Supplementary Material: Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments

J. M. Huguet*, M. Ribezzi-Crivellari*, C. V. Bizarro, F. Ritort

November 14, 2017

Contents

*Equally contributed

A Mean unzipping forces

The mean unzipping forces as a function of salt concentrationin in the Na^+ and Mg^{2+} cases are shown in Figure S1. The inset shows the expected logarithmic dependence and the validity of the approximate 100^{th} rule, recently reported in force experiments in DNA [13] and RNA [1], by which the non-specific binding affinity of divalent cations at a given concentration is equivalent to that of monovalent cations taken at 100-fold concentration.

B Abundances of different NNBP along the DNA sequence

Our 6.7kb sequence contains an even amount of all different NN motifs, to exclude misrepresentation of any of them (see Table S1).

C Extracting the Equilibrium Force-Distance Curve (FDC) in regions with hysteresis

In the absence of hysteretic effects between unzipping and rezipping we simply average the filtered unzipping and rezipping FDCs to extract the thermodynamic FDC. However, there are force rips in the FDC that exhibit

NNBP	Fraction (%)
AA/TT	16.4
AC/TG	11.2
AG/TC	10.9
AT/TA	7.6
CA/GT	14.1
CC/GG	9.1
CG/GC	6.0
GA/CT	12.8
GC/CG	6.5
TA/AT	5.4

Table S1: Abundances of the different NN motifs in the DNA sequence. Notice that abundances are similar. For those six motifs where they appear to be too high (AA/TT, AC/TG, AG/TC, CA/GT, GA/CT, CC/GG) there is the double degeneracy due to Watson-Crick bp symmetry (e.g. the fraction of AA/TT includes AA/TT and TT/AA).

hysteresis. In these regions, the pulling speed is too fast with respect to the equilibration time, so multiple transitions (corresponding to coexistence) between the open and closed states are not observed (Fig. S2a) and the unziping/rezipping process does not proceed under quasi-static conditions. A simple average of both (unzipping and rezipping) FDCs provides an estimate of the equilibrium FDC (orange curve in Fig. S2d). However such estimate strongly deviates from the expected theoretical FDC posing convergence problems to the fitting algorithm of the NN free energies. A different strategy is thus required to extract FDCs in regions with hysteresis.

First, we determine the start and end points of the regions where hysteresis is observed (Fig. S2a). Then, we calculate the work along the unzipping (zipping) process from the start (end) point up to a given intermediate distance x:

$$\begin{aligned} W_{\rm U}(x) &= \int_{x_1}^{x} f_{\rm U}(x) \cdot {\rm d}x \\ W_{\rm R}(x) &= \int_{x_2}^{x} f_{\rm R}(x) \cdot {\rm d}x. \end{aligned} \tag{S1}$$

The unzipping (rezipping) work is positive (negative) (Fig. S2b). We also

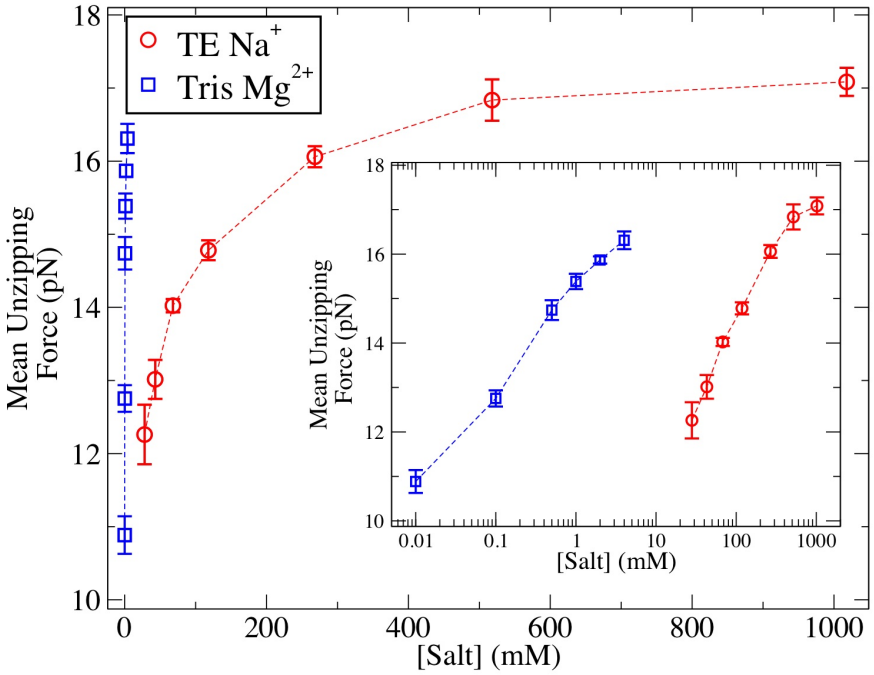

Figure S1: Mean unzipping forces from experimental data. Mean unzipping force as a function of salt concentration for varying Mg^{2+} and Na^{+} concentrations. The inset shows data in a linear-log scale and the approximated validity of the 100^{th} rule.

calculate the average work function, $W_{\rm avg}(x) = (W_{\rm U}(x) + W_{\rm R}(x))/2$ (orange curve in Fig. S2b) determining the coexistence point ($x_{\rm c}$) where the average work equals 0, $W_{\rm avg}(x_{\rm c}) = 0$. An estimate of the free energy along each branch is given by:

$$\begin{aligned} G_{\rm U}(x) &= G_{\rm U}(x_1) + W_{\rm U}(x) \\ G_{\rm R}(x) &= G_{\rm R}(x_2) + W_{\rm R}(x) \end{aligned} \tag{S2}$$

where $G_{\rm U}(x)$ ($G_{\rm R}(x)$) is the free energy of the unzipping (rezipping) branch at distance x. $G_{\rm U}(x_1)$ is the free energy of the molecule when it is closed at distance x_1; and $G_{\rm R}(x_2)$ is the free energy of the molecule when it is unzipped at distance x_2. Here we introduce a simplifying approximation i.e. that:

$$W_U(x_{\rm c}) = -W_R(x_{\rm c}) = \frac{G_{\rm R}(x_2) - G_{\rm U}(x_1)}{2}.$$

This is exact in cases with special symmetry and is a very good approximation in most other cases. Also, in the following, the value $G_{\rm U}(x_1)$ is set to zero to get simpler expressions. As a consequence we can rewrite Eqs. S2 in terms of quantites that can be obtained directly from the FDCs:

$$\begin{aligned} G_{\rm U}(x) &= W_{\rm U}(x) \\ G_{\rm R}(x) &= 2W_{\rm U}(x_{\rm c}) + W_{\rm R}(x) \end{aligned} \tag{S3}$$

Figure S2c shows the calculated free energy of both branches. The free energy of the system can be obtained from the partial free energies of the two metastable states as:

$$G_{\rm eq}(x) = -k_B T \cdot \ln\left(e^{-\beta G_{\rm U}(x)} + e^{-\beta G_{\rm R}(x)}\right) \tag{S4}$$

where $G_{\rm eq}(x)$ is the predicted equilibrium free energy, k_B is the Boltzmann constant, T is the temperature and $\beta = (k_B T)^{-1}$. Finally, the equilibrium FDC ($f_{\rm eq}(x)$) is obtained as:

$$f_{\rm eq}(x) = \frac{\partial\, G_{\rm eq}(x)}{\partial x} \tag{S5}$$

Figure S2d shows the estimated equilibrium FDC.

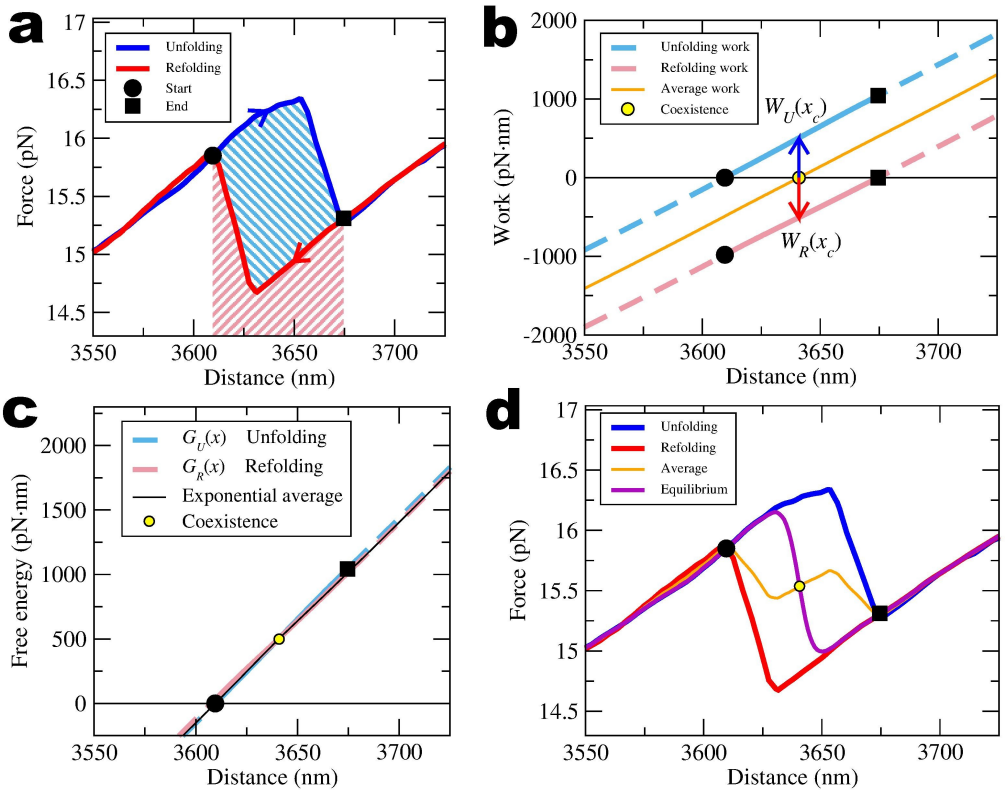

Figure S2: Estimation of the equilibrium FDC in force rips with hysteresis. (a) Unzipping and rezipping FDC in a region with hysteresis. The shaded areas show the unzipping and rezipping work. **(b)** $W_U(x)$, $W_R(x)$ and $W_{\rm avg}(x)$. Start (end) point for the integration are shown with a black solid circle (square). At the coexistence point x_c the average work is zero. **(c)** Unzipping and rezipping estimated free energy branches. The global free energy of the system at different positions is estimated as the potential of mean force of the two free energy branches. **(d)** Estimation of the equilibrium FDC. Notice the difference between the resulting FDC using the average work or the the potential of mean force.

D The self-consistent Nearest-Neighbor Base-Pair (NNBP) relations

The derivation of the nearest neighbor base-pair (NNBP) energies using either bulk or single molecule methods is often done in terms of a set of 10 parameters defining the nearest-neighbor (NN) model. However one can take advantage of the circular symmetry of the NN model[11, 12], which results into additional self-consistent relations between the NNBP energies. This reduces the total number of free parameters from 10 to 8.

