Supporting Information

Combined theoretical, bioinformatic, and biochemical analyses of RNA editing by adenine base editors

Kartik L. Rallapalli,[*,†] Brodie L. Ranzau,[*,†,‖] Kaushik R. Ganapathy,[‡,‖]
Francesco Paesani[*,†,¶,§] and Alexis C. Komor,[*,†]

†*Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California 92093, USA*

‡*Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, California 92093*

¶*Materials Science and Engineering, University of California San Diego, La Jolla, California 92093, USA*

§*San Diego Supercomputer Center, University of California San Diego, La Jolla, California 92093, USA*

‖*Contributed equally to this work*

E-mail: krallapa@ucsd.edu; fpaesani@ucsd.edu; akomor@ucsd.edu

Table of Contents:

1 Materials and Methods **2**
1.1 Data-driven statistical sequence analysis . . . 2
1.1.1 Data curation . . . 2
1.1.2 Entropy calculation . . . 2
1.1.3 Binned Entropy Calculation . . . 2
1.2 Molecular Dynamics Simulations . . . 3
1.2.1 System Setup . . . 3
1.2.2 Unbiased Molecular dynamics (MD) Simulations . . . 3
1.2.3 Binding energy Simulations . . . 4
1.2.4 QM/MM Simulations . . . 5
1.3 Analysis Protocol . . . 5

2 Supplementary Notes **7**
2.1 Supplementary Note 1 . . . 7
2.2 Supplementary Note 2 . . . 7

3 Supplementary Figures **8**

4 Supplementary Tables **25**

5 Supplementary Sequences **33**

6 Supplementary References **40**

1 Materials and Methods

1.1 Data-driven statistical sequence analysis

1.1.1 Data curation

Extant homologs were obtained using BLAST program[1] using *E. coli* wtTadA as the initial query sequence with an e-value cutoff of 0.1 in the SWISSPROT database.[2] This resulted in a dataset comprising of 75 homologs.

We then used special filters to reduce this dataset, by removing sequences with more than 40% gap percentage and to minimize redundant sequences with more than 95% identity to the query sequence. The final filtered dataset comprises of 35 homologs. Visual rationalization of the filtering process was performed with the help of dimensionality reduction through Principle Component Analysis (PCA), followed by K-means clustering. Given that each sequence in the aligned set of sequences contains 167 dimensions (length of query sequence is 167 amino acids), extracting the first two principle components from the PCA algorithm determines the 2-dimensional (2D) Cartesian coordinates of each of these sequences. Since the query of the multiple sequence alignment was *E. coli* wtTadA, it is important to realize that the coordinates of selected set of sequences are relative to the Cartesian coordinates of this sequence. We also calculated the distance of each sequence in 2D space from wtTadA. Additional phylogenetic analyses were performed in conjunction using ClustalW, and DENDROSCOPE[3] tool as well (Figure S1B).

To further substantiate the diversity of sequences selected by the filtering we implemented, a K-means clustering algorithm is run on the sequences upon the PCA dimensionality reduction. The optimal number of partitioning clusters (k) was determined to be four, through the conventional elbow method (Figure S1C).

1.1.2 Entropy calculation

The resultant dataset was used to calculate the sequence entropy score, defined as follows:

$$S_i \;=\; -\sum_{n=1}^{N} p(i_n)\, log_{20}\, p(i_n) \quad \text{for} \quad i \in \{1,\ldots,L\} \quad \forall \quad \text{amino acids} \tag{1}$$

$$G_i \;=\; \frac{g_i}{N} \quad i \in \{1,\ldots,L\} \quad \forall \quad \text{gaps} \tag{2}$$

$$H_i \;=\; S_i \;+\; G_i \tag{3}$$

where $p(i_n)$ refers to the statistical probability of having a particular amino acid n at site i and N is the total number of amino acids. g_i represented the total number of gaps at position i among all hits. The first term in the equation (S_i) represents the entropy value for amino acids at residue site i and the second term (G_i) represents the fraction of sequences with gaps (i.e. lack of any aligned amino acid after the MSA at that residue site). This treatment of the gap sites as a separate term (G_i) is done so as to retain the biochemical significance of presence of gaps, that such residue site are not functionally constrained. They can be (and have been) deleted from their parent sequences without damaging the activity of the enzyme.[4] The entropy scores obtained through Equation 3 were mapped on to the full-length structure of *E. coli* wtTadA that was generated by the missing C- and N-terminal residues using MODELLER[5] to PDB ID:1Z3A.[6]

1.1.3 Binned Entropy Calculation

Binned entropy calculations are performed using a similar approach to that described in Equation 3. However, these entropies are not calculated wrt individual amino acids but rather a collections of similar amino acids (bins) defined on the basis of their chemical properties (Figure S2A). Thus, the binned entropy scores are

evaluated as:

$$R_i \;=\; -\sum_{n=1}^{N} b(i_n)\, log_6\, b(i_n) \quad \text{for} \quad i \in \{1, \ldots, L\} \quad \forall \quad \text{bins of amino acids} \tag{4}$$

$$G_i \;=\; \frac{g_i}{N} \quad i \in \{1, \ldots, L\} \quad \forall \quad \text{gaps} \tag{5}$$

$$B_i \;=\; R_i \;+\; G_i \tag{6}$$

where $b(i_n)$ refers to the statistical probability of a residue at site i pertaining to bins shown in Figure S2A and N is the total number of amino acid bins. g_i represented the total number of gaps at position i among all hits.

1.2 Molecular Dynamics Simulations

1.2.1 System Setup

The TadA*0.1 model was built using the crystal structure of *E. coli* TadA (PDB ID: 1z3a).[6] Given the sequence homology between *S. aureus* TadA and *E. coli* TadA, we combined the *sa*TadA-RNA structure (PDB ID: 2B3J) with the TadA*0.1 model to build the TadA*0.1-RNA model.[7] The TadA*0.1 was transformed into the various ABE mutants using the `swapaa` command in Chimera.[8] For both apo-TadA* and TadA*-RNA models, all crystallographic water molecules within 3 Å distance of the surface of the protein or the RNA were preserved during the modeling procedure. All titratable residues were protonated using the H++ server using default settings.[9,10]

To parameterize the metal-containing active site of TadA* (comprising of Zn^{+2}, His^{57}, Cys^{87}, Cys^{90}, and an activated water molecule) we converted the two cysteines to their deprotonated state and used the MCPB.py approach at B3LYP/6-31G* level of theory.[11] The Zn^{+2} ion and the side chains of its coordinating residues, along with the active site water were used to determine the bond and angle parameters between the Zn^{+2} and the coordinating S^{γ} atoms, $N^{\delta 1}$ atom, and the oxygen of the activated water molecule. To obtain the partial charges on the active site residues, a larger sub-system was chosen around the Zn^{+2} ion. This larger sub-system was first optimized and then the optimized structure was used to obtain the partial charges through RESP fitting. The torsional terms involving the Zn^{+2} ion were ignored for simplicity. The resultant model considered the active site as a fully-bonded metal center, with the Zn^{+2} ion tetrahedrally linked to all its neighboring residues - including the activated water molecule.

To obtain a more realistic representation of the metal center and activated water molecule, we deleted the parameters associated with the Zn^{+2} and the activated water and scaled the charge on the Zn^{+2} ion appropriately. Hence, the ultimate model used in all the apo-TadA* and TadA*-RNA simulation involve a hybrid bonded- and non-bonded representation of the Zn^{+2} active site, where the activated water molecule is no different from the bulk water molecules and can freely diffuse away from the ion. The partial charges and force field parameters are provided in the supplementary information folder.

The rest of the protein was represented using Amber ff14SB[12] and the RNA was represented using RNA.OL3 force field.[13–15] LEap tool from AmberTools was used to immerse the apo-TadA* and TadA*-RNA complexes into a pre-equilibrated truncated octahedron box of explicit TIP3P water, with a 15 Å buffer distance.[16] Varying number of Na^{+} ions were added to each of the systems to maintain electroneutrality and the simulation cell was then replicated infinitely in three dimensions to impose periodic boundary conditions.

1.2.2 Unbiased Molecular dynamics (MD) Simulations

To relieve any bad contacts that may have been present in the crystal structures or may have been introduced during the system preparation stages, all the apo-TadA* and TadA*-RNA models were first subjected to a multi-step energy minimization procedure using a combination of steepest descent and conjugate gradient

algorithms. This minimization was carried out in four separate stages (((2000 steps of steepest descent + 3000 steps of conjugate gradient) × 4) total steps), each stage becoming less restrained than the previous one. During the first stage (5000 steps) only the solvent and the counter ions were allowed to move and all protein (or protein-RNA) atoms were restrained using a weight of 200 kcal/mol Å^2. During the second stage, all the heavy atoms of the protein (or protein-RNA) were restrained with a force of 200 kcal/mol Å^2, allowing only the hydrogen atoms and the solvent atoms to move freely. During the third stage, only the protein (or protein-RNA) backbone atoms were restrained with a force of 200 kcal/mol Å^2. During the fourth and final stage of minimization, all the restraints were removed and the entire system was allowed to relax freely.

The minimization was followed by heating to the temperature of 300 K, using a Langevin thermostat with the collision frequency of 2 ps^{-1} to assign initial velocities to all the atoms in the system. This heating was followed by the equilibration of the systems in an isothermal-isobaric (NPT) ensemble with the Brendsen barostat for pressure scaling. Similar to the minimization procedure described above, the equilibration of the systems was also done in a multi-step manner (10 ns × 4), where restraints were gradually decreased (from 2.0 kcal/mol Å^2 to 1.0 kcal/mol Å^2 to 0.5 kcal/mol Å^2 and finally to no restraints) from all the protein (or protein-RNA) atoms . Long-range electrostatics were evaluated using the Ewald method with the non-bonded energy cutoff at 10 Å.[17] The SHAKE algorithm was used for constraining all bond distances involing hydrogen atoms.[18]

These equilibrated structures were used to initiate the 1 μs simulations as well as the biased MD simulations used to calculate the binding affinity between the various protein-RNA complexes. All simulations were propagated in time using the velocity Verlet algorithm with a time step of 2 fs. All simulations were conducted using the CUDA accelerated version of PMEMD Amber18.[19–22]

1.2.3 Binding energy Simulations

Starting from the equilibrated structures obtained from the unbiased MD, as described above, we calculated the binding energy profiles for the TadA*-RNA complexes. This was done in two parts : first, a preliminary estimation was generated using steered MD (SMD) simulations,[23] which was followed by confirmatory umbrella sampling (US) calculations.[24]

The collective variable (ξ) used to monitor the binding or unbinding process was defined as the distance between the centers of mass (COM) of the TadA* and the RNA (Figure 5F). Using a constant rate of 0.1 Å/ns pulling (and pushing) the TadA*-RNA complexes were dissociated (and associated) beyond their equilibrated distance ($\approx$ 17 Å) by $\pm$ 10 Å.

Configurations from these SMD binding profiles were used as the seeds for longer US simulations. The coordinate space ($\xi \in$ [17, 37] Å; step 0.5 Å) was divided to generate a set of 41 discrete US windows. Each US window was subjected to a production stage for 5 ns under the umbrella restraints with a force constant, k, of 40 kcal/mol Å^2.

Four independent US simulations were carried out for each of the 41 windows for all the TadA*-RNA complexes, which were then used to determine the potentials of mean force (PMFs) representing the free energy profiles associated with the binding process along the ξ.

Thus, a total of 820 ns (20 ns × 41 windows) was sampled for each TadA*-RNA complex and this led to the accumulation of 100000 instantaneous values scanning along the ξ coordinate. Due to the external bias applied on the system through harmonic restraining , this biased probability distribution (Figure S7) is uniformly distributed along ξ. To reveal the true unbiased probability distribution and thus, the PMF along ξ, we made use of the WHAM algorithm,[25,26] with a convergence threshold of 10^{-8}.

With 4 independent PMF profiles for each of the TadA*-RNA complexes, the error in convergence due to the sampling was calculated as a the standard deviation of the 4 datasets. Additional error analysis was performed using the block averaging method to evaluate the uncertainty associated with the normalization procedure implemented within the WHAM algorithm.[27]

1.2.4 QM/MM Simulations

Through a hybrid quantum mechanical/molecular mechanical (QM/MM) approach the free energy changes for the deprotonation of the activated water molecule by the Glu^{59} residue for the TadA* (and TadA*-RNA) models were computed for the various mutants.[28]

As this first step involves the transfer of a proton from the water (activated by Zn^{+2}) to the Glu^{59}, the QM subsystem consisted of the side chains of the active site residues (His^{57}, Glu^{59}, Cys^{87}, and Cys^{90}), the Zn^{+2} ion, and the activated water for both the apo-TadA* and TadA*-RNA models. These QM atoms were treated using self-consistent charge density functional tight binding (SCC-DFTB) method implemented within Amber18, which has been shown to have good accuracy at nominal computational expense despite being a semiempirical model.[29,30] The atoms beyond this active site cluster were represented the MM subsystem and were treated using the force fields as in the unbiased MD.

The QM subsystem was electrostatically embedded into the MM subsystem, and the bonds spanning the two subsystems were capped using hydrogens as the link atoms. The net charge on the QM region was -1. The electrostatics of the QM region were cutoff beyond 9 Å distance. No SHAKE was implemented in the QM region as the reaction to be simulated involves a proton and the time step was reduced to 0.5 fs. All QM/MM simulations were conducted with the QM(DFTB)/MM implementation in the sander module of Amber18.

For the apo-TadA* systems we started from the equilibrated conformations extracted from the unbiased MD and for the TadA*-RNA systems we started from the conformations extracted from the minima in the binding energy profile, to initiate the catalytic deprotonation reaction. These structures were further equilibrated using the QM/MM scheme for 500 ps, which was followed by a QM/MM SMD simulation to explore the reaction profile of the deprotonation.

The difference of the distances between the Wat O and shared proton and the Glu^{59} O and shared proton, was chosen as the collective variable (ξ) to monitor the deprotonation. Using the QM/MM SMD scheme with a constant pulling speed of 0.0024 Å/ps and a force constant of 200 kcal/mol $Å^2$, the shared proton was forced away from the water and onto the Glu^{59} O.

From this preliminary estimate of the reaction profile of the system, snapshots were extracted at every 0.1 Å, from $\xi \in$ [-0.6, 0.6] Å, to seed 13 discrete windows, to conduct 1950 ps (13 windows $\times$ 150 ps) of simulations under umbrella restraints. This led to the accumulation of 300000 instantaneous values of the ξ for each individual window. To obtain the unbiased probability distribution from this raw data, and thus, the PMF along ξ, we again made use of the WHAM algorithm with a convergence threshold of 10^{-8} (SI Figure Figure S14 and Figure S13).

With 3 different PMF profiles for each of the TadA* and TadA*-RNA models, the error in the average reaction profile was estimated as the standard deviation of these sets.

1.3 Analysis Protocol

The trajectories from the simulations of the various apo-TadA* and TadA*-RNA systems were analyzed using the cpptraj tool.[31,32] For creating the asteroid plots, we first identified the amino acid residues within the primary interaction shell around the nucleotides in the acitve site (`-UACG-`) and then using these residues in the cpptraj distance-based `mask` we analyzed the trajectories of the unbiased MD. The atom-list per frame data obtained using the cpptraj module was re-normalized to give the percentage residue contacts, using the following formula:

$$\text{Percentage contact} = \frac{\text{Total atomic contact during all frames} \times 100}{\text{Number of atoms in the amino acid} \times \text{Total number of frames}} \tag{7}$$

The average number of hydrogen-bonding interactions between the residues in the primary interaction shell and the RNA bases was computed using the `hbond` feature in cpptraj with the default hydrogen-bond definition (3 Å distance between donor and acceptor atoms, and 135° angle between the donor, hydrogen, and acceptor atoms) (Figure S6).

To generate the modified chord diagrams, we tracked individual waters using a similar distance-based `mask` but this time restricted to only water oxygens and ignoring all the other atoms of the systems (SI Figure S10). The visualization of all the trajectories was rendered using Chimera,[8] the graphical data was plotted using Matplotlib,[33] and the curve-fittings (SI Figure S7) were done using.[34]

2 Supplementary Notes

2.1 Supplementary Note 1

To understand this non-additive behavior in ABE1.1(L84F) we again rely on our entropy-based analysis. As residues 84 and 108 are in direct contact (see the three-dimensional structure of TadA* in (Figure 3F)), we speculate that each of these residues may encode useful information regarding the nature of the other. To test this hypothesis, we define the covariance of two residue sites i and j using mutual information as:

$$\text{Mutual Information} \;=\; \text{MI}_{i,j} \;=\; H_i + H_j - H_{i,j} \tag{8}$$

where H_i and H_j are the self-entropy scores calculated using Equation 1, and $H_{i,j}$ is the joint entropy for the occurrence of a certain pair of amino acids at sites i and j. This computation leads to a mutual information score between residues 84 and 108, $MI_{84,108}$, equal to 0.245. In the context of all the MI values calculated for the entire sequence of TadA, $MI_{84,108}$ is in the 74th percentile (Figure S16), indicating a somewhat weak correlation between the two sites. However, it should be noted that in our dataset site 84 is never encountered to be a phenylalanine (Figure 3D). Thus, there is no combination of 84F and 108N in our dataset. This indicates that mutual information is not the appropriate metric to understand the nature of the interaction between residues 84 and 108.

