

PREPRINT

# Salt corrections for RNA secondary structures in the ViennaRNA package

Hua-Ting Yao<sup>1\*</sup>, Ronny Lorenz<sup>1</sup>, Ivo L. Hofacker<sup>1,2</sup> and Peter F. Stadler<sup>3,4,5,1,6,7\*</sup>

\*Correspondence:

htyao@tbi.univie.ac.at;  
studla@bioinf.uni-leipzig.de

<sup>1</sup>Department of Theoretical  
Chemistry, University of Vienna  
Währinger Straße 17, A-1090  
Vienna, Austria

<sup>3</sup>Bioinformatics Group,  
Department of Computer Science,  
and Interdisciplinary Center for  
Bioinformatics, Universität  
Leipzig, Härtelstraße 16–18,  
D-04107, Leipzig, Germany  
Full list of author information is  
available at the end of the article

## Abstract

**Background:** RNA features a highly negatively charged phosphate backbone that attracts a cloud of counter-ions that reduce the electrostatic repulsion in a concentration dependent manner. Ion concentrations thus have a large influence on folding and stability of RNA structures. Despite their well-documented effects, salt effects are not handled by currently available secondary structure prediction algorithms. Combining Debye-Hückel potentials for line charges and Manning's counter-ion condensation theory, Einert *et al.* [*Biophys. J.* **100**: 2745-2753 (2011)] modeled the energetic effects of monovalent cations on loops and helices.

**Results:** The model of Einert *et al.* is adapted to match the structure of the dynamic programming recursion of RNA secondary structure prediction algorithms. An empirical term describing the dependence of the duplex initiation energy on salt concentration is added to improve co-folding predictions for two or more RNA strands. The slightly modified model is implemented in the ViennaRNA package in such way that only the energy parameters but not the algorithmic structure is affected. A comparison with data from the literature shows that predicted free energies and melting temperatures are in reasonable agreement with experiments.

**Conclusion:** The new feature in the ViennaRNA package makes it possible to study effects of salt concentrations on RNA folding in a systematic manner. Strictly speaking, the model pertains only to mono-valent cations, and thus covers the most important parameter, i.e., the NaCl concentration. It remains a question for future research to what extent unspecific effects of bi- and tri-valent cations can be approximated in a similar manner.

**Availability:** Corrections for the concentration of monovalent cations are available in the ViennaRNA package starting from version 2.6.0.

**Keywords:** RNA Secondary Structure; Salt concentration; Debye-Hückel potential

## Introduction

Nucleic acids are highly negatively charged molecules since their phosphate backbone carries one negative charge per residue. Structure formation brings these charges into close proximity and thus incurs substantial electrostatic penalties on the secondary and tertiary structure level. Nucleic acid-ion interactions also provide large interaction energies and therefore contribute decisively to the folding of RNA and DNA and to their interactions with ligands and macromolecular partners [1]. Counter-ions reduce the electrostatic repulsion of the backbone. Cation concentrations determine the extent of this “charge screening” and thus strongly

influence RNA folding. Indeed, many functional RNAs will not fold under low salt conditions [2], and experimental investigations of the thermodynamics of RNA folding are mostly confined to high salt conditions. Energy parameters for RNA secondary predictions likewise pertain to 1M sodium concentrations, more precisely 1.021M while taking all  $\text{Na}^+$  ions in the buffer into account [3]. Although the importance of counter-ions for the RNA folding is well known, ion concentration, in contrast to temperature, is not a tunable parameter in currently available RNA secondary structure prediction tools. While temperature dependence is conceptually straightforward and can be easily modeled by splitting free energy contributions into enthalpic and entropic contributions [4], the energetics of ion–nucleic acid interactions are much more difficult to understand.

Cations affect RNA structure in two different ways. The electrostatic stabilization of the structure due to charge screening is at least conceptually independent of the chemical nature and charge of the cation. In addition, divalent cations and in particular  $\text{Mg}^{2+}$  can also strongly bind to specific, chelating sites [5, 6]. Quantitative salt effects on RNA folding have been studied extensively over the last decades, see [7] for a study that summarized much of the pertinent earlier literature. In the absence of a well-founded theoretical model, most authors resorted to describing the salt-dependence of RNA folding by means of simple heuristic function fitting the effects of changes in the sodium concentration on the free energy of folding or a melting temperature. Such empirical fits, however, are limited to handling salt effects close to standard conditions, and an approach that explains the functional form of salt effects is clearly preferable.

If three-dimensional structures are known, the nonlinear Poisson-Boltzmann equation can be solved to obtain electrostatic potentials of RNA molecules in solution [8–10]. This approach, however, appears to be too detailed to derive a practically manageable parametrization of salt effects at the level of secondary structure prediction algorithms. In order to handle counter-ions in RNA secondary structure prediction algorithms, the effects must be attributed to individual bases, base pairs, or loops (including the stacking of two consecutive base pairs). This is necessary because secondary structure prediction algorithms operate on these combinatorial substructures [11]. This, in particular, precludes models that explicitly require a detailed geometric description of three-dimensional structure of an RNA.

To derive a suitably simple model, Einert and Netz [12] proposed to represent the RNA backbone as a charged polymer that interacts by means of a Debye-Hückel potential [13] and treats single-stranded regions as freely jointed chains [14]. The non-linear screening effect of monovalent cations is incorporated using Manning’s approach to counter-ion condensation [15]. The formulation for loop contribution was originally developed to understand the salt-dependent modulation of nucleosomal structures [16]. The two strands of a helix are modeled as parallel rods that again interact via a Debye-Hückel potential governed by the screening length. Here, the theory yields a per-position contribution that is independent of sequence features and the position within the helix [12]. While the theory makes several approximations it has been shown by its authors to reproduce experimental data quite well. It also has the advantage that it has no free parameters other than well known generic geometric characteristics of RNA 3D structures such as distances between

nucleotides or the planes of stacked pairs. Of course, the secondary structure based approach has significant limitations. In particular, it is by design not suitable to model the site-specific binding of chelated  $\text{Mg}^{2+}$ , which in some cases is known to be crucial for tertiary structure formation and RNA function [17].

Here we implement in the **ViennaRNA** package [18] the Debye-Hückel/Manning model, which captures the energetic effects of electrostatic interactions of monovalent cations with structured RNA [12]. In the following theory section we briefly review the features of the energy model and show that it can be brought into a form where only the energy parameters but not the dynamic programming recursions are modified. We then evaluate the model on a collection of empirical data from the literature to show that use of this type of salt corrections has a significant beneficial effect.

## Theory

The model of [12] considers loops a freely jointed charged chains and helices as parallel rods interacting via Debye-Hückel potentials and uses Manning's framework to model counter-ions condensation. This results in distinct types of correction terms from loops and helices, which we describe separately in the following. In either case we focus on how the salt correction terms are incorporated into the dynamic programming schemes for RNA secondary structure prediction. As we shall see, the salt corrections can be phrased in such a way that they exclusively effect the energy parameters. The folding algorithms therefore remain unchanged. Owing to the architecture of the **ViennaRNA** package, it is therefore possible to handle salt effects exclusively as a pre-processing of the energy parameter set.

