Supplementary material for: Determining the effects of DNA sequence on Hel308 helicase translocation along single-stranded DNA using nanopore tweezers

Contents

1. Salt and voltage dependence of Hel308 translocation 5
2. Ion-current consensuses for heteropolymer measuring sequences 7
3. Supporting measurements of registration distance 9
4. Analysis of the [ATP]-dependent step 12
5. Protein sequence analysis of Hel308 14
6. References 16

Sequence Name	Sequence	N	Gen
homopolymer A	PTACTACTACATTACXXCTTTGTCGTTGTGCAGTCGTTAAAAAAAAAAA AAAAAAAAATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	62	1
homopolymer C	PTACTACTACATTACXXCTTTGTCGTTGTGCAGTCGTTCCCCCCCCCCC CCCCCCCCCCTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	41	1
homopolmyer T	PTACTACTACATTACXXCTTTGTCGTTGTGCAGTCGTTTTTTTTTTTT TTTTTTTTTTTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	99	1
dinucleotide CA	PTACTACTACATTACXXTGTAGTCGTTGTGCAGTCGTTCACACACACAC ACACACACACTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	90	2
dinucleotide AG	PTACTACTACATTACXXTGTAGTCGTTGTGCAGTCGTTAGAGAGAGAGA GAGAGAGAGATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	62	2
dinculeotide TC	PTACTACTACATTACXXTGTAGTCGTTGTGCAGTCGTTTCTCTCTCTCT CTCTCTCTCTTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	56	2
dinucleotide GT	PTACTACTACATTACXXTGTAGTCGTTGTGCAGTCGTTGTGTGTGTGTG TGTGTGTGTGTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	57	2
Spacer1	PTACTACTACATTACXXTGTAGTCGTTGTGCAGTCGTTCCTCAAAATCA GATCTCACTATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	95	2
Spacer2	PTACTACTACATTACXXTGTAGTCGTTGTGCAGTCGTTCCTCAAAATCA CATCTCACTATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	76	2
Spacer3	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTCCTCAAAATCC CATCTCACTATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	35	3
Spacer4	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTCCTCAAAATCC CCTCTCACTATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	52	3
Spacer5	PTACTACTACATTACXXTGTAGTCGTTGTGCAGTCGTTAGAGAGAGAGA CCCCCCCCCCTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	73	3
Spacer6	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTCCCCCCCCCCA CCCCCCCCCCTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	63	3
Grich	PTACTACTACATTACXXTGTAGTCGTTGTGCAGTCGTTGGGAGGTGGGC GGAGGGTGGCTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	29	2
Background 1	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTAAAAAAAAAGG GCCCCCCCCCTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	25	3
Background 2	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTCCCCCCCCCGG GAAAAAAAAATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	28	3
Heteropolymer 1	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTGTTTAGCGGAC GTTTAATACTTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	42	3
Heteropolymer 2	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTTGATCGTCCGG CATCCGCCGGTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	41	3
Heteropolymer 3	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTCCCGAGTAGCT TTGGGCGGAGTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	29	3
Heteropolymer 4	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTGTCCCGAGTTT CGGGCGAGCATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	25	3
Heteropolymer 5	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTCATCAATTAAT GTGTCTCCCCTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	21	3
Heteropolymer 6	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTTCCCCGAGCTA TCAGTAACTTTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	28	3
Heteropolymer 7	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTTGCCGGGCTTC GTAATTGTAATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	28	3
Heteropolymer 8	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTGCTCCCGGTGC AGCCGTGGGTTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	25	3
Heteropolymer 9	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTAGTCCTTATAA TGAAGGAATATGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	34	3
Heteropolymer 10	PTACTACTACATTACXXTGTAGTCGTAGTGCAGTCGTTGGGATTCCGCC GGTGGGCAAGTGGTATCTCACTATCGCATTCTCATGCAGGTCGTAGCC	40	3

TABLE S1. Experiment statistics. The first column lists the experiment name, the second column lists the corresponding DNA sequence. 'X' and 'P' refer to an abasic marker and phosphate group, respectively. All sequences are written 5′ to 3′. N is the number of analyzed recordings with hand-corrected alignments of each experiment (described in more detail in (13) SI Appendix 1). Gen refers to the generation of measuring sequence used (see Fig. s5).