2.2 Supplementary Note 2

An additional `-UACG-` motif was identified within the RNA editing Site 1 amplicon (referred to as Site 1′ henceforth) which was previously unreported.[35] Editing levels at this site are much lower than the other six sites, but otherwise follow the same patterns when treated with the different ABE variants (Figure S3. Interestingly, there is a drastic difference in editing levels at this site between ABE0.1 and ABE1.1. Introduction of the L84F mutation to ABE1.1 drastically lowers editing levels, and even lower editing is seen with ABE7.10. The secondary structure for Site 1′ is highlighted in Figure S4.

3 Supplementary Figures

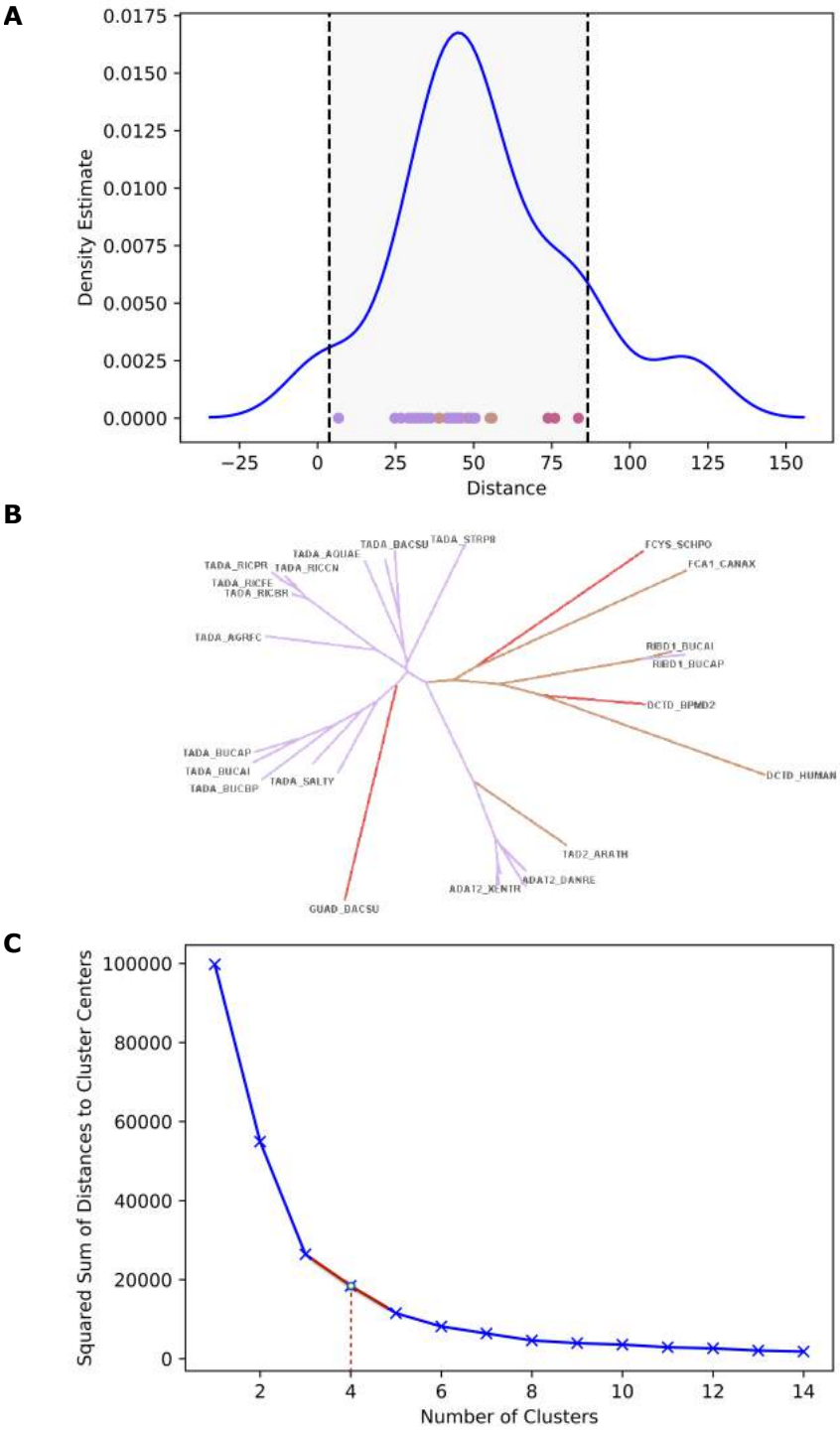


Figure S1: (A) Kernel density estimate (KDE) showing the distance of sequences from *E. coli* wtTadA from the first 2 principal components. (B) Phylogenetic analysis of the alignment for *E. coli* wtTadA. (C) Elbow plot to determine the optimal number of clusters to be used in k-means clustering.

A

Amino Acid	Category
Arginine, Histidine, Lysine	Positively charged
Aspartic Acid, Glutamic Acid	Negatively charged
Serine, Threonine, Aspargine, Glutamine	Polar uncharged
Alanine, Valine, Leucine, Isoleucine, Methionine	Hydrophobic-aliphatic
Phenylalanine, Tyrosine, Tryptophan	Hydrophobic-aromatic
Cysteine, Selenocysteine, Glycine, Proline	Special

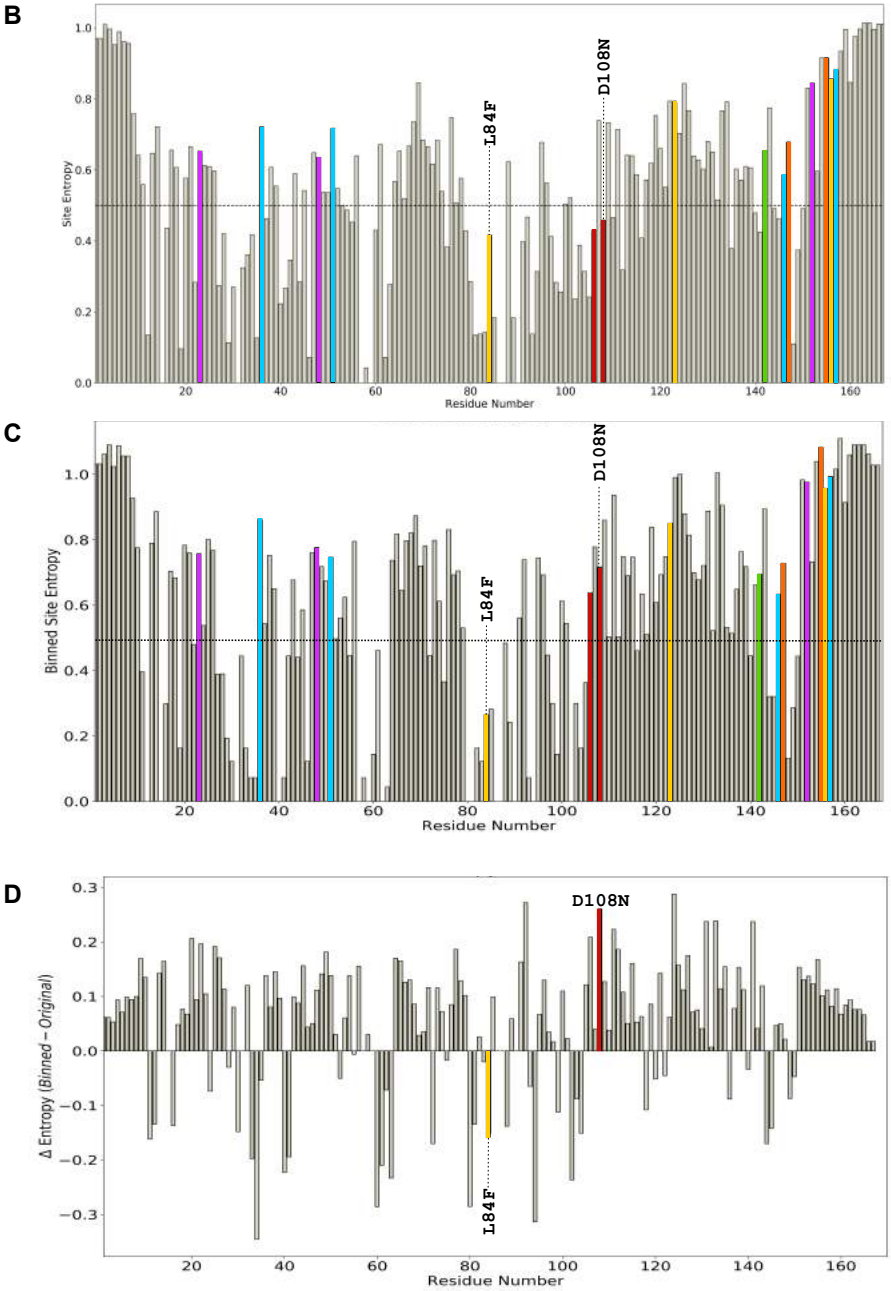


Figure S2: (A) Categories into which each of the 20 amino acid is placed for analyzing site-specific amino acid diversity. (B) Entropy-based analysis of ABE7.10 mutations. (C) Binned-entropy based analysis of ABE7.10 mutations. (D) Difference in the binned and regular entropy values to highlight the increase in entropy of site 108 and decrease in the entropy of site 84.

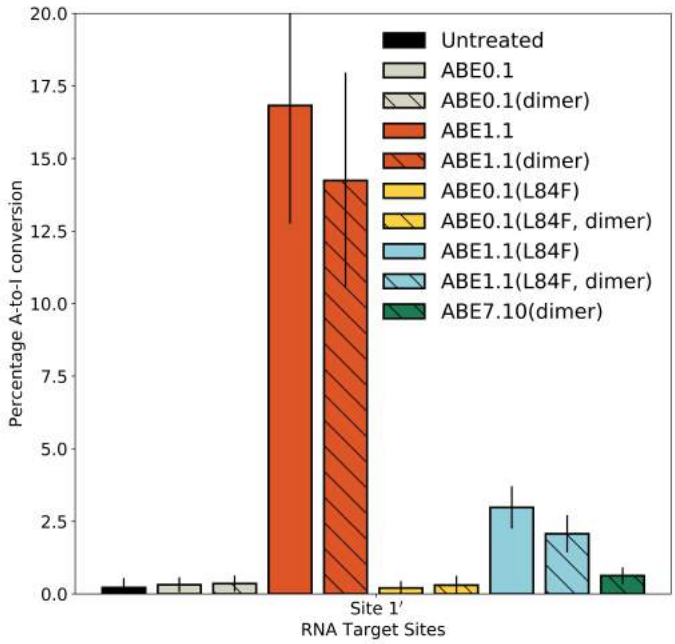


Figure S3: A-to-I base editing efficiencies in HEK293T cells by various ABE mutants at alternative RNA off-target site 1′. Values and error bars reflect the mean and SD of four independent biological replicates performed on different days.

Experimentally tested sites:

RNA site 1 (DNAJB1) (chr: 14518195)

GCGCTACCACCCGGACAAGAACAAGGAGCCCGGCGCCGAGGAGAAGTTCAAGGAGATCGCTGAGGCCTACGACGTGCTCAGCGACCCGC
GCAAGCGCGAGATCTTCGACCGCTACGGGGAGGAAGGCCTAAAGGGGAGTGGCCCCAGTGGCGGTAGCGGCGGTGGTGCCAATGGTACC
TCTTTCAGCTACACATTCCATGGAGACCCTCATGCCATG

RNA site 2 (MTA2) (chr: 62594034)

TCTGGCTTCAGGGATTCGTTCAAGCTCACAGCCAGCAGCCAAGCGTCAGAAACTAAACCCAGCTGATGCCCCCAATCCTGTGGTGTTTG
TGGCCACAAAGGATACCAGGGCCCTACGGAAGGCTCTGACCCATCTGGAAATGCGGCGAGCTGCTCGCCGACCCAACTTGCCCCTGAAG
GTGAAGCCAACGCTGATTGCAGTGCGGCCCCCTGTCCCTCTACCTGCACCCTCACATC

RNA site 3 (PTBP2) (chr: 96813053)

AGATTTTGGTAATTCCCCATTGCATCGTTTTAAGAAACCTGGATCCAAAAATTTTCAAAACATTTTTCCTCCTTCTGCCACCCTTCACC
TATCTAATATCCCTCCATCAGTAGCAGAAGAGGATCTACGAACACTGTTCGCTAACACTGGGGGCACTGTGAAAGCATTTAAGTTTTTT
CAAAGAGATCACAAAATGGCTCTTCTTCAGATGGCAACAGTGGAAGAAGCTATTCAGG

RNA site 4 (SAP30BP) (chr: 75703316)

CAGAACCCCCTGGCAGATGTTCAAATCACTTGCAAGACAAGATCCAGAAGCTTTATGAACGAAAGATAAAGGAGGGAATGGATATGAAC
TACATTATCCAAAGGAAGAAAGAATTTCGGAACCCTAGCATCTACGAGAAGCTGATCCAGTTCTGTGCCATTGACGAGCTTGGCACCAA
CTACCCAAAGGATATGTTTGATCCCCATGGCTGGTCTGAGGACTCCTACT

RNA site 5 (LCMT1) (chr: 25164711)

ATTGCCAACACTCCTGATAGCTGAATGTGTGCTGGTTTACATGACTCCAGAGCAGTCCGCAAACCTCCTGAAGTGGGCAGCCAACAGTT
TTGAGAGAGCCATGTTCATAAACTACGAACAGGTGAACATGGGTGATCGGTTTGGGCAGATCATGATTGAAAACCTGCGGAGACGCCAG

TGTGACCTGGCGGGAGTGGAGACCTGCAAGTC

RNA site 6 (SCAP) (chr: 47420696)

CCATTGACATTCGCCGGATGGAGCTAGCAGACCTGAACAAGCGACTGCCCCCTGAGGCCTGCCTGCCCTCAGCCAAGCCAGTGGGACAG
CCAACGCGCTACGAGCGGCAGCTGGCTGTGAGGCCGTCCACACCCCACACCATCACGTTGCAGCCGTCTTCCTTCCGAAACCTGCGGCT
CCCCAAGAGGCTGCGTGTTGTCTACTTC

Sequence used in the simulations:

UUGACUACGAUCAA

Figure S4: RNA sequences tested in the experimental analyses and simulation models. The amplicon sequences pertain to the cDNA used for sequencing analysis. Thus, every T should be considered a U in RNA. The consensus sequence for all sites tested is hence -UACG-, which is the same as the native RNA substrate sequence used in the simulation models.

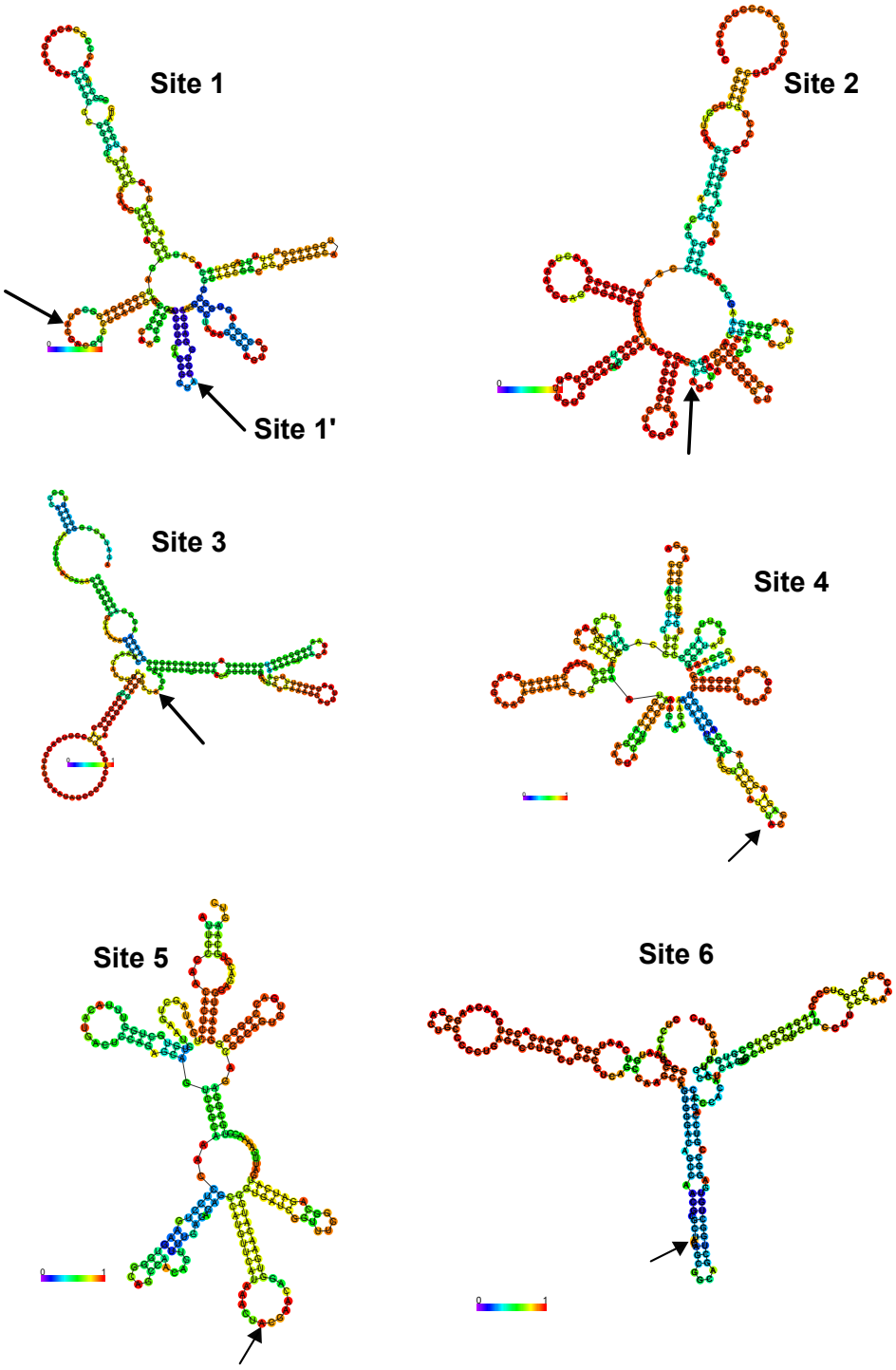


Figure S5: Secondary structures of the six RNA sites that are tested experimentally. The structure were determined using the minimum free energy method of the RNAfold webserver with the default parameters.[36] The arrows here indicate the Adenine base that is being edited by the various ABEs. The bases are colored by their base-pairing probabilities. For unpaired regions the color denotes the probability of being unpaired. All sites, expect site 6, predict the target adenine to occur as unpaired hairpin loop regions.

