ONLINE METHODS

Identification of a minimal catalytic core. Full-length *S. cerevisiae* Pol3 (residues 1-1097) tended to aggregate and eluted primarily in the void volume during the gel filtration step of the purification protocol. To define a more soluble and stable catalytic core, we undertook limited proteolysis of Pol3. Digestion with the non-specific protease papain and characterization of the large proteolytic fragment by N-terminal sequencing indicated cleavage at Met68. We made a 68-985 construct that retained both the polymerase and exonuclease domains of Pol3 and eliminated the C-terminal region containing two zinc fingers. $Pol3_{68-985}$ is catalytically active, similar to full length Pol3.

Protein and DNA preparation. The *S. cerevisiae* $Pol3_{68-985}$, $Pol3_{68-985}$ (exo-) and $E_{800}A$ $E_{802}A$ mutant $Pol3_{68-985}$ (exo-) proteins were expressed in the protease deficient yeast strain YRP654 as GST fusion proteins from plasmids pBJ1305, pBJ1657 and pBJ1661, respectively. The exonuclease deficient version of Pol3 was obtained by replacing the wild type fragment of the Pol3 gene with that obtained by PCR of genomic DNA from a yeast strain harboring the *pol3-01* mutation, which contains the exonuclease active site $D_{321}A$ $E_{323}A$ mutations. The E800A E802A mutation was generated by PCR using mutagenic oligonucleotides. Yeast cells were grown and GST tagged proteins were purified as described[56]. For crystallization the $Pol3_{68-985}$ protein was further purified over a gel filtration SD200 column and concentrated. The DNA duplex used for co-crystallization contained a 12-nt oligonucleotide primer harboring a dideoxycytosine at its 3' end (5'-ATCCTCCCCTAC^{dd}-3') annealed to a 16-nt template (5'-TAAGGTAGGGGAGGAT-3'), yielding a 12/16 primer-template.

Nucleotide incorporation assays and kinetic analysis. DNA synthesis assays were performed as described[56], using a 75mer oligonucleotide template (5'-AGCAAGTCAC CAATGTCTAA GAGTTCGTAT XATGCCTACA CTGGAGTACC GGAGCATCGT CGTGACTGGG AAAAC-3', where X denotes either a G or T residue) annealed to a 44mer oligonucleotide primer (5'-GTTTTCCCAG TCACGACGAT GCTCCGGTAC TCCAGTGTAG GCAT-3'). Reactions (5µl) contained 25 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 0.1 mg/ml BSA, 1 mM DTT, 10 % glycerol, 10 nM DNA substrate and varying amounts of nucleotide (0 - 1 mM dGTP, dATP, dTTP, or dCTP). Assays contained 10 nM of wild type or 20 nM of $E_{800}A$ $E_{802}A$ mutant protein and were carried out for 5-20 min at 30° C. Reaction products were separated on 10% TBE-PAGE gels containing 8M urea, and visualized by a phophorimager. Kinetic parameters were determined by plotting the rate of product formation versus dNTP concentration and fit to the Michealis-Menten equation as described[56].

Cocrystallization. The complex was prepared by incubating $Pol3_{68-985}$ with the 12/16 primer-template in a molar ratio of 1:1 with 5 mM dCTP, 10 mM $CaCl_2$, and then crystallized from solutions containing 10-14 % PEG 4000 (w/v), 100 mM sodium acetate (pH 5.5), and 200-240 mM calcium acetate. The cocrystals grow in the form of long thin plates/rods that emanate from a nucleation site. They belong to space group $P2_1$, with unit cell dimensions of $\underline{a}$ = 81.1Å, $\underline{b}$ = 85.9Å, $\underline{c}$ = 86.9 Å, and $\beta = 111.1^{O}$. For data collection, the cocrystals were cryoprotected by soaks in mother liquor solutions plus increasing amounts of ethylene glycol (0 to 25%, in 5% increments), and then flash frozen in liquid nitrogen.

Heavy atom derivatives. Cocrystals were soaked for periods ranging from minutes to days in various heavy atom solutions (0.1-10 mM). The most successful derivatives were obtained from

two separate soaks in $Au(CN)_2$, at concentrations of 5 mM (3 hrs; soak A) and 2 mM (14 hrs; soak B).

Structure determination. Two native data sets were measured at Brookhaven National Laboratory (BNL), one at λ = 1.0 Å extending to 2.0 Å resolution (beamline X12B), and another at λ = 1.50Å, extending to 2.8Å (beamline X29A). Derivative data (from the most promising soaks) were measured at BNL (beamline X29A) and at the Advanced Photon Source (APS, beamline 17-ID). Multiwavelength anomalous diffraction (MAD) data on the $Au(CN)_2$ soak B derivative (2.7 Å) were measured at BNL (beamline X29A) at three wavelengths, corresponding to the edge and peak of the Au L_{11I} edge absorption profile plus a remote point. We also measured data for $Au(CN)_2$ soak A derivative (2.7 Å) at λ = 1.0809Å. Two Au sites were found by inspection of isomorphous and anomalous difference Patterson maps. The native data collected at λ = 1.50Å were the most isomorphous to the $Au(CN)_2$ derivatives and were used for the calculation of Patterson maps and for phasing. The program SHARP[57] was used to locate and refine the sites from the two $Au(CN)_2$ soaks and to calculate the phases. The best phases were obtained with $Au(CN)_2$ soak A data, $Au(CN)_2$ soak B "peak" data (λ = 1.0394 Å), and native data at λ = 1.50Å. The subsequent electron density map, following solvent flattening with SOLOMON[58], was readily interpretable and showed, for example, well defined density for the DNA (Fig. 1B). An initial model of the ternary complex was built into the map using the program COOT[59], and subsequent iterative rounds of refinement with REFMAC[60], model building and water picking using ARP/wARP, lowered the R_{free} to 24.7%. At this stage, TLS refinement was performed with residues 96-171, 172-241, 242-577, 578-711, 712-838, 839-985 in $Pol3_{68-985}$ (identified using the TLSMD server) treated as rigid bodies, along with each DNA strand as a rigid body. TLS refinement dropped the R_{free} to 23.8% and R_{cryst} to 19.8%, after

which further rounds of building and improving the solvent model the R_{free} was reduced to 23.4% and R_{cryst} to 19.3%. The final model includes residues 95-490 and 497-985 (there is no density for residues 68-94 at the N-terminus and for a short region between Gln490 and Ser497 in the exonuclease domain), nucleotides 1-16 for the template strand, nucleotides 1-12 for the primer strand, incoming dCTP, five Ca^{2+} ions (three in the polymerase active site, one in the exonuclease active site and one bound in the major groove of the DNA away from the protein), one acetate ion and 536 water molecules. The model has excellent stereochemistry, with 97.4% of the residues in the most favored regions of the Ramachandran plot and 0% in the disallowed regions.