Let any adjacent stacks or dimer along a given DNA duplex be denoted by (i, j) where indices i, j take the possible values A, T, C, G along either of the two DNA strands. A pair (i, j) indicates that base j is stacked on top of base i as read from the 5' to 3'. For example AC indicates that the dinucleotide $5' - AC - 3'$ is hybridized with the complementary dinucleotide $5' - GT - 3'$. In the NN model the free energy of formation of an arbitrary DNA duplex can be written as,

$$E = \sum_{i,j=A,T,C,G} n_{ij}\epsilon_{ij} \tag{S6}$$

where ϵ_{ij} stands for the different NNBP energies between adjacent stacks or dimers (i, j) and n_{ij} is the occupancy of that dimer along the sequence. There are 16 elements in the matrix ϵ_{ij}. However, complementary strand symmetry imposes that only 10 out of the 16 energies are different (for instance, $\epsilon_{AC} = \epsilon_{GT}$). Finally, if we neglect end effects then the occupation numbers n_{ij} satisfy additional closure relations,

$$\sum_{i=A,T,C,G} n_{Ai} = \sum_{i=A,T,C,G} n_{iA} \tag{S7}$$

$$\sum_{i=A,T,C,G} n_{Ci} = \sum_{i=A,T,C,G} n_{iC} \tag{S8}$$

$$\sum_{i=A,T,C,G} n_{Gi} = \sum_{i=A,T,C,G} n_{iG} \tag{S9}$$

$$\sum_{i=A,T,C,G} n_{Ti} = \sum_{i=A,T,C,G} n_{iT} \tag{S10}$$

Due to complementary strand symmetry the four closure relations Eqs. S7,S8,S9,S10

can be reduced to two independent constraints:

$$2\,n_{CG} + n_{CA} + n_{TG} + n_{CT} + n_{AG} = 2\,n_{GC} + n_{TC} + n_{GA} + n_{AC} + n_{GT} \quad \text{(S11)}$$

$$2\,n_{AT} + n_{AG} + n_{CT} + n_{AC} + n_{GT} = 2\,n_{TA} + n_{GA} + n_{TC} + n_{TG} + n_{CA}. \quad \text{(S12)}$$

It is not difficult to verify that the previous closure relations impose equivalent self-consistent relations for the energies (see Section D.1 below),

$$2\,\epsilon_{CG} + \epsilon_{CA} + \epsilon_{AG} = 2\,\epsilon_{GC} + \epsilon_{GA} + \epsilon_{AC} \quad \text{(S13)}$$

$$2\,\epsilon_{AT} + \epsilon_{AG} + \epsilon_{AC} = 2\,\epsilon_{TA} + \epsilon_{GA} + \epsilon_{CA}. \quad \text{(S14)}$$

further reducing the set of 10 parameters to 8 independent energies. It is convenient to recast these relations in terms of the following ratios,

$$r_1 = \frac{\epsilon_{AT} - \epsilon_{TA} - \epsilon_{GC} + \epsilon_{CG}}{(\epsilon_{GA} - \epsilon_{AG})} \quad \text{(S15)}$$

$$r_2 = \frac{\epsilon_{AT} - \epsilon_{TA} + \epsilon_{GC} - \epsilon_{CG}}{(\epsilon_{CA} - \epsilon_{AC})} \quad \text{(S16)}$$

which are be equal to 1 if Eqs.(S13,S14) hold. Previous bulk and single molecule derivations of NNBP energies were performed with the 10 parameters. The UO data set gives $r_1 = 3.1, r_2 = 15.5$ at 25°C, 1M NaCl or $r_1 = 5.8, r_2 = 18.5$ at 37°C, 1M NaCl, quite far from 1. Interestingly, force unzipping measurements [9] give ratios typically between 0.5 and 1.5. This shows that, despite of the fact self-consistency relations were not used in Ref.[9] the minimization algorithm employed in the analysis of FDC data converged to the relevant 8-dimensional subspace of energy solutions. This is a consequence of the cicular symmetry, which is approximately valid for long linear molecules.

An alternative method to test the approximate validity of the circular symmetry in unzipping experiments is to analyze the Hessian of the mean-squared error function defined as:

$$\mathcal{E}(\{\epsilon_i\}) = \frac{1}{N}\sum_{i=1}^{N}\Big(f_i^{\rm exp} - f_i^{\rm the}(\{\epsilon_i\})\Big)^2 \quad \text{(S17)}$$

where $\{\epsilon_i\} = \{\epsilon_{\rm AA/TT}, \epsilon_{\rm AC/GT}, \epsilon_{\rm AG/CT}, \epsilon_{\rm AT/AT}, \epsilon_{\rm CA/GT}, \epsilon_{\rm CC/GG}, \epsilon_{\rm CG/CG}, \epsilon_{\rm GA/CT}, \epsilon_{\rm GG/CC}, \epsilon_{\rm TA/T}\}$ are the 10 NNBP energies, which are the fitting parameters,

hereafter denoted as ϵ; N is the number of sampling points of the experimental FDC; f_i^{exp} is the experimental force exerted on the molecule at distance d_i; and f_i^{the} is the theoretical FDC predicted by the mesoscopic model (see Section E). Fitting the model to the experimental data consists in minimizing E by tuning the values of $\{\epsilon_i\}$. At the minimum we can approximate $E(\{\epsilon_i\})$ as:

$$\Delta\mathcal{E} = \frac{1}{2}\delta\epsilon^{\mathrm{T}} \cdot H \cdot \delta\epsilon \tag{S18}$$

where $\Delta\mathcal{E} = \mathcal{E}(\epsilon) - \mathcal{E}_0$, $\mathcal{E}_0$ is the error at the minimum; ϵ_0 are the values of the parameters at the minimum and H is the Hessian matrix, given by $H_{ij} = \partial^2\mathcal{E}(\epsilon)/\partial\epsilon_i\partial\epsilon_j$. The diagonalization of the Hessian matrix gives:

$$H = V\Lambda V^{-1} \tag{S19}$$

where V is the matrix of eigenvectors $V = (v_1|v_2|...|v_{10})$; Λ is the diagonal matrix of eigenvalues λ_i. The approximate circular symmetry generates two zero modes in the 10-dimensional landscape defined by the error function, so that the Hessian matrix exhibits two eigenvectors with zero eigenvalue, $\lambda_1 = \lambda_2 = 0$. The corresponding eigenvectors v_1, v_2 span a two dimensional subspace defined by Eqs.(S11,S12). The two experimentally measured lowest eigenvalues, albeit small, should not be exactly zero because the circular symmetry is weakly broken due to end effects. As we can see in Figure S3a, the spectrum of eigenvalues clearly shows two small eigenvalues that differ by at least one order of magnitude with respect to the rest. Changing the energy parameters along the directions defined by the corresponding eigenvectors barely affects the error function. Good overlap is found between these experimental eigenvectors and the prediction based on the circular symmetry (see Fig. S3b). Figure S3c shows the reproducibility of these results for several molecules.

D.1 Self-consistent relations and eigenvectors of the Hessian

Here we show how the closure relations Eqs.(S7,S8) reduce, from 10 to 8, the number of energy parameters to be determined according to Eqs.(S11,S12).

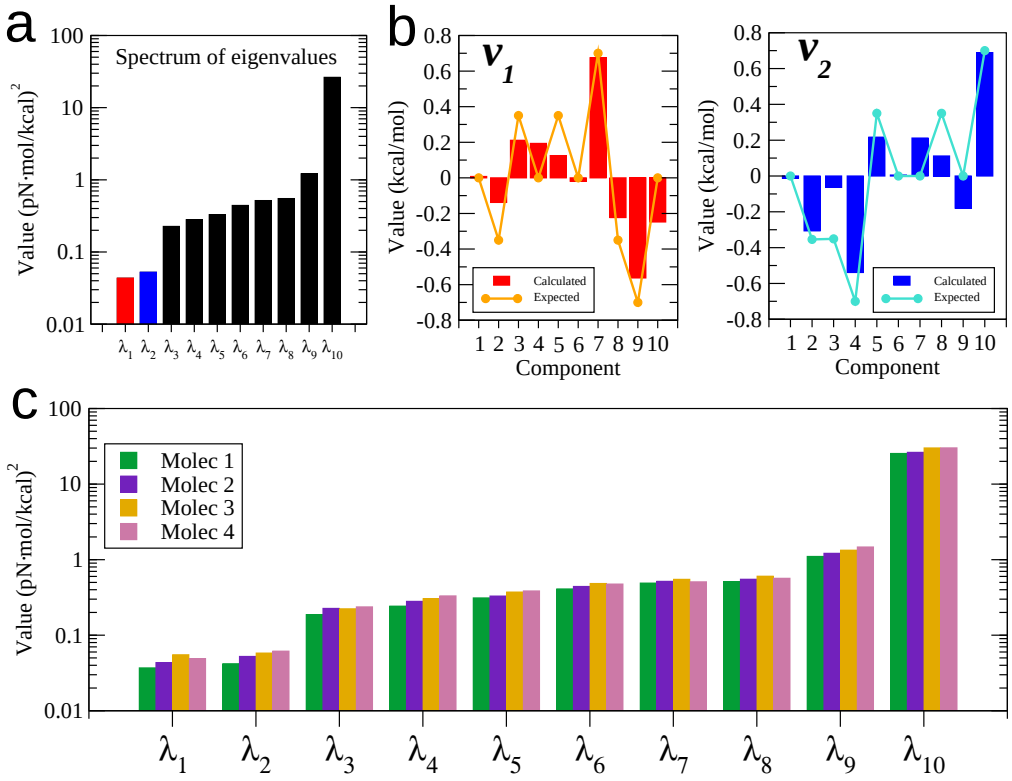

Figure S3: Eigenvalues and eigenvectors of the Hessian of the error function. **(a)** Spectrum of eigenvalues for one molecule. The two smaller eigenvalues (highlighted in red and blue) differ from the other ones by one order of magnitude. **(b)** The left panel shows the components of the eigenvector v_1, which has the lowest eigenvalue (depicted in red in panel a). The right panel, shows the components of the eigenvector v_2, which has the next-to-lowest eigenvalue (depicted in blue in panel a). The bars show the components of v_1 and v_2 after diagonalizing the Hessian matrix (Eq. S19). The lines show the theoretical prediction based on the circular symmetry. **(c)** Spectra of eigenvalues for 4 different molecules.

