



Supporting Online Material for

Real-Time DNA Sequencing from Single Polymerase Molecules

John Eid, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, David Rank, Primo Baybayan, Brad Bettman, Arkadiusz Bibillo, Keith Bjornson, Bidhan Chaudhuri, Frederick Christians, Ronald Cicero, Sonya Clark, Ravindra Dalal, Alex deWinter, John Dixon, Mathieu Foquet, Alfred Gaertner, Paul Hardenbol, Cheryl Heiner, Kevin Hester, David Holden, Gregory Kearns, Xiangxu Kong, Ronald Kuse, Yves Lacroix, Steven Lin, Paul Lundquist, Congcong Ma, Patrick Marks, Mark Maxham, Devon Murphy, Insil Park, Thang Pham, Michael Phillips, Joy Roy, Robert Sebra, Gene Shen, Jon Sorenson, Austin Tomaney, Kevin Travers, Mark Trulson, John Vieceli, Jeffrey Wegener, Dawn Wu, Alicia Yang, Denis Zaccarin, Peter Zhao, Frank Zhong, Jonas Korlach,* Stephen Turner*

*To whom correspondence should be addressed. E-mail: jkorlach@pacificbiosciences.com (J.K.); sturner@pacificbiosciences.com (S.T.)

Published 20 November 2008 on *Science Express*
DOI: 10.1126/science.1162986

This PDF file includes:

Materials and Methods

Figs. S1 to S8

Tables S1 to S3

References

Other Supporting Online Material for this manuscript includes the following:

Movie S1

Materials & Methods

Zero-mode waveguide (ZMW) fabrication. ZMW nanostructures were fabricated as previously described (1). ZMWs were arranged in a rectangular array of 93 rows and 33 columns, with 1.3 μm spacing between rows and 4.0 μm spacing between columns. A patterned “X” of missing ZMWs (Supplementary movie S1) aided spatial alignment in the instrument. Selective immobilization of DNA polymerase to the bottom glass surface of the ZMW was achieved by aluminum surface passivation using polyphosphonates (2), with an added deposition of Biotin-polyethyleneglycol-trimethoxysilane (Laysan Bio, Arab, AL) to enable specific attachment of the polymerase carrying an N-terminal Biotin-tag (3). Functionalized nanostructures were stored in vacuum until use.

Excitation volume comparison. The order of magnitude calculation for the TIRF volume used a 1/e evanescent field depth of 100 nm (4), but the lateral PSF in that reference was large (1.2 μm) compared to standard setups which achieve 1/e radius point spread function extensions of 0.4 μm (5). The calculated TIRF observation volume using $\pi r^2 h$ is thus 0.05 μm^3 . In comparison, a 100-nm diameter ZMW with a 1/e evanescent field depth of 0.03 μm yields a total volume of 0.0002 μm^3 .

DNA polymerase, DNA template and phospholinked dNTP preparation. DNA templates and DNA polymerase were generated as described (6). DNA template and primer sequences were 5'-GCTTCGTCTCAAAAAGAGAAGAGGATTTAATATACCA CACCAGAGAAGAGGATTTAATATACCAACACCAGAGAAGAGGATTTAATATAC CACACCAGAGAAGAGGATTTAATATACCAACCATCAGTCACGTCTAGATGC AGTCAGAT-3' and 5'-CTGACTGCATCTAGACGTGACT GA-3' for the experiment shown in Fig. 2; 5'-GATGTTGAAGTAGTAGTTGAAGATGTAGTAGATGTTCA ACAACTACTTCATCAACTACAACCTTCATCTACTA-3' and 5'-TACTACTTCAAC ATCTAGTAGATGAAGTTG-3' for Fig. 3; 5'-TTAAAATTTAAAGCAA

CTACAGGTTTGTTTAAAGATTTTATAGATAAATGGACGTACATCAAGACGAC
ATCAGAAGGAGCGATCAAGCAACTAGCAAACTGATGTTAAACAGTCTATAC
GGTAAATTCGCTAGTAACCCTGATGTTACAAACACCACCTACCACCTATCTAC
ATCACCA-3' and 5'-TGGTGATGTAGATAGGTGGTAGGTGGTGGTGT-3' for Fig. 4;
and 5'-GGGAGTTTATAGATAGGTAAATTTATGGATGGATGGAAGCCACCAAGC
AAGCCCAACAGCAAGCAAGACAAA-3' and 5'-ATCTATAAACTCCCTTTGTCTT
GCTTGC-3' for fig. S3.

Proprietary site-specific mutations were introduced to the DNA polymerase gene by total gene synthesis (DNA2.0, Menlo Park, CA) and site-directed mutagenesis (GenScript, Piscataway, NJ) to enhance the kinetic properties of the polymerase utilizing phospholinked dNTPs to approach native dNTP incorporation characteristics. Phospholinked dNTPs were synthesized as described (6), with the substitution of Fmoc-aminohexyl-triphosphate for Fmoc-aminohexyl-diphosphate in the condensation reaction with nucleoside triphosphate to yield nucleoside hexaphosphates. Molecular structures and normalized emission spectra for the four phospholinked dNTPs utilized in this report are shown in fig. S4. The Alexa Fluor 555 and 568 dyes (Invitrogen, Carlsbad, CA) are excited with the 532 nm laser line, while the Alexa Fluor 647 and 660 dyes are excited with the 643 nm laser line.

DNA sequencing assays. Biotinylated DNA polymerases were incubated at 4°C for ten minutes with 1.5x molar excess of primed DNA templates in loading buffer (50 mM MOPS, pH 7.5, 75 mM potassium acetate, 5 mM dithiothreitol and 0.05% Tween-20). Simultaneously, streptavidin (Invitrogen), at a 2-fold stoichiometric excess over polymerase, was incubated in the same buffer at 22°C on the ZMW array. The array was then washed five times with loading buffer. Thereafter, polymerase/template complexes were immobilized onto the arrays at 4°C for 15 minutes. Unbound complexes were removed by washing five times with reaction buffer (50 mM ACES, pH 7.1, 75 mM potassium acetate and 5 mM dithiothreitol) and the array was prepared for sequencing by adding an enzymatic oxygen scavenging system, triplet state quencher (1x protococatechuate dioxygenase, 4 mM protococatechuic acid and 6 mM nitrobenzoic acid (Sigma-Aldrich, St. Louis, MO)), and all phospholinked dNTPs (250 nM final concentration each), except the one corresponding to the first base to be incorporated into

the DNA template. Sequencing reactions were initiated by addition of this phospholinked dNTP and manganese acetate to final concentrations of 250 nM and 0.5 mM, respectively.

