por **Gerald Weber**, **Professor do Magistério Superior**, em 24/09/2021, às 11:52, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Luciana Márcia de Oliveira**, **Usuário Externo**, em 24/09/2021, às 11:53, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Anderson de Sá Pinheiro**, **Usuário Externo**, em 24/09/2021, às 12:54, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Mariana Torquato Quezado de Magalhaes**, **Professora do Magistério Superior**, em 24/09/2021, às 14:16, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Simone Silva Alexandre**, **Professora do Magistério Superior**, em 27/09/2021, às 09:39, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0984156** e o código CRC **530C4524**.

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

PARECER Nº **30/2021**
PROCESSO Nº 23072.250301/2021-77

FOLHA DE APROVAÇÃO

"Mesoscopic Models for RNA Salt Dependence and for Oncogene Probe Design with Locked Nucleic Acids"

Izabela Ferreira da Silva

Tese aprovada pela banca examinadora constituída pelos Professores:

Prof. Gerald Weber - Orientador
Universidade Federal de Minas Gerais

Profa. Mariana Torquato Quezado de Magalhães
Universidade Federal de Minas Gerais

Profa. Simone Silva Alexandre
Universidade Federal de Minas Gerais

Profa. Luciana Márcia de Oliveira
Institut national de recherche pour l'agriculture/France

Prof. Anderson de Sá Pinheiro
Universidade Federal do Rio de Janeiro

Belo Horizonte, 21 de setembro de 2021.

Documento assinado eletronicamente por **Gerald Weber**, **Professor do Magistério Superior**, em 24/09/2021, às 11:52, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Luciana Márcia de Oliveira**, **Usuário Externo**, em 24/09/2021, às 11:53, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Anderson de Sá Pinheiro**, **Usuário Externo**, em 24/09/2021, às 12:54, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Mariana Torquato Quezado de Magalhaes**, **Professora do Magistério Superior**, em 24/09/2021, às 14:16, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Simone Silva Alexandre**, **Professora do Magistério Superior**, em 27/09/2021, às 09:39, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0984165** e o código CRC **BD410D09**.

Project Identification

Title: Mesoscopic Models for RNA Salt Dependence and for Oncogene Probe Design with Locked Nucleic Acids

Key Words: Theoretical Models for DNA, H-bond in DNA Macromolecule, Morse potentials, DNA mesoscopic models, modified DNA, RNA, Locked Nucleic Acids.

Knowledge area: Molecular biophysics, bioinformatics, statistical physics.

Institution: Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte-MG.

Acknowledgements

My most sincere and profound gratitude goes to my advisor, Gerald Weber, that has been present in every moment and encouraged me to keep on going and persisting in times that I completely did not want to, this could have never been possible without you. I could not ever let to mention and be profoundly grateful to my co-advisor, Kira Astakhova, for all the incentives and guidance during this period.

To my parents, Juracy and Antônio, who have been supportive and understanding of my life choices, especially to the whole PhD pathway, I am thoroughly indebted.

I would like to acknowledge and be deeply grateful to every single teacher I had during this breathtaking journey called academic life, I could have never got this far without them. Also to all the friends I have been acquainted over the years and that kept being sympathetic to my crazy schedule. To the *Grupo de Biofísica Computacional* and the colleagues over the years, thank you for the great times.

I am also grateful to the institutions, *Universidade Federal de Minas Gerais*, **UFMG**, and *Danmarks Tekniske Universitet*, **DTU** for their thoughtful reception and support. To the *Programa Interunidades de Pós-graduação em Bioinformática* for all the kind attention and help during this path.

Lastly, I am indebted to the agencies whose have provided financial support — *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*, **CAPES** and to the *Fundação de Amparo à Pesquisa do Estado de Minas Gerais*, **Fapemig**. Special acknowledgements to the *Programa de Internacionalização* CAPES/PrInt UFMG which have conceded the necessary funding for my visit to DTU.

"In each moment the fire rages, it will burn away a hundred veils and carry you a thousand steps toward your goal."

Rumi

Abstract

The melting process of nucleic acids is suggested to be a first-order transition and has been described by different approaches over the years. One such approach is the Peyrard-Bishop (PB) model which mainly consists of describing the helix through the hydrogen bonds connecting each base pair and the intra-helix stacking interaction between adjacent bases. The PB model exhibited success and experimental accordance when applied to modified nucleic acids, such as Inosine and threose nucleic acid (TNA). Its computational feasibility and capability to derive intra-molecular parameters from melting temperatures provide us with a robust tool to reinterpret published experimental data and achieve new insights over hydrogen bonds and stacking interactions in oligonucleotides. Moreover, the PB model parameterization also allows us to use those derived parameters to predict the melting temperature of non-measured sequences.

One missing parameterization for the PB model is RNA at low salt concentrations, due to the limited amount of published melting temperatures. Although the PB model was found to be largely independent of strand concentrations, it requires that all temperatures are provided at the same strand concentrations. We adapted the PB model to handle multiple strand concentrations and in this way, we were able to make use of an experimental set of temperatures to model the hydrogen bond and stacking interactions at low and intermediate sodium concentrations. For the parameterizations, we make a distinction between terminal and internal base pairs, and the resulting potentials were qualitatively similar as we obtained previously for DNA. The main difference from DNA parameters, was the Morse potentials at low sodium concentrations for terminal r(AU) which is stronger than d(AT), suggesting higher hydrogen-bond strength.

Another open problem is the source of the stabilization provided by chemically modified base pairs. One such modification is the locked nucleic acid (LNA), which due to the methylene bridge addition in the sugar moiety mimics the conformation of an RNA helix. This change improves the overall stability of the helix and has been used in several applications such as in polymerase chain reaction (PCR), small interfering RNA (siRNA), and antisense oligonucleotides (ASOs). However, the source of the stability improvement is not clearly delineated yet. Studies have suggested a favorable change either in the entropy, enthalpy, or both, of the modified helix, which is mainly driven by an enhancement in the stacking interactions of neighboring base pairs. The major challenge lies in compile sets containing sufficient measurements of melting temperature at similar buffer conditions and in the model parameterization. Fortunately, due to LNA popularity, we were able to collect a data set of over 300 temperature measurements.

We have derived a complete set of parameters for the LNA insertion in DNA sequences and contrarily from the previous assumptions we have found stronger hydrogen bonding in the modified base pairs and their stacking interactions have shown little change. A complete parameterization of LNA base pairs allows the optimization of their use in oncogene probes and other types of applications. Therefore, we used the parameters to predict all possible LNA insertions in oncogene variants of *BRAF*, *KRAS* and *EGFR*. The probes were selected from the pool of temperature predictions, synthesized and their melting temperature measured, the accuracy of the measurements with the predictions was of 1 °C.

Resumo

O fenômeno de desnaturação dos ácidos nucléicos é sugerido como uma transição de primeira ordem e foi descrito por diferentes abordagens ao longo dos anos. Uma dessas abordagens é o modelo Peyrard-Bishop (PB), que consiste principalmente em descrever a dupla hélice através das ligações de hidrogênio que conectam cada par de base e a interação de empilhamento intra-hélice entre pares de base adjacentes. O modelo PB obteve sucesso e concordância experimental quando aplicado à ácidos nucleicos modificados, como Inosina e TNA. A sua viabilidade computacional e capacidade de derivar parâmetros intramoleculares de temperaturas de desnaturação provê-nos uma ferramenta robusta para reinterpretar dados experimentais publicados e obter novas percepções sobre ligações de hidrogênio e interações de empilhamento em oligonucleotídeos. Além disso, também podemos usar esses parâmetros para prever a temperatura de desnaturação de sequências que não foram medidas anteriormente.

Uma parametrização em falta para o modelo PB é para RNA em baixas concentrações de sal, devido à quantidade limitada de temperaturas de desnaturação publicadas. Embora o modelo PB tenha sido amplamente independente das concentrações de fita em parametrizações anteriores, requer-se que todas as temperaturas sejam fornecidas na mesma concentração de fita. Assim, adaptamos o modelo PB para lidar com múltiplas concentrações de fita e, dessa forma, pudemos fazer uso de um conjunto experimental de temperaturas para modelar a ligação de hidrogênio e as interações de empilhamento em concentrações baixas e intermediárias de sódio. Para as parametrizações, fizemos uma distinção entre pares de bases terminais e internos, e os potenciais resultantes foram qualitativamente semelhantes aos obtidos anteriormente para DNA. A principal diferença em relação aos parâmetros de DNA foi o potencial de Morse em baixas concentrações de sódio para o terminal r(AU), que é mais forte do que d(AT), sugerindo maior intensidade da ligação de hidrogênio.

Outro problema em aberto é a fonte da estabilização fornecida por pares de bases quimicamente modificados. Uma dessas modificações é o locked nucleic acid (LNA) que devido à adição da ponte de metileno em seu açúcar, conforma-se semelhantemente à uma hélice de RNA. Essa alteração melhora a estabilidade geral da hélice e têm sido usada em várias aplicações, como polymerase chain reaction (PCR), small interfering RNA (siRNA) e, antisense oligonucleotides (ASOs). No entanto, a fonte da melhoria da estabilidade ainda não é claramente compreendida. Estudos têm sugerido uma mudança favorável na entropia, entalpia ou em ambas, da hélice modificada, que é principalmente relacionada à um aumento nas interações de empilhamento de pares de bases vizinhos. Entretanto, o maior desafio está em compilar conjuntos contendo medidas suficientes de temperatura de desnaturacão em condições experimentais semelhantes e na parametrização do modelo. Felizmente, devido à popularidade do LNA conseguimos coletar um conjunto com mais de 300 medidas de temperatura.

Derivamos um conjunto completo de parâmetros para a inserção única de LNA em sequências de DNA e, ao contrário das suposições anteriores, encontramos ligações de hidrogênio mais fortes nos pares de bases modificados e suas interações de empilhamento mostraram pouca mudança. A parametrização completa dos pares de bases do LNA permite a otimização do seu uso em sondas de oncogene e outros tipos de aplicações. Portanto, nós utilizamos os parâmetros para prever todas as inserções de LNA possíveis em variantes dos oncogenes *BRAF*, *KRAS* e *EGFR*. As sondas foram selecionadas a partir do conjunto de predicões de temperatura, sintetizadas e a temperatura de desnaturacão medida, a precisão das temperaturas medidas comparadas com as temperaturas preditas foi menor que 1 ° C.

List of Figures

List of Acronyms

ASOs antisense oligonucleotides

BRAF v-raf murine sarcoma viral oncogene homolog B1

DNA deoxyribonucleic acid

DSC differential scanning calorimetry

dsDNA double stranded DNA

dsRNA double stranded RNA

EGFR epidermal growth factor receptor

FRET Fluorescence resonance energy transfer

HNA hexitol nucleic acid

KRAS Kirsten rat sarcoma viral oncogene homolog

LAMP loop-mediated isothermal amplification

LNA locked nucleic acid

MD molecular dynamics

miRNA microRNA

mRNA messenger RNA

NMR nuclear magnetic ressonance

NN nearest-neighbour

PB Peyrard-Bishop

PCR polymerase chain reaction

PNA peptide nucleic acid

RNA ribonucleic acid

siRNA small interfering RNA

TI transfer integral operator

TNA threose nucleic acid

UV ultra-violet

WLC worm like chain

Contents

Thesis structure

The thesis is subdivided into five main chapters, of which are, a brief introduction, two chapters presenting the research papers published during the PhD period, a chapter containing an integrative discussion and a last conclusion chapter.

(a) Chapter 1 shows an overview of DNA, RNA and modified nucleic acids. This is followed by a discussion about the thermal denaturation of nucleic acids and the theoretical models that are commonly used. Afterwards, the framework model, used in the published papers is discussed in details.

(b) Chapter 2 presents the work entitled "**Salt dependent mesoscopic model for RNA at multiple strand concentrations**" where we discuss a modification in the PB model to treat multiple strand concentration data sets. The modified model is applied to a data set comprising sequences measured at five different salt concentrations, where beyond saline effects we were also able to derive end effects.

(c) Chapter 3 presents the work entitled "**Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design**" where we studied a modified nucleic acid and derived a full set of parameters that can be used to predict features such as melting temperature and average opening profiles of modified oligonucleotide probes. On this way, we tested the derived parameters to predict all the possible sites of modification insertion and their respective melting temperature of a set of probes targetting oncogenes. From the pool of predictions were selected seven probes that were synthesized and had their melting temperatures measured.

(d) An integrative discussion correlating the work developed by the project and the contribution for the model as well as the future perspectives are enclosed in Chapter 4.

(e) Lastly, a brief conclusion is in Chapter 5.

1 Introduction

1.1 A brief introduction to DNA

Along with the propositions made by Chargaff[1] and the crucial work of X-ray crystallography carried out by the English researchers Rosalind Franklin and Maurice Wilkins,[2,3] James Watson and Francis Crick derived their three-dimensional and bihelicoidal model for the structure of the DNA.[4] This derivation was also only possible due to the recent advances in the construction of models or in the assembly of possible three-dimensional structures based on molecular distances and connection angles, an advanced technique proposed by the American biochemist Linus Pauling.[5]

There have been few modifications to the model proposed by Watson and Crick since their proposition in 1953. The four main fundamentals remain the same:

1. DNA is a double helix, with two strands connected by hydrogen bonds. It is a polynucleotide with four different types of nucleotides as basic units: adenine (A), thymine (T), cytosine (C) and guanine (G). The nucleotides are connected together by non-covalent bonds, specifically, hydrogen bonds. Following what is called Watson-Crick pairing, adenine usually makes two hydrogen bonds with thymine and cytosine makes three hydrogen bonds with guanine.

2. Most DNA double helices are dextrogyrous. That is, if we point the right hand with the thumb upwards, the fingers will bend along the thumb to the right. The thumb representing the helix axis and the fingers the sugar-phosphate skeleton. Such DNA helix is known as B-DNA.

3. The DNA helix is anti-parallel, which means that the 5' end of one strand is paired with the 3' end of the complementary strand (and vice versa). As shown in Fig.1.1, nucleotides are connected to each other by their phosphate groups, which link the 3' end of a sugar with the 5' end of the next sugar in the sequence.

4. Not only are the DNA base pairs connected by hydrogen bonds, but the ends containing nitrogenous bases are exposed and available for potential hydrogen bonds.

Furthermore, there is an interaction between the aromatic rings of the adjacent base pairs, which is known as $\pi - \pi$ or stacking $-\pi$ interaction. This interaction have a complex origin predominantly from the superposition of the π electrons of the bases and in part of the hydrophobic interactions — if water molecules penetrate the structure the energy cost would be very high. It also imposes a well-defined distance between the bases along the axis of the helix, explaining the high rigidity of the molecules along the same axis.[7] It is also known that the stacking energies are highly dependent on the base sequence.[8]

1.2 The subsequent study of RNA

Although there are several types of RNA molecules, the basic structure of each is similar, being a polymeric molecule produced by the union of individual ribonucleotides, always adding the 5'-phosphate group of a nucleotide to the 3'-hydroxyl group of the previous nucleotide. Like DNA, each RNA chain has the same basic structure, composed of nitrogenous bases covalently

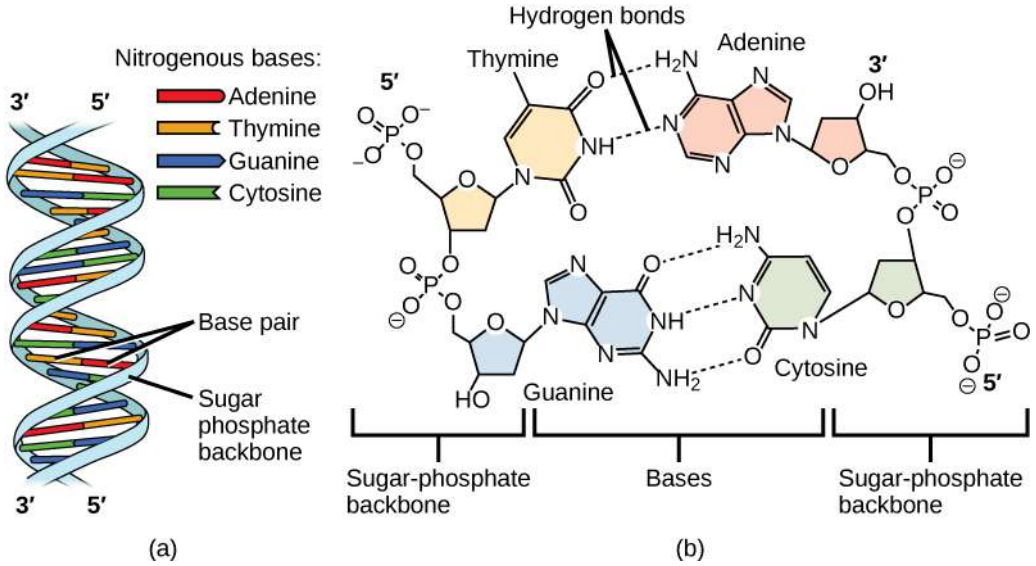


Figure 1.1
(a) DNA has a double helix in an anti-parallel structure, that is, a 3' end of one strand is connected to the 5' end of the other. Usually oriented to the right. (b) As proposed by Chargaff, each nucleotide of the DNA molecule is composed of a phosphate group, a sugar and one of the four base pairs (A, T, C, G). According to the pairing proposed by Watson and Crick, adenine and thymine are connected by two hydrogen bonds; while cytosine and guanine are connected by three hydrogen bonds. Figure extracted from Ref. 6.

linked to a main structure of sugar and phosphate, see Fig.1.2. However, unlike DNA, RNA is usually a single-stranded molecule. In addition, the sugar in the RNA is ribose instead of deoxyribose (ribose contains yet another hydroxyl group on the second carbon), which explains the name of the molecule.

RNA consists of four nitrogenous bases: adenine (A), cytosine (C), uracil (U) and guanine (G), uracil is a pyrimidine structurally similar to thymine, another pyrimidine found in DNA. Like thymine, uracil can pair with adenine.

1.2.1 Saline effects in dsRNA at multiple strand concentration

Despite being a single-stranded molecule, it was soon discovered that RNA could form double-stranded structures, important for its function. In 1956, Alexander Rich — an X-ray crystallographer and David Davies discovered that single strands of RNA can hybridize, uniting

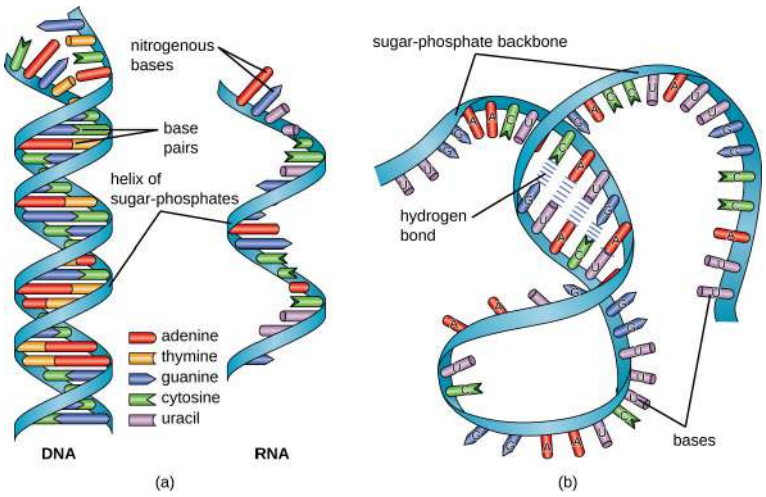


Figure 1.2
(a) Comparison of the secondary structure of DNA and RNA, RNA is usually found in single strands and in addition to the exchange of thymine for uracil, a similar pyrimidine, has a ribose sugar in each nucleotide — (carbon 2 of the ring is hydroxylated). (b) RNA can also fold over itself and generate other types of secondary structures, one of which is the formation of double helices. Figure extracted from Ref. 9.

to form a double helix structure.[10] Double-stranded (ds) RNAs are present in cells and perform a variety of biological functions.[11,12] Small non-coding dsRNAs can mediate neuronal differentiation,[13] segments of special lengths can inhibit the translation of mRNA molecules into proteins through attaching to mRNAs,[14,15] and dsRNAs of more than 30 base pairs (bp) can be key activators of the innate immune response against viral infections.[16]

Similarly to dsDNA, the interchain interactions stabilizing the structure of dsRNA are very sensitive to environmental conditions such as temperature and salt concentration.[17–20] For instance, the binding affinity between the protein kinase R (PKR) and the dsRNA is highly affected by changes in salt concentration and are directly correlated with an improvement in the recognition pathway.[21] One major difference between dsDNA and dsRNA is the helix conformation, due to the additional hydroxil dsRNA forms helices in an A-form, which has a deeper/narrower major groove and a wider minor groove than the dsDNA B-form.[22] This conformation difference confers a different surface electrostatic potential to dsRNA and have been suggested as responsible for the different condensation behaviour of multivalent ions[23] and different flexibilities between dsRNA and dsDNA.[24,25]

Thence, the presence of mono and divalent cations is essential to stabilize RNA secondary and tertiary structures by neutralizing the helix negative charges.[26,27] Even so magnesium ions concedes more stabilization to the helix structure due to their size and relative charge,[28–30] monovalent ions like sodium are crucial for folding and assembly of RNA tertiary structure due to their long-range interactions.[31,32] Several theoretical models have investigated the effects of monovalent ions in RNA, such as molecular dynamics – MD,[24,33–36] coarse-grained models,[37] Debye-Hückel models,[38] and tightly-bound ion theory.[39,40]

Furthermore, the process known as "base fraying", which is the breaking of base-pairing interactions at the termini of a RNA or DNA double helix have shown to retain a crucial role in the overall thermodynamics stability of the double-strand. For example - frayed states, are intermediaries in zipping and unzipping process[41–43] and have been suggested to be important on the interactions of RNA with proteins.[44]

Base fraying is an important aspect of RNA stability but still poorly understood, in particular it is unclear how fraying depends on salt concentration. Melting temperature measurements indicate that the 5' ends are substantially more stable when the purine is positioned at the 3' end, which determine the stability of sequential mismatches as well.[45] MD simulations encountered difficulties to deal with base fraying as existing force fields were inadequate for terminal AU bases.[46]

Previously we have investigated salt concentration and end-fraying effects in dsDNA,[47] and terminal AU base pairs at low salt concentration were found to be less stable than internal base pairs. Although dsDNA and dsRNA share numerous technical features we cannot assume that both will share the same dependence with salt concentration, requiring a more thorough inspection. However, due to a number of technical reasons, there are far less published melting temperatures for dsRNA than for dsDNA and we were not able to carry an investigation of salt and end-fraying effects in RNA on the same way.

More recently we had access to a data set of measured temperatures of dsRNA at multiple strands and salt concentrations, however, as will be further discussed in Section 1.4.2 the model we use to extract the parameters requires a data set at the same strand concentration, although the parameters are found to be fairly independent of strand concentration. In any case, we needed to assure this independence, therefore, we adapted the model to handle multiple strand concentrations and finally be able to confirm or refute if RNA and DNA show similar behaviors when it comes to salt and end-fraying effects.

1.3 Modified Nucleic Acids

In mid-1978, Zamecnik and Stephenson [48] were the first to draw attention to the great potential that oligonucleotides could have in controlling gene expression. Triggering then the synthesis of several analogs of both DNA and RNA, with the most diverse chemical characteristics in order to improve the potency, specificity and therapeutic efficacy. [49–51] Since even the modest improvement may affect its chemical characteristics, attention remains focused on the development of additional oligonucleotide analogs.

For this purpose, a multitude of nucleic acid analogs have been designed, synthesized and evaluated, with the aim of improving the effectiveness of these agents by changing their biophysical properties compared to native DNA and RNA. [52–54] Desirable properties include improved stability to nuclease enzymes and uptake into cells. Greater affinity, kinetics and specificity in base pairing after binding to nucleic acid targets can also improve the effectiveness of these agents.

Many of the chemical changes in oligonucleotides have been extensively reviewed. [55–57] The list of implemented changes only increases and there are several possibilities, from the replacement of the canonical bases by others, such as purina [58] and inosine, [59] changes in the way of matching these bases that are usually of the Watson-Crick type for Hoogsten-type pairings, or the formation of triplexes. [60,61] Changes in the phosphate-sugar structure are also extremely promising, and some examples are: peptide nucleic acid – PNA, [62] DNA phosphoramidate, [63] hexitol nucleic acid – HNA, [64] and morpholino oligomers. [65]

As an example, modifications of oligonucleotides could reduce the negative charge of phosphodiester bonds, increase resistance to nucleases, improve pharmacokinetic characteristics in vivo and facilitate cell uptake. One such analogue, DNA-phosphorothioate, provided the first approved oligonucleotide drug, Fomivirsen, [66] and the 2'-O-alkyl DNA-RNA chimera has been used in several ongoing clinical trials. [67]

There are countless possibilities to introduce changes, for example, in just a single article, Freier and Altmann [68] in 1997 report more than two hundred modifications in hybrid DNA/RNA structures. Mostly, these modifications maintain the same intended purpose when they were first proposed — use in the treatment of diseases that depend directly on the unwanted expression of genetic material, where the selection of these targets is governed mainly by gene sequence. Such knowledge can be used in the treatment of cancers, viral and bacterial infections, as well as other diseases. [69]

The use of oligonucleotides as research tools and diagnostic agents is so widespread that even modest improvements in their properties are likely to have a major impact. Optimized oligomers would not only be more effective antisense agents, but would also provide new options for diagnostic tests, improved probes for basic research and tools for analyzing genetic arrangements with improved sensitivity.

1.3.1 LNA - *Locked Nucleic Acids*

One of the modifications that have been related to an improvement in base pairing, increase in sequence affinity either for complementary RNA and DNA strands or duplexes, and an increase in melting temperature [70] are the so-called locked nucleic acid – LNA. [71] This base has a methylene bridge between the 2'-oxygen and 4'-carbon of the ribose ring locking it in a C3-endo/N conformation reducing its flexibility and mimicking a RNA helix. [72] We show a conformation comparison in the Fig.1.3. Tab. 1.1 shows a comparative of some modifications and their desireable properties, such as RNase H activation, affinity for the target sequence, clinical use, nuclease activity and comercial availability.

Among the remarkable properties of LNA is its ability to increase the stability of chimera duplexes, whether of DNA or RNA. This stability has a direct impact on the denaturation temperatures of the sequences. Several studies have shown a considerable increase in the melting

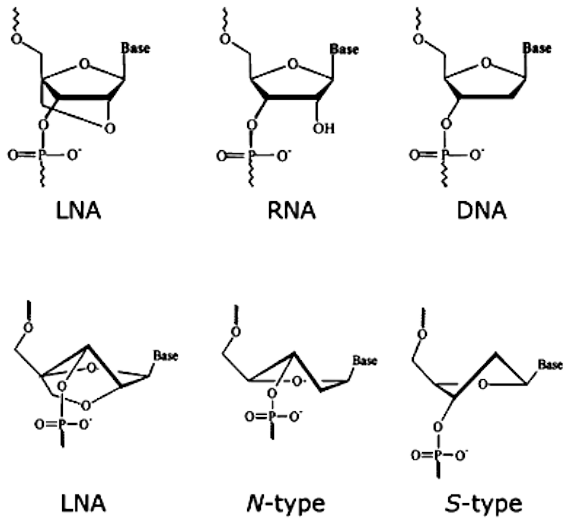


Figure 1.3
(a) Conformation of the furanose rings of DNA, RNA and LNA monomers. (b) The fixed N-type, (C3)-endo conformation, of the LNA nucleotide. Figure extracted from the reference Ref. 73.

temperature,[74–79] even reaching approximately 10°C per added modification. This stability is remarkable and the presence of three LNAs at the 5' and the 3' ends is enough to increase the half-life of the nucleotide.[80]

Nevertheless, the addition of LNA monomers has shown to improve mismatch discrimination, compatibility and specificity towards complementary DNA and RNA strands.[72,81,82] Also exhibit resistance to nuclease degradation, chemical stability and advantageously, convenient synthesis using standards reagents and coupling protocols, moreover, minimal nonspecific interactions with nucleic acid binding proteins.[83]

Several LNA applications have been described over the years. Its base pairing specificity and mismatch sensitivity makes it attractive for use in many applications, such as specific PCR detection and diagnosis,[84–86] high binding diagnostic probes,[87–89] stability improvement and hybridization efficiency in loop-mediated isothermal amplification – LAMP detection,[90] improvement of targeting, specificity and stability in antisense oligonucleotides – ASOs[91–93] and aptamers.[94–96] It is also used in DNAzymes and LNAzymes to improve targeting and cleavage efficiency[97–99] and in molecular beacons.[100] A detailed review of some of LNA applications can be found in the Refs. 72,101

Thermostability and affinity for complementar sequences makes it an ideal modification in molecular beacons,[100] as enhancers to RNA *in situ* hibridization[102,103] and as direct antagonist in miRNA silencing.[104–107] Furthermore, fully LNA modified probes can selectively capture genomic DNA sequences.[108]

The thermodynamic origin of LNAs enhanced base pairing stability has not been clearly delineated yet. It has been suggested that a decrease in the entropy of duplex formation and improved stacking in the duplex both play a part, and there is a negative contribution in enthalpy due to the disruption of the hydrogen bonds.[110,111] Additionally, it is also proposed an association between a favorable enthalpy increase and a more pronounced stacking interaction, however, this is context-dependent and might be influenced by flanking base-pairs.[79] Some studies also had suggested that the change in the helix conformation induces a negative change in the entropy variation, being localized at the level of individual base-pairs, consequently increasing the overall thermodynamic helix stability.[112,113]

Structurally, several studies addressed the source of the stabilizing effect provided by LNA incorporation and even well-established techniques such as NMR and X-ray diffraction have reached conflicting results. Some studies have accounted the stabilization as an improvement in the stacking interactions,[38,114,115] and a different NMR assay discarded an improvement either in stacking interaction or hydrogen-bonding suggesting more detailed studies with hybridized water.[116] Still, the effect is accounted as a localized change in the helix conformation due to

Modification	RNase H	Affinity	Clinical use	Nuclease	Comercial availability
PO-DNA	yes	no	no	no	yes
PO-RNA	no	no	no	no	yes
PS-DNA	yes	no	yes	yes	yes
2'-O-MeRNA	no	yes	no	yes	yes
2'-O-MOE-RNA	no	yes	no	yes	no
2'-O-MeRNA/DNA	yes	yes	yes	yes	yes
2'-O-MOE-RNA/DNA	yes	yes	yes	yes	no
PNA	no	yes	no	yes	yes
Phosphorothioate	no	yes	no	yes	yes
Morpholino	no	yes	no	yes	yes
LNA	yes	yes	no	yes	yes

Table 1.1
Comparison of the main desirable characteristics in modified nucleic acids and the LNA. Table extracted from the reference[109]

single locked nucleotides and is essentially limited to the immediate neighboring base-pairs.[117]

Nevertheless, the nearest-neighbour model that is usually applied to predict thermodynamic properties for LNA modified probes[110,113,118,119] uses empirical parameters and does not provide any theoretical insight over physical properties, like hydrogen bonds or stacking parameters. Mesoscopic models, like the model applied by us can derive easily such parameters[120] and exhibited success and experimental accordance when applied to modified nucleic acids.[121,122] In this way a parameterization of mesoscopic models to treat LNA could provide new insights about the physical interactions and point out the source of its stability. Furthermore, having LNA parameters at hand will allow new bioinformatics approaches, such as oncogene probe screening and tailored design and it may help to reduce experimental costs.

1.4 Methodology

1.4.1 The melting process of nucleic acids

Thermal unbinding of nucleic acids has a play role in a variety of biological phenomena including transcription, translating, and gene expression regulation. Several sophisticated applications rely on this fundamental interaction, including PCR, genetic sequencing, gene therapy, diagnostic tools,[123,124] use in super DNA networks[125] and digital circuits.[126] Although structurally stable, DNA is extremely dynamic in its bonds. Small temperature gradients can induce temporary bubbles known as breathing modes.[127] Furthermore, it can undergo important changes in its configuration during the transcription processes when forming DNA/RNA hybrids, or triplexes.

This process has been extensively studied for nearly four decades,[128–131] and still, the nature of its phase transition is controversial. The full comprehension and description of its intramolecular process remains as a challenge.[132] As an example is still not completely clear why the DNA melts in a narrow range of temperatures exhibiting behavior similar to a first-order phase transition.[133,134]

When there is an increase in temperature or pH variation, the double stranded DNA progressively denatures. Tipically the process starts in AT rich regions and subsequently moves to zones of increasing CG content. These so called bubbles then grow and merge until the helix is fully melted, the temperature at which this process occurs is experimentally defined as the temperature at which half of the DNA molecules have been denatured.[135]

The transitions in DNA duplex oligomers can be monitored by ultraviolet absorbance at 268

nm while temperature of a sample is linearly increased[136] or by measuring its heat capacity change from differential scanning calorimetry (DSC).[137] In principle, similar data could also be obtained through a fluorescence emission signal,[138] the intensity of an NMR peak,[139] by circular dichroism[140] or a Raman signal.[141]

In an attempt to describe the properties of DNA and its thermodynamics several theoretical approaches have been proposed whose can be mainly classified into three groups, depending on the amount of detail and computational complexity — macroscopic, microscopic and mesoscopic.

Overall, macroscopic models provide good results when describing phenomena at molecular scale, from which we can cite, the Worm Like Chain based models (WLC) and Nearest-Neighbour model (NN). The WLC model is a standard way of modeling the mechanical behavior of nano-scales filamentous structures in thermal equilibrium. These models present the interesting feature that a single parameter, i.e. the elastic parameter of bending modulus of the filament, is used.[142,143] The latter, NN model uses the measured Gibbs free-energy (ΔG_{37}) for each sequence in a considerable set of possibilities for each NN and then the individual thermodynamic data is extracted for each NN using a linear equation system.[144] Based on this principle, thermodynamic features for any set of sequences could be easily predicted.[145] However, the method is mostly empiric and can not give any information about intermolecular interactions in DNA and RNA.

Microscopic models aim to describe the local and specific phenomena of each base pair at an atomic level. We can cite as examples, the calculations of Density Functional Theory (DFT) and molecular dynamics (MD).[146,147] These models can provide a great amount of details in the description of phenomena, however this comes at a high computational cost, turning their use impractical for the study of large molecules.

Mesoscopic models, in general can provide more details about the transition than macroscopic models at a lower computational cost when compared to the detailed descriptions provided by microscopic approaches. The first mesoscopic models appeared in the 1960s, inspired by the Ising's physical-statistical models.[148] However, Ising's models could not describe the dynamics of the phenomenon, much less the large fluctuation amplitudes in the average opening of the duplexes.[149]

In the 1970s, Poland and Scheraga[128] implemented a more sophisticated version of the Ising model, where they added the entropy associated with each "bubble"[128,150] and modeled the DNA molecule as an alternating sequence of intact double-helical and denatured, single-stranded domains.

Nonetheless, there is another theoretical approach, the Peyrard-Bishop – PB model,[151] which manages to describe the main intermolecular interactions using simple potentials and predict melting temperatures in agreement with the experimental results. The PB model gives an example of the possibility of an entropy-driven transition in a one-dimensional system, showing the existence of a first-order temperature driven transition where the bulk entropy changes at the transition.[152] It has been used extensively and many applications of the PB model can be found in the literature.[153–156]

Nevertheless, it has been left aside due to its relative complexibility when compared with other methods. Weber[157] solved some problems of computational efficiency that had prevented its use, and with that, it became possible to use it in more practical situations. The detailed method is described in the Section 1.4.2.

1.4.2 The Peyrard-Bishop model

Firstly proposed in 1989, the PB model addresses the thermal melting using statistical mechanics, dodging the empiricity regarding the NN model and minimizing the number of freedom-degrees to be treated in a physical model. This made possible the extraction of intrinsic parameters regarding the transition as well as physical properties, which was only possible through MD simulations.

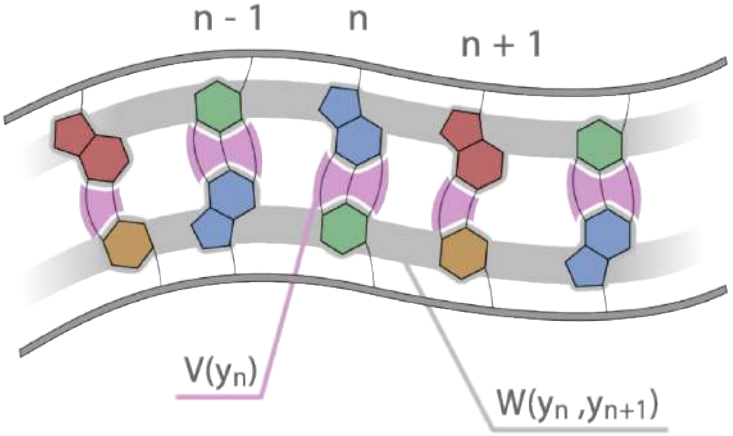


Figure 1.4
Scheme for Morse, $V(y_n)$, and stacking potentials, $W(y_n, y_{n+1})$, considered in the PB model. Figure extracted from Ref. 162

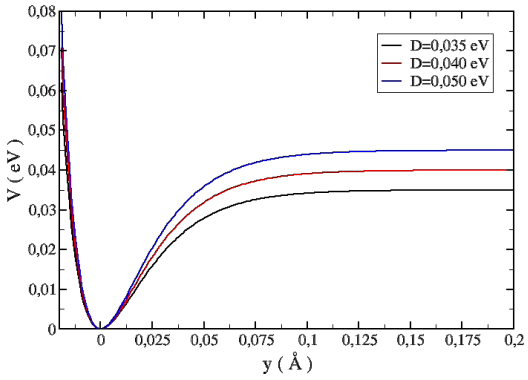


Figure 1.5
Morse potential with a fixed at $4,45$Å for three different values of D.

The PB model is capable of exhibit the traditional abrupt behavior observed in denaturation curves by doing a single molecule calculation and subsequent analytical investigations of its non-linear behavior suggested that energy locations intrinsic to the molecule could initiate the denaturation process. This was later proved by simulations of molecular dynamics.[158]

In the PB model, each base pair is described by two degrees of freedom, u_n and v_n, which is relative to the displacement of the bases from their equilibrium positions along the direction of their hydrogen bonds, which is approximated by a Morse potential,[159] as has been previously done in other studies.[160,161] It is an average non-linear potential representing the number of hydrogen bonds that connect the base-pair. Additionally, it assumes a harmonic coupling between adjacent base-pairs due to its stacking interaction.

Thereby, the hamiltonian for the system is written as:

$$H = \sum_n \frac{1}{2} m(\dot{u_n}^2 + \dot{v_n}^2) + W(u_n, v_n) + V(u_n - v_n) \tag{1.1}$$

where m is the mass of the base-pair. Here, it is assumed the reduced mass of the base-pairs. The potentials $W(u_n, v_n)$ e $V(u_n - v_n)$ are defined by:

$$W(u_n, v_n) = \frac{1}{2} k[(u_n - u_{n-1})^2 + (v_n - v_{n-1})^2] \qquad \text{e} \qquad V(u_n - v_n) = D(e^{-a(u_n - v_n)} - 1)^2 \tag{1.2}$$

Here, k defines an elastic constant for the harmonic potential, D represents the energy required to dissociate the base-pair, and a the reach of the last potential. In the Fig.1.5 we show the behavior of the Morse potential for a fixed value of $a = 4.45$Å and different values of the dissociation constant D.

u_n and v_n can be decoupled by considering the in-phase and out-phase movements, respectively:

$$x_n = \frac{1}{\sqrt{2}}(u_n + v_n) \qquad \text{e} \qquad y_n = \frac{1}{\sqrt{2}}(u_n - v_n), \tag{1.3}$$

Only the out-phase displacements, y_n, stretch the hydrogen bonds. The hamiltonian in Eq.(1.1) becomes:

$$H = H(x_n) + H(y_n) = \left[\sum_n \frac{p_{x_n}^2}{2m} + W(x_n, x_{n-1})\right] + \left[\sum_n \frac{p_{y_n}^2}{2m} + W(y_n, y_{n-1}) + V(y_n)\right], \tag{1.4}$$

where $p_{x_n} = m\dot{x}_n$, $p_{y_n} = m\dot{y}_n$ are the canonical momentum for the new variables x_n and y_n. The potentials $W(x_n, x_{n-1})$, $W(y_n, y_{n-1})$ and $V(y_n)$ remain the same as in Eq.(1.2).

Considering that those dissociation reactions take place in low species concentration (μM), the partition function for the system is therefore defined in the canonical ensemble. Moreover, for a chain containing N base pairs:

$$Z = \int \prod_{n=1}^{N} dp_{x_n} dx_n dp_{y_n} dy_n e^{-\beta H(p_{x_n}, x_n, p_{y_n}, y_n)} = \left(\frac{8\pi^3 m^2}{k\beta^3}\right)^{N/2} Z_y \tag{1.5}$$

where $\beta = 1/k_B T$ and k_B is the Boltzmann constant. Thus, the hamiltonian term that contains the integration of positions for the y_n coordinate is given by Z_y:

$$Z_y = \int \prod_{n=1}^{N} dy_n e^{-\beta(W(y_n, y_{n-1}) + V(y_n))} \tag{1.6}$$

The partition function Z_y can be exactly evaluated in the thermodynamic limit ($N \to +\infty$) using the eigenfunctions and eigenvalues of a transfer integral operator – TI.[163,164]

$$\int dy_{n-1} K(y_n, y_{n-1}) \varphi_i(y_{n-1}) = e^{-\beta\epsilon_i} \varphi_i(y_n), \tag{1.7}$$

where

$$K(y_n, y_{n-1}) = e^{-\beta\{W(y_n, y_{n-1}) + \frac{1}{2}[V(y_n) + V(y_{n-1})]\}} \tag{1.8}$$

Thus, the solution refering to Z_y is given by:

$$Z_y = e^{-\beta N \epsilon_0} \tag{1.9}$$

For instance, ϵ_0 is the lowest eigenvalue of a Schrödinger-type equation which determines the eigenfunctions of the transfer integral operator.

$$\frac{-1}{2\beta^2 k}\frac{\partial^2 \varphi_i(y)}{\partial y^2} + D\left(e^{-2\sqrt{2}ay} - 2e^{-a\sqrt{2}y}\right)\varphi_i(y) = (\epsilon_i - s_0 - D)\varphi_i(y) \tag{1.10}$$

com $s_0 = \frac{1}{2\beta}\ln\left(\frac{\beta k}{2\pi}\right)$.

The Eq.(1.10) is formally identical to the Schrödinger equation for a particle in a Morse potential, which can be solved exactly. Now the physical properties of the DNA thermal transition can be derived from its eigenfunctions and eigenvalues.

Furthermore, the mean-stretching of the hydrogen bonds $\langle y_m \rangle$, can be evaluated appling a similar procedure performed for Eq.(1.9):

$$\langle y_m \rangle = \frac{1}{Z_y}\int \prod_{n=1}^{N} y K(y_n, y_{n-1}) dy_n \approx \langle \varphi_0(y)|y|\varphi_0(y)\rangle = \int \varphi_0^2(y) y dy \tag{1.11}$$

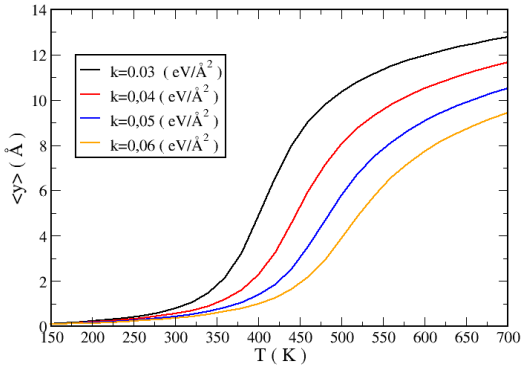


Figure 1.6
Variation of $\langle y_m \rangle$ as a function of temperature T for four different values of the stacking constant k. Were used, $D = 0.04$ eV, $a = 4.45$Å, $\theta = 0.01$ and we varied the values of k

for the normalized function $\varphi_0(y)$. This integral can be numerically evaluated and shows how the average lenght of the hydrogen bonds changes with the temperature for different stacking constants k, as can be seen in Fig. (1.6).

Despite exhibiting a thermal behavior in qualitative agreement with the experiments, it was not yet able to achieve the narrow temperature range in which the transition was experimentally observed.[165] Such behavior had been successfully described in a previous model using a description from Ising models,[166] considering that the extremely sharp state change was due to the effects of cooperativeness between base pairs. Roughly, this implies that a base pair in a bound-state next to another in an unbound-state is more likely to have its hydrogen bonds broken.[167] Which was similar to the idea of the entropy-driven transition previously proposed.[168] That is, a local unveiling of the helix contributes to its denaturation process. However, for these models, these concepts appeared through phenomenological parameters not related to the intrinsic physical properties of the molecule.

Over the years, several modifications were proposed to the model in order to better approximate the experimental behavior.[130,153,154,169–174]

Corrections for heterogeneous sequences

The great advantage of the model is that the configurational part of the Hamiltonian, Z_y can be solved rigorously by applying the TI technique, as shown above, granting a better understanding of the non-linear character of the transition.

However, the model in its original form had two main limitations - the solution applying the TI it is only valid for $N \to \infty$, that is, an infinite sequence of base pairs, and still, it must be a homogeneous sequence. Such problems were later circumvented by Zhang et al.[175] in order to make the model applicable to more realistic situations.

The first problem is solved by considering that the thermal denaturation has a dissociation balance between dsDNA (C_2) and ssDNA (C_1):

$$C_2 \rightleftarrows 2C_1 \tag{1.12}$$

Briefly, the scope between the "internal" and "external" movements is approximately separated by the use of a diameter d in the phase space.

The TI technique is applicable assuming that all potentials are in unbounded states, which is not true for the PB model since the strands are "bounded" to each other. Therefore, when we take the expansion of the nucleus across the $[+\infty; -\infty]$ phase space, both cases are considered. What the author does, is establish an upper limit for $\langle y \rangle$ avoiding the divergence of the integral.

The second is solved by expanding the TI kernel in an appropriate set of orthonormal bases.

Defining:

$$K^{(i,i+1)}(y_i, y_{i+1}) = e^{-\beta(W(y,y_{i+1})+\frac{1}{2}[V_i(y_i)+V_{i+1}(y_{i+1})])} \tag{1.13}$$

Hence, the partition function referring to Eq.(1.6) is rewritten as:

$$Z = \int dy_1 dy_2 \cdots dy_N K^{(1,2)}(y_1, y_2) K^{(2,3)}(y_2, y_3) \cdots K^{(N,1)}(y_N, y_1) \tag{1.14}$$

The boundary condition is incorporated by connecting the last base-pair of the sequence to the first. That is, $N + 1 \to 1$.

Here the different kernels are expanded by the same set of bases. $K^{(i,i+1)}(x, y)$ is defined in a limited space $\{a, b : a, b\}$ keeping the convergence of the integral. Nonetheless, any complete set of orthonormal basis functions, $\{\varphi_m(y)\}$ can be chosen to expand the cores of $K^{(i,i+1)}(x, y)$:

$$K^{(i,i+1)}(x, y) = \sum_{n,m=1}^{+\infty} C_{n,m}^{(i,i+1)} \varphi_n(x) \varphi_m(y), \tag{1.15}$$

and the coefficients $C_{n,m}^{(i,i+1)}$ are determined by:

$$C_{n,m}^{(i,i+1)} = \int\int dx dy K^{(i,i+1)}(x, y) \varphi_n(x) \varphi_m(y). \tag{1.16}$$

In practice, the Eq.(1.15) is truncated within the first M bases, so the equation is rewritten as:

$$K^{(i,i+1)}(x, y) \approx \sum_{n,m=1}^{M} C_{n,m}^{(i,i+1)} \varphi_n(x) \varphi_m(y). \tag{1.17}$$

Replacing the Eq.(1.17) in Eq.(1.5):

$$Z = Tr\left(\prod_{i=1}^{N} C^{(i,i+1)} \right). \tag{1.18}$$

Where each one of the matrices $C^{(i,i+1)}$ represent the interaction between adjacent base pairs, i and $i + 1$. The boundary condition is included in the last matrix, connecting the first and last base pair. Thence, the infinite sequence problem has been circumvented and it is now possible to treat sequences with realistic lengths. It also can be treated as a ring, in the case where the boundary condition is periodic or the stacking interaction between the first and last base pair are neglected to result in an "open" sequence.

Until now, no distinction has been made between base pairs. The four possible successive neighbor types for AT(w), 'w' for weak, and CG(s), 's' for strong. Their respective matrices can be defined, explicitly:

$$C^{(w,w)}, C^{(w,s)}, C^{(s,w)}, C^{(s,s)}.$$

As mentioned earlier, any set of orthonormal bases can be used, thus the functions obtained from the transfer operator are taken as the basis for the case of a homogeneous CG sequence. That is, the matrix $C^{(s,s)}$ which is simplified to a diagonal matrix Λ with eigenvalues given by the matrix λ_i.

Therefore, any $C^{(a,b)}$ matrix can be used:

$$C^{(a,b)} = \Lambda + \Delta^{(a,b)}, \tag{1.19}$$

representing the difference between the interaction of neighbors of type (a, b) and neighbors (s, s). In particular, $\Delta^{(s,s)} = 0$.

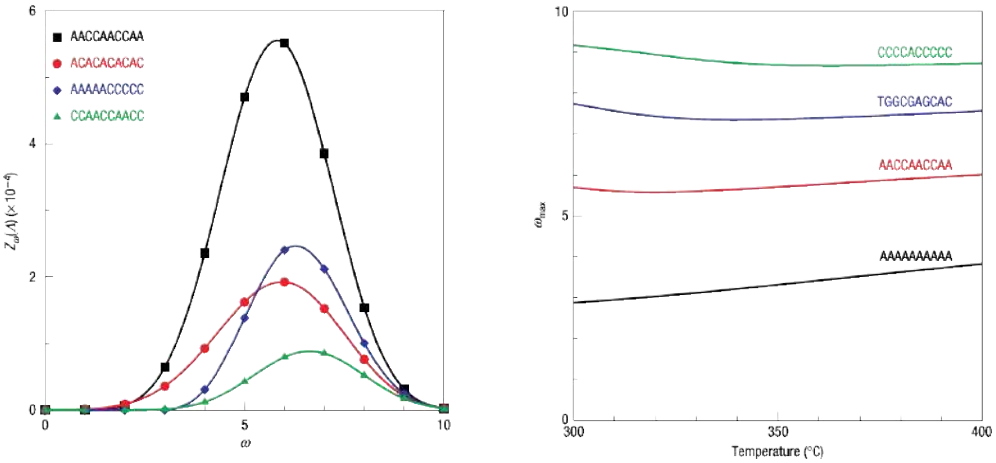


Figure 1.7
The figure on the left shows the Gaussian behavior between $Z_\omega(\Lambda)$ and ω for 4 sequences of 10 base pairs containing 40% to 60% of CG base pairs in temperature 370 K. The figure on the right shows the behavior of ω_{max} as a function of temperature for another 4 sequences with 10 base pairs. Figure extracted from Ref. 176

The partition function is rewritten as:

$$\begin{aligned} Z &= Tr[(\Lambda + \Delta^{(1,2)})(\Lambda + \Delta^{(2,3)}) \dots (\Lambda + \Delta^{(N,1)})] \\ &= \sum_{\omega=0}^{N} Z_\omega(\Lambda) = \sum_{\omega=0}^{N} Tr[M(\Lambda^\omega)], \end{aligned} \tag{1.20}$$

Where $M(\Lambda^\omega)$ are all terms containing ω multiplications of the matrix Λ. As pointed out by Weber [157] the analytical factorization of each Λ term is not possible, except for very small N.

However, terms related to sequence homogeneity are derived. For example, for a sequence containing only CG the only non-null term would have the order of Λ as all other Δ matrices are null. Lower orders of Λ are obtained depending on the number of non-null Δ matrices.

In Fig.1.7, the behavior of Λ can be observed in terms of ω for some sequences (Λ) and the fraction of CG present in the sequence.

Indubitably, the partition function has a strong correlation with temperature and could not be used to compare different sequences. However, when observing the behavior of ω_{max} obtained by interpolating the Gaussian function obtained previously, the dependence on temperature is no longer observed, see Fig. 1.8. Roughly, ω_{max} can be understood as an interpolation parameter between a complete CG homogeneous sequence ($\omega_{max} = N$) and a sequence consisting of only AT (lowest ω_{max}).

Thermodynamic Equivalence

When comparing how ω_{max} behaved as a function of temperature, see Fig. 1.8, they observed a strong linear correlation and the slope of the regression lines was related to the length of the sequence. The *melting index* is then defined as $\tau = {\omega_{max}}^{1/2}$, and through its equivalence with experimental temperatures, temperature predictions for denaturation of new sequences can be carried out. [176] This correlation with experimental temperatures is done at a single strand concentration since the PB model parameters are calculated in a single molecule model, and for this reason the melting temperatures must be given at a single strand concentration.

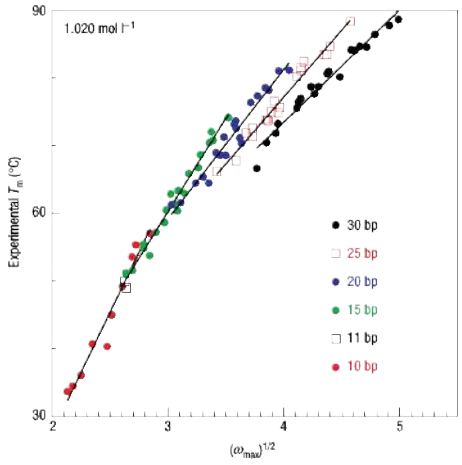


Figure 1.8
Experimental melting temperatures (T_m) as a function of the values of ω_{max} for several sequences. Figure extracted from Ref. 176

The correlation between predicted and measured melting temperatures is accomplished by taking experimental measurements of melting temperature and applying a multidimensional numerical method of parameter optimization,[177] which will be further revisited in details in the Section 1.4.3.

The coefficient τ_i, for the ith duplex in the data set provides the temperature prediction T_i':

$$T_i'(P) = a_0(N) + a_1(N)\tau_i(P) \tag{1.21}$$

where the coefficients $a_{0,1}$ are calculated via a linear regression of the experimental melting temperatures T_i at a single strand concentration C_t, and P is a set of tentative model parameters. For each group of sequence lenght, N, it is carried out a separated regression, as exemplified in the Fig.(1.8). The regression of the coefficients used in Eq.(1.21) is typically carried out at a single strand concentration C_t.

Therefore, if the set does not contain at least 3 different groups of sequence lenghts, only one regression is carried out for the whole set:

$$T_i'(P) = a_0 + a_1\tau_i(P), \tag{1.22}$$

Average opening profiles

In addition to the prediction of temperatures for new sequences, another parameter can be derived: the average opening for each base pair of the sequence using Eq.(1.11). Hereby, it is possible to detect regions prone to greater instability as well as the dynamics of the denaturation process. Furthermore, it can be used as a way of analyzing and validating the results when comparing with similar data experimentally measured.

This analysis step also allows us to determine if there were any problems during the optimization of the parameters. For example, an anomalous opening, that is, openings that were not observed in the experimental literature in similar systems, may indicate that the minimization was centered on a local minimum, far from the real minimum of the system, thus requiring a new processing of the parameters. Or a change in the minimization boundary variables.

1.4.3 Parameter optimization

The predicted temperature achieved by using the PB model is compared in several rounds of minimizations. A parameter optimization is carried out in several rounds of global minimizations, which in turn is composed of a certain amount of rounds of local minimizations.