We also show that the Hessian matrix of the energy function, defined in the subspace of 10 energies, has two eigenvectors of zero eigenvalue. To prove the first result we consider the relevant 10 occupation numbers n_{ij} and energies ϵ_{ij} as 10-dimensional vectors, $n_k^{10}, \epsilon_k^{10}\, k \in \{0, 9\}$, through the map:$\epsilon_{\rm AA} = \epsilon_0, \epsilon_{\rm AC} = \epsilon_1, \epsilon_{\rm AG} = \epsilon_2, \epsilon_{\rm AT} = \epsilon_3, \epsilon_{\rm CA} = \epsilon_4, \epsilon_{\rm CC} = \epsilon_5, \epsilon_{\rm CG} = \epsilon_6, \epsilon_{\rm GA} = \epsilon_7, \epsilon_{\rm GC} = \epsilon_8, \epsilon_{\rm TA} = \epsilon_9$ and similarly for n_{ij}. In this language the free energy of formation

of a DNA duplex can be written as:

$$E = \sum_i n_i \epsilon_i = n \cdot \epsilon, \tag{S20}$$

where by $\cdot$ we mean the scalar product in the 10-dimensional space. Also, in the same notation, the two closure relations Eqs.(S7,S8) are conveniently written as:

$$2\,n_8 = 2\,n_6 + n_4 + n_2 - n_7 - n_1 \tag{S21}$$

$$2\,n_9 = 2\,n_3 + n_2 + n_1 - n_7 - n_4. \tag{S22}$$

In this form the closure relations show how, once the first 8 occupation numbers are fixed ($i \in \{0,7\}$), the values of the last two follow. The 10-dimensional occupation number vector can be obtained from a reduced 8-dimensional vector, n^8_k, $k \in \{0,7\}$ by a linear operation:

$$n^{10} = A n^8, \tag{S23}$$

where

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -\frac{1}{2} & 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} \end{pmatrix}. \tag{S24}$$

Inserting Eq. S23 in Eq. S20 and using the usual property of the scalar product we get:

$$E = n^{10} \cdot \epsilon^{10} = \left(A n^8\right)^T \cdot \epsilon^{10} = n^8 \cdot \left(A^T \epsilon^{10}\right), \tag{S25}$$

where by A^T we mean the transpose of A.

Consider now two different energy vectors: ϵ^{10}, η^{10}, such that $\epsilon^{10} = \eta^{10} + \delta^{10}$ with

$$A^T \delta^{10} = 0. \tag{S26}$$

The total stacking energy computed using the two different energy vectors is the same:

$$\begin{aligned} n^{10} \cdot \epsilon^{10} &= n^8 \cdot \left(A^T \epsilon^{10}\right) = n^8 \cdot \left(A^T \left(\eta^{10} + \delta^{10}\right)\right) = \\ &= n^8 \cdot \left(A^T \eta^{10}\right) + n^8 \cdot \left(A^T \delta^{10}\right) = n^8 \cdot \left(A^T \eta^{10}\right) \quad . \end{aligned} \tag{S27}$$

Solutions to Eq. (S26) belong to a two dimensional vector subspace $\mathcal{V}$, spanned by the vectors

$$u = \frac{1}{\sqrt{8}} \left(0, 1, -1, 0, -1, 0, -2, 1, 2, 0\right), \; v = \frac{1}{\sqrt{8}} \left(0, -1, -1, -2, 1, 0, 0, 1, 0, 2\right). \tag{S28}$$

Energy vectors satisfying Eq. (S26) are "invisible" in the sense they do not contribute to the total energy. An energy vector is "visible" if it belongs to the complement of $\mathcal{V}$ i.e. to the space of vectors which have 0 projection on (or scalar product with) both u and v. Enforcing these conditions we get the defining equations for the 8-dimensional space of "visible vectors":

$$\epsilon_8^{10} = \frac{1}{2} \left(\epsilon_2^{10} + \epsilon_4^{10} - \epsilon_1^{10} - \epsilon_7^{10}\right) + \epsilon_6^{10} \tag{S29}$$

$$\epsilon_9^{10} = \frac{1}{2} \left(\epsilon_1^{10} + \epsilon_2^{10} - \epsilon_4^{10} - \epsilon_7^{10}\right) + \epsilon_3^{10}, \tag{S30}$$

which are equivalent to Eqs.(S13,S14). It is now easy to show that the Hessian of the mean squared error function has two null eigenvalues. Indeed the mean squared error function depends on ϵ only through the scalar products $n \cdot \epsilon$ where n does always fulfill Eqs.(S7,S8) to a good approximation. As a consequence, changing ϵ along along the subspace spanned by vectors u, v does not change neither E, Eq.(S20), nor $\mathcal{E}$, Eq.(S17). In other words the directional derivative of $E(\epsilon)$ along u or v is 0, and so is the Hessian. The vectors u and v will thus be eigenvectors of the Hessian with null eigenvalue.

E Fit of unzipping FDCs

The values of the free energies of the NN motifs mostly define the shape of the FDC, i.e. the location and the height of the force rips. The estimation of the free energy motifs is done by fitting the theoretical FDC to the experimental one. This is achieved by performing a minimization of the mean squared error function Eq.(S17).The theoretical FDC is calculated from the mesoscopic model introduced in Ref. [9]. In this model the free energy of each element involved in the unzipping experiment (elastic handles, bead in the optical trap, hybridization energy of the dsDNA and elastic energy of the partially unzipped DNA) is computed to calculate the total free energy ($G_{\rm tot}$) of the system:

$$G_{\rm tot}(x,n) = E_b(x_b) + 2G_h(x_h) + 2G_s(x_s,n) + G_{\rm DNA}(n) \qquad \text{(S31)}$$

where x is the total extension of the system and n is the number of open base pairs. $E_b(x_b)$ is the elastic energy of the bead in the optical trap, which depends on the bead to trap distance x_b and on the optical trap stiffness k according to $E_b(x_b) = kx_b^2/2$. The elastic free energy of one handle, $G_h(x_h)$, depends on the extension of the handle x_h and is calculated from the FDC ($f_h(x)$) as predicted by the Worm-Like Chain model with appropriate parameters (persistence and contour length). Such contribution is given by:

$$G_h(x_h) = \int_0^{x_h} f_h(x)\mathrm{d}x \quad . \qquad \text{(S32)}$$

Moreover $G_s(x_s,n)$ is the elastic free energy of the partially unzipped DNA, which depends on its extension x_s and on the number of open base pairs n (which determine the contour length L of the ssDNA according to $L = n \cdot b$, where b is the contour length of one base). The free energy $G_s(x_s,n)$ is then calculated from the FDC ($f_s(x)$) of a Worm-Like Chain or a Freely-Jointed-Chain model with the specific parameters of a ssDNA molecule,

$$G_s(x_s) = \int_0^{x_s} f_s(x)\mathrm{d}x \quad . \qquad \text{(S33)}$$

Finally, $G_{\rm DNA}(n)$ is the hybridization free energy of the DNA duplex with n partially unzipped base pairs, as given by the NN model Eq.(S20) plus

initiation terms:

$$G_{\rm DNA}(n) = {\rm IF}(n+1) + {\rm IF}(N) + \sum_{n}^{N-1} \Delta g_i \tag{S34}$$

where N is the total number of base pairs of the DNA duplex, ${\rm IF}(n+1)$ is the Initiation Factor of the first free base pair at one end of the duplex, ${\rm IF}(N)$ is the Initiation Factor of the last base pair at the other end of the duplex, and the summation runs over all formed base pairs of the partially unzipped DNA molecule (from now on we will use Δg_i instead of ϵ_i to denote the NN energies). The total energy of the system, $G_{\rm tot}(x,n)$ (Eq.(S31)), is constrained by the total distance (x), which is the control parameter in our experiments and constrains the extensions of the different elements according to: $x = x_b + 2x_h + 2x_s$. Then the partition function of the system ($Z(x)$) can be computed at every value of x by summing over all possible configurations:

$$Z(x) = \sum_{n} e^{-\beta G_{\rm tot}(x,n)} \tag{S35}$$

where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant and T is the temperature. From this expression the theoretical FDC can be calculated with the following relation:

$$f^{\rm the}(x) = -k_B T \frac{\partial \ln Z(x)}{\partial x} \tag{S36}$$

which is the prediction of the FDC that we use in Eq. S17.

For each unzipping experiment (i.e. one molecule at a given salt concentration) we have a FDC. Using this FDC, the value of the mean squared error Eq.(S17) is minimized with a Monte Carlo algorithm (see Ref. [9] for a detailed description) and the values of the fitting parameters (Δg_i) are obtained. Table S2 summarizes the results and the total amount of data available from Mg^{2+} unzipping experiments. Similarly, Table S3 shows the same results for the Na^{+} experiments carried out in Ref. [9]. Note that, compared to Ref. [9], here we improved our methodology by performing the minimization over the eight parameters (as discussed in the previous section) plus the initiation factors as described in the main text.

	Δg_i (kcal/mol)							
i	NNBP	10 μM	100 μM	500 μM	1 mM	2 mM	4 mM	10 mM
1	AA/TT	-0.70 (0.03)	-0.86 (0.04)	-1.09 (0.02)	-1.13 (0.03)	-1.16 (0.01)	-1.25 (0.02)	-1.24 (0.01)
2	AC/TG	-1.03 (0.02)	-1.22 (0.02)	-1.40 (0.02)	-1.45 (0.04)	-1.47 (0.01)	-1.52 (0.03)	-1.49 (0.05)
3	AG/TC	-1.01 (0.02)	-1.13 (0.03)	-1.31 (0.02)	-1.34 (0.04)	-1.38 (0.01)	-1.44 (0.01)	-1.47 (0.00)
4	AT/TA	-0.44 (0.02)	-0.73 (0.04)	-0.91 (0.06)	-0.94 (0.01)	-0.95 (0.02)	-1.01 (0.02)	-1.11 (0.01)
5	CA/GT	-1.22 (0.06)	-1.42 (0.05)	-1.65 (0.03)	-1.67 (0.02)	-1.65 (0.01)	-1.74 (0.02)	-1.74 (0.01)
6	CC/GG	-1.81 (0.06)	-1.89 (0.03)	-1.90 (0.02)	-2.01 (0.02)	-2.01 (0.03)	-2.03 (0.01)	-1.99 (0.01)
7	CG/GC	-1.93 (0.04)	-2.17 (0.04)	-2.22 (0.05)	-2.31 (0.05)	-2.31 (0.02)	-2.37 (0.02)	-2.29 (0.02)
8	GA/CT	-1.01 (0.04)	-1.15 (0.01)	-1.34 (0.03)	-1.40 (0.04)	-1.42 (0.02)	-1.52 (0.02)	-1.46 (0.02)
	ΔIF	-0.37 (0.32)	-0.38 (0.22)	-0.04 (0.38)	0.02 (0.25)	-0.19 (0.22)	0.02 (0.38)	0.47 (0.26)
	# molecules	6	4	12	8	12	11	8

Table S2: The free energies of the motifs obtained from the unzipping experiments with Mg^{2+} buffer. These free energy data were fit to a logarithmic salt dependency (Eq.1 in main text). Minimization was performed over the eight parameters plus the initiation factors.