Data collection & analysis. Data were collected on a highly parallel confocal fluorescence detection instrument, using prism-based dispersion optics and an electron-multiplying charge coupled device camera to spectrally resolve single-molecule emission (7). Supplementary Movie S1 shows an example of a measurement (100 Hz camera frame rate). Spectral information is spatially distributed in the horizontal dimension between each ZMW column, going from shorter to longer wavelengths from left to right.

Fluorescence pulse calling was performed by a threshold algorithm on the dye-weighted intensities using fluorescence emission calibration spectra for each of the phospholinked dNTPs. The spectrum that yields the minimum chi-squared difference when compared to the pulse spectrum identifies the pulse as a particular phospholinked dNTP.

The resulting reads were aligned using a Smith-Waterman algorithm (8, 9) allowing local alignment of sections of reads to portions of the template. Interpulse duration values were tabulated from all consecutive pairs of correctly aligning template positions (Fig. 4D). The consensus alignment was performed using the “center-star” multiple sequence alignment algorithm (10). Consensus accuracy and length values were obtained using a bootstrapping approach that randomly sampled the reads across the data set. The pattern of errors across the consensus was statistically assessed by applying a sub-sampling procedure to the read alignments. One hundred samples, each containing 15X average depth of coverage, were drawn from the observed reads. The histogram of per-position error counts for these samples is shown in Fig. 4F (grey bars). To investigate whether the observed consensus errors were due to systematic position-dependent sources or if they were consistent with a random error model based on the photophysics of the system, we simulated an equivalent number of reads from a random model which distributed errors according to the per-dye probabilities for mismatch, insertion, and deletion errors (but blind to sequence position). These reads were subjected to the same sub-sampling procedure as the true reads and the resulting distribution of consensus errors is also shown in Fig. 4F (black bars). The good match between the observed and

random distributions suggests that dye-specific errors are sufficient to explain the observed variation in the consensus accuracy, and no sequence-context specific bias is appreciably affecting these results.

For analysis of missing pulse probabilities, a semi-analytical model was created (fig. S6). The pulse signal (total intensity) distribution was determined empirically. From this distribution the overall signal-to-noise, SNR, was computed using:

$$SNR = \frac{signal}{\sqrt{2(signal + bkg)}};$$

where bkg is the baseline total intensity during an equivalent duration as the signal (derived empirically from the exponential of the pulse width distribution) and the factor of $\sqrt{2}$ is due to the electron-multiplication noise of the CCD detector (11). Taking the SNR and the threshold sigma, σ_{thresh} , used in the pulse calling, a false negative probability was calculated:

$$FNP = GaussianCDF\left(\frac{SNR}{\sigma_{thresh}}\right);$$

where *GaussianCDF* represents the Gaussian cumulative distribution function. The detected pulse probability is 1-FNP.

Bulk polymerase kinetics assays. Single nucleotide incorporation kinetics was measured using rapid chemical quench flow (12) (KinTek Corporation, Austin, TX) by the addition of nucleotide and metal to preformed polymerase-DNA complexes under the following conditions: 25 nM DNA template, 75 nM DNA polymerase, 15 μ M phospholinked dNTP, 50 mM ACES buffer, pH 7.1, 75 mM potassium acetate, 0.5 mM $MnCl_2$, and 300 mM EDTA quench (final concentrations).

DNA synthesis rates were measured using a 4-methylumbelliferyl (4-MU) coumarin (Invitrogen) fluorescence-based assay (13). This real-time, steady-state DNA polymerization assay was performed as a function of the phospholinked dNTP

concentration. The data were analyzed to extract the steady-state polymerization rate. The assay utilizes DNA polymerization on a primed linear single-stranded DNA template (the same as used in experiments for Fig. 4) at five-fold molar excess of the polymerase. Two of the nucleotides (dATP and dTTP) are phospholinked with 4-MU. Incorporation of the derivatized nucleotides releases the non-fluorescent pentaphosphate 4-MU. In a coupled reaction, shrimp alkaline phosphatase (SAP, USB Corp., Cleveland, OH) quickly hydrolyzes the pendant phosphates, creating the fluorescent 7-hydroxyl methylumbelliferyl coumarin. Thus, the increase in the fluorescent signal with time is proportional to the rate of DNA polymerization. A standard curve constructed from the free fluorophore was used to convert the rate of fluorescence change into moles of nucleotide incorporated per unit time. The conditions of this assay were matched to the single molecule experiment: 5 nM template (Fig. 2 template above), 25 nM polymerase, 10 μ M of the appropriate 4-MU phospholinked dNTPs, 50 mM ACES, pH 7.1, 75 mM potassium acetate, 5 mM DTT, 0.5 mM MnCl_2 , and 0.04 U/ μ l SAP.

Capillary electrophoresis assays were performed using an Applied Biosystems 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA) with G5 dye set and standard POP-6 Polymer, 50cm array Fragment Analysis run module. Reaction conditions were set to mimic the single molecule experiments as closely as possible: ACES buffer, pH 7.1, 500 nM phospholinked dNTPs, 400 nM DNA polymerase, and 25 nM DNA template/primer (sequence as in Fig. 4). Data shown in fig. S7 are from a two minute extension reaction time point. The primer was labeled at the 5' end with FAM for detection purposes on the 3130xl. Additionally, excess dideoxy-nucleotide terminated primer was used as a DNA trap to capture any polymerase that dissociated from the target template/primer complex during extension. Extension reactions were loaded onto the 3130xl in Hi-Di™ formamide with GeneScan 600 LIZ size standard. Sizing analysis was done using Applied Biosystems GeneMapper software v. 4.0.

Supporting References

1. M. Foquet *et al.*, *J. Appl. Phys.* **103**, 034301 (2008).
2. J. Korlach *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1176 (2008).
3. D. Beckett, E. Kovaleva, P. J. Schatz, *Protein Sci.* **8**, 921 (1999).

