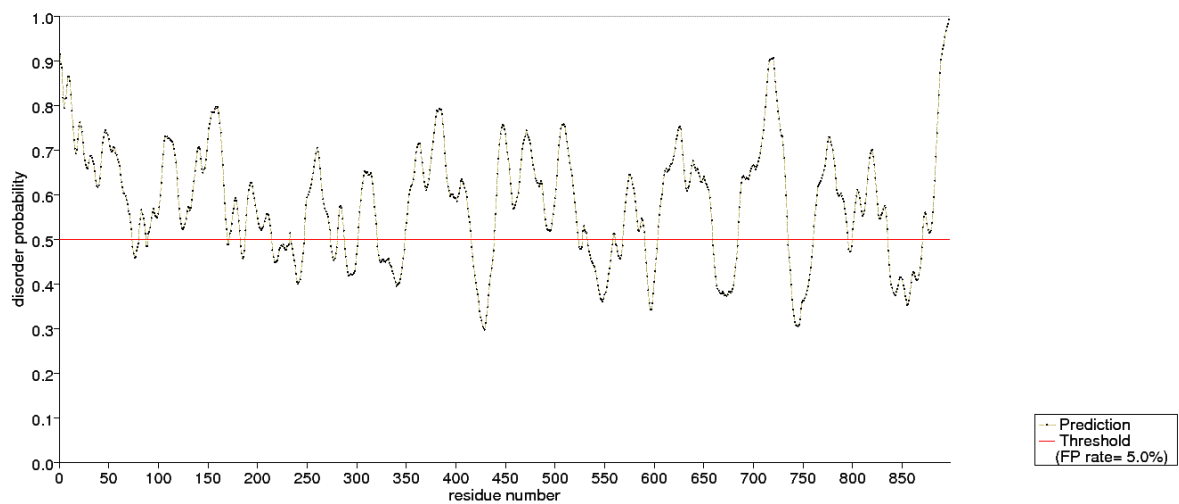


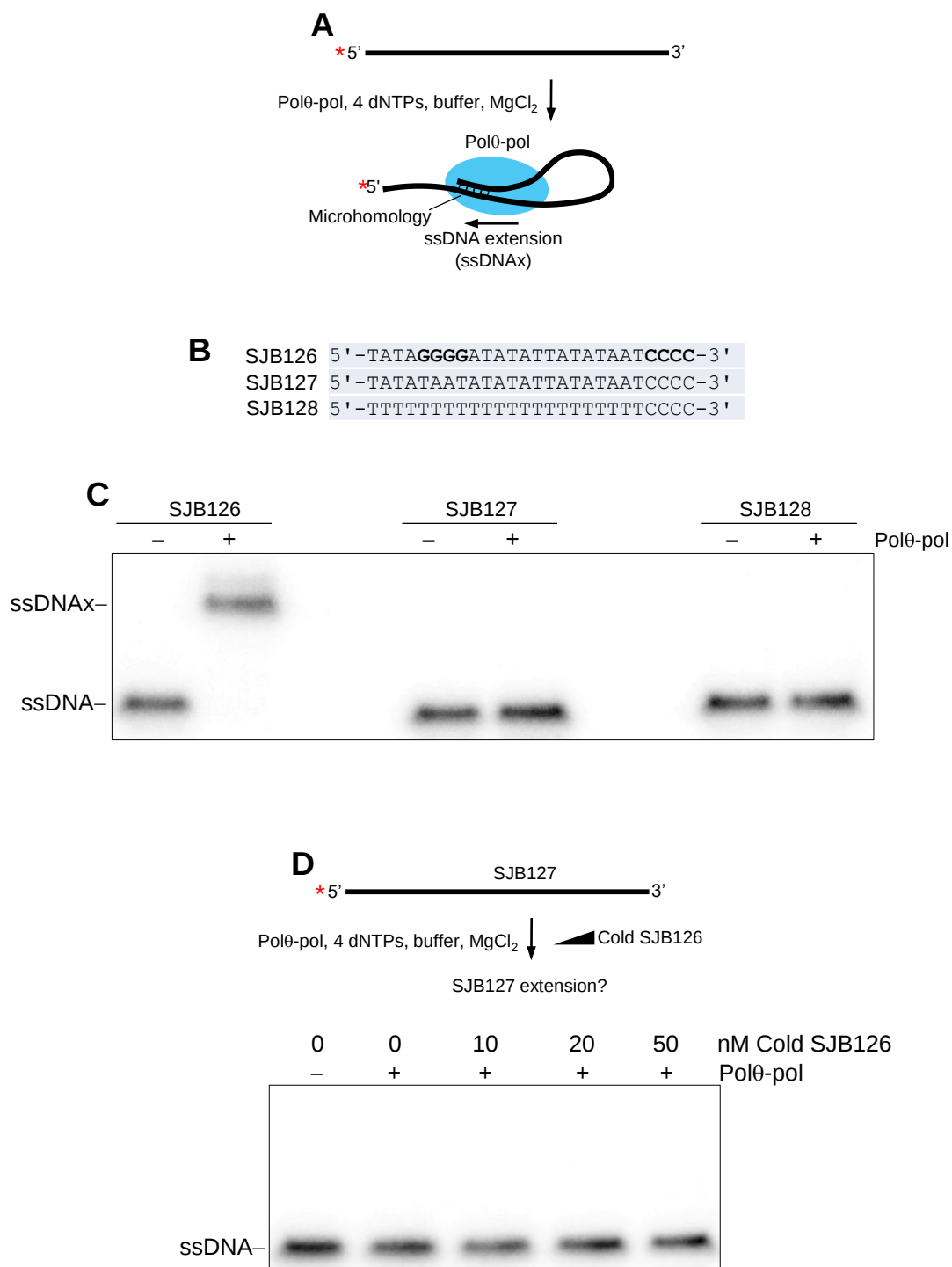
Molecular Basis of Microhomology-Mediated End-Joining by Purified Full-Length Polθ

