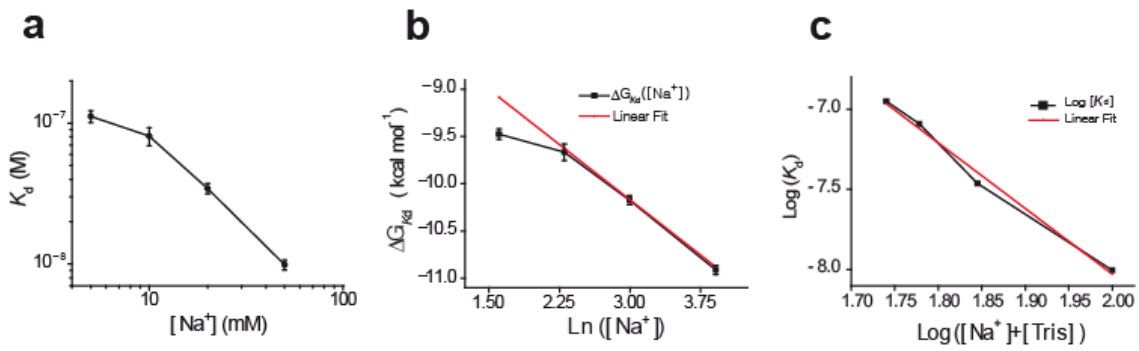
A rule of seven in Watson-Crick base pairing of mismatched sequences

Ibrahim I. Cisse[1,3], Hajin Kim[1,2], Taekjip Ha[1,2]

[1]Department of Physics and Center for the Physics of Living Cells, University of Illinois at Urbana- Champaign, Urbana, Illinois US 61801. [2]Howard Hughes Medical Institute, Urbana, Illinois US 61801. [3]Present address: 46 Rue d'Ulm, Laboratoire Kastler Brossel, département de Physique & Institue de biologie de l'Ecole Normale Superieure, Ecole Normale Supérieure, Paris 75005, France. Correspondence should be addressed to T.H. email: tjha@illinois.edu



Supplementary Figure 1: Salt dependence of K_d is consistent with the prediction from a thermodynamic model. (**a**) Log-Log representation of the salt dependence of K_d for the full 9 bp DNA. The tail-end (10 ~ 50 mM [Na^+]) shows a linear signature consistent with the salt-dependence in a nearest-neighbor thermodynamic model[1]. (**b**) Gibbs free energy associated with the measured K_d plotted against the natural logarithm of sodium concentration*. The tail end best fits the linear expression (plotted in red)**: - 7.83 - 0.78×Ln([Na^+])**, in good agreement with the nearest-neighbor thermodynamic model**. (**c**) Full linear dependence of K_d in Log-Log representation was obtained by considering contribution from the buffer electrolyte, suggesting the possible compensating ionic effect of the buffer electrolyte below 10 mM [Na^+]***.

* For K_d measured, the corresponding Gibbs free energy was calculated using the formula:

$$\Delta G = - RT \ln(1/K_d) = RT \ln(K_d),$$

where $R = 1.98588^* 10^{-3}$ kcal K^{-1} mol^{-1} is the molar gas constant, $T = 296$ K is room temperature (23 °C) in Kelvin, Ln denotes the natural logarithm. Here the equilibrium constant was estimated as ~ $1/K_d$. Error bars were propagated from the errors in K_d measurement (σ_{Kd}) using the estimation: $\sigma_{\Delta G} \sim RT\sigma_{Kd}/K_d$.

The general functional form of the salt dependence calculated from the thermodynamic data of 26 oligonucleotides (from 10 mM and 300 mM [NaCl]) at 37°C can be expressed in the linear form as[1]

$$\Delta G_{37\,^\circ C}([Na^+]) = \Delta G_{37\,^\circ C}(1\text{ M [NaCl]}) - 0.114\times N\times \ln([Na^+])\text{ (kcal mol}^{-1}),$$

where $\Delta G_{37\,^\circ C}$ (1 M [NaCl]) is the free energy at 37 °C, as predicted from the unified nearest-neighbor thermodynamic parameters at 1 M [NaCl], and N is the number of base pairs in the duplex.

