UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

MARIA IZABEL MUNIZ FOSCARINI

Mesoscopic model applied to DNA/TNA hybrids and DNA in solutions containing Mg^{2+}

BELO HORIZONTE
2022

Maria Izabel Muniz Foscarini

Mesoscopic model applied to DNA/TNA hybrids and DNA in solutions containing Mg^{2+}

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

Supervisor: Gerald Weber

Belo Horizonte

2022

Dados Internacionais de Catalogação na Publicação (CIP)

F747m Foscarini, Maria Izabel Muniz.
Mesoscopic model applied to DNA/TNA hybrids and DNA in solutions containing Mg2+ / Maria Izabel Muniz Foscarini. – 2022.
134f. : il.

Orientador: Gerald Weber.
Tese (doutorado) – Universidade Federal de Minas Gerais, Departamento de Física.
Bibliografia: f. 78-99.

1. Ácidos nucleicos. 2. DNA. 3. Biofísica. I. Título. II. Weber, Gerald. III. Universidade Federal de Minas Gerais, Departamento de Física.

CDU – 577.3 (043)

Ficha catalográfica elaborada por Romário Martins Ribeiro – CRB6 3595
Biblioteca Professor Manoel Lopes de Siqueira – Departamento de Física - UFMG

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

ATA DE DEFESA DE TESE

ATA DA SESSÃO DE ARGUIÇÃO DA 407ª TESE DO PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA, DEFENDIDA POR MARIA IZABEL MUNIZ FOSCARINI orientada pelo professor Gerald Weber, para obtenção do grau de **DOUTORA EM CIÊNCIAS, área de concentração Física**. Às 9 horas de dois de dezembro de dois mil e vinte e dois reuniu-se, por videoconferência, a Comissão Examinadora, composta pelos professores **Gerald Weber** (Orientador - Departamento de Física/UFMG), **Lívia Siman Gomes** (Departamento de Física/UFMG), **Mariana Torquato Quezado de Magalhães** (Instituto de Ciências Biológicas/UFMG), **Luciana Magalhães Rebêlo Alencar** (Departamento de Física/UFMA) e **Yraima Cordeiro** (Faculdade de Farmácia/UFRJ) para dar cumprimento ao Artigo 37 do Regimento Geral da UFMG, submetendo a Mestre **MARIA IZABEL MUNIZ FOSCARINI** à arguição de seu trabalho de Tese de Doutorado, que recebeu o título de **"Mesoscopic model applied to DNA/TNA hybrids and DNA in solutions containing Mg2+"**. A candidata fez uma exposição oral de seu trabalho durante aproximadamente 50 minutos. Após esta, os membros da comissão prosseguiram com a sua arguição, e apresentaram seus pareceres individuais sobre o trabalho, concluindo pela aprovação da candidata.

Belo Horizonte, 02 de dezembro de 2022.

Prof. Gerald Weber
Orientador da estudante
Departamento de Física/UFMG

Profa. Luciana Magalhães Rebêlo Alencar
DF/UFMA

Profa. Lívia Siman Gomes
Departamento de Física/UFMG

Profa. Yraima Cordeiro
Faculdade de Farmácia/UFRJ

Profa. Mariana Torquato Quezado de Magalhães
ICB/UFMG

Candidata: Maria Izabel Muniz Foscarini

Documento assinado eletronicamente por **Gerald Weber**, **Professor do Magistério Superior**, em 05/12/2022, às 10:45, 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 05/12/2022, às 11:59, 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 **Maria Izabel Muniz Foscarini**, **Usuário Externo**, em 05/12/2022, às 12:35, 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 **Livia Siman Gomes**, **Professora do Magistério Superior**, em 05/12/2022, às 14:00, 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 Magalhães Rebêlo Alencar**, **Usuária Externa**, em 07/12/2022, às 09:57, 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 **Yraima Moura Lopes Cordeiro**, **Usuária Externa**, em 07/12/2022, às 12:00, 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 **1939198** e o código CRC **CB0021AD**.

Referência: Processo nº 23072.270911/2022-78 SEI nº 1939198

Acknowledgements

To my father, Ailton, I thank for having supported me in good and bad times, and for having instructed me to endure the difficulties that life imposes.

To my grandparents, Antônio and Antônia, who I unfortunately did not have the opportunity to meet, but who I know are always looking out for me, and Aristóteles and Aparecida, for their love, support and for never giving up on me, you are greatly missed here.

To my husband, João Pedro, I thank you for always believing in me and showing me the great potential that I have.

I thank my friends, Bárbara and Marina, for always supporting me and being by my side, and also thank the friends that I made in UFMG, who have helped me a lot during this long journey.

To my colleagues of Computational Biophysics Laboratory, thank you for the discussions and the great times.

I thank Kira Astakhova from Technical University of Denmark for synthesizing new DNA sequences, and Hershel H. Lackey from University of Utah and Jennifer M. Heemstra from Emory University for providing their unpublished melting temperatures for TNA.

I thank Ana Maria, my first advisor on my academic life, whose teachings I will always take with me.

I thank my dear advisor, Gerald Weber, who had (a lot of) patience with my quirks. You inspire me and give me the strength to withstand all the pressure that the profession requires.

I thank the agencies that have provided me financial support: *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) and *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* (Fapemig).

"We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained."

Marie Curie

Resumo

Quando a dupla hélice dos ácidos nucleicos é influenciada por condições da solução, tais como um aumento na temperatura da solução, mudança no pH ou agentes químicos (formamida, etilenoglicol e ureia, por exemplo), essa estrutura estável pode se desenovelar devido à ruptura das ligações de hidrogênio entre os pares de bases, levando a duas fitas simples separadas. Esse fenômeno é chamado de desnaturação e pode ser descrito por um modelo físico estatístico e mesoscópico chamado modelo Peyrard-Bishop (PB). O modelo PB considera dois potenciais de interação: Morse e harmônico. O potencial de Morse é a conexão das ligações de hidrogênio em duas bases em fitas opostas formando pares de bases, e o potencial harmônico como a interação de empilhamento entre dois pares de bases vizinhos. Uma das vantagens desse modelo é que ele pode ser usado para determinar a intensidade das ligações de hidrogênio e das interações de empilhamento usando as temperaturas de desnaturação como entrada experimental. Aqui nós usamos essa aproximação para dois projetos, em um deles nós estudamos um ácido nucleico modificado e no outro investigamos os efeitos dos cátions divalentes na desnaturação do DNA.

No primeiro projeto nós usamos as temperaturas de desnaturação medidas para obter uma estimativa das ligações de hidrogênio e interações de empilhamento do ácido nucleico modificado chamado de ácido nucleico de treose (TNA) (ácido nucleico de α-L-(3′-2′)-treofuranosil) que tem uma espinha dorsal com um átomo a menos que o DNA e RNA, mas ainda assim pode formar duplexos antiparalelos. O TNA é de interesse prático, pois é resistente à degradação de nucleases e, portanto, de interesse para aplicações biotecnológicas. Os resultados indicam que os híbridos de TNA/DNA compartilham várias similaridades com RNA/DNA, como a tendência para formar hélices do tipo A e uma forte dependência de suas propriedades termodinâmicas na razão purina/pirimidina. Além disso, para os pares de base AT no DNA/TNA as forças das ligações de hidrogênio são quase idênticas as suas contrapartes em RNA/DNA, mas surpreendentemente o CG acabou sendo muito mais fraco, apesar da estabilidade semelhante. Por outro lado, as interações de empilhamento foram mais fortes para DNA/TNA do que para DNA/RNA.

No segundo projeto, nós avaliamos com o modelo mesoscópico as temperaturas de desnaturação de sequências de DNA em soluções de Mg^{2+} e $Mg^{2+} + K^{+}$, fazendo a distinção entre os pares de base internos e externos. Além dessa distinção da posição do par de base, nós investigamos via modelo PB múltiplas concentrações de Mg^{2+} para entender como elas afetam a estabilidade do DNA. Este estudo é de interesse para diagnósticos de PCR, que dependem de reações enzimáticas que requerem Mg^{2+}. Os resultados de Mg^{2+} e $Mg^{2+} + K^{+}$ são comparados aos cálculos para Na^{+} feitos previamente, em termos da concentração de sódio equivalente e força iônica, e nós mostramos que os potenciais das ligação de

hidrogênio resultantes podem ser relacionados pela concentração de sódio equivalente de Mg^{2+}, de modo que os potenciais de Morse são essencialmente constantes e não são afetados pelos cátions. Considerando o sódio equivalente, nós encontramos que a ligação de hidrogênio resultante não muda, independentemente da valência e concentração do íon. Para interações de empilhamento, por outro lado, encontramos uma clara dependência com a força iônica e a valência do cátion. As maiores variações de força iônica, tanto para ligações de hidrogênio quanto para interações de empilhamento, foram encontradas nos terminais da sequência.

Nesses projetos usando o modelo PB, conseguimos entender a estabilidade térmica de um ácido nucleico modificado e a influência do DNA por um íon divalente. Os resultados aprofundaram nossa compreensão sobre o TNA e responderam a várias questões sobre a influência do Mg^{2+} no DNA.

Palavras-chave: Modelos teóricos para DNA, modelos mesoscópicos, desnaturação, RNA, ácidos nucleicos modificados, potencial de Morse, potencial harmônico, estabilidade térmica.

Abstract

When the double-stranded helix of nucleic acids is influenced by conditions such as increase the temperature of the solution, change pH and chemical agents (formamide, ethylene glycol and urea for example), this stable structure may unwind, due to the disruption of hydrogen bonds between base pairs, leading to two separate single strands. This phenomenon is called denaturation, and can be described by physical statistical and mesoscopic model called Peyrard-Bishop (PB) model. The PB model considers two interaction potentials: Morse and harmonic potential. The Morse potential as the connecting hydrogen bonds in two bases on opposite strands forming base pairs, and the harmonic potential as the stacking interaction between two neighboring base pairs. One of the advantages of the model is that it can be used to determine the strength of the hydrogen bonds and stacking interactions using melting temperatures as experimental input. Here, we use this approach for two projects, in one we study a modified nucleic acid and in the other we investigate the effects of divalent cations in DNA melting.

In the first project, we used measured melting temperatures to obtain an estimate of hydrogen bonds and stacking interactions of a modified nucleic acid called threose nucleic acid (TNA) (α-L-(3′-2′)-threofuranosyl nucleic acid) which has a backbone that is one atom shorter than DNA and RNA, but can still form anti-parallel duplexes. TNA is of practical interest as it is resistant to nuclease degradation and therefore of interest for biotechnological applications. Our results indicated that TNA/DNA hybrids share several similarities with RNA/DNA, such as the tendency to form A-type helices and a strong dependency of their thermodynamic properties on purine/pyrimidine ratio. Furthermore, for AT base pairs in DNA/TNA have nearly identical hydrogen bond strengths than their counterparts in RNA/DNA, but surprisingly CG turned out to be much weaker despite similar stability. On the other hand, the stacking interactions were found to be stronger for DNA/TNA than for DNA/RNA.

In the second project, we evaluate sequences of DNA melting temperatures in Mg^{2+} and $Mg^{2+} + K^{+}$ buffers with a mesoscopic model, making a distinction between internal and terminal base pairs. In addition to this distinction of the base pair position, we investigate via PB model multiple concentrations of Mg^{2+} to understand how they affect DNA stability. This study is of interest for PCR diagnostics which rely on enzymatic reactions performed that require Mg^{2+}. The Mg^{2+} and $Mg^{2+} + K^{+}$ results are compared to previous calculations for Na^{+}, in terms of equivalent sodium concentration and ionic strength, and we show that the resulting hydrogen bond potentials can be related by the equivalent sodium concentration of Mg^{2+}, so the Morse potentials are essentially constant and unaffected by cation conditions. Considering the sodium equivalence, we found that the resulting

hydrogen bond do not change, regardless of ion valence and concentration. For stacking interactions on the other hand we find a clear dependence of ionic strength and cation valence. The highest ionic strength variations, for both hydrogen bonds and stacking interactions, was found at the sequence termini.

In these projects using the PB model, we were able to understand the thermal stability of a xeno nucleic acid and the influence of a divalent ion on DNA. Our results have deepened our understanding of TNA and answered several questions regarding the influence of Mg^{2+} in DNA.

Keywords: Theoretical models for DNA, mesoscopic models, denaturation, RNA, modified nucleic acids, Morse potential, harmonic potential, thermal stability.

List of Figures

List of Tables

List of abbreviations and acronyms

DD	DNA/DNA
DFT	density functional theory
DNA	deoxyribonucleic acid
DR	DNA/RNA hybrids
dsDNA	double-stranded helix DNA
dsRNA	double-stranded helix RNA
DT	DNA/TNA hybrids
MD	molecular dynamics
miRNA	microRNA
mRNA	messenger RNA
NMR	nuclear magnetic ressonance
NN	nearest-neighbor
PB	Peyrard-Bishop
PCR	polymerase chain reaction
RNA	ribonucleic acid
RT	reverse transcriptase
TI	transfer integral operator
TNA	threose nucleic acid
XNA	xeno nucleic acid

Contents

Thesis structure

The thesis is subdivided into five main chapters, of which are, an overview of nucleic acids and theoretical model, two chapters presenting the projects, and a last conclusion chapter.

(a) Chapter 1 shows an overview of natural (DNA and RNA) and modified nucleic acids (XNA). This is followed by ionic effects and the phenomenon of thermal denaturation of nucleic acids.

(b) Chapter 2 describes some types of theoretical models based on DNA denaturation, to study thermodynamic properties and mechanics from nucleic acids. The first model is nearest-neighbor (NN) model that considers an experimental data set from van't Hoff plot to calculate enthalpy, entropy, and free energy based on neighboring base pairs. Then we discuss the Peyrard-Bishop (PB) model, a physical statistical model, that is applied only to homogeneous sequences. The second is an adaptation of PB model for heterogeneous sequences. Finally, the last model includes the previous ones, and was developed by our laboratory, with a new parameter that represents the thermodynamic equivalence.

(c) Chapter 3 presents a study of a modified nucleic acid called threose nucleic acid (TNA). We use the mesoscopic approach to investigate the thermal stability of DNA/TNA hybrids, and compare with DNA/RNA hybrids.

(d) Chapter 4 discusses the effects of internal and terminal base pairs to divalent cations Mg^{2+}, as well as mixed mono- and divalent $Mg^{2+} + K^{+}$ buffers. These studies are compared to the previous studies on Na^{+}, and in addition to the interaction potentials, we also compared the results through ionic strength and equivalent sodium concentration. Thus, the multiple types of buffers allow us to understand how the ionic strength and sodium equivalent relate to the structural aspects of DNA, especially in terms of hydrogen bonding and base pair stacking, as well as how they affect the thermal stability of DNA and terminal base pairs.

(e) The main conclusions of each project and the future perspectives are presented in Chapter 5.

1 Nucleic acids phenomenology

In this chapter, we will discuss important concepts regarding nucleic acids: its structure, properties, ionic effects and the phenomenon of thermal denaturation (where the double-stranded helix is separated, leading to two single strands). Furthermore, we will also introduce the group of modified nucleic acids, that has extensive biological application, including for diagnostics and therapies.

1.1 Structure of the nucleic acids

In 1869, a young Swiss physician Friedrich Miescher was analyzing the chemical composition of leukocytes (white blood cells) in some experiments, when he observed an unknown substance [1]. During the isolation procedure and resistance to protease digestion, Miescher noticed that the properties of this substance were not compatible with a lipid. Furthermore, the elementary composition had large amounts of phosphorous and no sulphur, unlike proteins. After carrying out some experiments, Miescher realized that he had discovered a new molecule, and as he isolated it from the nuclei of cells, he named it nuclein, currently known as deoxyribonucleic acid (DNA), a type of natural nucleic acid.

Nucleic acids are large biomolecules (biopolymers) responsible for creating, encoding and storing genetic information in the nucleus of every living cell [2]. The most common types of nucleic acids are DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). Both DNA and RNA are formed by nucleotides. These nucleotides are formed by a phosphate group PO_4^- and a nucleoside, a nitrogenous base (nucleobase) and a pentose ring (sugar), as shown in Fig. 1.

There are two groups of nitrogenous bases, purine and pyrimidine, as shown in

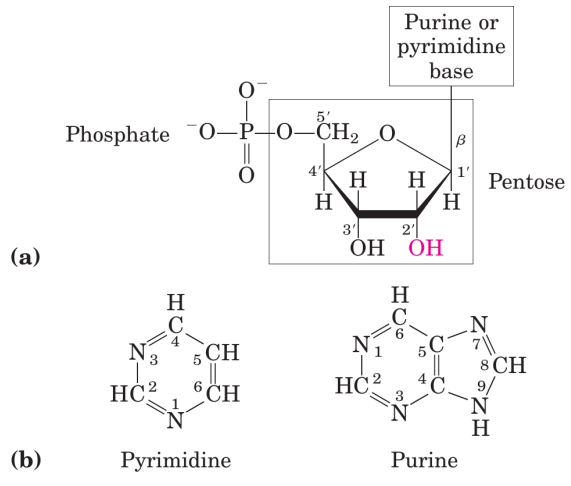

Figure 1 – Structure of nucleotides. (a) General structure of RNA containing a phosphate group, a nitrogenous base, and a pentose ring (sugar), with orientation of carbons. In DNA, the –OH group on the 2′ carbon (in pink) is replaced with –H. (b) The types of structures of nitrogenous bases. Figure taken from reference [3].

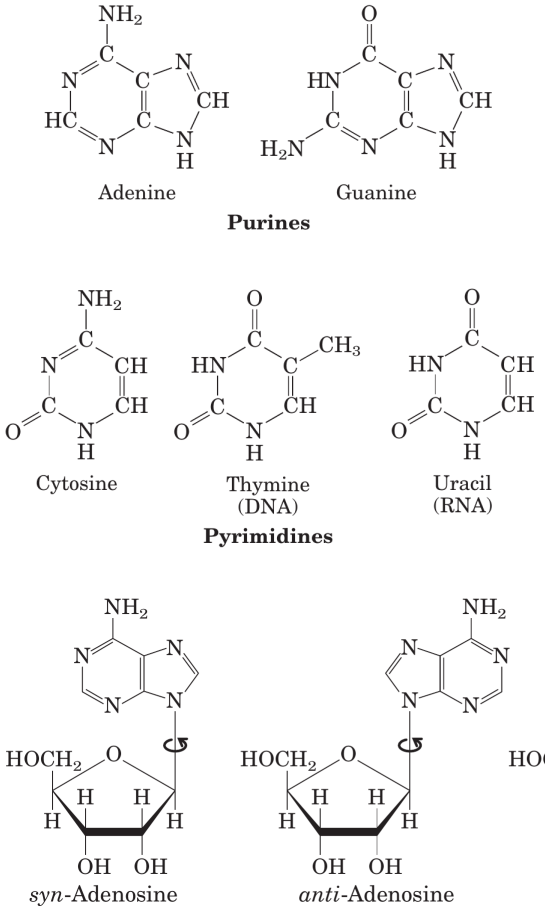

Figure 2 – The types of nitrogenous bases, purine: adenine and guanine, and pyrimidine: cytosine, uracil and thymine. Figure taken from reference [3].

Figure 3 – Conformations of nucleosides, for purine bases: *syn*-adenosine and *anti*-adenosine, and for pyrimidine bases: *anti*-cytidine. Figure taken from reference [3].

Figs. 1 and 2. The pyrimidine bases: cytosine, thymine, and uracil, are heterocyclic aromatic organic compounds that contain a single carbon-nitrogen ring [4]. Naturally occurring in DNA are the nucleobases cytosine and thymine, while in RNA the thymine is replaced by uracil. The purine bases: adenine and guanine, are heterocyclic aromatic organic compounds, consisting of a pyrimidine ring and an imidazole ring, that is, they consist of two carbon-nitrogen rings.

To understand how the nucleotides are linked to form the structure of nucleic acids, we show in Fig. 1 that the phosphate is esterified to the C-5′, and the nitrogenous bases are covalently linked with C-1′ of pentose, at N-1 of pyrimidines and N-9 of purines, by a β-glycosidic bond [3, 5]. This bond allows the free rotation of bases defined as two conformations, *syn* and *anti*, as shown in Fig. 3. The purine bases have both types, and pyrimidine bases have only *anti*, because of possible interference between the pentose and the carbonyl oxygen at C-2′ [3].

Apart from the nitrogenous bases, the nucleotides have two common types of pentoses that in DNA are defined as deoxyribose and ribose to RNA, which are both formed by five-carbon. These pentoses form an important structure with the phosphate group called backbone. Furthermore, the phosphate group for DNA and RNA is the same and has an important role of covalently linking the successive nucleotides, called

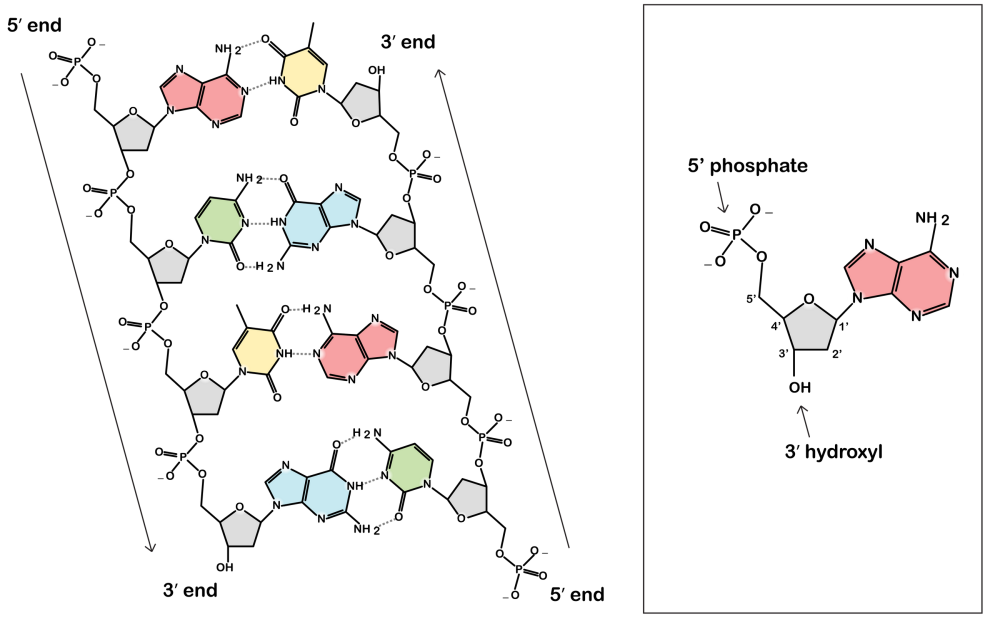

Figure 4 – The structure of DNA, with 5′ (phosphate-bearing end) to 3′ (hydroxyl-bearing end) orientation of the strands. Figure taken from reference [6].

phosphodiester linkage [3]. This occurs when the 5′-phosphate group of one nucleotide unit is joined to the 3′-hydroxyl group of the next nucleotide, as shown in Fig. 4. In DNA and RNA, all phosphodiester linkages have the same orientation, with 5′ to 3′ orientation of a strand refers to the ends of it.

In 1953 through the analysis of X-ray photographs, Linus Pauling and Robert B. Corey presented a possible structure of nucleic acids [7]. They proposed a three chain structure (triple-stranded helix) for the nucleic acids in which each chain is a helix and the dense core is formed by the phosphate groups, where these groups depend on hydrogen bonds between them. However, in the same year, Francis Crick and James Watson published an article about structure of DNA, introducing the double-stranded helix (duplex) [8], similar to the one shown in Fig. 5, different from that proposed by Pauling and Corey.

In the same edition of Nature in which Francis Crick and James Watson published the first DNA structure, we can find two other published articles to explain and prove the helical structure of DNA. The first was written by Wilkins et al. [10] in which they have explained the diffraction phenomenon using Bessel functions to describe the pattern formed in Fig. 6. Moreover, they also discussed the interpretation of the X-ray photograph. The second article was written by Rosalind E. Franklin and Raymond G. Gosling in which they found the helical structure of DNA in a sodium solution through X-ray diffraction [11], as shown in Fig. 6, in which the dark layer lines represent the base pairs in a helical configuration. This photograph is referred to as photograph 51, and it is about one of two structures of DNA, the structure B. This structure is the highly hydrated form of DNA, while the structure A is the drier form, as shown in Fig. 7.

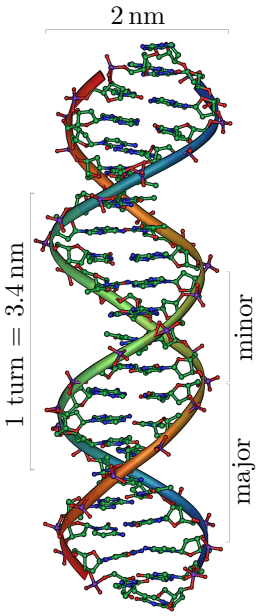

Figure 5 – Double-stranded helix DNA. Figure adapted from reference [9].

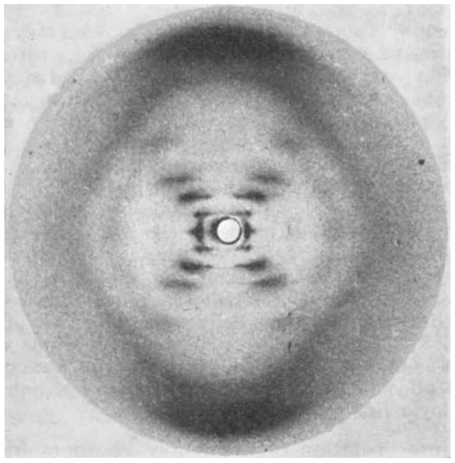

Figure 6 – Photograph of structure B of DNA in solution containing sodium formed by X-ray diagram. Figure taken from reference [11].

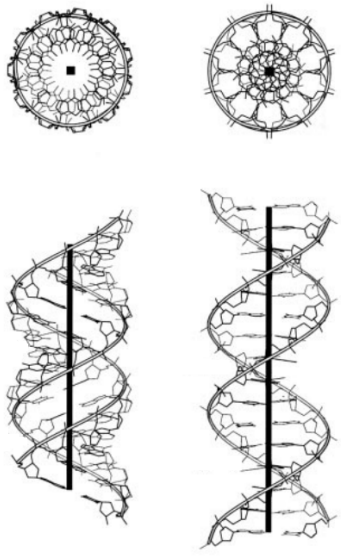

Figure 7 – Two structures of DNA, left: A-DNA and right: B-DNA. Figure taken from reference [12].

The dimensions described in Fig. 5 were confirmed by the X-ray diffraction through the photograph 51. Thus, we have that the diameter of the helix is about 2 nm, the distance between the adjacent base pairs is 0.34 nm and one turn, composed of 10 base pairs, is 3.4 nm. Moreover, we can see two types of grooves: minor and major. They are a consequence of the base pairs asymmetry to the backbone [13], and measure 1.2 nm and 2.2 nm, respectively [14]. In the major groove, the edges of the bases are more accessible to external ligands than in minor groove [15].

In 1954 Francis Crick and James Watson [16] published a more detailed description of the structure of DNA. They explained that the hydrogen bonds have a role of bonding the nitrogenous bases (adenine with thymine and cytosine with guanine) leading to base pairs formation, linking one strand with a complementary strand, which provides the characteristic shape of the helix and stabilizes the structure. In this article they stated that the interactions between adenine with thymine (AT) and cytosine with guanine (CG) were linked by two hydrogen bonds. They have also observed that the negatively charged phosphate groups formed a backbone located outside the double-stranded helix, while the base pairs were stacked inside the double-stranded helix [3]. Furthermore, according to Rosalind E. Franklin and Raymond G. Gosling, the backbone was hydrophilic, and the base pairs were hydrophobic [3, 17], which allow the DNA backbone to bond with the water molecules.

In contrast to Francis Crick and James Watson, Linus Pauling and Robert B. Corey influenced by the article written by Donohue [18], in which he discusses the types of hydrogen bonded interactions between purine and pyrimidine bases, published an article in 1956 describing a base pair GC with three hydrogen bonds [19]. The structures and dimensions of these bonds are described in Fig. 8, in which for AT base pair, the hydrogen bonds measure 0.28 nm (2.8 Å) and 0.3 nm (3.0 Å), and the distance between C-1′ of adenine to C-1′ thymine is 1.11 nm (11.1 Å). For GC base pair, two hydrogen bonds measure 0.29 nm (2.9 Å) and the other, 0.3 nm (3.0 Å), and the distance between C-1′ of guanine to C-1′ cytosine is 1.08 nm (10.8 Å). Currently, research shows the existence of a third hydrogen bond between adenine and thymine in both Watson-Crick and Hoogsteen base pairs in DNA [20], where the Hoogsteen base pair provide an alternating pairing geometry to Watson-Crick base pair [21], as shown in Figure 9. The adenine, as well as guanine, is flipped 180°, in relation to the helix axis, forming a conformation *syn* and a hydrogen bond is formed between N-3 and N-7, instead of N-1 and N-3 of the Watson-Crick base pair [22]. Thus, in this configuration these nitrogenous bases forming a completely new set of hydrogen bonds with thymine or cytosine.

The base pairs, in addition to hydrogen bonds, are also composed of non-covalent bonds between aromatic rings that form $\pi - \pi$ interactions [24]. The main function of these interactions is to establish a well-defined distance between the bases along the axis

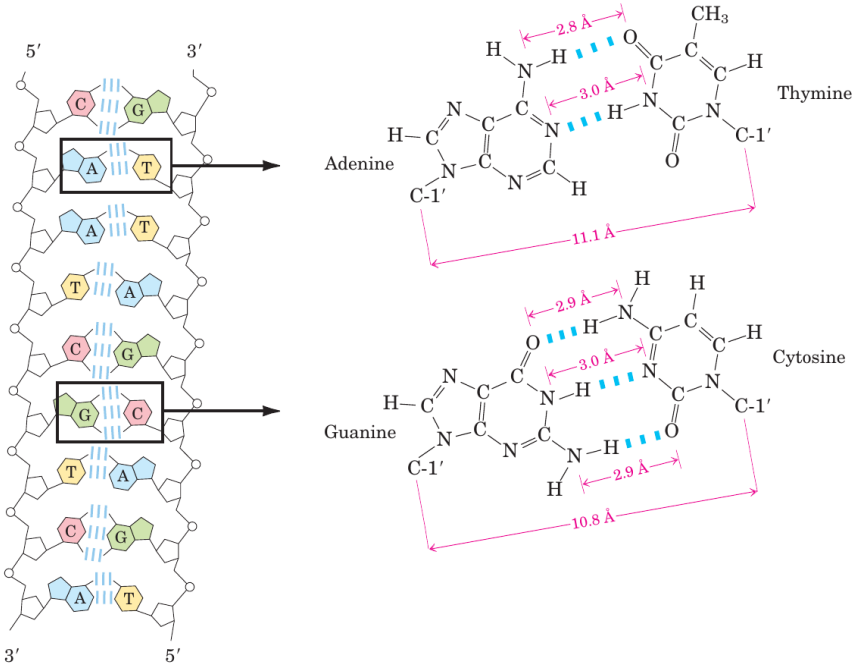

Figure 8 – Hydrogen bonds of Watson-Crick base pairs. Figure taken from reference [3].

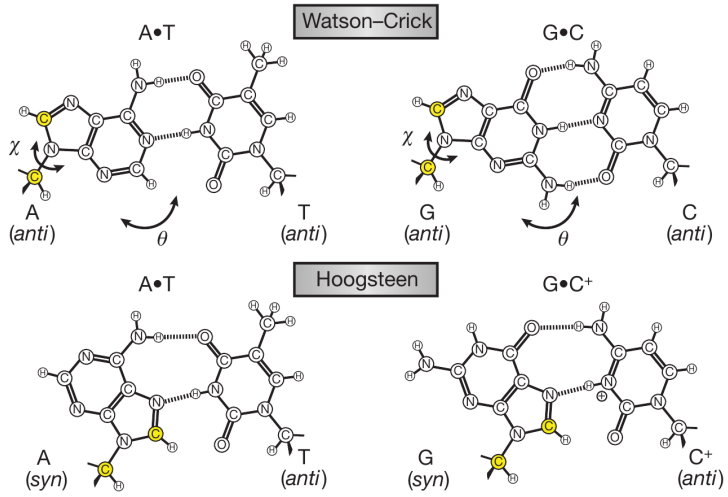

Figure 9 – Watson-Crick and Hoogsteen base pairs for AT and GC, with β-glycosidic bond and base-flipping represented by χ and θ, respectively. Figure taken from reference [23].

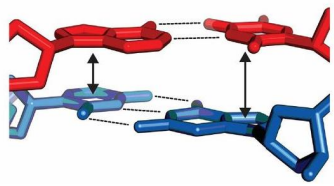

Figure 10 – The hydrogen bonds and stacking interaction between nitrogenous bases indicated by dashed lines and arrows, respectively. Figure taken from reference [28].

of the helix, resulting in a high rigidity of the molecules along the axis [25]. So this lead to stacking interaction between two adjacent base pairs inside the double-stranded helix, as shown in Fig. 10, that contribute significantly to the structure and stability of DNA and RNA [26, 27].

Due to the aromatic structure, when the stacking interaction is broken, the absorption of ultraviolet light occurs at a maximum wavelength of 260 nm [5]. This property

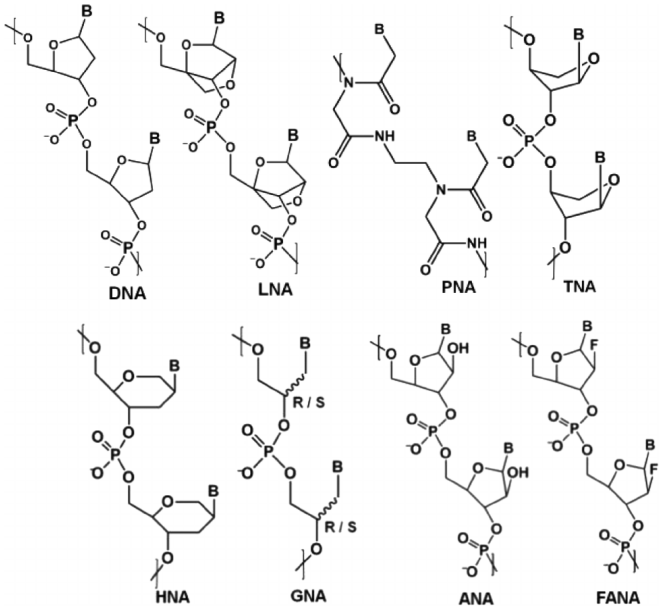

Figure 11 – Chemical structures of DNA and different types of XNAs, LNA (locked nucleic acid), PNA (peptide nucleic acid), TNA (threose nucleic acid), HNA (hexitol nucleic acid), GNA (glycol nucleic acid), ANA (arabinonucleic acid), and FANA (fluoroarabinonucleic acid), where B represents the nitrogenous bases. Figure taken from reference [43].

allows the identification of the nucleic acid and its concentration via spectrophotometry.

As the structure of nucleic acids is composed of highly negative charges, this makes transitions sensitive to ionic conditions [29]. The interaction between this structure and positive ions generates a reduction in the repulsion between the phosphate groups. Some ions are much more effective than others, for example, K^+ and Mg^{2+} are the most relevant cations for *in vivo* conditions [30]. For both ions, they interact with atoms and molecules through electrostatic forces. For metal ions, such as Mn^{2+} and Pb^{2+}, which have high polarizability, it leads to the formation of covalent bonding, and forms stronger bonds with nucleic acids than K^+ and Mg^{2+}. In Section 1.3, we will discuss in detail the effects of ions on the structure of nucleic acids.

1.2 Modified nucleic acids

In the 1980s, DNA and RNA, natural nucleic acids, began to be considered as therapeutics [31, 32] including a diverse class of drugs, such as antisense oligonucleotides (ASOs) [33, 34], microRNAs [35], small interfering RNAs (siRNAs) [36], aptamers [37], and others. However, the natural nucleic acids have poor chemical and biological stability, and due to these limitations, some chemical modifications in nucleobases, phosphodiester backbone and/or pentose [38] were proposed. The list of changes includes several possibilities and compose a group of modified (non-natural) nucleic acids [39–41] called XNA (xeno nucleic acid) as shown in Fig. 11. The modifications of the nucleobases can modulate base pairs strength and specificity, while in the phosphodiester backbone leads to larger degradation resistance by nucleases than in DNA and RNA [42]. For pentoses, the influence occurs on various nucleic acid properties such as duplex-forming ability, nuclease resistance and toxicity in cells.

The chemical modifications for nucleobases are possible to replace the canonical bases by others, such as 5-hydroxymethylcytosine and 7-methylguanine [44, 45], consequently, for nucleosides, we can have dihydrouridine and inosine [46, 47]. Modifications in backbone structure have promising therapeutic properties, such as TNA (threose nucleic acid), GNA (glycol nucleic acid), and LNA (locked nucleic acid) [48–50]. Even modified, some of these nucleic acids can still be linked via Watson-Crick base pairs with themselves, like TNA/TNA and LNA/LNA, and with DNA and RNA [51, 52]. Other possibility is to consider Hoogsteen base pair, that is responsible for connecting the third strand on the major groove of double-stranded helix forming triple-stranded helix structure [53, 54].

The thermodynamic properties are interesting features in modified nucleic acids. For example, LNA has greater thermal stability than RNA and DNA, and due to this feature it can be incorporated into siRNA duplexes to increase thermal duplex stability and half-life without compromising efficiency [55, 56]. For PNA (peptide nucleic acid), due to a flexible backbone, it can hybridize to DNA, RNA, and PNA with high thermal stability [57]. While for FANA (fluoroarabinonucleic acid) the high thermal stability is due to favorable enthalpy of hybridization and conformational pre-organization of the fluorinated sugar [58]. In addition to these chemical changes, the sequence composition can also influence the thermal stability, such as GNA, ZNA (zip nucleic acid), and SNA (serinol nucleic acid) that form homo-duplexes of high thermal stability, in contrast to hetero-duplexes [57].

It is known that therapeutic oligonucleotides composed of naturally occurring nucleotides are rapidly degraded *in vivo*, which makes them unsuitable for drug development [59]. For this reason, the main application of XNA is in biological applications for the field of medicine and drug discovery, due to additional assembly parameters and enhanced stability [60]. For example, PNA can be used to control cell growth, gene expression and growth phenotypes in the bacteria *Escherichia coli* [61], HNA and TNA are promising candidates for a treatment of the human immunodeficiency virus (HIV) [62, 63], and the modified nucleobase N1-methylpseudouridine is a key aspect of coronavirus disease (COVID-19) mRNA vaccines [64].

1.3 Ionic effects

The sugar-phosphate backbone of single-stranded DNA is negatively charged, therefore when two complementary single strands come together, an electrostatic repulsion occurs between them [65]. To reduce these repulsive Coulomb interactions, some type of salt is often added to solution facilitating hybridization and stabilizing the structure. Therefore, we can say that ions have important roles in the structure, dynamics, as well as the function of nucleic acids [66–73].

Among the many studies already carried out [74–82], we can discuss the study carried out by Cheng et al. [83] on the different ways in which the monovalent ions Na^+ and K^+ bind to DNA. In this study via a molecular dynamics (MD) simulation, they concluded that Na^+ binds preferentially with the phosphate groups, while K^+ interacts to the negatively charged sites of DNA base pairs in major and minor grooves. However, the results of MD simulations are sensitive to conditions, such as sequence composition, conformation, water contents, etc., as discussed by Várnai and Zakrzewska [84], in which K^+ mainly binds the major groove, and for Na^+ interact in both grooves, which is strongly sequence-dependent, and the preferred binding site is in the minor groove. Furthermore, in another approach by Savelyev and Papoian [85], they argue that Na^+ penetrates the DNA interior and also condenses around the DNA exterior to a significantly larger degree compared with K^+. This feature implies that in the presence of Na^+ closer to the DNA exterior surface, a stronger DNA electrostatic potential arises.

For divalent ions, according to Langlais et al. [86] in an analysis via Raman spectroscopy, Mg^{2+} and Ca^{2+} bind to phosphate groups of DNA, and Zn^{2+} and Cd^{2+} interact extensively with nitrogenous bases, contrary to the alkaline-earth metal ions which bind almost exclusively to the phosphate groups. Furthermore, a very small concentration of these metal ions causes a slight structural change in the nucleic acids, and resulting in increased base-stacking interactions. In another analysis by Hackl et al. [87] using infrared spectroscopy, it was shown that Cu^{2+}, Zn^{2+}, Mn^{2+} and Ca^{2+} bind both to DNA phosphate groups and bases, while Mg^{2+} only binds to phosphate groups of DNA. DNA undergoes structural transition into a compact form (a drastic decrease in the volume occupied by DNA molecules) due to interaction with these ions (except for Mg^{2+}), and the effectiveness to induce this form correlates with this affinity: $Cu^{2+} \gg Zn^{2+} > Mn^{2+} > Ca^{2+}$.

RNA as well as DNA is also influenced by several ions, as investigated to Kolev et al. [88], in experimental studies from X-ray crystallography and spectroscopic analysis (infrared, Raman and NMR). These methodologies suggest that Mg^{2+} and Ca^{2+} have high affinity to the phosphate groups, and for this reason, Mg^{2+} has a crucial role to preserve the tertiary structure of RNA, stabilizing the folding of the molecule. However, Na^+ and K^+ prefer to bind the nitrogenous bases and are often found in the groves of RNA (or DNA), which result in stabilization of tertiary ribosomal structure.

A way to quantify the concentration of ions in a solution is called the ionic strength (I), which describes the effect of charges and interionic interactions on electrolyte of various valence types [89]. The ionic strength is defined as

$$I = \frac{1}{2} \sum m_i z_i^2, \tag{1.1}$$

where m_i and z_i are the ionic concentration in M (M = mol/L), and the charge number on the ion, respectively [90, 91]. The ionic strength of mono- and divalent ions, according to Jambrec and Gebala [92], corresponds to $I = 3\,\text{mM}$ for $MgCl_2 \approx 100\,\text{mM}$ for NaCl,

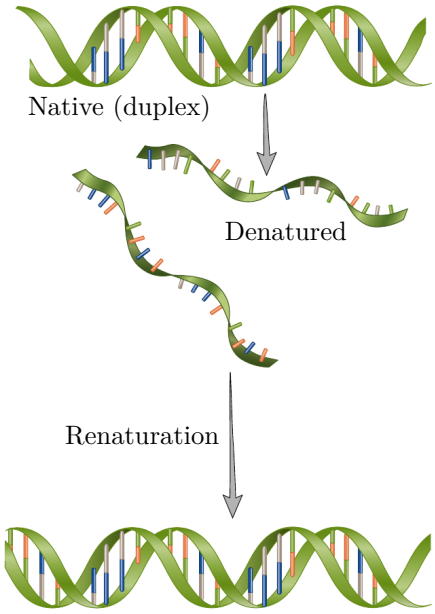

Figure 12 – Denaturation and renaturation processes of double-stranded helix. Figure adapted from reference [103].

although a 1 mM $MgCl_2$ has similar electrostatic screening properties as a 100 mM NaCl solution.

Due to the negatively charged structure of DNA, the strands repel each other more strongly when the ionic strength is low, but the repulsion decreases when the ionic strength increases through electrostatic screening, and this leads to a more flexible DNA chain [93]. This rule may not necessarily apply to XNAs with nonionic backbones, such as PNA and CNA (click nucleic acid), that can hybridize and form stable structures with DNA, even at low ionic strength [94, 95].

1.4 Thermal denaturation experiment

An important phenomenon in nucleic acids is called denaturation, in which the disruption of hydrogen bonds between base pairs, without affecting the phosphodiester bonds of the backbone [96], implies unwinding of double-stranded helix, and leads to two separate single strands [3], as shown in Fig. 12. The potential of this process is considered in several sophisticated techniques, such as high-resolution melting analysis (HRMA), a rapid, high-throughput, and robust method for detecting variants in the DNA sequence, and that includes gene scanning, genotyping, sequence matching, methylation analysis, and real-time polymerase chain reaction (PCR) [97–100]. HRMA has been effectively applied in clinical research and diagnostics, such as for viruses affecting humans, and phytopathogenic bacteria [101, 102].

The denaturation occurs when DNA is subjected to changes in buffer conditions. The most common condition is to increase the temperature of the solution, in which we consider melting temperature as the temperature at which precisely half the strands in

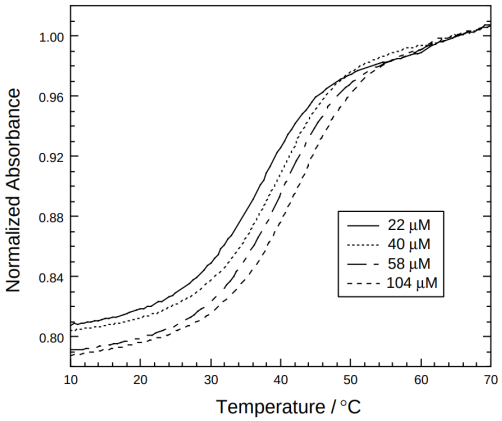

Figure 13 – Melting curves of absorbance *versus* temperature for four total concentrations of dCA_7G and dCT_7G. The curves are normalized to an absorbance of 1 at 65 °C. Figure taken from reference [110].

solution are dissociated (denatured) [104]. Another condition is through various chemical agents such as formamide, ethylene glycol, and urea [105–107], that influences electrostatic interactions leading to the destabilization of the negatively charged ions. The second type of chemical agent is used to change pH that leads to broken hydrogen bonds.

Denaturation has a reverse process, known as renaturation or hybridization, as shown in Fig. 12, and occurs when the heated solution of denatured double-stranded helix are slowly cooled upon below the melting temperature, until the unwound segments of the two strands rewind [3, 103], and after these, the original properties of double-stranded helix are restored [108]. Like denaturation, renaturation has important roles in other biological processes, such as recombination and design of oligonucleotide probes [109].

When the denaturation occurs due to an increase in temperature, the thermodynamic properties can be monitored by ultraviolet (UV) spectrophotometry [110]. In this type of experiment, when the hydrogen bonds linking two opposite strands are broken, the UV absorbance increases, mainly from 260 nm, due to nitrogenous bases are more exposed to environment [108]. The graph of UV absorbance *versus* temperature, is defined as melting curve. For instance, a thermal denaturation experiment measuring four total concentrations of dCA_7G and dCT_7G [110], the melting curves is shown in Fig. 13, where one can observe the regular sigmoidal form of the melting curves, on which the midpoint of the transition is the melting temperature (T_m). In T_m, the fraction of base pairs is $f = 1/2$, and is calculated via the maximum of the first derivative [110] of the melting curve.

As we have now established that the melting curve corresponds the transition between double-stranded helix o single strands, we define the two-state model for two strands (X and Y) [104], as

$$\mathrm{X} \cdot \mathrm{Y} \rightleftharpoons \mathrm{X} + \mathrm{Y}. \tag{1.2}$$

We define C_t as the sum of the single strand concentrations at high temperature in which

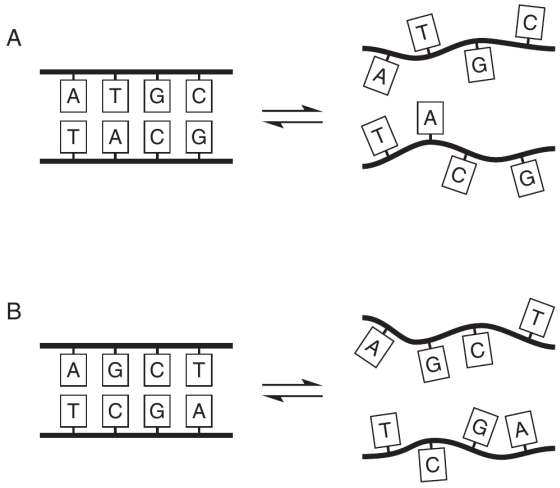

Figure 14 – The melting of short oligonucleotide chains, (A) a non-self-complementary duplex and (B) a self-complementary duplex. Figure taken from reference [104].

the duplex is completely denatured,

$$C_{\mathrm{t}} = [\mathrm{X}]_{\mathrm{tot}} + [\mathrm{Y}]_{\mathrm{tot}}, \tag{1.3}$$

which is also called the total strand concentration.

