Engineering an Enzyme for Direct Electrical Monitoring of Activity

Bintian Zhang[1,¶], Hanqing Deng[2,¶], Sohini Mukherjee[2], Weisi Song[1], Xu Wang[2] and Stuart Lindsay[1,2,3,*]

[1]Biodesign Institute, [2] School of Molecular Sciences, [3]Department of Physics, Arizona State University, Tempe, AZ 85287.

ABSTRACT

Proteins have been shown to be electrically-conductive if tethered to an electrode by means of a specific binding agent, opening the possibility of building electronic devices that exploit the remarkable chemical versatility of enzymes by direct read-out of activity. Single-molecule conductance increases tenfold if two specific contacts are made (as is possible with bivalent antibodies). Here, we address the problem of forming contacts with proteins that, unlike antibodies, do not possess a multiplicity of native binding sites. In particular, we have engineered contact points into a Φ29 polymerase, with the points chosen so as to leave the active site free. Contacts are formed by introducing biotinylatable peptide sequences into Φ29 and functionalizing electrodes with streptavidin. Φ29 connected by one biotinylated contact and one non-specific contact shows rapid small fluctuations in current when activated. Signals are greatly enhanced with two specific contacts, a feature in the conductance distribution changing by > 9 nS between the open and closed conformations of the polymerase. Polymerase activity is accompanied by large changes in conductance over the millisecond timescale.

Corresponding Author: *E-mail: stuart.lindsay@asu.edu.

KEYWORDS: Bioelectronics, polymerase activity, single molecule conductance

Proteins are widely assumed to be insulators. Reports of metallic conduction in bacterial wires[1] and long range conduction in protein multilayers[2] were thought to be exceptions. We recently demonstrated that a number of proteins, chosen only for their *redox inactivity* (i.e. no electron transport function, no redox active centers) conduct very well if contacted by binding agents that can inject charge carriers into their interiors. This conductance is limited to electrons (or holes) because the proteins we have studied to date act to passivate electrodes in cyclic voltammetry studies. Furthermore, the internal conductance of the protein is sensitive to conformational changes.[3] Key to these experiments was the use of antibodies together with electrodes that were functionalized with an epitope to which the two binding domains of the antibody could bind. We measured the spread of conductances (owing to the range of contact geometries) by recording current-voltage (IV) curves from ~ 1000 molecules, and plotting histograms of the conductances derived from the IV curves. Distributions for antibodies had two peaks, one at ~0.2 nS (similar to the value of the single peak observed for proteins bound by only one ligand) with a second peak at ~ 10x the conductance (~2nS). We interpreted the second peak as owing to specific binding events at each binding domain of the antibody, attached so that the antibody bridges the electrode gap. This interpretation was confirmed using Fab fragments (one binding head) or functionalizing only one of the two electrodes with an epitope.[3] We also found that the internal conductance of proteins contacted via specific bindings agent is higher than the conductance associated with contact regions, so that that the measured electronic properties do not change much with gap size, until contact is lost with the protein.[3, 4] Proteins are remarkably versatile molecular machines, capable of molecular recognition, highly selective catalysis, directional energy transfer, directed polymer synthesis and many other functions, so integration of proteins into bioelectronic devices has been a long sought-after goal,[5] and here we address the problem of making specific contacts to proteins that do not have multiple ligand binding sites, with the goal of demonstrating a device that transduces enzyme activity into electrical signals directly.

We chose as our test case Φ29 polymerase. It is a preferred polymerase for DNA sequencing applications because it is highly processive, has a high strand displacement activity (needing only a nick to initiate polymerization) and is very accurate (1 part in 10^6).[6, 7] It is also fast, turning over 100 nucleotides per second at saturating concentrations of nucleotidetriphosphates (dNTP) and template.[8] The major motion of the enzyme involves the closing of a "finger" domain into the "palm" and "thumb" once template and matching dNTP are assembled and ready for incorporation. At some point over the open-closed-open transition, the enzyme translocates the template so that the next available site is positioned for binding of the complementary dNTP.[9, 10]

Here, we report the engineering of specific binding sites for electrical contacts into Φ29 polymerase using insertion of the Avitag sequence into a clone for the polymerase, and subsequently biotinylating the lysine within the Avitag peptide sequence using the BirA enzyme.[11, 12] This then allows for strong, and specific binding to Streptavidin. We initially tested a Φ29 biotinylated just at the N terminus (Gen I), and bound to electrodes that had been covered with thiolated Streptavidin. Biotin-bound Streptavidin makes an excellent molecular wire[3] and also serves to keep the Φ29 (with its seven surface cysteines) away from the metal electrodes. We introduced a second contact about 5 nm distant from the first (Gen II), at a site in the (inactivated) exonuclease domain. A third type of polymerase (Gen III) was made by incorporating a flexible linker next to the N-terminal Avitag. Finally, we compared direct tethering of a thiolated Streptavidin to the electrodes, with indirect tethering, using a thiolated biotin molecule to functionalize the electrodes. We measured both conductance distributions (by

recording current-voltage, IV, curves) and dynamic responses by recording current vs. time (I(t)) for bound polymerases in a gap of nominally constant size. The introduction of two binding sites introduced a new high conductance feature in the conductance distribution, analogous to the second peak found naturally for antibodies with two binding domains. We also found that the indirect connection of streptavidin via thiolated biotin gives higher conductance than direct connection via surface thiolation of the streptavidin. The conductance of the Φ29 was found to increase greatly in the closed conformation. The change in DC conductance measured the between open and closed states is consistent with the scale of large dynamic swings in the current recorded through the enzyme as a function of time when the enzyme is active.

RESULTS AND DISCUSSION

Recombinant Φ29 DNA polymerase constructs with inserted Avitag

The starting enzyme was a Φ29 DNA polymerase, rendered exonuclease-deficient with D12A and D66A mutations. A Q5 site-directed mutagenesis kit (NEB) was used to insert the Avitag DNA sequence into a pET15b plasmid containing the mutant polymerase gene. The equivalent inserted peptide sequence is shown in blue, with flanking linker sequences in yellow, in Figure 1. The epsilon-amine of the central lysine (K, marked by red arrow on Fig. 1) was biotinylated using the BirA enzyme.[11] Three generations of modified enzyme were tested. The first (Gen 1) was biotinylated only at the N terminus of the protein. The second (Gen II) contained a second Avitag ~5nm from the N terminus between E279 and D280. This second site is located in the deactivated exonuclease domain and was chosen because its position does not change with respect to the N terminus over the open to closed transition.[9, 10] The third (Gen III) contained an additional flexible linking sequence (GNSTNGTSNGSS) adjacent to the N-terminal Avitag to allow for greater flexibility in contact geometry. Biotinylation was verified with a SDS-PAGE gel analysis of the free- and streptavidin bound polymerases. Figure 1b shows the increase in molecular weight that occurs as two streptavidins are bound to the Gen III polymerase (yet higher molecular weight features are likely polymer aggregates of alternating polymerase and streptavidin). The activity of the modified polymerases was verified in vitro using a rolling circle amplification assay (Figure 1c).

Scanning Tunneling Microscope Conductance Measurements.

Our measurements were made using an electrochemical scanning tunneling microscope (Pico STM, Agilent) with insulated palladium (Pd) probes[13] and a Pd substrate, both held under potential control using a salt-bridged reference electrode (Fig. 2a). Electrodes were modified with either a thiolated biotin (SH-biotin) or thiolated streptavidin, and then incubated with a solution of the biotinylated polymerase (Methods). Measurements were made in a reaction buffer containing $MgCl_2$ and tris(2-carboxyethyl)phosphine) (TCEP) to prevent polymerase oxidation. Nucleotide triphosphates were added to activate the polymerases. Current-voltage (IV) characteristics were measured using a fixed gap (no servo control). The gap remained constant to within about 0.1nm over ~ 1 minute, and the bias was swept between -0.2 and +0.2V and back again at a rate of 1V/s. No attempt was made to fish for proteins or to push or pull them. Thus the trapped proteins that yielded signals were molecules that bound across the gap in equilibrium, or near equilibrium conformations. After 1 minute, the gap was returned to the set-point value. This cycle was repeated to obtain further IV sweeps. 80% of these sweeps

reproduced exactly on reversing the sweep direction. (In a number of runs, we recorded the current before re-engaging the servo to check that there had been no significant drift.)