Black et al.



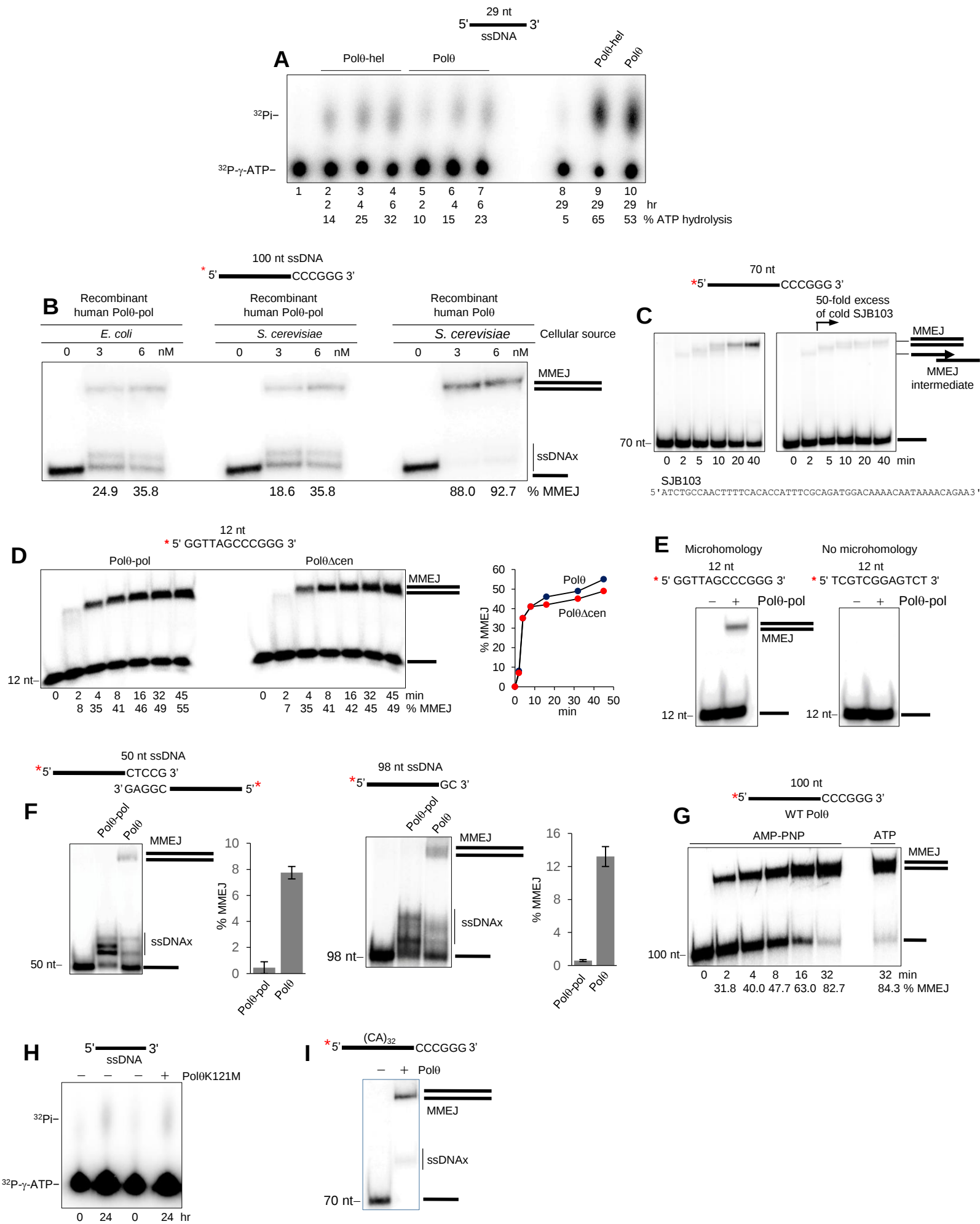
Supplementary Figure 1. Polθ-cen is highly disordered.

