

Supporting Information

Conformational Smear Characterization and Binning of Single-Molecule Conductance Measurements for Enhanced Molecular Recognition

Lee E. Korshoj,^{1,2} Sepideh Afsari,^{1,2} Anushree Chatterjee,^{1,3} Prashant Nagpal^{1,2,3,4*}

¹Department of Chemical and Biological Engineering, University of Colorado Boulder

²Renewable and Sustainable Energy Institute (RASEI), University of Colorado Boulder

³BioFrontiers Institute, University of Colorado Boulder

⁴Materials Science and Engineering, University of Colorado Boulder

*Email: pnagpal@colorado.edu

Sections:

1. STM break junction (STM-BJ) experiments
2. Data processing and constructing conductance histograms
3. Characterizing DNA adsorption on cysteamine
4. Single gold atom junctions and measuring instrument response function (IRF)
5. Calculating smear parameter S_p and binning
6. Landauer transmission model
7. Base calling/molecular recognition calculations

Figures:

Figure S1. Measuring single-molecule conductance with STM-BJ technique

Figure S2. Cysteamine and ethanedithiol conductance histograms

Figure S3. DNA surface adsorption on the cysteamine SAM

Figure S4. Gold atom conductance and IRF determination

Figure S5. Smear impact on conductance histogram signatures

Figure S6. Smear parameter S_p distributions and binning

Figure S7. Transmission pathways through nucleotides

Figure S8. Transmission calculations for single STM-BJ measurements

Figure S9. Junction verification with pH perturbation

Figure S10. Correlation coefficients and thresholds for base calling calculations

Figure S11. Detailed base calling output with 20× coverage for no SCRIB

Figure S12. Detailed base calling output with 20× coverage for SCRIB

Figure S13. Detailed base calling output with 7× coverage for low smear

Figure S14. Correct sequence of base calls for SCRIB analysis

1. STM-BJ experiments

As mentioned in the main text, the displacement of the STM tip in relation to the substrate gives rise to quasi-exponentially decaying traces without molecules, and current plateaus or steps in the current-distance traces when a molecular junction forms (Fig. S1). Histograms from STM-BJ measurements on the cysteamine SAM by itself (no DNA) showed a double conductance peak at -0.10 V bias, which was also seen for ethanedithiol (similar structure, with a second thiol replacing the amine group). The origin of multiple conductance peaks for thiol-terminated alkanes is likely due to the different available contact geometries between the anchoring groups of the molecules and the gold atoms of the electrodes, and it is an inherent feature in the formation of single-molecule junctions.¹ The peaks were not observed in the measurement window at -0.50 V bias, which we selected as the fixed bias for subsequent conductance measurements (Fig. S2).

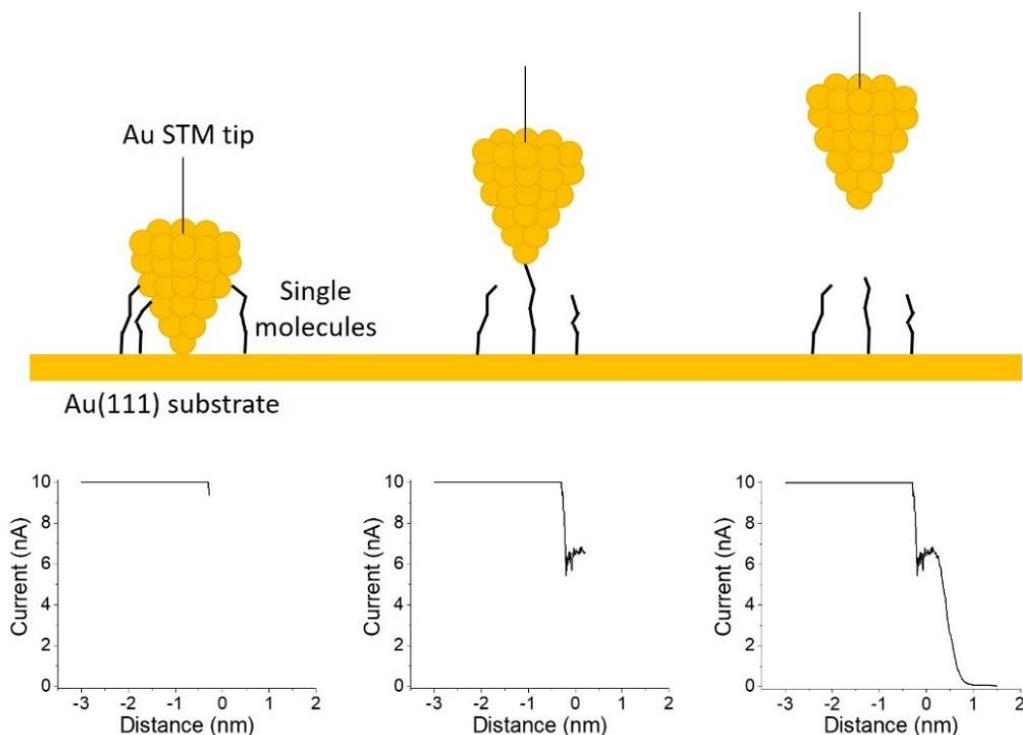


Figure S1. Measuring single-molecule conductance with STM-BJ technique. Schematic showing the process of forming molecular-junctions (top), and the corresponding current plateaus or steps in the current-distance traces observed when a molecular junction is formed (bottom).

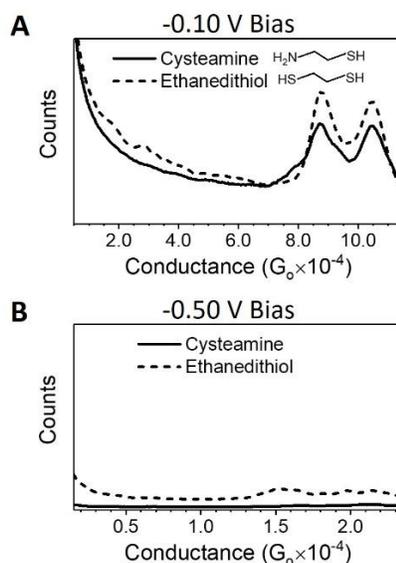


Figure S2. Cysteamine and ethanedithiol conductance histograms. (A) Conductance histograms from STM-BJ measurements of ethanedithiol (dotted line) and cysteamine (solid line) at -0.10 V bias, comprised of 1526 and 1439 individual current-distance curves, respectively (2509 and 4784 individual current-distance curves prior to filtering, respectively). Structures of both molecules are shown. (B) Conductance histogram from STM-BJ measurements of ethanedithiol (dotted line) and cysteamine (solid line) at -0.50 V bias, comprised of 2422 and 3034 individual current-distance curves, respectively.

2. Data processing and constructing conductance histograms

Sets of current-distance traces used to construct histograms shown in this report have been filtered to remove current-distance traces where a break junction was not formed (leading to noise in the histograms). A MATLAB (The Mathworks, Inc. MA, USA) program was written to perform this task. The program iterated through each current-distance curve and determined if a current plateau (signifying a molecular break junction) was formed. If there were no break junctions formed, then that particular current-distance curve was removed and did not contribute to the overall histogram. Note that this filtering procedure was not used for histograms where no features (or peaks) were observed, as the majority of current-distance traces would be removed (i.e., Fig. S2B for cysteamine and ethanedithiol). To construct signature histograms (from hundreds of spectra), a bin size of 0.02 nA was used. For the current range of 0-10 nA, this corresponds to 500 bins in each histogram. Additionally, the conductance histograms shown in the report are 16-point average smoothed.