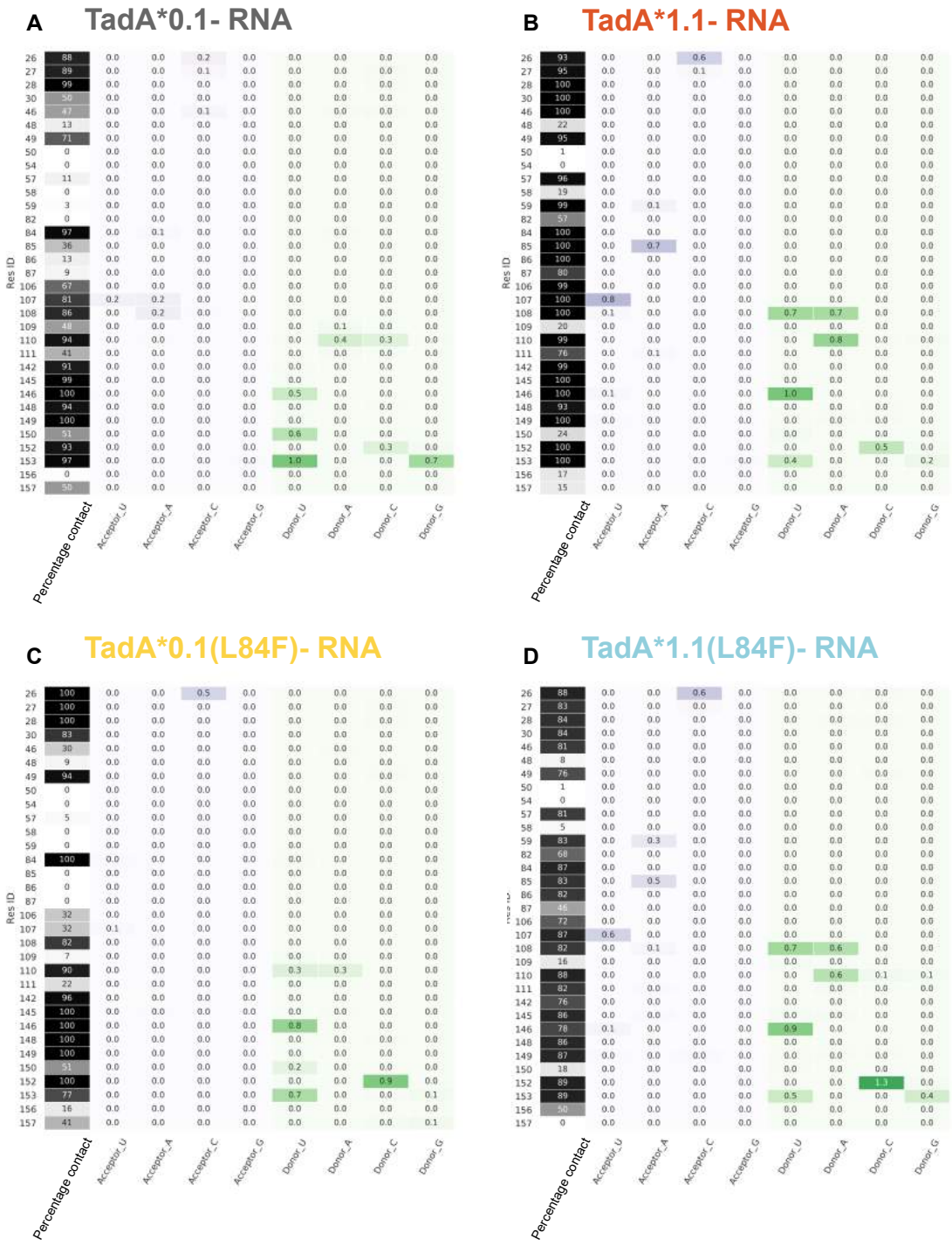


Figure S6: Percentage contact and the fractional H-bonding between the -UACG- consensus sequence nucleotides and the first interaction shell amino acids. For (A) TadA*0.1, (B) TadA*1.1, (C) TadA*0.1(L84F) and (D) TadA*1.1(L84F) in complex with RNA.The intensity of the colors in the columns signifies the magnitude of the percentage contact and the H-bonding strength.

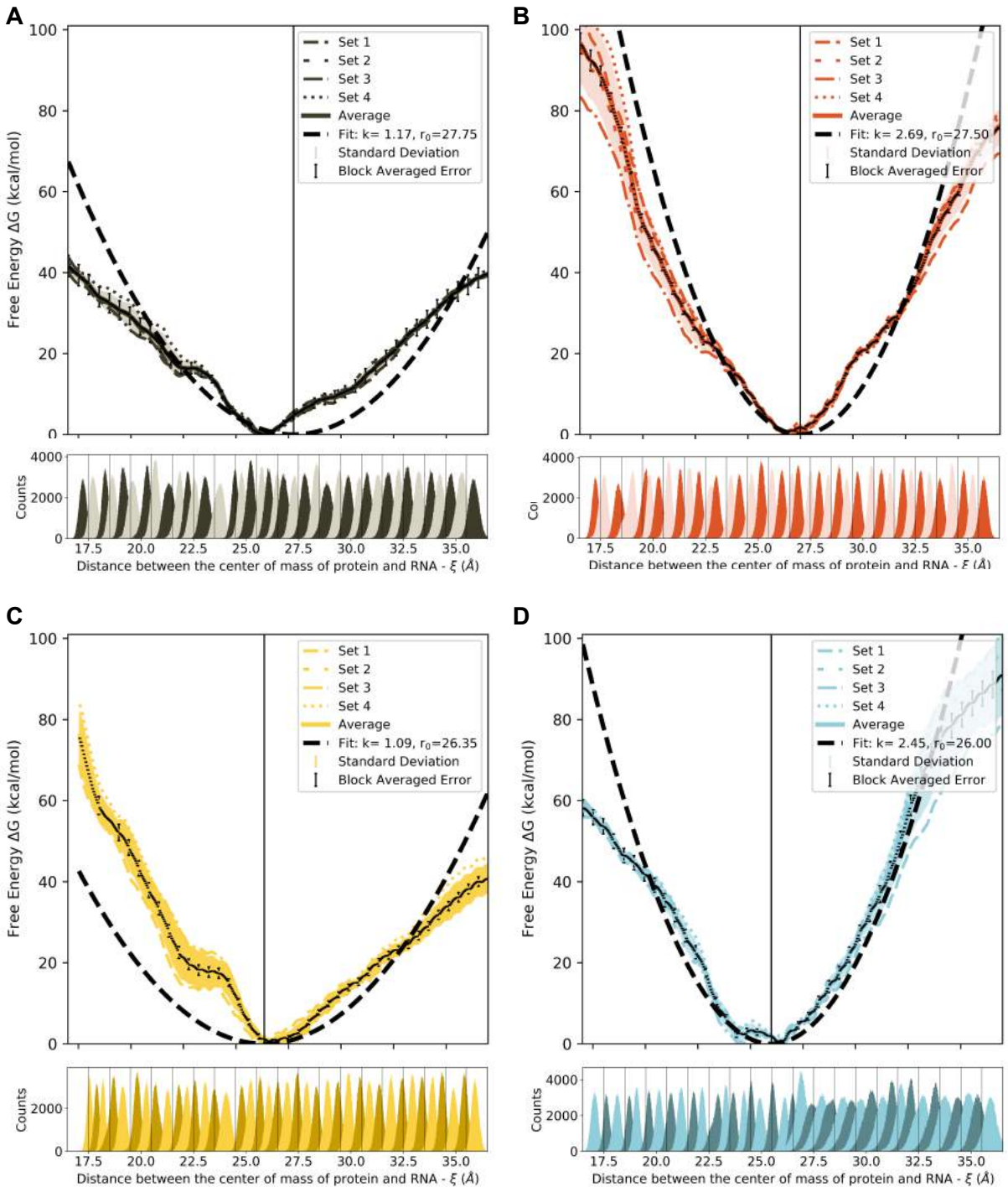


Figure S7: Umbrella sampling data and biased statistics. Individual pmfs associated to 4 independently conducted umbrella sampling simulations and the biased probability distributions (histograms) obtained from individual windows that stratify the ξ space, for TadA*0.1-RNA (A), TadA*1.1-RNA (B), TadA*0.1(L84F)-RNA (C), and, TadA*1.1(L84F)-RNA (D) complex. In each pmf profile the shaded region indicated the standard deviation of the individual pmfs and the error bars are the error calculated using block averaging method for individual windows.

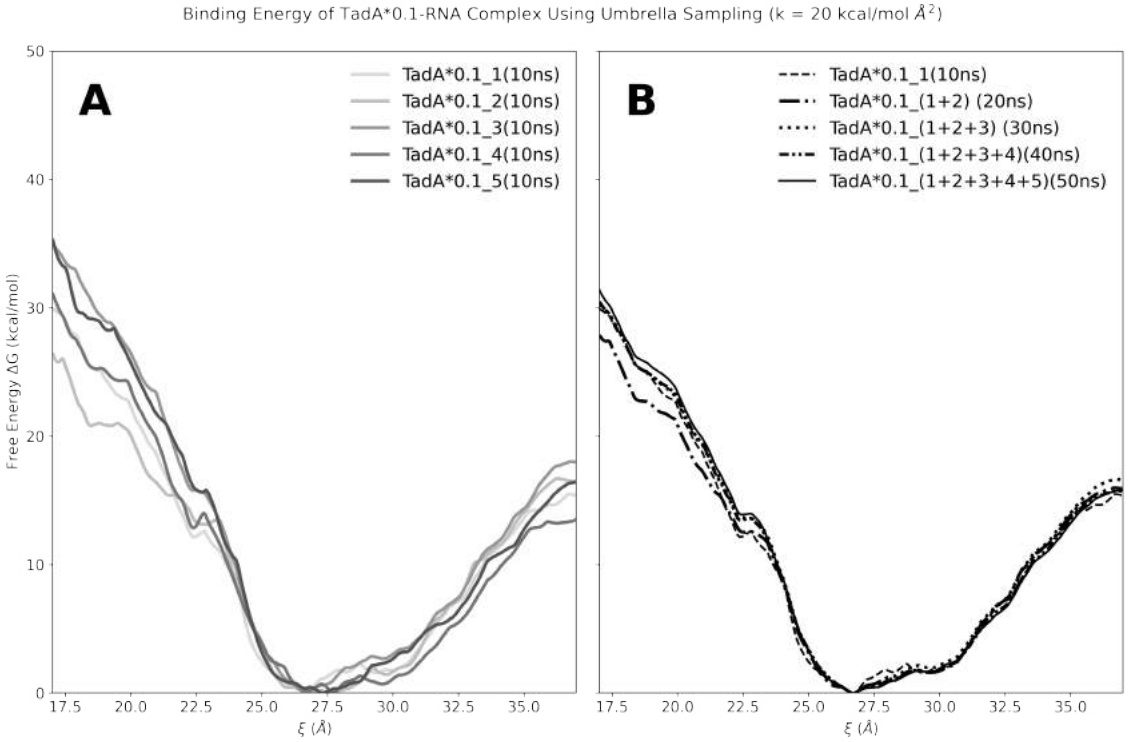


Figure S8: Analysis of the convergence for the local free energy changes in the TadA*0.1-RNA complex. (A) Free energy curves for 5 independent, 10 ns long, umbrella sampling simulations. (B) The running average of the free energy curves as these simulations are progressively combined together, from 10ns (TadA*0.1_1) to 50ns (TadA*0.1_1_2_3_4_5). This analysis shows that average free energy is converged on the time scale of umbrella sampling simulations used in this study.

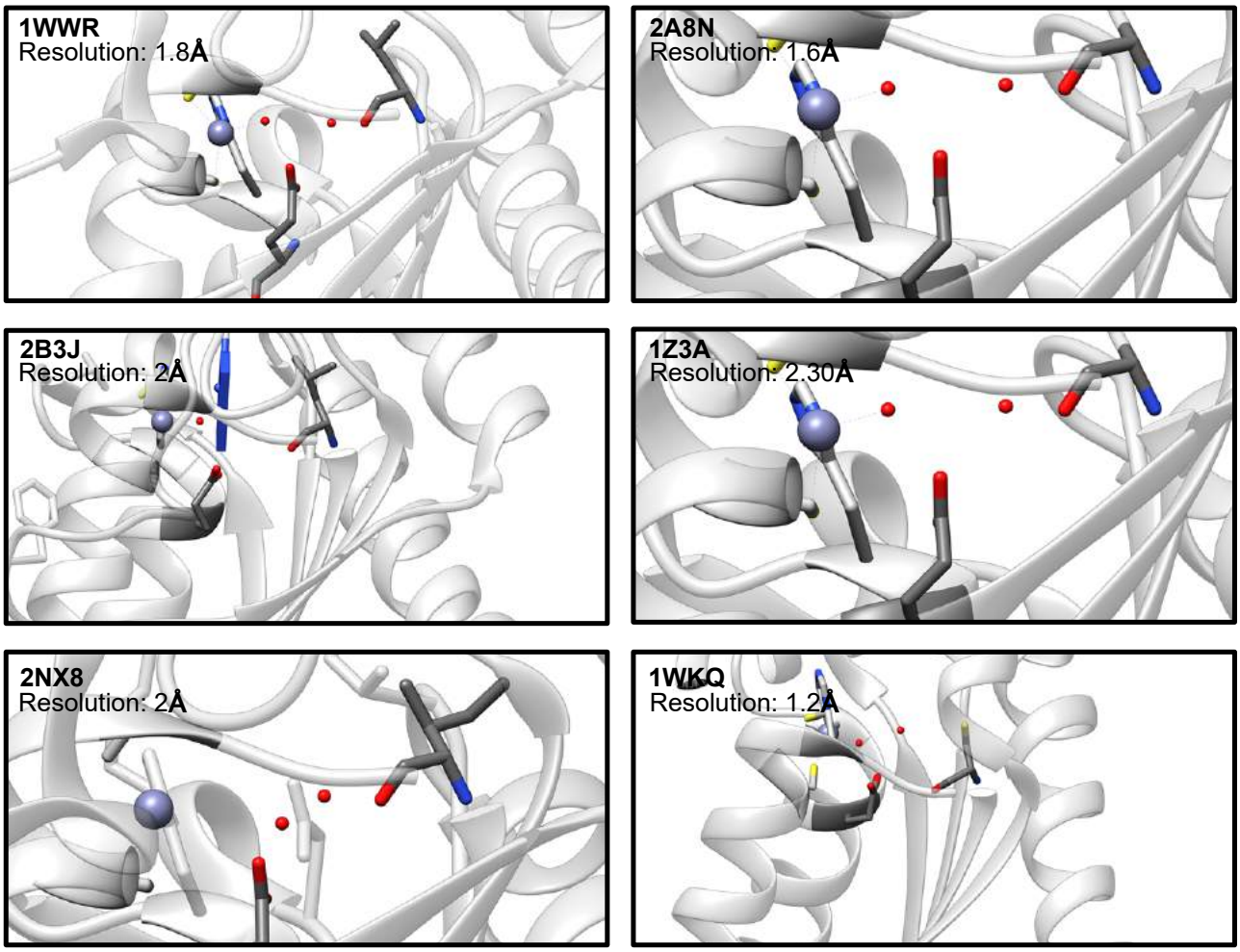


Figure S9: Crystal structures of *ec*TadA homolog highlighting the active site architecture and the relative position of the two critical waters, the activated water and the bridging water, hydrogen bonded to the same glutamate oxygen atom.

