Published in final edited form as:

Chem Phys. 2019 May 1; 521: 69–76. doi:10.1016/j.chemphys.2019.01.016.

Replacing salt correction factors with optimized RNA nearest-neighbour enthalpy and entropy parameters

Izabela Ferreira^a, Elizabeth A. Jolley^b, Brent M. Znosko^b, Gerald Weber^{a,*}

^aDepartamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil

^bDepartment of Chemistry, Saint Louis University, St. Louis, Missouri 63103, United States

Abstract

We calculate the nearest-neighbour enthalpies and entropies at 5 salt concentrations of 18 RNA sequences, each for at least 9 different species concentrations, totalling 757 melting temperatures, using a melting temperature optimization method. These new parameters do not need to be salt-corrected and are shown to provide overall improved melting temperature predictions. They show a marked quadratic dependence with salt concentrations which are compensated to form linear Gibbs free energies. Two different parameter schemes were tested, with fixed or variable initial parameters. We have found that using variable initial parameters provides better predictive results than using salt correction factors and that the prediction uncertainty is considerably reduced for a validation set of independent sequences. An interpolation scheme is introduced to generate model parameters for arbitrary salt concentrations which performs better against a validation set than predictions using salt corrections.

Keywords

RNA stability; Nearest-neighbour model; Melting temperature calculation

1. Introduction

RNA nearest neighbour (NN) parameters are instrumental for many applications in molecular biology, of which perhaps the best known is secondary structure prediction [1,2]. Sequence dependent transcription elongation models also require detailed knowledge for NN parameters [3] as well as target affinity [4]. Some mesoscopic theoretical models require NN enthalpy parameters for the calculation of thermodynamic properties of oligonucleotides [5,6].

*Corresponding author. gweber@ufmg.br (G. Weber).

Appendix A. Supplementary data

Supplementary Tables S1–S5 show the sequences and their measured temperatures at different species concentrations used to calculate the optimized parameters. The validation sequences are shown in Table S6. Regression coefficients for the quadratic interpolations are given in Supplementary Tables S7 and S8. Supplementary Figs. S1–S4 show the scatter plots of the predicted and experimental melting temperatures. Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.chemphys.2019.01.016>.

For obtaining NN enthalpy and entropy parameters the usual method [7,8] is to either derive first the total sequence enthalpy and entropy variation using $T_m \times \log C_t$ plots or by fitting the UV absorption curve. With the total enthalpy and entropy variation at hand the detailed NN are then calculated with algebraic methods. However, this procedure makes it difficult to obtain NN parameters if the total sequence enthalpy and entropy parameters are not available or can not be calculated with algebraic methods. Recently we developed a melting temperature optimization (MTO) method to obtain NN parameters for DNA directly from measured melting temperatures T_m [9]. The MTO method bypasses the intermediate step of calculating total sequence enthalpy and entropy parameters and we showed that for DNA it improves on melting temperature predictions, especially for higher temperatures. We also showed that we could calculate the NN parameters using a dataset for which total sequence enthalpy and entropy parameters were not given [9].

One possibility opened by the MTO method is to use directly the same temperatures that are commonly used for the $T_m \times \log C_t$ plots, avoiding the linear regression. Unfortunately, this type of data are rarely published and therefore this possibility has remained untested up to now. Here we apply the MTO method to a set of RNA melting temperatures with a large variation of species concentration C_t which allowed us to test the effectiveness to bypass the $T_m \times \log C_t$ fitting with the MTO method. We derived the NN enthalpy and entropy parameters as a function of salt concentration under two different parameter schemes where we tested the influence of optimizing or not the initiation parameters. A quadratic regression was calculated for all parameters which enables the prediction of parameters at arbitrary salt concentrations without the use of salt correction factors. We also tested the optimization of the initiation factors, where we adopted the same scheme as Xia et al. [10] for RNA; we have found that this optimization has important effects on the prediction quality.

To our knowledge, there are presently no salt-dependent NN model parameters available for RNA. The work around is to use the well known parameters from Xia et al. [10] and apply salt-correction factors [11]. It is possible to derive new model parameters at specific salt concentrations for which melting temperature data is available, but then remains the question of how to deal with intermediate concentrations where not datasets exists. Here, we test an interpolation scheme where we are able to generate new NN parameters at arbitrary salt concentrations. These interpolated parameters are tested against a validation set and are shown to perform better than predictions using salt corrections.