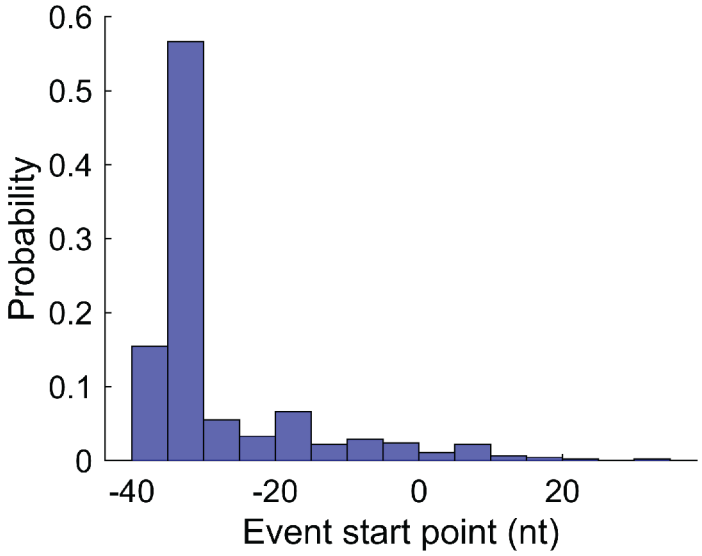

FIGURE S1. Distribution of event start points for 452 reads of Hel308 on the DNA sequence 'Registration Distance 1'. The position axis is displayed as in figure 1C of the main text. The DNA sequence extends to position -60 *nt*, however we do not observe the sequence between -60 *nt* and -40 *nt* because during event initialization, approximately 10 *nt* are bound by Hel308, and a further ≈ 15 *nt* are in the MspA vestibule. Because the DNA is pulled out of MspA during these experiments, these nucleotides never pass through the constriction. We therefore conclude that most events are observed starting close to the location at which Hel308 binds to the DNA.

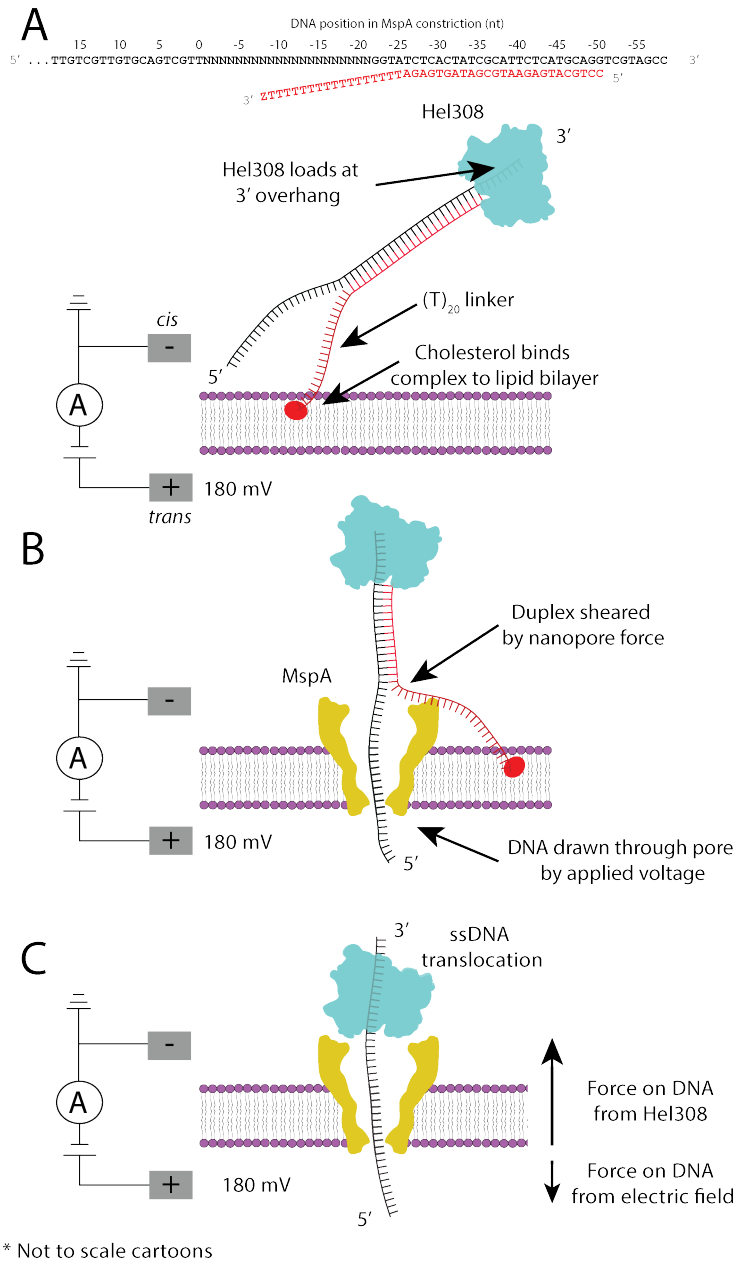

FIGURE S2. Cartoon schematic of Hel308 SPRNT experiment. (A) Hel308 binds to a free 3′ end of the template DNA strand. A cholesterol tag on the 3′ end of the non-template sequence anchors the complex to the bilayer, increasing the local density of DNA near the nanopore. (B) The 5′ end of the template DNA is drawn through the pore, with the duplex being sheared by the nanopore force. (C) Hel308 moves from 3′ to 5′, drawing ssDNA out of the nanopore. The arrows reflect the forces on the DNA during the experiment. The force on the DNA from Hel308 on the DNA must be larger than that of the electric field, since the average direction of DNA motion is out of the pore.

