ABSTRACT: Employing DNA as a high-density data storage medium has paved the way for next-generation digital storage and biosensing technologies. However, the multipart architecture of current DNA-based recording techniques renders them inherently slow and incapable of recording fluctuating signals with subhour frequencies. To address this limitation, we developed a simplified system employing a single enzyme, terminal deoxynucleotidyl transferase (TdT), to transduce environmental signals into DNA. TdT adds nucleotides to the 3'-ends of single-stranded DNA (ssDNA) in a template-independent manner, selecting bases according to inherent preferences and environmental conditions. By characterizing TdT nucleotide selectivity under different conditions, we show that TdT can encode various physiologically relevant signals such as Co²⁺, Ca²⁺, and Zn²⁺ concentrations and temperature changes in vitro. Further, by considering the average rate of nucleotide incorporation, we show that the resulting ssDNA functions as a molecular ticker tape. With this method we accurately encode a temporal record of fluctuations in Co²⁺ concentration to within 1 min over a 60 min period. Finally, we engineer TdT to allosterically turn off in the presence of a physiologically relevant concentration of calcium. We use this engineered TdT in concert with a reference TdT to develop a two-polymerase system capable of recording a single-step change in the Ca²⁺ signal to within 1 min over a 60 min period. This work expands the repertoire of DNA-based recording techniques by developing a novel DNA synthesis-based system that can record temporal environmental signals into DNA with a resolution of minutes.

INTRODUCTION

DNA is an attractive medium for both long-term data storage and for in vitro recording of molecular events due to its high information density and long-term stability. Molecular recording strategies write information into DNA by altering existing DNA sequences or adding new sequences. For example, systems have been developed that use methods including differential CRISPR spacer acquisition, enzymatic synthesis, and others. By connecting these DNA modifications to a user input (in the case of data storage) or environmental signal of interest (in the case of recording events), these strategies enable post hoc recovery of signal dynamics over time by DNA sequencing. To date, molecular recording systems, both in vitro and in vivo, have connected signals of interest to DNA recordings with transcriptional control, using signal-responsive promoters to drive the expression of molecular writers, such as base-editors, CRISPR-associated systems, and gene-circuits, to record changes in signal. These approaches have yielded accurate recordings; however, the time required to transduce signals through a recording apparatus that includes transcription, translation, and DNA modification fundamentally constrains the application of these methods to events on the time scales of hours or days. A recording mechanism that relies only on post-translational elements would be inherently faster, as signal transduction would only require subsecond conformational shifts in one enzyme.

In an effort to speed up DNA recording processes, we hypothesized that a DNA polymerase (DNAp), which continually incorporates bases, could serve as a candidate for post-translational molecular encoding. In such a system, a DNAp functions as a “ticker-tape” recorder, transforming changes in environmental signals into changes in the composition of the DNA it synthesizes (Figure 1A). Much faster than transcription and translation, nucleotide incorporo-
ration occurs on a time scale on the order of milliseconds to seconds,15 potentially enabling orders of magnitude improvements in the temporal accuracy and resolution of molecular recording. However, prototypical DNAPs replicate the contents of an existing strand, which would prevent recording of new information. A DNAP that does not simply replicate DNA but rather creates a de novo sequence could allow for DNA recording.

Terminal deoxynucleotidyl transferase (TdT) is a DNAP that can randomly incorporate bases to the 3′-position of a DNA strand with biases toward particular bases.16,17 Shifting the nucleotide bias of TdT could make it a prime candidate for post-translational control of DNA encoding. In fact, in vitro experiments have shown that cations (e.g., Co2+) can shift the bias of TdT.16,18 In addition, DNA is synthesized in a sequential manner, which provides an estimate of the time a particular base is added. We therefore hypothesized that the environment in which a TdT extends a DNA strand might be encoded by the average base composition of the extended DNA. Put another way, by a combination of the change in nucleotide bias of TdT in the presence of various environmental signals tested. Signal 0 was 10 mM Mg2+ at 37 °C for 1 h. Signal 1 was (i) 10 mM Mg2+ + 0.25 mM Co2+ at 37 °C for 1 h, A incorporation increased by 12.4%, while G decreased by 8.0%, and T and C decreased by 1.5% and 2.9%, respectively; (ii) 10 mM Mg2+ + 1 mM Ca2+ at 37 °C for 1 h, A increased by 1.5%, G decreased by 1.2%, T increased by 0.4%, and C decreased by 0.8%; (iii) 10 mM Mg + 20 μM Zn2+ at 37 °C for 1 h, A increased by 14.9%, G decreased by 8.8%, T decreased by 3.3%, and C decreased by 2.8%, and (iv) 10 mM Mg2+ at 20 °C for 1 h, 0.4% increase in A, 3.8% decrease in G, 1.5% increase in T, and 1.8% increase in incorporation of C. Error bars show two standard deviations of the mean. The statistical significance was assessed after first transforming the data into Aitchison space, which makes each dNTP frequency change statistically independent of the others (Figure S2).

**RESULTS**

TdT Can Encode Environmental Signals In Vitro via Changes in Base Selectivity. The cations present in the reaction environment of TdT affect the rate of incorporation for specific nucleotides.16 For example, previous studies18–20 and our experiments show that, when only one nucleotide is present, the incorporation rates of pyrimidines, dCTP and dTTP, increase in the presence of Co2+ (Figure S1).

We sought to examine if these Co2+-dependent changes in kinetics also occurred in the presence of all four nucleotides, dATP, dCTP, dGTP, and dTTP (hereafter referred to as A, C, G, and T). The nucleotide composition of ssDNA extended by bovine TdT in a cobalt-free reaction buffer or with cobalt added was determined by next-generation sequencing. In the presence of Co2+, A incorporation increased, while G, T, and C incorporation decreased (Figure 1B-i and Figure S2). Notably, the significant difference in the composition of DNA under each condition effectively encodes information about the environmental Co2+ concentration at the time of DNA synthesis.

Next, we were interested in understanding which conditions could be encoded by TURTLES. We examined Ca2+, Zn2+, and temperature. Ca2+ signaling is biologically ubiquitous and...
functions in neural firing, fertilization, and neurodevelopment, and Zn$^{2+}$ is an important signal in the development and differentiation of cells, and the temperature is relevant in many situations.

Each signal altered both the particular dNTPs affected and the magnitude of the change in dNTP selectivity. We were able to encode 1 mM Ca$^{2+}$, 20 μM Zn$^{2+}$, and a temperature of 20 °C (Figure 1B-ii–iv and Figure S2). Both cation addition and a temperature change also altered the lengths of ssDNA strands synthesized (Figures S3–S8). For each environmental condition tested, we observed significant differences in the composition of TdT-synthesized DNA. We conclude that input-dependent changes in TdT nucleotide selectivity can encode environmental information into DNA. For further analysis we chose to focus on Co$^{2+}$ as the candidate cationic signal due to the large difference in TdT selectivity.

**Recording a Single Step Change in Co$^{2+}$ Concentration with Resolution of Minutes.** Having shown nucleotide selectivity changes in the presence of Co$^{2+}$, we attempted to identify the time at which Co$^{2+}$ was added to a TdT-catalyzed ssDNA synthesis reaction on the basis of the composition of the nucleotides incorporated (Figure 2A). During a 60 min extension reaction, we created step transitions in cobalt concentration by adding 0.25 mM Co$^{2+}$ at 10, 20, and 45 min (hereafter referred to as 0 → 1 input, where 0 is Co$^{2+}$ free and 1 is with 0.25 mM added Co$^{2+}$) (Figure 2B, top). For each reaction, we analyzed approximately 500000 DNA strands by deep sequencing and calculated the dNTP incorporation frequencies over all reads. Because the change in dNTP selectivity is a compositional data type (i.e., all changes in base frequency sum to 0%), they are not independent and do not satisfy the independence assumption required for most statistical tests. Therefore, to perform hypothesis testing, the base composition was transformed into Aitchison space, where the proportion of each base becomes independent of the other three. The output signal for each reaction was calculated as the normalized distance in Aitchison space between the 0 and 1 controls.

After each sequence was normalized by its own length, the Aitchison distance between the 0 and 1 controls was calculated. We were able to accurately estimate the times of Co$^{2+}$ addition (0 → 1) and removal (1 → 0), simultaneously synthesizing >500000 strands of DNA is infeasible for certain applications. To determine the number of strands needed for a reasonable statistical certainty, we randomly sampled smaller groups of strands from the experiment and evaluated our ability to predict when Co$^{2+}$ was added (Figure 2C, top and bottom). With about 1000 strands, we still estimated the time of Co$^{2+}$ additions to within 2 min of actual input times (Figure 2C). Thus, TURTLES can be employed for recording temporal information even with a limited number of ssDNA substrates.