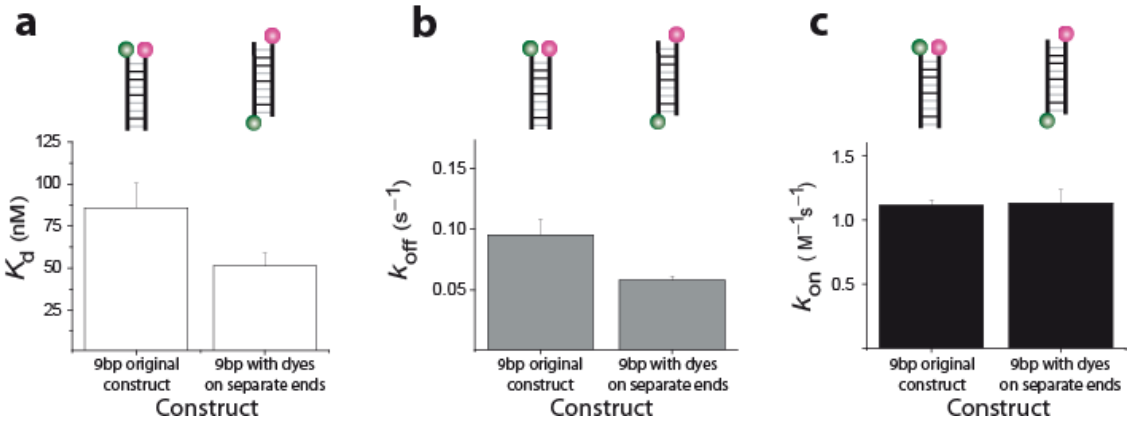
For our full 9 bp data at room temperature, fitting over the linear range (10 ~ 50 mM [Na^+]) yielded a similar form: $\Delta G_{23\,^\circ C}([Na^+]) = -7.83 - 0.78\times\ln([Na^+])$ (kcal mol^{-1}).

** We note that, despite the difference in temperature (room temperature vs 37 °C) and modifications for dye labeling, a good agreement was obtained through direct comparison of our fit and the unified thermodynamic function, in particular:

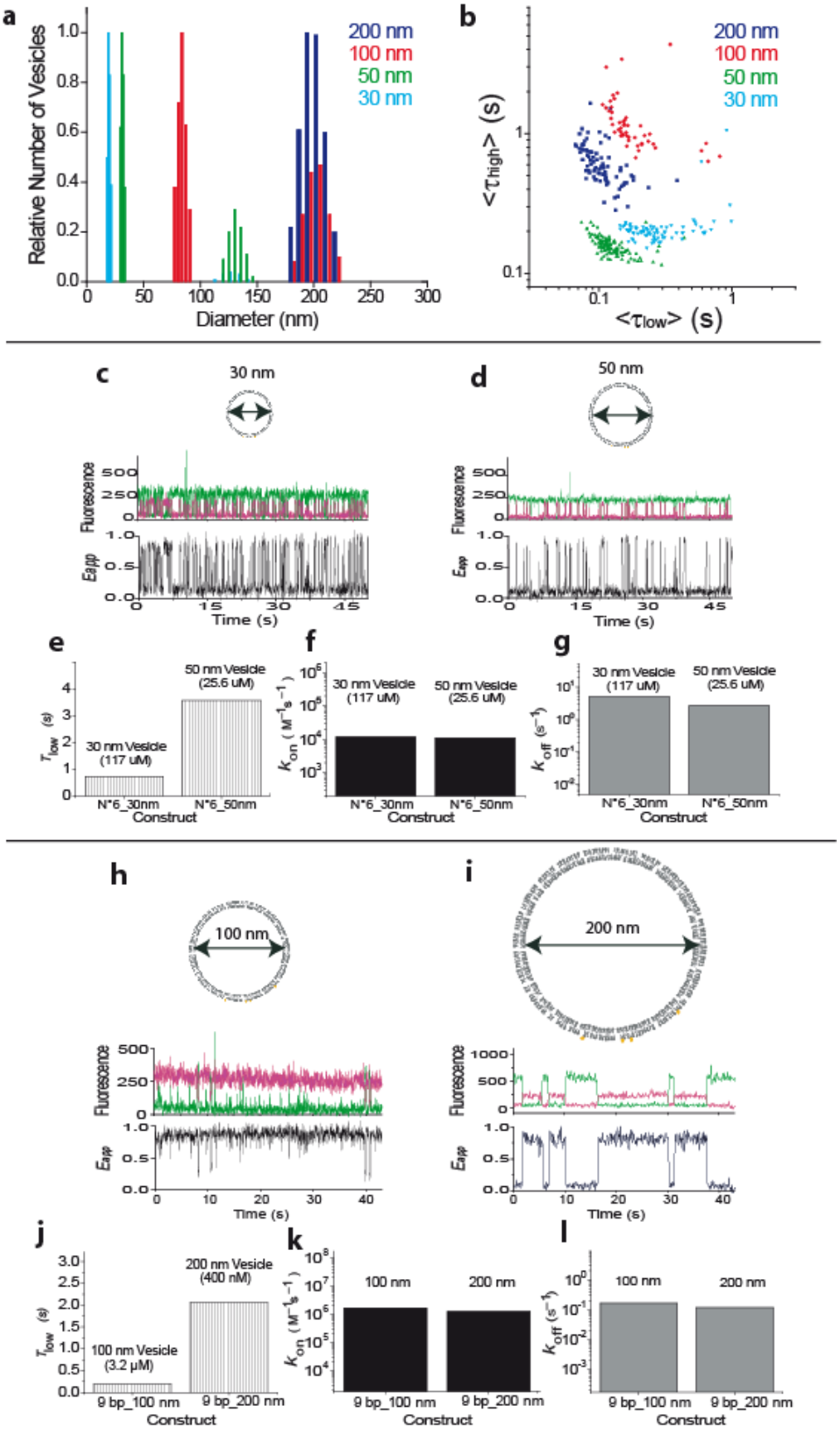
- The slope[1]: $(1/N)\times\partial\Delta G/\partial \mathrm{Ln}([Na^+])$ = - 0.78/9 = - 0.087 kcal mol^{-1} for our data at 23 °C is well in the range of - 0.114 ± 0.033 kcal mol^{-1} found from thermodynamic database at 37 °C.