3. Characterizing DNA adsorption on cysteamine

AFM studies were used to analyze DNA adsorption characteristics. DNA was adsorbed at varying concentrations onto a cysteamine monolayer on gold. The density of DNA molecules on the surface was estimated via AFM imaging and semi-automated image analysis. First, a number

of representative AFM images were collected on different areas of each sample. Next the images were leveled by mean plane subtraction and flattened line-by-line using Gwyddion image analysis software (<http://gwyddion.net/>). Then a mask was generated to separate all features over a minimum height threshold after removing single-pixel noise with a median filter. Some minor editing of the mask by hand was sometimes required to omit horizontal imaging ‘scars’ or separate closely-spaced but clearly distinguishable features. Then, the maximum height of each individual molecular feature was measured relative to the immediately surrounding local background height using a custom MATLAB script. Finally, features with a maximum height within the range of 0.3–2.0 nm were counted as individual DNA molecules. The mean surface-density was determined by averaging the number of molecules per area in a number of images corresponding to different areas of the surface. As mentioned in the main text, we found that the DNA surface density increased with increasing DNA concentration up to ~5 nM, at which point it appeared to saturate around 2000 molecules- μm^{-2} (Fig. S3). Adsorption is likely limited by inefficient packing of the DNA due to mutual electrostatic repulsion by the negatively-charged molecules. In order to increase the coverage, we extended the adsorption time to overnight, allowing the DNA in solution to bind well and the solvent to evaporate (which increases the DNA concentration as the solution dries) before rinsing off excess unbound DNA.

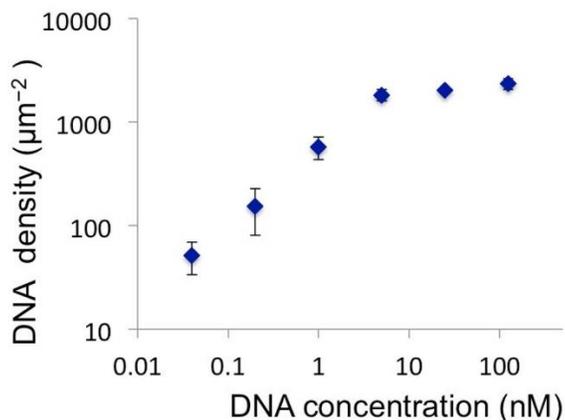


Figure S3. DNA surface adsorption on the cysteamine SAM

4. Single gold atom junctions and measuring IRF

In attempting to quantify molecular smear, it is important that we first characterized the instrument response function (IRF) of our STM-BJ system by collecting measurements on the bare Au(111) substrate. For creating the gold point contact (single gold atom junction) using STM-BJ, a sharp gold tip is driven into contact with a gold surface. After indentation, the tip is retracted and a connective neck with one or multiple atoms bridging between the tip and the surface is formed (Fig. S1). The single atom junction is broken eventually as the tip is retracted farther. As with molecules, the displacement of the two electrodes (the STM tip and the substrate) without formation of junctions gives rise to quasi-exponentially decaying traces. When

any junction forms, the current remains approximately constant as the electrode-electrode distance increases so that a current plateau, or step, appears in the traces. The current drops exponentially with distance when the junction breaks. The histogram generated from these gold metal-metal junctions contains a peak that can be ascribed to charge transport and conductance of a one-atom contact (Fig. S4). In this case, the gold atoms are treated as “hard spheres” and we can characterize the IRF to measure a baseline “zero smear” quality factor or percentage variance in conduction ($\Delta G_{0,FWHM}/G_0$, where $\Delta G_{0,FWHM}$ is the full width at half-maximum, FWHM, for the quantum conductance peak in the gold histogram, and G_0 is the quantum conductance $2e^2/h = 7.75 \times 10^{-5}$ S) in such molecular measurements. A Gaussian was fit to the histogram made from filtered spectra so that the IRF was comparable to filtered molecular measurements.

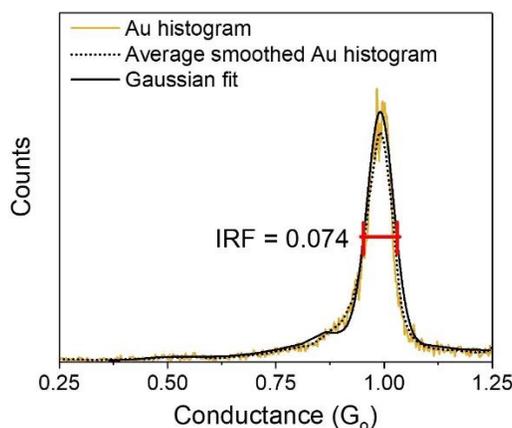


Figure S4. Gold atom conductance and IRF determination. OriginPro 2016 was used to fit a Gaussian to the quantum conductance peak in the filtered histogram generated from STM-BJ measurements on the bare Au(111) substrate.

5. Calculating smear parameter S_p and binning

A MATLAB program was written to calculate the step distance in each current-distance STM-BJ measurement in order to bin individual measurements into smear bins for SCRIB analysis. This step distance, or the distance over which the molecular junction remains intact, is referred to as our smear parameter S_p . We wrote an algorithm to identify the start and end positions of step-like features and extract the resulting distance over which the step is maintained (as shown in step 2 of Fig. 3D). After calculating S_p values for each measurement, they were able to be separated (binned) to generate new histogram signatures comprised on only spectra having similar S_p values and hence similar conformation on the surface. Binned histogram signatures for C are shown in Fig. 1C, and those for A, G, and T are shown in Fig. S5. From these binned histogram signatures, a clear direct correlation can be seen between S_p and the variance in conductance (or increasing FWHM), which is quantified into a smear factor S_F in Fig. 1D. Due to the different range of S_p values for each nucleotide (Fig. 1E and Fig. S6), bin sizes were

determined separately such that enough spectra for base calling analyses were contained in each bin. A breakdown of S_P distributions and bin sizes is given in Fig. S6.