Another relevant parameter is the equilibrium constant K_{eq}, which has a relation with C_{t} that depends on the type of duplex oligonucleotide: self- and non-self-complementary, as shown schematically in Fig. 14. The self-complementary sequences are two identical sequences capable of hybridizing to each other. In Fig. 14, we notice that for non-self-complementary duplex, the linked strands are different, 5′-ATGC-3′ is linked with 3′-TAGC-5′. In contrast, for self-complementary duplex, the strand 5′-AGCT-3′ is hybridized to another strand of identical composition in the opposite orientation, 3′-AGCT-5′. Furthermore, in both types, the duplexes are complementary, adenine is linked to thymine, and cytosine with guanine.

At any temperature in a non-self-complementary duplex oligonucleotide, we define

$$f = \frac{[\mathrm{X}]}{[\mathrm{X}\cdot\mathrm{Y}]_{\mathrm{init}}} = \frac{[\mathrm{Y}]}{[\mathrm{X}\cdot\mathrm{Y}]_{\mathrm{init}}}, \tag{1.4}$$

$$C_{\mathrm{t}} = [\mathrm{X}] + [\mathrm{Y}] + 2[\mathrm{X}\cdot\mathrm{Y}], \tag{1.5}$$

$$K_{\mathrm{eq}} = f^2 \frac{C_{\mathrm{t}}}{2(1-f)}. \tag{1.6}$$

These expressions are applied at the melting temperature ($f = 1/2$),

$$[\mathrm{X}]_m = [\mathrm{Y}]_m = [\mathrm{X}\cdot\mathrm{Y}]_m, \tag{1.7}$$

$$K_{\mathrm{eq}} = \frac{C_{\mathrm{t}}}{4}. \tag{1.8}$$

For self-complementary duplex oligonucleotide Eqs. (1.4-1.8) take a somewhat different form as now both strands are of type X. We set X = Y in the Eq. (1.2), and

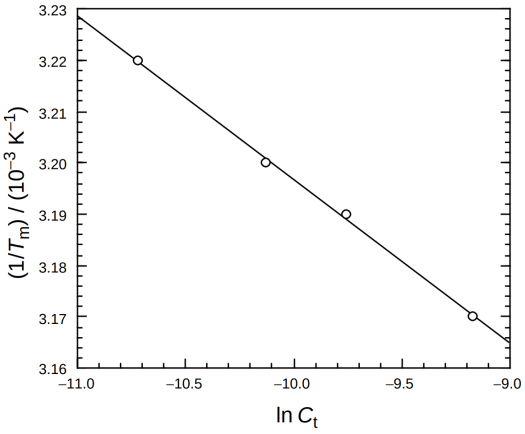

Figure 15 – An example of van't Hoff plot of $1/T_m$ *versus* $\ln C_\mathrm{t}$, where C_t is usually expressed in μM. Figure taken from reference [110].

define

$$f = \frac{[\mathrm{X}]}{2[\mathrm{X}\cdot\mathrm{X}]_\mathrm{init}}, \tag{1.9}$$

$$C_\mathrm{t} = [\mathrm{X}] + 2[\mathrm{X}\cdot\mathrm{X}], \tag{1.10}$$

$$K_\mathrm{eq} = 2f^2\frac{C_\mathrm{t}}{1-f}. \tag{1.11}$$

These expressions are applied at the melting temperature,

$$[\mathrm{X}]_m = 2[\mathrm{X}\cdot\mathrm{X}]_m, \tag{1.12}$$

$$K_\mathrm{eq} = C_\mathrm{t}. \tag{1.13}$$

In the two-state model, considering the previous equations in which we define K_eq, it is possible to perform a thermodynamic analysis, as discussed by van't Hoff [104]. The T_m is related to the standard Gibbs energy

$$\Delta G = -RT_m \ln K_\mathrm{eq}, \tag{1.14}$$

by substituting $\Delta G = \Delta H - T_m\Delta S$, and isolating $1/T_m$,

$$\frac{1}{T_m} = \begin{cases} -(R/\Delta H)\ln C_\mathrm{t} + (\Delta S - R\ln 4)/\Delta H & \text{non-self-complementary} \\ -(R/\Delta H)\ln C_\mathrm{t} + \Delta S/\Delta H & \text{self-complementary} \end{cases} \tag{1.15}$$

The van't Hoff plot in Fig. 15 shows the four concentrations for self-complementary strands $\mathrm{dCA_7G}$ and $\mathrm{dCT_7G}$. By using linear fit on the graph, we can obtain the values of the Eq. (1.15), ΔH and ΔS, and consequently, ΔG to a fixed temperature. If the strands contain a larger amount of cytosine and guanine (CG) than adenine and thymine (AT) base pairs, the value of T_m is higher than otherwise. This occurs because the CG has an extra hydrogen bond compared to AT.

1.5 Conclusions

In this chapter we revised the basic concepts of duplex nucleic acids, as well as of modified nucleic acids, and the definition of oligonucleotide melting temperatures. These definitions will be important for the next chapters: in Chapter 2 we discuss the theoretical models used in the projects described in Chapters 3 and 4 about TNA, a modified nucleic acid, and DNA in solutions containing Mg^{2+}, respectively.

2 Theoretical models

The thermodynamic properties of duplexes can be investigated through theoretical models, such as the nearest-neighbor (NN) model in which the free energy calculates depend on the identity of the two adjacent base pairs. Furthermore, the denaturation process was described by M. Peyrard and A. R. Bishop in a physical statistical and mesoscopic model called Peyrard-Bishop (PB) model. The first proposal for the PB model was originally limited to homogeneous sequences of DNA, but a few years later this was extended to heterogeneous sequences by Zhang et al. [111]. For the projects in this thesis, we considered the both types of sequences in an approach developed by our research group and discussed in Weber et al. [112]. In this methodology, a new parameter called melting index is introduced to represent thermodynamic equivalence, and maps the experimental DNA melting temperatures for short DNA sequences. In this chapter, we will review NN model, which is fundamental to understand the experimental results, as well as the various approaches to the PB model in more detail.

2.1 Nearest-neighbor model

To investigate nucleic acid duplex stability, we can use the nearest-neighbor (NN) model developed by Tinoco Jr et al. [113], that predicts thermodynamic properties (ΔG, ΔH, ΔS, and T_m) using the free energy of neighboring base pairs in a sequence [114] calculated from regression of duplex concentration versus melting temperature data for various sequences [115]. In the NN model, the sequence is subdivided into neighboring base pairs, as exemplified with the canonical DNA sequence,

$$\begin{matrix} 5'\text{-CGTATG-}3' \\ 3'\text{-GCATAC-}5' \end{matrix} \rightarrow \begin{matrix} 5'\text{-CG-}3' \\ 3'\text{-GC-}5' \end{matrix} + \begin{matrix} 5'\text{-GT-}3' \\ 3'\text{-CA-}5' \end{matrix} + \begin{matrix} 5'\text{-TA-}3' \\ 3'\text{-AT-}5' \end{matrix} + \begin{matrix} 5'\text{-AT-}3' \\ 3'\text{-TA-}5' \end{matrix} + \begin{matrix} 5'\text{-TG-}3' \\ 3'\text{-AC-}5' \end{matrix}$$

and for simplicity, we can use the notation

$$\begin{matrix} 5'\text{-CGTATG-}3' \\ 3'\text{-GCATAC-}5' \end{matrix} \rightarrow \text{CG/GC + GT/CA + TA/AT + AT/TA + TG/AC}$$

In this sequence of length 6 bp (base pair), the energy contributions corresponds to 5 neighboring base pairs. Thus, a sequence formed by N base pairs has $N-1$ energy contributions, and for canonical neighboring base pairs in DNA and RNA duplexes, we can have 10 combinations, while in DNA/RNA hybrids are 16 combinations.

The free energy depends on the interactions of neighboring base pairs is defined as,

$$\Delta G_i(\text{total}) = \sum_j n_{ij} \Delta G_j + \Delta G(\text{init}) + \Delta G_i(\text{sym}), \tag{2.1}$$

where ΔG_j are the free energies for the 10 canonical neighboring base pairs (AA/TT, AT/TA,TA/AT, CA/GT, GT/CA, CA/GT, GA/CT, CG/GC, GC/CG, and GG/CC), n_{ij} is the number of occurrences of each neighboring base pairs j in each sequence i, $\Delta G(\text{init})$ is the free energy of initiation (initial base pair), and $\Delta G_i(\text{sym})$ is $+0.4\,\text{kcal/mol}$ for a self-complementary duplex and 0 for a non-self-complementary [116, 117].

The NN model can use experimental data set from van't Hoff plot ($1/T_m$ *versus* $\ln C_\text{t}$) to calculate the total enthalpy and entropy of a given nucleic acid sequence. For each calculation, we use the Eq. (1.15), along with a linear regression of the curve. Furthermore, if the thermodynamic parameters (ΔH, ΔS, and ΔG) are known, the melting temperature can also be predicted using Eq. (1.15).

In an analysis of RNA duplexes done by Xia et al. [117] via NN model, they concluded that the terminal GC base pairs are more stable than duplexes with the same nearest-neighbors but for terminal AU base pairs. A possible reason for this difference can be attributed to the dependence of the numbers of Watson-Crick hydrogen bonds on base composition. As for DNA duplexes according to SantaLucia et al. [116], the trend in nearest-neighbor stability is defined as: GC/GC > CG/GC > GG/CC > GA/CT ≈ GT/CA ≈ CA/GT > CT/GA > AA/TT > AT/TA > TA/AT.

The NN model predicts thermodynamic parameters based on experimental melting temperature dataset. However, this model has some limitations, such as to obtain parameters accurately, it is necessary to have a sufficiently large sequence set, and it does not describe the physical properties related to the intramolecular interactions of the sequences [118]. Another important model is the Poland-Scheraga (PS) model which is a physical statistical model in which the denaturation is described as bound or open states, like the Ising model [119]. In the PS model, DNA duplex is described as a two-state sequence, where 1 represents a bound state, with an associated a Boltzmann weight $q = e^{-\epsilon/k_B T}$, while 0 represents an open state, that is, the nitrogenous bases do not form hydrogen bonds, with a purely entropic weight [120]. However, the PS model ignores the chemical composition, stiffness or torsion [121], and as the NN model, does not describe the intramolecular interactions of the sequences. These limitations motivated the development of other theoretical models to be able to describe thermal denaturation by explicitly taking into account the molecular interactions. So, in the next section, we will discuss the Peyrard-Bishop model that defines the intramolecular interactions as hydrogen bonds, and introduces the Morse potential.

2.2 Peyrard-Bishop model for homogeneous sequences

The physical statistical and mesoscopic model called Peyrard-Bishop (PB) model was developed in 1989 by M. Peyrard and A. R. Bishop, and describes the denaturation of

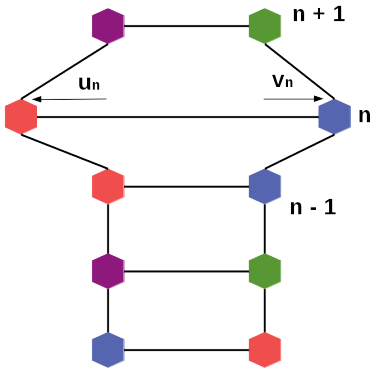

Figure 16 – The degrees of freedom u_n and v_n in the PB model. Figure taken from reference [123].

a double-stranded helix of DNA [122]. In the PB model each base pair has two degrees of freedom u_n and v_n, corresponding to the displacement of the bases from equilibrium positions along the direction of the hydrogen bonds that connect the two bases in a pair, as shown in Figure 16. They defined the harmonic potential as the stacking interaction between two neighboring pairs, and the Morse potential as the connecting hydrogen bonds in two bases on opposite strands.

The harmonic and Morse potentials are defined as the total energy of the system, that is represented by the Hamiltonian

$$H = \sum_n \frac{1}{2}m(\dot{u}_n^2 + \dot{v}_n^2) + \frac{1}{2}k[(u_n - u_{n+1})^2 + (v_n - v_{n+1})^2] + D\{e^{[-a(u_n - v_n)/\sqrt{2}]} - 1\}^2, \quad (2.2)$$

where the first term is the kinetic energy for bases of mass m, the second is the harmonic potential with coupling constant k, and the last is the Morse potential, in which the parameter D represents the energy required to separate the base pair, and a is the range of the potential. The Morse potential is an average, or effective, potential that represents the all bonds which connect the two bases in a pair [122].

The variables u_n and v_n are changed via a canonical transformation, so that the Hamiltonian of Eq. (2.2) is written with the decoupling of variables,

$$x_n = (u_n + v_n)/\sqrt{2}, \quad y_n = (u_n - v_n)/\sqrt{2}.$$

These representations describe the motion of strands in-phase and out-in-phase, respectively [122]. Furthermore, x_n is the center of mass corresponding to base pair and y_n is associated with DNA dynamics. The advantage of this transformation is that it allows the separation of Hamiltonian terms,

$$H = H(x) + H(y) = \sum_n \left[\frac{p_n^2}{2m} + W(x_n, x_{n+1})\right] + \sum_n \left[\frac{q_n^2}{2m} + W(y_n, y_{n+1}) + V(y_n)\right], \quad (2.3)$$

where $p_n = m\dot{x}_n$ and $q_n = m\dot{y}_n$. $W(x_n, x_{n+1})$ and $W(y_n, y_{n+1})$ are the harmonic potentials, and $V(y_n)$ is the Morse potential, shown in Figs. 17 and 18, and defined as

$$W(x_n, x_{n+1}) = \frac{1}{2}k(x_n - x_{n+1})^2, \quad W(y_n, y_{n+1}) = \frac{1}{2}k(y_n - y_{n+1})^2, \quad (2.4)$$

$$V(y_n) = D(e^{-ay_n} - 1)^2. \quad (2.5)$$

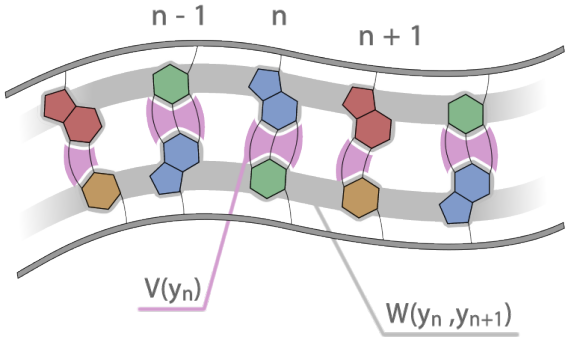

Figure 17 – The DNA base pairs are represented as $n-1$, n and $n+1$. The PB model describes the denaturation of double-stranded helix through the harmonic $W(y_n, y_{n+1})$ and Morse potentials $V(y_n)$. Figure taken from reference [124].

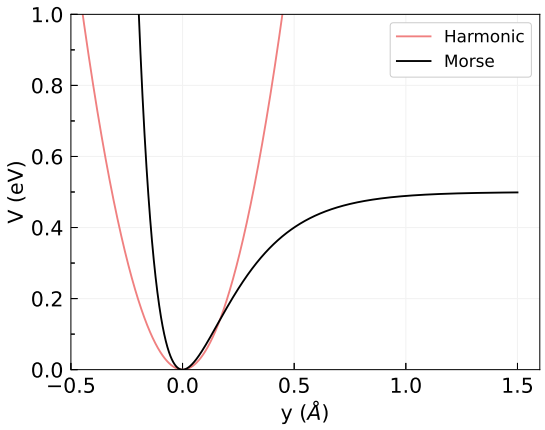

Figure 18 – Morse and harmonic potentials as a function of inter-base distance y.

The physical properties of DNA via PB model are investigate through canonical ensemble. This choice is due to the fact that a DNA strand is placed in a solution that will be heated until it reaches the melting temperature, which consists of a system where a molecule is placed in a thermal bath. For a system with N base pairs in a fixed temperature and using the Eq. (2.3), we define the partition function as

$$Z = \int_{-\infty}^{+\infty} \prod_{n=1}^{N} \mathrm{d}x_n \mathrm{d}y_n \mathrm{d}p_n \mathrm{d}q_n e^{-\beta H(p_n, x_n, q_n, y_n)} = Z_p Z_x Z_q Z_y, \tag{2.6}$$

where $\beta = 1/k_B T$ and k_B is the Boltzmann constant. The partition functions Z_p, Z_q and Z_x are solved using Gaussian integral

$$I = \int_{-\infty}^{\infty} e^{-a\xi^2} \mathrm{d}\xi = (\pi/a)^{1/2}, \tag{2.7}$$

where ξ is a generic variable. However, Z_y has two equations to integrate, making the solution more difficult. Therefore, the Eq. (2.6) is reduced to

$$\begin{aligned} Z &= \left(\frac{8\pi^3 m^2}{k\beta^3}\right)^{N/2} Z_y \\ &= \left(\frac{8\pi^3 m^2}{k\beta^3}\right)^{N/2} \int_{-\infty}^{+\infty} \prod_{n=1}^{N} \mathrm{d}y_n e^{-\beta[W(y_n, y_{n+1}) + V(y_n)]}. \end{aligned} \tag{2.8}$$

To simplify, the function partition Z_y is rewritten using the kernel function [125],

$$K(y_n, y_{n+1}) = e^{-\beta[W(y_n,y_{n+1})+V(y_n)/2+V(y_{n+1})/2]} = K(y_{n+1}, y_n), \tag{2.9}$$

$$\begin{aligned} Z_y &= \int_{-\infty}^{+\infty} \prod_{n=1}^{N} \mathrm{d}y_n K(y_n, y_{n+1}) \\ &= \int_{-\infty}^{+\infty} \mathrm{d}y_1 \mathrm{d}y_2 \ldots \mathrm{d}y_N K(y_1, y_2) K(y_2, y_3) \ldots K(y_N, y_1). \end{aligned} \tag{2.10}$$

The partition function Z_y can be evaluated in the thermodynamic limit ($N \to \infty$) using the eigenfunctions and eigenvalues of a transfer integral operator

$$\int \mathrm{d}y_{n+1} K(y_n, y_{n+1}) \varphi_i(y_{n+1}) = e^{-\beta\epsilon_i} \varphi_i(y_n), \tag{2.11}$$

and the result is $Z_y = e^{-N\beta\epsilon_0}$, where ϵ_0 is the lowest eigenvalue of a Schrödinger-type equation [122], which determines the eigenfunctions of the transfer integral operator,

$$-\frac{1}{\beta^2 k} \frac{\partial^2 \varphi_i(y)}{\partial y_i^2} + D(e^{-4ay} - 2e^{-2ay}) \varphi_i(y) = (\epsilon_i - s_0 - D) \varphi_i(y) \tag{2.12}$$

with $s_0 = (1/2\beta) \ln(\beta k / 2\pi)$. For a particle in a Morse potential, the Eq. (2.12) is formally identical to the Schrödinger equation, and can be solved exactly.

The eigenfunctions and eigenvalues are used to determine an important parameter of DNA denaturation, the mean relative stretching $\langle y \rangle$ of the hydrogen bonds shown in Figure 19. The value of $\langle y \rangle$ indicates whether the DNA is denatured or not. It is given by

$$\langle y_m \rangle = \frac{1}{Z_y} \int \prod_{n=1}^{N} y_m K(y_n, y_{n+1}) \mathrm{d}y_n. \tag{2.13}$$

This integral can be calculated via transfer integral method for the limit of large N,

$$\langle y \rangle = \langle \varphi_0(y) | y | \varphi_0(y) \rangle = \int \varphi_0^2(y) y \mathrm{d}y, \tag{2.14}$$

where $\varphi_0(y)$ is the normalized eigenfunction. The parameter $\langle y \rangle$ is used to analyze the amplitudes of the base pairs oscillations [126], that is, the expected base pair separation at certain temperatures.

The main limitation of the PB model involves treating homogeneous sequences of DNA, which are defined as strands composed of only one type of nucleobase, for example, a cytosine strand and a complementary guanine strand. However, a new technique were developed with different potentials [125,127] including heterogeneous sequences [111] which we will review next.

2.3 Peyrard-Bishop model for heterogeneous sequences

An approximation for dealing with heterogeneous sequences was developed by Zhang et al. [111]. These sequences are formed by both occurrences of AT and CG base

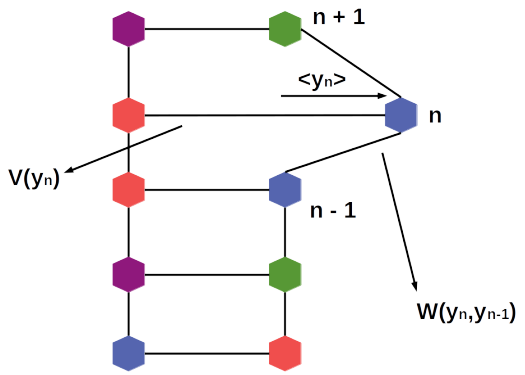

Figure 19 – The base displacement $\langle y \rangle$ in the direction of hydrogen bonds. Figure taken from reference [123].

pairs. So an adaptation is necessary due to the difference between these base pairs, for example AT and CG have two and three hydrogen bonds, respectively. The approach starts with a way to solve the function partition shown in Eq. (2.10). First, we define the integral equation involving the kernel,

$$\int K(y_n, y_{n+1})\varphi(y_{n+1})\mathrm{d}y_{n+1} = \lambda\varphi(y_n). \tag{2.15}$$

As seen in Eq. (2.9), the kernel has a symmetry. Therefore, with $K(y_n, y_{n+1}) > 0$, we can assume

$$\|K(y_n, y_{n+1})\| = \left\{ \iint [K(y_n, y_{n+1})]^2 \mathrm{d}y_n \mathrm{d}y_{n+1} \right\}^{1/2} < \infty. \tag{2.16}$$

This implies in a set of positive eigenvalues and orthonormal eigenfunctions [111]. The notation to describe eigenvalues and eigenfunctions are λ_1, λ_2, etc., and $\varphi_1(y_n), \varphi_2(y_n)$, etc., respectively. The eigenfunctions must obey the orthogonality and completeness relations, respectively,

$$\sum_{i=1}^{\infty} \varphi_i(y_n)\varphi_i(y_{n+1}) = \delta(y_n - y_{n+1}), \tag{2.17}$$

$$\int \mathrm{d}y_n \varphi_i(y_n)\varphi_j(y_n) = \delta_{ij}. \tag{2.18}$$

Therefore, the kernel $K(y_n, y_{n+1})$ can be expanded as

$$K(y_n, y_{n+1}) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(y_n)\varphi_i(y_{n+1}). \tag{2.19}$$

When we substitute the Eqs. (2.17), (2.18) and (2.19) in Eq. (2.10), the result for y_2 is

$$\begin{aligned} \int K(y_1, y_2)K(y_2, y_3)\mathrm{d}y_2 &= \int \sum_{i=1}^{\infty} \lambda_i \varphi_i(y_1)\varphi_i(y_2) \sum_{j=1}^{\infty} \lambda_j \varphi_j(y_2)\varphi_j(y_3)\mathrm{d}y_2 \\ &= \sum_i \sum_j \lambda_i \lambda_j \varphi_i(y_1)\varphi_j(y_3)\delta_{ij} \\ &= \sum_i \lambda_i^2 \varphi_i(y_1)\varphi_i(y_3). \end{aligned} \tag{2.20}$$

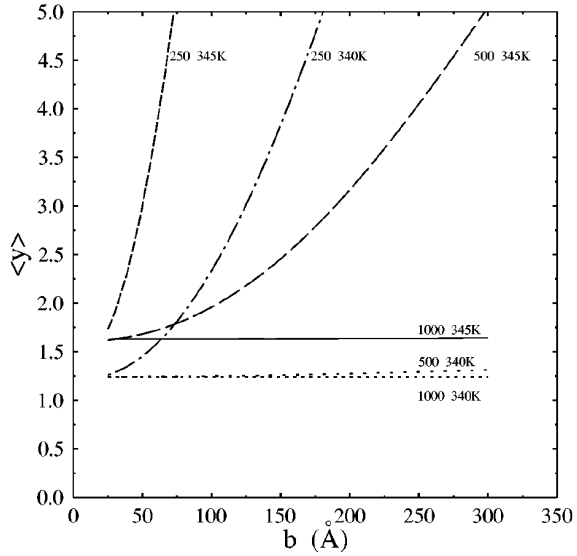

Figure 20 – Average relative displacement $\langle y \rangle$ as a function of the upper integration limit b, for three values of N, 250, 500, and 1000, at two different temperatures, 340 K and 345 K. Figure taken from reference [111].

Then, proceeding this way, the partition function is

$$\begin{aligned} Z_y &= \int \sum_{i=1}^{\infty} \lambda_i^N \varphi_i(y_1)\varphi_i(y_1)\mathrm{d}y_1 \\ &= \sum_{i=1}^{\infty} \lambda_i^N \lambda_{ii} = \sum_{i=1}^{\infty} \lambda_i^N. \end{aligned} \tag{2.21}$$

Furthermore, the Eq. (2.13) is rewritten as

$$\langle y \rangle = \frac{1}{Z_y} \sum_{i=1}^{\infty} \lambda_i^N \langle i|y|i\rangle = \frac{1}{Z_y} \sum_{i=1}^{\infty} \lambda_i^N \int \varphi_i^*(y) y \varphi_i(y)\mathrm{d}y. \tag{2.22}$$

The analytical calculation of the Eq. (2.22) is done by integrating the spatial coordinates over an interval $]-\infty, +\infty[$ [111]. When we perform the numerical integration, this equation is discretized and truncated to the limits of integration $[a, b]$. Analyzing the results, the model presents a divergence of $\langle y \rangle$ to $b \to +\infty$, as shown in Fig. 20. This problem can be solved with an introduction of a small torsion angle in the harmonic potential [112],

$$W(y_n, y_{n+1}) = \frac{k}{2}(y_n^2 - 2y_n y_{n+1}\cos\theta + y_{n+1}^2), \tag{2.23}$$

where θ is a small angle fixed at 0.01 rad.

To include heterogeneous sequences, we need to change the equations (2.9) and (2.10),

$$K^{(n,n+1)}(y_n, y_{n+1}) = e^{-\beta[W^{(n,n+1)}(y_n,y_{n+1})+V^n(y_n)/2+V^{n+1}(y_{n+1})/2]}, \tag{2.24}$$

$$Z_y = \int \mathrm{d}y_1\mathrm{d}y_2 \ldots \mathrm{d}y_N K^{(1,2)}(y_1, y_2)K^{(2,3)}(y_2, y_3)\ldots K^{(N,1)}(y_N, y_1), \tag{2.25}$$

where the indexes n and $n+1$ influence V and W and their parameters changes with the site n [128]. The simplification made previously in kernel can not be used anymore, so to expand it, the proposal is to consider the set of eigenfunctions of homogeneous sequences [111],

$$K^{(n,n+1)}(y_n, y_{n+1}) = \sum_{j,k=1}^{\infty} C_{jk}^{(n,n+1)} \varphi_j(y_n)\varphi_k(y_{n+1}), \tag{2.26}$$

where this is more general than Eq. (2.19). The coefficients $C_{jk}^{(n,n+1)}$ are defined as

$$C_{jk}^{(n,n+1)} = \iint \mathrm{d}y_n \mathrm{d}y_{n+1} K^{(n,n+1)}(y_n, y_{n+1}) \varphi_j(y_n) \varphi_k(y_{n+1}). \tag{2.27}$$

Considering these equations, the partition function can be written as

$$Z_y = \mathrm{Tr}(\mathbf{C}^{(1,2)} \mathbf{C}^{(2,3)} \dots \mathbf{C}^{(N,1)}), \tag{2.28}$$

where each matrix $\mathbf{C}^{(n,n+1)} = [C_{jk}^{(n,n+1)}]$ represents the interaction between neighboring base pairs n and $n+1$. In Eq. (2.28), the last matrix, $\mathbf{C}^{(N,1)}$, represents the boundary condition which links the first and the last base pair [128]. The condition can either be periodic, where the DNA sequence is considered as a ring, or open ended where the stacking interaction is neglected, that is, $W^{(N,1)} = 0$. Similarly, the mean stretching of hydrogen bonds is

$$\langle y \rangle = \frac{1}{Z_y} \mathrm{Tr}(\mathbf{C}^{(1,2)} \mathbf{C}^{(2,3)} \dots \mathbf{C}^{(n-1,n)} \mathbf{Y} \mathbf{C}^{(n,n+1)} \dots \mathbf{C}^{(N,1)}), \tag{2.29}$$

where the matrix elements $\mathbf{Y}$ are

$$Y_{jk} = \int \varphi_j(y_n) y_n \varphi_k(y_n) \mathrm{d}y_n. \tag{2.30}$$

2.4 Peyrard-Bishop model via thermodynamic equivalence

An alternative approach is proposed to investigate DNA sequences with similar thermodynamic properties [112], without the need to calculate melting temperatures and mean relative stretching. In this approach, we define the thermal equivalence as a physical quantity that allows us to map it to measure melting temperatures. The thermal equivalence is used as a melting index, that is not strongly dependent on the temperature for which the calculation is performed [128]. This methodology has computational efficiency for the optimization of parameters related to hydrogen bonding and stacking interactions.

The PB model determines the matrices $\mathbf{C}$ from the expansion of a homogeneous sequence [128]. We can assume a sequence formed by base pairs CG, that leads to the matrix $\mathbf{C}^{\mathrm{CG,GC}}$, a diagonal matrix which the elements are the eigenvalues λ_i of Eq. (2.21). Therefore, we define the others matrices as the sum of a diagonal Λ and non-diagonal $\Delta^{(n,n+1)}$ matrix,

$$\mathrm{C}^{(n,n+1)} = \Lambda + \Delta^{(n,n+1)}, \quad \Lambda = \mathrm{C}^{(\mathrm{CG,CG})}, \tag{2.31}$$

where $\Delta^{(n,n+1)}$ is the difference of the interaction between neighbors of type $(n, n+1)$ and (CG,CG) [128]. The partition function in Eq. (2.28) is rewritten as

$$Z_y = \mathrm{Tr}[(\Lambda + \Delta^{(1,2)})(\Lambda + \Delta^{(2,3)}) \dots (\Lambda + \Delta^{(N,1)})]. \tag{2.32}$$

When we carry out the matrix product in equation above, the result is a sum of $N+1$ terms as a function of Λ^{ω},

$$Z_y = \sum_{\omega=0}^{N} Z_{\omega}(\Lambda) = \sum_{\omega=0}^{N} \mathrm{Tr}[M(\Lambda^{\omega})], \tag{2.33}$$

where $M(\Lambda^{\omega})$ represents all terms containing ω multiplications of the matrix Λ. For homogeneous CG sequences, $\Delta^{(n,n+1)} = 0$ and

$$Z_y = \sum_{\omega=0}^{N} \mathrm{Tr}[M(\Lambda^{\omega})] = \sum_{i=1}^{\infty} \lambda_i^N, \tag{2.34}$$

the same result obtained in Eq. (2.21).

The calculated values of $Z_{\omega}(\Lambda)$ for several short sequences with CG content between 40% and 60%, as a function of the ω parameter are shown in Fig. 21. We notice a Gaussian shape for all curves of $Z_{\omega}(\Lambda)$, and it is evident that the values of ω are dependent on the structure of the sequence, that is, ω allows us to characterize each sequence individually. However, the maximal values of ω defined as $\omega_{\max}$, do not depend strongly on the temperature, as shown in Fig. 22. Simply, $\omega_{\max}$ can be interpreted as an interpolation parameter between a complete CG homogeneous sequence ($\omega_{\max} = N$) and a sequence consisting of only AT (lowest $\omega_{\max}$) [129]. Furthermore, analyzing the Fig. 23, we observe the linear relation between $\omega_{\max}^{1/2}$ and experimental melting temperatures. Therefore, $\omega_{\max}^{1/2}$ can be considered as a dimensionless value to the thermal equivalence and defined as melting index τ [128].

For the ith duplex, the relation between predicted melting temperatures $T_i'(P)$, and melting indexes τ_i is defined as

$$T_i'(P) = a_0(N) + a_1(N)\tau_i(P), \tag{2.35}$$

$$a_k = b_{0,k} + b_{1,k}N^{1/2}, \tag{2.36}$$

where N is the sequence length, a_0 and a_1 are the linear regression coefficients to each set N and at a single strand concentration C_t, P is a set of tentative model parameters, and $k = 0$ or 1. The advantage of the melting index is to optimize the parameters of PB model at a small computational cost [112].

The predicted melting temperature (T_i') is compared in several rounds of minimizations to test whether the value obtained is an occurrence of a local minimum. So the Hamiltonian parameters vary until the difference between the predicted and experimental melting temperature (T_i) is minimized,

$$\chi^2 = \sum_{i=1}^{N} [T_i - T_i'(\{p\}_k)]^2, \tag{2.37}$$

resulting from the set of tentative parameters $\{p\}_k$,

$$\{p\}_k = \{p_{k1}, p_{k2}, p_{k3}, \dots\}. \tag{2.38}$$

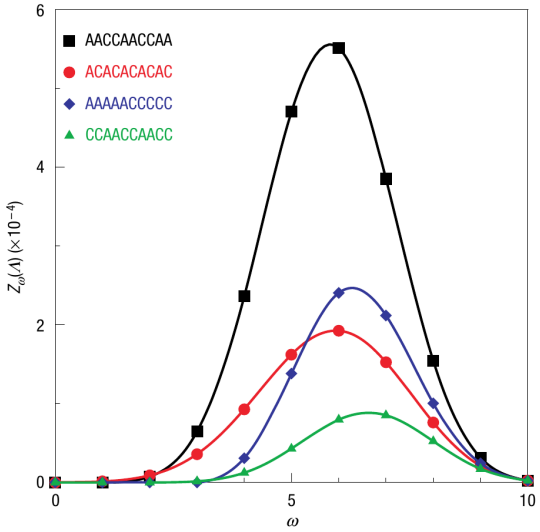

Figure 21 – Partition function Z_ω as a function of the order ω of the diagonal matrix Λ, for sequences of length 10 bp with CG content between 40% and 60%. Figure taken from reference [112].

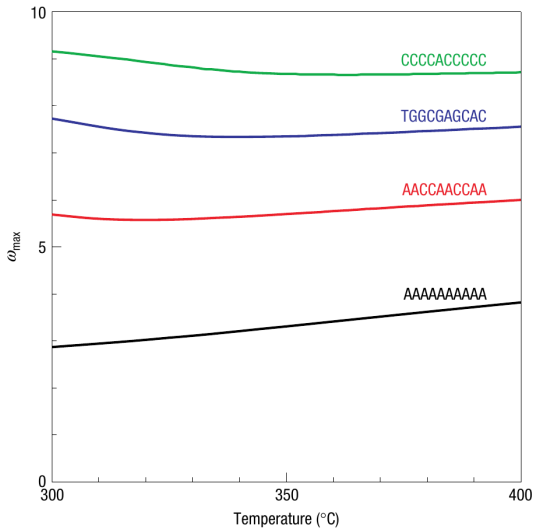

Figure 22 – $\omega_{\max}$ as a function of temperature for different sequences of length 10 bp. Figure taken from reference [112].

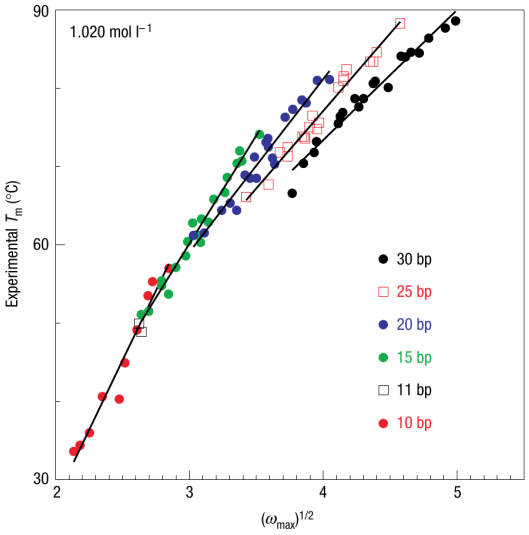

Figure 23 – Experimental melting temperatures T_m correlated to the order parameter $\omega_{\max}^{1/2}$. Figure taken from reference [112].

These calculations consider an initial set of parameters $\{p\}_0$, and these parameters are varied until we minimize the value of χ^2 [130]. After L steps of minimizations, $\{p\}_L$ contain a set of parameters which best reproduce the experimental melting temperatures.

In addition to Eq. 2.37, to ensure the quality of minimization, we also calculate the average absolute melting temperature deviation (average prediction difference)

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^{N} |T_i - T_i'(\{p\}_k)|. \tag{2.39}$$

Furthermore, root mean square of prediction differences is closely related to χ^2,

$$\Delta T_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [T_i - T_i'(\{p\}_k)]^2} = \sqrt{\frac{\chi^2}{N}}. \tag{2.40}$$

The minimization of Eq. (2.37) that leads to a low value of χ^2 and consequently, to a set of parameters $\{p\}_L$, is done via the multidimensional Nelder-Mead method (also called downhill simplex method) [131]. This method is a numerical method used to calculate the minimum or maximum of a function in a multidimensional space, and is often applied to nonlinear optimization problems, whose derivatives may not be known [131]. Each iteration begins with a simplex, a geometrical figure in n dimensions, of $n + 1$ points (or vertices), where these points are defined as possible values of the function, and all their interconnecting line segments, polygonal faces, etc [132]. In our calculations, each calculated parameter (Morse and stacking interaction potentials) is one dimension of the simplex, and the order of magnitude of these parameters represents the characteristic length of minimization [130]. Thus, the objective is to replace the worst point in the set, reducing the volume of simplex until to obtain the best (lower) value of the function.

The minimization procedure allows calculating the predicted melting temperatures, and inferring intramolecular properties for different configurations of nucleic acids. For example, analyzing canonical RNA sequences, the hydrogen bonds to AU (adenine and uracil) was found to be stronger than AT in DNA [126], which is consistent with experimental results from by Swart et al. [133].

There are several other mesoscopic models which consider additional degrees of freedom, for a comprehensive review see [134], however none of these were developed for high-throughput melting temperature calculation, and they would require additional structural parameters for which we do not have the necessary experimental data.

2.5 Conclusions

Here we revised the main models that are used in this thesis. We briefly revised the nearest-neighbor (NN) model that uses experimental data set to calculate the total entropy,

enthalpy, and free energy of each nucleic acid sequence, based on neighboring base pairs. However, the NN model do not provide information about intramolecular interactions of the sequences, so to describe these effects we choose the Peyrard-Bishop (PB) model. Next, we presented three approaches to the PB model for homogeneous and heterogeneous sequences, and via thermodynamic equivalence. This last approach is used in this project to calculate the Morse and stacking interaction potentials, as well as to predict the melting temperatures, and that will be presented in the next chapters.

3 DNA/TNA hybrids

TNA, α-L-(3′-2′)-threofuranosyl nucleic acid, is a type of modified nucleic acid. In this nucleic acid, the sugar is threose, whereas for DNA and RNA are deoxyribose and ribose, respectively. Furthermore, the orientation of strands is 3′ to 2′, while for DNA and RNA is 5′ to 3′, as shown in Fig 24. TNA was developed with the aim of being more stable and resistant to degradation under different physiological conditions. These features make TNA as a promising candidate for diagnostic and therapeutic applications. In this project, we use a mesoscopic analysis of measured melting temperatures to obtain an estimate of hydrogen bonds and stacking interactions, as well as to predict melting temperatures.

3.1 Introduction

The development of non-natural modified nucleic acid has gained traction in the early 2000s and are collectively known as xeno nucleic acid (XNA) [40, 135–137], that can have chemical modifications in nucleobases, phosphodiester backbone and/or pentose. Some XNAs with modifications in backbone structures result in nucleic acids with greater resistance to degradation by enzymes that capable of breaking the bonds between nucleotides, called nucleases [41, 138], when compared with the natural nucleic acid (DNA and RNA). The structures of DNA, RNA and an XNA called TNA (α-L-(3′-2′)-threofuranosyl nucleic acid) are compared in Fig 24. The essential difference is related to backbone composition, and this allows the TNA to have higher chemical stability, less reactive groups (a smaller number of hydroxyl groups), lower conformational flexibility, and resistant to degradation under physiological conditions [138, 139]. Despite having a shorter backbone, TNA is still able to form stable duplexes with DNA, RNA, and with itself, through Watson-Crick base pairing [140].

The sugar moiety [1] in TNA contains an unnatural four-carbon sugar of α-L-threose, composed by carbon atoms and a single oxygen atom, one atom shorter than DNA and RNA [143], which leads to a reduced conformational flexibility, making TNA more suitable

[1] The moiety is used to describe the larger and characteristic parts of organic molecules [142].

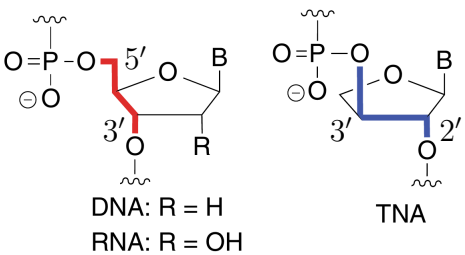

Figure 24 – Comparative between the chemical structures of DNA, RNA and TNA with B representing the nitrogenous bases. Figure taken from reference [141].

for information storage than DNA [139]. The phosphodiester groups are linked to the 3′ and 2′ positions of the threofuranose ring, without the presence of the methylene group $-CH_2$ [62], resulting in a shorter backbone than that of DNA [144], as shown in Fig. 24. Furthermore, TNA, like the others XNAs, can be formed by all five canonical purine and pyrimidine nucleobases.

Hybridization of DNA/TNA (DT) has the tendency to form A-type helices [145], similar to what is found in DNA/RNA (DR) [146–148]. This tendency was largely confirmed by nuclear magnetic resonance (NMR) and X-ray experiments [144, 145, 149]. These similarities between DT and DR have led to discussions if TNA might have been the natural precursor of RNA [150–152]. DT nucleotides were found to be in *anti* conformation in regard the β-glycosidic bond [153], but otherwise very little is known regarding their base-pair formation such as its hydrogen bonds and stacking interactions. The structure of TNA has already been studied by experimental techniques [141, 154, 155], however not yet by theoretical studies.

The reverse transcriptase (RT) is an enzyme that synthesizes a complementary DNA from natural RNA, which is used in some methods of molecular and synthetic biology, such as RT-qPCR (quantitative PCR with reverse transcription), RNA sequencing, and ribosome display [156]. The RT from the Moloney murine leukemia virus (MoMLV), known as Superscript II (SSII), can function as an efficient TNA-dependent DNA polymerase that could copy a short TNA template into DNA [157]. In another study, *Geobacillus stearothermophilus* (Bst) DNA polymerase was shown to be an efficient TNA-dependent DNA polymerase with activity superior to that of SSII [158]. The engineered TNA polymerase allows the isolation of functional TNA aptamers [2] with affinity to an RT found in the human immunodeficiency virus (HIV) [63]. These aptamers remain active in the presence of nucleases, exhibit higher thermal stability than monoclonal antibodies, and due to biological stability and high binding affinity, TNA aptamers are considered a powerful system for the development of diagnostic and therapeutic agents. A recent application for TNA aptamers is as an alternative molecular tool for the development of immunological inhibitors for cancer immunotherapy, due to significant tumor-growth inhibition [159, 160]. Furthermore, several ligases, enzymes that catalyze the bond between an oligonucleotide donor and an acceptor, can recognize TNA as the donor or acceptor strand with DNA, according to enzyme screening and reaction optimization [161].

3.2 Objective

TNA is an artificial genetic polymer with high potential for biological and biomedical applications and, therefore, the study of its structural properties is essential. For this, we

[2] Aptamer is a nucleic acid molecule that mimic antibodies and binds to specific targets such as nucleic acid or protein. [143].

Table 1 – Shown are the measured melting temperatures (T_i) from the references indicated (Ref.) and those marked with * are new unpublished sequences, and the predicted melting temperatures (T_i') in °C, for the DNA/TNA sequences (3′ to 2′) used in this work.

Sequence	Ref.	T_i	T_i'	Sequence	Ref.	T_i	T_i'
t(AGATACAA)	41	25.2	25.3	t(GAGGAATGACGT)	41	68.5	68.4
t(AATACAGA)	41	25.1	25.3	t(ATTCAGCG)	140	26.0	24.0
t(AAGCGTAG)	41	36.1	35.5	t(CGCTGAAT)	140	25.0	27.2
t(AGCGTAAG)	41	35.0	35.5	t(TTTTTTTTTTTTTTTT)	140	32.0	32.1
t(CTACGCTT)	41	20.3	21.2	t(TAATAATATAAATTTT)	140	47.0	42.8
t(CTTACGCT)	41	22.8	21.2	t(TTTTAAATATAATAAT)	140	43.0	42.8
t(AGTCCTGA)	41	20.0	21.3	t(AAAAAAAAAAAAAAAA)	140	68.0	68.0
t(CTGAGTCC)	41	22.9	21.8	t(AAAATTTATATTATTA)	140	41.0	40.8
t(GAGCCGTG)	41	40.8	40.0	t(ATTATTATATTTAAAA)	140	36.0	40.8
t(GCCGTGAG)	41	39.9	40.0	t(ATGGCGTGAC)	*	55.0	54.9
t(ACGTCATTCCTC)	41	44.6	45.0	t(CGCCTGTCTAGAAGTT)	*	62.0	60.7
t(GCAATGTTCAGC)	41	51.1	52.0	t(AACTTCTAGACAGGCG)	*	63.0	64.8
t(GCTGAACATTGC)	41	51.0	49.6				

consider the PB model via thermodynamic equivalence discussed in Section 2.4, which allows us to an estimate of hydrogen bonds and stacking interactions for DT hybrids based on the measured melting temperatures. We demonstrated that it is possible to run the model in reverse and to extract model parameters from the melting temperatures and in this way gain an insight of the intramolecular interactions [162]. With the new parameters obtained in this way we are able to predict melting temperatures of new sequences and also calculate a qualitative opening profile of the double-stranded helix as function of temperature.

3.3 Melting temperature data set

In this project, we used 8 DT sequences and melting temperature data set from Ref. 140, 14 from Ref. 41, and 3 new unpublished sequences. All data set was measured at 1.0 M NaCl (M = mol/L), 10 mM NaH_2PO_4, 0.1 mM EDTA, pH 7.0 and a strand concentration C_t of 10 μM. The complete list of sequences is shown in the Table 1.

Here, we consider only 25 DT sequences, and this limited data set is because TNA is not commercially available and is difficult to synthesize, and that is why there is relatively little information about its hybridization properties. The additional measurements (marked with * in Table 1) where necessary to increase the number of occurrences of base pairs, and were synthesized by Hershel H. Lackey and Jennifer M. Heemstra, our US international collaborators from University of Utah and Emory University, respectively. To consider the measured melting temperatures of DT hybrids from other studies, we need the sequences to be in a buffer identical to the one used in this project. For example, the study performed

by Yang et al. [151] cannot be considered, since the strand concentration is half that used. Furthermore, we analyze only DT hybrids due to the few denaturation measurements made with RNA/TNA and TNA/TNA, such as the study carried out by Schöning et al. [140] in which only 8 and 4 sequences were considered, respectively.