4. A. M. Lieto, R. C. Cush, N. L. Thompson, *Biophys J* **85**, 3294 (2003).
5. M. J. Lang, P. M. Fordyce, A. M. Engh, K. C. Neuman, S. M. Block, *Nat Methods* **1**, 133 (2004).
6. J. Korlach *et al.*, *Nucleos. Nucleot. Nucleic Acids* **27**, 1072 (2008).
7. P. M. Lundquist *et al.*, *Optics Letters* **33**, 1026 (2008).
8. O. Gotoh, *J. Mol. Biol.* **162**, 705 (1982).
9. G. S. Slater, E. Birney, *BMC Bioinformatics* **6**, 31 (2005).
10. D. Gusfield, *Algorithms on strings, trees, and sequences: computer science and computational biology* (Cambridge University Press, Cambridge [England]; New York, 1997), pp. 534.
11. B. K. Teo *et al.*, paper presented at the Nuclear Science Symposium Conference Record, 2005 IEEE 2005.
12. K. A. Johnson, *Methods Enzymol.* **249**, 38 (1995).
13. M. Kozlov, V. Bergendahl, R. Burgess, A. Goldfarb, A. Mustaev, *Anal Biochem* **342**, 206 (2005).

Figure S1. Pulse width distributions for the two phospholinked dNTPs used in the two-base signature sequence detection experiment (Fig. 2).

Figure S2. Analysis of kinetic substates of DNA polymerase. (A) Sections of fluorescence time traces (20 s each) obtained from the same DNA polymerase molecule undergoing rolling circle DNA synthesis as described in Fig. 3 of the paper, exhibiting two different DNA synthesis rates of ~2 bases/s (top) and ~4 bases/s (bottom). (B) Statistical analysis of interpulse durations and pulse widths for the two different states from six DNA polymerase molecules. The arrow indicates the presence of a state characterized by short interpulse durations in the faster kinetic state.

Figure S3. Long read-length DNA polymerase activity detection. A second polymerase mutant, exhibiting approximately half the average bulk DNA synthesis rate compared to the polymerase shown in Fig. 3, was used in conjunction with a different pair of dye-labeled nucleotides, A555-dATP and A647-dGTP, and a different circular DNA template, also designed to yield alternating periods of like pulses (see supplementary information for the template sequence). (A) Representative read from a single polymerase molecule and corresponding total length of synthesized DNA (top axis). (B) DNA synthesis rate profiles for several molecules.

Figure S4. (A) Molecular structures of the phospholinked nucleotides used in this study. The molecular structure of Alexa Fluor 660 is proprietary (Invitrogen Corp., Carlsbad, CA). (B) Normalized fluorescence emission spectra of the labeled nucleotides. The two laser excitation wavelengths of 532 nm and 643 nm are indicated by arrows.

Figure S5. HPLC analysis of the four phospholinked dNTPs used in single-molecule, four-color sequencing experiments (Fig. 4). Traces are shown with absorbance detected

at maximum absorption wavelengths of the fluorescent dyes (left panels) and nucleobases (right panels).

Figure S6. Pulse detection sensitivity analysis. (A) Overlay of the predicted pulse fraction detected as a function of pulse width (black, left axis), with the A555-dATP pulse width probability distribution function (blue, right axis). (B) Zoom of the predicted pulse fraction detected as a function of pulse width, with dashed lines representing the robust standard deviation.

Figure S7. (A) Possible hairpin structures and their $\Delta\Delta G$ values (in kcal/mol) shown for 25°C, $[\text{Na}^+] = 75 \text{ mM}$, and $[\text{Mg}^{2+}] = 0.7 \text{ mM}$. The energetically most likely hairpin is #3 at position 46. (B) DNA product distributions from a bulk extension reaction stopped after two minutes, analyzed by capillary electrophoresis (CE - top panel) under equivalent reaction conditions to the single molecule experiments that produced the interpulse durations shown in the bottom panel. Polymerase pause sites corresponding to regions with predicted stable secondary structures in the template lead to accumulation of products in the electropherogram.

Figure S8. Consensus accuracy as a function of average molecular fold coverage. At least 100 sub-samples of the 449 raw reads were used at each fold coverage. Median values and 25th and 75th percentile error bars are shown.

Table S1. Pulse metrics for the two-color sequence pattern experiment shown in Fig. 2. Median values are shown for 10 individual DNA polymerases (the molecule shown in Fig. 2 constitutes the second entry in the table), as well as for the entire dataset (averages of $n = 740$ ZMWs, errors are standard deviations). Abbreviations are: PW - pulse width, and SNR - signal-to-noise ratio. Read statistics for each ZMW include the DNA synthesis rate and total errors observed, broken down to three error types.

Table S2. Pulse statistics from two-base experiments performed with linear, single-stranded DNA template and two phospholinked nucleotides (A555-dCTP and A647-

dGTP), compared with values obtained by bulk measurements with equivalent reaction conditions. (A) Median pulse widths for pH 6.5 and pH 7.1 reaction buffer. (B) Median DNA synthesis rates for 100 nM and 250 nM phospholinked nucleotide concentrations.

Table S3. Pulse and trace statistics for the 4 color DNA sequencing reads. The first three columns contain per analog mean values and standard error of the means for brightness, pulse width, and interpulse duration. Two values for the A647-dGTP are listed for the interpulse duration: inclusion of all template locations, and excluding the pause site at base 40 (see Fig. 4D and text). The last three columns list the average proportion of each error type by analog.

Movie S1. Real-time video of single molecule DNA polymerase activity measured on an array of 3,000 ZMWs. The movie first shows the layout of the ZMW array as viewed in transmission light mode before switching to fluorescence excitation. The polymerization reaction is initiated after 1 second. The ZMW depicted in Fig. 2 of the paper is marked with a red rectangle. Scale bar = 10 μm .