Analysis of human Polθ central domain consisting of amino acids 895 to 1791 using Protein Disorder Prediction System (PrDOS). Plot shows disordered probability for each residue. Residue 1 refers to residue 895 in Polθ-cen. Ishida, T and Kinoshita, K, PrDOS: prediction of disordered protein regions from amino acid sequence., *Nucleic Acids Res*, **35**, Web Server issue, 2007.



Supplementary Figure 2. Polθ-pol only performs ssDNAx when intrastrand base-pairing is available.

(A) Schematic of Polθ-pol mediated ssDNAx. (B) Sequences of ssDNA substrates used in Polθ-pol mediated ssDNAx assay. (C and D) Denaturing gels showing ssDNAx by Polθ-pol on the indicated ssDNA. Polθ-pol fails to perform ssDNAx due to lack of intrastrand pairing.



Supplementary Figure 3. Control assays for WT Polθ and Polθ variants.

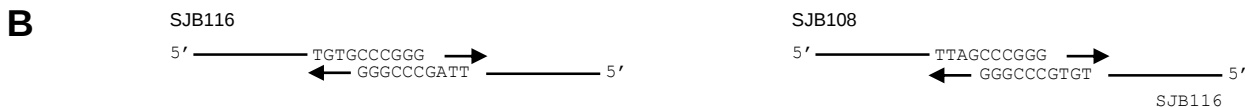
(A) Thin layer chromatography plate image showing a time course of Polθ and Polθ-hel ATPase activities in the presence of ssDNA. Lanes 8-10 demonstrate substantial ^{32}P - γ -ATP hydrolysis over 29 hr. (B) Non-denaturing gel showing MMEJ by Polθ and Polθ-pol purified from both *E. coli* and *S. cerevisiae* on the indicated 100 nt ssDNA. Activity of Polθ-pol is identical regardless of source. (C) Polθ MMEJ processivity assay. A 50-fold excess of the indicated unlabeled ssDNA was added after 2 min (right). Conversion of the intermediate MMEJ products to full-length MMEJ products following addition of the excess cold ssDNA demonstrates processive MMEJ after the initial extension step. (D) Non-denaturing gel showing a time course of MMEJ by the indicated proteins on the indicated 12 nt ssDNA with microhomology (left). Plot of MMEJ time courses (right). (E) Non-denaturing gel of Polθ-pol MMEJ reactions performed on 12 nt ssDNA with (left) and without (right) microhomology. (F) Non-denaturing gels showing MMEJ by the indicated proteins on the indicated ssDNA (left). Bar charts showing % MMEJ. $n = 3 \pm \text{s.d.}$ (right). (G) Non-denaturing gel showing MMEJ by Polθ on the indicated 100 nt ssDNA in the presence of AMP-PNP and ATP. Reactions were terminated at the indicated times. (H) PolθK121M lacks ATPase activity. Thin layer chromatography plate image showing ATP hydrolysis after the indicated time intervals in the presence of ssDNA and with and without PolθK121M. (I) Non-denaturing gel showing MMEJ by Polθ on the indicated 70 nt ssDNA which is not predicted to form secondary structures.

