

# Nearest-neighbor parameters for the prediction of RNA duplex stability in diverse *in vitro* and cellular-like crowding conditions

Saptarshi Ghosh<sup>1</sup>, Shuntaro Takahashi<sup>1</sup>, Dipanwita Banerjee<sup>1</sup>, Tatsuya Ohyama<sup>1</sup>, Tamaki Endoh<sup>1</sup>, Hisae Tateishi-Karimata<sup>1</sup> and Naoki Sugimoto<sup>1,2,\*</sup>

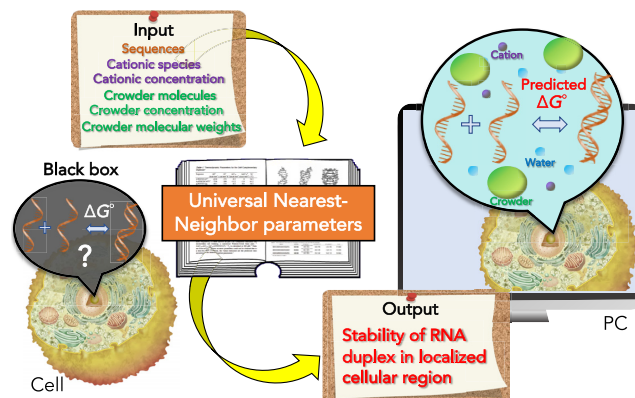
<sup>1</sup>Frontier Institute for Biomolecular Engineering Research (FIBER), Konan University, 7-1-20 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan and <sup>2</sup>Graduate School of Frontiers of Innovative Research in Science and Technology (FIRST), Konan University, 7-1-20 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan

Received November 24, 2022; Revised December 20, 2022; Editorial Decision January 03, 2023; Accepted January 11, 2023

## ABSTRACT

RNA performs various spatiotemporal functions in living cells. As the solution environments significantly affect the stability of RNA duplexes, a stability prediction of the RNA duplexes in diverse crowded conditions is required to understand and modulate gene expression in heterogeneously crowded intracellular conditions. Herein, we determined the nearest-neighbor (NN) parameters for RNA duplex formation when subjected to crowding conditions with an ionic concentration relevant to that found in cells. Determination of the individual contributions of excluded volume effect and water activity to each of the NN parameters in crowded environments enabled prediction of the thermodynamic parameters and their melting temperatures for plenty of tested RNA duplex formation *in vitro* and in cell with significant accuracy. The parameters reported herein will help predicting RNA duplex stability in different crowded environments, which will lead to an improved understanding of the stability-function relationship for RNAs in various cellular organelles with different molecular environments.

## GRAPHICAL ABSTRACT



## INTRODUCTION

RNA is one of the most important biomolecules present in cells and is associated with various biological processes, such as protein synthesis, biocatalysis, and enzymatic reactions (1,2). RNA can form various inter- or intramolecular Watson–Crick base-paired secondary structures sourced from the primary base sequences, and these structures often influence RNA functions. Among the possible RNA secondary structural motifs, the stem-loop is the most common, where the duplex stem is generally formed by Watson–Crick pairing between complementary bases. Such RNA secondary structures have been shown to affect the kinetics and yields of transcription and translation reactions significantly, depending on the stability of these structures (3–5). Furthermore, miRNA, which is a short RNA naturally transcribed in the cells, forms duplexes with mRNA via base pairings and regulates translation efficiency with an RNA-induced silencing complex (RISC) (6). From a biotechnological viewpoint, an siRNA that fully hybridizes with a

\*To whom correspondence should be addressed. Tel: +81 78 303 1147; Fax: +81 78 303 1495; Email: sugimoto@konan-u.ac.jp

target mRNA induces the silencing of gene expression (7). Moreover, RNA secondary structures have been widely utilized in the development of diagnosis, therapy, and targeted delivery (8). However, the determination of RNA secondary structures in cells is limited. Certain techniques can be used to probe the base pairings of RNA in cells using chemical footprinting (9), confocal microscopy (10), and in-cell NMR (11), although these techniques are often complicated and time-consuming. Therefore, computational prediction models are ideal and more achievable for studying RNA secondary structure formation in cells.

To predict the formation of an RNA base pair, the thermodynamic stability of Watson–Crick base pair formation is required. The most successful prediction method for duplex stability is based on the nearest-neighbor (NN) model, which assumes that the free energy change ( $\Delta G^\circ$ ) for the duplex formation comprises the sum of the  $\Delta G^\circ$  values of adjacent NN base pairs and helix initiation (12,13). Similarly, the relevant thermodynamic parameters, such as changes in enthalpy ( $\Delta H^\circ$ ) and entropy ( $\Delta S^\circ$ ), can also be predicted from the base sequence. Based on this model, parameters for predicting the thermodynamics of RNA, DNA and RNA/DNA hybrid duplex formations were developed (14–16). Due to the establishment of the NN parameters (14), RNA secondary structure prediction models such as Mfold (17), RNA fold (18), RNA structure (19) and the recently developed MXFold2 (20) and software for RNAs containing modified nucleobase (21) compare the minimum free energies for different possible conformations to obtain the most probable secondary structure. However, these methods have a critical limitation for their application in cells, because the classical NN parameters for RNA duplex formation were determined under conditions of 1 M NaCl (22), whereas the intracellular environment contains much lower and various cation concentrations (23). For example, cytosol consists not only monovalent cations like  $K^+$  and  $Na^+$  but also different divalent cations like  $Mg^{2+}$  and  $Ca^{2+}$  (23). Therefore, prediction parameters for RNA duplexes applicable for various cation conditions are required. Furthermore, intracellular environments remain populated with highly concentrated macromolecules termed ‘molecular crowding’ (24). As an example, nucleus contains nucleic acids and proteins such as histones and protamines as the main component of biomolecules, whereas, biomolecular composition in cytosol majorly consists of protein filaments, microtubules, vesicles and polysaccharides (23). Such molecular crowding in cell significantly affects nucleic acid structure and stability (25). To understand the behavior of nucleic acid structures in cells, synthetic cosolutes are often used to mimic intracellular crowding conditions *in vitro* whereby polyethylene glycols (PEGs) are predominantly used (26). Recently, we developed NN parameters for DNA duplexes available in various concentrations of cations and cosolutes (27). For RNA duplexes, however, NN parameters are only available in a specific crowding solution containing 1 M NaCl, limiting their applications for cellular conditions (28). Intracellular RNAs exist in various organelles, such as the cytosol, nucleus, and stress granules. Moreover, intracellular conditions vary dynamically. Therefore, a general method is required to utilize these parameters in varying concentrations of a cosolute, in the presence of

different cosolutes, and under physiological cation concentrations.

In this study, we analyzed the thermodynamics of formation of different RNA duplexes with varying lengths and base compositions in a physiologically relevant solution containing 100 mM NaCl and 40 wt% PEG with an average molecular weight of 200 g mol<sup>−1</sup> (PEG200). Based on these data, we determined the NN parameters of 10 possible propagating Watson–Crick NN base pairs and helix initiation factor for RNA duplexes, which predicted stability of RNA duplexes with high accuracy even in a solution containing exact intracellular cation composition. Furthermore, we expanded these NN parameters available under different cosolutes at varying concentrations by determining the contributions of the excluded volume effect and water activity under the cosolute-containing solutions on duplex stability, which permitted to predict stabilities of different RNA duplexes in diverse *in vitro* and intracellular crowding conditions with significant accuracy. The universal NN parameters applicable to different crowding conditions will be not only useful in RNA structure determination software when attempting to improve our understanding of RNA structures *in vivo*, but also in designing RNA-based therapeutics.

## MATERIALS AND METHODS

### Materials

All the synthetic RNA oligonucleotides used in this work, listed in Table S1 (in Supplementary Data), were purchased from Japan Bio Services Co. and purified by high-performance liquid chromatography (HPLC). Dialysis could be done to set the exact ionic concentrations of the studied solutions, however, we did not perform dialysis, except otherwise mentioned. This is due to the fact that we used quite short length RNAs with low concentrations for this study. Thus, even if there were some residual cations on RNAs after synthesis and purification, the effect on the stability should be negligible. RNA samples were prepared in Milli-Q water and stocked at  $-20^\circ\text{C}$  until use. The concentrations of the single-stranded oligonucleotides were determined by measuring the absorbance at 260 nm at  $90^\circ\text{C}$  using the extinction coefficients (29). Polyethylene glycols (PEGs), glycerol, ethylene glycol, 2-methoxy ethanol, 1,2-dimethoxy ethane and 1,3-propanediol were purchased from Wako Pure Chemical Industries, Japan and used without further purification. Disodium hydrogen phosphate ( $\text{Na}_2\text{HPO}_4$ ), Dipotassium hydrogen phosphate ( $\text{K}_2\text{HPO}_4$ ), sodium chloride (NaCl), potassium chloride (KCl), magnesium chloride ( $\text{MgCl}_2$ ) and calcium chloride ( $\text{CaCl}_2$ ) were purchased from Wako Pure Chemical Industries (Japan); disodium ethylenediaminetetraacetate ( $\text{Na}_2\text{EDTA}$ ) and dipotassium ethylenediaminetetraacetate ( $\text{K}_2\text{EDTA}$ ) were purchased from Dojindo Molecular Technologies (Japan) and all these chemicals were used as received.