**Recording Multiple Fluctuations in Co$^{2+}$ Concentration onto DNA with a Resolution of Minutes.** In contrast to current DNA-based recorders, which rely on time-integrated recording methods (i.e., accumulation of mutations) or slow signal transducing steps, DNA synthesis-based
approaches can record the dynamics of multiple fluctuations in real time. While accumulation can tell what fraction of the time a signal was present in a time period, the ability to record multiple temporal changes would enable new levels of insight into dynamic processes such as physiological signaling, which are poorly captured by time-integrated recording methods.

We used TURTLES to record a 0 → 1 input cobalt signal. The 0 condition was maintained for the first 20 min, 1 for the next 20 min, and 0 for the last 20 min of the extension reaction (Figure 2D, center). To account for the additional complexity of multiple fluctuations, we used an algorithm previously developed by Glaser et al. (Materials and Methods for details) to binarize the value of the output signal every 0.1 min. We were able to accurately reconstruct the input 0 → 1 → 0 signal, estimating transitions between the 0 and 1 signals occurring at 21 and 41 min on the basis of sequencing data (Figure 2D, bottom).

On the basis of the measured experimental parameters, we used in silico simulations to estimate the performance of TURTLES in more complex recording environments. We investigated how rapidly signals could change and still be detected and how many consecutive condition changes could be recorded accurately. By varying the length of time of each input condition (0 or 1) from 1 to 20 min (Figure S12A), we estimated that TURTLES can record 6 consecutive signal changes with >80% accuracy (Figure S12D). We thus show that TURTLES has the potential for high temporal precision and can decode signals across a range of frequencies.

**TdT Can Be Engineered to Allosterically Respond to and Encode Environmental Signals.** Unlike Co²⁺ and Zn²⁺, we observed that Ca²⁺ only modestly altered the dNTP selectivity of TdT, precluding temporal recordings of Ca²⁺ concentration. To show that TURTLES could be expanded to signals to which TdT was unresponsive or weakly responsive, we attempted to engineer a TdT to allosterically respond to Ca²⁺. The structural determinants of base selectivity in TdT are poorly understood, which ruled out directly increasing the dynamic range of Ca²⁺-responsive dNTP selectivity changes. Accordingly, we conceived a modular recording system based on two distinct TdT species with different inherent dNTP selectivity.

The two-TdT system, TURTLES-2, uses a reference TdT whose catalytic rate is unaffected by inputs and a sensor TdT that is allosterically activated or deactivated in response to input signals. By the choice of a pair of sensor and reference TdTs with distinct nucleotide selectivity, TURTLES-2 encodes environmental signals into changes in DNA composition on the basis of the differential activity of the two TdTs (Figure 3A). As the sensing and recording functions of the system are distributed between two TdT variants, TURTLES-2 is more accessible to tuning and engineering efforts.

We employed the natural calcium sensing protein calmodulin (CaM) and the cognate binding peptide M13 to generate a TdT with allosterically modulated activity. The calcium-dependent interaction between CaM and M13 has been previously utilized in the engineering of allosteric calcium biosensors and as a platform for generalizable ligand biosensors. The structure, function, and dynamics of these molecules are well understood, making them ideal for engineering applications. In the presence of Ca²⁺ only, the engineered TdT synthesizes DNA, while in the absence of Ca²⁺ both the engineered TdT and reference TdT are carrying out DNA synthesis. In the presence of Ca²⁺ only the reference TdT synthesizes DNA. Using the same methods as for the single-step transition, we calculated the output signal (Figure 2D, top). Using a modified version of the algorithm detailed in Glaser et al., the signal was deconvoluted into a binary response, with the predicted switch time of 30 min. Using a modified version of the algorithm detailed in Glaser et al., the signal was deconvoluted into a binary response, with the predicted switch time of 22 min.

**Figure 3.** Recording Ca²⁺ changes into ssDNA with TURTLES-2. (A.i) Representative calcium transition changing concentration from calcium-free to calcium-added conditions during a TdT-based DNA synthesis reaction. Mg²⁺ concentration and reaction temperature are held constant.

(A.ii) The CaM subunit (orange) of engineered TdT (teal) binds with the fused M13 peptide in the presence of Ca²⁺ to allosterically turn off DNA synthesis. The activity of the reference TdT (gray) is not affected by Ca²⁺. In the absence of Ca²⁺ both the engineered TdT and reference TdT are carrying out DNA synthesis. In the presence of Ca²⁺ only the reference TdT synthesizes DNA. (A.iii) This results in a change in the overall nucleotide incorporation preference upon a change in Ca²⁺. (B) Top: 100 μM CaCl₂ was added to the extension reaction at 30 min to generate a 0 → 1 transition. Center: Mean output signal across three biological replicates. Bottom: Using a modified version of the algorithm detailed in Glaser et al., the signal was deconvoluted into a binary response, with the predicted switch time of 30 min. (C) Top: 50 μM EGTA was added to the extension reaction at 30 min to generate a 1 → 0 transition. Center: Mean output signal across three biological replicates. Bottom: Using a modified version of the algorithm detailed in Glaser et al., the signal was deconvoluted into a binary response, with the predicted switch time of 22 min.
Here, we generated variants with M13 fused to one of four sites in mTdT that were predicted to minimize the structural disruption of inserting the M13 sequence using SCHEMA-RASPP	extsuperscript{31} (Figure S13 and Table S2). After an initial activity screening (Figure S14) of the variants, we observed that one variant, mTdT(M13-388), retained polymerase activity. CaM was subsequently fused to the N-terminus of mTdT(M13-388) via a linker. Primer extension reactions showed that the resulting CaM-mTdT(M13-388) variant was active under calcium-free conditions and inactive under calcium-added conditions (Figure S15). To confirm that the calcium-dependent interaction between CaM and M13 was responsible for the observed activity modulation, we mutated four essential Ca\textsuperscript{2+}-binding residues in CaM, which ablated the calcium sensitivity of CaM-mTdT(M13-388) (Figure S15).

Depending on the application, sensor polymerases with different calcium affinities may be useful to selectively record Ca\textsuperscript{2+} fluctuations exceeding threshold concentrations. We anticipated that the modular design of CaM-mTdT(M13-388) would allow the properties of the fusion to be rationally modified with CaM variants with known differences in Ca\textsuperscript{2+} affinity. The polymerase activity was determined by the length distributions of primer extensions, CaM-mTdT(M13-388) variants containing the CaM mutants D96V, D130G, and D142L, which reduce the calcium affinity of CaM.

We observed that all variants exhibited greater activity in comparison to the unmodified CaM-mTdT(m13-388) in the presence of low concentrations of calcium (Figure S16). Moreover, the increase in activity correlated with the reported effective Ca\textsuperscript{2+} \( K_D \) value of the variants, demonstrating that the calcium sensitivity of CaM-mTdT(M13-388) can be rationally tuned.

Next, we tested if the TURTLES-2 system could encode the Ca\textsuperscript{2+} state into DNA. CaM-mTdT(M13-388) was purified, and an NGS analysis of extension reactions performed with the polymerase confirmed the calcium-sensitive phenotype (Figures S17 and S18). We characterized CaM-mTdT(M13-388) in the context of the TURTLES-2 recording system by performing extensions with a mixture of purified bovine TdT and CaM-mTdT(M13-388) under calcium-free and calcium-added conditions. The two-polymerase system exhibited a significantly altered nucleotide selectivity under the calcium-added conditions (Figure S19). As expected, under the calcium-free conditions the overall base incorporation preference was approximately the average of the observed preferences of bovine TdT and CaM-mTdT(M13-388), whereas the overall base preference under the calcium-added conditions was nearly identical with that of bovine TdT (Figure S20). We conclude that the differential overall base selectivity of the TURTLES-2 system is capable of encoding the environmental calcium state into DNA.

**Recording a Single-Step Change in Ca\textsuperscript{2+} Concentration with a Resolution of Minutes with Two TdT Systems.** We next investigated if the differential calcium response of TURTLES-2 could be used to infer the time at which calcium concentrations changed in an extension reaction. During a 60 min extension we tested both 0 → 1 and 1 → 0 step transitions at 30 min, where 0 is calcium-free and 1 is calcium-added (Figure 3B, top, Figure 3C, top, Supplementary Note 6). Using a variation of the model developed by Glaser et al.,

we inferred transition times of 22 min for the 1 → 0 transition (Figure 3B, bottom) and 30 min for the 0 → 1 transition (Figure 3C bottom). The decreased accuracy of the estimated time for the calcium transitions did not correspond to an increase in measurement variability. We speculate that the offset may be due to the different kinetics of holo calmodulin binding M13 and apo calmodulin releasing M13. Time estimations for transitions in TURTLES-2 would likely improve with more sophisticated decoding algorithms and a deeper characterization of the transition behavior of CaM-mTdT(M13-388). We conclude that TURTLES-2 serves as a promising proof of concept for developing high-resolution temporal calcium signal encoding systems.

### Discussion

While many DNA-based biosensors have been deployed for studying physiological signals of interest, the scalability and spatial resolution of biosensors are intrinsically limiting in some applications.