The main step of the method is the minimization of the total squared difference between the

predicted T_i' and measured T_i:

$$\chi^2 = \sum_i \left[T_i - T_i'(\{p\}_k)\right]^2 \tag{1.23}$$

where T_i is the measured temperature for the ith sequence. T_i' is the corresponding predicted temperature resulting from the set of tentative parameters $\{p\}_k$:

$$\{p\}_k = \{p_{k1}, p_{k2}, p_{k3}\}, \tag{1.24}$$

In addition we use the average prediction difference

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^{N} \left| T_i - T_i'(\{p\}_k) \right|. \tag{1.25}$$

as a simple comparative parameter.

Seed parameters

In all optimizations the ith initial parameter p_i is varied randomly in an interval

$$p_i \in [(1-f)s_i, (1+f)s_i] \tag{1.26}$$

that is, within a fraction $\pm f$ of a seed value s_i.

For instance f=0.1 for PB minimizations, which results in the interval $[0.9s_i, 1.1s_i]$.

The core of the minimization is a Nelder-Mead simplex minimization [177] which will map the surface generated by the N parameters contained in the minimized set of sequences.

1.5 Objectives

1.5.1 General Objectives

Obtain a parameterization of the PB model for sequences at different strand concentrations. On the same way, also clarify the effect of strand concentration on the model and the melting temperature. Study LNA, using the PB model from published data. Furthermore, validate and apply the optimized paramaters in further probe predictions.

1.5.2 Specific objectives

1. RNA at different salt concentrations
   (a) Develop a theoretical approach of the PB model for sets with different strand concentration;
   (b) Apply this model on the parameterization of RNA at different strand and salt concentration;
2. Modified nucleic acids
   (a) Develop a theoretical approach of the PB model for modified nucleic acids;
   (b) Apply this new model on the parameterization of LNA;
   (c) Use the obtained parameters to predict optimized DNA+LNA:DNA probes.

2 Salt dependent mesoscopic model for RNA at multiple strand concentrations

Biophysical Chemistry 271 (2021) 106551

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: www.elsevier.com/locate/biophyschem

Salt dependent mesoscopic model for RNA at multiple strand concentrations

Izabela Ferreira [a,b], Tauanne D. Amarante [c], Gerald Weber [a,*]

[a] Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[b] Programa Interunidades de Pós-Graduação em Bioinformática, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[c] MRC Cancer Unit, University of Cambridge, Hutchison/MRC Research Centre, Cambridge Biomedical Campus, Cambridge, UK

ARTICLE INFO

Keywords:
Peyrard-Bishop
Mesoscopic model
RNA
Strand concentration

ABSTRACT

Mesoscopic models can be used for the description of the thermodynamic properties of RNA duplexes. With the use of experimental melting temperatures, its parametrization can provide important insights into its hydrogen bonds and stacking interactions as has been done for high sodium concentrations. However, the RNA parametrization for lower salt concentrations is still missing due to the limited amount of published melting temperature data. While the Peyrard-Bishop (PB) parametrization was found to be largely independent of strand concentrations, it requires that all temperatures are provided at the same strand concentrations. Here we adapted the PB model to handle multiple strand concentrations and in this way we were able to make use of an experimental set of temperatures to model the hydrogen bond and stacking interactions at low and intermediate sodium concentrations. For the parametrizations we make a distinction between terminal and internal base pairs, and the resulting potentials were qualitatively similar as we obtained previously for DNA. The main difference from DNA parameters, was the Morse potentials at low sodium concentrations for terminal r(AU) which is stronger than d (AT), suggesting higher hydrogen bond strength.

1. Introduction

RNA plays an essential role in many cellular processes such as transcription, translation, and conservation of genetic information. Double stranded (ds) RNAs are present in cells and perform a variety of biological functions [1,2]. For instance, small non-coding dsRNA that mediate neuronal differentiation [3], dsRNA segments of special lengths, known as siRNA, can inhibit the translation of mRNA molecules into proteins through attaching to mRNAs [4,5], and RNAs of more than 30 base pairs of length can be key activators of the innate immune response against viral infections [6].

Similarly to dsDNA, the interchain interactions stabilizing the structure of dsRNA are very sensitive to environmental conditions such as temperature and salt concentration [7–10]. For example, a reduction in salt concentration increases the binding affinity between the protein kinase R (PKR) and the dsRNA, improving the recognition pathway [11].

dsRNA form helices in an A-form which has a much deeper/narrower major groove and a wider minor groove than the B-form of dsDNA, which concedes a very different surface electrostatic potential for dsDNA and dsRNA [12]. These different ion binding modes for dsDNA and dsRNA have been suggested to be responsible for the different multivalent ion-dependent condensation behaviours [13] and flexibilities for dsDNA and dsRNA [14,15]. Therefore, the presence of mono and divalent cations plays a fundamental role in the stabilization of RNA secondary and tertiary structures by neutralizing the negative charge and reducing the repulsion of the phosphates [16,17]. Although magnesium ions are much more stabilizing [18–20], monovalent ions like sodium are important and the general conclusion is that sodium ions are essential as they mediate the long-range interactions that are crucial for folding and assembly of RNA tertiary structures [21,22].

There is some NMR evidence that group I monovalent ions, Na^+ and K^+ in particular, remain well hydrated in the presence of RNA [23] interacting with it in a diffuse way [24,25] or may even be chelated by irregular RNA structures [26,27]. Those factors may relate the sensibility of RNA tertiary structures to the size of the monovalent cations that are present, in contrast to the weak discrimination shown by DNA helices in their interaction with different group I ions [28]. Another aspect that affects the thermodynamic stability of the double stranded duplex is a process known as" base fraying", which is the breaking of base-pairing interactions at the termini of a RNA or DNA. Frayed states

* Corresponding author.
E-mail address: gweber@ufmg.br (G. Weber).

https://doi.org/10.1016/j.bpc.2021.106551
Received 23 December 2020; Received in revised form 19 January 2021; Accepted 19 January 2021
Available online 29 January 2021

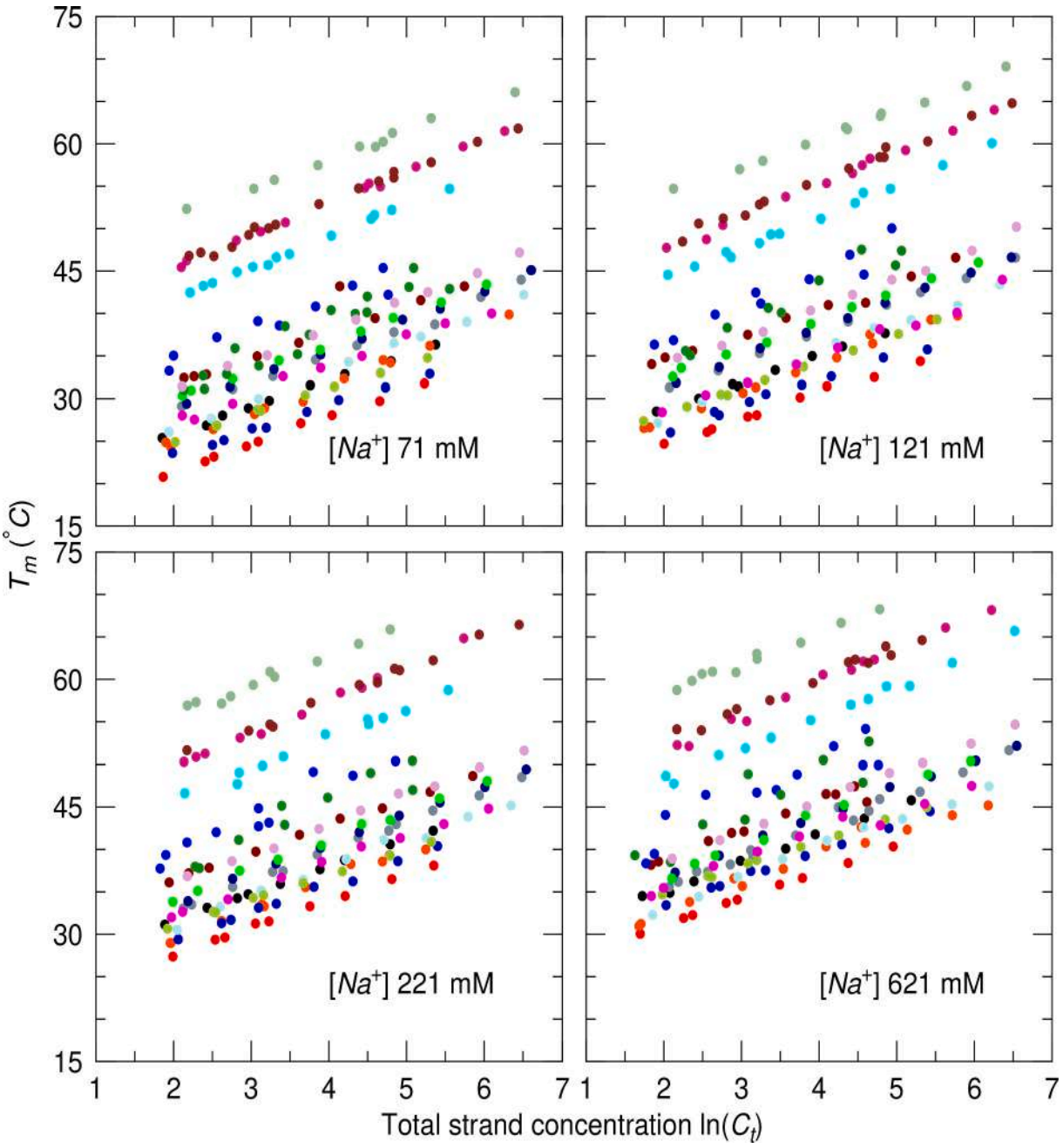


Fig. 1. Experimental melting temperatures from Ref. 52 as logarithmic function of total strand concentration ln(C_t). Each colour represents a specific sequence.

are intermediaries in zipping and unzipping processes and have been suggested to be important for the interactions of RNA with proteins [29,30], are required for secondary structure rearrangements for riboswitch function [31], and may be relevant for strand migration [32].

The effect of monovalent ions in RNA has been investigated with several theoretical methods, such as molecular dynamics (MD) [14,33–36] coarse-grained models [37], Debye-Hückel models [38], and tightly-bound ion theory [16,39]. For instance, MD simulations such as by Beššeová et al. [33,34] concluded that the force field and salt effects are sequence-dependent and the helix compactness is sensitive to the salt and water conditions. Salt effects and stability on the tridimensional structure of RNA were also explored by Monte-Carlo simulations and an increase in Na+ concentration tends to improve the folding of RNA hairpins, suggesting that the base-pair adjacent to the terminal is not stable due to the reduction of stacking [40]. Debye-Hückel models concluded that a decrease in salt concentration generally destabilize the folding of RNA and lowers its denaturation temperatures [37,38]. Mesoscopic modelling, based on the Peyrard-Bishop (PB) description, using experimental melting temperatures as input data, have been restricted to high sodium concentrations [41]. Existing RNA melting temperature data at lower sodium concentrations exists at varying strand concentrations, however the mesoscopic approach requires all temperatures to be at a single strand concentration [42]. Here, we extend this mesoscopic model to handle multiple strand concentration, thus overcoming the current limitations of this approach.

Base fraying is an important, yet still poorly understood aspect of RNA stability, in particular it is unclear how fraying depends on salt concentration. Melting temperature measurements indicate that the 5′ ends are substantially more stable when the purine is positioned at the 3′ end, which determine the stability of sequential mismatches as well [43]. This might be related to the significant overlap observed in a 3′ base and the little overlap observed on a 5′ purine which is not observed

on 5′ pyrimidine by NMR [44]. Other NMR measurements concluded that the opening and closing rates of r(AU) base-pairs are much larger than those observed for d(AT), despite comparable stability [45]. MD simulations have had difficulties to deal with base fraying as existing force fields were inadequate for terminal AU bases [46]. However, more recently this limitation seems to have been resolved and Pinamonti et al. [47] concluded that 5′ ends containing UApCG or AUpGC have a slower fraying due to a larger stability assigned to stacking interactions. This suggests that terminal adenine base pairs have stronger stacking interaction when compared with uracils [47]. In contrast to MD, for mesoscopic PB models [48] and coarse-grained models [49], end-fraying is well represented and they have in principle no difficulty in dealing with AU terminal pairs [41,50]. Nearest-neighbour (NN) models are typically limited to temperature prediction, and terminal effects are included as an energy penalty [51]. The salt dependence of these terminal factors were studied by us recently [52], and we observed a marked quadratic dependence in the enthalpies and entropies with salt concentration which are compensated to form almost linear Gibbs free energies [52].

Here, we adapt the mesoscopic PB model to RNA with varying salt dependence, multiple strand concentrations and including terminal effects. In part, we applied a similar approach as from our previous work on DNA salt-dependent terminal effects [53], which enables us to compare RNA and DNA terminal effects and discuss their differences. However, for RNA the available melting temperatures are scattered into a non-uniform range of strand concentrations, see Fig. 1 [52]. This represents a challenge for the mesoscopic model which usually requires that all temperatures are at the same concentration [54]. The reason for this is that the PB model is a single molecule calculation, and the melting temperatures are correlated to experimental values at a single strand concentration [55]. To work with the existing set of temperatures we adapted the model to handle multiple strand concentrations simultaneously. To achieve this we grouped the strand concentrations into logarithmic groups and then worked out the corresponding model parameters. We tested various levels of grouping and, surprisingly, the model parameters had very little dependence on the grouping factors. Once we established the best level of grouping we were able to compare the new salt-dependent parameters to our previous DNA parameters. In general, we found that the Morse potential representing the hydrogen bonds of RNA follows very closely that of DNA, except for low salt concentrations where d(AT) had an important reduction which we did not observe for r(AU). Recently, experimental and theoretical approaches have shown a reduction of persistence length in DNA and RNA with ionic strength [56–58]. The new mesoscopic parameters calculated here could be applied to the calculation of RNA persistence lengths using methods such as by Jeon et al. [59].

2. Methods

2.1. Model

The configurational part of the PB Hamiltonian is written as [48,60].

$$U_{i,i+1} = \frac{k_{\alpha,\beta}}{2}(y_i - y_{i+1})^2 + D_\alpha(e^{-y/\lambda_\alpha} - 1)^2, \quad (1)$$

which describes the interaction of a base pair of type α, at sequence position i, with its nearest-neighbour of type β at position $i+1$. The Morse potential, which describes the hydrogen bond between the base pairs, uses two more parameters to characterize its depth and width of the ith base pair of type α, D_α, λ_α, respectively. The stacking interaction between adjacent base-pairs or the nearest-neighbours is represented by an elastic constant $k_{\alpha,\beta}$, and the coordinate y represents the relative displacements between the bases. An extension of the PB model is the Peyrard-Bishop-Dauxois (PBD) model [61], which has an added anharmonic term. We are not using the anharmonic as it only increases the number of parameters to optimize without improving the quality of the melting temperature predictions [42].

Therefore, the sum for the Eq. (1) over all N base-pairs is carried out using its partition function:

$$Z_y = \int_{y_{min}}^{y_{max}} dy_1 \int_{y_{min}}^{y_{max}} dy_2 \cdots \int_{y_{min}}^{y_{max}} dy_N \int \prod_{n=1}^{N} e^{-\beta U(y_i, y_{i+1})} \quad (2)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant and T the absolute temperature. Subsequently, the integral over all possible configurations of base pair displacements, y_i is performed. Thus, all possible Morse potentials and stacking interactions are considered simultaneously during the evaluation. From the partition function, Eq. (2) an adimensional index τ is calculated and it is directly correlated to the experimental melting temperatures as we will see in the next sections [42].

Furthermore, the average base pair displacement, $\langle y_m \rangle$, at the mth position in the sequence can be obtained from

$$\langle y_m \rangle = \frac{1}{Z_y} \int_{y_{min}}^{y_{max}} dy_1 \int_{y_{min}}^{y_{max}} dy_2 \cdots \int_{y_{min}}^{y_{max}} dy_N y_m \int \prod_{n=1}^{N} e^{-\beta U(y_i, y_{i+1})} \quad (3)$$

2.2. Notation

To reliably distinguish terminal base pairs from internal base pairs we need to establish an unambiguous notation. Consider the following example sequence

$$5' - \text{AGAUAUCU} - 3'$$
$$3' - \text{UCUAUAGA} - 5'$$

where we separate the terminal and internal base pairs

$$\underbrace{\begin{matrix}\text{A}\\ \text{U}\end{matrix}}_{\text{terminal}} \cdot \underbrace{\text{GCAAGU}}_{\text{internal}} \cdot \underbrace{\begin{matrix}\text{U}\\ \text{A}\end{matrix}}_{\text{terminal}}$$

For the terminal base pairs we will use a superscript *, in our example this would be AU* at the 5′-side and UA* at the 3′-side. For Morse potentials, AU* is equivalent to UA*, as well as CG* is to GC*, and will share the same parameters D and λ, see Eq. (1).

For the nearest-neighbour (NN) stacking parameter k there will be a mixed notation of terminal and internal base pairs. The first NN pair of our example sequence would be AU*pGC, that is a terminal AU* followed by an internal GC. The AU*pGC pair is symmetric to CGpUA*:

$$5'\text{-AGAUAUCU} - 3' \leftrightarrow 5'\text{-AGAUAU}\mathbf{CU}\text{-}3'$$
$$3'\text{-UCUAUAGA} - 5' \quad\quad 3'\text{-UCUAUA}\mathbf{GA}\text{-}5'$$

As both can be described by the same stacking parameter k, we keep just the one that precedes alphabetically, in this case AU*pGC. Therefore, as stated above, the stacking parameter for AUpGC NN pairs will be divided into three separate parameters, namely AU*pGC for terminal AU*, AUpGC* for terminal CG* and the internals for which we maintain the original notation AUpGC.

Some base pairs will have only one additional terminal-related parameter as a result of the NN pair symmetry. Such as CGpGC which has only one terminal related NN CGpGC* since it is symmetric to CG*pGC.

2.3. Melting temperature sets

The melting temperatures used here fall into two very different categories: four are at lower sodium concentrations where all sequences are self-complementary and of the same length; and a single set at high sodium concentrations with a mix of self-complementary and non-self-complementary sequences and variable lengths. This requires different theoretical approaches depending on the type of temperature set, and therefore we will distinguish them by low salt (LS) and high salt (HS).

2.3.1. (LS) 71 to 621 mM [Na^+]

We used the set of RNA melting temperatures from Ferreira et al. [52], consisting of 18 RNA duplexes at four different [Na^+] concentrations (71, 121, 221, 621 mM). For each sequence and salt concentrations there are at least 9 measurements at different strand concentrations C_t in the range of 5 μM to 700 μM, see Fig. 1. All sequences are self-complementary and either are 6 or 8 base-pairs (bp) in length.

2.3.2. (HS) 1021 mM [Na^+]

For the higher salt concentration we used the melting temperature set from Xia et al. [51] which was complemented by two sequences from Chen and Znosko [62]. Unlike the LS data, they have varying lengths, include non-self-complementary sequences, are at the same strand concentration and originate from various sources.

2.4. Temperature correlation with melting index

The PB model describes the thermodynamics via a coefficient τ_i, obtained from the partition function Eq. (2), for the ith duplex in the data set which provides temperature prediction T_i'

$$T_i^{'}(P) = a_0 + a_1\tau_i(P) \tag{4}$$

where the coefficients $a_{0,\ 1}$ are calculated via a linear regression of the experimental melting temperatures T_i at a single strand concentration C_t, and P is a set of tentative model parameters. The regression of the coefficients used in Eq. (4) is typically carried out at a single strand concentration C_t. However, for the LS dataset, there are multiple strand concentrations which require a different approach as we will discuss next.

2.4.1. LS temperature regression

Previous studies have confirmed that the resulting model parameters are independent of the strand concentration [41,54]. However, when the melting temperature set involves multiple strand concentrations the regression of Eq. (4) needs to be carried separately for each concentration C_t, that is

$$T_i^{'}(P) = a_0(C_t) + a_1(C_t)\tau_i(P) \tag{5}$$

where the coefficients $a_{0,\ 1}$ are now functions of C_t, which requires a minimum amount of melting temperatures for each value of C_t as otherwise the regression calculation cannot be carried out. In other words, there needs to be subsets of melting temperatures grouped to the same C_t. However, here the dataset has measurements scattered over a wide range of C_t and there is no single subset that was measured at the same C_t, see Fig. 1. This does not represent a problem for the nearest-neighbour model [52], but for the PB model it becomes necessary to group the melting temperatures together to the closest value of C_t.

2.4.2. Strand concentration grouping

Since the melting temperatures scale with $\ln(C_t)$ [63], it makes sense to introduce a logarithmic group index

$$L_f = \frac{\mathrm{round}[f\ln(C_t/C_0)]}{f} \tag{6}$$

where f is a factor that controls the coarseness of groups, and C_0 is a fixed reference concentration taken as 1 μM to ensure that L_f is adimensional. As we will perform a linear regression for each group, we only consider groups with at least 3 elements. For each available melting temperature we work out to which group L_f belongs depending on its C_t and the coarseness factor f which results in n_f groups with a total of N_f members. A small f will create a small number of groups n_f with many elements, while a large f results in many groups with few elements. The upper limit of f is when there are too few melting temperatures per group to perform a meaningful linear regression (at least 3 elements), and the lower limit of f is when there is only a single group that contains all temperatures. Table 1 shows the summary of the logarithmic grouping L_f that is considered in this work for [Na^+] 71 mM. See supplementary tables S6, S7 and S8 for a summary of the remaining LS salt concentrations, and a detailed breakdown in supplementary tables S9–S12.

Table 1
Summary of logarithmic grouping for coarseness f. Shown are the number of groups n_f, the total number of grouped elements N_f, and the total number of ungrouped elements U_f for [Na^+] 71 mM.

f	n_f	N_f	U_f
5.0	22	179	6
4.0	18	180	5
3.0	14	182	3
2.0	10	184	1
1.5	8	184	1
1.4	7	184	1
1.3	7	183	2
1.2	7	184	1
1.1	6	184	1
1.0	5	182	3
0.9	5	184	1
0.8	5	182	3
0.7	4	184	1
0.6	4	184	1
0.5	3	184	1
0.4	3	185	0

Using the logarithmic grouping, we now replace Eq. (5) with

$$T_i^{'}(P) = a_0(L_f) + a_1(L_f)\tau_i(P) \tag{7}$$

and the regression coefficients are obtained independently for each group L_f.

2.5. HS temperature regression

For the HS data, which are all given at the same strand concentration and are available at varying sequence lengths N, the linear regression is

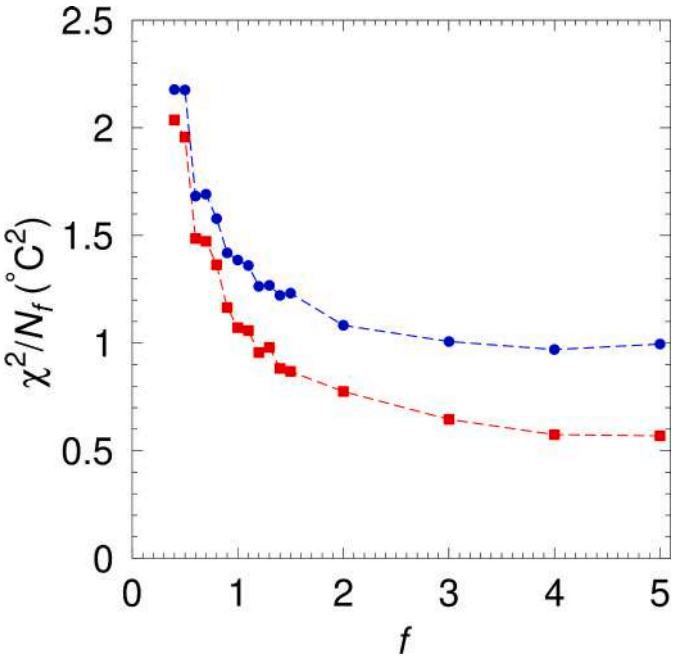


Fig. 2. Final merit function χ^2/N_f as a function of the grouping coarseness factor f for [Na^+] 121 mM. Red boxes and blue circles represent T/I and UN optimizations, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

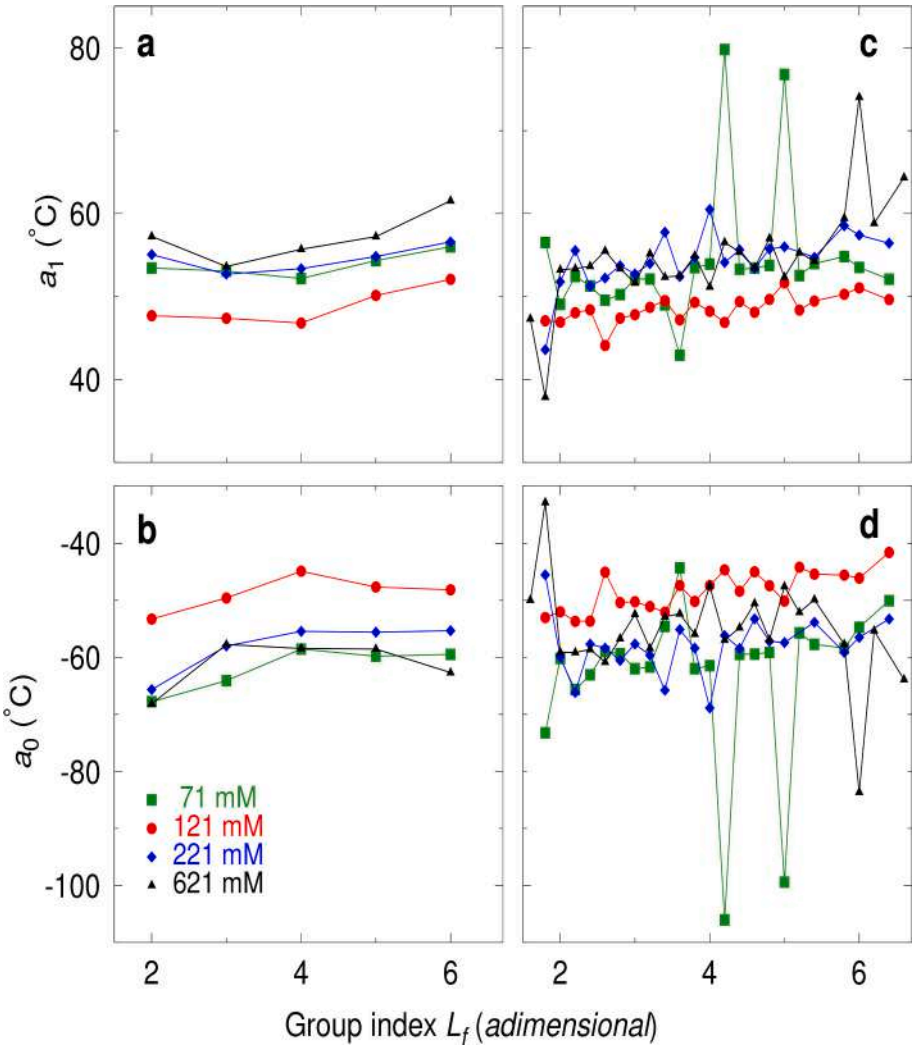


Fig. 3. Regression parameters for (a,b) $f = 1$ and (c,d) $f = 5$ for salt concentrations 71 mM (green boxes), 121 mM (red bullets), 221 mM (blue diamonds) and 621 mM (black triangles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

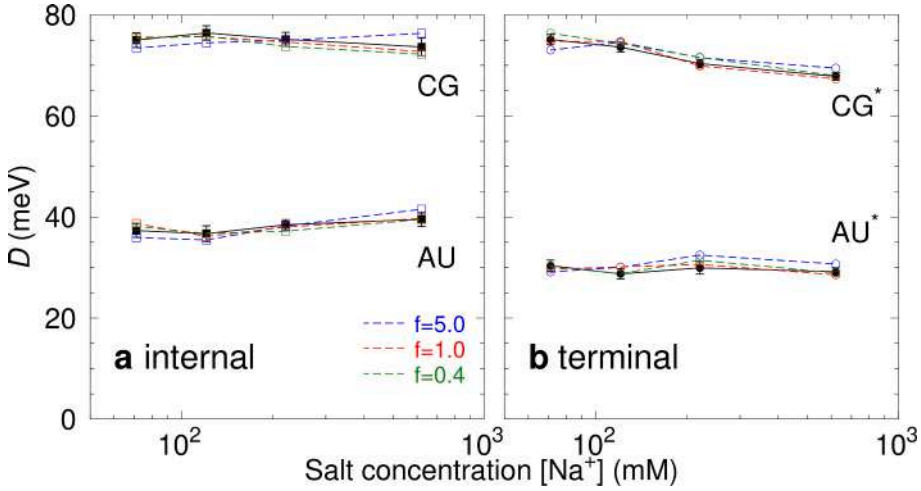


Fig. 4. Morse potentials averaged over all f (black squares) for (a) internal and (b) terminal base pairs. The error bars represent the standard deviation within the f sets. Specific results for $f = 0.4,1,5$ are shown in green, red and blue, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

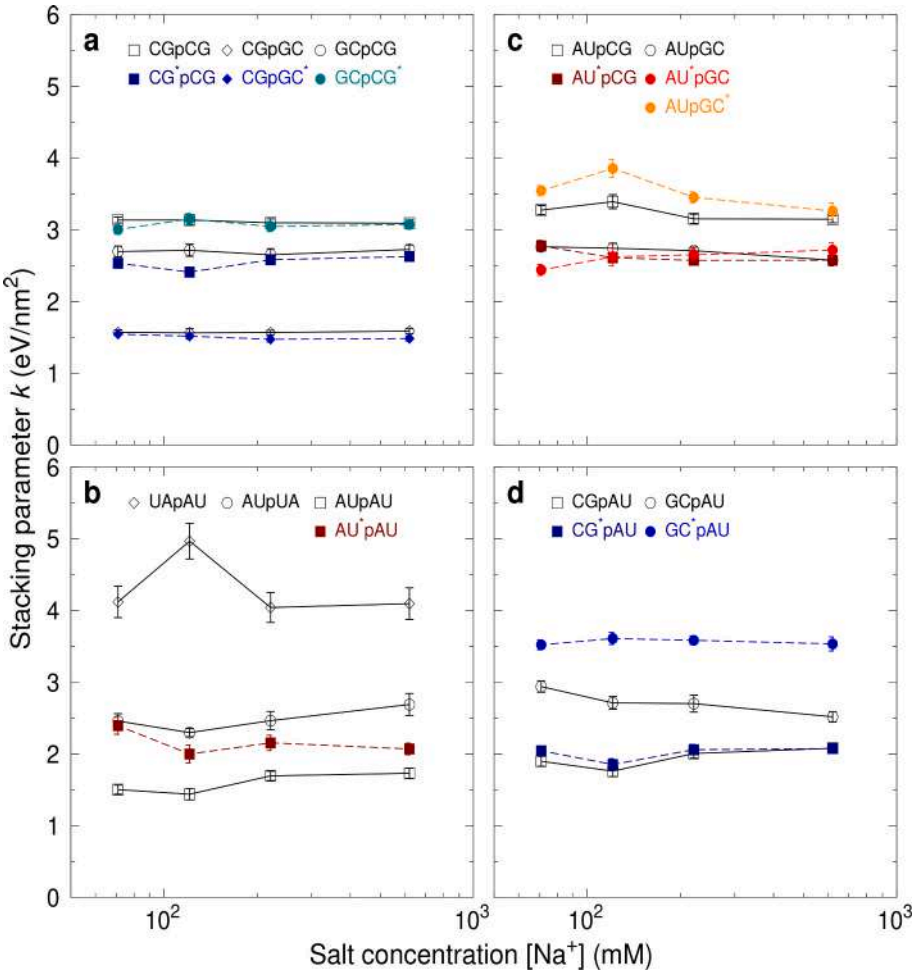


Fig. 5. Stacking potentials averaged over all *f*. Panels (a,b) show the symmetric NN and panels (c,d) the non-symmetric NN. The error bars represent the standard deviation within the *f* sets. Dashed lines are for NNs with terminal base pairs.

performed separately for each group of base pair length N [55].

$$T_i^{'}(P) = a_0(N) + a_1(N)\tau_i(P), \tag{8}$$

similarly as used in our previous work [41], and gives better results than the single regression Eq. (4).

2.6. Optimization

The parameter sets P needed for the calculation of the melting index $\tau_i(P)$, Eqs. (7) and (8), contains the model parameters used in the Hamiltonian Eq. (1) for each type of base pair and nearest-neighbour present in the sequence set. Therefore, we will need to find the optimal set of L parameters, $P_j = \{p_1, p_2, \ldots, p_L\}$ that will provide the temperature predictions $T_i'(P)$ that are closest to the experimental temperatures T_i. The P parameters are varied until we minimize the squared difference

$$\chi_j^2 = \sum_{i=1}^{M} \left[T_i^{'}(P_j) - T_i\right]^2. \tag{9}$$

where P_j is the jth tentative set of parameters and M is the number of experimental melting temperatures. Each parameter within the P_j is sampled between $0.1p_u$ and $1.1p_u$, where p_u is the uniform parameter calculated previously for high salt concentration [54]. For LS we use $M = N_f$ which is the total number of grouped temperatures for a given coarseness factor f. The numerical parameter optimization is performed by a downhill simplex multidimensional minimization algorithm [64].

We will also refer to another quality parameter which is average melting temperature deviation

$$\langle \Delta T \rangle = \frac{1}{M} \sum_{i=1}^{M} \left| T_i^{'}(P_j) - T_i \right| \tag{10}$$

As a result of the terminal/internal (T/I) notation, we will be dealing with 4 Morse potentials (2 internal, 2 terminal) and 26 NN stacking potentials (10 internal and 16 terminal), representing $L = 30$ parameters. For comparison, we will also perform all calculations without the distinction between terminal/internal which we will call uniform (UN) parameters and represents $L = 12$ variables. Next, we will detail the optimization steps used here.

2.6.1. MR1 (LS and HS)

The first minimization round (MR1) of the parameter optimization was performed by varying the initial Morse and stacking potentials [41,54] randomly over an interval which averages to the initial values. For the T/I scheme, initial parameters are assumed to have the same values, although designated by different variables (AU and AU*), so that they can vary separately. The minimization procedure was repeated 100 times for each f. The same procedure was carried out for HS, the only difference being the use of Eq. (8), applied during the minimization.

2.6.2. MR2 (LS and HS)

For the next round, we calculate the average of those parameters with lowest χ^2 from MR1 to be used as a new fixed initial set of parameters for a second round of minimizations (MR2), following the same

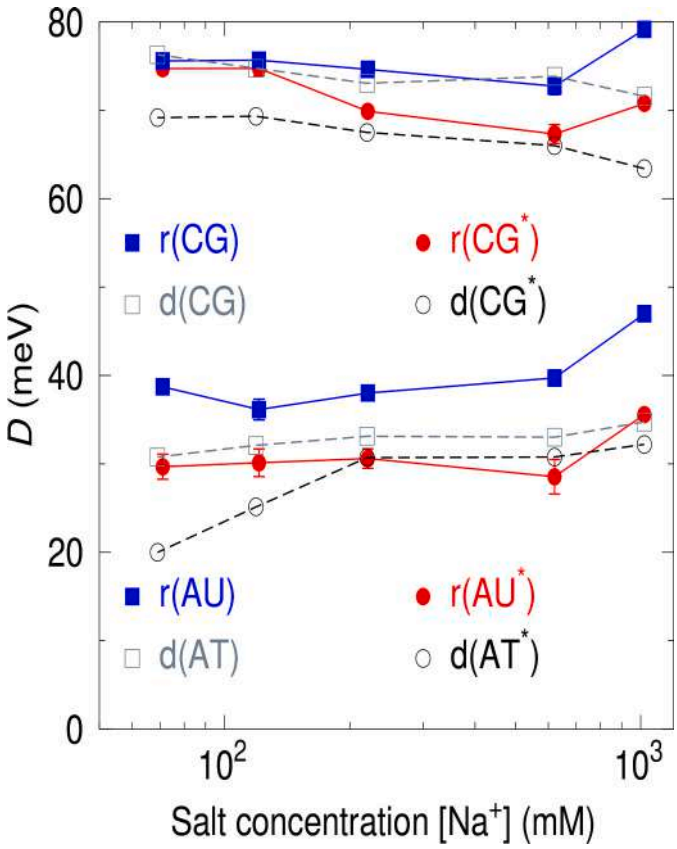


Fig. 6. Average Morse potentials as function of salt concentration. LS results are for $f = 1$. Error bars were estimated in the EU-LS/HS minimization round. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal) [53]. The lines connecting the data points are only intended as guide to the eyes.

procedure described for MR1. Here, that is a way to refine the parameters and reduce the difference between each minimization and consequently reduce the parameter standard deviation. Once more, this was repeated 100 times for HS and for each f (LS).

2.6.3. EU-HS

The last step is to evaluate the impact of the experimental uncertainty (EU) by changing the temperatures of the dataset by small amounts, such that the standard deviation between the original set and the optimized set approaches the declared experimental uncertainty. We then run again the minimization procedure, however, unlike for MR1/2 we keep the initial parameters fixed and only disturb the melting temperatures. A standard deviation of 0.5°C was considered and the minimization carried out for 100 rounds for each f.

2.6.4. EU-LS

For the LS-type datasets, in addition to the impact of the melting temperature uncertainty, we also need to evaluate the impact of the strand concentration grouping procedure described in section 2.4.2. For this, we proceeded in a very similar way as for the temperature perturbation described in the previous section: we disturb the C_t by small amounts and rerun the minimization again. The estimated uncertainty for C_t was reported as 5%, using absorbance reading at 260 nm at 80 °C [62]. Again, this was repeated 100 times for each f and gives us an estimate of the uncertainty over the calculated parameters. Therefore, the final results shown here are the averages over these minimizations. All those steps were carried out independently for each LS salt concentration.

2.7. Validation

For a validation set we collected 45 sequences and their melting temperatures at various salt concentrations and species concentrations from Refs. 65–71, which are shown in supplementary Table S13.

3. Results

3.1. Logarithmic groups

The available LS melting temperatures are scattered over a wide range of strand concentrations C_t. Here, we will attempt to group these temperatures, according to a logarithmic grouping scheme described in section 2.4.2, in a scheme which similar to a histogram. The first question we need to address is how this logarithmic grouping impacts the parameter optimization and what is best the coarseness factor f. If f is too small, the melting temperatures are separated into very few large groups, if it is too large they end up scattered into many sparsely populated groups. To answer this question, we performed all minimization independently for f ranging between 0.4 and 5.0, see Table 1 and supplementary tables S9–S12.

In Fig. 2 we show the final merit function χ^2 that was minimized during rounds MR1, MR2 and LS-EU for the UN (blue circles) and T/I minimization (red boxes). Both show the same behaviour as function of f. χ^2 levels off after $f = 2$ and there is little difference between 3 and 5. The regression coefficients, $a_{0,\,1}$, Eq. (7), for $f = 1$ and 5 are shown in Fig. 3. See Figs. S1— S14. At $f = 1$, both a_0 and a_1 show a relatively uniform behaviour for all salt concentrations, with a_1 increasing slightly with L_f. However, for the larger $f = 5$ this uniformity is lost due to the low number of melting temperatures in some L_f groups. This is especially evident for the lowest salt concentration 71 mM, see also the first column in supplementary table S9.

In Fig. 4 we show the Morse potentials for three f factors, namely 0.4 (3 groups), 1.0 (5 groups) and 5.0 (21 to 24 groups, depending on salt concentration). Comparatively we also show the Morse average over the results for all calculated f factors. The standard deviation within the sets is marginally small with $f = 5$ showing the most pronounced deviation from the average (5 meV). Moreover, for the stacking parameters the higher deviation occurs for the nearest-neighbour UApAU and the rest remain nearly equal within the average. We also compute the average for the stacking potentials which is shown in the Fig. 5. Even while displaying more unstable regression parameters, higher values of f still derive parameters consistent within the set and with previous works [52] and on the average produce similar results.

Finally, to answer the question of the most adequate coarseness factor f, it would seem that balancing a low merit factor χ^2 with uniform regression coefficients $a_{0,\,1}$ points toward an f around 1. It is desirable to deal with monotonic regression coefficients as they allow us to interpolate new coefficient for missing salt concentrations which is not possible for large f. On the other hand, for the optimized parameters shown in Figs. 4 and 5, the actual value of the coarseness factor f appears to be of little importance. Therefore, for the remainder of this article we will discuss the results for $f = 1$, unless noted otherwise.

Since the T/I minimization has a substantially lower merit factor χ^2 than the UN parameters, Fig. 2, there is a possibility of overfitting for the T/I minimization due to the larger number of parameters. To verify if overfitting may have occurred we apply these parameters to the prediction of melting temperatures of an independent validation set of sequences that was not used for the optimization, see supplementary table S13. Using UN parameters, for $f = 1$, we obtain $\langle \Delta T \rangle = 1.88$ °C and a $\chi^2 = 324.79$ °C^2. However, using the parameters derived from T/I minimization, also for $f = 1$, we obtain an important reduction, $\langle \Delta T \rangle = 1.73$ °C and a $\chi^2 = 270.10$ °C^2, which gives us confidence that no overfitting occurred for the T/I minimization.

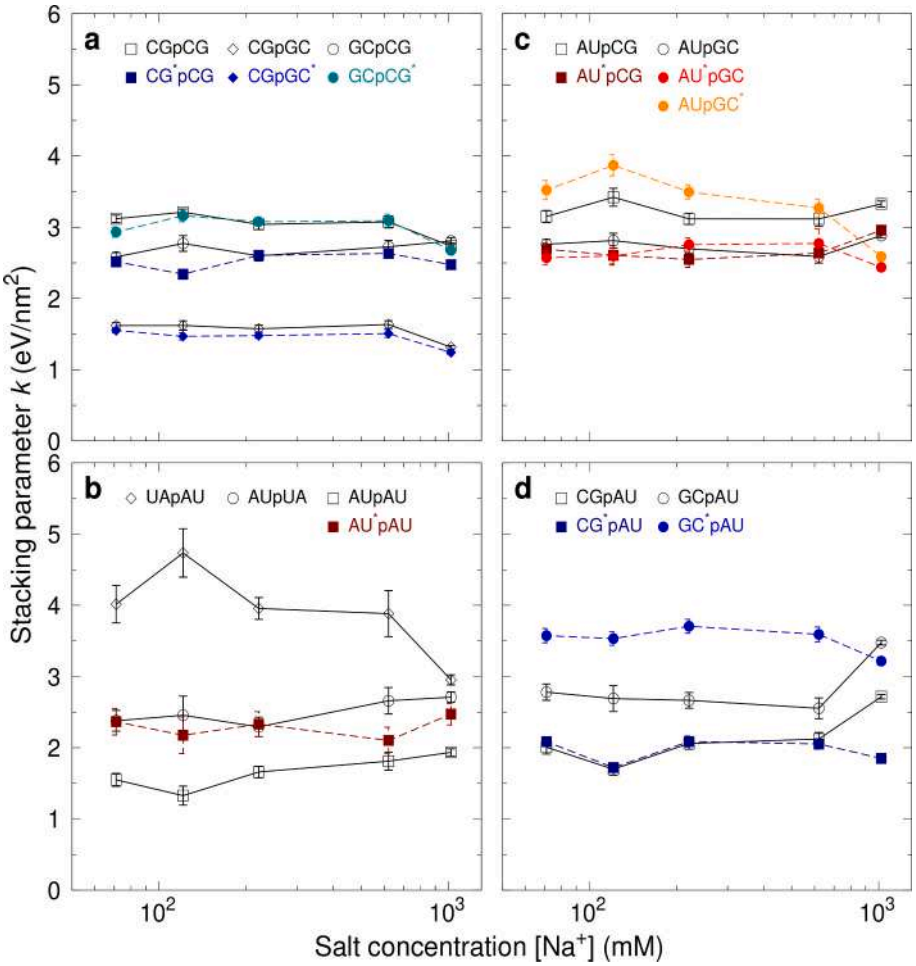


Fig. 7. Average stacking potentials as function of sodium concentration. LS results are for $f = 1$. Panels (a,b) show the symmetric NN and panels (c,d) the asymmetric NN. Error bars were estimated in the EU-LS/HS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs. The lines connecting the data points are only intended as guide to the eyes.

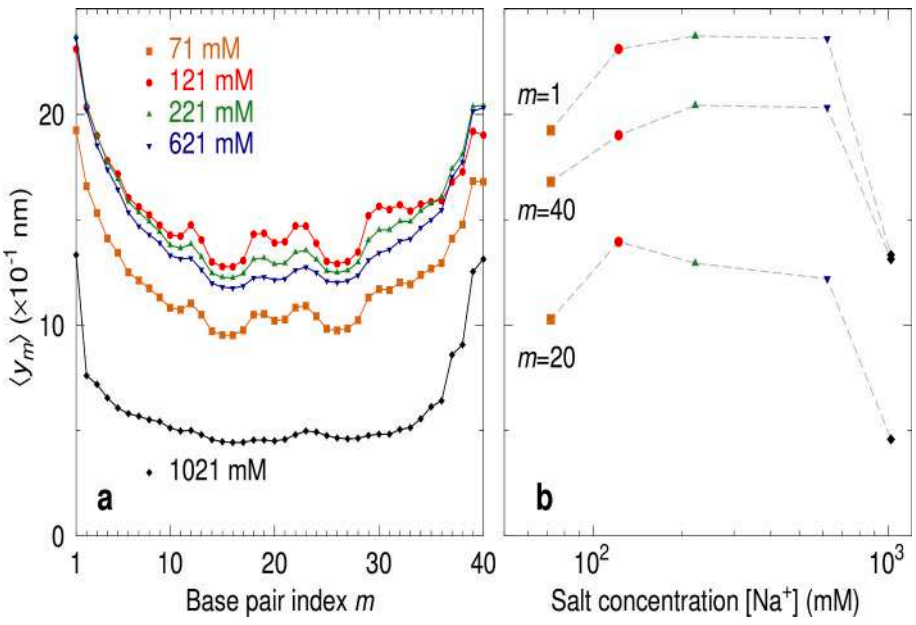


Fig. 8. (a) Average opening profile for the dsRNA sequence from Ref. 74. Squares, diamonds, triangle up, triangle down and bullets represent the average opening for RNA in 71, 121, 221, 621 and 1021 mM [Na^+] for the T/I calculation, respectively. Calculation was carried out at 300 K which has no relation to the melting temperature. (b) Comparative opening for the 5′ terminal ($m = 1$), central ($m = 20$) and 3′ terminal ($m = 40$) base-pair as function of salt concentration.

3.2. Parameters at $f = 1$

In Fig. 6 we show the final average Morse potentials for $f = 1$. For comparison, we also show previous salt-dependent results for DNA from Ref. 53. Some authors have suggested a theoretical logarithmic dependence of the Morse potential as function of sodium concentration [72], however this is not supported by our data, Fig. 6. The internal base-pair Morse potentials are always larger than those for the terminal base pairs, which is consistent with our previous calculations for DNA which were calculated at a very low strand concentration (2 μM). The major difference to the DNA results is that we do not observe a reduced Morse potential for terminal r(AU*) at very low salt concentrations. Therefore, it would seem that the hydrogen bonding of r(AU*) is less susceptible to the sodium concentration. However, there is still a considerable difference between internal r(AU) and terminal r(AU*) Morse potentials which makes the terminal base pairs even more vulnerable to end-fraying. The Morse potentials of internal r(AU) base pairs are consistently higher than their d(AT) counterparts, confirming our previous findings for high sodium concentrations [41]. For HS we observe an increase in all Morse potentials, however due to the very different compositions of the LS and HS data sets it is presently not possible to unambiguously attribute this increase to a property of the base pairs at high sodium concentration.

For r(CG*) Morse potentials we found similar values to r(CG) at very low salt concentrations which is not observed for their DNA analogs. In other words, RNA appears to be less susceptible to end-fraying than DNA at low salt concentrations. We attribute the shift toward higher Morse potentials for HS in Fig. 6 to the substantial difference between the LS and HS datasets, as described in the methods sections.

The calculated stacking parameters are shown in Fig. 7, grouped into symmetric and asymmetric NNs. Note that not all combinations of NNs with terminal base pairs were present in the dataset, therefore not all terminal analogues of internal NNs could be calculated. Except for UApAU NNs, most stacking interactions show little change with salt concentrations. Similarly, our previous results for DNA have shown little dependence of stacking with sodium except for AT*pAT, TApAT* and ATpGC* [53].

4. Discussion

Due to the non-linear Hamiltonian, Eq. (1), it is not straightforward to visualize how the Morse and stacking potentials will affect the opening of the base pairs. For this, we use the average displacement profiles using Eq. (3), that is the $\langle y_m \rangle$ where m is the base pair index. The average displacements indicate which base pairs are likely to open first at a given temperature and can be qualitatively related to the to the root-mean-squared distance (RMSD) or root-mean-squared fluctuation (RMSF) used in MD and coarse-grained simulations [73].

In Fig. 8a we show the average opening for the sequence used in coarse-grained calculations reported in Refs. 37, 74. Similarly to the coarse-grained calculations [37] we observe larger fluctuations at the terminal base pairs which increase up to intermediate salt concentrations. However, for higher salt concentrations we observe a saturation and even a substantial drop in $\langle y_m \rangle$ at 1021 mM [Na^+]. This saturation between 121 and 621 mM is better seen in Fig. 8b where we show $\langle y_m \rangle$ as a function of sodium concentration for three locations in the sequence. It is unclear if the reduction of $\langle y_m \rangle$ at HS in the Fig. 8a is due to the large difference between the melting temperature datasets or it is intrinsic of the duplex at this salt concentration. Nevertheless, it does not support the continuous increase in RMSF with salt concentration calculated by Jin et al. [37]. The terminal 5′ shows a considerably wider opening than the 3′ end, see Fig. 8b. This is dissimilar to the calculations by Jin et al. [37], yet consistent with results from O'Toole et al. [75]. Results from tightly bound ion models show an increase of ion binding which slowly saturates [39,76], which is consistent with the saturation observed in Fig. 8b at intermediate sodium concentrations. However, at high sodium concentrations there is a clear departure from the observed saturation at intermediate concentrations which might be attributable to the differences between the HS and LS data sets. The salt dependent behaviour of $\langle y_m \rangle$ is not sequence specific and can equally be observed for homogeneous sequences as exemplified in supplementary fig. S43.

In Fig. 9 we show an example for sequence II from Ref. 45, comparing RNA to the equivalent DNA sequence at 121 mM [Na^+]. The calculation temperature in this case was 180 K, which has no relation to the melting temperature correlation of Eqs. (7) and (8). Fig. 9 shows that for internal base pairs, $\langle y_m \rangle$ is somewhat larger at the r(AU) tract than the equivalent d(AT) tract, despite the larger r(AU) Morse potential. The reason for this is that the internal $\langle y_m \rangle$ is pushed up by the terminal r(CG*), which illustrates that the cooperativity of the base pairs at the termini affects the internal base pairs as well. In the specific case of sequence II from Snoussi and Leroy [45], their NMR measurements indicated a shorter r(AU) lifetime than for d(AT), which would be consistent with a larger displacement for r(AU) seen in Fig. 9. On the other hand, contrary to their results, we observe larger base-fraying for r(CG*) which can be understood from the larger difference between internal and terminal Morse potentials for CG at this salt concentration.

5. Conclusions

We introduce a new technique to parametrise the PB model at multiple strand concentrations by the use of a logarithmic groups. The resulting parameters show little dependence on the coarseness of the grouping which evidences that this technique is robust, and enabled us to make use of a large salt dependent RNA melting temperature dataset. We calculated new salt dependent PB parameters, including specific parameters for the sequence terminals. Unlike d(AT), the Morse potentials for r(AU), which are related to hydrogen bonding, showed no important reduction at low sodium concentrations. Most stacking interactions show little change with salt concentration, however for some terminal contexts stronger stacking interactions were found, similar to

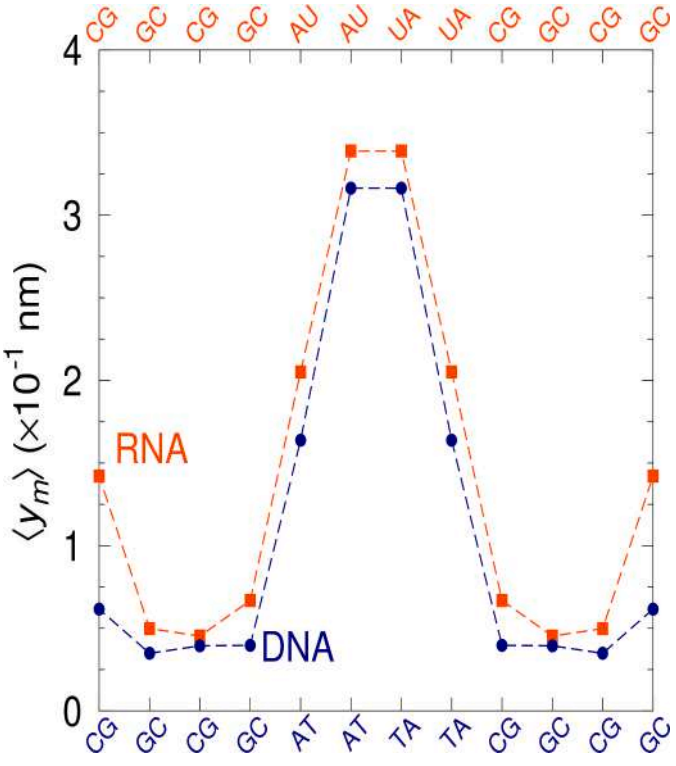


Fig. 9. Average opening profile for sequence IV from Ref. 45. Open and closed squares (bullets), represent the average opening for RNA (DNA) in 121 mM [Na^+], for the UN and T/I calculation, respectively.

our previous study for DNA [53].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the agencies that funded this research: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brasil, Finance Code 001).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bpc.2021.106551.