- The intercept: $\Delta G_{23\ ^\circ C}$ (1 M [Na^+]) = - 7.8 ±0.1 kcal mol^{-1} compares well to the estimated $\Delta G_{37\ ^\circ C}$ (1 M [NaCl]) = - 9.04 kcal mol^{-1} from the unified nearest-neighbor thermodynamics prediction solely using the 9 nt DNA sequence[3].

*** Below 10 mM [Na^+], there was a significant deviation from the thermodynamic model. We postulated that at such low salt concentrations, contribution from background electrolytes (for example buffer ions, which are typically ignored in thermodynamic models) could influence the measured kinetics. To check for a possible effect of the buffer ion (Tris) on the expected linear dependence of K_d in Log-Log plot, we first estimated the contribution of Tris ion to correspond roughly one-to-one to that of Na^+, based on a study quantifying the relative effect of various electrolytes[2]. When we investigated the dependence of K_d on the total cationic strength, [Na^+] + [Tris], we observed that the full range of measured data points falls on the same linear form: $\mathrm{Log}(K_d) = 0.2 - 4.1\times\mathrm{Log}([Na^+] + [\mathrm{Tris}])$. We conclude therefore that the apparent deviation at 5 mM [Na^+] could be due to the buffer electrolyte.

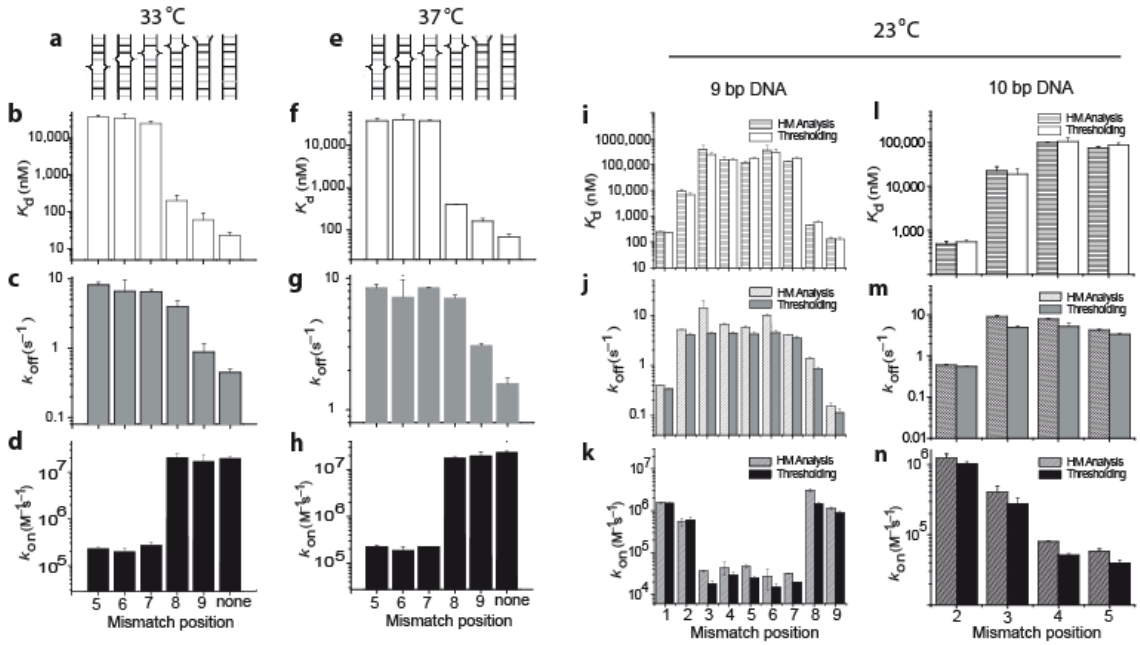


Supplementary Figure 2: Effect of dyes on measured kinetics. (**a**–**b**)For the full 9 bp DNA, lower K_d (**a**) and k_{off} (**b**) were obtained by labeling dyes on separate duplex ends compared to the original construct with the dyes on the same duplex end. (**c**)The annealing rate k_{on} is unaffected by dye position.



Supplementary Figure 3: Vesicle size characterization. (**a**) Estimated vesicle diameters measured by deconvolution from dynamic light scattering data of vesicles

prepared with various filter sizes; in some cases vesicle aggregates may appear as lower amplitude peaks at integer multiples of the estimated vesicle diameter. (**b**) Representative heterogeneity plots are illustrated in; each data point represents the average dwell times in high FRET and low FRET states, $<\tau_{high}>$ and $<\tau_{low}>$ respectively, measured from a single vesicle, and the color