By leveraging the post hoc recovery of biological data, optimized TURTLES systems may be capable of enabling otherwise inaccessible high-resolution spatial and temporal recordings of physiological signaling molecules that fluctuate on the time scale of \( 10^{-1} \)–\( 10^{-9} \) minutes. Such signals include slow calcium signaling that occurs in neurons\textsuperscript{35,41,58,59} and vertebrate development.

Additional optimization of the TURTLES system will be required to enable the spatiotemporal resolution for characterizing systems with shorter time scales.

Beyond biological applications, there has been a sustained interest in biosensors for testing environmental parameters such as water quality. For longer-term tracking of contaminating metal ions in water, one could use TURTLES to track the cobalt concentration over time.\textsuperscript{40} In concert with microfluidic reaction control, TURTLES can also serve as a competitive platform for enzymatic DNA synthesis for data archiving, which is an appealing alternative to phosphoramidite methods due to the low cost and reduced environmental impact.\textsuperscript{42} Although TURTLES was used in this work to record binary signals, it could also be used to record more than two signals in sequence. For example, in vitro addition of nucleotides via TdT with varied input nucleotide compositions could be used to encode data (Figure S21). While the information density of such a system would be lower than those using base-specific DNA synthesis, it would not require specialized substrates or complex reaction cycling and would thus be a competitive application for digital data storage.

Going forward, more sophisticated computational methods will improve the recording accuracy of TURTLES. In this study we utilized simple, intuitive models of TdT activity to transform sequence data into temporal information. By incorporating kinetic models of TdT activity or machine learning to classify signal changes along individual DNA strands, the accuracy of temporal estimations could likely be increased. These methods would also improve the robustness of TURTLES recordings by reducing the required sequencing depth from thousands to hundreds of reads. In this work, both the inputs and outputs for TURTLES were binarized; however, the underlying principle can be extended to record continuously varying analog signals with improved decoding algorithms.

The quality of TURTLES recordings may also be improved by engineering the properties of TdTs. In particular, the sequencing depth required to accurately decode recordings can be reduced by increasing the magnitude of changes in nucleotide selectivity in response to inputs. Likewise, reference TdTs that have a more distinct nucleotide selectivity from the...
CaM-mTdT(M13-388) sensor TdT can be engineered or identified among natural TdT diversity. Improvements to the temporal resolution of TURTLES systems can be accomplished by enhancing the nucleotide incorporation rate of TdT. In TURTLES-2, the structural optimization of CaM-mTdT-(M13-388) may improve temporal resolution by optimizing the kinetics of the CaM—M13 interaction. Notably, fluorescent biosensors based on CaM—M13 interactions can report calcium spikes on the order of seconds, suggesting that calcium sensing will not be limiting with respect to temporal resolution in an optimized system. The functionality of TURTLES-2 may be further expanded by employing generalizable sensors based on the CaM—M13 interaction or by probing TdT with sensing domains other than calmodulin such that new signals of interest can be encoded or recorded. In all, we have demonstrated a new methodology for recording dynamic, environmental information into DNA that relies only on allosteric regulation, enabling a resolution of minutes.

CONCLUSION

In this study, we demonstrated two DNA synthesis-based recording concepts that encode and record the temporal dynamics of fluctuating environmental signals with an accuracy of minutes. By coupling sensing and writing functions, TURTLES simplifies the recording apparatus to a post-translational system. This gives TURTLES distinct advantages over the temporal characteristics of existing tools, paving the way toward the development of tools with heretofore unprecedented temporal accuracy and resolution. While TURTLES can record several physiologically relevant signals, TURTLES-2 lends tunability to the recording system with simple rational engineering. Given the uncomplicated and genetically encodable design of TURTLES systems, we anticipate that TURTLES can be further developed for both in vitro and in vivo biorecording applications.

MATERIALS AND METHODS

Enzymes and ssDNA Substrate. Terminal deoxynucleotidyl polymerase, T4 RNA ligase 1, and Phusion High-Fidelity PCR Master Mix with HF Buffer were purchased from New England Biolabs (NEB). ssDNA substrates used for extension reactions were purchased from Integrated DNA Technologies (IDT). ssDNA substrates used for extension reactions were purchased from Integrated DNA Technologies (IDT) with standard NEB. Accyummy DNA polymerase (Bioline) using DNA primers P28-P28 given in Table S1. The amplification was then purified using a Cycle Pure Kit (Omega Biotek). The vector backbone fragment was amplified from pET28a-M-CaM-cdTTS plasmid DNA constructed above using pET Ultra II Hotstart PCR Master Mix using DNA primers P29—P36 given in Table S1. The PCR product was then digested with DpnI to remove the DNA template. The approximately 8100 bp ampiclon was purified using a gel extraction kit (Omega Biotek). DNA concentrations of both linear fragments were measured using a Take3 plate coupled with a Biotek Cytation 5 plate reader. The corresponding backbone and M13 fragments were then assembled using the replitQa HiFi Assembly Kit (Quanta bio), transformed into chemically competent DH5α, and selected on LB-Kanamycin (50 μg/mL) plates. Individual colonies were then screened via restriction digestion and verified using Sanger sequencing (CBFG, Miami University) with primers S1—S8 (Table S1).

CaM-TdT(M13-388) Expression and Purification. Purification optimizations determined that N-terminal MBP was unnecessary for expression and purification and was not included in the final expression construct. The expression construct (pET28a-CaM-mTdT(m13-388)) was transformed into chemically competent NEB T7Express cells, plated on kanamycin-selective plates, and incubated at 37 °C. The following day, a single colony was selected and inoculated into 5 mL of kanamycin-supplemented LB. The culture was incubated for 20 h at 37 °C. Four flasks with 120 mL of kanamycin-supplemented LB were inoculated 1/400 (v/v) with the overnight culture. The cultures were incubated with shaking at 250 rpm. Once the OD600 value was between 0.5 and 0.6, the cultures were cooled to room temperature and induced with 1 mM IPTG. Following induction, the cultures were incubated for 18 h at 15 °C. The cells were pelleted at 4 °C, and the supernatant was discarded. The decanted cell pellets were stored at −80 °C.

The cell pellets were thawed on ice. Lysis and affinity chromatography were performed using the Takara Bio HisTALON gravity column purification kit; all steps were performed according to the manufacturer’s native protein extraction protocol. Note that the cell pellets were treated with optional DNase I and lysozyme during lysis. A 1 mL portion of Takara Bio TALON metal affinity resin was used for affinity chromatography. All binding and washing steps were performed on ice with shaking at 250 rpm. Fifteen bed volumes of wash buffer were used for all washes. CaM-mTdT-(M13-388) was eluted from the resin in 10 500 μl fractions. Each fraction was analyzed by SDS-PAGE, and the total protein concentration in each fraction was measured by absorbance at 280 nm. The first five elution fractions, which contained the majority of eluted protein, were pooled.

The pooled fractions were diluted in binding buffer (20 mM Tris-HCl, pH 8.3) and further purified by anion exchange chromatography using a Cytiva HiTrap Q HP 5 mL column and a 40 CV gradient from 0 to 1 mM NaCl in binding buffer with a GE Healthcare AKTAExpress FPLC apparatus. The protein eluted in two fractions. Both elutions were buffer-exchanged by dialysis into a storage buffer consisting of 200 mM KH2PO4 and 100 mM NaCl at pH 6.5 and concentrated using Vivaspin 20 columns to a final concentration of 0.37 mg/mL for the first fraction and 0.98 mg/mL for the second fraction. The fractions were aliquoted and flash frozen on dry ice for storage at −80 °C. Notably, a PAGE analysis showed that the second elution contained a product at 25 kDa in addition to the expected fusion protein at approximately 70 kDa. Both CaM-mTdT(M13-388) elutions recapitulated the calcium-sensitive phenotype and exhibited similar nucleotide selectivities (Figures S18 and S19). As significantly more protein was recovered in the second elution, it was used for all subsequent experiments.

Cell-Free Protein Expression and Primer Extension Assay. For initial activity screening of fusion variants and CaM-mTdT(M13-
characterization, proteins were expressed in cell-free reactions. Variants were expressed using NEB PURExpress in 25 μL reactions containing 40% (v/v) PURExpress Solution A, 30% (v/v) PURExpress Solution B, 1.6 U/μL NEB RnaSe I, 10 ng/μL expression vector DNA, and dh2O to volume. The expression reaction mixtures were incubated for 4 h at 30 °C.

Primers extension reactions were prepared on ice. Primer extensions were performed in 25 μL reactions containing 1X NEB TdT Reaction buffer, 0.8 μM single-stranded, FAM-labeled substrate DNA FAM_NB (Table S2), 1 mM dNTPs, polymerase, and dh2O to 25 μL. For variants expressed in PURExpress, 2.5 μL of the expression reaction was used immediately after expression, 20U (approximately 0.2 μg) of the NEB bovine TdT was used for positive control reactions, and approximately 0.5 μg of purified CaM-tdT (M13-388) was used for activity validation reactions after purification.