### Salt corrections for loops

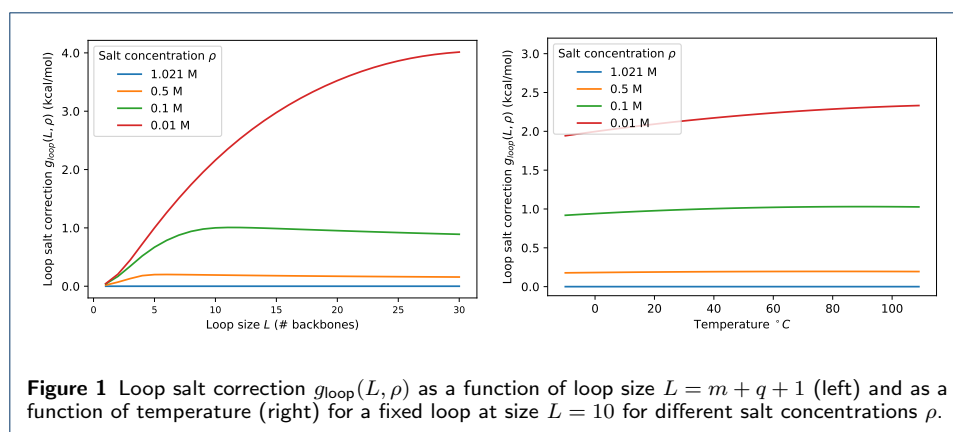
The electrostatic free energy contribution for a loop comprising  $L$  backbone bonds can be written, at the level of the Debye-Hückel approximation, in the form [16]:

$$\mathcal{G}_u(L) = RT \frac{\ell_B}{\kappa} \tau_{ss}^2 \Phi(y) \quad (1)$$

with  $y = \kappa l_{ss} L$  where  $l_{ss} = 6.4 \text{ \AA}$  is the length of an RNA backbone bond,  $\ell_B$  is the Bjerrum length and  $\kappa^{-1}$  is the Debye screening length, which depends on the ionic strength, and  $\tau_{ss} = \min(1/l_{ss}, 1/\ell_B z_c)$  accounts for the nonlinear electrostatic effects. For monovalent ions, the ionic strength equals the salt concentration  $\rho$  and thus we have  $\kappa = \kappa(\rho)$ . Following [16],  $\Phi$  is given by

$$\begin{aligned} \Phi(y) = & y \ln y + y(\gamma - \ln(\pi/2)) - \frac{y^2}{2} {}_1F_2 \left( \begin{matrix} 1/2 \\ 1, 3/2 \end{matrix} \middle| \frac{y^2}{4\pi^2} \right) \\ & - \frac{y^3}{2\pi^2} {}_2F_3 \left( \begin{matrix} 1, 1 \\ 3/2, 3/2, 2 \end{matrix} \middle| \frac{y^2}{4\pi^2} \right) + 1 - \exp(-y) + y\Gamma(0, y) \end{aligned} \quad (2)$$

Here,  ${}_1F_2$  and  ${}_2F_3$  are generalized hypergeometric functions and  $\Gamma(0, y)$  is the incomplete gamma function. Using  ${}_1F_2(\dots, 0) = {}_2F_3(\dots, 0) = 1$ , and  $y \ln y \rightarrow 0$  for  $y \rightarrow 0$ , we observe that  $\Phi(y) = 0$ . Equ.(A8) in [16] gives an expansion for small  $y$



of form  $\Phi(y)/y = (1 - \ln(\pi/2)) + O(y^2)$ , where the  $\ln y$  term and the logarithmic divergence of  $\Gamma(0, y)$  cancel. Thus,  $\Phi(y)$  increases linearly with  $y$ . Since both  $y$  and  $\kappa$  are proportional to  $\sqrt{\rho}$ , the  $\rho$ -dependence cancels and  $\mathcal{G}_u(L)$  approaches  $L$  times a constant for  $\rho \rightarrow 0$ .

Since  $\mathcal{G}_u(L)$  describes the salt-dependent electrostatic effects on loops, this term is already included in the empirical energy parameters of the RNA standard model for the standard conditions of  $T = 37^\circ\text{C}$  and 1M sodium concentration. Writing  $\hat{\mathcal{G}}_u(L)$  for the values on Eq.(1) at standard conditions, allows us to write the salt correction of a given loop as

$$g_{\text{loop}}(L, \rho) := \mathcal{G}_u(L) - \hat{\mathcal{G}}_u(L) \quad (3)$$

The ViennaRNA package quantifies the length of a loop by the number  $m$  of unpaired nucleotides rather than the number of bonds. For hairpin loops we have  $L = m + 1$ , while  $L = m + 2$  for interior loops. In multi-loops we have  $L = m + q + 1$ , where  $q$  is the degree (number of branches) of the multiloop. Setting  $q = 0$  for hairpin loops and  $q = 1$  for interior loops, the appropriate salt correction for a loop is therefore  $g_{\text{loop}}(m + q + 1, \rho)$ .

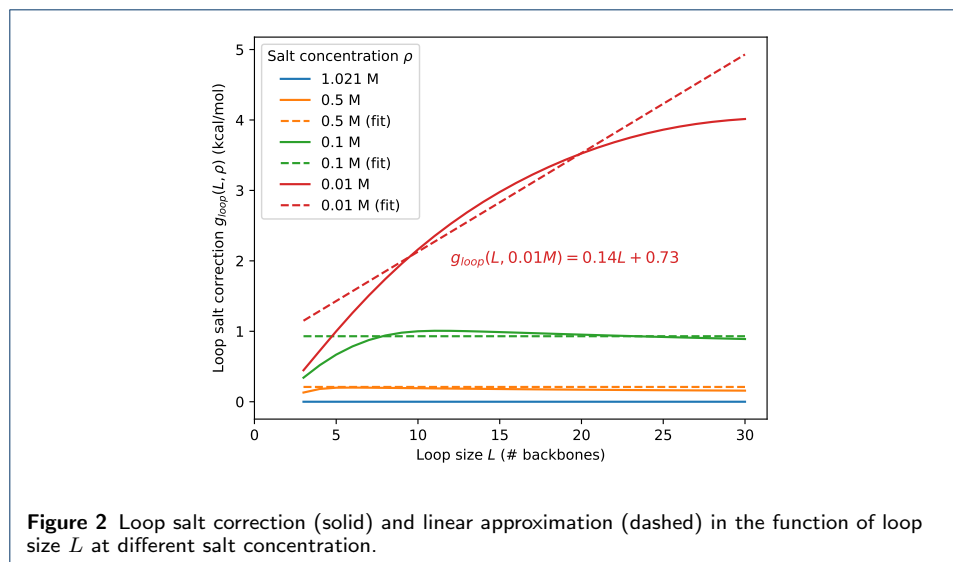
As seen in Fig. 1, the salt correction, as expected, increases while decreasing the salt concentration contributing to the destabilization of the structure in a low salt environment. In the usual temperature range, the plot shows that loop correction is closed to a constance with a slight increase at low concentration.