	Δg_i (kcal/mol)							
i	NNBP	10 mM	25 mM	50 mM	100 mM	250 mM	500 mM	1 M
1	AA/TT	-0.75 (0.02)	-0.77 (0.02)	-0.87 (0.01)	-0.91 (0.02)	-1.07 (0.02)	-1.16 (0.02)	-1.24 (0.03)
2	AC/TG	-1.09 (0.05)	-1.06 (0.06)	-1.22 (0.02)	-1.22 (0.02)	-1.36 (0.02)	-1.45 (0.02)	-1.49 (0.03)
3	AG/TC	-1.02 (0.04)	-1.04 (0.05)	-1.15 (0.03)	-1.11 (0.02)	-1.24 (0.02)	-1.31 (0.02)	-1.33 (0.02)
4	AT/TA	-0.69 (0.02)	-0.74 (0.05)	-0.77 (0.01)	-0.84 (0.05)	-0.95 (0.02)	-1.05 (0.02)	-1.07 (0.05)
5	CA/GT	-1.44 (0.04)	-1.46 (0.05)	-1.50 (0.02)	-1.48 (0.02)	-1.65 (0.03)	-1.72 (0.03)	-1.76 (0.05)
6	CC/GG	-1.66 (0.04)	-1.64 (0.04)	-1.76 (0.02)	-1.84 (0.01)	-1.86 (0.03)	-1.91 (0.03)	-1.92 (0.02)
7	CG/GC	-1.91 (0.08)	-1.89 (0.10)	-2.00 (0.06)	-2.22 (0.04)	-2.23 (0.08)	-2.30 (0.08)	-2.34 (0.05)
8	GA/CT	-0.99 (0.05)	-1.03 (0.05)	-1.17 (0.02)	-1.26 (0.02)	-1.32 (0.03)	-1.46 (0.03)	-1.52 (0.02)
	ΔIF	0.65 (0.07)	0.44 (0.06)	0.55 (0.12)	0.46 (0.08)	0.48 (0.03)	0.58 (0.02)	0.49 (0.06)
	# molecules	6	4	6	4	5	5	5

Table S3: The free energies of the motifs obtained from the unzipping experiments with Na^{+} buffer. These free energy data were fit to a logarithmic salt dependency (Eq.1 in main text). Minimization was performed over the eight parameters plus the initiation factors.

F Correlation of salt dependent correction factors with GC content

Salt dependent corrections are available for melting temperature and the entropy of duplex sequences (and for their free energy as the enthalpy is commonly taken as sequence independent). One of them is Eq.(22) in Ref. [16] for the entropy,

$$\Delta S^0([Mg^{2+}]) = \Delta S^0(1MNa^+) + \Delta H^0\Big(a + b\log([Mg^{2+}]) + f_{GC}(c + d\log([Mg^{2+}])) + \frac{e + f\log([Mg^{2+}]) + g\log^2([Mg^{2+}])}{2(N_{bp}-1)}\Big) \quad \text{(S37)}$$

with $\Delta S^0(1MNa^+)$ the reference entropy at 1M Na^+, ΔH_0 is the temperature and salt independent enthalpy, f_{GC} is the fraction of GC base pairs in the duplex and N_{bp} is total the number of base pairs and a, b, c, d, e, f, g are experimentally determined parameters (see Ref. [16]). A common aspect of Eq.(S37) and related ones is the fact that magnesium salt correction depends on the fraction of GC content, f_{GC}. We have looked whether such correlation exists by plotting the NN dependent salt correction parameter m as a function of the GC content of the NN motif which can be 0 (e.g. AA/TT), 0.5 (e.g. AC/TG), 1 (e.g. CC/GG). As we show in Figure S4 such correlation is clear for the magnesium case, but weaker (although still present) in the sodium case. It is interesting to note that the GC content dependence for m ($m = m_0 + \alpha f_{GC}$) implies a corresponding GC-content dependence for the salt correction to the entropy, $m\log([Mg^{2+}])/T_0 = \alpha f_{GC}\log([Mg^{2+}])/T_0$ where T_0 =298K. In the limit of very large N_{bp} only the linear logarithmic salt correction prevails in Eq.(S37) giving $\alpha = d\Delta H_0 T_0$. The linear fits in Figure S4 then give $\alpha = -0.04$kcal/mol (Mg^{2+}) whereas $d = 1.42 * 10^{-5}K^{-1}$ (see Table 2 in Ref. [16]) giving $\Delta H_0 = \alpha/(dT_0) = -9.45$kcal/mol per NN motif. This value should be compared with the average enthalpy per NN motif obtained from the 10 values of Δh_i^0 in our paper (third column of Table 2) which is equal to -9.16 kcal/mol, showing the compatibility of our results and those in Ref. [16]. A similar calculation for the sodium case gives $\Delta H_0 = \alpha/(d * T_0) = -1.95$kcal/mol per NN motif (note that in Na^+, $\alpha = -0.025$kcal/mol see Figure S4 , $d = 4.29 * 10^{-5}K^{-1}$ see Eq.(S37)) which

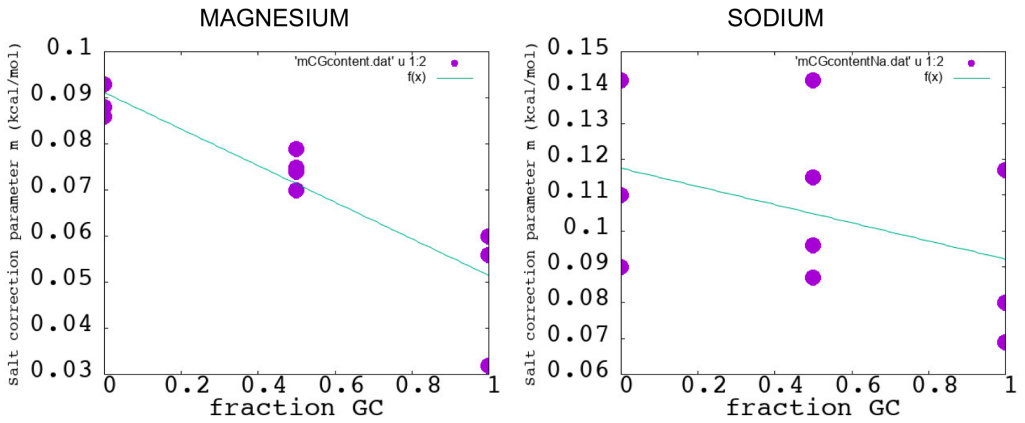

Figure S4: Salt correction parameter as a function of GC content of the different NN motifs. A correlation is clear for the magnesium case (left) as compared to sodium (right) except for the motif CC/GG that lies far from the correlation line. Fits to straight lines give $m = 0.09 - 0.04 f_{GC}$ ($\chi^2 = 7 * 10^{-5}$) for the magnesium case and $m = 0.12 - 0.025 f_{GC}$ ($\chi^2 = 6 * 10^{-4}$).

is completely off (the expected value is around -7.8kcal/mol). Summing up, our NN specific magnesium corrections to the free energy show a strong correlation with the GC content and are well compatible with the f_{GC} correction term in Eq.(S37) for the melting temperature correction formula of Ref. [16]. This is not true for sodium where there is a much weaker correlation between the NN specific sodium corrections to the free energy and the GC content. This suggests that, in the sodium case, either Eq.(22) in Ref. [17] or Eq.(4) in Ref. [17] do not capture the sequence dependence as it probably does the salt correction Eq.(22) in magnesium [16].

G An attempt to combine Na^+ and Mg^+ experiments into a single expression.

The correction formulas for Mg^{2+} buffers use the same generalized logarithmic expressions as for Na^+ buffers, with a simple adjustment wherein Mg^{2+} ion concentration is replaced by the equivalent Na^+ concentration, $[Na^+]_{eq} = \beta\sqrt{[Mg^{2+}]}$. The equivalent Na^+ concentration $[Na^+]_{eq}$ is defined as the con-

centration of sodium ions in a buffer that stabilizes duplexes to the same extent as the magnesium buffer does. The squared root factor term accounts for the law of mass action for cation binding activity: the binding rate of divalent cations to DNA is expected to scale proportionally to the square of the binding rate of monovalent cations. The following expression usually describes the combined effect of $[Na^+]$ and $[Mg^{2+}]$ ions:

$$[Mon^+]_{eq} = [Na^+]_{eq} + [Na^+] = \beta\sqrt{[Mg^{2+}]} + [Na^+] \tag{S38}$$

where $[Mon^+]_{eq}$ is the equivalent monovalent salt condition (as if all salt was $[Na^+]$). The Mfold server uses $\beta = 3.3$ when predicting the effect of the Mg^{2+} ions on DNA hybridization [20]. To account for the competitive effect between the different type of ions we generalize the previous expression and write:

$$[Mon^+]_{eq} = c \cdot [Mg^{2+}]^{\alpha} + [Na^+] \tag{S39}$$

with c a constant and α an exponent not necessarily equal to 1/2. Using this expression, we can write a single free energy salt dependency of the NN motifs (as in Eq.1 of main text) including contributions from both types of ions:

$$\Delta g_i = \Delta g_i^{\circ} - m_i \cdot \ln\left(c_i \cdot [Mg^{2+}]^{\alpha_i} + [Na^+]\right) \tag{S40}$$

where i=AA/TT, AC/TC,...,GA/CT is the index of the motif and c_i and α_i are motif dependent parameters.

If $[Mg^{2+}]$=0, Eq. S40 reduces to:

$$\Delta g_i = \Delta g_i^{\circ} - m_i \cdot \ln[Na^+] \tag{S41}$$

where Δg_i° and m_i are given in the left side of Table 1 (main text). On the other hand, if $[Na^+]$=0, then Eq. S40 is reduced to:

$$\begin{aligned} \Delta g_i &= \Delta g_i^{\circ} - m_i \cdot \ln\left(c_i \cdot [Mg^{2+}]^{\alpha_i}\right) \\ \Delta g_i &= \Delta g_i^{\circ} - m_i \cdot \ln c_i - m_i \cdot \alpha_i \cdot \ln [Mg^{2+}] \\ \Delta g_i &= \Delta g_i^{\circ}(\mathrm{Mg}) - m_i(\mathrm{Mg}) \cdot \ln [Mg^{2+}]) \end{aligned} \tag{S42}$$

where $\Delta g_i^{\circ}(\mathrm{Mg}) = \Delta g_i^{\circ} - m_i \cdot \ln c_i$ and $m_i(\mathrm{Mg}) = m_i \cdot \alpha_i$ are given in Table 1 (main text). Table S4 summarizes all the combined parameters. Equation (S40) should be seen as an attempt to interpolate all results of the

NNBP	Δg_i° (kcal/mol)	m_i (kcal/mol)	c_i	α_i
AA/TT	-1.24	0.142	24	0.60
AC/TG	-1.49	0.115	40	0.64
AG/TC	-1.33	0.087	254	0.81
AT/TA	-1.07	0.107	89	0.87
CA/GT	-1.76	0.096	73	0.83
CC/GG	-1.92	0.069	44	0.46
CG/GC	-2.34	0.117	14	0.49
GA/CT	-1.52	0.142	13	0.53
	ΔIF$^\circ$ (kcal/mol)	$m_{\rm IF}$ (kcal/mol)	$c_{\rm IF}$	$\alpha_{\rm IF}$
IF	0.40	0.05	0.003	-2.1