2. Materials and methods

2.1. Melting temperature dataset and notation

We used 18 RNA sequences measured at 4 different salt concentrations and 2 sequences at high salt concentration, where each sequence was measured at least for 9 different species concentrations, resulting in a total of 757 melting temperatures. This dataset was originally used to calculate the $T_m \times \log C_t$ regressions for the melting temperatures, published in Ref. [11], where the experimental details can be found. Following a similar notation as employed previously for DNA [9], we will refer to this dataset by the code D-RW, with RW standing for 'raw', meaning that these are the temperatures prior to the $T_m \times \log C_t$ regression. For each salt concentration we appended a number to the code representing the actual sodium

concentration in mM, that is, D-RW-[Na⁺]. For instance, the dataset at 121 mM will be referred to as D-RW-121. As the dataset for high salt concentration only contains 2 sequences, we complemented it with additional 109 sequences from Xia et al. [10]. Therefore, this extended dataset received the code D-RW-EXT1021 to highlight this fact. The melting temperatures for each sequence and concentration were obtained from a single measurement and an error of 0.5 °C is used in the our analysis. All these sequences and their respective melting temperatures are given in Supplementary Tables S1–S5.

For a validation set we collected 25 sequences and their melting temperatures at low salt concentrations and various species concentrations from Refs. [12–19], these are shown in Supplementary Table S6. This validation set, called D-RNAVL, was not used to calculate the nearest neighbour parameters reported in this work, but only to assess the overall prediction quality of the new parameters.

As we will compare our results with the calculated melting temperatures from Ref. [11], Table 1, we will employ the notation D-CH13-[Na⁺], and the temperatures obtained from curve fitting (CF), from the Supplementary files of Ref. [11], are referred to as D-CH13CF-[Na⁺].

Parameter codes are prefixed with P when retrieved from published sources, specifically, here we use P-XIA98 for the parameters published by Xia et al. [10]. Calculated parameters are prefixed with AOP (average optimized parameters) following the convention used in our work for DNA [9]. A third type of parameters is introduced here where we interpolate the AOP type parameters for a range of salt concentrations, these are prefixed with IP (interpolated parameters).

All datasets and parameters are available with the software package [9] and the files follow the same notation as used here.

2.2. Nearest-neighbour (NN) model

In the nearest-neighbour (NN) model the enthalpy and entropy variation of a given sequence, ΔH_i^{tot} and ΔS_i^{tot} , are subdivided into contributions of consecutive base-pairs

$$\Delta H_i^{\text{tot}} = \sum_{\alpha} n_{i\alpha} \Delta H_{\alpha} + \sum_c n_i^c \Delta H^c \quad (1)$$

$$\Delta S_i^{\text{tot}} = \sum_{\alpha} n_{i\alpha} \Delta S_{\alpha} + \sum_c n_i^c \Delta S^c \quad (2)$$

where $n_{i\alpha}$ are the occurrences of nearest-neighbour types α in the i th sequence. H^c and S^c are typically correction factors such as symmetry correction (S^{symm}) or terminal initiation (H^{term}).

The following examples illustrates the notation of nearest-neighbour types α

note that due to symmetry considerations CpA is equivalent to UpG. Here, given two equivalent possibilities, such as CpA and UpG, we will always use the one with alphabetical precedence for clarity, in this case this would be CpA.

Given the total entropy and enthalpy variations, ΔH_i^{tot} and ΔS_i^{tot} , for a sequence i the melting temperature can now be calculated from

$$T'_i = \frac{\Delta H_i^{\text{tot}}}{\Delta S_i^{\text{tot}} + R \ln \frac{C_t}{fC_0}} \quad (4)$$

where $f=1$ for self-complementary sequences and $f=4$ otherwise. C_t is the species concentration in μM and $C_0 = 1 \mu\text{M}$ is a reference concentration.