### UV melting measurement

Absorption spectra were measured on a Shimadzu 1800 spectrophotometer with a thermoprogrammer. All the ex-

periments were conducted in a buffer containing 10 mM  $\text{Na}_2\text{HPO}_4$ , 1 mM  $\text{Na}_2\text{EDTA}$  and 100 mM  $\text{NaCl}$ , in the presence of cosolutes with specific weight percentages. We adjusted the pH of the buffer to 7.0 after adding the cosolutes to maintain the pH of the buffer solution. For melting experiments, 10–12 fresh solutions of oligonucleotides were prepared by varying the concentrations over a 50–150-fold range. The RNA solutions were kept at 90°C for 5 min, followed by the decrease of temperature from 90°C to 0°C at a rate of 1°C min<sup>-1</sup> to anneal the duplexes. Thereafter, the samples were heated from 0°C to 90°C at a rate of 0.5°C min<sup>-1</sup> to melt the duplex after keeping them at 0°C for 5 min. Water condensation on the cuvette exterior at low temperature was avoided by flushing with a constant stream of dry  $\text{N}_2$  gas.

### Determination of thermodynamics for duplex formation

Thermodynamic parameters ( $\Delta H^\circ$ ,  $\Delta S^\circ$  and  $\Delta G^\circ_{37}$ ) for RNA duplexes were determined from the  $T_m^{-1}$  versus  $\ln(C_t/s)$  plots as described in our earlier studies (27,30,31). From the slope and intercept of the linear plots, the thermodynamic parameters were calculated using the following equations:

$$T_m^{-1} = R \ln(C_t/s) / \Delta H^\circ + \Delta S^\circ / \Delta H^\circ \quad (1)$$

$$\Delta G^\circ_{37} = \Delta H^\circ - 310.15 \bullet \Delta S^\circ \quad (2)$$

where  $R$  is the gas constant,  $C_t$  is the total strand concentration of the oligonucleotides, and  $s$  reflects the sequence symmetry of the self-complementary ( $s = 1$ ) or non-self-complementary strands ( $s = 4$ ). Following the standard practice for calculation of the thermodynamic parameters, we assumed the difference in heat capacities ( $\Delta C_p$ ) of the two states (single-strand and duplex) to be zero (14,22). Because the melting temperature ( $T_m$ ) values for most of the studied sequences in the crowding condition were not far from 37°C, zero  $\Delta C_p$  approximation should be acceptable for  $\Delta G^\circ_{37}$  calculation due to minimal extrapolations. Generally,  $\Delta G^\circ_{37}$  is relatively insensitive to  $\Delta C_p$  due to enthalpy–entropy compensation (22,32).

### Calculation of nearest-neighbor parameters

NN parameters were calculated using our original software written in Python based on the linear least square fitting algorithm as described in our earlier studies (27,31). Briefly, the program determined  $\Delta G^\circ_{37}$  and  $\Delta H^\circ$  using a set of 13 parameters (ten Watson-Crick NN base pairs, one initiation parameter, a parameter for terminal AU pair and a symmetry parameter for self-complementary sequences) by applying linear least square fitting to the measured values of the thermodynamic parameters of 45 RNA duplexes in 40 wt% PEG200 with 100 mM  $\text{NaCl}$ . The parameters for symmetry correction were set to the same value as that in the absence of cosolutes, since it does not depend on the solution condition, but on the complementarity of the sequences.  $\Delta S^\circ$  values were obtained from the determined respective values of  $\Delta G^\circ_{37}$  and  $\Delta H^\circ$  using Equation 2.

### Circular dichroism (CD) measurements

CD spectra were obtained on a JASCO J-1500 spectropolarimeter equipped with a temperature controller. The experimental temperature was 4°C. The cuvette-holding chamber was flushed with a constant stream of dry  $\text{N}_2$  gas to avoid water condensation on the cuvette exterior. The CD spectra were measured from 200 to 340 nm in 1.0 mm path-length cuvettes with a scan rate of 50 nm min<sup>-1</sup>. The concentration of the samples was 20  $\mu\text{M}$  in a buffer containing 100 mM  $\text{NaCl}$ , 10 mM  $\text{Na}_2\text{HPO}_4$  (pH 7.0) and 1 mM  $\text{Na}_2\text{EDTA}$  with or without 40 wt% PEG200.

### Water activity and dielectric constant measurement

The water activities were determined by vapor phase osmometry (a Wescor pressure osmometer) or freezing point depression osmometry (a Knauer osmometer). The water activity ( $a_w$ ) was calculated from the measured osmomolality (mmol kg<sup>-1</sup>) using the following equation (33)

$$\Psi = (RT/M_w) \ln a_w \quad (3)$$

where  $\Psi$  and  $M_w$  represent the water potential and the molecular weight of water (0.018 kg mol<sup>-1</sup>), respectively. The relationship between water potential and osmomolality is given by the following equation (33)

$$\Psi \text{ (MPa)} = \text{osmomolality (mmol kg}^{-1}\text{)} \times 10^3 / (-400) \quad (4)$$

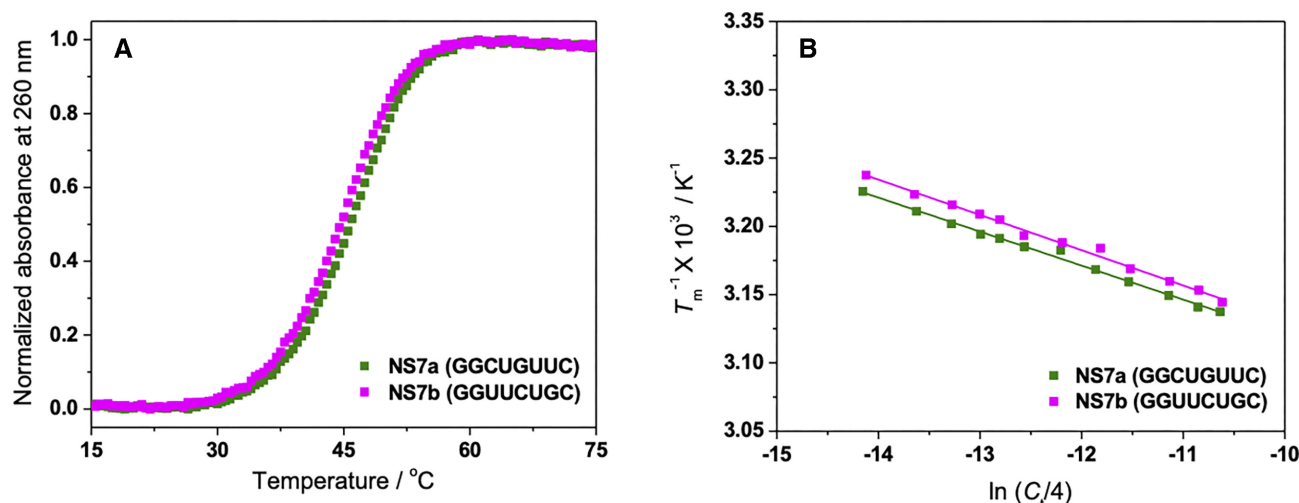
The dielectric constants were measured by using fluorescent probe ANS. The fluorescence maximum of ANS showed a hypsochromic shift in the media of low dielectricity. The spectra were recorded in fluorescence spectrophotometer (Jasco F-6500) exciting at 370 nm. The dielectric constants were calculated using a standard curve of organic solvents with known values.

## RESULTS AND DISCUSSION

### Validation of the NN model for RNA duplexes under physiological crowding condition

We selected RNA sequences ranging between 6- and 12-mers, ensuring sufficient stability and two-state transitions, since short sequences are too unstable to analyze, and longer sequences may show non-two-state transitions by forming intramolecular structures. The chosen sequences had different combinations of NN frequencies, covering all 10 NN base pairs for RNA duplex formation, without any bias to particular pairs (Table S1). We selected 25 non-self-complementary (NS1–NS16) and 20 self-complementary (S1–S15) RNA duplexes, as shown in Table S1. Among the duplexes investigated, nine pairs of non-self-complementary sequences (NS7–NS15) and five pairs of self-complementary sequences (S9–S12 and S15) had identical NN base sets (marked as ‘a’ and ‘b’ in Table S1). CD spectral data of representative RNA duplexes indicated that these sequences adopted the A-form duplex both in the absence or presence of PEG200 (Supplementary Figure S1). The slight changes in the peak positions and ellipticities in the PEG200 solution (Supplementary Figure S1) can be attributed to the different stabilities of the RNA duplexes under crowded conditions.





**Figure 1.** Representative UV melting curves of 100  $\mu$ M NS7a and NS7b in 40 wt% PEG200 with 100 mM NaCl (A) and  $T_m^{-1}$  versus  $\ln(C_i/4)$  plots for these sequences (B).