Table S4: Combined parameters for Na^+ and Mg^{2+} obtained from unzipping experiments.

unzipping experiments (Na^+ and Mg^{2+}) into a single expression valid for all salt concentrations. Figure S5 shows a three dimensional view of the free energy surface of different NN motifs plotted over the plane of monovalent-divalent salt concentrations and where the results from the unzipping experiments are shown as blue dots (Na^+ case) or red dots (Mg^{2+} case) connected by lines. Melting experiments suggest the existence of competitive effects between monovalent and divalent ions, which led Owczarzy and collaborators to define 4 different salt regimes [16], depending on the ratio $\sqrt{[Mg^{2+}]}/[Na^+]$. In contrast Mfold uses one single regime with homogeneous salt correction, an approximation which validity range is unknown. In contrast, our results can be described by a single logarithmic correction (Eq.(S40)) with heterogeneous salt corrections, which provides a simple framework accurately describing not only unzipping data but also melting data (Section H). If we define $F = (\Delta g_i - \Delta g_i^\circ)/m_i$ and we use Equation (S40), the resulting expression for F exhibits a logarithmic salt dependence for all the NN motifs at all salt concentrations: $F = -\ln[\mathrm{Mon}^+]_{\rm eq}$. Figure S6 shows the results of our experiments in an attempt to collapse all data into a single logarithmic curve. Although such relation captures the global trend of data, some NN motifs (e.g. CC/GG) appear to be incompatible with a hypothetical universal dependence of salt correction, as they deviate by a few standard

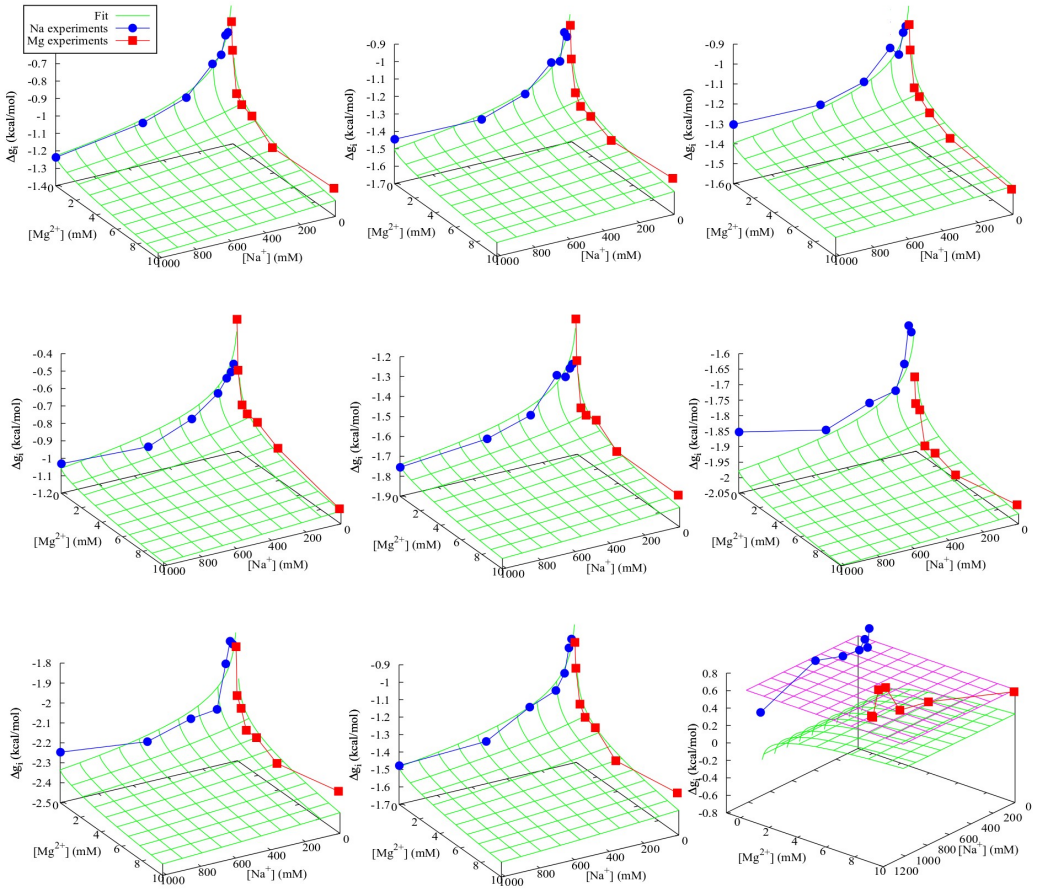

Figure S5: Explored salt regimes. NNB energies vs. different salt concentrations. In our unzipping experiments we have explored the boundaries ([Mg^{2+}]=0, [Na^{+}]=0). The green surface shows the extrapolation of our results to arbitrary monovalent and divalent salt condition.

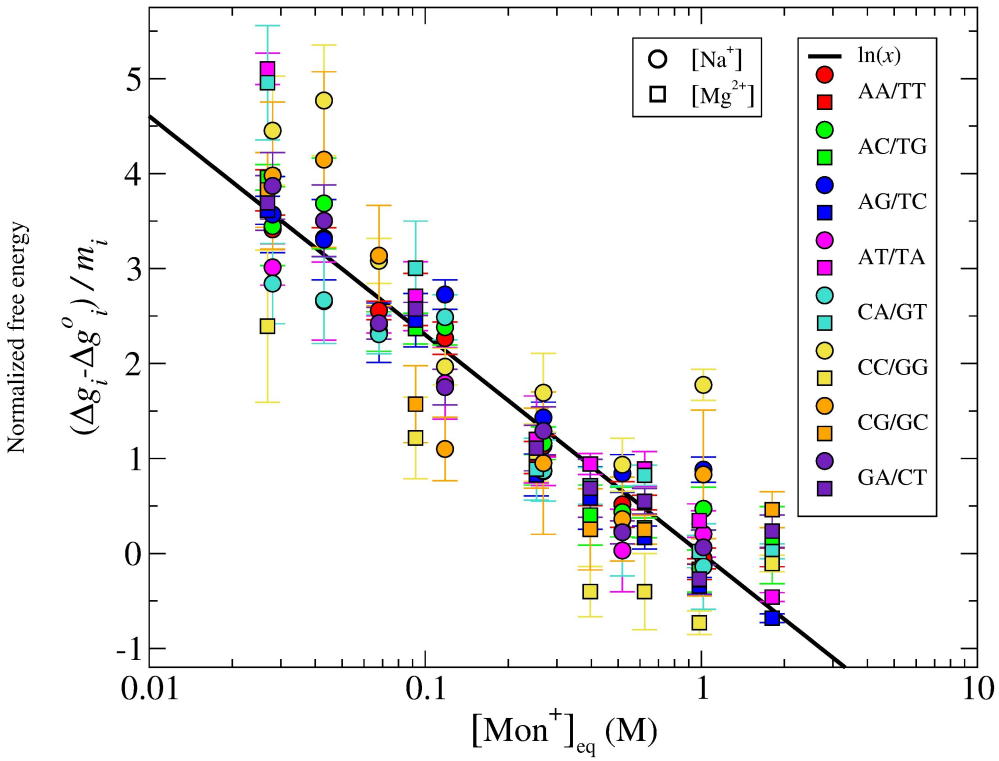

Figure S6: Normalized salt correction for both Mg^{2+} and Na^+. Some NNBP motifs do not collapse to a logarithm dependence.

deviations from the general rule.

H Fit of melting temperatures

To extract the enthalpies and entropies in Mg^{2+} of the NN motifs and initiation factors we have fit our predicted melting temperatures to the set of melting temperature data reported by Owczarzy and collaborators in Ref.[16]. As reported in that reference, melting temperatures increase monotonically with Mg^{2+} concentration reaching a maximum around 50mM and then decreasing again. Such trend cannot be reproduced within the scheme of linear logarithmic corrections of Eq.4 (main text). For this reason, our fits to melting temperature data are made only up to 10mM Mg^{2+} where the scheme of Eq.4 (main text) applies.

For that we calculate the melting temperature of a DNA duplex using the NN model and our thermodynamic values for the NN motifs. The enthaply (ΔH) and entropy (ΔS) contributions to the hybridization free energy are computed as:

$$\Delta H = \sum_i \Delta h_i + \mathrm{IF_h}(1) + \mathrm{IF_h}(N), \tag{S43}$$

$$\Delta S = \sum_i \Delta s_i + \mathrm{IF_s}(1) + \mathrm{IF_s}(N), \tag{S44}$$

where $\Delta h_i, \Delta s_i$ are the motif dependent NN enthalpy and entropy contributions and $\mathrm{IF_h}, \mathrm{IF_s}$ are the initiation factors as described in the main text. From these quantities the melting temperature, T_m, can be predicted as:

$$T_m = \frac{\Delta H}{\Delta S + R \ln[c_T/4]}, \tag{S45}$$

where c_T is the total oligo concentration. Using the previous calculation of T_m, we minimize the value of χ^2 in Eq. 3 in the main text. The minimization scheme we used is described in Eq.4 (main text) and illustrated in Table S5.

I Bootstrap re-sampling and independent validation dataset

Our enthalpy and entropy values are determined by combining our own free energy measurements with melting temperature data published in Ref. [16].