The procedure differs from that of our earlier study of bound proteins[3] in that both probe and substrate are functionalized with streptavidin. In consequence, two streptavidin molecules could contact each other across the gap. However, it appears that this non-bonded contact between two proteins does not conduct, as evidenced by frequent crashes in which the probe is physically damaged on approach to the surface. When contacts are successfully made with streptavidin, they recapitulate the distribution observed for single streptavidin molecules (Figure S1) and no signals are observed from streptavidin alone for gaps larger than 3.5 nm. However, when biotinylated polymerase is allowed to complex with the streptavidin, significant currents are measured out to gap distances >4.5 nm with conductance distributions that are quite distinct from that of streptavidin. The data reported in this paper were obtained with a gap of 4.5 nm unless noted otherwise.

Figure 2b shows a selection of IV curves for a Gen I and Gen II polymerase. Forward and reverse sweeps are superimposed. As observed for several other proteins,[3] the response is linear, with the exception of noise spikes that appear above ±100mV as a consequence of voltage-induced contact fluctuations. The slope of each trace was used to calculate a conductance, G=I/V and histograms of conductance (based on ~ 1000 measurements each) are shown for the three generations of polymerase in Figure 3.

The Gen I polymerase forms only one specific contact, but in contrast to what we have observed in the past for proteins with a single binding site,[3] there are two peaks in the conductance distribution, one at about 0.2 nS (characteristic of a weak, non-specific binding) and a second peak at ~ 1 nS. We speculate that the second peak is a consequence of additional surface linkages owing to the seven surface cysteines in this polymerase (conductance via a thiolated streptavidin is highly unlikely at this gap size - Figure S2). This second peak only appears when a polymerase is present, and the current distribution is quite different from that measured through streptavidin alone at a smaller gap size (Figure S1). As observed for other proteins, these distributions change very little with gap size until contact with the protein is lost (Figure S2). Figure 3d shows the conductance distribution measured for a Gen II with thiolated streptavidin on the surface but with a bare (unfunctionalized) probe. The 3 nS peak observed in Figure 3b is now absent. Taken together, these experiments show that an additional high conductance feature occurs in the conductance distribution when (a) the polymerase has two biotinylated contacts and (b) streptavidin is present on both the probe and substrate. We therefore attribute this highest conductance peak to biotin-mediated binding events that bridge the electrode gap.

Figure 3c show a conductance histogram for a Gen III polymerase, using a thiolated biotin molecule to attach Streptavidin to the electrode surface. The highest conductance feature has moved out to 5.6 nS. This is a consequence of the biotin-mediated attachment of the streptavidin to the electrodes. Figures S1c and d show conductance distributions for the Gen III polymerase as connected by thiolated streptavidin (S1c - larger peak = 3.7 nS), or wild type (WT) streptavidin bound to biotinylated electrodes (S1d - peak 3 = 7nS) showing the increase in conductance when a specific ligand is used for a contact, as previously reported.[3] In additon, we conclude that the introduction of the additional flexible linker has reduced the overall conductance of the complex a little for a given coupling chemistry (5.6 nS for Gen III - Figure 3c, vs 7 nS for Gen II - Figure S1d - with biotin mediated coupling to the electrodes in both cases).

These data demonstrate the successful engineering of a bis-functionalized polymerase that retains polymerase activity while showing increased electrical conductivity.

We turn now to the conformational sensitivity of the electronic signals. We had shown earlier that ligand binding changes the conductance distribution of a streptavidin molecule significantly: is the same true of a polymerase? In order to probe this, we generated a Gen III polymerase stabilized in the closed form by incubating a template-bound polymerase with nonhydrolzable nucleotide triphosphates (NH-dNTPs). The NH-dNTP binds the template-polymerase complex, closing the polymerase, which then remains in the closed configuration because of the non-hydrolyzable triphosphate.[14] Conductance distributions for the normal, open (-dNTP) and closed (+NHdNTP) forms are shown in Figures 4a and b. There is a very large change in the highest conductance feature (from 5.6 to 15 nS). The highest conductance feature is quite small, so the experiment was repeated to verify its existence (Figure S3). Interestingly, IV curves taken on the active form (using normal dNTPs) yield an intermediate conductance (Figure S4), implying that an active polymerase may stay in a high conductance state for long enough to yield increased conductance in an IV measurement (which takes 0.4 s). Taken together, these measurements indicate that the conductance distribution is affected by the conformation of the polymerase. One interesting consequence of measuring IV curves on an active enzyme (as in Figure S4) is that noise now appears on the curves in the usually quiet region below 100mV (Figure S5).

Taken together, these observations suggest that it may be possible to observe signs of polymerase activity in a recording of current taken over time (I(t)), as the polymerase incorporates incoming nucleotides. We recorded I(t) at a constant 50mV bias (below the threshold for contact noise) for a single stranded template primed by a 15 base pair hairpin (Methods). To do this, we followed the same procedure outline above for obtaining IV curves, but holding the bias constant, and recording current for 90s before withdrawing the probe and reestablishing servo control. We started with a gap of 2.5 nm under servo control, increasing the gap to 6 nm, and then brought the tip down to 4.5 nm and then recorded current for 60 s. Typically, no current was recorded for the first 10-20 s, after which a contact formed and an I(t) curve was obtained. Contacts were formed with molecules in >50% of these "fishing" attempts. The currents jumped suddenly on contact with the molecule, but then changes substantially as the contact point drifts. A typical current-time trace is shown in Figure 5a. By comparing the distribution of currents measured in current-time traces with the conductance distributions obtained from many repeated contacts with the molecule, we have shown that the variation in the current-time trace can be accounted for by variations in the contact point.[3]

Noise Analysis.

We first studied the Gen I polymerase with a single contact. Bursts of noise were clearly visible when the polymerase was activated on addition of the complementary dNTPs (Figure S6). However, currents were generally low (c.f., the conductance distribution in Figure 3a) as were the signal levels. It was for this reason that we constructed the bis-functionalized polymerases that yield the higher conductance features seen in Figure 3b and c. Figure 5a shows a typical I(t) trace for an $d(A)_{10}$ single stranded template in the presence of all four dNTPs (black trace, mainly obscured by the overlaid red trace) in the polymerization buffer (Methods). Large changes in the background current obscure the fluctuations, so we removed the background using an asymmetric least squares (ALS) fit.[15] The ALS accurately follows the background without distorting the noise signals (Figure S7). (Attempts to remove the background with a

median filter distorted the noise signals.) The ALS fit to the raw data (red trace) is superimposed on it in Figure 5a. The flattened noise trace, obtained by subtracting the ALS fit from the raw data, is shown in Figure 5b. The noise signals are quite large, in excess of a nA in this case, and occur in bursts with pauses in between them, though no discernable repeat pattern was found. We have selected a control signal of similar overall DC conductance (Figure 5c), taken in a buffer containing only dATP, dCTP and dGTP, lacking the necessary dTTP. Following the same subtraction procedure outlined for the active molecule, a noise trace is obtained, as shown in Figure 5d. The result is clearly not random noise, because it coincides exactly with the period for which the molecule is contacted. The same behavior was seen in eight other control experiments. So clearly, both active and inactive polymerases produce electrical noise. Given the variable background signal and the stochastic nature of the fluctuations, can we distinguish active from inactive polymerase? Inspection of the details of the noise signals at a higher time resolution offers a clue. Figure 5e shows a trace containing both large (LF) and small (SF) fluctuations. If we compare the size of the fluctuations from the control experiment (-dTTP) with the size distribution of the fluctuations for the active polymerase (+dTTP) (measured at about the same background current) we see that the current fluctuations in the active polymerase are much larger (Figure 5f). We chose to compare runs of similar background current, because inspection of many traces indicates that the size of the current fluctuations increases with the background current.