Thermodynamic data ( $\Delta H^\circ$ ,  $\Delta S^\circ$  and  $\Delta G^\circ_{37}$ ) of RNA duplex formation for the studied sequences under physiological crowding condition containing 40 wt% PEG200 with 100 mM NaCl were determined from the UV melting experiment, as depicted in Figure 1, for a pair of typical sequences. Absence of hysteresis between the denaturation and renaturation profiles for all sequences, as shown in Supplementary Figure S2 for representative sequences, indicated that the transition between duplex and single-strands was of a two-state nature. The obtained parameters are presented in Table S2, together with their stabilities in the absence of a cosolute. It should be noted that the actual  $\text{Na}^+$  concentration of the experimental solution was 122 mM considering the  $\text{Na}^+$  derived from the buffer components (10 mM  $\text{Na}_2\text{HPO}_4$  and 1 mM  $\text{Na}_2\text{EDTA}$ ) in addition to 100 mM NaCl. Consistent with previous reports, the stability of all the RNA duplexes decreased under crowding condition due to changes in the solution properties such as water activity and the dielectric constant of the medium in the presence of a cosolute (34,35). The extent of destabilization of RNA duplexes in the cosolute solution varied depending on the base sequences of the duplexes. Moreover, RNA sequences with identical NN sets showed similar melting profiles and  $T_m^{-1}$  versus  $\ln(C_i/4)$  plots under physiological crowding condition (Figure 1). As a result, the thermodynamic parameters obtained for the pairs of sequences with the same NN sets were also similar (Table S2). For example, in the case of sequence NS7a (GGCUGUUC) and NS7b (GGUUCUGC), the thermodynamic values were  $-79.8$  and  $-76.9$  kcal mol $^{-1}$  for  $\Delta H^\circ$ ,  $-71.1$  and  $-68.5$  kcal mol $^{-1}$  for  $T\Delta S^\circ$ , and  $-8.7$  and  $-8.4$  kcal mol $^{-1}$  for  $\Delta G^\circ_{37}$ , respectively. For the 14 pairs of RNA sequences with the same NN base sets (sequences indicated by ‘a’ and ‘b’ in Table S2), the average differences in the percentage of measured  $\Delta H^\circ$ ,  $T\Delta S^\circ$  and  $\Delta G^\circ_{37}$  were 5.2%, 5.8% and 4.0%, respectively. Since the NN model assumes the same stability for sequences with identical NN pairs (36), the results clearly showed the validity of the model for RNA duplexes under physiological crowding condition and the dependency of in-

dividual stability of the duplexes on the NN pairs present in that particular sequence.

#### NN parameters for RNA duplexes in the physiological crowding condition

The parameters for 10 possible Watson–Crick NN pairs ( $\Delta H^\circ_{\text{NN}}$ ,  $\Delta S^\circ_{\text{NN}}$  and  $\Delta G^\circ_{37, \text{NN}}$ ) and the penalties for helix initiation and terminal A–U pairs were calculated using a linear least-square fitting algorithm from the thermodynamic parameters of 45 RNA sequences (Table S2) (27). Table 1 lists the parameters determined in the presence of 40 wt% PEG200 at 100 mM NaCl. The qualitative trend observed in the order of increasing stability ( $-\Delta G^\circ_{37, \text{NN}}$ ) was:  $r(\text{AU}/\text{UA}) \approx r(\text{AA}/\text{UU}) < r(\text{UA}/\text{AU}) < r(\text{CU}/\text{GA}) \approx r(\text{GU}/\text{CA}) < r(\text{CA}/\text{GU}) \approx r(\text{CG}/\text{GC}) \approx r(\text{GA}/\text{CU}) < r(\text{GG}/\text{CC}) < r(\text{GC}/\text{CG})$ . The quantitative effect of molecular crowding on the individual NN pairs was estimated by subtracting parameters in the absence of a cosolute of the same cation concentration (37) (Table S3) from the obtained NN parameters in Table 1 and were shown in Table S4. As revealed by Table S4, some of the NN pairs were destabilized by the crowding, whereas others were minimally affected or conversely stabilized. This was likely due to the different states of hydration associated with each NN pair. Cosolutes disrupt the hydration network around nucleic acids and destabilize them (38). Each base pair exhibits variation among the helicities and geometries of donors and acceptors for hydrogen bonding, inducing different levels of hydration around the grooves for each NN pair (39). For four NN pairs [ $r(\text{AA}/\text{UU})$ ,  $r(\text{GU}/\text{CA})$ ,  $r(\text{GC}/\text{CG})$  and  $r(\text{GG}/\text{CC})$ ], associated water molecules may be available to interact with cosolutes, which would result in the reduction in the stability of the pairs. However, for five pairs [ $r(\text{AU}/\text{UA})$ ,  $r(\text{UA}/\text{AU})$ ,  $r(\text{CU}/\text{GA})$ ,  $r(\text{GA}/\text{CU})$  and  $r(\text{CG}/\text{GC})$ ], water molecules may be buried in deeper grooves, making them unavailable to interact with cosolutes and causing a nominal decrease in stability. For,  $r(\text{CA}/\text{GU})$  pair cosolutes may be assisting to maintain the hydration

**Table 1.** Nearest-neighbor parameters for RNA duplexes in the presence of 40 wt% PEG200 and 100 mM NaCl at 37°C<sup>a</sup>

Sequence	$\Delta H^{\circ}_{NN}$ (kcal mol <sup>-1</sup> )	$\Delta S^{\circ}_{NN}$ (cal mol <sup>-1</sup> K <sup>-1</sup> )	$\Delta G^{\circ}_{37, NN}$ (kcal mol <sup>-1</sup> )
r(AA/UU)	-10.0 ± 0.1	-30.4 ± 0.2	-0.57 ± 0.05
r(AU/UA)	-10.1 ± 0.1	-30.8 ± 0.1	-0.55 ± 0.03
r(UA/AU)	-11.1 ± 0.4	-31.5 ± 0.8	-1.33 ± 0.09
r(CA/GU)	-12.1 ± 0.3	-32.1 ± 0.6	-2.14 ± 0.08
r(GU/CA)	-10.7 ± 0.2	-28.7 ± 0.1	-1.80 ± 0.16
r(CU/GA)	-11.2 ± 0.3	-30.4 ± 0.3	-1.77 ± 0.17
r(GA/CU)	-11.7 ± 0.2	-30.5 ± 0.3	-2.24 ± 0.07
r(CG/GC)	-11.1 ± 0.5	-28.8 ± 1.1	-2.16 ± 0.16
r(GC/CG)	-13.8 ± 0.1	-34.6 ± 0.1	-3.07 ± 0.01
r(GG/CC)	-14.8 ± 0.1	-38.4 ± 0.1	-2.89 ± 0.06
Initiation	4.6 ± 2.0	-2.9 ± 6.1	5.50 ± 0.11
Per terminal AU	6.5 ± 0.1	18.2 ± 0.1	0.85 ± 0.92
Self-complementary	0	-1.4	0.43
Non-self-complementary	0	0	0

<sup>a</sup>Experiments were performed in 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1 mM Na<sub>2</sub>EDTA, 100 mM NaCl and 40 wt% PEG200 at pH 7.0.

via H-bonding interaction, resulting in stabilization of the pair dominated by favorable enthalpy contribution.

The effect of crowding was significant for helix initiation and terminal A–U pairs. Helix initiation and terminal A–U pairs were destabilized by 1.41 and 0.40 kcal mol<sup>-1</sup>, respectively, in the 40 wt% PEG200 solution (Table S4). This suggests that the formation of very short RNA duplexes under conditions of molecular crowding may be hindered due to the large unfavorable energy required for forming the first base pair in the duplex, while several NN pairs may be needed to compensate for this energy. The hydration around the terminal of a duplex is different from that of the stem as water molecules around the terminals are more accessible, resulting in greater disruption by the crowders. Therefore, the relative effect of crowders on NN pairs depends on the hydration around the particular NN pair.

The average differences in  $\Delta H^{\circ}$ ,  $\Delta S^{\circ}$ ,  $\Delta G^{\circ}_{37}$  and  $T_m$  between the predicted and measured values among the 45 studied RNA duplexes were 5.0%, 5.5%, 3.5% and 1.1°C, respectively (Table S5), which were well within the error limits reported in previous studies (22,28). Therefore, our derived parameters can be used to precisely predict the stability and thermodynamics of RNA duplex formation under physiological crowding condition.

### Development of generalized NN parameters for diverse crowding conditions

The molecular environments in cell vary spatiotemporally with the cell cycle, leading to different crowding densities (40). Moreover, cation concentration changes dynamically in different cellular compartments. Therefore, crowding conditions in different regions of a cell cannot be represented by a single cosolute at a particular cation concentration. Thus, to develop a generalized NN parameter set applicable to different cation and cosolute conditions, we separated each NN pair ( $\Delta G^{\circ}_{NN}$ ) into the individual contributions of the bulk interactions ( $\Delta G^{\circ}_{NN, bulk}$ ) and environmental factors ( $\Delta G^{\circ}_{NN, env}$ ) comprising of the energetic contribution of the cation ( $\Delta G^{\circ}_{NN, cation}$ ) and crowder ( $\Delta G^{\circ}_{NN, crowder}$ ). Therefore, each NN pair ( $\Delta G^{\circ}_{NN}$ ) under

crowding conditions can be expressed as follows:

$$\begin{aligned}\Delta G^{\circ}_{NN} &= \Delta G^{\circ}_{NN, bulk} + \Delta G^{\circ}_{NN, env} \\ &= \Delta G^{\circ}_{NN, bulk} + \Delta G^{\circ}_{NN, cation} + \Delta G^{\circ}_{NN, crowder}\end{aligned}\quad (5)$$

The bulk interactions consist of hydrogen bonding in base pairs, stacking interactions between neighboring base pairs, and conformational entropic penalty to bring two single strands into a helical structure (39). According to Equation 5,  $\Delta G^{\circ}_{NN}$  is equal to  $\Delta G^{\circ}_{NN, bulk}$  in the absence of contributions from  $\Delta G^{\circ}_{NN, cation}$  and  $\Delta G^{\circ}_{NN, crowder}$ . In the absence of a crowder,  $\Delta G^{\circ}_{NN}$  showed exponential decay with increasing concentration of Na<sup>+</sup> (Supplementary Figure S3). However, NN pairs with only A–U pairs deviated relatively more from the fitted line because of larger fluctuations in their reported  $\Delta G^{\circ}_{NN}$  values at different Na<sup>+</sup> concentrations (Supplementary Figure S3) (37). Theoretically  $\Delta G^{\circ}_{NN, bulk}$  should be calculated by extrapolating the individual  $\Delta G^{\circ}_{NN}$  to zero Na<sup>+</sup> concentration in the absence of a crowder, however, duplex formation without any cation is practically unrealistic. We found that stable oligomeric RNA duplexes with analyzable melting profile can be formed at minimum of 10 mM Na<sup>+</sup> concentration (Supplementary Figure S4). Thus,  $\Delta G^{\circ}_{NN}$  values at 10 mM Na<sup>+</sup> were approximated as  $\Delta G^{\circ}_{NN, bulk}$  for individual NN pairs and listed in Table 2, although numerical values at 10 mM Na<sup>+</sup> did not change much compared to the values at zero Na<sup>+</sup> concentration (Supplementary Figure S3). The  $\Delta G^{\circ}_{NN, cation}$  values can be obtained by subtracting  $\Delta G^{\circ}_{NN, bulk}$  from the reported  $\Delta G^{\circ}_{NN}$  for different cation concentrations in the absence of a crowder, as shown in Table 2, for 122 mM Na<sup>+</sup> (37). Because  $\Delta G^{\circ}_{NN}$  for RNA duplexes is available only for Na<sup>+</sup> (37), we can determine the  $\Delta G^{\circ}_{NN, cation}$  for different Na<sup>+</sup> concentrations. However, for other cations such as K<sup>+</sup> and Mg<sup>2+</sup>, some corrections are required to compensate for the differential effects from the Na<sup>+</sup> parameters (see later sections).