The Gibbs free energy parameters G_i at 37°C are given by

$$\Delta G_i(37^\circ\text{C}) = \Delta H_i - (273.15 + 37)\Delta S_i \quad (5)$$

2.3. Parameter optimization

The parameter optimization is carried out in several rounds of global minimizations, which in turn is composed of a certain amount of rounds of local minimizations. The aim of the method is to minimize the total squared difference between predicted T'_i and measured T_i

$$\chi^2 = \sum_i [T_i - T'_i(\{p\}_k)]^2 \quad (6)$$

where T_i is the measured temperature for the i th sequence. T'_i is the corresponding predicted temperature resulting from the set of tentative parameters

$$\{p\}_k = \{p_{k1}, p_{k2}, p_{k3}, \dots\}, \quad p = \Delta H_\alpha, \Delta S_\alpha, \Delta H^c, \Delta S^c \quad (7)$$

In addition we use the average prediction difference

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

as a simple comparative parameter.

2.3.1. Local minimizations with varying initial parameters (LMVIP)—We start with an initial set of parameters $\{p\}_0$ and let the parameters vary until we minimize as much as possible χ^2 . After L steps of minimizations the resulting set $\{p\}_L$ represent the parameters which best reproduce the melting temperatures for the model.

To minimize simultaneously all parameters we used the multi-dimensional Nelder-Mead or downhill simplex method [20] in a similar way as for the non-linear models [21]. Each local minimization was carried as follows:

1. define a set of initial parameters $\{p\}_0$ randomly selected within a range $\pm 20\%$ of the seed parameter. Here we used as seed parameters -30 cal/mol.K for entropies and -15 kcal/mol for enthalpies.
2. carry out the minimization with a characteristic length which should be roughly of the order of magnitude of the parameters as required by the downhill simplex method [20]. Here we used a value of 10 for the characteristic length.
3. stop the minimization either when the variation of χ^2 from one round to the next falls below 0.01 or when a limit of $L = 10^4$ steps is reached
4. use the newly obtained parameters $\{p\}$ as new initial parameters and repeat the whole procedure ten more times, starting from step 2.

2.3.2. Local minimizations with varying melting temperatures (LMVMT)—This type of minimization is similar to the previous one except that we always start with the same initial parameters but vary the experimental temperatures by random amounts. These random amounts are chosen such that the standard deviation between the original melting temperature set and the new perturbed set is 0.5 °C. This procedure allows us to estimate the error of our new parameters as a function of an experimental uncertainty.

2.4. Global minimization

The local minimizations, of type LMVIP, are repeated 2000 times, each time for a different set of initial parameters, and we record the average final parameters. We call the collective results of local minimizations the global minimization, that is, the attempt to arrive at the best overall solution. The reason for this procedure is that the minimization process mostly finds local minima and the algorithm runs into difficulty of getting away from such local minima. This is an inherent limitation of multi-dimensional parameter optimization and the typical strategy to circumvent this problem is to initiate the minimization from a variety of different parameters.

We perform a second round of global minimization using as input the initial parameters of the averaged results from the first round. However, this time using the local minimizations with varying melting temperatures (LMVMT). This is again performed 2000 times and we use the final averages over all these local minimizations as our final result. All minimizations are carried out independently for each salt concentration according to the procedure outlined in this section.

2.4.1. Initiation factor schemes—To evaluate the influence of the initiation parameters on the optimization we adopted two schemes. One in which these parameters change freely (variable initiation factors, VIF) and another where they are kept fixed at the same values as in Xia et al. [10] (fixed initiation factors, FIF). For both schemes, summarized in Table 1, the symmetry correction is kept constant. The reason for keeping S^{symm} fixed is that all sequences in the data set, except for D-RW-EXT1021, are self-complementary. This means that we have two entropy factors that apply to the whole dataset, S^{symm} and S^{init} , and we can not optimize both parameters simultaneously. As S^{symm} was defined on grounds of symmetry arguments [22] we have chosen to keep it constant and let only S^{init} vary freely.

2.5. Availability

All parameters, data sets and software used in this work, are freely available from.

- <https://download.opensuse.org/repositories/home:/drgweber/>
- <http://tinyurl.com/tfregufmg>
- via the GWDG-FTP service <https://tinyurl.com/yac7rkot>
- a web interface is available at <https://bioinf.fisica.ufmg.br/app/comparetm.pl>

3. Results

In this section we will present the new salt dependent parameters, after a brief discussion on some particularities of the datasets that are important for understanding the results. We focus on the question of how the initiation factors influence the resulting parameters and on the qualitative differences to the widely used P-XIA98 parameters. In the following section we will discuss the accuracy of the new parameters and discuss how they compare with the P-XIA98 parameters when used together with salt corrections factors.