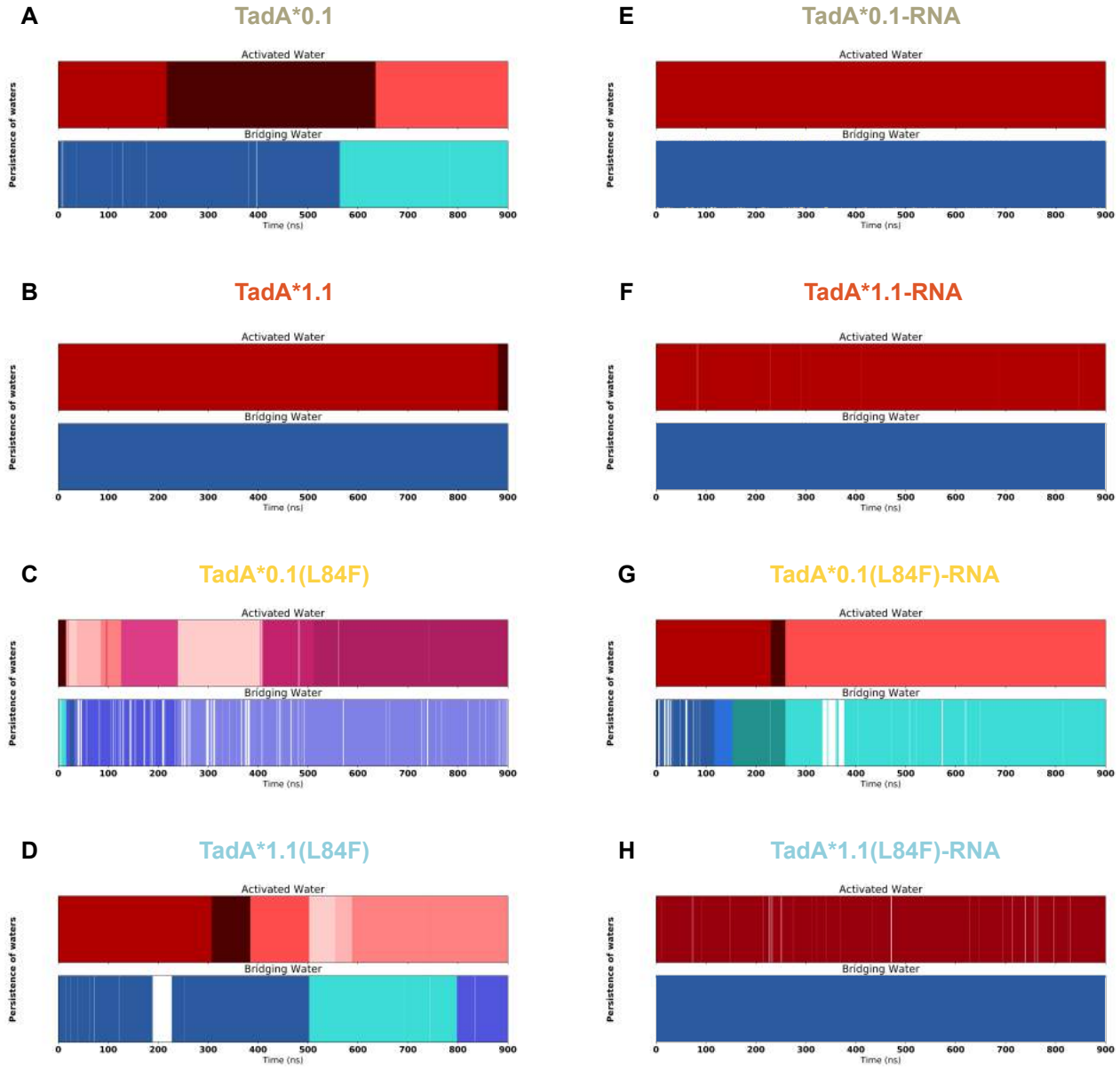
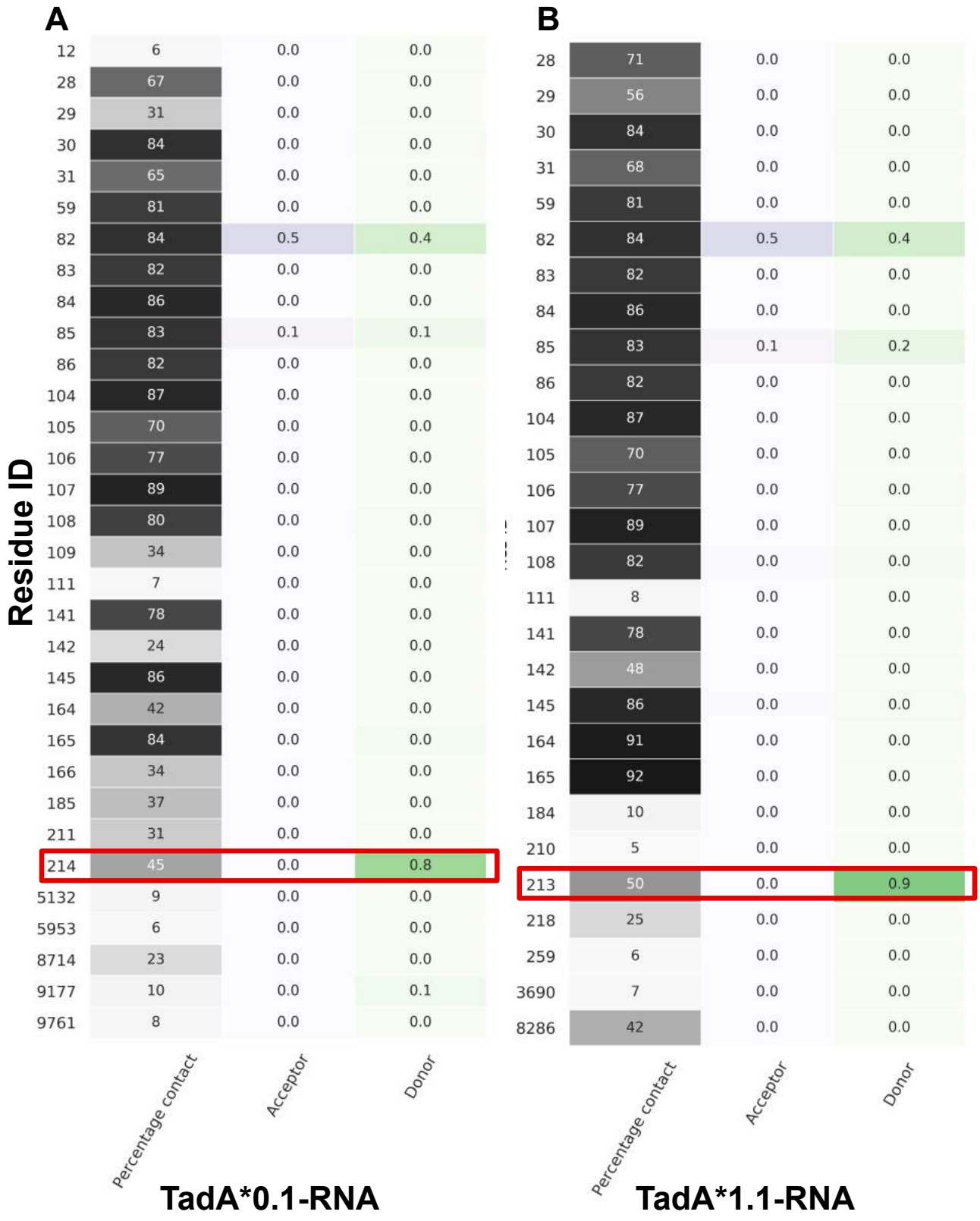


Figure S10: Persistence of the activated and bridging water at their respective positions calculated through the analysis of the unbaised MD trajectories of the apo-TadA* models ((A) to (D)) and for TadA*-RNA models ((E) to (G)). The different colors signify a unique water molecule that visited these critical water positions. Note that the data was obtained using a distance-based `mask` of 3.5 Å from the Zn^{+2} ion, for activated water and from the peptide backbone of the 84 residue, for the bridging water.


A
Residue ID
12 6 0.0 0.0
28 67 0.0 0.0
29 31 0.0 0.0
30 84 0.0 0.0
31 65 0.0 0.0
59 81 0.0 0.0
82 84 0.5 0.4
83 82 0.0 0.0
84 86 0.0 0.0
85 83 0.1 0.1
86 82 0.0 0.0
104 87 0.0 0.0
105 70 0.0 0.0
106 77 0.0 0.0
107 89 0.0 0.0
108 80 0.0 0.0
109 34 0.0 0.0
111 7 0.0 0.0
141 78 0.0 0.0
142 24 0.0 0.0
145 86 0.0 0.0
164 42 0.0 0.0
165 84 0.0 0.0
166 34 0.0 0.0
185 37 0.0 0.0
211 31 0.0 0.0
214 45 0.0 0.8
5132 9 0.0 0.0
5953 6 0.0 0.0
8714 23 0.0 0.0
9177 10 0.0 0.1
9761 8 0.0 0.0
Percentage contact
Acceptor
Donor
TadA*0.1-RNA
B
28 71 0.0 0.0
29 56 0.0 0.0
30 84 0.0 0.0
31 68 0.0 0.0
59 81 0.0 0.0
82 84 0.5 0.4
83 82 0.0 0.0
84 86 0.0 0.0
85 83 0.1 0.2
86 82 0.0 0.0
104 87 0.0 0.0
105 70 0.0 0.0
106 77 0.0 0.0
107 89 0.0 0.0
108 82 0.0 0.0
111 8 0.0 0.0
141 78 0.0 0.0
142 48 0.0 0.0
145 86 0.0 0.0
164 91 0.0 0.0
165 92 0.0 0.0
184 10 0.0 0.0
210 5 0.0 0.0
213 50 0.0 0.9
218 25 0.0 0.0
259 6 0.0 0.0
3690 7 0.0 0.0
8286 42 0.0 0.0
Percentage contact
Acceptor
Donor
TadA*1.1-RNA

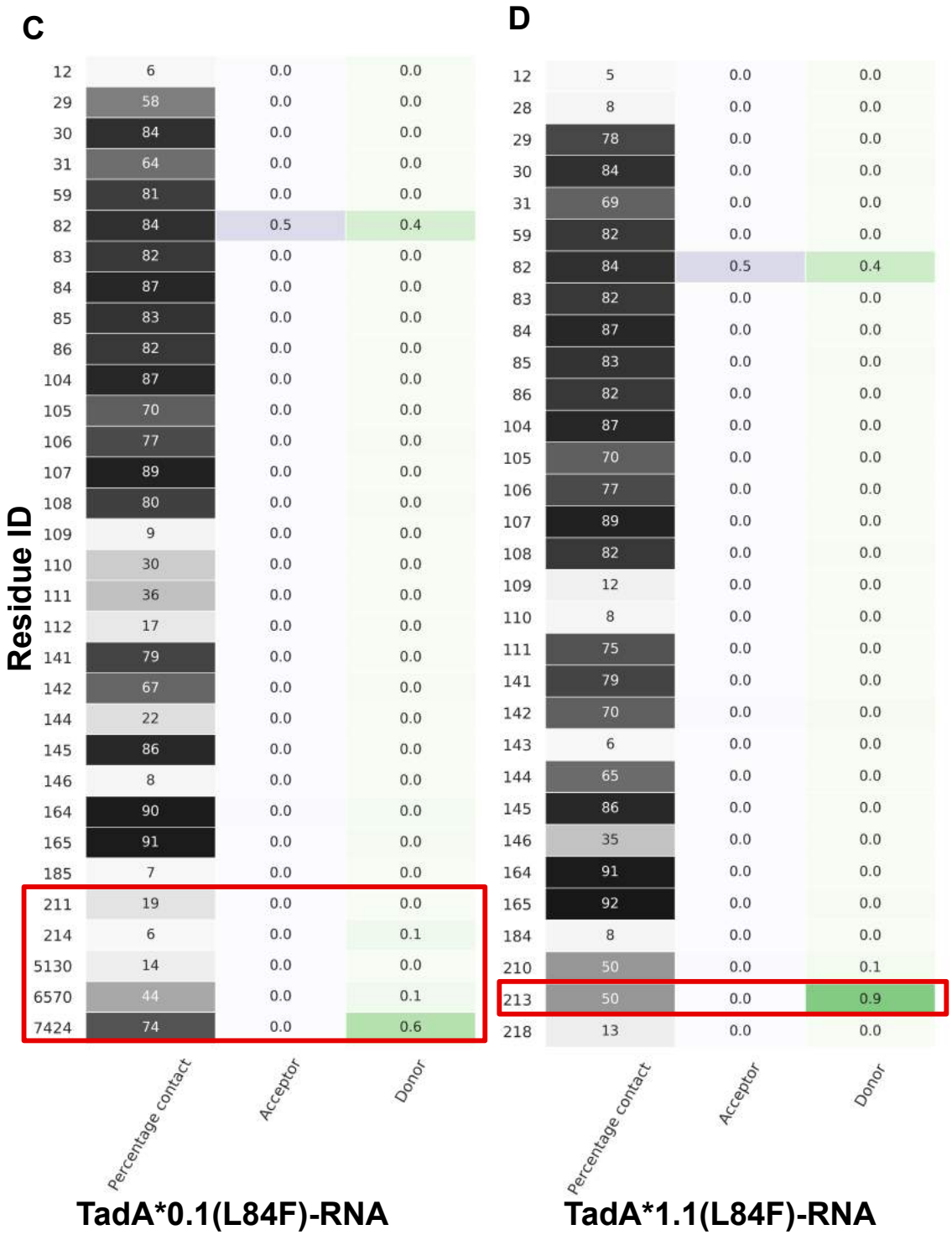


Figure S11: Percentage contact and the fractional H-bonding of the 84 residue. For (A) TadA*0.1, (B) TadA*1.1, (C) TadA*0.1(L84F) and (D) TadA*1.1(L84F) in complex with RNA.The intensity of the colors in the columns signifies the magnitude of the percentage contact and the H-bonding strength. The waters bridging (i.e. the bridging waters) the backbone of 84 residue and the Glu^{59} residue are highlighted for each mutant.

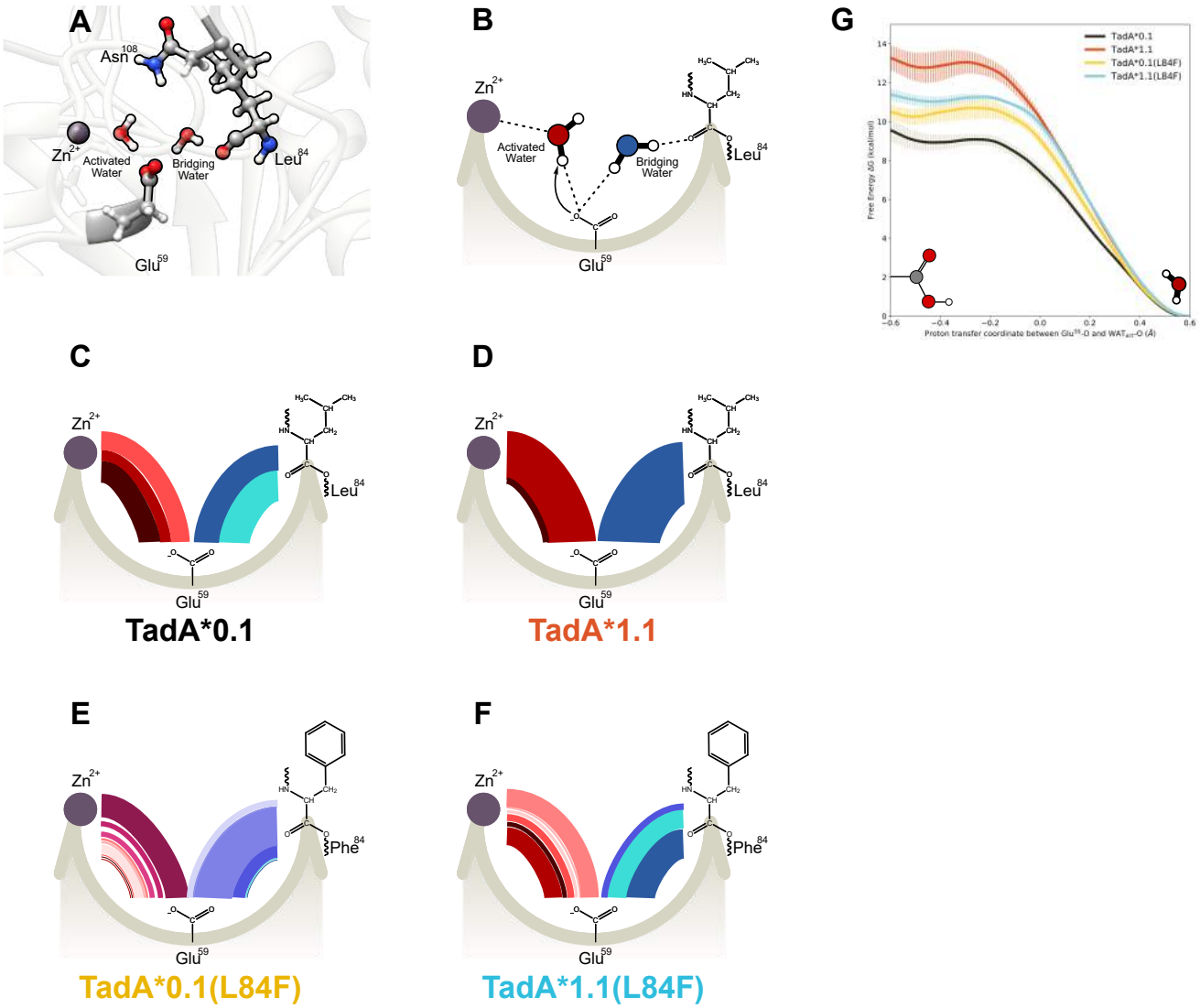


Figure S12: (A) Side view of the of the apo-TadA* models highlighting the location of the catalytically relevant residues. The Zn^{+2}ion is coordinated by His^{57},Cys^{87}, and Cys^{90} (not shown here for clarity) and a water molecule. This water molecule is activated by Glu^{59},which is also connected to another water molecule. This second water acts as a bridge between the Glu^{59} and the carbonyl backbone of residue 84. The target adenine is deep within the active site and residue 108 is farther away from the active site waters. (B) Simplified flat lay representation to highlight the interactions of active site waters. Modified chord diagrams to demonstrate the persistence of the active site waters for (C) TadA*0.1, (D) TadA*1.1, (E) TadA*0.1(L84F), and (F) TadA*1.1(L84F). The red chords connecting Glu^{59} with Zn^{+2} depict the stability of the activated water molecule. Similarly the blue chords connecting Glu^{59} with residue 84 depict the stability of the bridging water molecule. Different colors signify unique water molecules, with the thickness of individual chords being directly proportional to the total time these water molecules interact with the active site of TadA*–RNA during the simulation. (G) Reaction profile for the deprotonation of the activated water molecule the various TadA*-RNA systems.