### Linear approximation for multi-branch Loop

Although the conformation energy of a loop asymptotically depends on logarithmically on the length [19], this behavior is usually approximated for multi-loops by a linear function

$$E_{ml} = \alpha + \beta q + \gamma m \quad (4)$$

for the sake of computational efficiency. The parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  are the energy cost for having, respectively, the closing pair, branch, and unpaired base in a multiloop. Without this approximation, the dynamic programming recursions of the



folding algorithm become cubic in their memory consumption since loop lengths need to be tracked explicitly [20]. The linear approximation is further motivated by the empirical observation that models with logarithmic dependence are outperformed by the linear approximation in terms of accuracy of structure prediction [20].

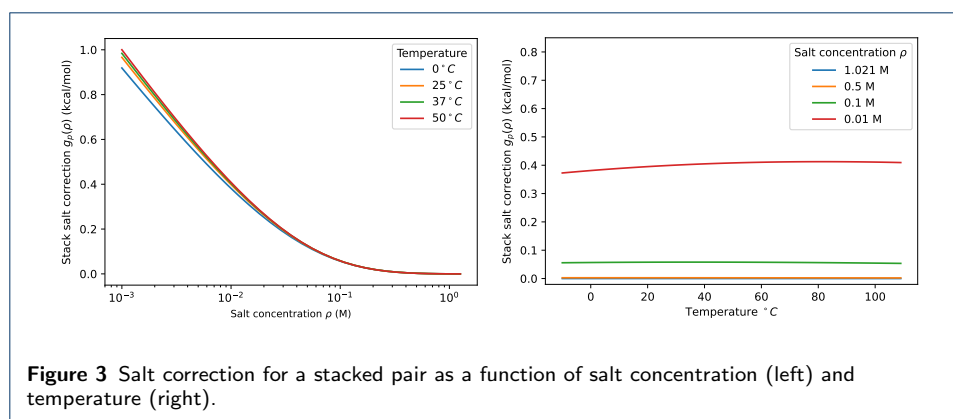
In order to handle multi-loops without abandoning the memory-efficient multi-loop decomposition, we retain the linear multiloop energies and employ a linear fit of the form

$$g_{\text{loop}}(m, \rho) \approx a_0(\rho) + a_1(\rho)L \quad (5)$$

to model the salt-corrections. Again using  $L = m + q + 1$ , this amounts to adding  $a_0 + a_1$  to the closing pair term  $\alpha$  of the multi-loop and  $a_1$  to both each unpaired base  $\gamma$  and each component  $\beta$  of the multi-loop. The salt-dependent multiloop models thus reads:

$$E_{ml} = (\alpha + a_0(\rho) + a_1(\rho)) + (\beta + a_1(\rho))q + (\gamma + a_1(\rho))m \quad (6)$$

In order to fit the multi-loop parameters in practise, we first investigated their size distribution in a sample of 5000 MFE structures for different structure sizes, see Fig. S1. Although multi-loops with size  $L = 3$  exist, very short multi-loops are rare. We argue that inaccuracies in these rare cases are likely acceptable. Very short loops presumable are also subject to specifically constrained three-dimensional structures and thus follow the model only to a crude approximation in the first place. In the current implementation, we use the loop size range  $L \in [6, 24]$  to obtain linear fits for  $a_0$  and  $a_1$  from  $g_{\text{loop}}(L, \rho)$ , see Fig. 2 for salt correction and their linear approximations. In general, the fit over-corrects for very small loops and, at low salt concentrations, also for very large loops. The maximal fitting errors are on the order of 1 kcal/mol, which is still within the ball-park of the rather large inaccuracies expected for multi-loop energies.



### Salt corrections for stacked base pairs

Describing the backbones of stacked pairs as rods with distance of  $d = 2\text{nm}$  interacting via a Debye-Hückel potential with screening length  $1/\kappa$  yields the following electrostatic energy for the interaction of a nucleotide with the other strand [12]:

$$\mathcal{G}_p = 2RT\tau_{ds}^2 l_{ds} \ell_B K_0(\kappa d) \quad (7)$$

Here the charge density  $\tau_{ds} = \min(1/l_{ds}, 1/(\ell_B z_c))$  is again estimated according to Manning's counterion condensation theory [15]. The length parameter  $l_{ds} = 3.4\text{\AA}$  is the helical rise per base pair and  $z_c = 1$  is the charge of the cation.  $K_0$  denotes the zeroth-order modified Bessel function of the 2nd kind, see e.g. [21] and Fig. S2. Since  $K_0(z)$  diverges like  $-\ln z$  for  $z \rightarrow 0$ , and  $\kappa \sim \sqrt{\rho}$ , the salt corrections  $\mathcal{G}_p$  diverges logarithmically for vanishing salt concentrations.

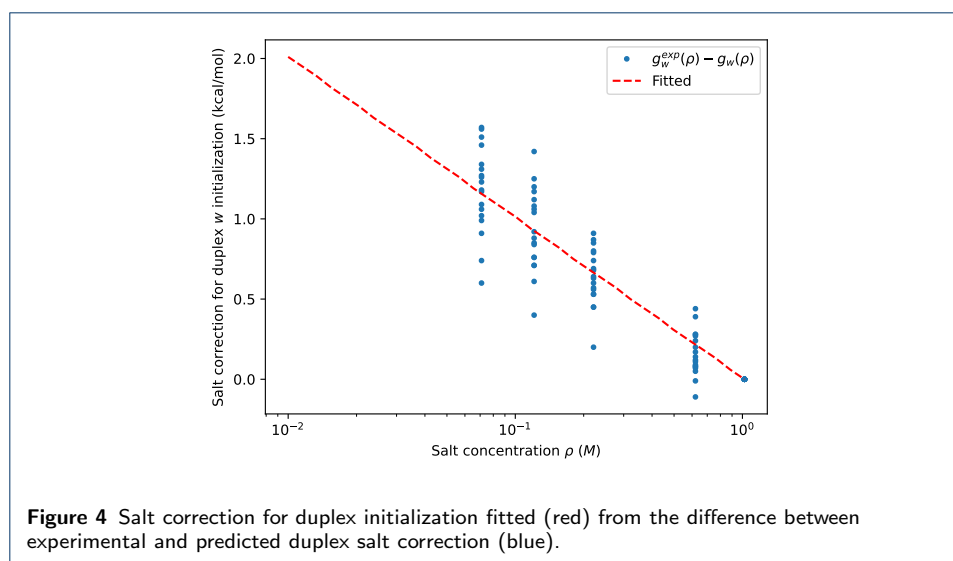
As in the case of loops, these electrostatic effects are already included in the empirical energy parameters for standard conditions. The relevant salt correction thus is given by the difference between the values  $\mathcal{G}_p$  at the current conditions and standard conditions.

$$g_p := \mathcal{G}_p - \hat{\mathcal{G}}_p \quad (8)$$

Fig. 3 shows the dependence of the position-wise salt correction for stacking energies as function of salt concentration and temperature. Similar to the loop correction, the stack correction is closed to a constance in the usual temperature range for a given salt concentration.