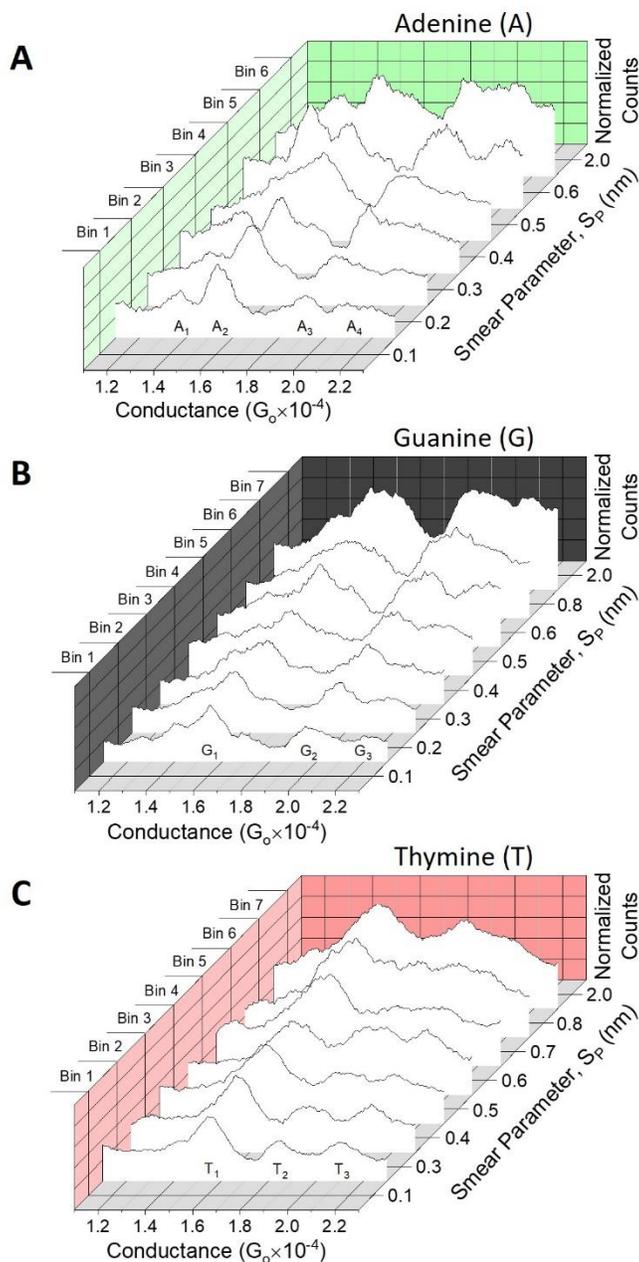


Figure S5. Smear impact on conductance histogram signatures. Observing the impact of smear on conductance signatures for (A) A, (B) G, and (C) T. As smear parameter S_P increases (increasing distance over which the molecular junction is maintained), the variance in the signature histogram peaks also increases. Measurements can be binned according to the S_P .

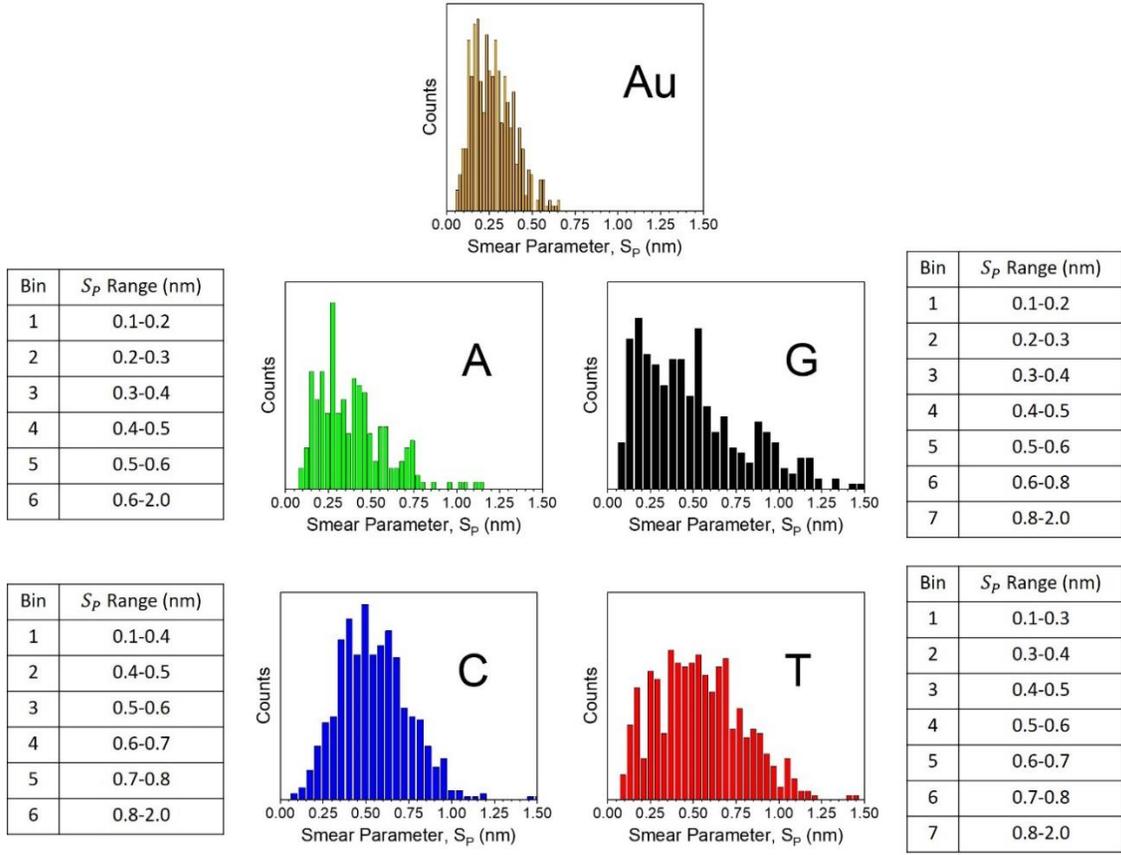


Figure S6. Smear parameter S_p distributions and binning. Histograms indicate distributions of S_p values from individual current-distance spectra. From these distributions, bins were set to contain sample sizes large enough to perform base calling analyses. The S_p distribution for Au measurements that were used to extract the IRF are also shown.

6. Landauer transmission model

We developed a theoretical formalism using Landauer transmission coefficients to describe charge conduction through the nucleotides in DNA.^{2,3} Schematics are provided in Fig. S7. Using the deoxyadenosine nucleotide as an example, we generated an expression for conductance corresponding to peak A_4 :

$$G_{A4} = G_0 \cdot T_{A4}(E_F)$$

Here, T_{A4} is the transmission function for the junction leading to conductance peak A_4 :

$$T_{A4}(E_F) = 16 \cdot \frac{|V_{sub,S}|^2}{\Gamma_{sub}(E_F)} \cdot \frac{|V_{S,C}|^2}{(E_S - E_F)^2} \cdot \frac{|V_{C,C}|^2}{(E_C - E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C - E_F)^2} \cdot \frac{|V_{N,O}|^2}{(E_N - E_F)^2} \cdot \frac{|V_{O,P}|^2}{(E_O - E_F)^2} \cdot \frac{|V_{P,O}|^2}{(E_P - E_F)^2} \cdot \frac{|V_{O,C}|^2}{(E_O - E_F)^2} \cdot \frac{|V_{C,C}|^2}{(E_C - E_F)^2} \cdot \frac{|V_{C,O}|^2}{(E_C - E_F)^2} \cdot \frac{|V_{O,C}|^2}{(E_O - E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C - E_F)^2} \cdot \frac{|V_{N,C}|^2}{(E_N - E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C - E_F)^2} \cdot \frac{|V_{N,tip}|^2}{(E_N - E_F)^2 \Gamma_{tip}(E_F)}$$