3.4 Notation

To facilitate the description about the hybrid DT duplex, we introduce the usual notation dA, dC, dG and dT for DNA bases and tA, tC, tG and tT for TNA bases. This is then used to describe the base pairs as dAtT, dTtA, dGtC and dCtG. The nearest-neighbor model allows us to define

$$\begin{matrix} {}^{3'}\mathrm{t(TA)}^{2'} \\ {}^{3'}\mathrm{d(AT)}^{5'} \end{matrix} \leftrightarrow \begin{matrix} {}^{3'}\mathrm{d(TA)}^{5'} \\ {}^{3'}\mathrm{t(AT)}^{2'} \end{matrix} \rightarrow \mathrm{dTtA\text{–}dAtT}$$

Note that due to symmetry considerations, dTtA–dAtT is same as tTdA–tAdT. Therefore, for simplicity, we always use the notation starting with deoxy (d). When grouping in terms of purine or pyrimidine bases, we use the notation dR (tR) for dA, dG (tA, tG) and dY (tY) for dT, dC (tT, tC). Note that some authors use Pu for R, and Py for Y.

3.5 Minimization procedure

The challenge of minimizing Eq. (2.37) is to find the optimal set P which in the case of DT consists of 20 parameters, 4 Morse and 16 harmonic (stacking interaction) potentials, using the Eqs. (2.5) and (2.23), respectively. The most important problem here is the occurrence of local minima of χ^2 and to overcome this a well tested approach is to perform the minimization procedure many times, each time starting from a different initial set of parameters, P_{init}, as discussed in Section 2.4.

The minimization procedure starts with the same generic parameters as used for DNA and RNA [126, 162], $D_{\mathrm{dAtT}} = D_{\mathrm{dTtA}} = 30\,\mathrm{meV}$, $D_{\mathrm{dCtG}} = D_{\mathrm{dGtC}} = 80\,\mathrm{meV}$, $\lambda_{\mathrm{dAtT}} = \lambda_{\mathrm{dTtA}} = 3.3333 \times 10^{-2}\,\mathrm{meV}$, $\lambda_{\mathrm{dCtG}} = \lambda_{\mathrm{dGtC}} = 1.25 \times 10^{-2}\,\mathrm{meV}$, and for all nearest-neighbors, $k = 2.5\,\mathrm{eV/nm^2}$. We call this set the seed parameters, P_{seed}, and for each new minimization we choose new initial set P_{init} where each initial parameter is randomly chosen within $\pm 10\%$ of their corresponding seed parameters. The λ parameters are kept at fixed values during all minimizations as we found that they have almost no influence over the final results [162], otherwise we would have to deal with yet another four parameters that would bring further difficulties for the convergence of the optimization procedure. In the initial minimization, we start from the generic seed parameters and only calculate the Morse parameters D, while fixing all others. We calculated the minimization starting with 1000 different sets of initial parameters, and the best value of χ^2 found

Table 2 – Merit parameters χ^2 and $\langle \Delta T \rangle$ in each minimization.

	χ^2 (°C^2)	$\langle \Delta T \rangle$ (°C)
1°	24068	26
2°	125	1.8
3°	67	1.1
4°	66	1.1

was 235 °C^2. In the second minimization, we calculated all parameters D and k, and use the Morse potentials obtained in previous minimization as seed parameters. Again, we started with 1000 different sets of initial parameters around ±10% of seed parameters. We then averaged the resulting D and k parameters and their quality parameters were $\chi^2 = 125\,°\text{C}^2$ and $\langle \Delta T \rangle = 1.8\,°\text{C}$. In the third minimization, we use the parameters from the second minimization as new seed values and calculated again 1000 rounds for further optimization. The quality parameters were reduced to $\chi^2 = 67\,°\text{C}^2$ and $\langle \Delta T \rangle = 1.1\,°\text{C}$. In the final minimization, we consider the influence of the experimental uncertainty on our optimized parameters. We modify the melting temperatures by small random amounts such that the standard deviation from the original set is 0.5 °C. Considering 1000 rounds of minimizations, we obtain the final average values of D and k as well as their relative uncertainty. The final quality parameters changed little compared to the previous round, $\chi^2 = 66\,°\text{C}^2$ and $\langle \Delta T \rangle = 1.1\,°\text{C}$.

3.6 Results

For the numerical optimization of the 20 new parameters: 4 Morse and 16 stacking interactions potentials, we had only 25 independent sequences and measured melting temperatures. For a linear system this would clearly be sufficient, but for a nonlinear model it is harder to establish *a priori* if overfitting is avoidable. The problem with overfitting is that the model performs better for the initial set of parameters but worse for new ones. A clear sign of overfitting is when the average temperature deviation of $\langle \Delta T \rangle$ results much smaller than the estimated experimental uncertainty. Fortunately, the resulting average temperature deviation was $\langle \Delta T \rangle = 1.1\,°\text{C}$, which compared to a typical uncertainty of 0.5 °C, gives us confidence that numerical overfitting has not occurred. Note that for the nearest-neighbor model, the amount of required parameters is 32 (16 enthalpy and 16 entropy variations) and calculating these would not be possible for such a small amount of melting temperatures [130]. In addition to these parameters, in each minimization, the melting temperatures are predicted, and in Fig. 25 we compare the predicted and measured melting temperatures in second and fourth minimization procedures. The difference between these melting temperatures becomes smaller with each new minimization, as demonstrated by merit functions χ^2 and $\langle \Delta T \rangle$, as shown in Table 2.

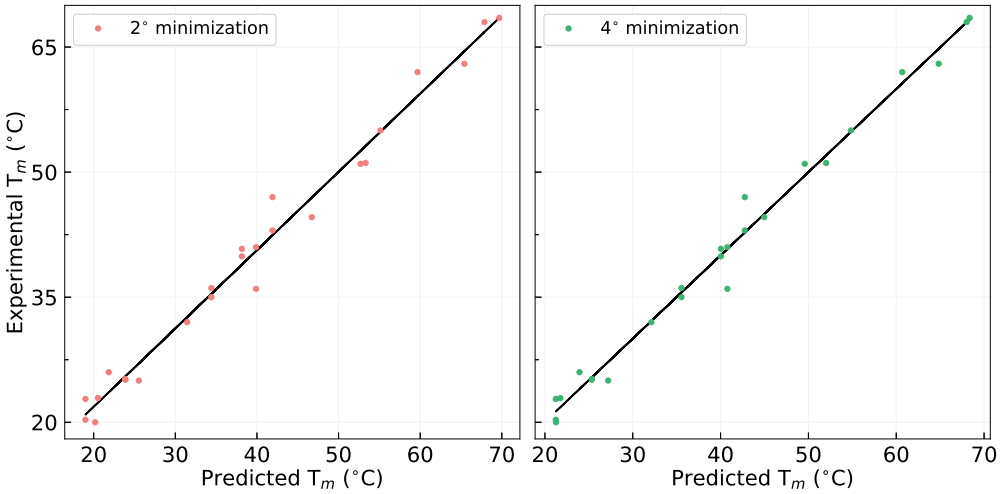

Figure 25 – Experimental and predicted melting temperature curves for DT hybrids in second and fourth minimization procedures.

In Table 3, we show the calculated Morse potentials, which can be associated to hydrogen bond strengths [162], for the DT hybrids. The uncertainties shown are calculated following the procedure described above, they are not related to the numerical inaccuracy of the minimization algorithm but represent the influence of the experimental uncertainty over the new parameters. The TNA purine base (tR) clearly has a stabilizing influence as shown by the much stronger Morse potential of dTtA and dCtG over the pyrimidine base (tY), which suggest that this known purine dependence [41] is primarily due to hydrogen bonding. Note that dGtC with a tY base has a Morse potential which is barely larger than the dTtA base pair, which indicates that for certain applications there might be a problem discriminating between these types of base pairs. Also shown in Table 3 are the Morse potentials obtained for DR hybrids in Ref. 148. The comparison between the DT and DR Morse potentials shows two interesting results. First, the very close similarity of the dTtA and dAtT Morse potentials to their DR counterparts, indicating that the dT and dA bases accommodate really well in the TNA strand. This is consistent with the results from [144] who observed this for dAtT and dTtA base pairs embedded in a B-DNA and A-DNA, respectively. However, the second result is probably more surprising: a comparatively weak Morse potential for the CG-type DT base pairs. Even the stronger dCtG potential falls well below the weaker DR dGrC potential. The dCtG and dGtC potentials obtained are unexpected, since the CG base pair is formed by three hydrogen bonds, which gives it greater thermal stability than AT. In addition, one of the characteristics of TNA that makes it a good candidate for therapeutic applications is its high biological stability.

The calculated stacking potential constants k are shown in Table 4, and their values range between 1.48 eV/nm^2 for dCtG-dAtT to 3.77 eV/nm^2 for dAtT-dCtG. These parameters represent the base stacking interactions that influence in nucleic acid stabilization, and demonstrate the sequence-dependence property [163]. When we compared the DT

Table 3 – Morse potential depth D in meV optimized for DT hybrids, compared with results for DR [148] base pairs. The standard deviation is displayed in parentheses in compact uncertainty notation.

type	DT	D (meV)	type	DR	D (meV)
dYtR	dTtA	38.8(4)	dYrR	dTrA	40(2)
dRtY	dAtT	30.7(8)	dRrY	dArU	28(3)
dYtR	dCtG	53.5(5)	dYrR	dCrG	74(1)
dRtY	dGtC	42.0(3)	dRrY	dGrC	63(1)

Table 4 – Harmonic potential coupling constant k in eV/nm^2 optimized for nearest-neighbors base pairs of DT hybrids. The standard deviation is displayed in parentheses in compact uncertainty notation.

DT groups	NN	k (eV/nm^2)	NN	k (eV/nm^2)
dRtY-dRtY	dAtT-dAtT	3.02(7)	dGtC-dAtT	1.81(6)
	dAtT-dGtC	2.17(7)	dGtC-dGtC	2.43(11)
dRtY-dYtR	dAtT-dCtG	3.77(15)	dGtC-dCtG	2.18(6)
	dAtT-dTtA	2.48(10)	dGtC-dTtA	3.01(10)
dYtR-dYtR	dCtG-dCtG	2.62(14)	dTtA-dCtG	2.47(6)
	dCtG-dTtA	1.62(5)	dTtA-dTtA	3.14(5)
dYtR-dRtY	dCtG-dAtT	1.48(7)	dTtA-dAtT	2.35(7)
	dCtG-dGtC	2.70(11)	dTtA-dGtC	2.94(8)

stacking potentials and the values of k for DR in Ref. 148, we found no correlation as expected, since TNA may be a potentially natural derivative of RNA due to the simplicity of its structure and ability to exchange genetic information with itself and with RNA [164]. Unlike the DR stacking potentials, which are quite small in some specific cases, we obtained no such weak potentials for DT. In fact, the higher DT stacking appears to counteract the weaker CG Morse potentials in several cases and provides additional thermal stability.

The interplay of the various potentials within the nonlinear PB model is better appreciated by examining the average displacement profiles for some example sequences using the Eq. (2.29). Average displacement profiles are obtained from calculating the expected value $\langle y_i \rangle$ of the y distance of Eq. (2.5) for the ith base pair [122]. Analyzing these profiles, we can detect regions prone to greater instability, along with the dynamics of the denaturation process [129], and they can still be interpreted as the expected base pair separation at certain temperatures. Unfortunately, as the PB approach is a 2D model, some relevant degree of freedoms are absent from the thermodynamics and for very short sequences one has to set unrealistically low calculation temperatures. Note that this calculation temperature is unrelated to the melting temperatures calculated from Eq. (2.35). To overcome the problem of short sequences we used the melting index τ to predict the melting temperatures at which the DNA sequence melts over a very short range of temperatures, that is, its melting can be described as a two-state helix denaturation [128].

Table 5 – DD, DR, and DT sequences from Ref. 165 used for the calculation of average displacement profiles in Fig. 26, where the upper (pink) and lower (black) axis represent the pyrimidine and purine rich target sequences, respectively.

Target	DD	DR	DT
Pyrimidine rich	d(TACCCTATAT)	r(UACCCUAUAU)	t(TACCCTATAT)
	d(ATGGGATATA)	d(ATGGGATATA)	d(ATGGGATATA)
Purine rich	d(ATGGGATATA)	r(AUGGGAUAUA)	t(ATGGGATATA)
	d(TACCCTATAT)	d(TACCCTATAT)	d(TACCCTATAT)

Therefore, the displacement profiles only give us a qualitative view of the expected base pair openings which, nevertheless, is helpful to understand the detailed influences the various potentials have over the stability of the duplex. For the average displacement profiles shown in Fig. 26, we show six sequences adapted from Table S1 of Ref. 165, all are hybridized to a probe DNA strand, and described in Table 5. The complementary DNA, RNA and TNA are target sequences which are either purine or pyrimidine rich. For DNA/DNA (DD) sequences, the purine content has little impact, as these are homo-duplexes, as seen in Fig. 26a, at higher temperatures the displacement profiles are still very similar. The situation changes completely for DR, at higher temperatures the purine rich target sequence is much more stable (lower $\langle y_i \rangle$), even more so than is DD equivalent, Fig. 26b. This purine/pyrimidine asymmetry is a well established property of DR hybrids [148, 166] and has important consequences in several biological processes as a key intermediate in replication and recombination [167]. For DT hybrids, the thermal stability shows a similar asymmetry to purine/pyrimidine content to DR as shown in Fig. 26c. However, unlike DR, the DT shows a much more flattened out profile, and this can be explained by the proximity of the stacking potentials, as shown in Table 4. While it is easy to make out the CG base pairs from the opening profile for DR, for DT it is much harder to tell where the CG base pairs are from the profile alone.

The other average displacement profiles are shown in Fig. 27, whose sequences are described in Table 6, where the percentage of deoxypyrimidine (dPy) content (cytosine and thymine) for DR and DT is varying from 0% to 100%. In DR profiles, we can see that the thermodynamic properties varying with dPy content, as confirmed by Suresh and Priyakumar [166] via molecular dynamics (MD) simulations. Furthermore, DR profiles has a characteristic bulged opening in the central positions due to AU base pairs. This behavior is related to the low value for dArU-dArU (0.9 eV/nm^2) [148], which generates a weak interaction between these base pairs, easily broken during denaturation. In contrast, for equivalent DT sequences the profiles are almost flat, without any such pronounced features. Also, for 0% dPy (100% tPy) content the DT sequence shows much higher opening implying in a much weaker thermal stability, which correlates to a very large dissociation rates observed for large tPy content [165].

We can also calculate the average displacement profiles for DT varying the cal-

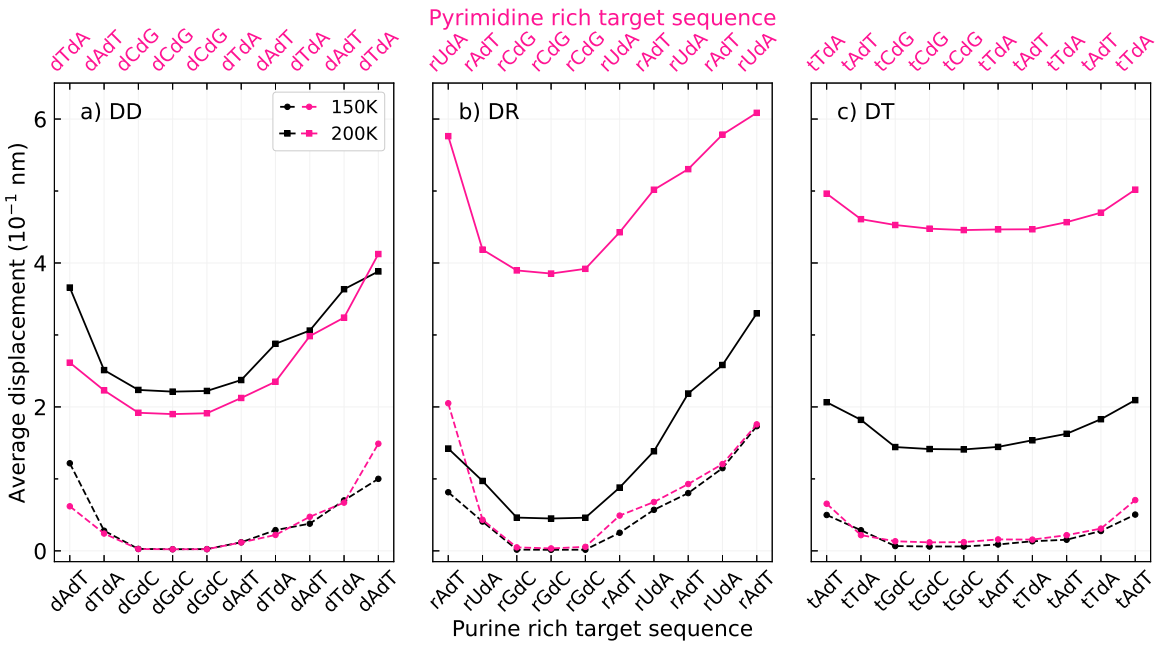

Figure 26 – Average displacement profiles with temperature dependence for (a) DD, (b) DR and (c) DT purine rich (black) and pyrimidine rich (pink) targets adapted from Ref. 165. Bullets connected by dashed lines are for a calculation temperature of 150 K and squares connected by full lines are for 200 K.

Table 6 – DR and DT sequences from Ref. 166 used for the calculation of average displacement profiles in Fig. 27.

dPy (%)	DR	DT
0	d(GGGAAAAAAGGG)	d(GGGAAAAAAGGG)
	r(CCCUUUUUUCCC)	t(CCCTTTTTTCCC)
25	d(CGCAAAAAAGCG)	d(CGCAAAAAAGCG)
	r(GCGUUUUUUCGC)	t(GCGTTTTTTCGC)
50	d(CGCAAATTTGCG)	d(CGCAAATTTGCG)
	r(GCGUUUAAACGC)	t(GCGTTTAAACGC)
75	d(GCGTTTTTTCGC)	d(GCGTTTTTTCGC)
	r(CGCAAAAAAGCG)	t(CGCAAAAAAGCG)
100	d(CCCTTTTTTCCC)	d(CCCTTTTTTCCC)
	r(GGGAAAAAAGGG)	t(GGGAAAAAAGGG)

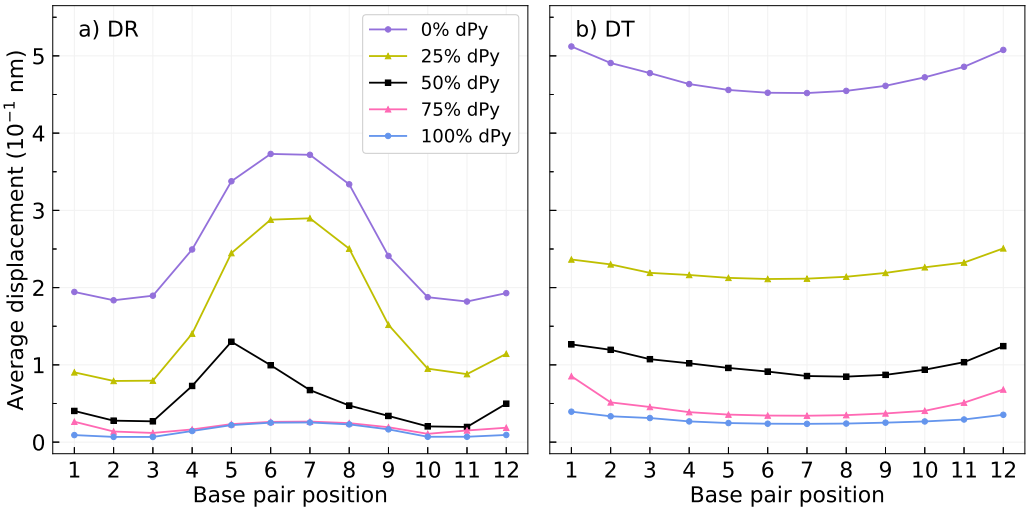

Figure 27 – Average displacement profiles for (a) DR and (b) DT with varying deoxypyrimidine (dPy) content, calculated at 200 K. Sequences were adapted from Ref. 166, and calculations for DR are from Ref. 148.

culation temperatures (parameter β in Eq. (2.6)), in a range of 150 to 200 K, as shown in Fig. 28. The complementary sequences (2′-3′) are `t(GGCATTACGG)` from Ref. 145, and `t(ACTGATGTTGA)` from Ref. 168. In Fig. 28a, the curves are almost flat even with increasing temperature. However, in Fig. 28b, we see a different behavior, and this may be due to the base pairs that form the sequence, as well as the values of the stacking potentials k of neighboring base pairs. For example, the last ks in the sequence are dAtT-dCtG (3.77 eV/nm^2) and dCtG-dTtA (1.62 eV/nm^2), whose difference in values exceeds a factor of two, in contrast to the first sequence, in which the interactions have very close values, and therefore we do not see significant changes. Thus, analyzing both sequences, independent of the position and type of base pairs, when the temperature increases, the shape of the curves is preserved, and at higher temperatures lead to large openings, due to high values of average displacements. This is a characteristic found in DT hybrids. In addition, these profiles allow us to determine if there were any inconsistency in our results, such as the opening of a base pair that has not yet been observed in other research in similar systems [129].

3.7 Conclusions

The DNA/TNA mesoscopic parameters, which can be associated to hydrogen bonding and stacking interactions, were calculated based on the measured melting temperatures. TNA has a backbone composition formed by a simple sugar with fewer reactive groups and less conformational flexibility, which contributes to its greater stability than DNA and RNA, that is, the interactions between the nitrogenous bases are significantly stronger, as demonstrated for the obtained results. Moreover, these results also confirm the expected asymmetry of the thermodynamic properties in regard to purine/pyrimidine content in

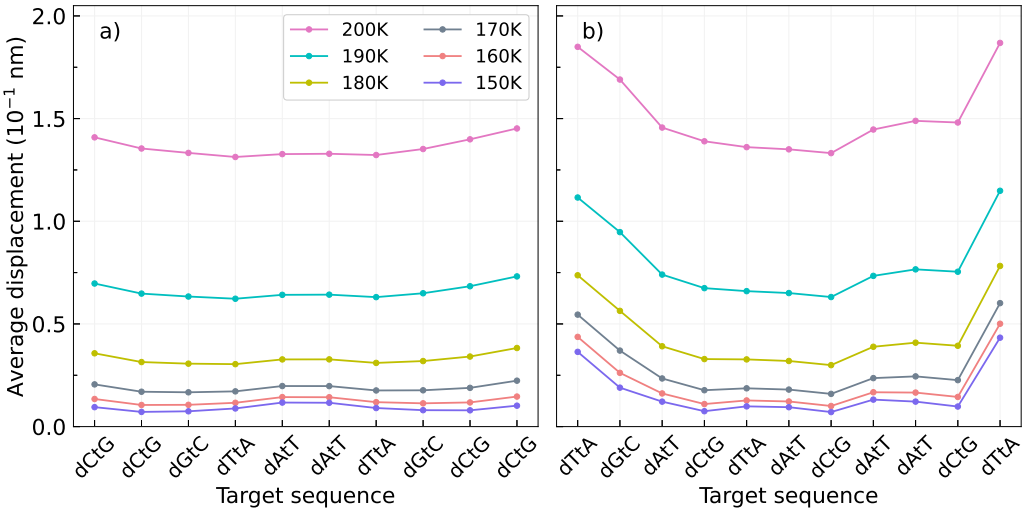

Figure 28 – Average displacement profiles for (a) `t(GGCATTACGG)` from Ref. 145, and (b) `t(ACTGATGTTGA)` from Ref. 168. For the same sequence, we have considered different temperatures, in a range of 150 to 200 K.

DNA/TNA, similar to those of DNA/RNA hybrids. However, the Morse potentials for CG base pairs for DNA/TNA is very small when compared with DNA/RNA. The consequence of these small potentials is that for DNA/TNA hybrids the duplex appears to dissociate much for uniformly and largely independent of sequence composition. Furthermore, to stacking potential, unlike the DNA/RNA, which are quite small in some specific cases, we obtained no such weak potentials for DNA/TNA. Our results demonstrate the effectiveness of the thermodynamic equivalence methodology to determine the properties and stability of DNA/TNA hybrids. As researchers continue to explore the use of TNA in synthetic biology and nanotechnology applications, these data provide important insight into the design of sequences and prediction of duplex stability. After we published our results [48] a number of articles have already cited our work Refs. 169–173.

4 DNA in solutions containing Mg^{2+}

Denaturation is a biological process considered in several sophisticated applications such as high-resolution melting analysis, PCR, and gene therapy. Thus, one important approach is to understand about the effects of various types of mono- and divalent cations in DNA denaturation. In this project we evaluate DNA melting temperatures in magnesium and magnesium-potassium buffers with a mesoscopic model via thermodynamic equivalence to estimate hydrogen bonds and stacking interaction potentials. These results are compared to previous calculations for buffers containing sodium ions, in terms of equivalent sodium concentration and ionic strength.

4.1 Introduction

Over the years, the effects of various cations in nucleic acids have been extensively studied via experimental and theoretical techniques, such as Fourier transform infrared spectroscopy [174], ultraviolet-visible (UV-Vis) spectroscopy [175], Raman microspectroscopy [176], optical tweezers [177, 178], Monte Carlo simulation [179], molecular dynamics (MD) [180, 181], Poisson-Boltzmann polyelectrolyte theory [182]. An example is the study carried out by Gebala and Herschlag [183] using the nonlinear Poisson-Boltzmann equation, which showed that RNA attracts more monovalent cations than DNA, and has a more intense magnesium ions (Mg^{2+}) interaction, implying a stronger electrostatic field for RNA than that DNA. These results are confirmed by Xi et al. [184] via MD simulations, and they also showed that the binding of sodium ions (Na^{+}) to RNA is slightly weaker than that to DNA. These differences can be attributed to the tendency of RNA duplex to form A-type helix, while DNA forms B-type, and the phosphoryl groups in DNA duplex are oriented towards the solvent, whereas for RNA they face inwards. We can also investigate where the cations are located in the double-stranded helix structure, as shown by Li et al. [185] in a MD simulation study for a specific 23 base pair (bp) DNA duplex. They observed that Mg^{2+} ions were found at the phosphate backbone and in the major groove of GC base pair, where the interaction occurs mainly through hydrogen bonds, and cannot penetrate in minor groove, because the molecule is large. In contrast, AT base pair have no good electrostatic environment for Mg^{2+}, leading to a low occupancy in both grooves. For Na^{+} ions, the interaction with CG and AT base pairs occur in both grooves, where the occupancy in major is larger than that minor groove, such as in CG which is larger than AT base pair. Other MD simulations, such as from Mukherjee and Bhattacharyya [186], concluded that Na^{+} binds largely to the minor groove, while potassium ions (K^{+}) binds to the major groove and closer to the center of the duplex. The MD simulations by Lavery

et al. [187] showed that K^+ is localized, most of the time, within the major groove. More recently Kolesnikov et al. [72], compare the MD calculations of Na^+ and K^+, showing that the binding affinity of monovalent ions has an important dependence on the solvent model when these ions are deeply bound. For Poisson-Boltzmann polyelectrolyte calculations by Misra and Draper [188], they observed an anticooperativity effect for Mg^{2+} concentrations higher than 10 mM caused by a saturation of the ion binding to DNA, and according to Gebala and Herschlag [183] Na^+ interacts with GC and AT base pairs in both grooves, being the occupancy in the major groove larger than the minor one. Furthermore, in relation to the structure, the binding of Mg^{2+} to DNA was found to be more rigid than for Na^+ bound to DNA [186]. In another study, Ferreira et al. [189] used the PB model via thermodynamic equivalence to separate the effects of internal and terminal base pairs. Differently from internal base pairs, terminal pairs do show some dependence on Na^+ concentrations. Being able to describe terminal base pairs is a unique feature of the mesoscopic model. In comparison, even the versatile MD method is limited in its ability to deal with terminal base pair interactions [190], especially terminal AT base pairs.

A motivation of these researches is due to the presence of cations in cells and in molecular biology techniques. For instance Mg^{2+} is the most abundant divalent cation in living organisms and essential for many biological processes [191], such as PCR, since DNA polymerase is an enzyme that is magnesium dependent [192], and as a cofactor for adenosine triphosphate (ATP) [193]. Furthermore, a biological process also influenced by the interaction of cations and nucleic acids is denaturation. According to Thomas [194] to maintain native DNA at room temperature, the concentration of NaCl should be 100 times higher than that of $MgCl_2$. As a result, the denaturation temperature in the presence of $MgCl_2$ is higher than NaCl at the same cation concentration [195]. This is due to Mg^{2+} reducing the repulsive Coulomb interactions between phosphate groups [196, 197], which neutralizes the negative charges on the DNA backbone [198]. Similar conclusions were reached in several comparative studies of how Na^+ and Mg^{2+} influences the denaturation [75, 199, 200].

Early UV-Vis spectroscopy [74] observed that the melting temperature of DNA increases with Mg^{2+} concentration and concluded that they bind to the phosphates. However, if the Mg^{2+} concentration exceeds the total strand concentration, a destabilization occurs as observed by Baba and Kagemoto [196] with differential scanning calorimetry (DSC). The infrared spectroscopy according to Serec et al. [174,201] concluded that high concentrations of Mg^{2+} do not affect stacking interactions. Every and Russu [202] in particular concluded from NMR measurements that Mg^{2+} may lead to increased GCpCG openings. Using buffer equilibration–atomic emission spectroscopy ion counting techniques, Bai et al. [203] analyzed competitive binding, in particular between monovalent Na^+ and divalent cations Mg^{2+}. They have shown that already at moderate concentration of Mg^{2+} of 10 mM the Na^+ have been out competed entirely, which was further confirmed by Xi et al. [184]. In comparison to RNA, DNA has a number of differences when it comes to ionic interactions.

4.2 Objective

Currently, there is no clear picture of how the ion valence affects the hydrogen bonds and stacking in DNA and some open questions remain. For instance, we have seen previously that monovalent Na^+ concentration has little effect on hydrogen bonding, would this be different for divalent cations? To answer this and other questions, we use this mesoscopic approach to study the effects of internal and terminal base pairs to divalent cations Mg^{2+}, as well as mixed mono- and divalent $Mg^{2+}+K^+$ buffers. These studies together with the previous studies on Na^+ developed by Ferreira et al. [189], allow us to draw a comprehensive picture of the differences between the various types of buffers and how they affect the thermal stability of DNA. Therefore, ionic strength is not regarded as a good predictor of melting temperatures. To overcome this problem, the concept of the sodium equivalent is used instead [204, 205], that is, the concentration of Na^+ required to result in a similar stabilization as Mg^{2+}. Therefore, using multiple types of buffers allow us to understand how the ionic strength and sodium equivalent relate to the hydrogen bonds and stacking interactions, as well as how they affect the terminal base pairs.

4.3 Melting temperature data set

For the results to be reliable, it is important to have an expressive set of experimental data. So in Refs. [205,206] we found several sequences in Mg^{2+} and $Mg^{2+}+K^+$ buffers. These data are interesting due to the composition of the buffers by various Mg^{2+} concentrations.

Mg^{2+} The main dataset related to variable Mg^{2+} concentrations was drawn from Refs. [205, 206] consisting of 92 sequences. In addition, 10 new unpublished sequences were synthesized by Adrian H. Bustos, Sofie Slott and Kira Astakhova from Technical University of Denmark, and measured such as to complement the existing dataset, especially in regard to terminal AT base pairs and the under-represented set of sequences of 11 bp length. These sequences were synthesized using automated solid-phase synthesis procedure in 200 nmol scale on ABI equipment (Expedite). After the synthesis, the sequences were cleaved with aqueous ammonia and analyzed by HPLC and MALDI MS. The purity of the sequences was $> 85\%$ by IE HPLC. Their melting temperatures were measured on a DU800 UV/VIS spectrophotometer equipped with a Beckman Coulter Performance Temperature Controller. Complementary strands (0.5 μM of each strand), in a 1×PBS were mixed, denatured 10 min at 90 °C and subsequently cooled to 15 °C. The measured melting temperature values present the maximum of the first derivative of the curve and are an average of the two measurements with melting temperature result deviation within 1 °C. The measured melting temperatures are summarized in Table A.1 for the new sequences in Appendix A. The 92 DNA sequences from Refs. [205, 206] and their melting temperature data, are all at 0.5–125 mM $MgCl_2$, 2 mM or 10 mM Tris-HCl, pH 8.3 and with total strand

concentration C_t of 2 μM. Their respective measured and predicted melting temperatures are shown in Table A.2 in Appendix A. We will refer to the combined data of Tables A.1 and A.2 as the Mg^{2+} set.

$Mg^{2+}+K^+$ The second melting temperature data set has 80 DNA sequences from Ref. [206], all at 0.5–20 mM $MgCl_2$, 10 mM Tris-HCl, 50 mM KCl, pH 8.3 and a strand concentration C_t of 2 μM. We will refer to these melting temperature data as the $Mg^{2+}+K^+$ set. The complete list of sequences and their respective measured and predicted melting temperatures are shown in Table A.3 in Appendix A.

4.4 Na^+ equivalence

The concept of equivalent or effective sodium concentration $[Na^+_{eq}]$ was introduced by Mitsuhashi [207] to convert the concentration of Na^+ in a buffer that stabilizes duplexes to the same extent as the Mg^{2+} buffer [205]. This relation can be expressed as

$$[Na^+_{eq}] = \beta\sqrt{[Mg^{2+}]} + [Mon^+] \tag{4.1}$$

where the parameter β has values between 3.3 to 4.0 $M^{1/2}$ (M = mol/L), depending on the type of salt correction that was applied [204, 205, 207], and $[Mon^+]$ is the sum of concentrations of monovalent cations. The squared root factor term is responsible for the law of mass action for cation binding, where the binding rate of monovalent cations is expected to increase proportionally to the square of the binding rate of divalent cations. This relation express that Mg^{2+} stabilize duplexes more effectively than Na^+ with the same concentration. $[Na^+_{eq}]$ are inserted into the T_m salt correction for Na^+ to correct melting temperatures for Mg^{2+} buffer. Here, we used three specific values of the equivalence parameter β: 3.3 $M^{1/2}$ [205], 3.79 $M^{1/2}$ [204] and 4 $M^{1/2}$ [207]. The $[Na^+_{eq}]$ values related to the Mg^{2+}, $Mg^{2+}+K^+$, and Na^+ sets are shown are in Table A.4 in Appendix A, along with ionic strength (I).

4.5 Notation

We describe the uniform, internal and terminal base pairs terminology. For uniform base pairs we make no distinction of their position in the sequence. In the example below we have a system with two base pairs (AT and CG) and 10 nearest-neighbor interactions as typically used in nearest-neighbor models [208], corresponding to two Morse potentials D and 10 stacking interaction parameters k. Terminal base pairs are considered as different from internal base pairs [189] and marked with * to differentiate from internal base pairs.

The following example illustrates both uniform and internal/terminal terminology,

$$\begin{array}{l} 5'\text{-AGCGTAAGTC-}3' \\ 3'\text{-TCGCATTCAG-}5' \end{array} \longleftrightarrow \begin{array}{c} \text{AGCGTAAGTC} \\ \underbrace{\text{TCGCATTCAG}}_{\text{uniform AT or CG}} \end{array} \longleftrightarrow \begin{array}{ccc} \text{A} & \text{GCGTAAGT} & \text{C} \\ \underbrace{\text{T}}_{\text{term. AT*}} & \underbrace{\text{CGCATTCA}}_{\text{internal AT/CG}} & \underbrace{\text{G}}_{\text{term. CG*}} \end{array}$$

where we have an AT* at the 5′-terminal and a CG* at the 3′-terminal. The internal/terminal scheme adds further 18 parameters to be optimized, of which two are Morse potentials D and 16 are stacking interaction parameters k, resulting in 30 parameters in total.

For Morse potential, the base pairs AT* and TA*, and CG* and GC* are considered symmetrical and consequently have the same parameters D and λ [189]. However, for the stacking interaction potential, we have a mixed notation of terminal and internal base pairs. In our example, for the stacking interaction parameter, the first nearest-neighbor is defined as AT*pGC with terminal AT and internal GC. The AT*pGC is symmetric to CGpTA*,

$$\begin{array}{l} 5'\text{-AG CGTAAG TC-}3' \\ \underbrace{3'\text{-TC}}_{\text{AT*pGC}}\ \text{GCATTC}\ \underbrace{\text{AG-}5'}_{\text{TApCG*}} \end{array} \longleftrightarrow \begin{array}{l} 5'\text{-GA CTTACG CT-}3' \\ \underbrace{3'\text{-CT}}_{\text{GC*pAT}}\ \text{GAATGC}\ \underbrace{\text{GA-}5'}_{\text{CGpTA*}} \end{array}$$

Thus, we will only consider the notations in alphabetical order. Furthermore, for the stacking interaction parameter ATpGC we have three notations, AT*pGC with terminal AT, ATpGC* with terminal GC, and ATpGC without distinguishing terminal and internal base pairs. For some cases due to symmetry of nearest-neighbor pair, as CGpGC, that has only one terminal, CGpGC*, symmetric to CG*pGC [189]. To ease the notational burden, occasionally we will refer to the various stacking configurations collectively, for instance by AT-AT we mean ATpAT, TApAT and ATpAT; by CG-CG we mean CGpCG, GCpCG and CGpGC; by AT-CG we mean ATpCG and ATpGC; by GC-AT we mean CGpAT and GCpAT.

4.6 Minimization procedure

To obtain the optimal set of parameters, P, to minimize the Eq. (2.37), we calculate 30 DNA parameters for each Mg^{2+} and $Mg^{2+}+K^{+}$ buffer: 4 Morse potentials and 26 stacking interaction potentials, where the potentials differ in the internal and terminal base pairs. The calculations are done using the downhill simplex algorithm, which has a local minima occurrence problem, and to solve it, we perform the minimization procedure many times, with different initial set of parameters, P_{init} in each round, as discussed in Section 2.4.

For the first round of minimizations (R1), all calculations draw their initial parameters $P_{\text{init}}^{\text{R1}}$ from the seed parameters shown in Tables A.5 and A.6 in the Supporting Material, in the following way: for each minimization we randomly chose a fresh initial set

Table 7 – Merit parameters $\langle \Delta T \rangle$ in °C for Mg^{2+} and $Mg^{2+}+K^+$ sets.

Mg^{2+} concentration:	0.5 mM	1.5 mM	3 mM	10 mM	20 mM	50 mM	125 mM
Mg^{2+} set	0.69	0.68	0.70	0.67	0.67	0.60	0.69
$Mg^{2+}+K^+$ set	0.82	0.74	0.70	0.62	0.66		

$P_{\text{init}}^{\text{R1}}$ from within ±20% of their corresponding seed parameters P_{seed}. This is performed 1000 times, and we take the resulting parameters and calculate their averages over all calculations, which we call resulting averaged parameter set $P_{\text{res}}^{\text{R1}}$. Next, we repeat this procedure for a second round (R2), but now we draw the new initial parameters $P_{\text{init}}^{\text{R2}}$ from within ±10% of $P_{\text{res}}^{\text{R1}}$. This is repeated another 1000 times and results in the averaged parameter set $P_{\text{res}}^{\text{R2}}$. To validate the set of parameters, it is important to consider the influence of the experimental uncertainty, so in the final minimization (R3) we include the standard deviation from the original set, reported in [205, 206] as 0.3 °C. This time the initial parameters are fixed at $P_{\text{res}}^{\text{R2}}$, and we vary the experimental temperatures T_i by small random amounts such that the standard deviation of the new set of temperatures falls within 0.3 °C of the original temperatures. R3 is repeated 1000 times and the final results presented here are the averaged parameters $P_{\text{res}}^{\text{R3}}$ and the error bars in the figures are the standard deviations of $P_{\text{res}}^{\text{R3}}$. Each of the 12 buffer conditions is evaluated independently through rounds R1–R3. Final R3 values of χ^2 and $\langle \Delta T \rangle$, from Eqs. (2.37) and (2.39), are shown in Tables A.7 and 7.

4.7 Results

The parameter optimization was carried out in a same way as in our previous work on Na^+ in DNA [189], except that this time we performed five time as many minimizations, which was made possible by code optimizations in the software that was used for the calculations. Final merit parameters $\langle \Delta T \rangle$, Eq. (2.39), are shown in Table 7 and are roughly twice as large as the reported experimental uncertainty of the data used [205], which is very similar to our previous results [162, 189]. Other melting temperature optimization approaches, such as by Freeman et al. [209] where coarse-grained models where employed with explicit ions, do not achieve this level of agreement to the experimental data. At each minimization we predict the melting temperatures, and this allows us to compare them with the experimental ones, as shown in Fig. 29 for Mg^{2+} and $Mg^{2+}+K^+$ buffers. Analyzing the curves of each buffer, the difference between the predicted and experimental melting temperatures is smaller with each new minimization, which demonstrates the accuracy of the methodology. Furthermore, in Fig. 30 (based on Fig. 23), we have these experimental melting temperatures as function of melting index ($\omega_{\text{max}}^{1/2}$) also for Mg^{2+} and $Mg^{2+}+K^+$ buffers. The points show linear relations through each group of the same

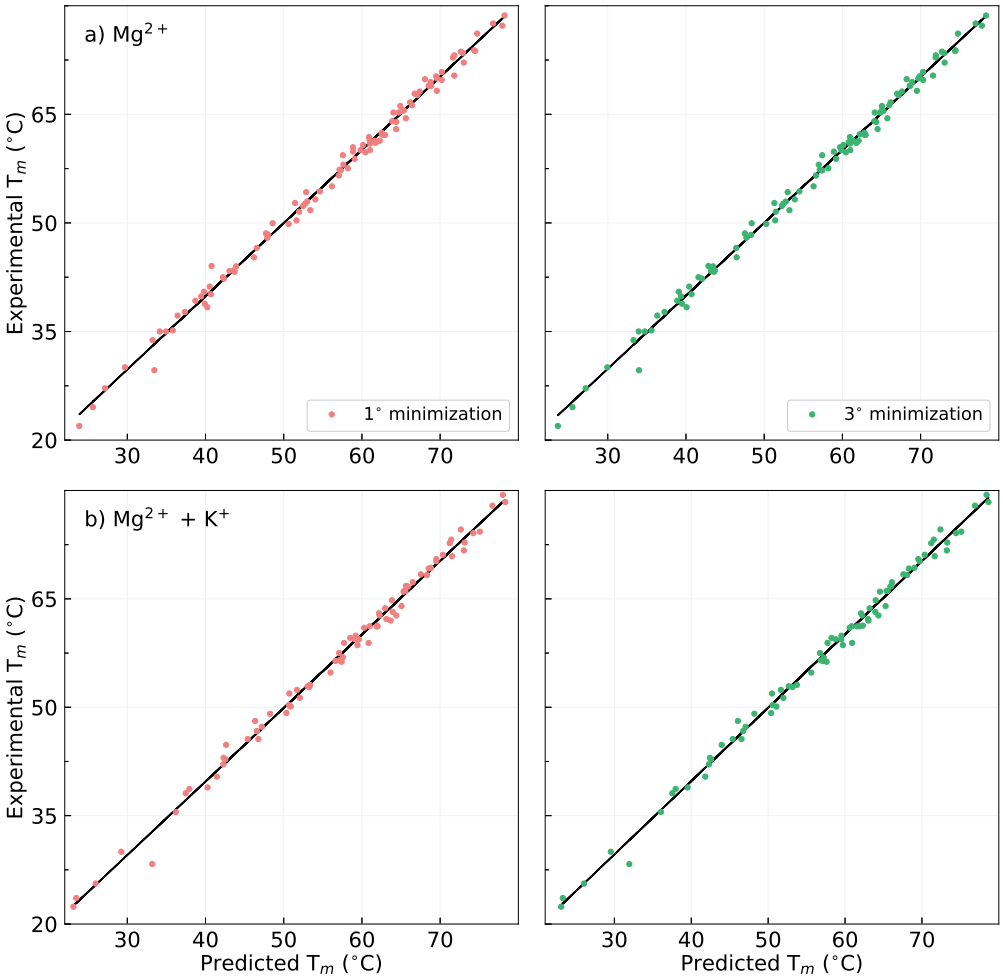

Figure 29 – Melting temperature experimental and predicted curves for 0.5 mM Mg^{2+} with 50 mM or no KCl.

sequence size, as well as a linear dependence with $\omega_{\max}^{1/2}$ [112], demonstrating that $\omega_{\max}^{1/2}$ is a convenient quantity for comparing DNA sequences.

One of the main questions we wish to address is how much the Morse potential depends on Mg^{2+} or $Mg^{2+}+K^+$ concentrations. Would it be similar to Na^+ or does cation valence play some role? To answer this question we show the results for Mg^{2+} and $Mg^{2+}+K^+$ alongside the previous results for Na^+ from [189] in Fig. 31. For CG base pairs we observe very little difference in regard of cation type, whether we plot the results as function of Na_{eq}^+ concentration Eq. (4.1) or as function of ionic strength Eq. (1.1). For CG* base pairs we observe a discrete reduction for all buffer types as shown in Figs. 31(b,d). Only terminal base pairs do display a moderate dependency on salt concentration, however terminal base pairs have a minor influence over the sequence in general. For AT base pairs we observe an overall similar dependence for all three buffer types, only for Mg^{2+} it appears slightly reduced for some concentrations for internal base pairs as shown in Figs. 31(a,c). In general, we may conclude that one may extrapolate the use of a constant D to any Na_{eq}^+ concentration and cation type. The weak dependence of internal D with any type

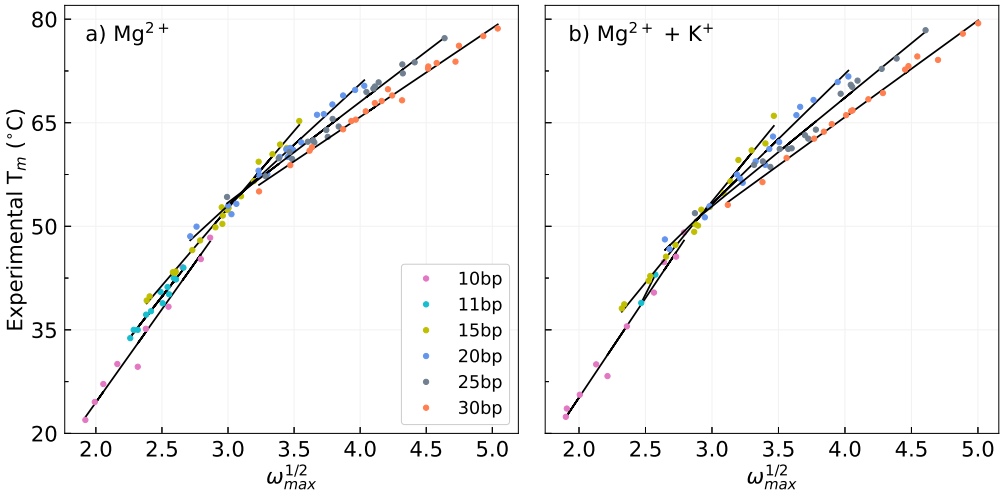

Figure 30 – Experimental melting temperatures as function of melting index ($\omega_{\max}^{1/2}$) for different sequences length in 0.5 mM Mg^{2+} buffer with 50 mM or no KCl.

of cation in DNA is different from Na^+ in RNA [210] where a larger dependence on salt concentration was found. The near complete independence of the Morse potential depth with valence and concentration suggest that there is no fundamental difference in the way Na^+ and Mg^{2+} influence the hydrogen bonds. This is consistent with molecular dynamic simulations which do not report important changes to the base pair hydrogen bonds Li et al. [185]. Furthermore, in Fig. 37, for uniform AT and CG base pairs as function of Na_{eq}^+ concentration, where we do not distinguish the position in the sequence, the behaviors are very similar to those shown by internal base pairs, and these are due to the fact that most base pairs are in internal positions. Note that, differently from simulations, our results are directly obtained from experimental data.