1. Salt and voltage dependence of Hel308 translocation

To test the dependence of Hel308 on salt concentration we reduced the *cis* [KCl] from 400 mM to 100 mM for the DNA sequence 'registration distance 1' (Table s1). Figure s3a shows the probability of a backwards step vs. DNA position. The backstepping shows the same pattern of backwards step, indicating independence of salt concentration between these conditions. Figure s3b shows the dwell time vs. DNA position. The two curves track each other well, with the dwell-time at 400 mM [KCl] being slightly longer. This deviation is consistent with experiment-to-experiment fluctuations in temperature. There appears to be minimal effects of [KCl] on Hel308 kinetics.

We also extended our previous study of voltage-dependence (13) of Hel308 kinetics by measuring Hel308 translocation over the measuring sequence 'registration distance 1' at a voltage of 60 mV. At this voltage the nanopore signal-to-noise ratio is substantially reduced and the registration distance changes by more than a nucleotide at these conditions (15), making step analysis difficult. We do note that on some parts of the DNA sequence we can still resolve some steps. However, the topography of the ion-current signal is maintained (Fig. s5), so that the series of peaks and troughs is still observable. We therefore measure the peak-to-peak duration between DNA position +7 nt and position +22 nt (Fig s5), which is still well-resolved. Figure s4 shows the distribution of total durations for 'registration distance 1' taken at 100 mM [KCl] for voltages 180 mV and 60 mV, corresponding to forces of approximately 35 pN and 12 pN, respectively. We used the 2-sample KS test to compare these distributions, and find that they are nearly indistinguishable ($p = 0.09$). There is an excess in low total durations in the data taken at 60 mV, however given the difficulty of accurately assigning DNA position at nucleotide resolution in these low force and [KCl] experiments, it may be that our selection of peak positions could be off by 1/2 - 1 nt, leading to an effective shorter read region, in nucleotides, leading to an excess of short total durations. Similarly, the most commonly observed error mode in the DNA sequences we read is nucleotide deletion, which at higher voltages can be easily identified by the absence of consecutive ion-current states. These events are normally excluded from the analysis, however in the low force experiments identifying such deletions is difficult. A deletion will result in a decrease in the number of nucleotides translocated between peaks, leading to an effective decrease in the dwell-time. Given the results presented here, we conclude that Hel308 kinetics are nearly independent of the applied force over the range of 12 pN to 65 pN.

Lastly, we note that we still observe Hel308 backwards steps in these conditions, although quantifying them in the same way as in the rest of the study requires development of the current-to-sequence maps and measurements of the registration distance at low force.

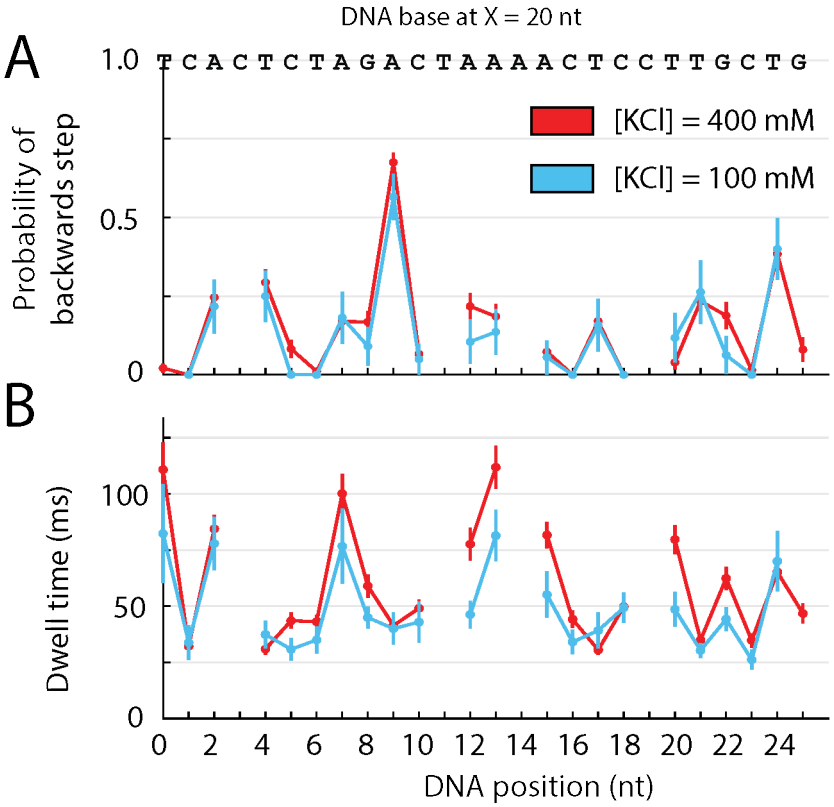

FIGURE S3. Salt dependence of Hel308 translocation. (A) Probability of a backwards step vs. DNA position for the DNA sequence 'registration distance 1' taken at 400 mM KCl (red, N = 95) and 100 mM KCl (blue, N = 23). (B) Dwell-time vs. DNA position for the same sequence and conditions as above.