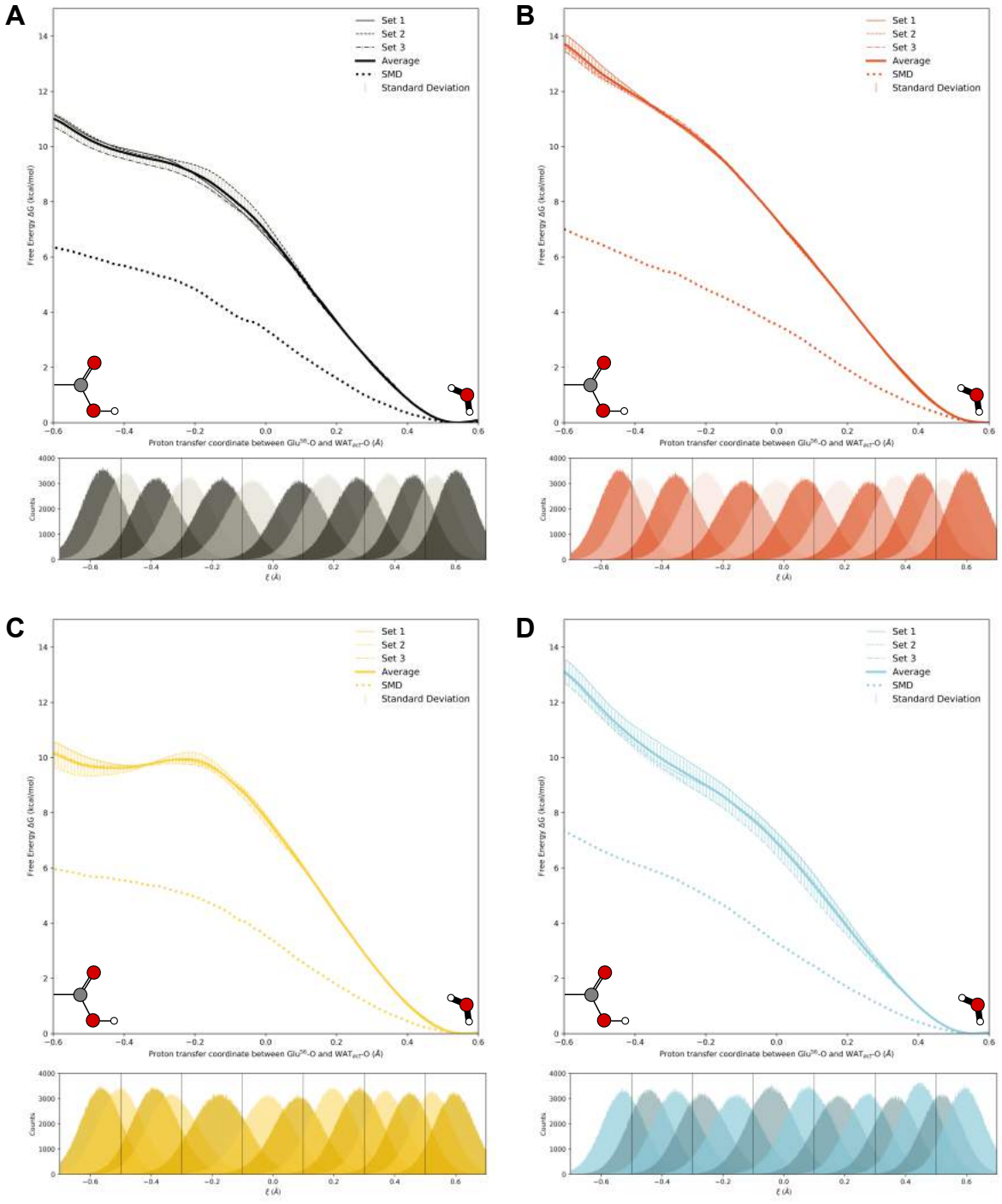


Figure S13: Steered MD, umbrella sampling data and biased statistics obtained from the QM/MM simulations of TadA*-RNA systems. 3 individual pmfs that were used to compute the average pmf for the deprotonation reaction along the ξ for (A) TadA*0.1-RNA, (B) TadA*1.1-RNA, (C) TadA*0.1(L84F)-RNA, and (D) TadA*1.1(L84F)-RNA. In each pmf profile the shaded region indicated the standard deviation of the average pmfs. The right extreme signifies that shared proton resides entirely on the activated water and the left extreme signifies that the shared proton resides entirely on the $Glu^{59}O$ atom.

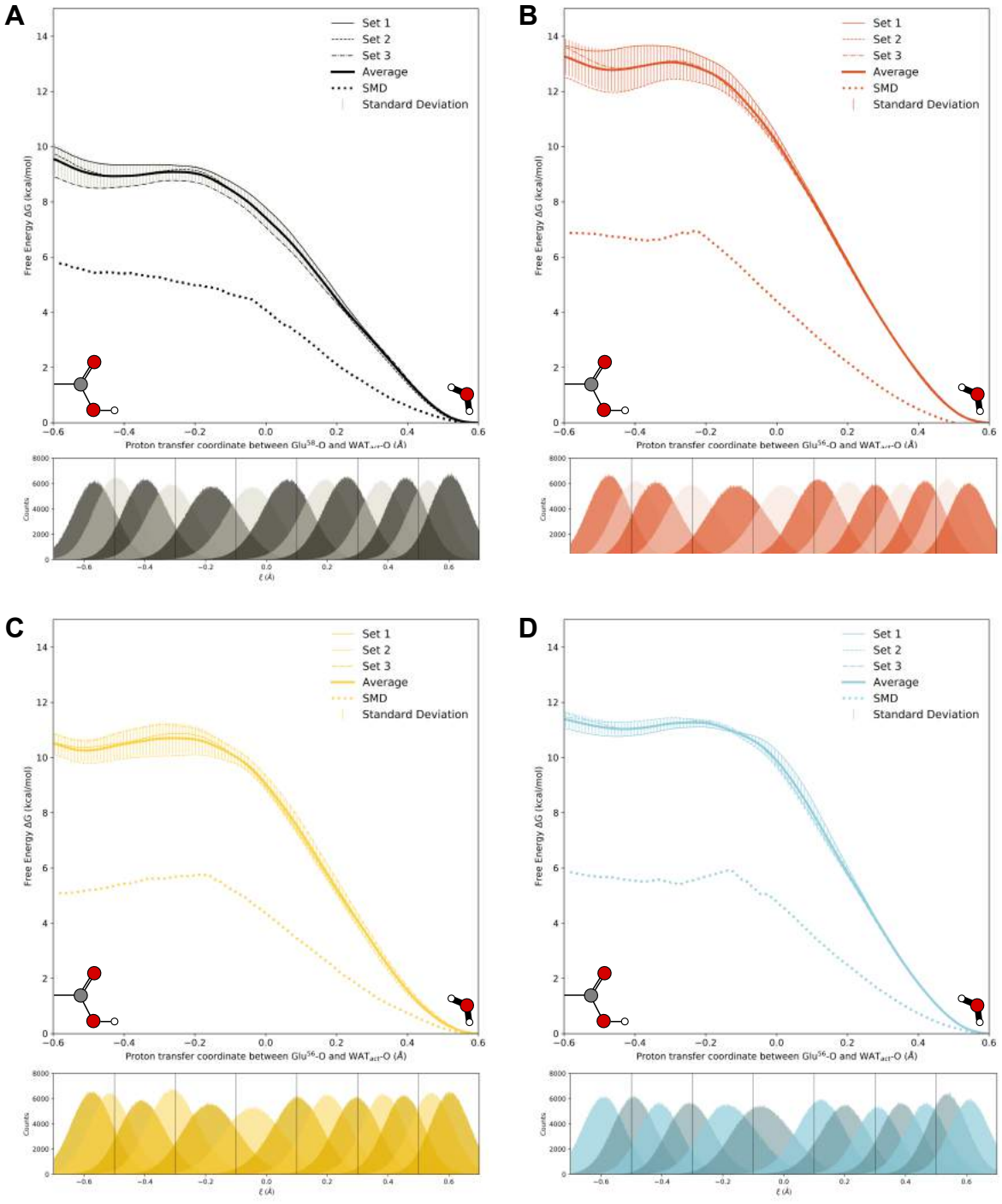


Figure S14: Steered MD,umbrella sampling data and biased statistics obtained from the QM/MM simulations of apo-TadA* systems. 3 individual pmfs that were used to compute the average pmf for the deprotonation reaction along the ξ for (A) TadA*0.1, (B) TadA*1.1, (C) TadA*0.1(L84F), and (D) TadA*1.1(L84F). In each pmf profile the shaded region indicated the standard deviation of the average pmfs. The right extreme signifies that shared proton resides entirely on the activated water and the left extreme signifies that the shared proton resides entirely on the Glu^{59} O atom.

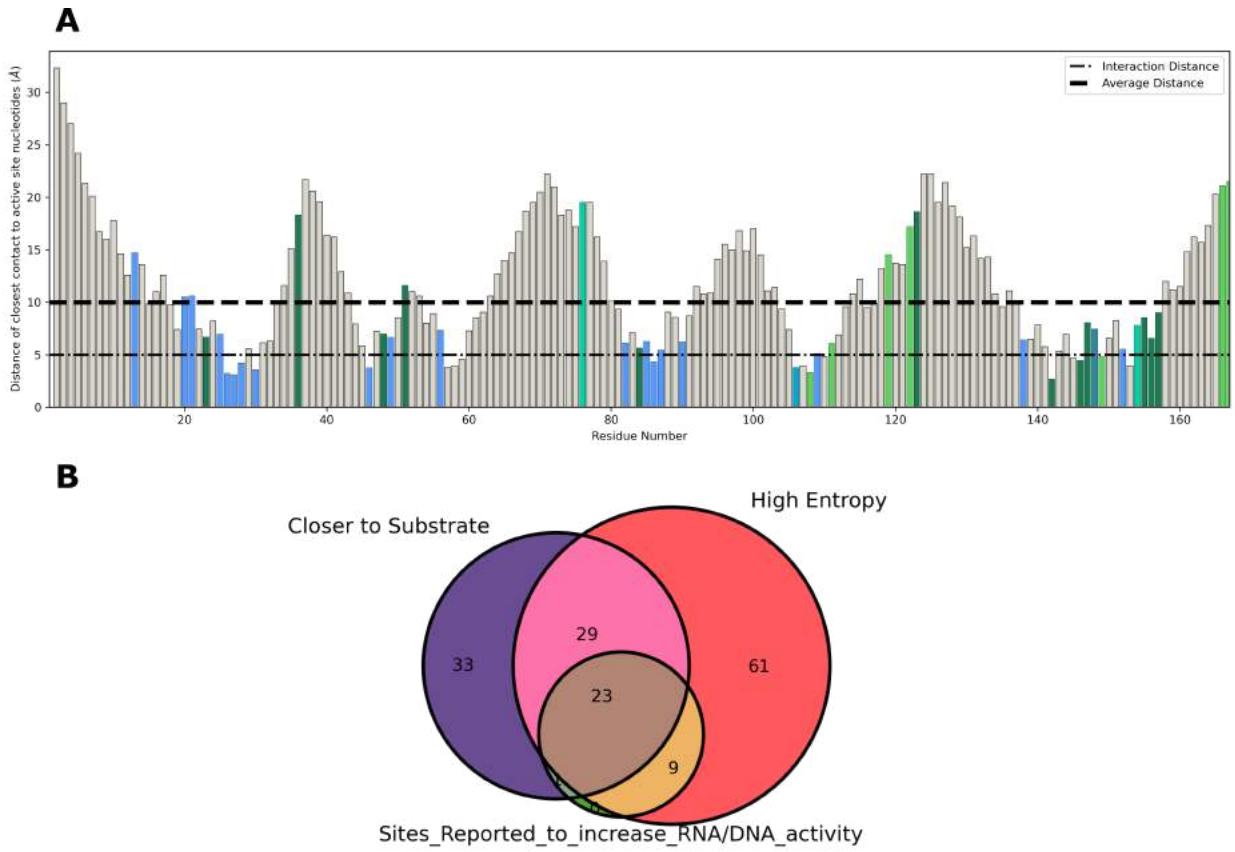


Figure S15: (A) Distance of closest contact between TadA* residues (Cα atom) and the target RNA nucleotide, highlighting the various positions that have mutational data associated to them. 91% of all mutations occur at residues with contact distance less than the average contact distance (10 Å) in TadA*-RNA structure. Moreover, 36.9% of the mutations lie within interaction distance (5 Å) of the nucleotide target. (B) Venn diagram depicting the intersections between the residues with high entropy, residues which have contact distance less than the protein average (10 Å, and the residues which have been shown to increase editing efficiency of TadA* upon mutagenesis.

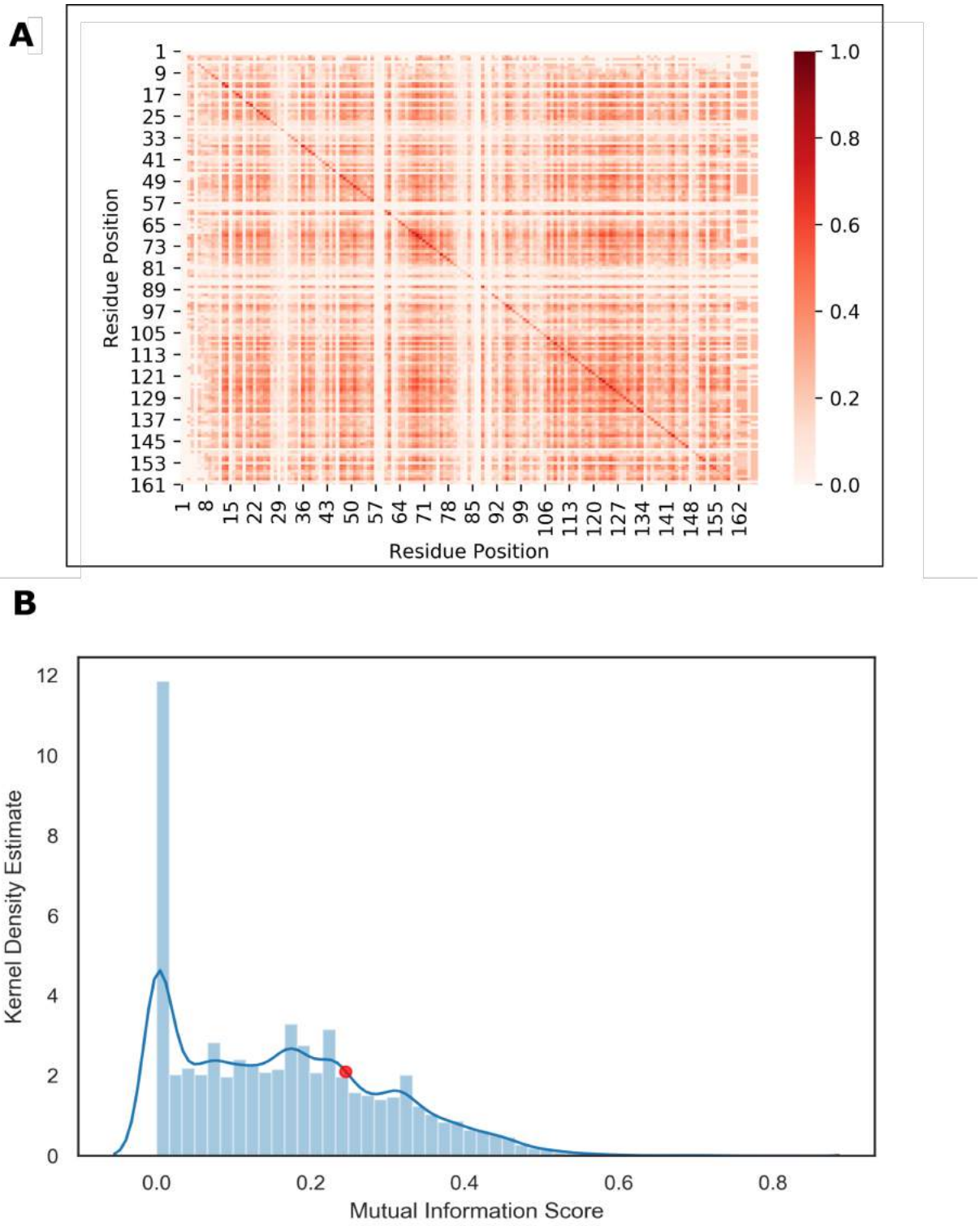


Figure S16: (A) Mutual information scores between pairs of residues for *E. coli* wtTadA. (B) Histogram showing the distribution of these mutual information scores. The value of the mutual information between residue 84 and 108 is indicated in red.