In order to quantify the relationship between the size of the fluctuations and the background current, we studied the distributions of both for a number of molecules with, and without dTTP present in the buffer solution. For each molecule, we binned the ALS fitted baseline currents as shown by the examples in Figures 6a,c. Many of these distributions could be fitted by a Gaussian (red curve). Many could not - for example the background can jump between two or three levels. In these cases we fitted the largest peak that was clear of the background. The peak of the fitted Gaussian, I_p, was then used to characterize the baseline for that run. Example of the binned noise signals are given in in Figures 6b,d. To characterize these we used a double exponential distribution:

$$N(i) = A_1 \exp\left(-\frac{i}{i_S}\right) + A_2 \exp\left(-\frac{i}{i_L}\right)$$

where *i* is the current in a given bin of the histogramed current distribution. In the case of the experiments in which dTTP was withheld (e.g., Figure 6b), the fits all converged to a single exponential ($i_s = i_L$). For recordings with dTTP present, most of the fits converged on the double exponential distribution with $i_s < i_L$ (4 of 13 molecules showed the small peaks only). The results are summarized in Figure 6e. Activated molecules (+dTTP) showed both large (red) and small (blue) fluctuations. The controls (-dTTP) showed only small fluctuations (green), essentially equal to the small fluctuations also seen in the active polymerases. There is an approximately linear relation between background current and fluctuation amplitude as shown by the three linear fits. For the large fluctuations, characteristic of active polymerases, $i_L = (0.25 \pm 0.026) i_p$ (*i* is the 1/e value of the fit, so larger and smaller spikes are observed also). For the small fluctuations, present in both active an inactive polymerases, $i_S = (0.06 \pm 0.01) i_p$. Thus the active state can be identified by the presence of fluctuations that are about 25% of the baseline current, while fluctuations in the inactive state are about 10% of the baseline current. Not all polymerase molecules contacted were active as indicated by the lack of large

fluctuations in 4 of the 13 molecules studied. On the other hand, none of the eight -dTTP control runs showed large fluctuations (Table 1).

This analysis was repeated using data obtained in 38 runs with a $d(ATC)_5$ template and 25 runs with a $d(C)_{10}$ template. The results are summarized in Figures S10a and b. The fitted amplitude distributions for the large fluctuations (LF) show considerable variation, but the trends observed for $d(A)_{10}$ (Figure 6) are reproduced well with $i_L = 0.27(\pm 0.03)I_p$ and $i_s = 0.04(\pm 0.01)I_p$ for $d(ATC)_5$ and $i_L = 0.32(\pm 0.03)I_p$ and $i_S = 0.05(\pm 0.007)I_p$ for $d(C)_{10}$.

In order to confirm the association of large fluctuations (by which we now mean 25%, or more, of the baseline current) with polymerase activity we carried out experiments with different templates in different conditions. In each case, the DNA template consisted of a 15 base double-stranded region connected by a 9 base loop (Methods and SI) with the ssDNA template extending from the 5' end of the hairpin. The ssDNA templates were $(dA)_{10}$, $(dC)_{10}$, $(dA)_{10}(dC)_{10}$ and $(dTAC)_5$. Positive experiments were carried out in the reaction buffer (1mM phosphate buffer, pH=7.4, 4 mM TCEP, 10mM $MgCl_2$ with 1 mM dNTPs and 1μM template). In the control experiments, one critical ingredient was withheld as listed below in Table 1 (where they are labeled as control, C). In addition, we carried out measurements using nonhydrolyzable dNTPs. Each of the runs was analyzed as described above for $(dA)_{10}$ ± dTTP.

Experimental Conditions	Control/Active	Fraction Large Fluctuations
$(dA)_{10}$ +dNTPs +Mg^{2+}	A	0.69
$(dA)_{10}$ +dATP, + dCTP, +dGTP, -dTTP, + Mg^{2+}	C	0
$(dC)_{10}$ +dNTPs +Mg^{2+}	A	0.52
$(dA)_{10}(dC)_{10}$ +dNTPs +Mg^{2+}	A	0.44
$(dATC)_5$ +dNTPs +Mg^{2+}	A	0.52
$(dA)_{10}(dC)_{10}$ - dNTPs + Mg^{2+}	C	0
$(dA)_{10}(dC)_{10}$ +dNTPs - Mg^{2+}	C	0
$(dA)_{10}(dC)_{10}$ +dTTP -dGTP, + Mg^{2+}	C	0.32
$(dA)_{10}(dC)_{10}$ NH-dNTPs +Mg^{2+}	C	0
No template +dNTPs +Mg^{2+}	C	0

Table 1: Occurrence of large fluctuations (fraction of measured molecules) for various experimental conditions. NH = non-hydrolyzable dNTPs. Large fluctuations are identified by a machine fit converging on a two component distribution of fluctuations.

Clearly, withholding any one of the ingredients critical to polymerase function abolishes the large fluctuations. An interesting exception was $(dA)_{10}(dC)_{10}$ in the presence of dTTP only. One expects that the polymerase would reach the end of the A tract and then stall for want of the missing dGTP nucleotide, but that is not the case. In order to investigate the behavior of the polymerase further, we ran denaturing gels of the polymerization reaction for all four templates. Because the primer is terminated in a hairpin (Figure S10), the denatured polymerization product produces a higher molecular weight feature on the gel, with no feature at the molecular weight of the primer if the reaction goes to completion. The results (Figure S8) are quite striking.

Though single-stranded homopolymers have been used for studies of a different polymerase motion successfully in the past[16] it is known that polymerases tend to stall as they encounter secondary structure in a single-stranded DNA template.[17] The gels show that the only template that is fully extended is $d(ATC)_5$. $d(A)_{10}$ is fully extended in a significant fraction of events. $d(C)_{10}$ does not appear to be extended at all, which is quite surprising because the noise signals from this template are quite dense (Figure S9). The $d(A)_{10}d(C)_{10}$ is extended in a fraction of cases, but only to the end of the A tract. Thus it is likely that the polymerase falls off this template, allowing for polymerization to start again in the presence dTTP only, accounting for the observation of activity in the absence of dGTP (Table 1). Longer pauses are observed in this experiment as might be consistent with the polymerase falling off and having to bind a new template.

Finally, we examined some of the details of the noise for two different sequences to investigate whether any of the noise features changed as the template sequence was changed. We cannot compare homopolymer templates because of the poor activity of the enzyme on C tracts, so we have compared $d(A)_{10}$ and $d(ATC)_5$. Figure 7 shows raw data for $d(A)_{10}$ (a) and $d(ATC)_5$ (e) with the extracted noise signature plotted below in blue. Although the flattening algorithm does not distort the signal significantly, the amplitude of fluctuations is affected by changes in the baseline current, as discussed above. For this reason, we searched for regions where the baseline was constant and the noise signal is dense. Two representative samples are shown in Figures 7 b and f. The difference in these noise signals is striking. All of the runs like this in the A or C homopolymers had significant amplitude fluctuations. Since the baseline current (and therefore the contact) was constant in the regions we examined, these fluctuations must reflect the behavior of the polymerase. On the other hand, the ATC polymer signals have many regions where the fluctuation amplitude (at constant baseline current) was constant, for times of up to 0.5 s, the longest constant baseline signal obtained. This difference may be related to the difference in processivity observed in the biochemical measurements (Figure S8). It is striking that signals obtained from the $d(C)_{10}$ template are similar to those from $d(A)_{10}$ (Figure S9) despite the fact that $d(C)_{10}$ is barely extended at all. So the rather chaotic signals observed in both A and C homopolymers may reflect stochastic motion of the polymerase on these templates. In contrast, the readily processed $d(ATC)_5$ yields regions of steady signal that might prove easier to interpret. The peak widths (Figures 7c and g) and the distribution of intervals between pulses (Figures 7d and h) show that there are some differences in the peak characteristics and peak frequencies between the two templates.