NNBP	Δg_i° (kcal/mol)	Δh_i° (kcal/mol)	Δs_i° (kcal·mol^{-1}·K^{-1})	m_i (kcal/mol)	m_h (kcal/mol)	m_{s_i} (kcal/mol)
aa/tt	$\Delta g^\circ_{\mathrm{aa/tt}}$	$\Delta h^\circ_{\mathrm{aa/tt}}$	$\Delta s^\circ_{\mathrm{aa/tt}} = \frac{\Delta h^\circ_{\mathrm{aa/tt}} - \Delta g^\circ_{\mathrm{aa/tt}}}{T}$	$m_{\mathrm{aa/tt}}$	$m_h = 0$	$m_{s_{\mathrm{aa/tt}}} = m_{\mathrm{aa/tt}}$
ac/tg	$\Delta g^\circ_{\mathrm{ac/tg}}$	$\Delta h^\circ_{\mathrm{ac/tg}}$	$\Delta s^\circ_{\mathrm{ac/tg}} = \frac{\Delta h^\circ_{\mathrm{ac/tg}} - \Delta g^\circ_{\mathrm{ac/tg}}}{T}$	$m_{\mathrm{ac/tg}}$	$m_h = 0$	$m_{s_{\mathrm{ac/tg}}} = m_{\mathrm{ac/tg}}$
ag/tc	$\Delta g^\circ_{\mathrm{ag/tc}}$	$\Delta h^\circ_{\mathrm{ag/tc}}$	$\Delta s^\circ_{\mathrm{ag/tc}} = \frac{\Delta h^\circ_{\mathrm{ag/tc}} - \Delta g^\circ_{\mathrm{ag/tc}}}{T}$	$m_{\mathrm{ag/tc}}$	$m_h = 0$	$m_{s_{\mathrm{ag/tc}}} = m_{\mathrm{ag/tc}}$
at/ta	$\Delta g^\circ_{\mathrm{at/ta}}$	$\Delta h^\circ_{\mathrm{at/ta}}$	$\Delta s^\circ_{\mathrm{at/ta}} = \frac{\Delta h^\circ_{\mathrm{at/ta}} - \Delta g^\circ_{\mathrm{at/ta}}}{T}$	$m_{\mathrm{at/ta}}$	$m_h = 0$	$m_{s_{\mathrm{at/ta}}} = m_{\mathrm{at/ta}}$
ca/gt	$\Delta g^\circ_{\mathrm{ca/gt}}$	$\Delta h^\circ_{\mathrm{ca/gt}}$	$\Delta s^\circ_{\mathrm{ca/gt}} = \frac{\Delta h^\circ_{\mathrm{ca/gt}} - \Delta g^\circ_{\mathrm{ca/gt}}}{T}$	$m_{\mathrm{ca/gt}}$	$m_h = 0$	$m_{s_{\mathrm{ca/gt}}} = m_{\mathrm{ca/gt}}$
cc/gg	$\Delta g^\circ_{\mathrm{cc/gg}}$	$\Delta h^\circ_{\mathrm{cc/gg}}$	$\Delta s^\circ_{\mathrm{cc/gg}} = \frac{\Delta h^\circ_{\mathrm{cc/gg}} - \Delta g^\circ_{\mathrm{cc/gg}}}{T}$	$m_{\mathrm{cc/gg}}$	$m_h = 0$	$m_{s_{\mathrm{cc/gg}}} = m_{\mathrm{cc/gg}}$
cg/gc	$\Delta g^\circ_{\mathrm{cg/gc}}$	$\Delta h^\circ_{\mathrm{cg/gc}}$	$\Delta s^\circ_{\mathrm{cg/gc}} = \frac{\Delta h^\circ_{\mathrm{cg/gc}} - \Delta g^\circ_{\mathrm{cg/gc}}}{T}$	$m_{\mathrm{cg/gc}}$	$m_h = 0$	$m_{s_{\mathrm{cg/gc}}} = m_{\mathrm{cg/gc}}$
ga/ct	$\Delta g^\circ_{\mathrm{ga/ct}}$	$\Delta h^\circ_{\mathrm{ga/ct}}$	$\Delta s^\circ_{\mathrm{ga/ct}} = \frac{\Delta h^\circ_{\mathrm{ga/ct}} - \Delta g^\circ_{\mathrm{ga/ct}}}{T}$	$m_{\mathrm{ga/ct}}$	$m_h = 0$	$m_{s_{\mathrm{ga/ct}}} = m_{\mathrm{ga/ct}}$
$\mathrm{IF_{at}}$	$\Delta g^\circ_{\mathrm{at}}$	$\Delta h^\circ_{\mathrm{at}}$	$\Delta s^\circ_{\mathrm{at}} = \frac{\Delta h^\circ_{\mathrm{at}} - \Delta g^\circ_{\mathrm{at}}}{T}$			
$\mathrm{IF_{cg}}$	$\Delta g^\circ_{\mathrm{cg}} = \Delta g^\circ_{\mathrm{at}} - \Delta g^\circ_{\mathrm{IF}}$	$\Delta h^\circ_{\mathrm{cg}}$	$\Delta s^\circ_{\mathrm{cg}} = \frac{\Delta h^\circ_{\mathrm{cg}} - \Delta g^\circ_{\mathrm{gc}}}{T}$			

Table S5: Dependencies of thermodynamic parameters. Red cells correspond to fixed values, as given by the unzipping experiments. Green cells correspond to fitting parameters, which are determined by minimizing the mean squared error in Eq. 3 (main text). Yellow cells correspond to dependent parameters, determined by the fixed (red) and the fitting parameters (green). Note that the free energy difference for the inititation factors, $\Delta g^\circ_{\mathrm{IF}}$, is marked as red because it has been measured in the unzipping experiments.

It is true that, as a result of such a fitting procedure, our newly determined enthalpies and entropies are prone to be biased towards predicting the main data published in that reference. It is therefore expected that predictions from unzipping data (red points in Figure 5b, main text) perform better than UO numbers (blue points in Figure 5b,main text). However our fits to melting temperatures are done by constraining the free energy of the different motifs, something that has not been done in bulk studies. To better understand how reliable is our approach we have carried out two additional tests. On the one hand we have carried out a bootstrap re-sampling of available data. On the other hand we have challenged our newly derived set of energy values with an independent set of 22 oligos at various salt conditions and a total of 58 melting temperatures.

I.1 Bootstrap re-sampling

We have performed 5744 bootstrap re-samplings of the data and our NN model has been fit to each subset of data. We have collected the fitting parameters (enthalpy, entropy and initiation factors) for all re-samplings and we have computed the mean and the standard deviation of all of them. The mean and standard deviation of the temperature errors for all the re-samplings is $3.2 \pm 0.2°C$. The table shiown in Fig. S7 summarizes the results:

We get an estimation of the error involved in the fit, which is approximately 0.3 kcal/mol for the enthalpies and 0.9 cal/(mol*K) for the entropies. Figures S8,S9 shows that the distribution of the fitting parameters after performing the bootstrap are nearly Gaussians centered on the mean providing an independent validation of our energy numbers.

I.2 Independent validation dataset

In order to check our energy numbers we have used a new independent validation datasets shown in Tables 4 and S2 (supplementary material) taken from Reference[16]. These tables combine sodium and magnesium, however the reliable data to us are those sequences where the parameter $R = \sqrt{[Mg^{2+}]}/[Mon^{+}]$ is greater than 0.22, a regime where magnesium competes

NNBP	Δg_i° (kcal/mol)	Δh_i° (kcal/mol)	Δs_i° (cal/mol/K)	m_i (kcal/mol)	m_h (kcal/mol)	m_{si} (kcal/mol)
AA/TT	-1.69	-8.4±0.3	-22.7±0.9	0.086	0	0.086
AC/TG	-1.91	-8.7±0.3	-22.6±0.9	0.074	0	0.074
AG/TC	-1.81	-7.6±0.3	-19.5±0.9	0.070	0	0.070
AT/TA	-1.55	-8.6±0.4	-23.7±0.9	0.093	0	0.093
CA/GT	-2.17	-10.0±0.2	-26.2±0.8	0.079	0	0.079
CC/GG	-2.18	-9.9±0.3	-25.9±0.9	0.032	0	0.032
CG/GC	-2.65	-11.0±0.3	-28.1±0.8	0.058	0	0.058
GA/CT	-1.88	-8.6±0.3	-22.5±0.9	0.075	0	0.075
GC/CG	-2.74	-11.2±0.3	-28.4±0.9	0.058	0	0.058
TA/AT	-1.38	-7.5±0.3	-20.4±0.9	0.088	0	0.088
IF_{at}	0.70±0.06	2.6±1.3	6±4			
IF_{cg}	0.76±0.06	2.5±1.2	6±4			

Figure S7: Bootstrap re-sampling analysis. New thermodynamic data for DNA in magnesium at 298K derived from bootstrap re-sampling from the melting temperature data corresponding to the 92 different oligos of length 10-30 bp at four different salt concentrations ($[Mg^{2+}]$=0.5, 1.5, 3, 10 mM) as described in the text.

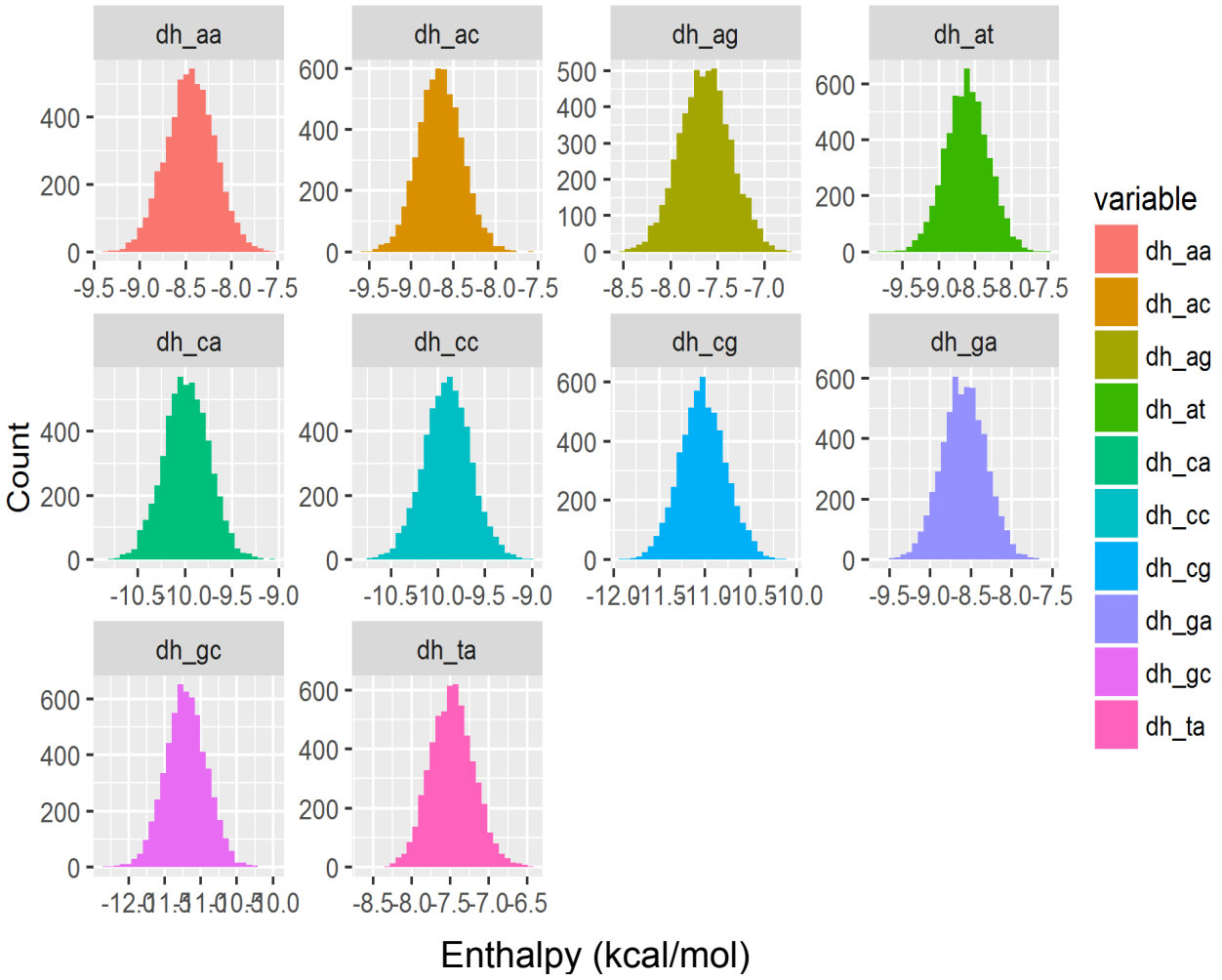

Figure S8: Distributions of enthalpies after bootstrap re-sampling.