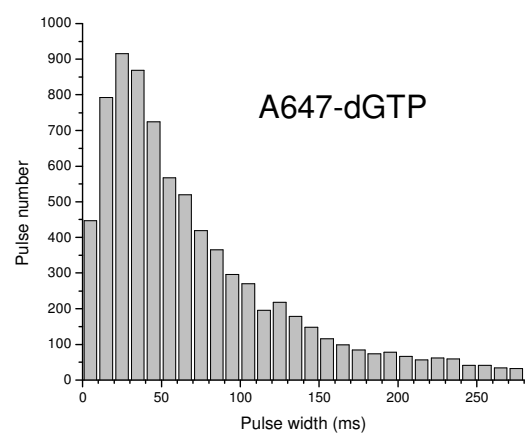
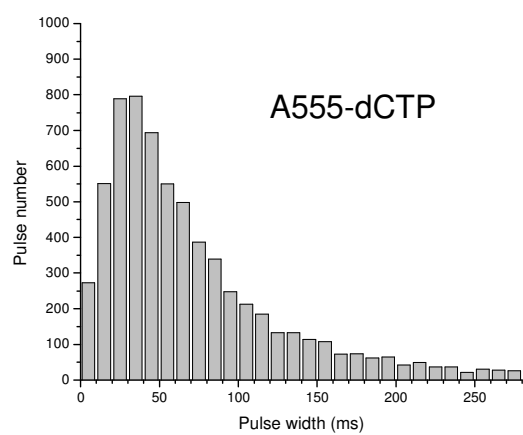
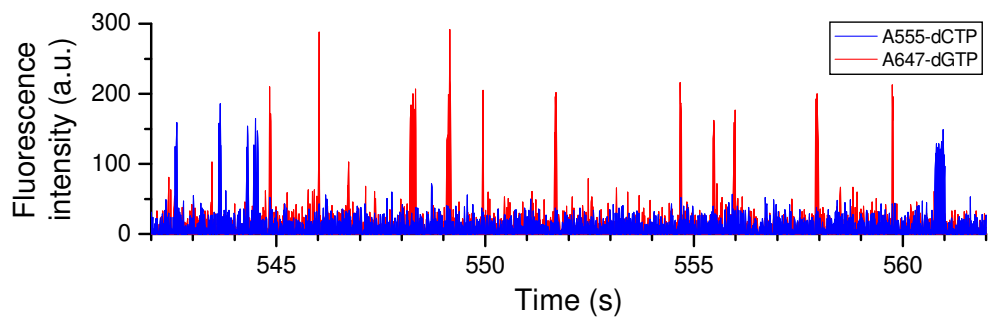


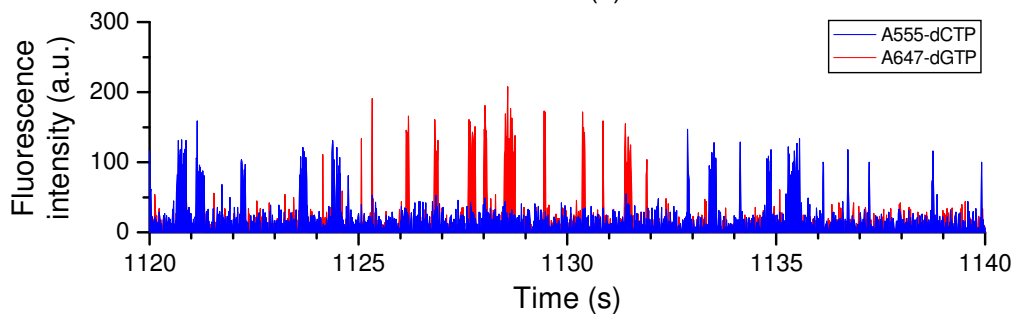
Figure S1

A

~2 bases/s:



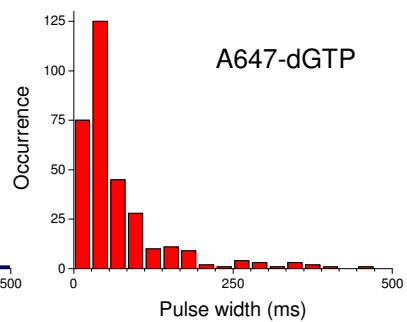
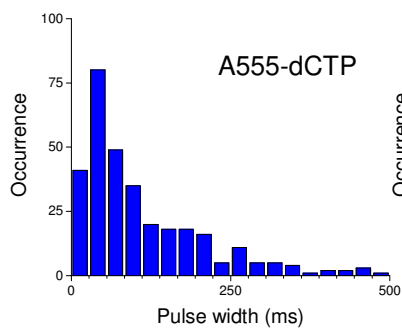
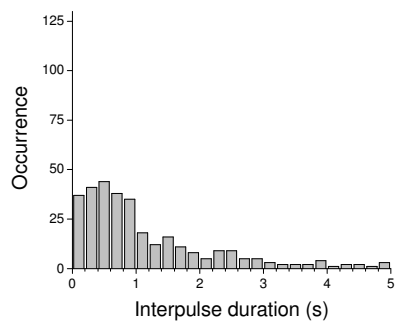
~4 bases/s:

**B**DNA
synthesis
rate

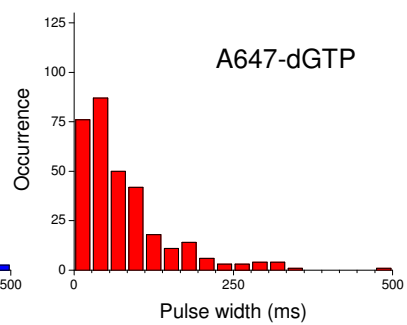
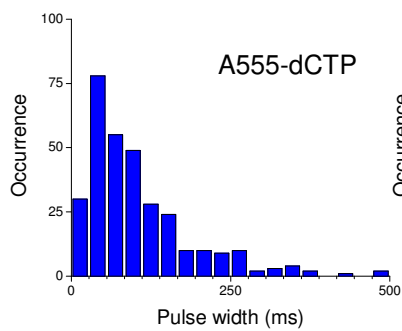
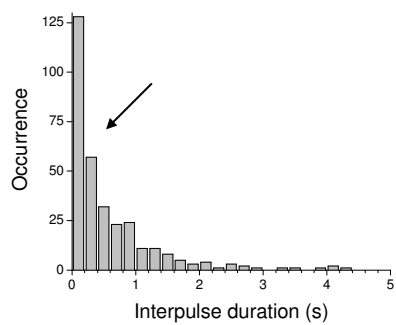
Interpulse duration