code corresponds to the filter size as in **a**; the DNA samples represented in **b** were the following: the full 9 bp at 37 °C, 200 mM [Na^+] for 200 nm vesicle diameter, N° 2 mismatch construct at 23 °C, 10 mM [Na^+] for 100 nm diameter, N° 7 mismatch construct at 33 °C, 150 mM [Na^+] for 50 nm vesicle diameter, and N° 5 (G-G) mismatch at 23 °C, 10 mM [Na^+] for 30 nm vesicle diameter. The variations in $<\tau_{low}>$ are similar to the variations in $<\tau_{high}>$ in Log-Log scale and the variations in $<\tau_{low}>$ are less than an order of magnitude for each sample. **(c–l)** Represented are comparisons of the measured kinetics in different vesicle sizes for the mismatched DNA construct N° 6 (**c**–**g**) and the full 9 bp construct (**h**–**l**); The average low FRET dwell times were much longer for larger vesicles (**e, j**) but after normalization by the effective local concentration C_{eff} the measured annealing rates, $k_{on} = (\tau_{low} C_{eff})^{-1}$, were identical for the same DNA construct (**f, k**). (**g, l**) k_{off} values were identical for the same DNA construct regardless of the vesicle size.



Supplementary Figure 4: Comparative analysis of kinetic rates dependence on mismatch position for different conditions (**a**–**h)** Kinetic rates were measured for constructs with a mismatch in positions 5 – 9 and for the full 9 bp DNA, as illustrated in panels **a** or **e** for the following salt and temperature conditions: (**b** through **d**) 33 °C, 150 mM [Na^+], (**f** through **h**) 37 °C, 200 mM [Na^+]. (**b**–**h**)Represented are K_d (**b**, **f**), k_{off} (**c**, **g**), and k_{on} (**d**, **h**), with the full 9 bp (without mismatch) marked as 'none'. (**c,g)** Similar to the room temperature data, k_{off} shows a gradual dependence on mismatch position, and (**d,h)** k_{on} shows a step function with about two orders of magnitude difference depending on whether the construct preserves at least 7 contiguous base pairs. (**i**–**n)** Comparative kinetic rate analysis with Hidden Markov modeling (dashed bars) vs. thresholding (solid bars) for 9 bp (**i**–**k**) and 10 bp (**l**–**n**) DNA constructs.

