Supplemental Text

Efficiency of halting

While halted transcription complexes contain a single RNA per DNA after washing, we see that during the transcription reaction, T7 RNAP is likely transcribing until it reaches Tus, whereupon it either halts or terminates. In the situations presented here, we have separated the transcription complexes from the other components of the transcription reaction that allow the cycle of halting and termination to continue; in the EMSA the unincorporated NTPs are removed from the complexes quickly due to their high mobility through the gel, in the cases of halting on beads, the beads were washed, and on the GAIIx, the flowcell was washed. In these cases, we have shown that halted complexes with functional RNAs are very stable. However, we see that during transcription, RNA produced by polymerase that terminates at the ter site accumulates. When we transcribe DNAs bound to beads, a large amount of transcript remains anchored to beads, but an even larger fraction is released into the supernatant (**Supplementary Fig. 1a**). Bacterial RNA polymerases have been shown to terminate through a cooperative forward translocation mechanism when presented by roadblocks[38]. Thus, removal of the components of the transcription reaction from the halted transcription complexes is likely to be at least partially responsible for the stability that we observe.

The experiment in **Supplementary Figure 1a** also demonstrated the polarity of Tus. In a parallel reaction, the templates used were identical to those in the first except that their ter sites were in the opposite orientation. In this case, T7 RNAP is able to break through the barrier of Tus and run off the end of the template in most cases; this template had two ter sites in the permissive orientation and only minor bands corresponding with each ter site are seen (**Supplementary Fig. 1a**). Therefore, when Tus is bound to DNA in the non-permissive orientation, nearly every T7 RNAP either halts or terminates, while in the opposite orientation, most polymerases proceed past the Tus/ter complex.

We believe that the technique for transcription halting presented in this work will be suited to a wide range of applications. It is the first technique, to our knowledge, capable of producing stable complexes containing both a DNA molecule and its RNA transcript. We have presented one case where it was coupled with high-throughput DNA sequencing technology to analyze RNA properties at the time of sequencing.

Interpreting HiTS-RAP Data

We subjected the large-scale data to the most logical filtering routine that we could devise. We have used Illumina's quality scores to eliminate so called mutants that are likely

caused by sequencing errors: a mutant cluster is included only if the mutated base has a quality score of 25 or greater, which corresponds to a <0.003 probability of an incorrect base call. This is a more stringent quality score filtering criteria than the common practice in SNP calling algorithms[34], which often use quality score >20 (<0.01 probability of an incorrect base call) as a threshold. Without this filter, the data do not reflect many of the important changes in affinity caused by mutations that we have found, as sequencing errors result in inclusion of miscalled wild type aptamers in the average binding curves of a mutant. This problem is exacerbated near the end of the sequencing read where the error rate is highest. Fortunately, we have found that quality filtering eliminates this problem, and we are confident in the mutants that we have called. However, we cannot absolutely rule out the contribution of sequencing errors to our analysis of mutant aptamer clusters.

Other considerations in judging the reliability of a measurement are the number of clusters identified for a mutant, and the K_d determined for it. Average binding curves are less smooth for mutants with fewer clusters. Thus, we see that the variation in lower copy number mutants is greater (both in the covariance of fitted K_ds and in variability between lanes). This and a concern about sequencing errors were the major factors in our decision to only examine mutants with at least ten clusters in a lane. The K_d of a mutant also limits our ability to make an accurate measurement of its affinity. Our binding curves only have seven points: we only expect to be able to fit good K_ds for mutants that bind in that range. The ranges used were decided based on the known K_ds of the two aptamers. We wanted our assay to be most accurate around those known K_ds. Mutants that saturate binding too early or late in the curve are more difficult to fit reliably. This is why special consideration had to be taken in analyzing GFPapt mutants, as many of these failed to bind in the range of concentrations that we assayed. For NELF-E, an RNA binding protein, non-specific RNA binding was high enough that failure to saturate signal in the range assayed was not a concern. Similarly, for both aptamers, attempting to assay a mutant whose K_d is closer to the lower end of our dynamic range, bound by the protein concentrations used, is problematic. In conclusion, both of these limitations indicate that the most accurate K_d measurements are for high copy number mutants whose affinity is near the center of the concentrations for which we have measured fluorescence intensities.

Strategies for solving K_ds

We have determined that a single weighted nonlinear regression to an average binding curve representative of a unique sequence is the best way to solve dissociation constants. In this way of fitting, the intensities of each individual clusters' binding curve are normalized for cluster size and position[14] by dividing by the average intensity of that cluster during sequencing. After this normalization, each average intensity is representative of each sequence's behavior at a given concentration. The standard deviation of the clusters' intensity gives a good measure of confidence in how that average intensity is truly representative of the RNA's behavior; thus, it is

used as the weight in the fit. We find that median and average binding curves give similar dissociation constants, especially considering that this is a case where we are measuring point mutants, so we expect many of the measured affinities to be near that of the canonical aptamer.

We have also used a method of fitting where we solve a dissociation constant for each cluster individually in a nonlinear regression. This method truly exploits the data to its full potential, as each cluster is itself an independent binding event. Rather than giving each unique sequence only one opportunity to converge on a good fit to the Hill equation, this method gives these sequences as many fits as there are clusters. The dissociation constant for a unique sequence is then the geometric mean of these fitted K_ds. While we find that the geometric mean of all fitted K_ds is usually accurate, individual clusters tend to contain a large amount of noise and therefore give poor fits. Because of this, we use only those whose fits return a low covariance when taking the mean for a unique sequence. A major disadvantage of this method, however, is that it is computationally expensive. When doing fits to average binding curves, we take only unique sequences with a certain number of clusters in the flowcell. This amounts to only several thousand fits per lane of the flowcell. Individual fits, however, require several million fits per lane. The fact that these two methods give similar results means the extra time and complication of individual fits is unnecessary.

HiTS-RAP as a Tool for Identification, Characterization, and Optimization of Aptamers

HiTS-RAP can be used to optimize aptamers that are selected by SELEX. Although they have enormous diversity (up to 10^{16} unique sequences), libraries used for SELEX are far from having complete coverage of the potential sequence space (70 nt random library = 1.4×10^{42} unique sequences). Therefore, SELEX identifies the highest affinity aptamers present within the starting library, but the true highest affinity aptamer could be a related sequence that was not present. In addition, even the most enriched aptamer pools at the end of a SELEX process with 10 rounds of selection often have a large complexity with thousands to millions of different aptamers[39]. The highest affinity aptamers in such pools are not necessarily the most abundant ones either, therefore, a high-throughput analysis method, like HiTS-RAP, would be useful in identifying the highest affinity aptamers for downstream applications. While large-scale, quantitative characterization of DNA aptamer libraries has been carried out[40], no such method has been developed for RNA libraries. HITS-RAP could be used to rapidly identify the true highest affinity RNA aptamers. This point is well demonstrated by identification of GFP aptamer mutants that have higher affinity than the canonical aptamer (i.e. C58U and U60A). The original selection of GFPapt was exceptionally difficult, requiring a laborious two stage selection process with 27 total rounds of selection, followed by mutational analysis, and structure minimization. Therefore, the starting, unmutated GFP aptamer that was used in this study had already undergone a substantial optimization. Also, HITS-RAP could be modified in ways to select aptamers for additional functions such as disruption of a specific interaction of the target molecule. Such applications would be valuable given the fact that many aptamers which disrupt their target proteins' function or interaction with other molecules are in clinical trials[41,42].

Supplementary References

38. Epshtein, V., Toulmé, F., Rahmouni, A. R., Borukhov, S. & Nudler, E. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J.* **22,** 4719–27 (2003).

39. Schütze, T. *et al.* Probing the SELEX process with next-generation sequencing. *PLoS One* **6,** e29604 (2011).

40. Cho, M. *et al.* Quantitative selection and parallel characterization of aptamers. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 18460–5 (2013).

41. Ni, X., Castanares, M., Mukherjee, A. & Lupold, S. E. Nucleic acid aptamers: clinical applications and promising new horizons. *Curr. Med. Chem.* **18,** 4206–14 (2011).

42. Sundaram, P., Kurniawan, H., Byrne, M. E. & Wower, J. Therapeutic RNA aptamers in clinical trials. *Eur. J. Pharm. Sci.* **48,** 259–71 (2013).