Since the part of this expression relating to tip and substrate anchoring, cysteamine, and the deoxyribose sugar are identical for all nucleobases, we lumped the terms into a single transmission, denoted $T1$:

$$T1 = 16 \cdot \frac{|V_{sub,S}|^2}{\Gamma_{sub}(E_F)} \cdot \frac{|V_{S,C}|^2}{(E_S-E_F)^2} \cdot \frac{|V_{C,C}|^2}{(E_C-E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C-E_F)^2} \cdot \frac{|V_{N,O}|^2}{(E_N-E_F)^2} \cdot \frac{|V_{O,P}|^2}{(E_O-E_F)^2} \cdot \frac{|V_{P,O}|^2}{(E_P-E_F)^2} \cdot \frac{|V_{O,C}|^2}{(E_O-E_F)^2} \cdot \frac{|V_{C,C}|^2}{(E_C-E_F)^2} \cdot \frac{|V_{C,O}|^2}{(E_C-E_F)^2} \cdot \frac{|V_{O,C}|^2}{(E_O-E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C-E_F)^2} \cdot \frac{|V_{N,tip}|^2}{(E_N-E_F)^2 \Gamma_{tip}(E_F)}$$

The remaining transmissions are related to the specific nucleobase structure. For A_4 , they are denoted $T2$ and $T3$:

$$T2 = \frac{|V_{N,C}|^2}{(E_N-E_F)^2}$$

$$T3 = \frac{|V_{C,N}|^2}{(E_C-E_F)^2}$$

In total, the conductance model for A_4 becomes

$$\frac{G_{A4}}{G_0} = T_{A4}(E_F) = T1 \cdot T2 \cdot T3 \quad (1)$$

In a similar fashion, the conductance models for A_1 , A_2 , and A_3 are

$$\frac{G_{A3}}{G_0} = T_{A3}(E_F) = T1 \cdot \frac{|V_{N,C}|^2}{(E_N-E_F)^2} \cdot \frac{|V_{C,C}|^2}{(E_C-E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C-E_F)^2} = T1 \cdot T2 \cdot T4 \cdot T3 \quad (2)$$

$$\frac{G_{A2}}{G_0} = T_{A2}(E_F) = T1 \cdot \frac{|V_{N,C}|^2}{(E_N-E_F)^2} \cdot T_{jump} \cdot \frac{|V_{C,N}|^2}{(E_C-E_F)^2} = T1 \cdot T2 \cdot T5 \cdot T3 \quad (3)$$

$$\frac{G_{A1}}{G_0} = T_{A1}(E_F) = T1 \cdot \frac{|V_{N,C}|^2}{(E_N-E_F)^2} \cdot T_{jump} \cdot \frac{|V_{C,N}|^2}{(E_C-E_F)^2} \cdot \frac{|V_{N,C}|^2}{(E_N-E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C-E_F)^2} = T1 \cdot T2 \cdot T5 \cdot T3 \cdot T2 \cdot T3 \quad (4)$$

For G, C, and T, we derive

$$\frac{G_{G3}}{G_0} = T_{G3}(E_F) = T1 \cdot \frac{|V_{N,C}|^2}{(E_N-E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C-E_F)^2} = T1 \cdot T2 \cdot T3 \quad (5)$$

$$\frac{G_{G2}}{G_0} = T_{G2}(E_F) = T1 \cdot \frac{|V_{N,C}|^2}{(E_N-E_F)^2} \cdot \frac{|V_{C,C}|^2}{(E_C-E_F)^2} \cdot \frac{|V_{C,N}|^2}{(E_C-E_F)^2} = T1 \cdot T2 \cdot T4 \cdot T3 \quad (6)$$

$$\frac{G_{G1}}{G_0} = T_{G1}(E_F) = T1 \cdot \frac{|V_{N,C}|^2}{(E_N - E_F)^2} \cdot T_{jump} \cdot \frac{|V_{C,N}|^2}{(E_C - E_F)^2} = T1 \cdot T2 \cdot T5 \cdot T3 \quad (7)$$

$$\frac{G_{C2}}{G_0} = T_{C2}(E_F) = T1 \cdot \frac{|V_{N,C=O}|^2}{(E_N - E_F)^2} \cdot \frac{|V_{C=O,N}|^2}{(E_{C=O} - E_F)^2} = T1 \cdot T6 \cdot T7 \quad (8)$$

$$\frac{G_{T2}}{G_0} = T_{T2}(E_F) = T1 \cdot \frac{|V_{N,C=O}|^2}{(E_N - E_F)^2} \cdot \frac{|V_{C=O,N}|^2}{(E_{C=O} - E_F)^2} = T1 \cdot T6 \cdot T7 \quad (9)$$

We would like to note that all bonds were conjugated here and to simplify the model, we assumed symmetric transmission coefficients in the two purine nucleobase rings. Only a single transmission exists for C and T even though multiple peaks are seen in each of the histogram signatures because the peaks arise due to multiple molecule-electrode contact geometries (or tautomers for T), not different anchoring groups or transmission pathways through the nucleobase. More details are provided in the following paragraph.

We verified this model with molecular junctions and transmission coefficients using single-molecule conductance measurements of individual nucleotides in DNA. The number of conductance peaks in the histogram for each nucleotide agrees with the proposed charge conduction pathways in our model. The two conductance peaks seen for C nucleotides (C_1 and C_2) match the observed conductance ratios ($\sim 1.21 \pm 0.01$) for the peaks in cysteamine and ethanedithiol (Fig. S2A), indicating that they likely arise from junctions on the same nitrogen in the nucleobase but different contact geometries with the gold electrodes. In support of this, we saw negative Pearson correlation coefficients for the peaks in ethanedithiol, cysteamine, and cytosine, indicating that each individual measurement only contributes to one of the two peaks in the histogram. For T nucleotides, peaks T_1 and T_2 occur due to the same reason. The closely-spaced peaks T_2 and T_3 occur due to the presence of keto-enol tautomers (Fig. S7D). For A and G nucleotide measurements, we do not see such distinct double peaks likely due to a larger number of transmission pathways leading to overlap in conduction peaks. Furthermore, all of the histogram peaks for G (G_1, G_2, G_3) and A (A_1, A_2, A_3, A_4) have positive Pearson correlation coefficients, with values increasing if the conduction happens in the same conjugated ring. To solve for transmission coefficients, we used the different nucleotide conductance peaks from the ensemble histograms containing spectra from all bins together (Fig. 1A). Since equations 1 & 5; 2 & 6; 3 & 7; and 8 & 9 are identical, we (as expected) observed similar conductance peaks for the different nucleotide measurements. Considering the identical equations, we are left with 5 distinct equations with 7 variables (transmission coefficients $T1, T2, T3, T4, T5, T6$, and $T7$). To calculate estimates for as many transmission coefficients as possible and verify our model, we combined some transmission coefficients. Using the ratio of equations 2 and 1, we get transmission across conjugated carbon-carbon ($T4, T_{C=C} = G_{A3}/G_{A4} = 0.88$, which is consistent with literature values for conjugated resonant carbon bonds.² Using the ratio of equations 7 and 6, $T5 = T_{jump} = (G_{G1}/G_{G2}) \cdot T4 = 0.78 \cdot T4 = 0.69$. We made a simplifying assumption $T2 \approx T3$ (also verified by numerical solution without the assumption) since $T2$ and

$T3$ ($T_{N=C}$ and $T_{C=N}$) are symmetric, and take the ratio of equations 4 and 3 to get $T2 \approx T3 = \sqrt{G_{A1}/G_{A2}} = 0.94$. Using equation 6, $T1 = (G_{G2}/G_0)/(T2 \cdot T3 \cdot T4) = 1.98 \cdot 10^{-4}$. While the values of transmission coefficients within the conjugated resonant ring ($T_{C=C}$, $T_{N=C}$, and $T_{C=N}$) should be close (0.88, 0.94, and 0.94), the value of $T1$ is lower mainly due to the transmission across the two anchoring groups and product of several transmission steps across the cysteamine molecule and the deoxyribose sugar bonds. Therefore, the number of peaks, correlation of proposed conduction pathways, and reasonable values for transmission coefficients compared to literature² all support the proposed model.