References

[1] F. Michelini, A.P. Jalihal, S. Francia, C. Meers, Z.T. Neeb, F. Rossiello, U. Gioia, J. Aguado, C. Jones-Weinert, B. Luke, et al., From "cellular" RNA to "smart" RNA: multiple roles of RNA in genome stability and beyond, Chem. Rev. 118 (8) (2018) 4365–4403.
[2] S. Nellimarla, K.L. Mossman, Extracellular dsRNA: its function and mechanism of cellular uptake, J. Interf. Cytokine Res. 34 (6) (2014) 419–426.
[3] T. Kuwabara, J. Hsieh, K. Nakashima, K. Taira, F.H. Gage, A small modulatory dsRNA specifies the fate of adult neural stem cells, Cell 116 (6) (2004) 779–793.
[4] G.J. Hannon, RNA interference, Nature 418 (6894) (2002) 244–251.
[5] G. Meister, T. Tuschl, Mechanisms of gene silencing by double-stranded RNA, Nature 431 (7006) (2004) 343–349.
[6] S. Akira, K. Takeda, Toll-like receptor signalling, Nat. Rev. Immunol. 4 (7) (2004) 499–511.
[7] J. Lipfert, S. Doniach, R. Das, D. Herschlag, Understanding nucleic acid–ion interactions, Annu. Rev. Biochem. 83 (2014) 813–841.
[8] S.A. Woodson, Metal ions and RNA folding: a highly charged topic with a dynamic future, Curr. Opin. Chem. Biol. 9 (2) (2005) 104–109.
[9] D.E. Draper, Folding of RNA tertiary structure: linkages between backbone phosphates, ions, and water, Biopoly. 99 (12) (2013) 1105–1113.
[10] E. Koculi, C. Hyeon, D. Thirumalai, S.A. Woodson, Charge density of divalent metal cations determines RNA stability, J. Am. Chem. Soc. 129 (9) (2007) 2676–2682.
[11] P.A. Lemaire, E. Anderson, J. Lary, J.L. Cole, Mechanism of PKR activation by dsRNA, J. Mol. Biol. 381 (2) (2008) 351–360.
[12] K. Xi, F.-H. Wang, G. Xiong, Z.-L. Zhang, Z.-J. Tan, Competitive binding of Mg^{2+} and Na^{+} ions to nucleic acids: from helices to tertiary structures, Biophys. J. 114 (8) (2018) 1776–1790.
[13] Y.-Y. Wu, Z.-L. Zhang, J.-S. Zhang, X.-L. Zhu, Z.-J. Tan, Multivalent ion-mediated nucleic acid helix-helix interactions: RNA versus DNA, Nucleic Acids Res. 43 (12) (2015) 6156–6165.
[14] S. Kirmizialtin, R. Elber, Computational exploration of mobile ion distributions around RNA duplex, J. Phys. Chem. B 114 (24) (2010) 8207–8220.
[15] A.V. Drozdetski, I.S. Tolokh, L. Pollack, N. Baker, A.V. Onufriev, Opposing effects of multivalent ions on the flexibility of DNA and RNA, Phys. Rev. Lett. 117 (2) (2016), 028101.
[16] Z.-J. Tan, S.-J. Chen, Importance of diffuse metal ion binding to RNA, Met. Ions Life Sci. 9 (2011) 101.
[17] E.D. Holmstrom, J.L. Fiore, D.J. Nesbitt, Thermodynamic origins of monovalent facilitated RNA folding, Biochem. 51 (18) (2012) 3732–3743.
[18] J.L. Chen, A.L. Dishler, S.D. Kennedy, I. Yildirim, B. Liu, D.H. Turner, M.J. Serra, Testing the nearest neighbor model for canonical RNA Base pairs: revision of GU parameters, Biochem. 51 (16) (2012) 3508–3522.
[19] L.G. Laing, T.C. Gluick, D.E. Draper, Stabilization of RNA structure by mg ions: specific and non-specific effects, J. Mol. Biol. 237 (5) (1994) 577–587.
[20] A. Pyle, Metal ions in the structure and function of RNA, JBIC J. Biol. Inorg. Chem. 7 (7–8) (2002) 679–690.
[21] D.E. Draper, A guide to ions and RNA structure, RNA 10 (3) (2004) 335–343.
[22] D. Lambert, D. Leipply, R. Shiman, D.E. Draper, The influence of monovalent cation size on the stability of RNA tertiary structures, J. Mol. Biol. 390 (4) (2009) 791–804.
[23] M. Egli, Structural aspects of nucleic acid analogs and antisense oligonucleotides, Angew. Chem., Int. Ed. Engl. 35 (17) (1996) 1894–1909.
[24] V.B. Chu, Y. Bai, J. Lipfert, D. Herschlag, S. Doniach, A repulsive field: advances in the electrostatics of the ion atmosphere, Curr. Opin. Struct. Biol. 12 (6) (2008) 619–625. ISSN 1367–5931, https://doi.org/10.1016/j.cbpa.2008.10.010, http://www.sciencedirect.com/science/article/pii/S1367593108001816. biopolymers/Model Systems.
[25] G.L. Conn, A.G. Gittis, E.E. Lattman, V.K. Misra, D.E. Draper, A compact RNA tertiary structure contains a buried backbone–K+ complex, J. Mol. Biol. 318 (4) (2002) 963–973.
[26] Y.V. Bukhman, D.E. Draper, Affinities and selectivities of divalent cation binding sites within an RNA tertiary structure, J. Mol. Biol. 273 (5) (1997) 1020–1031.
[27] K. Takamoto, Q. He, S. Morris, M.R. Chance, M. Brenowitz, Monovalent cations mediate formation of native tertiary structure of the *Tetrahymena thermophila* ribozyme, Nat. Struc. Biol. 9 (12) (2002) 928–933.
[28] R. Shiman, D.E. Draper, Stabilization of RNA tertiary structure by monovalent cations, J. Mol. Biol. 302 (1) (2000) 79–91.
[29] L.-T. Da, F. Pardo-Avila, L. Xu, D.-A. Silva, L. Zhang, X. Gao, D. Wang, X. Huang, Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue, Nature Comm. 7 (2016) 11244.
[30] J.F. Sydow, F. Brueckner, A.C. Cheung, G.E. Damsma, S. Dengl, E. Lehmann, D. Vassylyev, P. Cramer, Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA, Mol. Cell 34 (6) (2009) 710–721.
[31] A. Serganov, E. Nudler, A decade of riboswitches, Cell 152 (1–2) (2013) 17–24.
[32] W. Huang, J. Kim, S. Jha, F. Aboul-Ela, The impact of a ligand binding on strand migration in the SAM-I riboswitch, PLoS Comput. Biol. 9 (5) (2013), e1003069.
[33] I. Bešťeová, M. Otyepka, K. Réblová, J. Šponer, Dependence of A-RNA simulations on the choice of the force field and salt strength, Phys. Chem. Chem. Phys. 11 (45) (2009) 10701–10711.
[34] I. Bešťeová, P. Banáš, P. Kührová, P. Koşinová, M. Otyepka, J. Šponer, Simulations of A-RNA duplexes. The effect of sequence, solute force field, water model, and salt concentration, J. Phys. Chem. B 116 (33) (2012) 9899–9916.
[35] J. Virtanen, T. Sosnick, K. Freed, Ionic strength independence of charge distributions in solvation of biomolecules, J. Chem. Phys. 141 (22) (2014), 12B604_1.
[36] L. Bao, J. Wang, Y. Xiao, Dynamics of metal ions around an RNA molecule, Phys. Rev. E 99 (1) (2019), 012420.
[37] L. Jin, Y.-Z. Shi, C.-J. Feng, Y.-L. Tan, Z.-J. Tan, Modeling structure, stability, and flexibility of double-stranded RNAs in salt solutions, Biophys. J. 115 (8) (2018) 1403–1416.
[38] T.R. Einert, R.R. Netz, Theory for RNA folding, stretching, and melting including loops and salt, Biophys. J. 100 (11) (2011) 2745–2753.
[39] Z.-J. Tan, S.-J. Chen, RNA helix stability in mixed Na+/Mg2+ solution, Biophys. J. 92 (10) (2007) 3615–3632.
[40] Y.-Z. Shi, F.-H. Wang, Y.-Y. Wu, Z.-J. Tan, A coarse-grained model with implicit salt for RNAs: predicting 3D structure, stability and salt effect, J. Chem. Phys. 141 (10) (2014), 09B606_1.
[41] G. Weber, Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures, Nucleic Acids Res. 41 (2013) e30, https://doi.org/10.1093/nar/gks964. URL, http://nar.oxfordjournals.org/content/41/1/e30.
[42] G. Weber, N. Haslam, J.W. Essex, C. Neylon, Thermal Equivalence of DNA Duplexes for probe design, J. Phys. Condens. Matter 21 (2009) 034106, https://doi.org/10.1088/0953-8984/21/3/034106.
[43] K. Clanton-Arrowood, J. McGurk, S.J. Schroeder, 3′ terminal nucleotides determine thermodynamic stabilities of mismatches at the ends of RNA helices, Biochemistry 47 (50) (2008) 13418–13427.
[44] M.E. Burkard, R. Kierzek, D.H. Turner, Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs, J. Mol. Biol. 290 (5) (1999) 967–982.
[45] K. Snoussi, J.L. Leroy, Imino proton exchange and base-pair kinetics in RNA duplexes, Biochemistry 40 (30) (2001) 8898–8904.
[46] M. Zgarbová, M. Otyepka, J. Sponer, F. Lankas, P. Jurečka, Base pair fraying in molecular dynamics simulations of DNA and RNA, J. Chem. Theory Comput. 10 (8) (2014) 3177–3189, https://doi.org/10.1021/ct500120v.
[47] G. Pinamonti, F. Paul, F. Noé, A. Rodriguez, G. Bussi, The mechanism of RNA base fraying: molecular dynamics simulations analyzed with core-set Markov state models, J. Chem. Phys. 150 (15) (2019) 154123.
[48] Y.-L. Zhang, W.-M. Zheng, J.-X. Liu, Y.Z. Chen, Theory of DNA melting based on the Peyrard-Bishop model, Phys. Rev. E 56 (6) (1997) 7100–7115.
[49] P. Šulc, F. Romano, T.E. Ouldridge, L. Rovigatti, J.P. Doye, A.A. Louis, Sequence-dependent thermodynamics of a coarse-grained DNA model, J. Chem. Phys. 137 (13) (2012) 135101, https://doi.org/10.1063/1.4754132.
[50] P. Šulc, F. Romano, T.E. Ouldridge, J.P. Doye, A.A. Louis, A nucleotide-level coarse-grained model of RNA, J. Chem. Phys. 140 (23) (2014) 235102, https://doi.org/10.1063/1.4881424.
[51] T. Xia, J. SantaLucia Jr., M.E. Burkard, R. Kierzek, S.J. Schroeder, X. Jiao, C. Cox, D.H. Turner, Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick Base pairs, Biochem. 37 (1998) 14719–14735.
[52] I. Ferreira, E.A. Jolley, B.M. Znosko, G. Weber, Replacing salt correction factors with optimized RNA nearest-neighbour enthalpy and entropy parameters, Chem. Phys. 521 (2019) 69–76, https://doi.org/10.1016/j.chemphys.2019.01.016. URL, https://www.sciencedirect.com/science/article/abs/pii/S0301010418311200.
[53] I. Ferreira, T.D. Amarante, G. Weber, DNA terminal base pairs have weaker hydrogen bonds especially for AT under low salt concentration, J. Chem. Phys. 143 (2015) 175101, https://doi.org/10.1063/1.4934783.
[54] G. Weber, J.W. Essex, C. Neylon, Probing the Microscopic Flexibility of DNA from Melting Temperatures, Nat. Phys. 5 (2009) 769–773, https://doi.org/10.1038/nphys1371.

[55] G. Weber, N. Haslam, N. Whiteford, A. Prügel-Bennett, J.W. Essex, C. Neylon, Thermal equivalence of DNA duplexes without melting temperature calculation, Nat. Phys. 2 (2006) 55–59, https://doi.org/10.1038/nphys189.
[56] C.V. Bizarro, A. Alemany, F. Ritort, Non-specific binding of Na^{+} and Mg^{2+} to RNA determined by force spectroscopy methods, Nucleic Acids Res. 40 (14) (2012) 6922–6935.
[57] S. Guilbaud, L. Salomé, N. Destainville, M. Manghi, C. Tardin, Dependence of DNA persistence length on ionic strength and ion type, Phys. Rev. Lett. 122 (2) (2019), 028102.
[58] E. Trizac, T. Shen, Bending stiff charged polymers: the electrostatic persistence length, Europhys. Lett. 116 (1) (2016) 18007.
[59] J.-H. Jeon, W. Sung, F.H. Ree, A semiflexible chain model of local denaturation in double-stranded DNA, J. Chem. Phys. 124 (16) (2006) 164905.
[60] M. Peyrard, A.R. Bishop, Statistical mechanics of a nonlinear model for DNA denaturation, Phys. Rev. Lett. 62 (23) (1989) 2755–2757.
[61] T. Dauxois, M. Peyrard, A.R. Bishop, Entropy-driven DNA denaturation, Phys. Rev. E 47 (1) (1993) R44–R47.
[62] Z. Chen, B.M. Znosko, Effect of sodium ions on RNA duplex stability, Biochem. 52 (42) (2013) 7477–7485, https://doi.org/10.1021/bi4008275.
[63] S. Schreiber-Gosche, R.A. Edwards, Thermodynamics of oligonucleotide duplex melting, J. Chem. Educ. 86 (5) (2009) 644.
[64] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C, Cambridge University Press, Cambridge, 1988.
[65] S.-H. Hung, Q. Yu, D.M. Gray, R.L. Ratliff, Evidence from CD spectra that d (purine)-r(pyrimidine) and r(purine)-d(pyrimidine) hybrids are in different structural classes, Nucleic Acids Res. 22 (20) (1994) 4326–4334.
[66] S. Nakano, M. Fujimoto, H. Hara, N. Sugimoto, Nucleic acid duplex stability: influence of base composition on cation effects, Nucleic Acids Res. 27 (14) (1999) 2957.
[67] J. Lisowiec-Wachnicka, B.M. Znosko, A. Pasternak, Contribution of 3′T and 3′TT overhangs to the thermodynamic stability of model siRNA duplexes, Biophys. Chem. 246 (2019) 35–39.
[68] J.I. Gyi, G.L. Conn, A.N. Lane, T. Brown, Comparison of the thermodynamic stabilities and solution conformations of DNA RNA hybrids containing purine-rich and pyrimidine-rich strands with DNA and RNA duplexes, Biochem. 35 (38) (1996) 12538–12548, https://doi.org/10.1021/bi960948z.
[69] A. Pasternak, J. Wengel, Thermodynamics of RNA duplexes modified with unlocked nucleic acid nucleotides, Nucleic Acids Res. 38 (19) (2010) 6697–6706, https://doi.org/10.1093/nar/gkq561.
[70] S. Wang, E.T. Kool, Origins of the large differences in stability of DNA and RNA helixes: C-5 methyl and 2′-hydroxyl effects, Biochem. 34 (12) (1995) 4125–4132.
[71] S.J. Schroeder, D.H. Turner, Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA, Biochem. 39 (31) (2000) 9257–9274.
[72] A. Singh, N. Singh, Effect of Salt Concentration on the Stability of Heterogeneous DNA, Phys. A (Amsterdam, Neth.) 419 (2015) 328–334.
[73] L.M. Oliveira, A.S. Long, T. Brown, K.R. Fox, G. Weber, Melting temperature measurement and mesoscopic evaluation of single, double and triple DNA mismatches, Chem. Sci. 11 (2020) 8273–8287, https://doi.org/10.1039/d0sc01700k. URL, https://pubs.rsc.org/en/content/articlelanding/2020/SC/D0SC01700K.
[74] L. Bao, X. Zhang, Y.-Z. Shi, Y.-Y. Wu, Z.-J. Tan, Understanding the relative flexibility of RNA and DNA duplexes: stretching and twist-stretch coupling, Biophys. J. 112 (6) (2017) 1094–1104.
[75] A.S. O'Toole, S. Miller, M.J. Serra, Stability of 3′ double nucleotide overhangs that model the 3′ ends of siRNA, RNA 11 (4) (2005) 512–516.
[76] Z.-J. Tan, S.-J. Chen, Nucleic acid helix stability: effects of salt concentration, cation valence and size, and chain length, Biophys. J. 90 (4) (2006) 1175–1190.

3 Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

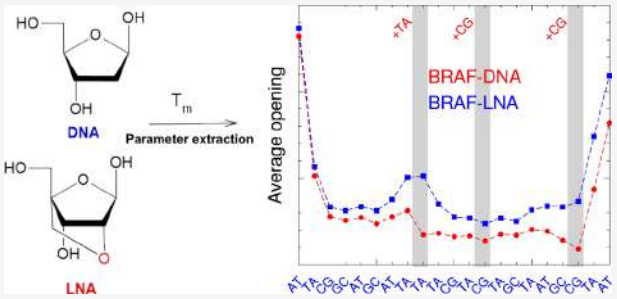
Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design

Izabela Ferreira, Sofie Slott, Kira Astakhova,* and Gerald Weber*

Cite This: *J. Chem. Inf. Model.* 2021, 61, 3615−3624 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The use of mesoscopic models to describe the thermodynamic properties of locked nucleic acid (LNA)-modified nucleotides can provide useful insights into their properties, such as hydrogen-bonding and stacking interactions. In addition, the mesoscopic parameters can be used to optimize LNA insertion in probes, to achieve accurate melting temperature predictions, and to obtain duplex opening profiles at the base-pair level. Here, we applied this type of model to parameterize a large set of melting temperatures for LNA-modified sequences, from published sources, covering all possible nearest-neighbor configurations. We have found a very large increase in Morse potentials, which indicates very strong hydrogen bonding as the main cause of improved LNA thermodynamic stability. LNA-modified adenine−thymine (AT) was found to have similar hydrogen bonding to unmodified cytosine−guanine (CG) base pairs, while for LNA CG, we found exceptionally large hydrogen bonding. In contrast, stacking interactions, which were thought to be behind the stability of LNA, were similar to unmodified DNA in most cases. We applied the new LNA parameters to the design of *BRAF*, *KRAS*, and *EGFR* oncogene variants by testing all possible LNA modifications. Selected sequences were then synthesized and had their hybridization temperatures measured, achieving a prediction accuracy within 1 °C. We performed a detailed base-pair opening analysis to discuss specific aspects of these probe hybridizations that may be relevant for probe design.



INTRODUCTION

Locked nucleic acid, or LNA, are DNA/RNA analogues with a modification that adds a methylene bridge between the 2′-oxygen and 4′-carbon of the ribose sugar, locking it in a C3′-endo/N conformation reducing its flexibility, resembling an RNA helix. They were synthesized in the late 1990s by several groups.[1−6] One of the most remarkable properties of LNA is the ability to increase the stability of chimera duplexes in both DNA and RNA. This stability has a direct impact on the denaturation temperatures of the sequences. Several studies have shown a considerable increase in the melting temperature[7−10] reaching up to 10° C per added modification. The presence of three LNAs at the 5′ and 3′ ends is enough to increase the half-life of the nucleotide,[11] and LNA-modified probes can selectively capture genomic DNA sequences.[12]

Its base pairing specificity and mismatch sensitivity make it attractive for use in many applications, such as polymerase chain reaction (PCR) detection and diagnosis,[13−15] high binding diagnostic probes,[16−18] stability improvement and hybridization efficiency in loop-mediated isothermal amplification (LAMP) detection,[19] improvement of targeting, specificity and stability in antisense oligonucleotides (ASOs),[20−22] aptamers,[23−25] and siRNA approaches.[26,27] LNA is also used in DNAzymes and LNAzymes to improve targeting and cleavage efficiency,[28−30] molecular beacons,[31] and as enhancers to RNA *in situ* hybridization[32,33] and as direct antagonist in miRNA silencing.[34−36]

One application of LNA that is of particular interest is to detect cancer mutations directly in DNA present in patients' blood samples (so-called circulating tumor DNA, or ctDNA) and in cellular mRNA.[37] Recently, it was shown that mutations in oncogene regions of ctDNA are attractive biomarkers for early diagnostics and monitoring of different cancer diseases.[38] Thus, the ability to effectively detect these oncogenes in ctDNA would open up new opportunities for point-of-care diagnostics and treatment of cancer. However, ctDNAs are typically present in the very complex blood medium at extremely low concentrations. This makes ctDNAs challenging for detection and require ultrasensitive and specific probes.[37] In turn, mRNA detection using new probes could provide valuable insights into cancer development.[39]

Received: April 27, 2021
Published: July 12, 2021

The design of ultrasensitive and specific probes with LNA modifications requires a good understanding of intramolecular interactions, specifically hydrogen bonding and stacking, that gives LNA its increased thermodynamic stability. Yet, despite the intensive use of LNA over the last two decades, the origin of its enhanced base pairing stability is still not sufficiently understood, and some conflicting explanations have been given. It was suggested by thermodynamic studies that a decrease in the entropy of duplex formation and improved stacking on the duplex may both play a part, and there may be a negative contribution in enthalpy due to the disruption of the hydrogen bonds.[40,41] Additionally, an association between a favorable enthalpy increase and a more pronounced stacking interaction was proposed; however, this is context-dependent and might be influenced by flanking base pairs.[10] Some studies suggested that the conformation change induces a reduction in entropy, being localized at the level of individual base pairs, consequently increasing the overall thermodynamic helix stability.[42,43] Several studies attempted to address the source of the stabilizing effect provided by LNA incorporation from a structural point of view. For instance, NMR and X-ray diffraction studies accounted for the stabilization in a stacking enhancement induced by the modification.[44−47] However, another NMR assay ruled out both stacking interaction or hydrogen bonding as a source of the stabilization effect, calling for more detailed studies with hybridized water, that is, the water molecule bridging two hydrogen bonds simultaneously.[48]

Nearest-neighbor (NN) models were extensively used to predict the thermodynamic properties for LNA-modified probes.[41,43,49,50] However, these types of models provide little insight into the intramolecular interactions and were unable to indicate the physical source of the increased LNA stability. On the other hand, mesoscopic models make use of the same melting temperature data as NN models, but have the potential to provide insights into the hydrogen bonds and stacking interactions.[51] We have shown that it is possible to obtain a good insight into these interactions from melting temperatures,[51] and we have validated the method over the past 10 years for a variety of different oligonucleotides, such as RNA,[52,53] DNA mismatches,[54] and DNA analogues such as TNA.[55] A recent study by our group,[56] using this type of mesoscopic modeling, found that the increased stability of LNA is largely hydrogen-bond-driven. Yet we also found that in certain stacking configurations, a decrease of stabilization may occur. These results were based on a small set of experimental melting temperatures, and we could not cover all possible stacking configurations. The sequences also had additional elements such as overhangs of variable length and, in some cases, fluorophores, which added to the uncertainty of our results. Therefore, it became necessary to recourse to a much larger set of melting temperatures, preferably without overhangs and fluorophores, which would enable us to identify the sources of LNA stabilization. Fortunately, the existing published data of single LNA modifications is unusually abundant,[43,49,50] providing over 300 sequences and their melting temperatures and allowing us to cover all single LNA:DNA/DNA stacking interactions.

For probe design, one aims at maximizing the difference in probe hybridization temperatures between the mutated and wild-type genes. LNA enrichment allows us to further increase this difference,[57] but poorly placed LNA modifications may be of little help as we have shown recently.[56] The full mesoscopic parameterization of LNA modifications opens the possibility to optimize the LNA inserts for oncogene probes. An improved stability and specificity of the probes is also highly desirable since it allows the detection of small quantities of the mutation, favoring an early diagnosis and detection. In terms of hybridization temperature prediction, the mesoscopic model has a similar predictive capability to the NN model;[58] however, it also allows stability analysis at a base-pair level, especially as it takes into account the nonlinear interactions between neighboring base pairs.

Here, we apply the new mesoscopic parameters to the high-throughput prediction of hybridization temperatures of the oncogene variants *BRAF*, *KRAS*, and *EGFR*, which were reported to be present in over 30% of all solid tumors, including breast, colon, and lung cancers.[59] Additionally, *KRAS* variants have been imposing a challenge in PCR and sequencing assays, resulting in smaller sensitivities.[60] It was suggested that the poor discrimination in *KRAS* probes is due to a CG internal region resulting in poor efficiency even in LNA-enriched probes.[56] First, we selected candidate probes for these three variants and then tested all possible combinations of one, two, and three LNA modifications, making use of the new mesoscopic parameters. From these, we selected six candidate probes, which were synthesized, had their melting temperatures measured, and we obtained an average prediction accuracy of 1°C.

MATERIALS AND METHODS

Model. The Peyrard−Bishop (PB) mesoscopic model uses two different potentials in its Hamiltonian: a Morse potential representing the hydrogen bonds that connect each base and a harmonic potential describing the stacking interaction between adjacent base pairs

$$U_{i,i+1} = \frac{k_{\alpha,\gamma}}{2}(y_i - y_{i-1})^2 + D_\alpha(e^{-y_i/\lambda_\alpha} - 1)^2 \quad (1)$$

which describes the interaction of the ith base pair of type α with its nearest neighbor $i + 1$ of type γ. The parameters D_α and λ_α describe the depth and width of the potential well for the ith base pair of type α, respectively.

During our calculations, the potential width λ is kept constant; therefore, the potential D completely represents the Morse potential. For the remainder of this work, we use $\lambda = 3.2 \times 10^{-2}$ and 0.97×10^{-2} nm for base pairs adenine−thymine (AT) and cytosine−guanine (CG),[51] respectively. Modified LNA base pairs use the same values for λ as their canonical analogues.

An elastic constant $k_{\alpha,\gamma}$ is used to describe the coupling between nearest-neighbor base pairs, and the coordinate y represents the relative displacements between the bases.

Equation 1 is used to calculate the partition function over all possible energetic configurations of N base pairs

$$Z_y = \int_{y_{min}}^{y_{max}} dy_1 \int_{y_{min}}^{y_{max}} dy_2 \ldots \int_{y_{min}}^{y_{max}} dy_N \times \prod_{n=1}^{N} e^{-\beta U(y_i, y_{i+1})} \quad (2)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the absolute temperature. The integral over all possible configurations of base-pair displacements, y_i, is performed and all possible Morse potentials, and stacking interactions are

https://doi.org/10.1021/acs.jcim.1c00470
J. Chem. Inf. Model. 2021, 61, 3615−3624

considered simultaneously during the evaluation. The reduced degrees of freedom of the model represented by eq 2, while allowing for an efficient numerical evaluation, has a consequence of resulting in transition temperatures, which are very low for short sequences. To overcome this problem, we introduced an adimensional index τ, which is calculated from eq 2, and it is directly correlated with the experimental melting temperatures.[61]

Moreover, the average relative displacement, $\langle y_m \rangle$, at the mth position in the sequence can be derived from

$$\langle y_m \rangle = \frac{1}{Z_y} \int_{y_{min}}^{y_{max}} dy_1 \int_{y_{min}}^{y_{max}} dy_2 \ldots \int_{y_{min}}^{y_{max}} dy_N \times y_m \prod_{n=1}^{N} e^{-\beta U(y_i, y_{i+1})} \quad (3)$$

Note that for the calculation of $\langle y_m \rangle$, we will use temperatures that are much lower than the actual melting temperatures.

Here, eqs 2 and 3 are evaluated via the transfer integral technique for heterogeneous sequences,[62] and we use $y_{min} = -0.1$ nm and $y_{max} = 40$ nm, and a calculation temperature $T = 370$ K, which has no relation to the melting temperatures. The relative displacement y can be negative, meaning that the bases are moving toward each other. However, for negative y, the Morse potential is strongly repulsive, which is why a short $y_{min} = -0.1$ nm is sufficient for numerical convergence. On the other hand, when the bases move away from each other (positive y), the Morse potential becomes flat and a large value of y is necessary to achieve numerical convergence.[62] The index τ is largely temperature-independent, and $T = 370$ K was found to be adequate for its calculation from the classical partition function; for details, see ref 61.

Temperature Prediction. Considering a set

$$P = \{p_1, p_2, \ldots, p_F\} \quad (4)$$

where p_i are Morse potentials, D_α, and stacking parameters, $k_{\alpha,\gamma}$. An index $\tau_i(P)$ is calculated for each sequence i using the partition function, eq 2, for the PB Hamiltonian. Thereon, the melting temperature, $T_i'(P)$, for each parameter set, P, is derived from the equation

$$T_i'(P) = a_0(N) + a_1(N)\tau \quad (5)$$

where N is the length of the sequence. The calculation of the regression coefficients a_0 and a_1 is carried out at a total strand concentration C_t. The melting temperatures, $T_i'(P)$, required for the parameterization were obtained from published sources as described in the next section.

Notation. LNA modifications are preceded by a plus sign and marked in bold. Since antiparallel strands of native DNA duplexes exhibit structural symmetry, only two Morse potentials (AT,CG) are required to represent the hydrogen bonds between base pairs. Such symmetry does not apply to modified LNA base pairs, which leads to four new Morse parameters, A+**T**, T+**A**, C+**G**, and G+**C**, if we consider only base pairs of the type DNA+LNA:DNA or DNA:DNA+LNA. A similar situation occurs for the stacking parameters, some nearest-neighbor doublets are identical, for instance, ATpCG and GCpTA, such that 10 nearest-neighbor parameters are sufficient to represent internal DNA:DNA doublets. For example, the doublet GCpT+**A** is symmetrical to the +**A**TpCG doublet. The set of sequences collected in the literature contain all of the 32 possible single LNA modifications, and the symmetry is reduced similarly for LNA NN and we always keep the NN that precede alphabetically. A full breakdown of the NN occurrences in the dataset is shown in Supporting Information Table S1.

As a practical example of the notation, consider the following sequence

$$5' - \text{CTAACG} + \mathbf{G}\text{ATGC} - 3'$$
$$3' - \text{GAT} + \mathbf{T}\text{GCCTACG} - 5'$$

in which case four parameters are required to represent the Morse potentials: AT, CG, A+**T**, and +**G**C, and 10 stacking potentials: CGpTA, TApAT, ATpA+**T**, A+**T**pCG, CGpGC, GCp+**G**C, +**G**CpAT, ATpTA, TApGC, and GCpCG.

Melting Temperature Set for Parametrization. The requirement for the sequences to be used for this work is that they contain only single LNA modification, that is, an LNA:DNA base pair. Further condition is that they should not have fluorophores attached and that any LNA:DNA nearest-neighbor configuration should be present in multiple sequences. Pure DNA:DNA control sequences, coming from the same LNA data sets were also used to adjust DNA parameters as we will describe in the next section. Existing data sets with tandem LNA modifications, such as Hughesman et al.,[42] or with attached fluorophores, such as Owczarzy et al.,[41] were not used for these reasons. In view of these requirements, we used the published melting temperatures for sequences measured at a high salt concentration (1021 mM Na^+) collected from refs 43, 49, 50. A total of 306 sequences meet our requirements, ranging between 8 and 23 base pairs in length.

The regression scheme of the PB model, eq 5, requires that all melting temperatures should be at equivalent strand concentrations. Therefore, we used the strand concentration of 2 μM, used in ref 50, as a reference. For the remaining sequences from refs 43, 49, we recalculated the melting temperatures using their respective reported total enthalpies and entropies. All sequences and their respective measured (or recalculated) and predicted melting temperatures are shown in Table S2.

Minimization Procedure. *Optimization.* To obtain an improved set of parameters for the PB model, we use an optimization method, detailed in refs 58, 63. Each parameter set P_j, eq 5, consisting of the parameters for the PB model will result in a melting temperature prediction $T_i'(P_j)$. This result is then compared with the experimental temperature T_i and the parameter set is varied until the squared difference is minimized

$$\chi_j^2 = \sum_{i=1}^{N} [T_i'(P_j) - T_i]^2 \quad (6)$$

Seed Parameters. This variation is initiated considering an initial set of i parameters, p_i. Those are varied in an interval

$$p_i \in [(1 - f)s_i, (1 + f)s_i] \quad (7)$$

that is, the value is sampled within a fraction $\pm f$ of a seed value for the parameter s_i. In this work, $f = 0.1$ results in the interval $[0.9s_i, 1.1s_i]$. The minimization procedure is numerically implemented using a downhill simplex method[63,64] using eq 6 as the merit function, and its minimum is searched in the multidimensional space defined by the parameter set P_j. The

https://doi.org/10.1021/acs.jcim.1c00470
J. Chem. Inf. Model. 2021, 61, 3615−3624

melting temperature deviation between the predicted and experimental temperatures is defined as

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^{N} |T_i' - T_i| \quad (8)$$

Initial Minimization (IM). In the first part of the minimization, IM, we kept constant all parameters without LNA modifications using previously published results for DNA at a high salt concentration, 1021 mM Na^+,[63] that is, considering that DNA canonical parameters will remain approximately the same in the presence of LNA. Parameters with the modification are considered as having the same initial value as its nonmodified analogues. We let the minimization proceed for the 4 Morse and 32 stacking parameters that contain LNA modifications. A large number of minimization rounds did converge poorly; for this reason, we kept only 100 out of 300 rounds with the lowest χ^2. Before this first optimization, the dataset had quality factors χ^2 = 4665.5 °C^2 and $\langle \Delta T \rangle$ = 2.9 °C. After optimization, they decreased to χ^2 = 1584.4°C^2, $\langle \Delta T \rangle$ = 1.7°C.

DNA/LNA Minimization (DL). In the previous steps, we have considered that the unmodified DNA bases keep the same parameters as from previous work.[51] In this new minimization, we used IM as seed parameters, but now we also let the DNA base parameters vary and we will call this the DNA/LNA minimization (DL). That was followed for 300 minimization rounds, and the 100 with the lowest χ^2 were averaged as the final result.

Influence of the Experimental Error (EE-DL). To simulate the influence of experimental error (EE) associated with the temperature measurement on our new parameters, we change the temperature by small random amounts such that the standard deviation between the original set and the optimized set approaches the declared experimental uncertainty. Here, we used as the initial set of parameters the results from the minimization round DL, and therefore, this round is called EE-DL. As the whole dataset comes from three different sources, we used the highest declared uncertainty, 0.8° C, from Fakhfakh et al.[43] Again, the minimization procedure was followed 300 times, and we averaged the 100 with the lowest χ^2 as the final result. We use the standard deviation of the lowest 100 rounds to represent the uncertainty estimate of our new parameters. Each minimization step took approximately 7 days in a server with 400 cores, that is, the whole minimization took of the order of 21 days. A summary of the quality parameters for each minimization round is shown in Table 1.

Table 1. Summary of the Quality Parameters for the Minimization Rounds

round	$\langle \Delta T \rangle$ (°C)	χ^2 (°C^2)
IM	1.75	1584.4
DL	1.50	1053.7
EE-DL	1.47	1013.6

Capture/Linker Probe Design. Candidate DNA/DNA probes for *BRAF*, *EGFR*, and *KRAS* oncogenes were designed using public available DNA human genome sequencing data (NCBI) with GC content in the range of 38−55% and melting temperature above 45 °C according to the initial NN prediction for unmodified DNA in medium salt buffer. The latter criterion was applied to secure adequate binding properties in most hybridization assays.[65,66] Probe design was followed similarly as done before.[67,68] Briefly, the assembly of the human oncogenes and their respective NCBI code are described in Table S4.

Mutated oncogenes were assessed using dbSNP base (rs113488022) (*BRAF V600E*); rs121913529 (*KRAS G12D*); rs112445441 (*KRAS G13D*); rs121434568 (*EGFR L858R*). Oligonucleotide probes were designed to be complementary to the position of the corresponding gene bearing the mutation (capture probe), and downstream the gene with a gap of >20 nt from the position of capture probe, for linker probe. The length of each probe was determined using probe uniqueness software described in ref 69.

Once the candidate capture (cap) and linker (lin) probes were established, we used the optimized parameters DNA +LNA:DNA parameters to calculated all possible configurations with one, two, or three LNA modifications for each of the selected probes. The melting temperatures of both, modified and nonmodified probes, were evaluated, and a salt correction from Owczarzy et al.[70] to a salt concentration of 137 mM Na^+ was applied to the results. We calculated between 834 and 2345 different LNA-modified sequences, depending on length, for each candidate probe. Supporting Information Tables S5−S11 show the 30 highest and 30 lowest melting temperature predictions for *BRAF*, *EGFR*, and *KRAS* linker and capture variants, respectively.

We selected the probes with the lowest possible number of LNA incorporation per sequence to achieve the highest overall T_m and the highest discrimination T_m full match vs T_m mismatch for the probes binding to the position of mutation in the corresponding oncogene. Selected candidate probes were synthesized and their melting temperatures were measured as described in the next section.

Oligonucleotide Synthesis and UV Melting Procedure. The LNA/DNA oligonucleotides were synthesized on an ASM-800 ET synthesizer from BIOSSET Ltd. Commercial phosphoramidites from Sigma-Aldrich (dA(Bz), dC(Bz), dG(ib), dT) and Qiagen (LNA-A(Bz), LNA-T, LNA-mC(Bz), LNA-G(dmf)) were dissolved in dry acetonitrile to a concentration of 0.07 M. Reagents for solid-phase synthesis were purchased from Sigma-Aldrich: TCA Deblock, DCI activator 0.25 M, Oxidizer 0.02 M, Cap A, and Cap B. The phosphoamidites were loaded onto the synthesizer following the manufacture standard protocol. All of the oligonucleotides were synthesized on controlled pore glass (CPG 1000 Å) universal support purchased from Sigma-Aldrich in a 1 μmol scale using a double-coupling protocol. DMT-off mode was applied. The coupling rate during the synthesis was estimated based on absorbance measurements of the DMT cation using TM800 software.

After synthesis, the columns were dried by nitrogen purge and then transferred to 5 mL tubes. For cleavage from the universal support, aqueous ammonia (28−30%; 1 mL) was added and the samples were placed at 55 °C for 12 h. Next, the samples were cooled at −20 °C for about 10 min and filtrated into 2 mL Eppendorf tubes. The oligonucleotides were evaporated to dryness and redissolved in 1000 μL of Milli-Q water.

The identity of oligonucleotides was confirmed by mass spectrometry (MS) using an Autoflex speed MALDI-TOF mass spectrometer (Bruker Daltonics, Germany). Matrix-assisted laser desorption ionization−mass spectrometry (MALDI-MS) of purified oligonucleotides and their respective

https://doi.org/10.1021/acs.jcim.1c00470
J. Chem. Inf. Model. 2021, 61, 3615−3624

purities are shown in Supporting Information Table S12. Representative MALDI-MS spectra for each probe are shown in Supporting Information Figures S2−S14. Samples were co-spotted with 3-hydroxypicolinic acid as matrix on an MTP AnchorChip target plate for the analysis. The obtained mass spectra were recorded by the flexControl 3.4 (Bruker Daltonics, Germany) software. The oligonucleotides were purified on an Ultimate 3000 UHPLC (Dionex, United States) using a DNA-Pac RP (Thermo Fisher Scientific, United States) column (4 μm, 3.0 × 100 mm) with a gradient of 5−25% buffer B in A over 30 min at 60 °C (buffer A: 0.05 M TEAA, buffer B: 25% A in acetonitrile). The peaks were monitored at 260 nm. Representative ultrahigh-performance liquid chromatography (UHPLC) traces for each probe are shown in Supporting Information Figures S1−S13.

The melting temperature studies were performed on a DU800 UV/VIS spectrophotometer equipped with a Beckman Coulter Performance Temperature Controller. Complementary strands (0.5 μM of each strand), in 1 × phosphate-buffered saline (PBS) were mixed, denatured for 10 min at 90 °C, and subsequently cooled to 15 °C, the temperature at which the experiment was started. Reported melting temperatures present the maximum of the first derivative of the curve and are an average of the two measurements within a deviation of 1 °C. T_m curves for each probe are shown in Supporting Information Figures S15−S20.

RESULTS AND DISCUSSION

All results presented here refer to the final round, EE-DL, unless noted otherwise.

Morse Potentials. The Morse potential has a long history for the use of modeling hydrogen bonding in molecular dynamics,[71] coarse-grained models,[72] and for mesoscopic models.[73] Other model potentials, such as the Lennard-Jones potential, could, in principle, be used, yet no appreciable difference was noted compared to the Morse potential in the PB model.[74] For the remainder of this discussion, we will refer to the potential depth D simply as the Morse potential.

In Figure 1, we show the final average parameter D of the Morse potentials and, for comparison, the previous results for canonical DNA.[51] Modified base pairs consistently show higher Morse potentials than their analogues. This suggests that, in comparison to the unmodified DNA, new hydrogen bonds between the base pairs may have been formed due to the LNA modification, or that the existing ones were strengthened in some way. Morse potentials of unmodified DNA base pairs, AT and CG, remained nearly the same as that of previous calculations,[51,75] shown as dashed lines in Figure 1, which is consistent with the similarity of buffer conditions between all data sets. As we are dealing with single LNA modification in a DNA duplex, little influence of LNA on the remaining base pairs was expected. The small change of the canonical base-pair parameters also confirms this and is consistent with atomistic molecular dynamics simulations, which found that the change induced by single LNA modification in a DNA oligonucleotide is very localized and essentially limited to the immediate neighboring base pairs.[76] It was suggested by a thermodynamic study a similar stability ordering as the one we have found here for Morse potentials, +**A** ≤ +**T** ≤ +**G** ≤ +**C**.[49] Although we have found a stronger Morse potential for the modified guanine, +**G**. However, we cannot establish a direct comparison since the stability on the PB model is also related to the stacking parameters. One of the consequences of the stronger Morse potentials is a reduction in entropy, which can be calculated from the partition function eq 2. This is consistent with the entropy reduction seen for nearest-neighbor analysis.[42,43] Some examples of calculated configurational entropies with LNA-modified sequences are shown in Supporting Information Figure S25.

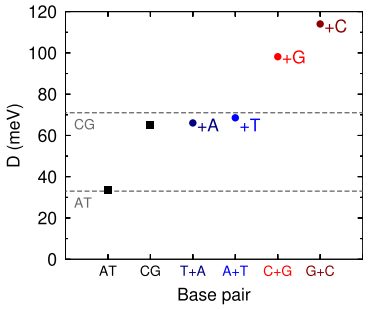


Figure 1. Average D parameter of the Morse potential for LNA-modified (bullets) and unmodified DNA (boxes) base pairs. LNA modifications are shown next to symbols. Error bars were omitted as they are smaller than the symbol size. For comparison, we show the published parameters[51] for unmodified DNA base pairs, AT and CG, as gray dashed lines. For clarity, LNA monomers are separated by color, and will use this color scheme in the next figure.

There is an ongoing debate regarding the intramolecular origin of the increased stability of LNA modifications. On the one hand, there have been proposals that LNA stability is due to increased stacking interactions.[44,45,49,77] On the other hand, Jensen et al.[48] concluded against either increased stacking or hydrogen bonding and argued that hybridized water may be the cause of LNA stability. An NMR study by Egli et al.[46] established that the extra oxygen atoms in LNA monomers are engaged in water-bridged hydrogen bonds. This may account for an increased thermodynamic stability as an extensive hydration of hydrogen-bond acceptors and donors is associated with a stability increase.[46] In another molecular dynamics study,[78] modified LNA duplexes were found forming even four water-bridged hydrogen bonds. In this study,[78] LNA helices were found to be less hydrated than DNA or RNA, which was hypothesized to be due to its shorter intrastrand phosphate distances and lower backbone flexibility, allowing the formation of water-mediated hydrogen bonds providing extra stability to LNA base pairs.

Our results undoubtedly support the idea of increased hydrogen bonding, as shown in Figure 1, which is also in line with our earlier findings[56] for a much smaller dataset. It is particularly interesting to note that both AT modifications, A +T and T+A, have become almost identical in strength to an unmodified CG base pair (Figure 1). There are suggestions that AT may have a third weak bond C−H···O,[79−81] although it may not be a hydrogen bond but a van der Waals interaction.[82] While our calculations cannot pinpoint the exact nature of this interaction, one might speculate that the C−H···O contact may have been strengthened due to the LNA modification, for both +A and +T.

Stacking Potentials. The calculated stacking parameters are shown in Figure 2. Higher fluctuations from the canonical values depend on the modified base-pair location and direction on the strand. Similarly to previous results for DNA,[51] NNs containing only AT base pairs (Figure 2a) present larger stacking variations. Less fluctuation is seen for mixed NNs

https://doi.org/10.1021/acs.jcim.1c00470
J. Chem. Inf. Model. 2021, 61, 3615−3624

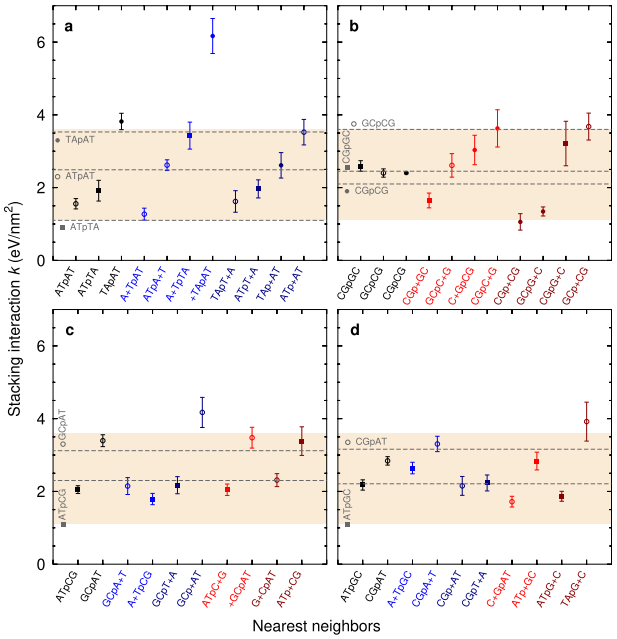


Figure 2. Average stacking potentials for the LNA-modified and canonical nearest neighbors. Results are separated into panels grouped by NN patterns: (a) for ATpTA (boxes), ATpAT (circles), and TApAT (bullets); (b) for CGpCG (bullets), CGpCG (boxes), and GCpCG (circles); (c) for ATpCG (boxes) and GCpAT (circles); (d) for ATpGC (boxes) and CGpAT (circles). Error bars were estimated in the EE-DL minimization round. For comparison, we show the published stacking potentials for canonical DNA[51] as gray dashed lines. The shaded region shows the range of all possible stacking parameters for canonical DNA,[51] with the minimum being the stacking parameter for ATpTA and maximum for GCpCG (the same on all panels). LNA monomers are separated by color and follow the same color scheme as Figure 1.

(Figure 2c,d), which tend to fluctuate around the stacking parameters of their unmodified analogues. Molecular dynamics studies have established a local disturbance due to LNA base pairs affecting the first nearest neighbors,[76] which is similarly related to the little change in canonical interactions found in our results. Moreover, LNA enrichment does not have a strong effect on canonical stacking interactions, except for CG shown in Figure 2b, as they keep nearly the same within the standard deviation during the minimization. Only one LNA-modified NN, +TApAT, displayed significantly larger stacking, completely out of the region of canonical stacking (shaded area in Figure 2a).

Figure 2 shows that very few LNA modifications have stacking interactions that are larger than typical DNA stacking, that is, that lie above the shaded area. Only one type of modification, +TApAT, has a very strong stacking interaction, which may indicate a specially modified stacking arrangement[54] or perhaps to a departure from the *anti*−*anti* configuration.[83] In fact, a considerable number of LNA:DNA configurations decrease their stacking interactions in comparison to their unmodified analogues. Therefore, we conclude that, in general, stacking does not account for the increased stability of LNA, and in many cases, it appears to be even a destabilizing factor. Incidentally, the single sequence analyzed by Nielsen et al.,[44] who concluded stacking as a major contributing factor to LNA stability, contains +TApAT, which is the only strongly stacked modification that we observed (Figure 2). Therefore, even though we conclude against stacking as a major factor of LNA:DNA stability, our results are not at odds with the conclusion of Nielsen et al.[44] for +TApAT.

Probe Design, Synthesis, and Evaluation. The calculation of the LNA:DNA parameters is time-consuming and requires considerable computational resources. However, once they become available, it is straightforward to calculate melting temperature predictions for a large set of sequences. For prediction, the computational efficiency of the PB regression scheme is comparable to the NN model.[84,85] First, candidate probes were selected from genomic analysis for *BRAF*, *EGFR*, and *KRAS* oncogenes. Then, with the new LNA:DNA parameters at hand, we calculated the melting temperatures of LNA modifications at all positions of *BRAF*, *EGFR*, and *KRAS* capture and linker probes. The resulting temperatures, which are valid for high sodium concentrations, were recalculated by applying a salt correction factor from Owczarzy et al.,[70] which was shown to be adequate for LNA.[50] An excerpt of the predicted melting temperatures for up to three candidate LNA modifications per probe is shown in Supporting Information Tables S5−S11.

One probe was selected from the list of the candidate probes for each capture and linker pair, using additional criteria of the lowest possible number of LNA incorporations per sequence to achieve the highest overall T_m and the highest discrimination T_m full match vs T_m mismatch for the probes binding to the position of mutation in the corresponding oncogene. The selected probes were synthesized, and their

Table 2. Measured T_{exp}, Control T_{ctrl}, and Predicted Temperatures T_{pred} for the Selected Sequences from the Probe Prediction[a]

ID	synthesized probe with LNA modifications	T_{ctrl}	T_{exp}	T_{pred}
BRAFcap-M1225	ATCGAGAT+TTCT+CTGTAG+CTA	59.57	64.50	64.92
BRAFlin-M416	CAA+CTGTT+CAAA+CTGAT	50.75	62.20	60.75
KRAS12cap-M1344	GCACTCTTGCCTACCCA+ATC	64.07	64.70	65.42
KRAS13cap-M1304	GCACTCTTGCCTA+CGCA+TTC	64.15	68.90	68.83
KRASlin-M616	TGAAGT+CA+CA+TTATATA	48.15	55.00	58.17
EGFRcap-M943	GAG+AAAAAGTTTCTCA+TGTA+CAGT	57.40	61.2	61.76
EGFRlin-M478	TTG+TTGGAT+CATA+TTCGT	54.45	61.50	61.52
			$\langle T_{exp} - T_{pred} \rangle$ = 0.91 °C	

[a]Control T_{ctrl} refers to the unmodified DNA/DNA probes. The probe identification (ID) refers to the candidate probe list in Supporting Information Tables S5−S11. The average deviation $\langle T_{exp} - T_{pred} \rangle$ was calculated between the available predicted and measured temperatures. Probe sequences are shown in the 5′ → 3′ direction.

https://doi.org/10.1021/acs.jcim.1c00470
J. Chem. Inf. Model. 2021, 61, 3615−3624

melting temperatures were measured as described in the Materials and Methods section. Table 2 shows the predicted and measured melting temperatures for each probe, and their average deviation was found to be of the order of 1 °C. While this represents an overall satisfactory prediction, it is worthwhile to have a closer look at the probes that fell short of their prediction target. For this purpose, we calculated the opening profiles of these sequences using eq 3, which provides an intuitive way to visualize localized instabilities in the sequence. For instance, KRASlin-M616 was predicted to melt at 3 °C higher than measured. Figure 3 shows how this probe

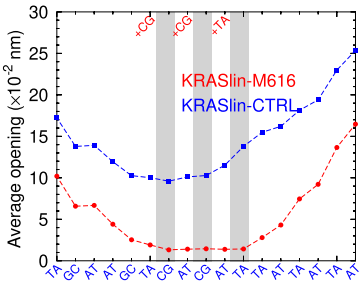


Figure 3. Average opening profile for the LNA-modified KRASlin-M616 probe (red bullets) and its corresponding unmodified control probe (blue squares). The location of the LNAs is indicated by the gray shaded area. Calculation was carried out at 220 K, eq 3, which has no relation to the melting temperature.

stabilizes (lower openings) in comparison to its control sequence (larger openings). However, the added LNA modifications provided only moderate additional stabilization for its 3′ AT-tail. One possibility for the lower measured melting temperature is that the high salt parameters overestimate the stability of AT base pairs at 3′ sequence ends. In previous studies,[75] we observed that AT base pairs at the sequence end tended to have much lower Morse potentials for low salt concentrations; it is therefore possible that the lower-than-predicted temperature has a similar effect. For BRAFlin-M416, the measured temperature was 1 °C higher than the predicted value. While this is within the measurement uncertainty, Figure 4 shows that the distribution of +C along

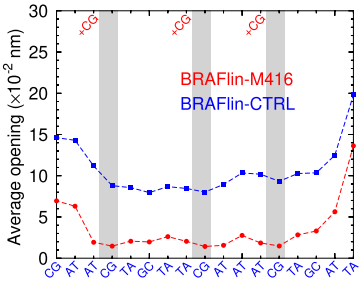


Figure 4. Average opening profile for the LNA-modified BRAFlin-M416 probe (red bullets) and its corresponding unmodified control probe (blue squares). Figure elements are the same as for Figure 3.

the sequence has a very large stabilizing effect. KRAS12cap-M1344 was selected among the lowest-stabilizing LNA modifications, and indeed its measured melting temperature is only marginally larger than that of the control sequence. Figure 5 shows a small stabilization end toward the 3′ end, which is already capped by a CG base pair, and comparing the +AT openings, one notices its similarity to an unmodified CG.

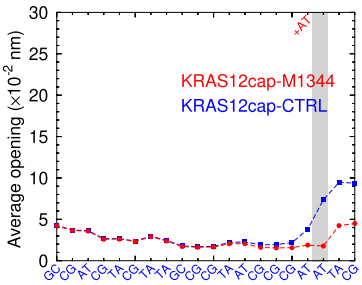


Figure 5. Average opening profile for the LNA-modified KRAS12cap-M1344 (red bullets) and its corresponding unmodified control probe (blue squares). Figure elements are the same as for Figure 3.

CONCLUSIONS

We have successfully applied a mesoscopic model to parameterize a large set of published melting temperatures for LNA-modified sequences measured at a high salt concentration. We have found a substantial increase in Morse potentials, indicating stronger hydrogen bonding for LNA:DNA base pairs. This confirms previous results at low salt concentrations[56] and establishes hydrogen bonding as the main source of LNA:DNA stabilization. The large number of sequences allowed us to fully parameterize all 32 possible LNA:DNA/DNA nearest-neighbor stacking interaction. Only one case of increased stacking for +TApAT was found, while the remaining ones are similar to unmodified DNA with a few cases of reduced stacking interactions. The new parameters were used to predict probes hybridization, targeting the oncogene variants *BRAF*, *KRAS*, and *EGFR* in a medium salt buffer, as a proof of concept. For each probe, we tested all combinations of up to three LNA modifications and selected six candidate probes for synthesis and melting temperature measurements. We have found a good agreement in the predictions, after applying established salt correction factors.[70] The few discrepancies between prediction and measured temperatures were analyzed using opening profiles at a base-pair level.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00470.

Number of modified and nonmodified NN (Table S1); sequences used in the minimization and their respective measured and predicted temperatures (Table S2); measured and predicted melting temperatures for the validation set of sequences containing LNA tandem modifications (Table S3); summary of the oncogenes used for the probe design and their respective human assemblies (Table S4); melting temperature predictions for *BRAF*, *EGFR*, and *KRAS* variants (Tables S5−S11); MALDI-MS and purities of the measured probes (Tables S12); HPLC and MALDI-MS of the measured probes (Figures S1−S14); representative T_m curve of the measured probes (Figures S15−S20); average opening profiles for the LNA-modified probes measured in this study (Figures S21−S24); and some examples of

https://doi.org/10.1021/acs.jcim.1c00470
J. Chem. Inf. Model. 2021, 61, 3615−3624

calculated entropy for LNA-modified sequences (Figure S25) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Kira Astakhova − *Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark;* orcid.org/0000-0003-4878-0301; Email: kiraas@kemi.dtu.dk

Gerald Weber − *Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil;* orcid.org/0000-0002-2935-1571; Email: gweber@ufmg.br

Authors

Izabela Ferreira − *Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil; Programa Interunidades de Pós-Graduação em Bioinformática, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil*

Sofie Slott − *Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c00470

Notes

The authors declare no competing financial interest.

All sequences, their respective melting temperatures, and all calculated parameters are included in our updated software package TfReg (https://bioinf.fisica.ufmg.br/software) and available as buildable packages in the OpenSuse repository (https://build.opensuse.org/package/show/home:drgweber/TfReg) and are free of charge. All results presented here can be verified with this software. This package also calculates opening profiles, such as those of Figures 3−5, which take only a few seconds to run on standard computers with the provided parameters.

ACKNOWLEDGMENTS

I.F. and G.W. were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Brazil, Finance Code 001). K.A. and S.S. were supported by DTU Ellite Ph.D. Scholarship and DTU Proof of concept programme.

REFERENCES

(1) Obika, S.; Nanbu, D.; Hari, Y.; Morio, K.-i.; In, Y.; Ishida, T.; Imanishi, T. Synthesis of 2′-O, 4′-C-methyleneuridine and-cytidine. Novel bicyclic nucleosides having a fixed C3-endo sugar puckering. *Tetrahedron Lett.* **1997**, *38*, 8735−8738.

(2) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* **1998**, *54*, 3607−3630.

(3) Kumar, R.; Singh, S. K.; Koshkin, A. A.; Rajwanshi, V. K.; Meldgaard, M.; Wengel, J. The first analogues of LNA (locked nucleic acids): phosphorothioate-LNA and 2′-thio-LNA. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2219−2222.

(4) Obika, S.; Nanbu, D.; Hari, Y.; Andoh, J.-i.; Morio, K.-i.; Doi, T.; Imanishi, T. Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation, 2′-O, 4′-C-methyleneribonucleosides. *Tetrahedron Lett.* **1998**, *39*, 5401−5404.