### Calculation of the excluded volume effect by the crowders

For  $\Delta G^{\circ}_{NN, crowder}$ , individual contributions from the excluded volume effect ( $\Delta G^{\circ}_{NN, ev}$ ) and water activity ( $\Delta G^{\circ}_{NN, wa}$ ) should be quantified since those are the dominant physical and chemical factors, respectively, for the duplex stability in cosolute solutions (25,27,41). By summing the contributions from the physicochemical properties,  $\Delta G^{\circ}_{NN, crowder}$  can be represented as follows:

$$\Delta G^{\circ}_{NN, crowder} = \Delta G^{\circ}_{NN, ev} + \Delta G^{\circ}_{NN, wa} \quad (6)$$

Replacing  $\Delta G^{\circ}_{NN, crowder}$  in Equation 5 by Equation 6 results in the following relation:

$$\Delta G^{\circ}_{NN} = \Delta G^{\circ}_{NN, bulk} + \Delta G^{\circ}_{NN, cation} + \Delta G^{\circ}_{NN, ev} + \Delta G^{\circ}_{NN, wa} \quad (7)$$

The volume excluded by crowders influences the stability of nucleic acids by shifting the equilibrium towards the formation of more compact structures with reduced volume (42). As an example, duplex formation is stabilized in the presence of larger biomolecules and cosolutes (43,44). The excluded volume effect is purely a physical interaction and has no dependence on the chemical nature of the cosolute or nucleic acid, supporting the additive nature of Equation 6.

**Table 2.** Parameters for  $\Delta G^{\circ}_{37, \text{NN, bulk}}$ ,  $\Delta G^{\circ}_{37, \text{NN, [122 mM Na}^+ \text{]}}$ ,  $\Delta G^{\circ}_{37, \text{NN, ev}}$  and  $\Delta G^{\circ}_{37, \text{NN, wa}}$  in 40 wt% PEG200 along with prefactors ( $m_{\text{cs}}$ ) for different cosolutes<sup>a</sup>

Sequence	$\Delta G^{\circ}_{37, \text{NN, bulk}}$ (kcal mol <sup>-1</sup> )	$\Delta G^{\circ}_{37, \text{NN, [122 mM Na}^+ \text{]}}$ <sup>b</sup> (kcal mol <sup>-1</sup> )	$\Delta G^{\circ}_{37, \text{NN, ev [40 wt% PEG200]}}$ <sup>c</sup> (kcal mol <sup>-1</sup> )	$\Delta G^{\circ}_{37, \text{NN, wa [40 wt% PEG200]}}$ (kcal mol <sup>-1</sup> )	$m_{\text{PEG/2-ME/1,2 DME}}$ (kcal mol <sup>-1</sup> )	$m_{\text{EG/Gly/1,3 PDO}}$ (kcal mol <sup>-1</sup> )
r(AA/UU)	-0.70	-0.00	-0.22	0.35	7.1	2.9
r(AU/UA)	-0.22	-0.30	-0.22	0.19	3.9	1.6
r(UA/AU)	-1.22	-0.08	-0.22	0.19	3.9	1.6
r(CA/GU)	-1.69	-0.09	-0.22	-0.14	-2.9	-1.2
r(GU/CA)	-1.73	-0.41	-0.22	0.56	11.4	4.7
r(CU/GA)	-1.52	-0.28	-0.22	0.25	5.1	2.1
r(GA/CU)	-2.06	-0.16	-0.22	0.20	4.1	1.7
r(CG/GC)	-2.00	-0.18	-0.22	0.24	4.9	2.0
r(GC/CG)	-2.72	-0.58	-0.22	0.45	9.2	3.8
r(GG/CC)	-2.92	-0.18	-0.22	0.43	8.8	3.6
Initiation	4.09	0.0 <sup>d</sup>	-0.22	1.63	33.3	13.7
Per terminal AU	0.45	0.0 <sup>d</sup>	NA <sup>e</sup>	0.40	8.2	3.4

<sup>a</sup>Correction factor for self-complementary sequences is the same (0.43 kcal mol<sup>-1</sup>) for all cosolutes because it is independent of the solution condition. <sup>b</sup> $\Delta G^{\circ}_{37, \text{NN, cat}}$  for any concentration of Na<sup>+</sup> can be obtained from the Na<sup>+</sup> dependency of individual NN pairs, as mentioned in Supplementary Figure S3. <sup>c</sup>Averaged over all 45 sequences used in this study. <sup>d</sup>Terminal pairs were assumed to be independent of Na<sup>+</sup> concentration following the report by Huguet *et al.* (47) and thus, cation-induced stabilization was not added to terminal pairs. <sup>e</sup>Excluded volume effect for terminal AU pairs was not considered to avoid overestimation of the effect, as it was already considered for initiation.

Excluded volume depends on length of the sequence as well as size of the cosolute. In the presence of significantly higher concentration of cosolute compared to nucleic acid, contribution of excluded volume to duplex stability ( $\Delta G^{\circ}_{\text{ev, dup}}$ ) was determined using the following equation: (45)

$$\Delta G^{\circ}_{\text{ev, dup}} = RT \bullet \rho \bullet \Delta V \bullet C \tag{8}$$

where  $R$  is the gas constant in kcal mol<sup>-1</sup> K<sup>-1</sup>,  $T$  is the temperature in kelvin (K),  $\rho$  is the density of water in kg L<sup>-1</sup>,  $\Delta V$  is the change in the excluded volume between the nucleic acid and the cosolute upon duplex formation in L mol<sup>-1</sup>, and  $C$  is the molal concentration of the cosolute in mol kg<sup>-1</sup> (45).  $\Delta V$  for duplex formation can be represented as follows:

$$\Delta V = V_{\text{dup}} - 2V_{\text{ss}} \tag{9}$$

where  $V_{\text{dup}}$  and  $V_{\text{ss}}$  are the volumes excluded by the duplex and single strands, respectively.  $V$  was determined from the following relation, considering RNAs as cylinders and cosolutes as polymeric chains (46);

$$V = N \bullet l^2 \bullet N_A \bullet k \bullet 10^{-27} \tag{10}$$

where  $N$  is the number of monomers present in the cosolute,  $l$  is the statistical segment length of the polymer in Å,  $N_A$  is the Avogadro's number (6.022 × 10<sup>23</sup> mol<sup>-1</sup>), and  $k$  is a geometrical factor derived from the length and radius of the RNA in Å (Table S6). The factor 10<sup>-27</sup> is used for the conversion of Å<sup>3</sup> to L. Using the parameters listed in Table S6, we calculated  $V/N$  for different lengths of RNA in PEG using Equation 10 (Table S6). From the values of  $V/N$  in Table S6,  $\Delta V/N$  and  $\Delta G^{\circ}_{37, \text{ev, dup}}/C \bullet N$  were calculated using Equations 9 and 8, respectively, for RNA-PEG interactions for different lengths of RNA duplexes (Table S7).  $\Delta G^{\circ}_{37, \text{ev, dup}}/C \bullet N$  decreased as the length of the RNA increased (Figure 2A), and showed a good linear correlation with the length of the duplex, producing the following equation:

$$\Delta G^{\circ}_{37, \text{ev, dup}}/C \bullet N = -0.019 \bullet n - 0.070 \tag{11}$$

where  $n$  denotes the number of base pairs. The strong linear correlation (correlation coefficient of 0.999) suggested