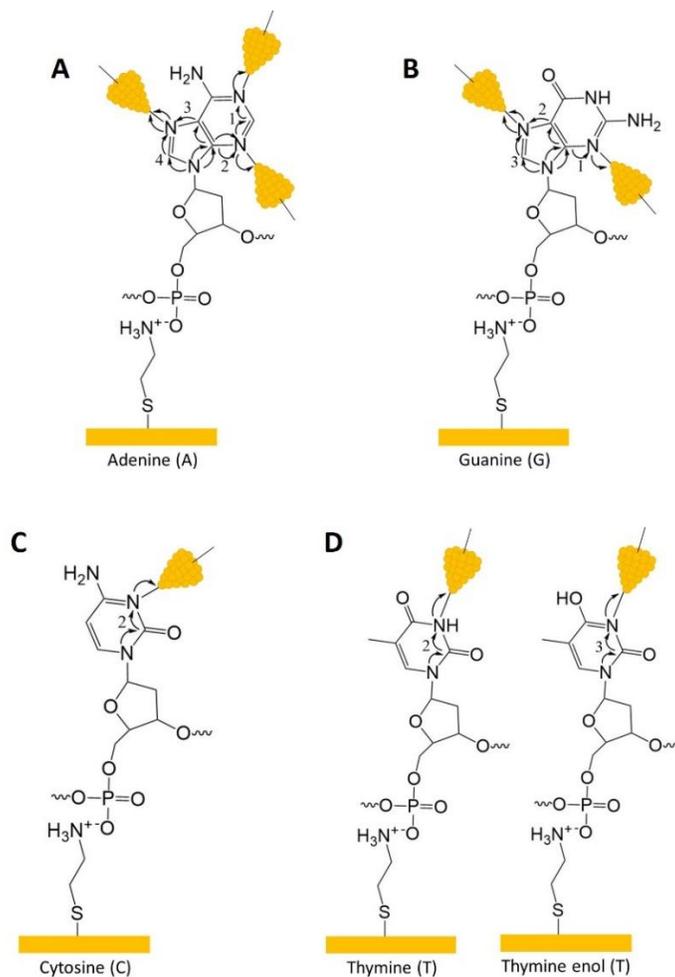


Figure S7. Transmission pathways through nucleotides. (A) Adenine (A) – 4 transmissions. (B) Guanine (G) – 3 transmissions. (C) Cytosine (C) – 1 transmission. (D) Thymine (T) – 2 transmissions (one for each keto-enol tautomer). Numbers for each pathway correspond to peaks in the histogram signatures (Fig. 1A).

As described in the main text, we used single STM-BJ current-distance measurements to calculate a distribution of transmissions for varying smear parameter S_p bins as a way to verify molecular smear. The schematic detailing this process with an example shown for deoxyadenosine is shown in Fig. S8.

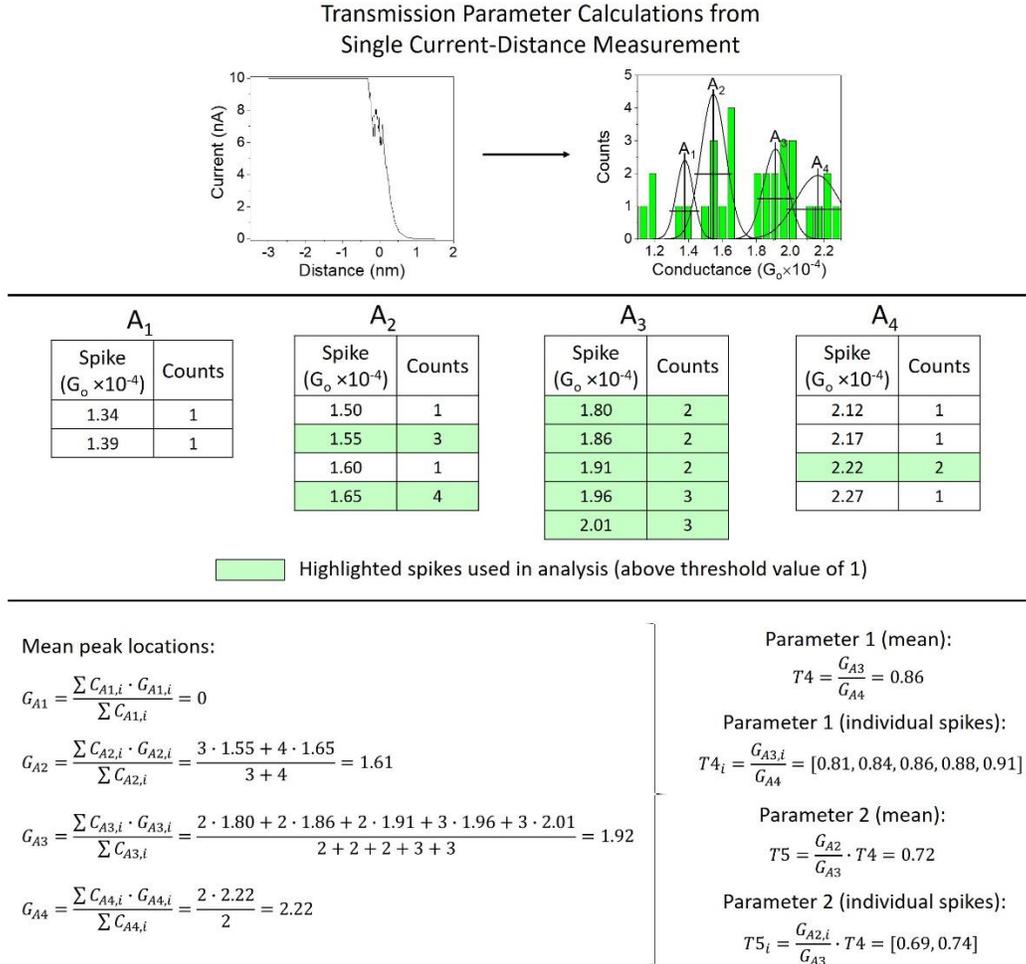


Figure S8. Transmission calculations for single STM-BJ measurements. After creating a histogram from a single current-distance measurement, we identified conductance values from spikes greater than one count in the histogram within \pm FWHM of the known signature peak positions for A (A₁, A₂, A₃, and A₄). Then, we calculated a mean peak location weighted by the counts for each spike, e.g., $G_{A4} = \sum C_{A4,i} \cdot G_{A4,i} / \sum C_{A4,i}$. From the weighted mean peak locations, we then obtained a mean transmission coefficient value (e.g., $T4 = G_{A3}/G_{A4}$ and $T5 = (G_{A2}/G_{A3}) \cdot T4$) and also a transmission coefficient value for individual spikes (e.g., $T4_i = G_{A3,i}/G_{A4}$ with count $C_{A3,i}$ and $T5_i = (G_{A2,i}/G_{A3}) \cdot T4$ with count $C_{A2,i}$). The transmission coefficient values for individual spikes for all respective deoxyadenosine nucleotide measurements were combined to give the distributions shown in Fig. 2B.