4 Supplementary Tables

Table S1: RNA sequences:

RNA Site	Gene name	Amplicon
1	DNAJB1	GCGCTACCACCCGGACAAGAACAAGGAGCCCGGCGCCGAGGAGAAGTTCAAGGAGATCGC TGAGGCCTACGACGTGCTCAGCGACCCGCGCAAGCGCGAGATCTTCGACCGCTACGGGGA GGAAGGCCTAAAGGGGAGTGGCCCCAGTGGCGGTAGCGGCGGTGGTGCCAATGGTACCTC TTTCAGCTACACATTCCATGGAGACCCTCATGCCATG
2	MTA2	TCTGGCTTCAGGGATTCGTTCAAGCTCACAGCCAGCAGCCAAGCGTCAGAAACTAAACCC AGCTGATGCCCCCAATCCTGTGGTGTTTGTGGCCACAAAGGATACCAGGGCCCTACGGAA GGCTCTGACCCATCTGGAAATGCGGCGAGCTGCTCGCCGACCCAACTTGCCCCTGAAGGT GAAGCCAACGCTGATTGCAGTGCGGCCCCCTGTCCCTCTACCTGCACCCTCACATC
3	PTBP2	AGATTTTGGTAATTCCCCATTGCATCGTTTTAAGAAACCTGGATCCAAAAATTTTCAAAA CATTTTTCCTCCTTCTGCCACCCTTCACCTATCTAATATCCCTCCATCAGTAGCAGAAGA GGATCTACGAACACTGTTCGCTAACACTGGGGGCACTGTGAAAGCATTTAAGTTTTTTCA AAGAGATCACAAAATGGCTCTTCTTCAGATGGCAACAGTGGAAGAAGCTATTCAGG
4	SAP30BP	CAGAACCCCCTGGCAGATGTTCAAATCACTTGCAAGACAAGATCCAGAAGCTTTATGAAC GAAAGATAAAGGAGGGAATGGATATGAACTACATTATCCAAAGGAAGAAAGAATTTCGGA ACCCTAGCATCTACGAGAAGCTGATCCAGTTCTGTGCCATTGACGAGCTTGGCACCAACT ACCCAAAGGATATGTTTGATCCCCATGGCTGGTCTGAGGACTCCTACT
5	LCMT1	ATTGCCAACACTCCTGATAGCTGAATGTGTGCTGGTTTACATGACTCCAGAGCAGTCCGC AAACCTCCTGAAGTGGGCAGCCAACAGTTTTGAGAGAGCCATGTTCATAAACTACGAACA GGTGAACATGGGTGATCGGTTTGGGCAGATCATGATTGAAAACCTGCGGAGACGCCAGTG TGACCTGGCGGGAGTGGAGACCTGCAAGTC
6	SCAP	CCATTGACATTCGCCGGATGGAGCTAGCAGACCTGAACAAGCGACTGCCCCCTGAGGCCT GCCTGCCCTCAGCCAAGCCAGTGGGACAGCCAACGCGCTACGAGCGGCAGCTGGCTGTGA GGCCGTCCACACCCCACACCATCACGTTGCAGCCGTCTTCCTTCCGAAACCTGCGGCTCC CCAAGAGGCTGCGTGTTGTCTACTTC
gRNA sequence		GGTATTACTGATATTGGTGGG

Table S2: Primers:

RNA Site	Primer 1	Primer 2
1	CATGGCATGAGGGTCTCCATGG	GCGCTACCACCCGGACAAG
2	GATGTGAGGGTGCAGGTAGAGGG	TCTGGCTTCAGGGATTCGTTCAAG
3	AGATTTTGGTAATTCCCCATTGCATCG	CCTGAATAGCTTCTTCCACTGTTGCC
4	CAGAACCCCCTGGCAGATGTTC	AGTAGGAGTCCTCAGACCAGCC
5	ATTGCCAACACTCCTGATAGCTGAATG	GACTTGCAGGTCTCCACTCCCG
6	GAAGTAGACAACACGCAGCCTCTTG	CCATTGACATTCGCCGGATGGAG

I5 extension on primer 1s: ACACTCTTTCCCTACACGACGCTCTTCCGATCT

I7 extension on primer 2s: GACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Table S3: Summary of the systems modeled and the types of simulations conducted in this study.

System	Number of Atoms	Simulation Type	Simulation Length
TadA*0.1-RNA	49461	Unbiased MD	1 μs
TadA*1.1-RNA	49471		
TadA*0.1(L84F)-RNA	49456		
TadA*1.1(L84F)-RNA	49463		
TadA*0.1-RNA	49461	Steered MD	200 ns
TadA*1.1-RNA	49471		
TadA*0.1(L84F)-RNA	49456		
TadA*1.1(L84F)-RNA	49463		
TadA*0.1-RNA	49461	Umbrella Sampling	41 windows $\times$ 5 $\frac{ns}{window}$ $\times$ 4 sets
TadA*1.1-RNA	49471		
TadA*0.1(L84F)-RNA	49456		
TadA*1.1(L84F)-RNA	49463		
TadA*0.1-RNA	49461	Umbrella Sampling	41 windows $\times$ 10 $\frac{ns}{window}$ $\times$ 5 sets
TadA*0.1	45752	QM/MM Steered MD	500 ps
TadA*1.1	45740		
TadA*0.1(L84F)	45693		
TadA*1.1(L84F)	45693		
TadA*0.1-RNA	49461		
TadA*1.1-RNA	49471		
TadA*0.1(L84F)-RNA	49456		
TadA*1.1(L84F)-RNA	49463		
TadA*0.1-RNA	45486	QM/MM Umbrella Sampling	13 windows $\times$ 150 $\frac{ps}{window}$ $\times$ 3 sets
TadA*1.1-RNA	45693		
TadA*0.1(L84F)-RNA	49456		
TadA*1.1(L84F)-RNA	49463		
TadA*0.1-RNA	49461		
TadA*1.1-RNA	49471		
TadA*0.1(L84F)-RNA	49456		
TadA*1.1(L84F)-RNA	49463		

Table S4: Analysis of the novelty of mutations in experiments by detecting frequency of prevalence across all hits in the filtered multiple sequence alignment output. * denotes that the residue site was mutated after the ABE7.10 mutations. Sites which have been mutated multiple times producing conflicting editing outcomes have been colored gray to highlight only the outcome that represents the chemically conserved mutation.

TadA* Mutation (Laboratory evolution)	DNA-editing outcome	RNA-editing outcome	Literature Reference	Entropy Score (H_i)	Frequency of occurrence (within the dataset of natural TadA homologs)
R13A	Active	Active	[35]	0.65	3
K20A	Active	Active	[35]	0.58	0
R21A	Active	Active	[35]	0.66	0
W23R	Active	Active	[37]	0.65	1
W23L	Active	Active	[37]	0.65	12
R23A*	Active	Active	[35]	0.65	1
E25A	Active	Active	[35]	0.61	2

R26A	Active	Active	[35]	0.60	6
E27A	Inactive	Inactive	[35]	0.27	0
V28G	Inactive	Inactive	[35]	0.41	1
V30G	Inactive	Inactive	[35]	0.27	0
H36L	Active	Active	[37]	0.72	1
N46A	Inactive	Inactive	[35]	0.07	0
R47Q	Active	Active	[38]	0.65	0
R47F	Active	Active	[38]	0.65	0
R47W	Active	Active	[38]	0.65	0
R47M	Active	Active	[38]	0.65	2
P48A	Active	Active	[37]	0.63	0
P48S	Active	Active	[37]	0.63	3
A48G*	Active	Active	[35]	0.63	0
I49A	Active	Active	[35]	0.54	2
R51L	Active	Active	[37]	0.72	3
D53E	Active	Active	[39]	0.5	0
A56G	Active	Active	[35]	0.64	0
E59A	Inactive	Inactive	[38]	0	0
I76Y	Active	Active	[40]	0.75	0
V82G	Active	Inactive	[35]	0.13	0
V82W	Inactive	Inactive	[35]	0.13	0
V82S	Active	Active	[40]	0.13	0
L84F	Active	Active	[37]	0.42	0
E85A	Inactive	Inactive	[35]	0.18	0
P86A	Inactive	Inactive	[35]	0	0
C87A	Inactive	Inactive	[35]	0	0
C90A	Inactive	Inactive	[35]	0	0
A106V	Active	Active	[37]	0.41	0
V106G*	Active	Inactive	[35]	0.41	0
V106W*	Active	Inactive	[38]	0.41	0
V106Q*	Active	Inactive	[38]	0.41	0
V106F*	Active	Inactive	[38]	0.41	0
V106M*	Active	Inactive	[38]	0.41	0
D108N	Active	Active	[37]	0.46	12
N108A*	Active	Active	[35]	0.46	0
N108Q*	Inactive	Inactive	[38]	0.46	0
N108F*	Inactive	Inactive	[38]	0.46	0
N108W*	Inactive	Inactive	[38]	0.46	0
N108M*	Inactive	Inactive	[38]	0.46	0
N108K*	Inactive	Inactive	[38]	0.46	0
A109G	Active	Active	[35]	0.73	0
A109S	Active	Active	[41]	0.73	7
T111A	Active	Active	[35]	0.71	1
T111R	Active	Active	[41]	0.71	0
D119N	Active	Active	[41]	0.75	8
H122N	Active	Active	[41]	0.79	8
H123Y	Active	Active	[37]	0.79	0
Y123H*	Active	Active	[40]	0.79	5
A138G	Active	Active	[35]	0.61	0
A142L	Active	Active	[37]	0.65	0
A142G	Active	Active	[35]	0.65	2

A143G	Active	Active	[35]	0.77	1
S146C	Active	Active	[37]	0.59	0
D147Y	Active	Active	[37]	0.68	0
D147R*	Active	Active	[40]	0.68	0
Y147D*	Active	Active	[41]	0.68	5
F148A	Active	Active	[35]	0.11	0
F149A	Active	Active	[35]	0.37	0
F149Y	Active	Active	[41]	0.37	13
R152P	Active	Active	[37]	0.82	0
P152A*	Active	Active	[35]	0.82	0
Q154R	Active	Active	[40]	0.92	0
E155V	Active	Active	[37]	0.92	1
V155G*	Active	Active	[35]	0.92	0
V155W*	Active	Active	[35]	0.92	0
I156F	Active	Active	[37]	0.85	0
K157N	Active	Active	[37]	0.88	0
T166I	Active	Active	[41]	1.0	1
D167N	Active	Active	[41]	1.0	1

Table S5: Amino acid distribution for individual residue sites of TadA in the naturally occurring homologs of the enzyme. The wild type amino acid in *E. coli* TadA have been highlighted. '-' represents a gap in the alignment.

Residue Site	**Entropy Score (H_i)**	**Amino Acid Distribution**
1	0.97	'-': 32, **'M': 4**
2	0.97	'-': 32, **'S': 4**
3	1.01	'-': 32, 'D': 2, **'E': 1**, 'S': 1
4	1.00	'-': 31, **'V': 3**, 'C': 1, 'T': 1
5	0.95	'-': 31, **'E': 5**
6	0.99	'-': 30, 'L': 3, **'F': 2**, 'D': 1
7	0.96	'-': 28, **'S': 4**, 'D': 3, 'N': 1
8	0.96	'-': 28, **'H': 5**, 'D': 2, 'E': 1
9	0.76	'-': 18, **'E': 13**, 'K': 3, 'Q': 1, 'N': 1
10	0.64	**'Y': 16**, '-': 8, 'K': 5, 'T': 2, 'F': 2, 'L': 1, 'H': 1, 'A': 1
11	0.56	'F': 13, **'W': 11**, '-': 6, 'Y': 4, 'M': 1, 'G': 1
12	0.13	**'M': 31**, 'L': 5
13	0.65	'E': 8, 'K': 7, **'R': 6**, 'Q': 6, 'A': 3, 'T': 2, 'H': 2, 'G': 2
14	0.72	'E': 10, 'Q': 5, 'V': 4, **'H': 3**, 'I': 3, 'K': 3, 'R': 3, 'L': 2, 'Y': 1, 'C': 1, 'F': 1
15	0.00	**'A': 36**
16	0.44	**'L': 20**, 'I': 6, 'F': 5, 'M': 3, 'V': 1, 'T': 1
17	0.66	'K': 15, **'T': 4**, 'E': 3, 'Q': 3, 'R': 2, 'H': 2, 'D': 2, 'L': 2, 'F': 1, 'V': 1, 'A': 1
18	0.61	'E': 9, **'L': 8**, 'Q': 7, 'M': 6, 'Y': 2, 'S': 2, 'V': 1, 'A': 1
19	0.10	**'A': 33**, 'S': 3
20	0.58	**'K': 15**, 'E': 6, 'Q': 6, 'R': 3, 'Y': 2, 'D': 1, 'G': 1, 'V': 1, 'C': 1
21	0.67	'K': 9, **'R': 7**, 'L': 5, 'I': 4, 'E': 4, 'D': 3, 'Y': 2, 'S': 1, 'Q': 1
22	0.28	**'A': 23**, 'S': 10, 'G': 3
23	0.65	'L': 12, 'F': 8, 'E': 4, **'W': 3**, 'K': 3, 'A': 1, 'G': 1, 'V': 1, 'Y': 1, 'R': 1, 'D': 1
24	0.61	**'D': 11**, 'E': 9, 'Q': 6, 'N': 3, 'S': 2, 'T': 2, 'A': 1, 'M': 1, 'C': 1
25	0.61	'K': 13, 'N': 7, **'E': 6**, 'L': 2, 'A': 2, 'T': 2, 'P': 2, 'I': 1, 'R': 1