Supplementary Figure Legends:

Supplementary Figure 1: Transcription halts or terminates at Tus-bound ter sites in the non-permissive orientation

a. Radiolabeled *in vitro* transcription of Tus-bound DNA templates. Biotinlyated GFP aptamer DNA templates contained two ter sites, either in the non-permissive (NP) or permissive (P) orientation. The DNA templates were bound to streptavidin-agarose beads (Pierce) and then bound by Tus protein. After transcription for 30 minutes at 37^{o}C in the presence of radiolabeled CTP, beads and supernatant were separated and run on an 8% denaturing PAGE (7 M urea). Most transcription does not proceed past ter site in the non-permissive orientation. 16% of the transcripts produced in the non-permissive ter reaction remained bound to beads in halted transcription complexes, while the rest terminated and released to the supernatant. RNA transcripts of a single, distinct length, consistent with halting or termination at the ter site but not before, produced in these reactions can be explained by forced termination of leading T7 RNAP(s) stopped at the ter site by the trailing ones, thus leaving a single halted T7 RNAP on the template at the ter site[38]. RNAs in the supernatant were produced by polymerases that terminated at the Tus-bound ter site or passed through and ran off the end of the template. In the permissive orientation, the vast majority of the transcripts produced are run off transcripts, and thus are found in the supernatant. Only 6% of the transcripts remained bound to the beads, and most of these are run off transcripts, indicating that they are likely due to non-specific binding to beads by full-length RNAs which are no longer engaged in halted transcription complexes. **b.** EMSA of halted transcription complexes with radiolabeled DNA. Complexes produced after transcription halting contain RNA. This experiment is identical to that of Figure 1, except that no SUPERase In (Ambion) was used and the addition of a lane that was treated with RNase cocktail (Ambion) after transcription halting. The RNase treatment caused the band of intermediate mobility created by transcription halting to slow to a mobility even less than that of Tus bound DNA, presumably a consequence of the loss of full length charged RNA and presence of bound T7 RNAP (the lanes are all from a single gel, thus mobilities can be directly compared).

Supplementary Figure 2: EGFP only binds to transcription complexes with full-length GFP aptamer RNA

GFP aptamer halting templates with two ter sites were immobilized on glutathione beads through the GST tagged Tus protein. Single round transcription was initiated by incubating beads in a transcription solution lacking CTP. One sample was then chased with all four nucleotides including CTP, allowing T7 RNAP to transcribe up to the Tus bound ter site, while the other sample was not, leaving an initiated polymerase with an 11 nt transcript. The transcription complexes present in both samples contain identical components of DNA template, GST-tagged Tus protein, T7 RNAP, and the glutathione beads, but differ in that the ‘initiated control” has

only an 11 nt transcript while the 'chased' has the full aptamer RNA. EGFP only binds to 'chased' sample that presents the full-length GFP aptamer RNA. Thus, EGFP labeling of halted transcription complexes is due to an interaction with the full-length RNA transcript.

Supplementary Figure 3: Schematic of HiTS-RAP templates
Different sequence elements needed in DNA templates for HiTS-RAP are labeled. Illumina adaptors located at the extreme 5' and 3' ends are used to generate clusters on the sequencer. The sequence of interest in flanked by a T7 RNA polymerase promoter at the 5'end, and the Illumina sequencing primer and ter site at the 3' end. Thus, during transcription, T7 RNAP initiates at its promoter transcribes through the sequence of interest (the GFP aptamer in this case), and then begins to transcribe the Illumina sequencing primer before it encounters the non-permissive Tus bound ter site and halts transcription. Thus, the entire GFP aptamer RNA should have emerged from the polymerase. The approximate positions on the template of Tus protein and T7 RNAP in a halted transcription complex are shown.

Supplementary Figure 4: Corrections applied to HiTS-RAP data.
a. Stability of halted transcription complexes modeled by linear approximation of exponential decay. GFP aptamer RNA presenting clusters were imaged at equilibrium with 625 nM EGFP-mOrange nine successive times following a single transcription halting reaction. The average raw, non-normalized fluorescence intensity of bound EGFP-mOrange at canonical GFP aptamer clusters in three lanes (~2.5 million clusters per lane) are plotted in blue, with error bars representing the standard deviation between the three lanes. The reported decay rate is the average ± standard deviation of the fits from the three lanes. The fitted line is plotted as a straight blue line.
Decay rate corrected intensities (see Online Methods) were plotted in red together with a line fit to these intensities (red straight line). **b-d.** HiTS-RAP binding curves for GFPapt and the eight mutants verified by EMSA (see **Supplementary Fig. 5**) in each of the three lanes (L005-L007). Not every mutant is plotted in every lane: it is excluded if scipy.optimize.curve_fit failed to converge on a fit. Each sequence is colored the same in each lane. The number of clusters and percent increase in signal is reported for each lane. Percent increase is the percent increase in fluorescence signal between the average of the first and last two measured intensities in the average binding curve. Intensities decrease across the flowcell due to photobleaching. Sequences with lower affinity tend to saturate at lower fluorescence intensities. These data were used to determine the threshold for signal increase considered as not binding, for assigning sequences in each lane affinities of >125 nM. G8A_U56C (violet), a sequence which was confirmed as having an affinity of >125 nM by EMSA has a 2.8% increase in signal in lane 5 and a 2.6% decrease in lane 7 (and was not present in lane 6). G68C (purple), a mutant with a confirmed affinity of 75 nM, increased by 12.4, 5.8, and 2.5 % in lanes 5, 6, and 7. Thus, we determined that sequences with an increase of less than 3% in signal should be scored as not binding within the detection limit our assay given the concentrations of protein used: therefore

they were assigned a K_d of >125 nM (the second highest protein concentration used). If this occurred in only a subset of lanes, it was assigned a K_d of 125 nM for the lanes where it did not bind, and averaged with the rest.

Supplementary Figure 5: Confirmation of HiTS-RAP measured affinities by EMSA
a. EMSA of fluorescently-labeled canonical GFP aptamer and eight mutants with varying affinities. Duplicate EMSA (left panels) were done with varying GST-EGFP concentration in three-fold increments from 0.03 to 625 nM. The bottom band corresponds to unbound RNA, and the top band to EGFP bound RNA. A representative scan of the GST-EGFP used in EMSA, visualized by EGFP fluorescence, is shown for the last EMSA gel (bottom left panel). Fraction of aptamer bound by EGFP, quantified from gels, are plotted (black dots in right panels) and fitted to the Hill equation (red solid line) to determine the K_ds. Minimum and maximum fraction of aptamer bound, Hill coefficient and the K_d of the fits are reported in the figure legend for wt (canonical), C58U, U60A, G7A, C76U, U34C, A66U, G68C, G8A_U56C mutant GFP aptamers.
b. Agreement between HiTS-RAP and EMSA. Affinities measured by EMSA in (a) are plotted against their HiTS-RAP counterpart as log(K_d, nM). Mutants scored as not binding in at least one lane (or in all lanes as is the case with G8A_U56C) are plotted as open circles. The r^2 of 0.64 is for all eight mutants and the GFP aptamer.

Supplementary Figure 6: Secondary structure predictions of GFP aptamer single point mutants with higher or lower affinities
a) Predicted secondary structures of the two highest affinity GFP aptamer single point mutants. C58U (1.21 ×/ 1.03 nM K_d) and U60A (1.71 ×/ 1.13 nM K_d) mutants are predicted to have structures that are very similar to that of wild-type GFP aptamer. Compared to the wild-type aptamer (Inset, at bottom), U60A mutant has a slight expansion of the terminal loop (6 nt vs. 4 nt) and a concurrent shortening of the preceding stem (4 bp vs. 5 bp) in stem-loop#3. b) Predicted secondary structures of the two lower affinity GFP aptamer single point mutants. Both G45C (>125 nM K_d) and C57A (>125 nM K_d) single point mutants are predicted to have substantially altered secondary structures when compared to the wild-type aptamer structure (**Fig. 3b**). Mutated nucleotides are colored in red. Folded structures and the folding free energies are predicted by mFold.