Further verification of the proposed transmission model is seen by adjusting pH to attenuate molecular junction formation. As demonstrated for A and C, by reducing the pH below the pKa (4.1 for A and 4.4 for C), the nucleobase becomes protonated.⁴ Without the available lone pair electrons, the gold tip can no longer form molecular junctions. Evidence can be seen in the reduction of peaks in the conductance histograms (Fig. S9). This supports the junction locations noted in the transmission model above.

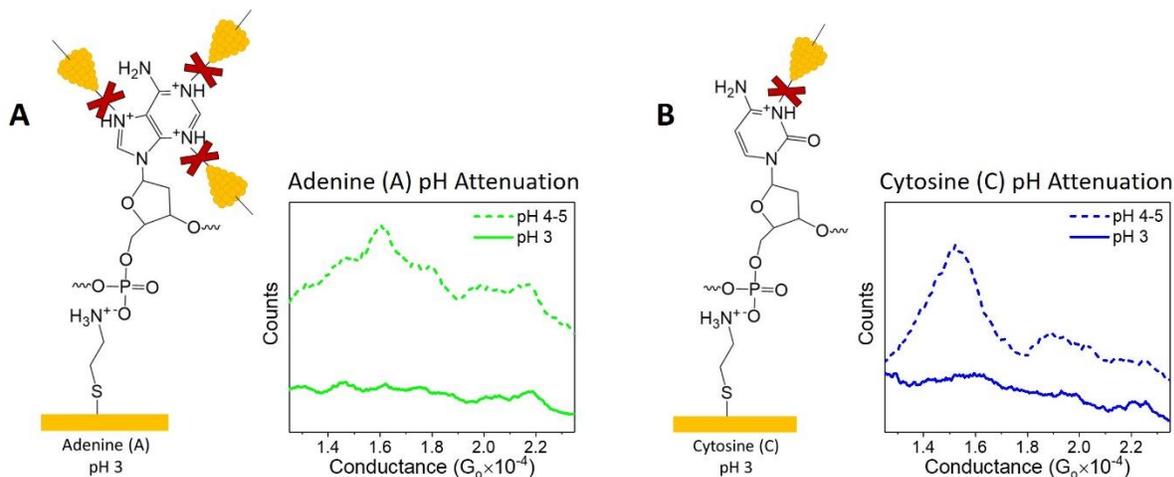


Figure S9. Junction verification with pH perturbation. The histogram peaks seen for (A) A and (B) C are attenuated at pH 3 due to protonation of the nucleobase, as shown in the schematics. This supports the proposed transmission model verifying that junctions are formed between the gold tip and nitrogen groups within the nucleobase.

7. Base calling/molecular recognition calculations

Calculations were performed to determine the base calling/molecular recognition capabilities of our designed algorithm (equation 5 in the main text, results in Figs. 4 and 5). The conductance datasets for STM-BJ measurements on homologous sequences were randomly split into thirds. For each combination of two-thirds, reference signatures were developed as described in the main text; peak locations and their FWHM (from Gaussian fitting in OriginPro 2016), Pearson correlation coefficients, and thresholds were determined. Values for Pearson correlation coefficients and thresholds are shown in Fig. S10 for calculations with no SCRIB and with SCRIB. Each of the one-third sections of STM-BJ spectra were then introduced into a MATLAB program implementing the base calling algorithm, with the opposite two-thirds sections used as reference signatures. By using separate testing and training datasets, the calculations benchmark our algorithm on its ability to accurately identify signals from unknown STM-BJ measurements. Results are shown in Fig. 4 and 5, with accuracy values calculated from 200 random combinations of a variable number of single STM-BJ spectra for each nucleobase (800 total base calls per x-axis value in Fig. 4). The accuracy reported for each nucleobase is the percent recall from a confusion matrix analysis: True Positives/(True positives + False Negatives)·100%. As

seen in Fig. 4, accuracy increases as more spectra are used per base call, analogous to coverage reported by other sequencing technologies. Detailed output from the base calling algorithm (including probability values, confidence of base calling, and accuracy from the oligomer measurements) are shown in Figs. S11-13 at 20× coverage for no SCRIB and SCRIB, and at 7× coverage for low smear measurements only. These plots show all 800 base calls, whereas Fig. 5A-C only shows a subset of the base calls. The correct sequence of calls is given in Fig. S14.



Figure S10. Correlation coefficients and thresholds for base calling calculations. Color coded matrices are Pearson correlation coefficients (ranging from -1 to 1, with extreme values indicating perfectly negative and positive linear correlations, respectively). White tables are threshold values.