CONCLUSIONS

Engineering two contact points into a polymerase yields features in the conductance distribution that are approximately 3 to 10 times larger than those observed with only one engineered contact and a second, non-specific, contact. The conductance of the complex of streptavidin and doubly biotinylated Φ29 is further increased if biotin is used to anchor the streptavidin to the electrodes in place of the usual anchoring by means of thiolating surface lysines. There are significant changes in the conductivity as polymerase Φ29 undergoes the open to closed transition and these changes can be detected dynamically as the occurrence of rapid noise spikes that have an amplitude of about 25% (or more) of the background current. There are similar, but much smaller, signals generated when the polymerase is not active. These small signals are also proportional to the background current (~6%) and are present in both resting

and active polymerase. This observation of a proportionality between the signal amplitude and the background current has implications for models of the mechanism of signal generation. Measurements of the electronic decay length in protein molecules[4] show that the overall conductance has a large contribution from a small contact conductance, which limits the overall conductance significantly. The measurements of electronic decay length also show that the decay length is on the order of the protein size, so that it is impossible to distinguish between a slow exponential decay and a resistive model in which the decay of current scales with 1/(length). The two models predict different dependencies of the measured current on internal changes in the molecule. In a resistor model, where the contact resistance is R_C , the internal molecular resistance is R and the change associated with a conformational transition is δR, the fractional change in current on undergoing the transition is $\delta i/i = \delta R/(R_C + R + \delta R)$. In a model based on transition probability, the current fluctuation is directly proportional to the change associated with the conformational transmission multiplied by the probability of charge injection, so again proportional to the baseline current. However, the resistive model implies a non-linear relationship between δi and i because of the dependence of the slope on R_c. We do have some indication of the relative contributions of the contact resistance and internal resistance in these measurements[4] but the scatter on our data are too large to draw any conclusion at this point. An even more pressing open question is the mechanism whereby a conformational transition could cause such large electronic changes.

At this stage, we have no clear understanding of the relationship between incorporation kinetics and signal features. One complication is presented by our original choice of single-stranded templates. Stalling and falling-off are clearly significant problems which makes it difficult to discriminate between the end of a polymerization event and a stall, and this greatly complicates the problem of assigning a single nucleotide incorporation event to particular features in a cluster of signals. If, as appears to be the case, the continuous and rather constant signals obtained from the $d(ATC)_5$ template do represent continuous incorporation of nucleotides, then the interval between peaks (~0.3 ms) is much faster than the rate of nucleotide incorporation found in bulk kinetic studies.[8] Either the bulk studies are averages over very fast kinetics interspersed with long pauses, or, as seems more likely, these electrical measurements are detecting more than one signal spike per nucleotide incorporation.

Further investigations require improvements to be made in future studies. The use of a fixed-gap device should greatly reduce fluctuations in the contact geometry and allow for data accumulation over much longer times. A constant baseline current would enable separation of the large fluctuations (which may carry sequence information) from the small fluctuations (which probably do not) as each peak is acquired. Careful design of double-stranded templates should eliminate many of the problems posed by secondary structure in single stranded templates, and yield runs of constant signal (as observed here for the $d(ATC)_5$ template).

While we have clearly shown that polymerase activity can be detected electrically, the data set obtained to date cannot give a definitive answer to the interesting question of whether or not nucleotides could be identified as they are incorporated. There are reasons why no differences might be expected: Kool has shown that the major factor in polymerase selectivity is stacking between the penultimate base and the dNTP being incorporated[18] so that contacts with the polymerase do not change much between the four nucleotides. But there do appear to be sequence-specific differences in incorporation kinetics. Xia et al. have shown that there are measurable differences in the incorporation kinetics of the various combinations of dNTP and

penultimate base,[19] as a consequence of base-stacking interactions in the polymerase binding pocket as the incoming dNTP is loaded on top of the previously incorporated nucleotide. Thus, with adequate data, combined with studies of chemical kinetics, and templates that probe all 16 base-stacking combinations, insight might be gained into the sequence sensitivity of these signals.

METHODS

Expression and purification of Avitag φ29 DNA polymerase

E. coli strain BL21 (DE3) (Novagen) was transformed with different versions of the φ29 gene embedded in pET15b plasmid and grown on an LB agar plate (1% tryptone, 0.5% yeast extract, 0.5% NaCl, and 1.5% agar) containing ampicillin (50 µg/ml) to select the transformants. Cells were grown in LB medium (20 ml) at 37ºC with shaking for 12 hours. A part of the culture was diluted 1:1000 into fresh LB medium (1L) and grown at 37ºC with shaking. The 1L culture was induced by adding 0.5 mM IPTG when the OD_{600} reached 0.6 and kept shaking overnight at 18ºC. Cells were harvested in a 1L centrifuge bottle by centrifugation at 5,710g for 20 minutes at 4ºC and stored at -80ºC until required. The purification is essentially as described elsewhere.[20]

Biontinylation of Avitag φ29 DNA polymerase

For subsequent BirA biotinylation, the purified protein was exchanged to a buffer containing 20 mM potassium phosphate buffer pH 7.0, 200 mM L-Glutamic acid potassium salt and 1 mM DTT. The *in vitro* enzymatic biotinylation was performed by incubating 100 µg of polymerase in the same buffer with 10 mM ATP, 10 mM $Mg(OAc)_2$, 50 mM biotin and 15 units of BirA (Avidity) for 1 hour at 30ºC. Free biotin was removed by a desalting column (GE Life Sciences).

Rolling circle replication (RCR) ssDNA template and primer

Linear single strand oligonucleotides RCR (5'- p - CCGTACGATTCGTATCTACTATCGTTCGATTCGCATCATCTA -3') was used to form circular RCR template by enzymatic self-ligation with Circligase (Epicentre). 0.1 nmol linear single strand RCR DNA was mixed with 100 Units Circligase in 1 X reaction buffer containing 50 µM ATP and 2.5 mM $MnCl_2$. After 2 hours incubation at 60ºC, the product was heated up to 80ºC for 10 minutes to inactivate the Circligase. The linear ssDNA left in the solution was digested by Exo I (NEB). The RCR template was analyzed by electrophoresis on a denaturing gel containing 8 M urea and 20% polyacrylamide for quality control. 2.5 pmols of RCR template was annealed with 50 pmols RCR primer (5'- GGCATGCTAAGCATAGATGAT -3') by heating up to 95ºC for 5 minutes and gradually cooling down to room temperature (decreasing 0.1ºC/s) and stored at -20ºC for later use.

Activity assay of Avitag φ29 DNA polymerase

A rolling circle replication reaction was performed for the activity test of all versions of φ29 DNA polymerase. 1.25 pmols RCR template and primer complex was mixed with 500 µM dNTP and 4 pmols φ29 DNA polymerase in 1 X reaction buffer containing 50 mM Tris-HCl ph7.5, 10 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 4 mM DTT. The mixture was incubated at 30ºC for 1 hour. The product was visualized on 0.8% agarose gel by GelRed (Biotium) staining.

Functionalizing substrates and STM probes

Palladium substrates for STM measurement were prepared by evaporating a 200 nm palladium film onto a silicon wafer using an electron-beam evaporator (Lesker PVD 75), with a 10 nm titanium adhesion layer. The substrates were treated with a hydrogen flame immediately before functionalizing and then immersed in solutions of thiolated streptavidin (ProteinMods) or thiolated biotin overnight. The thiolated biotin was prepared as described elsewhere[3] and dissolved in freshly degassed pure ethanol to a final concentration of 50 μM. 1 μM thiolated streptavidin solutions in 1 mM PB buffer were used for substrate functionalization. All the buffers and solutions were prepared in Milli-Q water with a conductivity of 18.2 MΩ. For all measurements, the 1 mM PB buffer (pH 7.4) was degassed with argon to avoid interference from oxygen. The polymerization buffer was 1mM phosphate buffer, pH=7.4, 4 mM TCEP, 10mM $MgCl_2$ with 1 mM dNTPs and 1 μM template. Sequences of the self-priming templates used are shown in Figure S11. Substrate functionalization was characterized by Fourier transform infrared (FTIR) spectroscopy (Fig. S12). STM probes were etched from a 0.25mm Pd wire (California Fine Wires) by an AC electrochemical method. To avoid current leakage, probes were insulated with high-density polyethylene following the method described earlier for gold probes.[13] Each probe was tested by STM in 1 mM phosphate buffer (pH 7.4) at +0.5 V bias to ensure the leakage current was <1pA. For functionalization, the probe was immersed in ligand solutions for 4 h or overnight. After that, it was taken out, rinsed with water, gently blown dry with nitrogen gas, and used immediately. Cyclic voltammetry was used to check that the potential regions used (+50 to -50mV vs Ag/AgCl, 10 mM KCl) were free from Faradaic currents in the presence of the various components of the assembly (Figures S13-15).