The parameter optimization was carried out independently for each salt concentration, as described in the methods section. In this section we present the final averaged parameters and their estimated uncertainty for the two schemes using fixed (FIF) and variable (VIF) initiation factors, see Table 1. To understand the results presented here it is important to keep in mind that the high salt D-RW-EXT1021 dataset has some key differences compared to the lower-salt datasets. The datasets at lower salt concentrations (D-RW-71 to D-RW-621) are composed of 18 different sequences and for each of these sequences there are multiple melting temperatures at various species concentrations C_p , shown in Supplementary Tables S1–S4. However, at 1021 mM there are only two sequences with multiple melting temperatures which is not enough for calculating all parameters as these two sequences do

not cover all the NN configurations. Therefore, this dataset was extended with the sequences from Xia et al. [10] resulting in the dataset D-RW-EXT1021, shown in Supplementary Table S5. Unfortunately, there is only one melting temperature for each sequence from the dataset by Xia et al. [10]. As we will see, while the extended set D-RW-EXT1021 allowed us to perform the parameter calculations, some of the results do not correlate well with the parameters from the low-salt concentration data.

Fig. 1 shows the averaged optimized parameters for fixed initial factors (AOP-FIF). For clarity, figure panels were separated for strongly hydrogen bonded nearest neighbours [panels (a) and (d)], moderately hydrogen bonded [panels (b) and (e)] and weakly hydrogen bonded [panels (c) and (f)]. For strongly hydrogen bonded nearest neighbours we only observe an important salt concentration dependence for CpG, while CpG and CpC follow very closely the P-XIA98 parameters which are shown as dotted lines. It is interesting to note that these CG-type parameters are almost identical to the P-XIA98 parameters at the extreme values of salt concentration. The correspondence of these CG-type parameters to the P-XIA98 parameters, happened despite the fact that we started the minimization from completely arbitrary initial parameters. Also, note that the minimizations were carried out independently, always starting from generic parameters, for each salt concentration. In other words, the results at high salt concentrations have no influence over the parameters at low salt concentration. The lines shown in Fig. 1 are quadratic interpolations to the averaged parameters as a function of $\ln[\text{Na}^+]$; the regression coefficients are given in Supplementary Table S7. We will call these interpolated FIF-type parameters as IP-FIF and their accuracy will be discussed in the next section. The error bars are the standard deviation of the parameters resulting from the 2000 rounds of minimization of type LMVMT. For CG-type parameters these error bars are quite small, but they do increase considerably especially for weakly hydrogen bonded AU-type parameters as shown in Fig. 1c and f. The AU-type parameters are markedly different to the P-XIA98, especially at high salt concentration where a greater similarity would have been expected.

For variable initial factors (AOP-VIF), shown in Fig. 2, we observe a much more pronounced deviation from the P-XIA98 parameters. This indicates that the initiation factors have an important role in the final minimization results. In particular, this means that the good correspondence between the CG-type FIF parameters and the P-XIA98, as shown in Fig. 1, is influenced by the fact that we used the same initiation factors. Note that the initiation factors from P-XIA98 were used as seed parameters and therefore had an influence over the final AOPVIF parameters. Another interesting consequence of the initiation factors are the ApU parameters, Fig. 2c and g, which show a nearly linear dependence with $\ln[\text{Na}^+]$, in contrast to their FIF counterparts which are quadratic.

In Fig. 3 we show the Gibbs free energy parameters G_{37}° calculated from Eq. (5) for all FIF-type nearest neighbours. The enthalpies and entropies are clearly compensating each other and the resulting Gibbs free energy parameters are almost linear with salt concentration. Enthalpy-entropy compensation is to be expected as a strict correlation between these two physical quantities is enforced through Eq. (4) [23,24], yet a nearly salt-independent G_{37}° is a surprising outcome.