26	0.60	'G': 13, 'N': 7, 'A': 6, **'R': 4**, 'T': 2, 'D': 1, 'I': 1, 'L': 1, 'S': 1
27	0.27	**'E': 28**, 'S': 3, 'G': 2, 'P': 2, 'Q': 1
28	0.42	**'V': 21**, 'I': 8, '-': 3, 'N': 2, 'G': 1, 'H': 1
29	0.11	**'P': 33**, 'Q': 2, 'K': 1
30	0.27	**'V': 23**, 'I': 11, 'F': 2
31	0.00	**'G': 36**
32	0.32	**'A': 18**, 'C': 15, 'V': 2, 'S': 1
33	0.36	**'V': 22**, 'L': 6, 'I': 5, 'C': 3
34	0.42	'I': 18, **'L': 9**, 'M': 6, 'V': 2, 'F': 1
35	0.13	**'V': 33**, 'T': 1, 'L': 1, 'I': 1
36	0.72	'K': 10, 'Y': 7, **'H': 3**, 'D': 3, 'F': 3, 'E': 3, 'N': 2, 'I': 1, 'L': 1, 'C': 1, 'S': 1, 'Q': 1
37	0.46	**'N': 16**, 'D': 13, 'S': 2, 'K': 1, 'Q': 1, 'E': 1, 'H': 1, 'G': 1
38	0.61	**'N': 11**, 'G': 10, 'Q': 5, 'D': 3, 'H': 2, 'E': 2, 'A': 1, 'K': 1, 'R': 1
39	0.55	'E': 12, 'K': 10, **'R': 7**, 'T': 2, 'Q': 2, 'A': 1, 'I': 1, 'N': 1
40	0.22	**'I': 22**, 'V': 14
41	0.27	**'I': 25**, 'V': 9, 'L': 1, 'R': 1
42	0.35	**'G': 24**, 'A': 6, 'V': 3, 'S': 2, 'M': 1
43	0.59	'K': 9, 'R': 8, **'E': 7**, 'S': 6, 'I': 3, 'T': 2, 'Q': 1
44	0.28	**'G': 26**, 'S': 5, 'A': 4, 'T': 1
45	0.54	'H': 13, **'W': 8**, 'R': 7, 'Y': 4, 'G': 2, 'L': 1, 'Q': 1
46	0.07	**'N': 34**, 'H': 2
47	0.65	'E': 8, **'R': 6**, 'A': 6, 'N': 6, 'S': 4, 'L': 3, 'M': 2, 'K': 1
48	0.64	'V': 9, 'R': 8, **'P': 6**, 'T': 5, 'S': 3, 'I': 2, 'Q': 1, '-': 1, 'K': 1
49	0.54	'E': 13, 'N': 9, **'I': 8**, 'A': 2, 'R': 1, 'T': 1, 'V': 1, 'Y': 1
50	0.54	'E': 16, **'G': 8**, 'Q': 4, 'T': 3, 'S': 2, 'V': 1, 'L': 1, 'P': 1
51	0.72	'K': 8, 'T': 7, 'S': 6, **'R': 3**, 'L': 3, 'Q': 2, 'E': 2, 'C': 2, 'N': 1, 'D': 1, 'A': 1
52	0.55	'N': 15, 'K': 8, **'H': 5**, 'S': 3, 'Q': 2, 'R': 1, 'G': 1, 'A': 1
53	0.50	**'D': 13**, 'N': 12, 'Q': 6, '-': 2, 'R': 1, 'S': 1, 'G': 1
54	0.49	'A': 17, **'P': 10**, 'S': 2, 'V': 2, '-': 2, 'D': 2, 'C': 1
55	0.45	**'T': 19**, 'I': 7, 'L': 5, '-': 2, 'V': 2, 'S': 1
56	0.64	**'A': 12**, 'R': 6, 'M': 5, 'C': 4, 'Y': 3, 'Q': 2, '-': 2, 'L': 1, 'T': 1
57	0.00	**'H': 36**
58	0.04	**'A': 35**, 'G': 1
59	0.00	**'E': 36**
60	0.43	**'I': 15**, 'M': 15, 'L': 2, 'H': 1, 'V': 1, 'T': 1, 'A': 1
61	0.67	'I': 9, **'M': 8**, 'V': 6, 'N': 5, 'L': 2, 'K': 1, 'A': 1, 'T': 1, 'E': 1, 'S': 1, 'R': 1
62	0.07	**'A': 34**, 'V': 2
63	0.28	'I': 23, **'L': 11**, '-': 1, 'V': 1
64	0.57	**'R': 11**, 'N': 11, 'D': 6, '-': 5, 'E': 2, 'K': 1
65	0.65	'E': 12, **'Q': 9**, '-': 5, 'N': 3, 'R': 2, 'M': 2, 'G': 1, 'T': 1, 'K': 1
66	0.52	'A': 19, **'G': 6**, '-': 6, 'V': 2, 'D': 2, 'S': 1
67	0.67	**'G': 10**, 'C': 9, 'N': 5, '-': 5, 'A': 3, 'R': 1, 'L': 1, 'W': 1, 'D': 1
68	0.74	**'L': 8**, 'K': 7, 'A': 5, 'N': 4, 'R': 3, '-': 2, 'M': 2, 'E': 1, 'S': 1, 'D': 1, 'C': 1, 'W': 1
69	0.84	**'V': 6**, 'H': 5, 'N': 4, 'L': 4, 'I': 3, 'A': 2, 'P': 2, 'S': 2, 'F': 1, 'K': 1, 'R': 1, 'W': 1, 'E': 1, 'C': 1, '-': 1, 'Y': 1
70	0.68	'L': 9, 'I': 8, 'E': 6, **'M': 3**, 'P': 2, 'C': 2, 'K': 2, 'S': 1, 'G':1, 'T': 1, '-': 1
71	0.67	'G': 9, 'S': 7, 'K': 6, **'Q': 4**, 'N': 4, 'D': 2, 'E': 1, 'H': 1, 'A':1, '-': 1
72	0.62	**'N': 13**, 'S': 9, 'T': 4, 'Q': 3, 'V': 1, 'E': 1, 'W': 1, 'A': 1, 'G': 1, '-': 1, 'R': 1
73	0.68	**'Y': 9**, 'W': 7, 'K': 7, 'E': 4, 'V': 2, 'T': 2, 'F': 1, 'C': 1, 'H':1, 'S': 1, 'R': 1
74	0.54	**'R': 17**, 'K': 6, 'N': 5, 'D': 3, 'Y': 1, 'E': 1, '-': 1, 'G': 1, 'Q': 1
75	0.38	**'L': 25**, 'F': 3, '-': 2, 'A': 2, 'V': 2, 'Q': 1, 'S': 1

76	0.75	'L': 10, 'E': 6, 'N': 4, 'S': 3, 'K': 3, **'I': 2**, '-': 2, 'V': 1, 'P': 1, 'R': 1, 'T': 1, 'D': 1, 'Q': 1
77	0.51	**'D': 16**, 'G': 9, 'N': 4, 'H': 3, 'K': 2, 'E': 1, 'P': 1
78	0.58	'T': 13, 'C': 8, 'Y': 5, **'A': 4**, 'S': 3, 'V': 1, 'I': 1, 'G': 1
79	0.43	**'T': 20**, 'D': 6, 'V': 6, 'S': 2, 'E': 1, 'I': 1
80	0.29	**'L': 26**, 'I': 6, 'A': 2, 'M': 2
81	0.13	**'Y': 31**, 'F': 5
82	0.14	**'V': 32**, 'T': 3, 'I': 1
83	0.14	**'T': 32**, 'S': 2, 'A': 2
84	0.42	**'L': 21**, 'V': 6, 'I': 5, 'C': 2, 'T': 1, 'R': 1
85	0.18	**'E': 31**, 'Q': 2, 'F': 2, 'S': 1
86	0.00	**'P': 36**
87	0.00	**'C': 36**
88	0.62	'I': 10, **'V': 9**, 'A': 6, 'P': 4, 'T': 2, 'N': 2, 'M': 1, 'S': 1, 'Y': 1
89	0.18	**'M': 31**, 'P': 2, 'E': 2, 'A': 1
90	0.00	**'C': 36**
91	0.40	**'A': 21**, 'S': 8, 'C': 4, 'Y': 1, 'L': 1, 'T': 1
92	0.47	**'G': 19**, 'A': 8, 'S': 3, 'N': 2, 'K': 2, 'Y': 1, 'D': 1
93	0.14	**'A': 32**, 'L': 3, 'T': 1
94	0.31	'I': 23, 'L': 9, **'M': 3**, 'V': 1
95	0.68	**'I': 8**, 'R': 6, 'V': 5, 'G': 5, 'A': 4, 'S': 4, 'L': 2, 'Y': 1, 'F':1
96	0.56	'L': 14, **'H': 10**, 'Q': 4, 'F': 2, 'W': 2, 'N': 1, 'M': 1, 'K': 1, 'A': 1
97	0.41	**'S': 19**, 'A': 9, 'L': 4, 'M': 3, 'Y': 1
98	0.28	**'R': 24**, 'G': 8, 'K': 4
99	0.26	**'I': 28**, 'L': 5, 'V': 1, 'P': 1, 'F': 1
100	0.50	'K': 14, 'P': 12, **'G': 3**, 'E': 3, 'R': 2, 'S': 1, 'N': 1
101	0.52	**'R': 18**, 'H': 5, 'L': 5, 'E': 3, 'K': 2, 'F': 1, 'A': 1, 'N': 1
102	0.24	**'V': 25**, 'L': 10, 'M': 1
103	0.39	**'V': 18**, 'I': 10, 'F': 6, 'Y': 2
104	0.31	'Y': 21, **'F': 12**, 'I': 2, 'M': 1
105	0.24	**'G': 29**, 'S': 3, 'M': 2, 'A': 1, '-': 1
106	0.43	**'A': 21**, 'C': 7, '-': 3, 'S': 3, 'E': 1, 'N': 1
107	0.74	'S': 12, **'R': 5**, 'Q': 4, '-': 3, 'N': 2, 'L': 2, 'H': 2, 'D': 1, 'F': 1, 'K': 1, 'E': 1, 'P': 1, 'I': 1
108	0.46	**'D': 17**, 'N': 12, '-': 3, 'C': 1, 'Y': 1, 'H': 1, 'V': 1
109	0.73	'S': 7, 'E': 6, 'P': 5, 'Q': 5, **'A': 4**, 'N': 2, 'K': 2, '-': 2, 'Y': 1, 'T': 1, 'D': 1
110	0.47	**'K': 20**, 'R': 6, 'D': 4, 'N': 2, '-': 2, 'S': 1, 'F': 1
111	0.71	'F': 12, 'H': 5, **'T': 4**, 'G': 3, 'L': 2, 'P': 2, '-': 2, 'E': 2, 'Y': 1, 'S': 1, 'Q': 1, 'A': 1
112	0.32	**'G': 26**, 'K': 4, 'A': 3, '-': 2, 'D': 1
113	0.64	'G': 13, **'A': 9**, 'N': 2, 'C': 2, 'V': 2, '-': 2, 'T': 2, 'H': 1, 'K': 1, 'E': 1, 'I': 1
114	0.64	**'A': 10**, 'V': 8, 'C': 7, 'S': 4, '-': 2, 'I': 1, 'L': 1, 'R': 1, 'N': 1, 'E': 1
115	0.59	**'G': 16**, 'E': 6, 'D': 5, '-': 2, 'A': 2, 'Q': 1, 'F': 1, 'P': 1, 'V': 1, 'K': 1
116	0.41	**'S': 25**, '-': 3, 'K': 2, 'R': 2, 'F': 1, 'T': 1, 'L': 1, 'N': 1
117	0.57	**'L': 15**, 'V': 7, 'N': 5, 'G': 3, '-': 3, 'R': 1, 'Y': 1, 'I': 1
118	0.62	'L': 14, 'I': 5, 'Y': 5, **'M': 3**, '-': 3, 'F': 2, 'K': 2, 'V': 2
119	0.75	'N': 8, 'R': 6, **'D': 5**, 'Q': 5, '-': 5, 'L': 2, 'H': 1, 'C': 1, 'K': 1, 'S': 1, 'E': 1
120	0.66	'I': 10, 'Y': 7, **'V': 6**, '-': 6, 'L': 4, 'F': 2, 'C': 1
121	0.55	**'L': 13**, 'F': 11, '-': 5, 'A': 4, 'S': 2, 'K': 1
122	0.79	'N': 8, 'T': 5, '-': 5, **'H': 4**, 'S': 4, 'D': 3, 'Q': 2, 'C': 1, 'I': 1, 'G': 1, 'K': 1, 'R': 1
123	0.79	'D': 8, **'H': 5**, 'S': 5, '-': 4, 'A': 3, 'Q': 2, 'E': 2, 'K': 2, 'M': 2, 'N': 1, 'L': 1, 'I': 1
124	0.70	**'P': 12**, 'E': 6, '-': 5, 'S': 4, 'L': 2, 'K': 2, 'A': 2, 'Y': 1, 'F': 1, 'N': 1

125	0.84	**‘G’: 7**, ‘R’: 6, ‘-’: 5, ‘N’: 3, ‘A’: 3, ‘T’: 2, ‘S’: 2, ‘D’: 2, ‘K’: 1, ‘F’: 1, ‘Q’: 1, ‘E’: 1, ‘V’: 1, ‘H’: 1
126	0.77	‘L’: 6, ‘C’: 6, ‘T’: 6, ‘-’: 5, **‘M’: 4**, ‘I’: 2, ‘K’: 2, ‘F’: 2, ‘V’: 2, ‘S’: 1
127	0.64	**‘N’: 14**, ‘G’: 6, ‘F’: 5, ‘-’: 5, ‘T’: 2, ‘D’: 1, ‘H’: 1, ‘E’: 1, ‘S’: 1
128	0.63	**‘H’: 17**, ‘-’: 4, ‘Y’: 3, ‘T’: 3, ‘K’: 2, ‘R’: 2, ‘F’: 2, ‘E’: 1, ‘N’: 1, ‘V’: 1
129	0.60	**‘R’: 17**, ‘K’: 5, ‘P’: 3, ‘-’: 3, ‘A’: 2, ‘G’: 2, ‘T’: 1, ‘Q’: 1,‘S’: 1, ‘E’: 1
130	0.68	**’V’: 10**, ’P’: 6, ’F’: 6, ’L’: 3, ’I’: 3, ’-’: 3, ’A’: 2, ’K’: 2,’Y’: 1
131	0.65	**‘E’: 11**, ‘Q’: 8, ‘K’: 6, ‘-’: 4, ‘D’: 2, ‘F’: 2, ‘S’: 1, ‘I’: 1,‘C’: 1
132	0.52	**‘I’: 13**, ‘V’: 11, ‘C’: 6, ‘-’: 5, ‘Y’: 1
133	0.77	‘I’: 7, ‘Y’: 6, ‘E’: 5, **‘T’: 4**, ‘-’: 4, ‘K’: 3, ‘V’: 2, ‘P’: 2,‘ D’: 1, ‘W’: 1, ‘R’: 1
134	0.79	‘S’: 8, ‘-’: 6, ‘R’: 5, **‘E’: 4**, ‘K’: 3, ‘P’: 3, ‘G’: 3, ‘Q’: 1,‘A’: 1, ‘N’: 1, ‘T’: 1
135	0.38	**‘G’: 26**, ‘-’: 6, ‘E’: 1, ‘N’: 1, ‘H’: 1, ‘Y’: 1
136	0.60	**‘I’: 8**, ‘L’: 8, ‘V’: 7, ‘Y’: 7, ‘-’: 6
137	0.57	**‘L’: 18**, ‘-’: 5, ‘M’: 4, ‘R’: 4, ‘Q’: 2, ‘P’: 1, ‘F’: 1, ‘A’: 1
138	0.61	**‘A’: 17**, ‘-’: 5, ‘S’: 4, ‘R’: 3, ‘E’: 3, ‘Q’: 1, ‘K’: 1, ‘L’: 1,‘N’: 1
139	0.61	‘E’: 16, ‘A’: 5, ‘-’: 5, **‘D’: 4**, ‘S’: 2, ‘K’: 2, ‘R’: 1, ‘N’: 1
140	0.48	**‘E’: 18**, ‘D’: 8, ‘K’: 4, ‘-’: 3, ‘H’: 2, ‘A’: 1
141	0.42	**‘C’: 17**, ‘A’: 8, ‘S’: 8, ‘-’: 3
142	0.65	**‘A’: 11**, ‘S’: 7, ‘V’: 6, ‘R’: 3, ‘-’: 3, ‘G’: 2, ‘I’: 2, ‘K’: 1, ‘Q’: 1
143	0.77	‘E’: 8, ‘N’: 6, ‘K’: 4, ‘T’: 3, ‘L’: 3, ‘-’: 3, ‘Q’: 2, ‘D’: 2, ‘F’: 2, **‘A’: 1**, ‘G’: 1, ‘S’: 1
144	0.49	**‘L’: 16**, ‘I’: 10, ‘M’: 4, ‘-’: 3, ‘Y’: 2, ‘K’: 1
145	0.46	**‘L’: 17**, ‘M’: 11, ‘-’: 3, ‘N’: 2, ‘V’: 2, ‘F’: 1
146	0.59	‘K’: 13, **‘S’: 8**, ‘Q’: 6, ‘I’: 3, ‘-’: 3, ‘T’: 1, ‘R’: 1, ‘A’: 1
147	0.68	‘T’: 13, **‘D’: 5**, ‘E’: 5, ‘G’: 3, ‘-’: 3, ‘N’: 2, ‘S’: 1, ‘A’: 1, ‘Q’: 1, ‘C’: 1, ‘K’: 1
148	0.11	**‘F’: 33**, ‘-’: 3
149	0.37	**‘F’: 23**, ‘Y’: 7, ‘-’: 5, ‘I’: 1
150	0.49	‘K’: 15, **‘R’: 11**, ‘-’: 6, ‘Q’: 4
151	0.83	‘-’: 14, ‘Q’: 6, ‘K’: 5, **‘M’: 4**, ‘R’: 4, ‘E’: 2, ‘N’: 1
152	0.84	‘-’: 14, **‘R’: 5**, ‘K’: 5, ‘G’: 5, ‘I’: 4, ‘L’: 1, ‘M’: 1, ‘E’: 1
153	0.60	**‘R’: 19**, ‘-’: 15, ‘N’: 1, ‘P’: 1
154	0.92	‘-’: 19, ‘E’: 6, **‘Q’: 4**, ‘K’: 2, ‘S’: 2, ‘N’: 1, ‘T’: 1, ‘P’: 1
155	0.92	‘-’: 19, ‘R’: 5, **‘E’: 4**, ‘N’: 4, ‘T’: 1, ‘V’: 1, ‘A’: 1, ‘D’: 1
156	0.86	‘-’: 21, **‘I’: 7**, ‘K’: 6, ‘L’: 1, ‘A’: 1
157	0.88	‘-’: 24, **‘K’: 9**, ‘R’: 1, ‘S’: 1, ‘P’: 1
158	0.93	‘-’: 25, ‘I’: 5, **‘A’: 4**, ‘S’: 1, ‘K’: 1
159	1.00	‘-’: 25, ‘A’: 4, ‘L’: 2, **‘Q’: 1**, ‘E’: 1, ‘N’: 1, ‘H’: 1, ‘P’: 1
160	0.85	‘-’: 25, **‘K’: 10**, ‘R’: 1
161	0.98	‘-’: 31, **‘K’: 4**, ‘V’: 1
162	1.00	‘-’: 31, **‘A’: 3**, ‘S’: 1, ‘R’: 1
163	1.01	‘-’: 31, ‘D’: 2, **‘Q’: 1**, ‘T’: 1, ‘K’: 1
164	1.01	‘-’: 31, ‘R’: 2, **‘S’: 1**, ‘N’: 1, ‘D’: 1
165	1.00	‘-’: 32, **‘S’: 2**, ‘A’: 2
166	1.01	‘-’: 34, **‘T’: 1**, ‘I’: 1
167	1.01	‘-’: 34, **‘D’: 1**, ‘N’: 1

Table S6: Dynamics of the activated and bridging water molecules for various TadA* and TadA*–RNA systems.