(5) Singh, S. K.; Koshkin, A. A.; Wengel, J.; Nielsen, P. LNA (Locked nucleic acids) synthesis and high-affinity nucleic acid recognition. *Chem. Commun.* **1998**, *4*, 455−456.

(6) Wang, G.; Gunic, E.; Girardet, J.-L.; Stoisavljevic, V. Conformationally locked nucleosides. Synthesis and hybridization properties of oligodeoxynucleotides containing 2′,4′-C-bridged 2′-deoxynucleosides. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1147−1150.

(7) Obika, S.; Hari, Y.; Morio, K.-i.; Imanishi, T. Triplex formation by an oligonucleotide containing conformationally locked C-nucleoside, 5-(2-O, 4-C-methylene-β-Dribofuranosyl) oxazole. *Tetrahedron Lett.* **2000**, *41*, 221−224.

(8) Wahlestedt, C.; Salmi, P.; Good, L.; Kela, J.; Johnsson, T.; Hökfelt, T.; Broberger, C.; Porreca, F.; Lai, J.; Ren, K. Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5633−5638.

(9) Pasternak, A.; Wengel, J. Thermodynamics of RNA duplexes modified with unlocked nucleic acid nucleotides. *Nucleic Acids Res.* **2010**, *38*, 6697−6706.

(10) Kaur, H.; Arora, A.; Wengel, J.; Maiti, S. Thermodynamic, counterion, and hydration effects for the incorporation of locked nucleic acid nucleotides into DNA duplexes. *Biochemistry* **2006**, *45*, 7347−7355.

(11) Kurreck, J.; Wyszko, E.; Gillen, C.; Erdmann, V. A. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.* **2002**, *30*, 1911−1918.

(12) Jacobsen, N.; Bentzen, J.; Meldgaard, M.; Jakobsen, M. H.; Fenger, M.; Kauppinen, S.; Skouv, J. LNA-enhanced detection of single nucleotide polymorphisms in the apolipoprotein E. *Nucleic Acids Res.* **2002**, *30*, No. e100.

(13) Josefsen, M. H.; Löfström, C.; Sommer, H. M.; Hoorfar, J. Diagnostic PCR: comparative sensitivity of four probe chemistries. *Mol. Cell. Probes* **2009**, *23*, 201−203.

(14) Morandi, L.; De Biase, D.; Visani, M.; Cesari, V.; De Maglio, G.; Pizzolitto, S.; Pession, A.; Tallini, G. Allele specific locked nucleic acid quantitative PCR (ASLNAqPCR): an accurate and cost-effective assay to diagnose and quantify KRAS and BRAF mutation. *PLoS One* **2012**, *7*, No. e36084.

(15) Fontanilles, M.; Marguet, F.; Ruminy, P.; Basset, C.; Noel, A.; Beaussire, L.; Viennot, M.; Viailly, P.-J.; Cassinari, K.; Chambon, P. Simultaneous detection of EGFR amplification and EGFRvIII variant using digital PCR-based method in glioblastoma. *Acta Neuropathol. Commun.* **2020**, *8*, No. 52.

(16) Astakhova, I. V.; Ustinov, A. V.; Korshun, V. A.; Wengel, J. LNA for optimization of fluorescent oligonucleotide probes: improved spectral properties and target binding. *Bioconjugate Chem.* **2011**, *22*, 533−539.

(17) Østergaard, M. E.; Cheguru, P.; Papasani, M. R.; Hill, R. A.; Hrdlicka, P. J. Glowing locked nucleic acids: brightly fluorescent probes for detection of nucleic acids in cells. *J. Am. Chem. Soc.* **2010**, *132*, 14221−14228.

(18) Hussung, S.; Follo, M.; Klar, R. F.; Michalczyk, S.; Fritsch, K.; Nollmann, F.; Hipp, J.; Duyster, J.; Scherer, F.; von Bubnoff, N. Development and clinical validation of discriminatory multi-target digital droplet PCR assays for the detection of hot spot KRAS and NRAS mutations in cell-free DNA. *J. Mol. Diagn.* **2020**, *22*, 943−956.

(19) Bakthavathsalam, P.; Longatte, G.; Jensen, S. O.; Manefield, M.; Gooding, J. J. Locked nucleic acid molecular beacon for multiplex detection of loop mediated isothermal amplification. *Sens. Actuators, B* **2018**, *268*, 255−263.

(20) Bagheri, P.; Sharifi, M.; Ghadiri, A. Downregulation of MIR100HG Induces Apoptosis in Human Megakaryoblastic Leukemia Cells. *Indian J. Hematol. Blood Transfus.* **2021**, *37*, 232−239.

(21) Zhang, K.; Zheludev, I. N.; Hagey, R. J.; Wu, M. T.-P.; Haslecker, R.; Hou, Y. J.; Kretsch, R.; Pintilie, G. D.; Rangan, R.; Kladwang, W. Cryo-electron Microscopy and Exploratory Antisense

Targeting of the 28-kDa Frameshift Stimulation Element from the SARS-CoV-2 RNA Genome. *bioRxiv* **2020**, *395*, No. 1245.
(22) Lim, K. R. Q.; Maruyama, R.; Echigoya, Y.; Nguyen, Q.; Zhang, A.; Khawaja, H.; Sen Chandra, S.; Jones, T.; Jones, P.; Chen, Y.-W.; Yokota, T. Inhibition of DUX4 expression with antisense LNA gapmers as a therapy for facioscapulohumeral muscular dystrophy. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 16509−16515.
(23) Schmidt, K. S.; Borkowski, S.; Kurreck, J.; Stephens, A. W.; Bald, R.; Hecht, M.; Friebe, M.; Dinkelborg, L.; Erdmann, V. A. Application of locked nucleic acids to improve aptamer in vivo stability and targeting function. *Nucleic Acids Res.* **2004**, *32*, 5757−5765.
(24) Darfeuille, F.; Reigadas, S.; Hansen, J. B.; Orum, H.; Di Primo, C.; Toulmé, J.-J. Aptamers targeted to an RNA hairpin show improved specificity compared to that of complementary oligonucleotides. *Biochemistry* **2006**, *45*, 12076−12082.
(25) Wojtyniak, M.; Schmidtgall, B.; Kirsch, P.; Ducho, C. Towards Zwitterionic Oligonucleotides with Improved Properties: the NAA/LNA-Gapmer Approach. *ChemBioChem* **2020**, *21*, 3234−3243.
(26) Braasch, D. A.; Jensen, S.; Liu, Y.; Kaur, K.; Arar, K.; White, M. A.; Corey, D. R. RNA interference in mammalian cells by chemically-modified RNA. *Biochemistry* **2003**, *42*, 7967−7975.
(27) Elmén, J.; Thonberg, H.; Ljungberg, K.; Frieden, M.; Westergaard, M.; Xu, Y.; Wahren, B.; Liang, Z.; Ørum, H.; Koch, T. Locked nucleic acid (LNA) mediated improvements in siRNA stability and functionality. *Nucleic Acids Res.* **2005**, *33*, 439−447.
(28) Vester, B.; Lundberg, L. B.; Sørensen, M. D.; Babu, B. R.; Douthwaite, S.; Wengel, J. LNAzymes: Incorporation of LNA-Type Monomers into DNAzymes Markedly Increases RNA Cleavage. *J. Am. Chem. Soc.* **2002**, *124*, 13682−13683.
(29) Jakobsen, M. R.; Haasnoot, J.; Wengel, J.; Berkhout, B.; Kjems, J. Efficient inhibition of HIV-1 expression by LNA modified antisense oligonucleotides and DNAzymes targeted to functionally selected binding sites. *Retrovirology* **2007**, *4*, No. 29.
(30) Jadhav, V. M.; Scaria, V.; Maiti, S. Antagomirzymes: oligonucleotide enzymes that specifically silence microRNA function. *Angew. Chem., Int. Ed.* **2009**, *48*, 2557−2560.
(31) Wang, L.; Yang, C. J.; Medley, C. D.; Benner, S. A.; Tan, W. Locked nucleic acid molecular beacons. *J. Am. Chem. Soc.* **2005**, *127*, 15664−15665.
(32) Wienholds, E.; Kloosterman, W. P.; Miska, E.; Alvarez-Saavedra, E.; Berezikov, E.; de Bruijn, E.; Horvitz, H. R.; Kauppinen, S.; Plasterk, R. H. MicroRNA expression in zebrafish embryonic development. *Science* **2005**, *309*, 310−311.
(33) Thomsen, R.; Nielsen, P. S.; Jensen, T. H. Dramatically improved RNA in situ hybridization signals using LNA-modified probes. *RNA* **2005**, *11*, 1745−1748.
(34) Elmén, J.; Lindow, M.; Schütz, S.; Lawrence, M.; Petri, A.; Obad, S.; Lindholm, M.; Hedtjärn, M.; Hansen, H. F.; Berger, U. LNA-mediated microRNA silencing in non-human primates. *Nature* **2008**, *452*, 896−899.
(35) Huang, S.; Ichikawa, Y.; Yoshitake, K.; Igarashi, Y.; Kinoshita, S.; Asaduzzaman, M.; Omori, F.; Maeyama, K.; Nagai, K.; Watabe, S. Potential silencing of gene expression by PIWI-interacting RNAs (piRNAs) in somatic tissues in mollusk. *bioRxiv* **2020**, No. 199877.
(36) Kalinina, M.; Skvortsov, D.; Kalmykova, S.; Ivanov, T.; Dontsova, O.; Pervouchine, D. D. Multiple competing RNA structures dynamically control alternative splicing in the human ATE1 gene. *Nucleic Acids Res.* **2021**, *49*, 479−490.
(37) Ignatiadis, M.; Sledge, G. W.; Jeffrey, S. S. Liquid biopsy enters the clinic−implementation issues and future challenges. *Nat. Rev. Clin. Oncol.* **2021**, 297−312.
(38) Van Der Pol, Y.; Mouliere, F. Toward the early detection of cancer by decoding the epigenetic and environmental fingerprints of cell-free DNA. *Cancer Cell* **2019**, *36*, 350−368.
(39) Liu, X.-P.; Yin, X.-H.; Meng, X.-Y.; Yan, X.-H.; Wang, F.; He, L. Development and validation of a 9-gene prognostic signature in patients with multiple myeloma. *Front. Oncol.* **2019**, *8*, No. 615.
(40) Bruylants, G.; Boccongelli, M.; Snoussi, K.; Bartik, K. Comparison of the thermodynamics and base-pair dynamics of a full LNA:DNA duplex and of the isosequential DNA:DNA duplex. *Biochemistry* **2009**, *48*, 8473−8482.
(41) Owczarzy, R.; You, Y.; Groth, C.; Tataurov, A. Stability and Mismatch Discrimination of Locked Nucleic Acid-DNA Duplexes. *Biochemistry* **2011**, *50*, 9352−9367.
(42) Hughesman, C. B.; Turner, R. F.; Haynes, C. A. Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified LNA. *Biochemistry* **2011**, *50*, 5354−5368.
(43) Fakhfakh, K.; Marais, O.; Cheng, X. B. J.; Castañeda, J. R.; Hughesman, C. B.; Haynes, C. Molecular thermodynamics of LNA:LNA base pairs and the hyperstabilizing effect of 5′-proximal LNA:DNA base pairs. *AIChE J.* **2015**, *61*, 2711−2731.
(44) Nielsen, C. B.; Singh, S. K.; Wengel, J.; Jacobsen, J. P. The solution structure of a locked nucleic acid (LNA) hybridized to DNA. *J. Biomol. Struct. Dyn.* **1999**, *17*, 175−191.
(45) Nielsen, K. E.; Singh, S. K.; Wengel, J.; Jacobsen, J. P. Solution structure of an LNA hybridized to DNA: NMR study of the d(CTLGCTLTLCTLGC):d(GCAGAAGCAG) duplex containing four locked nucleotides. *Bioconjugate Chem.* **2000**, *11*, 228−238.
(46) Egli, M.; Minasov, G.; Teplova, M.; Kumar, R.; Wengel, J. X-ray crystal structure of a locked nucleic acid (LNA) duplex composed of a palindromic 10-mer DNA strand containing one LNA thymine monomer. *Chem. Commun.* **2001**, 651−652.
(47) Eichert, A.; Behling, K.; Betzel, C.; Erdmann, V. A.; Fürste, J. P.; Förster, C. The crystal structure of an "All Locked" nucleic acid duplex. *Nucleic Acids Res.* **2010**, *38*, 6729−6736.
(48) Jensen, G. A.; Singh, S. K.; Kumar, R.; Wengel, J.; Jacobsen, J. P. A comparison of the solution structures of an LNA:DNA duplex and the unmodified DNA:DNA duplex. *J. Chem. Soc., Perkin Trans. 2* **2001**, *2*, 1224−1232.
(49) McTigue, P. M.; Peterson, R. J.; Kahn, J. D. Sequence-dependent thermodynamic parameters for Locked Nucleic Acid (LNA)-DNA duplex formation. *Biochemistry* **2004**, *43*, 5388−5405.
(50) You, Y.; Moreira, B. G.; Behlke, M. A.; Owczarzy, R. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res.* **2006**, *34*, No. e60.
(51) Weber, G.; Essex, J. W.; Neylon, C. Probing the Microscopic Flexibility of DNA from Melting Temperatures. *Nat. Phys.* **2009**, *5*, 769−773.
(52) Amarante, T. D.; Weber, G. Evaluating hydrogen bonds and base stackings of single, tandem and terminal GU in RNA mismatches with a mesoscopic model. *J. Chem. Inf. Model.* **2016**, *56*, 101−109.
(53) Ferreira, I.; Amarante, T. D.; Weber, G. Salt dependent mesoscopic model for RNA with multiple strand concentrations. *Biophys. Chem.* **2020**, *271*, No. 106551.
(54) Oliveira, L. M.; Long, A. S.; Brown, T.; Fox, K. R.; Weber, G. Melting temperature measurement and mesoscopic evaluation of single, double and triple DNA Mismatches. *Chem. Sci.* **2020**, *11*, 8273−8287.
(55) Muniz, M. I.; Lackey, H.H.; Heemstra, J. M.; Weber, G. DNA/TNA mesoscopic modeling of melting temperatures suggest weaker hydrogen bonding of CG than in DNA/RNA. *Chem. Phys. Lett.* **2020**, No. 137413.
(56) Domljanovic, I.; Taskova, M.; Miranda, P.; Weber, G.; Astakhova, K. Nucleic acid probes—Optical and theoretical study reveals new details on strand Recognition. *Commun. Chem.* **2020**, *3*, No. 111.
(57) Tolstrup, N.; Nielsen, P. S.; Kolberg, J. G.; Frankel, A. M.; Vissing, H.; Kauppinen, S. OligoDesign: optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling. *Nucleic Acids Res.* **2003**, *31*, 3758−3762.
(58) Weber, G. TfReg: Calculating DNA and RNA Melting Temperatures and Opening Profiles with Mesoscopic Models. *Bioinformatics* **2013**, *29*, 1345−1347.
(59) Cheng, M. L.; Pectasides, E.; Hanna, G. J.; Parsons, H. A.; Choudhury, A. D.; Oxnard, G. R. Circulating tumor DNA in advanced

https://doi.org/10.1021/acs.jcim.1c00470
J. Chem. Inf. Model. 2021, 61, 3615−3624

solid tumors: Clinical relevance and future directions. *Ca-Cancer J. Clin.* **2020**, *71*, 176−190.
(60) Yamada, T.; Iwai, T.; Takahashi, G.; Kan, H.; Koizumi, M.; Matsuda, A.; Shinji, S.; Yamagishi, A.; Yokoyama, Y.; Tatsuguchi, A. Utility of KRAS mutation detection using circulating cell-free DNA from patients with colorectal cancer. *Cancer Sci.* **2016**, *107*, 936−943.
(61) Weber, G.; Haslam, N.; Whiteford, N.; Prügel-Bennett, A.; Essex, J. W.; Neylon, C. Thermal Equivalence of DNA Duplexes Without Melting Temperature Calculation. *Nat. Phys.* **2006**, *2*, 55−59.
(62) Zhang, Y.-L.; Zheng, W.-M.; Liu, J.-X.; Chen, Y. Z. Theory of DNA melting based on the Peyrard-Bishop model. *Phys. Rev. E* **1997**, *56*, 7100−7115.
(63) Weber, G.; Haslam, N.; Essex, J. W.; Neylon, C. Thermal Equivalence of DNA Duplexes for Probe Design. *J. Phys.: Condens. Matter* **2009**, *21*, No. 034106.
(64) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: Cambridge, 1988.
(65) Taskova, M.; Uhd, J.; Miotke, L.; Kubit, M.; Bell, J.; Ji, H. P.; Astakhova, K. Tandem oligonucleotide probe annealing and elongation to discriminate viral sequence. *Anal. Chem.* **2017**, *89*, 4363−4366.
(66) Taskova, M.; Astakhova, K. Fluorescent Oligonucleotides with Bis(prop-2-yn-1-yloxy)butane-1,3-diol Scaffold Rapidly Detect Disease-Associated Nucleic Acids. *Bioconjugate Chem.* **2019**, *30*, 3007−3012.
(67) Taskova, M.; Barducci, M. C.; Astakhova, K. Environmentally sensitive molecular probes reveal mutations and epigenetic 5-methyl cytosine in human oncogenes. *Org. Biomol. Chem.* **2017**, *15*, 5680−5684.
(68) Taskova, M.; Astakhova, K. Fluorescent Oligonucleotides with Bis (prop-2-yn-1-yloxy) butane-1, 3-diol Scaffold Rapidly Detect Disease-Associated Nucleic Acids. *Bioconjugate Chem.* **2019**, *30*, 3007−3012.
(69) Miotke, L.; Barducci, M.; Astakhova, K. Novel Signal-enhancing approaches for optical detection of nucleic acids−Going beyond target amplification. *Chemosensors* **2015**, *3*, 224−240.
(70) Owczarzy, R.; You, Y.; Moreira, B. G.; Manthey, J. A.; Huang, L.; Behlke, M. A.; Walder, J. A. Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry* **2004**, *43*, 3537−3554.
(71) Drukker, K.; Wu, G.; Schatz, G. C. Model simulations of DNA denaturation dynamics. *J. Chem. Phys.* **2001**, *114*, 579−590.
(72) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, No. 02B627.
(73) Peyrard, M.; Bishop, A. R. Statistical Mechanics of a Nonlinear Model for DNA denaturation. *Phys. Rev. Lett.* **1989**, *62*, 2755−2757.
(74) Cuesta-López, S.; Peyrard, M.; Graham, D. J. Model for DNA hairpin denaturation. *Eur. Phys. J. E* **2005**, *16*, 235−246.
(75) Ferreira, I.; D Amarante, T.; Weber, G. DNA terminal base pairs have weaker hydrogen bonds especially for AT under low salt concentration. *J. Chem. Phys.* **2015**, *143*, No. 175101.
(76) Ivanova, A.; Rösch, N. The structure of LNA:DNA hybrids from molecular dynamics simulations: the effect of locked nucleotides. *J. Phys. Chem. A* **2007**, *111*, 9307−9319.
(77) Vester, B.; Wengel, J. LNA (Locked Nucleic Acid) High-Affinity Targeting of Complementary RNA and DNA. *Biochemistry* **2004**, *43*, 13233−13241.
(78) Pande, V.; Nilsson, L. Insights into structure, dynamics and hydration of locked nucleic acid (LNA) strand-based duplexes from molecular dynamics simulations. *Nucleic Acids Res.* **2008**, *36*, 1508−1516.
(79) Leonard, G. A.; McAuley-Hecht, K.; Brown, T.; Hunter, W. N. Do C−H···O hydrogen bonds contribute to the stability of nucleic acid base pairs? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *51*, 136−139.
(80) Asensio, A.; Kobko, N.; Dannenberg, J. Cooperative hydrogen-bonding in adenine-thymine and guanine-cytosine base pairs. Density functional theory and Møller−Plesset molecular orbital study. *J. Phys. Chem. A* **2003**, *107*, 6441−6443.
(81) Grunenberg, J. Direct assessment of interresidue forces in Watson-Crick base pairs using theoretical compliance constants. *J. Am. Chem. Soc.* **2004**, *126*, 16310−16311.
(82) Srinivasadesikan, V.; Sahu, P. K.; Lee, S.-L. Spectroscopic probe on N−H···N, N−H···O and controversial C−H···O contact in A−T base pair: A DFT study. *Spectrochim. Acta, Part A* **2014**, *120*, 542−547.
(83) Maximiano, R. V.; Weber, G. Deoxyinosine Mismatch Parameters Calculated with a Mesoscopic Model Result in Uniform Hydrogen Bonding and Strongly Variable Stacking Interactions. *Chem. Phys. Lett.* **2015**, *631*−*632*, 87−91.
(84) Weber, G. Mesoscopic Model Parametrization of Hydrogen Bonds and Stacking Interactions of RNA from Melting Temperatures. *Nucleic Acids Res.* **2013**, *41*, No. e30.
(85) Weber, G. Optimization Method for Obtaining Nearest-Neighbour DNA Entropies and Enthalpies Directly from Melting Temperatures. *Bioinformatics* **2015**, *31*, 871−877.

https://doi.org/10.1021/acs.jcim.1c00470
J. Chem. Inf. Model. 2021, 61, 3615−3624

4 Integrative Discussion

Mesoscopic models have been applied by our group and others in the prediction of nucleic acid properties. This type of model can be applied in the study of thermodynamic stability and can also be used to extract information on molecular interactions from experimental data. The great advantage of such models is their computational efficiency, once the appropriate parameters are determined, several quantities of interest can be calculated in a few seconds on personal computers. However, as the model relies on published data, the great scientific challenge lies in parameterizing the model for different experimental conditions which requires the gathering of several sequences and their melting temperatures.

Currently, the model is parameterized for several types of nucleic acids. For instance, canonical DNA at different salt conditions,[176] DNA considering terminal effects,[47] DNA mismatches,[178] DNA with attached fluorophores,[179] RNA at high salt concentration,[180] DNA/RNA hybrids,[181] among others.[121,122,182,183] If we aim to predict thermodynamic properties for sets that contain non-parameterized nucleic acid types, a new parameter extraction is required, and oftentimes this also means an adaptation in the model.

The model does not have a unique potential to model salt or end fraying, which are studied indirectly by using data sets measured at different salt conditions or by introducing different potentials at the terminals. The adjustment of the model to perform on such conditions allows us to study problems such as the effects of sodium on duplex stability and end fraying. These are important problems in nucleic acid biophysics as they play an important part in the hybridization and denaturation of the duplexes. Furthermore, end fraying is a difficult problem, even for well-established theoretical methods such as molecular dynamics, and it is still largely unknown how sodium ions affect end fraying. Such assessment was performed before by us for canonical DNA, and we know that for DNA the terminal AT base pairs have weaker hydrogen bonds at low salt concentrations.

However, we had no idea if that would also be the case for terminal AU base pairs in RNA. To make matters worse, for a number of technical reasons, such as RNA higher degradability and susceptivity to form secondary structures, there are far less melting temperature data publicly available for RNA than for DNA, and until now, this type of analysis was essentially out of our reach due to a lack of data.

This changed when recently we had access to a rich data set of RNA melting temperatures under varying salt concentrations obtained through a different collaboration.[184] However, this data set contained melting temperatures scattered into a non-uniform range of strand concentrations, which constitutes a challenge for mesoscopic models. This is because the models perform a single molecule calculation and the melting temperature is correlated with the experimental data at a single strand concentration.

From this point of view, in our paper enclosed in Chapter 2, we adapted the model to handle multiple strand concentrations. In the regression correlation, instead of grouping by the sequence length, we grouped the strand concentration into factorized logarithmic groups. This was not a trivial problem as we had to evaluate a wide range of factors to ensure the feasibility of the technique, increasing appreciably the number of minimizations and data we had to deal with. Surprisingly, the model parameters had very little dependence on the grouping factors.

Having solved the problem of multiple strand concentrations, we found ourselves in a position to address RNA in the same way as we did for DNA. That is, consider both, salt and end fraying effects. It turns out that, unlike DNA, terminal AU does not seem to weaken their hydrogen bonds for low salt concentrations, which is yet another difference between DNA and RNA with

important consequences for thermodynamic stability. We also have found that the derived parameters are largely independent of the coarseness factor and have small variability and high robustness when dealing with oligomer concentration. Furthermore, this model modification allows us to treat data sets that were not previously available due to the strand concentration limitation, increasing the scope of problems that could be approached by the method.

Besides model adaptations to handle different experimental conditions, we frequently see ourselves challenged to deal with nucleic acid modifications, demanding a completely independent parameterization. As it happened, when our group had to model a small set of probes containing LNA modified base-pairs for a different collaboration. LNA is short for locked nucleic acid, an RNA modification that attracted considerable attention due to its ability to improve oligonucleotide stability, still keeping the major features desired in nucleic acid modifications. The problem is, they had access to a small set of sequences to train the model that also comprised other types of modifications, therefore; a full characterization was not feasible due to a lack of data. To worsen things, they have found stronger hydrogen bonding for LNA base-pairs, a major contradiction with the current understanding that stacking should be the dominant LNA stabilizer. Clearly, a careful validation based on a larger set of data and free of additional modifications was needed to confirm or refute their findings.

For this reason, we extended this collaboration, and the principal investigator of the project, Prof. Kira Astakhova, became the co-advisor of this thesis. Her projects often take advantage of the LNA stabilizing effect on oncogenic and diagnostic oligonucleotide probes. However, not every probe design of LNA gives good results, and some even fail terribly. To date, even careful NMR and X-ray diffraction experiments have given conflicting results about the origin of LNA's outstanding stability, and without answering these fundamental questions, LNA probe design remains mostly an expensive trial-and-error business. In this way, a better understanding of LNA intramolecular properties could shed some light on what should be addressed in the design of LNA probes and save experimental efforts.

Favorably, due to LNA's popularity, a large volume of published experimental data could be used in a more thorough study. We collected over 300 experimental LNA melting temperatures, which was not trivial as these data came from various sources and had to be carefully curated before proper parameterization. In addition, the presence of LNA in a DNA oligonucleotide increases the number of possible combinations dramatically, increasing the number of parameters demanding substantial computational resources. Nonetheless, we were able to derive a set of parameters for all possible single LNA modifications, which were used to test all possible LNA insertions and predict the melting temperature of ultrasensitive probes for oncogene detection of *BRAF*, *KRAS* and *EGFR* variants.

This was followed by a curated selection of each variant in the pool of predictions, based on their stability and number of inserted modifications, which is directly related to their melting temperature. These probes were synthesized and had their melting temperature measured by Astakhova's group. The agreement between the predictions and the experimentally measured probes was outstanding of 1 ° C. Furthermore, our parameterization provides a tailored selection of candidate probes for synthesis; possibly minimizing the final experimental costs. Also, our results completely confirm that increased hydrogen bonding is the major source of LNA stability and stacking interactions have a very small influence. Basically, LNA-modified AT becomes as strong as a normal CG, while a modified CG hydrogen bond increases notably. This work is discussed in detail in our manuscript enclosed in Chapter 3. The project was partially developed during my exchange period at the Danmarks Tekniske Universitet — DTU in the Astakhova group and also helped guide other projects.

The complete parameterization of different types of nucleic acids by mesoscopic models and their adaptation to treat any strand concentration allows an application in bioinformatics that would not be possible by using traditional methods of atomistic simulation or other types of intense computational treatments. Modeling of LNAs is of technological interest, for example,

isothermal amplification is a diagnostic tool that uses LNA to increase the thermal stability and thereby improve the results of the technique.[185] The parameters could also be used to computationally estimate the binding efficiency of millions of candidate probes that could bring great savings in the development of diagnostic systems.[90] It also opens up the possibility of performing screening for drugs based on oligonucleotides[186] and even opens product development perspectives as in the case of the Komiya et al.[187] patent. Also, from a scientific point of view, as we treat new types of nucleic acids with the method we end up consolidating the technique and expanding the scope of where mesoscopic models can be applied.

5 Conclusion

The applicability of mesoscopic models such as the one employed by us has been seizing attention in the design of technological approaches based on nucleic acids. As the model can give useful insights about hydrogen bonds and stacking interactions, the main sources of the cohesion of nucleic acids by using thermodynamic experimental data as an input.

However, there are some constraints, as the model relies on experimental data we need to characterize it for each experimental situation. In this way as further as we address different data sets and deal with different obstacles, further we can transfer the basic science knowledge embodied by the model to scientific applications.

Here, we developed an adaptation in the model to deal with multiple strand concentrations which in summary brings to the scope of the model data sets that were not previously available due to data restrictions in mesoscopic models as they are mainly single molecule calculations. This adjustment enabled us to model an RNA data set at different salt concentrations that were out of our reach as the measurements were made at different strand concentrations. It also demonstrates the very small dependence of the model parameters on the strand concentration increasing the scope of data accessible to the model.

Having this at hand, enabled us to model not only saline effects in RNA but also terminal effects, two characterizations that lacked into the model. It turns out that, unlike DNA, an AU base pair at the terminal is not as sensitive to salt concentrations as a terminal AT and the behaviour of the parameters within salt concentration is largely uniform.

Furthermore, we applied the model in a modified nucleic acid, LNA, a modification that is able to increase the stability and improve the binding of oligonucleotide probes. LNA's outstanding stability has been often attributed to an improvement in stacking interactions or still to a strengthening of hydrogen bonds and has been achieving conflicting results even in well-established techniques such as NMR and X-ray diffraction experiments. The use of mesoscopic models could shed light on the intramolecular interactions, for instance, hydrogen bonding and stacking interactions, which could give better insights into the source of LNA increased stability.

As a matter of fact, we were able to successfully derive a complete parameterization of single LNAs, which was used to predict tailored sites to LNA insertion in oncogenic probes. In agreement with our previous assumptions, the LNA stability source is mainly from an improvement in hydrogen bonds and the dependence with stacking is barely significant.

Not to mention these two characterizations bring a widening to the model scope and open opportunities to further collaborations and experimental applications, consolidating the model applicability.

Bibliography

[1] Erwin Chargaff. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, 6(6):201–209, 1950.

[2] Rosalind E Franklin and Raymond G Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.

[3] M. H. F. Wilkins, R. G. Gosling, and W. E. Seeds. Nucleic acid: an extensible molecule? *Nature*, 167:759–760, 1951.

[4] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[5] Linus Pauling and Robert B Corey. A proposed structure for the nucleic acids. *Proc. Natl. Acad. Sci. USA*, 39(2):84, 1953.

[6] OpenStax Biology. DNA structure, May 2012. URL `http://cnx.org/contents/GFy_h8cu@10.53:rZudN6XP@2/Introduction`.

[7] M. Peyrard. Nonlinear dynamics and statistical physics of DNA. *Nonlinearity*, 17:R1, 2004.

[8] Wolfram Saenger. *Principles of nucleic acid structure*. Springer Science & Business Media, 2013.

[9] OpenStax Biology. RNA structure, Jan 2013. URL `https://cnx.org/contents/XcbB5HTY@5.8:hRuNts1p@5/Structure-and-Function-of-RNA`.

[10] Alexander Rich and David R Davies. A new two stranded helical structure: polyadenylic acid and polyuridylic acid. *Journal of the American Chemical Society*, 78(14):3548–3549, 1956.

[11] Ignacio Tinoco Jr and Carlos Bustamante. How RNA folds. *Journal of molecular biology*, 293(2):271–281, 1999.

[12] Pan TX Li, Jeffrey Vieregg, and Ignacio Tinoco Jr. How RNA unfolds and refolds. *Annu. Rev. Biochem.*, 77:77–100, 2008.

[13] Tomoko Kuwabara, Jenny Hsieh, Kinichi Nakashima, Kazunari Taira, and Fred H Gage. A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell*, 116(6): 779–793, 2004.

[14] Gregory J Hannon. RNA interference. *Nature*, 418(6894):244–251, 2002.

[15] Gunter Meister and Thomas Tuschl. Mechanisms of gene silencing by double-stranded RNA. *Nature*, 431(7006):343–349, 2004.

[16] Shizuo Akira and Kiyoshi Takeda. Toll-like receptor signalling. *Nature Reviews Immunology*, 4(7):499–511, 2004.

[17] Jan Lipfert, Sebastian Doniach, Rhiju Das, and Daniel Herschlag. Understanding nucleic acid–ion interactions. *Annual Review of Biochemistry*, 83:813–841, 2014.

[18] Sarah A Woodson. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.*, 9(2):104–109, 2005.

[19] David E Draper. Folding of RNA tertiary structure: linkages between backbone phosphates, ions, and water. *Biopoly.*, 99(12):1105–1113, 2013.

[20] Eda Koculi, Changbong Hyeon, D Thirumalai, and Sarah A Woodson. Charge density of divalent metal cations determines RNA stability. *J. Am. Chem. Soc.*, 129(9):2676–2682, 2007.

[21] Peter A Lemaire, Eric Anderson, Jeffrey Lary, and James L Cole. Mechanism of PKR activation by dsRNA. *J. Mol. Biol.*, 381(2):351–360, 2008.

[22] Kun Xi, Feng-Hua Wang, Gui Xiong, Zhong-Liang Zhang, and Zhi-Jie Tan. Competitive binding of Mg^{2+} and Na^{+} ions to nucleic acids: from helices to tertiary structures. *Biophys. J.*, 114(8):1776–1790, 2018.

[23] Yuan-Yan Wu, Zhong-Liang Zhang, Jin-Si Zhang, Xiao-Long Zhu, and Zhi-Jie Tan. Multivalent ion-mediated nucleic acid helix-helix interactions: RNA versus DNA. *Nucleic Acids Res.*, 43(12):6156–6165, 2015.

[24] Serdal Kirmizialtin and Ron Elber. Computational exploration of mobile ion distributions around RNA duplex. *J. Phys. Chem. B*, 114(24):8207–8220, 2010.

[25] Aleksander V Drozdetski, Igor S Tolokh, Lois Pollack, Nathan Baker, and Alexey V Onufriev. Opposing effects of multivalent ions on the flexibility of DNA and RNA. *Phys. Rev. Lett.*, 117(2):028101, 2016.

[26] Zhi-Jie Tan and Shi-Jie Chen. Salt contribution to RNA tertiary structure folding stability. *Biophys. J.*, 101(1):176–187, 2011.

[27] Erik D Holmstrom, Julie L Fiore, and David J Nesbitt. Thermodynamic origins of monovalent facilitated RNA folding. *Biochem.*, 51(18):3732–3743, 2012.

[28] Jonathan L Chen, Abigael L Dishler, Scott D Kennedy, Ilyas Yildirim, Biao Liu, Douglas H Turner, and Martin J Serra. Testing the nearest neighbor model for canonical RNA base pairs: Revision of GU parameters. *Biochem.*, 51(16):3508–3522, 2012.

[29] Lance G Laing, Thomas C Gluick, and David E Draper. Stabilization of RNA structure by Mg ions: specific and non-specific effects. *J. Mol. Biol.*, 237(5):577–587, 1994.

[30] Anna Pyle. Metal ions in the structure and function of RNA. *JBIC Journal of Biological Inorganic Chemistry*, 7(7-8):679–690, 2002.

[31] David E Draper. A guide to ions and RNA structure. *RNA*, 10(3):335–343, 2004.

[32] Dominic Lambert, Desirae Leipply, Ross Shiman, and David E Draper. The influence of monovalent cation size on the stability of RNA tertiary structures. *J. Mol. Biol.*, 390(4): 791–804, 2009.

[33] Ivana Beššeová, Michal Otyepka, Kamila Réblová, and Jiří Šponer. Dependence of A-RNA simulations on the choice of the force field and salt strength. *Phys. Chem. Chem. Phys.*, 11(45):10701–10711, 2009.

[34] Ivana Beššeová, Pavel Banáş, Petra Kührová, Pavlína Koşinová, Michal Otyepka, and Jiří Šponer. Simulations of A-RNA duplexes. the effect of sequence, solute force field, water model, and salt concentration. *J. Phys. Chem. B*, 116(33):9899–9916, 2012.

[35] JJ Virtanen, TR Sosnick, and KF Freed. Ionic strength independence of charge distributions in solvation of biomolecules. *J. Chem. Phys.*, 141(22):12B604_1, 2014.

[36] Lei Bao, Jun Wang, and Yi Xiao. Dynamics of metal ions around an RNA molecule. *Phys. Rev. E*, 99(1):012420, 2019.

[37] Lei Jin, Ya-Zhou Shi, Chen-Jie Feng, Ya-Lan Tan, and Zhi-Jie Tan. Modeling structure, stability, and flexibility of double-stranded RNAs in salt solutions. *Biophys. J.*, 115(8): 1403–1416, 2018.

[38] Thomas R Einert and Roland R Netz. Theory for RNA folding, stretching, and melting including loops and salt. *Biophys. J.*, 100(11):2745–2753, 2011.

[39] Zhi-Jie Tan and Shi-Jie Chen. RNA helix stability in mixed Na+/Mg2+ solution. *Biophys. J.*, 92(10):3615–3632, 2007.

[40] Zhi-Jie Tan and Shi-Jie Chen. Importance of diffuse metal ion binding to RNA. *Met. Ions Life Sci.*, 9:101, 2011.

[41] Francesco Colizzi and Giovanni Bussi. RNA unwinding from reweighted pulling simulations. *J. Am. Chem. Soc.*, 134(11):5173–5179, 2012.

[42] K. Snoussi and J. L. Leroy. Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry*, 40(30):8898–8904, 2001.

[43] Lin-Tai Da, Fátima Pardo-Avila, Liang Xu, Daniel-Adriano Silva, Lu Zhang, Xin Gao, Dong Wang, and Xuhui Huang. Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nature Comm.*, 7:11244, 2016.

[44] MD Betterton and Frank Jülicher. Opening of nucleic-acid double strands by helicases: active versus passive opening. *Phys. Rev. E*, 71(1):011904, 2005.

[45] Koree Clanton-Arrowood, John McGurk, and Susan J Schroeder. $3'$ terminal nucleotides determine thermodynamic stabilities of mismatches at the ends of RNA helices. *Biochemistry*, 47(50):13418–13427, 2008.

[46] Marie Zgarbová, Michal Otyepka, Jiri Sponer, Filip Lankas, and Petr Jurečka. Base pair fraying in molecular dynamics simulations of DNA and RNA. *J. Chem. Theory Comput.*, 10(8):3177–3189, 2014. doi:10.1021/ct500120v.

[47] Izabela Ferreira, Tauanne D. Amarante, and Gerald Weber. DNA terminal base pairs have weaker hydrogen bonds especially for AT under low salt concentration. *J. Chem. Phys.*, 143:175101, 2015. doi:10.1063/1.4934783.

[48] Paul C Zamecnik and Mary L Stephenson. Inhibition of rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proceedings of the National Academy of Sciences*, 75(1):280–284, 1978.

[49] C A. Stein. The experimental use of antisense oligonucleotides: a guide for the perplexed. *The Journal of clinical investigation*, 108,5:641–4, 2001. doi:10.1172/JCI13885.

[50] S T Crooke. Therapeutic applications of oligonucleotides. *Annual Review of Pharmacology and Toxicology*, 32(1):329–376, 1992. doi:10.1146/annurev.pa.32.040192.001553. URL `https://doi.org/10.1146/annurev.pa.32.040192.001553`. PMID: 1605571.

[51] Potaman V.N Soyfer V.N. *The Discovery of Triple-Stranded Nucleic Acids. In: Triple-Helical Nucleic Acids.* Springer, New York, NY, 1996. ISBN 978-1-4612-3972-7. doi:10.1007/978-1-4612-3972-7_1.

[52] Eugen Uhlmann and Anusch Peyman. Antisense oligonucleotides: a new therapeutic principle. *Chemical Reviews*, 90(4):543–584, 1990. doi:10.1021/cr00102a001. URL `https://doi.org/10.1021/cr00102a001`.

[53] John F. Milligan, Mark D. Matteucci, and John C. Martin. Current concepts in antisense drug design. *Journal of Medicinal Chemistry*, 36(14):1923–1937, 1993. doi:10.1021/jm00066a001. URL `https://doi.org/10.1021/jm00066a001`.

[54] P E Nielsen. DNA analogues with nonphosphodiester backbones. *Annual Review of Biophysics and Biomolecular Structure*, 24(1):167–183, 1995. doi:10.1146/annurev.bb.24.060195.001123. URL `https://doi.org/10.1146/annurev.bb.24.060195.001123`. PMID: 7663113.

[55] Olga Khorkova and Claes Wahlestedt. Oligonucleotide therapies for disorders of the nervous system. *Nature biotechnology*, 35(3):249, 2017.

[56] Rosemary Kanasty, Joseph Robert Dorkin, Arturo Vegas, and Daniel Anderson. Delivery materials for siRNA therapeutics. *Nature materials*, 12(11):967–977, 2013.

[57] Anastasia Khvorova and Jonathan K Watts. The chemical evolution of oligonucleotide therapies of clinical utility. *Nature biotechnology*, 35(3):238, 2017.

[58] Elizabeth A Jolley and Brent M Znosko. The loss of a hydrogen bond: Thermodynamic contributions of a non-standard nucleotide. *Nucleic Acids Res.*, 45(3):1479–1487, 2016. doi:10.1093/nar/gkw830.

[59] Francis H Martin, Miguel M Castro, Fareed Aboul-ela, and Ignacio Tinoco. Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Research*, 13(24): 8927–8938, 1985.

[60] Nguyen T. Thuong and Claude Hélène. Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angewandte Chemie International Edition in English*, 32(5):666–690, 1993. doi:10.1002/anie.199306661. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.199306661`.

[61] C. Giovannangeli and C. Hélene. Progress in developments of triplex-based strategies. *Antisense and Nucleic Acid Drug Development*, 7(4):413–421, 1997. doi:10.1089/oli.1.1997.7.413. URL `https://doi.org/10.1089/oli.1.1997.7.413`. PMID: 9303193.

[62] H Jakob Larsen, Thomas Bentin, and Peter E Nielsen. Antisense properties of peptide nucleic acid. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1489 (1):159–166, 1999.

[63] Sergei M Gryaznov. Oligonucleotide N3' -¿ P5' phosphoramidates as potential therapeutic agents. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1489(1): 131–140, 1999.

[64] Eveline Lescrinier, Robert Esnouf, Jan Schraml, Roger Busson, HA Heus, CW Hilbers, and Piet Herdewijn. Solution structure of a HNA/RNA hybrid. *Chemistry & biology*, 7 (9):719–731, 2000.

[65] Robert M Hudziak, JAMES SUMMERTON, DWIGHT D WELLER, and PATRICK L IVERSEN. Antiproliferative effects of steric blocking phosphorodiamidate morpholino antisense agents directed against c-myc. *Antisense and Nucleic Acid Drug Development*, 10(3):163–176, 2000.

[66] Charles Marwick. First antisense drug will treat CMV retinitis. *Jama*, 280(10):871–871, 1998.

[67] Richard I Hogrefe. An antisense oligonucleotide primer. *Antisense and Nucleic Acid Drug Development*, 9(4):351–357, 1999.

[68] S. M. Freier and K.-H. Altmann. The ups and downs of nucleic acid duplex stability: structure-stability studies in chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.*, 25(22):4429–4443, 1997.

[69] Irina Kira Astakhova, Evgeniya Samokhina, B Ravindra Babu, and Jesper Wengel. Novel (phenylethynyl) pyrene–LNA constructs for fluorescence SNP sensing in polymorphic nucleic acid targets. *ChemBioChem*, 13(10):1509–1519, 2012.

[70] J. Bhattacharyya, S. Maiti, S. Muhuri, S. Nakano, D. Miyoshi, and N. Sugimoto. Effect of locked nucleic acid modification on thermal stability of non-canonical DNA structure. *Biochemistry*, 50:7414–7425, 2011.

[71] Alexei A Koshkin, Sanjay K Singh, Poul Nielsen, Vivek K Rajwanshi, Ravindra Kumar, Michael Meldgaard, Carl Erik Olsen, and Jesper Wengel. LNA (locked nucleic acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron*, 54(14):3607–3630, 1998.

[72] Birte Vester and Jesper Wengel. LNA (locked nucleic acid) high-affinity targeting of complementary RNA and DNA. *Biochemistry*, 43(42):13233–13241, 2004.

[73] Asli Silahtaroglu, Henrik Pfundheller, Alexei Koshkin, Niels Tommerup, and Sakari Kauppinen. LNA-modified oligonucleotides are highly efficient as FISH probes. *Cytogenetic and genome research*, 107:32–7, 02 2004. doi:10.1159/000079569.

[74] Satoshi Obika, Yoshiyuki Hari, Ken-ichiro Morio, and Takeshi Imanishi. Triplex formation by an oligonucleotide containing conformationally locked c-nucleoside, 5-(2-O, 4-C-methylene-β-dribofuranosyl) oxazole. *Tetrahedron Letters*, 41(2):221–224, 2000.

[75] Claes Wahlestedt, Peter Salmi, Liam Good, Johanna Kela, Thomas Johnsson, Tomas Høkfelt, Christian Broberger, Frank Porreca, Josephine Lai, Kunkun Ren, et al. Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proceedings of the National Academy of Sciences*, 97(10):5633–5638, 2000.

[76] Henrik Ørum, Mogens H Jakobsen, Troels Koch, Jens Vuust, and Martin B Borre. Detection of the factor V Leiden mutation by direct allele-specific hybridization of PCR amplicons to photoimmobilized locked nucleic acids. *Clinical chemistry*, 45(11):1898–1905, 1999.

[77] Sanjay K. Singh, Alexei A. Koshkin, Jesper Wengel, and Poul Nielsen. LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chem. Commun.*, pages 455–456, 1998. doi:10.1039/A708608C. URL `http://dx.doi.org/10.1039/A708608C`.

[78] A. Pasternak and J. Wengel. Thermodynamics of RNA duplexes modified with unlocked nucleic acid nucleotides. *Nucleic Acids Res.*, 38(19):6697–6706, 2010. doi:10.1093/nar/gkq561.

[79] Harleen Kaur, Amit Arora, Jesper Wengel, and Souvik Maiti. Thermodynamic, counterion, and hydration effects for the incorporation of locked nucleic acid nucleotides into DNA duplexes. *Biochemistry*, 45(23):7347–7355, 2006.

[80] Jens Kurreck, Eliza Wyszko, Clemens Gillen, and Volker A Erdmann. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic acids research*, 30(9):1911–1918, 2002.

[81] Kelly L Robertson and Gary J Vora. Locked nucleic acid and flow cytometry-fluorescence in situ hybridization for the detection of bacterial small noncoding RNAs. *Applied and environmental microbiology*, 78(1):14–20, 2012.

[82] Y Zhang, Z Qu, S Kim, V Shi, B Liao, P Kraft, R Bandaru, Y Wu, LM Greenberger, and ID Horak. Down-modulation of cancer targets using locked nucleic acid (LNA)-based antisense oligonucleotides without transfection. *Gene therapy*, 18(4):326–333, 2011.

[83] Prabodhika Mallikaratchy, Jeffery Gardner, Lars Ulrik R Nordstrøm, Nicholas J Veomett, Michael R McDevitt, Mark L Heaney, and David A Scheinberg. A self-assembling short oligonucleotide duplex suitable for pretargeting. *Nucleic acid therapeutics*, 23(4):289–299, 2013.

[84] Mathilde H Josefsen, Charlotta Løfstrøm, Helle M Sommer, and Jeffrey Hoorfar. Diagnostic PCR: comparative sensitivity of four probe chemistries. *Molecular and cellular probes*, 23(3-4):201–203, 2009.

[85] Luca Morandi, Dario De Biase, Michela Visani, Valentina Cesari, Giovanna De Maglio, Stefano Pizzolitto, Annalisa Pession, and Giovanni Tallini. Allele specific locked nucleic acid quantitative PCR (ASLNAqPCR): an accurate and cost-effective assay to diagnose and quantify KRAS and BRAF mutation. *PloS one*, 7(4):e36084, 2012.

[86] Maxime Fontanilles, Florent Marguet, Philippe Ruminy, Carole Basset, Adrien Noel, Ludivine Beaussire, Mathieu Viennot, Pierre-Julien Viailly, Kevin Cassinari, Pascal Chambon, et al. Simultaneous detection of EGFR amplification and EGFRvIII variant using digital pcr-based method in glioblastoma. *Acta neuropathologica communications*, 8(1): 1–10, 2020.

[87] Irina V Astakhova, Alexey V Ustinov, Vladimir A Korshun, and Jesper Wengel. LNA for optimization of fluorescent oligonucleotide probes: improved spectral properties and target binding. *Bioconjugate chemistry*, 22(4):533–539, 2011.

[88] Michael E Østergaard, Pallavi Cheguru, Madhusudhan R Papasani, Rodney A Hill, and Patrick J Hrdlicka. Glowing locked nucleic acids: brightly fluorescent probes for detection of nucleic acids in cells. *Journal of the American Chemical Society*, 132(40):14221–14228, 2010.

[89] Saskia Hussung, Marie Follo, Rhena FU Klar, Sandra Michalczyk, Kornelia Fritsch, Friederike Nollmann, Julian Hipp, Justus Duyster, Florian Scherer, Nikolas von Bubnoff, et al. Development and clinical validation of discriminatory multi-target digital droplet PCR assays for the detection of hot spot KRAS and NRAS mutations in cell-free DNA. *The Journal of Molecular Diagnostics*, 2020.

[90] Padmavathy Bakthavathsalam, Guillaume Longatte, Slade O Jensen, Mike Manefield, and J Justin Gooding. Locked nucleic acid molecular beacon for multiplex detection of loop mediated isothermal amplification. *Sensors and Actuators B: Chemical*, 268:255–263, 2018.

[91] Parisa Bagheri, Mohammadreza Sharifi, and Ava Ghadiri. Downregulation of MIR100HG induces apoptosis in human megakaryoblastic leukemia cells. *Indian Journal of Hematology and Blood Transfusion*, pages 1–8, 2020.

[92] Kaiming Zhang, Ivan N Zheludev, Rachel J Hagey, Marie Teng-Pei Wu, Raphael Haslecker, Yixuan J Hou, Rachael Kretsch, Grigore D Pintilie, Ramya Rangan, Wipapat Kladwang, et al. Cryo-electron microscopy and exploratory antisense targeting of the 28-kda frameshift stimulation element from the SARS-CoV-2 RNA genome. *bioRxiv*, 2020.

[93] Kenji Rowel Q. Lim, Rika Maruyama, Yusuke Echigoya, Quynh Nguyen, Aiping Zhang, Hunain Khawaja, Sreetama Sen Chandra, Takako Jones, Peter Jones, Yi-Wen Chen, and Toshifumi Yokota. Inhibition of DUX4 expression with antisense LNA gapmers as a therapy for facioscapulohumeral muscular dystrophy. *Proceedings of the National Academy of Sciences*, 117(28):16509–16515, 2020. ISSN 0027-8424. doi:10.1073/pnas.1909649117. URL `https://www.pnas.org/content/117/28/16509`.

[94] Kathrin S Schmidt, Sandra Borkowski, Jens Kurreck, Andrew W Stephens, Rolf Bald, Maren Hecht, Matthias Friebe, Ludger Dinkelborg, and Volker A Erdmann. Application of locked nucleic acids to improve aptamer in vivo stability and targeting function. *Nucleic acids research*, 32(19):5757–5765, 2004.

[95] Fabien Darfeuille, Sandrine Reigadas, Jens Bo Hansen, Henrik Orum, Carmelo Di Primo, and Jean-Jacques Toulmé. Aptamers targeted to an RNA hairpin show improved specificity compared to that of complementary oligonucleotides. *Biochemistry*, 45(39):12076–12082, 2006.

[96] Melissa Wojtyniak, Boris Schmidtgall, Philine Kirsch, and Christian Ducho. Towards zwitterionic oligonucleotides with improved properties: the NAA/LNA-gapmer approach. *ChemBioChem*, n/a(n/a), 2020. doi:10.1002/cbic.202000450.

[97] Birte Vester, Lars Bo Lundberg, Mads D. Sørensen, B. Ravindra Babu, Stephen Douthwaite, and Jesper Wengel. LNAzymes: incorporation of LNA-type monomers into DNAzymes markedly increases RNA cleavage. *Journal of the American Chemical Society*, 124(46):13682–13683, 2002. PMID: 12431091.

[98] Martin R Jakobsen, Joost Haasnoot, Jesper Wengel, Ben Berkhout, and Jørgen Kjems. Efficient inhibition of HIV-1 expression by LNA modified antisense oligonucleotides and DNAzymes targeted to functionally selected binding sites. *Retrovirology*, 4(1):1–13, 2007.

[99] Vaibhav M Jadhav, Vinod Scaria, and Souvik Maiti. Antagomirzymes: oligonucleotide enzymes that specifically silence microRNA function. *Angewandte Chemie International Edition*, 48(14):2557–2560, 2009.

[100] Lin Wang, Chaoyong James Yang, Colin D Medley, Steven A Benner, and Weihong Tan. Locked nucleic acid molecular beacons. *Journal of the American Chemical Society*, 127 (45):15664–15665, 2005.

[101] H Orum and J Wengel. Locked nucleic acids: a promising molecular family for gene-function analysis and antisense drug development. *Current opinion in molecular therapeutics*, 3(3):239–243, 2001.

[102] Erno Wienholds, Wigard P Kloosterman, Eric Miska, Ezequiel Alvarez-Saavedra, Eugene Berezikov, Ewart de Bruijn, H Robert Horvitz, Sakari Kauppinen, and Ronald HA Plasterk. MicroRNA expression in zebrafish embryonic development. *Science*, 309(5732): 310–311, 2005.

[103] Rune Thomsen, Peter Stein Nielsen, and Torben Heick Jensen. Dramatically improved RNA in situ hybridization signals using LNA-modified probes. *RNA*, 11:1745–1748, 2005. doi:10.1261/rna.2139705.

[104] Joacim Elmén, Morten Lindow, Sylvia Schutz, Matthew Lawrence, Andreas Petri, Susanna Obad, Marie Lindholm, Maj Hedtjarn, Henrik Frydenlund Hansen, Urs Berger, et al. LNA-mediated microRNA silencing in non-human primates. *Nature*, 452(7189):896–899, 2008.

[105] Maurice B. Fluitt, Narayan Shivapurkar, Manju Kumari, Sarojini Singh, Lijun Li, Swasti Tiwari, and Carolyn M. Ecelbarger. Systemic inhibition of miR-451 increases fibrotic signaling and diminishes autophagic response to exacerbate renal damage in tallyho/jng mice. *American Journal of Physiology-Renal Physiology*, 0(0):null, 0. doi:10.1152/ajprenal.00594.2019. URL `https://doi.org/10.1152/ajprenal.00594.2019`. PMID: 32715758.

[106] Songqian Huang, Yuki Ichikawa, Kazutoshi Yoshitake, Yoji Igarashi, Shigeharu Kinoshita, Md Asaduzzaman, Fumito Omori, Kaoru Maeyama, Kiyohito Nagai, Shugo Watabe, et al. Potential silencing of gene expression by PIWI-interacting RNAs (piRNAs) in somatic tissues in mollusk. *bioRxiv*, 2020.

[107] Marina Kalinina, Dmitry Skvortsov, Svetlana Kalmykova, Timofei Ivanov, Olga Dontsova, and Dmitri Pervouchine. Multiple competing RNA structures dynamically control alternative splicing in human ATE1 gene. *bioRxiv*, 2020.

[108] Nana Jacobsen, Joan Bentzen, Michael Meldgaard, Mogens Havsteen Jakobsen, Mogens Fenger, Sakari Kauppinen, and Jan Skouv. LNA-enhanced detection of single nucleotide polymorphisms in the apolipoprotein e. *Nucleic acids research*, 30(19):e100–e100, 2002.

[109] Dwaine A Braasch and David R Corey. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chemistry & biology*, 8(1):1–7, 2001.