* Thresholding analysis includes a two-state cut-off at the apparent FRET efficiency $E_{app} = 0.5$, and a minimum of three-points are needed to validate a transition in order to decrease the likelihood of picking false transitions (for example due to photophysics). For constructs where the optimal vesicle size results in a transition with a dwell time of the order of our camera's acquisition time, missed events may increase the apparent dwell times and affect the apparent rate. Such events can be factored into the rate determination in the cut-off-free hidden Markov modeling[3]. The comparison between hidden Markov modeling and threshold based analysis on the same sets of data shows small variations that do not affect our conclusions based on 10 ~ 100 fold mismatch-dependent variations.

Supplementary Table 1: Oligonucleotides sequences and design

5'- /5Cy5/AGGACTTGT -3'	* 9 nt Cy5 DNA strand
5'- /5Cy5/AGGAGTTGT -3'	9 nt Cy5 DNA for #5 mismatch (G-G) if complemented to the full 9bp Cy3 strand
5'- ACAAGTCCT/3Cy3Sp/ -3'	* Full 9 bp Cy3 strand
5'- TCAAGTCCT/3Cy3Sp/ -3'	N° 1 mismatch strand for 9 nt DNA
5'- AGAAGTCCT/3Cy3Sp/ -3'	N° 2 mismatch strand for 9 nt DNA
5'- ACTAGTCCT/3Cy3Sp/ -3'	N° 3 mismatch strand for 9 nt DNA
5'- ACATGTCCT/3Cy3Sp/ -3'	N° 4 mismatch strand for 9 nt DNA
5'- ACAACTCCT/3Cy3Sp/ -3'	N° 5 mismatch strand (C-C) for 9 nt DNA
5'- ACAAGACCT/3Cy3Sp/ -3'	N° 6 mismatch strand for 9 nt DNA
5'- ACAAGTGCT/3Cy3Sp/ -3'	N° 7 mismatch strand for 9 nt DNA
5'- ACAAGTCGT/3Cy3Sp/ -3'	N° 8 mismatch strand for 9 nt DNA
5'- ACAAGTCCA/3Cy3Sp/ -3'	N° 9 mismatch strand for 9 nt DNA
5'- rUrCrCrCrUrGrArG/3Cy5Sp/ -3'	** Cy5 RNA strand (8 nt seed sequence of *Lin-4 miRNA /miR125*)
5'- /5Cy3/rCrUrCrArGrGrGrU -3'	N°1RNA_7cont.bp strand (target in *TP53*)
5'- /5Cy3/rCrUrCrArGrGrUrA -3'	N°2RNA_6cont.bp strand
5'- /5Cy3/ACAAGTCCT -3'	9 nt strand for dyes on separate ends when complemented to the 9 nt Cy5 DNA strand
5'- /5Cy5/GTCTAGTGAT -3'	*** 10 nt Cy5 DNA strand
5'- AACACTAGAC/3Cy3Sp/ -3'	N° 2 mismatch strand for 10 nt DNA
5'- ATGACTAGAC/3Cy3Sp/ -3'	N° 3 mismatch strand for 10 nt DNA
5'- ATCTCTAGAC/3Cy3Sp/ -3'	N° 4 mismatch strand for 10 nt DNA
5'- ATCAGTAGAC/3Cy3Sp/ -3'	N° 5 mismatch strand for 10 nt DNA

* The 9 nt DNA sequence was designed to melt at room temperature. Melting temperature T_m was determined to be 23.3 °C using the formula $(T_m + 273) = (R/\Delta H \ln C + \Delta S/\Delta H)^{-1}$ where C = 400 nM corresponds to the concentration of one molecule in the vicinity of another within a vesicle of 200 nm diameter, with ΔH and ΔS corresponding to the sequence dependent enthalpies and entropies at 50 mM $[Na^+]$ as predicted using the database of parameters from nearest-neighbor thermodynamic models for DNA oligonucleotides[4].
** As described in the main text, the RNA sequences represent the miR125 seed sequence and its corresponding target in human tumor protein *TP53* gene.
*** The 10 nt sequence was designed similarly to the 9 nt DNA as explained above, with a predicted melting temperature of 23.0 °C.

All oligonucleotides were purchased from **Integrated DNA Technology (Coralville, IA 52241**), with the company's modification-specific nomenclature exactly as presented in the first column of the Table above.

References

1. SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* **95**, 1460-5 (1998).
2. Bai, Y. et al. Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *Journal of the American Chemical Society* **129**, 14981-14988 (2007).
3. McKinney, S.A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical journal* **91**, 1941-1951 (2006).
4. Allawi, H.T. & SantaLucia, J., Jr. Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry* **36**, 10581-94 (1997).