STM Measurements

STM measurements were carried out on a PicoSPM scanning probe microscope (Agilent Technologies), using a DAQ card (PCI-6821 or PCIE-7842R, National Instruments) for data acquisition. The Teflon cell, in which buffer solution and analytes were added, was cleaned with Piranha solution and then sonicated in Milli-Q water three times to remove residues (Note that Piranha solution is highly corrosive and must be handled with extreme care). In order to better control the surface potential, an Ag/AgCl reference electrode with a 10 mM KCl salt bridge was connected onto the substrate. The probe was firstly engaged to 4 pA setpoint current with a bias of -0.2 V and then left to stabilize for 2 h before measurement. For STM IV sweep measurements, the servo system was first turned off and the probe was retracted by ΔZ nm with a speed of 1 nm/s. After that, the probe was suspended at that height for 1 minute, during which a custom Labview program was used to monitor the current change. Once the current exceeded a threshold of 50 pA, we considered it as a binding event and IV sweeps were started from -0.2 V to +0.2 V and then back, with a sweep rate of 1 V/s, followed by a 0.2 s resting. Subsequently, the current was checked again. If the current was still more than twice the noise level (6 pA), IV curves were continuously recorded until the bound protein molecule escaped. After one-minute of measurement, the servo system was turned on to re-engage the probe and then the whole process repeated. In each measurement, at least 1000 IV curves were collected, from which curves that reproduced on the up sweep and down sweep were selected (80% of the total) to construct the conductance distribution histogram. Current vs. time traces were recorded by another Labview program with a similar procedure except that the bias was held

constant during the probe holding process. The analog-to-digital sampling rate is 50 KHz, about 5 times the intrinsic response of the STM electronics.

ASSOCIATED CONTENT

Online supporting Information: Figures S1 to S15.

Author Contributions

¶ These authors contributed equally to this work.

Notes

The authors declare the following competing financial interest(s): SL is a cofounder of a company with interests in this the subject matter of the present paper. SL, BZ and HD are named as co-inventors on patent applications.

ACKNOWLEDGMENT This work was supported by grants HG006323 and HG010522 from the National Human Genome Research Institute, by Recognition AnalytiX Corp and the Edward and Nadine Carson Endowment.

REFERENCES

1. Adhikari, R. Y.; Malvankar, N. S.; Tuominen, M. T.; Lovley, D. R. *RSC Advances* **2016,** 6, 8354-8357.
2. Nadav Amdursky; Debora Marchak; Lior Sepunaru; Israel Pecht; Mordechai Sheves; Cahen, D. *Advanced Materials* **2014,** 26, 7142–7161.
3. Zhang, B.; Song, W.; Pang, P.; Lai, H.; Chen, Q.; Zhang, P.; Lindsay, S. *Proc Natl Acad Sci U S A* **2019,** https://doi.org/10.1073/pnas.1819674116.
4. Zhang, B.; Lindsay, S. *Nano letters* **2019,** doi.org/10.1021/acs.nanolett.9b01254.
5. Bostick, C. D.; Mukhopadhyay, S.; Pecht, I.; Sheves, M.; Cahen, D.; Lederman, D. *Reports on Progress in Physics* **2018,** 81, 026601.
6. Blanco, L.; Salas, M. *J Biol Chem* **1996,** 271, (15), 8509-12.
7. Esteban, J. A.; Salas, M.; Blanco, L. *J Biol Chem* **1993,** 268, (4), 2719-26.
8. Morin, J. A.; Cao, F. J.; Lazaro, J. M.; Arias-Gonzalez, J. R.; Valpuesta, J. M.; Carrascosa, J. L.; Salas, M.; Ibarra, B. *Nucleic acids research* **2015,** 43, (7), 3643-52.
9. Berman, A. J.; Kamtekar, S.; Goodman, J. L.; Lazaro, J. M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T. A. *EMBO J* **2007,** 26, (14), 3494-505.
10. Kamtekar, S.; Berman, A. J.; Wang, J.; Lazaro, J. M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T. A. *Mol Cell* **2004,** 16, (4), 609-18.
11. Fairhead, M.; Howarth, M. *Methods Mol Biol* **2015,** 1266, 171-84.
12. Beckett, D.; Kovaleva, E.; Schatz, P. J. *Protein Sci* **1999,** 8, (4), 921-9.
13. Tuchband, M.; He, J.; Huang, S.; Lindsay, S. *Rev Sci Instrum* **2012,** 83, (1), 015102.
14. Beard, W. A.; Shock, D. D.; Batra, V. K.; Pedersen, L. C.; Wilson, S. H. *J Biol Chem* **2009,** 284, (46), 31680-9.
15. Peng, J.; Peng, S.; Jiang, A.; Wei, J.; Li, C.; JieTana. *Analytica Chimica Acta* **2010,** 683, 63-68.

16. Olsen, T. J.; Choi, Y.; Sims, P. C.; Gul, O. T.; Corso, B. L.; Dong, C.; Brown, W. A.; Collins, P. G.; Weiss, G. A. *Journal of the American Chemical Society* **2013,** 135, (21), 7855-60.
17. Maier, B.; Bensimon, D.; Croquette, V. *Proc Natl Acad Sci U S A* **2000,** 97, (22), 12002-7.
18. Kool, E. T. *Annu Rev Biophys Biomol Struct* **2001,** 30, 1-22.
19. Xia, S.; Vashishtha, A.; Bulkley, D.; Eom, S. H.; Wang, J.; Konigsberg, W. H. *Biochemistry* **2012,** 51, (24), 4922-31.
20. Lazaro, J. M.; Blanco, L.; Salas, M. *Methods Enzymol* **1995,** 262, 42-9.

