SUPPLEMENTARY DATA

A scanning-to-incision switch in TFIIH-XPG induced by DNA damage licenses nucleotide excision repair

Amer Bralić[1], Muhammad Tehseen[1], Mohamed A. Sobhy[1], Chi-Lin Tsai[2], Lubna Alhudhali[1], Gang Yi[1], Jina Yu[3], Chunli Yan[3], Ivaylo Ivanov[3], Susan Tsutakawa[4], John A. Tainer[4,5*], and Samir M. Hamdan[1*]

[1]Bioscience Program, Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

[2]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA.

[3]Department of Chemistry, Georgia State University, Atlanta, GA, 30302 USA; Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA, 30302 USA.

[4]Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA.

[5]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA; Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA.

*Correspondence: J.A.T. (email: jtainer@mdanderon.org) and S.M.H (email: samir.hamdan@kaust.edu.sa)

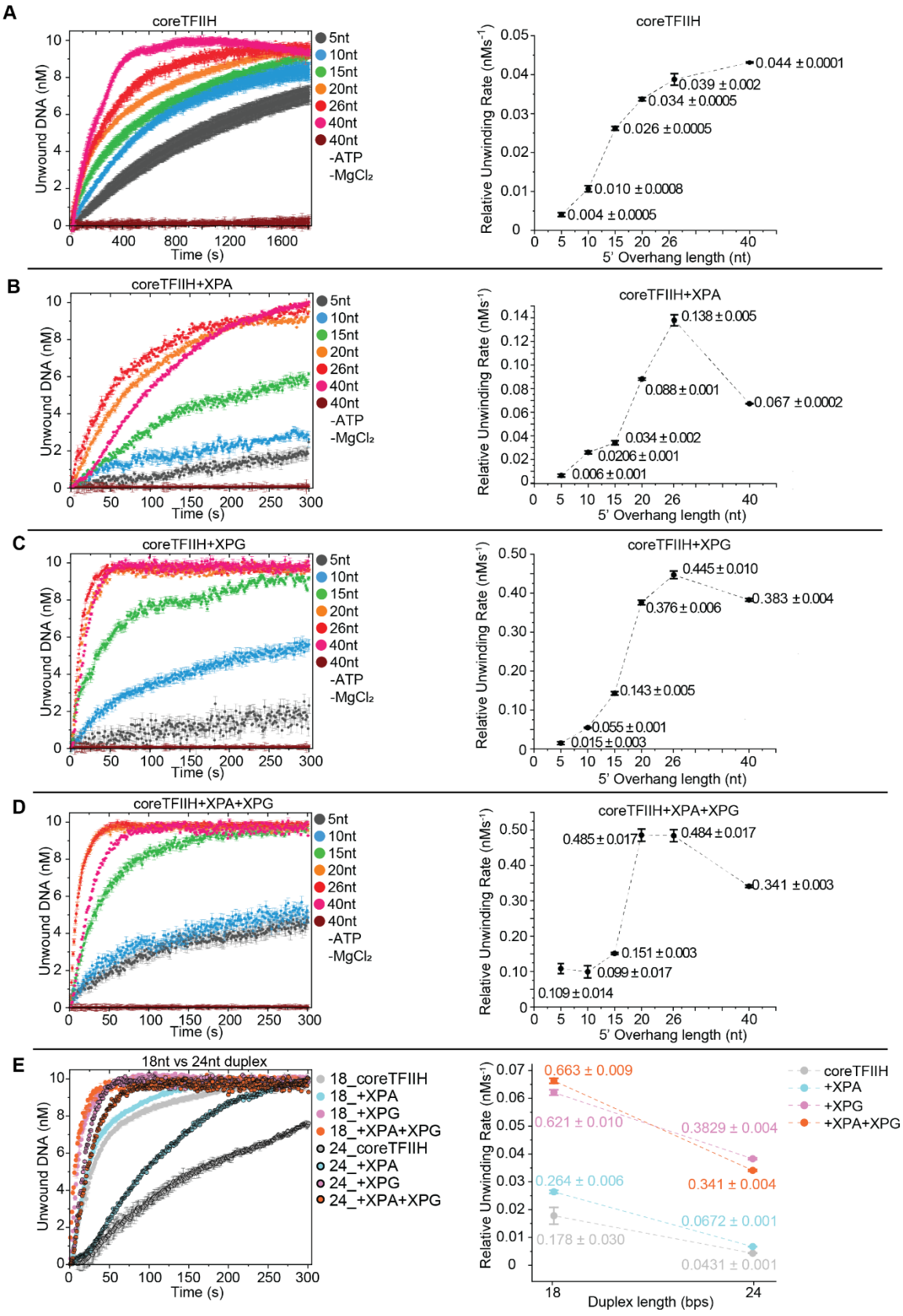
A
coreTFIIH
Unwound DNA (nM)
Time (s)
5nt
10nt
15nt
20nt
26nt
40nt
40nt
-ATP
-MgCl₂
coreTFIIH
Relative Unwinding Rate (nMs⁻¹)
0.044 ±0.0001
0.039 ±0.002
0.034 ±0.0005
0.026 ±0.0005
0.010 ±0.0008
0.004 ±0.0005
5' Overhang length (nt)
B
coreTFIIH+XPA
0.138 ±0.005
0.088 ±0.001
0.067 ±0.0002
0.034 ±0.002
0.0206 ±0.001
0.006 ±0.001
C
coreTFIIH+XPG
0.445 ±0.010
0.376 ±0.006
0.383 ±0.004
0.143 ±0.005
0.055 ±0.001
0.015 ±0.003
D
coreTFIIH+XPA+XPG
0.485 ±0.017
0.484 ±0.017
0.341 ±0.003
0.151 ±0.003
0.099 ±0.017
0.109 ±0.014
E
18nt vs 24nt duplex
18_coreTFIIH
18_+XPA
18_+XPG
18_+XPA+XPG
24_coreTFIIH
24_+XPA
24_+XPG
24_+XPA+XPG
0.663 ± 0.009
0.621 ± 0.010
0.3829 ± 0.004
0.341 ± 0.004
0.264 ± 0.006
0.0672 ± 0.001
0.178 ± 0.030
0.0431 ± 0.001
coreTFIIH
+XPA
+XPG
+XPA+XPG
Duplex length (bps)