### Salt corrections for duplex initialization

The formation of a double strand from two RNA molecules in solution is associated with an additional initialization energy  $E_{\text{init}}$  in the Turner energy model. One expects that duplex formation becomes more difficult due to electrostatic repulsion at low salt concentrations. The initialization energy thus should also depend on  $\rho$ . In addition, the distance between two single strands changes during formation, which was neglected in the theory as discussed in [12]. Indeed, as we will see later in the result section, the duplex free energies are systematically overestimated compared to the experimental data.



Due to the lack of theoretical support, we propose a salt corrections for duplex initialization  $g_{\text{init}}(\rho)$  derived from the prediction and the experimental data taken from [7]. Let  $g_w(\rho)$  and  $g_w^{\text{exp}}(\rho)$  be the predicted and experimental salt correction at concentration  $\rho$  from the standard condition for a give duplex  $w$ . Fitting the difference  $g_w^{\text{exp}}(\rho) - g_w(\rho)$  of 18 duplexes at four non-standard sodium concentrations yields, as plotted in Fig. 4, the salt corrections for duplex initialization

$$g_{\text{init}}(\rho) = a \log\left(\frac{\rho}{\rho_0}\right) \quad (9)$$

with  $a = -100.14 \text{ kcal/mol}$ .

## Implementation in ViennaRNA

The extension of **ViennaRNA** provides access to four user-defined parameters: the concentration of the monovalent cation  $\rho$ , the bounds  $L_1$  and  $L_2$  delimiting the interval of loop length that is used to fit the two multiploop parameters  $a_0$  and  $a_1$  for given salt concentration, and the user-provided salt correction  $g$  for duplex initialization. Default values are  $L_1 = 6$ ,  $L_2 = 24$ , salt concentration  $\rho = 1.021 \text{ M}$ , and  $g = 0$ . In **ViennaRNA**, these parameters are appended in the model object `vrna_md_s` as `salt` for  $\rho$ , `saltMLLower` and `saltMLUpper` for  $L_1$  and  $L_2$ , and `saltDPXInit` for  $g$ . For the salt concentration, **ViennaRNA** assumes the standard conditions of the Turner energy model, *i.e.*,  $\rho = 1.021 \text{ mol/l}$ . Thus no salt corrections apply by default. If a different concentration  $\rho$  is requested, first the value of  $g_p(\rho)$  for the stack and the array of values  $g_{\text{loop}}(L, \rho)$  for loop of different size  $L \in [1, 31]$  is computed. Note that these energy contribution depend on the temperature  $T$  and thus are recomputed if the user sets a different temperature. In addition, the use of duplex initialization salt correction  $g_{\text{init}}(\rho)$  for duplex is turned off if users set `saltDPXInit` to zero.

The array  $g_{\text{loop}}$  is then used to determine  $a_0(\rho)$  and  $a_1(\rho)$  by linear regression. Subsequently, the free energy parameters are set as sums of the default values  $E(0)$

for the given temperature and the salt corrections:

$$\begin{aligned}
 E^{bp} &= E^{bp}(0) + g_p(\rho) \\
 E_m^{HL} &= E_m^{HL}(0) + g_{\text{loop}}(m+1, \rho) \\
 E_m^{IL} &= E_m^{IL}(0) + g_{\text{loop}}(m+2, \rho) \\
 \alpha &= \alpha(0) + a_0(\rho) + a_1(\rho) \\
 \beta &= \beta(0) + a_1(\rho) \\
 \gamma &= \gamma(0) + a_1(\rho) \\
 E_{\text{init}} &= E_{\text{init}} + g_{\text{init}}(\rho)
 \end{aligned} \tag{10}$$

Here  $E^{bp}$ ,  $E_m^{HL}$ ,  $E_m^{IL}$ , and  $E_{\text{init}}$  refer to *all* parameters for stacked pairs, hairpin loops of length  $m$ , interior loops of length  $m$ , and duplex initialization, respectively. Dangling end and coaxial stacking contributions, on the other hand, remain unchanged.

The Bessel functions  $K_0$  is computed as in `scipy`, which in turn used the `cephes` mathematical function library described in [22]. The function  $\Phi$  in loop salt correction can be approximated as given in [12]:

$$\begin{aligned}
 \Phi(y) &= y \ln y + y(\gamma - \ln(\pi/2)) + y \frac{(2\pi)^6}{y^6 + (2\pi)^6} \left( \frac{y^4}{36\pi^4} - \frac{y^3}{24\pi^2} + \frac{y^2}{2\pi^2} - \frac{y}{2} \right) \\
 &+ \left( 1 - \frac{(2\pi)^6}{y^6 + (2\pi)^6} \right) (y(\log(2\pi) - 1.96351) - y \log y) + (1 - e^{-y}) + y\Gamma(0, y)
 \end{aligned} \tag{11}$$

## Parameters

The key physical parameter containing the salt correction terms are the Debye screening length  $\kappa^{-1}$ . It is convenient to express  $\kappa^{-1}$  in terms of Bjerrum length  $\ell_B$  and the ionic strength  $I$ :

$$\begin{aligned}
 \ell_B &= \frac{e_0^2}{4\pi k_B T \epsilon_0 \epsilon_r(T)} \approx (167100 \text{ ÅK}) \frac{1}{T \epsilon_r(T)} \\
 \kappa^{-1} &= \sqrt{\frac{k_B T \epsilon_0 \epsilon_r(T)}{2N_A e_0^2 I}} = \frac{1}{\sqrt{8\pi N_A}} \frac{1}{\sqrt{\ell_B I}} \approx \left( 8.1284 \text{ Å}^{3/2} \text{l}^{-1/2} \text{mol}^{1/2} \right) \frac{1}{\sqrt{\ell_B I}}
 \end{aligned} \tag{12}$$

The constants appearing in these expressions are the unit charge  $e_0 \approx 1.602 \cdot 10^{-19} \text{C}$ , Avogadro's constant  $N_A \approx 6.022 \cdot 10^{23} \text{mol}^{-1}$ , and the vacuum permittivity  $\epsilon_0 \approx 8.854 \cdot 10^{-22} \text{C}^2/\text{JÅ}$ . The salt concentration enters only through the ionic strength  $I := \frac{1}{2} \sum_i \rho_i z_i^2$ , where  $\rho_i$  is the concentration and  $z_i$  is the charge of ion-species  $i$ . For monovalent ions, as in the case of NaCl, the ionic strength reduces to the salt concentration, i.e.,  $I = \rho$ . The salt concentration  $\rho$  is expressed as molarity, i.e., in units of mol/l, whereas the Debye screening length  $\kappa^{-1}$  and the Bjerrum length  $\ell_B$  are conveniently expressed in Å. Note that the conversion factor for the length units (l=dm<sup>3</sup> versus Å) is absorbed into the numerical constant.