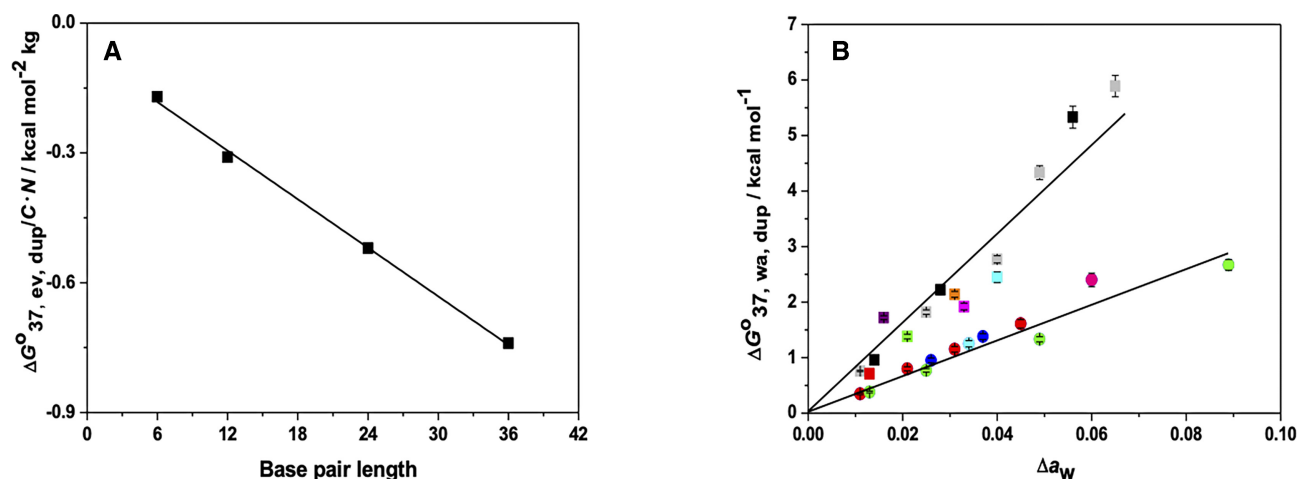
that Equation 11 can be used to calculate  $\Delta G^{\circ}_{37, \text{ev, dup}}$  for any given length of RNA duplex and cosolute concentration. Assuming that all NN pairs have the same volume effect, the contribution of the excluded volume effect for each NN pair and initiation factor ( $\Delta G^{\circ}_{37, \text{NN, ev}}$ ) was determined by the relation below (Equation 12). The values for 40 wt% PEG200 are shown in Table 2.

$$\Delta G^{\circ}_{37, \text{NN, ev}} = \Delta G^{\circ}_{37, \text{ev, dup}}/n \tag{12}$$

**Determination of the contribution of water activity of the cosolute solutions**

To explore the relationship between the RNA duplex stability and the water activity of the solution, we measured the stability of a test RNA sequence in solutions with different water activities as obtained by different cosolutes at varying concentrations (Table S8). We selected GAUUACGC-CUG as a test sequence because it contained all 10 NN base pairs with a single occurrence, which eliminated bias toward any particular NN pair. The contribution of water activity to the duplex ( $\Delta G^{\circ}_{37, \text{wa, dup}}$ ) for the test sequence in different cosolute solutions was calculated by subtracting the stability in the absence of cosolute (equivalent to  $\Delta G^{\circ}_{37, \text{bulk, dup}} + \Delta G^{\circ}_{37, \text{cation, dup}}$ ) and  $\Delta G^{\circ}_{37, \text{ev, dup}}$  from the measured  $\Delta G^{\circ}_{37}$  (Table S8). For example, in 10 wt% EG solution at 100 mM NaCl,  $\Delta G^{\circ}_{37, \text{wa, dup}}$  for GAUUACGC-CUG was calculated to be 0.95 kcal mol<sup>-1</sup> by as follows:  $\Delta G^{\circ}_{37, \text{wa, dup}} = \Delta G^{\circ}_{37} - \Delta G^{\circ}_{37, \text{no cosolute}} - \Delta G^{\circ}_{37, \text{ev, dup}}$ , where  $\Delta G^{\circ}_{37}$ ,  $\Delta G^{\circ}_{37, \text{no cosolute}}$  and  $\Delta G^{\circ}_{37, \text{ev, dup}}$  were -13.6, -14.1 and -0.45 kcal mol<sup>-1</sup>, respectively (Table S8). The plot of  $\Delta G^{\circ}_{37, \text{wa, dup}}$  against the changes in water activity,  $\Delta a_w (\Delta a_w = a_{\text{no cosolute}} - a_{\text{cosolute}})$  revealed two different linear correlations depending on the chemical structures of the cosolutes (Figure 2B). Cosolutes lacking vicinal hydroxyl groups, such as PEGs, 2-methoxy ethanol (2-ME), and 1,2-dimethoxyethane (1,2-DME), showed a higher order of destabilization. On the other hand, the extent of destabilization was reduced in the presence of ethylene glycol (EG), glycerol (Gly) and 1,3-propanediol (1,3-PDO) with hydroxyl groups in the vicinity. The destabilization of the duplex by the addition of a cosolute was assumed





**Figure 2.** (A) Plot of excluded volume effect of RNA-PEG interaction against the base pair length of RNA duplexes. Data are provided in Table S7. (B) Plot of the contribution of water activity on the stability of GAUUACGCCUG against  $\Delta a_w$  in EG (blue circles), Gly (red circles), 1,3-PDO (green circles), PEG200 (grey squares), 2-ME (green square), 1,2-DME (orange square), PEG400 (purple square), PEG600 (red square) at 100 mM NaCl and in EG (pink circle), 1,3-PDO (cyan circle), PEG200 (black squares), 2-ME (cyan square), 1,2-DME (magenta square) at 1 M NaCl. EG, Gly and 1,3-PDO belong to the same group (circles) whereas PEGs, 2-ME and 1,2-DME are in another group (squares). Data are presented in Table S8.

to be due to the disruption of the ordered water network around the duplex (38). However, the cosolutes containing hydroxyl groups in close vicinity, such as EG, Gly and 1,3-PDO, form H-bonds with the water molecules surrounding the duplex, which can help to maintain the ordered water structure around the duplex even under crowding conditions, which reduces destabilization. As shown in Figure 2B, the lowest destabilization was observed in the 1,3-PDO solutions (green circles). This is due to the suitable positioning of the two hydroxyl groups in 1,3-PDO for the H-bonding interaction with water molecules (48), resulting in higher stabilities dominated by favorable enthalpy contributions as compared to the solution without any cosolute (Table S8). Interestingly, we observed the same correlation for cosolutes in both 1 M and 100 mM NaCl solutions, suggesting that  $\Delta G_{37}^{\circ, \text{wa, dup}}$  is independent of the ionic concentration of the solution, thus validating the additive nature of Equation 7. Because we found a linear relationship between  $\Delta G_{37}^{\circ, \text{wa, dup}}$  and  $\Delta a_w$ , it is reasonable to assume that each of the 10 NN base pairs and terminal parameters ( $\Delta G_{37}^{\circ, \text{NN, wa}}$ ) is also linearly correlated with the water activity of the solution by the following relation:

$$\Delta G_{37}^{\circ, \text{NN, wa}} = m_{\text{cs}} \cdot \Delta a_w \quad (13)$$

where  $m_{\text{cs}}$  is a prefactor for cosolutes, which is equivalent to the energy parameter for NN base pairs in the presence of a cosolute, dependent on the individual NN base pair and cosolute. Subtracting the contributions of bulk ( $\Delta G_{37}^{\circ, \text{NN, bulk}}$ ), cation ( $\Delta G_{37}^{\circ, \text{NN, [122 mM Na}^+]}$ ) and excluded volume effect ( $\Delta G_{37}^{\circ, \text{NN, ev}}$ ) from the  $\Delta G_{37}^{\circ, \text{NN}}$  parameters of Table 1 provided  $\Delta G_{37}^{\circ, \text{NN, wa}}$  for 40 wt% PEG200, as shown in Table 2. The  $m_{\text{cs}}$  values for PEG ( $m_{\text{PEG}}$ ) were calculated from the  $\Delta G_{37}^{\circ, \text{NN, wa [40 wt% PEG200]}$  values using Equation 13 (Table 2). The  $m_{\text{cs}}$  values of the other cosolutes were calculated as follows:

$$m_{\text{cs}} = m_{\text{PEG}} \cdot (S_{\text{CS}} / S_{\text{PEG}}) \quad (14)$$

where  $S_{\text{CS}}$  and  $S_{\text{PEG}}$  are the slopes for the cosolute of interest and PEG, respectively, in the  $\Delta G_{37}^{\circ, \text{wa, dup}}$  vs.  $\Delta a_w$  plot (Figure 2B).  $\Delta G_{37}^{\circ, \text{NN, wa}}$  for different cosolutes can be calculated by multiplying  $m_{\text{cs}}$  (Table 2) and  $\Delta a_w$ . The water activities ( $a_w$ ) in different cosolute solutions are available in the literature or can be measured by osmometry, as mentioned in the Supplementary data. Therefore, the stability of RNA duplexes in different crowded environments can be predicted from their base sequences, using the universal parameters listed in Table 2.

#### Validation of the generalized NN parameters for different *in vitro* crowded conditions

Although the NN parameters were established from the experimental data of  $\text{Na}^+$ -containing solutions, the intracellular environment contained  $\text{K}^+$  as the major monovalent cation (49,50). Therefore, considering the utility of the parameters for intracellular reactions, we verified the suitability of our parameters for the prediction of RNA stability in 100 mM KCl under the same crowding condition with 16 RNA duplexes selected from Table S2 differing in length and base composition. The measured  $\Delta G_{37}^{\circ}$  and  $T_m$  values indicated that, under the molecular crowding induced by 40 wt% PEG200, stabilities of RNA duplexes in  $\text{Na}^+$  were higher than those in  $\text{K}^+$  with the average difference of 12.1%, (Table S9). This could be as a result of the different binding modes of the two cations with the RNA (51);  $\text{K}^+$  interacts with the major grooves, whereas  $\text{Na}^+$  preferentially binds to highly charged phosphates, leading to more efficient charge screening. These differences were found significant when the dielectric constant was dramatically dropped, such as in 40 wt% PEG200 where dielectricity reduced to almost 50% compared to the non-crowding solution (Table S10), favoring electrostatic interaction between negatively charged phosphates and  $\text{Na}^+$ . However, in the absence of cosolute and in 40 wt% EG, where dielectricity did not decrease much, RNA duplexes showed only a marginal in-

crease in their stability in the presence of  $\text{Na}^+$  compared with  $\text{K}^+$  with an average differences of 3.1% and 4.2%, respectively (Tables S11 and S12). We investigated the quantitative relationship of RNA stability between  $\text{K}^+$  and  $\text{Na}^+$  in 40 wt% PEG200 to obtain duplex stability in the presence of  $\text{K}^+$  from their values in the  $\text{Na}^+$  solution. We observed a good linear relationship between the measured stability in  $\text{K}^+$  and the predicted stability in the same concentration of  $\text{Na}^+$  solution using our derived parameters (Supplementary Figure S5A) to produce following equation:

$$\Delta G_{37}^{\circ} (\text{measured in } \text{K}^+) = 0.93 \Delta G_{37}^{\circ} (\text{predicted in } \text{Na}^+) + 0.45 \quad (15)$$