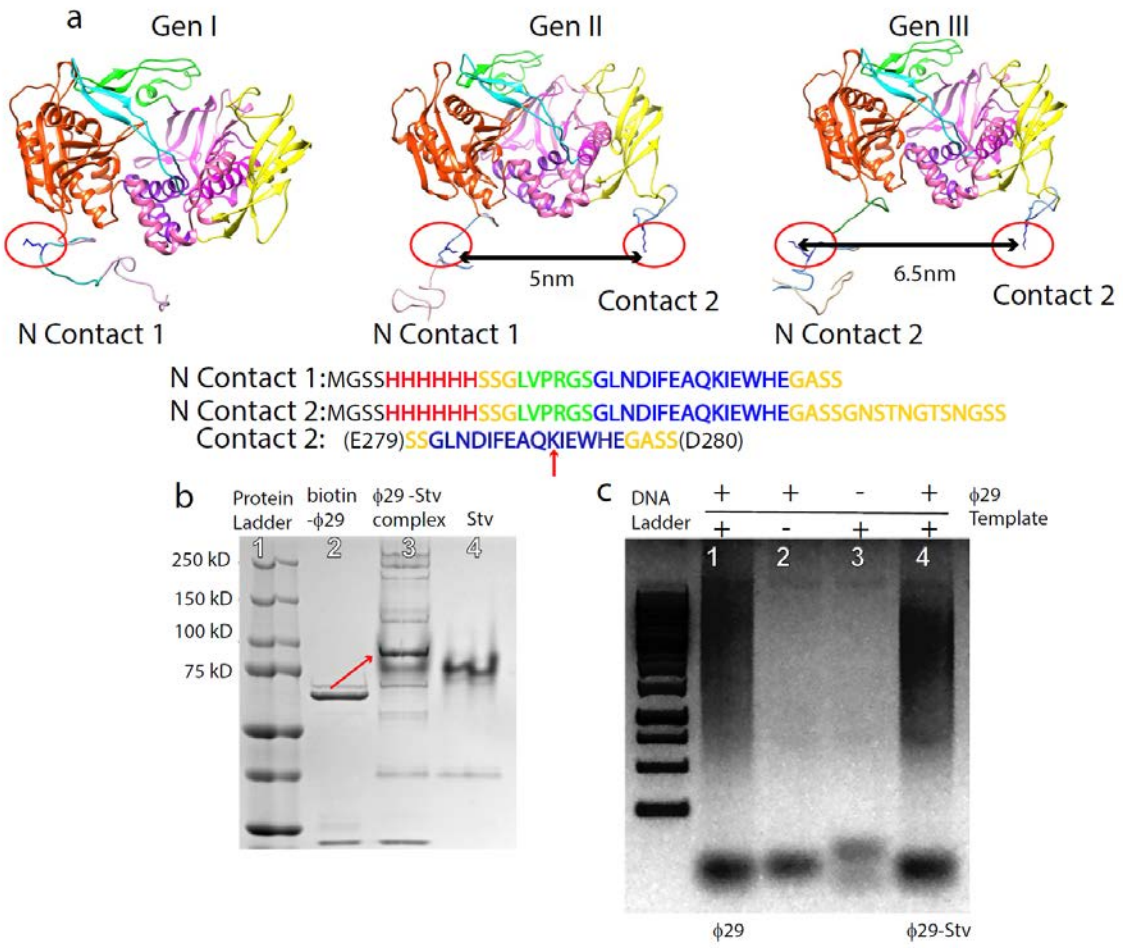

Figure 1: Engineering a polymerase with two contacts. (a-i) Φ29 polymerase with a single Avitag at the N terminus (sequence N Contact 1). (a-ii) Φ29 polymerase with a second Avitag inserted between E279 and D280 (sequence Contact 2). (a-iii) Φ29 polymerase with the two Avitags and an additional flexible linker in the N terminus sequence (sequence N Contact 2). (b) SDS-PAGE gel showing the Gen III modified polymerase in lane 2. On incubating with strepavidin, the major product (red arrow) is a complex with a mass approximately equal to the polymerase plus two streptavidin molecules. Lane 4 is WT streptavidin. (c) Agarose DNA gel showing elongation products from a rolling circle polymerization assay using (lane 1) the Gen III polymerase and (lane 4) the polymerase complexed with streptavidin.

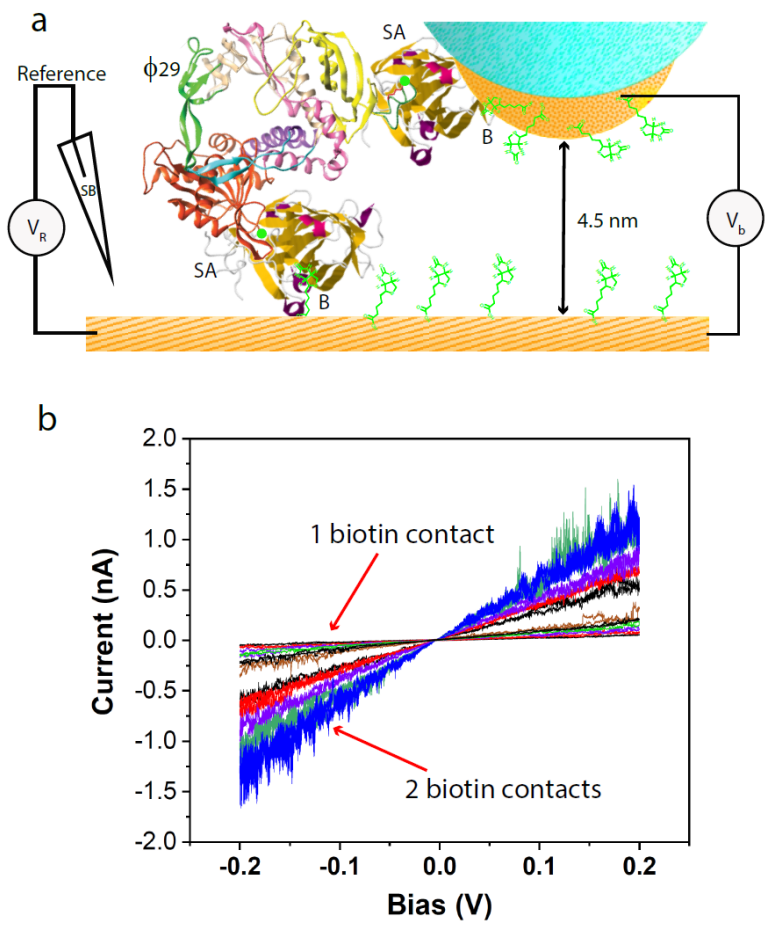

Figure 2: Bioelectronic circuit including a polymerase. (a) An STM probe is held ~4.5 nm above a conducting substrate, immersed in electrolyte and under potential control via a salt bridge (SB) connected to an Ag/AgCl reference. Potentials are maintained so as to avoid Faradaic currents. The electrodes are functionalized with thiolated biotin which captures streptavidin molecules (SA) which then trap a polymerase (Φ29). (b) Typical current-voltage curves from trapped single molecule of polymerase. Each trace is for a sweep from -0.2V to +0.2V and then back to -0.2V. Trace and retrace superimpose, demonstrating reproducibility. The conductances for individual molecules are obtained from the slope of these traces. Currents are increased by about an order of magnitude for Gen III polymerase (2 biotin contacts) compared to Gen I polymerase (1 biotin contact).

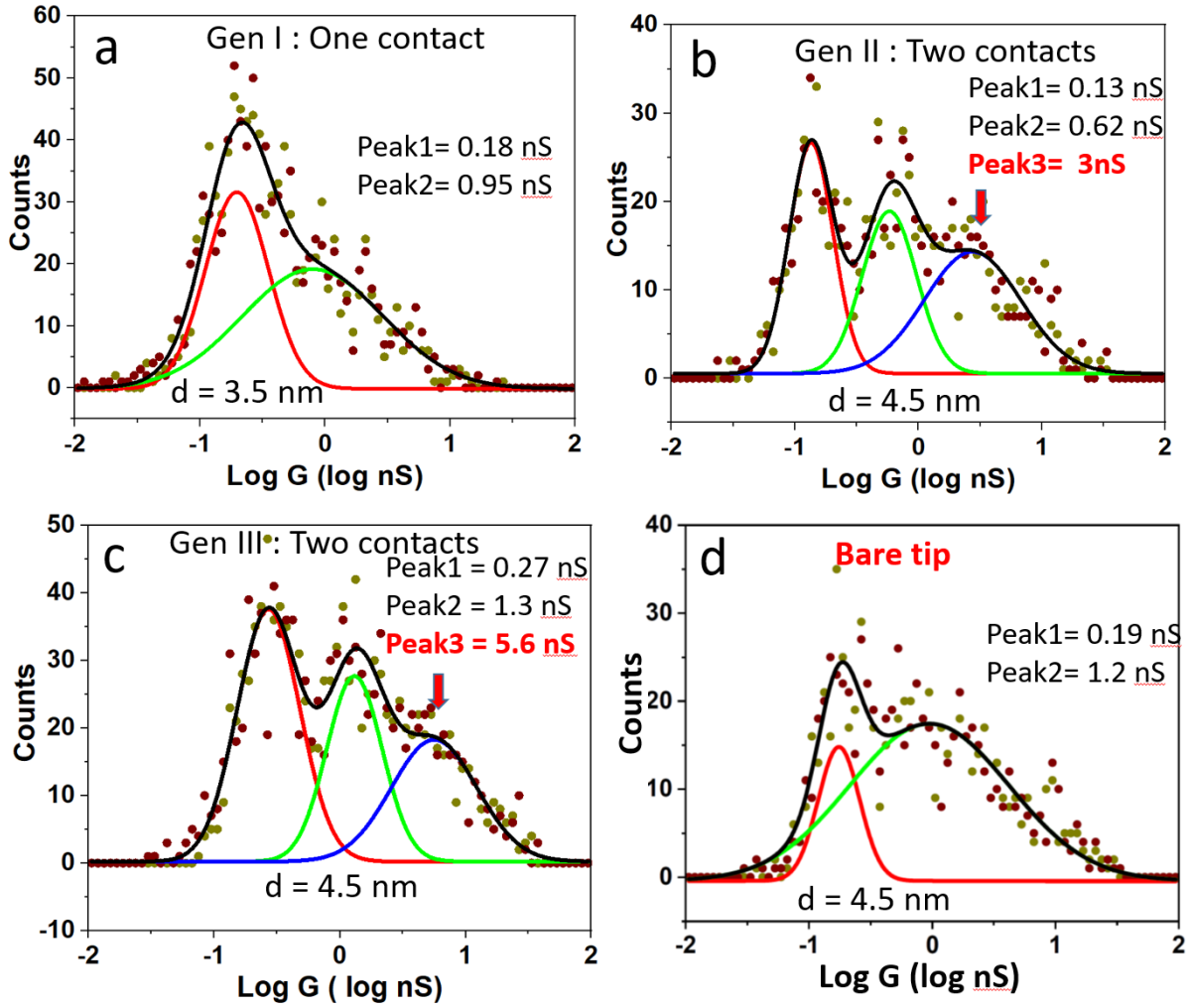

Figure 3: **Conductance distributions for the three generations of polymerase.** (a) Gen I - one contact yields 2 peaks presumably because of additional contacts vis the surface cystienes on Φ29. (b) Gen II connected via thiolated streptavidins. (c) Gen III connected via WT streptavidin connected to the electrodes by thiolated biotin molecules. The red arrow points to the additional high conductance peak that appears as a consequence of the second contact point. (d) When a bare tip is used (Gen II, thio-streptavidin on the substrate) Peak 3 is no longer observed.