For the VIF-type parameters shown in Fig. 4 the enthalpy-entropy compensation is not quite so evident as for the FIF parameters, but still happens except for the 1021 mM G_{37}° parameters. Also the terminal-AU G_{37}° parameters are compensating as shown in Fig. 4d. The initiation parameters are markedly different for the 1021 mM salt concentration whereas for the lower salt concentration they vary over a narrower range of values. As we discussed at the beginning of this section, the 1021 mM dataset has some key differences to the datasets at lower salt temperatures, especially the fact that it is the only one that contains non-self-complementary sequences. This difference in the initiation parameters may have arisen as a compensation to the fixed symmetry factor. In our previous work on DNA salt dependence we have found that the symmetry factor does deviate from the generally accepted value of $S^{\text{symm}} = -1.4 \text{ cal}/(\text{mol K})$ [9] when this factor is allowed to vary.

4. Discussion

Since the primary purpose of the new parameter sets are the prediction of melting temperatures without the need for extra salt correction factors, let's briefly review some previous results from Ref. [11], which are the datasets of type D-CH13. Table 2 shows a breakdown of the accuracy for the best salt correction factor reported in Eq. (22) of Ref. [11]. For each salt concentration $[\text{Na}^+]_1$ we applied the salt correction for another salt concentration $[\text{Na}^+]_2$ and compared those with the measured T_m . For instance, if we apply a correction to the whole dataset D-CH13-71 to a salt concentration $[\text{Na}^+]$ of 1021 mM and compare it with the dataset D-CH13-1021 we obtain an average temperature difference of 0.80°C . On the other hand the inverse results in a slightly smaller difference of 0.76°C . In general, the average temperature difference $\langle \Delta T \rangle$ increases in proportion to the difference between the base concentration $[\text{Na}^+]_1$ and target concentration $[\text{Na}^+]_2$. While this may sound straightforward, providing a detailed quantification of this difference will be helpful in the following discussion.

In Table 3 we show the differences between the measured D-RW-type melting temperatures for the different parameters under discussion. For the original P-XIA98 parameters [10], the prediction accuracy decreases dramatically for lower salt concentrations if no salt correction is applied, which is of course expected. However, when salt correction is applied the merit factors improve considerably, especially when Eq. (22) from Ref. [11] is used. For the optimized AOP-FIF and AOP-VIF parameters a substantial reduction in the average temperature difference as well as in the squared difference was obtained. Overall, AOP-VIF parameters appear to be marginally better than their fixed initiation factor counterparts AOP-FIF. The interpolated parameters, IP-FIF and IP-VIF, perform similarly as the AOP-type parameters, which means that they are equally acceptable for performing melting temperature predictions. In general, we can say that both the AOP and IP parameters perform better than the combination of P-XIA98 and salt correction, except for the higher salt data set D-RW-EXT1021. For this extended set we only observe a reduction in the χ^2 merit factor. As discussed previously, the extended set D-RW-EXT1021 is very different from the lower salt sets for various reasons, it is the only one to contain non-self-complementary sequences and only two sequences are at multiple species concentrations. For the D-CH13 dataset, which is the published table from Ref. [11], Table 4, we observe a

similar reduction in the merit factors, here the average difference in melting temperatures reduces to similar values as in Table 2.

In Fig. 5a we show the scatter plots comparing the measured temperatures in dataset D-RW-121 to the predicted temperatures using P-XIA98 parameters and applying the best salt correction from [11]. Note that the deviations from the straight line are consistently larger towards lower temperatures. For the interpolated parameters IP-VIF the points are much closer to the straight line, Fig. 5b, as expected; yet the larger deviation for lower temperatures persists. This trend is observed for all predictions and all datasets, including the extended data set D-RW-EXT1021 which includes additional data by Xia et al. [10], shown in Supplementary Fig. S1. The cause of this trend is unclear and was not observed for DNA [9]. The scatter plots for all datasets are shown in Supplementary Figs. S1–S4.

So far the predictions of the optimised parameters are clear improvements over the existing scheme of using the P-XIA98 parameters together with salt correction factors. However, the results in Tables 3 and 4, are essentially self-predictions, that is, they are predicting the same data that was used to obtain the parameters. In Table 5 we show the predictions for D-CH13CF, which are melting temperatures obtained from curve fitting parameters which were not used for deriving the optimized parameters. Here again, we observe the same encouraging reductions of merit factors. However, while Table 5 is not strictly a self-prediction, the sequence compositions are the same as for the D-RW datasets. For this reason, we performed a cross-validation with 25 sequences from various sources [12–19], which were specifically selected for low salt concentration, with $[\text{Na}^+]$ ranging between 10 mM and 132 mM. The actual sequences and their experimental details are shown in Supplementary Table S6. When the original P-XIA98 parameters [10] were used together with the salt correction of Eq. (22) from Ref. [11], we obtained an average difference between measured and predicted temperatures of 4.60 °C. For the IP-FIF parameters we obtained value of 3.96 °C and for VIF this was further reduced to 3.62 °C. The reduction of squared differences χ^2 was even more pronounced, reducing from 728.44 °C² to 473.23 °C² for IP-VIF parameters.