The slope of the equation is defined as the ratio of the change in  $\Delta G_{37}^{\circ}$  values between the  $\text{K}^+$  and  $\text{Na}^+$  solutions. The value of 0.93 indicated that  $\Delta G_{37}^{\circ}$  decreased by 7% in the  $\text{K}^+$  condition compared to the  $\text{Na}^+$  solution because of the lower charge screening of RNA duplexes in the  $\text{K}^+$  medium. Henning-Knechtel et al. reported a 6% lesser charge neutralization by  $\text{K}^+$  than by  $\text{Na}^+$  for HIV-1 TAR RNA hairpins, as determined from molecular dynamics simulations (51). The intercept of the linear equation indicates the excess energy contribution in  $\text{K}^+$  medium when  $\Delta G_{37}^{\circ} (\text{Na}^+)$  is zero. This excess energy contribution is mainly affected by structural factors, such as duplex terminals. Positive intercept suggested that, under crowding conditions,  $\text{K}^+$  destabilizes the duplex terminals on average by 0.45 kcal mol<sup>-1</sup> compared to  $\text{Na}^+$ . Using Equation 15, the calculated  $\Delta G_{37}^{\circ}$  of the 16 duplexes in the  $\text{K}^+$  solution showed an average difference of only 4.3% from the measured values (Table S13), suggesting that, without calculating the NN parameters in the KCl condition, our derived parameters in 100 mM NaCl solution can provide an accurate estimation of RNA stability in 100 mM KCl following a simple linear relation.

We verified the applicability of the generalized parameters for different crowding conditions using a large number of RNA sequences with different NN compositions and lengths. We used the parameters in Table 2 to predict RNA duplex stability at different concentrations of a cosolute, different types of cosolutes, and different NaCl concentrations (Table S14). The reported stabilities of 38 RNA duplexes (28) in 20 vol% PEG200 at 1 M NaCl were also predicted (Table S14). Although NN parameters reported by Adams and Znosko predicted the measured thermodynamic parameters for 38 sequences with accuracy (28), unlike the current generalized parameters, those parameters are only applicable for a specific crowding condition. The calculation of  $\Delta G_{37}^{\circ}$  for a sequence in a particular cosolute solution using generalized parameters is illustrated in Figure 3. The average difference between the predicted and measured  $\Delta G_{37}^{\circ}$  values for the 52 sequences listed in Table S14 was 7.3%. Therefore, these results indicate that the generalized parameters can predict RNA duplex stability under different NaCl conditions and in various cosolutes, differing largely in their molecular weights, with significant accuracy.

## Stability prediction of RNA structures in local intracellular conditions

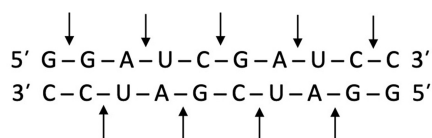
For the established parameters to predict RNA duplex stabilities in cells, which contain monovalent cations such as  $\text{Na}^+$  and  $\text{K}^+$ , and small amounts of divalent cations such as  $\text{Mg}^{2+}$ , the approximation of  $\Delta G_{\text{NN, cation}}^{\circ}$  is needed because the database of  $\Delta G_{\text{NN, cation}}^{\circ}$  values was available for a single component of the cation. To verify the applicability of the derived parameters under intracellular cation conditions, we measured thermodynamic parameters for the same 16 RNA duplexes in a buffer solution containing the exact intracellular cation composition (140 mM  $\text{K}^+$ , 10 mM  $\text{Na}^+$ , 0.5 mM  $\text{Mg}^{2+}$  and 0.0001 mM  $\text{Ca}^{2+}$ ) (49,50) under 40 wt% PEG200 at a pH of 7.2. The measured values were in agreement with the thermodynamic parameters and  $T_m$  of the duplexes measured in 122 mM  $\text{Na}^+$  with 40 wt% PEG200 (Table S15), indicating that the combined effect of monovalent and divalent cations on the duplex stability at the intracellular concentration is similar to that of the  $\text{Na}^+$  condition used for determining the NN parameters. When we plotted the measured  $\Delta G_{37}^{\circ}$  in the 40 wt% PEG200 with the intracellular cation concentration against the corresponding predicted values in the  $\text{Na}^+$  solution with 40 wt% PEG200 (Figure S5B), we obtained a straight line with the following equation.

$$\Delta G_{37}^{\circ} (\text{measured in intracellular cation conc.}) = 0.92 \Delta G_{37}^{\circ} (\text{predicted in } \text{Na}^+) - 0.12 \quad (16)$$

Equation 16 revealed a slope of 0.92, similar to the slope of Equation 15, suggesting a similar extent of charge neutralization of RNAs by intracellular cation composition and 122 mM  $\text{K}^+$ . However, contrary to Equation 15, the linear plot produced a slightly negative intercept of 0.12 kcal mol<sup>-1</sup>, suggesting that duplex terminals were stabilized by intracellular cation concentration as  $\text{Mg}^{2+}$  can bind to the N7 atom of the exposed purine bases at terminals (52). The combined effect of charge neutralization and terminal stabilization by the intracellular cation composition resulted in a similar stabilization to the  $\text{Na}^+$  condition used for determining the NN parameters, as evident from the data in Table S15. As a result, the prediction errors showed average differences in  $\Delta H^{\circ}$ ,  $\Delta S^{\circ}$ ,  $\Delta G_{37}^{\circ}$  and  $T_m$  of 7.4%, 7.7%, 6.7% and 1.9°C, respectively (Table S16), validating the applicability of the derived NN parameters for intracellular cation conditions. Therefore, molecular crowding conditions with 122 mM  $\text{Na}^+$  should be ideal as a model solution for studying physiological crowding in cells for physicochemical measurements.

According to the approximation of physiological cation condition as above, we demonstrated stability prediction of RNA duplexes in cells. It should be remembered that experimental data on nucleic acid stability in cells are limited, owing to the complexity of measurements and sophisticated instrumentation, as well as the crowding condition inside a cell is dynamic and heterogeneous. Thus, we can validate only few available data for RNA structures to demonstrate the applications of the established NN parameters for in-cell stability prediction. Here, we applied to predict RNA duplexes under nucleolus-like conditions in cells. In the





$$\Delta G^{\circ}_{37, \text{dup}} = 2 \Delta G^{\circ}_{37} r(\text{GG/CC}) + 4 \Delta G^{\circ}_{37} r(\text{GA/CU}) + 2 \Delta G^{\circ}_{37} r(\text{AU/UA}) + \Delta G^{\circ}_{37} r(\text{CG/GC}) + \Delta G^{\circ}_{37, \text{initiation}} + \text{symmetry factor}$$

$$\Delta G^{\circ}_{37, \text{NN}} (122 \text{ mM Na}^+, 20 \text{ wt\% EG}) = \Delta G^{\circ}_{37, \text{NN}} (\text{no cosolute}, 122 \text{ mM Na}^+) + \Delta G^{\circ}_{37, \text{NN, ev}} (20 \text{ wt\% EG}) + \Delta G^{\circ}_{37, \text{NN, wa}} (20 \text{ wt\% EG})$$

$$\Delta G^{\circ}_{37, \text{NN}} (\text{no cosolute}, 122 \text{ mM Na}^+) = \Delta G^{\circ}_{37, \text{NN, bulk}} + \Delta G^{\circ}_{37, \text{NN}, 122 \text{ mM Na}^+}$$

$$\Delta G^{\circ}_{37, \text{dup}} (\text{pred.}) = \Delta G^{\circ}_{37, \text{dup}} (\text{no cosolute}, 122 \text{ mM Na}^+) + \Delta G^{\circ}_{37, \text{ev, dup}} (20 \text{ wt\% EG}) + \Delta G^{\circ}_{37, \text{wa, dup}} (20 \text{ wt\% EG})$$

$$\begin{aligned} \Delta G^{\circ}_{37, \text{dup}} (\text{no cosolute}, 122 \text{ mM Na}^+) &= 2(-3.10) + 4(-2.22) + 2(-0.52) + (-2.18) + (4.09) + (0.43) \\ &= -13.78 \text{ kcal mol}^{-1} \end{aligned}$$

$$\Delta G^{\circ}_{37, \text{ev, dup}} (20 \text{ wt\% EG}) = -0.84 \text{ kcal mol}^{-1}$$

$$\begin{aligned} \Delta G^{\circ}_{37, \text{wa, dup}} (20 \text{ wt\% EG}) &= 2(3.6 \cdot 0.047) + 4(1.7 \cdot 0.047) + 2(1.6 \cdot 0.047) \\ &\quad + (2.0 \cdot 0.047) + (13.7 \cdot 0.047) \\ &= 1.55 \text{ kcal mol}^{-1} \end{aligned}$$

$$\begin{aligned} \Delta G^{\circ}_{37, \text{dup}} (\text{pred.}) &= (-13.78) + (-0.84) + (1.55) \\ &= -13.1 \text{ kcal mol}^{-1} \end{aligned}$$

$$\Delta G^{\circ}_{37, \text{dup}} (\text{exp.}) = -12.7 \text{ kcal mol}^{-1}$$