For calcium-added conditions, CaCl2 was added to the reaction mixtures to a final concentration of 1 mM. Extension reaction mixtures were incubated for 2 h at 37 °C.

Completed extensions were analyzed by urea-PAGE. A 8 μL portion of each completed primer extension reaction was combined with 12 μL of BioRad 2x TBE urea sample buffer and boiled for 10 min. The boiled samples were loaded onto a 10% polyacrylamide TBE urea gel (BioRad 4566036), and 200V was applied to the gel for 40 min. The gels were imaged on a GE Healthcare Typhoon 9400 laser scanner using a 200 μm pixel size and λem = 488 nm and λex = 520 nm at 40 nm BP40. The imaging gain was adjusted for each experiment to avoid saturation.

**Extension Reaction for Calculating the Effects of Ca2+, Zn2+, and Temperature on the Overall dNTP Preference of TdT.** Each extension reaction consisted of a final concentration of 10 μM ssDNA substrate CS1 (Table S2), 1 mM dNTP mix (each dNTP at 1 mM final concentration), 1.4X NEB TdT reaction buffer, and 10 units of TdT to a final volume of 50 μL. When the effect of cations was tested, CoCl2 was added at a final concentration of 0.25 mM, CaCl2 at 2 mM, or Zn(Ac)2 at 20 μM. It is important to note that reaction initiation was done by adding TdT to the ssDNA substrate mix (ssDNA substrate mix consisted of the ssDNA substrate, dNTPs, and the cation). Prior to reaction initiation, the ssDNA substrate mix and TdT were stored in separate PCR strip tubes at 0 °C (on ice). The reaction was run for 1 h at 37 °C in a Bio-Rad PCR block. When the effect of temperature was tested, the same reaction mix was run on a Bio-Rad PCR block set at the tested temperatures for 1 h. Reactions were stopped by freezing at −20 °C. For initial testing, reactions were analyzed by urea-PAGE; 2 μL of the reaction mixture was mixed with 12 μL of TBE-Urea (Bio-Rad) loading dye and boiled for 10 min at 100 °C. All of the diluted extension reaction mixtures was then loaded onto 30 μL, 10-well 10% TBE-Urea Gel (Bio-Rad) and run for 40 min at 200V. Immediately after the run was over, the gel was stained with Sybr Gold for 15 min and imaged on an ImageQuant BioRad instrument.

**TURTLES 0 → 1 Extension Reactions.** Mg2+ only for 1 h (signal 0) and Mg2+ + Co2+ for 1 h (signal 1) were set up as the regular extension reactions mentioned above. The 0 → 1 reactions where the signal changed from 0 to 1 at various times during the 1 h extension in some replicates (data not included). We presume this is due to some ethanol getting carried forward into the eluted ssDNA. Thus, we extended the dry spin time on the basis of a suggestion from Zymo Research to 4 min. We also utilized two other ways to evaporate any remaining ethanol after the column dry spin step based on the protocol mentioned in Cold Spring Harbor Protocols.4 We either kept the columns open in a biohazard for 15 min to allow for evaporation or after elution of ssDNA we kept the 1.5 mL Eppendorf tubes containing the eluted ssDNA open at 45 °C for 3 min. Both methods gave better ethanol removal than just dry spin, and they were tried in triplicate and averaged and plotted for the time prediction analysis (Figure 2D).

**Illumina Library Preparation and Sequencing.** Our sample preparation pipeline for NGS was adapted from a previous protocol.4 After an extension reaction, 2 μL of the product was utilized for a ligation reaction. A 22bp universal tag, common sequence 2 (CS2) of the Fluidigm Access Array Barcode Library for Illumina Sequencers (Fluidigm), synthesized as ssDNA with a 5′-phosphate modification and PAGE purified (Integrated DNA Technologies), was blunt-end ligated to the 3′-end of extended products using T4 RNA ligase. Ligation reactions were carried out in 20 μL volumes and consisted of 2 μL of the extension reaction mixture, 1 μM of CS1 ssDNA, 1X T4 RNA Ligase Reaction Buffer (NEB), and 10 units of T4 RNA Ligase 1 (NEB). Ligation reaction mixtures were incubated at 25 °C for 16 h.ligated products were stored at −20 °C until PCR, which was carried out on the same day. Ligation products were not stored at −20 °C for more than 24 h. PCR was performed with barcoded primer sets from the Access Array Barcode Library for Illumina Sequencers (Fluidigm) to label extension products from up to 96 individual reactions. Each PCR primer set contained a unique barcode in the reverse primer. From S′-
the forward PCR primer (PE1 CS1) contained a 25-base paired-end Illumina adapter 1 sequence followed by CS1. The binding target of the forward PCR primer was the reverse complement of the CS1 tag that was used as the starting DNA substrate. From 5′-3′ the reverse PCR primer (PE2 BC CS2) consisted of a 24-base paired-end Illumina adapter 2 sequence (PE2), a 10-base Fluidigm barcode (BC), and the reverse complement of CS2. CS2 DNA that had been ligated onto the 3′-end of extended products served as the reverse PCR primer-binding site. Each PCR reaction consisted of 2 μL of ligation product, IX Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB), and 400 nM forward and reverse Fluidigm PCR primers in a 20 μL reaction volume. Products were initially denatured for 30 s at 98 °C, followed by 20 cycles of 10 s at 98 °C (denaturation), 30 s at 60 °C (annealing), and 30 s at 72 °C (extension). Final extensions were performed at 72 °C for 10 min. Amplified products were stored at −20 °C until clean up and pooling. QC for individual sequencing libraries was performed as follows. A 2 μL portion of each library was pooled into a QC pool, and the size and approximate concentration were determined using an Agilent 4200 Tapestation. The pool concentration was further determined using Qubit and qPCR methods. Sequencing was performed on an Illumina MiniSeq Mid Output flow cell, and sequencing was initiated using custom sequencing primers targeting the CS1 and CS2 conserved sites in the library linkers. Additionally, the phiX control library was spiked into the run at 15–20% to increase the diversity of the library clustering across the flow cell. After demultiplexing, the percent seen for each sample was used to calculate a new volume to pool for a final sequencing run with evenly balanced indexing across all samples. This pool was sequenced with metrics identical with those of the QC pool. Library preparation and sequencing were performed at the University of Illinois at Chicago Sequencing Core (UIUCSCQ).

NGS Data Preprocessing. For each sample, the NGS reads were first trimmed and filtered using cutadapt (v1.16). Only NGS reads pairs with both Illumina Common Sequence adapters, CS1 and CS2, were kept. Of these, CS2 was trimmed off each R1 sequence and CS1 was trimmed off each R2 sequence. Cutadapt parameters were set as follows: a minimum quality cutoff (-q) of 30, a maximum error rate (-e) of 0.05, a minimum overlap (-m) of 1. The minimum overlap was set to be higher than the default value of 3 because extended sequences in this case are random, and we did not want to filter out sequences where the final 1–10 bases just happen to look like the first 10 bases of CS2 (the read must still contain a full CS2 sequence for it to be kept and subsequently trimmed, however). The 5′(-a) adapter trimmed from the R1 reads was 5′AGACCAACTCTTCAGCTACGTTAGS′ (CS2 reverse complement), and the 5′(-A) adapter trimmed from the R2 reads was 5′TCTGTAACAGCTCTGTCAGTGTT3′ (CS1 reverse complement). FastQC was used to quickly inspect the output trimmed fastq files before downstream analysis. See filter and trim TdT.sh at https://github.com/ryo-nu/turtles for an example preprocessing script. All runs were trimmed using this script. All initial preprocessing was done on Quest, Northwestern University’s high-performance computing facility, using a node running Red Hat Enterprise Linux Server release 7.5 (Maipo) with 4 cores and 4 GB of RAM, although only 1 core was used. Preprocessing took between 5 and 30 min depending on the number of conditions, replicates, and reads per replicate in a given run. Finally, for each analysis, we did further preprocessing locally. We cut off bases that were still present in the reads but not added during the experiment. Degenerate bases (if any) that are part of the 5′-ssDNA substrate (at its 3′-end before the extension) were removed from the beginning of each sequence. Then, we cut off 5.8 bases off the end of every sequence because we found that, on average, 5.8 bases were being added after the extension reaction during the 16 h ligation step (Figure S9). Because 5.8 is not an integer value, we cut 5 bases off of 80% of the sequences and 6 bases off of 20% of the sequences. We also filtered out sequences with lengths of less than 6 bases.