[110] R. Owczarzy, Y. You, C.L. Groth, and A.V. Tataurov. Stability and mismatch discrimination of locked nucleic acid-DNA duplexes. *Biochemistry*, 50:9352–9367, 2011.

[111] Gilles Bruylants, Marina Boccongelli, Karim Snoussi, and Kristin Bartik. Comparison of the thermodynamics and base-pair dynamics of a full LNA: DNA duplex and of the isosequential DNA: DNA duplex. *Biochemistry*, 48(35):8473–8482, 2009.

[112] Curtis B Hughesman, Robin FB Turner, and Charles A Haynes. Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified lna. *Biochemistry*, 50(23):5354–5368, 2011.

[113] Kareem Fakhfakh, Olivia Marais, Xin Bo Justin Cheng, Jorge Real Castañeda, Curtis B Hughesman, and Charles Haynes. Molecular thermodynamics of lna: LNA base pairs and the hyperstabilizing effect of 5'-proximal lna: DNA base pairs. *AIChE Journal*, 61(9): 2711–2731, 2015.

[114] Christina B Nielsen, Sanjay K Singh, Jesper Wengel, and Jens Peter Jacobsen. The solution structure of a locked nucleic acid (LNA) hybridized to DNA. *Journal of Biomolecular Structure and Dynamics*, 17(2):175–191, 1999.

[115] Elzbieta Kierzek, Anna Pasternak, Karol Pasternak, Zofia Gdaniec, Ilyas Yildirim, Douglas H Turner, and Ryszard Kierzek. Contributions of stacking, preorganization, and hydrogen bonding to the thermodynamic stability of duplexes between RNA and 2'-o-methyl RNA with locked nucleic acids. *Biochemistry*, 48(20):4377–4387, 2009.

[116] Gitte A. Jensen, Sanjay K. Singh, Ravindra Kumar, Jesper Wengel, and Jens Peter Jacobsen. A comparison of the solution structures of an LNA:DNA duplex and the unmodified DNA:DNA duplex. *J. Chem. Soc., Perkin Trans. 2*, pages 1224–1232, 2001. doi:10.1039/B008431J. URL `http://dx.doi.org/10.1039/B008431J`.

[117] Anela Ivanova and Notker Rösch. The structure of LNA: DNA hybrids from molecular dynamics simulations: the effect of locked nucleotides. *The Journal of Physical Chemistry A*, 111(38):9307–9319, 2007.

[118] Patricia M McTigue, Raymond J Peterson, and Jason D Kahn. Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation. *Biochemistry*, 43(18):5388–5405, 2004.

[119] Yong You, Bernardo G Moreira, Mark A Behlke, and Richard Owczarzy. Design of LNA probes that improve mismatch discrimination. *Nucleic acids research*, 34(8):e60–e60, 2006.

[120] Gerald Weber, Jonathan W. Essex, and Cameron Neylon. Probing the microscopic flexibility of DNA from melting temperatures. *Nat. Phys.*, 5:769–773, 2009. doi:10.1038/nphys1371.

[121] Rodolfo Vieira Maximiano and Gerald Weber. Deoxyinosine mismatch parameters calculated with a mesoscopic model result in uniform hydrogen bonding and strongly variable stacking interactions. *Chem. Phys. Lett.*, 631–632:87–91, 2015. doi:10.1016/j.cplett.2015.04.045.

[122] Maria Izabel Muniz, Hershel H. Lackey, Jennifer M. Heemstra, and Gerald Weber. DNA/TNA mesoscopic modeling of melting temperatures suggest weaker hydrogen bonding of CG than in DNA/RNA. *Chem. Phys. Lett.*, page 137413, 2020. doi:10.1016/j.cplett.2020.137413. URL `https://www.sciencedirect.com/science/article/abs/pii/S0009261420303286`.

[123] Jesse L Montgomery, Lindsay N Sanford, and Carl T Wittwer. High-resolution DNA melting analysis in clinical research and diagnostics. *Expert Review of Molecular Diagnostics*, 10 (2):219–240, 2010. doi:10.1586/erm.09.84. URL `https://doi.org/10.1586/erm.09.84`.

[124] Panagiotis Madesis, Ioannis Ganopoulos, Argiriou Anagnostis, and Athanasios Tsaftaris. The application of Bar-HRM (barcode DNA-High Resolution Melting) analysis for authenticity testing and quantitative detection of bean crops (leguminosae) without prior DNA purification. *Food Control*, 25(2):576 – 582, 2012. ISSN 0956-7135. doi:https://doi.org/10.1016/j.foodcont.2011.11.034. URL `http://www.sciencedirect.com/science/article/pii/S0956713511005160`.

[125] Robert J Macfarlane, Byeongdu Lee, Matthew R Jones, Nadine Harris, George C Schatz, and Chad A Mirkin. Nanoparticle superlattice engineering with DNA. *Science*, 334(6053): 204–208, 2011. doi:10.1126/science.1210493.

[126] Lulu Qian and Erik Winfree. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332(6034):1196–1201, 2011.

[127] Peter H von Hippel, Neil P Johnson, and Andrew H Marcus. Fifty years of DNA "breathing": Reflections on old and new approaches. *Biopolymers*, 99(12):923–954, 2013.

[128] Douglas Poland and Harold A Scheraga. Kinetics of the helix—coil transition in polyamino acids. *The Journal of Chemical Physics*, 45(6):2071–2090, 1966.

[129] JL Lebowitz, JK Percus, and L Verlet. Ensemble dependence of fluctuations with application to machine computations. *Physical Review*, 153(1):250, 1967.

[130] Simona Cocco and Remi Monasson. Statistical mechanics of torque induced denaturation of DNA. *Phys. Rev. Lett.*, 83:5178–81, 1999.

[131] Nikos Theodorakopoulos, Thierry Dauxois, and Michel Peyrard. Order of the phase transition in models of DNA thermal denaturation. *Phys. Rev. Lett.*, 85(1):6–9, 2000.

[132] M. Peyrard. Melting the double helix. *Nat. Phys.*, 2:13–14, 2006. doi:10.1038/nphys197.

[133] Yariv Kafri, David Mukamel, and Luca Peliti. Why is the DNA denaturation transition first order? *Phys. Rev. Lett.*, 85(23):4988–4991, 2000.

[134] Paolo Grinza and Alessandro Mossa. Topological origin of the phase transition in a model of DNA denaturation. *Phys. Rev. Lett.*, 92(15):158102, 2004.

[135] Sherrie Schreiber-Gosche and Robert A Edwards. Thermodynamics of oligonucleotide duplex melting. *J. Chem. Educ.*, 86(5):644, 2009.

[136] René Thomas. The denaturation of DNA. *Gene*, 135(1-2):77–79, 1993.

[137] Tigran V Chalikian, Jens Vølker, G Eric Plum, and Kenneth J Breslauer. A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques. *Proceedings of the National Academy of Sciences*, 96(14):7853–7858, 1999.

[138] Jean-Louis Mergny and Jean-Claude Maurizot. Fluorescence resonance energy transfer as a probe for g-quartet formation by a telomeric repeat. *ChemBioChem*, 2(2):124–132, 2001.

[139] Perrine Cahen, Michel Luhmer, Catherine Fontaine, Claude Morat, Jacques Reisse, and Kristin Bartik. Study by 23Na-NMR, 1H-NMR, and ultraviolet spectroscopy of the thermal stability of an 11-basepair oligonucleotide. *Biophysical journal*, 78(2):1059–1069, 2000.

[140] Qiu Guo, Min Lu, and Neville R Kallenbach. Adenine affects the structure and stability of telomeric sequences. *Journal of Biological Chemistry*, 267(22):15293–15300, 1992.

[141] John G Duguid, Victor A Bloomfield, James M Benevides, and GJ Thomas. DNA melting investigated by differential scanning calorimetry and raman spectroscopy. *Biophysical Journal*, 71(6):3350–3360, 1996.

[142] Arthur A Evans and Alex J Levine. High-energy deformation of filaments with internal structure and localized torque-induced melting of DNA. *Physical Review E*, 85(5):051915, 2012.

[143] Jae-Yeol Kim, Jae-Hyung Jeon, and Wokyung Sung. A breathing wormlike chain model on DNA denaturation and bubble: Effects of stacking interactions. *J. Chem. Phys.*, 128: 055101, 2008.

[144] K. J. Breslauer, R Frank, H Blocker, and L. A. Marky. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA*, 83(11):3746–3750, 1986.

[145] John SantaLucia, Jr., H T Allawi, and P A Seneviratne. Improved nearest-neighbour parameters for predicting DNA duplex stability. *Biochem.*, 35:3555–3562, 1996.

[146] Célia Fonseca Guerra, F Matthias Bickelhaupt, Jaap G Snijders, and Evert Jan Baerends. Hydrogen bonding in DNA base pairs: reconciliation of theory and experiment. *J. Am. Chem. Soc.*, 122(17):4117–4128, 2000.

[147] Angelo Di Garbo. Anharmonic longitudinal motion of bases and dynamics of nonlinear excitation in DNA. *Biophysical chemistry*, 208:76–83, 2016.

[148] R. M. Wartell and A. S. Benight. Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.*, 126(2):67–107, 1985.

[149] Jean Louis Leroy, Michel Kochoyan, Tam Huynh-Dinh, and Maurice Guéron. Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *Journal of molecular biology*, 200(2):223–238, 1988.

[150] D. Poland and H. A. Scheraga. Occurence of a phase transition in nucleic acids models. *J. Chem. Phys.*, 45(5):1464–1469, 1966.

[151] M. Peyrard and A. R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, 62(23):2755–2757, 1989.

[152] T. Dauxois and M. Peyrard. Entropy-driven transition in a one-dimensional system. *Phys. Rev. E*, 51(5):4027–4040, 1995.

[153] T. Dauxois, M. Peyrard, and A. R. Bishop. Entropy-driven DNA denaturation. *Phys. Rev. E*, 47(1):R44–R47, 1993.

[154] S Behnia, A Akhshani, M Panahi, A Mobaraki, and M Ghaderian. Multifractal analysis of thermal denaturation based on the Peyrard-Bishop-Dauxois model. *Phys. Rev. E*, 84 (3):031918, 2011.

[155] S. Behnia, A. Akhshani, M. Panahi, A. Mobaraki, and M. Ghaderian. Multifractal properties of denaturation process based on Peyrard–Bishop model. *Phys. Lett. A*, 376:2538–2547, 2012.

[156] K. A. Velizhanin, C. C. Chien, Y. Dubi, and M. Zwolak. Driving denaturation: Nanoscale thermal transport as a probe of DNA melting. *Phys. Rev. E*, 83(5):050906, 2011.

[157] Gerald Weber. Sharp DNA denaturation due to solvent interaction. *Europhys. Lett.*, 73 (5):806–811, 2006. doi:10.1209/epl/i2005-10466-6.

[158] M Peyrard, T Dauxois, and H Hoyet. Dynamics of DNA melting. *Nanobiology*, 1(3):313, 1992.

[159] Philip M Morse. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Physical Review*, 34(1):57, 1929.

[160] Y Gao, KV Devi-Prasad, and EW Prohofsky. A self-consistent microscopic theory of hydrogen bond melting with application to poly (d g)·poly (d c). *The Journal of Chemical Physics*, 80(12):6291–6298, 1984.

[161] Y Gao and EW Prohofsky. A modified self-consistent phonon theory of hydrogen bond melting. *The Journal of chemical physics*, 80(5):2242–2243, 1984.

[162] Rodolfo Vieira Maximiano. Estudos das propriedades de inosina em DNA através do modelo peyrard-bishop e análise dos parâmetros termodinâmicos utilizados na predição de estruturas secundárias de RNA. *Universidade Federal de Minas Gerais*, 2018.

[163] D. J. Scalapino, M. Sears, and R. A. Ferrell. Statistical mechanics of the one-dimensional Ginzburg-Landau fields. *Phys. Rev. B*, 6(9):3409–3416, 1972.

[164] JA Krumhansl and JR Schrieffer. Dynamics and statistical mechanics of a one-dimensional model hamiltonian for structural phase transitions. *Physical Review B*, 11(9):3535, 1975.

[165] Wolfram Saenger. Defining terms for the nucleic acids. In *Principles of nucleic acid structure*, pages 9–28. Springer, 1984.

[166] Douglas C. Poland and Harold A. Scheraga. Statistical mechanics of noncovalent bonds in polyamino acids. ix. the two-state theory of protein denaturation. *Biopolymers*, 3(4): 401–419, 1965. doi:10.1002/bip.1965.360030405. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.1965.360030405`.

[167] M Ya Azbel. Generalized one-dimensional ising model for polymer thermodynamics. *The Journal of Chemical Physics*, 62(9):3635–3641, 1975.

[168] J Day and CR Willis. Determination of the single chain density of states for membrane systems. *Journal of Theoretical Biology*, 88(4):693–718, 1981.

[169] Maria Barbi, Simona Cocco, and Michel Peyrard. Helicoidal model for DNA opening. *Phys. Lett. A*, 253:358–369, 1999.

[170] Marc Joyeux and Sahin Buyukdagli. Dynamical model based on finite stacking enthalpies for homogeneous and inhomogeneous DNA thermal denaturation. *Phys. Rev. E*, 72:051902, 2005.

[171] S. Buyukdagli and M. Joyeux. Theoretical investigation of finite size effects at DNA melting. *Phys. Rev. E*, 76(2):021917, 2007.

[172] Marc Joyeux and Ana-Maria Florescu. Dynamical versus statistical mesoscopic models for DNA denaturation. *Journal of Physics: Condensed Matter*, 21(3):034101, 2008.

[173] Michel Peyrard, Santiago Cuesta-López, and Guillaume James. Modelling DNA at the mesoscale: a challenge for nonlinear science? *Nonlinearity*, 21(6):T91, 2008.

[174] Slobodan Zdravković. Helicoidal peyrard–bishop model of DNA dynamics. *Journal of Nonlinear Mathematical Physics*, 18(supp02):463–484, 2011.

[175] Yong-Li Zhang, Wei-Mou Zheng, Ji-Xing Liu, and Y. Z. Chen. Theory of DNA melting based on the Peyrard-Bishop model. *Phys. Rev. E*, 56(6):7100–7115, 1997.

[176] Gerald Weber, Niall Haslam, Jonathan W. Essex, and Cameron Neylon. Thermal equivalence of DNA duplexes for probe design. *J. Phys.: Condens. Matter*, 21:034106, 2009. doi:10.1088/0953-8984/21/3/034106.

[177] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.

[178] Luciana M. Oliveira, Adam S. Long, Tom Brown, Keith R. Fox, and Gerald Weber. Melting temperature measurement and mesoscopic evaluation of single, double and triple DNA mismatches. *Chem. Sci.*, 11:8273–8287, 2020. doi:10.1039/d0sc01700k. URL `https://pubs.rsc.org/en/content/articlelanding/2020/SC/D0SC01700K`.

[179] Pâmella Miranda, Luciana M. Oliveira, and Gerald Weber. Mesoscopic modelling of Cy3 and Cy5 dyes attached to DNA duplexes. *Biophys. Chem.*, 230C:62–67, 2017. ISSN 0301-4622. doi:10.1016/j.bpc.2017.08.007. URL `http://www.sciencedirect.com/science/article/pii/S0301462217302831`.

[180] Gerald Weber. Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures. *Nucleic Acids Res.*, 41:e30, 2013. doi:10.1093/nar/gks964. URL `http://nar.oxfordjournals.org/content/41/1/e30`.

[181] Erik de Oliveira Martins and Gerald Weber. An asymmetric mesoscopic model for single bulges in RNA. *J. Chem. Phys.*, 147:155102, 2017. doi:10.1063/1.5006948.

[182] Tauanne D. Amarante and Gerald Weber. Evaluating hydrogen bonds and base stackings of single, tandem and terminal GU in RNA mismatches with a mesoscopic model. *J. Chem. Inf. Model.*, 56(1):101–109, 2016. doi:10.1021/acs.jcim.5b00571. URL `http://dx.doi.org/10.1021/acs.jcim.5b00571`.

[183] Ivana Domljanovic, Maria Taskova, Pâmella Miranda, Gerald Weber, and Kira Astakhova. Nucleic acid probes – optical and theoretical study reveals new details on strand recognition. *Communications Chemistry*, 3:111, 2020. doi:10.1038/s42004-020-00362-5. URL `https://www.nature.com/articles/s42004-020-00362-5`.

[184] Izabela Ferreira, Elizabeth A. Jolley, Brent M. Znosko, and Gerald Weber. Replacing salt correction factors with optimized RNA nearest-neighbour enthalpy and entropy parameters. *Chem. Phys.*, 521:69–76, may 2019. doi:10.1016/j.chemphys.2019.01.016. URL `https://www.sciencedirect.com/science/article/abs/pii/S0301010418311200`.

[185] Kazutaka Yamada, Takeshi Terahara, Shinya Kurata, Toyokazu Yokomaku, Satoshi Tsuneda, and Shigeaki Harayama. Retrieval of entire genes from environmental DNA by inverse PCR with pre-amplification of target genes using primers containing locked nucleic acids. *Environmental Microbiology*, 10(4):978–987, 2008. doi:10.1111/j.1462-2920.2007.01518.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2007.01518.x`.

[186] Paloma H Giangrande, Francis Miller, and Kevin Urak. Nucleic acid aptamers to treat histone-induced disease states, May 17 2018. US Patent App. 15/538,106.

[187] Ken Komiya, Makoto Komori, and Toru Yoshimura. Sequence conversion and signal amplifier DNA having locked nucleic acids and detection methods using same, April 14 2016. US Patent App. 14/882,124.

Appendices

A Supplementary materials

A.1 Salt dependent mesoscopic model for RNA at multiple strand concentrations

Supplementary information: Salt dependent mesoscopic model for RNA at multiple strand concentrations

Izabela Ferreira[a,b], Tauanne D. Amarante[c], Gerald Weber[a,*]

[a]*Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil.*
[b]*Programa Interunidades de Pós-Graduação em Bioinformática, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil*
[c]*MRC Cancer Unit, University of Cambridge, Hutchison/MRC Research Centre, Cambridge Biomedical Campus, Cambridge, UK*

Supplementary Tabs. S1— S5 show the number of internal and terminal parameters. Supplementary Tabs. S6— S8 show the summary of logarithmic grouping for coarseness f. Supplementary Tabs. S9— S12 show the number of elements per group and its respective C_t. The validation sequences are shown in Tab. S13. Figs. S1— S14 contain the average regression parameters for the rounds EU for each L_f. Figs. S15— S28 show the Morse potentials for each L_f over the range of salt concentration. Figs. S29— S42 show the stacking potentials for each L_f over the range of salt concentration.

*Corresponding author
Email addresses: izabelaferreira13@gmail.com (Izabela Ferreira), tauamarante@gmail.com (Tauanne D. Amarante), gweberbh@gmail.com (Gerald Weber)

Table S1: Number of occurrences of base pairs and NN pairs contained in the dataset of Ref. [1] for [Na^+] 71mM.

internal	occurrences	terminal	occurrences	terminal	occurrences
AU	292	AU*	122		
CG	520	CG*	248		
AUpAU	18	AU*pAU	18		
AUpCG	60	AU*pCG	54		
AUpGC	106	AUpGC*	18	AU*pGC	50
AUpUA	45				
CGpAU	54	CG*pAU	40		
CGpCG	60	CG*pCG	36		
CGpGC	101	CGpGC*	48		
GCpAU	18	GC*pAU	36		
GCpCG	111	GCpCG*	70		
UApAU	54				

Table S2: Number of occurrences of base pairs and NN pairs contained in the dataset of Ref. [1] for [Na^+] 121mM.

internal	occurrences	terminal	occurrences	terminal	occurrences
AU	288	AU*	106		
CG	486	CG*	246		
AUpAU	18	AU*pAU	16		
AUpCG	56	AU*pCG	54		
AUpGC	106	AUpGC*	18	AU*pGC	36
AUpUA	45				
CGpAU	54	CG*pAU	40		
CGpCG	62	CG*pCG	36		
CGpGC	90	CGpGC*	48		
GCpAU	18	GC*pAU	36		
GCpCG	96	GCpCG*	68		
UApAU	53				

Table S3: Number of occurrences of base pairs and NN pairs contained in the dataset of Ref. [1] for [Na^+] 221mM.

internal	occurrences	terminal	occurrences	terminal	occurrences
AU	296	AU*	110		
CG	476	CG*	240		
AUpAU	18	AU*pAU	18		
AUpCG	56	AU*pCG	56		
AUpGC	108	AUpGC*	24	AU*pGC	36
AUpUA	45				
CGpAU	60	CG*pAU	38		
CGpCG	62	CG*pCG	36		
CGpGC	83	CGpGC*	44		
GCpAU	18	GC*pAU	36		
GCpCG	93	GCpCG*	62		
UApAU	54				

Table S4: Number of occurrences of base pairs and NN pairs contained in the dataset of Ref. [1] for [Na^+] 621mM.

internal	occurrences	terminal	occurrences	terminal	occurrences
AU	302	AU*	110		
CG	492	CG*	252		
AUpAU	18	AU*pAU	16		
AUpCG	56	AU*pCG	58		
AUpGC	108	AUpGC*	30	AU*pGC	36
AUpUA	45				
CGpAU	66	CG*pAU	42		
CGpCG	64	CG*pCG	38		
CGpGC	86	CGpGC*	46		
GCpAU	18	GC*pAU	34		
GCpCG	98	GCpCG*	62		
UApAU	54				

Table S5: Number of occurrences of base pairs and NN pairs contained in the dataset of Ref. [2, 3] for [Na^+] 1021mM.

internal	occurrences	terminal	occurrences	terminal	occurrences
AU	244	AU*	87		
CG	287	CG*	135		
AUpAU	37	AUpAU*	10	AU*pAU	8
AUpCG	41	AUpCG*	17	AU*pCG	9
AUpGC	55	AUpGC*	14	AU*pGC	14
AUpUA	30	AUpUA*	10		
CGpAU	57	CGpAU*	13	CG*pAU	15
CGpCG	33	CGpCG*	10	CG*pCG	10
CGpGC	42	CGpGC*	21		
GCpAU	49	GCpAU*	17	GC*pAU	18
GCpCG	49	GCpCG*	30		
UApAU	27	UApAU*	6		

Table S6: Number of groups n_f , total number of grouped elements N_f, and total number of ungrouped elements U_f for [Na^+] 121mM

f	5	4	3	2	1.5	1.4	1.3	1.2	1.1	1	0.9	0.8	0.7	0.6	0.5	0.4
n_f	22	20	15	10	8	7	7	7	6	5	5	5	5	4	3	3
N_f	170	176	174	174	176	174	175	176	176	174	176	176	176	176	176	176
U_f	6	0	2	2	0	2	1	0	0	2	0	0	0	0	0	0

Table S7: Number of groups n_f , total number of grouped elements N_f, and total number of ungrouped elements U_f for [Na^+] 221mM.

f	5.0	4.0	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.0	0.9	0.8	0.7	0.6	0.5	0.4
n_f	21	18	14	10	8	7	7	7	6	5	5	4	5	4	3	3
N_f	170	173	172	175	175	175	174	175	175	173	175	174	175	175	175	175
U_f	5	2	3	0	0	0	1	0	0	2	0	1	0	0	0	0

Table S8: Number of groups n_f , total number of grouped elements N_f, and total number of ungrouped elements U_f for [Na^+] 621mM.

f	5.0	4.0	3.0	2.0	1.5	1.4	1.3	1.2	1.1	1.0	0.9	0.8	0.7	0.6	0.5	0.4
n_f	24	20	16	11	8	8	7	7	6	5	5	5	5	4	3	3
N_f	179	181	181	181	180	181	180	181	181	181	180	181	181	181	181	181
U_f	2	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0

Table S9: Number of elements per group for each L_f for [Na^+] 71mM.

L_f Groups																																			
L_5	C_t	L_4	C_t	L_3	C_t	L_2	C_t	$L_{1.5}$	C_t	$L_{1.4}$	C_t	$L_{1.3}$	C_t	$L_{1.2}$	C_t	$L_{1.1}$	C_t	L_1	C_t	$L_{0.9}$	C_t	$L_{0.8}$	C_t	$L_{0.7}$	C_t	$L_{0.6}$	C_t	$L_{0.5}$	C_t	$L_{0.4}$	C_t				
1.80	3	2.00	12	2.00	16	2.00	22	2.00	24	2.14	32	1.54	3	1.67	9	1.82	24	2.00	32	2.22	49	2.50	66	1.43	16	1.67	32	2.00	55	2.50	92				
2.00	7	2.25	11	2.33	16	2.50	23	2.67	31	2.86	39	2.31	40	2.50	43	2.73	45	3.00	53	3.33	49	3.75	46	2.86	69	3.33	75	4.00	93	5.00	84				
2.20	14	2.50	15	2.67	20	3.00	29	3.33	33	3.57	30	3.08	41	3.33	40	3.64	36	4.00	37	4.44	50	5.00	53	4.29	63	5.00	63	6.00	36	7.50	9				
2.40	8	2.75	12	3.00	16	3.50	18	4.00	23	4.29	26	3.85	25	4.17	32	4.55	43	5.00	43	5.56	28	6.25	17	5.71	31	6.67	14								
2.60	12	3.00	16	3.33	17	4.00	17	4.67	37	5.00	33	4.62	39	5.00	38	5.45	22	6.00	17	6.67	8	7.14	5												
2.80	8	3.25	13	3.67	11	4.50	26	5.33	19	5.71	15	5.38	21	5.83	14	6.36	14																		
3.00	13	3.50	6	4.00	11	5.00	19	6.00	11	6.43	9	6.15	14	6.67	8																				
3.20	12	3.75	11	4.33	15	5.50	15	6.67	6																										
3.40	8	4.00	7	4.67	19	6.00	7																												
3.60	5	4.25	9	5.00	10	6.50	8																												
3.80	11	4.50	13	5.33	14																														
4.00	4	4.75	19	5.67	5																														
4.20	5	5.00	5	6.00	6																														
4.40	12	5.25	10	6.33	6																														
4.60	9	5.50	6																																
4.80	15	5.75	3																																
5.00	4	6.00	6																																
5.20	8	6.50	6																																
5.40	7																																		
5.80	3																																		
6.00	6																																		
6.40	5																																		

Table S10: Number of elements per group for each L_f for [Na^+] 121mM.

L_f Groups																																			
L_5	C_t	L_4	C_t	L_3	C_t	L_2	C_t	$L_{1.5}$	C_t	$L_{1.4}$	C_t	$L_{1.3}$	C_t	$L_{1.2}$	C_t	$L_{1.1}$	C_t	L_1	C_t	$L_{0.9}$	C_t	$L_{0.8}$	C_t	$L_{0.7}$	C_t	$L_{0.6}$	C_t	$L_{0.5}$	C_t	$L_{0.4}$	C_t				
1.80	5	1.75	4	1.67	3	2.00	19	2.00	23	2.14	28	1.54	7	1.67	13	1.82	21	2.00	30	2.22	43	1.25	4	1.43	17	1.67	30	2.00	51	2.50	82				
2.00	9	2.00	12	2.00	15	2.50	20	2.67	28	2.86	31	2.31	29	2.50	36	2.73	37	3.00	46	3.33	48	2.50	54	2.86	61	3.33	69	4.00	88	5.00	86				
2.20	9	2.25	9	2.33	12	3.00	24	3.33	28	3.57	32	3.08	39	3.33	33	3.64	37	4.00	38	4.44	48	3.75	49	4.29	61	5.00	62	6.00	37	7.50	8				
2.40	7	2.50	9	2.67	17	3.50	17	4.00	24	4.29	26	3.85	26	4.17	36	4.55	44	5.00	40	5.56	28	5.00	49	5.71	33	6.67	15								
2.60	6	2.75	14	3.00	11	4.00	20	4.67	36	5.00	28	4.62	38	5.00	34	5.45	24	6.00	20	6.67	9	6.25	20	7.14	4										
2.80	13	3.00	10	3.33	18	4.50	22	5.33	17	5.71	20	5.38	19	5.83	16	6.36	13																		
3.00	8	3.25	15	3.67	11	5.00	21	6.00	14	6.43	9	6.15	17	6.67	8																				
3.20	12	3.50	6	4.00	12	5.50	12	6.67	6																										
3.40	7	3.75	11	4.33	15	6.00	11																												
3.60	4	4.00	8	4.67	15	6.50	8																												
3.80	12	4.25	9	5.00	13																														
4.00	4	4.50	12	5.33	12																														
4.20	6	4.75	16	5.67	7																														
4.40	12	5.00	7	6.00	6																														
4.60	8	5.25	7	6.33	7																														
4.80	13	5.50	7																																
5.00	6	5.75	5																																
5.20	4	6.00	6																																
5.40	9	6.25	4																																
5.80	7	6.50	5																																
6.00	4																																		
6.40	5																																		

Table S11: Number of elements per group for each L_f for [Na^+] 221mM.

L_f Groups																															
L_5	C_t	L_4	C_t	L_3	C_t	L_2	C_t	$L_{1.5}$	C_t	$L_{1.4}$	C_t	$L_{1.3}$	C_t	$L_{1.2}$	C_t	$L_{1.1}$	C_t	L_1	C_t	$L_{0.9}$	C_t	$L_{0.8}$	C_t	$L_{0.7}$	C_t	$L_{0.6}$	C_t	$L_{0.5}$	C_t	$L_{0.4}$	C_t
1.80	3	2.00	13	2.00	15	2.00	23	2.00	28	2.14	31	1.54	4	1.67	12	1.82	23	2.00	31	2.22	46	2.50	60	1.43	15	1.67	31	2.00	54	2.50	89
2.00	9	2.25	14	2.33	15	2.50	20	2.67	26	2.86	36	2.31	36	2.50	40	2.73	44	3.00	53	3.33	52	3.75	50	2.86	69	3.33	73	4.00	91	5.00	81
2.20	14	2.50	11	2.67	17	3.00	28	3.33	32	3.57	33	3.08	44	3.33	37	3.64	35	4.00	36	4.44	47	5.00	50	4.29	61	5.00	59	6.00	30	7.50	5
2.40	5	2.75	13	3.00	19	3.50	18	4.00	25	4.29	25	3.85	24	4.17	33	4.55	43	5.00	40	5.56	25	6.25	14	5.71	26	6.67	12				
2.60	9	3.00	9	3.33	17	4.00	19	4.67	34	5.00	30	4.62	37	5.00	35	5.45	19	6.00	13	6.67	5	7.14	4								
2.80	12	3.25	18	3.67	9	4.50	21	5.33	16	5.71	15	5.38	17	5.83	13	6.36	11														
3.00	9	3.50	6	4.00	11	5.00	19	6.00	9	6.43	5	6.15	12	6.67	5																
3.20	14	3.75	11	4.33	16	5.50	14	6.67	5																						
3.40	9	4.00	6	4.67	16	6.00	8																								
3.60	5	4.25	9	5.00	11	6.50	5																								
3.80	9	4.50	14	5.33	13																										
4.00	4	4.75	14	5.67	3																										
4.20	7	5.00	8	6.00	7																										
4.40	11	5.25	9	6.33	3																										
4.60	7	5.50	5																												
4.80	14	5.75	3																												
5.00	6	6.00	6																												
5.40	11	6.50	4																												
5.80	3																														
6.00	6																														
6.40	3																														

Table S12: Number of elements per group for each L_f for [Na^+] 621mM.

L_f Groups																															
L_5	C_t	L_4	C_t	L_3	C_t	L_2	C_t	$L_{1.5}$	C_t	$L_{1.4}$	C_t	$L_{1.3}$	C_t	$L_{1.2}$	C_t	$L_{1.1}$	C_t	L_1	C_t	$L_{0.9}$	C_t	$L_{0.8}$	C_t	$L_{0.7}$	C_t	$L_{0.6}$	C_t	$L_{0.5}$	C_t	$L_{0.4}$	C_t
1.60	3	1.75	9	1.67	6	1.50	5	2.00	29	1.43	6	1.54	10	1.67	18	1.82	28	2.00	38	2.22	51	1.25	9	1.43	22	1.67	38	2.00	63	2.50	95
1.80	7	2.00	12	2.00	17	2.00	22	2.67	33	2.14	32	2.31	38	2.50	40	2.73	45	3.00	48	3.33	51	2.50	63	2.86	70	3.33	72	4.00	90	5.00	82
2.00	9	2.25	11	2.33	15	2.50	25	3.33	29	2.86	40	3.08	37	3.33	37	3.64	36	4.00	41	4.44	50	3.75	48	4.29	61	5.00	59	6.00	28	7.50	4
2.20	9	2.50	14	2.67	16	3.00	26	4.00	27	3.57	26	3.85	27	4.17	38	4.55	44	5.00	38	5.56	21	5.00	45	5.71	24	6.67	12				
2.40	10	2.75	11	3.00	19	3.50	17	4.67	34	4.29	34	4.62	41	5.00	30	5.45	17	6.00	13	6.67	7	6.25	16	7.14	4						
2.60	10	3.00	15	3.33	13	4.00	18	5.33	13	5.00	22	5.38	16	5.83	14	6.36	11	7.00	3												
2.80	9	3.25	11	3.67	15	4.50	28	6.00	11	5.71	14	6.15	11	6.67	4																
3.00	13	3.50	9	4.00	9	5.00	16	6.67	4	6.43	7																				
3.20	11	3.75	10	4.33	17	5.50	12																								
3.40	5	4.00	8	4.67	18	6.00	8																								
3.60	7	4.25	10	5.00	10	6.50	4																								
3.80	10	4.50	16	5.33	10																										
4.00	6	4.75	12	5.67	4																										
4.20	7	5.00	6	6.00	5																										
4.40	11	5.25	7	6.33	4																										
4.60	13	5.50	4	6.67	3																										
4.80	9	5.75	5																												
5.00	5	6.00	4																												
5.20	4	6.25	3																												
5.40	7	6.50	4																												
5.80	4																														
6.00	4																														
6.20	3																														
6.60	3																														

Table S13: Validation sequences from references 4–10. Shown are the experimental temperatures T_i and corresponding predictions T_i' for T/I and UN optimizations. All temperatures are in °C. Also shown are total squared difference χ^2 and the average difference of predicted and measured temperatures $\langle \Delta T \rangle$, Eqs. (9) and (10) in the main text.

Sequence	C_t (μM)	[Na^+] (mM)	exp. T_i	T/I T_i'	UN T_i'
1 AGGAGGAGGAGGAGGAGGAGGAGG [4]	50	51	77.6	79.7669	79.821
2 AAGAAGAAGAAGAAGAAGAAGAAG [4]	50	51	53.5	55.9059	55.6176
3 AGAGAGAGAGAGAGAGAGAGAGAG [4]	50	51	69.5	66.2503	67.2657
4 AAGGAAGGAAGGAAGGAAGGAAGG [4]	50	51	70.9	69.5769	68.7957
5 AAGUGAUC [5]	8	121	32.9	32.2563	32.3436
6 CGCUGUAA [5]	8	121	36.3	37.1733	36.8021
7 CACGGCUC [5]	8	121	45.5	44.8347	43.9326
8 GCCAGUUAA [5]	8	121	40.2	40.8365	40.2593
9 CGCUGUUAC [5]	8	121	37.9	40.7518	41.7481
10 ACCUGCAGGU [6]	50	121	63.7	63.4022	62.4274
11 UUGUGCACAA [6]	50	121	52.7	50.3499	51.0857
12 UUCUGCAGAA [6]	50	121	50.6	51.3189	51.5516
13 CAUUGCAAUG [6]	50	121	43.4	44.6428	43.6885
14 GAAGAGAAGC [7]	100	121	48.85	53.0902	51.6894
15 CAAUGCAUUG [6]	50	121	43.9	44.6428	43.6885
16 ACGUGCACGU [6]	50	121	61	59.7286	59.5454
17 CGUUGCAACG [6]	50	121	55.8	55.8179	54.6327
18 GGUUGCAACC [6]	50	121	58	59.3881	59.5375
19 ACUUGCAAGU [6]	50	121	49	49.368	50.2168
20 UGUUGCAACA [6]	50	121	51.7	50.286	51.0857
21 GCUUGCAAGC [6]	50	121	57.4	60.1403	58.9575
22 GUCUGCAGAC [6]	50	121	59	57.1146	59.7828
23 CUGGUCGCAUC [8]	100	121	66.8	67.1182	64.7904
24 ACGCCACGUGA [8]	100	121	69	67.6822	68.4456
25 ACGCGACGUGA [8]	100	121	67.6	66.0614	66.9848
26 ACGCUAGGUGA [8]	100	121	64.5	65.1202	64.7812
27 ACGCUAAGUGA [8]	100	121	56.4	57.3426	57.709
28 ACGCUACGUGA [8]	100	121	62.5	62.9918	63.0451
29 ACGCAACGUGA [8]	100	121	61.2	60.0745	61.5503
30 ACGCUAUGUGA [8]	100	121	56.7	56.8261	57.2203
31 AUUGGAUACAAA [5]	8	121	43.7	40.6474	41.4142
32 UUCUUUCUUUUC [9]	100	121	46.5	43.7427	43.8339
33 UCCGCGCA [10]	100	1000	71.2	72.4918	74.1847
34 CUCGCACA [10]	100	1000	62.8	63.2097	63.7433
35 ACCUUUGG [5]	100	1000	56.3	55.7896	57.5649
36 GUCGCAGC [10]	100	1000	66.1	69.7672	71.7399
37 CGACGCAG [10]	100	1000	64.8	66.3058	66.7014
38 GAGCCGAC [10]	100	1000	68.5	70.9651	72.3809
39 UGCAAGGA [10]	100	1000	58.5	59.6788	61.6528
40 AGGCCGGA [10]	100	1000	72.9	75.3854	77.5035
41 UCACCUGA [10]	100	1000	56.9	59.1385	61.7791
42 ACCUUUGC [10]	100	1000	54.4	56.2168	58.3864
43 UUCCAACCUU [10]	100	1000	64.4	62.4806	63.0923
44 AAGGCCGGAA [10]	100	1000	78.6	73.7921	74.5616
45 AUUGGAUACAAA [5]	100	1000	55.4	53.8979	57.0607
$\langle \Delta T \rangle$				1.73 °C	1.88 °C
χ^2				270.10 °C^2	324.79 °C^2

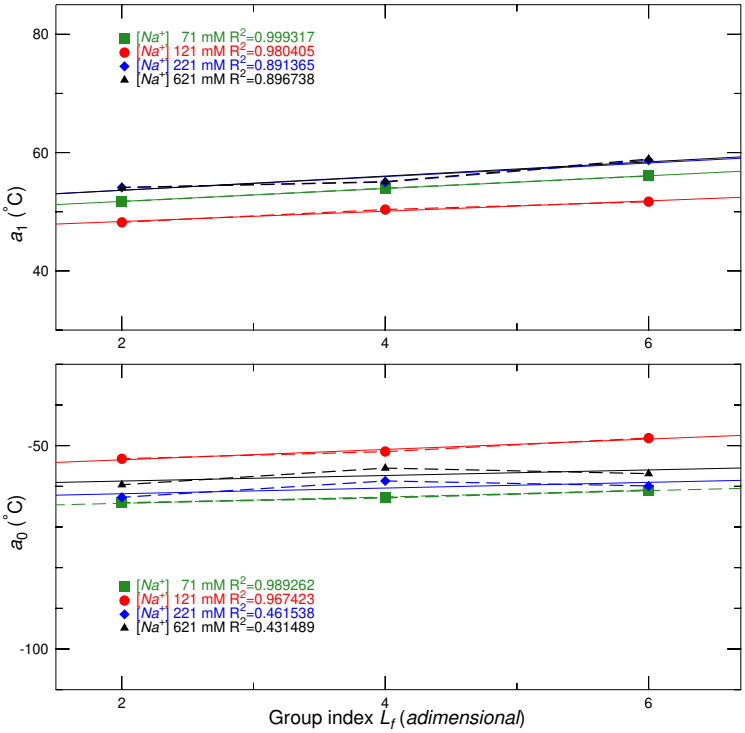


Figure S1: Regression parameters for $f = 0.5$ as function of the index L_f.

.

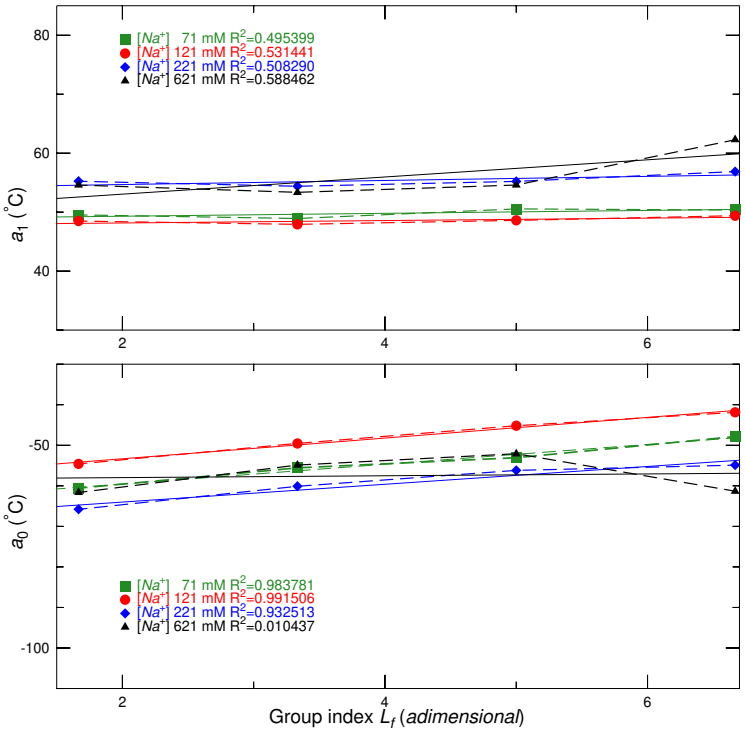


Figure S2: Regression parameters for $f = 0.6$ as function of the index L_f.

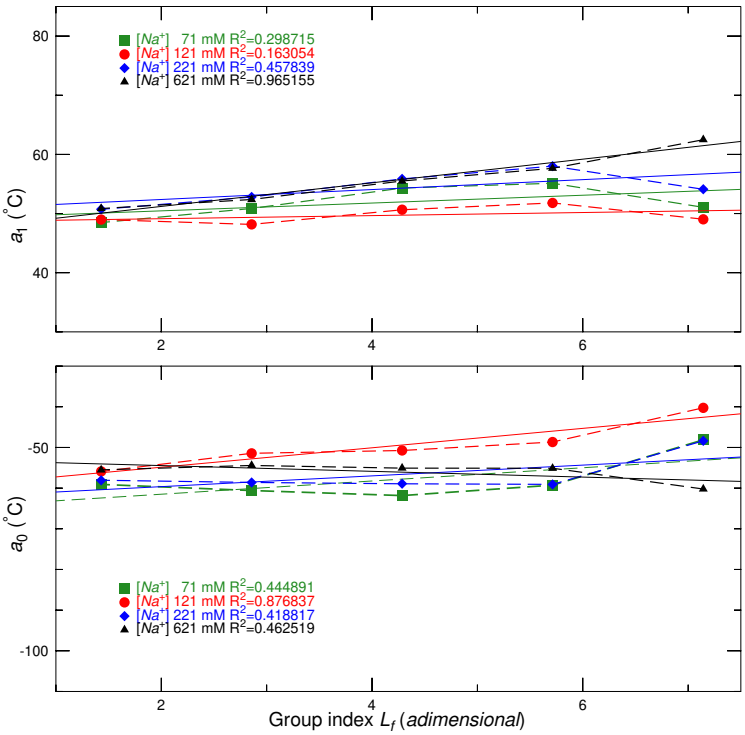


Figure S3: Regression parameters for $f = 0.7$ as function of the index L_f.

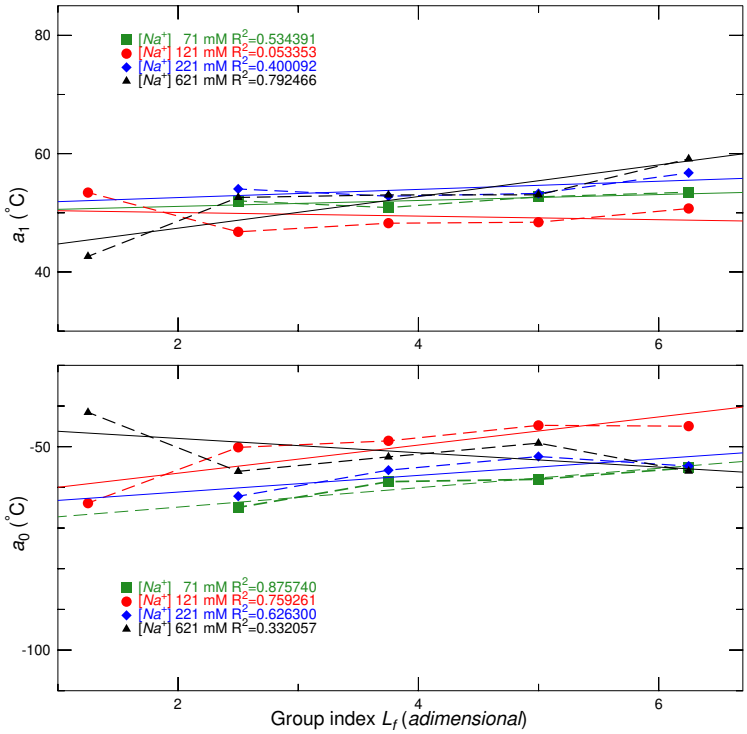


Figure S4: Regression parameters for $f = 0.8$ as function of the index L_f.

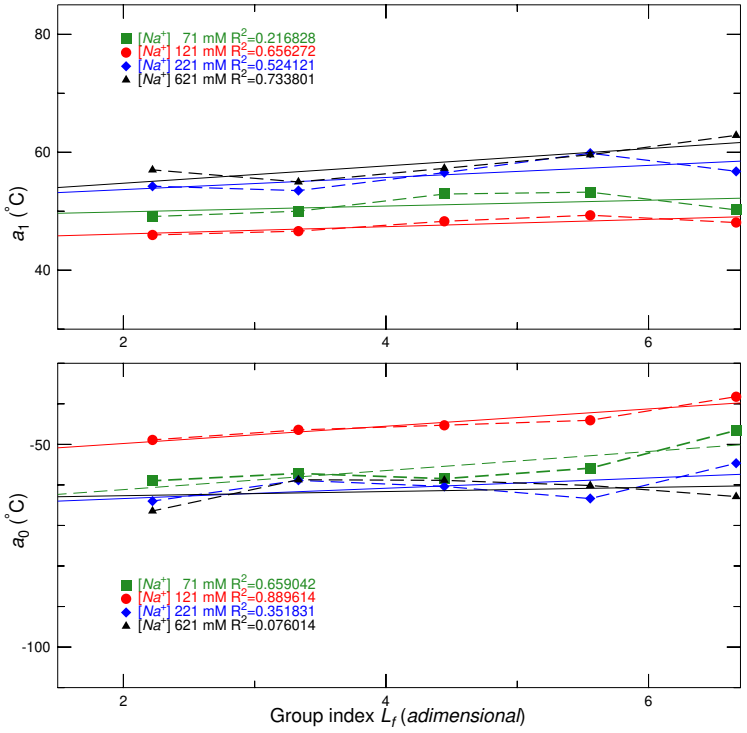


Figure S5: Regression parameters for $f = 0.9$ as function of the index L_f.

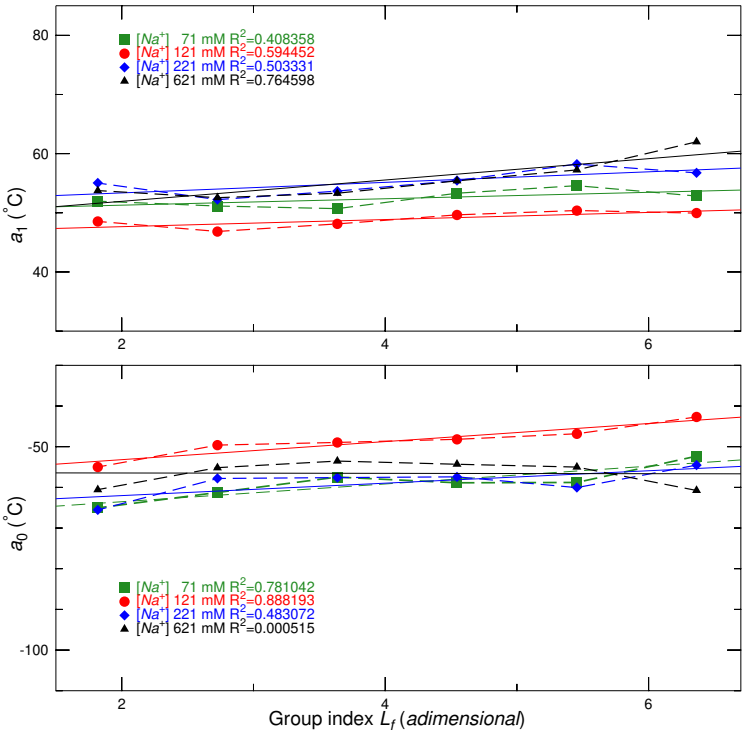


Figure S6: Regression parameters for $f = 1.1$ as function of the index L_f.

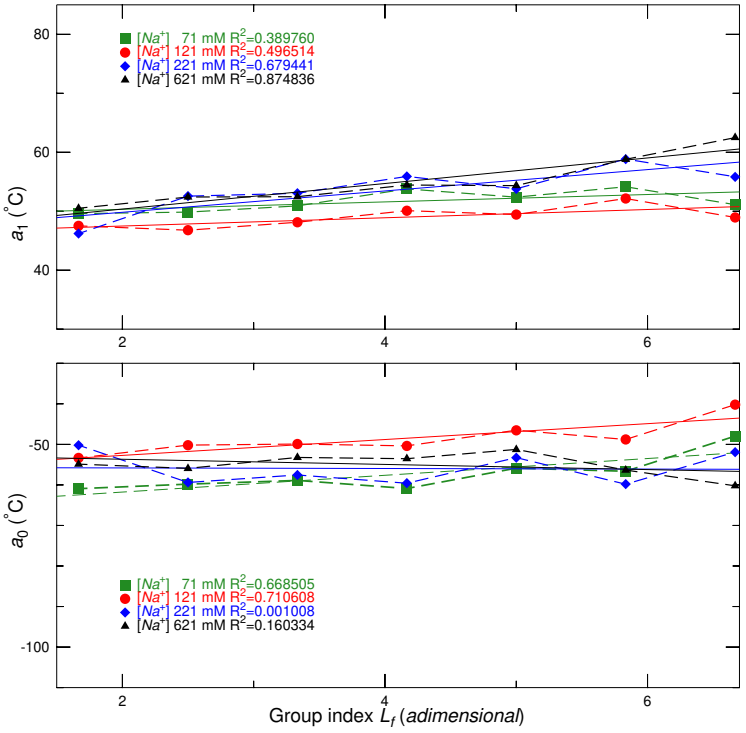


Figure S7: Regression parameters for $f = 1.2$ as function of the index L_f.

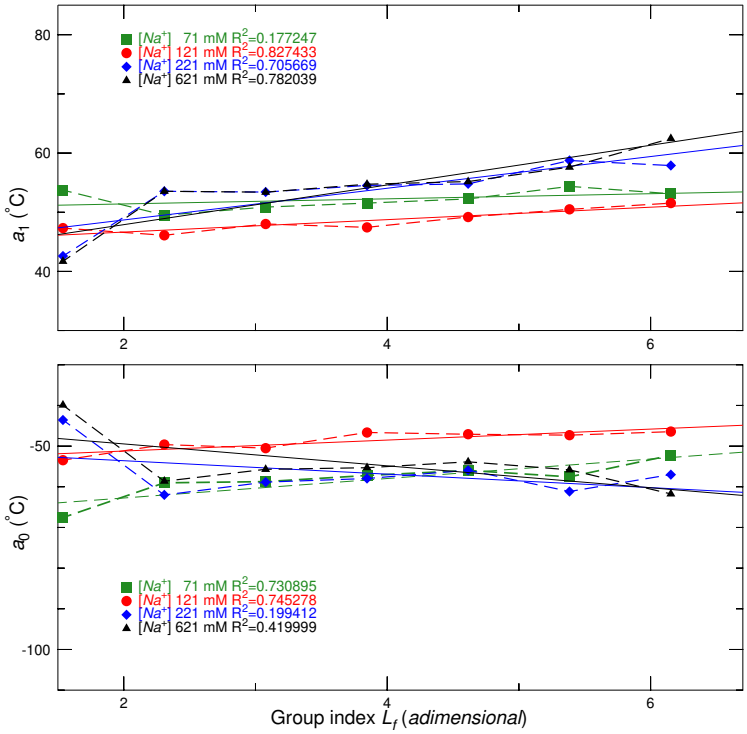


Figure S8: Regression parameters for $f = 1.3$ as function of the index L_f.

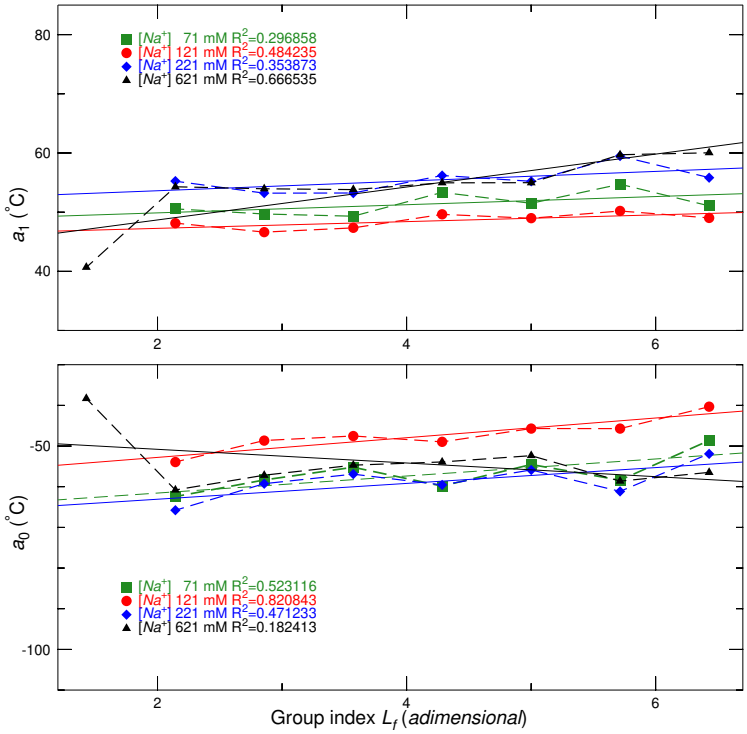


Figure S9: Regression parameters for $f = 1.4$ as function of the index L_f.

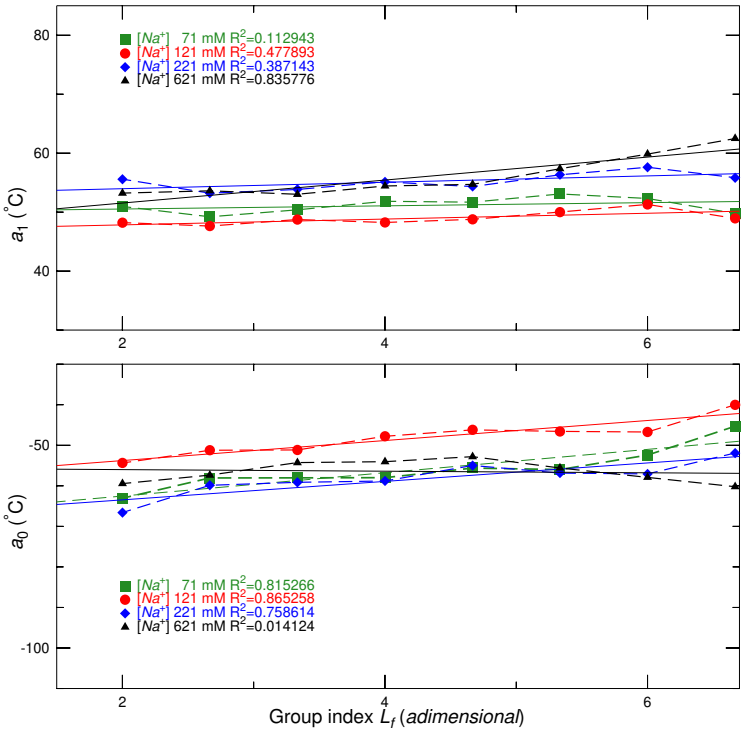


Figure S10: Regression parameters for $f = 1.5$ as function of the index L_f.