**Figure 3.** Prediction of  $\Delta G^{\circ}_{37}$  of sequence GGAUCGAUCC in solution containing 122 mM Na<sup>+</sup> and 20 wt% EG. Each arrow represents one NN set.  $\Delta G^{\circ}_{37, \text{NN, bulk}}$  and  $\Delta G^{\circ}_{37, \text{NN}, 122 \text{ mM Na}^+}$  values from Table 2 were used.  $\Delta G^{\circ}_{37, \text{ev, dup}}$  was calculated using Equation 11 and the  $\Delta G^{\circ}_{37, \text{wa, duplex}}$  was determined using the  $m_{\text{cs}}$  values obtained from Table 2, using Equation 13.  $\Delta a_w$  value of the solution was 0.047. This duplex is self-complementary, with symmetry factor of 0.43 kcal mol<sup>-1</sup>.

report by Nott et al. (53),  $\Delta \Delta G^{\circ}_{25}$  (difference in  $\Delta G^{\circ}_{25}$  between inside and outside the nucleolus) for two RNA duplexes, (ACUG)<sub>3</sub> and (ACUG)<sub>6</sub> were measured to be 3.0 and 3.3 kcal mol<sup>-1</sup> at 25°C, respectively. Since, these destabilizations were mainly due to the crowding condition inside the nucleolus, it can be considered as  $\Delta G^{\circ}_{\text{[crowder]}}$  at 25°C. We used our generalized  $\Delta G^{\circ}$  parameters adjusted for 25°C using  $\Delta H^{\circ}$  and  $\Delta S^{\circ}$  values (see Table S17) to determine the cosolute condition that imposes similar destabilizations, as reported for the two duplexes. We found that using our generalized parameters the measured destabilizations were matching with the predicted  $\Delta G^{\circ}_{25, \text{[crowder]}}$  in 40~50 wt% PEG200 solution (see Table S18), which is the same condition predicted using our DNA NN parameters for mimicking nucleolus (27). This result suggested the reliability of our newly derived RNA generalized parameters in predicting diverse crowding conditions both *in vitro* and intracellular environments.

Moreover, we demonstrated the prediction of hairpin RNA (Im-4U), mutated from *Salmonella* fourU-type RNA thermometer (4U) hairpin sequence located in 5'-UTR of the *agsA* gene (54). In Im-4U, G-U wobble base pairs, a hairpin loop, and an internal loop exist (Supplementary Figure S6). For G-U wobble base pairs, we estimated the stability of NN parameters including G-U pair according to the rules of the changes in NN pairs containing A-U pair because both base pairs have two hydrogen bonds, that are approximated to have same energy contribution

of A-U pair (Table S19) (55). For loops, as the destabilization effect of the hairpin tetraloop was reduced by 58% in 40 wt% PEG200 (Table S20) (56), we formulated an empirical relation for loop stability in the cosolute solutions ( $\Delta G^{\circ}_{37, \text{loop, cosolute}}$ ) from their stability in the absence of cosolute ( $\Delta G^{\circ}_{37, \text{loop, no cosolute}}$ ) and change in the water activity of cosolute solution ( $\Delta a_w$ ) as follows:

$$\Delta G^{\circ}_{37, \text{loop, cosolute}} = \Delta G^{\circ}_{37, \text{loop, no cosolute}} \cdot (1 - 11.8 \cdot \Delta a_w) \quad (17)$$

The details of these approximations were described in Supplementary Data. The excluded volume effect for the hairpin was determined following the same procedure as described earlier; only parameters for the hairpin were different from duplex (Table S6). From these parameters and relationship (Equation 17), the stability of the Im-4U in 40 wt% PEG200 at 100 mM NaCl was estimated to be 1.1 kcal mol<sup>-1</sup> (Supplementary Figure S6). The predicted value complied satisfactorily with the measured stability of 0.4 kcal mol<sup>-1</sup>, compared to the value of -2.7 kcal mol<sup>-1</sup>, predicted using classical NN parameters. The positive  $\Delta G^{\circ}_{37}$  of the predicted value indicates that destabilization by the loop was estimated well by the developed parameters, as other stem regions should stabilize the structure that negatively contributes to  $\Delta G^{\circ}_{37}$  (Table 1). Thus, this result suggests the dependability of the assumptions for the stability of G-U wobble pairs and loops under crowded condition.

Finally, we used the expanded parameters to estimate the crowding conditions in nucleus and cytosol.  $\Delta G^{\circ}_{37}$  of

Im-4U in the nucleus and cytosol of live HeLa cells were reported to be  $-1.7$  and  $-2.6$  kcal mol $^{-1}$ , respectively (54). The predicted  $\Delta G^{\circ}_{37}$  of Im-4U in PEG200 showed a significant destabilizing effect on the structure (1.1 kcal mol $^{-1}$  in 40 wt% PEG200) and did not match the reported  $\Delta G^{\circ}_{37}$  inside HeLa cells. Thus, we used less destabilizing cosolutes (EG, Gly, and 1,3 PDO) to estimate the intracellular crowding *in vitro*. We identified that the  $\Delta G^{\circ}_{37}$  of Im-4U in the nucleus and cytosol of HeLa cells were successfully predicted to be  $-1.9$  and  $-2.7$  kcal mol $^{-1}$ , respectively, using our parameters in 30 wt% and 40 wt% 1,3-PDO conditions at 100 mM NaCl (Table S21). These solution conditions were different from that of the nucleolus, as mentioned above, indicating that our prediction is versatile in different localized cellular regions. In contrast, it should be noted that the experimental  $\Delta G^{\circ}_{37}$  exhibited a wide distribution of about 2.0 kcal mol $^{-1}$  in the measured values in the nucleus and cytosol (54). This wide distribution suggests that some Im-4U strands further localized in environments such as liquid-liquid separated regions in the cytosol and nucleus, which were significantly different from other regions, and gained more stability in the cytosol than in the nucleus or vice versa. Therefore, our determined parameters will provide accurate stability of the secondary RNA structures in cells when the molecular environments of cellular compartments are well resolved; this requires high-resolution stability data for RNA structures in individual localized regions of a cell.

## CONCLUSION AND GUIDE FOR USERS

We developed prediction parameters of *in vitro* and in-cell stability of RNA duplex formation based on the NN model available in diverged crowding conditions. Our method proposed a general approach to obtain the NN parameters for different types of commonly used cosolutes, which is a sum of four separated NN parameters: 1)  $\Delta G^{\circ}_{37, \text{NN, bulk}}$  from Table 2, 2)  $\Delta G^{\circ}_{37, \text{NN, Na}^+}$  from Table 2, 3)  $\Delta G^{\circ}_{37, \text{NN, ev}}$  from Table 2 calculated using Equations 12, and 4)  $\Delta G^{\circ}_{37, \text{NN, wa}}$  determined from the  $m_{\text{CS}}$  values in Table 2 using Equation 13. For example, by changing  $\Delta G^{\circ}_{37, \text{NN, ev}}$  and  $\Delta G^{\circ}_{37, \text{NN, wa}}$  terms, specific NN parameters for the nucleolus condition using 40 or 50 wt% PEG200 in 100 mM NaCl solution and the cytosol condition using 40 wt% 1,3 PDO with 100 mM NaCl can be obtained. The users can freely arrange the parameters for predictions in their specific conditions containing different concentrations of Na $^+$  and crowders using these table and equations (as demonstrated in Figure 3).

A prediction of the thermodynamic stability of RNA duplexes under diverse crowding conditions is a basis to determine RNA folded structures in spatiotemporal area in cells. Further collection of parameters of various structural motifs like bulges, mismatches, and dangling ends will enable to predict large RNA structures *in vitro* and in cells. Our approach should be beneficial for not only investigating biological processes but also for designing RNA structures for RNA interference, antisense oligonucleotide therapy, and mRNA vaccines, which work efficiently in specific environments of different cells. Moreover, our method can provide the information of in-cell secondary structures which is currently unknown to develop RNA targeting drugs for RNA viruses including SARS-CoV-2.

## DATA AVAILABILITY

All relevant data are included in the paper and/or its supplementary data.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

Ministry of Education, Culture, Sports, Science and Technology (MEXT); Japan Society for the Promotion of Science (JSPS) [JP17H06351, 18KK0164, 19K05723, 21B208]; Grant-in-Aid for Scientific Research (S) [22H04975]; JSPS Core-to-Core Program [JPJSCCA20220005]; The Hirao Taro Foundation of Konan Gakuen for Academic Research; The Asahi Glass Foundation; Chubei Itoh Foundation. The open access publication charge for this paper has been waived by Oxford University Press – NAR Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.

*Conflict of interest statement.* None declared.