Pulse widths

~2 bases/s



~4 bases/s

**Figure S2**

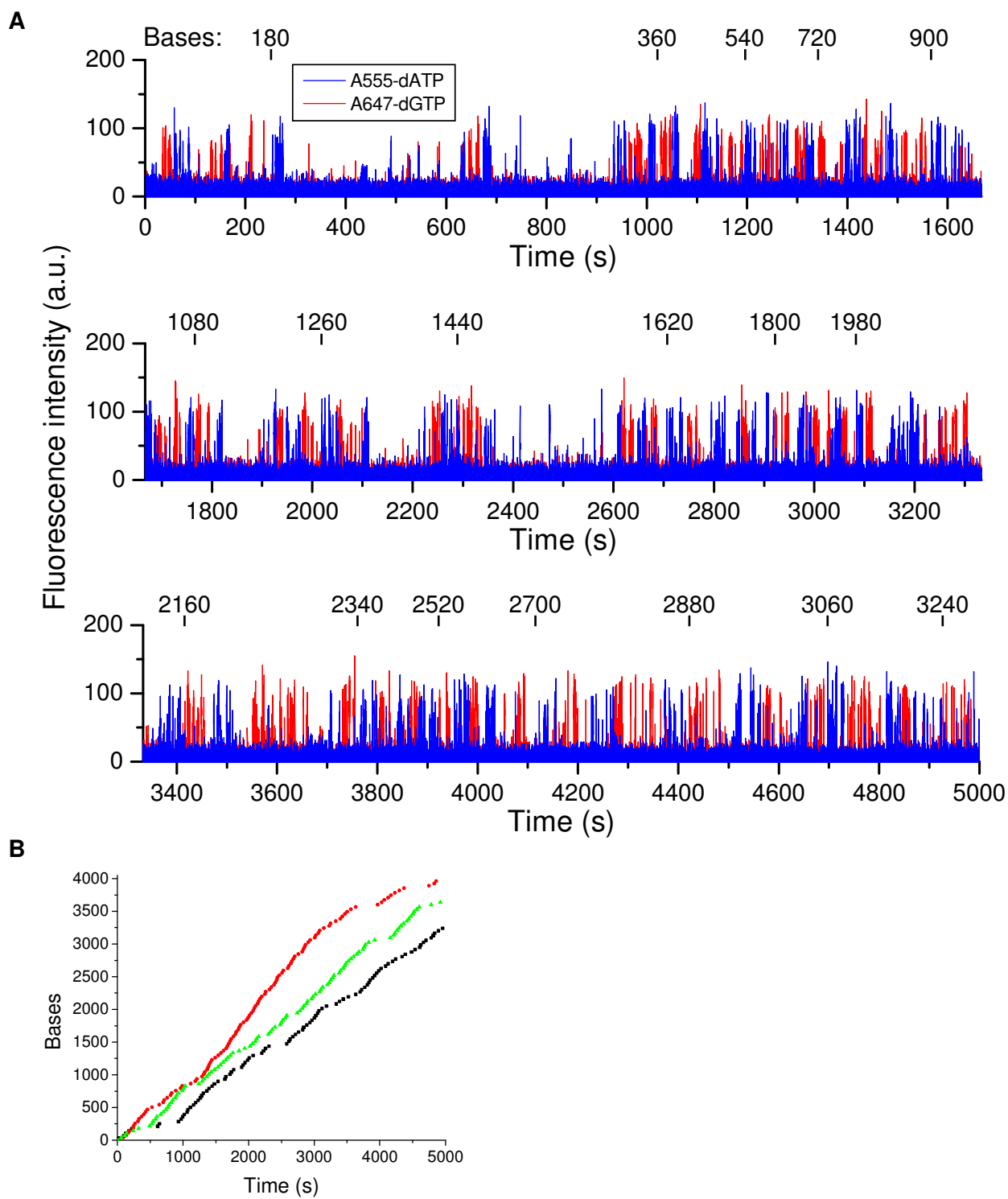
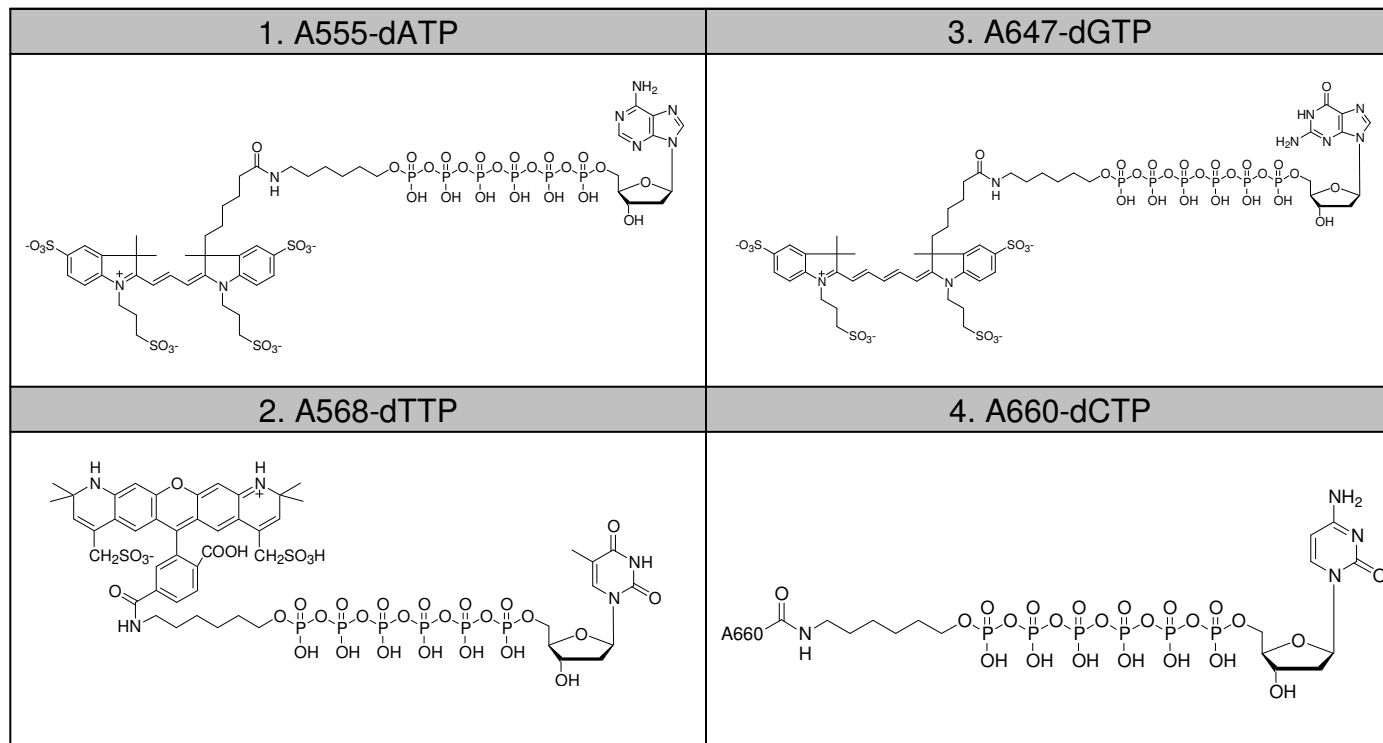
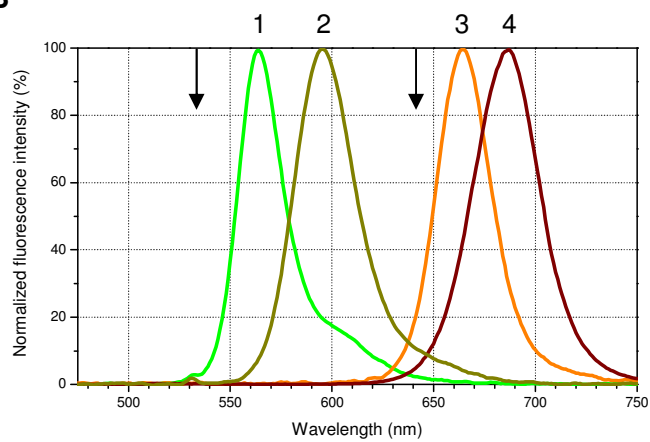
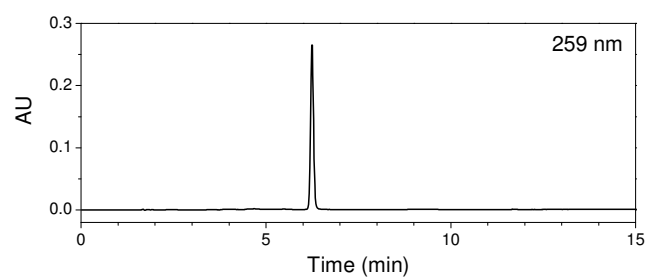
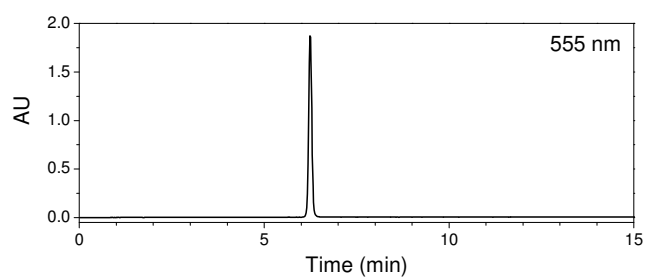