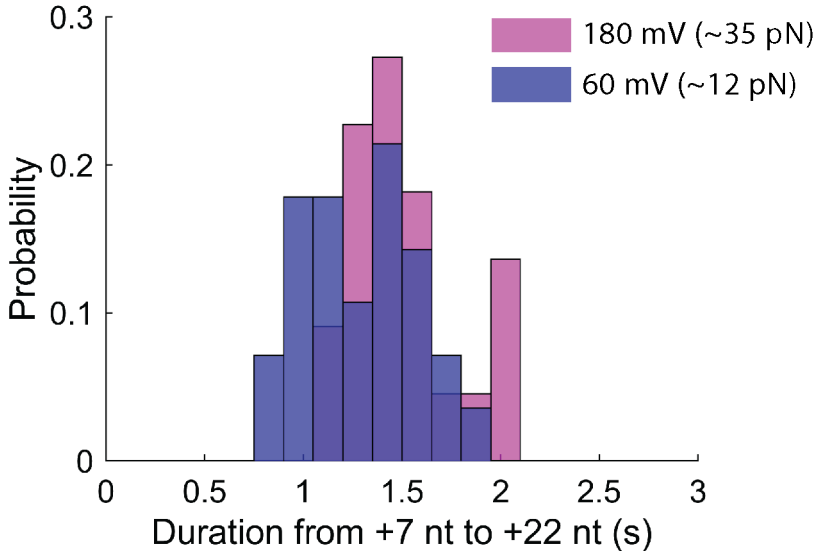

FIGURE S4. Voltage dependence of Hel308 translocation. The distribution of total durations between DNA position +7 and +22 for Hel308 translocation over the DNA sequence 'resgistration distance 1' at 180 mV applied voltage (pink, N = 28) and 60 mV (blue, N = 22).

2. Ion-current consensuses for heteropolymer measuring sequences

Ion current consensuses were generated using established methods (13, 15, 16) and are shown for the heteropolymer section of the DNA in figure s5. As we performed experiments, we improved the heteropolymer design to produce higher-contrast ion-current states. The 'generation' in table s1 refers to the three different heteropolymer sequences used in this manuscript. We previously verified that the kinetics were independent of the DNA in the constriction(13), and subsequent generations of measuring sequences were made to improve the sensitivity of SPRNT measurements.

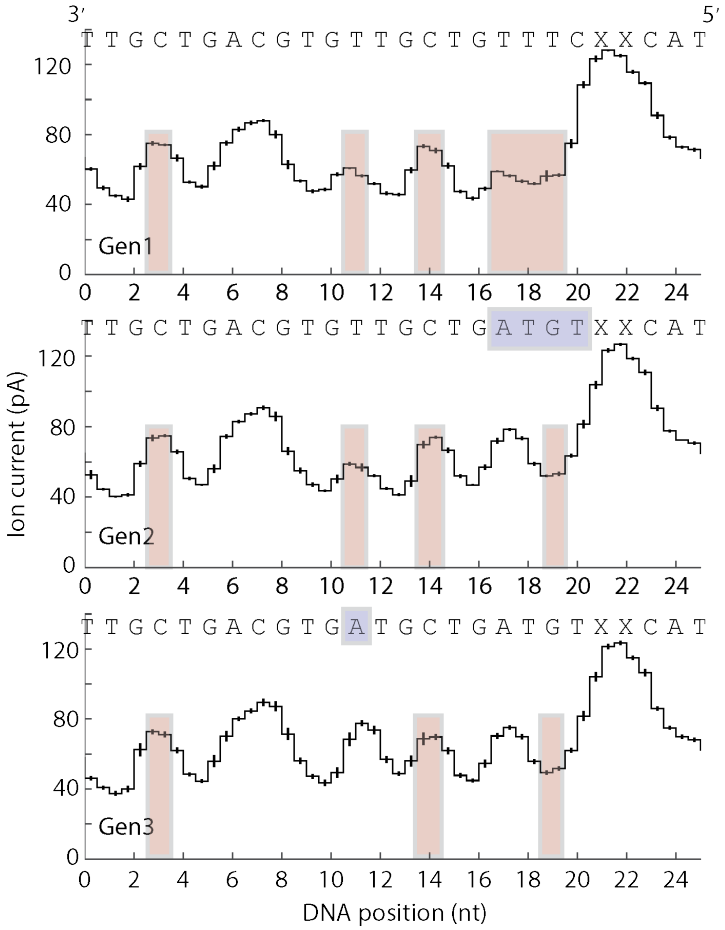

FIGURE S5. Ion current vs. DNA position for each 'measuring sequence' generation. (top) generation 1, (middle) generation 2, and (bottom) generation 3. Blue shaded letters indicate a change in the measuring sequence from the previous generation. Red shaded sections indicate ion-current states omitted from kinetic analysis. The gaps in the main text correspond to these sections. Error bars are S.E.