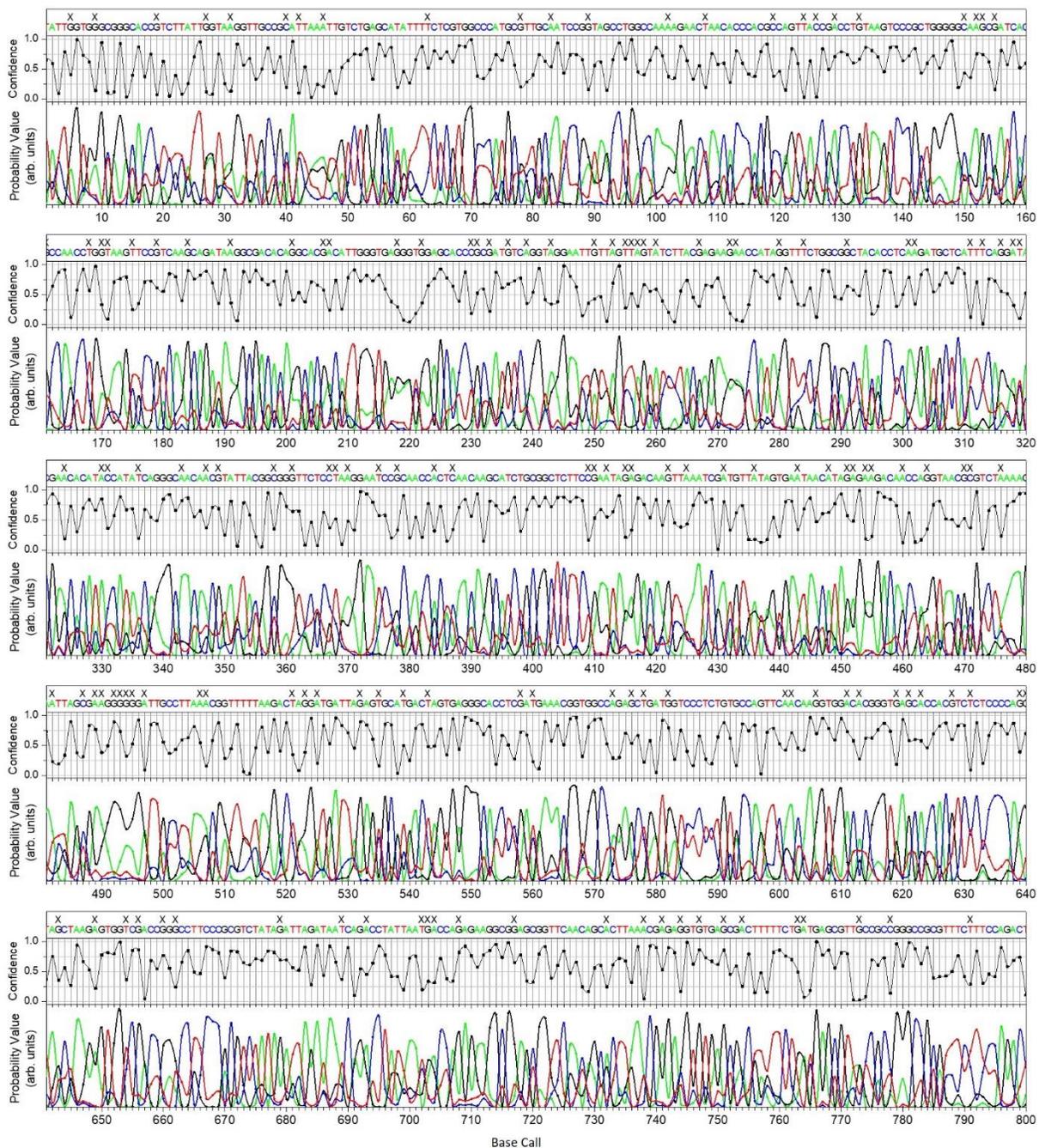


Figure S11. Detailed base calling output with 20× coverage for no SCRIB. Probability values (obtained from the base calling algorithm), confidence of base calling, and accuracy (X indicates incorrect calls) for the complete set of 800 base calls using 20× coverage for no SCRIB (for one of the three repeated calculations with results in Figs. 4A and 5A).

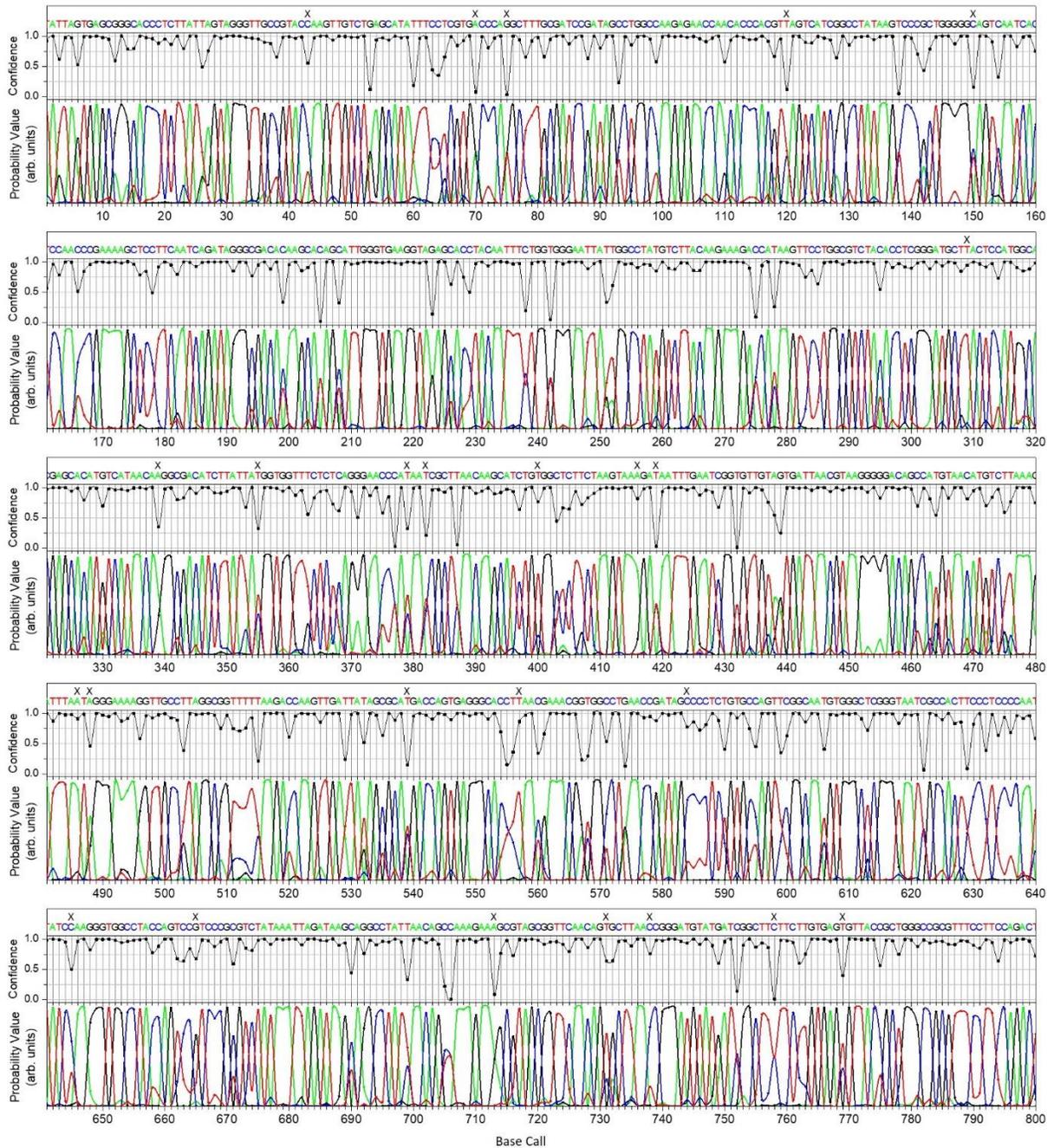


Figure S12. Detailed base calling output with 20× coverage for SCRIB. Probability values (obtained from the base calling algorithm), confidence of base calling, and accuracy (X indicates incorrect calls) for the complete set of 800 base calls using 20× coverage for SCRIB (for one of the three repeated calculations with results in Figs. 4B and 5B).