Figure S1: Bulk fluorescence unwinding assays of different 5' overhang substrates

Substrates with a 5 nt (grey), 10 nt (blue), 15 nt (green), 20 nt (orange), 26 nt (red), and 40 nt (magenta) 5' overhang were tested for unwinding under conditions with **(A)** coreTFIIH only, **(B)** coreTFIIH+XPA, **(C)** coreTFIIH+XPG, and **(D)** coreTFIIH+XPA+XPG. The coreTFIIH-only condition required a longer reaction period to obtain a plateau; hence, measurements were obtained for 1800 s. For the other conditions, a period of 300 s was sufficient to observe a plateau. A plateau represents 100% of unwound substrate as determined from the fluorescence intensity of an equimolar amount of free fluorescent oligo. A negative control with the 40 nt overhang but without Mg and ATP is shown in maroon. Each data point represents the average of three independent experiments, and the error bars represent the SEM. The y-axis in the time plots (left panels) is expressed in nM corresponding to 100% fluorescence intensity emitted when all of the DNA is unwound (10 nM in a 12 µL reaction volume). The panels on the right contain the relative unwinding rates using 400 nM protein obtained from linear fits of the linear regions in the time plots. The relative rate was plotted against the 5' overhang length. **(E)** Time plots for DNA unwinding under different protein conditions were performed with DNA duplexes of 24 bp or 18 bp, both with a 5' overhang of 40 nt. The relative unwinding rates are indicated in the right panel.