Another interesting finding is that the difference between CG and AT Morse potentials remains of the order of 40 meV for all cations and all concentrations. This can be verified in Table A.8 in the Supporting Material which shows the numerical values of the Morse potentials depth D in terms of Mg^{2+} and K^+ concentrations. Compared to our previous results [162, 189, 211], the difference between CG and AT Morse potentials remains the same regardless of buffer type. This is now the combined result of 18 independent calculations comprising over 1100 melting temperature measurements from various sources [162, 189, 211]. In this sense this now firmly establishes a relation of 0.47 (35/75) between the internal AT and CG hydrogen bond strength for the mesoscopic model. Note that if one would naively just take the number of bonds this relation would be much larger, 0.67. However, DFT calculations, revised in Guerra et al. [212], place this relation between 0.48 and 0.525, which is much closer to our result.

The next questions we wish to discuss is: does it matter whether we represent the Morse potential as function of Na_{eq}^+ concentration or as function of ionic strength? Or, to put it differently, would either one provide a better or perhaps a more useful description?

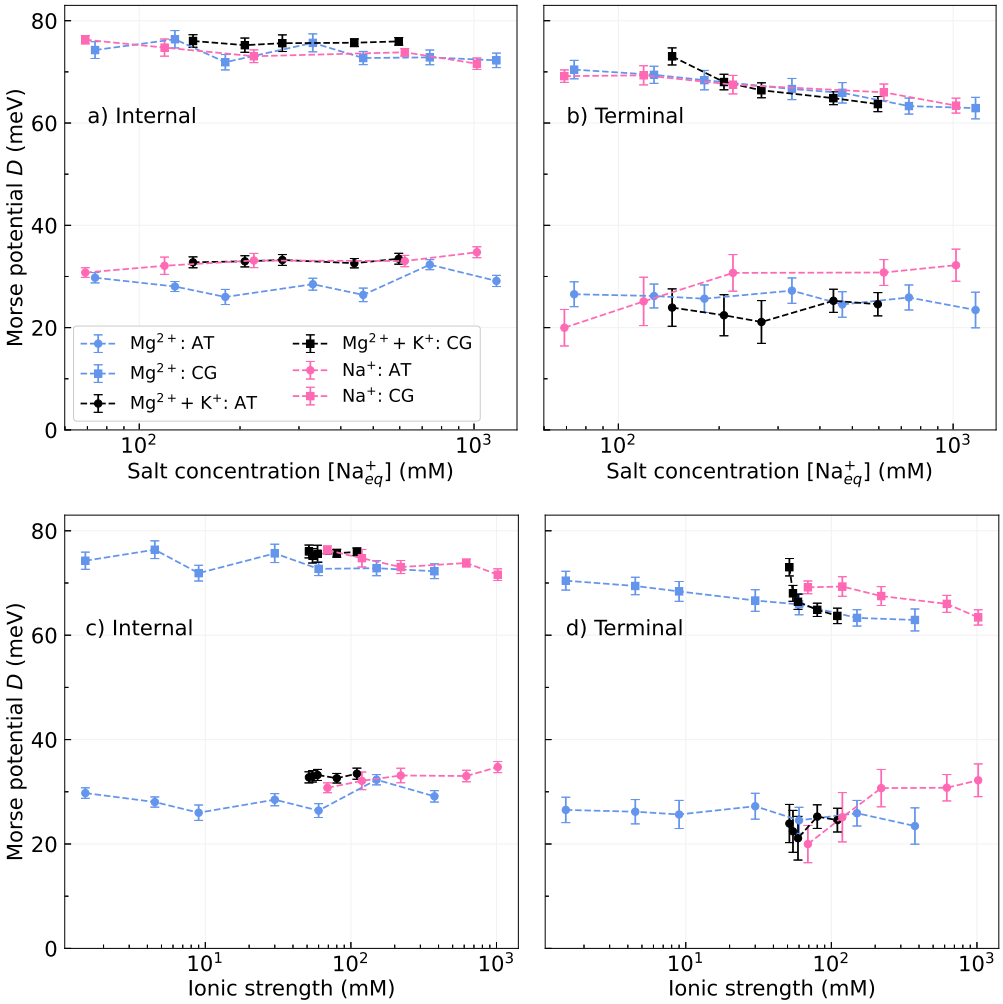

Figure 31 – Average Morse potentials D as function of the logarithm of equivalent sodium concentration $[Na^+_{eq}]$ and the ionic strength I, $\beta = 3.3\,M^{1/2}$, for (a,c) internal and (b,d) terminal of AT (circles) and CG (bullets) base pairs, in Mg^{2+} (blue) and $Mg^{2+}+K^+$ (black). Results for Na^+ (pink) from Ref. [189] are included for comparison. Error bars are displayed only when larger than symbol size.

Before we can proceed with this discussion we need to comment on the fact that Na^+_{eq} concentration, Eq (4.1), depends on an empirical β factor for which there is no clear consensus. In Fig. 31 we show the results using $\beta = 3.3\,M^{1/2}$. However, for different values of β we observed very little change, see Figs. 38 ($\beta = 3.79\,M^{1/2}$) and 39 ($\beta = 4\,M^{1/2}$). While the results for the various β factors a very similar, $\beta = 3.3\,M^{1/2}$ has a slightly better superposition for the various cations, especially for terminal base pairs at lower Na^+_{eq} concentration. Therefore, it would appear that the lower $\beta = 3.3\,M^{1/2}$ better fulfills the idea of equivalence. For the plots as function of ionic strength the representation changes substantially. The results for the Mg^{2+} are concentrated at the lower end of the ionic strength scale, while those for Na^+ at the higher part as shown in Figs. 31(c,d). Basically, the results run almost monotonically with ionic strength, with the results for Na^+ continuing trend of Mg^{2+}, and even the $Mg^{2+}+K^+$ results do overlap nicely. Considering

that ionic strength is a more fundamental and well established concept, it would be sensible to favor it in our analysis in comparison to the lesser known Na^+_{eq}, especially as it does not depend on adjustable parameters. Yet, as we will see next, the analysis in terms of ionic strength becomes even more interesting when applied to the stacking interaction.

The internal and terminal stacking parameters are shown in Figs. 32–35 as function of ionic strength, and in Figs. 40–43 as function of equivalent sodium concentration, and their numerical values are given in Table A.9 in the Appendix A. It is immediately evident that stacking parameters have a variety of different dependencies with ionic strength, which we can broadly classify into a few groups. In one group the stacking of the $Mg^{2+}+K^+$ buffer overlaps with the Mg^{2+} buffer, while the stacking for Na^+ is either much higher or smaller, these are shown in TApAT (Fig. 32a), ATpTA* (Fig. 32d), ATpAT* (Fig. 32f) and CG*pCG (Fig. 33f) where the presence of K^+ plays no role at all. In contrast, there is a single case where the stacking of the $Mg^{2+}+K^+$ overlaps with Na^+ and is detached from Mg^{2+}, shown in Fig. 32c. Another group has a dependency that resembles that of the Morse potentials, where the stacking all three buffer types appear to be a continuous monotonic dependence with ionic strengths, these are the cases in Figs. 33(a,c,e), Figs. 34(a,b), and Fig. 35d. Perhaps the most interesting case is where the stacking for $Mg^{2+}+K^+$ starts at Na^+ and gradually increases to Mg^{2+}, that is, where initially the monovalent cation K^+ predominates and gradually the divalent Mg^{2+} takes over. Clearly, cation valence plays an important role for stacking but so does the nearest-neighbor configuration. In several cases, we notice a trend of the $Mg^{2+}+K^+$ stacking parameters to start from the Na^+ parameters at low ionic strength and then tend towards the Mg^{2+} parameters at higher concentrations. For several nearest-neighbor configurations, as Mg^{2+} in the $Mg^{2+}+K^+$ buffer starts to increase we retrieve similar results to the Mg^{2+} buffer. This is very clearly the case for ATpAT, Fig. 32e, ATpGC, Fig. 34d and CGpAT Fig. 35a. For these three nearest-neighbors the cation valence is the property that defines the stacking, as evidenced by the mixed $Mg^{2+}+K^+$. We interpret the trend of $Mg^{2+}+K^+$ as competition between Mg^{2+} and K^+, with Mg^{2+} gradually displacing K^+. For some other nearest-neighbor configurations, the $Mg^{2+}+K^+$ mostly coincides with the Mg^{2+} buffer, see for instance Fig. 32a, which means that even for small concentrations, Mg^{2+} completely dominates the binding. There is only one case, ATpTA Fig. 35c, where the $Mg^{2+}+K^+$ stacking coincides with Na^+, however the stacking potentials are all very close to each other, therefore no clear conclusion can be drawn here. For CG-CG nearest-neighbors the $Mg^{2+}+K^+$, Mg^{2+} and Na^+ stacking potentials are in general very similar, with Na^+ appearing as a continuation of the Mg^{2+} stacking. In this case, the screening due to the cations seems to be very similar in all cases, and it is possible that the Mg^{2+} does not enter the major groove sufficiently enough to cause a major difference in the stacking. The stacking interaction parameter k shows little variation with Mg^{2+} concentration which is consistent with the findings from infrared spectroscopy by Serec et al. [201].

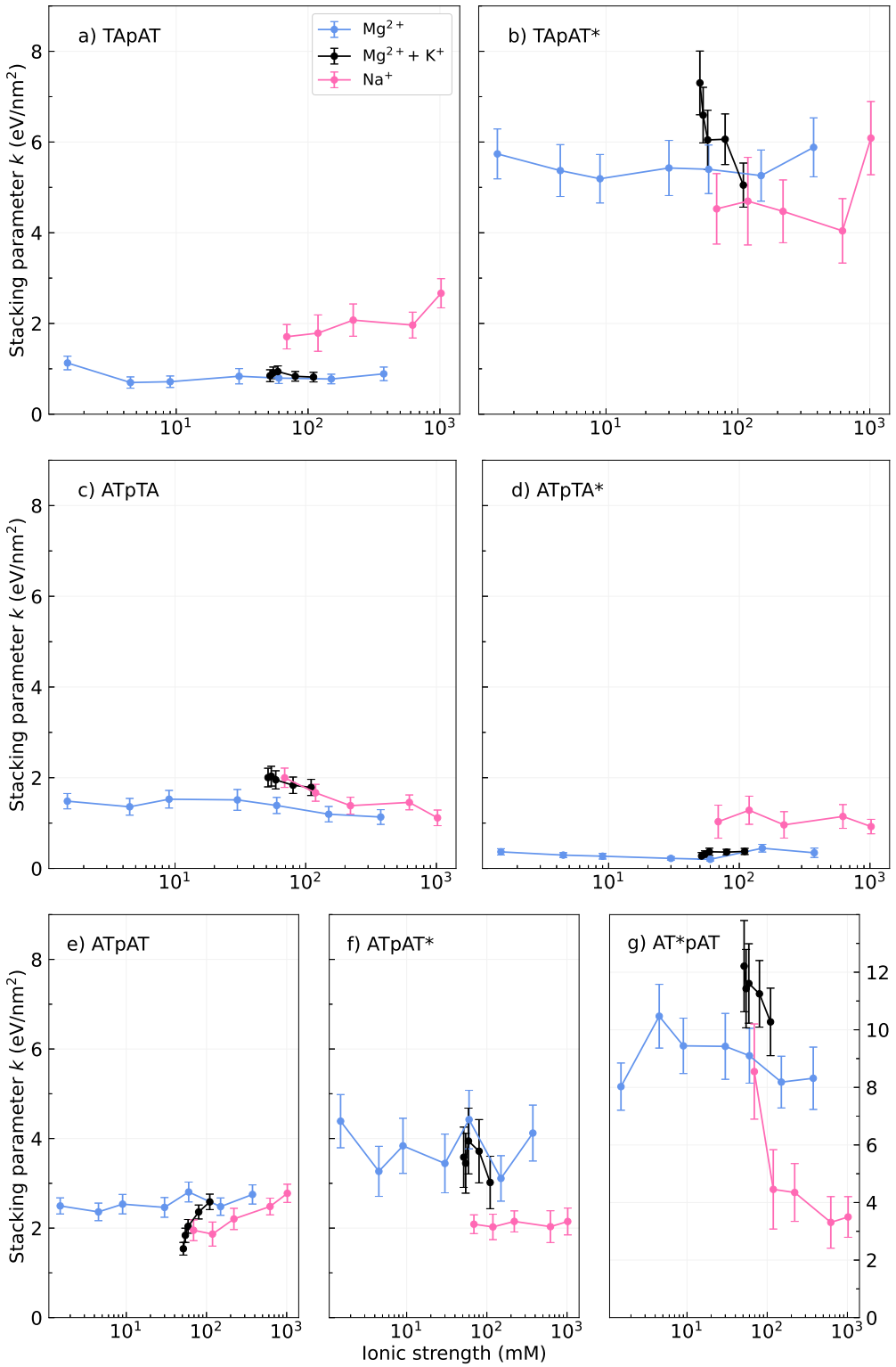

Figure 32 – Stacking parameters k as function of the logarithm of ionic strength I, for (a,b) TApAT, (c,d) ATpTA and (e,f,g) ATpAT nearest-neighbors at internal (panels a,c,e) and terminal (panels b,d,f,g) positions. Results for Na^+ (pink) from Ref. [189] are included for comparison.

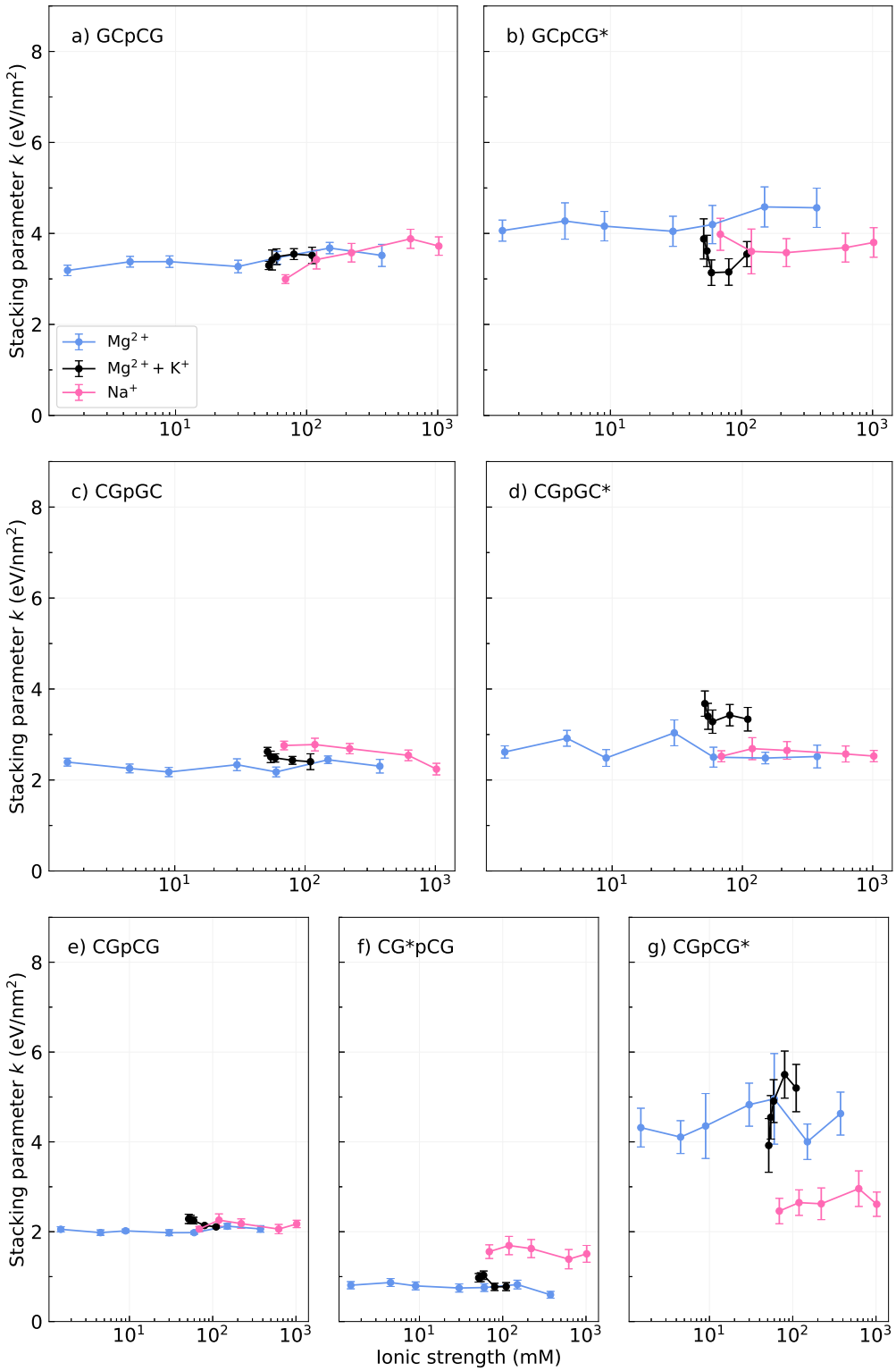

Figure 33 – Stacking parameters k as function of the logarithm of ionic strength I, for (a,b) GCpCG, (c,d) CGpGC and (e,f,g) CGpGC nearest-neighbors at internal (panels a,c,e) and terminal (panels b,d,f,g) positions. Results for Na^+ (pink) from Ref. [189] are included for comparison.

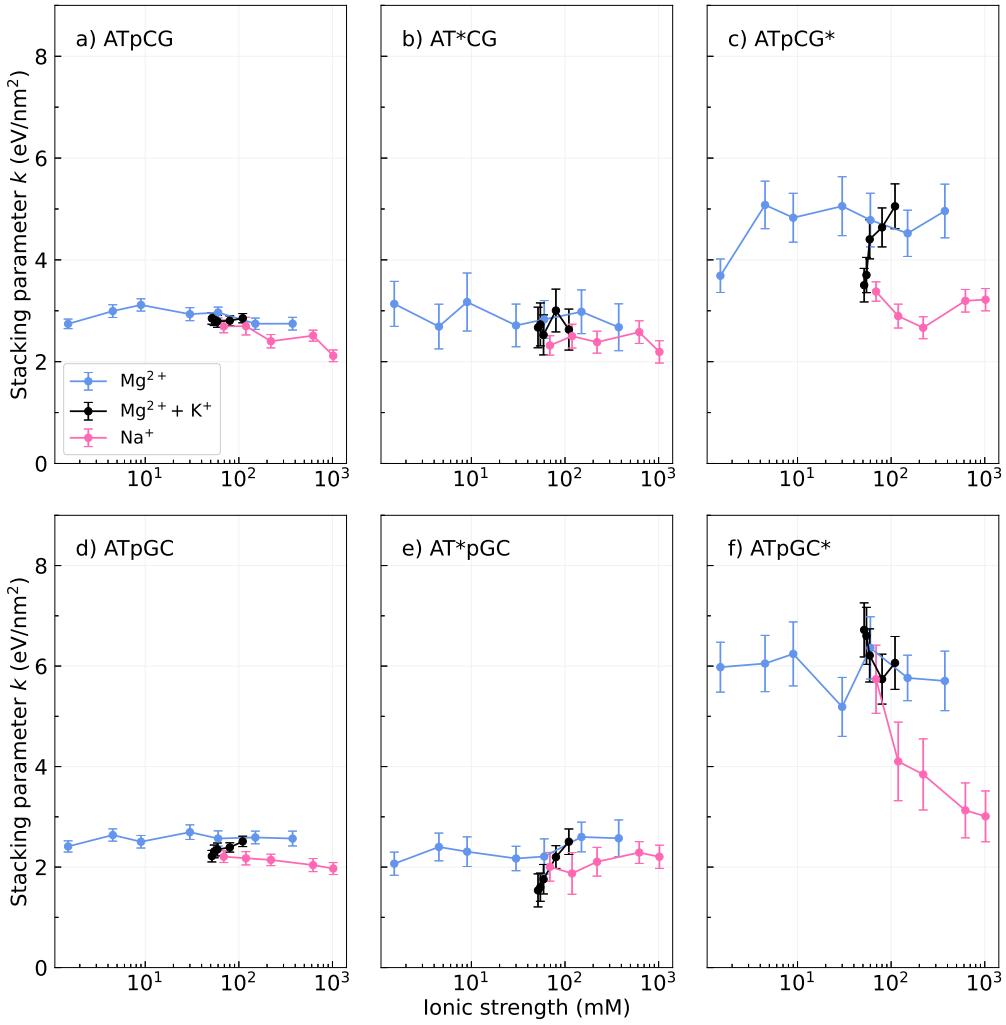

Figure 34 – Stacking parameters k as function of the logarithm of ionic strength I, for (a,b,c) ATpCG and (d,e,f) ATpGC nearest-neighbors at internal (panels a,d) and terminal (panels b,c,e,f) positions. Results for Na^+ (pink) from Ref. [189] are included for comparison.

Terminal stacking parameters differ considerably when compared to internal stacking, with no cases of simultaneous superposition of the three buffer types. However, in most cases the $Mg^{2+}+K^+$ parameters tends towards those of the Mg^{2+}, and in general stacking differences are much larger compared to the internal nearest-neighbors. Some configurations have very large stacking differences between the various buffers, for instance CGpCG* Fig. 33g and ATpGC* Fig. 34f. Every and Russu [202] has suggested that increased Mg^{2+} concentrations would lead to an increased opening of GCpCG pairs that should be reflected in smaller stacking parameters. However, from Fig. 33a we observe no decrease in stacking that would support this conclusion.

In other analysis, we have the internal and terminal stacking parameters described as function of equivalent sodium concentration $[Na^+_{eq}]$ in Figs. 40–43. The most Mg^{2+} and $Mg^{2+}+K^+$ curves of internal base pairs are close to Na^+. However, for terminal we have noticeable differences, but that do not make them very distant, this behavior may be due

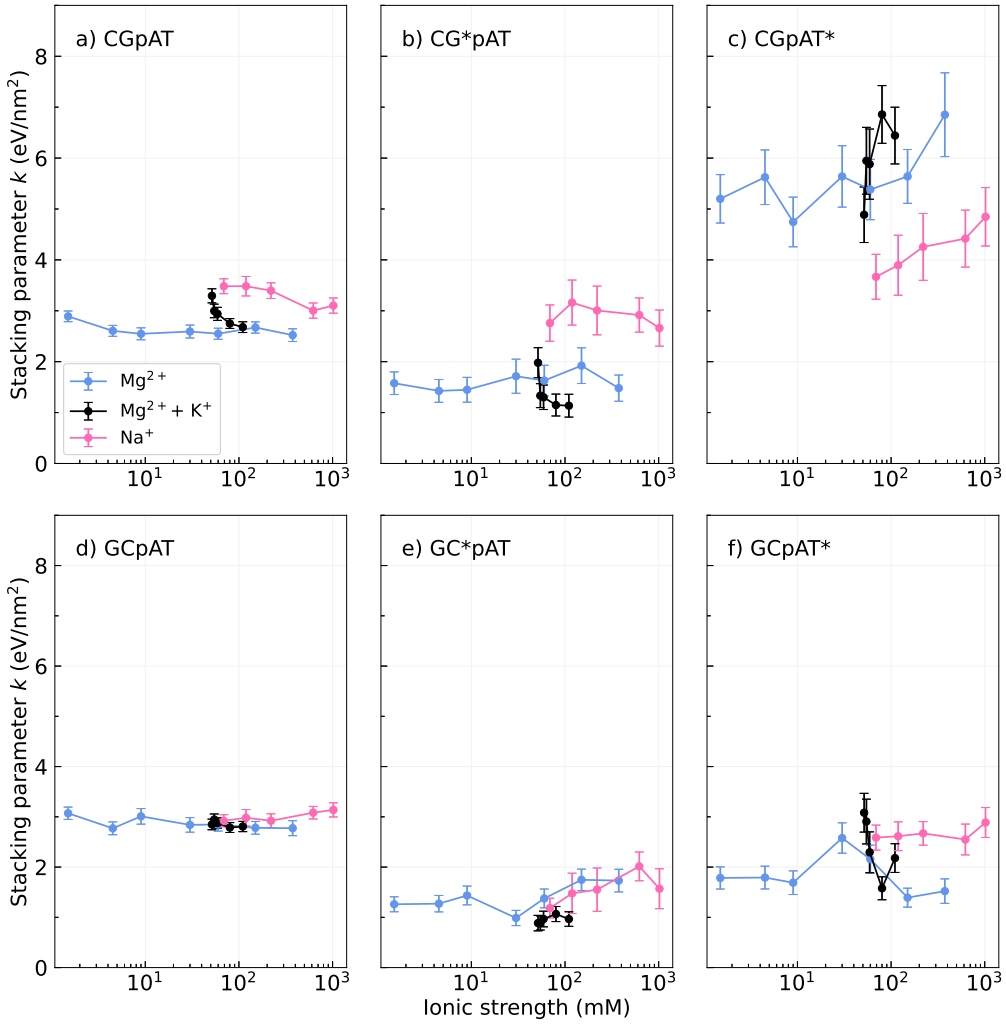

Figure 35 – Stacking parameters k as function of the logarithm of ionic strength I, for (a,b,c) CGpAT and (d,e,f) GCpAT nearest-neighbors at internal (panels a,d) and terminal (panels b,c,e,f) positions. Results for Na^+ (pink) from Ref. [189] are included for comparison.

to the different ways that the ions bind to the terminal base pairs. These results allow us to infer that $[Na^+_{eq}]$, calculated via Eq. (4.1), accurately describes buffers composed by mono- and divalent ions.

The interplay of the Morse and stacking potentials can be investigated by examining the average displacement profiles for some example sequences using the Eq. (2.29). So in Fig. 36 we show the average displacement profiles for two DNA duplexes (5′-3′), `d(CCAACGTTGG)` from Ref. [213] (above) and `d(CGCGAATTCGCG)` from Ref. [186] (below), respectively, for Na^+, Mg^{2+}, and $Mg^{2+}+K^+$ buffers. The smooth differences in Mg^{2+} and $Mg^{2+}+K^+$ buffers are due to the proximity in the values of Morse and stacking interaction potentials. In Na^+ buffer there is a similarity with the curves presented by $Mg^{2+}+K^+$ buffer, as shown in Figs. 36(b,c) and Figs. 36(e,f). This can be explained by the proximity of the values of the Morse and stacking interaction parameters for the neighboring base pairs combinations in each sequence, as well as the presence of K^+, a monovalent ion

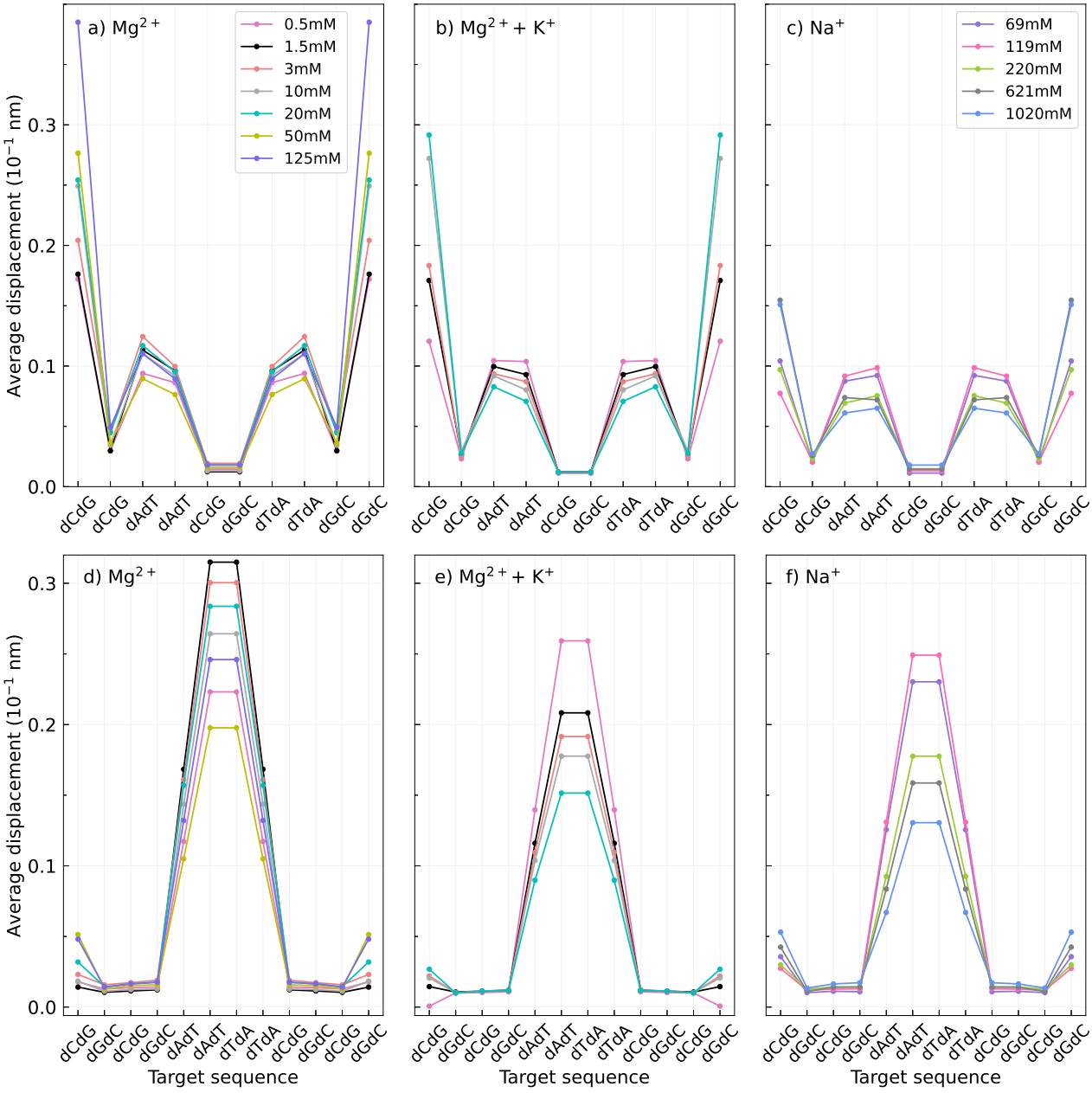

Figure 36 – Average displacement profiles calculated at 150 K for (a,d) Mg^{2+}, (b,e) $Mg^{2+}+K^{+}$, and (c,f) Na^{+} buffers, where (a,b,c) and (d,e,f) represent `d(CCAACGTTGG)` from Ref. [213] and `d(CGCGAATTCGCG)` from Ref. [186], respectively.

like Na^{+}. In addition, these profiles demonstrate the ionic potential of Mg^{2+} on DNA duplex stability, which even at low concentrations has a similar influence at high Na^{+} concentrations.

4.8 Conclusions

We performed a comparative study of DNA Morse potentials and stacking interactions in varying concentrations of Mg^{2+} and $Mg^{2+}+K^{+}$, including previous results in Na^{+} buffers, and considering different base pairs terminology. Based on measured melting

temperatures, our results confirm that the internal AT and CG Morse potentials are effectively constant for any cation valence and concentration. Therefore, the internal hydrogen bonds, which here are represented by Morse potentials, appear to be shielded from the ion interaction regardless of valence. We analyzed the Morse potential as function of the ionic strength and Na^+_{eq} concentration, and concluded that ionic strength provides an overall better description of the effects of cations on hydrogen bonding and stacking interactions, and has the potential to be extrapolated to other types of ions. Based on this observation we extended the ionic strength analysis also to the stacking interaction, where we obtained a distinct valence signature and signs of Mg^{2+} and K^+ ion competition in the mixed $Mg^{2+}+K^+$ buffer. We believe that, in view of our results, detailed temperature measurements could be used to evaluate the ionic charge distribution as well as to understand the ion competition in DNA. The strong cation dependency of terminal stacking and hydrogen bonding suggest that it could be used for the design of DNA liquid crystal structures where the end-to-end interaction plays an important role in nematic ordering [214]. This work has been accepted for publication in Biophysical Chemistry [215].

5 Conclusions

The PB model via thermodynamic equivalence is used to describe the intramolecular interactions of modified nucleic acids. In the project describes in Chapter 3, we investigate the thermal stability of DNA/TNA hybrids, where TNA is threose nucleic acid, a type of XNA. The sugar moiety in TNA contains an unnatural four-carbon threose sugar, composed by carbon atoms and a single oxygen atom, one atom shorter than DNA and RNA. TNA is resistant to biological degradation, has higher chemical stability, and therefore it is a promising candidate for diagnostic and therapeutic applications, such as biocompatible antisense oligonucleotides to suppress the gene expression in living environments [216]. The mesoscopic parameters from DNA/TNA were compared to DNA/RNA, and they confirm the expected asymmetry of the thermodynamic properties in regard to purine/pyrimidine content in DNA/TNA, similar to those of DNA/RNA hybrids. The Morse potentials for CG base pairs for DNA/TNA is very small when compared with DNA/RNA. The consequence of these small potentials is that for DNA/TNA hybrids the duplex appears to dissociate much for uniformly and largely independent of sequence composition. Furthermore, to stacking potential, unlike the DNA/RNA, which are quite small in some specific cases, we obtained no such weak potentials for DNA/TNA. This work was published in Chemical Physics Letters B.

In the second project, described in Chapter 4, we use the mesoscopic approach to study the effects of internal and terminal base pairs to divalent cations Mg^{2+}, as well as mixed mono- and divalent $Mg^{2+} + K^{+}$ buffers. These studies are compared to the previous studies on Na^{+} allow us to draw a comprehensive picture of the differences between the various types of buffers and how they affect the thermal stability of DNA. In addition to the Morse and stacking interaction potentials, we also use the ionic strength and equivalent sodium (Na^{+}_{eq}) concentration. This concentration is defined as the concentration of Na^{+} in a buffer that stabilizes duplexes to the same extent as the Mg^{2+} buffer. Thus, using multiple types of buffers allow us to understand how the ionic strength and sodium equivalent relate to the structural aspects of DNA, especially in terms of hydrogen bonding and base pair stacking, as well as how they affect the terminal base pairs. Our results confirm that the internal AT and CG Morse potentials are effectively constant for any cation valence and concentration. Therefore, the internal hydrogen bonds, which here are represented by Morse potentials, appear to be shielded from the ion interaction regardless of valence. We analyzed the Morse potential as function of the ionic strength and the equivalent Na^{+}_{eq} concentration, and concluded that ionic strength provides an overall better description and has the potential to be extrapolated to other types of ions. Based on this observation we extended the ionic strength analysis also to the stacking interaction, where we obtained

a distinct valence signature and signs of Mg^{2+} and K^+ ion competition in the mixed $Mg^{2+}+K^+$ buffer. These results allow experimental researchers to know which sequences have low or high thermal stability without the need to synthesize all combinations. This work has been accepted for publication in Biophysical Chemistry, and a preprint is shown in Appendix C.

5.1 Other projects and perspective

Another project is under way about the single internal mismatches in RNA/DNA duplexes, where mismatch is defined as non-Watson-Crick base pair, such as rAdA, rCdT, rUdC, and rGdG. The applications of mismatches include CRISPR-Cas probes, a forefront technology for genome editing used as a diagnostic tool [217, 218]. The experimental data set is composed by 108 sequences: 13 only with Watson-Crick (canonical) base pairs and 95 containing single internal mismatches. However, there are not enough sequences in the published literature that cover all neighboring base pairs combinations. For this reason we established a collaboration with Marco Buscaglia from University of Milan, Italy, our international collaboration, to perform additional measurements such as to fill in the missing combinations. They already have provided us with 15 new melting temperatures and the calculations are in progress. In preliminary results, we observed that the Morse potentials for dTrG and dUrG are the most stable mismatches base pairs, in contrast to dCrC and dArA. Furthermore, for stacking parameters, the most stable combinations are dArU-dTrU and dArU-dTrG, while dTrA-dCrC and dTrA-dArA are unstable. While dArU-dTrG is the greatest value, dTrG-dArU is 6 times smaller, proving that the order of the base pairs influence the stacking parameter. All results highlight the influence of neighboring base pairs on the mismatch stability.

Appendix

APPENDIX A – DNA in solutions containing Mg^{2+}: Tables and graphs

Table A.1 – Shown are the measured melting temperatures (T_i) and the predicted melting temperatures (T_i') in °C, for the 10 new unpublished DNA sequences (5′ to 3′) used in this work in solutions containing magnesium ions.

	0.5 mM		1.5 mM		3 mM		10 mM		20 mM		50 mM		125 mM	
Sequence	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'
ACGTCATGACG	42.5	41.6	45.7	45.0	48.0	46.8	49.0	48.8	49.2	48.9	50.5	49.4	50.4	49.6
AGAGAGAGAGA	37.2	36.3	41.2	40.6	42.5	42.2	46.0	45.4	45.0	45.3	46.4	46.1	47.0	46.5
AGTCTATCTCC	37.7	37.3	41.0	41.0	44.0	43.1	46.0	45.4	46.0	45.9	47.0	46.1	47.7	46.9
CAGATCAGACT	35.0	34.7	39.0	38.6	42.0	40.5	44.0	43.1	44.0	43.3	46.0	44.7	45.0	44.7
CTACGTTACAT	33.8	33.3	38.0	37.2	40.0	38.9	42.0	41.1	43.0	41.9	44.0	42.8	44.0	43.1
CTATGGATCCC	42.3	42.0	45.5	45.0	48.3	47.0	50.0	48.6	50.8	49.5	50.0	49.3	50.0	50.1
GTACATTCACA	35.0	34.0	39.0	38.2	40.0	39.9	44.0	42.6	43.0	42.9	44.0	43.8	45.0	44.7
TCGACTCATCA	40.5	39.1	43.0	42.9	44.8	44.2	49.3	47.5	48.0	47.4	49.0	47.8	51.0	48.6
TGGACTTGACC	44.0	43.5	48.6	47.1	50.3	48.9	51.0	50.9	53.0	51.7	52.0	51.5	53.2	52.7
TTGATGCAACC	41.2	40.4	43.5	43.3	45.0	45.5	47.0	47.4	49.0	48.6	49.0	48.5	50.6	49.2

Table A.2 – Shown are the measured melting temperatures from Ref. [206] (T_i) and the predicted melting temperatures (T_i') in °C, for the 92 DNA sequences (5′ to 3′) used in this work in solutions containing magnesium ions.

	0.5 mM		1.5 mM		3 mM		10 mM		20 mM		50 mM		125 mM	
Sequence	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'
ATCAATCATA	21.9	23.6	25.6	27.0	27.4	29.0	29.9	32.6	30.9	32.3	33.0	34.5	33.1	34.5
TTGTAGTCAT	27.1	27.2	30.9	31.0	32.4	32.8	34.9	36.1	36.0	36.2	37.2	37.6	37.5	38.0
GAAATGAAAG	24.6	25.5	27.4	28.8	30.6	31.6	31.4	33.3	33.9	34.8	34.6	36.3	35.4	36.6
CCAACTTCTT	30.1	30.0	35.2	34.8	36.5	36.4	39.4	39.1	39.8	39.4	40.5	40.6	40.2	40.4
ATCGTCTGGA	35.1	35.6	39.0	39.7	40.5	41.0	43.9	44.4	43.4	44.2	43.8	44.8	44.4	45.1
AGCGTAAGTC	29.6	34.0	36.1	39.0	37.5	40.6	39.4	42.8	40.9	43.5	42.1	44.7	42.5	45.2
CGATCTGCGA	38.4	40.1	43.2	44.2	44.0	45.3	47.6	48.5	47.6	48.3	47.2	48.1	47.5	48.6
TGGCGAGCAC	45.2	46.5	49.6	51.1	50.5	52.1	52.8	54.1	53.2	54.9	54.1	55.1	54.5	54.7
GATGCGCTCG	44.0	42.9	47.8	47.0	48.9	48.3	50.6	50.0	51.2	50.8	51.5	50.1	51.9	51.7
GGGACCGCCT	48.4	48.3	52.0	52.5	53.8	54.1	55.5	55.4	55.9	56.4	55.5	56.0	56.0	56.9
CGTACACATGC	40.1	40.7	44.2	44.4	45.6	45.9	48.0	47.7	48.0	48.2	48.1	48.6	48.6	49.0
CCATTGCTACC	38.9	39.5	42.6	43.1	44.1	44.8	46.4	46.8	47.0	47.7	47.2	47.6	47.5	48.0
TACTAACATTAACTA	39.2	38.9	42.9	42.5	44.4	44.1	47.1	46.6	47.5	47.2	48.4	48.3	49.0	48.9
ATACTTACTGATTAG	39.9	39.4	43.5	42.6	45.0	44.3	46.8	46.3	47.9	47.2	48.6	48.2	49.2	48.7
GTACACTGTCTTATA	43.4	43.2	47.1	46.8	48.6	48.4	50.5	50.5	51.2	51.0	52.1	51.9	52.5	52.4
GTATGAGAGACTTTA	43.5	43.7	47.0	47.1	48.5	48.8	51.0	50.8	51.5	51.4	52.2	52.4	52.9	52.9
TTCTACCTATGTGAT	43.2	43.6	46.0	46.7	47.9	48.2	50.2	50.3	50.4	50.9	51.0	51.7	51.5	52.2
AGTAGTAATCACACC	46.5	46.4	49.5	49.8	51.0	51.5	53.0	53.2	53.5	54.1	54.0	54.1	54.5	54.9
ATCGTCTCGGTATAA	48.0	47.7	51.0	50.7	52.6	52.1	54.8	54.1	55.1	54.6	55.5	55.1	56.4	55.7
ACGACAGGTTTACCA	51.5	51.5	54.5	54.8	56.0	56.2	58.1	57.8	58.5	58.5	59.0	58.9	59.5	59.6
CTTTCATGTCCGCAT	52.8	51.3	55.1	54.1	56.5	55.6	58.4	56.7	59.0	57.9	59.5	58.5	59.6	58.7
TGGATGTGTGAACAC	49.9	50.3	53.5	53.5	54.8	55.0	56.6	56.5	57.2	57.2	57.9	57.8	58.2	58.4
ACCCCGCAATACATG	52.6	52.4	55.4	55.4	56.6	56.8	58.5	58.3	59.0	59.1	58.9	59.5	59.2	59.6
GCAGTGGATGTGAGA	52.4	52.3	55.1	55.2	56.5	56.6	58.0	58.2	58.6	58.9	59.2	59.1	59.0	59.4
GGTCCTTACTTGGTG	50.4	51.4	53.4	54.5	54.8	56.2	56.5	57.6	57.1	58.6	57.5	58.5	57.5	59.1
CGCCTCATGCTCATC	54.4	54.5	57.2	57.6	58.6	58.8	60.1	59.9	60.6	60.9	61.2	61.2	61.0	61.5
AAATAGCCGGGCCGC	60.5	59.7	63.5	63.1	64.5	64.2	66.2	65.1	66.3	66.2	67.0	66.3	67.2	66.7
CCAGCCAGTCTCTCC	56.5	56.6	59.5	59.7	60.5	61.0	62.0	62.0	62.5	63.1	63.0	62.9	63.0	63.2
GACGACAAGACCGCG	59.4	57.4	61.8	60.6	62.9	61.9	64.2	62.8	64.5	63.7	65.2	63.8	65.0	64.2
CAGCCTCGTCGCAGC	61.9	61.0	64.2	63.9	65.3	65.1	66.8	66.0	67.2	67.1	67.5	66.6	67.3	67.4
CTCGCGGTCGAAGCG	61.4	62.1	64.0	64.7	65.2	66.1	66.2	66.5	66.8	67.8	67.2	67.6	66.8	67.9
GCGTCGGTCCGGGCT	65.2	64.1	67.5	67.2	68.5	68.4	69.5	68.8	70.3	69.9	70.5	69.8	70.2	70.2
TATGTATATTTTGTAATCAG	48.5	47.5	51.4	49.9	52.5	51.6	54.4	53.1	55.0	54.1	55.8	54.9	56.2	55.4
TTCAAGTTAAACATTCTATC	50.0	48.4	52.5	51.0	54.0	52.7	55.6	54.2	56.5	55.3	57.2	56.2	57.9	56.8
TGATTCTACCTATGTGATTT	51.8	53.2	54.5	56.1	55.8	57.4	57.6	59.1	58.2	59.7	59.2	60.3	59.5	60.8
GAGATTGTTTCCCTTTCAAA	53.2	53.9	56.1	56.4	57.5	58.0	59.2	59.2	60.1	60.4	60.9	61.0	61.2	61.6
ATGCAATGCTACATATTCGC	57.4	57.1	60.0	60.0	61.0	60.5	62.5	62.4	63.1	62.7	63.9	63.6	63.8	63.7
CCACTATACCATCTATGTAC	53.0	52.8	55.5	55.7	56.6	56.9	58.0	58.4	58.8	59.0	59.1	59.4	59.5	59.7
CCATCATTGTGTCTACCTCA	58.0	57.0	60.5	59.6	61.5	60.7	62.9	62.1	63.5	62.7	64.0	63.3	64.0	63.5
CGGGACCAACTAAAGGAAAT	57.5	58.2	60.1	61.0	61.5	62.1	62.9	63.4	63.5	64.2	64.0	64.4	64.5	65.0
TAGTGGCGATTAGATTCTGC	60.0	59.9	62.5	62.4	63.5	63.6	64.8	64.8	65.7	65.5	66.2	66.0	66.5	66.3
AGCTGCAGTGGATGTGAGAA	61.0	61.8	63.6	64.3	64.8	65.5	66.3	66.5	66.8	67.4	67.7	67.9	67.2	68.0
TACTTCCAGTGCTCAGCGTA	62.1	62.9	64.7	65.6	65.8	66.6	67.2	67.7	67.7	68.6	68.2	69.0	68.3	69.2
CAGTGAGACAGCAATGGTCG	61.4	61.4	63.9	63.9	65.0	65.0	66.2	66.1	66.8	66.8	67.2	67.2	67.0	67.2
CGAGCTTATCCCTATCCCTC	60.0	61.0	62.5	63.7	63.5	64.8	65.0	65.9	65.5	66.6	66.0	66.8	65.8	67.2

(continued on next page)