A SJB108
5' GTTCTTCGGTCTCGAGGTGACTACAAGGATGACGACGACAAGGGCACTGTGAGCTTAGGGTTAGCCCGGG 3'

SJB108 reverse complement
3' CCCGGGCTAACCCCTAAGCTCACAGTGCCCTTGTCTGTCGTCATCCTTGTAGTCACCTCGAGACCGAAGAAC 5'

SJB116
5' CACTGTGAGCTTAGGGTTAGGCGGCTTGACAGACACAGAGCCGAGAAATGTGCTCTAGATTCCGATGCTGACTTGCTGGGTATTATATGTGTGCCCGGG 3'

SJB116 reverse complement
3' CCCGGGCACACATATAATACCCAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGCCTAACCCCTAAGCTCACAGTG 5'



↓ Polθ MMEJ *in vitro*, purification of MMEJ products
↓ dA tailing of Polθ MMEJ products using NEB kit
↓ Cloning of MMEJ products into TOPO TA cloning vector (Invitrogen) followed by vector sequencing

C

Clone 1: SJB116 - SJB108 reverse complement sequencing

NNNNNNNNNTAGGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGT
GTGATGGATATCTGCAGAATTCGCCCTTCACTGTGAGCTTAGGGTTAGGCGGCTT
GCAGAGCACAGAGGCCGAGAAATGTGCTCTAGATTCCGATGCTGACTTGTGGGT
ATTATATGTGTGCCCGGGCTAACCCCTAAGCTCACAGTGCCCTTGTCTGTCGTCATC
CTTGTAGTCACCTCGAGACCGAAACAAAGGGCGAATTCAGCACACTGGCGGCC
GTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGGCGTAATCATGGTCATAGCTG

Clone 2: SJB108 - SJB116 reverse complement sequencing

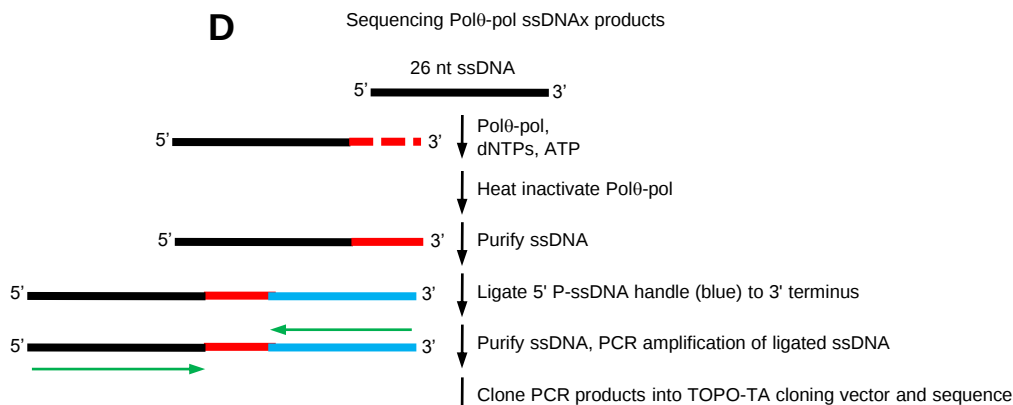
NNNNNNNNNNAGGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGT
GTGATGGATATCTGCAGAATTCGCCCTTGTCTTCGGTCTCGAGGTGACTACAAG
GATGACGACGACAAGGGCACTGTGAGCTTAGGGTTAGCCCGGGCACACATATAAT
ACCCAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGC
AAGCCGCTTAACCCCTAAGCTCACAGTGAAGGGCGAATTCAGCACACTGGCGGCC
GTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGGCGTAATCATGGTCATAGCTG

Clone 18: SJB116 - SJB108 reverse complement sequencing

NNNNNNNNNNNGGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAG
TGTGATGGATATCTGCAGAATTCGCCCTTCACTGTGAGCTTAGGGTTAGCGGGCT
TGCAGAGCACAGAGGCCGAGAAATGTGCTCTAGATTCCGATGCTGACTTGTGGG
TATTATATGTGTGCCCGGGCTAACCCCTAAGCTCACAGTGCCCTTGTCTGTCGTCAT
CCTTGTAGTCACCTCGAGACCGAAGAACAAGGGCGAATTCAGCACACTGGCGGCC
CGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGGCGTAATCATGGTCATAGCTG

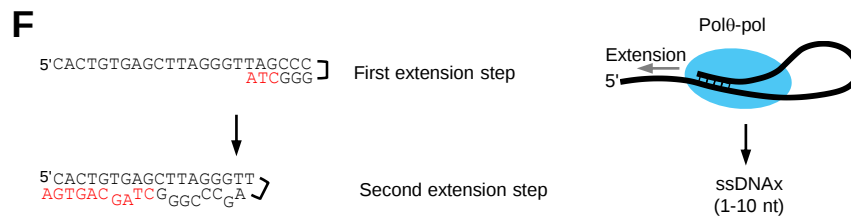
Clone 4: SJB108 - SJB116 reverse complement sequencing

NNNNNNNNNNNGGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAG
TGTGATGGATATCTGCAGAATTCGCCCTTGTCTTCGGTCTCGAGGTGACTACAA
GGATGACGACGACAAGGGCACTGTGAGCTTAGGGTTAGCCCGGGCACACATATAA
TACCCAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTG
CAAGCCGCCTAACCCCTAAGCTCACAGTGAAGGGCGAATTCAGCACACTGGCGGCC
CGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGGCGTAATCATGGTCATAGCTG



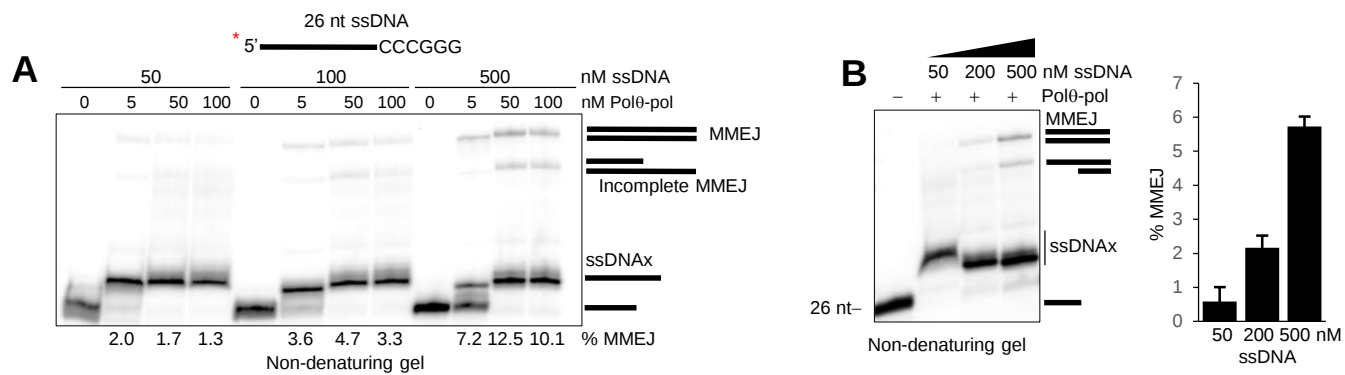
E

ssDNA sequence (5'-3')	ssDNAx
CACTGTGAGCTTAGGGTTAGCCCGGGCTAGCAGTGA	
CACTGTGAGCTTAGGGTTAGCCCGGGTCA	
CACTGTGAGCTTAGGGTTAGCCCGGGTGA	
CACTGTGAGCTTAGGGTTAGCCCGGGT	
CACTGTGAGCTTAGGGTTAGCCCGGGT	
CACTGTGAGCTTAGGGTTAGCCCGGGT	
CACTGTGAGCTTAGGGTTAGCCCGGGT	