The temperature dependence of  $\epsilon_r$  can be fitted from empirical data. In the current implementation we use the function proposed in [23]:

$$\epsilon_r(T) = 5321T^{-1} + 233.76 - 0.9297T + 1.417 \cdot 10^{-3}T^2 - 0.8292 \cdot 10^{-6}T^3 \quad (13)$$

with temperature in Kelvin. The temperature-dependence of  $\epsilon_r$  ensures that the Bjerrum length  $\ell_B$  is longer than the backbone bond length  $l_{ss}$  in the entire temperature range, see Fig. S3. The nonlinear electrostatic effects on unpaired bases thus become  $\tau_{ss} = 1/(\ell_B)$  for the monovalent cations considered here.

## Comparison with experimental data

### Available data sets

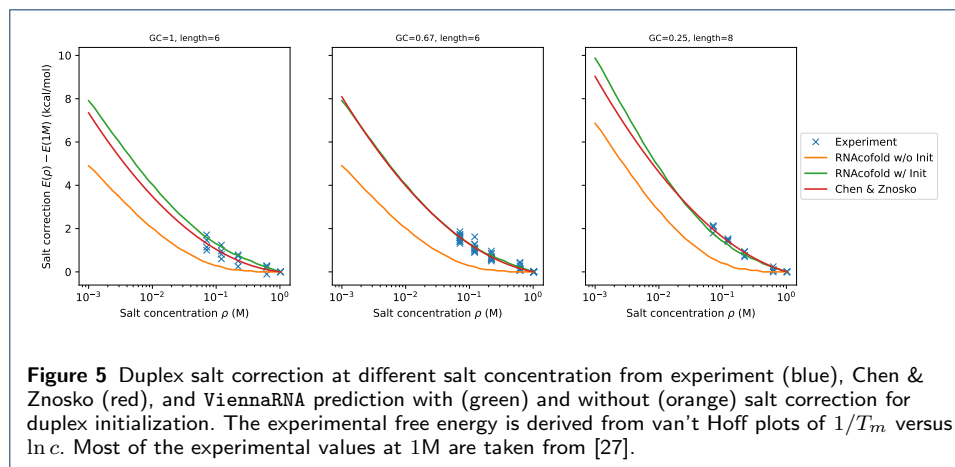
Even though the dependence of RNA structures on salt concentrations is of considerable practical interest, systematic data sets suitable for benchmarking salt correction models are by no means abundant. Most of the direct evidence for the salt dependence derives from studies of short duplexes and hairpins. Here, we analyzed the following three datasets:

- 1 Melting experiment of 18 RNA self-complementary duplexes of length 6 at the 8 different sodium concentration, 0.071, 0.121, 0.221, 0.621, and 1.021M, with different species concentration  $c$  were reported in [7, 24]. Both melting temperatures  $T_m$  and free energies at  $37^\circ C$  were obtained from  $1/T_m$  vs.  $\ln c$  plots. The same data were also used by the same lab [24] to obtain optimised thermodynamic parameters, entropy and enthalpy, at a given salt concentration.
- 2 A different set of 8 self-complementary duplexes of length 10, 12, and 14 was reported in [25]. The melting temperature  $T_m$  is derived from the melting curve as the temperature at which half of the dimer dissolved. The data set covers two different species concentration  $c$ ,  $100\mu M$  and  $2\mu M$  and two sodium concentrations,  $1.0002M$  and  $0.0102M$ .
- 3 The free energies of forming hairpins at different sodium concentrations were measured in [26]. The thermodynamics parameters at different salt concentration, different sodium concentration with an additional of  $0.011M$  cations in the buffer, were obtained from melting experiments. Hairpins consist of a helix of length 5 and a hairpin loop of size 8 or 10.

### Comparison of duplex free energies

Free energies were computed using RNAcofold [28, 29] with a self-complementary correction of  $RT \ln 2$  added for all sequences that coincided with their reverse complement. Since the predicted free energy at the standard condition differs from the experimental values, we are interested in comparing the salt correction  $g_w(\rho) = E_w(\rho) - E_w(\rho_0)$  at concentration  $\rho$  from the standard concentration  $\rho_0 = 1.021M$ , where  $E_w(\rho)$  is the free energy of duplex  $w$  at concentration  $\rho$ . Let  $l_w$  and  $gc_w$  be the length and the fraction of GC of duplex  $w$ . Then the salt correction for RNAcofold is then given by.

$$g_w^{\text{RNAcofold}}(\rho) = (l_w - 1) \times g_p(\rho) + g_{\text{init}}(\rho).$$



The salt correction proposed by Chen & Znosko [7] is

$$g_w^{C\&Z}(\rho) = (0.324gc_w - 0.486) \ln(\rho/\rho_0) + 0.133(\ln(\rho/\rho_0))^2.$$

The computational results are summarized in Fig. 5. Not surprisingly, the Chen & Znosko salt correction provides a slightly better fit to the data because the empirical formula for  $g_w^{C\&Z}$  was obtained by fitting to the same data set. In contrast, only the duplex initialization  $g_{\text{init}}(\rho)$  is fitted in our model. The largest deviations are observed for GC-only sequences.

### Comparison of duplex melting temperatures

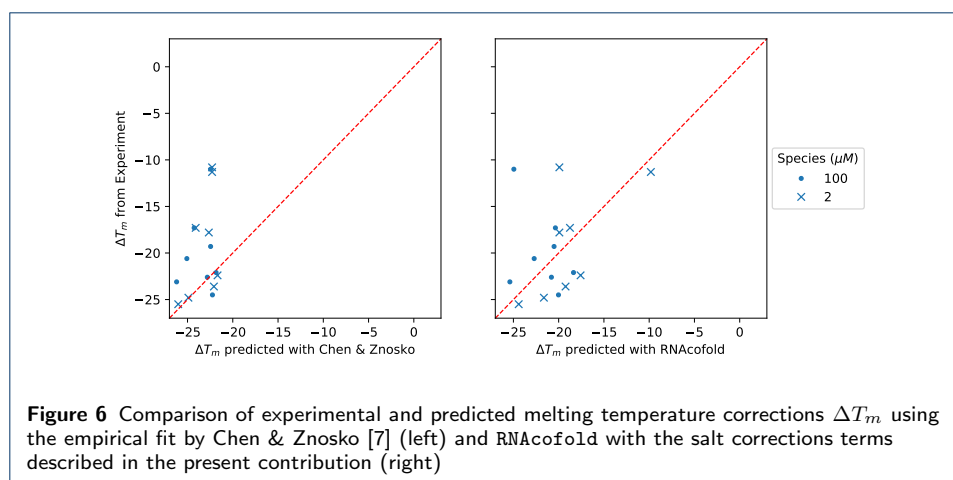
Let  $A$  be a self-complementary RNA sequence and  $AA$  the corresponding dimer. The dimerization reaction is  $2A \rightleftharpoons AA$ . The corresponding concentrations are denoted by  $[A]$  and  $[AA]$ , respectively. In equilibrium, we have

$$\frac{[AA]}{[A][A]} = \frac{Z_{AA}}{Z_A Z_A} = e^{(2G_A - G_{AA})/RT} \quad (14)$$