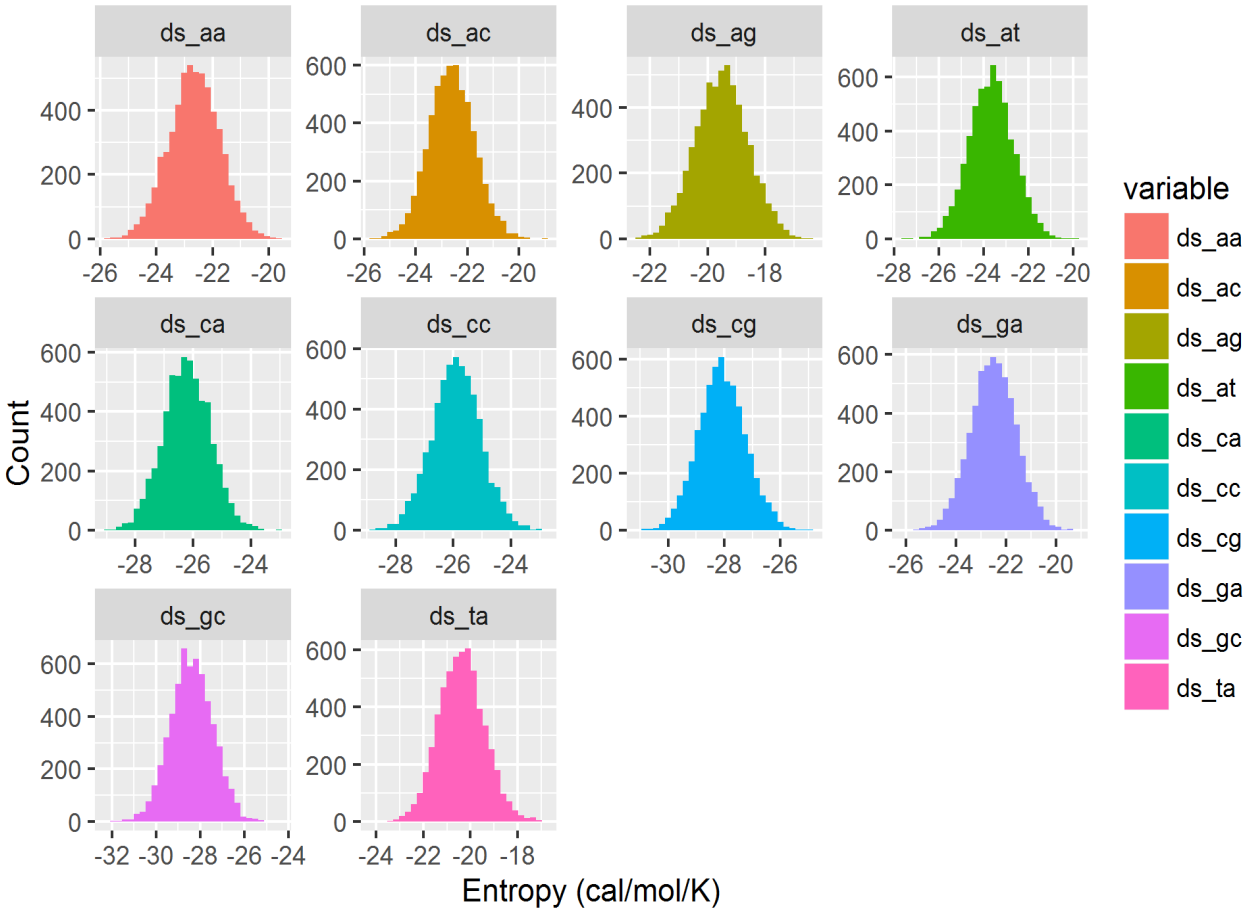

Figure S9: Distributions of entropies after bootstrap re-sampling.

with or dominates sodium. This regime corresponds also to our own experimental range of magnesium concentrations so we can use our energy numbers to predict the values of T_m in such regime. The results are shown in the Table displayed in Figure S10 and exhibit excellent prediction accuracy (χ^2 being lower than 3, see Figure S11). In Figure S12 we plot the experimentaql versus predicted melting temperatures according to salt concentration (upper panel) or source (lower panel).

Sequence (C_T=2uM) (Ref. a)	Mon(mM)	0.5mM	Pred	1.5mM	Pred	3mM	Pred	10mM	Pred
TTCTACCTATGTGAT	5	42.1	43.3	45.6	46.8	47.6	49.1	49.5	53.0
GCAGTGGATGTGAGA	5	51.9	51.7	54.7	55.0	56.2	57	58.2	60.6
CAGCCTCGTCGCAGC	5	61.1	60.3	64.2	63.2	65.1	65.1	67.2	68.4
TGATTCTACCTATGTGATTT	5	51.0	50.4	54.5	54.0	55.8	56.4	58.1	60.5
AGCTGCAGTGGATGTGAGAA	5	60.7	60.0	63.9	63.4	65.1	65.5	66.9	69.4
CAGCCTCGTTCGCACAGCCC	5	68.7	66.5	71.3	69.4	72.3	71.3	73.8	74.7
GTTCTATACTCTTGAAGTTGATTAC	5	56.3	53.8	59.4	57.7	60.7	60.2	62.6	64.6
CTGGTCTGGATCTGAGAACTTCAGG	5	65.1	62.1	67.6	65.6	68.5	67.8	70.1	71.8
CAGTGGGCTCCTGGGCGTGCTGGTC	5	73.3	71.1	75.9	74.0	76.6	75.9	78.0	79.2
CTTAAGATATGAGAACTTCAACTAATGTGT	5	60.5	57.5	63.2	61.4	64.4	63.9	65.9	68.4
AGTCTGGTCTGGATCTGAGAACTTCAGGCT	5	68.8	66.5	71.6	70.0	72.3	72.3	73.8	76.3
GACCTGACGTGGACCGCTCCTGGGCGTGGT	5	75.9	73.7	78.1	76.6	79.0	78.5	79.8	82
OTHER LABS		**Exp**	**Pred**	**Exp**	**Pred**				
GCCAGTTAA (C_T=8uM) (Ref.b)	10	29.5 (1mM)	33.1	35.4 (10mM)	40.4				
GTAGATCACT (C_T=7.3uM) (Ref.d)	10	38.7 (5mM)	37.6						
TCTTCTCTTTCT (C_T=0.1uM) (Ref.g)	25	47.8 (10mM)	46.8						
TTTCCCTTTCTT (C_T=0.1uM)(Ref.g)	25	48.2 (10mM)	48.4						
CTGACGACAAGACT (C_T=2uM)(Ref.a)	5	56.3 (10mM)	57						
AAAAAAAATATATAT (C_T=1.4uM)(Ref.i)	25	42.7 (15mM)	43						
TCTCAATGGTGTTACG (C_T=2uM)(Ref.a)	5	49.0 (.5mM)	49.7						
TAATTTAAAATTTTTAAAAAA (C_T=2uM) (Ref.k)	10	46.8 (2mM)	47.6						
AAAAAAAAAATAATTTTAAATATTT (C_T=2uM) (Ref.k)	10	50.8 (2mM)	50.9						
TTTTTTTTTTATTAAAATTTATAAA (C_T=2uM) (Ref.k)	10	50.3 (2mM)	50.8						

Figure S10: Independent validation dataset. (Caption next page.)

Figure S10: (Previous page.) **Independent validation dataset.** Melting temperature prediction from our unzipping energy numbers using melting temperatures from an independent validation dataset. Melting temperatures of different oligo sequences (first column) were measured in buffers containing a given amount of monovalent salt (second column) at different magnesium concentrations (0.5mM, third column; 1.5mM, fifth column; 3mM, seventh column; 10mM, ninth column). All sequences have $R = \sqrt{[Mg^{2+}]}/[Mon^{+}] > 0.22$, a regime where magnesium competes with or dominates sodium. This regime corresponds also to our own experimental range of magnesium concentrations. The first 12 sequences were obtained from Table 4 from Ref. [16] at a total oligo concentration of $C_T \sim 2um$. The rest 10 sequences are given in Table S2 of Ref. [16] and are obtained at different salt conditions and C_T (value indicated in the first column) in different labs (as indicated by letters a-k). References: a. [Owc08]. b. Nakano, S., Fujimoto, M., Hara, H., and Sugimoto, N. (1999) Nucleic Acids Res. 27, 2957-2965. c. Hudson, R. H. E., Uddin, A. H., and Damha, M. J. (1995) J. Am. Chem. Soc. 117, 12470-12477. d. Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P. E., Norden, B., and Graslund, A. (1996) J. Am. Chem. Soc. 118, 5544-5552. e. Gryaznov, S., and Schultz, R. G. (1994) Tetrahedron Lett. 35, 2489-2492. f. Sugiyama, T., Schweinberger, E., Kazimierczuk, Z., Ramzaeva, N., Rosemeyer, H., and Seela, F. (2000) Chem. Eur. J. 6, 369-378. g. Sugimoto, N., Wu, P., Hara, H., and Kawamoto, Y. (2001) Biochemistry 40, 9396-9405. h. Sund, C., Puri, N., and Chattopadhyaya, J. (1996) Tetrahedron 52, 12275-12290. i. Germann, M. W., Kalisch, B. W., and van de Sande, J. H. (1988) Biochemistry 27, 8302-8306. j. Hou, M-H., Lin, S-B., Yuann, J-M. P., Lin, W-C., Wang, A. H-J., and Kan, L-S. (2001) Nucleic Acids Res. 29, 5121-5128. k. Jovin, T. M., Rippe, K., Ramsing, N. B., Klement, R., Elhorst, W., and Vojtiskova, M. (1990) Structure and Methods 3, 155-174. i. Jain, A., Rajeswari, M. R., and Ahmed, F. (2002) J. Biomol. Struct. Dyn. 19, 691-699

Sample	χ^2	N
New data from Ref.[Owc08]	2.7	48
Others (from Refs[a-i], Table R3)	1.7	10
All new validation dataset	2.6	58

Figure S11: χ^2 analysis of the different melting temperature data prediction shown in previous Table. N is the total number of melting temperatures.