Supplementary Figure 7: Secondary structure predictions of GFP aptamer double point mutants with substantially higher or lower affinities than predicted by the single point mutants

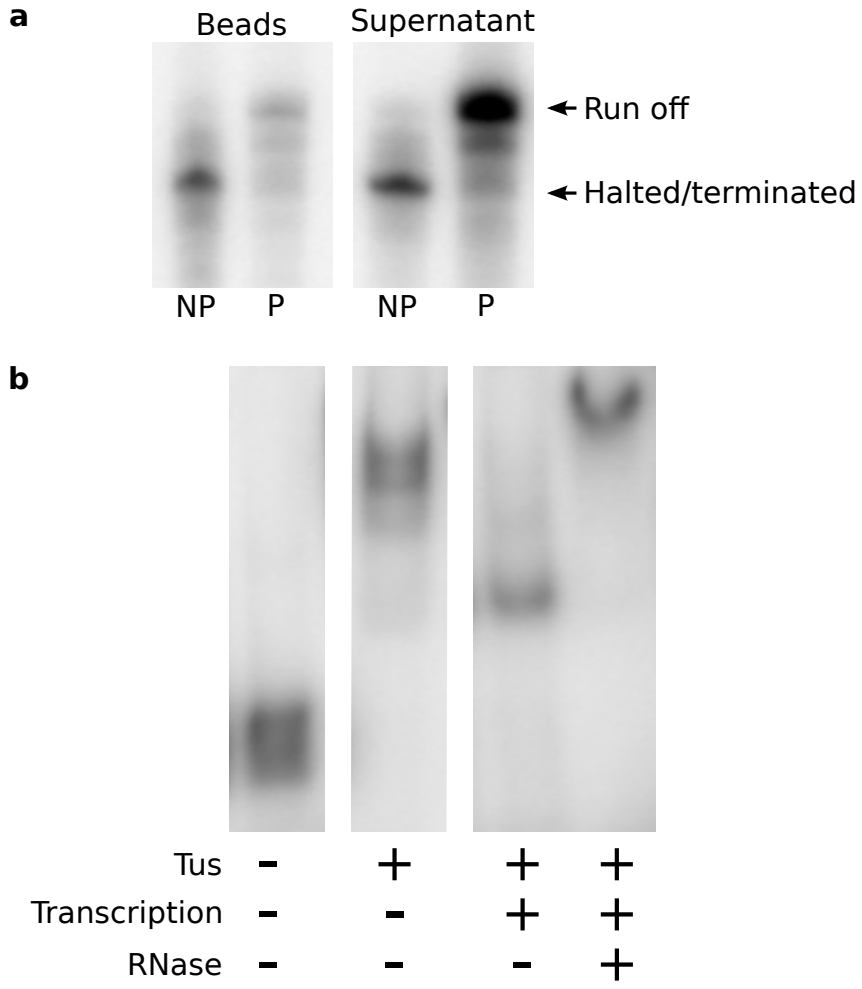
a) Predicted secondary structures of the two single point mutants and the corresponding double mutant GFP aptamer with better affinity than predicted from single point mutants. A23G single point mutant (5.87 ×/ 1.23 nM K_d) and A23G_U34C double point mutant (1.23 ×/ 1.95 nM K_d) are predicted to have wild-type GFP aptamer-like structures, whereas U34C single point mutant (24.21 ×/ 1.58 nM K_d) is predicted to have altered structure. A23G_U34C double point mutant is expected to have 33.3 ×/ 1.6 nM K_d based on A23G and U34C single point mutants' effect on affinity. Predicted secondary structure of wild type GFPapt. (inset, at the bottom) is included for comparison) b) Predicted secondary structures of the two single point mutants and the corresponding double mutant GFP aptamer with worse affinity than predicted from single point mutants. U34C single point mutant (24.21 ×/ 1.58 nM K_d) and U34C_C58U double point mutant (68.24 ×/ 1.63 nM K_d) are predicted to have altered secondary structures, whereas C58U single point mutant (1.21 ×/ 1.03 nM K_d) has a wild-type GFP aptamer-like structure. U34C_C58U double mutant is expected to have 6.85 ×/ 1.28 nM K_d based on U34C and C58U single point mutants' effect on affinity. Mutated nucleotides are colored in red. Folded structures and the folding free energies are predicted by mFold.

Supplementary Figure 8: EMSA confirmation of HiTS-RAP affinities for NELFapt.
Affinities of two NELFapt mutations within the k-turn were verified by EMSA. These experiments were carried out using the minimal (35 nt) aptamer with one GC base pair added due to the need for a G rich region at the beginning of a T7 transcription template, as was done in the original publication[17]. The minimal aptamer binds with lower affinity than observed by HiTS-RAP when measured by EMSA (50 nM for the minimal aptamer vs 8.5 nM for the full length by EMSA). Consistent with HiTS-RAP, A59G binds NELF with lower affinity (48 nM for the full length by HiTS-RAP, 360 nM for the minimal version by EMSA), and G58C binds with even lower affinity (65 nM for the full length by HiTS-RAP, 640 nM for the minimal version by EMSA). Thus, these two bases are as important as those within the NBE.

Supplementary Figure 9: Secondary structure predictions of NELF-E aptamer double point mutants withsubstantially higher or lower affinities than predicted by the effects of single point mutants.
a) Predicted secondary structures of the two single point mutants and the corresponding double mutant NELF-E aptamer with better affinity than predicted from single point mutants. A39G single point mutant (8.97 nM K_d) and A39G_U61C double point mutant (8.89 nM K_d) are predicted to have wild-type NELF-E aptamer-like structures, whereas U61C single point mutant (63.68 nM K_d) is predicted to have altered structure where 3'-end of the NBE and the k-turn inducing GA nucleotides, colored in blue, are in a stem. A39G_U61C double mutant is expected to have 109.04 nM K_d based on A39G and U61C single point mutants' effect on affinity. Predicted secondary structure of wild type NELFapt. (inset, at the bottom) is included for

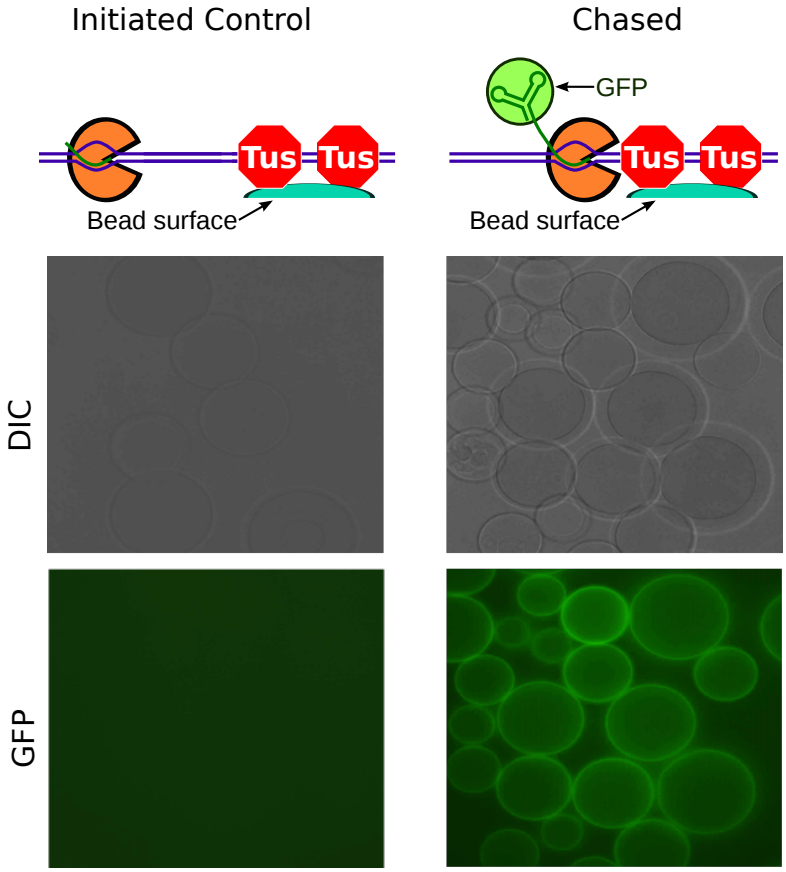
comparison) **b)** Predicted secondary structures of the two single point mutants and the corresponding double mutant NELFapt with worse affinity than predicted from single point mutants. A39G single point mutant (8.97 nM K_d) is predicted to have wild-type NELF-E aptamer-like structure. Both G63A single point mutant (12.81 nM K_d) and A39G_G63A double point mutant (87.06 nM K_d) are predicted to have altered secondary structures where the NBE and the k-turn GA are partially or completely in a stem. A39G_G63A double mutant is expected to have 21.93 nM K_d based on A39G and G63A single point mutants' effect on affinity. Mutated nucleotides are colored in red, and NBE and k-turn inducing nucleotides are colored in cyan. Folded structures and the folding free energies are predicted by mFold.