5. Conclusions

We have obtained the optimized nearest-neighbour parameters for RNA duplexes as a function of salt concentration which do not require salt correction factors. We found an important influence of the initiation parameters that reduces the self-prediction uncertainty considerably. Cross-validation with an independent data set at low salt concentrations shows that the interpolated parameter scheme with variable initial parameters has the best overall accuracy for melting temperature prediction at arbitrary sodium concentrations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors acknowledge the agencies that funded this research: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brasil, Finance Code 001) and National Institutes of Health (NIH grant 2R15GM085699).

References

- [1]. Mathews DH, Sabina J, Zuker M, Turner DH, Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure, *J. Mol. Biol* 288 (5) (1999) 911–940. [PubMed: 10329189]
- [2]. Lu ZJ, Turner DH, Mathews DH, A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation, *Nucleic Acids Res* 34 (17) (2006) 4912–4924. [PubMed: 16982646]
- [3]. Costa PR, Acencio ML, Lemke N, Cooperative RNA polymerase molecules behavior on a stochastic sequence-dependent model for transcription elongation, *PLoS One* 8 (2) (2013) e57328. [PubMed: 23437369]
- [4]. Mathews DH, Burkard ME, Freier SM, Wyatt JR, Turner DH, Predicting oligonucleotide affinity to nucleic acid targets, *RNA* 5 (11) (1999) 1458–1469. [PubMed: 10580474]
- [5]. Joyeux M, Buyukdagli S, Dynamical model based on finite stacking enthalpies for homogeneous and inhomogeneous DNA thermal denaturation, *Phys. Rev. E* 72 (2005) 051902.
- [6]. Weber G, Finite enthalpy model parameters from DNA melting temperatures, *Europhys. Lett* 96 (2011) 68001, 10.1209/0295-5075/96/68001.
- [7]. Breslauer KJ, Frank R, Blocker H, Marky LA, Predicting DNA duplex stability from the Base sequence, *Proc. Natl. Acad. Sci. U.S.A* 83 (11) (1986) 3746–3750. [PubMed: 3459152]
- [8]. SantaLucia J Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. U.S.A* 95 (4) (1998) 1460–1465 URL:<http://www.pnas.org/cgi/content/abstract/95/4/1460>. [PubMed: 9465037]
- [9]. Weber G, Optimization method for obtaining nearest-neighbour DNA entropies and enthalpies directly from melting temperatures, *Bioinformatics* 31 (6) (2015) 871–877, 10.1093/bioinformatics/btu751 URL:<http://bioinformatics.oxfordjournals.org/content/31/6/871>. [PubMed: 25391397]
- [10]. Xia T, SantaLucia J Jr., Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH, Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs, *Biochemistry* 37 (1998) 14719–14735. [PubMed: 9778347]
- [11]. Chen Z, Znosko BM, Effect of sodium ions on RNA duplex stability, *Biochemistry* 52 (42) (2013) 7477–7485, 10.1021/bi4008275. [PubMed: 24106785]
- [12]. Walker GT, Deoxycytidine methylation does not affect DNA-RNA hybrid formation or B–A transitions of (dG)n.(dC)n sequences, *Nucleic Acids Res* 16 (7) (1988) 3091–3099. [PubMed: 3368317]
- [13]. Mikulecky PJ, Feig AL, Heat capacity changes associated with DNA duplex formation: salt- and sequence-dependent effects, *Biochemistry* 45 (2) (2006) 604–616. [PubMed: 16401089]
- [14]. Nakano S, Fujimoto M, Hara H, Sugimoto N, Nucleic acid duplex stability: influence of base composition on cation effects, *Nucleic Acids Res* 27 (14) (1999) 2957. [PubMed: 10390539]
- [15]. Pasternak A, Wengel J, Thermodynamics of RNA duplexes modified with unlocked nucleic acid nucleotides, *Nucleic Acids Res* 38 (19) (2010) 6697–6706, 10.1093/nar/gkq561. [PubMed: 20562222]
- [16]. Gyi JI, Conn GL, Lane AN, Brown T, Comparison of the thermodynamic stabilities and solution conformations of DNA-RNA hybrids containing purine-rich and pyrimidine-rich strands with DNA and RNA duplexes, *Biochemistry* 35 (38) (1996) 12538–12548. [PubMed: 8823191]
- [17]. Wang S, Kool ET, Origins of the large differences in stability of DNA and RNA helices: C-5 methyl and 2'-hydroxyl effects, *Biochemistry* 34 (12) (1995) 4125–4132. [PubMed: 7535100]