System	Activated water		Bridging water	
	Max. Persistence (ns)	Unique waters	Max. Persistence (ns)	Unique waters
TadA*0.1	347.7	3	525.9	2
TadA*1.1	778.5	2	887.6	1
TadA*0.1(L84F)	375.4	10	162.0	8
TadA*1.1(L84F)	291.9	6	472.9	4
TadA*0.1–RNA	1000	1	1000	1
TadA*1.1–RNA	1000	1	1000	1
TadA*0.1(L84F)–RNA	642.2	3	667.6	5
TadA*1.1(L84F)–RNA	1000	1	1000	1

Table S7: Distances describing the conformational changes in the active site of Tad*–RNA complexes during the transition state formation. The $Zn^{+}2$-Target A distances are measured from the activated water to the C6 atom of target A base and the Res84-target A distances were measured by considering only the side change of residue 84 and the nucleobase of target A. The values here are the averages and S.D. corresponding to these distances in the umbrella sampling window with ξ=-0.5Å.

TadA*-RNA complex	TadA*0.1	TadA*1.1	TadA*0.1(L84F)	TadA*1.1(L84F)
Zn^{2+}-Target A distance (Å)	4.25±0.64	3.34±0.32	6.65±0.61	3.60±0.31
Res84 -Target A distance (Å)	6.34±0.28	5.19±0.21	8.07±0.33	5.71±0.19

5 Supplementary Sequences

TADA_ECOLI

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVM
CAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

TADA_SALTY

MSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVM
CAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRA--

TADA_SALTI

MSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVM
CAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRA--

TADA_HAEIN

--------EKMMRYALELADKAEALGEIPVGAVLVDDARIIGEGWNLSIVQSDPTAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTM
CAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEVTSGVLAEECSQKLSTFFQKRREEKKIEK-------

TADA_BUCBP ------SDKYFMKCAIFLAKISEMIGEVPVGAVLVFNNTIIGKGLNSSILNHDPTAHAEIKALRNGAKFLKNY
RLLHTTLYVTLEPCIMCYGAIIHSRISRLVFG---AKYKNLQKYICCKNHFFINKKISITQEVLESECSNLLSSFFKRKRK--------
-----

TADA_BUCAI

----------WMKIALKYAYYAKEKGEIPIGAILVFKERIIGIGWNSSISKNDPTAHAEIIALRGAGKKIKNYRLLNTTLYVTLQPCIM
CCGAIIQSRIKRLVFGANCNSSDHRFSLKNLFCDPQKDYKLDIKKNVMQRECSDILINFFQKKRKN------------

TADA_BUCAP

---------YWMKIALKYAYYAEENGEVPIGAILVFQEKIIGTGWNSVISQNDSTAHAEIIALREAGRNIKNYRLVNTTLYVTLQPCMM
CCGAIINSRIKRLVFGASYKDLKKNPFLKKIFINLEKN-KLKIKKHIMRNECAKILSNFFKNKR--------------

TADA_STAAM

---------YFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVM
CAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFK----NLRANKKSTN---

TADA_BACSU

--------ELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPM
CAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFR-----------------

TADA_STRPQ

---------YFMQEALKEAEKSLQKAEIPIGCVIVKDGEIIGRGHNAREESNQAIMHAEMMAINEANAHEGNWRLLDTTLFVTIEPCVM
CSGAIGLARIPHVIYGASNQKFGGADSLYQILTDERLNHRVQVERGLLAADCANIMQTFFRQGRERKKIAK-------

TADA_STRP6

---------YFMQEALKEAEKSLQKAEIPIGCVIVKDGEIIGRGHNAREESNQAIMHAEMMAINEANAHEGNWRLLDTTLFVTIEPCVM
CSGAIGLARIPHVIYGASNQKFGGADSLYQILTDERLNHRVQVERGLLAADCANIMQTFFRQGRERKKIAK-------

TADA_STRP3

---------YFMQEALKEAEKSLQKAEIPIGCVIVKDGEIIGRGHNAREESNQAIMHAEMMAINEANAHEGNWRLLDTTLFVTIEPCVM
CSGAIGLARIPHVIYGASNQKFGGADSLYQILTDERLNHRVQVERGLLAADCANIMQTFFRQGRERKKIAK-------

TADA_STRP8

---------YFMQEALKEAEKSLQKAEIPIGCVIVKDGEIIGRGHNAREESNQAIMHAEMMAINEANAHEGNWRLLDTTLFVTIEPCVM
CSGAIGLARIPHVIYGASNQKFGGADSLYQILTDERLNHRVQVERGLLAADCANIMQTFFRQGRERKK----------

TADA_STRP1

---------YFMQEALKEAEKSLQKAEIPIGCVIVKDGEIIGRGHNAREESNQAIMHAEMMAINEANAHEGNWRLLDTTLFVTIEPCVM
CSGAIGLARIPHVIYGASNQKFGGVDSLYQILTDERLNHRVQVERGLLAADCANIMQTFFRQGRERKKIAK-------

TADA_AQUAE

--------EYFLKVALREAKRAFEKGEVPVGAIIVKEGEIISKAHNSVEELKDPTAHAEMLAIKEACRRLNTKYLEGCELYVTLEPCIM
CSYALVLSRIEKVIFSALDKKHGGVVSVFNILDEPTLNHRVK-WEYYPLEEASELLSEFFKKLRNNI-----------

TADA_RICBR

-----------MREALKQAEIAFSKNEVPVGAVIVENQKIISKSYNNTEEKNNALYHAEIIAINEACRIISSKNLSDYDIYVTLEPCAM
CAAAIAHSRLKRLFYGASDSKHGAVESNLRYFNSKACFHRPEIYSGIFAEDSALLMKGFFKKIR--------------

TADA_RICTY

-----------MEQALKQARLAFDKNEVPVGVVIVYNQKIIVSSHNNIEEKNNALCHAEIIAINEACNLISSKNLNDYDIYVTLEPCAM
CASAISHSRLKRLFYGASDSKQGAVESNLRYFNSSACFHRPEIYSGILSEHSRFLMKEFFQKMRSTI-----------

TADA_RICCN

-----------MEQALKQAKIAFDKNEVPVGAVVVDHQKIIASTHNNTEEKNNALYHAEIIAINEACNLISSKNLNDYDIYVTLEPCAM
CAAAIAHSRLKRLFYGASDSKHGVVESNLRYFNSSACFHRPEIYSGILAEDSGLLMKEFFKRIRTVISSHR-------

TADA_AGRFC

---------HFMELALVEARSAGERDEVPIGAVLVLDGRVIARSGNRTRELNDVTAHAEIAVIRMACEALGQERLPGADLYVTLEPCTM
CAAAISFARIRRLYYGAQDPKGGAVESGVRFFSQPTCHHAPDVYSGLAESESAEILRQFFREKR--------------

TADA_RICPR

-----------MEQALKQARLAFDKNEVPVGVVIVCNQKIIVSSHNNIEEKKNPLCHAEIIAINTACNLISSKNLNDYDIYVTLEPCAM
CASAISHSRLKRLFYGASDSKHGAVESNLRYFNSNSCFYRPEIYSGILSEHSRFLMQEFFQRIRSAI-----------

TADA_RICFE

-----------MEQALKQAGIAFDKNEVPVGAVIVDNQKIIVSSHNNTEEKNNALYHAEIIAINEACNLISSKNLNDYDIYVTLEPCAM
CAAAIAHSRLKRLFYGASDSKHGAVESNLRYFNSSVCFYRPEIYSGILAEDSRLLMKEFFKRIR--------------

ADAT2_BOVIN

--------EKWMEQAMQMAKDALDNTEVPVGCLMVYNNEVVGKGRNEVNQTKNATRHAEMVAIRRGRSPSEVFE--HTVLYVTVEPCIM
CAAALRLMRIPLVVYGCQNERFGGCGSVLDIASAPSTGKPFQCTPGYRAEEAVEMLKTFYK-----------------

ADAT2_HUMAN

--------EKWMEEAMHMAKEALENTEVPVGCLMVYNNEVVGKGRNEVNQTKNATRHAEMVAIRQSGKSPSEV-FEHTVLYVTVEPCIM
CAAALRLMKIPLVVYGCQNERFGGCGSVLNIASAPNTGRPFQCIPGYRAEEAVEMLKTFYK-----------------

ADAT2_DANRE

--------QTWMAKAFDMAVEALENGEVPVGCLMVYNNEIIGKGRNEVNETKNATRHAEMVALDQ---VLDWCRLRETVLYVTVEPCIM
CAAALRLLRIPFVVYGCKNERFGGCGSVLDVSHLPHTGTSFKCIAGYRAEEAVEMLKTFYKQENPNAPKPKVRKDSIN

ADAT2_MOUSE

--------EKWMEEAMRMAKEALENIEVPVGCLMVYNNEVVGKGRNEVNQTKNATRHAEMVAIDQVLDWCHQHGQSPSTLYVTVEPCIM
CAAALRLMKIPLVVYGCQNERFGGCGSVLNIASAPNTGRPFQCIPGYRAEEAVELLKTFYK-----------------

ADAT2_XENLA

----------WMHKAFQMAQDALNNGEVPVGCLMVYGNQVVGKGRNEVNETKNATQHAEMVAIDQDWCEMNSKKSTDVVLYVTVEPCIM
CAGALRLLKIPLVVYGCRNERFGGCGSVLNVSGDPDTGTKFKCIGGYQAEKAIELLKTFYK-----------------

ADAT2_XENTR

----------WMHKAFQMAQDALNNGEVPVGCLMVYDNQVVGKGRNEVNETKNATRHAEMVAIDQVDWCEKNSKFENIVLYVTVEPCIM
CAGALRLLKIPLVVYGCRNERFGGCGSVLNVANIPDTGTEFKYIGGYQAEKAVELLKTFYK-----------------

GUAD_BACSU

------NHETFLKRAVTLACEGVNAGGGPFGAVIVKDGAIIAEGQNNVTTSNDPTAHAEVTAIRKACKVLGAYQLDDCILYTSCEPCPM
CLGAIYWARPKAVFYAAEHTDAAEAG----------

TAD2_ARATH

---CEDSHNY-MGFALHQAKLALEALEVPVGCVFLEDGKVIASGRNRTNETRNATRHAEMEAIDQDGLSPSQVKFSKCVLYVTCEPCIM
CASALSFLGIKEVYYGCPNDKFGGCGSILSL--HLGSEEAYKCRGGIMAEEAVSLFKCFY----------

FCA1_CANAX

-----FDDKKGLQVALDQAKKSYSEGGIPIGSCIISDDTVLGQGHNERIQKHSAILHGEMSALENAGLPGKTYK--DCTIYTTLSPCSM
CTGAILLYGFKRVVMGENVNFLGNEKLLIE--------NGVEVVN--LNDECIDLMAKFIKEKPQD----------

RIBD1_BUCAI

---------FYMKRAIELSKLGFTTAPNPVGCVIVKNNIIVGEGWHEQAGKN----HAEINA-----LIMAGEKAQGGTAYVTLEPCPP
CCNALIKSGINRVVISNIDPNPKISGNGILYLKK----HGICVKTGLLSKESKQYNKGFFK----------

FCYS_SCHPO

MSSTELSEKAYLREAIKVSQQARDEGQHPFGCIIVENDNVIMSAGNR-VPDGDVTQHAETRAV---GLITKTRRLEKCTLYTSTEPCAM
CSGAIFWSGIRRMIFG----------

RIBD1_BUCAP

---------FYMTRAIKLSKLGFTTSPNPVGCVIVQNKKIVGEGWHKKYGEN----HAEINALNMAG-----EKAKGSTAYITLEPCPP
CCDAIIQSGIKNVIISSLDPNPKVSGKGVLYLRKKGISVKI----GLMSKESQKYNKGFFRMR----------

DCTD_BPMD2

--------EYFLGIATAAAQRSDCERS-KVGAVVVKDRRVRGTGYNAPAGAAGCSTHAEANAL----LYCDREDLIGATLYVTREPCYA
CSNLIAASGIERVVY----------

DCTD_HUMAN

--------EYFMAVAFLSAQRSKDPNS-QVGACIVNSNKIVGIGYNMPNGCSDDVCHAELNAI----MNKNSTDVKGCSMYVALFPCNE
CAKLIIQAGIKEVIFMSHDSDEATAARLL--FNMAGVTFRKFIP------KCSKIVIDF----------

DCTD_PONAB

--------EYFMAVAFLSAQRSKDPNS-QVGACIVNSNKIVGIGYNMPNGCSDDVCHAELNAI----MNKNSTDVKGCSMYVALFPCNE
CAKLIIQAGIKEVIFMSHDSDEATAARLL--FDMAGVTFRKFIP------KCSKIVIDF----------

---

red: TadA; purple: mutations of interest; blue: Cas9

ABE0.1 monomer

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL
PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSI
DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG

KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFEPKKK
RKV

ABE0.1 dimer

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADE
CAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GDSGGSKRTADGSEFEPKKKRKV

ABE1.1 monomer

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL
PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSI
DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFEPKKK
RKV

ABE1.1 dimer

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGR

HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILADE
CAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GDSGGSKRTADGSEFEPKKKRKV

ABE0.1(L84F) monomer

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL
PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSI
DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFEPKKK
RKV

ABE0.1(L84F) dimer

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADE
CAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV

VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GDSGGSKRTADGSEFEPKKKRKV

ABE1.1(L84F) monomer

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL
PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSI
DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFEPKKK
RKV

ABE1.1(L84F) dimer

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGARNAKTGAAGSLMDVLHHPGMNHRVEITEGILADE
CAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GDSGGSKRTADGSEFEPKKKRKV

ABE7.10

MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQ
NYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQK
KAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK
AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GDSGGSKRTADGSEFEPKKKRKV

6 Supplementary References

(1) Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T. L. *Nucleic Acids Res.* **2008**, *36*, W5–W9.
(2) Bairoch, A.; Apweiler, R. *Nucleic Acids Res.* **2000**, *28*, 45–48.
(3) Huson, D. H.; Scornavacca, C. *Syst. Biol.* **2012**, *61*, 1061–1067.
(4) Valdar, W. *Proteins Struct. Funct. Bioinf.* **2002**, *48*, 227.
(5) Šali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779–815.
(6) Kim, J.; Malashkevich, V.; Roday, S.; Lisbin, M.; Schramm, V. L.; Almo, S. C. *Biochemistry* **2006**, *45*, 6407–6416.
(7) Losey, H. C.; Ruthenburg, A. J.; Verdine, G. L. *Nat. Struct. Mol. Biol.* **2006**, *13*, 153–159.
(8) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
(9) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. *Nucleic Acids Res.* **2005**, *33*, W368–W371.
(10) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. *Nucleic Acids Res.* **2012**, *40*, W537–W541.
(11) Li, P.; Merz Jr, K. M. MCPB.py: A python based metal center parameter builder, 2016.
(12) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
(13) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
(14) Banás, P.; Hollas, D.; Zgarbová, M.; Jurecka, P.; Orozco, M.; Cheatham III, T. E.; Sponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3836–3849.
(15) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham III, T. E.; Jurecka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.
(16) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(17) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(18) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.
(19) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198–210.
(20) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
(21) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
(22) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M., et al. *University of California, San Francisco.*
(23) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(24) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.
(25) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
(26) Grossfield, A. WHAM: the weighted histogram analysis method.
(27) Zhu, F.; Hummer, G. *J. Comput. Chem.* **2012**, *33*, 453–465.
(28) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
(29) Seabra, G. d. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. *J. Phys. Chem. A* **2007**, *111*, 5655–5664.
(30) Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 29–41.
(31) Roe, D. R.; Cheatham III, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
(32) Roe, D. R.; Cheatham III, T. E. *J. Comput. Chem.* **2018**, *39*, 2110–2117.
(33) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
(34) Virtanen, P. et al. *Nat. Methods* **2020**, *17*, 261–272.
(35) Grünewald, J.; Zhou, R.; Iyer, S.; Lareau, C. A.; Garcia, S. P.; Aryee, M. J.; Joung, J. K. *Nat. Biotechnol.* **2019**, *37*, 1041–1048.

(36) Gruber, A. R.; Lorenz, R.; Bernhart, S. H.; Neuböck, R.; Hofacker, I. L. *Nucleic Acids Res.* **2008**, *36*, W70–W74.
(37) Gaudelli, N. M.; Komor, A. C.; Rees, H. A.; Packer, M. S.; Badran, A. H.; Bryson, D. I.; Liu, D. R. *Nature* **2017**, *551*, 464–471.
(38) Rees, H. A.; Wilson, C.; Doman, J. L.; Liu, D. R. *Sci. Adv.* **2019**, *5*, eaax5717.
(39) Zhou, C.; Sun, Y.; Yan, R.; Liu, Y.; Zuo, E.; Gu, C.; Han, L.; Wei, Y.; Hu, X.; Zeng, R., et al. *Nature* **2019**, *571*, 275–278.
(40) Gaudelli, N. M.; Lam, D. K.; Rees, H. A.; Solá-Esteves, N. M.; Barrera, L. A.; Born, D. A.; Edwards, A.; Gehrke, J. M.; Lee, S.-J.; Liquori, A. J., et al. *Nat. Biotechnol.* **2020**, 1–9.
(41) Richter, M. F.; Zhao, K. T.; Eton, E.; Lapinaite, A.; Newby, G. A.; Thuronyi, B. W.; Wilson, C.; Koblan, L. W.; Zeng, J.; Bauer, D. E., et al. *Nat. Biotechnol.* **2020**, 1–9.