Supplementary Figure 1 | Transcription halts or terminates at Tus-bound ter sites in the non-permissive orientation



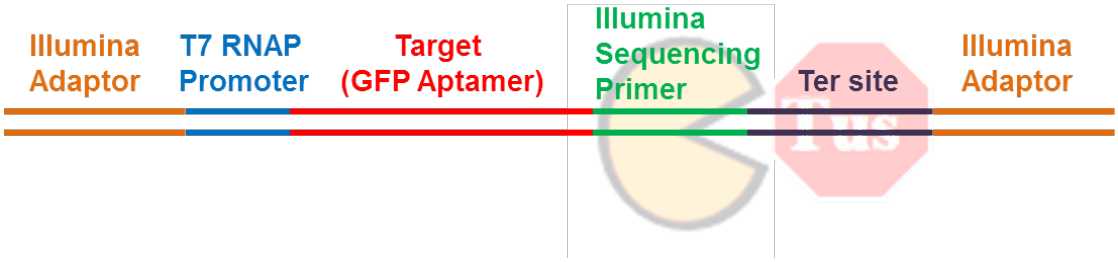
a. Radiolabeled in vitro transcription of Tus-bound DNA templates. Biotinlyated GFP aptamer DNA templates contained two ter sites, either in the non-permissive (NP) or permissive (P) orientation. The DNA templates were bound to streptavidin-agarose beads (Pierce) and then bound by Tus protein. After transcription for 30 minutes at 37oC in the presence of radiolabeled CTP, beads and supernatant were separated and run on an 8% denaturing PAGE (7 M urea). Most transcription does not proceed past ter site in the non-permissive orientation. 16% of the transcripts produced in the non-permissive ter reaction remained bound to beads in halted transcription complexes, while the rest terminated and released to the supernatant. RNA transcripts of a single, distinct length, consistent with halting or termination at the ter site but not before, produced in these reactions can be explained by forced termination of leading T7 RNAP(s) stopped at the ter site by the trailing ones, thus leaving a single halted T7 RNAP on the template at the ter site38. RNAs in the supernatant were produced by polymerases that terminated at the Tus-bound ter site or passed through and ran off the end of the template. In the permissive orientation, the vast majority of the transcripts produced are run off transcripts, and thus are found in the supernatant. Only 6% of the transcripts remained bound to the beads, and most of these are run off transcripts, indicating that they are likely due to non-specific binding to beads by full-length RNAs which are no longer engaged in halted transcription complexes.
b. EMSA of halted transcription complexes with radiolabeled DNA. Complexes produced after transcription halting contain RNA. This experiment is identical to that of Figure 1, except that no SUPERase In (Ambion) was used and the addition of a lane that was treated with RNase cocktail (Ambion) after transcription halting. The RNase treatment caused the band of intermediate mobility created by transcription halting to slow to a mobility even less than that of Tus bound DNA, presumably a consequence of the loss of full length charged RNA and presence of bound T7 RNAP (the lanes are all from a single gel, thus mobilities can be directly compared).

Supplementary Figure 2 | EGFP only binds to transcription complexes with full-length GFP aptamer RNA



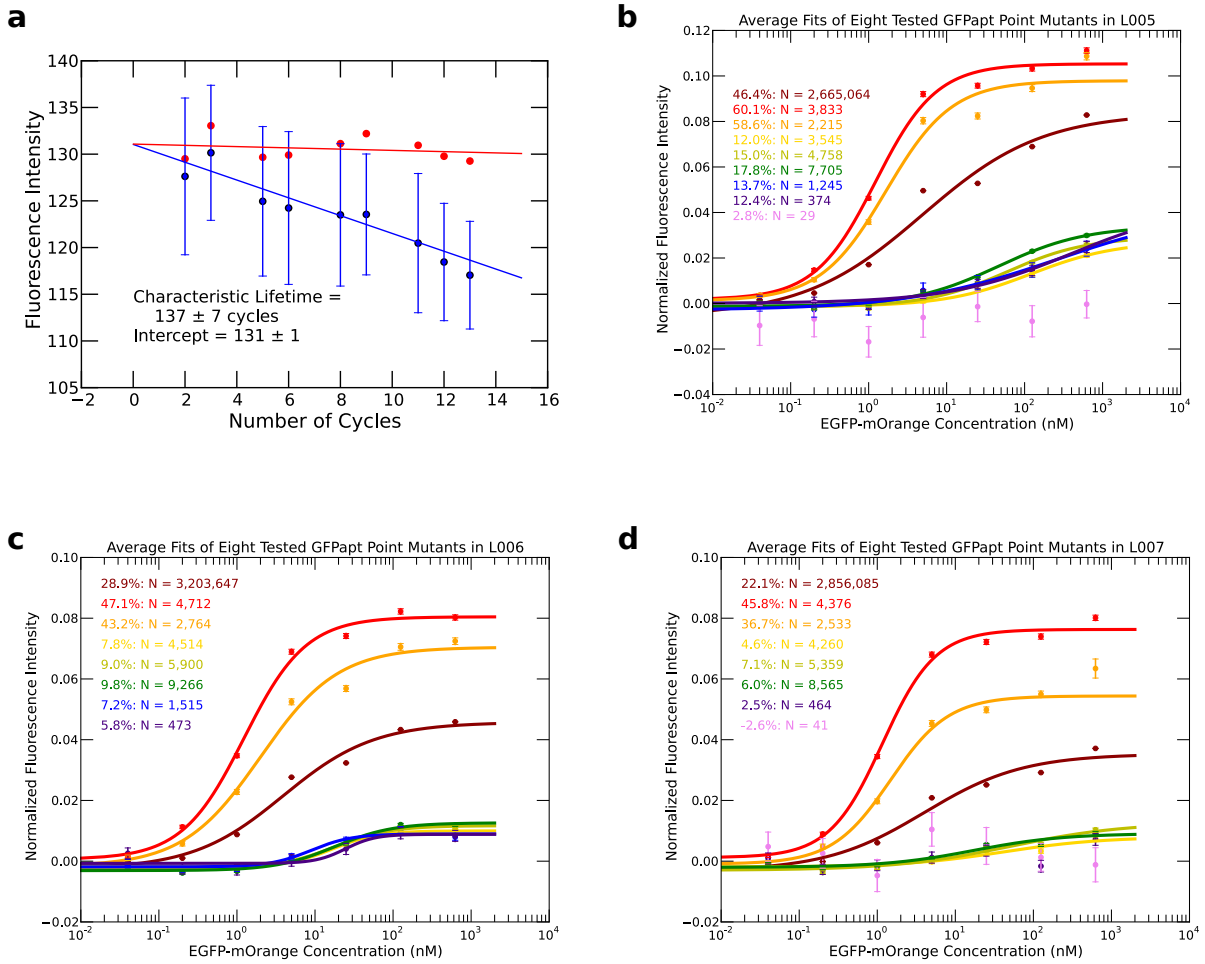
GFP aptamer halting templates with two ter sites were immobilized on glutathione beads through the GST tagged Tus protein. Single round transcription was initiated by incubating beads in a transcription solution lacking CTP. One sample was then chased with all four nucleotides including CTP, allowing T7 RNAP to transcribe up to the Tus bound ter site, while the other sample was not, leaving an initiated polymerase with an 11 nt transcript. The transcription complexes present in both samples contain identical components of DNA template, GST-tagged Tus protein, T7 RNAP, and the glutathione beads, but differ in that the ‘initiated control” has only an 11 nt transcript while the ‘chased’ has the full aptamer RNA. EGFP only binds to ‘chased’ sample that presents the full-length GFP aptamer RNA. Thus, EGFP labeling of halted transcription complexes is due to an interaction with the full-length RNA transcript.