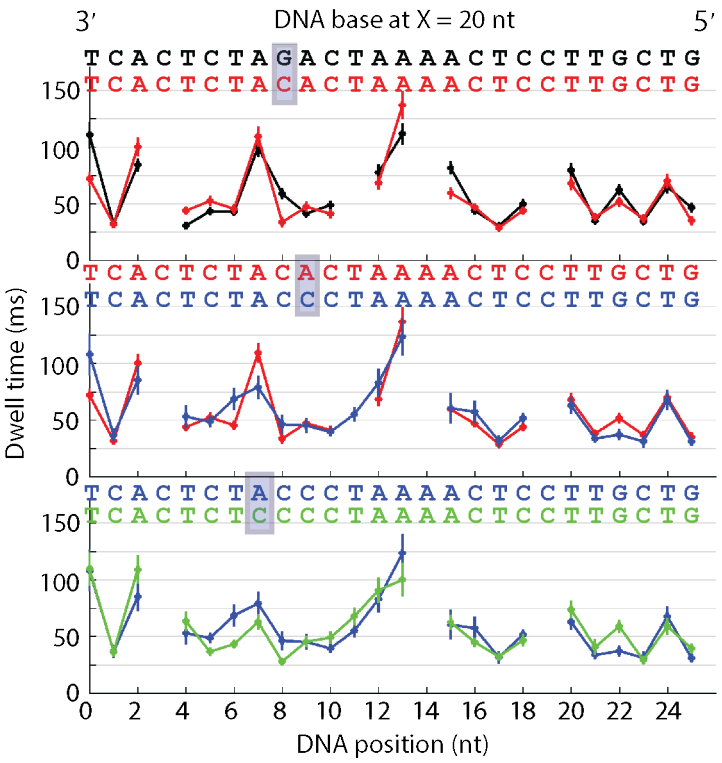

FIGURE S6. Dwell-time vs. DNA position for the same three DNA sequences shown in main text figure 3. Each panel is a sequential mutation of the DNA sequence 3′ ...AGA... 5′.

3. SUPPORTING MEASUREMENTS OF REGISTRATION DISTANCE

We performed two additional experiments to confirm the registration distance of 20 nt. In the first we used a mixed test sequence of the form 3′...CCCCCAGAG... 5′ (Fig. s7A). We determined the registration by offsetting each DNA sequence to the pattern of backwards steps among the three sequences. We found the probability of a backwards step to be small when cytosines were at X = 20 nt, and large when the adenines were at X = 20 nt. Second, we substituted a single adenine in a homopolymer cytosine background. Comparing this sequence to the homopolymer cytosine, the probability of backwards step and dwell-time both change when the registration is set at 20 nt (Fig. s7B).