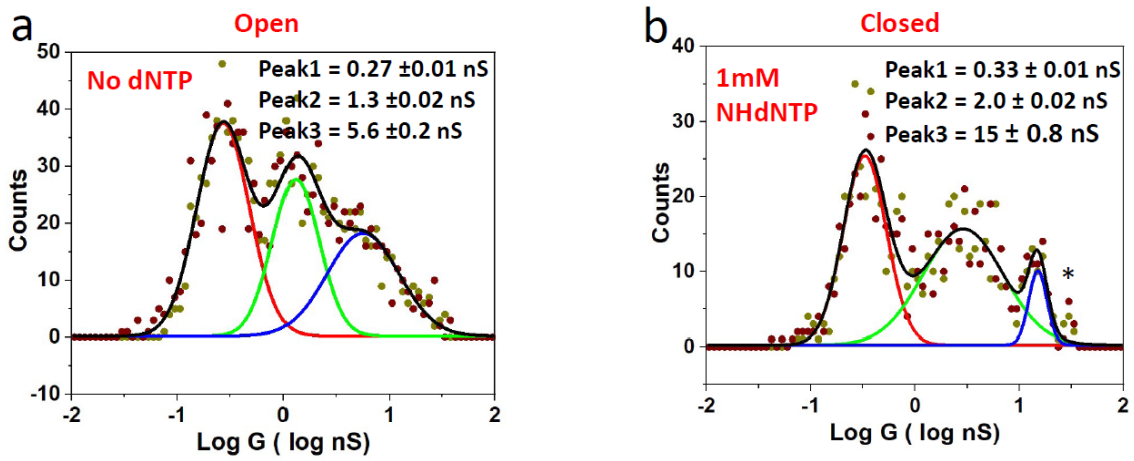

Figure 4: The closed to open transition changes polymerase conductance. (a) Distribution in the absence of dNTPs but with bound template. The polymerase is largely open. (b) In the presence of bound non-hydrolyzable dNTPs - the polymerase is locked in the close configuration. The highest conductance peak moves from 5.6 to ~15 nS.

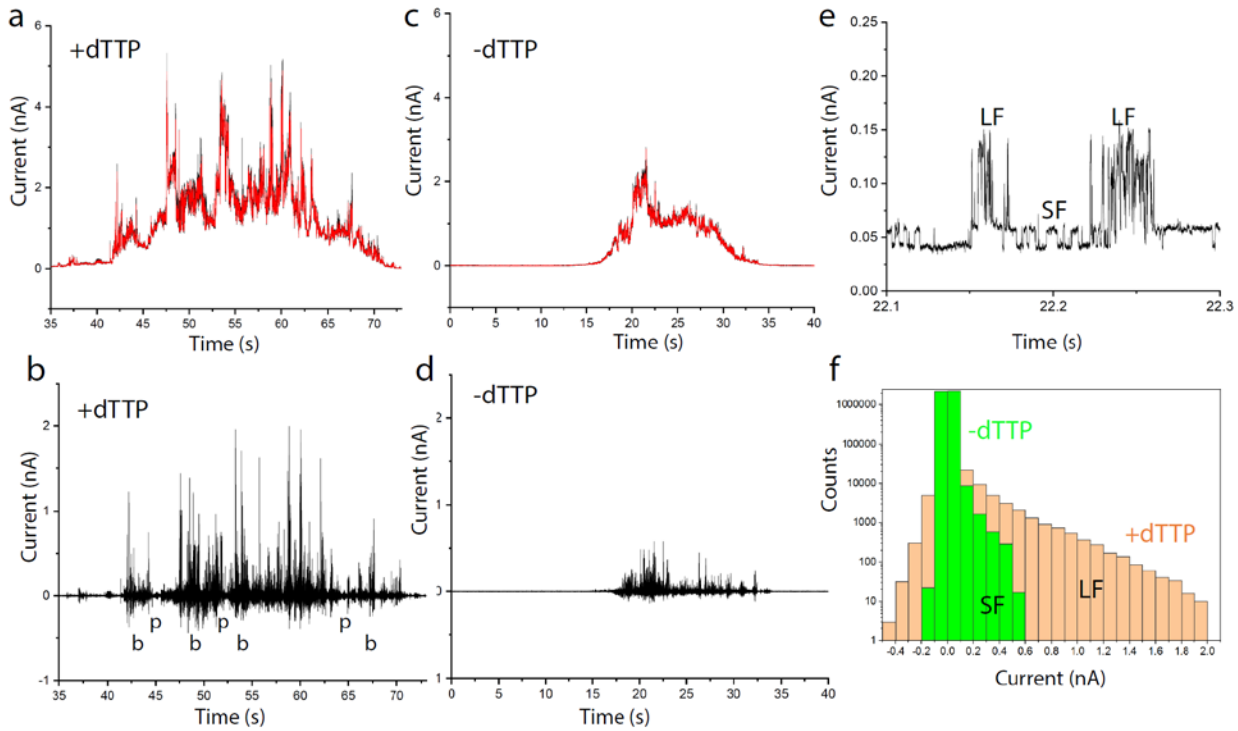

Figure 5: Characterizing the signals generated when polymerase is activated. (a) The current through the polymerase changes markedly over time as the STM probe drifts (black curve). This drifting baseline current is fitted with an asymmetric least squares (ALS) procedure to yield a smoothed background current (red curve superimposed on black curve). (b) Subtraction of the fitted background shows the rapid changes in current that occur in an activated polymerase. Typically the dynamic signals occur in bursts (b) interspersed with pauses (p). (c) Shows the current recorded vs time in the absence of the complementary nucleotide triphosphate, and the ALS fit. These data were chosen to have about the same DC conductance as the active polymerase in (a) and (b). (d) The baseline subtracted signal shows that there is noise in the inactive system. The noise coincides with the formation of a contact to the polymerase in all cases, so it is not random. (e) Shows an expanded portion of the noise signal measured from an activated polymerase showing two distinct noise components - large fluctuations (LF) and small fluctuations (SF). (f) Is histogram comparing the distribution of spike heights from the active (+dTTP) and inactive (-dTTP) polymerase.