Supplementary Figure 3 | Schematic of HiTS-RAP templates



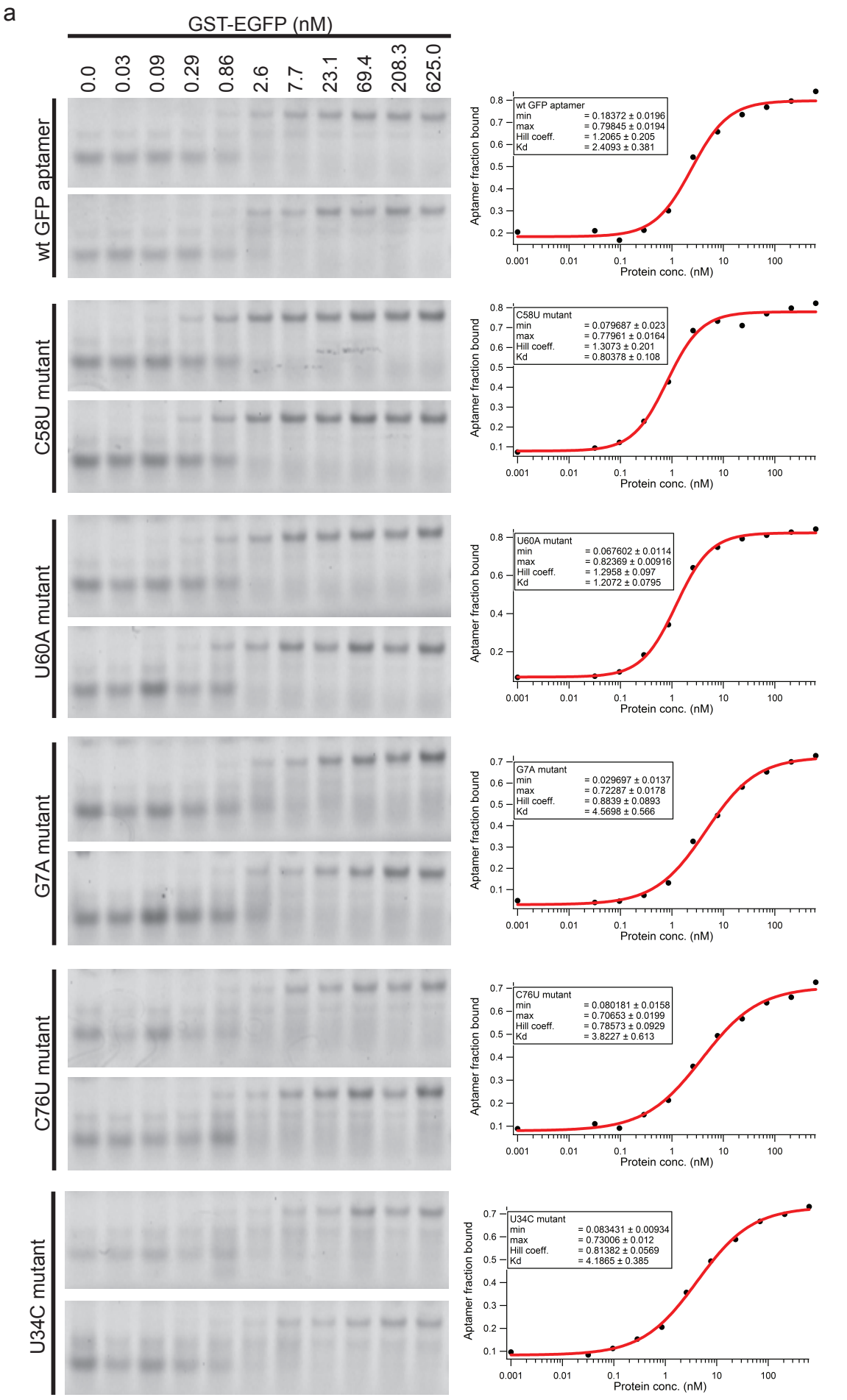
Different sequence elements needed in DNA templates for HiTS-RAP are labeled. Illumina adaptors located at the extreme 5' and 3' ends are used to generate clusters on the sequencer. The sequence of interest in flanked by a T7 RNA polymerase promoter at the 5'end, and the Illumina sequencing primer and ter site at the 3' end. Thus, during transcription, T7 RNAP initiates at its promoter transcribes through the sequence of interest (the GFP aptamer in this case), and then begins to transcribe the Illumina sequencing primer before it encounters the non-permissive Tus bound ter site and halts transcription. Thus, the entire GFP aptamer RNA should have emerged from the polymerase. The approximate positions on the template of Tus protein and T7 RNAP in a halted transcription complex are shown.

Supplementary Figure 4 | Corrections applied to HiTS-RAP data.



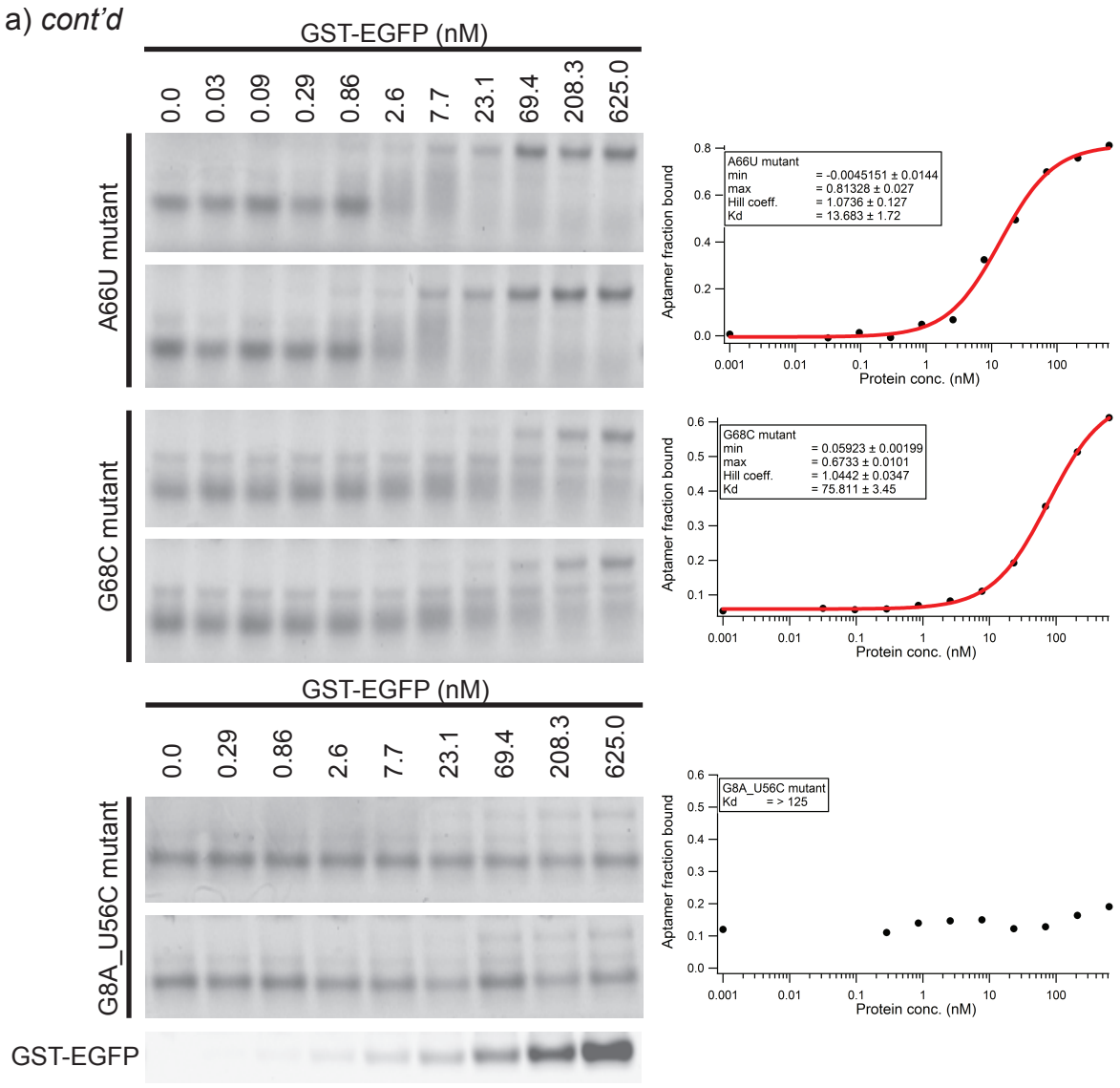
a. Stability of halted transcription complexes modeled by linear approximation of exponential decay. GFP aptamer RNA presenting clusters were imaged at equilibrium with 625 nM EGFP-mOrange nine successive times following a single transcription halting reaction. The average raw, non-normalized fluorescence intensity of bound EGFP-mOrange at canonical GFP aptamer clusters in three lanes (~2.5 million clusters per lane) are plotted in blue, with error bars representing the standard deviation between the three lanes. The reported decay rate is the average ± standard deviation of the fits from the three lanes. The fitted line is plotted as a straight blue line.
Decay rate corrected intensities (see Online Methods) were plotted in red together with a line fit to these intensities (red straight line). **b-d**. HiTS-RAP binding curves for GFPapt and the eight mutants verified by EMSA (see **Supplementary Fig. 5**) in each of the three lanes (L005-L007). Not every mutant is plotted in every lane: it is excluded if scipy.optimize.curve_fit failed to converge on a fit. Each sequence is colored the same in each lane. The number of clusters and percent increase in signal is reported for each lane. Percent increase is the percent increase in fluorescence signal between the average of the first and last two measured intensities in the average binding curve. Intensities decrease across the flowcell due to photobleaching. Sequences with lower affinity tend to saturate at lower fluorescence intensities. These data were used to determine the threshold for signal increase considered as not binding, for assigning sequences in each lane affinities of >125 nM. G8A_U56C (violet), a sequence which was confirmed as having an affinity of >125 nM by EMSA has a 2.8% increase in signal in lane 5 and a 2.6% decrease in lane 7 (and was not present in lane 6). G68C (purple), a mutant with a confirmed affinity of 75 nM, increased by 12.4, 5.8, and 2.5 % in lanes 5, 6, and 7. Thus, we determined that sequences with an increase of less than 3% in signal should be scored as not binding within the detection limit our assay given the concentrations of protein used: therefore they were assigned a Kd of >125 nM (the second highest protein concentration used). If this occurred in only a subset of lanes, it was assigned a Kd of 125 nM for the lanes where it did not bind, and averaged with the rest.