where  $Z_A$  and  $Z_{AA}$  are the partition functions of the monomer and the dimer, respectively. Here  $G_A = -RT \ln Z_A$  and  $G_{AA} = -RT \ln Z_{AA}$  are the *ensemble* free energies of  $A$  and  $AA$ , respectively. Note that in a pure two-state system, we can replace  $G_A$  and  $G_{AA}$  by the corresponding minimum free energies  $E_A$  and  $E_{AA}$ , respectively. We define the melting temperature  $T_m$  as the temperature at which half of  $A$  forms the dimer  $AA$ , i.e., where  $[AA] = c/4$  and  $[A] = c/2$ . Equ.(14) then yields

$$T_m = \frac{2G_A(T_m) - G_{AA}(T_m)}{-R \ln c}, \quad (15)$$

where we write  $G_A(T_m)$  and  $G_{AA}(T_m)$  to emphasize the temperature dependence of the ensemble free energies. The ensemble free energy  $G_{AA}$  for the pure dimer state is accessible directly via the function `fc.dimer_pf()` within the ViennaRNA library. The correction for self-complementary is already taken into account during the



computation of the the partition function [29]. Since the ensemble free energy is also a function of temperature, we use a binary search to find the melting temperature  $T_m$ .

For the data set comprising 18 short duplexes, experimental melting temperatures are available for several distinct species concentrations  $c$ . Van't Hoff's equation

$$\frac{1}{T_m} = \frac{R}{\Delta H} \ln c + \frac{\Delta S}{\Delta H}$$

implies a linear relationship between changes in  $1/T_m$  and changes in RNA concentration. We compare the predicted and experimental “van't Hoff” plots, i.e., diagrams of  $1/T_m$  versus  $\ln c$ , see Fig. S4. Overall, we observe an excellent agreement on the slope between RNAcofold prediction and experiment. Predicted and experimental intercepts are slightly more different for few of the duplexes.

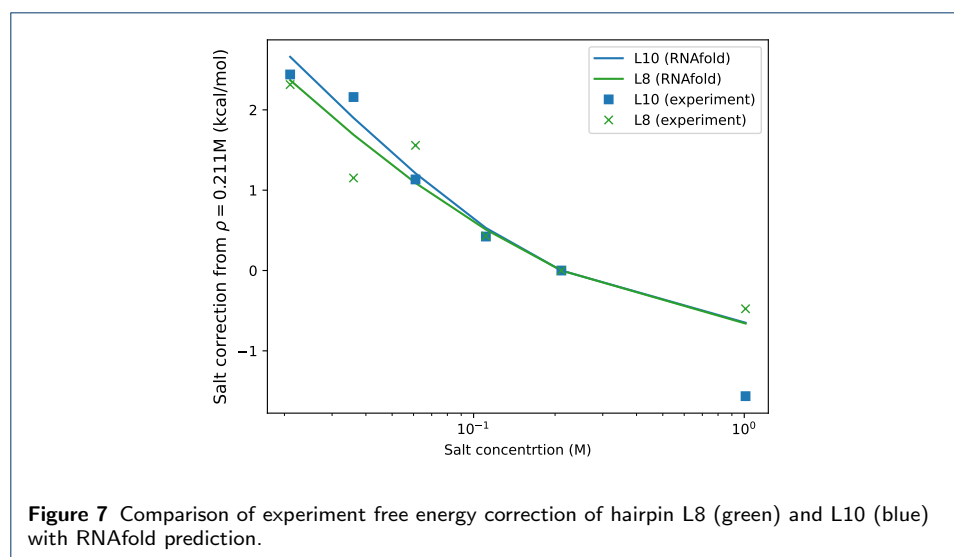
For the second dataset consisting of longer duplexes, we are interested in melting temperature correction  $\Delta T_m(0.01M)$  from the standard condition  $1M$  since the data is only available for two species concentrations. Chen & Znosko [7] proposed the following fit

$$\Delta T_m^{C\&Z}(\rho) = (-1.842gc_w + 2.675) \ln \rho - 0.7348(\ln \rho)^2$$

where  $gc_w$  is the GC fraction of duplex  $w$ . Fig. 6 shows the experimental melting temperature correction compared with  $\Delta T_m^{C\&Z}$  and the one computed by RNAcofold. Overall, the salt corrections described above show similar agreement (Pearson correlation  $r = 0.329$ ) with the experiment as the empirical fit ( $r = 0.341$ ) from [7].

### Comparison of hairpin free energies

For the hairpins L8 and L10 described in [26], entropy, enthalpy, and free energy were derived from melting experiments. For comparison, we computed the free energy using RNAfold at  $37^\circ C$  for different salt concentrations  $\rho$ . Fig. 7 summarizes the experimental and predicted free energy corrections from 0.211 M. They are in very good agreement in particular for lower salt concentrations.



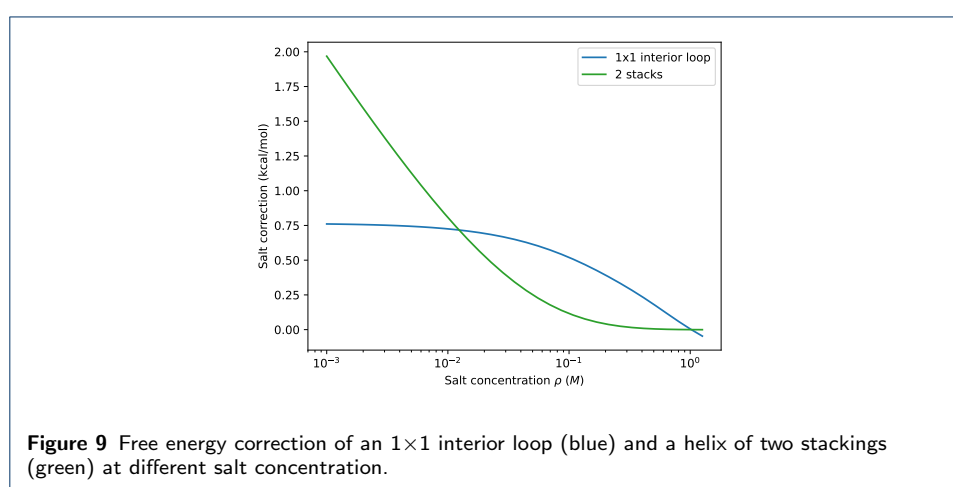
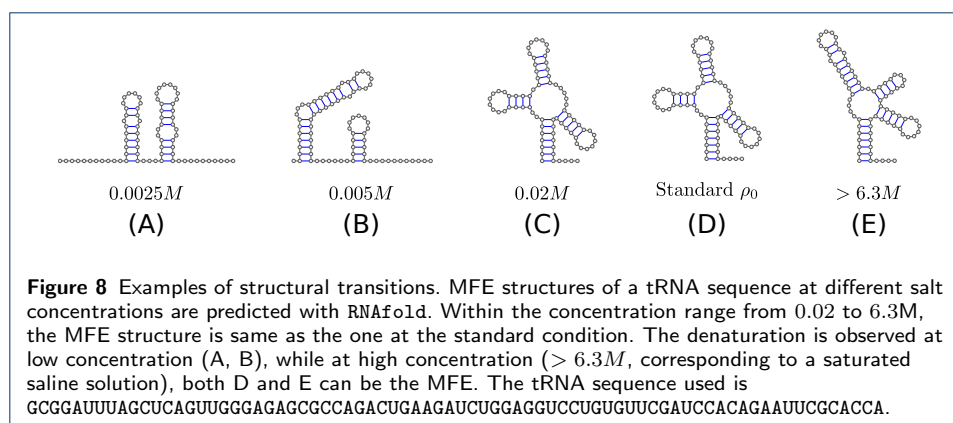
## Discussion