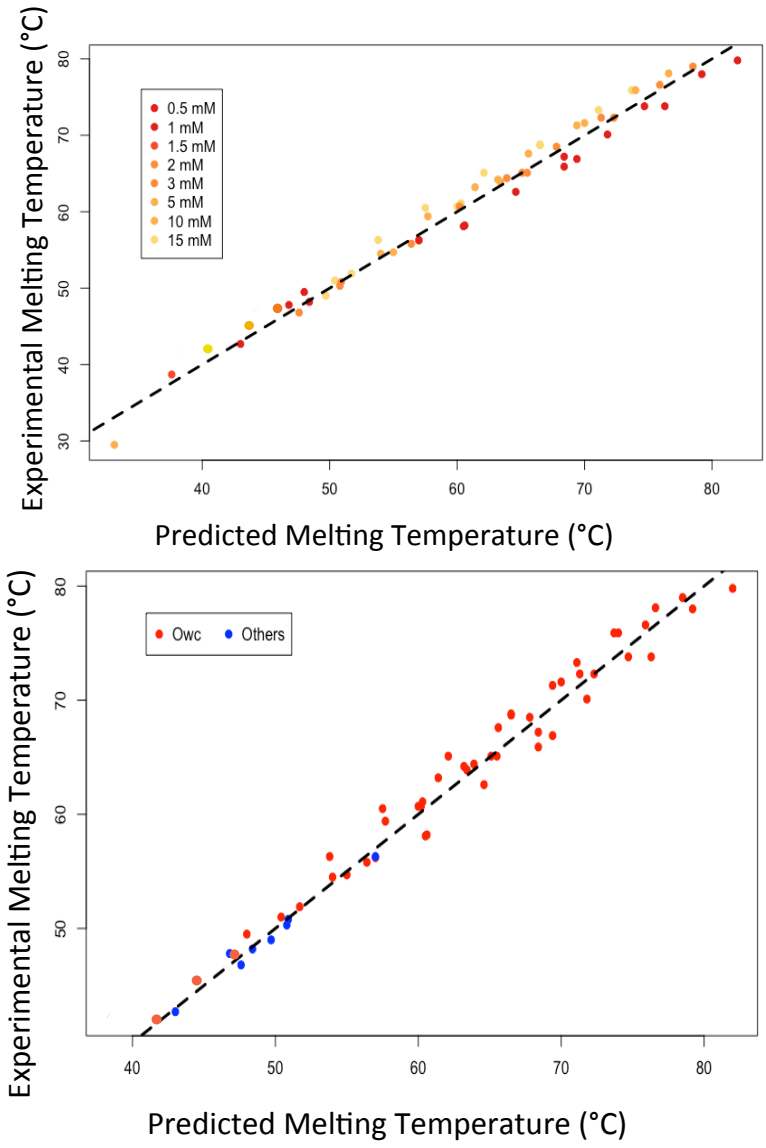

Figure S12: Experimental versus predicted melting temperature for the dataset of 58 independent measurements shown in the independent validation dataset table (Upper: classified by salt condition; Lower: classified from data source according to table shown in Figure S10 . Owc stands for the first 12 sequences obtained from Table 4 from Ref. [16] at a total oligo concentration of $C_T \sim 2um$ (see top table in Fig. S10). The rest 10 sequences are given in Table S2 of Ref. [16] and are obtained at different salt conditions in different labs (as indicated by letters a-k in caption of Figure S10, see "OTHER LABS" bottom table in Fig. S10).

J Derivation of enthalpies and entropies from melting data in Na^+ buffers.

The thermodynamic values for DNA duplex formation in Na^+ salt buffers are summarized in Table S6. These values were already obtained from unzipping experiments in Ref.[9] using the NN model with ten parameters. Table S6 shows the new results with eight parameters and initiation factors. As for the Mg^{2+} case the fits in the Na^+ case follow the scheme of Eq.4 (main text) illustrated in Table S5. Figure S13 shows a comparison between melting temperature data from Ref. [17] and prediction results from the UO values dataset of Santalucia [15], the ten NN energy values from unzipping experiments Ref.[9] and the new eight plus initiation factors values shown in Table S6.

The new results improve melting temperature data prediction over the numbers given in our previous work in Ref.[9] indicating better accuracy of the new scheme with eight parameters and initiation factors.

NNBP	Δg_i° (kcal/mol)	Δh_i° (kcal/mol)	Δs_i° (cal·mol^{-1}·K^{-1})	m_i (kcal/mol)
aa/tt	-1.24	-7.6	-21.2	0.142
ac/tg	-1.49	-7.1	-18.8	0.115
ag/tc	-1.33	-6.2	-16.4	0.087
at/ta	-1.07	-7.1	-20.4	0.110
ca/gt	-1.76	-9.0	-24.2	0.096
cc/gg	-1.92	-8.5	-22.0	0.069
cg/gc	-2.34	-9.6	-24.3	0.117
ga/ct	-1.52	-7.6	-20.3	0.142
GC/CG	-2.38	-9.8	-25.0	0.08
TA/AT	-0.98	-5.45	-15.3	0.09
IF_{at}	0.75	2.4	5.6	0
IF_{cg}	0.25	-2.8	-10.4	0

Table S6: Values of all the thermodynamic parameters after fitting melting temperatures of oligos in Na^+ buffer. These values were obtained from unzipping experiments using the NN model with eight parameters and initiation factors. The salt dependencies for enthalpies and entropies follows the scheme shown in Eq. 4 (main text).

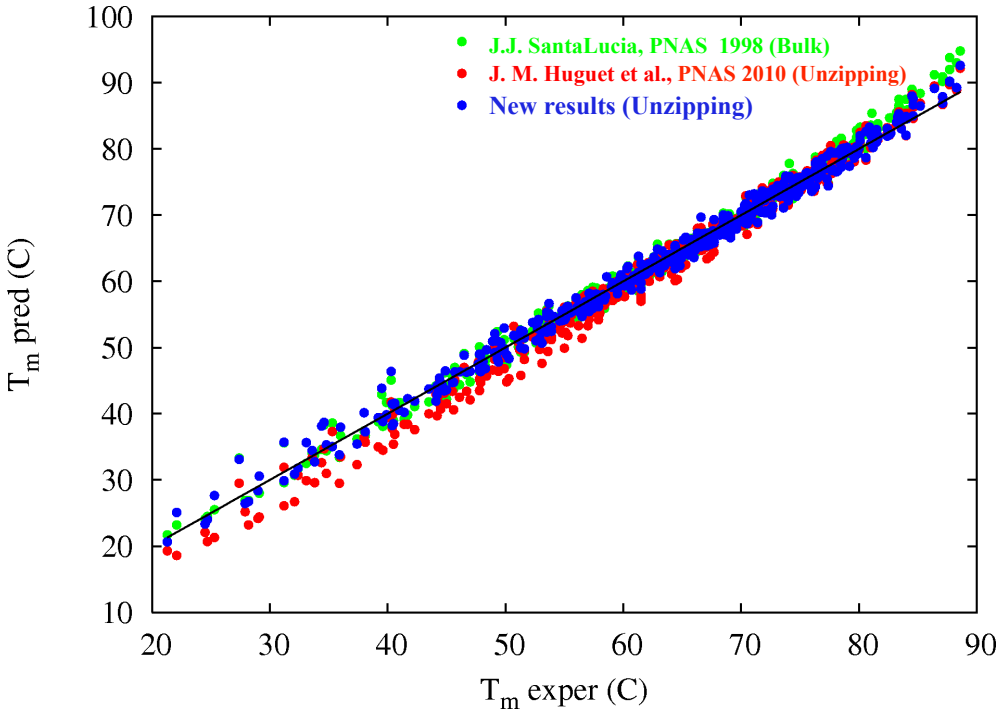

Figure S13: Improved melting temperature prediction in Na^+ buffer. Data shows experimental melting temperature data taken from Ref. [17] compared to predictions from the UO dataset from SantaLucia [15] (green points), from unzipping data using the NN model with ten parameters [9] (red points) and the new improved energy values reported in Table S6 obtained from the NN model with eight parameters and initiation factors obtained from unzipping experiments (blue points).

References

[1] Alemany, A Bizarro, C. V. Ritort, F. (2012) Non-specific binding of Na^+ and Mg^{2+} to RNA determined by force spectroscopy methods. *Nucleic Acids Research 40*, 6922-6935

[2] Shoemaker, D. D. et al. (2001) Experimental annotation of the human genome using microarray technology. *Nature 409*, 922-927

[3] Rothemund, P. W. K. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature 440*, 297-302.

[4] Douglas, S. M., et al. (2009) Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature 459*, 414-418.

[5] Dorsett, Y. and Tuschl, T. (2004) siRNAs: applications in functional genomics and potential as therapeutics. *Nat. Rev. Drug. Discov 3*, 318-329.

[6] Nykypanchuk, D., Maye, M. M., van der Lelie, D., Gang O. (2008) DNA-guided crystallization of colloidal nanoparticles. *Nature 451*, 549-552.

[7] Devoe, H., Tinoco, I. J. (1962) The stability of helical polynucleotides: base contributions. *J. Mol. Biol. 4*, 500-517.

[8] Crothers, D. M. and Zimm, B. H. (1964) Theory of the melting transition of synthetic polynucleotides: evaluation of the stacking free energy. *J. Mol. Biol. 9*, 1-9.

[9] Huguet J. M., Bizarro C. V., Forns N., Smith S. B., Bustamante C. and Ritort F. (2010) Single-molecule derivation of salt-dependent base-pair free energies in DNA it Proc. Nat. Acad. Sci. (USA) 107, 15431-15436

[10] Holbrook, J. A., Capp, M. W., Saecker, R. M. and Record, M. T. J. (1999) Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry 38*, 8409-8422.

[11] Goldstein, R.F. and Benight A.S. (1992) How many numbers are required to specify sequence-dependent properties of polynucleotides? *Biopolymers 32*, 1679-1693

[12] Licinio, P. and Guerra, J. C. O. (2007) Irreducible representation for nucleotide sequence physical properties and self-consistency of nearest-neighbor dimer sets. *Biophys. J. 92*, 2000-2006.

[13] De Lorenzo, S. Ribezzi-Crivellari, M Smith, S. B. Ritort, F. (2015) A Temperature-Jump Optical Trap for Single-Molecule Manipulation *Biophys. J, 108*, 2854-2864.

[14] Ritort F. (2006) Single-molecule experiments in biological physics: methods and applications *J. Phys.: Condens. Matter 18*, R531-R583.

[15] SantaLucia Jr., J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA 95*, 1460-1465.

[16] Owczarzy, R. Moreira, B. You, Y. Behlke, M. and Walder, J. (2008) Predicting Stability of DNA Duplexes in Solutions Containing Magnesium and Monovalent Cations *Biochemistry 47*, 5336-5353.

[17] Owczarzy, R. Moreira, B. You, Y. Manthey, J.A. Huang, L., Behlke, M.A. and and Walder, J. (2004) Effects of4 sodium ions on DNA duplex oligomers: improved predictions of melting temperatures *Biochemistry 43*, 3537-3554.

[18] Rouzina, I. and Bloomfield, V. (1999) Heat Capacity Effects on the Melting of DNA. 1. General Aspects. *Biophysical Journal 77*, 3242-3251.

[19] Junier, I. Mossa, A. Manosas, M. and Ritort. F (2009) Recovery of the free energy of branches in single molecule experiments. *Phys. Rev. Let. 102*, 070602.

[20] Peyret, N. (2000) Prediction of Nucleic Acid Hybridization: Parameters and Algorithms, *Ph.D. Thesis, Section 5.4.2*, 128. Wayne State University, Detroit, MI.