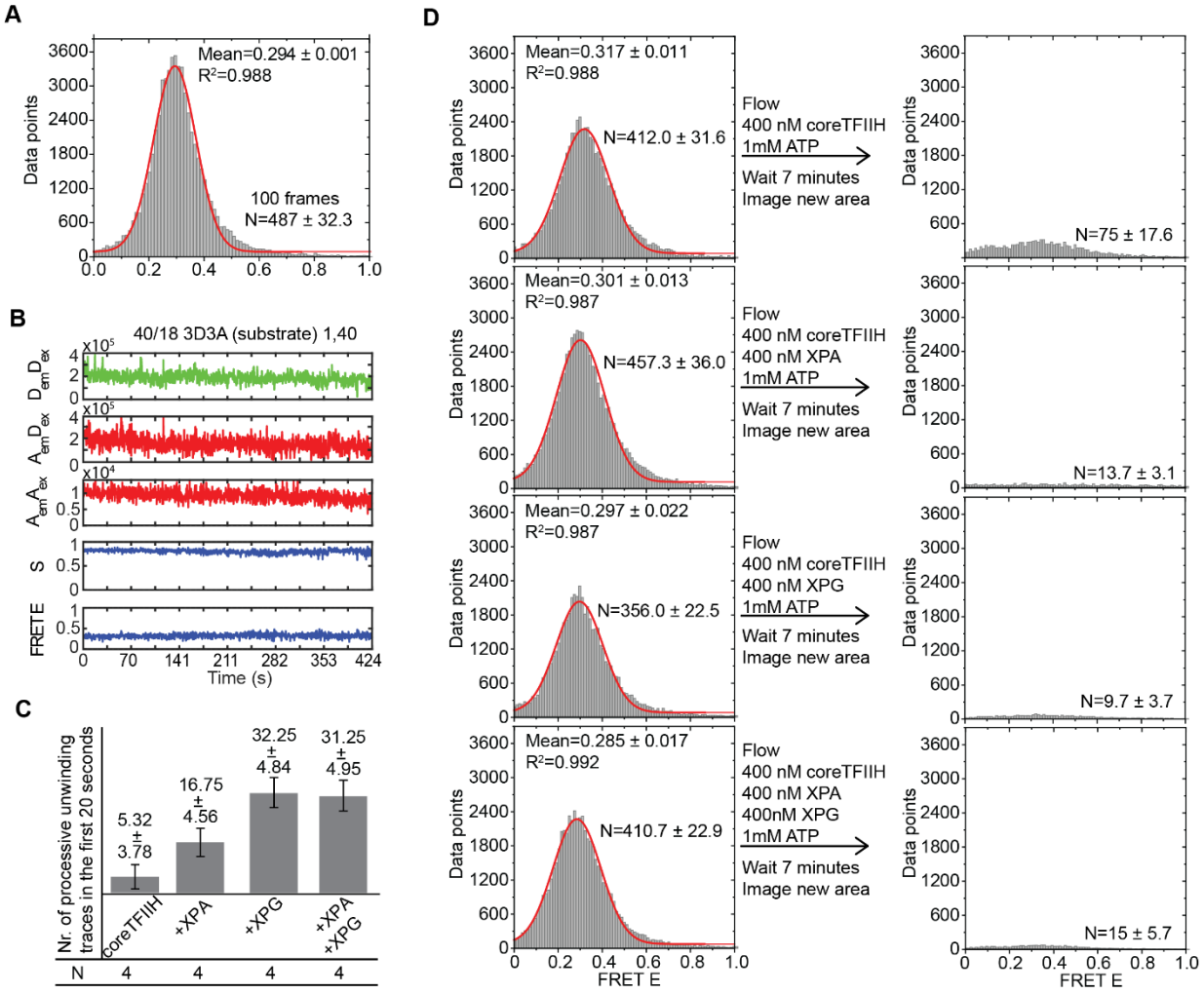

Figure S2: Relative initial unwinding efficiency using smFRET and characterization of the 40 ss/18 ds 3D3A substrate