Time Point Prediction for 0 → 1 Single Step Change Experiment. All further analysis was done in Python using Jupyter Notebooks. You can find all the Jupyter Notebooks used for this publication at https://github.com/ryo-nu/turtles. The following algorithm was applied in order to (1) read and normalize each sequence by its own length, (2) calculate a distance metric using the relative dATP, dCTP, dGTP, and dTTP percent incorporation changes between each condition and the 0 control, and (3) transform distances for all conditions into 0 → 1 space on the basis of the 0 and 1 control distance values.

We first normalized each sequence by length, such that all bases in each sequence were counted across 1000 bins. For example, for a sequence of length 10, the first base would get counted in the first 100 bins, the next base in bins 100–200, and so on. We then calculated the base composition, X0,i in the sequence for condition i at each bin with position j, using the formula for a closure (eq 1). Note that i is unique for each (condition, replicate) pair if multiple replicates are present for a given experimental condition.

\[ X_{0,i} = \frac{n_{i0}}{\sum_{k \in N} n_{ijk}} \]  

Here, \( n_{i0} \) is the total count of dATP, dCTP, dGTP, or dTTP depending on the value of k (k∈N = {A,C,G,T}) across all sequences for condition i at bin j.

To calculate the distance between two compositions at a given bin location (e.g., between the 0 and 1 controls at every bin), we had to first transform the compositional data. We could not simply take the L2 norm difference of each compositional element because the elements of a composition violate the principle of normality due to the total sum rule (all elements add up to 100%). Thus, the data were first transformed by using the center log-ratio (clr) transformation, which maps this 4-component composition from a 3-dimensional space to a 4-dimensional space. We then took the L2 norm of these transformed normal elements. This distance metric is known as the Aitchison distance, which was used here to calculate the base composition distance, \( d(0,i) \), from the 0 control to each condition i at each bin j (eq 2).

\[ d(0,i) = \sqrt{\sum_{k \in N} \left( \frac{X_{0,k}}{g(X_{0})} - \frac{X_{k,i}}{g(X_{i})} \right)^2} \]  

\( g(X) = \prod_{k \in N} X_{k} \) (eq 3).

For condition i and bin j, the output signal \( s_{ij} \) is calculated as

\[ s_{ij} = d(0,i) - d(0,0) = \frac{d(0,i)}{d(0,1)} \]  

where \( d(0,1) \) is the Aitchison distance between the 0 control base composition and 1 control base composition at bin j. \( d(0,0) = 0 \) for all j. If there were multiple replicates for the 0 control, their average composition was used for \( X_{0,i} \) and \( X_{0} \) in eq 2. If there were multiple replicates for the 1 control, their average composition was similarly used to calculate \( d(0,1) \) in eq 4.

Next, the switch times were estimated for each condition i, which contains a change in output signal \( s_{ij} \) (e.g., via addition of Co halfway through the reaction). For experiments with more than one change (e.g., 0 → 1 → 0), a more sophisticated approach was used and is detailed below. However, the following simpler, more intuitive approach was used to predict switch times for 0 → 1 and 1 → 0.

Switch times were estimated for a given condition, i by (1) finding \( t_{0,i} \), the average location across all the sequences (bin position j) at which half of the 1 control output signal is reached (i.e., \( s_{ij} = 0.5 \)), (2) calculating \( r \), the ratio of the average rate of nucleotide addition for the 0 and 1 controls, and (3) using \( t_{0,i} \) and \( r \) to calculate the switch time, \( t_{s,i} \), using eqs 5 and 6. For a derivation of eq 5, see Supplementary Note 3.
Experiment. To predict the Co2+ condition in the output signal of the simulated nucleotides, we used the algorithm we developed from Glaser et al. for decoding continuous concentrations. The input to this algorithm is the amount of output signal on every nucleotide. Here, the output signal is \( s_j \) from the previous section. The algorithm uses this information to predict continuous values of Co2+ between 0 and 1 for all time points that are most likely to produce the amount of output signal on the nucleotides. To binarize these predictions, we then set a threshold of 0.5. To be able to predict the values of Co2+, the algorithm requires a knowledge of the rate of nucleotide addition. Here, we use the normalized signal. The algorithm also requires a knowledge of the rate of nucleotide addition. Here, we fit an inverse Gaussian distribution to the average experimental dNTP addition rate distribution (the distribution of the sequence lengths divided by the experiment time) from the control experiments. Note that this algorithm also assumes that the rate of dNTP addition is independent of the cation concentration. Thus, when making predictions in the 0 → 1 → 0 experiment, we do not account for differences in the rate of dNTP addition distributions between the 0 and 1 conditions.

In Silico Simulations of Recording Faster and Higher Numbers of Input Signal Changes. Using the average dNTP incorporation rate from experiments, and the amount of output signal in the control conditions, we simulated additional experiments in silico. Each simulated experiment had at least 6 signal changes (instances of a single signal change from 0 → 1 or 1 → 0), where each condition was randomly chosen to be 0 or 1. All nucleotides that were added during the 0 or 1 condition had the signal associated with these control conditions. More specifically, to account for the experimental variability in signals within a given control condition, nucleotide signals were sampled from a normal distribution determined by the experimental variability of nucleotide signals within the control conditions. We calculated the variability in two ways, corresponding to the two representative curves in Figure S12A,C. In the first, the variability was calculated across the first 100 nucleotides, in which there were at least 2000 recordings of all base numbers. In the second, the variability was calculated across the first 50 nucleotides, in which there were at least 60000 recordings of all base numbers. Using the output signal of the simulated nucleotides, we used the algorithm we developed from Glaser et al. for decoding binary concentrations. The accuracy corresponds to the percentage of conditions correctly classified as 0 or 1 over the duration of the entire recording experiment.

Time Point Estimation for 0 → 1 and 1 → 0 Single-Step Changes for TURTLES-2 Using an Inverse Model. To predict the Co2+ condition in the 0 → 1 → 0 experiment, we used a variation of the algorithm we developed from Glaser et al. for decoding continuous concentrations. This algorithm will predict continuous values of Co2+ between 0 and 1 for all time points that are most likely to produce the amount of signal. Here, instead of using the amount of signal on every nucleotide to predict the continuous concentrations, we use the normalized signal.

Let \( s_j \) be the signal as a function of the condition \( i \) and the normalized position \( j \). Let \( \gamma_j(i) \) be the probability that a nucleotide corresponding to the normalized position \( j \) was written at time \( i \). Let \( \mathbf{C}_i \) be the normalized cation concentration for condition \( i \). As in ref 13, our model is that \( s_j \approx \sum \gamma_j(i)\mathbf{C}_i \). We use maximum likelihood estimation to find a \( \mathbf{C}_i \) value that minimizes \( \sum (s_j - \sum \gamma_j(i)\mathbf{C}_i)^2 \) subject to \( \mathbf{C}_i \in \{0.1\} \) and the given condition \( i \). To binarize the predictions, we then set a threshold for \( \mathbf{C} \) of 0.5.

Here, for an experiment of duration \( t_{\text{exp}} \) (e.g., 60 min), we let \( \gamma_j(t_{\text{exp}}) = N((j - 0.5)/t_{\text{exp}}, \sigma) \), where \( N \) renormalizes the probability distribution after values outside the domain of \([0, t_{\text{exp}}]\) are set to 0. We set \( \sigma \) for each experiment so that \( \gamma_j(t_{\text{exp}}) \) is equal to the frequency of strands with a single nucleotide divided by \( t_{\text{exp}} \) (because a normalized position of 1 would generally only be written in the \( t_{\text{exp}} \)th minute when there is a single nucleotide strand). Note that future work that more accurately models the kinetics of the polymerase to get a more accurate estimate of \( \gamma \) will provide improved results.

When running this algorithm on the Ca2+ data which have different rates when Ca2+ is or is not present, following the prediction of Ca2+ over time with the above algorithm, we used the ratio of incorporation rates between the 0 and 1 conditions, as described by eqs 5 and 6, to rescale the results.

ASSOCIATED CONTENT
Supporting Information
The Supporting Information includes an which are referred to in the main text. The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c07331. Extended descriptions of methods and supplementary figures and tables as described in the text (PDF)

AUTHOR INFORMATION
Corresponding Author
Keith E. J. Tyo — Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States; Email: k-tyo@northwestern.edu; Fax: +1 847 491 3728

Authors
Namita Bhan — Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States; Mitolab, Cambridge, Massachusetts 02139, United States
Alec Callisto — Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States; orcid.org/0000-0002-9556-7389
Jonathan Strutz — Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States; orcid.org/0000-0003-0934-4283
Joshua Glaser — Center for Theoretical Neuroscience, Columbia University, New York, New York 10027, United States
Reza Kalhor — Department of Biomedical Engineering, Center for Epigenetics, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, United States; orcid.org/0000-0002-5558-7545
Edward S. Boyden — Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; McGovern Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, United States

https://doi.org/10.1021/jacs.1c07331

16638
George Church – Department of Biomedical Engineering, Center for Epigenetics, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, United States

Konrad Kording – Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c07331

Author Contributions

N.B. and A.C. contributed equally to this work.