## REFERENCES

- Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
- Mustoe, A.M., Brooks, C.L. and Al-Hashimi, H.M. (2014) Hierarchy of RNA functional dynamics. *Ann. Rev. Biochem.*, **83**, 441–466.
- Nagatoishi, S., Ono, R. and Sugimoto, N. (2012) The yields of transcripts for a RNA polymerase regulated by hairpin structures in nascent RNAs. *Chem. Commun.*, **48**, 5121–5123.
- Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M.J. *et al.* (2019) mRNA structure regulates protein expression through changes in functional half-life. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 24075–24083.
- Chemla, Y., Peeri, M., Heltberg, M.L., Eichler, J., Jensen, M.H., Tuller, T. and Alfanta, L. (2020) A possible universal role for mRNA secondary structure in bacterial translation revealed using a synthetic operon. *Nat. Commun.*, **11**, 4827.
- Iwakawa, H.-o. and Tomari, Y. (2013) Molecular insights into microRNA-mediated translational repression in plants. *Mol. Cell*, **52**, 591–601.
- Lu, Z.J. and Mathews, D.H. (2008) Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res.*, **36**, 640–647.
- Guo, P. (2010) The emerging field of RNA nanotechnology. *Nat. Nanotechnol.*, **5**, 833–842.
- Tijerina, P., Mohr, S. and Russell, R. (2007) DMS footprinting of structured RNAs and RNA–protein complexes. *Nat. Protoc.*, **2**, 2608–2623.
- Wirth, R., Gao, P., Nienhaus, G.U., Sunbul, M. and Jäschke, A. (2019) SiRA: a silicon rhodamine-binding aptamer for live-cell super-resolution RNA imaging. *J. Am. Chem. Soc.*, **141**, 7562–7571.
- Hänsel, R., Foldynová-Trantířková, S., Löhr, F., Buck, J., Bongartz, E., Bamberg, E., Schwalbe, H., Dötsch, V. and Trantířek, L. (2009) Evaluation of parameters critical for observing nucleic acids inside living *Xenopus laevis* oocytes by in-cell NMR spectroscopy. *J. Am. Chem. Soc.*, **131**, 15761–15768.
- Tinoco, I. Jr, Uhlenbeck, O.C. and Levine, M.D. (1971) Estimation of secondary structure in ribonucleic acids. *Nature*, **230**, 362–367.
- Tinoco, I. Jr, Borer, P.N., Dengler, B., Levin, M.D., Uhlenbeck, O.C., Crothers, D.M. and Bralla, J. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nat. New Biol.*, **246**, 40–41.
- Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 9373–9377.

15. Breslauer, K.J., Frank, R., Blöcker, H. and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 3746–3750.
16. Sugimoto, N., Nakano, S.-i., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M. and Sasaki, M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211–11216.
17. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
18. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
19. Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinform.*, **11**, 129.
20. Sato, K., Akiyama, M. and Sakakibara, Y. (2021) RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat. Commun.*, **12**, 941.
21. Kierzek, E., Zhang, X., Watson, R.M., Kennedy, S.D., Szabat, M., Kierzek, R. and Mathews, D.H. (2022) Secondary structure prediction for RNA sequences including N6-methyladenosine. *Nat. Commun.*, **13**, 1271.
22. Xia, T., SantaLucia, J. Jr, Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719–14735.
23. Harvey Lodish, A.B., Lawrence Zipursky, S., Paul, M., Baltimore, D. and Darnell, J. (2000) In: *Molecular Cell Biology*. 4th edn., NY.
24. Kuznetsova, I.M., Turoverov, K.K. and Uversky, V.N. (2014) What macromolecular crowding can do to a protein. *Int. J. Mol. Sci.*, **15**, 23090–23140.
25. Nakano, S., Miyoshi, D. and Sugimoto, N. (2014) Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. *Chem. Rev.*, **114**, 2733–2758.
26. Takahashi, S. and Sugimoto, N. (2020) Stability prediction of canonical and non-canonical structures of nucleic acids in various molecular environments and cells. *Chem. Soc. Rev.*, **49**, 8439–8468.
27. Ghosh, S., Takahashi, S., Ohyama, T., Endoh, T., Tateishi-Karimata, H. and Sugimoto, N. (2020) Nearest-neighbor parameters for predicting DNA duplex stability in diverse molecular crowding conditions. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 14194–14201.
28. Adams, M.S. and Znosko, B.M. (2019) Thermodynamic characterization and nearest neighbor parameters for RNA duplexes under molecular crowding conditions. *Nucleic Acids Res.*, **47**, 3658–3666.
29. Fasman, G.D. (1976) In: *Handbook of Biochemistry and Molecular Biology*. CRC Press, Cleveland, Ohio.
30. Ghosh, S., Takahashi, S., Endoh, T., Tateishi-Karimata, H., Hazra, S. and Sugimoto, N. (2019) Validation of the nearest-neighbor model for Watson-Crick self-complementary DNA duplexes in molecular crowding condition. *Nucleic Acids Res.*, **47**, 3284–3294.
31. Banerjee, D., Tateishi-Karimata, H., Ohyama, T., Ghosh, S., Endoh, T., Takahashi, S. and Sugimoto, N. (2020) Improved nearest-neighbor parameters for the stability of RNA/DNA hybrids under a physiological condition. *Nucleic Acids Res.*, **48**, 12042–12054.
32. SantaLucia, J. Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 1460–1465.
33. Goobes, R., Kahana, N., Cohen, O. and Minsky, A. (2003) Metabolic buffering exerted by macromolecular crowding on DNA-DNA interactions: origin and physiological significance. *Biochemistry*, **42**, 2431–2440.
34. Gu, X., Nguyen, M.-T., Overacre, A., Seaton, S. and Schroeder, S.J. (2013) Effects of salt, polyethylene glycol, and locked nucleic acids on the thermodynamic stabilities of consecutive terminal adenosine mismatches in RNA duplexes. *J. Phys. Chem. B*, **117**, 3531–3540.
35. Leamy, K.A., Yennawar, N.H. and Bevilacqua, P.C. (2017) Cooperative RNA folding under cellular conditions arises from both tertiary structure stabilization and secondary structure destabilization. *Biochemistry*, **56**, 3422–3433.
36. Kierzek, R., Caruthers, M.H., Longfellow, C.E., Swinton, D., Turner, D.H. and Freier, S.M. (1986) Polymer-supported RNA synthesis and its application to test the nearest-neighbor model for duplex stability. *Biochemistry*, **25**, 7840–7846.
37. Ferreira, L., Jolley, E.A., Znosko, B.M. and Weber, G. (2019) Replacing salt correction factors with optimized RNA nearest-neighbour enthalpy and entropy parameters. *Chem. Phys.*, **521**, 69–76.
38. Nakano, M., Tateishi-Karimata, H., Tanaka, S., Tama, F., Miyashita, O., Nakano, S.-i. and Sugimoto, N. (2015) Thermodynamic properties of water molecules in the presence of cosolute depend on DNA structure: a study using grid inhomogeneous solvation theory. *Nucleic Acids Res.*, **43**, 10114–10125.
39. Victor, A., Bloomfield, D.M.C. and Ignacia Tinoco, JR. (2000) In: *Nucleic Acids: Structures, Properties, and Functions*. University Science Book.
40. Tzur, A., Kafri, R., LeBleu Valerie, S., Lahav, G. and Kirschner Marc, W. (2009) Cell growth and size homeostasis in proliferating animal cells. *Science*, **325**, 167–171.
41. Spink, C.H. and Chaires, J.B. (1999) Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA. *Biochemistry*, **38**, 496–508.
42. Kilburn, D., Roh, J.H., Guo, L., Briber, R.M. and Woodson, S.A. (2010) Molecular crowding stabilizes folded RNA structure by the excluded volume effect. *J. Am. Chem. Soc.*, **132**, 8690–8696.
43. Nakano, S.-i., Karimata, H., Ohmichi, T., Kawakami, J. and Sugimoto, N. (2004) The effect of molecular crowding with nucleotide length and cosolute structure on DNA duplex stability. *J. Am. Chem. Soc.*, **126**, 14330–14331.
44. Woolley, P. and Wills, P.R. (1985) Excluded-volume effect of inert macromolecules on the melting of nucleic acids. *Biophys. Chem.*, **22**, 89–94.
45. Knowles, D.B., LaCroix, A.S., Deines, N.F., Shkel, I. and Record, M.T. (2011) Separation of preferential interaction and excluded volume effects on DNA duplex and hairpin stability. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 12699–12704.
46. Hermans, J. (1982) Excluded-volume theory of polymer-protein interactions based on polymer chain statistics. *J. Chem. Phys.*, **77**, 2193–2203.
47. Huguette, J.M., Ribezzi-Crivellari, M., Bizarro, C.V. and Ritort, F. (2017) Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments. *Nucleic Acids Res.*, **45**, 12921–12931.
48. Gilligan, J.J., Vieira, N.E. and Yergey, A.L. (2004) Solvation of propanediol ions by water molecules in the gas phase. *J. Am. Soc. Mass Spectrom.*, **15**, 1123–1127.
49. Hille, B. (2001) In: *Ion Channels of Excitable Membranes*. 3rd edn., Sinauer Associates, Inc, Sunderland, MA.
50. Boyle, J. (2008) In: Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (eds). *Molecular Biology of the Cell*. 5th edition, Vol. **36**, pp. 317–318.
51. Henning-Knechtel, A., Thirumalai, D. and Kirmizialtin, S. (2022) Differences in ion-RNA binding modes due to charge density variations explain the stability of RNA in monovalent salts. *Sci. Adv.*, **8**, eabo1190.
52. Halder, A., Roy, R., Bhattacharyya, D. and Mitra, A. (2018) Consequences of Mg<sup>2+</sup> binding on the geometry and stability of RNA base pairs. *Phys. Chem. Chem. Phys.*, **20**, 21934–21948.
53. Nott, T.J., Craggs, T.D. and Baldwin, A.J. (2016) Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. *Nat. Chem.*, **8**, 569–575.
54. Gao, M., Gnutt, D., Orban, A., Appel, B., Righetti, F., Winter, R., Narberhaus, F., Müller, S. and Ebbinghaus, S. (2016) RNA hairpin folding in the crowded cell. *Angew. Chem. Int. Ed.*, **55**, 3224–3228.
55. Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. Edited by I. Tinoco. *J. Mol. Biol.*, **288**, 911–940.
56. Leonard, K.N. and Bloise, J.M. (2018) Effects of osmolytes and macromolecular crowders on stable GAAA tetraloops and their preference for a CG closing base pair. *PeerJ*, **6**, e4236.