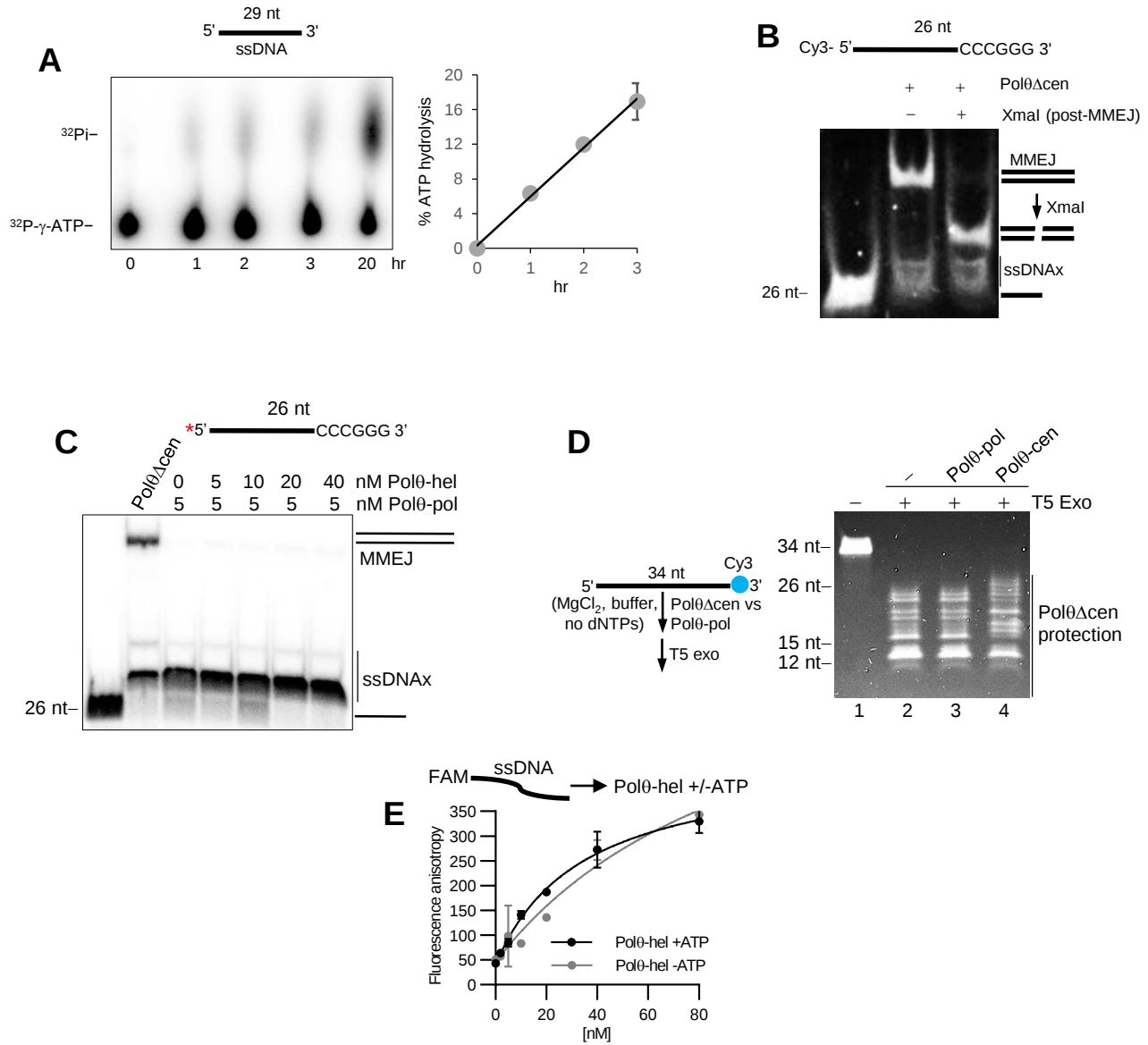
Supplementary Figure 4. Sequencing of Polθ MMEJ products and Polθ-pol ssDNA extension products.

(A) Sequences (and reverse complements) of ssDNA substrates (SJB108, SJB116) used in Polθ MMEJ assay. (B) Schematic of expected MMEJ via the 6 bp (5'-CCCGGG-3') microhomology tract (top left and right). Schematic of procedure used for sequencing Polθ MMEJ products (bottom). (C) Representative plasmid clones containing expected Polθ MMEJ products. Bold black, red and blue text corresponds to the colored sequences in panel a. Yellow highlight indicates mutation. (D) Schematic of method used for sequencing Polθ-pol ssDNAx products. (E) Representative sequences of Polθ-pol ssDNAx products. Initial ssDNA substrate sequence (black). Nucleotides transferred to the 3' terminus of ssDNA by Polθ-pol (red; ssDNAx). (F) Model of Polθ-pol ssDNAx via intrastrand pairing and extension. The model predicts how the first sequence in panel e is generated by Polθ-pol. Because Polθ-pol is highly promiscuous various intrastrand non-canonical base-pairing options may lead to the observed ssDNAx events in panel e. Figure S2 unequivocally shows that Polθ-pol ssDNAx is dependent on intrastrand pairing *in cis*.