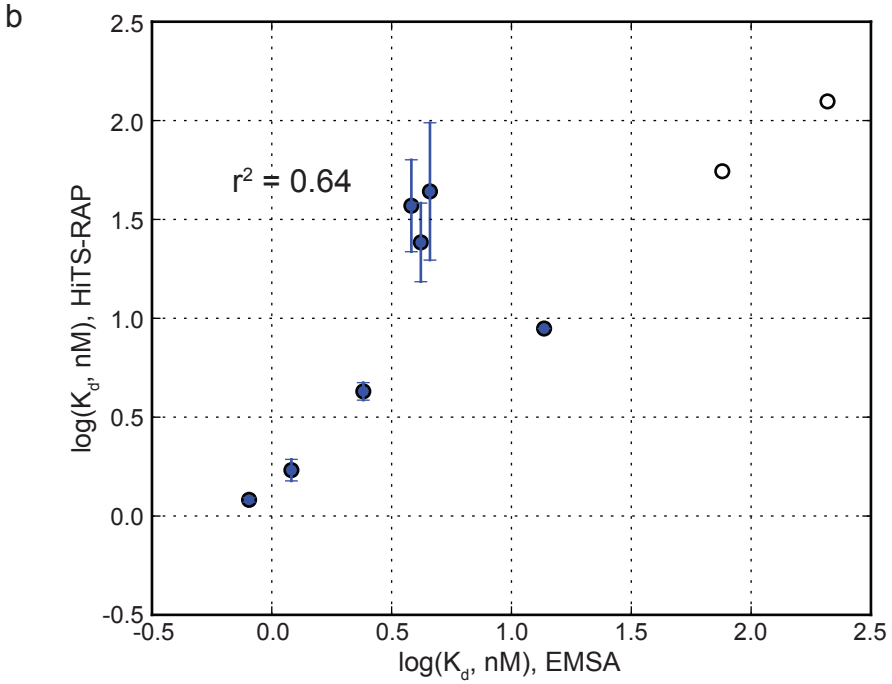
Supplementary Figure 5 | Confirmation of HiTS-RAP measured affinities by EMSA.



cont'd on next page

Supplementary Figure 5 | Confirmation of HiTS-RAP measured affinities by EMSA.



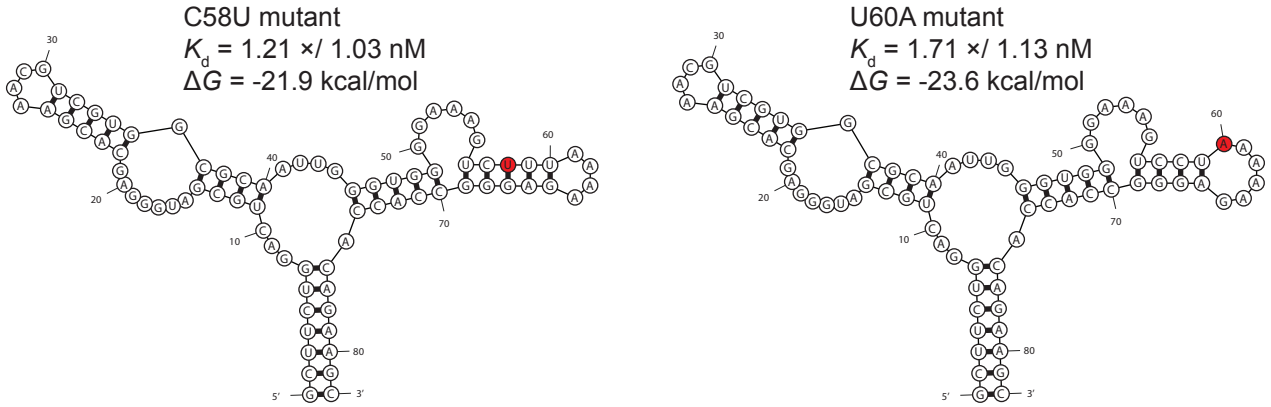


a) EMSA of fluorescently-labeled canonical GFP aptamer and nine mutants with varying affinities. Duplicate EMSA (left panels) were done with varying GST-EGFP concentration in three-fold increments from 0.03 to 625 nM. The bottom band corresponds to unbound RNA, and the top band to EGFP bound RNA. A representative scan of GST-EGFP used in EMSA, visualized by EGFP fluorescence, is shown for the last EMSA gel (bottom left panel). Fraction of aptamer bound by EGFP, quantified from gels, are plotted (black dots in right panels) and fitted to the Hill equation (red solid line) to determine the K_ds. Minimum and maximum fraction of aptamer bound, Hill coefficient and the K_d of the fits are reported in the figure legend for wt (canonical), C58U, U60A, G7A, C76U, U34C, A66U, G68C, G8A_U56C mutant GFP aptamers.

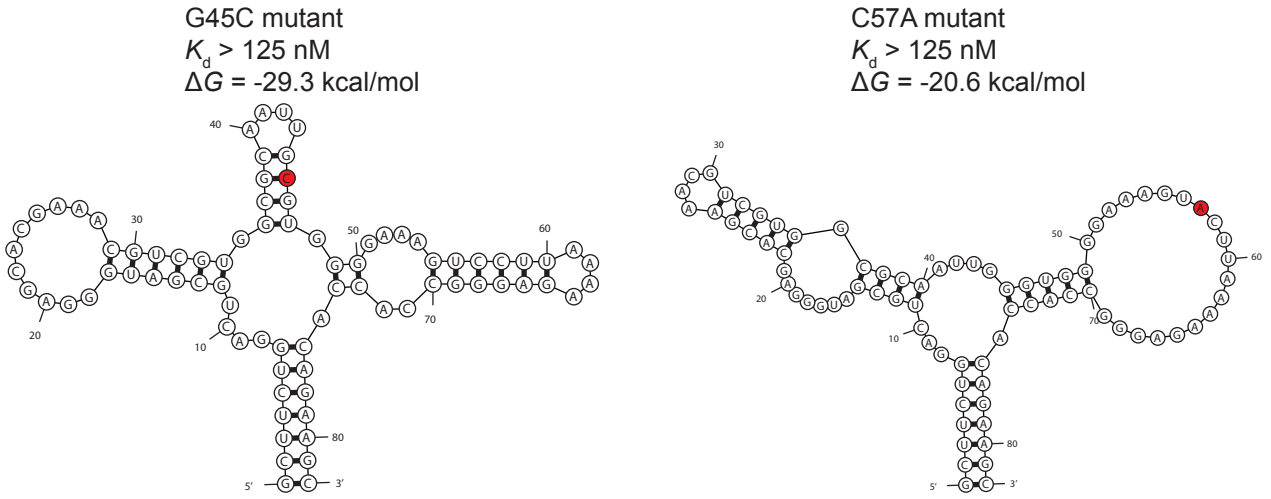
b) Agreement between HiTS-RAP and EMSA. Affinities measured by EMSA in (a) are plotted against their HiTS-RAP counterparts as log(K_d, nM). Mutants scored as not binding in at least one lane (or in all lanes as is the case with G8A_U56C) are plotted as open circles. The r^2 of 0.64 is for all eight mutants and the GFP aptamer.