(A) Average FRET E of the 3D3A substrate (donor and acceptor are located at the 3' ends) under no-flow conditions and in the presence of 1 mM ATP. The peak center is at ~0.3 and the error bars represent the SE of the fit. The histogram contains combined data from the first 100 frames for 1461 particles from three independent experiments. **(B)** Example traces of the 3D3A substrate in the presence of 1 mM ATP. Channels from top to bottom include Donor emission due to donor excitation (Dem-Dex); Acceptor emission due to donor excitation (Aem-Dex); Acceptor emission due to Acceptor excitation (Aem-Aex); Stoichiometry (S); and FRET efficiency (FRET E). **(C)** The average number of processive unwinding traces in the first 20 s after protein entry during smFRET. The initial rate from smFRET was approximated by considering the number of processive traces observed in the first 20 s. The complete movie length was 424.8 s. The numbers indicate the average from movies recorded from three independent experiments. The error bars represent the SEM. **(D)** Average FRET E distribution at the start and end of the unwinding smFRET experiments. The histogram plots show 100 bars whose width corresponds to

0.01 of FRET E (x-axis). The histograms include the FRET E values of the first 50 frames of a respective particle. N represents the average number of identified particles on the surface at the beginning of protein flow (left panels), and after 7 min had elapsed (right panels). The y-axis represents the cumulative number of frames with the respective FRET E value indicated on the x-axis. All histograms were generated using data from 3 different experiments. The values show the mean and SE of the fitted Gaussian distribution curve. The histograms in the right panels show an area of the coverslip which was not imaged before; therefore, it was not subject to bleaching while recording the particles for the left panels. Both histograms were generated from data obtained under flow. The histograms from top to bottom represent the conditions with coreTFIIH, coreTFIIH+XPA, coreTFIIH+XPG, and coreTFIIH+XPA+XPG, respectively.