Salt concentrations significantly influence folding and thermodynamics of nucleic acids. In this contribution we report on the implementation of an approximate physical model proposed by Einert and Netz [12] that represent the RNA backbone as a charged polymer interacting by means of a Debye-Hückel potential. The model was adapted to preserve the linear multi-loop model required for computational efficiency and extended by an empirical initiation term for duplex formation. While not perfect, the model reproduces experimental thermodynamic data on the NaCl dependence of folding energies and melting temperatures with reasonable accuracy. We note that the salt-dependent energies for stacks diverge logarithmically for vanishing salt concentrations  $\rho$ . The model thus cannot be used if cations are virtually absent.

The approach taken here has the practical advantage that it does not require any changes in the folding algorithms. The modification of the energy parameters is sufficient. The **ViennaRNA** packages handles this step during preprocessing. As a consequence, the computational performance of the folding routines remain unchanged. Moreover, the salt corrections are consistently applied to all variants of the folding algorithms, e.g. minimum energy and partition function computations, consensus computations from alignments, and co-folding of two or more components.

Changes in salt concentration also affect the predicted secondary structures, see Fig. 8 for an example. To our knowledge, there are no detailed experimental data that document structural changes as a function of NaCl concentration so that a directed validation of structures predicted for low salt concentrations remains a topic for future research.

The proposed salt correction diverges when the salt concentration becomes close to 0. The same behavior is also observed in our linear fitted correction for duplex initialization. To address it, we fitted a correction function with one parameter more comparing to Equ.(9) that converges at low concentration. Fig. S5 plots the predicted melting temperature correction versus the experiments of longer duplex, which shows a better agreement with Pearson correction  $r = 0.54$ . Due to the lack



of data to validate the hypothesis of convergence at low concentration, we propose the simpler Equ.(9) as the main correction for duplex initialization.

In the current model, a basepair mismatch in a helix is treated as an 1×1 interior loop and thus is associated with the salt correction for loops at non-standard salt concentrations. However, such a mismatch is likely to result in a slightly distorted helix that could still be seen as two parallel charged rods. One could therefore argue, that the salt correction for 2 stacked pairs rather than for a loop should be applied. Fig. 9 shows the difference of these two cases as a function of salt concentration. To our knowledge, there are no experimental data that could be used to decide which approach is more appropriate.

The approach taken here does not account for all effects of ions on RNA folding. Most importantly, it covers only unspecific interactions and thus does not describe specific interactions e.g. of  $Mg^{2+}$  with specific binding sites. Even for unspecific interactions, the validity of the model can be argued stringently only for mono-valent ions [12]. At present, available data, e.g. [30], are not sufficient to test whether replacing  $\rho$  by the ionic strength is sufficient to reasonably account for mixtures of mono-valent and di-valent cations.

The model described here does not account for salt effects of RNA structures that are neither loops nor stacked base pairs. In particular it does not apply to G quadruplexes [31], which optionally can be included in secondary structure predic-

tions [32]. Separate models for the ion dependencies of such features will need to be derived that account e.g. for the  $K^+$ -dependent stabilization of RNA quadruplexes.

#### **Acknowledgements**

##### **Funding**

This research was funded in part by the Austrian Science Fund (FWF), grant no. I 4520 and F 80, as well as the German Research Foundation (DFG), grant no. STA 850/48-1.

##### **Availability of data and materials**

The implementation is available in the ViennaRNA Package starting with version 2.6.0 available at <https://www.tbi.univie.ac.at/RNA>.

##### **Ethics approval and consent to participate**

Not applicable.

##### **Competing interests**

The authors declare that they have no competing interests.

##### **Consent for publication**

Not applicable.

##### **Authors' contributions**

All authors contributed to the design of the study and the writing of the manuscript. HTY implemented the model and conducted the computational evaluation, RL integrated the new features into the ViennaRNA package.

# Additional Files

Additional File 1 — List of supplementary figures

Figure S1 — Length distribution of multiloops

Distribution of multiloop size  $L$ , number of backbones, among MFE structures of 5 000 uniformly selected sequences at varied length.

Figure S2 — Approximation Error for  $K_0$

In [12] an approximation for the difference of  $K_0$  at a given concentration and  $1M$  was proposed. However, we noticed that this approximation yields a non-vanishing salt correction at  $1M$ . We therefore used the Cephes library to compute  $K_0$  directly. The panel shows the salt correction of base pair stack at  $37^\circ C$  in the function of salt concentration using the approximation (blue) and the precise computation implemented in ViennaRNA (orange).

Figure S3 — Nonlinear electrostatic effects  $\tau_{ss}$

In [12], the permittivity (relative dielectric constant)  $\epsilon_r$  of water  $\epsilon_r \approx 80$  is assumed to be temperature independent. This assumption results in a discontinuity of  $\tau_{ss}$  at around  $53.3^\circ C$ . Incorporating the empirical temperature dependence of  $\epsilon_r$  in equ.(13) [23] results in  $1/\ell_B < 1/l_{ss}$ .

Figure S4 — Van t'Hoff plots for 18 duplexes.

Plotting  $1/T_m$  versus  $\ln c$  shows a generally good agreement of between predictions and the experimental data from [7].

Figure S5 — Converged salt correction for duplex initialization.

Converged correction function fitted (left) to the difference  $g_w^{exp}(\rho) - g_w(\rho)$  of 18 duplexes data [7]. The plot (right) of the predicted melting temperature correction versus the experiments of longer duplexes [25] shows a better agreement with Pearson correction  $r = 0.54$ .

# Author details

<sup>1</sup>Department of Theoretical Chemistry, University of Vienna Währinger Straße 17, A-1090 Vienna, Austria.

<sup>2</sup>Research group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Währingerstraße 29, A-1090, Vienna, Austria. <sup>3</sup>Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16–18, D-04107, Leipzig, Germany.

<sup>4</sup>Competence Center for Scalable Data Services and Solutions Dresden/Leipzig, Interdisciplinary Center for Bioinformatics, German Centre for Integrative Biodiversity Research (iDiv), and Leipzig Research Center for Civilization Diseases, Universität Leipzig, Augustusplatz 12, D-04107 Leipzig, Germany. <sup>5</sup>Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04109 Leipzig, Leipzig, Germany. <sup>6</sup>Facultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Ciudad Universitaria, COL-111321 Bogotá, D.C., Colombia. <sup>7</sup> Santa Fe Institute, 1399 Hyde Park Rd., NM87501 Santa Fe, USA.