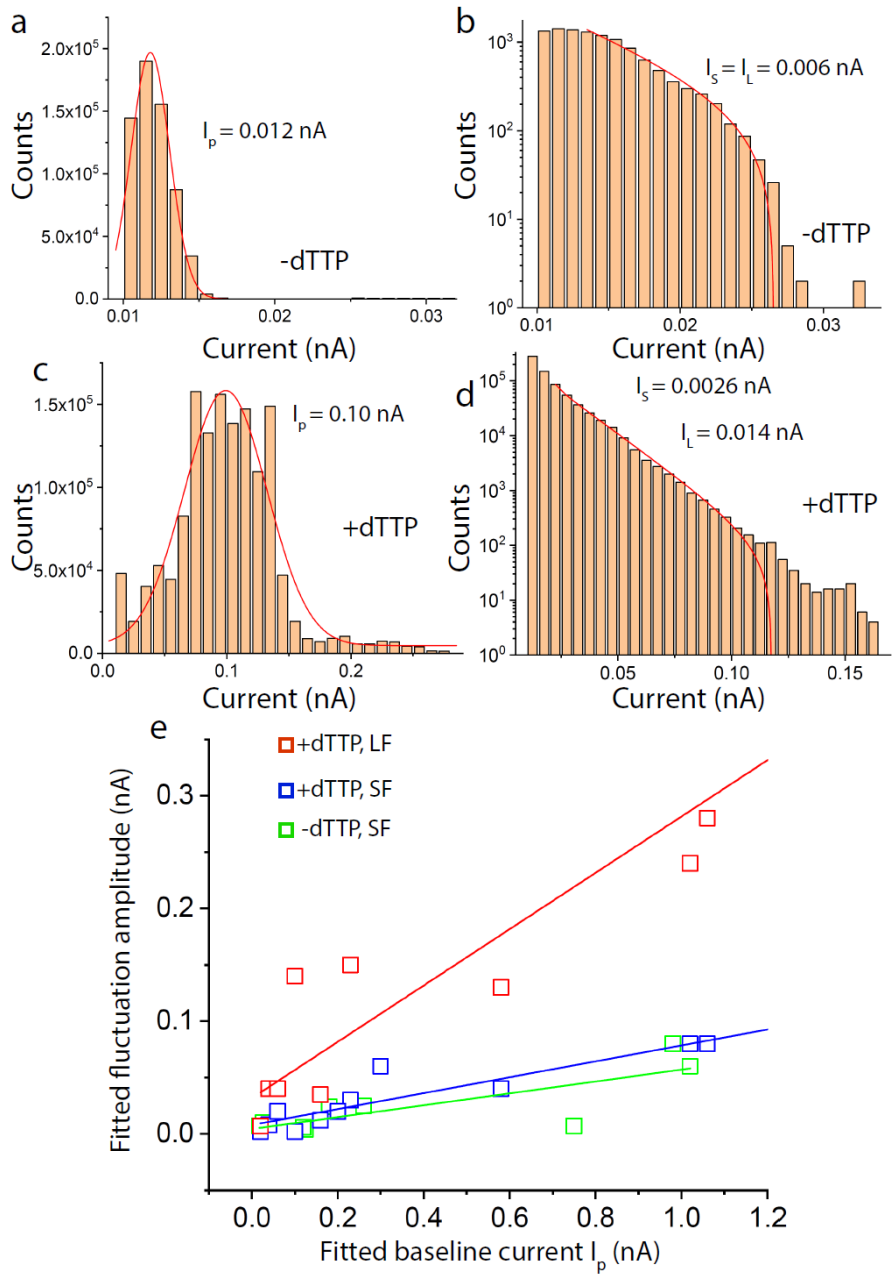

Figure 6: Small and large fluctuations defined in terms of baseline current. Larger signals are obtained with the higher-conductance contacts. To quantify this, the baseline levels in a given run are quantified by fitting a Gaussian model to the distribution of measured currents, characterized by Gaussian peak current, I_P (a, c). A two exponential fit (I_S, I_L) is used to model the distribution of noise signal amplitudes. When the complementary base is absent (b) the fit converges to one value, $I_S = I_L$ in all cases. (d) In the presence of the complementary nucleotide, 9 out of 13 molecules showed a bimodal distribution of spike heights with $I_S << I_L$. (e) Values of I_L (red) and I_S (blue) plotted vs the associated value of baseline current (I_P). Values of the single exponential distribution fitted for the -dTTP experiments are shown in blue. Lines are linear fits yielding $L_F = 0.25(\pm0.03)I_p$ and L_S $0.07(\pm0.01)I_p$. The small fluctuations have

essentially the same dependence on I_p in the control experiments (-dTTP). Analysis of data obtained from $d(ATC)_5$ is shown in Figure S10.

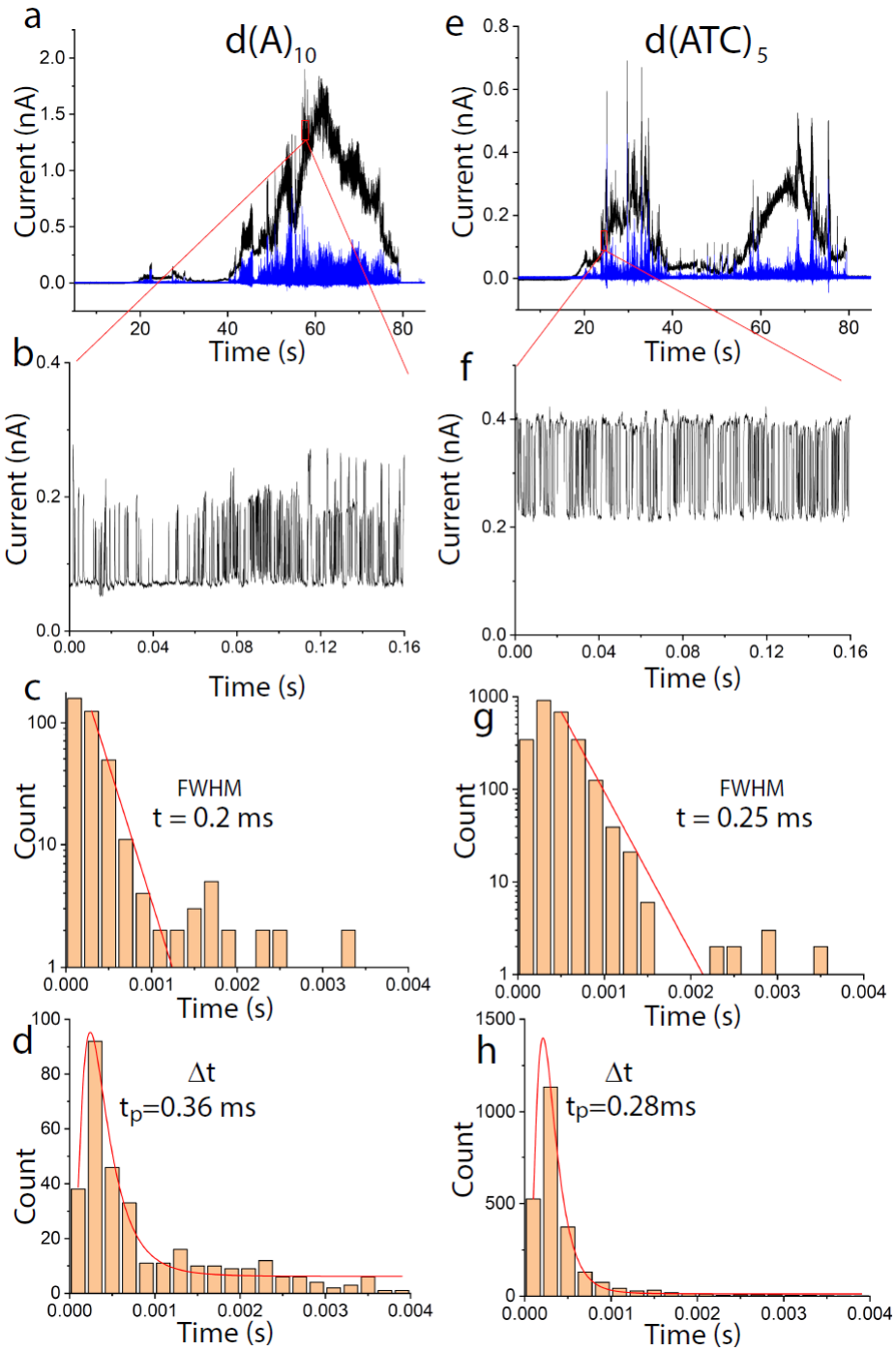

Figure 7: Comparing Noise characteristics from two sequences. Representative raw (black) and background subtracted signals (blue) for $d(A)_{10}$ (a) and $d(ATC)_5$ (e). 160 ms samples of noise are shown in (b) and (f). These are raw data selected for having a constant background

current. They illustrate the "25% rule" approximately, though small fluctuations appear to be absent in the run of $d(ATC)_5$ data shown. (c) and (g) FWHM distributions for ~ 1s of signal traces (exponential fits are shown in red). (d) and (h) Interval between signals spikes, fitted with log-normal distributions.