Notes

The authors declare the following competing financial interest(s): A utility patent has been filed for some of the inventions described in this work.

ACKNOWLEDGMENTS

The authors acknowledge Marija Milisavljevic for help with some experiments and Bradley Biggs for helpful discussions and comments on the manuscript. This research was supported in part through the computational resources and staff contributions provided for the Quest high-performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. All next-generation sequencing was done with the help of the Next Generation Sequencing Core facility at the University of Illinois at Chicago. Sanger sequencing was supported by the Northwestern University Keck Biophysics Facility and a Cancer Center Support Grant (NCI CA060533). The Keck Biophysics Facility’s Azure Sapphire Imager was funded by a 1S10OD026963-01 NIH grant. Protein purification was supported by the Northwestern University Recombinant Protein Production Core. This work was funded by the National Institutes of Health grants R01MH1103910 (to K.E.J.T., K.K., E.S.B., and G.C.), and U1FNS107697 (to K.E.J.T., K.K., and E.S.B.) and a National Institutes of Health Training Grant (T32GM008449) through UF1NS107697 (to K.E.J.T., K.K., and E.S.B.), and was funded by the National Institutes of Health grants NS066342 and R01MH103910 (to K.E.J.T., K.K., E.S.B., and G.C.).

REFERENCES


Recording Temporal Signals with Minutes Resolution Using Enzymatic DNA Synthesis

Namita Bhan1,10,*, Alec Callisto1,*, Jonathan Strutz1, Joshua Glaser2, Reza Kalhor3, Edward Boyden4,5,6, George Church3,4, Konrad Kording2, Keith E.J. Tyo1‡

1Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, 60208, USA
2Center for Theoretical Neuroscience, Columbia University, New York, NY, 10027, USA
3Center for Epigenetics, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA
4Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, 02139, USA
5McGovern Institute, MIT, Cambridge, MA, 02139, USA
6Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical School, Boston, MA, 02115, USA
7Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA
8Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, 02115, USA
9Department of Neuroscience, University of Pennsylvania, Philadelphia, PA, 19104, USA
10Mitolab, Cambridge, MA, 02139, USA

*These authors contributed equally to this work

‡Corresponding author: Keith E.J. Tyo
Telephone: +1 847 868 0319
Fax: +1 847 491 3728
Email: k-tyo@northwestern.edu

Competing Financial Interests
A utility patent application has been filed for some of the developments contained in this article by KT, NB, AC, JS, JG and KK.

Classification: Biological Sciences/Applied Biological Sciences

Keywords: DNA-based recording, enzymatic DNA synthesis, physiological signal recording, synthetic biology

This file includes:
Supplementary methods 1-6
Supplementary figures S1-S21
Supplementary tables S1-S2
**Supplementary Methods**: Extended description of methods

1. **Extension reaction with individual dNTPs for testing effect of Co\(^{2+}\):**

For initial testing to show Co\(^{2+}\) dependent dNTP preference change the ssDNA substrate used was AMD006 (Table S2). Total reaction volume was 25 μL with 0.1 μM ssDNA substrate, 1x NEB TdT reaction buffer, and 0.1 mM of each dNTP tested. Final concentration of CoCl\(_2\) in the test reaction was 0.25 mM. Reactions were initiated by addition of 5 units of TdT per reaction. Reactions were run for 30 minutes at 37 °C and stopped by boiling at 70 °C for 10 minutes. Then, 8 μL of the reaction was mixed with 12 μL of TBE-Urea loading dye and boiled for 10 minutes at 100 °C. All of the diluted extension reaction was then loaded onto 30 μL, 10 well 10% TBE-Urea Gel (Bio-Rad) and run for 40 minutes at 200 V. Immediately after the run was over, the gel was stained with Sybr Gold for 15 minutes and imaged on ImageQuant BioRad.

2. **Extension reactions for 1→0 set-up:**

Mg\(^{2+}\) only for 1 hour (signal 0) and Mg\(^{2+}\)Co\(^{2+}\) for 1 hour (signal 1) were set-up as regular extension reactions mentioned in Materials and Methods. The 1→0 reactions where the signal changed from 1 to 0 at 40 minutes were put through a ssDNA wash step at 40 minutes. ssDNA wash to remove cations, TdT and dNTPs was done exactly as mentioned in Materials and Methods. Reactions were all run for 1 hour in triplicates. Signal 0 and signal 1 controls were run for 1 hour for each set-up in triplicates and also put through the ssDNA wash step at 40 minutes.

3. **Derivation of Equation 5**

We start by deriving the equations for the average rate before the switch \((r_a)\) and after the switch \((r_b)\) for condition, \(i\):

\[
r_a,i = \frac{j_i^*}{t_i^*} \quad (1a)
\]

\[
r_b,i = \frac{1 - j_i^*}{t_{expt} - t_i^*} \quad (2a)
\]

where \(j_i^*\) is the average location in the sequences (length fraction, 0 to 1) at which the output signal, \(s_{ij}\), reaches 0.5 (Equation 4), \(t_i^*\) is the switch time, and \(t_{expt}\) is the total duration of the experiment. Because we can estimate \(r_{a,i}\) and \(r_{b,i}\) from average rates of the 0 and 1 controls across replicates \((\bar{r}_{a,ctrl}\) and \(\bar{r}_{b,ctrl}\)), we can use their ratio to combine equation 1a and 2a, above to write

\[
\frac{\bar{r}_{a,ctrl}}{\bar{r}_{b,ctrl}} \approx \frac{r_{a,i}}{r_{b,i}} = \frac{j_i^*}{t_i^*} \left( \frac{t_{expt} - t_i^*}{1 - j_i^*} \right) \quad (3a)
\]

Solving for \(t_i^*\), we get equation 5:

\[
{t_i^*} = \frac{\alpha t_{expt}}{j_i^* + \alpha - 1} \quad (5)
\]

where
\[
\alpha = \frac{\overline{r_{b,ctrl}}}{\overline{r_{a,ctrl}}} \quad (4a)
\]

We use equation 5 for time prediction \((t_i^*)\) after calculating \(j_i^*\) for a given condition and \(\alpha\) from the 0 and 1 controls. In equation 4a, \(\alpha\) is the first condition before the switch (0 or 1) and \(b\) is the condition after the switch (1 or 0).

4. Extensions reaction set-up for calculating rate of dNTP addition:

Each extension reaction consisted of a final concentration of 10 μM initiating ssDNA substrate, 1 mM dNTP mix (each dNTP at 1 mM final concentration), 1.4x NEB TdT reaction buffer, and 10 units of TdT to a final volume of 50 μL. The ssDNA substrate used for this extension reaction was CS1_5N. We have shown (data not included) that the identity of the last 5 bases on the 3’ end of the substrate affects the identity of the dNTP added to the ssDNA substrate. Thus, we purchased a ssDNA substrate (CS1_5N) with the last 5 bases having the base composition same as TdT dNTP preference under signal 0 (25% dATP, 15% dCTP, 45% dGTP and 15% dTTP). The reactions were initiated upon addition of TdT and run at 37 °C for 2 hours. 2 μL of sample was collected and immediately frozen (on ice, 0 °C) at 30 s, 1 min, 2 min, 3 min, 4 min, 5 min, 10 min, 20 min, 30 min, 45 min, 60 min, 92 min and 120 min. Subsequently, each sample was put through the ligation and Illumina library generation process as mentioned in Materials and Methods.

5. Test set-up for checking ssDNA clean-up kit bias:

Mg\(^{2+}\) only for 1 hour (signal 0) and Mg\(^{2+}\)+Co\(^{2+}\) for 1 hour (signal 1) were set up as regular extension reactions mentioned in Materials and Methods. The 0→1 reactions where the signal changed from 0 to 1 during the 1 hour extension were run starting with 45 μL with Mg\(^{2+}\) only. 5 μL of 2.5 mM CoCl\(_2\) was added at 10 min. Reactions were all run for 1 hour in triplicates. Fresh signal 0 and signal 1 controls were run for 1 hour with each set-up. 2 μL of extension reaction was used for ligation (“No Wash” set of samples). Ligation and subsequent PCR steps for Illumina library generation were followed as mentioned in Materials and Methods. Rest of the 48 μL of extension reaction was washed using the ssDNA clean-up kit. Protocol was followed as mentioned in the kit. ssDNA was eluted into 25 μL of ddH\(_2\)O and 2 μL of that was used for ligation (“Wash” set of samples). Ligation and subsequent PCR steps for Illumina library generation were followed as mentioned in Materials and Methods. Data obtained from Illumina sequencing was analyzed for the “No Wash” and “Wash” set of samples. Further, switch time calculations were carried out as mentioned previously (Fig. S11).