	0.5 mM		1.5 mM		3 mM		10 mM		20 mM		50 mM		125 mM	
Sequence	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'
CGTACTAGCGTTGGTCATGG	61.1	60.8	63.4	63.6	64.3	64.5	65.8	65.6	66.2	66.3	66.5	66.4	66.5	66.5
AAGGCGAGTCAGGCTCAGTG	66.2	65.1	68.5	67.8	69.8	68.8	70.7	69.8	71.2	70.6	71.5	71.0	71.5	70.9
ACCGACGACGCTGATCCGAT	66.2	66.0	68.7	68.1	69.5	69.3	71.0	70.2	71.0	70.8	71.7	71.1	71.2	71.1
AGCAGTCCGCCACACCCTGA	67.7	67.2	70.0	69.8	70.7	70.7	72.2	71.7	72.5	72.5	72.8	72.5	72.3	72.6
CAGCCTCGTTCGCACAGCCC	69.0	68.7	71.3	70.8	72.2	72.1	73.5	72.7	74.0	73.8	74.2	73.8	73.8	73.8
GTGGTGGGCCGTGCGCTCTG	69.8	70.3	72.0	72.8	73.2	73.8	74.3	74.3	74.7	75.3	75.0	75.0	74.5	75.3
GTCCACGCCCGGTGCGACGG	70.3	71.6	72.5	73.9	73.8	74.9	74.0	75.2	74.5	76.1	74.7	76.0	74.2	75.8
GATATAGCAAAATTCTAAGTTAATA	54.2	53.0	56.9	55.7	58.0	56.8	59.8	58.6	60.9	59.4	61.2	60.3	61.9	60.8
ATAACTTTACGTGTGTGACCTATTA	59.9	58.9	62.4	61.4	63.5	62.4	65.0	64.0	65.5	64.5	66.2	65.2	66.3	65.4
GTTCTATACTCTTGAAGTTGATTAC	57.2	57.4	59.6	60.0	60.9	61.3	62.5	62.7	63.0	63.3	63.8	64.0	64.0	64.4
CCCTGCACTTTAACTGAATTGTTTA	60.8	60.1	63.1	62.6	64.2	63.6	65.7	65.0	66.0	65.7	66.8	66.4	67.0	66.4
TAACCATACTGAATACCTTTTGACG	59.8	60.4	62.0	62.8	63.0	63.8	64.5	65.3	65.0	65.7	65.7	66.2	65.8	66.5
TCCACACGGTAGTAAAATTAGGCTT	62.5	62.8	64.8	65.5	65.7	66.3	67.0	67.9	68.0	68.3	68.2	68.6	68.2	68.8
TTCCAAAAGGAGTTATGAGTTGCGA	62.1	63.0	64.3	65.1	65.5	66.2	66.8	67.7	67.3	68.2	68.0	68.6	68.0	68.8
AATATCTCTCATGCGCCAAGCTACA	64.5	65.8	67.0	68.2	67.8	68.8	69.2	70.2	69.5	70.6	70.2	71.3	70.0	71.2
TAGTATATCGCAGCATCATACAGGC	63.0	64.5	65.0	66.6	66.2	67.5	68.0	68.8	67.8	69.2	68.3	69.8	68.3	69.8
TGGATTCTACTCAACCTTAGTCTGG	62.2	62.2	64.5	64.6	65.5	65.6	66.8	67.0	67.3	67.4	67.8	68.0	67.8	68.2
CGGAATCCATGTTACTTCGGCTATC	64.0	64.3	66.2	66.5	67.2	67.5	68.5	68.7	69.0	69.1	69.5	69.6	69.5	69.6
CTGGTCTGGATCTGAGAACTTCAGG	65.5	65.1	67.7	66.9	68.7	68.1	69.8	68.9	70.3	69.6	70.7	70.2	70.7	70.0
ACAGCGAATGGACCTACGTGGCCTT	70.0	69.7	72.0	71.8	73.0	72.7	74.0	73.7	74.2	74.1	74.3	74.6	74.2	74.3
AGCAAGTCGAGCAGGGCCTACGTTT	70.2	69.9	72.2	72.3	73.3	73.1	74.2	74.1	74.8	74.6	75.0	75.0	75.0	74.8
GCGAGCGACAGGTTACTTGGCTGAT	69.5	68.9	71.5	71.0	72.5	71.7	73.3	72.8	74.0	73.2	74.2	73.9	74.2	73.6
AAAGGTGTCGCGGAGAGTCGTGCTG	70.8	70.3	73.2	72.4	74.0	73.3	75.2	74.4	75.7	74.7	75.8	75.3	75.5	74.9
ATGGGTGGGAGCCTCGGTAGCAGCC	72.2	73.1	74.5	74.9	75.2	75.8	76.5	76.7	76.7	77.2	76.8	77.4	76.8	77.2
CAGTGGGCTCCTGGGCGTGCTGGTC	73.5	73.0	75.3	75.0	76.2	75.8	77.2	76.6	77.5	77.1	78.2	77.6	78.5	77.2
GCCAACTCCGTCGCCGTTCGTGCGC	73.8	74.4	75.7	76.7	76.5	77.7	77.2	78.2	77.5	78.7	77.7	78.6	77.5	78.7
ACGGGTCCCCGCACCGCACCGCCAG	77.2	77.8	79.0	79.2	79.8	80.2	80.2	80.5	80.5	81.1	80.8	80.8	80.3	80.8
TTATGTATTAAGTTATATAGTAGTAGTAGT	55.0	56.3	57.5	58.9	58.5	59.7	60.0	61.7	60.5	62.0	61.4	62.6	61.8	63.0
ATTGATATCCTTTTCTATTCATCTTTCATT	58.9	59.2	61.2	61.3	62.4	62.3	63.8	64.0	64.3	64.3	65.0	65.2	65.5	65.3
AAAGTACATCAACATAGAGAATTGCATTTC	61.5	61.2	63.8	63.4	64.8	64.4	66.0	65.9	66.5	66.2	67.2	67.2	67.5	67.1
CTTAAGATATGAGAACTTCAACTAATGTGT	61.0	61.0	63.1	62.9	64.0	64.0	65.5	65.5	66.0	65.8	67.2	66.7	67.2	66.6
CTCAACTTGCGGTAAATAAATCGCTTAATC	64.0	64.1	66.2	66.0	67.2	67.0	68.3	68.3	68.8	68.8	69.5	69.6	69.8	69.4
TATTGAGAACAAGTGTCCGATTAGCAGAAA	65.2	64.8	67.5	66.7	68.3	67.6	69.7	69.2	70.0	69.3	70.7	70.2	70.8	70.0
GTCATACGACTGAGTGCAACATTGTTCAAA	65.5	65.2	67.7	67.0	68.5	68.0	69.8	69.5	70.0	69.6	70.7	70.5	70.8	70.2
AACCTGCAACATGGAGTTTTTGTCTCATGC	67.8	67.0	70.0	68.8	70.8	69.8	71.7	71.1	72.3	71.2	72.8	72.1	72.8	71.6
CCGTGCGGTGTGTACGTTTTATTCATCATA	66.7	66.2	68.8	68.0	69.5	68.8	70.8	70.3	71.2	70.3	71.7	71.1	71.5	70.7
GTTCACGTCCGAAAGCTCGAAAAAGGATAC	68.2	67.6	70.2	69.5	71.2	70.5	72.2	71.9	72.8	71.9	73.2	72.7	73.5	72.4
AGTCTGGTCTGGATCTGAGAACTTCAGGCT	69.8	68.2	71.7	70.0	72.5	71.0	73.5	72.4	74.0	72.3	74.2	73.4	74.5	72.8
TCGGAGAAATCACTGAGCTGCCTGAGAAGA	69.0	68.6	71.0	70.4	72.0	71.2	72.8	73.0	73.3	72.8	73.8	73.5	73.7	73.0
CTTCAACGGATCAGGTAGGACTGTGGTGGG	68.2	69.5	70.0	71.0	71.0	72.0	72.0	73.2	72.5	73.1	73.0	74.0	73.0	73.3
ACGCCCACAGGATTAGGCTGGCCCACATTG	73.2	71.9	75.0	73.5	75.8	74.3	76.7	75.7	77.0	75.5	77.3	76.4	77.2	75.6
GTTATTCCGCAGTCCGATGGCAGCAGGCTC	72.8	71.9	74.8	73.6	75.5	74.2	76.7	75.7	77.0	75.5	77.3	76.5	77.0	75.7
TCAGTAGGCGTGACGCAGAGCTGGCGATGG	73.7	72.7	75.8	74.2	76.0	74.9	77.0	76.5	77.8	76.2	78.0	77.0	77.8	76.1
CGCGCCACGTGTGATCTACAGCCGTTCGGC	73.8	74.5	75.2	75.7	76.0	76.6	77.0	77.8	77.0	77.5	77.5	78.4	77.0	77.5
GACCTGACGTGGACCGCTCCTGGGCGTGGT	76.2	74.8	77.8	76.1	78.5	77.0	79.5	78.2	79.7	77.9	80.0	78.8	79.8	77.9
GCCCCTCCACTGGCCGACGGCAGCAGGCTC	77.5	77.0	79.0	78.3	79.8	78.8	80.5	80.2	80.8	79.7	81.0	80.8	80.7	79.7
CGCCGCTGCCGACTGGAGGAGCGCGGGACG	78.7	78.4	80.0	79.5	80.8	80.2	81.0	81.6	81.7	81.0	81.7	81.9	81.2	80.7

Table A.3 – Shown are the measured melting temperatures from Ref. [206] (T_i) and the predicted melting temperatures (T_i') in °C, for the 80 DNA sequences (5′ to 3′) used in this work in solutions containing magnesium ions (shown in the table header) and 50 mM KCl.

	0.5 mM		1.5 mM		3 mM		10 mM		20 mM	
Sequence	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'
ATCAATCATA	22.4	23.1	24.3	25.4	25.9	27.2	29.5	30.8	30.4	31.9
TTGTAGTCAT	25.6	26.0	28.1	28.3	30.2	30.4	33.3	33.8	34.6	35.0
GAAATGAAAG	23.6	23.3	25.7	25.7	27.4	27.7	30.8	31.3	32.3	33.0
CCAACTTCTT	30.0	29.5	32.2	31.9	34.5	33.9	37.4	37.0	38.9	38.4
ATCGTCTGGA	35.5	36.1	37.8	38.2	39.2	39.4	41.8	42.0	43.2	43.5
AGCGTAAGTC	28.3	31.9	30.9	34.4	33.2	36.6	38.0	40.7	39.3	42.1
CGATCTGCGA	40.4	41.8	41.9	43.4	42.8	44.4	45.4	46.7	46.9	47.9
TGGCGAGCAC	45.6	46.5	47.8	49.1	49.2	50.8	52.4	53.8	52.8	54.5
GATGCGCTCG	44.8	44.0	46.7	45.4	47.9	46.8	50.4	49.4	50.9	50.1
GGGACCGCCT	49.1	48.2	50.9	50.4	52.6	52.0	55.1	55.0	56.0	55.7
CGTACACATGC	43.0	42.4	43.4	43.3	44.2	44.1	46.6	46.4	47.6	47.4
CCATTGCTACC	38.9	39.5	41.2	41.6	43.0	43.5	45.6	45.9	46.3	46.7
TACTAACATTAACTA	38.1	37.5	40.8	40.3	42.6	42.1	46.3	45.7	47.4	46.8
ATACTTACTGATTAG	38.7	38.0	41.3	40.7	43.2	42.5	45.9	45.4	47.5	46.9
GTACACTGTCTTATA	42.8	42.6	45.3	45.0	47.1	46.9	49.9	49.7	51.0	50.9
GTATGAGAGACTTTA	42.1	42.3	45.1	45.1	47.1	47.0	50.2	50.0	51.4	51.3
AGTAGTAATCACACC	45.6	45.4	48.1	48.0	49.7	49.8	52.6	52.8	53.6	53.8
ATCGTCTCGGTATAA	47.3	47.1	49.8	49.6	51.4	51.3	54.2	53.7	55.1	54.5
ACGACAGGTTTACCA	50.3	50.6	53.1	53.3	54.6	54.7	57.8	57.8	58.5	58.5
CTTTCATGTCCGCAT	52.4	51.7	54.6	53.8	56.3	55.2	58.5	57.2	59.5	58.1
TGGATGTGTGAACAC	49.2	50.4	51.9	52.8	53.7	54.3	56.5	56.8	57.6	57.6
ACCCCGCAATACATG	52.8	53.2	55.0	54.9	56.2	56.2	58.5	58.4	59.2	58.9
GGTCCTTACTTGGTG	50.1	51.1	52.2	53.2	53.9	54.8	56.3	57.2	56.9	58.0
CGCCTCATGCTCATC	54.8	55.6	57.0	57.6	58.4	58.9	60.5	60.6	61.2	61.1
AAATAGCCGGGCCGC	61.0	60.6	63.5	62.6	64.5	63.7	66.3	65.7	67.0	66.3
CCAGCCAGTCTCTCC	56.5	56.8	58.9	59.2	60.3	60.6	62.3	62.2	63.1	62.9
GACGACAAGACCGCG	59.6	58.2	61.4	60.0	62.7	61.2	64.4	63.2	65.1	63.8
CTCGCGGTCGAAGCG	62.0	63.0	64.3	64.8	65.2	65.7	66.8	67.3	67.4	68.0
GCGTCGGTCCGGGCT	66.0	64.5	68.2	67.3	68.9	68.1	70.3	69.7	70.5	69.8
TATGTATATTTTGTAATCAG	46.7	46.7	49.5	49.3	51.4	50.9	54.0	53.5	55.2	54.5
TTCAAGTTAAACATTCTATC	48.1	46.1	50.8	49.1	52.8	51.2	55.4	54.1	56.5	55.3
GAGATTGTTTCCCTTTCAAA	51.3	52.0	54.2	54.9	56.3	56.8	59.3	59.4	60.3	60.4
ATGCAATGCTACATATTCGC	56.9	57.2	59.3	59.3	60.7	60.6	62.8	62.7	63.6	63.6
CCACTATACCATCTATGTAC	52.9	52.7	55.0	54.9	56.4	56.6	58.3	58.4	59.1	59.3
CCATCATTGTGTCTACCTCA	57.5	56.7	60.1	59.2	61.2	60.6	63.0	62.6	64.0	63.3
CGGGACCAACTAAAGGAAAT	56.3	57.6	59.5	60.1	60.8	61.6	63.2	63.8	64.1	64.7
TAGTGGCGATTAGATTCTGC	59.4	59.5	61.8	61.7	63.1	62.9	65.1	64.9	66.1	65.7
TACTTCCAGTGCTCAGCGTA	62.2	63.0	64.6	65.3	65.6	66.5	67.4	68.4	68.2	69.0

(continued on next page)

	0.5 mM		1.5 mM		3 mM		10 mM		20 mM	
Sequence	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'	T_i	T_i'
CAGTGAGACAGCAATGGTCG	63.0	62.1	63.9	63.8	64.9	65.0	66.8	66.6	67.4	67.3
CGAGCTTATCCCTATCCCTC	58.9	60.9	61.9	63.3	63.4	64.6	65.2	66.4	65.9	67.0
CGTACTAGCGTTGGTCATGG	61.2	61.5	63.4	63.4	64.6	64.6	66.1	66.2	66.9	66.8
AAGGCGAGTCAGGCTCAGTG	66.1	65.5	68.3	67.6	69.5	68.9	71.0	70.4	71.5	71.0
ACCGACGACGCTGATCCGAT	67.3	66.1	69.3	68.3	70.1	69.4	71.9	71.0	72.3	71.5
AGCAGTCCGCCACACCCTGA	68.3	68.1	70.3	70.1	71.2	71.1	72.7	72.4	73.1	73.1
GTGGTGGGCCGTGCGCTCTG	70.9	71.7	72.6	73.3	73.8	74.3	75.2	75.3	75.7	75.8
GTCCACGCCCGGTGCGACGG	71.7	73.2	73.3	74.9	74.5	76.0	75.1	76.4	75.3	76.7
GATATAGCAAAATTCTAAGTTAATA	51.9	50.5	54.7	53.7	56.8	55.7	59.5	58.7	60.6	59.7
ATAACTTTACGTGTGTGACCTATTA	58.9	57.7	61.4	60.3	62.8	61.9	65.1	64.2	66.1	65.0
CCCTGCACTTTAACTGAATTGTTTA	59.4	58.8	61.9	61.4	63.5	63.1	65.7	65.3	66.6	66.2
TAACCATACTGAATACCTTTTGACG	58.6	59.7	61.0	62.1	62.7	63.5	65.3	65.7	65.6	66.4
TCCACACGGTAGTAAAATTAGGCTT	61.3	62.3	64.0	64.8	65.4	66.2	67.3	68.1	68.2	68.9
TTCCAAAAGGAGTTATGAGTTGCGA	61.2	61.9	63.7	64.5	65.2	65.9	67.2	67.8	68.0	68.7
AATATCTCTCATGCGCCAAGCTACA	64.0	65.3	66.4	67.8	67.7	69.2	70.2	71.0	70.2	71.5
TAGTATATCGCAGCATCATACAGGC	62.7	64.3	64.9	66.4	66.4	67.5	67.7	69.2	68.4	69.7
TGGATTCTACTCAACCTTAGTCTGG	61.2	60.9	63.8	63.7	65.3	65.1	67.2	67.2	68.0	68.0
CGGAATCCATGTTACTTCGGCTATC	63.2	63.9	65.5	66.2	67.1	67.5	69.0	69.2	69.6	69.8
ACAGCGAATGGACCTACGTGGCCTT	70.2	69.7	72.0	71.9	73.2	73.0	75.3	74.5	75.2	74.8
AGCAAGTCGAGCAGGGCCTACGTTT	70.5	69.5	72.5	71.9	73.4	73.1	74.6	74.8	75.3	75.3
GCGAGCGACAGGTTACTTGGCTGAT	69.2	68.3	71.2	70.5	72.5	71.6	74.0	73.1	74.6	73.8
AAAGGTGTCGCGGAGAGTCGTGCTG	71.1	70.4	73.2	72.4	74.3	73.5	75.6	74.9	76.2	75.4
ATGGGTGGGAGCCTCGGTAGCAGCC	72.8	73.3	74.7	75.5	75.8	76.4	76.9	77.6	77.6	77.9
GCCAACTCCGTCGCCGTTCGTGCGC	74.3	75.1	76.2	77.2	76.8	77.8	77.5	78.8	78.0	79.0
ACGGGTCCCCGCACCGCACCGCCAG	78.4	78.7	80.1	80.3	80.8	81.0	81.5	81.8	81.5	81.9
TTATGTATTAAGTTATATAGTAGTAGTAGT	53.1	53.7	55.8	56.7	57.7	58.6	60.4	61.4	61.1	62.2
ATTGATATCCTTTTCTATTCATCTTTCATT	56.4	57.2	59.5	60.1	61.3	61.8	63.6	64.2	64.7	65.1
AAAGTACATCAACATAGAGAATTGCATTTC	59.9	59.5	62.6	62.1	64.2	63.8	66.1	66.1	66.9	66.9
CTCAACTTGCGGTAAATAAATCGCTTAATC	62.7	62.2	65.2	64.8	66.6	66.4	68.7	68.5	69.3	69.2
TATTGAGAACAAGTGTCCGATTAGCAGAAA	63.7	63.2	66.4	65.7	67.9	67.2	70.0	69.2	70.6	69.9
GTCATACGACTGAGTGCAACATTGTTCAAA	64.8	64.0	67.1	66.3	68.4	67.8	70.2	69.6	71.0	70.3
AACCTGCAACATGGAGTTTTTGTCTCATGC	66.8	66.0	69.0	68.2	70.1	69.3	71.8	71.1	72.5	71.8
CCGTGCGGTGTGTACGTTTTATTCATCATA	66.1	65.4	68.3	67.5	69.6	68.8	71.3	70.5	72.0	71.0
GTTCACGTCCGAAAGCTCGAAAAAGGATAC	66.7	65.9	69.0	68.4	70.5	69.9	72.0	71.8	73.2	72.5
TCGGAGAAATCACTGAGCTGCCTGAGAAGA	68.4	67.6	70.8	70.0	72.0	71.1	73.6	72.6	74.4	73.4
CTTCAACGGATCAGGTAGGACTGTGGTGGG	69.3	69.0	71.6	71.0	72.4	72.1	73.6	73.3	74.4	73.8
ACGCCCACAGGATTAGGCTGGCCCACATTG	73.2	71.5	75.3	73.5	76.5	74.5	77.2	75.8	77.8	76.2
GTTATTCCGCAGTCCGATGGCAGCAGGCTC	72.7	71.2	74.7	73.3	76.0	74.5	77.1	75.8	77.7	76.2
TCAGTAGGCGTGACGCAGAGCTGGCGATGG	74.6	72.4	76.3	74.4	77.2	75.4	78.3	76.4	78.4	76.9
CGCGCCACGTGTGATCTACAGCCGTTCGGC	74.1	74.4	75.6	76.1	76.6	76.9	77.8	78.0	77.9	78.4
GCCCCTCCACTGGCCGACGGCAGCAGGCTC	77.9	76.9	79.9	78.7	80.5	79.4	81.0	80.2	81.5	80.5
CGCCGCTGCCGACTGGAGGAGCGCGGGACG	79.4	78.4	81.0	80.1	81.5	80.8	81.9	81.7	82.3	82.0

Table A.4 – Sodium equivalent concentrations $Na^+_{eq}$and ionic strengths I for Mg^{2+} and $Mg^{2+}+K^+$ sets. All concentrations are in mM. Note that for the Na^+ buffer $[Na^+] = [Na^+_{eq}] = I$.

		Mg^{2+}		$Mg^{2+}+K^+$	
$[Na^+]$	$[Mg^{2+}]$	$[Na^+_{eq}]$	I	$[Na^+_{eq}]$	I
69	0.5	74	1.5	134	51.5
119	1.5	128	4.5	188	54.5
220	3	181	9	241	59
621	10	330	30	390	80
1020	20	467	60	527	110
—	50	738	150	—	—
—	125	1167	375	—	—

Table A.5 – Generic seed parameters related to Morse potential, D and λ.

Base pair	D (meV)	λ (nm)
AT	32	3.6294×10^{-2}
AT*	32	3.6294×10^{-2}
CG	73	1.0156×10^{-2}
CG*	73	1.0156×10^{-2}

Table A.6 – Generic seed parameters of harmonic potential k.

NN	k (eV/nm^2)	NN	k (eV/nm^2)	NN	k (eV/nm^2)
ATpAT	2.41	ATpTA	1.84	GCpAT	2.80
ATpAT*	2.41	ATpTA*	1.84	GCpAT*	2.80
AT*pAT	2.41	CGpAT	3.44	GCpCG	3.36
ATpCG	2.56	CGpAT*	3.44	GCpCG*	3.36
ATpCG*	2.56	CGpCG	2.06	TApAT	2.42
AT*pCG	2.56	CGpCG*	2.06	TApAT*	2.42
ATpGC	2.25	CGpGC	2.73	CG*pCG	2.23
ATpGC*	2.25	CGpGC*	2.73	GC*pAT	2.54
AT*pGC	2.25	CG*pAT	2.34		

Table A.7 – Merit parameters χ^2 in °C^2 for Mg^{2+} and $Mg^{2+}+K^+$ sets.

Buffer	0.5 mM	1.5 mM	3 mM	10 mM	20 mM	50 mM	125 mM
Mg^{2+}	84	73	75	81	73	58	77
$Mg^{2+}+K^+$	79	69	64	50	56	—	—

Table A.8 – Morse potential depth D in meV optimized for DNA duplexes in solutions containing different concentration of magnesium, in buffers with 50 mM or no KCl. The standard deviation is displayed in parentheses in compact uncertainty notation.

Mg^{2+}	0.5 mM		1.5 mM		3 mM		10 mM		20 mM		50 mM	125 mM
K^+	—	50 mM	—	50 mM	—	50 mM	—	50 mM	—	50 mM	—	—
AT	30(1)	33(1)	28(1)	33(1)	26(1)	33(1)	28(1)	33(1)	26(1)	33(1)	32(1)	29(1)
AT*	27(2)	24(4)	26(2)	22(4)	26(3)	21(4)	27(2)	25(2)	25(3)	25(2)	26(2)	23(3)
CG	74(2)	76(1)	76(2)	75(1)	72(1)	76(2)	76(2)	76(1)	73(1)	76(1)	73(1)	72(1)
CG*	70(2)	73(2)	69(2)	68(2)	68(2)	66(1)	67(2)	65(1)	66(2)	64(1)	63(2)	63(2)

Table A.9 – Harmonic potential, coupling constant k in eV/nm^2 optimized for nearest-neighbors (NN) for solutions containing different concentrations of magnesium, in buffers with 50 mM or no KCl. The standard deviation is displayed in parentheses in compact uncertainty notation.

	0.5 mM		1.5 mM		3 mM		10 mM		20 mM		50 mM	125 mM
NN	—	50 mM	—	50 mM	—	50 mM	—	50 mM	—	50 mM	—	—
AT*pAT	8.03(82)	12.21(1.58)	10.50(1.11)	11.43(1.36)	9.44(96)	11.61(1.38)	9.42(1.14)	11.25(1.16)	9.10(95)	10.27(1.17)	8.18(90)	8.32(1.08)
AT*pCG	3.14(44)	2.67(40)	2.69(44)	2.73(42)	3.17(57)	2.53(40)	2.71(42)	3.01(42)	2.83(37)	2.63(40)	2.98(43)	2.68(46)
AT*pGC	2.07(23)	1.54(33)	2.40(28)	1.60(28)	2.31(30)	1.76(29)	2.17(24)	2.20(23)	2.21(35)	2.50(25)	2.60(30)	2.57(37)
ATpAT	2.50(18)	1.54(14)	2.36(20)	1.84(16)	2.53(22)	2.04(15)	2.46(22)	2.36(16)	2.81(22)	2.59(17)	2.48(19)	2.75(21)
ATpAT*	4.39(59)	3.58(67)	3.27(56)	3.45(67)	3.84(62)	3.94(73)	3.44(66)	3.72(71)	4.42(65)	3.02(59)	3.11(51)	4.12(62)
ATpCG	2.74(9)	2.85(11)	2.99(12)	2.82(11)	3.11(12)	2.79(12)	2.93(13)	2.80(9)	2.97(11)	2.85(9)	2.75(11)	2.75(12)
ATpCG*	3.69(33)	3.50(33)	5.08(47)	3.70(35)	4.83(48)	4.40(38)	5.05(58)	4.64(38)	4.78(53)	5.05(44)	4.52(45)	4.96(53)
ATpGC	2.41(11)	2.22(11)	2.64(12)	2.31(13)	2.50(12)	2.35(13)	2.69(15)	2.39(9)	2.57(15)	2.51(10)	2.59(13)	2.57(15)
ATpGC*	5.98(50)	6.72(54)	6.05(56)	6.60(57)	6.24(64)	6.21(53)	5.19(59)	5.74(50)	6.36(62)	6.06(53)	5.76(45)	5.70(60)
ATpTA	1.48(17)	2.00(21)	1.36(18)	2.03(22)	1.53(19)	1.95(20)	1.51(23)	1.83(18)	1.39(18)	1.79(18)	1.20(17)	1.13(16)
ATpTA*	0.37(7)	0.27(8)	0.29(5)	0.32(7)	0.27(6)	0.37(8)	0.22(4)	0.36(6)	0.20(4)	0.38(7)	0.45(8)	0.35(10)
CG*pAT	1.58(22)	1.98(30)	1.43(22)	1.33(24)	1.45(24)	1.30(24)	1.71(34)	1.15(22)	1.63(30)	1.14(22)	1.92(35)	1.48(26)
CG*pCG	0.81(8)	0.97(9)	0.87(9)	0.99(10)	0.79(9)	1.03(9)	0.75(9)	0.77(8)	0.76(9)	0.78(9)	0.82(10)	0.59(8)
CGpAT	2.89(11)	3.29(14)	2.61(11)	3.00(13)	2.55(12)	2.94(13)	2.59(13)	2.75(9)	2.55(11)	2.68(11)	2.67(11)	2.52(12)
CGpAT*	5.20(48)	4.89(54)	5.62(54)	5.95(66)	4.74(49)	5.88(69)	5.64(60)	6.86(57)	5.38(59)	6.44(56)	5.64(53)	6.85(82)
CGpCG	2.05(6)	2.28(10)	1.98(6)	2.27(9)	2.02(4)	2.25(7)	1.98(7)	2.14(3)	1.98(4)	2.11(3)	2.12(6)	2.06(7)
CGpCG*	4.32(43)	3.92(60)	4.11(37)	4.55(48)	4.35(72)	4.91(48)	4.83(48)	5.50(52)	4.96(10)	5.20(53)	4.01(39)	4.63(48)
CGpGC	2.39(9)	2.62(9)	2.25(10)	2.51(13)	2.18(10)	2.48(9)	2.34(13)	2.43(9)	2.18(11)	2.40(17)	2.45(8)	2.30(15)
CGpGC*	2.62(13)	3.68(28)	2.92(17)	3.40(29)	2.48(18)	3.28(26)	3.04(28)	3.43(23)	2.50(22)	3.34(26)	2.48(13)	2.52(25)
GC*pAT	1.26(15)	0.88(15)	1.27(16)	0.89(14)	1.44(19)	0.97(16)	0.99(15)	1.07(15)	1.38(19)	0.97(14)	1.75(21)	1.73(23)
GCpAT	3.07(12)	2.85(11)	2.77(13)	2.95(11)	3.01(16)	2.88(10)	2.84(15)	2.79(10)	2.85(13)	2.81(10)	2.78(13)	2.77(15)
GCpAT*	1.78(22)	3.08(39)	1.79(23)	2.91(45)	1.69(24)	2.29(41)	2.58(30)	1.58(23)	2.16(28)	2.18(29)	1.39(19)	1.52(24)
GCpCG	3.19(11)	3.30(9)	3.38(12)	3.42(22)	3.38(13)	3.49(17)	3.27(14)	3.55(12)	3.47(14)	3.52(18)	3.68(12)	3.52(24)
GCpCG*	4.06(23)	3.88(44)	4.27(40)	3.61(34)	4.16(32)	3.14(28)	4.05(33)	3.15(29)	4.20(42)	3.55(28)	4.58(44)	4.56(43)
TApAT	1.13(15)	0.85(13)	0.70(12)	0.91(13)	0.72(13)	0.94(12)	0.84(17)	0.84(11)	0.80(12)	0.82(11)	0.78(11)	0.89(15)
TApAT*	5.74(55)	7.30(70)	5.37(57)	6.59(61)	5.19(53)	6.05(65)	5.43(61)	6.06(56)	5.40(56)	5.05(49)	5.26(56)	5.88(65)

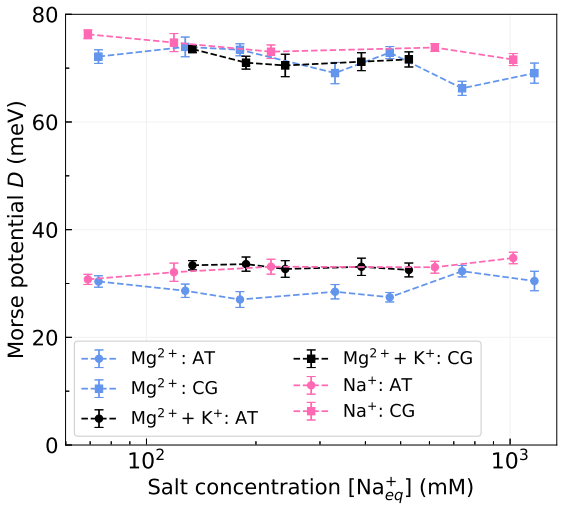

Figure 37 – Average Morse potentials D as function of the logarithm of equivalent sodium concentration $[Na^+_{eq}]$, $\beta = 3.3\,M^{1/2}$, for uniform AT (circles) and CG (bullets) base pairs, in Mg^{2+} (blue) and $Mg^{2+}+K^+$ (black). Results for Na^+ (pink) from Ref. [189] are included for comparison. Error bars are displayed only when larger than symbol size.

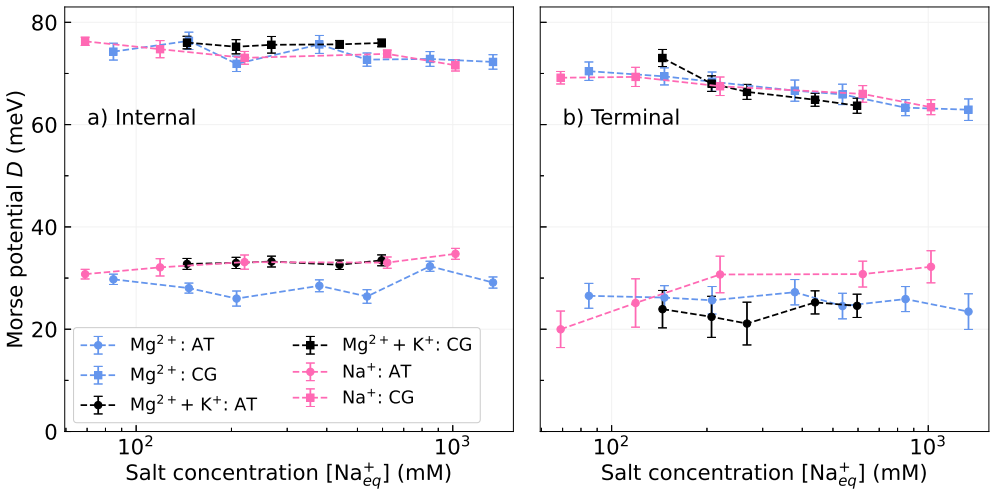

Figure 38 – Average Morse potentials D as function of the logarithm of equivalent sodium concentration $[Na^+_{eq}]$, $\beta = 3.79\,M^{1/2}$, for (a) internal and (b) terminal AT (circles) and CG (bullets) base pairs, in Mg^{2+} (blue) and $Mg^{2+}+K^+$ (black). Results for Na^+ (pink) from Ref. [189] are included for comparison. Error bars are displayed only when larger than symbol size.

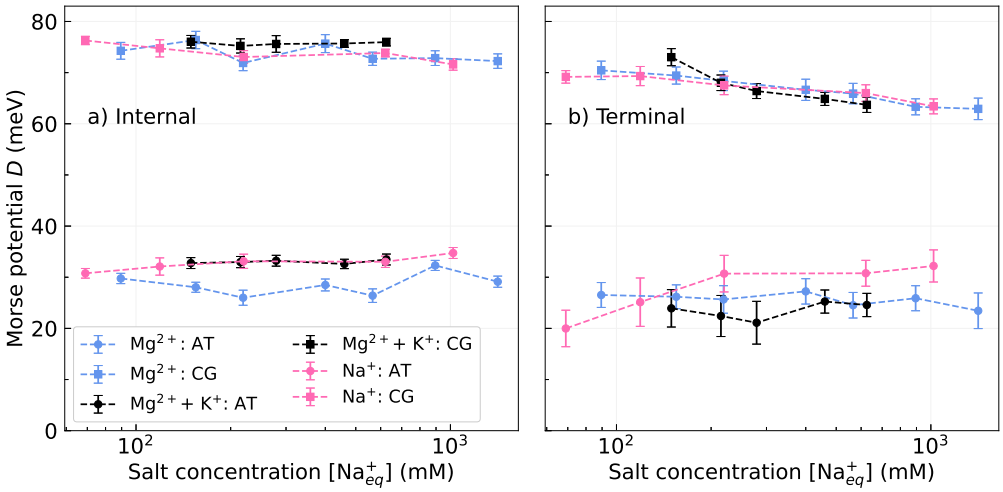

Figure 39 – Average Morse potentials D as function of the logarithm of equivalent sodium concentration $[Na^+_{eq}]$, $\beta = 4\,M^{1/2}$, for (a) internal and (b) terminal AT (circles) and CG (bullets) base pairs, in Mg^{2+} (blue) and $Mg^{2+}+K^+$ (black). Results for Na^+ (pink) from Ref. [189] are included for comparison. Error bars are displayed only when larger than symbol size.

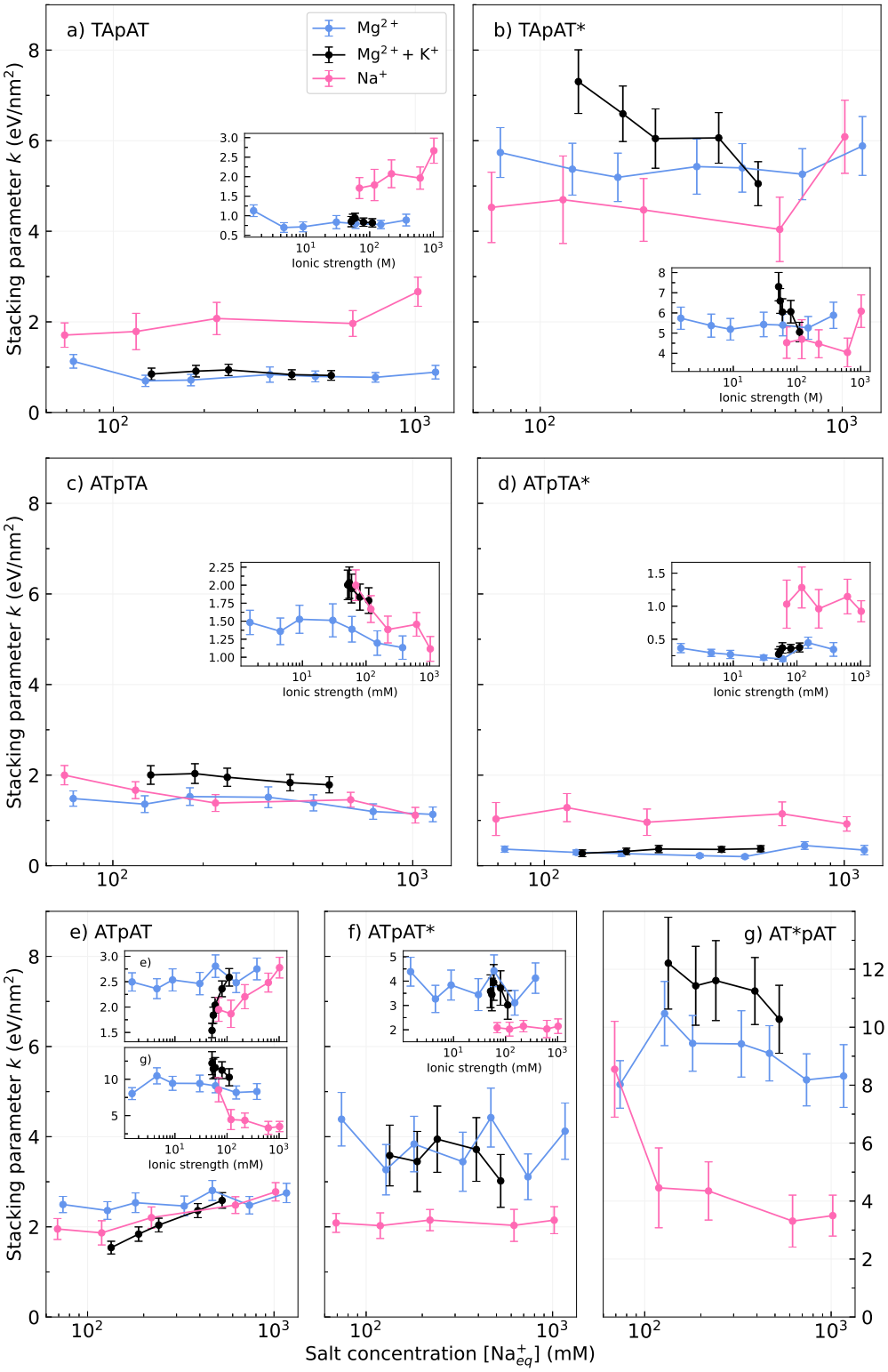

Figure 40 – Stacking parameters k as function of the logarithm of equivalent sodium concentration [Na^+_{eq}], $\beta = 3.3\,M^{1/2}$, for for (a,b) TApAT, (c,d) ATpTA and (e,f,g) ATpAT nearest-neighbors at internal (panels a,c,e) and terminal (panels b,d,f,g) positions. Results for Na^+ (pink) from Ref. [189] are included for comparison. Insets show the stacking parameters as function of ionic strength. Note that, for clarity, the inset for panel (g) was moved to panel (e, bottom inset).

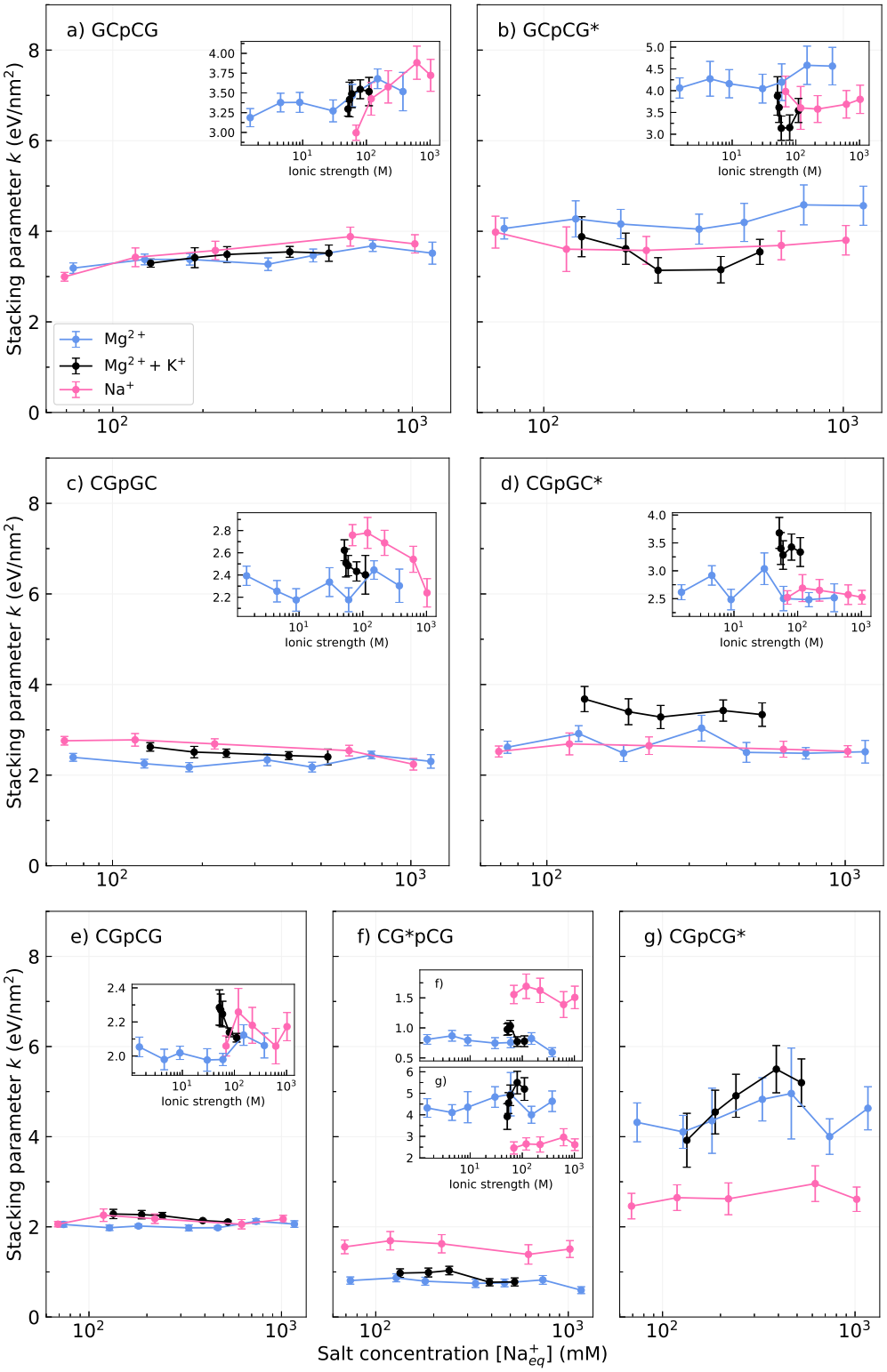

Figure 41 – Stacking parameters k as function of the logarithm of equivalent sodium concentration $[Na^+_{eq}]$, $\beta = 3.3\,M^{1/2}$, for (a,b) GCpCG, (c,d) CGpGC and (e,f,g) CGpGC; nearest-neighbors at internal (panels a,c,e) and terminal (panels b,d,f,g) positions. Results for Na^+ (pink) from Ref. [189] are included for comparison. Insets show the stacking parameters as function of ionic strength. Note that, for clarity, the inset for panel (g) was moved to panel (f, bottom inset).

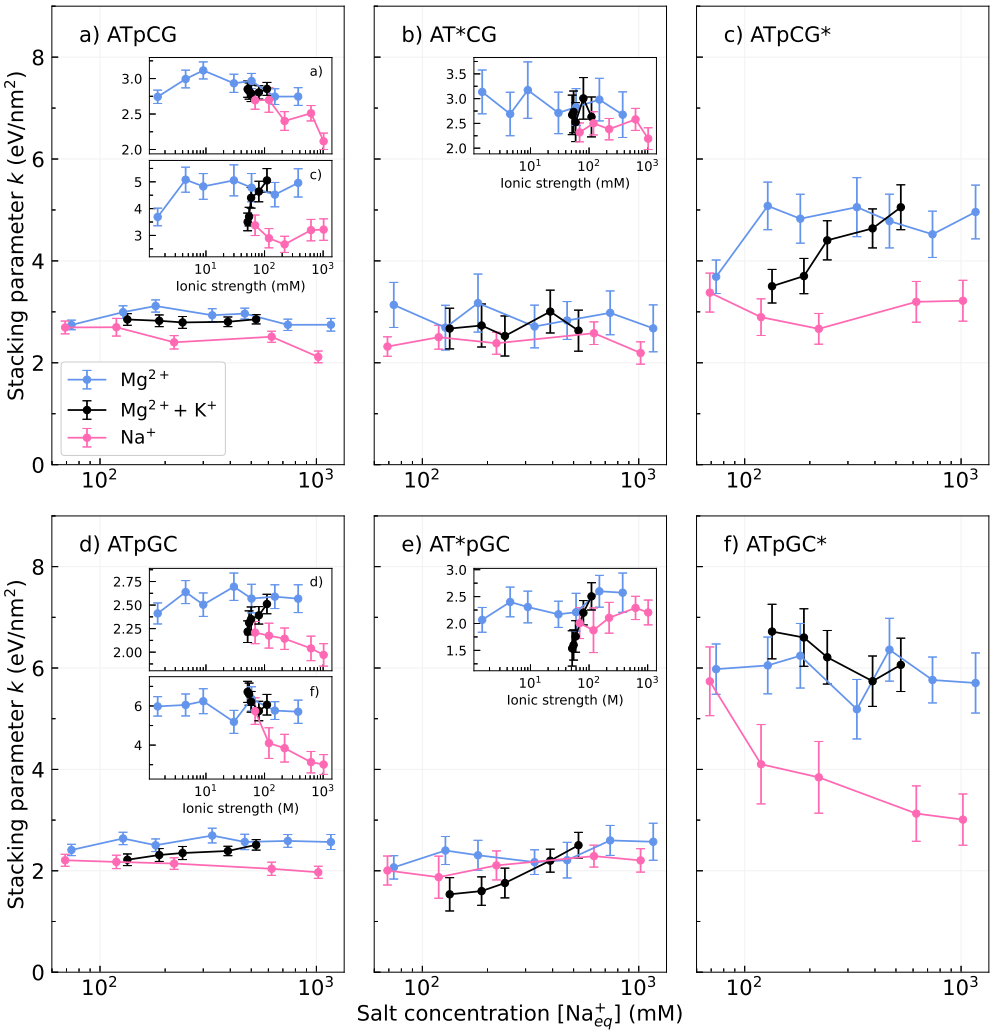

Figure 42 – Stacking parameters k as function of the logarithm of equivalent sodium concentration $[Na^+_{eq}]$, $\beta = 3.3\,M^{1/2}$, for (a,b,c) ATpCG and (d,e,f) ATpGC nearest-neighbors at internal (panels a,d) and terminal (panels b,c,e,f) positions. Results for Na^+ (pink) from Ref. [189] are included for comparison. Insets show the stacking parameters as function of ionic strength. Note that, for clarity, the inset for panel (c) was moved to panel (a, bottom inset), and for panel (f) was moved to panel (d, bottom inset).