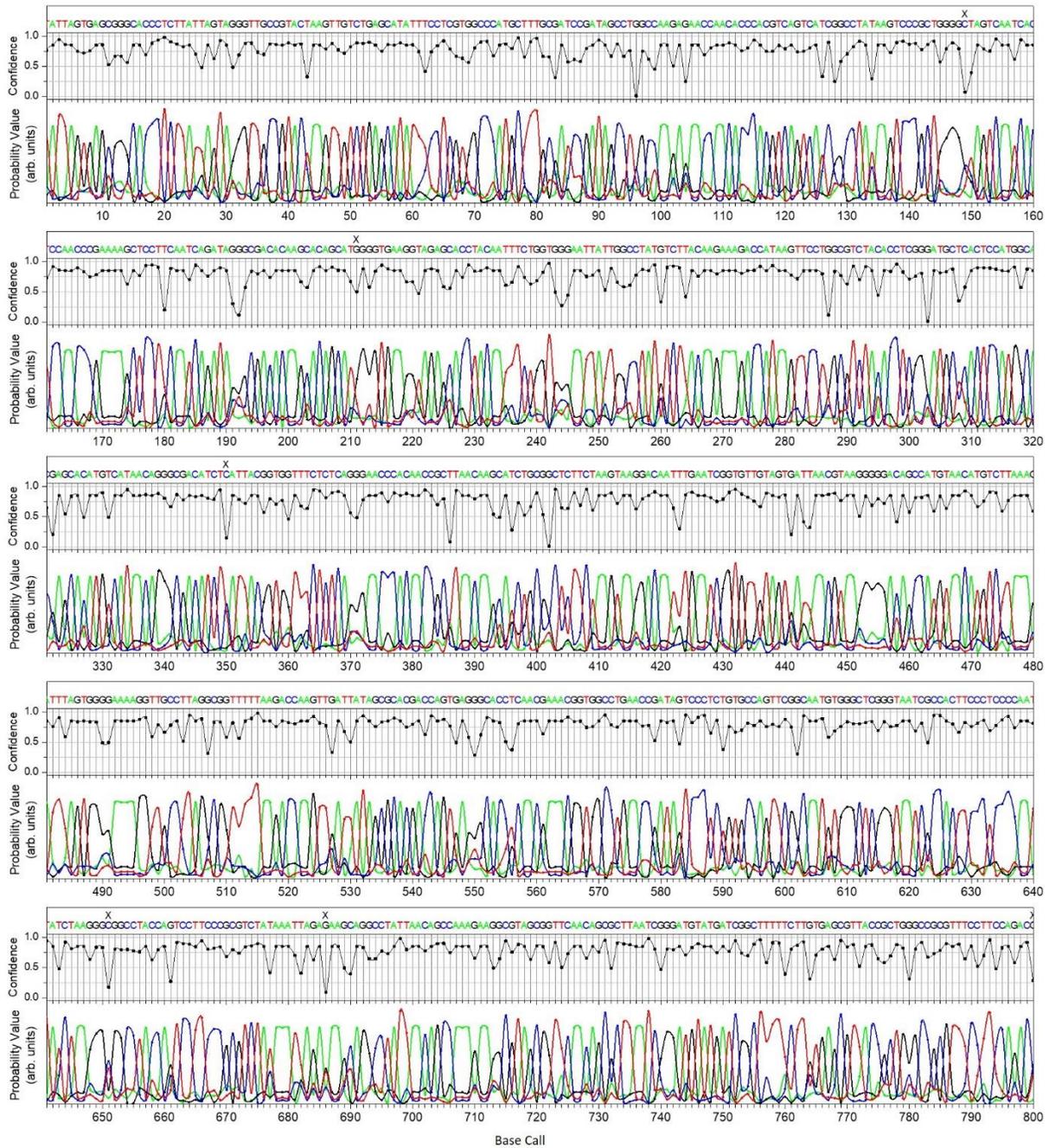


Figure S13. Detailed base calling output with 7× coverage for low smear. Probability values (obtained from the base calling algorithm), confidence of base calling, and accuracy (X indicates incorrect calls) for the complete set of 800 base calls using 7× coverage for low smear (for one of the three repeated calculations with results in Figs. 4C and 5C).

T	A	T	T	A	G	T	G	A	G	C	G	G	C	A	C	C	C	T	C	T	T	A	T	T	A	G	T	A	G	G	T	T	G	C	C	G	T				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40		
A	C	T	A	A	G	T	T	G	T	C	T	G	A	G	C	A	T	A	T	T	T	C	C	T	C	G	T	G	G	C	C	C	A	T	G	C	T	T	T		
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80		
G	C	G	A	T	C	C	G	A	T	A	G	C	C	T	G	G	C	C	A	A	G	A	G	A	A	C	C	A	A	C	A	C	C	C	A	C	C	G	T	C	
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120		
A	G	T	C	A	T	C	G	G	C	C	T	A	T	A	A	G	T	C	C	C	G	C	T	G	G	G	T	A	G	T	C	A	A	T	C	A	C	A	C		
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160		
T	C	C	A	A	C	C	C	G	A	A	A	A	G	C	T	C	C	T	T	C	A	A	T	C	A	G	A	T	A	G	G	G	C	G	A	C	A	C	A		
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200		
A	G	C	A	C	A	G	C	A	T	T	G	G	G	T	G	A	A	G	T	A	G	G	A	G	C	A	C	C	T	A	C	A	A	T	T	T	C	T	G		
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240		
G	T	G	G	G	A	A	T	T	A	T	T	G	G	C	C	T	A	C	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280		
G	T	T	C	C	T	G	G	C	G	T	C	T	A	C	A	C	C	T	C	G	G	G	A	T	G	C	T	C	A	C	T	C	C	A	T	G	G	C	A		
281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320		
C	G	A	G	C	A	C	A	T	G	T	C	A	T	A	A	C	A	G	G	C	A	G	G	A	C	A	T	C	T	T	A	T	T	A	C	G	G	T	G	G	
321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360		
T	T	T	C	T	C	T	C	A	G	G	G	A	A	C	C	C	A	C	A	A	C	C	C	A	A	C	C	T	T	A	A	A	A	G	C	A	T	C	T	G	C
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400		
G	G	C	T	C	T	T	C	T	A	A	G	T	A	A	G	G	A	C	A	A	T	T	T	G	A	A	T	C	G	G	T	G	T	T	G	T	A	G	T		
401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440		
G	A	T	T	A	A	C	G	T	A	A	G	G	G	G	G	A	C	A	G	C	C	A	T	G	T	A	A	C	A	T	G	T	C	T	T	A	A	A	A	G	
441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480		
A	T	T	T	A	G	T	G	G	G	G	A	A	A	A	G	G	T	T	G	C	C	T	T	A	G	G	C	G	G	T	T	T	T	T	A	A	G	A	C		
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520		
C	A	A	G	T	T	G	A	T	T	A	T	A	G	C	G	C	A	C	G	A	C	C	A	G	T	G	A	G	G	G	C	A	C	C	T	C	A	A	C		
521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560		
G	A	A	A	C	G	G	T	G	G	C	C	T	G	A	A	C	C	G	A	T	A	G	T	C	C	C	T	C	T	G	T	G	C	C	A	G	T	T	C		
561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600		
G	G	C	A	A	T	G	T	G	G	G	C	T	C	G	G	G	T	A	A	T	C	G	C	C	A	C	T	T	C	C	C	T	C	C	C	A	A	T			
601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640		
T	A	T	C	T	A	A	G	G	G	T	G	G	C	C	T	A	C	C	A	G	T	C	C	T	T	C	C	C	G	C	G	T	C	T	A	T	A	A	A		
641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680		
T	T	A	G	A	T	A	A	G	C	A	G	G	C	C	T	A	T	T	A	A	C	A	G	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720		
G	G	T	T	C	A	A	C	A	G	C	G	C	T	T	A	A	T	C	G	G	A	T	G	T	A	T	G	A	T	C	G	G	C	T	T	T	T	T	T		
721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760		
C	T	T	G	T	G	A	G	C	G	T	T	A	C	C	G	C	T	G	G	C	C	G	C	G	T	T	T	C	C	T	T	C	C	A	G	A	C	T			
761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800		

Figure S14. Correct sequence of base calls for SCRIB analysis. Displaying the correct nucleobase sequence for the base calls in Figs. S11-13.

References

- (1) Li, X.; He, J.; Hihath, J.; Xu, B.; Lindsay, S. M.; Tao, N. *J. Am. Chem. Soc.* **2006**, *128* (6), 2135.
- (2) Engelkes, V. B.; Beebe, J. M.; Frisbie, C. D. *J. Am. Chem. Soc.* **2004**, *126* (43), 14287.
- (3) Landauer, R. *Phys. Lett. A* **1981**, *85* (2), 91.
- (4) Verdolino, V.; Cammi, R.; Munk, B. H.; Schlegel, H. B. *J. Phys. Chem. B* **2008**, *112* (51), 16860.