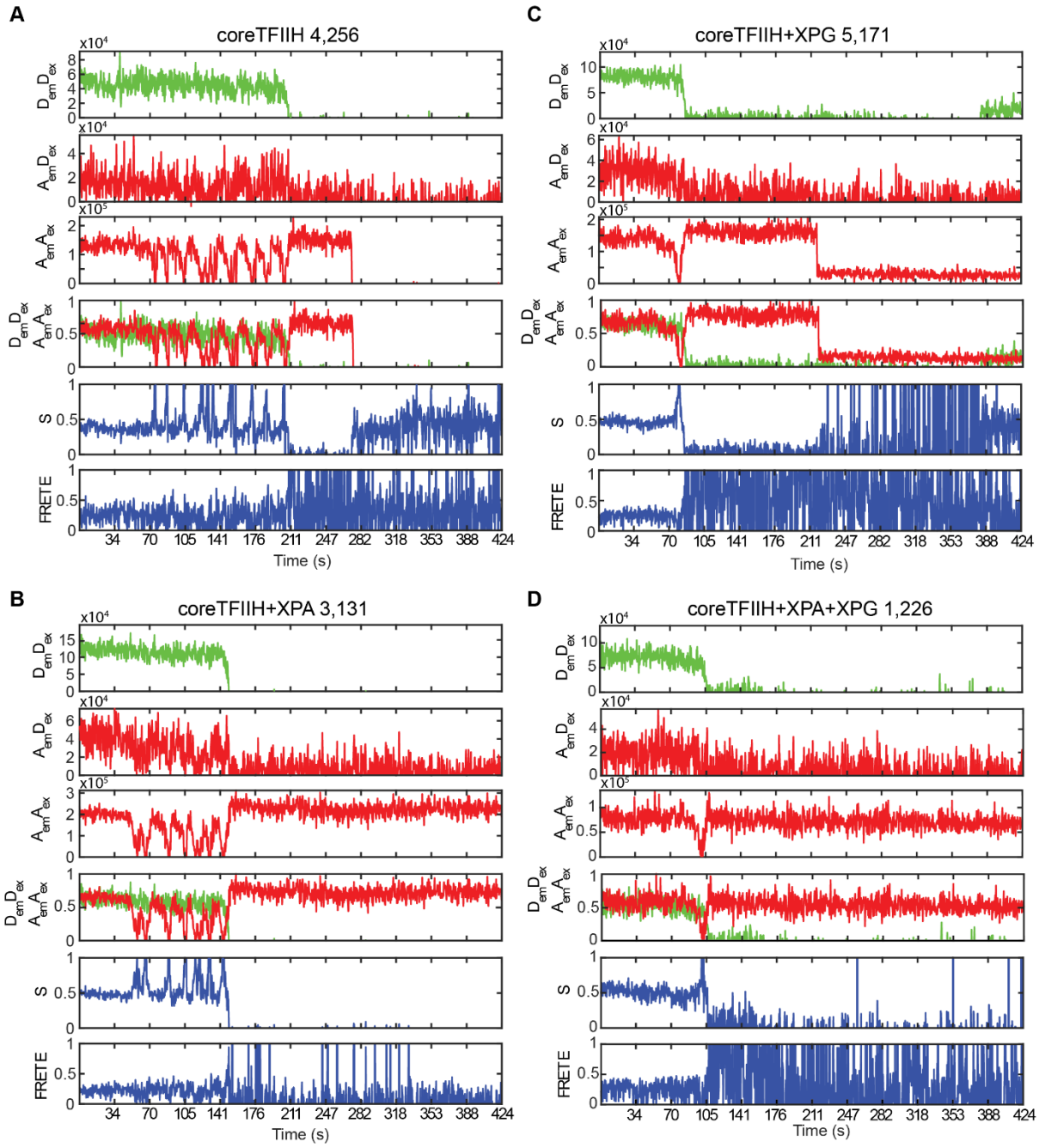

Figure S3: Full-length unwinding traces for different conditions in the presence of coreTFIIH

For all panels, channels from top to bottom include Donor emission due to donor excitation (Dem-Dex); Acceptor emission due to donor excitation (Aem-Dex); Acceptor emission due to Acceptor excitation (Aem-Aex); Stoichiometry (S); and FRET efficiency (FRET E). The x-axis shows the time from 0 s to 424.8 s. **(A)–(D)** Traces for coreTFIIH, coreTFIIH+XPA, coreTFIIH+XPG, and coreTFIIH+XPA+XPG, respectively.