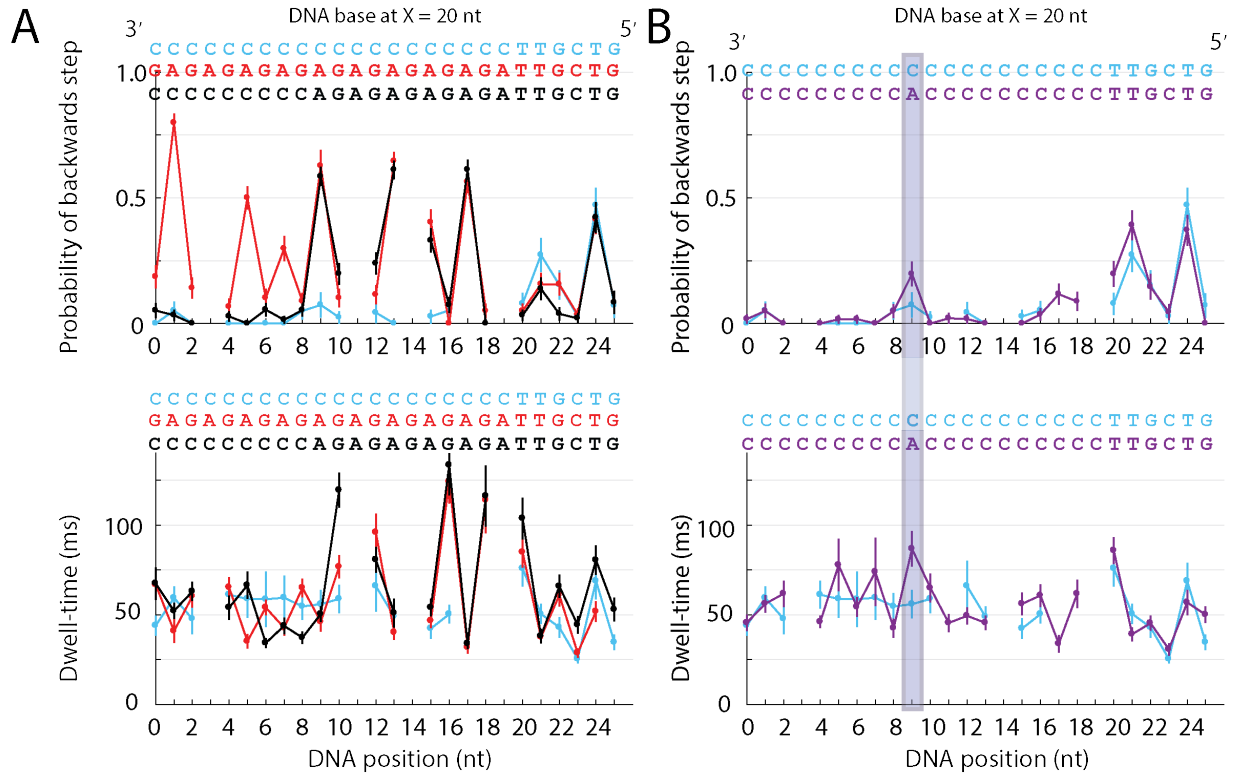

FIGURE S7. Analysis of the registration distance. (A, top) Probability of a backwards step vs. DNA position for three DNA sequences: homopolymer C (blue), dinucleotide AG (red) and a concatenation of the two (black). (A, bottom) Dwell-time vs. DNA position for the same sequences as above. (B, top) Probability of a backwards step vs. DNA position for two DNA sequences: homopolymer C (blue) and homopolymer C with a single adenine substitution (purple). (B, bottom) Dwell-time vs. DNA position for the sequences shown above. The registration distance has been set at X = 20 nt.

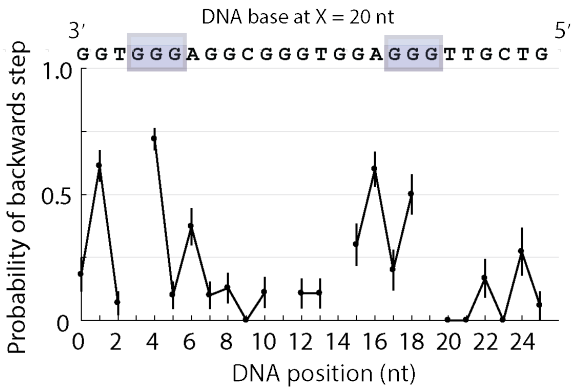

FIGURE S8. Probability of a backwards step vs. DNA position for a G-rich DNA sequence. The registration distance has been set at X = 20 nt. 'GGG' motifs are highlighted in blue. The central G in each triplex causes a large probability of a backwards step.

4. Analysis of the [ATP]-dependent step

In a previous study we found that the Michaelis-menten kinetics of the [ATP]-dependent step varied along a heterogeneous DNA sequence. Figure s9 shows the average dwell-time of the [ATP]-dependent step at saturating [ATP] conditioned on the X^{th} nucleotide above the MspA constriction, as in figure 5 in the main text. Unlike the [ATP]-independent step, the [ATP]-dependent step shows no obvious dependence on DNA sequence.

As a simple test of whether DNA sequence was affecting the [ATP]-dependent step kinetics we returned to the homopolymer and dinucleotide test sequences used in figure 2. We partitioned the data sets into even and odd half integer positions for each homopolymer and dinucleotide sequence (that is, data set 1 contains positions 1.5, 3.5, 5.5, ... and data set 2 contains positions 2.5, 4.5, 6.5, ...). We did not compare across experiments because of the possibility of fluctuations in ATP concentration. We used the two sample KS-test to compare each pair of distribution functions (figure 9), and found that for each of the homopolymer experiments the two data sets were indistinguishable at $p > 0.05$ confidence, whereas three of the four the dinucleotide experiments showed statistically significant differences between the two distributions, suggesting that the DNA sequence in Hel308 also modifies the [ATP]-dependent step kinetics.