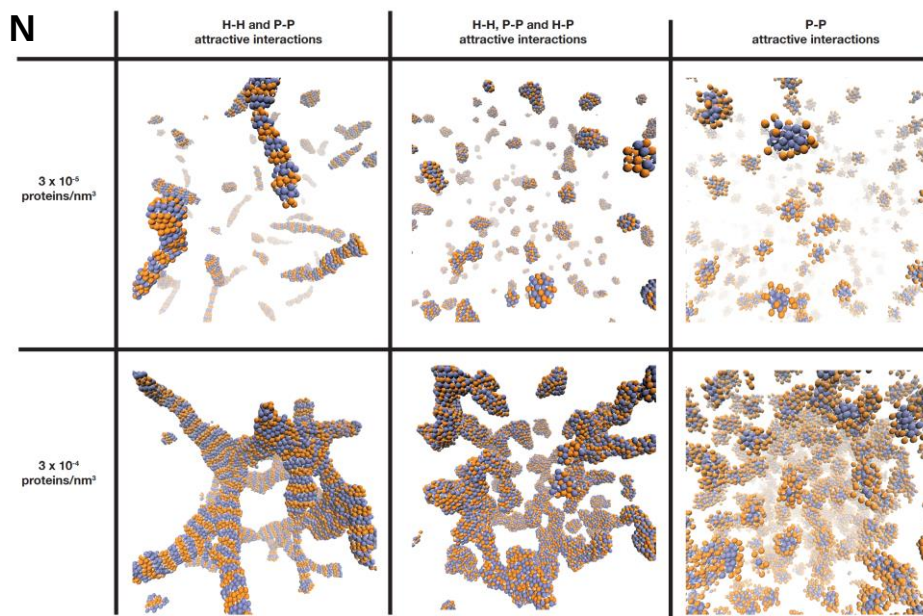
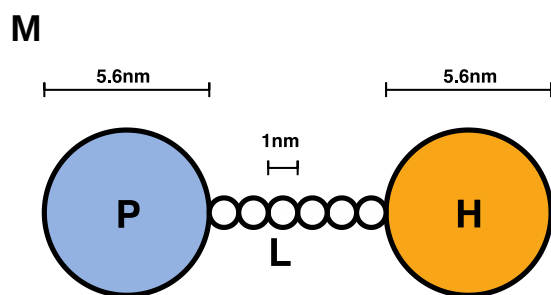
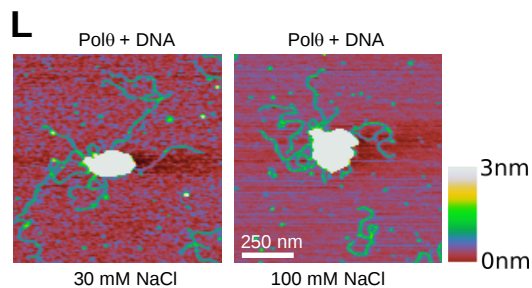
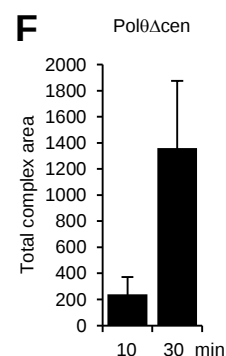
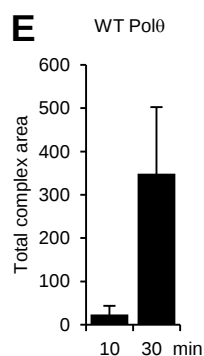
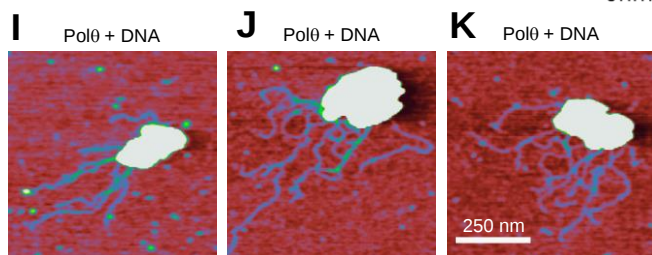
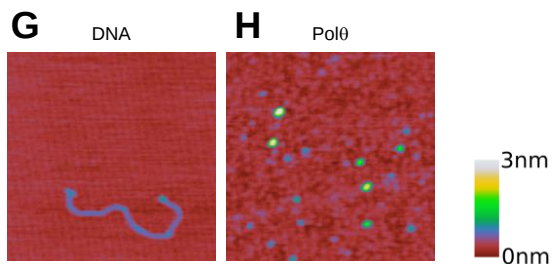
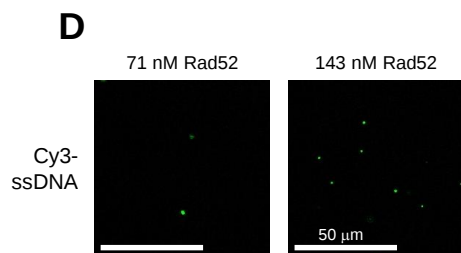
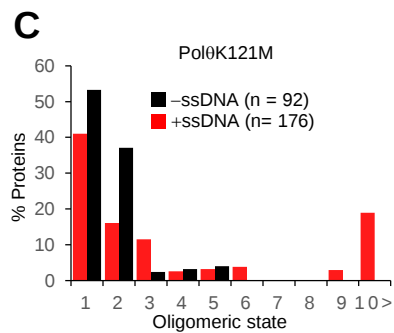
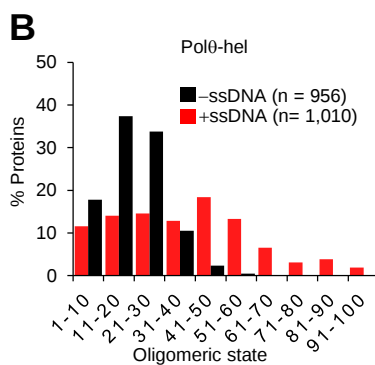
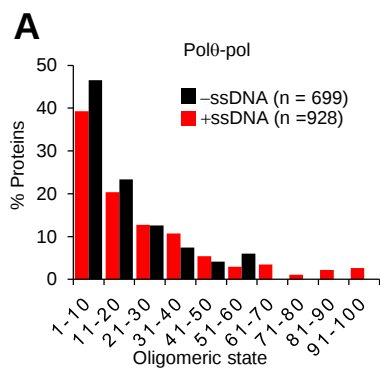
Supplementary Figure 5. Increasing concentrations of ssDNA slightly stimulates MMEJ by Polθ-pol.