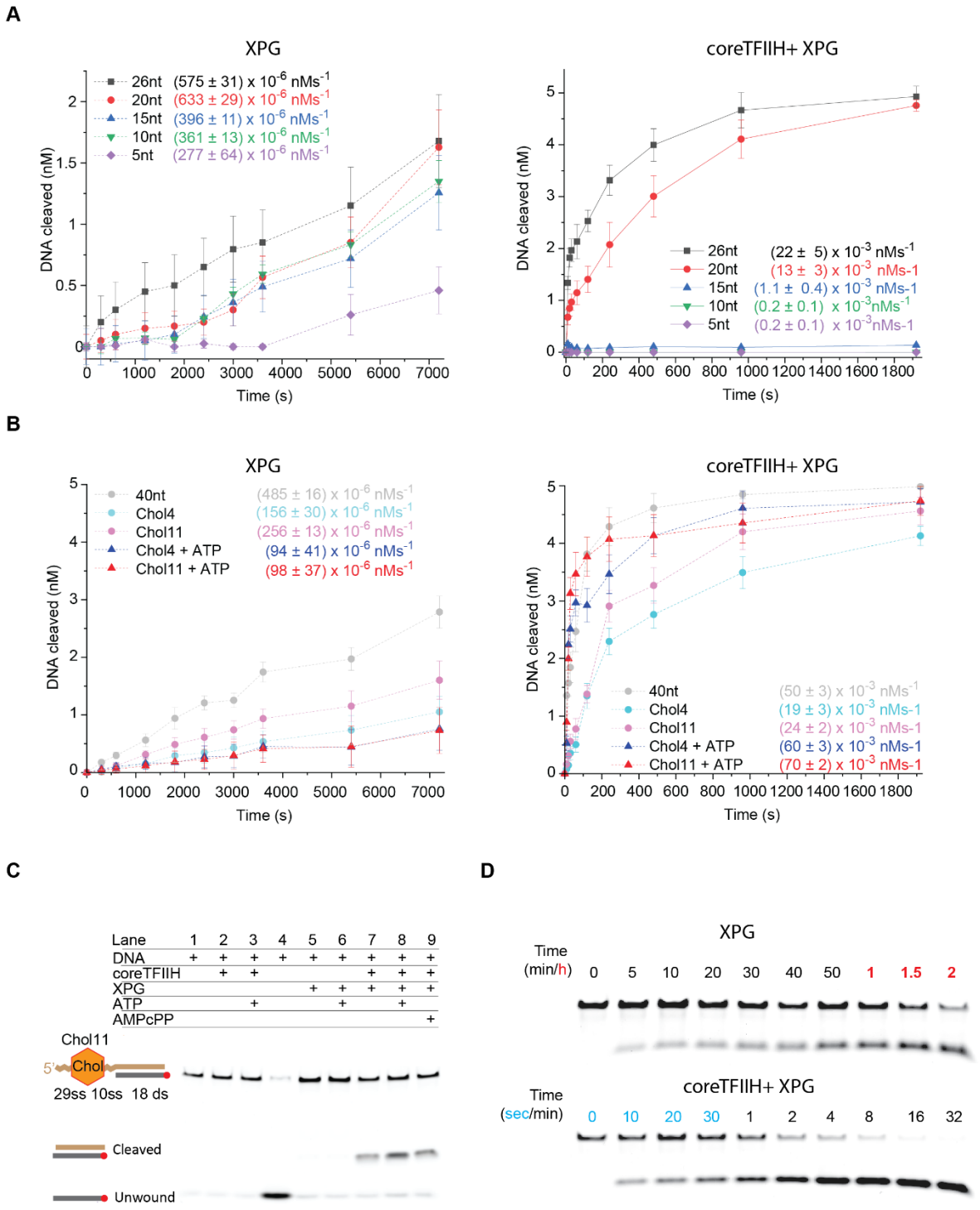

Figure S4: Study of the relative cleavage rates of 5' overhang substrates under XPG and coreTFIIH+XPG conditions

(A) Cleavage time plots of 5' overhang substrates with ssDNA lengths of 26 nt (black squares), 20 nt (red circles), 15 nt (blue triangles), 10 nt (green triangles), and 5 nt (purple

squares). Each data point represents the average amount of cleaved DNA from three independent experiments. Error bars represent the SEM. XPG cleavage is slow; hence, the incubation time was up to 7200 s (2 h). coreTFIIH-XPG cleaves considerably faster; hence, only 1920 s (32 min) were required to obtain 100% of the cleaved substrate. The y-axis shows the amount of cleaved substrate DNA in nM, where 100% corresponds to 5 nM. The relative cleavage rates obtained via linear fits of the linear regions are color-coded. **(B)** Cleavage time plots of the 40/18, Chol4, and Chol11 substrates to assess the effect of lesions on the cleavage rate in the absence or presence of ATP. **(C)** Example of cleavage and unwinding conditions for Chol11 resolved on a PAGE gel. The reactions were incubated for 2 min at 37 °C. Lane 4 contains DNA heated for 10 min to 95° C. **(D)** Example gels showing cleavage of the 40/18 substrate with XPG alone, and coreTFIIH+XPG. Note the different timescales. Seconds are shown in light blue, minutes in black, and hours in red.