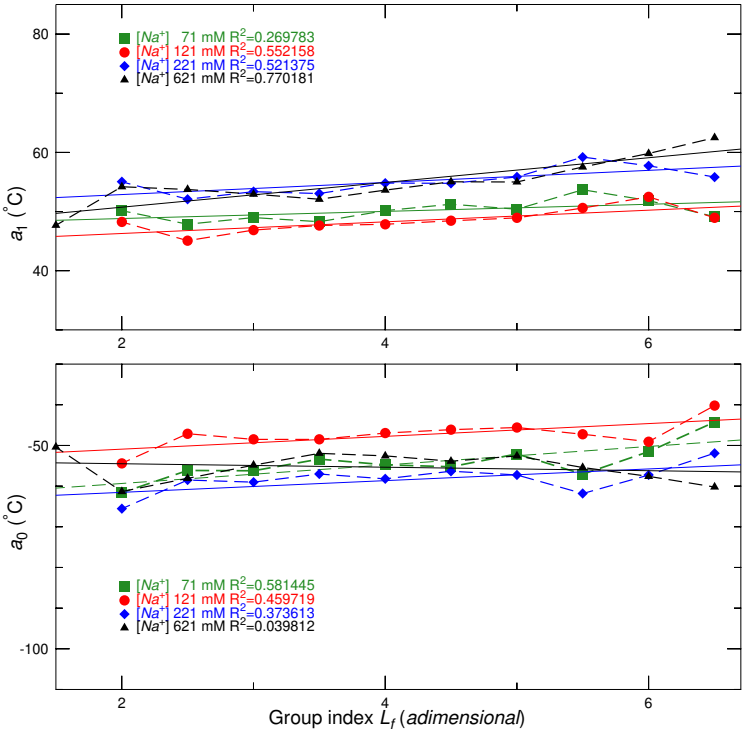


Figure S11: Regression parameters for $f = 2$ as function of the index L_f.

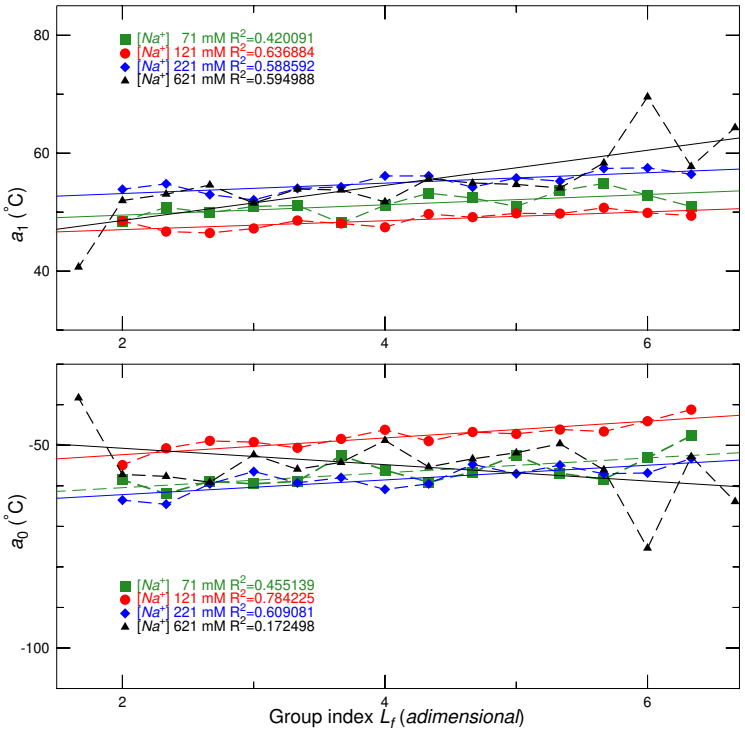


Figure S12: Regression parameters for $f = 3$ as function of the index L_f.

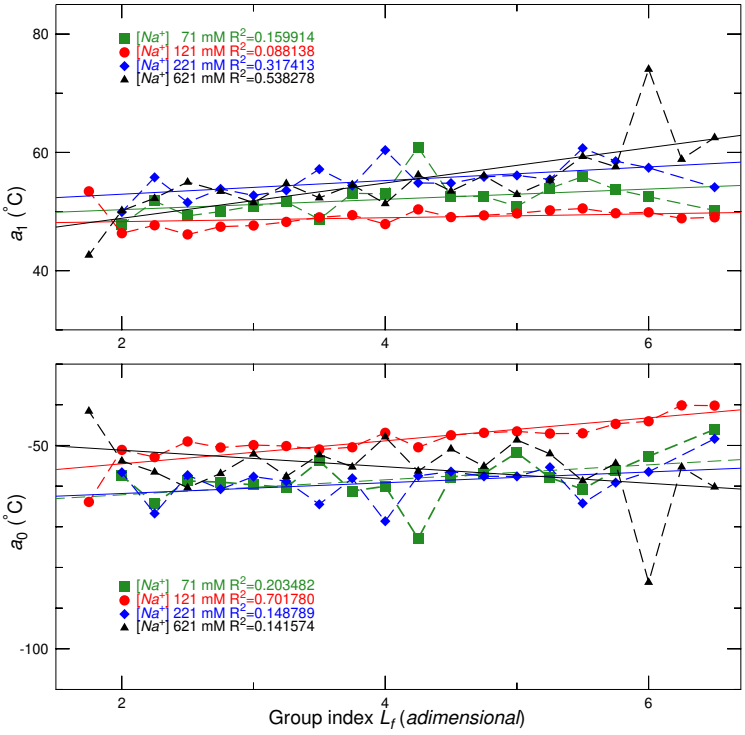


Figure S13: Regression parameters for $f = 4$ as function of the index L_f.

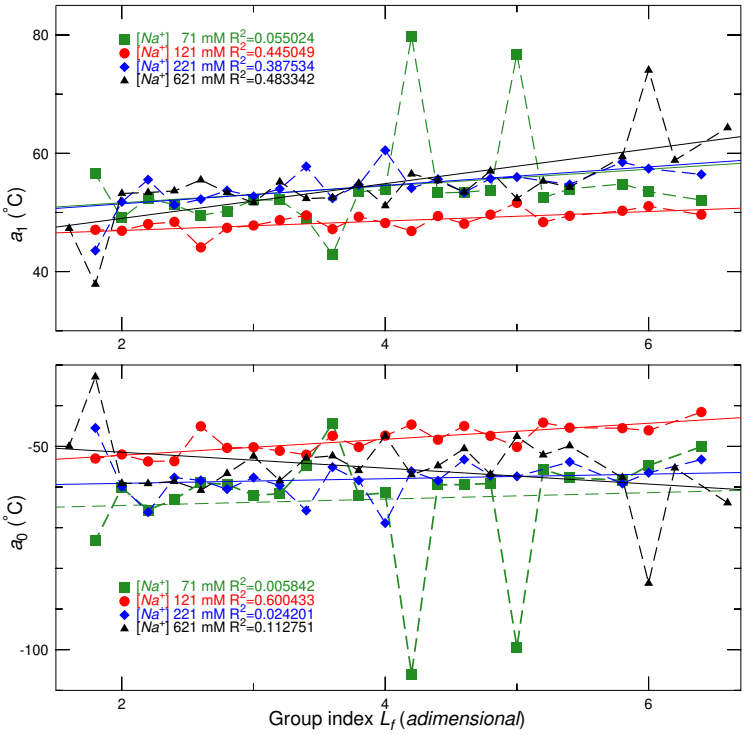


Figure S14: Regression parameters for $f = 5$ as function of the index L_f.

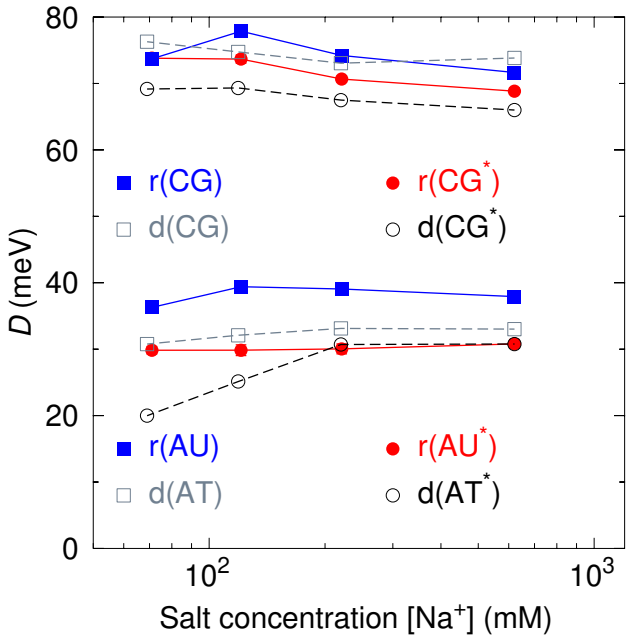


Figure S15: Average Morse potentials for $f = 0.5$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

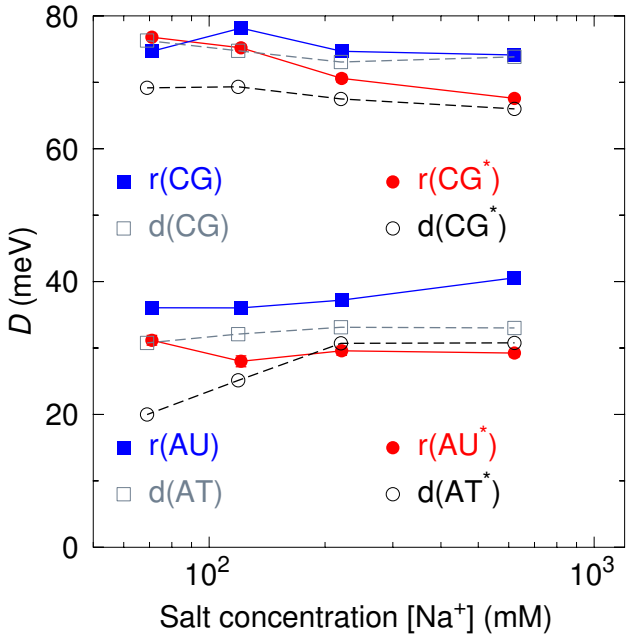


Figure S16: Average Morse potentials for $f = 0.6$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

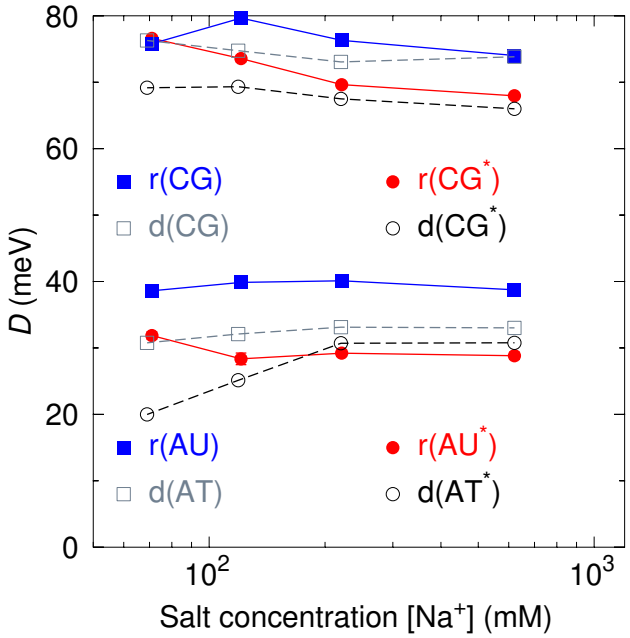


Figure S17: Average Morse potentials for $f = 0.7$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

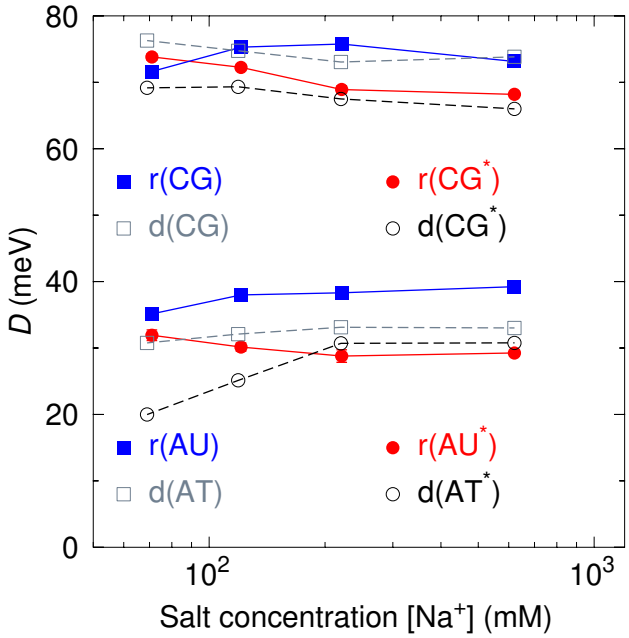


Figure S18: Average Morse potentials for $f = 0.8$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

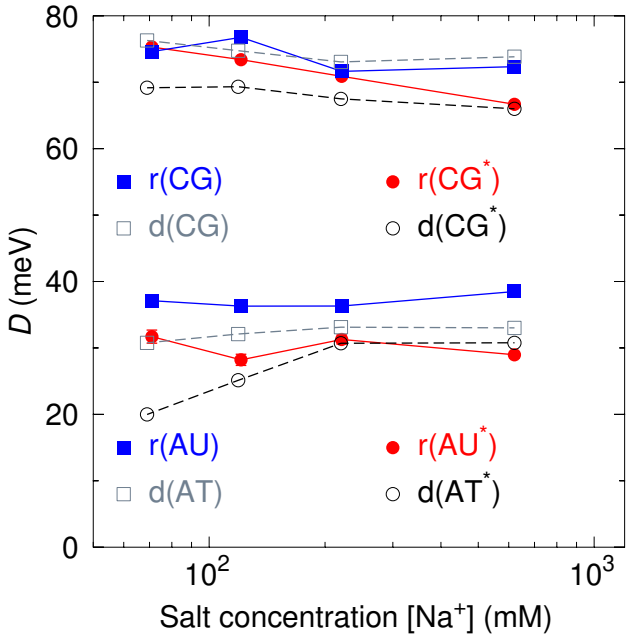


Figure S19: Average Morse potentials for $f = 0.9$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

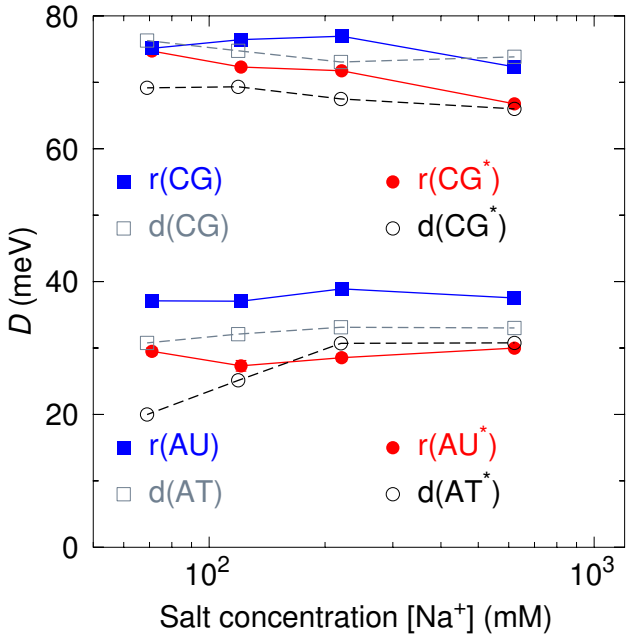


Figure S20: Average Morse potentials for $f = 0.9$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

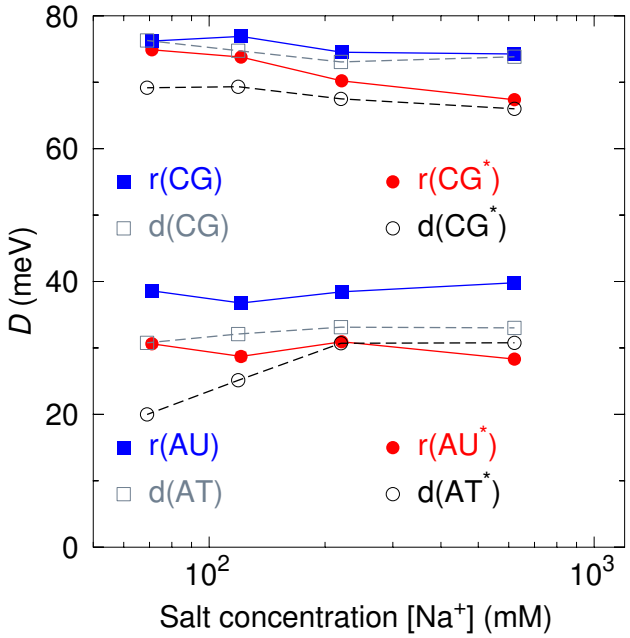


Figure S21: Average Morse potentials for $f = 1.2$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

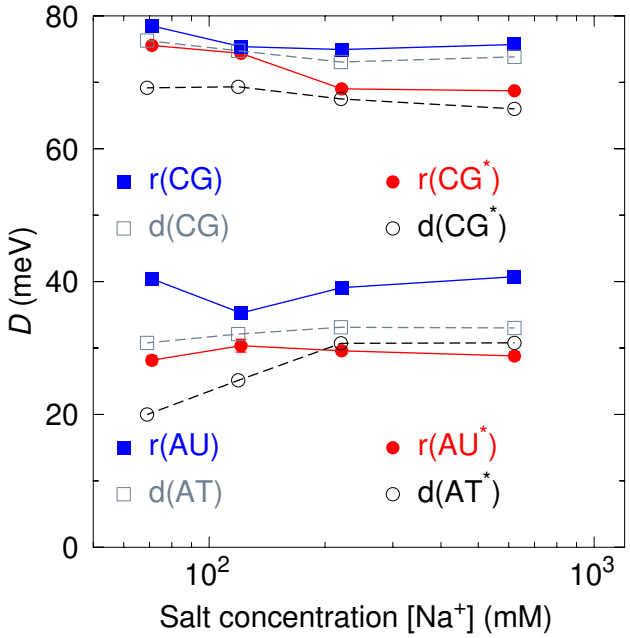


Figure S22: Average Morse potentials for $f = 1.3$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

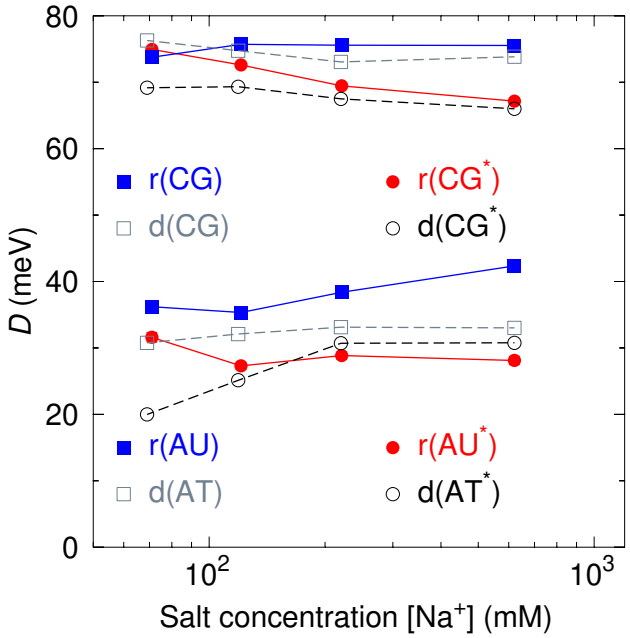


Figure S23: Average Morse potentials for $f = 1.4$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

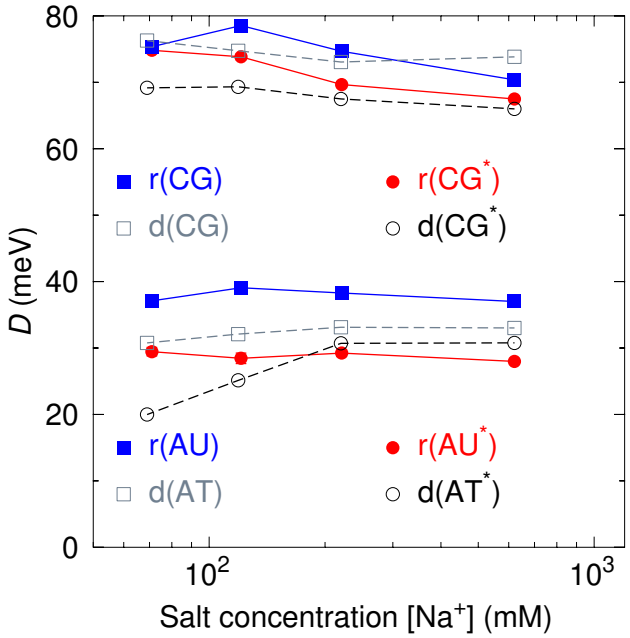


Figure S24: Average Morse potentials for $f = 1.5$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

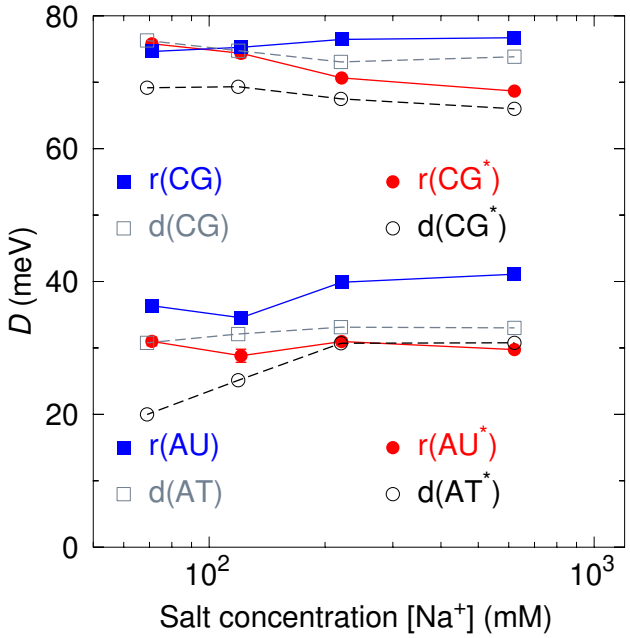


Figure S25: Average Morse potentials for $f = 2$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

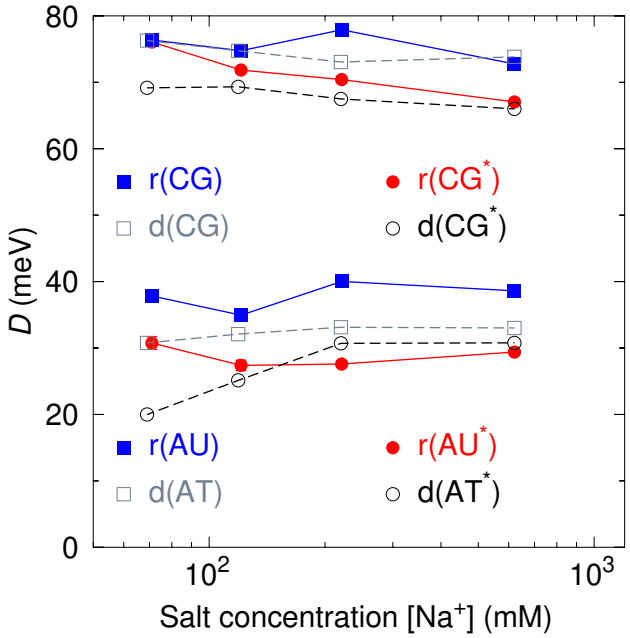


Figure S26: Average Morse potentials for $f = 3$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

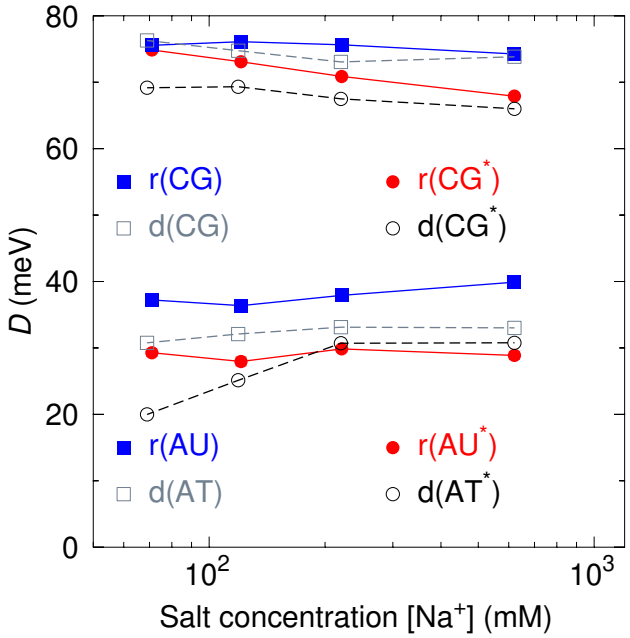


Figure S27: Average Morse potentials for $f = 4$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

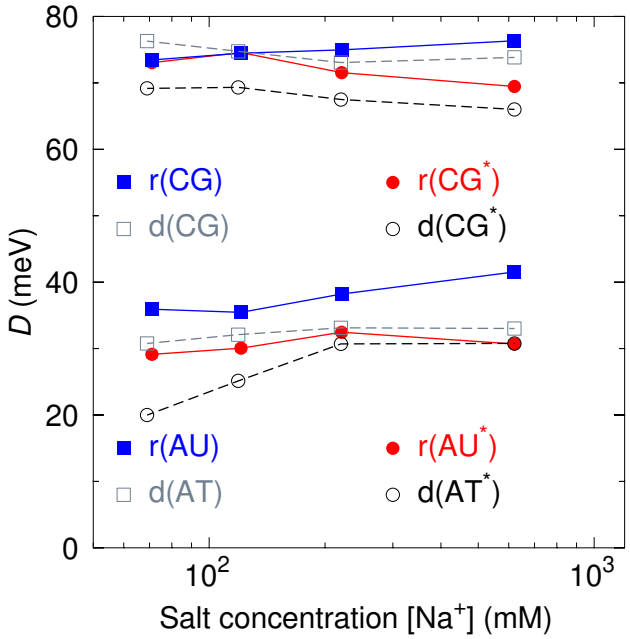


Figure S28: Average Morse potentials for $f = 5$ as function of salt concentration. For comparison, we show the analogous DNA parameters as grey boxes (internal) and black circles (terminal)[11]

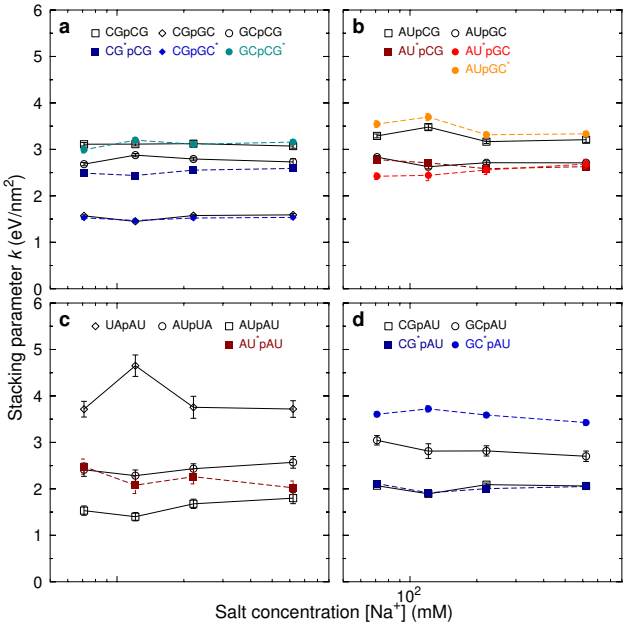


Figure S29: Average stacking potentials as function of sodium concentration for $f = 0.5$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

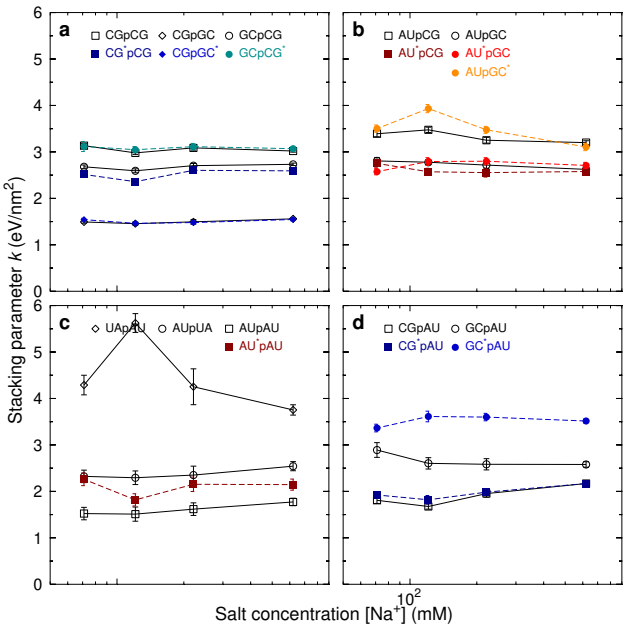


Figure S30: Average stacking potentials as function of sodium concentration for $f = 0.6$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

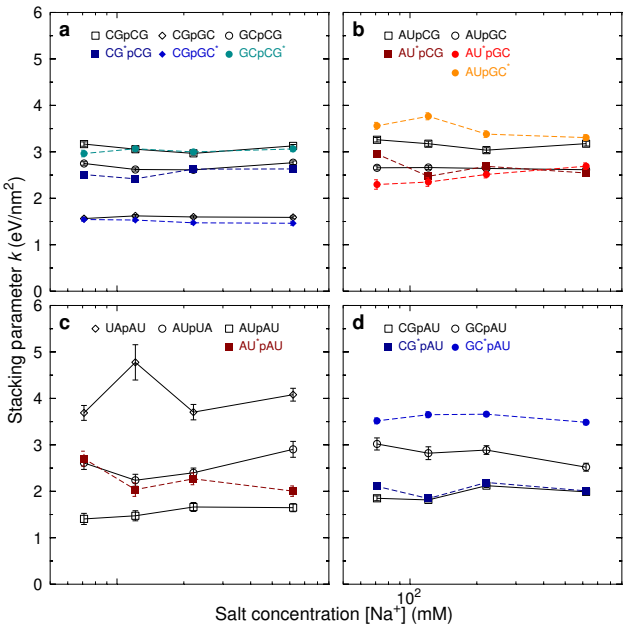


Figure S31: Average stacking potentials as function of sodium concentration for $f = 0.7$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

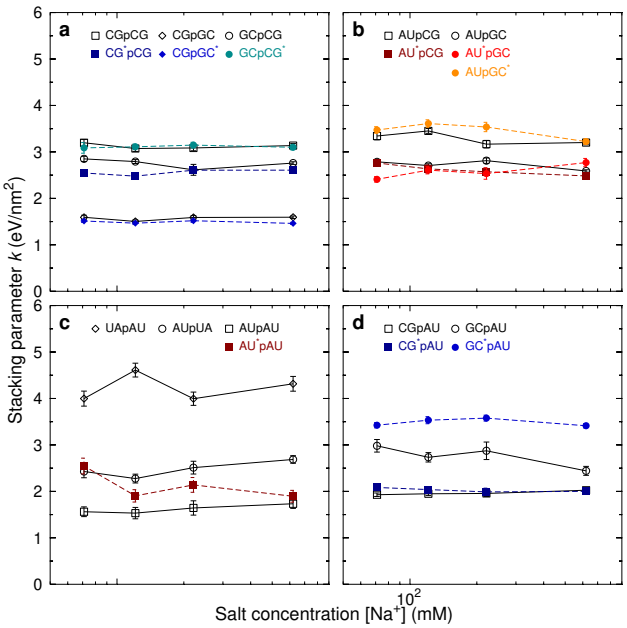


Figure S32: Average stacking potentials as function of sodium concentration for $f = 0.8$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

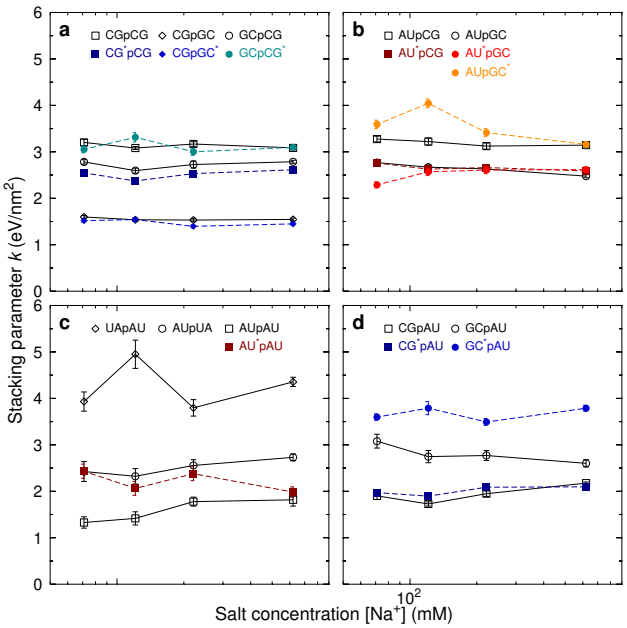


Figure S33: Average stacking potentials as function of sodium concentration for $f = 0.9$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

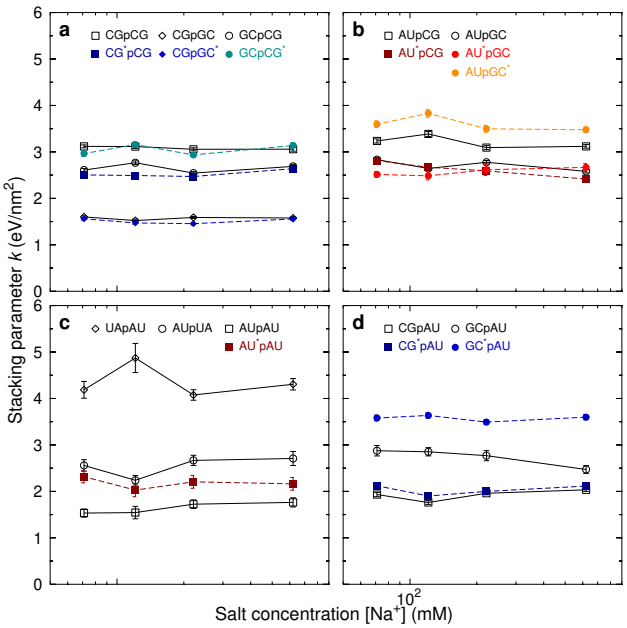


Figure S34: Average stacking potentials as function of sodium concentration for $f = 1.1$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

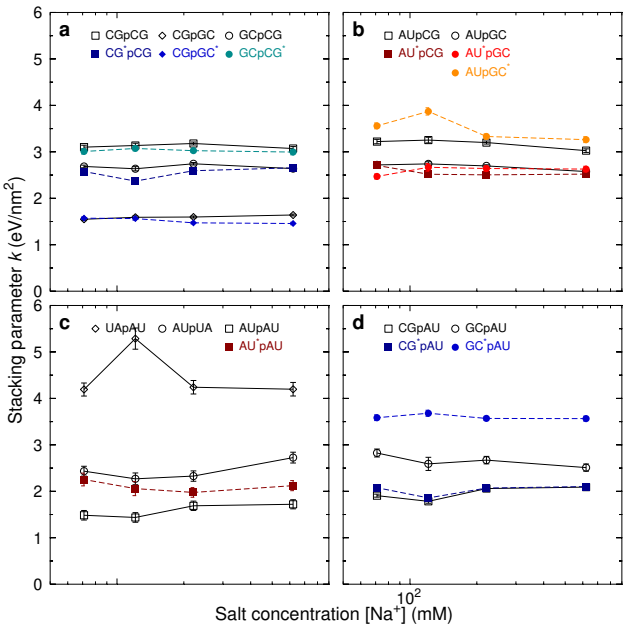


Figure S35: Average stacking potentials as function of sodium concentration for $f = 1.2$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

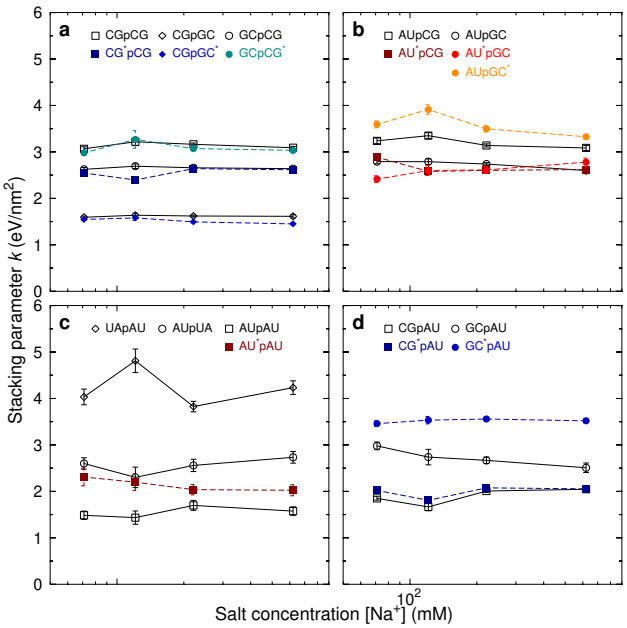


Figure S36: Average stacking potentials as function of sodium concentration for $f = 1.3$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

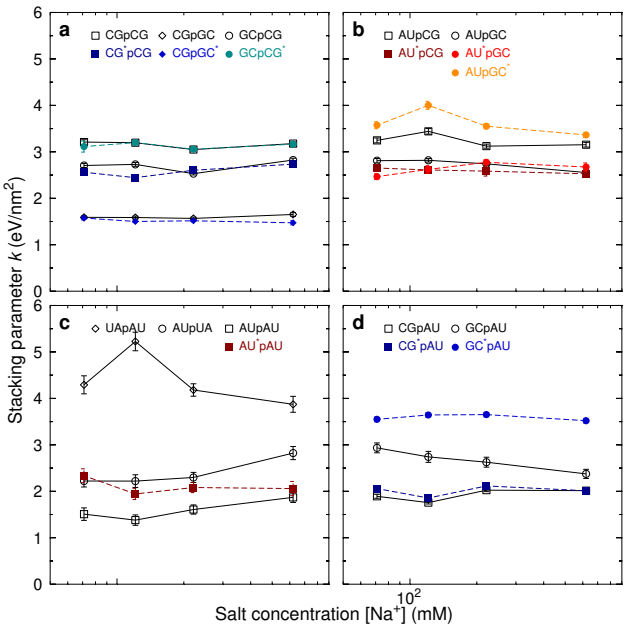


Figure S37: Average stacking potentials as function of sodium concentration for $f = 1.4$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

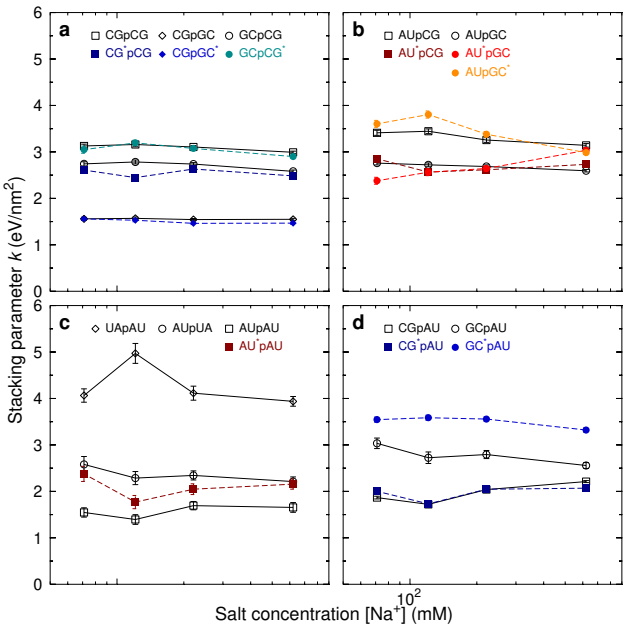


Figure S38: Average stacking potentials as function of sodium concentration for $f = 1.5$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

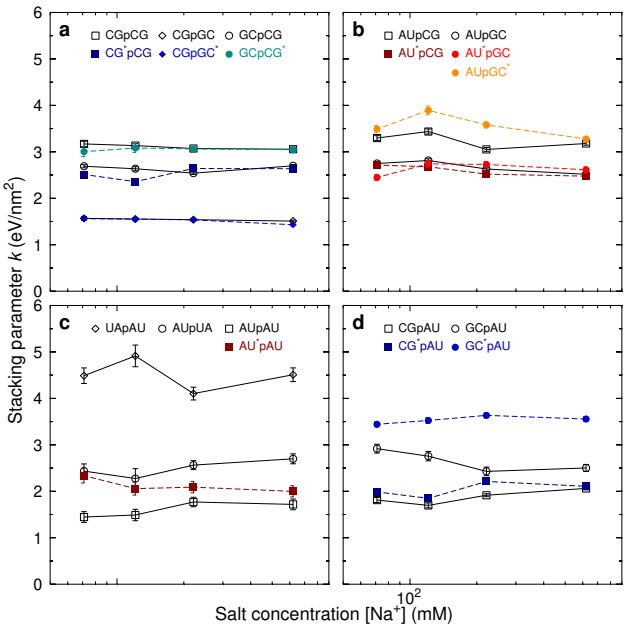


Figure S39: Average stacking potentials as function of sodium concentration for $f = 2$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

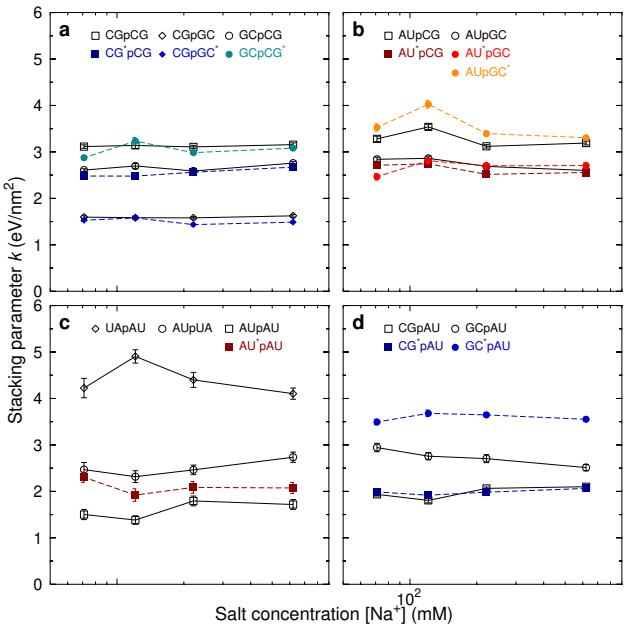


Figure S40: Average stacking potentials as function of sodium concentration for $f = 3$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

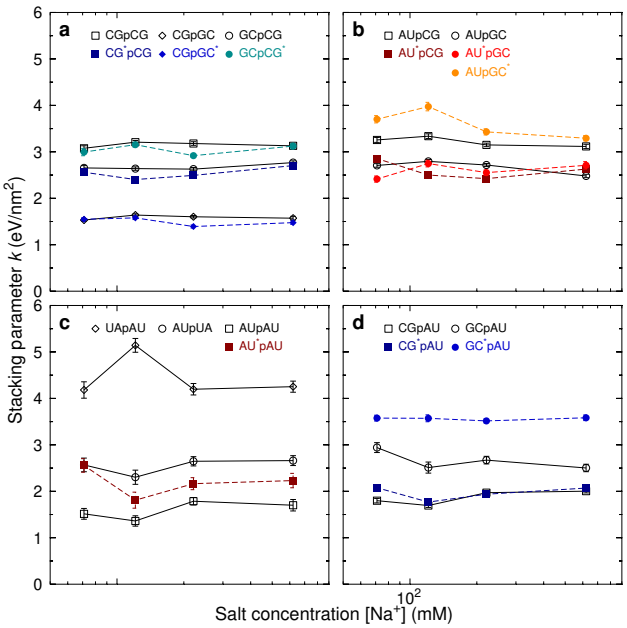


Figure S41: Average stacking potentials as function of sodium concentration for $f = 4$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

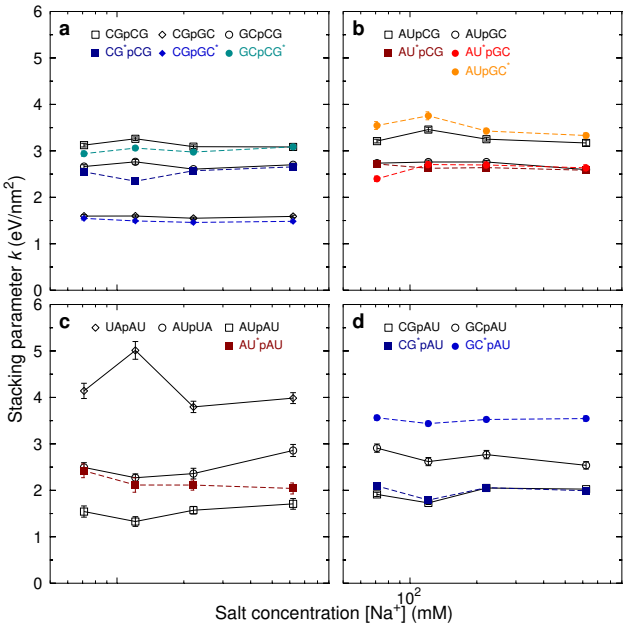


Figure S42: Average stacking potentials as function of sodium concentration for $f = 5$. Panels (a,c) show the symmetric NN and panels (b,d) the asymmetric NN. Error bars were estimated in the EU-LS minimization round. Solid lines are for internal NNs and dashed lines for NNs with terminal base pairs.

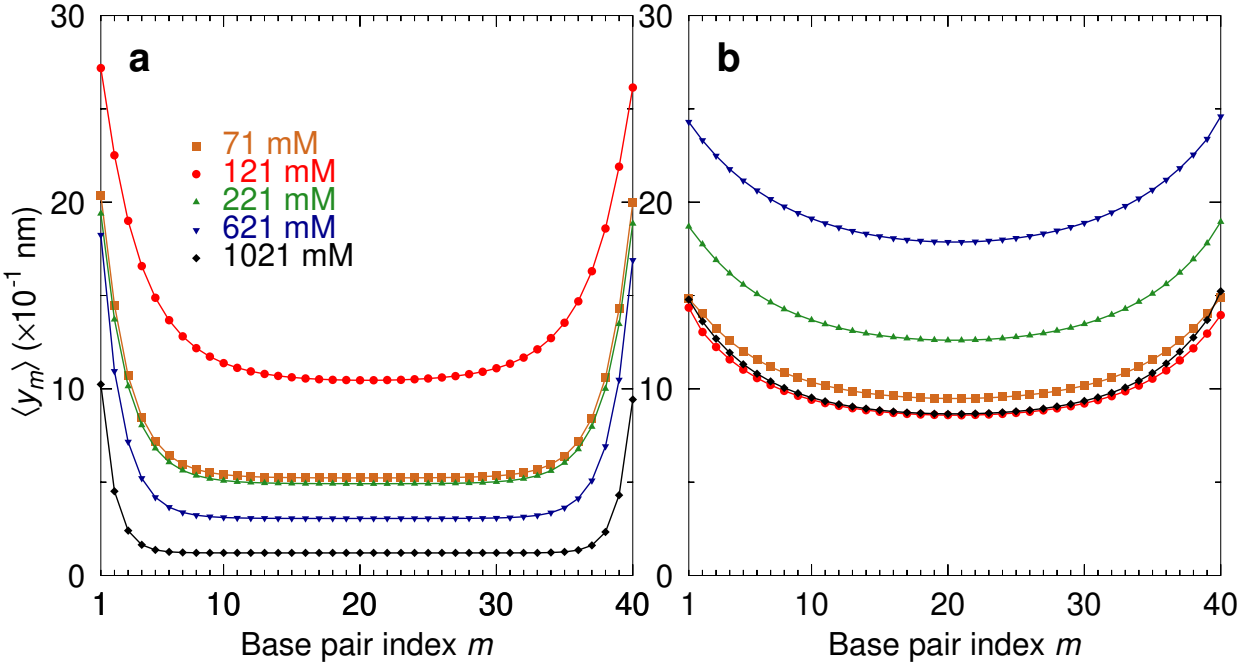


Figure S43: Average T/I opening profiles for (a) poly-A/poly-T and (b) poly-C/poly-G homogeneous dsRNA sequences at [Na^+] concentrations of 71 mM (squares), 121 mM (diamonds), 221 mM (triangle up), 621 mM (triangle down) and 1021 mM (bullets). Calculations were carried out at (a) 180 K and (b) 330 K. These temperatures have no relation to the melting temperature.

[1] I. Ferreira, E. A. Jolley, B. M. Znosko, G. Weber, Replacing salt correction factors with optimized RNA nearest-neighbour enthalpy and entropy parameters, Chem. Phys. 521 (2019) 69–76, doi:10.1016/j.chemphys.2019.01.016, URL https://www.sciencedirect.com/science/article/abs/pii/S0301010418311200.

[2] T. Xia, J. SantaLucia, Jr., M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, D. H. Turner, Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs, Biochem. 37 (1998) 14719–14735.

[3] Z. Chen, B. M. Znosko, Effect of Sodium Ions on RNA Duplex Stability, Biochem. 52 (42) (2013) 7477–7485, doi:10.1021/bi4008275.

[4] S.-H. Hung, Q. Yu, D. M. Gray, R. L. Ratliff, Evidence from CD spectra that d(purine)-r(pyrimidine) and r(purine)-d(pyrimidine) hybrids are in different structural classes, Nucleic Acids Res. 22 (20) (1994) 4326–4334.

[5] S. Nakano, M. Fujimoto, H. Hara, N. Sugimoto, Nucleic acid duplex stability: influence of base composition on cation effects, Nucleic Acids Res. 27 (14) (1999) 2957.

[6] J. Lisowiec-Wachnicka, B. M. Znosko, A. Pasternak, Contribution of 3′T and 3′TT overhangs to the thermodynamic stability of model siRNA duplexes, Biophys. Chem. 246 (2019) 35–39.

[7] J. I. Gyi, G. L. Conn, A. N. Lane, T. Brown, Comparison of the thermodynamic stabilities and solution conformations of DNA·RNA hybrids containing purine-rich and pyrimidine-rich strands with DNA and RNA duplexes, Biochem. 35 (38) (1996) 12538–12548, doi:10.1021/bi960948z.

[8] A. Pasternak, J. Wengel, Thermodynamics of RNA duplexes modified with unlocked nucleic acid nucleotides, Nucleic Acids Res. 38 (19) (2010) 6697–6706, doi:10.1093/nar/gkq561.

[9] S. Wang, E. T. Kool, Origins of the large differences in stability of DNA and RNA helixes: C-5 methyl and 2′-hydroxyl effects, Biochem. 34 (12) (1995) 4125–4132.

[10] S. J. Schroeder, D. H. Turner, Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA, Biochem. 39 (31) (2000) 9257–9274.

[11] I. Ferreira, T. D. Amarante, G. Weber, DNA terminal base pairs have weaker hydrogen bonds especially for AT under low salt concentration, J. Chem. Phys. 143 (2015) 175101, doi:10.1063/1.4934783.

A.2 Complete Mesoscopic Parameterization of Single LNA Modifications in DNA Applied to Oncogene Probe Design

Supplementary information: Complete mesoscopic parametrization of single LNA modifications in DNA applied to oncogene probe design

Izabela Ferreira[a,b], Sofie Slott[c], Kira Astakhova[c], Gerald Weber[a,*]

[a]*Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil.*
[b]*Programa Interunidades de Pós-Graduação em Bioinformática, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil*
[c]*Department of Chemistry, Technical University of Denmark, Kemitorvet, Bygning 207, 2800 Kgs. Lyngby, Denmark*

Supplementary Tab. S1 shows the number of modified and non-modified NN. Supplementary Tab. S2 shows the sequences used in the minimization and their respective measured and predicted temperature. Supplementary Tab. S3 shows the measured and predicted melting temperatures for the validation set of sequences containing LNA tandem modifications. Supplementary Tab. S4 shows a summary of the oncogenes used for the probe design and their respective human assemblies. Supplementary Tabs.S5–S11 show the melting temperature predictions for *BRAF*, *EGFR* and *KRAS* variants, respectively. Supplementary Tab.S12 shows the MALDI-MS and purities of the measured probes. Figs. S1— S14 show the HPLC and MALDI-MS of the measured probes. Figs. S15— S20 show the representative T_m curve of the measured probes. Figs. S21— S24 show the average opening profiles for the LNA modified probes measured in this study. Fig. S25 shows the calculated configurational entropy S_y comparing an unmodified DNA sequence to LNA modifications.

*Lead author

Email addresses: izabelaferreira13@gmail.com (Izabela Ferreira), sofslo@kemi.dtu.dk (Sofie Slott), kiraas@kemi.dtu.dk (Kira Astakhova), gweberbh@gmail.com (Gerald Weber)

Table S1: Number of occurrences of base pairs and NN pairs contained in the final minimized dataset of Refs. [1–3]. LNA modifications are preceded by a plus sign and marked in **bold**.

canonical	occurrences	modified	occurrences	modified	occurrences	modified	occurrences	modified	occurrences
ATpAT	253	A**+T**pAT	28	ATpA**+T**	26	ATp**+AT**	18	TApT**+A**	14
ATpTA	155	A**+T**pTA	27	ATpT**+A**	27				
TApAT	164	TAp**+AT**	20	**+T**ApAT	21				
ATpCG	386	A**+T**pCG	18	ATpC**+G**	27	ATp**+C**G	18	GCpT**+A**	18
ATpGC	345	A**+T**pGC	24	ATpG**+C**	30	ATp**+G**C	12	CGpT**+A**	19
CGpAT	348	CGpA**+T**	31	CGp**+A**T	18	C**+G**pAT	25	TApG**+C**	17
GCpAT	360	GCpA**+T**	19	GCp**+A**T	22	G**+C**pAT	30	**+G**CpAT	18
CGpCG	287	CGpC**+G**	16	CGp**+C**G	19	C**+G**pCG	26	GCpG**+C**	20
CGpGC	191	CGpG**+C**	21	CGp**+G**C	20				
GCpCG	195	GCpC**+G**	22	GCp**+C**G	21				

Table S2: Sequences used for the minimization procedure. Shown are the experimental temperatures T_i and corresponding predictions T_i'. All temperatures are in °C. Also shown are total squared difference $\chi 2$ and the average difference of predicted and measured temperatures $\langle \Delta T \rangle$, Eqs. (6) and (8) in the main text. LNA modifications are preceded by a plus sign and marked in **bold**.