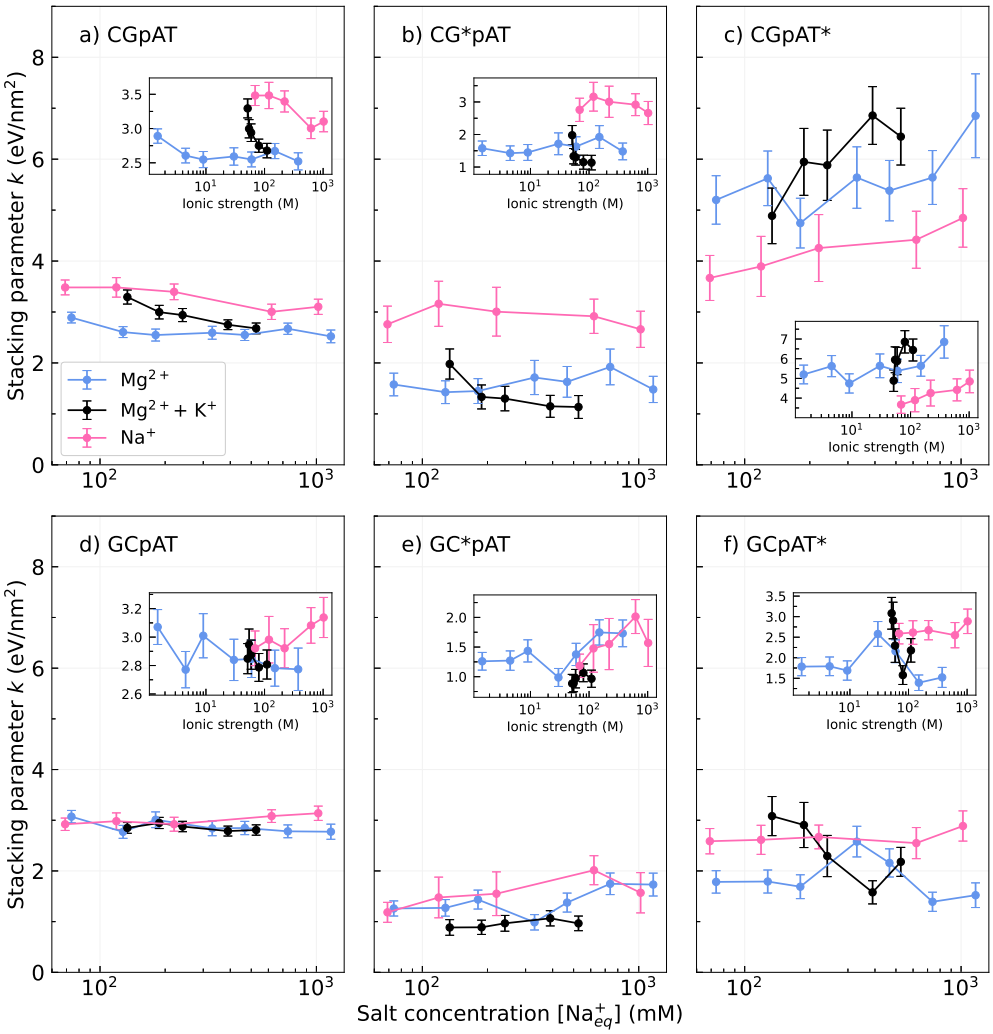

Figure 43 – Stacking parameters k as function of the logarithm of equivalent sodium concentration $[Na^+_{eq}]$, $\beta = 3.3\,M^{1/2}$, for (a,b,c) CGpAT and (d,e,f) GCpAT nearest-neighbors at internal (panels a,d) and terminal (panels b,c,e,f) positions. Results for Na^+ (pink) from Ref. [189] are included for comparison. Insets show the stacking parameters as function of ionic strength.

APPENDIX B – DNA/TNA mesoscopic modeling of melting temperatures suggests weaker hydrogen bonding of CG than in DNA/RNA

Chemical Physics Letters 749 (2020) 137413

Contents lists available at ScienceDirect

Chemical Physics Letters

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

Research paper

DNA/TNA mesoscopic modeling of melting temperatures suggests weaker hydrogen bonding of CG than in DNA/RNA

Maria Izabel Muniz[a], Hershel H. Lackey[b], Jennifer M. Heemstra[c], Gerald Weber[a,*]

[a] Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil
[b] Department of Chemistry, University of Utah, Salt Lake City, UT 84112, United States
[c] Department of Chemistry, Emory University, Atlanta, GA 30322, United States

HIGHLIGHTS

- Mesoscopic parameters of DNA/TNA hybrids were obtained from melting temperatures and compared to DNA/RNA.
- Base pairs involving AT in DNA/RNA were found to have similar hydrogen bonding strength as for DNA/RNA.
- For CG, base pairing strength is significantly reduced, but this is compensated for by a stronger stacking interaction.
- Opening profiles for DNA/TNA are mostly flatted out with little difference between AT and CG regions.

ARTICLE INFO

Keywords:
DNA/TNA hybrid
Peyrard-Bishop DNA model
DNA/TNA stability
Hydrogen bonds
Stacking interaction

ABSTRACT

TNA/DNA hybrids share several similarities to RNA/DNA, such as the tendency to form A-type helices and a strong dependency of their thermodynamic properties on purine/pyrimidine ratio. However, unlike RNA/DNA, not much is known about the base-pair properties of TNA. Here, we use a mesoscopic analysis of measured melting temperatures to obtain an estimate of hydrogen bonds and stacking interactions. Our results reveal that the AT base pairs in TNA/DNA have nearly identical hydrogen bond strengths than their counterparts in RNA/DNA, but surprisingly CG turned out to be much weaker despite similar stability.

1. Introduction

TNA (α-L-(3′-2′)-threofuranosyl nucleic acid) is a non-natural DNA analog that is capable of forming anti-parallel duplexes with both DNA and RNA [1], and was found to be more stable and resistant to degradation under physiological conditions [2]. Hybridization of DNA/TNA (DT) was found to be similar to that of DNA/RNA (DR) in a number of ways [3], especially the strong dependence of thermal stability on purine/pyrimidine ratio [4]. Another important similarity is the tendency of DT to form A-type helices, similar to what is found in DR [5,6]. These similarities have led to discussions if TNA could have been the natural precursor of RNA [7,8].

The sugar moiety in TNA contains carbon atoms and a single oxygen atom, and the phosphodiester groups are linked to the 2′ and 3′ positions of the threofuranose ring, as a result its backbone is shorter than that of DNA [9]. Nuclear magnetic resonance (NMR) and X-ray experiments largely confirmed the tendency of DT hybrids to form A-type helices [9,10,11]. This is hypothesized to result from the ability of DNA to adjust its conformation to TNA, reducing the inter-phosphate distance in a similar way as DR hybrids [9]. DT nucleotides were found to be in *anti* conformation in regard the glycosidic bond [12], but otherwise very little is known regarding their base-pair formation such as its hydrogen bonds and stacking interactions. Recently, we established a mesoscopic model to infer some of the structural information of oligonucleotides from melting temperatures, for instance we were able to show the stronger hydrogen bond in RNA [13], and also modeled DR [4] where we confirmed the deoxypyrimidine dependence on thermal stability. Here we apply the mesoscopic model to study the hydrogen bonds and stacking interactions for the DT hybrids based on melting temperatures.

The mesoscopic approach we use is the Peyrard-Bishop (PB) model [14], a simplified way to include the two main intra-molecular interactions of oligonucleotide duplexes which are the hydrogen bonds and stacking parameters. The PB model has found numerous applications, such as modeling G-quadruplexes [15], chaotic properties of promoter sequences [16], phonon modes [17] and DNA as a thermal transistor [18].

To calculate the melting temperatures we use an index calculated

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

https://doi.org/10.1016/j.cplett.2020.137413
Received 4 March 2020; Received in revised form 24 March 2020; Accepted 25 March 2020
Available online 10 April 2020

from the PB model and a correlation to measured temperatures [19]. We demonstrated that it is possible to run the model in reverse and to extract model parameters from the melting temperatures and in this way gain an insight of the intra-molecular interactions [20]. With the new parameters obtained in this way we are able to predict melting temperatures of new sequences and also calculate a qualitative opening profile of the double helix as function of temperature.

2. Materials and methods

2.1. Notation

To facilitate the description about the hybrid DNA/TNA (DT) duplex, we introduce the usual notation dA, dC, dG and dT for DNA bases and tA, tC, tG and tT for TNA bases. This is then used to describe the base pairs as dAtT, dTtA, dGtC and dCtG. The nearest-neighbor model allows us to define

$$\begin{matrix} {}^{3'}\mathrm{t(TA)}^{2'} \\ {}^{3'}\mathrm{d(AT)}^{5'} \end{matrix} \leftrightarrow \begin{matrix} {}^{3'}\mathrm{d(TA)}^{5'} \\ {}^{3'}\mathrm{t(AT)}^{2'} \end{matrix} \rightarrow \mathrm{dTtA\text{–}dAtT}. \quad (1)$$

Note that due to symmetry considerations, dTtA–dAtT is same as tTdA–tAdT. Therefore, for simplicity, we always use the notation starting with deoxy. When grouping in terms of purine or pyrimidine bases, we use the notation dR (tR) for dA, dG (tA, tG) and dY (tY) for dT, dC (tT, tC). Note that some authors use Pu for R, and Py for Y.

2.2. Melting temperature data set

In this work, we use 8 DT sequences and melting temperature data set from Schöning et al. [1], 14 from Lackey et al. [3] and 3 new as yet unpublished sequences, all at 1.0 M NaCl, 10 mM NaH_2PO_4, 0.1 mM EDTA, pH = 7.0 and at strand concentration C_t of 10 μM. The new sequences were synthesized and measured under the same conditions as in Ref. [3]. The complete list of sequences is shown in the Table 1.

2.3. Mesoscopic model

The Peyrard–Bishop (PB) model is a physical statistical model, in which the potential for describing the hydrogen bonds of two bases on opposite strands is a Morse potential [14], which defined for ith base pair as

$$V(y_i) = D_\alpha (e^{-y_i/\lambda_\alpha} - 1)^2, \quad (2)$$

where the parameter D_α represents the strength of the potential and λ_α is the potential width of a base-pair of type α. The stacking interaction between two neighboring base pairs is described by a harmonic potential, modified to consider a small twist angle θ [19,20]

$$W(y_i, y_{i-1}) = \frac{k_\beta}{2}(y_i^2 - 2y_i y_{i-1}\cos\theta + y_{i-1}^2), \quad (3)$$

where k_β is the coupling constant of nearest neighbors of type β, and the angle is fixed at $\theta = 0.01$ rad. We did not use a anharmonic potential [21], also known as the Peyrard-Bishop-Dauxois (PBD) model, as this would only increase the number of parameters without improving the parameter optimization, as shown in Ref. [20]. Nevertheless, the new parameters can be used together with the PBD model without problems.

With the two potentials of Eqs. (2) and (3) we can calculate the classical partition function from which we obtain an adimensional index τ_i, which is proportional to the melting temperatures. Additionally, we can calculate the average strand displacement $\langle y_i \rangle$ for each base pair i by calculating the expected value of y from the partition function [14,22].

2.4. Calculation of melting temperatures

The measured melting temperatures T_i are linearly correlated to the melting index $\tau_i(P)$ within groups of sequences of length N and for a given set of parameters P. This correlation allows us to calculate the linear regression coefficients $a_0(N)$ and $a_1(N)$ from which we can predict the melting temperature $T'_i(P)$ for any sequence

$$T'_i(P) = a_0(N) + a_1(N)\tau_i(P), \quad (4)$$

where $P = \{p_1, p_2, ...p_L\}$ is a set of L. The coefficients $a_{0,1}(N)$ follow a linear dependence with $N^{1/2}$

$$a_k(N) = b_{0,k} + b_{1,k}N^{1/2}, \quad k = 0, 1. \quad (5)$$

2.5. Thermal equivalence optimization

For the same ith sequence there is difference between the measured temperature T_i and the predicted temperature $T'_i(P)$ which depends on the parameter set P used to calculate the melting index $\tau_i(P)$. Our aim is to find a new set of parameters P that minimizes this difference

$$\chi_j^2(P_j) = \sum_{i=1}^{N} [T_i - T'_i(P_j)]^2, \quad (6)$$

where P_j is the jth tentative set of parameters [20].

Another equation considered in our discussion is the average absolute melting temperature deviation

$$\langle \Delta T \rangle = \frac{1}{N}\sum_{i=1}^{N} |T_i - T'_i(P_j)|. \quad (7)$$

Table 1
DNA/TNA sequences used in this work, of which only the TNA strand is shown, from 3′ to 2′. T_i are the measured, and T'_i the calculated melting temperatures in °C. The experimental temperatures T_i were obtained from the references indicated, those marked with ∗ are reported in this work.

Sequence	Reference	T_i	T'_i	Sequence	Reference	T_i	T'_i
t(TTTTTTTTTTTTTTTT)	[1]	32	32.1193	t(CTTACGCT)	[3]	22.8	21.2382
t(TAATAATATAAATTTT)	[1]	47	42.755	t(AGTCCTGA)	[3]	20	21.2645
t(TTTTAAATATAATAAT)	[1]	43	42.755	t(CTGAGTCC)	[3]	22.9	21.7645
t(CGCTGAAT)	[1]	25	27.1983	t(GAGCCGTG)	[3]	40.8	40.0352
t(AAAAAAAAAAAAAAAA)	[1]	68	68.0211	t(GCCGTGAG)	[3]	39.9	40.0352
t(AAAATTTATATTATTA)	[1]	41	40.7773	t(ACGTCATTCCTC)	[3]	44.6	44.9676
t(ATTATTATATTTAAAA)	[1]	36	40.7773	t(GCAATGTTCAGC)	[3]	51.1	52.014
t(ATTCAGCG)	[1]	26	23.9367	t(GCTGAACATTGC)	[3]	51	49.5808
t(AGATACAA)	[3]	25.2	25.3398	t(GAGGAATGACGT)	[3]	68.5	68.3511
t(AATACAGA)	[3]	25.1	25.3398	t(ATGGCGTGAC)	∗	55	54.8502
t(AAGCGTAG)	[3]	36.1	35.5452	t(CGCCTGTCTAGAAGTT)	∗	62	60.6895
t(AGCGTAAG)	[3]	35	35.5452	t(AACTTCTAGACAGGCG)	∗	63	64.8335
t(CTACGCTT)	[3]	20.3	21.2382				

Table 2
The parameter of Morse potential depth D optimized for DT hybrids, compared with previous results for related DR [4] base pairs. The standard deviation is displayed in parenthesis in compact uncertainty notation.

type	DT	D (meV)	type	DR	D (meV)
dYtR	dTtA	38.8(4)	dYrR	dTrA	40(2)
dRtY	dAtT	30.7(8)	dRrY	dArU	28(3)
dYtR	dCtG	53.5(5)	dYrR	dCrG	74(1)
dRtY	dGtC	42.0(3)	dRrY	dGrC	63(1)

2.6. Minimization procedure

The challenge of minimizing Eq. (6) is to find the optimal set P which in the case of DT is comprised of 20 parameters, 4 Morse potentials and 16 stacking interaction potentials. The most important problem here is the occurrence of local minima of χ^2 and to overcome this, a well tested approach is to perform the minimization procedure many times, each time starting from a different initial set of parameters, P_{init}. The downhill simplex algorithm was used for the numerical multi-dimensional minimization. The implementation of this algorithm was extensively verified with numerical test functions and also compared to continuous scale calculations of the PB parameters [23].

Here, we start with the same generic parameters as used for DNA and RNA [13,20], $D_{\text{dAtT}} = D_{\text{dTtA}} = 30$ meV, $D_{\text{dCtG}} = D_{\text{dGtC}} = 80$ meV, $\lambda_{\text{dAtT}} = \lambda_{\text{dTtA}} = 3.3333 \times 10^{-2}$ nm, $\lambda_{\text{dCtG}} = \lambda_{\text{dGtC}} = 1.25 \times 10^{-2}$ nm, and for all nearest-neighbors, $k = 2.5$ eV/nm^2. We call this set the seed parameters, P_{seed}, and for each new minimization we choose new initial set P_{init} where each initial parameter is randomly chosen within ± 10% of their corresponding seed parameters. The λ are kept at fixed values during all minimizations as we found that they have almost no influence over the final results [20], otherwise we would have to deal with yet another four parameters that would bring further difficulties for the convergence of the optimization procedure.

Initial minimization, D only. In this minimization, we start from the generic seed parameters and only calculate the Morse parameters D, while fixing all others. We calculated the minimization starting with 1000 different sets of initial parameters, and the best value of χ^2 found was 235 °C^2. The new Morse potentials were used as seed parameters for the next minimization step.

Second minimization, D and k. Here, we now calculated all parameters D and k, using as seed parameters the Morse potentials from the previous round. Again, we started with 1000 different sets of initial parameters around ± 10% of the previous seed parameters. We then averaged the resulting D and k parameters and their quality parameters were $\chi^2 = 125$ °C^2 and $\langle \Delta T \rangle = 1.8$ °C.

Third minimization, improving χ^2. We use the parameters from the second minimization as new seed values and calculated again 1000 rounds for further optimization. The quality parameters were reduced to $\chi^2 = 67$ °C^2 and $\langle \Delta T \rangle = 1.1$ °C.

Final minimization, estimate uncertainty. Lastly, we consider the influence of the experimental uncertainty on our optimized parameters. Instead of changing the initial parameters, we now modify the melting temperatures by small random amounts such that the standard deviation from the original set is 0.5 °C. Considering 1000 rounds of minimizations, we obtain the final average values of D and k as well as their relative uncertainty. The final quality parameters changed little compared to the previous round, $\chi^2 = 66$ °C^2 and $\langle \Delta T \rangle = 1.1$ °C.

2.7. Availability

The DT parameters calculated were included in the latest version of our free `TfReg` software [23] which can be used to verify our results. The software and the parameters are available at http://tinyurl.com/tfregufmg.

3. Results and discussion

For the numerical optimization of the 20 new parameters (4 Morse and 16 stacking potentials) we had only 25 independent sequences and melting temperatures. The limited data set is because TNA is not commercially available and is difficult to synthesize, and that is why there is relatively little information about its hybridization properties. For a linear system this would clearly be sufficient, but for a nonlinear model it is harder to establish *a priori* if overfitting is avoidable. A clear sign of overfitting is when the average temperature deviation of $\langle \Delta T \rangle$ results much smaller than the estimated experimental uncertainty. Fortunately, the resulting average temperature deviation was $\langle \Delta T \rangle = 1.1$ °C, which compared to a typical uncertainty of 0.5 °C, gives us confidence that numerical overfitting has not occurred. Note that for the nearest-neighbor model, the amount of required parameters is 32 (16 enthalpy and 16 entropy variations) and calculating these would not be possible for such a small amount of melting temperatures [24].

In Table 2 we show the calculated Morse potentials, which can be associated to hydrogen bond strengths [20], for the DT hybrids. The uncertainties shown are calculated following the procedure outlined in Section 2.6, they are not related to the numerical inaccuracy of the minimization algorithm but represent the influence of the experimental uncertainty over the new parameters. The TNA purine base (tR) clearly has a stabilizing influence as shown by the much stronger Morse potential of dTtA and dCtG over the pyrimidine base (tY), which suggest that this known purine dependence [3] is primarily due to hydrogen bonding. Note that dGtC with a tY base has a Morse potential which is barely larger than the dTtA base pair, which indicates that for certain applications there might be a problem discriminating between these types of base pairs. Also shown in Table 2 are the Morse potentials we obtained for DR hybrids in a previous work [4]. The comparison between the DT and DR Morse potentials shows two interesting results. First, the very close similarity of the dTtA and dAtT Morse potentials to their DR counterparts, indicating that the dT and dA bases accommodate really well to the TNA strand. This is consistent with the results from Pallan et al. [9] which observed this for dAtT and dTtA base pairs embedded in a B-DNA and A-DNA, respectively. However, the second result is probably more surprising: a comparatively weak Morse potential for the CG-type DT base pairs. Even the stronger dCtG potential falls well below the weaker DR dGrC potential. But the DT duplexes are generally not less stable than their DR equivalents, so how should we understand this results? As we will see next, it is the stacking potentials that will provide the additional stability.

The calculated stacking potential constants k are shown in Table 3 and their values range between 1.48 eV/nm^2 for dCtG-dAtT to 3.77 eV/nm^2 for dAtT-dCtG. We found no correlation between the DT stacking potentials and their DR counterparts, which are also shown in Table 3. Unlike the DR stacking potentials, which are quite small in some specific cases, we obtained no such weak potentials for DT. In fact, the higher DT stacking appears to counteract the weaker CG Morse potentials in several cases and provides for additional thermal stability.

The interplay of the various potentials within the nonlinear PB model is better appreciated by examining the average displacement profiles for some example sequences. Average displacement profiles are obtained from calculating the the expected value $\langle y_i \rangle$ of the y distance of Eq. (2) for the ith base pair [14]. These profiles can be interpreted as the expected base pair separation at certain temperatures. Unfortunately, as the PB approach is a 2D model, some of the relevant degrees of freedom are absent from the thermodynamics and for very short sequences one has to set unrealistically low calculation temperatures. Note that this calculation temperature is unrelated to the melting temperatures calculated from Eq. (4). There are several other mesoscopic models which consider additional degrees of freedom, for a comprehensive review see [25], however none of these were developed for high-throughput melting temperature calculation and they would require additional structural parameters for which we do not have the necessary

Table 3
The parameter of harmonic potential, coupling constant k optimized for DT hybrids. The standard deviation is displayed in parenthesis in compact uncertainty notation. Also shown, for comparison are the equivalent stacking potentials of DR hybrid k_{DR} from Ref [4]. All stacking potentials are given in eV/nm^2.

DT groups	NN	k_{DT}	k_{DR}	NN	k_{DT}	k_{DR}
dRtY-dRtY	dAtT-dAtT	3.02(7)	0.9(5)	dGtC-dAtT	1.81(6)	2.4(4)
	dAtT-dGtC	2.17(7)	2.8(4)	dGtC-dGtC	2.43(11)	2.6(2)
dRtY-dYtR	dAtT-dCtG	3.77(15)	2.8(3)	dGtC-dCtG	2.18(6)	2.8(2)
	dAtT-dTtA	2.48(10)	3.1(6)	dGtC-dTtA	3.01(10)	4.3(4)
dYtR-dYtR	dCtG-dCtG	2.62(14)	3.1(3)	dTtA-dCtG	2.47(6)	2.5(3)
	dCtG-dTtA	1.62(5)	3.1(3)	dTtA-dTtA	3.14(5)	2.4(2)
dYtR-dRtY	dCtG-dAtT	1.48(7)	2.6(3)	dTtA-dAtT	2.35(7)	0.8(4)
	dCtG-dGtC	2.70(11)	1.6(1)	dTtA-dGtC	2.94(8)	2.2(2)

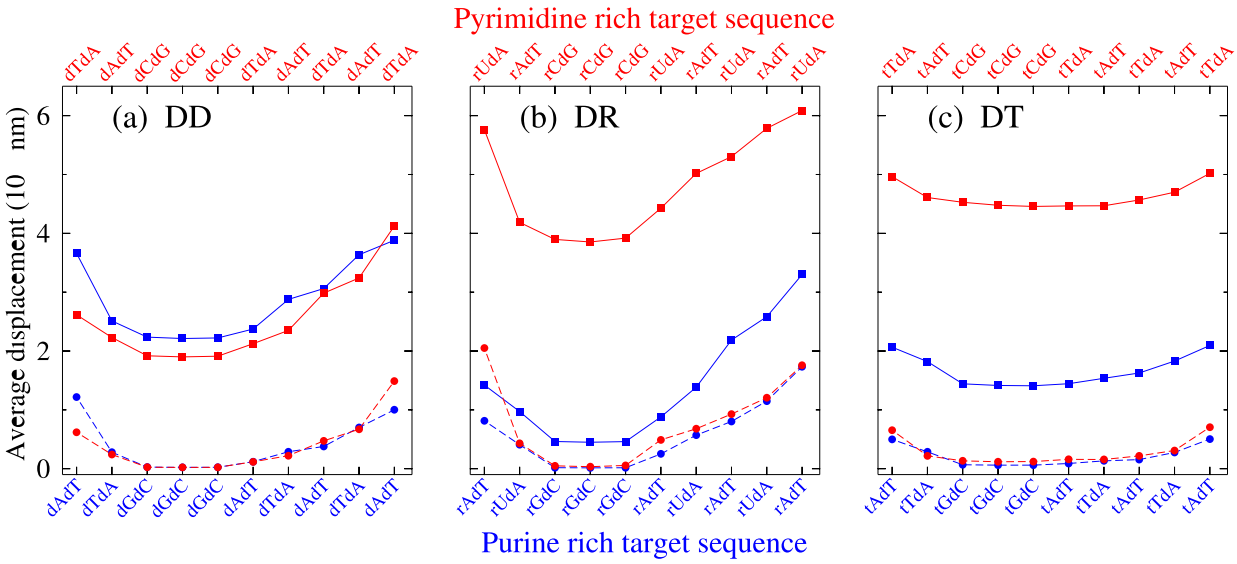

Fig. 1. Temperature dependence of average displacement profiles for (a) DD, (b) DR and (c) DT purine rich (blue) and pyrimidine rich (red) targets adapted from Ref. [26]. Bullets connected by dashed lines are for a calculation temperature of 150 K and squares connected by full lines are for 200 K. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

experimental data. Therefore, the displacement profiles of the PB model only give us a qualitative view of the expected base pair openings which, nevertheless, is helpful to understand the detailed influences the various potentials have over the stability of the duplex. In Fig. 1 we show six sequences adapted from Table S1 of Ref. [26], all are hybridized to a probe DNA strand. The complementary DNA, RNA and TNA are target sequences which are either purine or pyrimidine rich. For DNA/DNA (DD) sequences, the purine content has little impact, as these are homoduplexes, as seen in Fig. 1a, at higher temperatures the displacement profiles are still very similar. The situation changes completely for DNA/RNA (DR), at higher temperatures the purine rich target sequence is much more stable (lower $\langle y_i \rangle$), even more so than is DD equivalent, Fig. 1b. This purine/pyrimidine asymmetry is a well established property of DR hybrids [4,27] and has important biological consequences [28]. For DT hybrids, the thermal stability shows a similar asymmetry to purine/pyrimidine content to DR as shown in Fig. 1c. However, unlike DR, the DT shows a much more flattened out profile. While it is easy to make out the CG base pairs from the opening profile for DR, for DT it is much harder to tell where the CG base pairs are from the profile alone. Another example is shown in Fig. 2 where we show sequences with varying deoxypyrimidine (dPy) content for DR and DT. The DR profiles show a characteristic bulged opening in the central positions due to AT or AU base pairs. In contrast, for equivalent DT sequences the profiles are almost flat, without any such pronounced features. Also, for 0% dPy (100% tPy) content the DT sequence shows much higher opening implying in a much weaker thermal stability, which correlates to a very large dissociation rates observed for large tPy content [26].

4. Conclusion

We calculated the mesoscopic parameters from DNA/TNA melting temperatures which can be associated to hydrogen bonding and stacking interactions. Our results confirm the expected asymmetry of

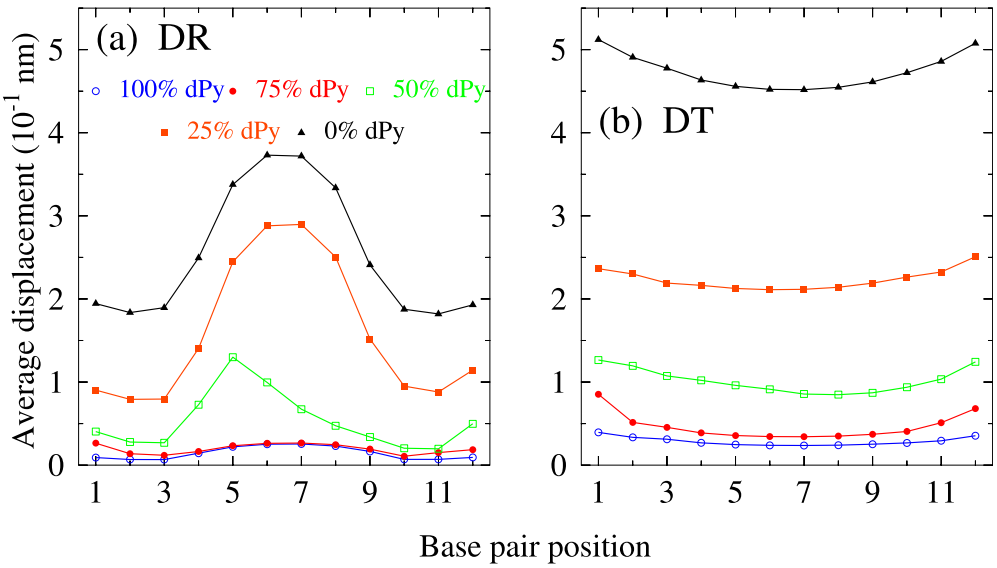

Fig. 2. Average displacement profiles for (a) DR and (b) DT with varying deoxypyrimidine (dPy) content, calculated at 200 K. Sequences were adapted from Ref. [27], calculations for DR are from Ref. [4].

the thermodynamic properties in regard to purine/pyrimidine content, similar to those of DNA/RNA hybrids. However, unlike DNA/RNA those the Morse potentials for CG base pairs turned out very small for DNA/TNA. The consequence of these small potentials is that for DNA/TNA hybrids the duplex appears to dissociate much for uniformly and largely independent of sequence composition. As researchers continue to explore the use of TNA in synthetic biology and nanotechnology applications, these data provide important insight into the design of sequences and prediction of duplex stability.

CRediT authorship contribution statement

Maria Izabel Muniz: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Visualization. **Hershel H. Lackey:** Conceptualization, Validation, Investigation. **Jennifer M. Heemstra:** Conceptualization, Resources, Supervision, Project administration. **Gerald Weber:** Conceptualization, Methodology, Supervision, Writing - original draft, Visualization, Project administration.

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

MIM and GW were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Brazil, Finance Code 001). JMH was supported by the National Science Foundation (CHE-1904885). HHL received funding from the Air Force Institute of Technology Civilian Institute Program.

References

[1] K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, Chemical etiology of nucleic acid structure: the α-threofuranosyl-(3′→2′) oligonucleotide system, Science 290 (5495) (2000) 1347–1351.
[2] M.C. Culbertson, K.W. Temburnikar, S.P. Sau, J.-Y. Liao, S. Bala, J.C. Chaput, Evaluating TNA stability under simulated physiological conditions, Bioorg. Med. Chem. Lett. 26 (10) (2016) 2418–2421.
[3] H.H. Lackey, E.M. Peterson, Z. Chen, J.M. Harris, J.M. Heemstra, Thermostability trends of TNA:DNA duplexes reveal strong purine dependence, ACS Synthetic Biol. 8 (5) (2019) 1144–1152.
[4] E.d.O. Martins, V.B. Barbosa, G. Weber, DNA/RNA hybrid mesoscopic model shows strong stability dependence with deoxypyrimidine content and stacking interactions similar to RNA/RNA, Chem. Phys. Lett. 715C (2019) 14–19, https://doi.org/10.1016/j.cplett.2018.11.015.
[5] M. Salazar, O.Y. Fedoroff, J.M. Miller, N.S. Ribeiro, B.R. Reid, The DNA strand in DNA.RNA hybrid duplexes is neither B-form nor A-form in solution, Biochem. 32 (16) (1993) 4207–4215.
[6] A. Noy, A. Pérez, M. Márquez, F.J. Luque, M. Orozco, Structure, recognition properties, and flexibility of the DNA.RNA hybrid, J. Am. Chem. Soc. 127 (13) (2005) 4910–4920.
[7] P. Herdewijn, TNA as a potential alternative to natural nucleic acids, Angew. Chem., Int. Ed. Engl. 40 (12) (2001) 2249–2251.
[8] Y.-W. Yang, S. Zhang, E.O. McCullum, J.C. Chaput, Experimental evidence that GNA and TNA were not sequential polymers in the prebiotic evolution of RNA, J. Mol. Evol. 65 (3) (2007) 289–295.
[9] P.S. Pallan, C.J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser, M. Egli, Why does TNA cross-pair more strongly with RNA than with DNA? An answer from X-ray analysis, Angew. Chem., Int. Ed. Engl. 42 (47) (2003) 5893–5895.
[10] K.-U. Schöning, P. Scholz, X. Wu, S. Guntha, G. Delgado, R. Krishnamurthy, A. Eschenmoser, The α-L-Threofuranosyl-(3′→2′)-oligonucleotide System ('TNA'): Synthesis and Pairing Properties, Helv. Chim. Acta 85 (12) (2002) 4111–4153.
[11] I. Anosova, E.A. Kowal, N.J. Sisco, S. Sau, J.-Y. Liao, S. Bala, E. Rozners, M. Egli, J.C. Chaput, W.D. Van Horn, Structural insights into conformation differences between DNA/TNA and RNA/TNA chimeric duplexes, ChemBioChem 17 (18) (2016) 1705–1708.
[12] C.J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser, M. Egli, Crystal structure of a B-form DNA duplex containing (L)-α-threofuranosyl (3′→2′) nucleosides: a four-carbon sugar is easily accommodated into the backbone of DNA, J. Am. Chem. Soc. 124 (46) (2002) 13716–13721.
[13] 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 http://nar.oxfordjournals.org/content/41/1/e30.
[14] M. Peyrard, A.R. Bishop, Statistical mechanics of a nonlinear model for DNA denaturation, Phys. Rev. Lett. 62 (23) (1989) 2755–2757.
[15] A. Bergues-Pupo, F. Falo, A. Fiasconaro, Modelling the DNA topology: the effect of the loop bending on G-quadruplex stability, J. Stat. Mech.: Theory Exp. 2019 (9) (2019) 094004.
[16] M. Hillebrand, G. Kalosakas, A. Schwellnus, C. Skokos, Heterogeneity and chaos in the Peyrard-Bishop-Dauxois DNA model, Phys. Rev. E 99 (2) (2019) 022213.
[17] L.B. Alexandrov, K.Ø. Rasmussen, A.R. Bishop, B.S. Alexandrov, Evaluating the role of coherent delocalized phonon-like modes in DNA cyclization, Sci. Rep. 7 (1) (2017) 9731.
[18] S. Behnia, R. Panahinia, Molecular thermal transistor: dimension analysis and mechanism, Chem. Phys. 505 (2018) 40–46.
[19] 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.
[20] 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.

[21] T. Dauxois, M. Peyrard, A.R. Bishop, Entropy-driven DNA denaturation, Phys. Rev. E 47 (1) (1993) R44–R47.
[22] 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.
[23] G. Weber, TfReg: calculating DNA and RNA melting temperatures and opening profiles with mesoscopic models, Bioinformatics 29 (2013) 1345–1347, https://doi.org/10.1093/bioinformatics/btt133 http://bioinformatics.oxfordjournals.org/content/29/10/1345.
[24] G. Weber, Optimization method for obtaining nearest-neighbour DNA entropies and enthalpies directly from melting temperatures, Bioinformatics 31 (6) (2015) 871–877, https://doi.org/10.1093/bioinformatics/btu751 http://bioinformatics.oxfordjournals.org/content/31/6/871.
[25] M. Manghi, N. Destainville, Physics of base-pairing dynamics in DNA, Phys. Rep. 631 (2016) 1–41.
[26] H.H. Lackey, Z. Chen, J.M. Harris, E.M. Peterson, J.M. Heemstra, Single-Molecule Kinetics Show DNA Pyrimidine Content Strongly Affects RNA:DNA and TNA:DNA Heteroduplex Dissociation Rates, ACS Synth. Biol. doi: 10.1021/acssynbio.9b00471.
[27] G. Suresh, U.D. Priyakumar, DNA–RNA hybrid duplexes with decreasing pyrimidine content in the DNA strand provide structural snapshots for the A-to B-form conformational transition of nucleic acids, Phys. Chem. Chem. Phys. 16 (34) (2014) 18148–18155.
[28] S. Hamperl, K.A. Cimprich, The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability, DNA Repair 19 (2014) 84–94.

APPENDIX C – Cation valence dependence of hydrogen bond and stacking potentials in DNA mesoscopic models

Cation valence dependence of hydrogen bond and stacking potentials in DNA mesoscopic models

Maria Izabel Muniz[a], Adrian H. Bustos[b], Sofie Slott[b], Kira Astakhova[b], Gerald Weber[a,1]

[a]*Departamento de Física, Universidade Federal de Minas Gerais, 31270-901, Belo Horizonte-MG, Brazil*
[b]*Department of Chemistry, Technical University of Denmark, 206-207 Kemitorvet, 2800 Kongens Lyngby, Denmark*

Abstract

Monovalent and divalent cations play a crucial role in living cells and for molecular techniques such as PCR. Here we evaluate DNA melting temperatures in magnesium (Mg^{2+}) and magnesium-potassium ($Mg^{2+}+K^{+}$) buffers with a mesoscopic model that allows us to estimate hydrogen bonds and stacking interaction potentials. The Mg^{2+} and $Mg^{2+}+K^{+}$ results are compared to previous calculations for sodium ions (Na^{+}), in terms of equivalent sodium concentration and ionic strength. Morse potentials, related to hydrogen bonding, were found to be essentially constant and unaffected by cation conditions. However, for stacking interactions we find a clear dependence with ionic strength and cation valence. The highest ionic strength variations, for both hydrogen bonds and stacking interactions, was found at the sequence termini. This suggests that end-to-end interactions in DNA will be strongly dependent on cation valence and ionic strength.

Introduction

DNA is strongly influenced by the presence of cations, such as sodium (Na^{+}) and magnesium ions (Mg^{2+}), which stabilize the double helix and influence its structural folding. These cations are present in cells and are used in molecular biology techniques. For instance Mg^{2+} is required for PCR, since DNA polymerase is an enzyme that is magnesium dependent [1]. Mg^{2+} has a much stronger stabilization on the double helix than Na^{+}. This is due to magnesium ions reducing the repulsive Coulomb interactions between phosphate groups [2], which neutralizes the negative charges on the DNA backbone [3]. As a result, the denaturation temperature in the presence of $MgCl_2$ is higher than NaCl at the same cation concentration [4]. This was already seen in early reports by Thomas [5] who found that to maintain native DNA at room temperature, the concentration of NaCl should be 100 times higher than that of $MgCl_2$. Similar conclusions were reached in several comparative studies of how Na^{+} and Mg^{2+} influences the denaturation [6–8]. Therefore, ionic strength is not regarded as a good predictor of melting temperatures. To overcome this problem, the concept of the sodium equivalent is used instead [9–11], that is, the concentration of Na^{+} required to result in a similar stabilization as Mg^{2+}. The sodium equivalent concept works well for melting temperature corrections where it provides simple way to relate measurements at varying buffer concentrations, but it is unclear if it could be extended to model potentials used mesoscopic descriptions of DNA. Here, we use a mesoscopic approach,

called Peyrard-Bishop (PB) model [12, 13], to understand how salt buffers of different types influence the hydrogen bonds and base pair stacking in DNA. In particular, we are interested in understanding how the ionic strength and equivalent sodium relate to these structural aspects of DNA, especially in terms of hydrogen bonding and base pair stacking

Ionic effects in DNA and RNA have a long history of experimental and theoretical studies, of which we will give a brief overview with a few examples. Early ultraviolet-visible (UV-Vis) spectroscopy [14] observed that the melting temperature of DNA increases with Mg^{2+} concentration and concluded that they bind to the phosphates. However, if the Mg^{2+} concentration exceeds the total strand concentration, a destabilization occurs as observed by Baba and Kagemoto [15] with differential scanning calorimetry (DSC). The infrared spectroscopy according to Serec et al. [16, 17] concluded that high concentrations of Mg^{2+} do not affect stacking interactions. Every and Russu [18] in particular concluded from NMR measurements that Mg^{2+} may lead to increased GCpCG openings. Using buffer equilibration–atomic emission spectroscopy ion counting techniques, Bai et al. [19] analyzed competitive binding, in particular between monovalent Na^+ and divalent cations Mg^{2+}. They have shown that already at moderate concentration of Mg^{2+} of 10 mM the Na^+ have been out competed entirely, which was further confirmed by Xi et al. [20]. In comparison to RNA, DNA has a number of differences when it comes to ionic interactions. For instance, using ion counting Gebala and Herschlag [21] showed that the electrostatic field around DNA is weaker than in RNA duplexes. One of the main reasons for this difference is that phosphoryl groups in double-stranded DNA are oriented towards the solvent, whereas for RNA they face inwards [21].

Theoretical approaches include methods such as Monte Carlo simulation [22], molecular dynamics (MD) [23–27], Poisson-Boltzmann polyelectrolyte (PBP) theory [28–31], and counterion condensation polyelectrolyte (CCP) theory [20, 32]. A key finding of atomistic MD simulations, such as from Li et al. [25], is that Mg^{2+} ions are positioned at the phosphate backbone and in the major groove of GC base pair, concluding that Mg^{2+} binding should be sequence-specific. Other MD simulations, such as from Mukherjee and Bhattacharyya [33], concluded that Na^+ binds largely to the minor groove, while potassium ions (K^+) binds to the major groove and closer to the center of the duplex. The MD simulations by Lavery et al. [34] showed that K^+ is localized, most of the time, within the major groove. More recently Kolesnikov et al. [35], compare the MD calculations of Na^+ and K^+, showing that the binding affinity of monovalent ions has an important dependence on the solvent model when these ions are deeply bound. PBP calculations from Misra and Draper [28] observed an anticooperativity effect for Mg^{2+} concentrations higher than 10 mM caused by a saturation of the ion binding to DNA. In other PBP calculations, Gebala and Herschlag [21] found that Na^+ interacts with GC and AT base pairs in both grooves, being the occupancy in the major groove larger than the minor one. Furthermore, in relation to the structure, the binding of Mg^{2+} to DNA was found to be more rigid than for Na^+ bound to DNA [33].

Mesoscopic models, such as the Peyrard-Bishop (PB) model, use a simplified representation of hydrogen bonds and stacking to describe the thermodynamic stability of DNA [36]. Along with experimental melting temperatures, these models were able to show how the intramolecular interactions are affected by Na^+ [13]. In particular, it was

found that the hydrogen bonding has little dependence with Na^+ concentrations. In another study, Ferreira et al. [37] used this model to separate the effects of internal and terminal base pairs. Differently from internal base pairs, terminal pairs do show some dependence on Na^+ concentrations. Being able to describe terminal base pairs is a unique feature of the mesoscopic model. In comparison, even the versatile MD method is limited in its ability to deal with terminal base pair interactions [38], especially terminal AT base pairs. Here, we extend these studies to include divalent cations Mg^{2+}, as well as mixed mono- and divalent $Mg^{2+}+K^+$ buffers. Together with our previous studies on Na^+, this allows us to draw a comprehensive picture of the differences between the various types of buffers and how they affect the thermal stability of DNA. Using multiple types of buffers allows us to understand how the ionic strength and sodium equivalent relate to the hydrogen bonds and stacking interactions, as well as how they affect the terminal base pairs. For our theoretical calculations, we use the published melting temperatures from Owczarzy et al. [11, 39], complemented with 10 new sequences which are being reported here. The new sequences were used to ensure a better representation of certain terminal base pairs as well as to complement under-represented length groups.

Materials and methods

Melting temperature data set

Mg^{2+} set. The main dataset related to variable Mg^{2+} concentrations was drawn from Refs. [11, 39] consisting of 92 sequences. In addition, 10 new as yet unpublished sequences were synthesized and measured such as to complement the existing dataset, especially in regard to terminal AT base pairs and the under-represented set of sequences of 11 bp length. These sequences were synthesized using automated solid-phase synthesis procedure in 200 nmol scale on ABI equipment (Expedite). After the synthesis, the sequences were cleaved with aqueous ammonia and analyzed by HPLC and MALDI MS. The purity of the sequences was >85% by IE HPLC. Their melting temperatures were measured on a DU800 UV/VIS spectrophotometer equipped with a Beckman Coulter Performance Temperature Controller. Complementary strands (0.5 μM of each strand), in a 1×PBS were mixed, denatured 10 min at 90 °C and subsequently cooled to 15 °C. The measured melting temperature values present the maximum of the first derivative of the curve and are an average of the two measurements with melting temperature result deviation within 1 °C. The measured melting temperatures are summarized in Table S1 for the new sequences. The 92 DNA sequences from Refs. [11, 39] and their melting temperature data, are all between 0.5 mM and 125 mM $MgCl_2$, and at 2 mM or 10 mM Tris-HCl, pH 8.3 and with total strand concentration C_t of 2 μM. Their respective measured and predicted melting temperatures are shown in Table S2. We will refer to the combined data of Tables S1 and S2 as the Mg^{2+} set.

$Mg^{2+}+K^+$ set. The second melting temperature data set has 80 DNA sequences from Ref. [39], all between 0.5 mM and 20 mM $MgCl_2$, 10 mM Tris-HCl, 50 mM KCl, pH 8.3 and a strand concentration C_t of 2 μM. We will refer to these melting temperature data as the $Mg^{2+}+K^+$ set. The complete list of sequences and their respective measured and predicted melting temperatures are shown in Table S3 in the Supporting Material.

Na^+ Equivalence

The equivalent, or effective, sodium concentration is defined as the equivalence between two different buffers: the one with magnesium ions [Mg^{2+}] and other monovalent ions [Mon^+], and another with sodium only at [Na^+_{eq}] that stabilizes duplexes to the same extent [11, 40]. This relation can be expressed as follows

$$[Na^+_{eq}] = \beta \sqrt{[Mg^{2+}]} + [Mon^+] \tag{1}$$

where the adjustable parameter β has values typically between 3.3 to 4.0 $M^{1/2}$ (M = mol/L), depending on the type of salt correction that was applied [9–11]. This relation expresses the fact that the magnesium ions stabilize duplexes much more effectively than sodium ions. Here, we used three specific values of the equivalence parameter β: 3.3 $M^{1/2}$ [11], 3.79 $M^{1/2}$ [10] and 4 $M^{1/2}$ [40].

Ionic strength

The ionic strength is defined as

$$I = \frac{1}{2} \sum m_i z_i^2 \tag{2}$$

where m_i and z_i are the ionic concentration in M (the same unit of I), and the charge number on the ion, respectively [41–43]. In this work, the values of I, related to the Mg^{2+} and $Mg^{2+} + K^+$ sets, are shown are in Table S4 in the Supporting Material, along with [Na^+_{eq}].

Mesoscopic model

Here we use the well tested mesoscopic Peyrard-Bishop (PB) model [36] which describes the denaturation of double-stranded helix through the use of interaction potentials. In this model, the stacking interaction between two neighboring base pairs, i and $i-1$, is described by a harmonic potential $W(y_i, y_{i-1})$, and the hydrogen bonding between base pairs is described by a Morse potential $V(y_i)$ [44],

$$W(y_i, y_{i-1}) = \frac{k_\alpha}{2}(y_i - y_{i-1})^2, \tag{3}$$

$$V(y_i) = D_\gamma \left(e^{-y_i/\lambda_\gamma} - 1\right)^2, \tag{4}$$

where k_α is the coupling constant of the nearest-neighbors of type α, and D_γ and λ_γ are the strength and width of Morse potential of a base pair of type γ, respectively. Another variation of the PB model is the Peyrard-Bishop-Dauxois (PBD) model, also called anharmonic PB model [44, 45], where the harmonic potential is replaced by anharmonic potential. However, for our theoretical calculations, this potential increases the number of parameters, leading to greater difficulty of convergence of calculations and without the benefit of a better agreement with experimental measurements as discussed by Weber et al. [13].

Melting temperature as a function of thermal equivalence

In the thermal equivalence approach we consider the potentials described in Eqs. (3) and (4) to calculate the classical partition function. This allows us to obtain an adimensional melting index τ_i, proportional to the experimental meting temperatures and dependent on the sequence length, for procedural details please see Ref. [12]. One of the advantages of this approach is the relatively small computational cost due to the optimization of the PB model parameters. Furthermore, it is possible to estimate the mean relative stretching $\langle y_i \rangle$ of hydrogen bonds by calculating the expected value of y from the partition function [36, 46]. The relation between melting index τ_i and melting temperatures [12], within groups of sequences of length N, for a given set of parameters P is

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

where $T_i'(P)$ is the predicted melting temperature, $P = \{p_1, p_2, \ldots p_L\}$ is a set of L, and $a_0(N)$ and $a_1(N)$ are the linear regression coefficients calculated as function of N,

$$a_k(N) = b_{0,k} + b_{1,k}N^{1/2}, \quad k = 0, 1. \tag{6}$$

To apply Eqs. (5) and (6) it is necessary to split the sequences in the dataset into groups of same length N. In particular, to calculate meaningful regression coefficients in Eq. (6) a group with specific length should contain at least three sequences.