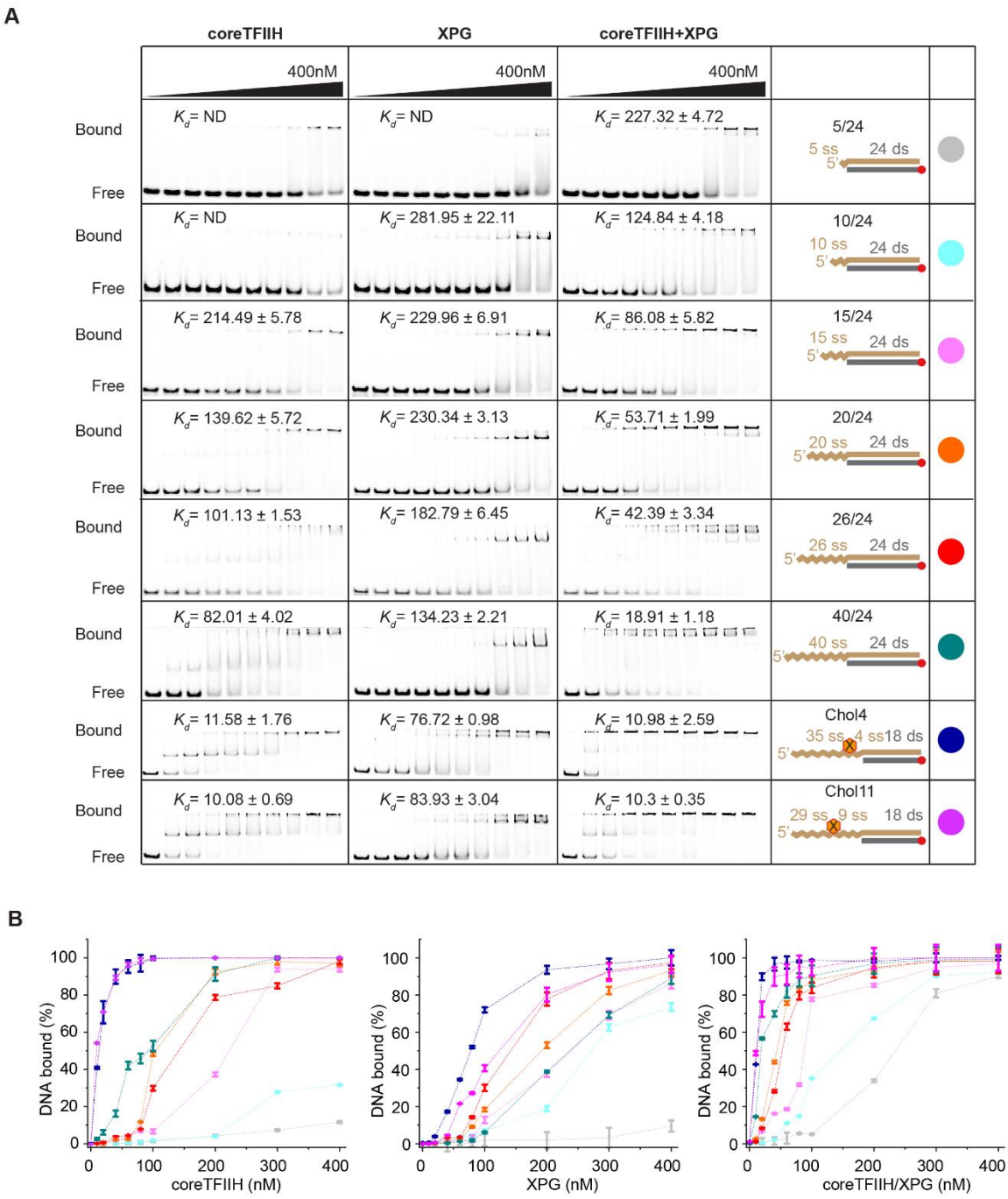

Figure S5: Determination of coreTFIIH and XPG dissociation constants on 5' overhang DNA