5' → 3'/3' → 5'	T_i (°C)	T_i' (°C)
1 CACGGCTC GTGCCGAG [1]	40.0	40.3
2 GGTGCCAA CCACGGTT [1]	36.1	35.8
3 TGCACGCTA ACGTGCGAT [1]	45.4	42.8
4 ACGTCTTCG TGCAGAAGC [1]	39.7	39.7
5 GCAGGTCTGC CGTCCAGACG [1]	52.0	50.7
6 GTCGAACAGC CAGCTTGTCG [1]	49.5	46.2
7 CTAAATAGCG GATTTATCGC [1]	38.7	37.6
8 GGACCTCGAC CCTG+**G**AGCTG [3]	57.7	56.1
9 GGACCTCGAC CCTGG+**A**GCTG [3]	57.2	56.9
10 GGACCTCGAC CCTGGAGCTG [3]	50.9	51.7
11 GGACCTCGAC CCT+**G**GAGCTG [3]	55.2	56.6
12 TTCATAGCCGT AA+**G**TATCGGCA [3]	51.0	52.1
13 TTCATAGCCGT AAGT+**A**TCGGCA [3]	51.8	51.2
14 TTCATAGCCGT AAGTAT+**C**GGCA [3]	57.4	55.1
15 TTCATAGCCGT AAGTATCGGCA [3]	50.1	49.0
16 TTCATAGCCGT AAGTA+**T**CGGCA [3]	58.6	54.8
17 TTCATAGCCGT AAG+**T**ATCGGCA [3]	53.1	54.2
18 TTCATAGCCGT AAGTATC+**G**GCA [3]	54.0	52.3
19 TTGGGAGTAGC AACCCTCATCG [1]	48.3	50.9
20 ACTGGCATCTG TGACCGTAGAC [1]	51.2	50.2
21 CTATTGGCGAC GATAACCGCTG [3]	52.4	50.2
22 CTATTGGCGAC GATAA+**C**CGCTG [3]	56.1	54.6
23 ACAAGCGACTC TGTTCGCTGAG [1]	52.5	50.0
24 CTACGCATTCC GATGCG+**T**AAGG [3]	54.2	55.3
25 CTACGCATTCC GAT+**G**CGTAAGG [3]	52.4	53.1
26 CTACGCATTCC GATGCGTA+**A**GG [3]	52.7	54.2
27 CTACGCATTCC GATGCGTAAGG [3]	50.1	50.2
28 CTACGCATTCC GATGC+**G**TAAGG [3]	52.1	53.5
29 GTAGCAGGAGT CATCGTCCTCA [1]	49.1	49.8
30 CGCTGTTACGC GCGACAATGCG [1]	53.4	53.7
31 CCATTGCTACC GGTAACGATGG [1]	48.9	49.5
32 GTAGCGATGTA CATCGCTACAT [1]	48.9	47.6
33 ATTTGACTCAG TAAACTGAGTC [1]	43.5	40.8
34 GTGGATCTTTA CACCTAGAAAT [1]	43.7	42.4
35 GTATCAAGTCT CATAGT+**T**CAGA [3]	47.2	46.5
36 GTATCAAGTCT CATAGTTCAGA [3]	43.9	41.2
37 GTATCAAGTCT CATAGTT+**C**AGA [3]	49.9	49.4
38 GTATCAAGTCT CATAG+**T**TCAGA [3]	48.2	45.3
39 CTGAAGTCCGC GACTTC+**A**GGCG [3]	59.7	58.3
40 CTGAAGTCCGC GACTTCA+**G**GCG [3]	59.7	58.3

Continued on next page

Table S2 – *Continued from previous page*

5' → 3'/3' → 5'	T_i (°C)	T_i' (°C)
41 CTGAAGTCCGC GACTTCAG**+G**CG [3]	57.2	57.5
42 CTGAAGTCCGC GACTTCAGGCG [3]	54.9	53.5
43 CTGAAGTCCGC GACTT**+C**AGGCG [3]	61.4	59.6
44 CGGTTGTGGCG GCCAACACCGC [1]	57.8	57.5
45 TGCGGATAAGT ACGCCT**+A**TTCA [3]	51.6	51.2
46 TGCGGATAAGT ACGCCTATTCA [3]	50.7	49.0
47 TGCGGATAAGT ACGCCTAT**+T**CA [3]	52.5	53.8
48 ATCTATCCGGC TAGATAGGCCG [1]	50.9	52.9
49 CGAACGTCTAT GCTTGCAGATA [1]	46.2	47.2
50 CCTGCGATGAC GGACGCTA**+C**TG [3]	60.0	59.5
51 CCTGCGATGAC GGACGCT**+A**CTG [3]	56.6	56.9
52 CCTGCGATGAC GGACGC**+T**ACTG [3]	58.8	57.7
53 CCTGCGATGAC GGAC**+G**CTACTG [3]	56.6	57.5
54 CCTGCGATGAC GGACGCTACTG [3]	54.9	54.3
55 CCTGCGATGAC GGACG**+C**TACTG [3]	59.8	58.8
56 CTAACGGATGC GA**+T**TGCCTACG [3]	53.6	54.1
57 CTAACGGATGC GATTGCCTACG [3]	50.2	50.2
58 CTAACGGATGC GATTG**+C**CTACG [3]	55.6	54.0
59 CTAACGGATGC GATTGC**+C**TACG [3]	55.2	53.1
60 CTAACGGATGC GATTGCC**+T**ACG [3]	53.9	53.8
61 CTAACGGATGC GAT**+T**GCCTACG [3]	53.7	54.0
62 CTAACGGATGC GATT**+G**CCTACG [3]	54.0	53.1
63 TTGCTCGATGT AACGAGCTACA [1]	50.3	48.5
64 GGAACAAGATGC CCTTG**+T**TCTACG [3]	56.7	55.1
65 GGAACAAGATGC CCTTGT**+T**CTACG [3]	54.9	56.1
66 GGAACAAGATGC CCTTGTT**+C**TACG [3]	57.4	56.6
67 GGAACAAGATGC CCTTGTTCTACG [3]	52.4	51.7
68 GGAACAAGATGC CCTT**+G**TTCTACG [3]	56.1	54.3
69 GGTCCTTGCTTGGTG CCAGGAACGAACCAC [2]	67.8	65.5
70 GGTCCTTTCTTGGTG CCAGGAAAGAACCAC [2]	65.3	62.8
71 GGTCCTTCCTTGGTG CCAGGAAGGAACCAC [2]	67.4	65.8
72 GGTCCTTACTTGGTG CCAGGAATGAACCAC [2]	63.6	62.8
73 ACGACCAGAGTTACAG TGCTGGTCTC**+A**ATGTC [3]	68.1	67.4
74 ACGACCAGAGTTACAG TGCTGGTCTCA**+A**TGTC [3]	66.8	65.8
75 ACGACCAGAGTTACAG TGCTGGTCTCAATGTC [3]	65.5	64.6
76 ACGACCAGAGTTACAG TGCTGGTCTCAA**+T**GTC [3]	68.3	67.5
77 ACGACCAGAGTTACAG TGCTG**+G**TCTCAATGTC [3]	69.2	67.5
78 ACGACCAGAGTTACAG TGCT**+G**GTCTCAATGTC [3]	68.5	67.8
79 ACGACCAGAGTTACAG TGCTGGT**+C**TCAATGTC [3]	69.0	67.9
80 CACGG**+C**TC GTGCCGAG [1]	47.5	45.6
81 GGTG**+C**CAA CCACGGTT [1]	41.5	39.1
82 GGT**+G**CCAA CCACGGTT [1]	38.4	39.8
83 GGTGC**+C**AA CCACGGTT [1]	39.1	40.6
84 CA**+C**GGCTC GTGCCGAG [1]	46.8	47.2

Continued on next page

Table S2 – *Continued from previous page*

5' → 3'/3' → 5'	T_i (°C)	T_i' (°C)
85 CAC**+G**GCTC GTGCCGAG [1]	43.6	44.7
86 GG**+T**GCCAA CCACGGTT [1]	42.2	41.9
87 CACG**+G**CTC GTGCCGAG [1]	47.3	46.1
88 ACGTCT**+T**CG TGCAGAAGC [1]	42.1	42.2
89 ACGT**+C**TTCG TGCAGAAGC [1]	46.9	44.9
90 TGCAC**+G**CTA ACGTGCGAT [1]	47.5	46.1
91 TGC**+A**CGCTA ACGTGCGAT [1]	46.8	47.3
92 AC**+G**TCTTCG TGCAGAAGC [1]	40.8	43.3
93 TG**+C**ACGCTA ACGTGCGAT [1]	47.9	47.1
94 ACGTC**+T**TCG TGCAGAAGC [1]	43.2	45.6
95 ACG**+T**CTTCG TGCAGAAGC [1]	42.9	43.4
96 CTAAA**+T**AGCG GATTTATCGC [1]	41.8	44.3
97 CTAA**+A**TAGCG GATTTATCGC [1]	40.0	42.4
98 GGACCT**+C**GAC CCTGGAGCTG [3]	55.7	55.7
99 CT**+A**AATAGCG GATTTATCGC [1]	37.7	40.1
100 GC**+A**GGTCTGC CGTCCAGACG [1]	55.4	54.7
101 GTCGAA**+C**AGC CAGCTTGTCG [1]	53.9	54.8
102 GTCGA**+A**CAGC CAGCTTGTCG [1]	51.5	51.6
103 TATTAAGCGACCACACATAA ATAATTCGCTGGTGTGTATT [1]	69.1	67.2
104 GTCG**+A**ACAGC CAGCTTGTCG [1]	51.3	50.8
105 GCAG**+G**TCTGC CGTCCAGACG [1]	55.9	55.4
106 GGA**+C**CTCGAC CCTGGAGCTG [3]	57.2	56.9
107 GCA**+G**GTCTGC CGTCCAGACG [1]	58.3	56.8
108 GTC**+G**AACAGC CAGCTTGTCG [1]	48.3	50.1
109 ATTATGCTCCAATCATGTCG TAATACGAGGTTAGTACAGC [1]	68.1	68.5
110 GGAC**+C**TCGAC CCTGGAGCTG [3]	55.3	55.9
111 GGACC**+T**CGAC CCTGGAGCTG [1]	53.3	55.8
112 CTA**+A**ATAGCG GATTTATCGC [1]	41.1	42.2
113 CTA**+A**CGGATGC GATTGC**+C**TACG [3]	61.9	59.0
114 CTA**+A**CGGATGC GATTG**+C**CTACG [3]	60.1	57.8
115 CTA**+A**CGGATGC GATTGCCTACG [3]	52.7	55.0
116 CCTGC**+G**ATGAC GGACGCTACTG [3]	57.5	57.4
117 CTAAC**+G**GATGC GATTGCCTACG [3]	53.8	53.2
118 CTAAC**+G**GATGC GAT**+T**GCCTACG [3]	56.1	58.0
119 CTAAC**+G**GATGC GATTGCC**+T**ACG [3]	60.0	57.7
120 CTACGCA**+T**TCC GATGCGTAAGG [3]	53.2	54.6
121 CTGA**+A**GTCCGC GACTTCA**+G**GCG [3]	65.4	62.6
122 CTGA**+A**GTCCGC GACTTCAGGCG [3]	60.5	58.1
123 AT**+T**TGACTCAG TAAACTGAGTC [1]	45.1	43.9
124 TTC**+A**TAGCCGT AAGTA**+T**CGGCA [3]	58.8	57.4
125 TTC**+A**TAGCCGT AAGTATCGGCA [3]	52.4	51.8
126 TTC**+A**TAGCCGT AAGTATC**+G**GCA [3]	55.2	56.3
127 TTC**+A**TAGCCGT AAGTAT**+C**GGCA [3]	58.4	55.6
128 TTGGGAG**+T**AGC AACCCTCATCG [1]	53.2	55.7

Continued on next page

Table S2 – *Continued from previous page*

5' → 3'/3' → 5'	T_i (°C)	T_i' (°C)
129 GTGGATCT**+T**TA CACCTAGAAAT [1]	45.1	45.5
130 CCTGCGA**+T**GAC GGACGCTACTG [3]	59.1	59.3
131 CGGTT**+G**TGGCG GCCAACACCGC [1]	60.0	60.1
132 TTGCTCGA**+T**GT AACGAGCTACA [1]	54.2	53.8
133 ATTT**+G**ACTCAG TAAACTGAGTC [1]	45.2	44.4
134 CGCTGTT**+A**CGC GCGACAATGCG [1]	55.0	56.6
135 CTG**+A**AGTCCGC GACTTCAGGCG [3]	59.7	57.5
136 CTG**+A**AGTCCGC GACTTC**+A**GGCG [3]	64.3	61.7
137 CGGTTGTG**+G**CG GCCAACACCGC [1]	60.8	61.5
138 ATCTA**+T**CCGGC TAGATAGGCCG [1]	52.6	56.4
139 CTAA**+C**GGATGC GATTGCC**+T**ACG [3]	60.4	60.4
140 CTAA**+C**GGATGC GATTGCCTACG [3]	54.2	55.0
141 CTAA**+C**GGATGC GATTGC**+C**TACG [3]	61.9	59.7
142 CC**+T**GCGATGAC GGACGCTACTG [3]	58.4	59.4
143 CG**+G**TTGTGGCG GCCAACACCGC [1]	61.4	61.3
144 CGAACGTC**+T**AT GCTTGCAGATA [1]	49.9	53.1
145 CCAT**+T**GCTACC GGTAACGATGG [1]	51.9	53.2
146 CTGAA**+G**TCCGC GACTTCA**+G**GCG [3]	64.3	62.4
147 CTGAA**+G**TCCGC GACTTCAGGCG [3]	59.3	58.0
148 CTGAA**+G**TCCGC GACTTCAG**+G**CG [3]	62.8	61.2
149 CTACGC**+A**TTCC GATGCGTA**+A**GG [3]	57.6	56.9
150 CTACGC**+A**TTCC GATGCGTAAGG [3]	51.2	52.9
151 ATCTATC**+C**GGC TAGATAGGCCG [1]	53.0	54.4
152 GT**+A**GCAGGAGT CATCGTCCTCA [1]	50.1	52.9
153 GT**+G**GATCTTTA CACCTAGAAAT [1]	46.0	47.1
154 TTGGG**+A**GTAGC AACCCTCATCG [1]	54.3	55.9
155 GTAGC**+G**ATGTA CATCGCTACAT [1]	51.3	51.3
156 GTGGA**+T**CTTTA CACCTAGAAAT [1]	45.6	46.3
157 GTAGCAGG**+A**GT CATCGTCCTCA [1]	52.6	55.0
158 CCATTG**+C**TACC GGTAACGATGG [1]	53.8	55.9
159 AC**+T**GGCATCTG TGACCGTAGAC [1]	53.4	55.7
160 ATTTGACT**+C**AG TAAACTGAGTC [1]	47.0	48.2
161 ACTGGCAT**+C**TG TGACCGTAGAC [1]	54.8	55.5
162 CTAACGG**+A**TGC GATT**+G**CCTACG [3]	54.5	56.9
163 CTAACGG**+A**TGC GATTGCCTACG [3]	53.4	54.3
164 CTAACGG**+A**TGC GATTG**+C**CTACG [3]	57.5	55.2
165 AC**+A**AGCGACTC TGTTCGCTGAG [1]	52.5	52.6
166 CTATTGG**+C**GAC GATAA**+C**CGCTG [3]	59.1	59.2
167 CTATTGG**+C**GAC GATAACCGCTG [3]	56.7	53.4
168 TT**+G**CTCGATGT AACGAGCTACA [1]	51.9	51.7
169 CGCT**+G**TTACGC GCGACAATGCG [1]	56.3	56.4
170 CT**+G**AAGTCCGC GACTTC**+A**GGCG [3]	61.0	60.6
171 CT**+G**AAGTCCGC GACTT**+C**AGGCG [3]	63.3	60.3
172 CT**+G**AAGTCCGC GACTTCAGGCG [3]	56.0	56.6

Continued on next page

Table S2 – *Continued from previous page*

5' → 3'/3' → 5'	T_i (°C)	T_i' (°C)
173 TTGCT**+C**GATGT AACGAGCTACA [1]	53.4	53.4
174 CTAACG**+G**ATGC GATT**+G**CCTACG [3]	56.7	54.0
175 CTAACG**+G**ATGC GAT**+T**GCCTACG [3]	57.4	58.8
176 CTAACG**+G**ATGC GATTGCCTACG [3]	55.3	53.9
177 ACTGG**+C**ATCTG TGACCGTAGAC [1]	58.0	55.3
178 CTACG**+C**ATTCC GATGCGTAAGG [3]	56.4	54.8
179 CTACG**+C**ATTCC GAT**+G**CGTAAGG [3]	59.3	59.0
180 CCTGCG**+A**TGAC GGACGCTACTG [3]	57.9	58.4
181 TTCAT**+A**GCCGT AAGTATCGGCA [3]	53.0	52.1
182 TTCAT**+A**GCCGT AAGTATC**+G**GCA [3]	59.9	56.6
183 TTCAT**+A**GCCGT AAG**+T**ATCGGCA [3]	55.2	57.1
184 CCA**+T**TGCTACC GGTAACGATGG [1]	51.5	54.0
185 GCTACTC**+C**CAA CGATGAGGGTT [1]	52.3	54.3
186 CC**+A**TTGCTACC GGTAACGATGG [1]	48.3	52.3
187 CGC**+T**GTTACGC GCGACAATGCG [1]	58.7	58.7
188 ATTTG**+A**CTCAG TAAACTGAGTC [1]	46.6	46.6
189 GTA**+T**CAAGTCT CATAGT**+T**CAGA [3]	51.8	50.1
190 GTA**+T**CAAGTCT CATAGTTCAGA [3]	47.2	45.2
191 CGAAC**+G**TCTAT GCTTGCAGATA [1]	49.0	50.4
192 CCT**+G**CGATGAC GGACGCTACTG [3]	56.2	57.3
193 TTCA**+T**AGCCGT AAGTATCGGCA [3]	54.5	54.5
194 TTCA**+T**AGCCGT AAGTATC**+G**GCA [3]	60.9	56.9
195 TTCA**+T**AGCCGT AAGTAT**+C**GGCA [3]	63.1	59.9
196 GTA**+G**CGATGTA CATCGCTACAT [1]	50.9	52.9
197 TGCGGA**+T**AAGT ACGCCTATTCA [3]	52.9	54.5
198 TGCGGA**+T**AAGT ACGCCTAT**+T**CA [3]	59.0	58.9
199 ACAAGCGA**+C**TC TGTTCGCTGAG [1]	57.5	56.7
200 TTCATA**+G**CCGT AAG**+T**ATCGGCA [3]	55.8	56.5
201 TTCATA**+G**CCGT AAGTATCGGCA [3]	52.6	53.1
202 TTCATA**+G**CCGT AAGT**+A**TCGGCA [3]	55.1	55.8
203 ACAAG**+C**GACTC TGTTCGCTGAG [1]	57.1	53.2
204 TTGG**+G**AGTAGC AACCCTCATCG [1]	52.6	55.4
205 GT**+A**TCAAGTCT CATAGTTCAGA [3]	46.8	43.6
206 GT**+A**TCAAGTCT CATAG**+T**TCAGA [3]	51.6	47.6
207 GTATC**+A**AGTCT CATAGTTCAGA [3]	46.2	44.2
208 CCTG**+C**GATGAC GGACGCTACTG [3]	59.6	60.2
209 ATTTGA**+C**TCAG TAAACTGAGTC [1]	49.1	48.9
210 TTGGGAGT**+A**GC AACCCTCATCG [1]	52.0	53.9
211 TT**+C**ATAGCCGT AAGTATCGGCA [3]	54.7	54.6
212 TT**+C**ATAGCCGT AAGTA**+T**CGGCA [3]	61.1	59.5
213 TT**+C**ATAGCCGT AAGT**+A**TCGGCA [3]	60.1	57.3
214 TT**+C**ATAGCCGT AAGTAT**+C**GGCA [3]	62.2	59.3
215 CG**+A**ACGTCTAT GCTTGCAGATA [1]	48.9	51.4
216 ATCTAT**+C**CGGC TAGATAGGCCG [1]	53.1	56.6

Continued on next page

Table S2 – *Continued from previous page*

5' → 3'/3' → 5'	T_i (°C)	T_i' (°C)
217 TGCGGATA**+A**GT ACGCCT**+A**TTCA [3]	55.2	55.8
218 TGCGGATA**+A**GT ACGCCTATTCA [3]	54.6	53.8
219 CTA**+C**GCATTCC GATGCG**+T**AAGG [3]	61.3	58.1
220 CTA**+C**GCATTCC GATGC**+G**TAAGG [3]	60.6	59.0
221 CTA**+C**GCATTCC GATGCGTAAGG [3]	54.6	55.0
222 GTAG**+C**GATGTA CATCGCTACAT [1]	53.6	53.8
223 GTAGC**+A**GGAGT CATCGTCCTCA [1]	51.9	53.6
224 TTCATAG**+C**CGT AAGTA**+T**CGGCA [3]	60.0	57.2
225 TTCATAG**+C**CGT AAGTATCGGCA [3]	55.1	53.4
226 TTCATAG**+C**CGT AAGT**+A**TCGGCA [3]	57.3	56.6
227 TTGGGA**+G**TAGC AACCCTCATCG [1]	52.9	55.5
228 TT**+G**GGAGTAGC AACCCTCATCG [1]	51.3	54.1
229 CGCTG**+T**TACGC GCGACAATGCG [1]	56.8	57.4
230 CGCTGT**+T**ACGC GCGACAATGCG [1]	57.0	57.4
231 ATC**+T**ATCCGGC TAGATAGGCCG [1]	54.8	58.5
232 CCATT**+G**CTACC GGTAACGATGG [1]	50.5	52.9
233 ATCT**+A**TCCGGC TAGATAGGCCG [1]	52.5	55.0
234 TTG**+G**GAGTAGC AACCCTCATCG [1]	53.1	56.2
235 CTACGCAT**+T**CC GATGCG**+T**AAGG [3]	56.7	57.2
236 CTACGCAT**+T**CC GATGCGTAAGG [3]	54.1	52.2
237 GTAT**+C**AAGTCT CATAGTTCAGA [3]	48.3	48.7
238 GTAGCGA**+T**GTA CATCGCTACAT [1]	54.7	52.9
239 GGA**+A**CAAGATGC CCTTGTTCTACG [3]	54.0	56.2
240 GGAACAAG**+A**TGC CCTTGTTCTACG [3]	54.9	55.7
241 GGAACA**+A**GATGC CCTTGTTCTACG [3]	55.9	56.1
242 GGAAC**+A**AGATGC CCTTGTTCTACG [3]	53.9	54.2
243 GGAACAA**+G**ATGC CCTTGTTCTACG [1]	56.9	56.4
244 CGGTCATCCACGTAGCCAGGTCA GCCAGTAGGTGCATCGGTCCAGT [2]	77.7	75.1
245 CGGTCATCAACGTAGCAAGGTCA GCCAGTAGTTGCATCGTTCCAGT [2]	75.8	72.8
246 CGGTCATCTACGTAGCTAGGTCA GCCAGTAGATGCATCGATCCAGT [2]	74.1	73.0
247 CGGTCATCGACGTAGCGAGGTCA GCCAGTAGCTGCATCGCTCCAGT [2]	78.8	75.4
248 GGTCCT**+T**GCTTGGTG CCAGGAACGAACCAC [2]	66.2	67.9
249 GGTCCTT**+A**CTTGGTG CCAGGAATGAACCAC [2]	60.9	64.8
250 GGTCCTTT**+C**TTGGTG CCAGGAAAGAACCAC [2]	64.4	66.4
251 GGTCCTTG**+C**TTGGTG CCAGGAACGAACCAC [2]	67.4	69.8
252 GGTCCTTA**+C**TTGGTG CCAGGAATGAACCAC [2]	63.6	67.8
253 GGTCCT**+T**TCTTGGTG CCAGGAAAGAACCAC [2]	62.5	64.7
254 GGTCCTT**+C**CTTGGTG CCAGGAAGGAACCAC [2]	65.6	68.7
255 GGTCCTT**+T**CTTGGTG CCAGGAAAGAACCAC [2]	62.9	64.1
256 GGTCCT**+T**CCTTGGTG CCAGGAAGGAACCAC [2]	64.3	67.1
257 GGTCCTTC**+C**TTGGTG CCAGGAAGGAACCAC [2]	65.6	68.7
258 GGTCCTT**+G**CTTGGTG CCAGGAACGAACCAC [2]	65.1	67.9
259 GGTCCT**+T**ACTTGGTG CCAGGAATGAACCAC [2]	62.0	65.4
260 ACGACCAGAG**+T**TACAG TGCTGGTCTCAATGTC [3]	68.1	66.8

Continued on next page

Table S2 – *Continued from previous page*

5' → 3'/3' → 5'	T_i (°C)	T_i' (°C)
261 ACG**+A**CCAGAGTTACAG TGCTGGTCTCAATGTC [3]	67.5	67.8
262 ACGACCAGAGT**+T**ACAG TGCTGGTCTCAATGTC [3]	68.4	66.9
263 ACGACCAG**+A**GTTACAG TGCTGGTCTCAATGTC [3]	69.5	67.8
264 ACGACCAGA**+G**TTACAG TGCTGGTCTCAATGTC [3]	67.9	67.4
265 ACGAC**+C**AGAGTTACAG TGCTGGTCTCAATGTC [3]	68.6	68.4
266 AC**+G**ACCAGAGTTACAG TGCTGGTCTCAATGTC [3]	66.2	66.9
267 ACGA**+C**CAGAGTTACAG TGCTGGTCTCAATGTC [3]	69.3	68.6
268 AT**+T**ATGCTCCAATCATGTCG TAATACGAGGTTAGTACAGC [1]	70.0	70.0
269 ATTAT**+G**CTCCAATCATGTCG TAATACGAGGTTAGTACAGC [1]	69.8	70.1
270 ATTATGCT**+C**CAATCATGTCG TAATACGAGGTTAGTACAGC [1]	71.2	70.5
271 TATTAAGCGACCACACA**+T**AA ATAATTCGCTGGTGTGTATT [1]	70.4	69.4
272 TATTAAGCGAC**+C**ACACATAA ATAATTCGCTGGTGTGTATT [1]	71.0	69.8
273 TATTAAGCGACCAC**+A**CATAA ATAATTCGCTGGTGTGTATT [1]	70.2	68.8
274 ATTATGCTCCAATC**+A**TGTCG TAATACGAGGTTAGTACAGC [1]	69.8	69.6
275 GGAC**+C**T**+C**GAC CCTGGAGCTG [3]	61.2	57.4
276 ATTATGCTCCA**+A**TCATGTCG TAATACGAGGTTAGTACAGC [1]	69.9	70.0
277 TA**+T**TAAGCGACCACACATAA ATAATTCGCTGGTGTGTATT [1]	69.3	69.0
278 TATTA**+A**GCGACCACACATAA ATAATTCGCTGGTGTGTATT [1]	70.5	69.1
279 ATTATGCTCCAATCATG**+T**CG TAATACGAGGTTAGTACAGC [1]	71.0	69.7
280 TATTAAGC**+G**ACCACACATAA ATAATTCGCTGGTGTGTATT [1]	69.9	68.8
281 GGAACA**+A**G**+A**TGC CCTTGTTCTACG [3]	59.3	59.8
282 GGA**+AC+A**AGATGC CCTTGTTCTACG [3]	58.0	58.5
283 GGTCCT**+T**G**+C**TTGGTG CCAGGAACGAACCAC [2]	69.5	71.3
284 GGTCCT**+T**C**+C**TTGGTG CCAGGAAGGAACCAC [2]	68.1	69.9
285 GGTCCT**+T**A**+C**TTGGTG CCAGGAATGAACCAC [2]	65.7	70.1
286 GGTCCT**+T**T**+C**TTGGTG CCAGGAAAGAACCAC [2]	65.8	68.2
287 G**+G**TC**+C**TT**+G**CT**+T**GG**+T**G CCAGGAACGAACCAC [2]	76.4	74.8
288 G**+G**TC**+C**TT**+T**CT**+T**GG**+T**G CCAGGAAAGAACCAC [2]	73.3	74.3
289 G**+G**TC**+C**TT**+A**CT**+T**GG**+T**G CCAGGAATGAACCAC [2]	72.8	72.6
290 G**+G**TC**+C**TT**+C**CT**+T**GG**+T**G CCAGGAAGGAACCAC [2]	76.5	74.8
291 G**+G**T**+C**C**+T**T**+T**C**+T**T**+G**G**+T**G CCAGGAAAGAACCAC [2]	79.0	77.9
292 G**+G**T**+C**C**+T**T**+A**C**+T**T**+G**G**+T**G CCAGGAATGAACCAC [2]	79.7	76.8
293 G**+G**T**+C**C**+T**T**+G**C**+T**T**+G**G**+T**G CCAGGAACGAACCAC [2]	83.5	78.3
294 G**+G**T**+C**C**+T**T**+C**C**+T**T**+G**G**+T**G CCAGGAAGGAACCAC [2]	83.3	79.6
$\langle \Delta T \rangle$		1.48 °C
χ^2		1013.60 °C^2

Table S3: Predicted temperatures for sequences bearing LNA tandem modifications from Ref.[2–4] using the optimized parameters from EE-DL minimization round. Also shown are total squared difference $\chi 2$ and the average difference of predicted and measured temperatures $\langle \Delta T \rangle$.

5' → 3'/3' → 5'	T_i (°C)	T_i' (°C)
1 GGACCTCGAC/CCT**+G+G**AGCTG [3]	60.9	59.7
2 TTCATAGCCGT/AA**+G+T**ATCGGCA [4]	57.9	55.1
3 TTCATAGCCGT/AAGTATC**+G+G**CA [4]	56.9	54.9
4 TTCATAGCCGT/AAGT**+A+T**CGGCA [4]	56.8	56.7
5 TTCATAGCCGT/AAGTA**+T+C**GGCA [4]	60.7	58.8
6 CTACGCATTCC/GATGCGT**+A+A**GG [4]	52.8	58.0
7 CTAACGGATGC/GAT**+T+G**CCTACG [4]	52.7	57.7
8 CTAACGGATGC/GATTGC**+C+T**ACG [4]	57.6	56.5
9 GGAACAAGATGC/CCTTG**+T+T**CTACG [3]	59.2	60.4
10 GGA**+C+C**TCGAC/CCTGGAGCTG [3]	59.3	58.3
11 CTA**+A+C**GGATGC/GATTGCCTACG [4]	56.5	60.5
12 TT**+C+A**TAGCCGT/AAGTATCGGCA [4]	55.5	54.3
13 CTA**+C+G**CATTCC/GATGCGTAAGG [4]	58.2	58.4
14 TTCATAG**+C+C**GT/AAGTATCGGCA [4]	59.3	55.1
15 TTCAT**+A+G**CCGT/AAGTATCGGCA [4]	55.3	56.1
16 CTAACG**+G+A**TGC/GATTGCCTACG [4]	56.4	54.4
17 CTAC**+G+C**ATTCC/GATGCGTAAGG [4]	56.9	57.0
18 CTACGC**+A+T**TCC/GATGCGTAAGG [4]	53.5	55.5
19 CTACGCA**+T+T**CC/GATGCGTAAGG [4]	57.0	57.7
20 GGAAC**+A+A**GATGC/CCTTGTTCTACG [3]	57.1	58.6
21 GGTCCT**+T+T**CTTGGTG/CCAGGAAAGAACCAC [2]	64.3	66.8
22 GGTCCT**+T+G**CTTGGTG/CCAGGAACGAACCAC [2]	69.7	70.5
23 GGTCCT**+T+A**CTTGGTG/CCAGGAATGAACCAC [2]	65.6	67.4
24 GGTCCT**+T+C**CTTGGTG/CCAGGAAGGAACCAC [2]	67.7	71.0
$\langle \Delta T \rangle$		1.94 °C
χ^2		139.50 °C^2

Table S4: Summary of the oncogenes used for the probe design and their respective human assemblies.

Oncogene	Gene ID	HGNC	Ensembl	MIM	GRCh38.p13
BRAF	673	1097	ENSG00000157764	164757	GCF_000001405.39
KRAS	3845	6407	ENSG00000133703	190070	GCF_000001405.39
EGFR	1956	3236	ENSG00000146648	131550	GCF_000001405.39

Table S5: Predicted temperatures for *BRAF* capture LNA modified probe at medium salt concentration using the calculated parameters at high salt concentration from EE-DL minimization and applying salt correction for DNA[5] sequences. LNA modifications are preceded by a plus sign and marked in **bold**. A cut-off was made in the tables to show only the 30 sequences with the highest and lowest temperature difference between the modified and non-modified probe, the total number of predicted sequences was 1562. As comparative, we also show the ΔT_m of each modified sequence and its non-modified analog. The chosen probe is highlighted in yellow.

T_m	ΔT_m	Probe (5' → 3')	ID
59.5693	0	ATCGAGATTTCTCTGTAGCTA	BRAFcap-CTRL
65.6223	6.053	AT+**C**GAGA+**T**TTCTCTGTAG+**C**TA	BRAFcap-M481
65.5789	6.010	ATCGA+**G**ATTTCTCTGTAG+**C**+**T**A	BRAFcap-M982
65.4673	5.898	ATCGAGATTTCTCTG+**T**AG+**C**+**T**A	BRAFcap-M1532
65.4294	5.860	A+**T**+**C**GAGATTTCTCTGTA+**G**CTA	BRAFcap-M228
65.4029	5.834	ATCGAGATTT+**C**TC+**T**GTAG+**C**TA	BRAFcap-M1356
65.3801	5.811	ATCGAGATTT+**C**T+**C**TGTAG+**C**TA	BRAFcap-M1348
65.3745	5.805	A+**T**+**C**GAGATTTCTCTGTAGC+**T**A	BRAFcap-M230
65.3631	5.794	ATCGAGA+**T**TTCTCTGTAG+**C**+**T**A	BRAFcap-M1180
65.327	5.758	AT+**C**GAGATTTCTCTGTA+**G**C+**T**A	BRAFcap-M567
65.2719	5.703	AT+**C**GAGATTTC+**T**+**C**TGTAGCTA	BRAFcap-M521
65.2643	5.695	A+**T**+**C**GAGATTTCTC+**T**GTAGCTA	BRAFcap-M224
65.2586	5.689	ATCGA+**G**ATTT+**C**TCTGTAG+**C**TA	BRAFcap-M929
65.2567	5.687	ATCGA+**G**ATTTCT+**C**TGTAG+**C**TA	BRAFcap-M948
65.2472	5.678	ATCG+**A**GATTTCTC+**T**GTAG+**C**TA	BRAFcap-M835
65.2339	5.665	A+**T**+**C**GAGATTTCT+**C**TGTAGCTA	BRAFcap-M223
65.232	5.663	A+**T**+**C**GAGATTT+**C**TCTGTAGCTA	BRAFcap-M221

Continued on next page

Table S5 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
65.2263	5.657	ATCG+**A**GATTT+**C**TCTGTAG+**C**TA	BRAFcap-M808
65.2244	5.655	ATCG+**A**GATTTCT+**C**TGTAG+**C**TA	BRAFcap-M827
65.2244	5.655	ATCGA+**G**ATTTCTC+**T**GTAG+**C**TA	BRAFcap-M956
65.1787	5.609	ATCGAG+**A**TTTCTCTGTAG+**C**+**T**A	BRAFcap-M1088
65.1711	5.602	AT+**C**GAGATTTCTC+**T**GTAGC+**T**A	BRAFcap-M545
65.1635	5.594	AT+**C**GAGA+**T**TTCTCTGTAGC+**T**A	BRAFcap-M482
65.1635	5.594	AT+**C**GAGATTTCT+**C**+**T**GTAGCTA	BRAFcap-M531
65.1559	5.587	AT+**C**GAGATTT+**C**TCTGTAGC+**T**A	BRAFcap-M518
65.154	5.585	AT+**C**GAGATTTCT+**C**TGTAGC+**T**A	BRAFcap-M537
65.1349	5.566	ATCGAGATTTC+**T**CTGTAG+**C**+**T**A	BRAFcap-M1428
65.133	5.564	AT+**C**GAGATTTCTC+**T**GTA+**G**CTA	BRAFcap-M543
64.9154	5.346	ATCGAGAT+**T**TCT+**C**TGTAG+**C**TA	BRAFcap-M1225
61.2724	1.703	ATC+**G**AGATTTCTCTGTAGCT+**A**	BRAFcap-M728
61.2047	1.635	ATCGAGATT+**T**CTCT+**G**TAGCTA	BRAFcap-M1303
61.1965	1.627	+**A**TCGAGATT+**T**CTCTGTAGCT+**A**	BRAFcap-M145
61.1636	1.594	ATCGAGATTTCTCTG+**T**AGCTA	BRAFcap-M1521
61.1471	1.578	A+**T**CGAGATTTCTCTGTAGCT+**A**	BRAFcap-M402
61.1451	1.576	+**A**+**T**CGAGATTTCTCTGTAGCT+**A**	BRAFcap-M21
61.0854	1.516	+**A**TCGAGATTTCTCTGT+**A**GCTA	BRAFcap-M197
61.0565	1.487	ATCGAG+**A**TTTCTCTGTAGCTA	BRAFcap-M987
61.0132	1.444	+**A**TCGAGATTTCTCT+**G**TAGCTA	BRAFcap-M184
60.9926	1.423	+**A**TCGAGAT+**T**TCTCTGTAGCTA	BRAFcap-M121
60.9926	1.423	ATCGAGA+**T**TTCTCTGTAGCTA	BRAFcap-M1093
60.972	1.403	ATCGAGATTTCTCTGT+**A**GCT+**A**	BRAFcap-M1547
60.9141	1.345	ATCGAGATTTC+**T**CTGTAGCTA	BRAFcap-M1387
60.9017	1.332	ATCGAGATTTCTCT+**G**TAGCT+**A**	BRAFcap-M1520

Continued on next page

Table S5 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
60.8893	1.320	ATC+**G**AGATTTCTCTGTAGCTA	BRAFcap-M575
60.8852	1.316	ATCGAGAT+**T**TCTCTGTAGCT+**A**	BRAFcap-M1263
60.7651	1.196	A+**T**CGAGATTTCTCTGTAGCTA	BRAFcap-M212
60.7631	1.194	+**A**TCGAGATT+**T**CTCTGTAGCTA	BRAFcap-M134
60.7112	1.142	+**A**+**T**CGAGATTTCTCTGTAGCTA	BRAFcap-M2
60.6531	1.084	ATCGAGATT+**T**CTCTGTAGCT+**A**	BRAFcap-M1330
60.597	1.028	ATCGAGATTTCTCTGT+**A**GCTA	BRAFcap-M1537
60.522	0.953	+**A**TCGAGATTTCTCTGTAGCT+**A**	BRAFcap-M211
60.522	0.953	ATCGAGATTTCTCT+**G**TAGCTA	BRAFcap-M1499
60.5033	0.934	ATCGAGAT+**T**TCTCTGTAGCTA	BRAFcap-M1185
60.2715	0.702	ATCGAGATT+**T**CTCTGTAGCTA	BRAFcap-M1264
60.0743	0.505	+**A**TCGAGATTTCTCTGTAGCTA	BRAFcap-M1
59.9627	0.393	ATCGAGATTTCTCTGTAGCT+**A**	BRAFcap-M1561
59.5693	0.000	ATCGAGATTTCTCTGTAGCTA	BRAFcap-CTRL

Table S6: Predicted temperatures for *BRAF* linker LNA modified probe at medium salt concentration using the calculated parameters at high salt concentration from EE-DL minimization and applying salt correction for DNA[5] sequences. LNA modifications are preceded by a plus sign and marked in **bold**. A cut-off was made in the tables to show only the 30 sequences with the highest and lowest temperature difference between the modified and non-modified probe, the total number of predicted sequences was 834. As comparative, we also show the ΔT_m of each modified sequence and its non-modified analog. The chosen probe is highlighted in yellow.

T_m	ΔT_m	Probe (5' → 3')	ID
50.7521	0	CAACTGTTCAAACTGAT	BRAFlin-CTRL
61.2632	10.511	+**C**AA+**C**TGTTCAAA+**C**TGAT	BRAFlin-M42
60.9684	10.216	CAA+**C**TGTT+**C**AAAC+**T**GAT	BRAFlin-M417
60.9594	10.207	CAAC+**T**GTT+**C**AAA+**C**TGAT	BRAFlin-M495
60.896	10.144	+**C**AACTGTT+**C**AAA+**C**TGAT	BRAFlin-M97
60.8749	10.123	+**C**AA+**C**TGTT+**C**AAACTGAT	BRAFlin-M38
60.7933	10.041	CAA+**C**TG+**T**TCAAA+**C**TGAT	BRAFlin-M397
60.754	10.002	CAA+**C**TGTT+**C**AAA+**C**TGAT	BRAFlin-M416
60.7268	9.975	CAA+**C**TGTTCA+**A**A+**C**TGAT	BRAFlin-M431
60.6239	9.872	CAA+**C**TGTTCAAA+**C**TG+**A**T	BRAFlin-M445
60.539	9.787	CA+**A**CTGTT+**C**AAA+**C**TGAT	BRAFlin-M324
60.5299	9.778	CAA+**C**TGTT+**C**AA+**A**CTGAT	BRAFlin-M415
60.4935	9.741	CAA+**C**TGTT+**C**A+**A**ACTGAT	BRAFlin-M414
60.4722	9.720	CAACTGTT+**C**A+**A**A+**C**TGAT	BRAFlin-M716
60.4358	9.684	CA+**A**+**C**TGTTCAAA+**C**TGAT	BRAFlin-M269
60.4327	9.681	CAA+**C**TGTTCAA+**A**+**C**TGAT	BRAFlin-M437
60.3598	9.608	CAA+**C**+**T**GTTCAAAC+**T**GAT	BRAFlin-M375
60.3598	9.608	CAAC+**T**GTTCAAA+**C**+**T**GAT	BRAFlin-M522
60.3415	9.589	CAA+**C**TGTTCAAA+**C**TGA+**T**	BRAFlin-M446
60.3172	9.565	CAA+**C**T+**G**TTCAAA+**C**TGAT	BRAFlin-M386
60.2532	9.501	CAA+**C**TGTT+**C**AAACTG+**A**T	BRAFlin-M419
60.2319	9.480	CAACTGTT+**C**AAA+**C**TG+**A**T	BRAFlin-M730

Continued on next page

Table S6 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
60.2197	9.468	C+**A**A+**C**TGTTCAAA+**C**TGAT	BRAFlin-M163
60.2197	9.468	CAA+**C**TGTTC+**A**AA+**C**TGAT	BRAFlin-M424
60.1831	9.431	+**C**AA+**C**TGTTCAA+**A**CTGAT	BRAFlin-M41
60.1801	9.428	+**C**A+**A**CTGTTCAAA+**C**TGAT	BRAFlin-M28
60.174	9.422	CA+**A**C+**T**GTTCAAA+**C**TGAT	BRAFlin-M282
60.174	9.422	CAA+**C**TGTTCAA+**A**C+**T**GAT	BRAFlin-M438
53.9496	3.197	CAAC+**T**GTTCAAACTGAT	BRAFlin-M457
53.9462	3.194	CAACTGTTCAAAC+**T**GAT	BRAFlin-M820
53.926	3.174	CAACT+**G**TTCAAACTGA+**T**	BRAFlin-M602
53.926	3.174	CAACTGT+**T**CA+**A**ACTGAT	BRAFlin-M677
53.9125	3.160	+**C**AACTGTTCAAACTGAT	BRAFlin-M1
53.8856	3.133	CAACTGTTCAAACTG+**A**+**T**	BRAFlin-M832
53.8519	3.100	C+**A**ACTGTTCAAACTGA+**T**	BRAFlin-M258
53.8451	3.093	CAACTGTTC+**A**AACTGA+**T**	BRAFlin-M770
53.808	3.056	C+**A**ACT+**G**TTCAAACTGAT	BRAFlin-M181
53.808	3.056	CAACT+**G**TTC+**A**AACTGAT	BRAFlin-M567
53.7709	3.019	CAACTGT+**T**CAAACT+**G**AT	BRAFlin-M699
53.7304	2.978	C+**A**ACTGTTC+**A**AACTGAT	BRAFlin-M223
53.5139	2.762	CAACTGT+**T**CAAACTGA+**T**	BRAFlin-M704
53.4732	2.721	CAACT+**G**T+**T**CAAACTGAT	BRAFlin-M548
53.4189	2.667	CA+**A**CTGTTCAAACTGAT	BRAFlin-M259
53.4189	2.667	CAACTGTTCAA+**A**CTGAT	BRAFlin-M793
53.3985	2.646	C+**A**ACTGT+**T**CAAACTGAT	BRAFlin-M204
53.3951	2.643	CAACTGT+**T**C+**A**AACTGAT	BRAFlin-M669
53.2489	2.497	CAACTGTTCAAACTG+**A**T	BRAFlin-M831
52.88	2.128	CAACTG+**T**TCAAACTGAT	BRAFlin-M603

Continued on next page

Table S6 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
52.8149	2.063	CAACTGTTCA+**A**ACTGAT	BRAFlin-M771
52.6432	1.891	CAACTGTTCAAACT+**G**AT	BRAFlin-M827
52.3846	1.633	CAACTGTTCAAACTGA+**T**	BRAFlin-M833
52.3431	1.591	CAACT+**G**TTCAAACTGAT	BRAFlin-M536
52.267	1.515	CAACTGTTC+**A**AACTGAT	BRAFlin-M742
52.2636	1.511	C+**A**ACTGTTCAAACTGAT	BRAFlin-M138
51.9233	1.171	CAACTGT+**T**CAAACTGAT	BRAFlin-M659
50.7521	0.000	CAACTGTTCAAACTGAT	BRAFlin-CTRL

Table S7: Predicted temperatures for *EGFR* capture LNA modified probe at medium salt concentration using the calculated parameters at high salt concentration from EE-DL minimization and applying salt correction for DNA[5] sequences. LNA modifications are preceded by a plus sign and marked in **bold**. A cut-off was made in the tables to show only the 30 sequences with the highest and lowest temperature difference between the modified and non-modified probe, the total number of predicted sequences was 2325. As comparative, we also show the ΔT_m of each modified sequence and its non-modified analog. The chosen probe is highlighted in yellow.

T_m	ΔT_m	Probe (5' → 3')	ID
57.4035	0	GAGAAAAAGTTTCTCATGTACAGT	EGFRcap-CTRL
62.3721	4.969	GAGAAAAAGTTT+**C**T+**C**ATGTA+**C**AGT	EGFRcap-M2045
62.3707	4.967	G+**A**GAAAAAGTTTCT+**C**ATGTA+**C**AGT	EGFRcap-M483
62.2985	4.895	GA+**G**AAAAAGTTTCT+**C**ATGTA+**C**AGT	EGFRcap-M715
62.2958	4.892	GAGAAAAAGTTTCT+**C**A+**T**GTA+**C**AGT	EGFRcap-M2164
62.2006	4.797	GAGAAAAA+**G**TTTCT+**C**ATGTA+**C**AGT	EGFRcap-M1701
62.1858	4.782	GAGAAAAAGTTTCT+**C**ATGTA+**C**A+**G**T	EGFRcap-M2188
62.1764	4.773	GAGAAAA+**A**GTTTCT+**C**ATGTA+**C**AGT	EGFRcap-M1580
62.1656	4.762	GAGAAAAAG+**T**TTCT+**C**ATGTA+**C**AGT	EGFRcap-M1807
62.1629	4.759	GAGAAAAAGTTTCT+**C**ATG+**T**A+**C**AGT	EGFRcap-M2177
62.1455	4.742	GAGA+**A**AAAGTTTCT+**C**ATGTA+**C**AGT	EGFRcap-M1117
62.1441	4.741	GAGAA+**A**AAGTTTCT+**C**ATGTA+**C**AGT	EGFRcap-M1289
62.1428	4.739	GAGAAA+**A**AGTTTCT+**C**ATGTA+**C**AGT	EGFRcap-M1443
62.1172	4.714	G+**A**GAAAAAGTTT+**C**TCATGTA+**C**AGT	EGFRcap-M462
62.1118	4.708	GAGAAAAAGTTTC+**T**+**C**ATGTA+**C**AGT	EGFRcap-M2101
62.0862	4.683	GAG+**A**AAAAGTTTCT+**C**ATGTA+**C**AGT	EGFRcap-M926
62.0822	4.679	GA+**G**AAAAAGTTT+**C**TCATGTA+**C**AGT	EGFRcap-M694
62.0607	4.657	GAGAAAAAGTTTCT+**C**ATGTA+**C**+**A**GT	EGFRcap-M2187
62.0539	4.650	GAGAAAAAGTTT+**C**TCA+**T**GTA+**C**AGT	EGFRcap-M2062
62.0054	4.602	G+**A**GAAAAAGTTTCTCA+**T**GTA+**C**AGT	EGFRcap-M500
61.9905	4.587	GA+**G**AAAAAGTTTCTCA+**T**GTA+**C**AGT	EGFRcap-M732
61.9662	4.563	GAGAAAAA+**G**TTT+**C**TCATGTA+**C**AGT	EGFRcap-M1680

Continued on next page

Table S7 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
61.9554	4.552	GAGAAAAAGTTTCT+**C**AT+**G**TA+**C**AGT	EGFRcap-M2171
61.9527	4.549	GAGAAAAAGTTT+**C**TCATGTA+**C**A+**G**T	EGFRcap-M2086
61.9446	4.541	GAGAAAA+**A**GTTT+**C**TCATGTA+**C**AGT	EGFRcap-M1559
61.9405	4.537	GAGAAAAAGT+**T**TCT+**C**ATGTA+**C**AGT	EGFRcap-M1899
61.9405	4.537	GAGAAAAAGTTT+**C**TCATG+**T**A+**C**AGT	EGFRcap-M2075
61.9324	4.529	G+**A**GAAAAAGTTTCTCATG+**T**A+**C**AGT	EGFRcap-M513
61.7603	4.357	GAG+**A**AAAAGTTTCTCA+**T**GTA+**C**AGT	EGFRcap-M943
58.6744	1.271	GAGAAAAAGTTTCTCA+**T**GTACAGT	EGFRcap-M2233
58.6478	1.244	GAGAAAAAGTT+**T**CTCATGT+**A**CAGT	EGFRcap-M2012
58.6227	1.219	GAGAAAAAGTTTCTCAT+**G**TACAG+**T**	EGFRcap-M2283
58.6183	1.215	GAGAAAAAGT+**T**TCTCATGTACAG+**T**	EGFRcap-M1947
58.6139	1.210	+**G**AGAAAAAGTT+**T**CTCATGTACAGT	EGFRcap-M187
58.6109	1.207	GAGAAAAAGTTTCTC+**A**TGTACAG+**T**	EGFRcap-M2232
58.5858	1.182	GAGAAAAAGTT+**T**CTCAT+**G**TACAGT	EGFRcap-M1999
58.5725	1.169	GAGAAAAAGTT+**T**CTC+**A**TGTACAGT	EGFRcap-M1982
58.5592	1.156	GAGAAAAAGTTTCTCATG+**T**ACAGT	EGFRcap-M2284
58.5444	1.141	GAGAAAAA+**G**TTTCTCATGTACAGT	EGFRcap-M1629
58.5415	1.138	GAGAAAAAGTTTCTCATGTACA+**G**T	EGFRcap-M2322
58.537	1.133	GAGAAAA+**A**GTTTCTCATGTACAGT	EGFRcap-M1492
58.463	1.059	GAG+**A**AAAAGTTTCTCATGTACAGT	EGFRcap-M764
58.4481	1.045	GAGAAAAAGTT+**T**CTCATGTACAG+**T**	EGFRcap-M2026
58.3798	0.976	GAGAAAAAGTTTC+**T**CATGTACAGT	EGFRcap-M2094
58.3144	0.911	GAGAAAAAG+**T**TTCTCATGTACAGT	EGFRcap-M1750
58.3099	0.906	GAGAAAAAGTTTCTCATGTAC+**A**GT	EGFRcap-M2318
58.2816	0.878	GAGAAA+**A**AGTTTCTCATGTACAGT	EGFRcap-M1338
58.2801	0.877	GAGA+**A**AAAGTTTCTCATGTACAGT	EGFRcap-M975

Continued on next page

Table S7 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
58.2801	0.877	GAGAA+**A**AAGTTTCTCATGTACAGT	EGFRcap-M1166
58.1547	0.751	GAGAAAAAGTTTCTCATGT+**A**CAGT	EGFRcap-M2300
58.1158	0.712	+**G**AGAAAAAGTTTCTCATGTACAGT	EGFRcap-M1
58.0889	0.685	GAGAAAAAGTTTCTCAT+**G**TACAGT	EGFRcap-M2262
58.0829	0.679	GAGAAAAAGT+**T**TCTCATGTACAGT	EGFRcap-M1856
58.0784	0.675	GAGAAAAAGTTTCTC+**A**TGTACAGT	EGFRcap-M2196
57.9508	0.547	GAGAAAAAGTTTCTCATGTACAG+**T**	EGFRcap-M2324
57.9131	0.510	GAGAAAAAGTT+**T**CTCATGTACAGT	EGFRcap-M1948
57.4035	0.000	GAGAAAAAGTTTCTCATGTACAGT	EGFRcap-CTRL

Table S8: Predicted temperatures for *EGFR* linker LNA modified probe at medium salt concentration using the calculated parameters at high salt concentration from EE-DL minimization and applying salt correction for DNA[5] sequences. LNA modifications are preceded by a plus sign and marked in **bold**. A cut-off was made in the tables to show only the 30 sequences with the highest and lowest temperature difference between the modified and non-modified probe, the total number of predicted sequences was 988. As comparative, we also show the ΔT_m of each modified sequence and its non-modified analog. The chosen probe is highlighted in yellow.