Notation

Here, we describe the internal versus terminal base pairs notation. In the example below we have a system with two base pairs (AT and CG) and 10 nearest-neighbor interactions as typically used in nearest-neighbor models [47], corresponding to two Morse potentials D and 10 stacking interaction parameters k. Terminal base pairs are marked with * to distinguish hem from internal base pairs,

$$\begin{array}{l} 5'\text{-AGCGTAAGTC-}3' \\ 3'\text{-TCGCATTCAG-}5' \end{array} \longleftrightarrow \begin{array}{lll} \text{A} & \text{GCGTAAGT} & \text{C} \\ \underbrace{\text{T}}_{\text{term. AT*}} & \underbrace{\text{CGCATTCA}}_{\text{internal AT or CG}} & \underbrace{\text{G}}_{\text{term. CG*}} \end{array}$$

where we have AT* at the 5′-terminal and CG* at the 3′-terminal. The internal/terminal scheme adds further 18 parameters to be optimized, of which two are Morse potentials D and 16 are stacking interaction parameters k, resulting in a total of 30 model parameters.

For Morse potentials, Eq. (4), the base pairs AT* and TA*, and CG* and GC* take the same parameter D [37], that is $D_{\text{AT*}} = D_{\text{TA*}}$ and $D_{\text{CG*}} = D_{\text{GC*}}$. However, for the stacking interaction potential, we have a mixed notation of terminal and internal base pairs. Considering our previous example, the first nearest-neighbor is AT*pGC with terminal AT* and internal GC, which is symmetric and therefore equivalent to CGpTA* as shown here,

$$\begin{array}{llll} 5'\text{-AG} & \text{CGTAAG} & \text{TC-}3' \\ \underbrace{3'\text{-TC}}_{\text{AT*pGC}} & \text{GCATTC} & \underbrace{\text{AG-}5'}_{\text{TApCG*}} \end{array} \longleftrightarrow \begin{array}{llll} 5'\text{-GA} & \text{CTTACG} & \text{CT-}3' \\ \underbrace{3'\text{-CT}}_{\text{GC*pAT}} & \text{GAATGC} & \underbrace{\text{GA-}5'}_{\text{CGpTA*}} \end{array}$$

In view of this equivalence, we will write the notations always in lexical order, that is we will write AT*pGC and avoid writing CGpTA*. To ease the notational burden, occasionally we will refer to the various stacking configurations collectively, for instance by AT-AT we mean ATpAT, TApAT and ATpAT; by CG-CG we mean CGpCG, GCpCG and CGpGC; by AT-CG we mean ATpCG and ATpGC; by GC-AT we mean CGpAT and GCpAT.

Optimization

In the theoretical calculations, our aim is to minimize the parameter χ^2, whose role is to compare the experimental (T_i) and predicted melting temperatures (T_i'), until we find the best set of parameters adjusted P to the experimental data,

$$\chi_j^2(P_j) = \sum_{i=1}^{N} [T_i'(P_j) - T_i]^2, \tag{7}$$

where P_j is the jth tentative set of parameters [13], and the parameter set are the Morse potentials and stacking interactions. To minimize the Eq. (7) and find the minimum of a function we use the downhill simplex method as described in [48]. We also consider in our discussion the equation of average absolute melting temperature deviation

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

which, for the sake of the discussion, is easier to follow than Eq. (7).

Minimization procedure

To obtain the optimal set of parameters, P, by minimizing Eq. (7), we calculate 30 DNA parameters for each Mg^{2+} and Mg^{2+}+ K^+ buffer: 4 Morse potentials and 26 stacking interaction potentials, where the potentials differ in the internal and terminal base pairs. All minimization algorithms share an intrinsic local minima occurrence problem, that is, the algorithm may get 'stuck' in a minimum that is not really the global minimum of Eq. (7). The algorithm that is used in our case is no different in this respect, and to mitigate this problem we perform the minimization procedure many times, each time with different initial set of parameters, P_{init} in each round. Next, we will describe how this was done in our calculations. For the first round of minimization (R1), all calculations draw their initial parameters $P_{\text{init}}^{\text{R1}}$ from the seed parameters shown in Tables S5 and S6 in the Supporting Material, in the following way: for each minimization we randomly chose a fresh initial set $P_{\text{init}}^{\text{R1}}$ from within ±20% of their corresponding seed parameters P_{seed}. This is performed 1000 times, and we take the resulting parameters and calculate their averages over all calculations, which we call resulting averaged parameter set $P_{\text{res}}^{\text{R1}}$. Next, we repeat this procedure for a second round (R2), but now we draw the new initial parameters $P_{\text{init}}^{\text{R2}}$ from within ±10% of $P_{\text{res}}^{\text{R1}}$. This is repeated another 1000 times and results in the averaged parameter set $P_{\text{res}}^{\text{R2}}$. A final round (R3) is performed to evaluate the influence of the experimental uncertainty by considering the standard deviation of the experimental melting temperatures from the original set, reported in [11, 39] as 0.3 °C. This time the initial parameters are fixed at $P_{\text{res}}^{\text{R2}}$, and we vary the experimental temperatures T_i by small random amounts such that the standard deviation of the new set of temperatures

Table 1: Final merit parameters $\langle \Delta T \rangle$ in °C for the two sets.

Mg^{2+} concentration:	0.5 mM	1.5 mM	3 mM	10 mM	20 mM	50 mM	125 mM
Mg^{2+} set	0.69	0.68	0.70	0.67	0.67	0.60	0.69
Mg^{2+} + K^+ set	0.82	0.74	0.70	0.62	0.66		

falls within 0.3 °C of the original temperatures. R3 is repeated again 1000 times and the final results presented here are the averaged parameters $P_{\text{res}}^{\text{R3}}$ and the error bars in the figures are the standard deviations of $P_{\text{res}}^{\text{R3}}$. Each of the 12 buffer conditions is evaluated independently through rounds R1–R3. Final R3 values of $\langle \Delta T \rangle$ and χ^2, from Eqs. (8) and (7), are shown in Tables 1 and S7, respectively.

Availability

The new parameters were included in the latest version of our free `TfReg` software [49] which can be used to verify our results. The software and the parameters are available at `http://tinyurl.com/tfregufmg`, see Supporting Information for other download sites.

Results and discussion

The parameter optimization was carried out in a same way as in our previous work on Na^+ in DNA [37], except that this time we performed five time as many minimizations, which was made possible by code optimizations in the software that was used for the calculations (see section Availability). Final merit parameters $\langle \Delta T \rangle$, Eq. (8), are shown in Table 1 and are roughly twice as large as the reported experimental uncertainty of the data used [11], which is very similar to our previous results [13, 37]. Other melting temperature optimization approaches, such as by Freeman et al. [50] where coarse-grained models where employed with explicit ions, do not achieve this level of agreement to the experimental data.

One of the main questions we wish to address is how much the Morse potential depends on Mg^{2+} or Mg^{2+} + K^+ concentrations. Would it be similar to Na^+ or does cation valence play some role? To answer this question we show the results for Mg^{2+} and Mg^{2+} + K^+ alongside the previous results for Na^+ from [37] in Fig. 1. For CG base pairs we observe very little difference in regard of cation type, whether we plot the results as function of the equivalent Na^+_{eq} concentration Eq. (1) or as function of ionic strength. For terminal CG* we observe a discrete reduction for all buffer types as shown in Fig. 1b. Only terminal base pairs do display a moderate dependency on salt concentration, however terminal base pairs have a minor influence over the sequence in general. For AT base pairs we observe an overall similar dependence for all three buffer types, only for Mg^{2+} it appears slightly reduced for some concentrations for internal base pairs as shown in Fig. 1a. In general, we may conclude that one may extrapolate the use of a constant D to any Na^+_{eq} concentration and cation type. The weak dependence of internal D with any type of cation in DNA is different from Na^+ in RNA [51] where a larger dependence on salt concentration was found. The near complete independence

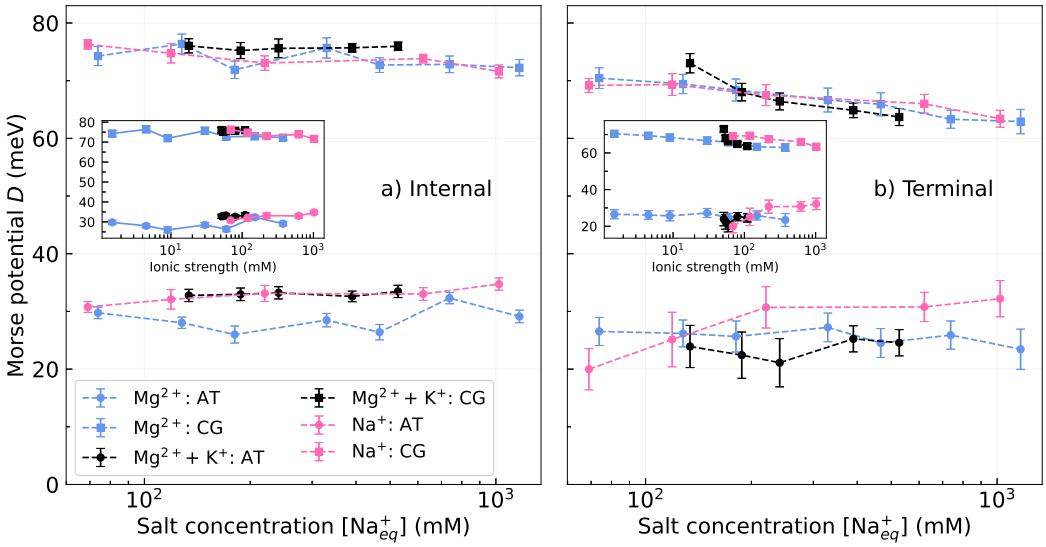

Figure 1: Average Morse potentials D as function of the logarithm of equivalent sodium concentration $[Na^+_{eq}]$, $\beta = 3.3\,M^{1/2}$, for (a) internal and (b) terminal of AT (circles) and CG (bullets) base pairs, in Mg^{2+} (blue) and $Mg^{2+} + K^+$ (black). Results for Na^+ (pink) from Ref. [37] are included for comparison. Insets show the Morse potentials as function of ionic strength. Error bars are displayed only when larger than symbol size.

of the Morse potential depth with valence and concentration suggest that there is no fundamental difference in the way Na^+ and Mg^{2+} influence the hydrogen bonds. This is consistent with molecular dynamic simulations which do not report important changes to the base pair hydrogen bonds Li et al. [25]. Note that, differently from simulations, our results are directly obtained from experimental data.

Another interesting finding is that the difference between CG and AT Morse potentials remains of the order of 40 meV for all cations and all concentrations. This can be verified in Table S8 in the Supporting Material which shows the numerical values of the Morse potentials depth D in terms of Mg^{2+} and K^+ concentrations. Compared to our previous results [13, 37, 52], the difference between CG and AT Morse potentials remains the same regardless of buffer type. This is now the combined result of 18 independent calculations comprising over 1100 melting temperature measurements from various sources [13, 37, 52]. In this sense this now firmly establishes a relation of 0.47 (35/75) between the internal AT and CG hydrogen bond strength for the mesoscopic model. Note that if one would naively just take the number of bonds this relation would be much larger, 0.67. However, DFT calculations, revised in Fonseca Guerra et al. [53], place this relation between 0.48 and 0.525, which is much closer to our result.

The next questions we wish to discuss is: does it matter whether we represent the Morse potential as function of equivalent Na^+_{eq} or as function of ionic strength? Or, to put it differently, would either one provide a better or perhaps a more useful description? Before we can proceed with this discussion we need to comment on the fact that the equivalent Na^+_{eq} concentration, Eq (1), depends on an empirical β factor for which there is no clear consensus. In Fig. 1 we show the results using $\beta = 3.3\,M^{1/2}$. However, for different values of β we observed very little change, see Figs. S1 ($\beta = 3.79\,M^{1/2}$) and S2 ($\beta = 4\,M^{1/2}$). While the results for the various β factors a very similar, $\beta =$

$3.3\,\mathrm{M}^{1/2}$ has a slightly better superposition for the various cations, especially for terminal base pairs at lower Na^+_{eq} concentration. Therefore, it would appear that the lower $\beta = 3.3\,\mathrm{M}^{1/2}$ better fulfills the idea of equivalence. For the plots as function of ionic strength the representation changes substantially. The results for the Mg^{2+} are concentrated at the lower end of the ionic strength scale, while those for Na^+ at the higher part as shown in the inset of Fig. 1. Basically, the results run almost monotonically with ionic strength, with the results for Na^+ continuing trend of Mg^{2+}, and even the $Mg^{2+}+K^+$ results do overlap nicely. Considering that ionic strength is a more fundamental and well established concept, it would be sensible to favor it in our analysis in comparison to the lesser known equivalent Na^+_{eq}, especially as it does not depend on adjustable parameters. Yet, as we will see next, the analysis in terms of ionic strength becomes even more interesting when applied to the stacking interaction.

The internal and terminal stacking parameters are shown in Figs. 2–5 as function of ionic strength, and in Figs. S3–S6 as function of equivalent sodium concentration, and their numerical values are given in Table S9 in the Supporting Material. It is immediately evident that stacking parameters have a variety of different dependencies with ionic strength, which we can broadly classify into a few groups. In one group the stacking of the $Mg^{2+}+K^+$ buffer overlaps with the Mg^{2+} buffer, while the stacking for Na^+ is either much higher or smaller, these are shown in TApAT (Fig. 2a), ATpTA* (Fig. 2d), ATpAT* (Fig. 2f) and CG*pCG (Fig. 3f) where the presence of K^+ plays no role at all. In contrast, there is a single case where the stacking of the $Mg^{2+}+K^+$ overlaps with Na^+ and is detached from Mg^{2+}, shown in Fig. 2c. Another group has a dependency that resembles that of the Morse potentials, where the stacking all three buffer types appear to be a continuous monotonic dependence with ionic strengths, these are the cases in Fig. 3a,c,e; Fig. 4a,b.; Fig. 5d. Perhaps the most interesting case is where the stacking for $Mg^{2+}+K^+$ starts at Na^+ and gradually increases to Mg^{2+}, that is, where initially the monovalent cation K^+ predominates and gradually the divalent Mg^{2+} takes over. Clearly, cation valence plays an important role for stacking but so does the nearest-neighbor (NN) configuration. In several cases, we notice a trend of the $Mg^{2+}+K^+$ stacking parameters to start from the Na^+ parameters at low ionic strength and then tend towards the Mg^{2+} parameters at higher concentrations. For several nearest-neighbor configurations, as Mg^{2+} in the $Mg^{2+}+K^+$ buffer starts to increase we retrieve similar results to the Mg^{2+} buffer. This is very clearly the case for ATpAT, Fig. 2e, ATpGC, Fig. 4d and CGpAT Fig. 5a. For these three nearest-neighbors the cation valence is the property that defines the stacking, as evidenced by the mixed $Mg^{2+}+K^+$. We interpret the trend of $Mg^{2+}+K^+$ as competition between Mg^{2+} and K^+, with Mg^{2+} gradually displacing K^+. For some other NN configurations, the $Mg^{2+}+K^+$ mostly coincides with the Mg^{2+} buffer, see for instance Fig. 2a, which means that even for small concentrations, Mg^{2+} completely dominates the binding. There is only one case, ATpTA Fig. 5c, where the $Mg^{2+}+K^+$ stacking coincides with Na^+, however the stacking potentials are all very close to each other, therefore no clear conclusion can be drawn here. For CG-CG NNs the $Mg^{2+}+K^+$, Mg^{2+} and Na^+ stacking potentials are in general very similar, with Na^+ appearing as a continuation of the Mg^{2+} stacking. In this case, the screening due to the cations seems to be very similar in all cases, and it is possible that the Mg^{2+} does not enter the major groove sufficiently enough to cause a major difference in the stacking. The stacking interaction parameter k shows little variation with Mg^{2+} concentration which is consistent with the findings from infrared spectroscopy by

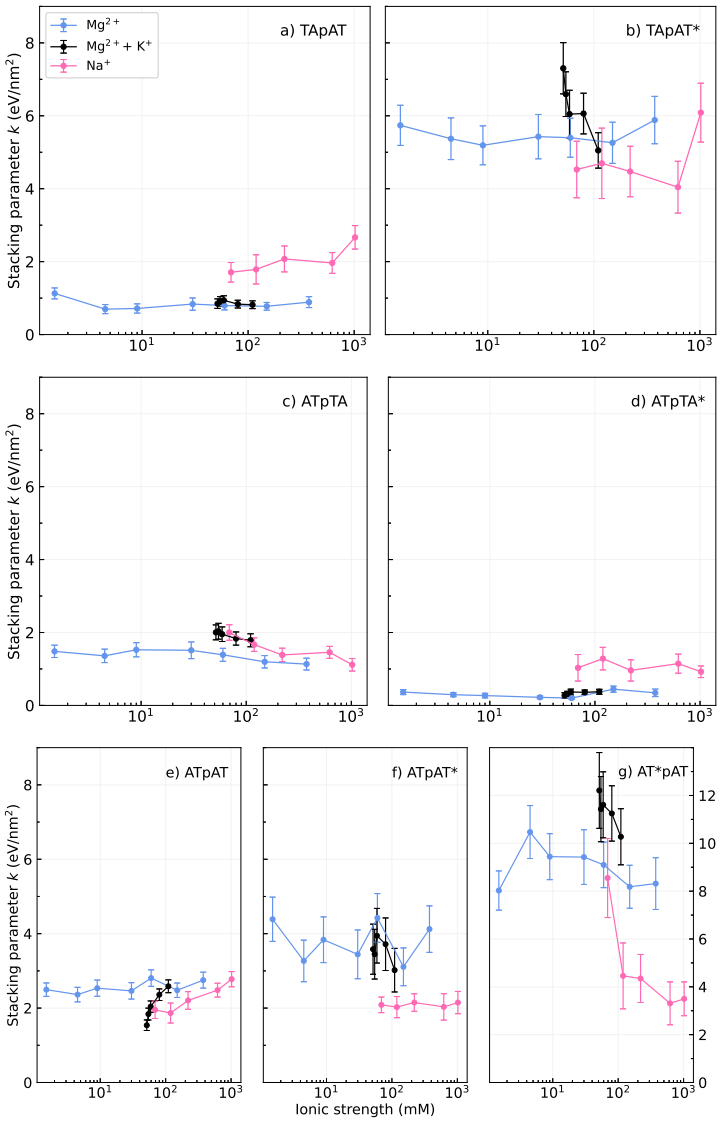

Figure 2: Stacking parameters k as function of ionic strength I, for (a,b) TApAT, (c,d) ATpTA and (e,f,g) ATpAT nearest-neighbors at internal (panels a,c,e) and terminal (panels b,d,f,g) positions. Results for Na^+ (pink) from Ref. [37] are included for comparison.

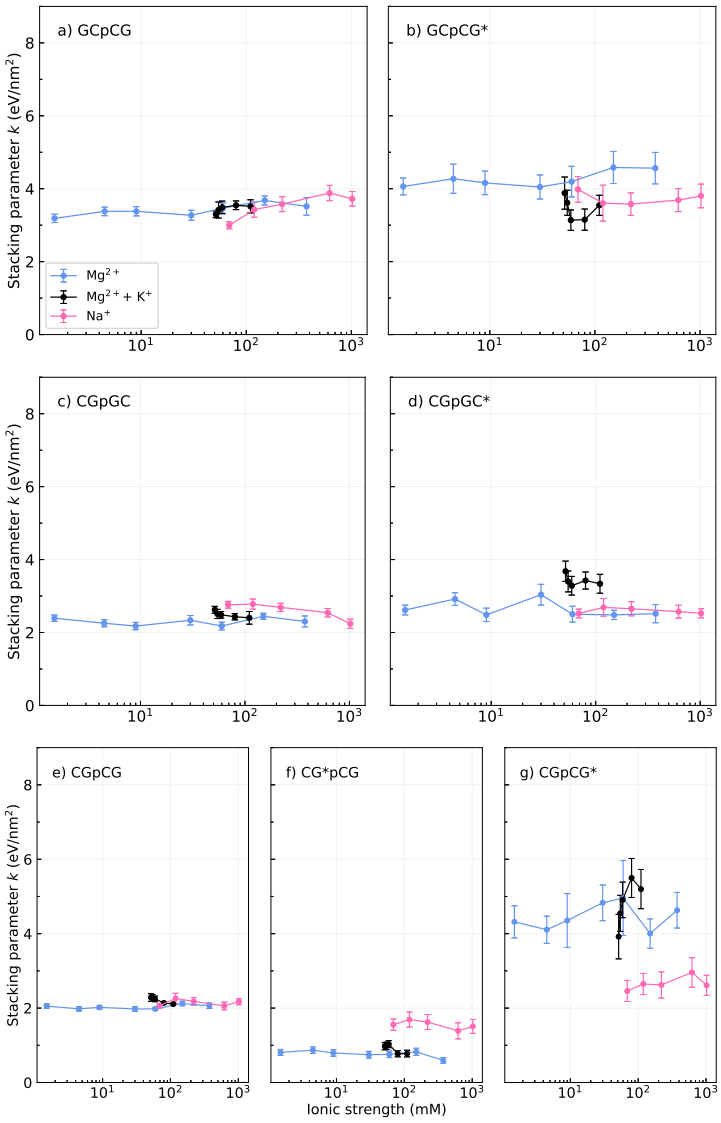

Figure 3: Stacking parameters k as function of as function of ionic strength I, for (a,b) GCpCG, (c,d) CGpGC and (e,f,g) CGpGC nearest-neighbors at internal (panels a,c,e) and terminal (panels b,d,f,g) positions. Results for Na^+ (pink) from Ref. [37] are included for comparison.

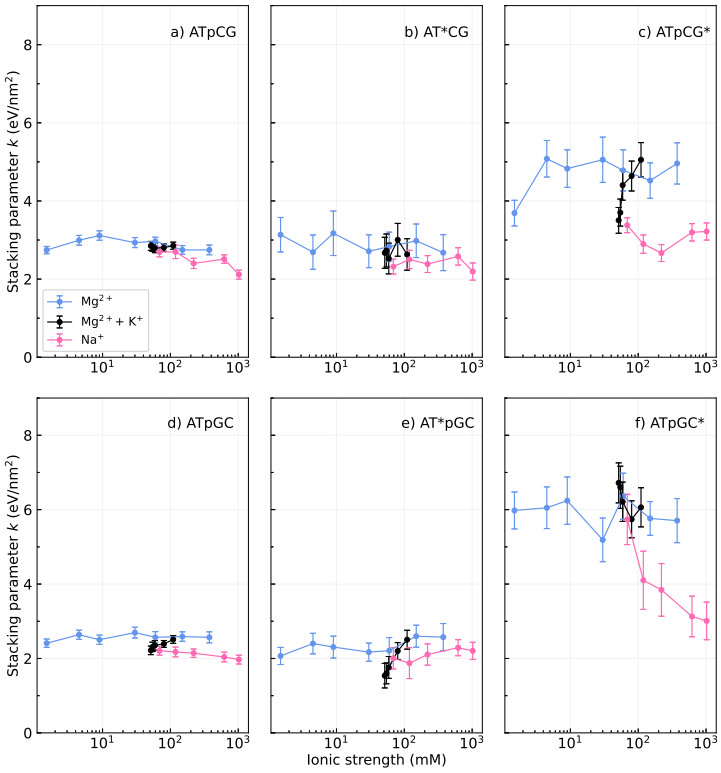

Figure 4: Stacking parameters k as function of ionic strength I, for (a,b,c) ATpCG and (d,e,f) ATpGC nearest-neighbors at internal (panels a,d) and terminal (panels b,c,e,f) positions. Results for Na^+ (pink) from Ref. [37] are included for comparison.

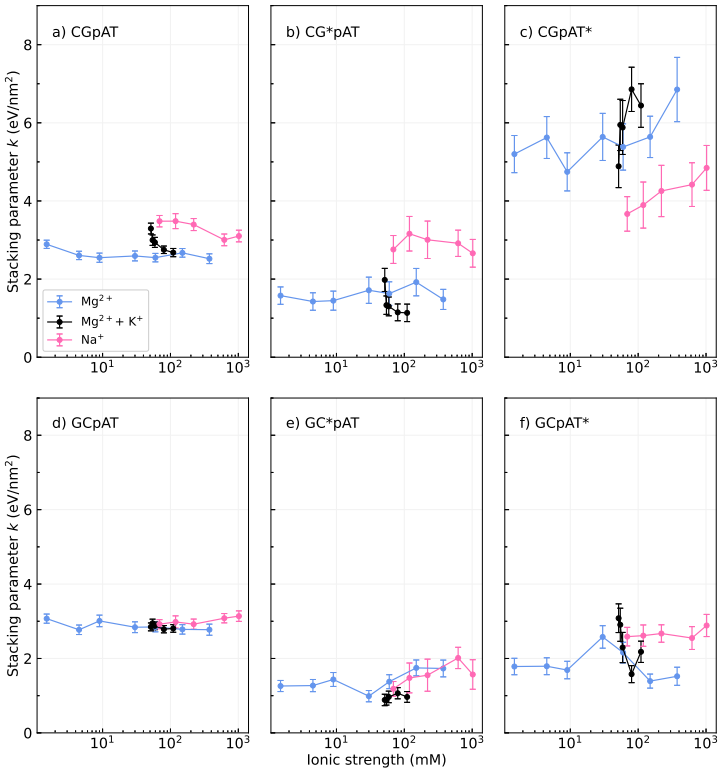

Figure 5: Stacking parameters k as function of ionic strength I, for (a,b,c) CGpAT and (d,e,f) GCpAT nearest-neighbors at internal (panels a,d) and terminal (panels b,c,e,f) positions. Results for Na^+ (pink) from Ref. [37] are included for comparison.

Serec et al. [16].

Terminal stacking parameters differ considerably when compared to internal stacking, with no cases of simultaneous superposition of the three buffer types. However, in most cases the $Mg^{2+}+K^{+}$ parameters tends towards those of the Mg^{2+}, and in general stacking differences are much larger compared to the internal nearest-neighbors. Some configurations have very large stacking differences between the various buffers, for instance CGpCG* Fig. 3g and ATpGC* Fig. 4f. Every and Russu [18] has suggested that increased Mg^{2+} concentrations would lead to an increased opening of GCpCG pairs that should be reflected in smaller stacking parameters. However, from Fig. 3a we observe no decrease in stacking that would support this conclusion.

Conclusion

We performed a comparative study of DNA Morse potentials and stacking interactions in varying concentrations of Mg^{2+} and $Mg^{2+}+K^{+}$, including previous results in Na^{+} buffers. Our results confirm that the internal AT and CG Morse potentials are effectively constant for any cation valence and concentration. Therefore, the internal hydrogen bonds, which here are represented by Morse potentials, appear to be shielded from the ion interaction regardless of valence. We analyzed the Morse potential as function of the ionic strength and the equivalent Na^{+}_{eq} concentration, and concluded that ionic strength provides an overall better description and has the potential to be extrapolated to other types of ions. Based on this observation we extended the ionic strength analysis also to the stacking interaction, where we obtained a distinct valence signature and signs of Mg^{2+} and K^{+} ion competition in the mixed $Mg^{2+}+K^{+}$ buffer. We believe that, in view of our results, detailed temperature measurements could be used to evaluate the ionic charge distribution as well as to understand the ion competition in DNA. The strong cation dependency of terminal stacking and hydrogen bonding suggest that it could be used for the design of DNA liquid crystal structures where the end-to-end interaction plays an important role in nematic ordering [54].

Author contributions

MIM and GW performed the calculations and theoretical analysis; AHB and SS conducted the synthesis and characterization of oligonucleotides; KA conducted the measurements.

Acknowledgements

MIM and GW were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Brazil, Finance Code 001). AHB, SS and KA were supported by DTU Elite PhD fellowship programme and Innoexplorer grant no. 41033 by Innovation Fond Denmark.

Supplementary data

Tables S1–S3 shows the sequences used and their experimental and predicted melting temperatures. Table S4 shows the sodium equivalent concentrations and ionic strength concentrations. Tables S5 and S6 show the seed parameters. Table S7 shows the final merit values. Table S8 and S9 show the calculated model parameters. Figures S1 and S2 show the Morse potentials displayed for different values of β. Figures S3–S6 show the stacking parameters as function of sodium equivalent.

References

[1] P. Markoulatos, N. Siafakas, M. Moncany, Multiplex polymerase chain reaction: a practical approach, J. Clin. Lab. Anal. 16 (2002) 47–51. doi:10.1002/jcla.2058.

[2] I. Sissöeff, J. Grisvard, E. Guillé, Studies on metal ions-DNA interactions: Specific behaviour of reiterative DNA sequences, Prog. Biophys. Mol. Biol. 31 (1976) 165–199. doi:10.1016/0079-6107(78)90008-1.

[3] Z. Tan, S. Chen, Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length, Biophys. J. 90 (2006) 1175–1190. doi:10.1529/biophysj.105.070904.

[4] J. W. Lyons, L. Kotin, The effects of magnesium ion on the secondary structure of deoxyribonucleic acid, J. Am. Chem. Soc. 87 (1965) 1781–1785. doi:10.1021/ja01086a030.

[5] R. Thomas, Recherches sur la d'enaturation des acides desoxyribonucléiques, Biochim. Biophys. Acta, 14 (1954) 231–240. URL: https://www.sciencedirect.com/science/article/abs/pii/0006300254901638. doi:10.1016/0006-3002(54)90163-8.

[6] W. F. Dove, N. Davidson, Cation effects on the denaturation of DNA, J. Math. Biol. 5 (1962) 467–478. doi:10.1016/S0022-2836(62)80119-3.

[7] M. T. Record, Jr, Effects of Na^+ and Mg^{++} ions on the helix–coil transition of DNA, Biophys. J. 14 (1975) 2137–2158. doi:10.1002/bip.1975.360141012.

[8] Y. P. Blagoi, V. Sorokin, V. Valeyev, S. Khomenko, G. Gladchenko, Magnesium ion effect on the helix-coil transition of DNA, Biopoly. 17 (1978) 1103–1118. doi:10.1002/bip.1978.360170502.

[9] C. Schildkraut, S. Lifson, Dependence of the melting temperature of DNA on salt concentration, Biopoly. 3 (1965) 195–208. doi:10.1002/bip.360030207.

[10] N. von Ahsen, C. T. Wittwer, E. Schütz, Oligonucleotide melting temperatures under PCR conditions: Nearest-neighbor corrections for Mg^{2+}, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas, Clin. Chem. 47 (2001) 1956–1961. doi:10.1093/clinchem/47.11.1956.

[11] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, J. A. Walder, Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations, Biochem. 47 (2008) 5336–5353. doi:10.1021/bi702363u.

[12] 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. doi:10.1038/nphys189.

[13] G. Weber, J. W. Essex, C. Neylon, Probing the microscopic flexibility of DNA from melting temperatures, Nat. Phys. 5 (2009) 769–773. doi:10.1038/nphys1371.

[14] G. L. Eichhorn, Y. A. Shin, Interaction of metal ions with polynucleotides and related compounds. XII. The relative effect of various metal ions on dna helicity, J. Am. Chem. Soc. 90 (1968) 7323–7328. doi:10.1021/ja01028a024.

[15] Y. Baba, A. Kagemoto, Influence of magnesium ions on helix-coil transition of DNA determined by modified differential scanning calorimeter, Biopoly. 13 (1974) 339–344. doi:10.1002/bip.1974.360130209.

[16] K. Serec, S. D. Babić, R. Podgornik, S. Tomić, Effect of magnesium ions on the structure of DNA thin films: an infrared spectroscopy study, Nucleic Acids Res. 44 (2016) 8456–8464. doi:10.1093/nar/gkw696.

[17] K. Serec, S. D. Babić, S. Tomić, Magnesium ions reversibly bind to DNA double stranded helix in thin films, Spectrochim. Acta, Part A 268 (2022) 120663. URL: https://doi.org/10.1016%2Fj.saa.2021.120663. doi:10.1016/j.saa.2021.120663.

[18] A. E. Every, I. M. Russu, Influence of magnesium ions on spontaneous opening of DNA base pairs, J. Phys. Chem. B 112 (2008) 7689–7695. doi:10.1021/jp8005876.

[19] Y. Bai, M. Greenfeld, K. J. Travers, V. B. Chu, J. Lipfert, S. Doniach, D. Herschlag, Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids, J. Am. Chem. Soc. 129 (2007) 14981–14988. URL: https://doi.org/10.1021/ja075020g. doi:10.1021/ja075020g. arXiv:https://doi.org/10.1021/ja075020g, pMID: 17990882.

[20] 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 (2018) 1776–1790. doi:10.1016/j.bpj.2018.03.001.

[21] M. Gebala, D. Herschlag, Quantitative studies of an RNA duplex electrostatics by ion counting, Biophys. J. 117 (2019) 1116–1124. doi:10.1016/j.bpj.2019.08.007.

[22] N. Korolev, A. P. Lyubartsev, A. Rupprecht, L. Nordenskiöld, Competitive binding of Mg^{2+}, Ca^{2+}, Na^{+}, and K^{+} ions to DNA in oriented DNA fibers: Experimental and Monte Carlo simulation results, Biophys. J. 77 (1999) 2736–2749. doi:10.1016/S0006-3495(99)77107-9.

[23] J. Alexander D. MacKerell, Influence of magnesium ions on duplex DNA structural, dynamic, and solvation properties, J. Phys. Chem. B 101 (1997) 646–650. doi:10.1021/jp9622795.

[24] N. Korolev, A. P. Lyubartsev, A. Laaksonen, L. Nordenskiöld, On the competition between water, sodium ions, and spermine in binding to DNA: A molecular dynamics computer simulation study, Biophys. J. 82 (2002) 2860–2875. doi:10.1016/S0006-3495(02)75628-2.

[25] W. Li, L. Nordenskiöld, Y. Mu, Sequence-specific Mg^{2+}–DNA interactions: A molecular dynamics simulation study, J. Phys. Chem. B 115 (2011) 14713–14720. doi:10.1021/jp2052568.

[26] H. Xu, N. Zhang, M. Li, F. Zhang, Comparison of the ionic effects of Ca^{2+} and Mg^{2+} on nucleic acids in liquids, J. Mol. Liq. 344 (2021) 117781. doi:10.1016/j.molliq.2021.117781.

[27] J. Xue, P. Wang, X. Li, R. Tan, W. Zong, Transformation characteristics of a-DNA in salt solution revealed through molecular dynamics simulations, Biophys. Chem. 288 (2022) 106845. URL: https://doi.org/10.1016%2Fj.bpc.2022.106845. doi:10.1016/j.bpc.2022.106845.

[28] V. K. Misra, D. E. Draper, The interpretation of Mg^{2+} binding isotherms for nucleic acids using poisson-boltzmann theory, J. Mol. Biol. 294 (1999) 1135–1147. URL: https://doi.org/10.1006%2Fjmbi.1999.3334. doi:10.1006/jmbi.1999.3334.

[29] N. Korolev, A. P. Lyubartsev, L. Nordenskiöld, Application of the Poisson Boltzmann polyelectrolyte model for analysis of thermal denaturation of DNA in the presence of Na^{+} and polyamine cations, Biophys. Chem. 104 (2003) 55–66. doi:10.1016/S0301-4622(02)00338-1.

[30] V. B. Chu, Y. Ba, J. Lipfert, D. Herschlag, S. Doniach, Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory, Biophys. J. 93 (2007) 3202–3209. doi:10.1529/biophysj.106.099168.

[31] B. P. Fingerhut, J. Schauss, A. Kundu, T. Elsaesser, Contact pairs of RNA with magnesium ions-electrostatics beyond the Poisson-Boltzmann equation, Biophys. J. 120 (2021) 5322–5332. doi:10.1016/j.bpj.2021.10.029.

[32] C. F. Anderson, J. M. Thomas Record, Polyelectrolyte theories and their applications to DNA, Annu. Rev. Phys. Chem. 33 (1982) 191–222. doi:10.1146/annurev.pc.33.100182.001203.

[33] S. Mukherjee, D. Bhattacharyya, Influence of divalent magnesium ion on DNA: molecular dynamics simulation studies, J. Biomol. Struct. Dyn. 31 (2013) 896–912. doi:10.1080/07391102.2012.713780.

[34] R. Lavery, J. H. Maddocks, M. Pasi, K. Zakrzewska, Analyzing ion distributions around DNA, Nucleic Acids Res. 42 (2014) 8138–8149. URL: https://doi.org/10.1093%2Fnar%2Fgku504. doi:10.1093/nar/gku504.

[35] E. S. Kolesnikov, I. Y. Gushchin, P. A. Zhilyaev, A. V. Onufriev, Similarities and differences between na^{+} and k^{+} distributions around DNA obtained with three popular water models, J. Chem. Theory Comput. 17 (2021) 7246–7259. URL: https://doi.org/10.1021%2Facs.jctc.1c00332. doi:10.1021/acs.jctc.1c00332.

[36] M. Peyrard, A. R. Bishop, Statistical mechanics of a nonlinear model for DNA denaturation, Phys. Rev. Lett. 62 (1989) 2755–2757. doi:10.1103/PhysRevLett.62.2755.

[37] 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.

[38] 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 (2014) 3177–3189. doi:10.1021/ct500120v.

[39] R. Owczarzy, B. Moreira, Y. You, M. Behlke, J. Walder, Method for estimating a melting temperature of a nucleic acid in buffers containing magnesium ions, 2012. US Patent 20,120,123,751.

[40] M. Mitsuhashi, Technical report: Part 1. Basic requirements for designing optimal oligonucleotide probe sequences, Journal of Clinical Laboratory Analysis 10 (1996) 277–284. doi:10.1002/(SICI)1098-2825(1996)10:5<277::AID-JCLA8>3.0.CO;2-5.

[41] G. N. Lewis, M. Randall, The activity coefficient of strong electrolytes, J. Am. Chem. Soc. 43 (1921) 1112–1154. doi:10.1021/ja01438a014.

[42] K. S. Pitzer, Gilbert n. lewis and the thermodynamics of strong electrolytes, J. Chem. Educ. 61 (1984) 104. doi:10.1021/ed061p104.

[43] T. Solomon, The definition and unit of ionic strength, J. Chem. Educ. 78 (2001) 1691. doi:10.1021/ed078p1691.

[44] T. Dauxois, M. Peyrard, A. R. Bishop, Entropy-driven DNA denaturation, Phys. Rev. E 47 (1993) R44–R47. doi:10.1103/PhysRevE.47.R44.

[45] S. Zdravković, Helicoidal Peyrard–Bishop model of DNA dynamics, Journal of Nonlinear Mathematical Physics 18 (2011) 463–484. doi:10.1142/S1402925111001635.

[46] 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 (1997) 7100–7115. doi:10.1103/PhysRevE.56.7100.

[47] K. J. Breslauer, R. Frank, H. Blocker, L. A. Marky, Predicting DNA duplex stability from the base sequence, Proc. Natl. Acad. Sci. USA 83 (1986) 3746–3750. doi:10.1073/pnas.83.11.3746.

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

[49] G. Weber, TfReg: Calculating DNA and RNA melting temperatures and opening profiles with mesoscopic models, Bioinformatics 29 (2013) 1345–1347. URL: http://bioinformatics.oxfordjournals.org/content/29/10/1345. doi:10.1093/bioinformatics/btt133.

[50] G. S. Freeman, D. M. Hinckley, J. J. de Pablo, A coarse-grain three-site-per-nucleotide model for DNA with explicit ions, J. Chem. Phys. 135 (2011) 165104. doi:10.1063/1.3652956.

[51] I. Ferreira, T. D. Amarante, G. Weber, Salt dependent mesoscopic model for RNA with multiple strand concentrations, Biophys. Chem. 271 (2020) 106551. doi:10.1016/j.bpc.2021.106551.

[52] 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. URL: https://pubs.rsc.org/en/content/articlelanding/2020/SC/D0SC01700K. doi:10.1039/d0sc01700k.

[53] C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, Hydrogen bonding in DNA base pairs: reconciliation of theory and experiment, J. Am. Chem. Soc. 122 (2000) 4117–4128. doi:10.1021/ja993262d.

[54] M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J. O. Cross, R. Pindak, T. Bellini, N. A. Clark, End-to-end stacking and liquid crystal condensation of 6–to 20–base pair DNA duplexes, Science 318 (2007) 1276–1279. doi:10.1126/science.1143826.

Bibliography

[1] Dahm, Ralf: *Discovering DNA: Friedrich Miescher and the early years of nucleic acid research*. Hum. Genet., 122(6):565–581, 2007. Cited on the page 19.

[2] Wikipedia: *Nucleic acid*, accessed in January 2020. `https://en.wikipedia.org/wiki/Nucleic_acid`. Cited on the page 19.

[3] Nelson, David L. and Michael M. Cox: *Lehninger Principles of Biochemistry*. W. H. Freeman and Company, 2008. Cited 7 times on the pages 19, 20, 21, 23, 24, 28, and 29.

[4] CNX, OpenStax: *Biology: Nucleic acids*, accessed in March 2020. `https://cnx.org/contents/GFy_h8cu@9.87:yxeAKc4X@8/Nucleic-Acids`. Cited on the page 20.

[5] Blanco, Antonio and Gustavo Blanco: *Medical Biochemistry*. Academic Press, 2017. Cited 2 times on the pages 20 and 24.

[6] Academy, Khan: *Discovery of the structure of DNA*, accessed in June 2022. `https://www.khanacademy.org/science/biology/dna-as-the-genetic-material/dna-discovery-and-structure/a/discovery-of-the-structure-of-dna`. Cited on the page 21.

[7] Pauling, Linus and Robert B. Corey: *A proposed structure for the nucleic acids*. Proc. Natl. Acad. Sci. USA, 39(2):84–97, 1953. Cited on the page 21.

[8] Watson, J. D. and F. H. C. Crick: *Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid*. Nature, 171:737–738, 1953. Cited on the page 21.

[9] Wikipedia: *Ficheiro: DNA Overview*, accessed in April 2020. `https://pt.wikipedia.org/wiki/Ficheiro:DNA_Overview.png`. Cited on the page 22.

[10] Wilkins, M. H. F., A. R. Stokes, and H. R. Wilson: *Molecular structure of deoxypentose nucleic acids*. Nature, 171:738–740, 1953. Cited on the page 21.

[11] Franklin, Rosalind E. and R. G. Gosling: *Molecular configuration in sodium thymonucleate*. Nature, 171(4356):740–741, 1953. Cited 2 times on the pages 21 and 22.

[12] Lu, Xiang-Jun and Wilma K. Olson: *3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures*. Nucleic Acids Res., 31(17):5108–5121, 2003. Cited on the page 22.

[13] Reference, Oxford: *DNA grooves*, accessed in June 2020. `https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095723885`. Cited on the page 23.

[14] Wikipedia: *DNA*, accessed in January 2020. `https://en.wikipedia.org/wiki/DNA`. Cited on the page 23.

[15] Wikipedia: *Nucleic acid double helix*, accessed in February 2020. `https://en.wikipedia.org/wiki/Nucleic_acid_double_helix`. Cited on the page 23.

[16] Crick, F. H. C. and J. D. Watson: *The complementary structure of deoxyribonucleic acid*. Proc. R. Soc. London, Ser. A, 223(1152):80–96, 1954. Cited on the page 23.

[17] Franklin, Rosalind E. and R. G. Gosling: *The structure of sodium thymonucleate fibres. I. The Influence of Water Content*. Acta Crystallogr., 6:673–677, 1953. Cited on the page 23.

[18] Donohue, Jerry: *Hydrogen-bonded helical configurations of polynucleotides*. Proc. Natl. Acad. Sci. USA, 42(2):60–65, 1956. Cited on the page 23.

[19] Pauling, Linus and Robert B. Corey: *Specific hydrogen-bond formation between pyrimidines and purines in deoxyribonucleic acids*. Archives of Biochemistry and Biophysics, 65(1):164–181, 1956. Cited on the page 23.

[20] Fick, Robert J., Amy Y. Liu, Felix Nussbaumer, Christoph Kreutz, Atul Rangadurai, Yu Xu, Roger D. Sommer, Honglue Shi, Steve Scheiner, and Allison L. Stelling: *Probing the hydrogen-bonding environment of individual bases in DNA duplexes with isotope-edited infrared spectroscopy*. J. Phys. Chem. B, 125(28):7613–7627, 2021. Cited on the page 23.

[21] Zhou, Huiqing, Bradley J. Hintze, Isaac J. Kimsey, Bharathwaj Sathyamoorthy, Shan Yang, Jane S. Richardson, and Hashim M. Al-Hashimi: *New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey*. Nucleic Acids Res., 43(7):3420–3433, 2015. Cited on the page 23.

[22] Portugal, José: *Do Hoogsteen base pairs occur in DNA?* Trends Biochem. Sci., 14(4):127–130, 1989. Cited on the page 23.

[23] Nikolova, Evgenia N., Eunae Kim, Abigail A. Wise, Patrick J. O'Brien, Ioan Andricioaei, and Hashim M. Al-Hashimi: *Transient Hoogsteen base pairs in canonical duplex DNA*. Nature, 470:498–502, 2011. Cited on the page 24.

[24] Fu, Xiao-Xiao, Jian-Fu Li, and Rui-Qin Zhang: *Strong Orbital Interaction in pi-pi Stacking System*. arXiv, pages 1–16, 2016. Cited on the page 23.

[25] Peyrard, M.: *Nonlinear dynamics and statistical physics of DNA*. Nonlinearity, 17(2):R1–R40, 2004. Cited on the page 24.

[26] Alhambra, Cristóbal, Francisco J. Luque, Federico Gago, and Modesto Orozco: *Ab initio study of stacking interactions in A- and B-DNA*. J. Phys. Chem. B, 101(19):3846–3853, 1997. Cited on the page 24.

[27] Zacharias, Martin: *Base-pairing and base-stacking contributions to double-stranded DNA formation*. J. Phys. Chem. B, 124(46):10345–10352, 2020. Cited on the page 24.

[28] Kilchherr, Fabian, Christian Wachauf, Benjamin Pelz, Matthias Rief, Martin Zacharias, and Hendrik Dietz: *Single-molecule dissection of stacking forces in DNA*. Science, 353(6304):aaf5508, 2016. Cited on the page 24.

[29] Draper, David E., Dan Grilley, and Ana Maria Soto: *Ions and RNA Folding*. Annu. Rev. Biophys. Biomol. Struct., 34:221–243, 2005. Cited on the page 25.

[30] Draper, David E.: *A guide to ions and RNA structure*. RNA, 10(3):335–343, 2004. Cited on the page 25.

[31] Khvorova, Anastasia and Jonathan K. Watts: *The chemical evolution of oligonucleotide therapies of clinical utility*. Nat. Biotechnol., 35(3):238–248, 2017. Cited on the page 25.

[32] Duffy, Karen, Sebastian Arangundy-Franklin, and Philipp Holliger: *Modified nucleic acids: replication, evolution, and next-generation therapeutics*. BMC Biol., 18:112, 2020. Cited on the page 25.

[33] Yamamoto, Tsuyoshi, Moeka Nakatani, Keisuke Narukawa, and Satoshi Obika: *Antisense drug discovery and development*. Future Med. Chem., 3(3):339–365, 2011. Cited on the page 25.

[34] Shen, Xiulong and David R. Corey: *Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs*. Nucleic Acids Res., 46(4):1584–1600, 2018. Cited on the page 25.

[35] Rupaimoole, Rajesha and Frank J. Slack: *MicroRNA therapeutics: towards a new era for the management of cancer and other diseases*. Nat. Rev. Drug Discov., 16:203–222, 2017. Cited on the page 25.

[36] Burnett, John C. and John J. Rossi: *RNA-based therapeutics: current progress and future prospects*. Chem. Bio., 19(1):60–71, 2012. Cited on the page 25.

[37] Zhou, Jiehua and John Rossi: *Aptamers as targeted therapeutics: current potential and challenges*. Nat. Rev. Drug Discov., 16:181–202, 2016. Cited on the page 25.