(A) Non-denaturing gel showing MMEJ by Polθ-pol on ssDNA with increasing concentrations of substrate and enzyme. (B) Non-denaturing gel showing MMEJ by Polθ-pol (left). Bar chart showing % MMEJ. $n = 3 \pm \text{s.d.}$ (right).



Supplementary Figure 6. Control assays for PolθΔcen and Polθ-hel.

(A) Thin layer chromatography plate image showing a time course of PolθΔcen ATPase activity in the presence of ssDNA (left). Plot of PolθΔcen ATPase activity. $n = 3 \pm$ s.d. (right). (B) Control showing MMEJ of 26 nt ssDNA by PolθΔcen. Non-denaturing gel showing PolθΔcen MMEJ of the indicated ssDNA (lane 2), followed by digestion of the MMEJ product by XmaI (lane 3) which cleaves CCCGGG double-strand DNA. (C) Non-denaturing gel showing MMEJ reactions performed with the indicated proteins on 26 nt ssDNA with the indicated 6 bp microhomology. (D) Exonuclease footprinting of PolθΔcen on ssDNA. Schematic of assay (left). Denaturing gel showing T5 exonuclease (0.5 U/ul) digestion of the indicated ssDNA with 10 nM Polθ-pol (lane 3), 10 nM PolθΔcen (lane 4), or no protein (lane 2). (E) ATP has no significant effect on Polθ-hel ssDNA binding. Plot showing fluorescence anisotropy in the presence of the indicated amounts of Polθ-hel with and without ATP. $n = 3 \pm$ s.d.



Supplementary Figure 7. Supplementary SFM, confocal imaging, and computational modeling.

(A-C) Bar charts showing oligomeric states of indicated proteins with (red) and without (black) 26 nt ssDNA determined by SFM volume measurements. (D) Confocal microscopy images of 46 nt Cy3-ssDNA with Rad52 at 71 nM and 143 nM (as heptamer). Scale bar = 50 μ m. (E and F) Bar charts showing total WT Pol θ (E) and Pol $\theta\Delta$ cen (F) Cy3-ssDNA complex area formed after indicated times. $n = 8 \pm$ s.d. (G-K) SFM images of DNA (G), Pol θ (H), and Pol θ -DNA complexes (I-K). DNA length, 1.8 kb. Scale bar = 250 nm. (L) SFM images of Pol θ -DNA complexes in 30 mM and 100 mM NaCl. DNA length, 1.8 kb. Scale bar = 250 nm. (M) Physical model for Pol $\theta\Delta$ cen is modelled by two large spheres, representing the polymerase [P] and helicase [H] domains, connected by a short flexible polymer chain made by 6 beads, representing the (GGGS)³ linker [L]. (N) Helicase or polymerase attractive interactions. Assembly differences at two protein densities using selectively attractive interactions between H-H, P-P and H-P. H-H and P-P attractive interactions yield elongated characteristic shapes, which are less abundant adding H-P attraction. Having only P-P attraction yield no gel formation at the chosen densities.