T_m	ΔT_m	Probe (5' → 3')	ID
54.4472	0	TTGTTGGATCATATTCGT	EGFRlin-CTRL
62.3575	7.910	TTGTTGGA+**T**+**C**A+**T**ATTCGT	EGFRlin-M816
62.2444	7.797	TTGTTGGAT+**C**ATA+**T**T+**C**GT	EGFRlin-M883
62.1364	7.689	TTGTTGGATCA+**T**A+**T**T+**C**GT	EGFRlin-M934
62.131	7.684	TTGTTGGAT+**C**A+**T**A+**T**TCGT	EGFRlin-M870
62.1013	7.654	TTGTTG+**G**A+**T**+**C**ATATTCGT	EGFRlin-M703
62.0635	7.616	TTGTTG+**G**AT+**C**A+**T**ATTCGT	EGFRlin-M714
62.0554	7.608	TTGTTGG+**A**T+**C**ATATT+**C**GT	EGFRlin-M774
62.0472	7.600	TTGTTGG+**A**T+**C**A+**T**ATTCGT	EGFRlin-M770
61.9606	7.513	TTG+**T**TGGAT+**C**A+**T**ATTCGT	EGFRlin-M476
61.9254	7.478	TTG+**T**TGGATCA+**T**ATT+**C**GT	EGFRlin-M495
61.882	7.435	TTGTT+**G**GAT+**C**ATATT+**C**GT	EGFRlin-M651
61.8495	7.402	TTGTTGGA+**T**+**C**ATA+**T**TCGT	EGFRlin-M818
61.8332	7.386	TTG+**T**TGGAT+**C**ATATT+**C**GT	EGFRlin-M480
61.8115	7.364	TTGTTGGA+**T**CA+**T**ATT+**C**GT	EGFRlin-M835
61.7925	7.345	TTGTTG+**G**ATCA+**T**ATT+**C**GT	EGFRlin-M733
61.7843	7.337	T+**T**GTTGGAT+**C**A+**T**ATTCGT	EGFRlin-M249
61.7653	7.318	TTGT+**T**GGAT+**C**A+**T**ATTCGT	EGFRlin-M568
61.7545	7.307	TTGTTGGAT+**C**ATAT+**T**+**C**GT	EGFRlin-M887
61.7164	7.269	T+**T**GTTGGAT+**C**ATATT+**C**GT	EGFRlin-M253
61.7083	7.261	TTGT+**T**GGAT+**C**ATATT+**C**GT	EGFRlin-M572
61.7083	7.261	TTGTTGG+**A**TCATA+**T**T+**C**GT	EGFRlin-M800

Continued on next page

Table S8 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
61.6511	7.204	TTGTTGGAT+**C**A+**T**ATTC+**G**T	EGFRlin-M873
61.6321	7.185	TTG+**T**+**T**GGAT+**C**ATATTCGT	EGFRlin-M419
61.5694	7.122	TTGTT+**G**GATCA+**T**ATT+**C**GT	EGFRlin-M666
61.5667	7.119	TTGTTGG+**A**T+**C**ATA+**T**TCGT	EGFRlin-M772
61.5422	7.095	TTGTTGG+**A**TCA+**T**ATT+**C**GT	EGFRlin-M789
61.5176	7.070	TTG+**T**TGGAT+**C**ATA+**T**TCGT	EGFRlin-M478
56.8219	2.375	TT+**G**TTGGATCAT+**A**TTCGT	EGFRlin-M392
56.816	2.369	TT+**G**TTGGATCATATTCG+**T**	EGFRlin-M412
56.8072	2.360	TT+**G**TTGGATCATAT+**T**CGT	EGFRlin-M403
56.7866	2.339	TTGTTGGATC+**A**T+**A**TTCGT	EGFRlin-M904
56.7836	2.336	TTGTT+**G**GATCATATTCGT	EGFRlin-M611
56.7836	2.336	TTGTTGGATC+**A**TATTCG+**T**	EGFRlin-M924
56.7748	2.328	TTGTTGGATC+**A**TAT+**T**CGT	EGFRlin-M915
56.7335	2.286	TTGTTGGATCAT+**A**+**T**TCGT	EGFRlin-M948
56.6568	2.210	+**T**TGTTGGATCATATTCG+**T**	EGFRlin-M154
56.5917	2.145	TTGTTGG+**A**TCATATTCGT	EGFRlin-M757
56.5622	2.115	+**T**TGTTGGATCAT+**A**TTCGT	EGFRlin-M134
56.5474	2.100	+**T**TGTTGGATCATAT+**T**CGT	EGFRlin-M145
56.5266	2.079	TTGTTGGATCATA+**T**TCGT	EGFRlin-M963
56.4645	2.017	TTGTTGGATCAT+**A**TTCG+**T**	EGFRlin-M962
56.4556	2.008	TTGTTGGATCAT+**A**T+**T**CGT	EGFRlin-M953
56.4437	1.996	TTGTTGGATCATAT+**T**CG+**T**	EGFRlin-M980
56.2775	1.830	T+**T**GTTGGATCATATTCGT	EGFRlin-M155
56.2745	1.827	TTGT+**T**GGATCATATTCGT	EGFRlin-M519
56.2715	1.824	TTG+**T**TGGATCATATTCGT	EGFRlin-M413
56.1882	1.741	TTGTTGGA+**T**CATATTCGT	EGFRlin-M813

Continued on next page

Table S8 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
56.06	1.613	TTGTTGGATCATATTC+**G**T	EGFRlin-M985
55.8238	1.377	TT+**G**TTGGATCATATTCGT	EGFRlin-M292
55.7999	1.353	TTGTTGGATC+**A**TATTCGT	EGFRlin-M896
55.5626	1.115	+**T**TTGTTGGATCATATTCGT	EGFRlin-M1
55.4752	1.028	TTGTTGGATCAT+**A**TTCGT	EGFRlin-M947
55.4662	1.019	TTGTTGGATCATATTCG+**T**	EGFRlin-M987
55.4572	1.010	TTGTTGGATCATAT+**T**CGT	EGFRlin-M974
54.4472	0.000	TTGTTGGATCATATTCGT	EGFRlin-CTRL

Table S9: Predicted temperatures for *KRAS12D* capture LNA modified probe at medium salt concentration using the calculated parameters at high salt concentration from EE-DL minimization and applying salt correction for DNA[5] sequences. LNA modifications are preceded by a plus sign and marked in **bold**. A cut-off was made in the tables to show only the 30 sequences with the highest and lowest temperature difference between the modified and non-modified probe, the total number of predicted sequences was 1351. As comparative, we also show the ΔT_m of each modified sequence and its non-modified analog. The chosen probe is highlighted in yellow.

T_m	ΔT_m	Probe (5' → 3')	ID
64.0694	0	GCACTCTTGCCTACCCAATC	KRAS12cap-CTRL
71.3693	7.300	G+**C**A+**C**TCTTGCC+**T**ACCCAATC	KRAS12cap-M219
70.7611	6.692	G+**C**A+**C**T+**C**TTGCCTACCCAATC	KRAS12cap-M213
70.5106	6.441	GCA+**C**TC+**T**TGCCTA+**C**CCAATC	KRAS12cap-M557
70.4782	6.409	G+**C**A+**C**TC+**T**TGCCTACCCAATC	KRAS12cap-M214
70.4741	6.405	G+**C**ACTCTTGCC+**T**ACCCA+**A**TC	KRAS12cap-M325
70.4538	6.384	G+**C**ACTCTTGCC+**T**A+**C**CCAATC	KRAS12cap-M321
70.326	6.257	G+**C**A+**C**TCTTGCCTACCCA+**A**TC	KRAS12cap-M225
70.2813	6.212	GCA+**C**T+**C**TTGCC+**T**ACCCAATC	KRAS12cap-M541
70.2468	6.177	GCA+**C**TCTTGCCTA+**C**CCAAT+**C**	KRAS12cap-M633
70.2081	6.139	GCA+**C**TC+**T**TGCC+**T**ACCCAATC	KRAS12cap-M555
70.1572	6.088	G+**C**ACTC+**T**TGCCTACCCAAT+**C**	KRAS12cap-M272
70.1532	6.084	GCA+**C**TC+**T**TGCCTACC+**C**AATC	KRAS12cap-M559
70.1185	6.049	GCA+**C**TCTTGCCTA+**C**CCA+**A**TC	KRAS12cap-M631
70.1165	6.047	G+**C**ACT+**C**TTGCCTA+**C**CCAATC	KRAS12cap-M252
70.1145	6.045	G+**C**A+**C**TCTTGCCTACCCAA+**T**C	KRAS12cap-M226
70.1022	6.033	G+**C**ACTC+**T**TGCCTACCCA+**A**TC	KRAS12cap-M270
70.088	6.019	G+**C**A+**C**TCTTGC+**C**TACCCAATC	KRAS12cap-M218
70.0839	6.014	G+**C**A+**C**TCTTGCCTACCCAAT+**C**	KRAS12cap-M227
70.0819	6.013	G+**C**ACTCTTGCCTA+**C**C+**C**AATC	KRAS12cap-M338
70.0615	5.992	G+**C**ACTCTTG+**C**+**C**TACCCAATC	KRAS12cap-M299
70.0615	5.992	GCA+**C**TCT+**T**GCCTA+**C**CCAATC	KRAS12cap-M570

Continued on next page

Table S9 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
70.0594	5.990	G+**C**ACT+**C**TTGCC+**T**ACCCAATC	KRAS12cap-M250
70.0533	5.984	GCA+**C**T+**C**+**T**TGCCTACCCAATC	KRAS12cap-M536
70.0472	5.978	G+**C**AC+**T**C+**T**TGCCTACCCAATC	KRAS12cap-M230
70.0451	5.976	GCA+**C**TCTTGCC+**T**ACCCAAT+**C**	KRAS12cap-M618
70.0349	5.965	GCACTC+**T**TGCC+**T**A+**C**CCAATC	KRAS12cap-M931
70.0207	5.951	GCA+**C**TCTTGC+**C**+**T**ACCCAATC	KRAS12cap-M601
66.4677	2.398	GCACTCTTGCCTACCC+**A**AT+**C**	KRAS12cap-M1343
66.433	2.364	GCAC+**T**CTTGCCTACCC+**A**ATC	KRAS12cap-M766
66.433	2.364	GCACTCTTGCCTACCC+**A**+**A**TC	KRAS12cap-M1338
66.4308	2.361	GCACTCTTGCCT+**A**CCCAA+**T**C	KRAS12cap-M1285
66.4265	2.357	GCACTCTT+**G**CCTACCC+**A**ATC	KRAS12cap-M1110
66.42	2.351	+**G**CACTCTTGCCTACCC+**A**ATC	KRAS12cap-M182
66.3528	2.283	GC+**A**CTCTTGCCTACCC+**A**ATC	KRAS12cap-M508
66.3528	2.283	GCACTCTTGCCTACC+**C**AATC	KRAS12cap-M1326
66.3441	2.275	GCACTCT+**T**GCCTACCC+**A**ATC	KRAS12cap-M1043
66.2812	2.212	GCACTCTTGCCTACCC+**A**A+**T**C	KRAS12cap-M1341
66.1268	2.057	GCACTCTTGCC+**T**ACCCAATC	KRAS12cap-M1222
66.1115	2.042	GCACTCTTGCCT+**A**CCC+**A**ATC	KRAS12cap-M1278
66.0788	2.009	GCACTCTTGCCTACCCA+**A**+**T**C	KRAS12cap-M1345
66.057	1.988	GCACT+**C**TTGCCTACCCAATC	KRAS12cap-M776
65.7641	1.695	GCACTC+**T**TGCCTACCCAATC	KRAS12cap-M882
65.729	1.660	GCACTCTTGC+**C**TACCCAATC	KRAS12cap-M1176
65.538	1.469	GCACTCTTGCCTACCCAAT+**C**	KRAS12cap-M1350
65.527	1.458	GCACTCTT+**G**CCTACCCAATC	KRAS12cap-M1053
65.516	1.447	+**G**CACTCTTGCCTACCCAATC	KRAS12cap-M1
65.5027	1.433	GCAC+**T**CTTGCCTACCCAATC	KRAS12cap-M655

Continued on next page

Table S9 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
65.4785	1.409	GCACTCTTGCCTAC+**C**CAATC	KRAS12cap-M1310
65.419	1.350	GC+**A**CTCTTGCCTACCCAATC	KRAS12cap-M364
65.4168	1.347	GCACTCTTGCCTACCCA+**A**TC	KRAS12cap-M1344
65.4146	1.345	GCACTCT+**T**GCCTACCCAATC	KRAS12cap-M974
65.3595	1.290	GCACTCTTGCCTACCCAA+**T**C	KRAS12cap-M1348
65.1804	1.111	GCACTCTTGCCT+**A**CCCAATC	KRAS12cap-M1259
65.0319	0.962	GCACTCTTGCCTACCC+**A**ATC	KRAS12cap-M1337
64.0694	0.000	GCACTCTTGCCTACCCAATC	KRAS12cap-CTRL

Table S10: Predicted temperatures for *KRAS13D* capture LNA modified probe at medium salt concentration using the calculated parameters at high salt concentration from EE-DL minimization and applying salt correction for DNA[5] sequences. LNA modifications are preceded by a plus sign and marked in **bold**. A cut-off was made in the tables to show only the 30 sequences with the highest and lowest temperature difference between the modified and non-modified probe, the total number of predicted sequences was 1351. As comparative, we also show the ΔT_m of each modified sequence and its non-modified analog. The chosen probe is highlighted in yellow.

T_m	ΔT_m	Probe (5' → 3')	ID
64.1551	0	GCACTCTTGCCTACGCATTC	KRAS13cap-CTRL
70.8901	6.735	GCA+**C**TCTTGCC+**T**ACG+**C**ATTC	KRAS13cap-M614
70.8861	6.731	G+**C**A+**C**TCTTGCC+**T**ACGCATTC	KRAS13cap-M219
70.7793	6.624	G+**C**ACTCTTGCCTA+**C**GCA+**T**TC	KRAS13cap-M340
70.7309	6.576	GCACTCTTGCCTA+**C**G+**C**A+**T**TC	KRAS13cap-M1297
70.7067	6.552	GCA+**C**TCTTGCC+**T**A+**C**GCATTC	KRAS13cap-M612
70.6663	6.511	GCA+**C**T+**C**TTGCCTACG+**C**ATTC	KRAS13cap-M545
70.6481	6.493	G+**C**A+**C**T+**C**TTGCCTACGCATTC	KRAS13cap-M213
70.6158	6.461	GCA+**C**T+**C**TTGCCTA+**C**GCATTC	KRAS13cap-M543
70.5227	6.368	GCACTCTTGCC+**T**A+**C**GCA+**T**TC	KRAS13cap-M1235
70.464	6.309	G+**C**A+**C**TCTTGCCTACGCA+**T**TC	KRAS13cap-M225
70.4579	6.303	GCACT+**C**TTGCCTA+**C**GCA+**T**TC	KRAS13cap-M858
70.4518	6.297	GCA+**C**TCTTGCCTACG+**C**A+**T**TC	KRAS13cap-M642
70.3869	6.232	G+**C**ACTCTTGCCTACG+**C**ATT+**C**	KRAS13cap-M353
70.3727	6.218	GCA+**C**TCTTGCCTA+**C**GCA+**T**TC	KRAS13cap-M631
70.3605	6.205	G+**C**ACTCTTGCCTA+**C**GCATT+**C**	KRAS13cap-M342
70.3342	6.179	GCACTCTTGCCTA+**C**G+**C**ATT+**C**	KRAS13cap-M1299
70.3281	6.173	GCA+**C**TC+**T**TGCCTACG+**C**ATTC	KRAS13cap-M559
70.3078	6.153	G+**C**A+**C**TC+**T**TGCCTACGCATTC	KRAS13cap-M214
70.2996	6.144	G+**C**ACTC+**T**TGCCTACGCA+**T**TC	KRAS13cap-M270
70.2895	6.134	GCA+**C**T+**C**TTGCC+**T**ACGCATTC	KRAS13cap-M541
70.2854	6.130	G+**C**AC+**T**CTTGCCTA+**C**GCATTC	KRAS13cap-M237

Continued on next page

Table S10 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
70.2773	6.122	G+**C**AC+**T**CTTGCCTACG+**C**ATTC	KRAS13cap-M239
70.2712	6.116	GCACTC+**T**TGCCTACG+**C**A+**T**TC	KRAS13cap-M961
70.261	6.106	GCA+**C**TC+**T**TGCC+**T**ACGCATTC	KRAS13cap-M555
70.2468	6.092	GCA+**C**TC+**T**TGCCTA+**C**GCATTC	KRAS13cap-M557
70.2102	6.055	GCA+**C**TCTTGCC+**T**ACGCA+**T**TC	KRAS13cap-M616
70.1959	6.041	GCACTCT+**T**GCC+**T**A+**C**GCATTC	KRAS13cap-M1010
68.8296	4.674	GCACTCTTGCCTA+**C**GCA+**T**TC	KRAS13cap-M1304
66.4698	2.315	GC+**A**CTCTTGCCTACGC+**A**TTC	KRAS13cap-M508
66.4612	2.306	GCACTCT+**T**GCCTACGC+**A**TTC	KRAS13cap-M1043
66.4503	2.295	GCACTCTTGCCTAC+**G**CAT+**T**C	KRAS13cap-M1323
66.3051	2.150	GCAC+**T**CTTGCCTACGCAT+**T**C	KRAS13cap-M773
66.2703	2.115	GCACTCTT+**G**CCTACGCAT+**T**C	KRAS13cap-M1117
66.2616	2.106	+**G**CACTCTTGCCTACGCAT+**T**C	KRAS13cap-M189
66.2268	2.072	GCACTCTTGCCT+**A**CGC+**A**TTC	KRAS13cap-M1278
66.2247	2.070	GC+**A**CTCTTGCCTACGCAT+**T**C	KRAS13cap-M515
66.2181	2.063	GCACTCT+**T**GCCTACGCAT+**T**C	KRAS13cap-M1050
66.2073	2.052	GCACTCTTGCC+**T**ACGCATTC	KRAS13cap-M1222
66.142	1.987	GCACT+**C**TTGCCTACGCATTC	KRAS13cap-M776
66.0745	1.919	GCACTCTTGCCTACGC+**A**+**T**TC	KRAS13cap-M1338
65.9828	1.828	GCACTCTTGCCT+**A**CGCAT+**T**C	KRAS13cap-M1285
65.8714	1.716	GCACTCTTGCCTACGC+**A**T+**T**C	KRAS13cap-M1341
65.8451	1.690	GCACTC+**T**TGCCTACGCATTC	KRAS13cap-M882
65.8079	1.653	GCACTCTTGC+**C**TACGCATTC	KRAS13cap-M1176
65.6983	1.543	GCACTCTTGCCTACGCA+**T**TC	KRAS13cap-M1344
65.6764	1.521	GCACTCTTGCCTAC+**G**CATTC	KRAS13cap-M1310
65.6215	1.466	GCACTCTTGCCTACGCATT+**C**	KRAS13cap-M1350

Continued on next page

Table S10 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
65.5995	1.444	+**G**CACTCTTGCCTACGCATTC	KRAS13cap-M1
65.5863	1.431	GCAC+**T**CTTGCCTACGCATTC	KRAS13cap-M655
65.5027	1.348	GC+**A**CTCTTGCCTACGCATTC	KRAS13cap-M364
65.5005	1.345	GCACTCT+**T**GCCTACGCATTC	KRAS13cap-M974
65.4939	1.339	GCACTCTT+**G**CCTACGCATTC	KRAS13cap-M1053
65.2623	1.107	GCACTCTTGCCT+**A**CGCATTC	KRAS13cap-M1259
65.1517	0.997	GCACTCTTGCCTACGC+**A**TTC	KRAS13cap-M1337
64.8986	0.743	GCACTCTTGCCTACGCAT+**T**C	KRAS13cap-M1348
64.1551	0.000	GCACTCTTGCCTACGCATTC	KRAS13cap-CTRL

Table S11: Predicted temperatures for *KRAS* linker LNA modified probe at medium salt concentration using the calculated parameters at high salt concentration from EE-DL minimization and applying salt correction for DNA[5] sequences. LNA modifications are preceded by a plus sign and marked in **bold**. A cut-off was made in the tables to show only the 30 sequences with the highest and lowest temperature difference between the modified and non-modified probe, the total number of predicted sequences was 834. As comparative, we also show the ΔT_m of each modified sequence and its non-modified analog. The chosen probe is highlighted in yellow.

T_m	ΔT_m	Probe (5' → 3')	ID
48.1502	0	TGAAGTCACATTATATA	KRASlin-CTRL
58.8304	10.680	TGAAGT+**C**A+**C**ATTATA+**T**A	KRASlin-M621
58.8152	10.665	TGAAGT+**C**A+**C**ATTA+**T**ATA	KRASlin-M619
58.544	10.394	TGAAGTCA+**C**ATTA+**T**A+**T**A	KRASlin-M734
58.3759	10.226	TGAA+**G**T+**C**A+**C**ATTATATA	KRASlin-M472
58.3636	10.213	TGAAG+**T**+**C**A+**C**ATTATATA	KRASlin-M539
58.333	10.183	TGA+**A**GT+**C**A+**C**ATTATATA	KRASlin-M393
58.2932	10.143	TGAA+**G**TCA+**C**ATTATA+**T**A	KRASlin-M498
58.2778	10.128	TGAA+**G**TCA+**C**ATTA+**T**ATA	KRASlin-M496
58.2594	10.109	TGA+**A**GTCA+**C**ATTATA+**T**A	KRASlin-M419
58.241	10.091	TGA+**A**GTCA+**C**ATTA+**T**ATA	KRASlin-M417
58.2226	10.072	TGAAGTCA+**C**A+**T**TATA+**T**A	KRASlin-M719
58.1981	10.048	TGAAGTCA+**C**A+**T**TA+**T**ATA	KRASlin-M717
58.1705	10.020	TGAAGT+**C**A+**C**A+**T**TATATA	KRASlin-M616
58.109	9.959	TGAAGTCA+**C**A+**T**+**T**ATATA	KRASlin-M715
58.0936	9.943	TG+**A**AGTCA+**C**ATTATA+**T**A	KRASlin-M327
58.0813	9.931	TG+**A**AGTCA+**C**ATTA+**T**ATA	KRASlin-M325
58.0383	9.888	TGAAGTCA+**C**AT+**T**ATA+**T**A	KRASlin-M725
58.0167	9.867	TGAAGTCA+**C**AT+**T**A+**T**ATA	KRASlin-M723
57.9551	9.805	TG+**A**AGT+**C**A+**C**ATTATATA	KRASlin-M301
57.9181	9.768	TGA+**A**+**G**TCA+**C**ATTATATA	KRASlin-M370
57.912	9.762	TGA+**A**GTCA+**C**A+**T**TATATA	KRASlin-M414

Continued on next page

Table S11 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
57.8781	9.728	TG+**A**+**A**GTCA+**C**ATTATATA	KRASlin-M265
57.8565	9.706	TGAAGT+**C**A+**C**ATTAT+**A**TA	KRASlin-M620
57.8503	9.700	TGAAGT+**C**A+**C**ATT+**A**TATA	KRASlin-M618
57.7484	9.598	TGAAGT+**C**A+**C**AT+**T**ATATA	KRASlin-M617
57.7267	9.577	TGA+**A**GTCA+**C**AT+**T**ATATA	KRASlin-M415
57.6989	9.549	TGAA+**G**TCA+**C**A+**T**TATATA	KRASlin-M493
51.1563	3.006	TGAAGTCACATT+**A**T+**A**T+**A**	KRASlin-M816
51.053	2.903	TGAAG+**T**CACATT+**A**TATA	KRASlin-M588
51.053	2.903	TGAAG+**T**CACATTAT+**A**TA	KRASlin-M597
51.0012	2.851	TGAAGTCACAT+**T**ATAT+**A**	KRASlin-M808
50.9529	2.803	TGAAGTC+**A**CATTATAT+**A**	KRASlin-M704
50.908	2.758	TGAAGTCAC+**A**TTAT+**A**TA	KRASlin-M765
50.9045	2.754	TGAAGTCAC+**A**TT+**A**TATA	KRASlin-M756
50.8319	2.682	TGAA+**G**TCACATTATATA	KRASlin-M457
50.8181	2.668	TGA+**A**GTCACATTATATA	KRASlin-M365
50.7765	2.626	+**T**GAAGTCACATTATAT+**A**	KRASlin-M137
50.7315	2.581	T+**G**AAGTCACATTATAT+**A**	KRASlin-M258
50.61	2.460	TG+**A**AGTCACATTATATA	KRASlin-M259
50.5996	2.449	TGAAGTCACA+**T**TATATA	KRASlin-M771
50.5474	2.397	TGAAG+**T**CACATTATAT+**A**	KRASlin-M602
50.5266	2.376	TGAAGTCACATT+**A**T+**A**TA	KRASlin-M814
50.4012	2.251	TGAAGTCAC+**A**TTATAT+**A**	KRASlin-M770
50.3593	2.209	TGAAGTCACAT+**T**ATATA	KRASlin-M793
50.307	2.157	TGAAGTC+**A**CATTATATA	KRASlin-M659
50.0725	1.922	T+**G**AAGTCACATTATATA	KRASlin-M138
50.0199	1.870	TGAAGTCACATT+**A**TAT+**A**	KRASlin-M819

Continued on next page

Table S11 – *Continued from previous page*

T_m	ΔT_m	Probe (5' → 3')	ID
50.0164	1.866	TGAAGTCACATTAT+**A**T+**A**	KRASlin-M830
50.0093	1.859	+**T**GAAGTCACATTATATA	KRASlin-M1
49.8899	1.740	TGAAG+**T**CACATTATATA	KRASlin-M536
49.7455	1.595	TGAAGTCAC+**A**TTATATA	KRASlin-M742
49.3597	1.209	TGAAGTCACATT+**A**TATA	KRASlin-M809
49.3597	1.209	TGAAGTCACATTAT+**A**TA	KRASlin-M827
48.8353	0.685	TGAAGTCACATTATAT+**A**	KRASlin-M833
48.1502	0.000	TGAAGTCACATTATATA	KRASlin-CTRL

Table S12: MALDI-MS of purified oligonucleotides and their respective purities. LNA modifications are preceded by a plus sign and marked in **bold**. The 'target-binding, measured probe, is marked in red.

Kas code	Name	Sequences (5'→ 3')	MW [g/mol]	MS [m/z]	Purity [%]
KAS209	BRAFcap	TTTTTTTTTTCCCCCCCCCCATCGAGAT+TTCT+CTGTAG+CTA	12457.1	12455.9	90.6
KAS210	EGFRcap	TTTTTTTTTTCCCCCCCCCCGAG+AAAAAGTTTCTCA+TGTA+CAGT	13431.7	13432.8	98.3
KAS211	KRAS12cap	TTTTTTTTTTCCCCCCCCCCGCACTCTTGCCTACCCA+ATC	11934.7	11933.8	99.6
KAS212	KRAS13cap	TTTTTTTTTT CCCCCCCCCCGCACTCTTGCCTA+CGCA+TTC	12007.8	12008.3	99.2
KAS213	BRAFlink	GA+TGG+GAATA+CCAGAC+CAC+CTGTTTTCAA+CTGTT+CAAA+CTGAT	13493.9	13493.0	97.8
KAS214	EGFRlink	GA+TGG+GAATA+CCAGAC+CAC+CTGTTTTTGAAGT+CA+CA+TTATATA	13518.9	13519.3	95.2
KAS215	KRASlink	GA+TGG+GAATA+CCAGAC+CAC+CTGTTTTTTG+TTGGAT+CATA+TTCGT	13823.0	13822.4	98.3

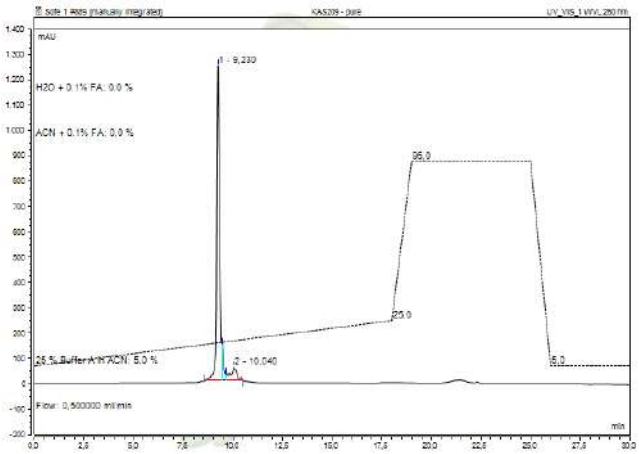


Figure S1: Representative UHPLC traces of *BRAF* capture probe.

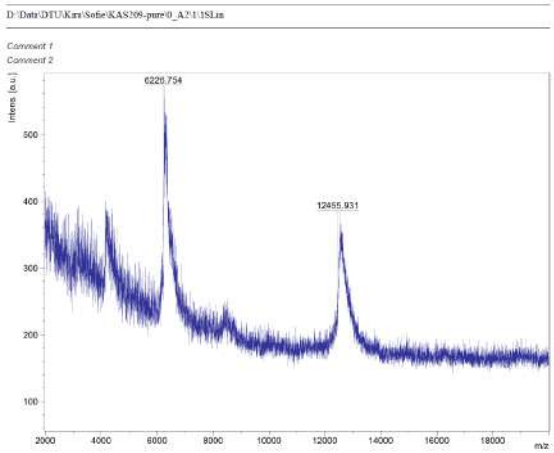


Figure S2: Representative MALDI-MS spectrum of *BRAF* capture probe.

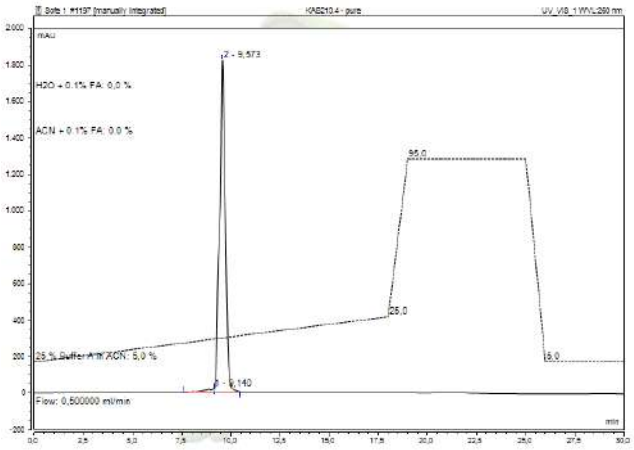


Figure S3: Representative UHPLC traces of *EGFR* capture probe.

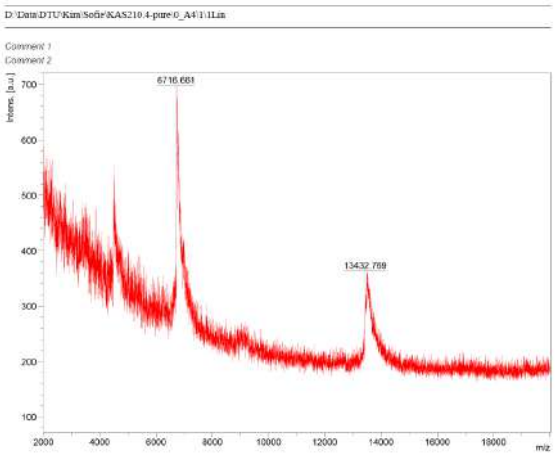


Figure S4: Representative MALDI-MS spectrum of *EGFR* capture probe.

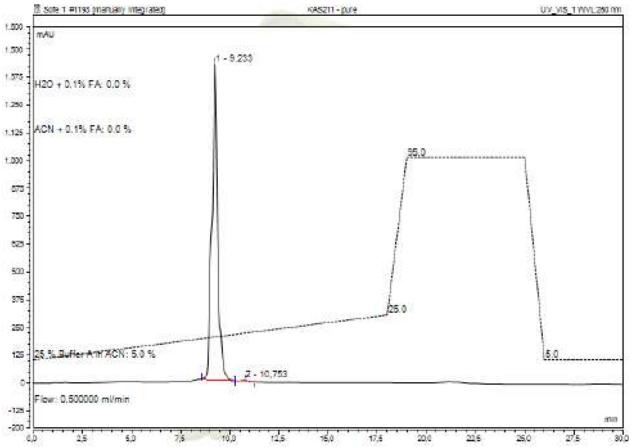


Figure S5: Representative UHPLC traces of *KRAS12D* capture probe.

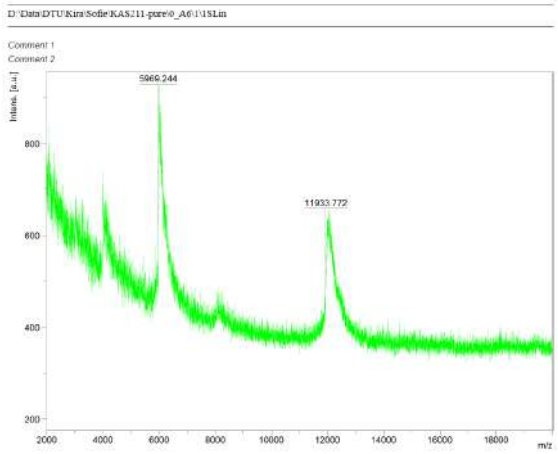


Figure S6: Representative MALDI-MS spectrum of *KRAS12D* capture probe.

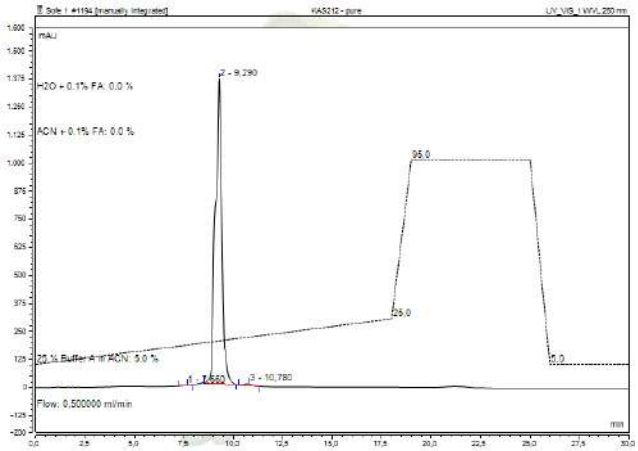


Figure S7: Representative UHPLC traces of *KRAS13D* capture probe.

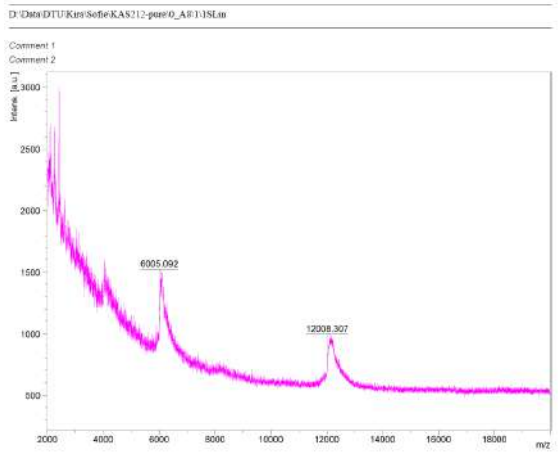


Figure S8: Representative MALDI-MS spectrum of *KRAS13D* capture probe.

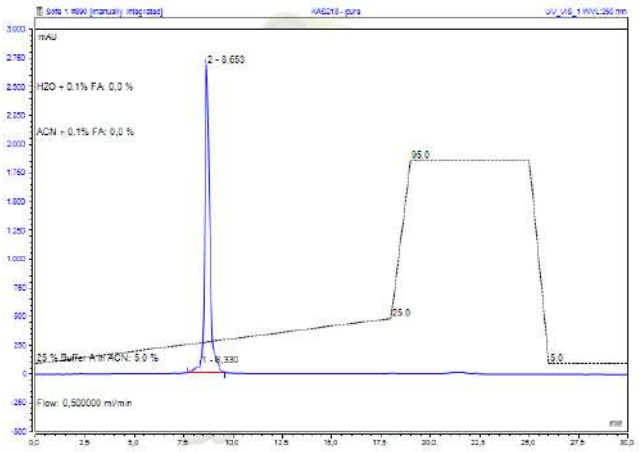


Figure S9: Representative UHPLC traces of *BRAF* linker probe.

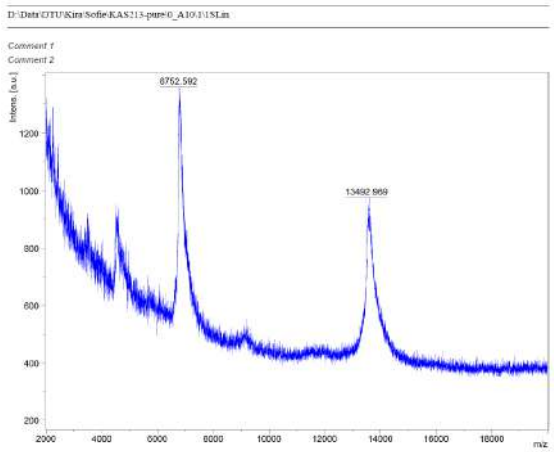


Figure S10: Representative MALDI-MS spectrum of *BRAF* linker probe.

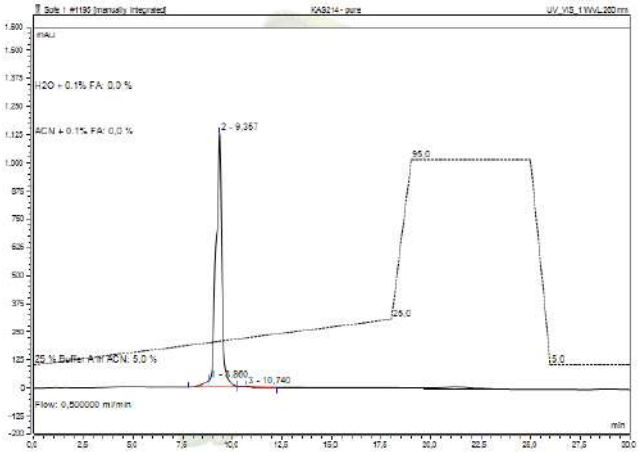


Figure S11: Representative UHPLC traces of *EGFR* linker probe.

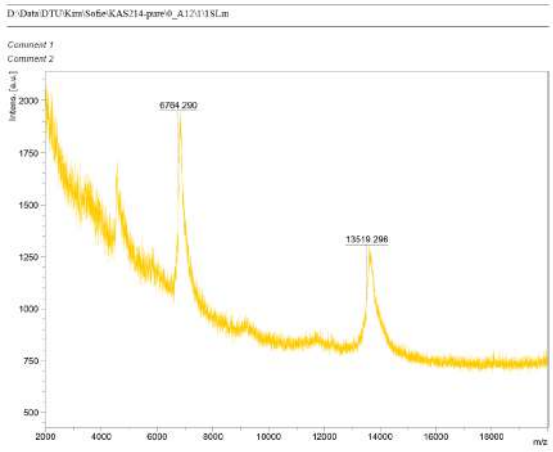


Figure S12: Representative MALDI-MS spectrum of *EGFR* linker probe.

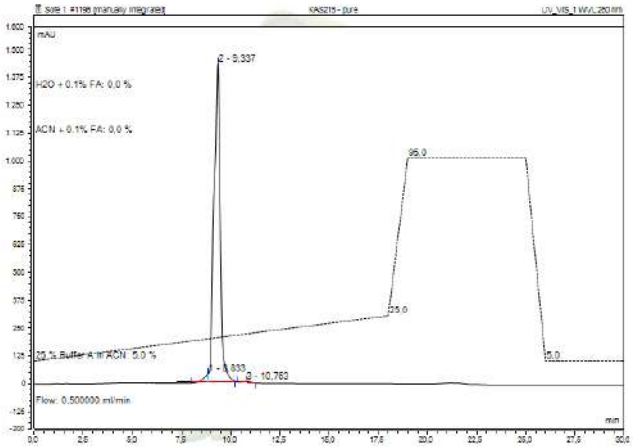


Figure S13: Representative UHPLC traces of *KRAS* linker probe.

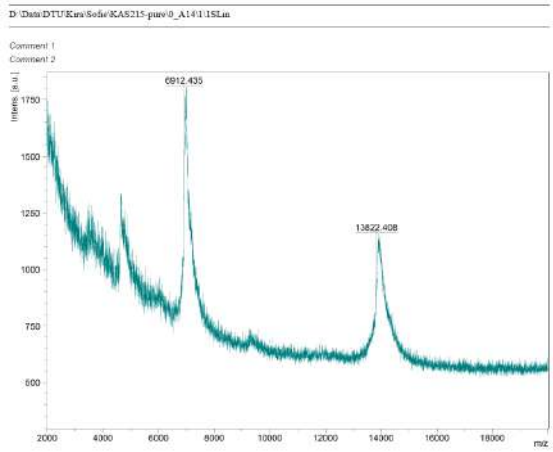


Figure S14: Representative MALDI-MS spectrum of *KRAS* linker probe.

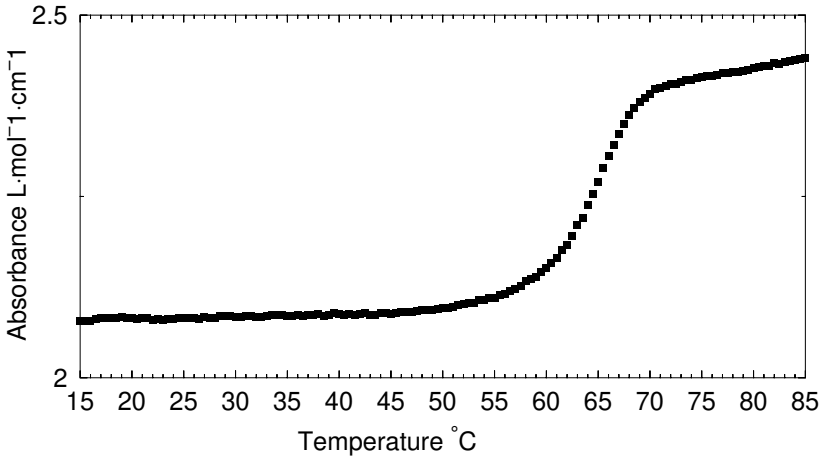


Figure S15: Representative T_m curve of *BRAF* capture probe.

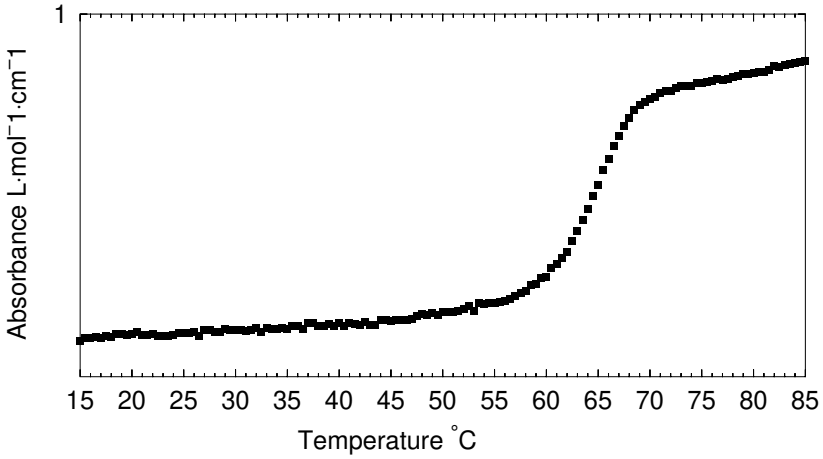


Figure S16: Representative T_m curve of *KRAS12D* capture probe.

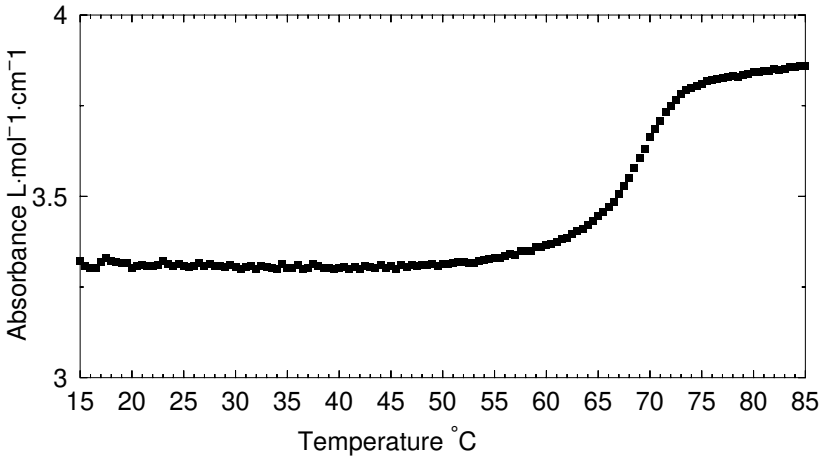


Figure S17: Representative T_m curve of *KRAS13D* capture probe.

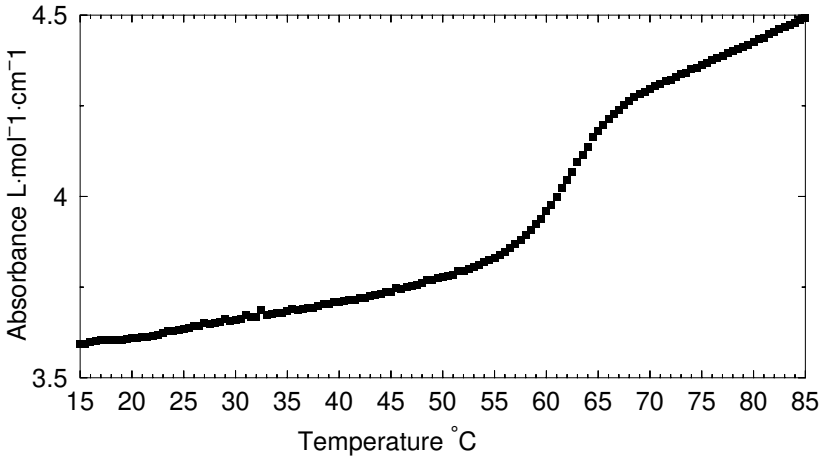


Figure S18: Representative T_m curve of *BRAF* linker probe.

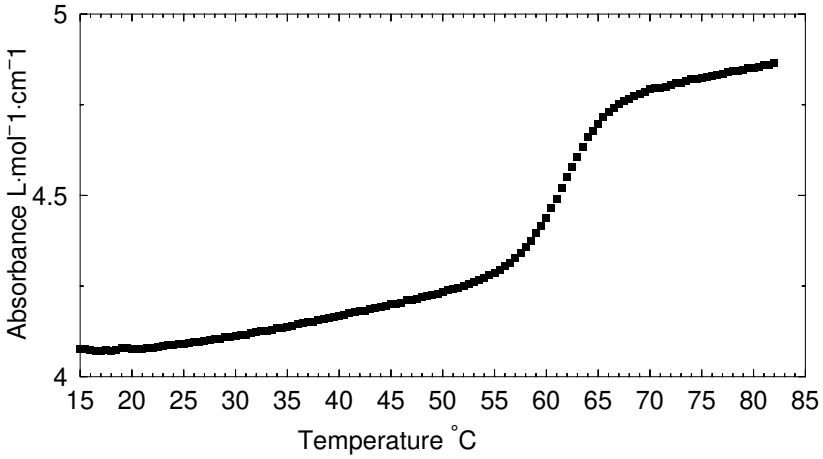


Figure S19: Representative T_m curve of *EGFR* linker probe.

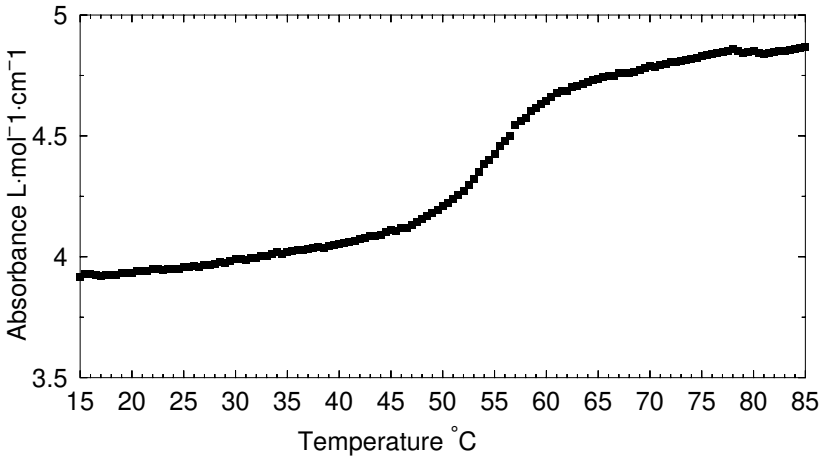


Figure S20: Representative T_m curve of *KRAS* linker probe.

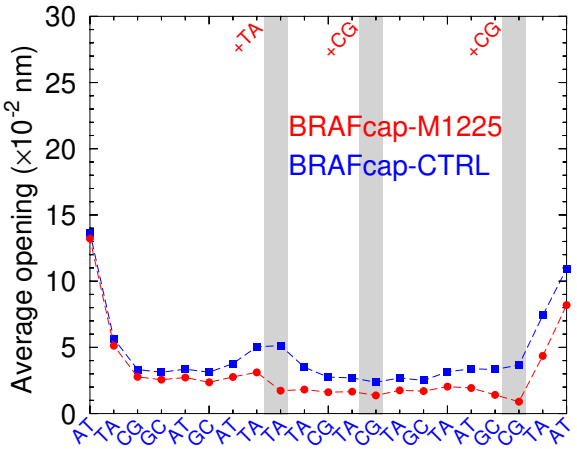


Figure S21: Average opening profile for *BRAF* capture probe. Modified and non-modified probe are shown in red bullets and blue squares, respectively. LNA modifications are preceded by a plus sign and marked in **bold**. The location of the LNAs is indicated by the gray shaded area. Calculation was carried out at 220 K which has no relation to the melting temperature.

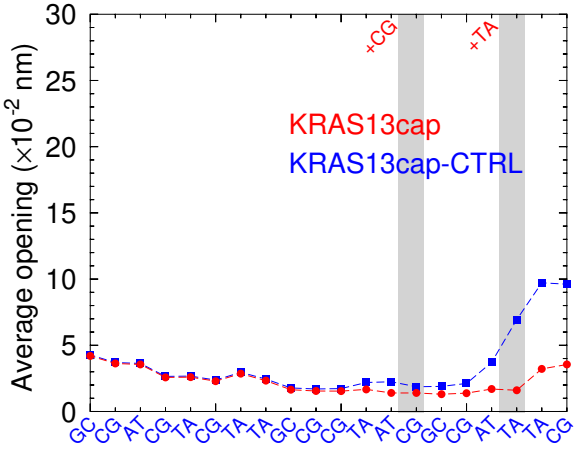


Figure S22: Average opening profile for *KRAS13D* capture probe. Modified and non-modified probe are shown in red bullets and blue squares, respectively. LNA modifications are preceded by a plus sign and marked in **bold**. The location of the LNAs is indicated by the gray shaded area. Calculation was carried out at 220 K which has no relation to the melting temperature.

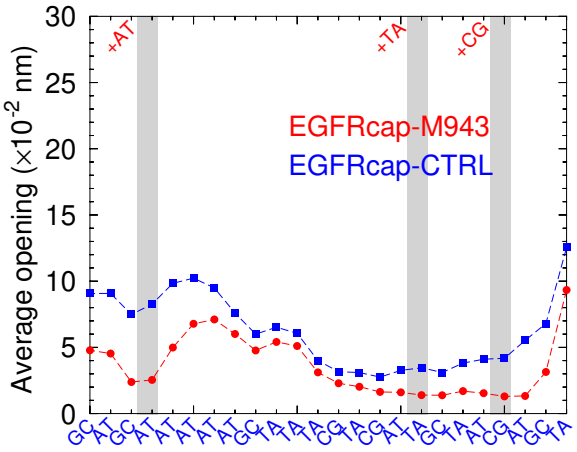


Figure S23: Average opening profile for *EGFR* capture probe. Modified and non-modified probe are shown in red bullets and blue squares, respectively. LNA modifications are preceded by a plus sign and marked in **bold**. The location of the LNAs is indicated by the gray shaded area. Calculation was carried out at 220 K which has no relation to the melting temperature.

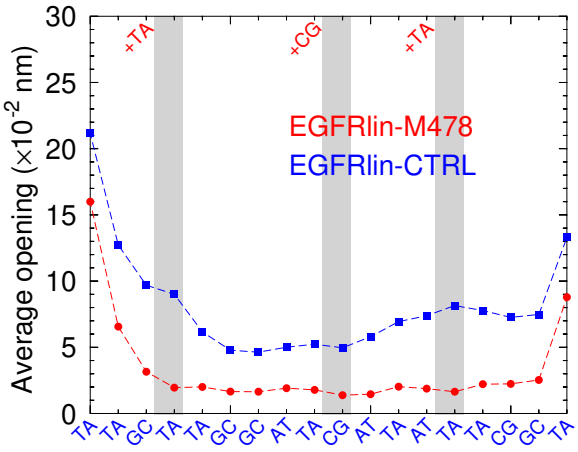


Figure S24: Average opening profile for *EGFR* linker probe. Modified and non-modified probe are shown in red bullets and blue squares, respectively. LNA modifications are preceded by a plus sign and marked in **bold**. The location of the LNAs is indicated by the gray shaded area. Calculation was carried out at 220 K which has no relation to the melting temperature.

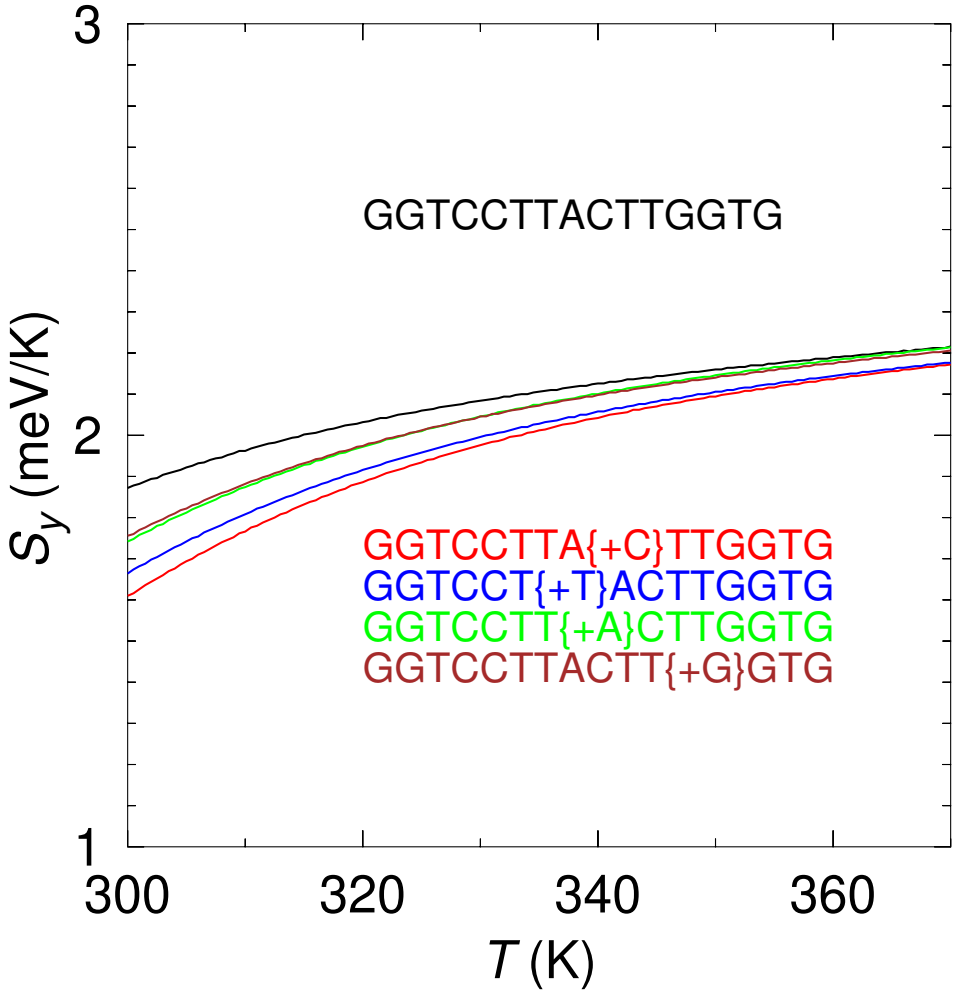


Figure S25: Calculated configurational entropy S_y comparing an unmodified DNA sequence (black curve) to LNA modifications +C (red), +T (blue) +A (green) and +G (brown). The entropy is calculated from the first derivative of the Helmholtz free energy $S_y = -\partial F_y/\partial T$, which in turn is obtained from $F_y = -k_B T \ln Z_y$, where Z_y is the configurational partition function Eq. (2) of the main text.

References

[1] Patricia M McTigue, Raymond J Peterson, and Jason D Kahn. Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation. *Biochem.*, 43(18):5388–5405, 2004. doi: 10.1021/jp073198j.

[2] Yong You, Bernardo G Moreira, Mark A Behlke, and Richard Owczarzy. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res.*, 34(8):e60, 2006. doi: 10.1093/nar/gkl175.

[3] Kareem Fakhfakh, Olivia Marais, Xin Bo Justin Cheng, Jorge Real Castañeda, Curtis B Hughesman, and Charles Haynes. Molecular thermodynamics of LNA:LNA base pairs and the hyperstabilizing effect of 5′-proximal LNA:DNA base pairs. *AIChE J.*, 61(9):2711–2731, 2015. doi: 10.1002/aic.14916.

[4] Curtis B Hughesman, Robin FB Turner, and Charles A Haynes. Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified LNA. *Biochem.*, 50(23):5354–5368, 2011. doi: 10.1021/bi200223s.

[5] Richard Owczarzy, Yong You, Bernardo G. Moreira, Jeffrey A. Manthey, Lingyan Huang, Mark A. Behlke, and Joseph A. Walder. Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochem.*, 43:3537–3554, 2004. doi: 10.1021/bi034621r.