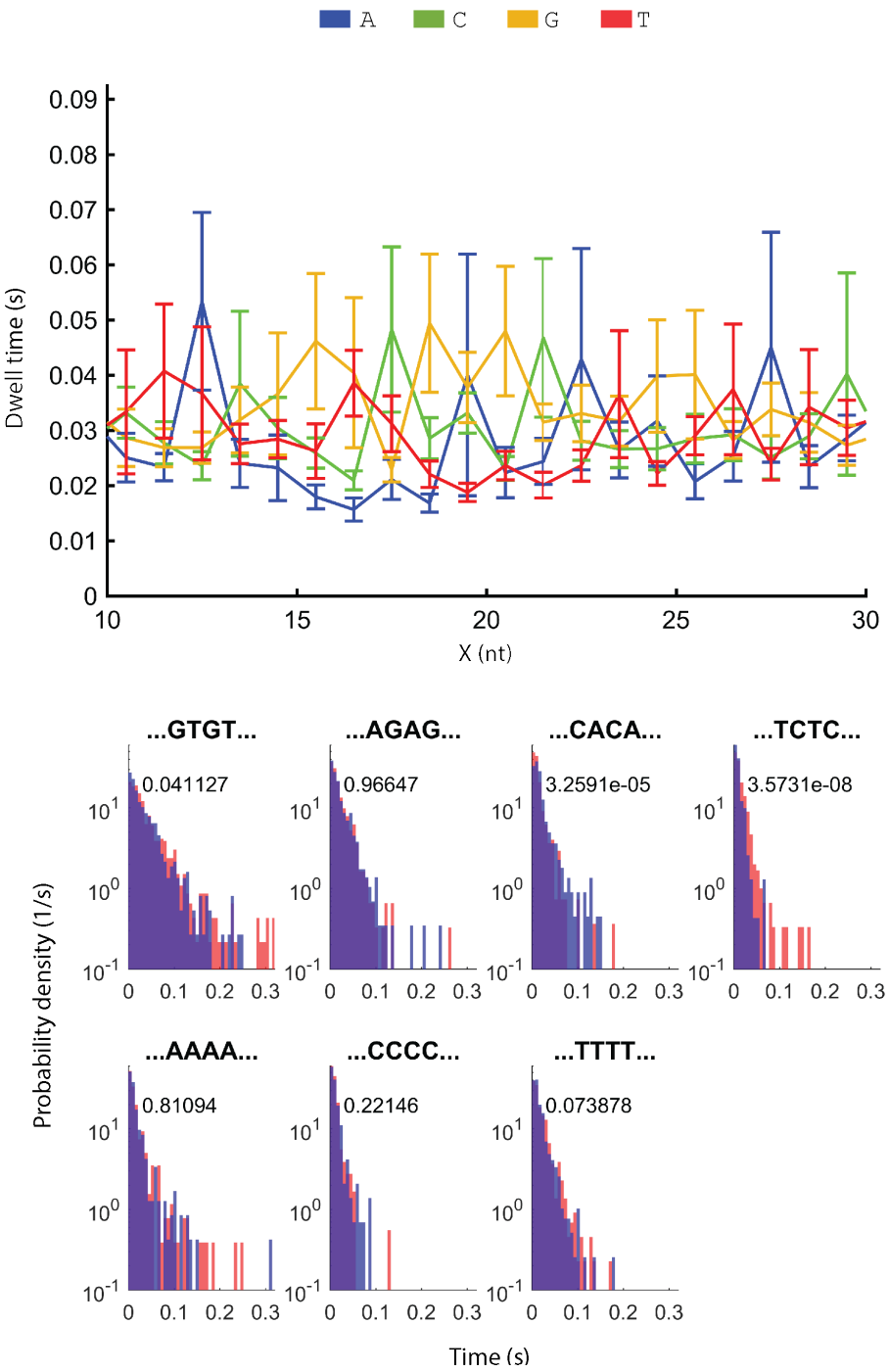

FIGURE S9. (top) Average dwell-time of [ATP]-dependent steps conditioned on the X^{th} nucleotide above the MspA constriction. (bottom)Dwell-time distributions for DNA positions 1.5, 3.5, 5.5 (blue) and 2.5, 4.5, and 6.5 (red) for the corresponding dinucleotide test sequences. The p-value for the two-sample KS test comparing the histograms is shown on each panel.In the dinucleotide repeat test sequences, separating the data into even and odd half-integer positions leads to significantly different distributions ($p < 0.05$) in 3/4 cases. In the homopolymer test sequences, separating the data into these bins has no effect, but in the dinucleotide experiments we see differences, indicating that the DNA sequence does effect kinetics in the [ATP]-dependent step.

5. Protein sequence analysis of Hel308

We compared the protein sequences of Hel308 from *Archaeoglobus fulgidus* and *Thermococcus gammatolerans* to see if the residues interacting with the DNA bases as imaged by X-ray crystallography in *A. fulgidus* (23) were conserved between the two sequences. Figure s10 shows the alignment of the two protein sequences to one another. We color the alignment at positions where the amino acid residues interact with the DNA as determined in (23, their Fig. 1C). In spite of having just 56% sequence similarity, 22/24 (91%) positions that interact with the DNA in *Archaeoglobus fulgidus* are similar in *T. gammatolerans*, suggesting that we can use the structural results from *A. fulgidus* to inform future mutational studies of Hel308 from *T. gammatolerans*.