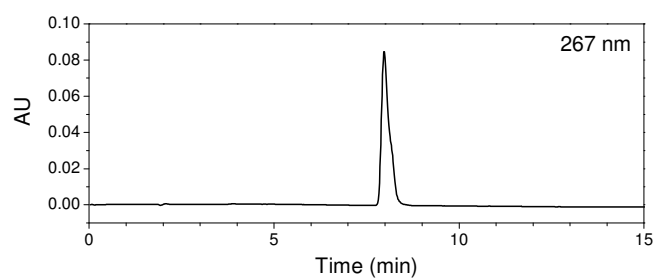
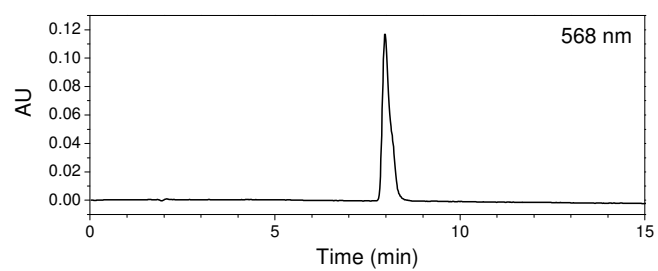
Figure S3

A**B****Figure S4**

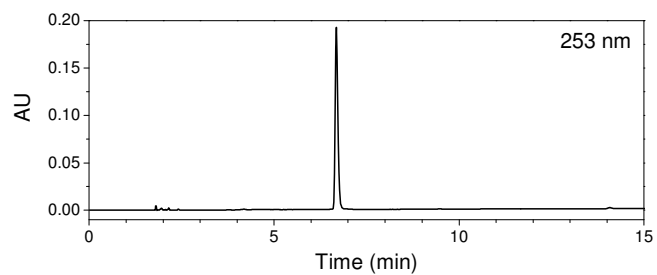
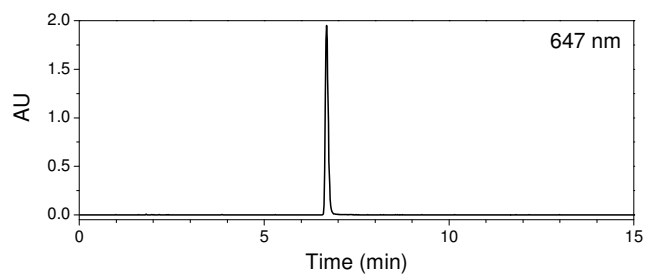
A555-dATP:



A568-dTTP:



A647-dGTP:



A660-dCTP:

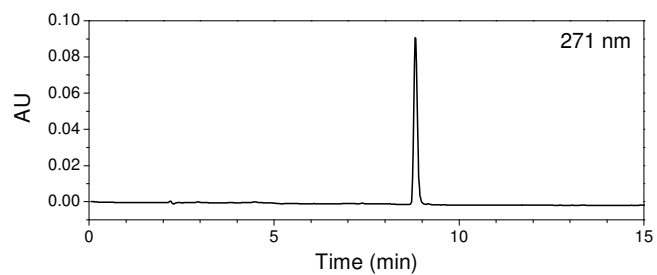
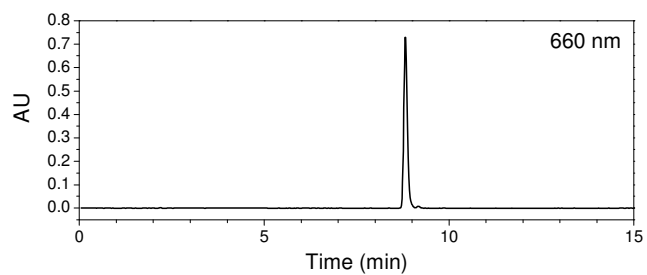
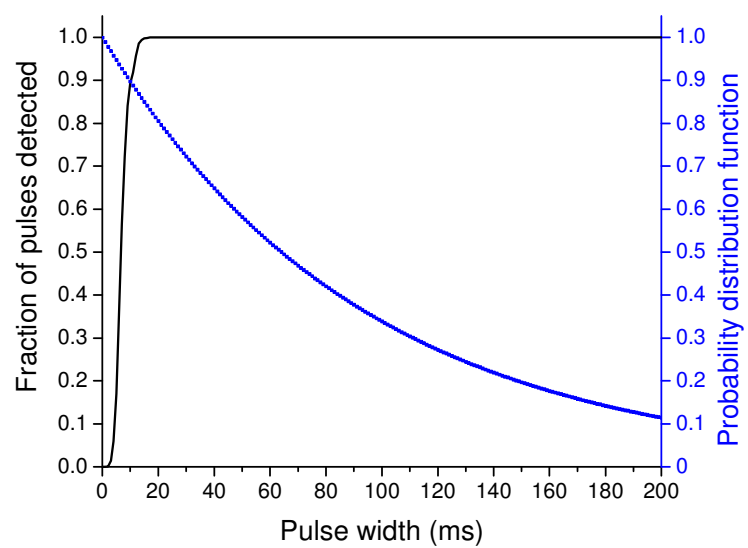
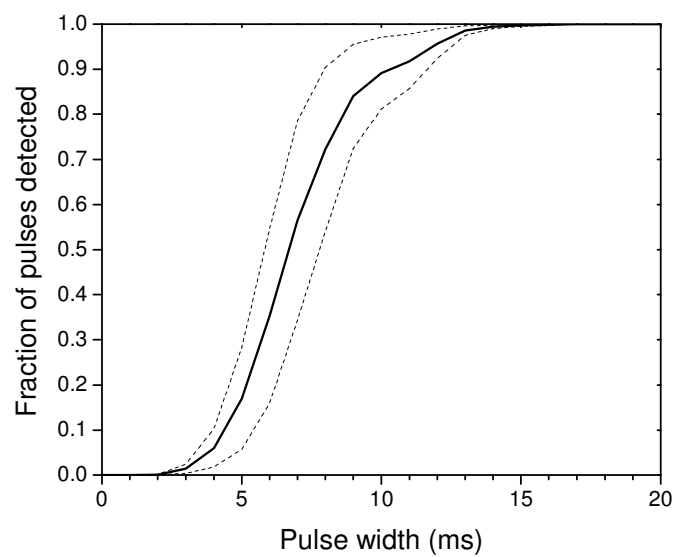


Figure S5

A**B****Figure S6**

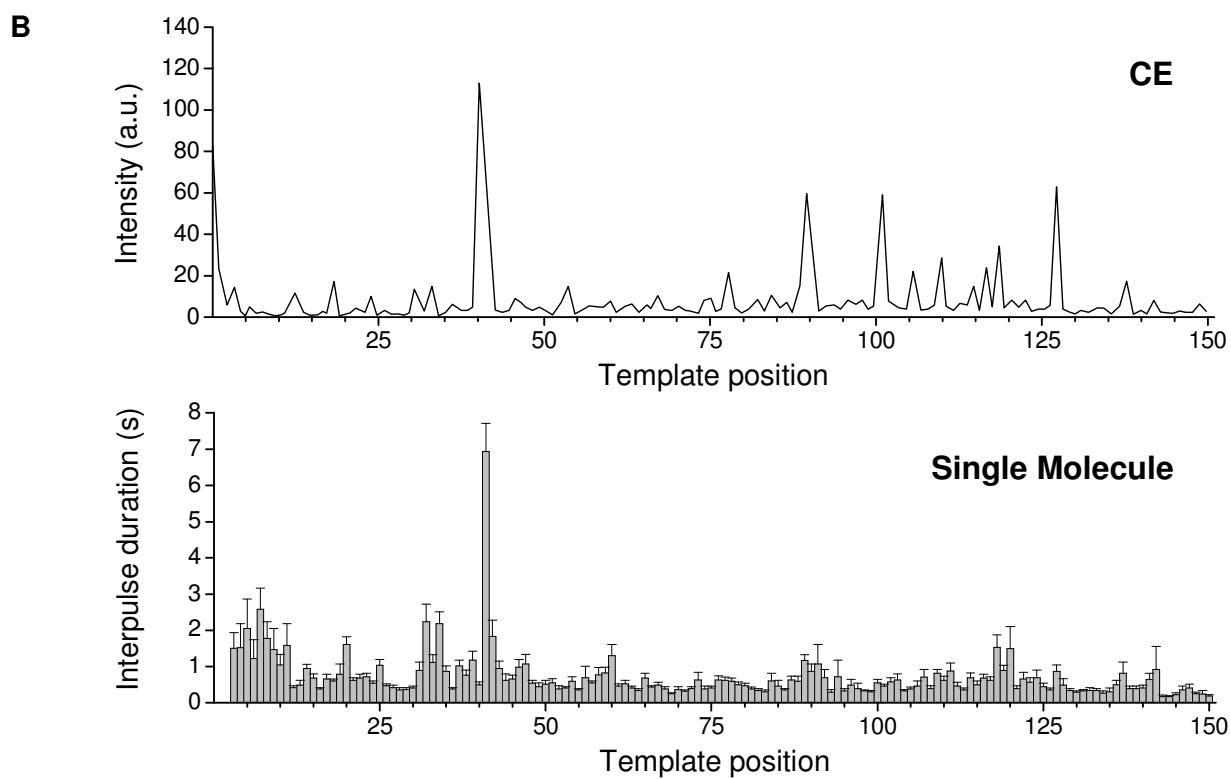
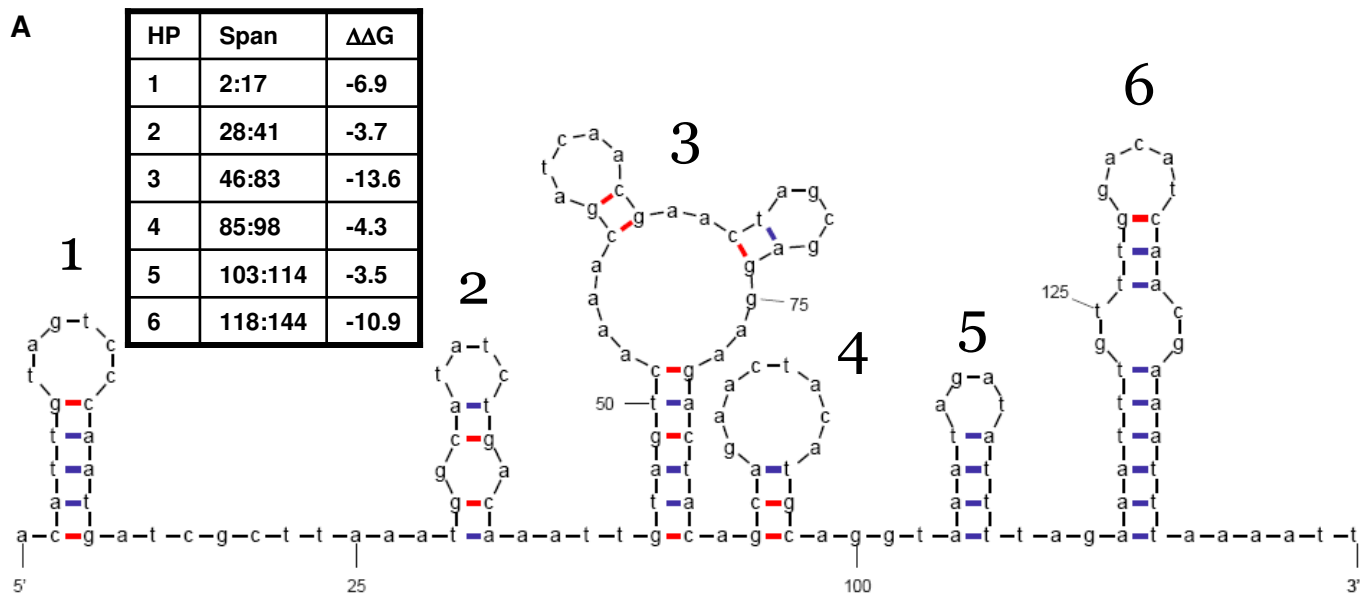