- [18]. Clanton-Arrowood K, McGurk J, Schroeder SJ, 3' terminal nucleotides determine thermodynamic stabilities of mismatches at the ends of RNA helices, *Biochemistry* 47 (50) (2008) 13418–13427. [PubMed: 19053257]
- [19]. Hara RI, Kageyama M, Arai K, Uchiyama N, Wada T, Synthesis of 2'-O-monohaloethoxymethyl-modified RNAs and their duplex formation ability, *RSC Adv* 7 (65) (2017) 41297–41303.
- [20]. Press WH, Teukolsky SA, Vetterling WT, Flannery BP, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1988.
- [21]. Weber G, Essex JW, Neylon C, Probing the microscopic flexibility of DNA from melting temperatures, *Nat. Phys* 5 (2009) 769–773, 10.1038/nphys1371.
- [22]. SantaLucia J Jr., Allawi HT, Seneviratne PA, Improved Nearest-Neighbour parameters for predicting DNA duplex stability, *Biochemistry* 35 (1996) 3555–3562. [PubMed: 8639506]
- [23]. Petruska J, Goodman MF, Enthalpy-entropy compensation in DNA melting thermodynamics, *J. Biol. Chem* 270 (2) (1995) 746–750. [PubMed: 7822305]
- [24]. Starikov EB, Norden B, DNA duplex length and salt concentration dependence of enthalpy-entropy compensation parameters for DNA melting, *J. Phys. Chem. B* 113 (33) (2009) 11375–11377. [PubMed: 19719257]

Fig. 1.

Averaged optimized parameters as a function of salt concentration for minimization scheme FIF. Top panels (a–c) show entropy variations and bottom panels (d–f) show enthalpy variations. Dotted lines show the equivalent parameters P-XIA98 [10] at high salt concentration. The curves are quadratic interpolations to the average data points representing the interpolated IP-FIF parameters. Their regression coefficients are given in Supplementary Table S7.

Fig. 2.

Averaged optimized parameters as a function of salt concentration for minimization scheme VIF. Top panels (a–d) show entropy variations and bottom panels (e–h) show enthalpy variations. Dotted lines show the equivalent parameters P-XIA98 [10] at high salt concentration. The curves are quadratic interpolations to the average data points representing the interpolated IP-VIF parameters. Their regression coefficients are given in Supplementary Table S8.

Fig. 3.
Gibbs free energy parameters at 37 °C as a function of salt concentration in minimization scheme FIF (dashed lines connecting the points are intended as guides to the eye).

Fig. 4. Gibbs free energy parameters at 37 °C as a function of salt concentration in minimization scheme VIF (dashed lines connecting the points are intended as guides to the eye). Note that panel (d) is at a different vertical scale (blue axis at right) than panels (a), (b).

Fig. 5. Scatter plots of the predicted and measured temperature for the dataset D-RW-71 with (a) P-XIA98 parameters together with salt correction Eq. (19) from [11] and (b) IP-VIF parameters.

Table 1

Initialization factors used in the optimisation. Enthalpies are given in kcal/ mol and entropies in cal/(mol K). Note that the symmetry factor was kept constant for both FIF and VIF as the datasets only contain self-complementary sequences.

	FIF	VIF
$H_{\text{AU}}^{\text{term}}$	3.72	Variable
$S_{\text{AU}}^{\text{term}}$	10.5	Variable
H^{init}	-1.5	Variable
S^{init}	3.61	Variable
S^{symm}	-1.4	-1.4

Table 2

Average temperature differences $\langle T \rangle$ merit factors χ^2 (in parenthesis) for the salt correction factors proposed in Ref. [11], Eq. (22). Base temperatures were taken from the datasets D-CH13-[Na +]₁ at the shown salt concentrations (first column) and recalculated with Eq. (22) from Ref. [11] and compared again to the measured temperatures of D-CH13-[Na +]₂ at the target salt concentrations.