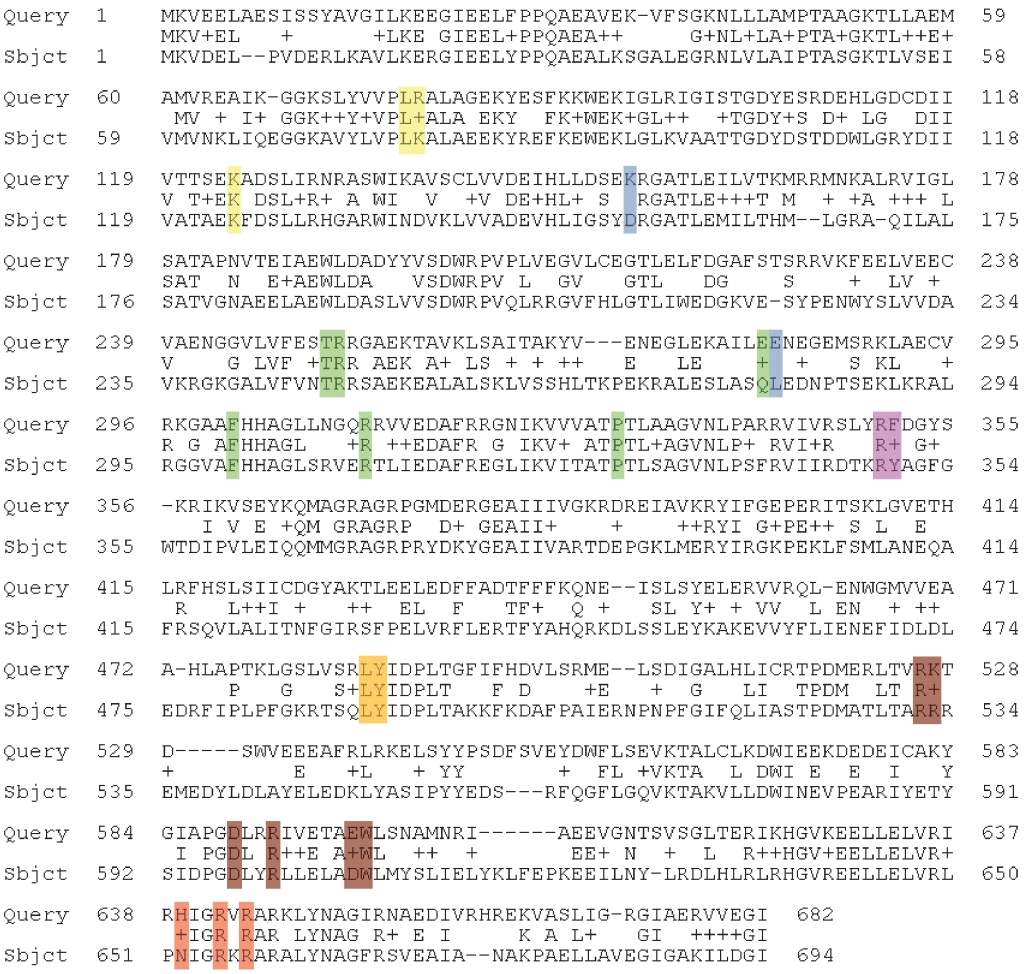

FIGURE S10. Protein BLAST alignment of Hel308 sequences from *Archaeoglobus fulgidus* (query, top row) and *Thermococcus gammatolerans* (subject, bottom row). Colors are as follows: (light yellow) domain 1 match, (light green) domain 2 match (light blue) mismatch (pink) beta-hairpin match (light orange) domain 3 match (brown) domain 4 match (light red) domain 5 match.

6. References

- (13) Craig JM, et al. (2017) Revealing dynamics of helicase translocation on single-stranded DNA using high-resolution nanopore tweezers. Proc Natl Acad Sci 114(45):11932–11937.
- (15) Derrington IM, et al. (2015) Subangstrom single-molecule measurements of motor proteins using a nanopore. Nat Biotechnol 33(10):1073–1075.
- (16) Laszlo AH, Derrington IM, Gundlach JH (2016) MspA nanopore as a single-molecule tool: From sequencing to SPRNT. Methods.
- (17) Butler TZ, Pavlenok M, Derrington IM, Niederweis M, Gundlach JH (2008) Single-molecule DNA detection with an engineered MspA protein nanopore. Proc Natl Acad Sci 105(52):20647–20652.
- (23) Büttner K, Nehring S, Hopfner K-P (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. Nat Struct Mol Biol 14(7):647–652.