6. High calcium conditions for TURTLES-2 reactions:

High calcium conditions for 0→1 and 1→0 reactions had different compositions to enable transitions without employing intermediate column washes. Although commercial TdT reaction buffer contains no added calcium, we observed that Ca\(_{2-}\)-mTdT(M13-388) was inactivated in reactions that were not supplemented with at least 50 μM EGTA (Figure S17), very likely due to calcium present in water or other reagents used. By titrating EGTA into a TdT extension reaction supplemented with 7 μM fura-2 until the fura-2 signal plateaued, we estimated that most free calcium could be eliminated from the reactions by the addition of 50 μM EGTA. By employing the un-supplemented reaction buffer as the high calcium condition for 1→0 reactions, we were able to transition to a low calcium condition with the addition of 50 μM EGTA.
Figure S1: Testing change in individual dNTP preference upon Co\(^{2+}\) addition

ssDNA substrate extensions carried out by TdT using just dATP, dTTP, dGTP, or dCTP in presence of Mg\(^{2+}\)+Co\(^{2+}\) (first 4 lanes) or in presence of just Mg\(^{2+}\) (next 4 lanes) were run on a gel. “L” is ssDNA size marker. Reactions were carried out as mentioned in supplementary text.
Figure S2: Aitchison distance for dATP, dCTP, dGTP and dTTP incorporation by TdT in the presence or absence of various signals. Signal 0 is always 10 mM Mg$^{2+}$ at 37 °C for 1 hour. Signal 1 was, going from left to right: (1) 10 mM Mg$^{2+}$ + 0.25 mM Co$^{2+}$ at 37 °C for 1 hour; (2) 10 mM Mg$^{2+}$ + 1 mM Ca$^{2+}$ at 37 °C for 1 hour; (3) 10 mM Mg + 20 μM Zn$^{2+}$ at 37 °C for 1 hour; and (4) 10 mM Mg$^{2+}$ at 20 °C for 1 hour. Error bars show two standard deviations of the mean Aitchison distance. Statistical significance was assessed after first transforming the data into Aitchison space which makes each dNTP frequency change statistically independent of the others. All base incorporation changes were found to be statistically significant, as shown by asterisks, ($\alpha = 0.01$) except for the change in dATP for 20 °C signal (p-value = 0.019).
Figure S3: Length distribution of extensions upon addition of Co$^{2+}$ based on NGS data. We calculated the mean frequency distribution of extension lengths for each condition (three biological replicates for each condition). Addition of Co$^{2+}$ did not change the length distribution significantly.
Figure S4: Length distribution of extensions upon addition of Zn\(^{2+}\) as seen on ssDNA gel

Extension reactions were run as mentioned in Materials and Methods section. Two biological replicates per test condition were then loaded onto a ssDNA gel (Mg\(^{2+}\) on left and Mg\(^{2+}\)+Zn\(^{2+}\) on right). Addition of Zn\(^{2+}\) increases the overall lengths of the extensions.
Figure S5: Length distribution of extensions upon addition of Zn\textsuperscript{2+} based on NGS data
We calculated the mean frequency distribution of extension lengths for each condition (three biological replicates for each condition). Addition of Zn\textsuperscript{2+} caused a shift in probability distribution toward longer lengths.
Figure S6: Length distribution of extensions upon addition of Ca$^{2+}$ as seen on ssDNA gel

Extension reactions were run as described in Materials and Methods section. Three biological replicates per test condition were then loaded on a ssDNA gel (Mg$^{2+}$ on left and Mg$^{2+}$+Ca$^{2+}$ on right). Addition of Ca$^{2+}$ decreases the overall lengths of the extensions.
Figure S7: Length distribution of extensions upon addition of Ca\(^{2+}\) based on NGS data.

We calculated the mean frequency distribution of extension lengths for each condition (three biological replicates for each condition). Addition of Ca\(^{2+}\) caused a shift toward shorter lengths.
Figure S8: Length distribution of extensions upon using temperature as a signal based on NGS data

We calculated the mean frequency distribution of extension lengths for each condition (three biological replicates for each condition). Reducing the temperature of the extension reaction to 20 °C caused a shift toward shorter lengths.
Figure S9: Anomalous dNTP composition initially found at the end of reads and rate of reaction measured for extensions with only Mg$_{2+}$ present

We observed a significant change in the individual dNTP frequency towards the ends of the ssDNA sequences synthesized. (A) Presents the significant change observed near the end of all reads with all the signals tested. Since we directly use 2 µL of extension reaction for ligation, the diluted TdT seems to be adding dNTPs to the ssDNA after the recording experiment, during the 16-hour ligation step. (B) To prove that these dNTPs were not added during the extension reaction (i.e. after the reaction), we sampled extension reactions (with Mg$_{2+}$ only) at several time points (Supplementary Text). We then calculated the mean extension length at each timepoint and applied a linear regression. The R$^2$ value of 0.96 for a straight line indicates that our assumption of constant rate (assuming input signal does not change) is valid. The slope of 0.17 reveals an average incorporation rate of 0.17 dNTPs/minute for this condition. Most importantly, the intercept of 5.82 indicates addition of 5.82 dNTPs (on average) either before or after the extension reaction. These are almost certainly being added after the extension reaction during the ligation step, which we conclude based on the anomalous behavior we see at the end of sequences in Panel A. (C) We created plots of the data from Panel A after trimming off last few dNTPs. See Materials and Methods for details on how these 5.8 bases were trimmed from the end of all sequences before further analysis.
Figure S10: Recording a single 1→0 step change in Co$^{2+}$ concentration onto ssDNA *in vitro*. 0.25 mM Co$^{2+}$ was removed at 40 minutes to generate a 1→0 transition. This mean output signal across 6 biological replicates shows there is a difference in the preference of dNTP incorporated by TdT in the Mg$^{2+}$ (green) and Mg$^{2+}$+Co$^{2+}$ (orange) control conditions. The plot further shows the changes from 1→0 for Co$^{2+}$ removed at 40 minutes (blue). We were able to get a switch time of 36.7 minutes with a std. dev. of 9.9 minutes using methods used for 0→1 switch time predictions. We suspect the possible reason for higher variance in time prediction for this set-up was due to the ssDNA wash step at 40 minutes as discussed in Fig. S11.
Figure S11: dNTP bias & variability introduced by ssDNA wash columns
We present here a comparison of the composition of sequences retained when the extension reactions were directly used for ligation (“No Wash”) vs. when the same extensions were put through a ssDNA wash kit (“Wash”). (A), (B), (C) and (D) show individual plots of each nucleotide frequency seen in extension reactions between No Wash vs Wash conditions for just Mg$^{2+}$ extensions. (E), (F), (G) and (H) show individual plots of each nucleotide frequency seen in extension reactions between No Wash vs Wash conditions for Mg$^{2+}$+ Co$^{2+}$ extensions. We observed a bias in overall dNTP content introduced by the columns used for ssDNA clean-up, when the reactions were washed after the recording experiment was stopped. ssDNA sequences with certain dNTP compositions were preferentially retained on the columns. (I) and (J) are plots for time prediction for No Wash and Wash condition respectively. We carried out an input signal of Co$^{2+}$ 0→1 at 10 minutes for a 1-hour extension. We obtained a time prediction of 12.8 minutes with 1.8 min std. dev. for No Wash condition. We obtained a time prediction of 12.4 min with a std. dev. of 1.2 min for the Wash condition. While the time predictions were very similar, there is a clear increase in variability (std. dev.) for the later part of the output signal recorded in (J) as compared to (I) (shown with a red arrow). Taken together, such biases and variability when introduced during the wash step for 0→1→0 experiment at 40 minutes for replacing +Co$^{2+}$ buffers with −Co$^{2+}$ buffers (See Materials and Methods: Extension reactions for 0→1→0 set-up) would cause more noise for the final 20 minutes of the recording.
Figure S12: *In silico* characterization of shortest resolvable input signal changes and highest number of consecutive signal changes that could be resolved using TURTLES.