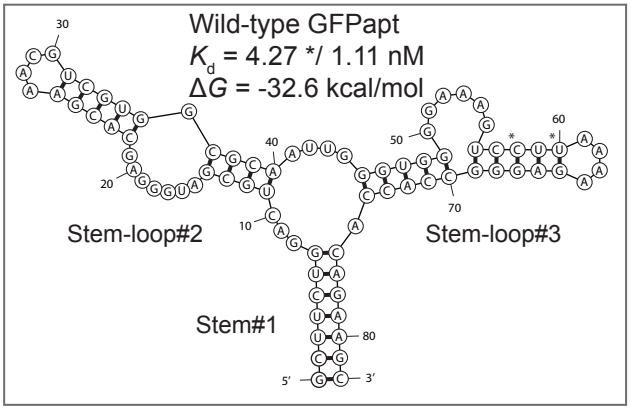
Supplementary Figure 6 | Secondary structure predictions of GFP aptamer single point mutants with higher or lower affinities

a



b

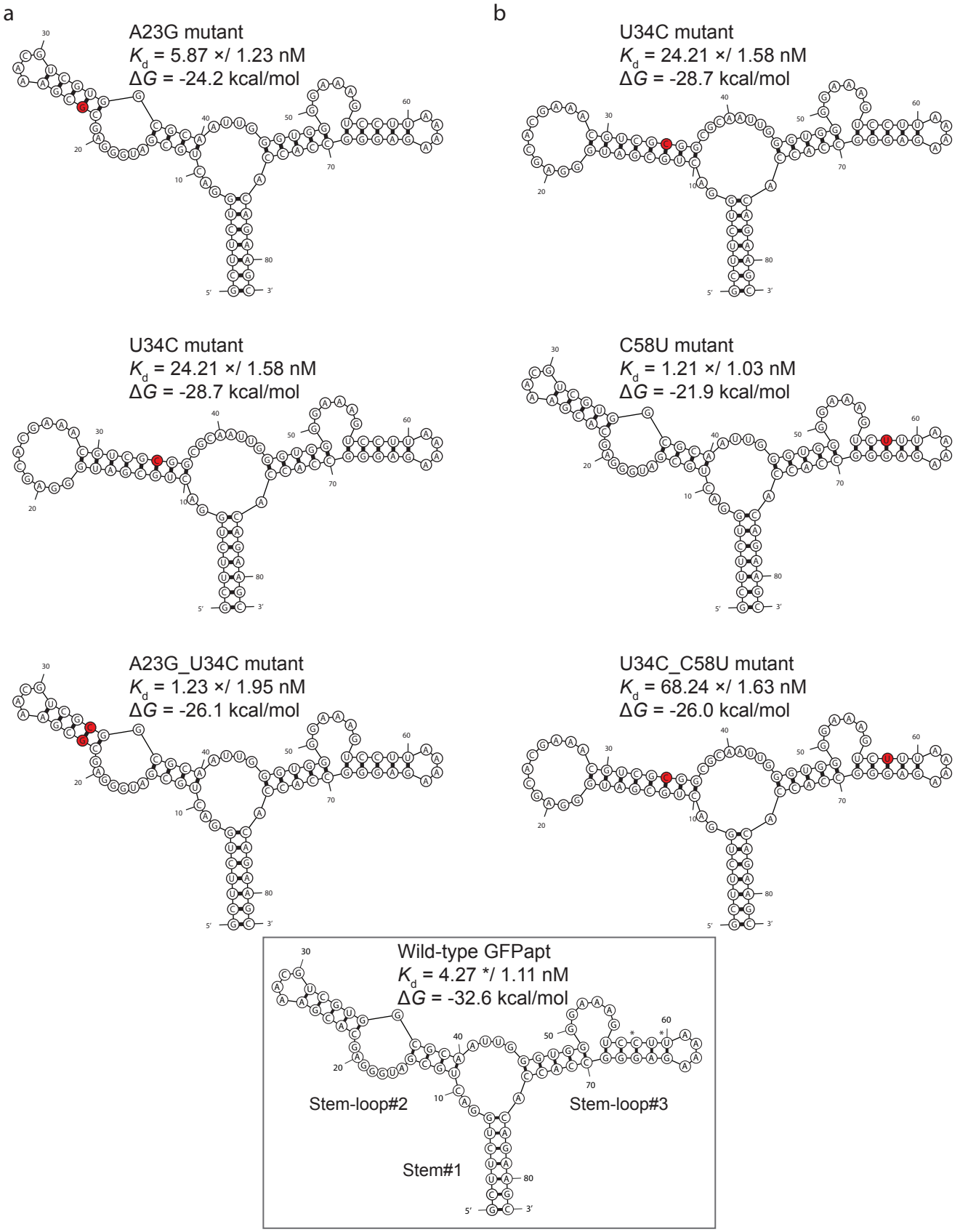




a) Predicted secondary structures of the two highest affinity GFP aptamer single point mutants. C58U (1.21 ×/ 1.03 nM K_d) and U60A (1.71 ×/ 1.13 nM K_d) mutants are predicted to have structures that are very similar to that of wild-type GFP aptamer. Compared to the wild-type aptamer (inset, at the bottom), U60A mutant has a slight expansion of the terminal loop (6 nt vs. 4 nt) and a concurrent shortening of the preceding stem (4 bp vs. 5 bp) in stem-loop#3.

b) Predicted secondary structures of the two lower affinity GFP aptamer single point mutants. Both G45C (>125 nM K_d) and C57A (>125 nM K_d) single point mutants are predicted to have substantially altered secondary structures when compared to the wild-type GFP aptamer (inset, at the bottom). Mutated nucleotides are colored in red. Folded structures and the folding free energies are predicted by mFold.

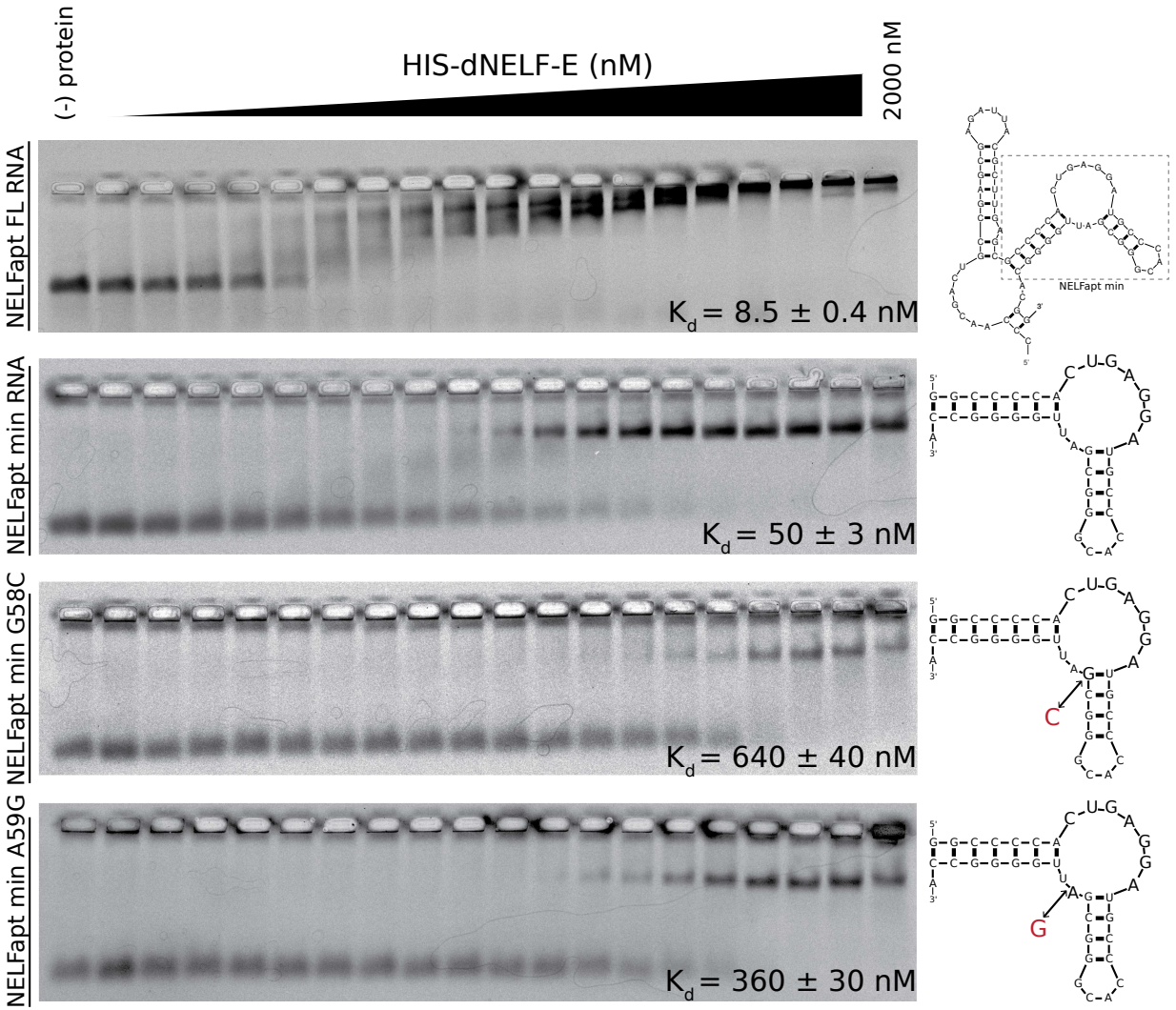
Supplementary Figure 7 | Secondary structure predictions of GFP aptamer double point mutants with substantially higher or lower affinities than predicted by the single point mutants.