Figure S7

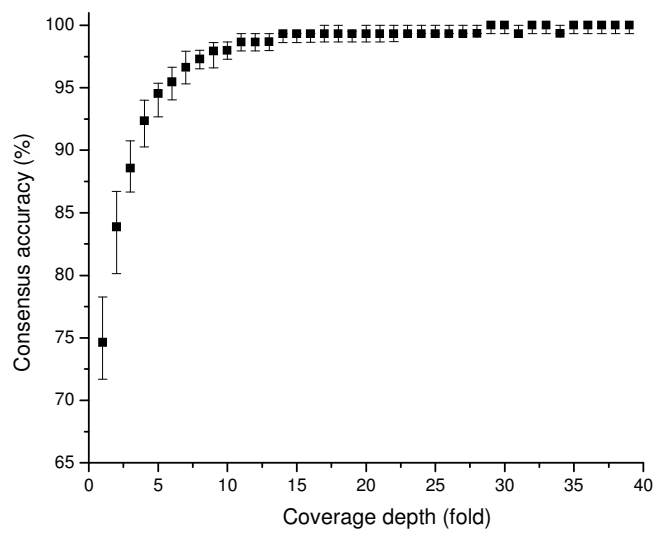


Figure S8

ZMW	A555-dCTP			A647-dGTP			Errors			Synthesis rate (bases/s)
	PW (ms)	Brightness (photons/s)	SNR	PW (ms)	Brightness (photons/s)	SNR	Mismatches	Insertions	Deletions	
1	60	4457	16	80	4274	14	2	2	1	2.9
2	60	8016	26	40	13361	30	2	3	0	4.7
3	65	8435	24	50	4937	14	1	2	2	4.5
4	80	9091	28	40	5678	12	2	3	0	2.5
5	65	6014	21	50	4176	12	1	3	0	3.4
6	35	9167	22	50	6042	17	0	1	3	5.1
7	55	4073	13	50	11258	24	3	2	0	4.0
8	50	6068	16	55	5445	16	2	2	2	3.0
9	40	7349	19	60	5728	17	0	0	4	5.7
10	90	7331	26	45	8512	21	2	1	0	5.5
All (n=740)	77 ± 30	7353 ± 2970	24 ± 10	70 ± 27	8408 ± 3381	25 ± 10	0.6 ± 0.5	1.3 ± 0.8	2.1 ± 1.0	4.7 ± 1.7

Table S1

A		Single molecule	Bulk
		Pulse width (ms)	Incorporation time (ms)
pH 6.5	A555-dCTP	115 ± 6	250 ± 21
	A647-dGTP	168 ± 5	211 ± 13
pH 7.1	A555-dCTP	94 ± 14	72 ± 5
	A647-dGTP	126 ± 10	55 ± 4

B		Single molecule	Bulk
		DNA synthesis rate (bases/s)	
100 nM		1.40 ± 0.03	1.10 ± 0.04
250 nM		3.75 ± 0.10	2.20 ± 0.10

Table S2

	Brightness (photons/s)	Pulse width (ms)	Interpulse duration (ms)	Fractional errors by analog (%)		
				Mismatches	Insertions	Deletions
A555-dATP	6446 ± 109	133 ± 22	770 ± 250	20	22	12
A568-dTTP	2781 ± 39	91 ± 13	670 ± 220	20	19	27
A647-dGTP	4865 ± 92	117 ± 14	960 ± 210 (650 ± 180)	16	36	19
A660-dCTP	2691 ± 41	96 ± 10	790 ± 230	43	23	42
Total error:				100%	100%	100%

Table S3