(A) Select electrophoretic mobility shift assay gels with coreTFIIH (left), XPG (middle), and coreTFIIH+XPG (right) panels. The substrates used were the 5, 10, 15, 20, 26, and 40 nt overhang substrates with a 24 base pair dsDNA region. Two 40ss/18ds substrates with a cholesterol lesion at 4 nt (Chol4) or 11 nt (Chol11) away from the ss/ds junction in the ss overhang. The free and protein-bound species were resolved using 6% TBE gels.

The percentage of the bound species in each lane was plotted relative to the concentration, and the resulting curve was fit to a sigmoidal curve to obtain the K_d values. The K_d values are indicated above the gel. The uncertainty of each value represents the standard error of the fit, and each value was determined from three independent experiments. The concentration of protein in each gel image from left to right was 0, 5, 10, 20, 40, 60, 80, 100, 200, 300, and 400 nM. The DNA concentration was 5 nM. **(B)** Binding curves for the aforementioned conditions and substrates. The data points are color-coded to indicate the respective substrates listed in panel A. The plots from left to right contain binding data for coreTFIIH, XPG, and coreTFIIH+XPG, respectively. For the coreTFIIH+XPG condition, both proteins are at the indicated concentration, i.e., 400 nM. coreTFIIH+XPG contains 400 nM coreTFIIH and 400 nM XPG.

Table S1: Oligonucleotide sequences and substrates used in the study.

Oligo nr.	Oligonucleotide sequence	Substrate	Oligos used
1	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA/OnyxQD/	OQ 40ss/24ds	1+7
2	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA/OnyxQD/	OQ 26ss/24ds	2+7
3	5'-TTTTTTTTTTTTTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA/OnyxQD/	OQ 20ss/24ds	3+7
4	5'-TTTTTTTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA/OnyxQD/	OQ 15ss/24ds	4+7
5	5'-TTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA/OnyxQD/	OQ 10ss/24ds	5+7
6	5'-TTTTTGCTGTTGCCCGTCTCACTGGTGAA/OnyxQD/	OQ 5ss/24ds	6+7
7	5'/Cy5/-TTCACCAGTGAGACGGGCAACAGC	OQ 40-5ss/24ds	1-6 + 7
8	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCTCGCTGCCGTCGCCA/OnyxQD/	OQ 40ss/18ds	8+9
9	5'-/Cy5/TGGCGACGGCAGCGAGGC		
10	5-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCTCGCTGCCGTCGCCA/Cy3/	40ss/18ds 3'D3'A	10+11
11	5'-TGGCGACGGCAGCGAGGC/Cy5/		
12	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA	3'Cy5 40ss/24ds	12+18
13	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA	3'Cy5 26ss/24ds	13+18
14	5'-TTTTTTTTTTTTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA	3'Cy5 20ss/24ds	14+18
15	5'-TTTTTTTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA	3'Cy5 15ss/24ds	15+18
16	5'-TTTTTTTTTTGCTGTTGCCCGTCTCACTGGTGAA	3'Cy5 10ss/24ds	16+18
17	5'-TTTTTGCTGTTGCCCGTCTCACTGGTGAA	3'Cy5 5ss/24ds	17+18
18	5'-/Cy5/-TTCACCAGTGAGACGGGCAACAGC	3'Cy5 40–5ss/24ds	12–17 + 18
19	5'TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT/Chol/TTTGCCTCGCTGCCGTCGCCA	Chol4	19+21
20	5'TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT/Chol/TTTTTTTTTTGCCTCGCTGCCGTCGCCA	Chol11	20+21
21	5'-/Cy5/-TGGCGACGGCAGCGAGGC	Chol4/Chol11	19/20 + 21