# References

1. Lipfert, J., Doniach, S., Das, R.D., Herschlag, D.: Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* **83**, 813–841 (2014). doi:10.1146/annurev-biochem-060409-092720
2. Brion, P., Westhof, E.: Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113–137 (1997). doi:10.1146/annurev.biophys.26.1.113
3. Turner, D.H., Mathews, D.H.: NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucl. Acids Res.* **38**, 280–282 (2010). doi:10.1093/nar/gkp892
4. Lu, Z.J., Turner, D.H., Mathews, D.H.: A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Res.* **34**, 4912–4924 (2006). doi:10.1093/nar/gkl472
5. Draper, D.E.: A guide to ions and RNA structure. *RNA* **10**, 335–343 (2004). doi:10.1261/rna.5205404
6. Draper, D.E.: RNA folding: Thermodynamic and molecular descriptions of the roles of ions. *Biophysical J.* **95**, 5489–5495 (2008). doi:10.1529/biophysj.108.131813
7. Chen, Z., Znosko, B.M.: Effect of sodium ions on RNA duplex stability. *Biochemistry* **52**(42), 7477–7485 (2013). doi:10.1021/bi4008275
8. Chin, K., Sharp, K.A., Honig, B., Pyle, A.M.: Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol.* **6**, 1055–1061. doi:10.1038/14940
9. Misra, V.K., Draper, D.E.: The interpretation of Mg(2+) binding isotherms for nucleic acids using Poisson-Boltzmann theory. *J Mol Biol* **294**, 1135–1147 (1999). doi:10.1006/jmbi.1999.3334
10. Misra, V.K., Draper, D.E.: The linkage between magnesium binding and RNA folding. *J Mol Biol* **317**, 507–521 (2002). doi:10.1016/j.jmbi.2002.5422
11. Zuker, M., Stiegler, P.: Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 133–148 (1981). doi:10.1093/nar/9.1.133
12. Einert, T.R., Netz, R.R.: Theory for RNA folding, stretching, and melting including loops and salt. *Biophys. J.* **100**, 2745–2753 (2011). doi:10.1016/j.bpj.2011.04.038
13. Debye, E. Peter und Hückel: Zur Theorie der Elektrolyte. *Physik. Z.* **24**, 185–206 (1923)
14. Mazars, M.: Statistical physics of the freely jointed chain. *Phys. Rev. E* **53**, 6297–6319 (1996). doi:10.1103/PhysRevE.53.6297
15. Manning, G.S.: Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties. *J. Chem. Phys.* **51**, 924–933 (1969). doi:10.1063/1.1672157
16. Kunze, K.-K., Netz, R.R.: Complexes of semiflexible polyelectrolytes and charged spheres as models for salt-modulated nucleosomal structures. *Phys. Rev. E* **66**, 011918. doi:10.1103/PhysRevE.66.011918

17. Grilley, D., Soto, A.M., Draper, D.E.:  $Mg^{2+}$ -RNA interaction free energies and their relationship to the folding of RNA tertiary structures. *Proc. Natl. Acad. Sci. USA* **103**, 14003–14008 (2006). doi:10.1073/pnas.06064091
18. Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L.: ViennaRNA Package 2.0. *Alg. Mol. Biol.* **6**, 26 (2011). doi:10.1186/1748-7188-6-26
19. Jacobson, H., Stockmayer, W.H.: Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys.* **18**, 1600–1606 (1950). doi:10.1063/1.1747547
20. Ward, M., Datta, A., Wise, M., Mathews, D.H.: Advanced multi-loop algorithms for RNA secondary structure prediction reveal that the simplest model is best. *Nucleic Acids Res.* **45**, 8541–8550 (2017). doi:10.1093/nar/gkx512
21. Yang, Z.-H., Chu, Y.-M.: On approximating the modified bessel function of the second kind. *J. Inequal Appl.* **2017**, 41 (2017). doi:10.1186/s13660-017-1317-z
22. Moshier, S.L.: *Methods and Programs for Mathematical Functions*. Prentice-Hall, Hoboken, NJ (1989)
23. Catenaccio, A., Daruich, Y., Magallanes, C.: Temperature dependence of the permittivity of water. *Chem Phys Letters* **367**, 669–671 (2003)
24. Ferreira, I., Jolley, E.A., Znosko, B.M., Weber, G.: Replacing salt correction factors with optimized RNA nearest-neighbour enthalpy and entropy parameters. *Chem. Phys.* **521**, 69–76 (2019). doi:10.1016/j.chemphys.2019.01.016
25. Nakano, S.-i., Kirihata, T., Fujii, S., Sakai, H., Kuwahara, M., Sawai, H., Sugimoto, N.: Influence of cationic molecules on the hairpin to duplex equilibria of self-complementary DNA and RNA oligonucleotides. *Nucleic Acids Res.* **35**(2), 486–494 (2007). doi:10.1093/nar/gkl1073
26. Williams, D.J., Hall, K.B.: Thermodynamic comparison of the salt dependence of natural RNA hairpins and RNA hairpins with non-nucleotide spacers. *Biochemistry* **35**(46), 14665–14670 (1996). doi:10.1021/bi961654g
27. Xia, T., SantaLucia Jr, J., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., Turner, D.H.: Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37**(42), 14719–14735 (1998). doi:10.1021/bi9809425
28. Bernhart, S.H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P.F., Hofacker, I.L.: Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.* **1**, 3 (2006). doi:10.1186/1748-7188-1-3
29. Lorenz, R., Flamm, C., Hofacker, I.L., Stadler, P.F.: Efficient algorithms for co-folding of multiple RNAs. In: Ye, X., Soares, F., De Maria, E., Gómez Vilda, P., Cabitza, F., Fred, A., Gamboa, H. (eds.) *Biomedical Engineering Systems and Technologies. BIOSTEC 2020. Comm. Comp. Inf. Sci.*, vol. 1400, pp. 10–1007978303072379810. Springer, Cham
30. Nguyen, H.T., Hori, N., Thirumalai, D.: Theory and simulations for RNA folding in mixtures of monovalent and divalent cations. *Proc. Natl. Acad. Sci.* **116**, 21022–21030 (2019). doi:10.1073/pnas.1911632116
31. Bhattacharyya, D., Arachchilage, G.M., Basu, S.: Metal cations in G-quadruplex folding and stability. *Front. Chem. Sec. Chemical Biology* **4** (2016). doi:10.3389/fchem.2016.00038
32. Lorenz, R., Bernhart, S.H., Qin, J., Höner zu Siederdissen, C., Tanzer, A., Amman, F., Hofacker, I.L., Stadler, P.F.: 2D meets 4G: G-quadruplexes in RNA secondary structure prediction. *IEEE Trans. Comp. Biol. Bioinf.* **10**, 832–844 (2013). doi:10.1109/TCBB.2013.7