[38] Samson, Camille, Pierre Legrand, Mustafa Tekpinar, Jef Rozenski, Mikhail Abramov, Philipp Holliger, Vitor B. Pinheiro, Piet Herdewijn, and Marc Delarue: *Structural studies of HNA substrate specificity in mutants of an archaeal DNA polymerase obtained by directed evolution.* Biomolecules, 10(12):1647, 2020. Cited on the page 25.

[39] Schmidt, Markus: *Xenobiology: A new form of life as the ultimate biosafety tool.* BioEssays, 32(4):322–331, 2010. Cited on the page 25.

[40] Chaput, John C. and Piet Herdewijn: *What is XNA?* Angew. Chem., Int. Ed. Engl., 58(34):11570–11572, 2019. Cited 2 times on the pages 25 and 45.

[41] Lackey, Hershel H., Eric M. Peterson, Zhe Chen, Joel M. Harris, and Jennifer M. Heemstra: *Thermostability trends of TNA:DNA duplexes reveal strong purine dependence.* ACS Synth. Biol., 8(5):1144–1152, 2019. Cited 4 times on the pages 25, 45, 47, and 50.

[42] Morihiro, Kunihiko, Yuuya Kasahara, and Satoshi Obika: *Biological applications of xeno nucleic acids.* Mol. BioSyst., 13(2):235–245, 2016. Cited on the page 25.

[43] Lahiri, Hiya, Sourav Mishra, and Rupa Mukhopadhyay: *Nanoscale nucleic acid recognition at the solid-liquid interface using xeno nucleic acid probes.* Nucleic Acids Res., 35(27):8875–8888, 2019. Cited on the page 25.

[44] Borgaro, Janine G. and Zhenyu Zhu: *Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases.* Nucleic Acids Res., 41(7):4198–4206, 2013. Cited on the page 26.

[45] Nilov, Dmitry, Natalya Maluchenko, Tatyana Kurgina, Sergey Pushkarev, Alexandra Lys, Mikhail Kutuzov, Nadezhda Gerasimova, Alexey Feofanov, Vytas Švedas, Olga Lavrik, and Vasily M. Studitsky: *Molecular mechanisms of PARP-1 inhibitor 7-methylguanine.* Int. J. Mol. Sci., 21(6):2159, 2020. Cited on the page 26.

[46] Dalluge, J. J., T. Hashizume, A. E. Sopchik, J. A. McCloskey, and D. R. Davis: *Conformational flexibility in RNA: The role of dihydrouridine.* Nucleic Acids Res., 24(6):1073–1079, 1996. Cited on the page 26.

[47] Jr, Norman E. Watkins and John SantaLucia Jr: *Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes.* Nucleic Acids Res., 33(19), 2005. Cited on the page 26.

[48] Muniz, Maria Izabel, Hershel H. Lackey, Jennifer M. Heemstra, and Gerald Weber: *DNA/TNA mesoscopic modeling of melting temperatures suggests weaker hydrogen bonding of CG than in DNA/RNA.* Chem. Phys. Lett., 749:137413, 2020. Cited 2 times on the pages 26 and 55.

[49] Johnson, Andrew T., Mark K. Schlegel, Eric Meggers, Lars Oliver Essen, and Olaf Wiest: *On the structure and dynamics of duplex GNA*. J. Org. Chem., 76(19):7964–7974, 2011. Cited on the page 26.

[50] Bhattacharyya, Jhimli, Souvik Maiti, Sanjukta Muhuri, Shu-ichi Nakano, Daisuke Miyoshi, and Naoki Sugimoto: *Effect of locked nucleic acid modifications on the thermal stability of noncanonical DNA structure*. Biochem., 50(34):7414–7425, 2011. Cited on the page 26.

[51] Ebert, Marc Olivier, Christian Mang, Ramanarayanan Krishnamurthy, Albert Eschenmoser, and Bernhard Jaun: *The structure of a TNA-TNA complex in solution: NMR study of the octamer duplex derived from α-(L)-threofuranosyl-(3′-2′)-CGAATTCG*. J. Am. Chem. Soc., 130(45):15105–15115, 2008. Cited on the page 26.

[52] Koshkin, Alexei A., Poul Nielsen, Michael Meldgaard, Vivek K. Rajwanshi, Sanjay K. Singh, and Jesper Wengel: *LNA (Locked Nucleic Acid): An RNA mimic forming exceedingly stable LNA:LNA duplexes*. J. Am. Chem. Soc., 120(50):13252–13253, 1998. Cited on the page 26.

[53] Felsenfeld, G., David R. Davies, and Alexander Rich: *Formation of a three-stranded polynucleotide molecule*. J. Am. Chem. Soc., 79(8):2023–2024, 1957. Cited on the page 26.

[54] Takahashi, Shuntaro and Naoki Sugimoto: *Watson–Crick versus Hoogsteen base pairs: Chemical strategy to encode and express genetic information in life*. Acc. Chem. Res., 54(9):2110–2120, 2021. Cited on the page 26.

[55] Braasch, Dwaine A., Susan Jensen, Yinghui Liu, Kiran Kaur, Khalil Arar, Michael A. White, and David R. Corey: *RNA interference in mammalian cells by chemically-modified RNA*. Biochem., 42(26):7967–7975, 2003. Cited on the page 26.

[56] Jepsen, Jan Stenvang, Mads D. Sørensen, and Jesper Wengel: *Locked nucleic acid: A potent nucleic acid analog in therapeutics and biotechnology*. Oligonucleotides, 14(2):130–146, 2004. Cited on the page 26.

[57] Murayama, Keiji and Hiroyuki Asanuma: *Design and hybridization properties of acyclic xeno nucleic acid oligomers*. ChemBioChem, 22(15):2507–2515, 2021. Cited on the page 26.

[58] Wang, Qian, Xiaoxing Chen, Xintong Li, Dongfan Song, Jintao Yang, Hanyang Yu, and Zhe Li: *2′-fluoroarabinonucleic acid nanostructures as stable carriers for cellular delivery in the strongly acidic environment*. ACS Appl. Mater. Interfaces, 12(48):53592–53597, 2020. Cited on the page 26.

[59] Chakravarthy, Madhuri, Suxiang Chen, Peter R. Dodd, and Rakesh N. Veedu: *Nucleic acid-based theranostics for tackling Alzheimer's disease.* Theranostics, 7(16):3933–3947, 2017. Cited on the page 26.

[60] Chandrasekaran, Arun Richard, Johnsi Mathivanan, Parisa Ebrahimi, Javier Vilcapoma, Alan A. Chen, Ken Halvorsen, and Jia Sheng: *Hybrid DNA/RNA nanostructures with 2′-5′ linkages.* Nanoscale, 12(42):21583–21590, 2020. Cited on the page 26.

[61] Good, Liam and Peter E. Nielsen: *Peptide nucleic acid (PNA) antisense effects in Escherichia coli.* Curr. Issues. Mol. Biol., 1(2):111–116, 1999. Cited on the page 26.

[62] Appella, Daniel H.: *Non-natural nucleic acids for synthetic biology.* Curr. Opin. Chem. Biol., 13(5–6):687–696, 2009. Cited 2 times on the pages 26 and 46.

[63] Dunn, Matthew R., Cailen M. McCloskey, Patricia Buckley, Katherine Rhea, and John C. Chaput: *Generating biologically stable TNA aptamers that function with high affinity and thermal stability.* J. Am. Chem. Soc., 142(17):7721–7724, 2020. Cited 2 times on the pages 26 and 46.

[64] Nance, Kellie D. and Jordan L. Meier: *Modifications in an emergency: The role of N1-Methylpseudouridine in COVID-19 vaccines.* ACS Cent. Sci., 7(5):748–756, 2021. Cited on the page 26.

[65] Traeger, Jeremiah C. and Daniel K. Schwartz: *Interplay of electrostatic repulsion and surface grafting density on surface-mediated DNA hybridization.* J. Colloid Interface Sci., 566:369–374, 2020. Cited on the page 26.

[66] Alexander D. MacKerell, Jr.: *Influence of magnesium ions on duplex DNA structural, dynamic, and solvation properties.* J. Phys. Chem. B, 101(4):646–650, 1997. Cited on the page 26.

[67] Anastassopoulou, Jane: *Metal-DNA interactions.* J. Mol. Struct., 651–653:19–26, 2003. Cited on the page 26.

[68] Chu, Vincent B., Yu Ba, Jan Lipfert, Daniel Herschlag, and Sebastian Doniach: *Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory.* Biophys. J., 93(9):3202–3209, 2007. Cited on the page 26.

[69] Tan, Zhi-Jie and Shi-Jie Chen: *RNA helix stability in mixed Na^+/Mg^{2+} solution.* Biophys. J., 92(10):3615–3632, 2007. Cited on the page 26.

[70] Guéroult, Marc, Olivier Boittin, Oliver Mauffret, Catherine Etchebest, and Brigitte Hartmann: *Mg^{2+} in the major groove modulates B-DNA structure and dynamics.* PLoS ONE, 7(7):e41704, 2012. Cited on the page 26.

[71] Thazhathveetil, Arun Kalliat, Anton Trifonov, Michael R. Wasielewski, and Frederick D. Lewis: *Effect of Mg^{2+} cations on the dynamics and efficiency of hole transport in DNA*. J. Phys. Chem. A, 118(45):10359–10363, 2014. Cited on the page 26.

[72] Kolesnikov, Egor S., Ivan Yu. Gushchin, Petr A. Zhilyaev, and Alexey V. Onufriev: *Similarities and differences between Na^+ and K^+ distributions around DNA obtained with three popular water models*. J. Chem. Theory Comput., 17(11):7246–7259, 2021. Cited 2 times on the pages 26 and 57.

[73] Xu, Hui-Ting, Nan Zhang, Ming-Ru Li, and Feng-Shou Zhang: *Comparison of the ionic effects of Ca^{2+} and Mg^{2+} on nucleic acids in liquids*. J. Mol. Liq., 344:117781, 2021. Cited on the page 26.

[74] Eichhorn, Gunther L. and Yong Ae Shin: *Interaction of metal ions with polynucleotides and related compounds. XII. The relative effect of various metal ions on DNA helicity*. J. Am. Chem. Soc., 90(26):7323–7328, 1968. Cited 2 times on the pages 27 and 57.

[75] Blagoi, Yu P., V. A. Sorokin, V. A. Valeyev, S. A. Khomenko, and G. O. Gladchenko: *Magnesium ion effect on the helix-coil transition of DNA*. Biopoly., 17(5):1103–1118, 1978. Cited 2 times on the pages 27 and 57.

[76] Bleam, M. Louise, Charles F. Anderson, and Jr. M. Thomas Record: *Relative binding affinities of monovalent cations for double-stranded DNA*. Proc. Natl. Acad. Sci. USA, 77(6):3085–3089, 1980. Cited on the page 27.

[77] Burda, Jaroslav V., Jiří Šponer, Jerzy Leszczynski, and Pavel Hobza: *Interaction of DNA base pairs with various metal cations (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Cu^+, Ag^+, Au^+, Zn^{2+}, Cd^{2+}, and Hg^{2+}): Nonempirical ab initio calculations on structures, energies, and nonadditivity of the interaction*. J. Phys. Chem. B, 101(46):9670–9677, 1997. Cited on the page 27.

[78] Korolev, Nikolay, Alexander P. Lyubartsev, Aatto Laaksonen, and Lars Nordenskiöld: *Application of the Poisson Boltzmann polyelectrolyte model for analysis of equilibria between single-, double-, and triple-stranded polynucleotides in the presence of K^+, Na^+, and Mg^{2+} ions*. J. Biomol. Struct. Dyn., 20(2):275–290, 2002. Cited on the page 27.

[79] Andresen, K., R. Das, H. Y. Park, H. Smith, L. W. Kwok, J. S. Lamb, E. J. Kirkland, D. Herschlag, K. D. Finkelstein, and L. Pollack: *Spatial distribution of competing ions around DNA in solution*. Phys. Rev. Lett., 93(24):248103, 2004. Cited on the page 27.

[80] Fingerhut, Benjamin Philipp, Jakob Schauss, Achintya Kundu, and Thomas Elsaesser: *Contact pairs of RNA with magnesium ions-electrostatics beyond the Poisson-Boltzmann equation.* Biophys. J., 120(23):5322–5332, 2021. Cited on the page 27.

[81] Wong, Kingsley L. and Juewen Liu: *Factors and methods to modulate DNA hybridization kinetics.* Biotechnol. J., 16(11):2000338, 2021. Cited on the page 27.

[82] Largy, Eric, Alexander König, Anirban Ghosh, Debasmita Ghosh, Sanae Benabou, Frédéric Rosu, and Valérie Gabelica: *Mass spectrometry of nucleic acid noncovalent complexes.* Chem. Rev., 122(8):7720–7839, 2022. Cited on the page 27.

[83] Cheng, Yuhua, Nikolay Korolev, and Lars Nordenskiöld: *Similarities and differences in interaction of K^+ and Na^+ with condensed ordered DNA. A molecular dynamics computer simulation study.* Nucleic Acids Res., 34(2):686–696, 2006. Cited on the page 27.

[84] Várnai, Péter and Krystyna Zakrzewska: *DNA and its counterions: a molecular dynamics study.* Nucleic Acids Res., 32(14):4269–4280, 2004. Cited on the page 27.

[85] Savelyev, Alexey and Garegin A. Papoian: *Electrostatic, steric, and hydration interactions favor Na^+ condensation around DNA compared with K^+.* J. Am. Chem. Soc., 128(45):14506–14518, 2006. Cited on the page 27.

[86] Langlais, Marc, Heidar All Tajmir-Riahi, and Rodrigue Savoie: *Raman spectroscopic study of the effects of Ca^{2+}, Mg^{2+}, Zn^{2+}, and Cd^{2+} ions on calf thymus DNA: Binding sites and conformational changes.* Biopoly., 30(7–8):743–752, 1990. Cited on the page 27.

[87] Hackl, Elene V., Svetlana V. Kornilova, and Yurij P. Blagoi: *DNA structural transitions induced by divalent metal ions in aqueous solutions.* Int. J. Biol. Macromol., 35(3):175–191, 2005. Cited on the page 27.

[88] Kolev, Stefan K., Petko St. Petkov, Miroslav A. Rangelov, Dimitar V. Trifonov, Teodor I. Milenov, and Georgi N. Vayssilov: *Interaction of Na^+, K^+, Mg^{2+} and Ca^{2+} counter cations with RNA.* Metallomics, 10(5):659–678, 2018. Cited on the page 27.

[89] Solomon, Theodros: *The definition and unit of ionic strength.* J. Chem. Educ., 78(12):1691, 2001. Cited on the page 27.

[90] Lewis, Gilbert N. and Merle Randall: *The activity coefficient of strong electrolytes.* J. Am. Chem. Soc., 43(5):1112–1154, 1921. Cited on the page 27.

[91] Pitzer, Kenneth S.: *Gilbert N. Lewis and the thermodynamics of strong electrolytes.* J. Chem. Educ., 61(2):104, 1984. Cited on the page 27.

[92] Jambrec, Daliborka and Magdalena Gebala: *DNA electrostatics: from theory to application*. ChemElectroChem, 9(4):e202101415, 2021. Cited on the page 27.

[93] Brunet, Annaël, Catherine Tardin, Laurence Salomé, Philippe Rousseau, Nicolas Destainville, and Manoel Manghi: *Dependence of DNA persistence length on ionic strength of solutions with monovalent and divalent salts: A joint theory-experiment study*. Macromolecules, 48(11):3641–3652, 2015. Cited on the page 28.

[94] Wang, Joseph, Emil Palecek, Peter E. Nielsen, Gustavo Rivas, Xiaohua Cai, Haruki Shiraishi, Narasaiah Dontha, Denbai Luo, and Percio A. M. Farias: *Peptide nucleic acid probes for sequence-specific DNA biosensors*. J. Am. Chem. Soc., 118(33):7667–7670, 1996. Cited on the page 28.

[95] Culver, Heidi R., Jasmine Sinha, Tania R. Prieto, Christopher J. Calo, Benjamin D. Fairbanks, and Christopher N. Bowman: *Click nucleic acid–DNA binding behavior: Dependence on length, sequence, and ionic strength*. Biomacromolecules, 21(10):4205–4211, 2020. Cited on the page 28.

[96] Glick, Bernard R., Jack J. Pasternak, and Cheryl L. Patten: *Molecular biotechnology: Principles and applications of recombinant DNA*. ASM Press, 2010. Cited on the page 28.

[97] Reed, Gudrun H., Jana O. Kent, and Carl T. Wittwer: *High-resolution DNA melting analysis for simple and efficient molecular diagnostics*. Pharmacogenomics, 8(6):597–608, 2007. Cited on the page 28.

[98] Vossen, Rolf H. A. M., Emmelien Aten, Anja Roos, and Johan T. den Dunnen: *High-resolution melting analysis (HRMA) – More than just sequence variant screening*. Hum. Mutat., 30(6):860–866, 2009. Cited on the page 28.

[99] Montgomery, Jesse L., Lindsay N. Sanford, and Carl T. Wittwer: *High-resolution DNA melting analysis in clinical research and diagnostics*. Expert Rev. Mol. Diagn. Lett., 10(2):219–240, 2010. Cited on the page 28.

[100] Bester, Rachelle, Anna E. C. Jooste, Hans J. Maree, and Johan T. Burger: *Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated virus 3 variant groups I, II, III and VI*. Virol. J., 9(219):1–11, 2012. Cited on the page 28.

[101] Toi, Cheryl S. and Dominic E. Dwyer: *Differentiation between vaccine and wild-type varicella-zoster virus genotypes by high-resolution melt analysis of single nucleotide polymorphisms*. J. Clin. Virol., 43(1):18–24, 2008. Cited on the page 28.

[102] Gori, Andrea, Matteo Cerboneschi, and Stefania Tegli: *High-resolution melting analysis as a powerful tool to discriminate and genotype pseudomonas savastanoi pathovars and strains.* PLoS ONE, 7(1):1–13, 2012. Cited on the page 28.

[103] Allison, Lizabeth Ann: *Fundamental Molecular Biology.* Blackwell Publishing Limited, 2007. Cited 2 times on the pages 28 and 29.

[104] Schreiber-Gosche, Sherrie and Robert A. Edwards: *Thermodynamics of oligonucleotide duplex melting.* J. Chem. Educ., 86(5):644–650, 2009. Cited 3 times on the pages 29, 30, and 31.

[105] Blake, R. D. and Scott G. Delcourt: *Thermodynamic effects of formamide on DNA stability.* Nucleic Acids Res., 24(11):2095–2103, 1996. Cited on the page 29.

[106] Hammouda, Boualem and David Worcester: *The denaturation transition of DNA in mixed solvents.* Biophys. J., 91(6):2237–2242, 2006. Cited on the page 29.

[107] Priyakumar, U. Deva, Changbong Hyeon, D. Thirumalai, and Alexander D. MacKerell Jr.: *Urea destabilizes RNA by forming stacking interactions and multiple hydrogen bonds with nucleic acid bases.* J. Am. Chem. Soc., 131(49):17759–17761, 2009. Cited on the page 29.

[108] Zaha, Arnaldo, Henrique Bunselmeyer, and Luciane M. P. Passaglia: *Biologia Molecular Básica.* Artmed Editora Ltda., 2014. Cited on the page 29.

[109] Niranjani, Gnanapragasam and Rajamanickam Murugan: *Theory on the mechanism of DNA renaturation: Stochastic nucleation and zipping.* PLoS ONE, 11(4):1–28, 2016. Cited on the page 29.

[110] Howard, Kathleen P.: *Thermodynamics of DNA duplex formation: A biophysical chemistry laboratory experiment.* J. Chem. Educ., 77(11):1469–1471, 2000. Cited 2 times on the pages 29 and 31.

[111] Zhang, Yong-li, 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. Cited 4 times on the pages 33, 37, 38, and 39.

[112] Weber, Gerald, Niall Haslam, Nava Whiteford, Adam Prügel-Bennett, Jonathan W. Essex, and Cameron Neylon: *Thermal equivalence of DNA duplexes without melting temperature calculation.* Nat. Phys., 2:55–59, 2006. Cited 6 times on the pages 33, 39, 40, 41, 42, and 62.

[113] Jr, Ignacio Tinoco, Olke C. Uhlenbeck, and Mark D. Levine: *Estimation of secondary structure in ribonucleic acids.* Nature, 230:362–367, 1971. Cited on the page 33.

[114] Huguet, Josep Maria, Marco Ribezzi-Crivellari, Cristiano Valim Bizarro, and Felix Ritort: *Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments.* Nucleic Acids Res., 45(22):12921–12931, 2017. Cited on the page 33.

[115] Borer, Philip N., Barbara Dengler, Ignacio Tinoco Jr, and Olke C. Uhlenbeck: *Stability of ribonucleic acid double-stranded helices.* J. Mol. Biol., 86(4):843–853, 1974. Cited on the page 33.

[116] SantaLucia, John, Hatim T. Allawi, and P. Ananda Seneviratne: *Improved nearest-neighbor parameters for predicting DNA duplex stability.* Biochem., 35(11):3555–3562, 1996. Cited on the page 34.

[117] Xia, Tianbing, Jr. John SantaLucia, Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoqi Jiao, Christopher Cox, and Douglas H. Turner: *Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs.* Biochem., 37(42):14719–14735, 1998. Cited on the page 34.

[118] Oliveira Martins, Erik de: *Desenvolvimento de modelos mesoscópicos assimétricos para oligonucleotídeos e sua aplicação a defeitos tipo single-bulges em RNAs.* PhD thesis, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, 2018. `https://repositorio.ufmg.br/handle/1843/SMRA-BBLJN7`. Cited on the page 34.

[119] Poland, Douglas C. and Harold A. Scheraga: *Comparison of theories of the helix-coil transition in polypeptides.* J. Chem. Phys., 43(6):2071, 1965. Cited on the page 34.

[120] Baiesi, Marco and Enrico Carlon: *Models of DNA denaturation dynamics: universal properties.* arXiv, pages 1–10, 2018. Cited on the page 34.

[121] Richard, C. and A. J. Guttmann: *Poland-Scheraga models and the DNA denaturation transition.* J. Stat. Phys., 115(314):925–947, 2004. Cited on the page 34.

[122] Peyrard, M. and A. R. Bishop: *Statistical mechanics of a nonlinear model for DNA denaturation.* Phys. Rev. Lett., 62(23):2755–2758, 1989. Cited 3 times on the pages 35, 37, and 51.

[123] Moura, Pâmella Miranda de: *Desenvolvimento de uma representação mesoscópica para marcadores fluorescentes em DNA e predição de intensidade de fluorescência.* Master's thesis, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, 2020. `https://repositorio.ufmg.br/handle/1843/32788`. Cited 2 times on the pages 35 and 38.

[124] Maximiano, Rodolfo Vieira: *Estudos das propriedades da inosina em DNA através do modelo Peyrard-Bishop e análise dos parâmetros termodinâmicos utilizados na predição de estruturas secundárias de RNA*. PhD thesis, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, 2017. `https://repositorio.ufmg.br/handle/1843/SMRA-BBTQR7`. Cited on the page 36.

[125] Dauxois, Thierry and Michel Peyrard: *Entropy-driven transition in a one-dimensional system*. Phys. Rev. E, 51(5):4027–4040, 1995. Cited on the page 37.

[126] Weber, Gerald: *Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures*. Nucleic Acids Res., 41(1):e30, 2013. Cited 3 times on the pages 37, 43, and 48.

[127] Dauxois, Thierry, M. Peyrard, and A. R. Bishop: *Entropy-driven DNA denaturation*. Phys. Rev. E, 47(1):R44–R47, 1993. Cited on the page 37.

[128] Weber, Gerald, Niall Haslam, Jonathan W. Essex, and Cameron Neylon: *Thermal equivalence of DNA duplexes for probe design*. J. Phys.: Condens. Matter, 21(3):034106, 2009. Cited 4 times on the pages 39, 40, 41, and 51.

[129] Silva, Izabela Ferreira da: *Mesoscopic models for RNA salt dependence and for oncogene probe design with locked nucleic acids*. PhD thesis, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, 2021. `https://repositorio.ufmg.br/handle/1843/38260`. Cited 3 times on the pages 41, 51, and 54.

[130] Weber, Gerald: *Optimization method for obtaining nearest-neighbour DNA entropies and enthalpies directly from melting temperatures*. Bioinformatics, 31(6):871–877, 2015. Cited 2 times on the pages 43 and 49.

[131] Wikipedia: *Nelder-Mead method*, accessed in July 2020. `https://en.wikipedia.org/wiki/Nelder-Mead_method`. Cited on the page 43.

[132] Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery: *Numerical Recipes in C*. Cambridge University Press, 1992. Cited on the page 43.

[133] Swart, Marcel, Célia Fonseca Guerra, and F. Matthias Bickelhaupt: *Hydrogen bonds of RNA are stronger than those of DNA, but NMR monitors only presence of methyl substituent in uracil/thymine*. J. Am. Chem. Soc., 126(51):16718–16719, 2004. Cited on the page 43.

[134] Manghi, Manoel and Nicolas Destainville: *Physics of base-pairing dynamics in DNA*. Phys. Rep., 631:1–41, 2016. Cited on the page 43.

[135] Wikipedia: *Xeno nucleic acid*, accessed in July 2020. `https://en.wikipedia.org/wiki/Xeno_nucleic_acid`. Cited on the page 45.

[136] Braasch, Dwaine A. and David R. Corey: *Locked nucleic acid (LNA): Fine-tuning the recognition of DNA and RNA*. Chemistry & Biology, 8(1):1–7, 2001. Cited on the page 45.

[137] Chaput, John C. and Jack W. Szostak: *TNA synthesis by DNA polymerases*. J. Am. Chem. Soc., 125(31):9274–9275, 2003. Cited on the page 45.

[138] Culbertson, Michelle C., Kartik W. Temburnikar, Sujay P. Sau, Jen Yu Liao, Saikat Bala, and John C. Chaput: *Evaluating TNA stability under simulated physiological conditions*. Bioorg. Med. Chem. Lett., 26(10):2418–2421, 2016. Cited on the page 45.

[139] Sun, Zhaoxi and John Z. H. Zhang: *Thermodynamic insights of base flipping in TNA duplex: Force fields, salt concentrations, and free-energy simulation methods*. CCS Chem., 3(2):1026–1039, 2020. Cited 2 times on the pages 45 and 46.

[140] Schöning, K. -U., P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, and A. Eschenmoser: *Chemical etiology of nucleic acid structure: The α-threofuranosyl-($3' \rightarrow 2'$) oligonucleotide system*. Science, 290(5495):1347–1351, 2000. Cited 3 times on the pages 45, 47, and 48.

[141] Chim, Nicholas, Changhua Shi, Sujay P. Sau, Ali Nikoomanzar, and John C. Chaput: *Structural basis for TNA synthesis by an engineered TNA polymerase*. Nature Comm., 8(1810):1–11, 2017. Cited 2 times on the pages 45 and 46.

[142] Wikipedia: *Moiety (chemistry)*, accessed in November 2022. `https://en.wikipedia.org/wiki/Moiety_(chemistry)`. Cited on the page 45.

[143] Dunn, Matthew R., Randi M. Jimenez, and John C. Chaput: *Analysis of aptamer discovery and technology*. Nat. Rev. Chem., 1(76):1–16, 2017. Cited 2 times on the pages 45 and 46.

[144] Pallan, Pradeep S., Christopher J. Wilds, Zdzislaw Wawrzak, Ramanarayanan Krishnamurthy, Albert Eschenmoser, and Martin Egli: *Why does TNA cross-pair more strongly with RNA than with DNA? An answer from x-ray analysis*. Angew. Chem., Int. Ed. Engl., 42(47):5893–5895, 2003. Cited 2 times on the pages 46 and 50.

[145] Anosova, Irina, Ewa A. Kowal, Nicholas J. Sisco, Sujay Sau, Jen-yu Liao, Saikat Bala, Eriks Rozners, Martin Egli, John C. Chaput, and Wade D. Van Horn: *Structural insights into conformation differences between DNA/TNA and RNA/TNA chimeric duplexes*. ChemBioChem, 17(18):1705–1708, 2016. Cited 3 times on the pages 46, 54, and 55.

[146] Salazar, Miguel, Oleg Y. Fedoroff, Julie M. Miller, N. Susan Ribeiro, and Brian R. Reid: *The DNA strand in DNA·RNA hybrid duplexes is neither B-form nor A-form in solution.* Biochem., 32(16):4207–4215, 1993. Cited on the page 46.

[147] Noy, Agnes, Alberto Pérez, Manuel Márquez, F. Javier Luque, and Modesto Orozco: *Structure, recognition properties, and flexibility of the DNA·RNA hybrid.* J. Am. Chem. Soc., 127(13):4910–4920, 2005. Cited on the page 46.

[148] Oliveira Martins, Erik de, Vivianne Basílio Barbosa, and Gerald Weber: *DNA/RNA hybrid mesoscopic model shows strong stability dependence with deoxypyrimidine content and stacking interactions similar to RNA/RNA.* Chem. Phys. Lett., 715:14–19, 2019. Cited 5 times on the pages 46, 50, 51, 52, and 54.

[149] Schöning, Kai-Uwe, Peter Scholz, Xiaolin Wu, Sreenivasulu Guntha, Guillermo Delgado, Ramanarayanan Krishnamurthy, and Albert Eschenmoser: *The α-L-threofuranosyl-(3′ → 2′)-oligonucleotide system ('TNA'): Synthesis and pairing properties.* Helv. Chim. Acta, 85(12):4111–4153, 2002. Cited on the page 46.

[150] Herdewijn, Piet: *TNA as a potential alternative to natural nucleic acids.* Angew. Chem., Int. Ed. Engl., 40(12):2249–2251, 2001. Cited on the page 46.

[151] Yang, Ying-Wei, Su Zhang, Elizabeth O. McCullum, and John C. Chaput: *Experimental evidence that GNA and TNA were not sequential polymers in the prebiotic evolution of RNA.* J. Mol. Evol., 65:289–295, 2007. Cited 2 times on the pages 46 and 48.

[152] Wang, Yao, Yueyao Wang, Dongfan Song, Xin Sun, Ze Zhang, Xintong Li, Zhe Li, and Hanyang Yu: *A threose nucleic acid enzyme with RNA ligase activity.* J. Am. Chem. Soc., 143(21):8154–8163, 2021. Cited on the page 46.

[153] Wilds, Christopher J., Zdzislaw Wawrzak, Ramanarayanan Krishnamurthy, Albert Eschenmoser, and Martin Egli: *Crystal structure of a B-form DNA duplex containing (L)-α-threofuranosyl (3′ →2′) nucleosides: A four-carbon sugar is easily accommodated into the backbone of DNA.* J. Am. Chem. Soc., 124(46):13716–13721, 2002. Cited on the page 46.

[154] Horhota, Allen, Keyong Zou, Justin K. Ichida, Biao Yu, Larry W. McLaughlin, Jack W. Szostak, and John C. Chaput: *Kinetic analysis of an efficient DNA-dependent TNA polymerase.* J. Am. Chem. Soc., 127(20):7427–7434, 2005. Cited on the page 46.

[155] Dunn, Matthew R., Andrew C. Larsen, Walter J. Zahurancik, Nour Eddine Fahmi, Madeline Meyers, Zucai Suo, and John C. Chaput: *DNA polymerase-mediated synthesis of unbiased threose nucleic acid (TNA) polymers requires 7-deazaguanine to*

suppress G:G mispairing during TNA transcription. J. Am. Chem. Soc., 137(12):4014–4017, 2015. Cited on the page 46.

[156] Houlihan, Gillian, Sebastian Arangundy-Franklin, Benjamin T. Porebski, Nithya Subramanian, Alexander I. Taylor, and Philipp Holliger: *Discovery and evolution of RNA and XNA reverse transcriptase function and fidelity.* Nat. Chem., 12:683–690, 2020. Cited on the page 46.

[157] Yu, Hanyang, Su Zhang, Matthew R. Dunn, and John C. Chaput: *An efficient and faithful in vitro replication system for threose nucleic acid.* J. Am. Chem. Soc., 135(9):3583–3591, 2013. Cited on the page 46.

[158] Dunn, Matthew R. and John C. Chaput: *Reverse transcription of threose nucleic acid by a naturally occurring DNA polymerase.* ChemBioChem, 17(19):1804–1808, 2016. Cited on the page 46.

[159] Li, Xintong, Zhe Li, and Hanyang Yu: *Selection of threose nucleic acid aptamers to block PD-1/PD-L1 interaction for cancer immunotherapy.* Chem. Commun., 56(93):14653–14656, 2020. Cited on the page 46.

[160] Wang, Fei, Ling Sum Liu, Cia Hin Lau, Tristan Juin Han Chang, Dick Yan Tam, Hoi Man Leung, Chung Tin, and Pik Kwan Lo: *Synthetic α-L-threose nucleic acids targeting BcL-2 show gene silencing and in vivo antitumor activity for cancer therapy.* ACS Appl. Mater. Interfaces, 11(42):38510–38518, 2019. Cited on the page 46.

[161] McCloskey, Cailen M., Jen-Yu Liao, Saikat Bala, and John C. Chaput: *Ligase-mediated threose nucleic acid synthesis on DNA templates.* ACS Synth. Biol., 8(2):282–286, 2019. Cited on the page 46.

[162] Weber, Gerald, Jonathan W. Essex, and Cameron Neylon: *Probing the microscopic flexibility of DNA from melting temperatures.* Nat. Phys., 5:769–773, 2009. Cited 5 times on the pages 47, 48, 50, 61, and 63.

[163] Zhang, Tian-biao, Chang-lin Zhang, Zai-li Dong, and Yi-fu Guan: *Determination of base binding strength and base stacking interaction of DNA duplex using atomic force microscope.* Sci. Rep., 5:9143, 2015. Cited on the page 50.

[164] Yu, Hanyang, Su Zhang, and John C. Chaput: *Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor.* Nat. Chem., 4:183–187, 2012. Cited on the page 51.

[165] Lackey, Hershel H., Zhe Chen, Joel M. Harris, Eric M. Peterson, and Jennifer M. Heemstra: *Single-molecule kinetics show DNA pyrimidine content strongly affects RNA:DNA and TNA:DNA heteroduplex dissociation rates.* ACS Synth. Biol., 9(2):249–253, 2020. Cited 2 times on the pages 52 and 53.

[166] Suresh, Gorle and U. Deva Priyakumar: *DNA-RNA hybrid duplexes with decreasing pyrimidine content in the DNA strand provide structural snapshots for the A- to B-form conformational transition of nucleic acids.* Phys. Chem. Chem. Phys., 16(34):18148–18155, 2014. Cited 3 times on the pages 52, 53, and 54.

[167] Hamperl, Stephan and Karlene A. Cimprich: *The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability.* DNA Repair, 19:84–94, 2014. Cited on the page 52.

[168] Wang, Fei, Ling Sum Liu, Pan Li, Hoi Man Leung, Dick Yan Tam, and Pik Kwan Lo: *Biologically stable threose nucleic acid-based probes for real-time microRNA detection and imaging in living cells.* Mol. Ther.–Nucleic Acids, 27:787–796, 2022. Cited 2 times on the pages 54 and 55.

[169] Hillebrand, M., G. Kalosakas, Ch. Skokos, and A. R. Bishop: *Distributions of bubble lifetimes and bubble lengths in DNA.* Phys. Rev. E, 102:062114, 2020. Cited on the page 55.

[170] Su, Shenyang, Guotao Sun, Xiuning Liang, and Hua Fang: *Effectively controlling the ESIPT behavior and fluorescence feature of 2-(2′-hydroxyphenyl)-4-chloromethylthiazole by changing its π-conjugation: A theoretical exploration.* J. Photochem. Photobiol. A, 422:113548, 2022. Cited on the page 55.

[171] Wang, Fei, Ling Sum Liu, Pan Li, Cia Hin Lau, Hoi Man Leung, Y Rebecca Chin, Chung Tin, and Pik Kwan Lo: *Cellular uptake, tissue penetration, biodistribution, and biosafety of threose nucleic acids: Assessing in vitro and in vivo delivery.* Mater. Today Bio, 15:100299, 2022. Cited on the page 55.

[172] Wang, Fei, Pan Li, Hoi Man Leung, and Pik Kwan Lo: *Nucleic acids and their analogues for biomedical applications.* Biosensors, 12(2):93, 2022. Cited on the page 55.

[173] Wang, Xiaoming, Ghazala Akram, Maasoomah Sadaf, Hajra Mariyam, and Muhammad Abbas: *Soliton solution of the Peyrard-Bishop-Dauxois model of DNA dynamics with M-truncated and β-fractional derivatives using Kudryashov's R function method.* Fractal Fract., 6(10):616, 2022. Cited on the page 55.

[174] Serec, Kristina, Sanja Dolanski Babić, and Silvia Tomić: *Magnesium ions reversibly bind to DNA double stranded helix in thin films.* Spectrochim. Acta, Part A, 268:120663, 2022. Cited 2 times on the pages 56 and 57.

[175] Sorokin, V. A., V. A. Valeev, E. L. Usenko, and V. V. Andrushchenko: *Divalent metal ion effect on helix-coil transition of high molecular weight DNA in neutral and*

alkaline solutions. Int. J. Biol. Macromol., 48(2):369–374, 2011. Cited on the page 56.

[176] Muntean, C. M., G. J. Puppels, J. Greve, G. M. J. Segers-Nolten, and S. Cinta-Pinzaru: *Raman microspectroscopic study on low-pH-induced DNA structural transitions in the presence of magnesium ions.* J. Raman Spec., 33(10):784–788, 2002. Cited on the page 56.

[177] Fu, H., H. Chen, C. G. Koh, and C. T. Lim: *Effects of magnesium salt concentrations on B-DNA overstretching transition.* Eur. Phys. J. E, 29:45–49, 2009. Cited on the page 56.

[178] Bizarro, C. V., A. Alemany, and F. Ritort: *Non-specific binding of Na^+ and Mg^{2+} to RNA determined by force spectroscopy methods.* Nucleic Acids Res., 40(14):6922–6935, 2012. Cited on the page 56.

[179] Korolev, Nikolay, Alexander P. Lyubartsev, Allan Rupprecht, and Lars Nordenskiöld: *Competitive binding of Mg^{2+}, Ca^{2+}, Na^+, and K^+ ions to DNA in oriented DNA fibers: Experimental and Monte Carlo simulation results.* Biophys. J., 77(5):2736–2749, 1999. Cited on the page 56.

[180] Korolev, Nikolay, Alexander P. Lyubartsev, Aatto Laaksonen, and Lars Nordenskiöld: *On the competition between water, sodium ions, and spermine in binding to DNA: A molecular dynamics computer simulation study.* Biophys. J., 82(6):2860–2875, 2002. Cited on the page 56.

[181] Xue, Jingjing, Peng Wang, Xinpeng Li, Rongri Tan, and Wenjun Zong: *Transformation characteristics of A-DNA in salt solution revealed through molecular dynamics simulations.* Biophys. Chem., 288:106845, 2022. Cited on the page 56.

[182] Chen, Shu-wen W. and Barry Honig: *Monovalent and divalent salt effects on electrostatic free energies defined by the nonlinear Poisson-Boltzmann equation: Application to DNA binding reactions.* J. Phys. Chem. B, 101(44):9113–9118, 1997. Cited on the page 56.

[183] Gebala, Magdalena and Daniel Herschlag: *Quantitative studies of an RNA duplex electrostatics by ion counting.* Biophys. J., 117(6):1116–1124, 2019. Cited 2 times on the pages 56 and 57.

[184] Xi, Kun, , 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. Cited 2 times on the pages 56 and 57.

[185] Li, Weifeng, Lars Nordenskiöld, and Yuguang Mu: *Sequence-specific Mg^{2+}-DNA interactions: A molecular dynamics simulation study.* J. Phys. Chem. B, 115(49):14713–14720, 2011. Cited 2 times on the pages 56 and 63.

[186] Mukherjee, Sanchita and Dhananjay Bhattacharyya: *Influence of divalent magnesium ion on DNA: molecular dynamics simulation studies.* J. Biomol. Struct. Dyn., 31(8):896–912, 2013. Cited 4 times on the pages 56, 57, 69, and 70.

[187] Lavery, Richard, John H. Maddocks, Marco Pasi, and Krystyna Zakrzewska: *Analyzing ion distributions around DNA.* Nucleic Acids Res., 42(12):8138–8149, 2014. Cited on the page 57.

[188] Misra, Vinod K. and David E. Draper: *The interpretation of Mg^{2+} binding isotherms for nucleic acids using Poisson-Boltzmann theory.* J. Mol. Biol., 294(5):1135–1147, 1999. Cited on the page 57.

[189] Ferreira, Izabela, 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(17):175101, 2015. Cited 18 times on the pages 57, 58, 59, 60, 61, 62, 63, 64, 66, 67, 68, 69, 84, 85, 86, 87, 88, and 89.

[190] Zgarbová, Marie, Michal Otyepka, Jiří Šponer, Filip Lankaš, and Petr Jurečka: *Base pair fraying in molecular dynamics simulations of DNA and RNA.* J. Chem. Theory Comput., 10(8):3177–3189, 2014. Cited on the page 57.

[191] Jin, Fei, Yichen Huang, and Motoyuki Hattori: *Recent advances in the structural biology of Mg^{2+} channels and transporters.* J. Mol. Biol., 434(19):167729, 2022. Cited on the page 57.

[192] Markoulatos, P., N. Siafakas, and M. Moncany: *Multiplex polymerase chain reaction: A practical approach.* J. Clin. Lab. Anal., 16(1):47–51, 2002. Cited on the page 57.

[193] Li, Feng-Yen, Benjamin Chaigne-Delalande, Chrysi Kanellopoulou, Jeremiah C. Davis, Helen F. Matthews, Daniel C. Douek, Jeffrey I. Cohen, Gulbu Uzel, Helen C. Su, and Michael J. Lenardo: *Second messenger role for Mg^{2+} revealed by human T-cell immunodeficiency.* Nature, 475:471–476, 2011. Cited on the page 57.

[194] Thomas, René: *Recherches sur la dénaturation des acides desoxyribonucléiques.* Biochim. Biophys. Acta,, 14:231–240, 1954. Cited on the page 57.

[195] Lyons, John W. and Leonard Kotin: *The effects of magnesium ion on the secondary structure of deoxyribonucleic acid.* J. Am. Chem. Soc., 87(8):1781–1785, 1965. Cited on the page 57.

[196] Baba, Yoshihiro and Akihiro Kagemoto: *Influence of magnesium ions on helix-coil transition of DNA determined by modified differential scanning calorimeter.* Biopoly., 13:339–344, 1974. Cited on the page 57.

[197] Sissöeff, I., J. Grisvard, and E. Guillé: *Studies on metal ions-DNA interactions: Specific behaviour of reiterative DNA sequences.* Prog. Biophys. Mol. Biol., 31(2):165–199, 1976. Cited on the page 57.

[198] Tan, Zhi-Jie and Shi-Jie Chen: *Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length.* Biophys. J., 90(4):1175–1190, 2006. Cited on the page 57.

[199] Dove, William F. and Norman Davidson: *Cation effects on the denaturation of DNA.* J. Math. Biol., 5(5):467–478, 1962. Cited on the page 57.

[200] M. Thomas Record, Jr.: *Effects of Na^+ and Mg^{++} ions on the helix-coil transition of DNA.* Biopoly., 14(10):2137–2158, 1975. Cited on the page 57.

[201] Serec, Kristina, Sanja Dolanski Babić, Rudolf Podgornik, and Silvia Tomić: *Effect of magnesium ions on the structure of DNA thin films: an infrared spectroscopy study.* Nucleic Acids Res., 44(17):8456–8464, 2016. Cited 2 times on the pages 57 and 65.

[202] Every, Alicia E. and Irina M. Russu: *Influence of magnesium ions on spontaneous opening of DNA base pairs.* J. Phys. Chem. B, 112(25):7689–7695, 2008. Cited 2 times on the pages 57 and 68.

[203] Bai, Yu, Max Greenfeld, Kevin J. Travers, Vincent B. Chu, Jan Lipfert, Sebastian Doniach, and Daniel Herschlag: *Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids.* J. Am. Chem. Soc., 129(48):14981–14988, 2007. Cited on the page 57.

[204] Ahsen, Nicolas von, Carl T. Wittwer, and Ekkehard Schütz: *Oligonucleotide melting temperatures under PCR conditions: Nearest-neighbor corrections for Mg^{2+}, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas.* Clin. Chem., 47(11):1956–1961, 2001. Cited 2 times on the pages 58 and 59.

[205] Owczarzy, Richard, Bernardo G. Moreira, Yong You, Mark A. Behlke, and Joseph A. Walder: *Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations.* Biochem., 47(19):5336–5353, 2008. Cited 3 times on the pages 58, 59, and 61.

[206] Owczarzy, Richard, Bernardo G. Moreira, Yong You, Mark A. Behlke, and Joseph A. Walder: *Method for estimating a melting temperature of a nucleic acid in buffers*

containing magneisum ions, 2012. US Patent 20,120,123,751. Cited 5 times on the pages 58, 59, 61, 77, and 79.

[207] Mitsuhashi, Masato: *Technical report: Part 1. Basic requirements for designing optimal oligonucleotide probe sequences.* J. Clin. Lab. Anal., 10(5):277–284, 1996. Cited on the page 59.

[208] Breslauer, K. J., 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. Cited on the page 59.

[209] Freeman, Gordon S., Daniel M. Hinckley, and Juan J. de Pablo: *A coarse-grain three-site-per-nucleotide model for DNA with explicit ions.* J. Chem. Phys., 135(16):165104, 2011. Cited on the page 61.

[210] Ferreira, Izabela, Tauanne D. Amarante, and Gerald Weber: *Salt dependent mesoscopic model for RNA with multiple strand concentrations.* Biophys. Chem., 271:106551, 2021. Cited on the page 63.

[211] Oliveira, Luciana M., 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. Cited on the page 63.

[212] Guerra, Célia Fonseca, 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. Cited on the page 63.

[213] Chiu, Thang Kien and Richard E. Dickerson: *1Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium.* J. Math. Biol., 301(4):915–945, 2000. Cited 2 times on the pages 69 and 70.

[214] Nakata, Michi, Giuliano Zanchetta, Brandon D. Chapman, Christopher D. Jones, Julie O. Cross, Ronald Pindak, Tommaso Bellini, and Noel A. Clark: *End-to-end stacking and liquid crystal condensation of 6–to 20–base pair DNA duplexes.* Science, 318(5854):1276–1279, 2007. Cited on the page 71.

[215] Muniz, Maria Izabel, Adrian H. Bustos, Sofie Slott, Kira Astakhova, and Gerald Weber: *Cation valence dependence of hydrogen bond and stacking potentials in DNA mesoscopic models.* Accepted for publication in Biophysical Chemistry in December 2022. Cited on the page 71.

[216] Liu, Ling Sum, Hoi Man Leung, Dick Yan Tam, Tsz Wan Lo, Sze Wing Wong, and Pik Kwan Lo: *$\alpha-$L-threose nucleic acids as biocompatible antisense oligonucleotides for suppressing gene expression in living cells.* ACS Appl. Mater. Interfaces, 10(11):9736–9743, 2018. Cited on the page 72.

[217] Mitchell, Brandon P., Rohaine V. Hsu, Marco A. Medrano, Nehemiah T. Zewde, Yogesh B. Narkhede, and Giulia Palermo: *Spontaneous embedding of DNA mismatches within the RNA:DNA hybrid of CRISPR-Cas9*. Front. Mol. Biosci., 7(39):1–9, 2020. Cited on the page 73.

[218] Kim, Seohyun, Sangmin Ji, and Hye Ran Koh: *CRISPR as a diagnostic tool*. Biomolecules, 11(8):1162, 2021. Cited on the page 73.