a) Predicted secondary structures of the two single point mutants and the corresponding double mutant GFP aptamer with better affinity than predicted from single point mutants. A23G single point mutant (5.87 ×/ 1.23 nM K_d) and A23G_U34C double point mutant (1.23 ×/ 1.95 nM K_d) are predicted to have wild-type GFP aptamer-like structures, whereas U34C single point mutant (24.21 ×/ 1.58 nM K_d) is predicted to have altered structure. A23G_U34C double point mutant is expected to have 33.30 ×/ 1.55 nM K_d based on A23G and U34C single point mutants' effect on affinity. Predicted secondary structure of wild-type GFPapt (inset, at the bottom) is included for comparison.

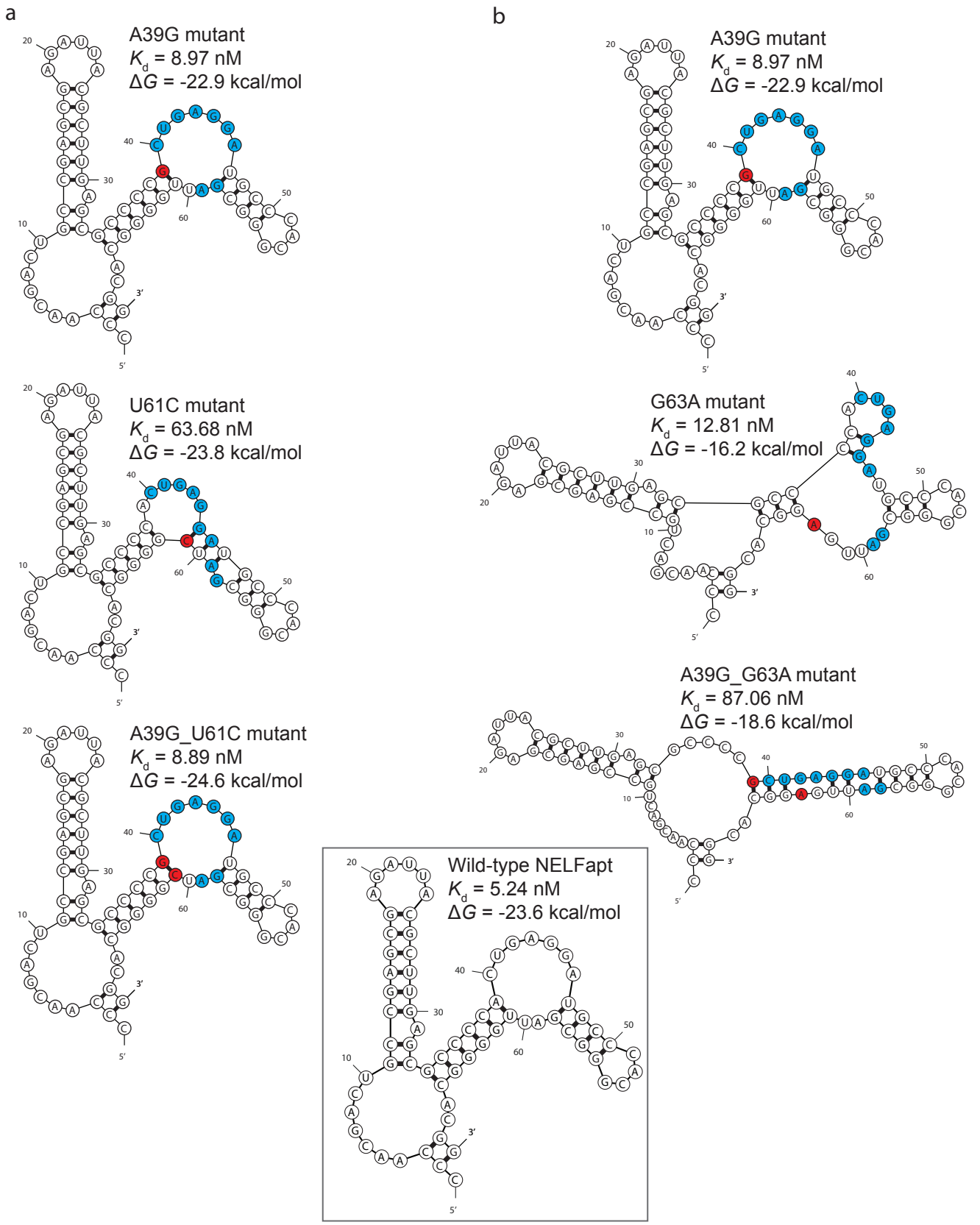
b) Predicted secondary structures of the two single point mutants and the corresponding double mutant GFP aptamer with worse affinity than predicted from single point mutants. U34C single point mutant (24.21 ×/ 1.58 nM K_d) and U34C_C58U double point mutant (68.24 ×/ 1.63 nM K_d) are predicted to have altered secondary structures, whereas C58U single point mutant (1.21 ×/ 1.03 nM K_d) has a wild-type GFP aptamer-like structure. U34C_C58U double mutant is expected to have 6.85 ×/ 1.28 nM K_d based on U34C and C58U single point mutants' effect on affinity. Mutated nucleotides are colored in red. Folded structures and the folding free energies are predicted by mFold.

Supplementary Figure 8 | EMSA confirmation of HiTS-RAP affinities for NELFapt.



EMSA verification of affinities of NELFapt and two mutations within the k-turn. The top gel shows the full length aptamer; this is the full sequence assayed by HiTS-RAP. The bottom three experiments were carried out using the minimal (35 nt) aptamer with one GC base pair added due to the need for a G rich region at the beginning of a T7 transcription template, as was done in the original publication[19]. The minimal version of the aptamer (second from top) binds with lower affinity than the full-length (50 nM versus 8.5 nM). Thus, versions of the minimal aptamer carrying mutations are expected to bind with lower affinity than the full-length version tested by HiTS-RAP. Consistent with HiTS-RAP, A59G binds NELF with lower affinity (48 nM for the full length by HiTS-RAP, 360 nM for the minimal version by EMSA), and G58C binds with even lower affinity (65 nM for the full length by HiTS-RAP, 640 nM for the minimal version by EMSA), confirming that these two bases are important for the interaction of NELFapt with NELF-E.

Supplementary Figure 9 | Secondary structure predictions of NELF-E aptamer double point mutants with substantially higher or lower affinities than predicted by the effects of single point mutants.



a) Predicted secondary structures of the two single point mutants and the corresponding double mutant NELF-E aptamer with better affinity than predicted from single point mutants. A39G single point mutant (8.97 nM K_d) and A39G_U61C double point mutant (8.89 nM K_d) are predicted to have wild-type NELF-E aptamer-like structures, whereas U61C single point mutant (63.68 nM K_d) is predicted to have altered structure where 3'-end of the NBE and the k-turn inducing GA nucleotides, colored in blue, are in a stem. A39G_U61C double mutant is expected to have 109.04 nM K_d based on A39G and U61C single point mutants' effect on affinity. Predicted secondary structure of wild-type NELFapt (inset, at the bottom) is included for comparison.

b) Predicted secondary structures of the two single point mutants and the corresponding double mutant NELF-E aptamer with worse affinity than predicted from single point mutants. A39G single point mutant (8.97 nM K_d) is predicted to have wild-type NELF-E aptamer-like structure. Both G63A single point mutant (12.81 nM K_d) and A39G_G63A double point mutant (87.06 nM K_d) are predicted to have altered secondary structures where the NBE and the k-turn GA are partially or completely in a stem. A39G_G63A double mutant is expected to have 21.93 nM K_d based on A39G and G63A single point mutants' effect on affinity. Mutated nucleotides are colored in red, and NBE and k-turn inducing nucleotides are colored in cyan. Folded structures and the folding free energies are predicted by mFold.