*In silico* simulations based on experimental TdT parameters. For all simulations, we calculated the variability in two ways. The variability was either calculated across the first 100 nucleotides, in which there were at least 2000 recordings of all base numbers (red curve) or across the first 50 nucleotides, in which there were at least 60000 recordings of all base numbers (blue curve). (A) Representative signal for rapidly changing input from 0→1 or 1→0. (B) Output of *in silico* characterization of the signals represented by (A). We tried to find the shortest times of input changes from 0→1 or 1→0 that could be resolved using TURTLES. Simulations were carried out where the number of input signal changes were kept constant at 6 (e.g., 0→1→0→1→0→1→0) while the length of time for each input condition (0 or 1) was varied. Even with just 2000 strands of ssDNA recorded (each 100bp in length), input signal changing every 1 minute was resolved with > 75% accuracy, and input signal changing every 10 minutes was resolved with > 90% accuracy. (C) Representative signal for multiple changes in input signal from 0→1 or 1→0. (D) Output of *in silico* characterization of the signals represented by (C). Simulations were carried out with the duration of each input condition (0 or 1) kept constant at 10 minutes while varying the number of input signal changes. The accuracy of resolving this output signal was calculated across simulations. Even with just 2000 strands of ssDNA recorded (each 100bp in length), 3 input signal changes could be resolved with almost 100% accuracy and 10 input signal changes of 10 minutes each could be resolved at about 80% accuracy.
Figure S13: CaM-TdT(M13) design and SCHEMA crossover site frequency and locations. (A) schematic representation of CaM-TdT(M13) variants. (B) the number of times each crossover site appeared in the SCHEMA/RASPP analysis. Sites selected for fusions are colored in red. (C) ribbon diagram of the catalytic core of the short isoform of murine TdT (PDB ID: 4i27) with fusion site residues colored in green. Fusion sites were selected to minimize steric interference and target polymerase regions required for DNA binding (282,284,287) or catalysis (388).
Figure S14: Primer extension activity of M13 TdT fusion proteins. Fusion proteins were screened by assaying the activity of the M13 fusion variants without N-terminal CaM to ensure that mTdT remained active with the M13 fusion. Fusions were expressed in PURExpress and their activity was screened by primer extension. Primer extension reactions were performed without added CaCl₂ (indicated by "-"), and with 1mM CaCl₂ (indicated by "+"). The reaction products were visualized by urea-PAGE (see methods). Fusions are labelled as mTdT(M13-###), where the number indicates the residue immediately preceding the GS-M13-GS fusion.
Figure S15: Primer extension activity of CaM-mTdT(M13-388), dCaM-mTdT(M13-388), and mTdT(M13-388). The activity of each polymerase was tested in three conditions where “0” indicates reactions with no added calcium or EGTA, “+EGTA” indicates reactions supplemented with 1 mM EGTA, and “+Ca” indicates reactions supplemented with 1 mM CaCl2. dCaM-mTdT(M13-388) contains CaM mutations D20A, D56A, D93A, and D129A that ablate the calcium-binding capability of CaM(49). CaM-mTdT consists only of mTdT with N-terminal CaM fusion.
Figure S16: Primer extension activity of CaM affinity variants. The activity of calcium affinity variants was tested by primer extension and visualized by urea PAGE in non-supplemented ("0"), 1 mM CaCl$_2$ added ("C"), and 1mM EGTA added conditions ("E"). Crotti et al. report effective calcium affinities of 38 μM, 150 μM, and 15μM respectively as compared to 2.8μM for wildtype CaM.
Figure S17: Primer extension activity of purified CaM-mTdT(M13-388). A) Anion exchange elutions 1 (p.1) and 2 (p.2), visualized by PAGE. CaM-mTdT(M13-388) appears at approximately 70kDa. (B) Following buffer exchange, the activity of both fractions was assayed by primer extension and compared to the wildtype polymerase and MBP-CaM-mTdT(M13-388) (from PURExpress). 1 mM CaCl₂ added ("C"), and 1mM EGTA added conditions ("E").
Figure S18. Comparison of purified CaM-mTdT(M13-388) fractions. Base incorporation frequencies of CaM-mTdT(m13-388) elution fractions 1 and 2 in primer extension reactions.
**Figure S19: Incorporation preference of two-polymerase recording system.** The difference in base selectivity of TURTLES-2 in the presence of calcium as compared to calcium-free conditions. Individual points (black dots) are added to the right plot to highlight the variability between each replicate’s 1 frequency and the mean 0 frequency of each base.
Figure S20: TURTLE-2 Selectivity comparison. Incorporation frequencies of each base for the two-polymerase TURTLES-2 system, bovine TdT, and purified CaM-mTdT(M13-388).
Figure S21: Digital data storage capacity of a TURTLES-based system that employs changes in concentration of nucleotides supplied over time. The data density of recordings is determined by the number of possible input concentrations of dNTPs (ratios) that yield distinguishable outputs. For example, with 4 unique bases (A, C, G, and T) and a set of ratios built in increments of 5% (e.g. 35% A, 20% C, 15% G, 30% T), the maximum amount of data stored per ratio is 11 bits, as there are $2^{11}$ possible compositions of A, C, G, and T that add to 100% and use 5% increments for each nucleotide. Even greater densities are possible with smaller increments or use of unnatural bases. The main challenge of this system would be inferring back out the ratio at each timepoint from a population of sequences, especially at fine-grained ratio increments (e.g. at 5% increments). Some potential strategies to address these issues include reserving a base to act as a buffer between signals, increasing the length of each encoded signal, and using error correcting codes.
## Supplementary Tables

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS1</td>
<td>ACACTGACGACATGGTTCTACA</td>
<td></td>
</tr>
<tr>
<td>CS1_5N</td>
<td>ACACTGACGACATGGTTCTACANNNNN</td>
<td>Degenerate &quot;N &quot; bases are distributed 25% A, 15%C, 45%G, 15% T</td>
</tr>
<tr>
<td>FAM_NB</td>
<td>(6'-FAM)A<em>C'A</em>C'TGACGACATGGTTCTACAACCGGTATGTACGGCGGTTATCGTA</td>
<td>&quot;*&quot; follows phosphorothioate d bases</td>
</tr>
<tr>
<td>AMD00</td>
<td>AGGCTAGTCGTCGCTGATAAG</td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>CTCAGAGGCGACAGGCTGACAGGCTGACAAATTAACCGAAGAAC</td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>TGTCCTGCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>AGGCCACAGAGAAGGACAGAGGCTGCTGACGAATTAACCGAAGAAC</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>AACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P5</td>
<td>CTCGAGAAGTGCGTCTGCTGACGACAAATTAACCGAAGAAC</td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>CGCGTTGAGCAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>AGGTAAAGGCGGCTGCTGACTGACAAATTAACCGAAGAAC</td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>ACACCGATAGCTTTGCTGCTCCTTGGCTGCTACATCATCTGCT</td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>ATTCGACGACGACGACAAAGCTGACACAAATTAACCGAAGAACAAATC</td>
<td></td>
</tr>
<tr>
<td>P10</td>
<td>TGCGAATCTTTTATTACCAGCAGAACGACTCCCTCCCTCCAGAACC</td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>TATGACACTAGGCGACAGGAAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG</td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>GTTAAATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>TATGACACTAGGCGACAGGAAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG</td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>GTTAAATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P15</td>
<td>TATGACAGGACAGGAGCAGAGCAGGACAGGAGCAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG</td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td>GTTAAATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P17</td>
<td>TGAGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P18</td>
<td>GTTAAATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P19</td>
<td>GTTACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P20</td>
<td>TGATGCTTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P21</td>
<td>AGATCCAAAGTGACAAAGCAGGCTGCTGACGCGCTGAAAGTGATCCGTACGATG</td>
<td></td>
</tr>
<tr>
<td>P22</td>
<td>TGATCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P23</td>
<td>AAAAGTCGACAAAGCAGGCTGCTGCTGACGCGCTGAAAGTGATCCGTACGATG</td>
<td></td>
</tr>
<tr>
<td>P24</td>
<td>GGTCTTTGCTATTTCAGGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P25</td>
<td>AAAGCCGATAGCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P26</td>
<td>AAAACCCACGCTTCTTTGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P27</td>
<td>AAAGGTACCTCTGAGAAGTTGCGGTCTGCTGACGCGCTGAAAGTGATCCGTACGATG</td>
<td></td>
</tr>
<tr>
<td>P28</td>
<td>ACTTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P29</td>
<td>AGATCCAAAGTGACAAAGCAGGCTGCTGACGCGCTGAAAGTGATCCGTACGATG</td>
<td></td>
</tr>
<tr>
<td>P30</td>
<td>TGTTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P31</td>
<td>AAAAGTCGACAAAGCAGGCTGCTGCTGACGCGCTGAAAGTGATCCGTACGATG</td>
<td></td>
</tr>
<tr>
<td>P32</td>
<td>GGTCTTTGCTATTTCAGGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P33</td>
<td>AAAGCCGATAGCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P34</td>
<td>AAAACCCACGCTTCTTTGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>P35</td>
<td>AAAGGTACCTCTGAGAAGTTGCGGTCTGCTGACGCGCTGAAAGTGATCCGTACGATG</td>
<td></td>
</tr>
<tr>
<td>P36</td>
<td>ACTTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
<td></td>
</tr>
</tbody>
</table>
S1    GCATTCTGGTATGCCGTCCG
S2    GGTGGACGAAATGATCCGCG
S3    GTTCGTGTGTCGCTAGCGG
S4    GGCTGAAGCGGTAAGTATGTTGG
S5    CGAGGAGGAAATCCGCGAAGC
S6    GATCACTTCTCCAGGAGCCACAG
S7    GAGCCTTCATTCCCTATTACCAG
S8    GTCCGCGTAGAGGATGAG
S9    CACTTGAAGAAGGGACTGAAGGCC

Table S1: DNA substrate sequences and primer sequences
Table S2: Amino acid sequences of TdT-CaM fusions

Expression vector sequences available upon request