[Na +] ₁	[Na +] ₂				
	71	121	221	621	1021
71	*	0.35 (6.13)	0.37 (5.47)	0.42 (7.27)	0.80 (15.6)
121	0.34 (5.93)	*	0.35 (3.59)	0.44 (5.36)	0.72 (18.1)
221	0.36 (5.13)	0.34 (3.48)	*	0.28 (2.75)	0.62 (12.2)
621	0.40 (6.54)	0.43 (4.99)	0.28 (2.65)	*	0.73 (15.3)
1021	0.76 (13.9)	0.69 (16.7)	0.60 (11.6)	0.73 (15.1)	*

Table 3

Average temperature differences $\langle T \rangle$ merit factors χ^2 (in parenthesis) for datasets D-RW-[Na⁺]. Note that the P-XIA98 parameters were obtained at 1021mM [10], therefore no salt correction was applied (table cells marked with *).

Parameters	Salt correction	D-RW-[Na ⁺]					
		71	121	221	621	1021	EXT1021
P-XIA98	none	8.9 (15300)	6.2 (7390)	3.8 (2940)	1.5 (658)	1.0 (73.2)	1.3 (417)
P-XIA98	Eq. (19) from [11]	1.5 (731)	1.3 (558)	1.4 (628)	1.3 (603)	*	*
P-XIA98	Eq. (22) from [11]	1.3 (585)	1.3 (524)	1.2 (469)	1.3 (532)	*	*
AOP-FIF	none	0.78 (279)	0.76 (182)	0.71 (158)	0.90 (256)	0.90 (60.7)	1.3 (357)
AOP-VIF	none	0.77 (277)	0.72 (176)	0.69 (155)	0.86 (250)	0.87 (56.0)	1.2 (336)
IP-FIF	none	0.79 (286)	0.77 (192)	0.76 (167)	0.93 (278)	0.84 (51.9)	1.3 (380)
IP-VIF	none	0.78 (285)	0.76 (186)	0.73 (167)	0.91 (276)	0.81 (48.1)	1.3 (403)

Table 4
Average temperature differences $\langle T \rangle$ merit factors χ^2 (in parenthesis) for datasets D-CHI3-[Na⁺].

Parameters	D-CHI3-[Na ⁺]				
	71	121	221	621	1021
P-XIA98 and Eq. (22)	1.1 (38.8)	1.2 (52.8)	1.2 (45.7)	1.1 (45.1)	1.4 (58.4)
AOP-FIF	0.55 (11.8)	0.66 (16.3)	0.57 (12.5)	0.68 (16.3)	1.3 (52.2)
AOP-VIF	0.55 (11.9)	0.63 (16.0)	0.54 (12.4)	0.65 (15.7)	1.3 (49.3)
IP-FIF	0.52 (11.7)	0.68 (17.5)	0.57 (13.2)	0.70 (19.0)	1.2 (44.7)
IP-VIF	0.55 (11.9)	0.68 (17.1)	0.57 (13.5)	0.68 (18.5)	1.2 (42.3)

Table 5

Average temperature differences $\langle T \rangle$ merit factors χ^2 (in parenthesis) for datasets D-CH13CF-[Na⁺].

Parameters	D-CH13CF-[Na ⁺]				
	71	121	221	621	1021
P-XIA98 and Eq. (22)	1.2 (40.5)	1.3 (50.4)	1.1 (41.8)	1.0 (37.1)	1.2 (44.1)
AOP-FIF	0.53 (11.9)	0.68 (16.1)	0.62 (13.5)	0.71 (16.3)	1.2 (39.7)
AOP-VIF	0.52 (11.8)	0.63 (15.7)	0.57 (13.0)	0.67 (15.4)	1.2 (37.3)
IP-FIF	0.53 (12.1)	0.71 (17.0)	0.66 (14.5)	0.68 (17.3)	1.1 (32.8)
IP-VIF	0.52 (11.6)	0.69 (16.7)	0.60 (14.4)	0.66 (16.6)	1.0 (30.8)