Published in final edited form as:

Anal Biochem. 2012 October 15; 429(2): 132–139. doi:10.1016/j.ab.2012.07.016.

A strategy for the expression of recombinant proteins traditionally hard to purify

Ekaterina G. Frank¹, John P. McDonald¹, Kiyonobu Karata², Donald Huston, and Roger Woodgate*

Laboratory of Genomic Integrity, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

We have developed a series of plasmid vectors for the soluble expression and subsequent purification of recombinant proteins that have historically proven to be extremely difficult to purify from *Escherichia coli*. Instead of dramatically overproducing the target protein, it is expressed at a low basal level that facilitates the correct folding of the recombinant protein and increases its solubility. Highly active recombinant proteins that are traditionally difficult to purify are readily purified using standard affinity tags and conventional chromatography. To demonstrate the utility of these vectors, we have expressed and purified full-length human DNA polymerases η , ι , and ν from *E. coli* and show that the purified DNA polymerases are catalytically active in vitro.

Keywords

Recombinant protein expression; Protein purification; DNA polymerase η ; DNA polymerase ι ; DNA polymerase ν

Because of its ease of growth and generally low cost to cultivate, *Escherichia coli* is often employed as the host for vectors expressing recombinant proteins. In an ideal situation, the recombinant protein is expressed from a strong promoter, highly soluble, and recovered in high yield and activity. Unfortunately, it is quite common that the overproduced recombinant protein is either detrimental to the cell or simply compartmentalized into insoluble inclusion bodies [1]. In some cases, the recombinant protein can be recovered in an active form after denaturation and subsequent renaturation [2]. However, this is less than desirable because it is often uncertain whether the refolded protein has regained full function. In an attempt to increase the solubility of recombinant proteins, they have often been coexpressed in the presence of chaperones [1] or at low temperature [3]. Here, we report an alternate method for increasing the solubility of recombinant proteins that, when normally expressed in *E. coli*, are traditionally hard to purify. Instead of expressing the recombinant protein at high levels, the protein is expressed at low basal levels without any induction from a low-copy-number vector. By adding an affinity tag to the recombinant protein, the low level of soluble protein is readily isolated from crude cell lysates and can be purified to homogeneity using conventional chromatographic methods.

*Corresponding author. Fax: +1 301 217 5815. woodgate@nih.gov (R. Woodgate).

¹The first two authors should be regarded as joint first authors.

²Current address: Molecular Biology of Infection Phenomena, Graduate School of Pharmacology Science, Chiba University, Chuo-Ku, Chiba 260-8675, Japan.

Materials and methods

Bacterial strains

The two expression strains used in this study are RW644 ($\Delta polB::spec \Delta dinB::zeo, \Delta umuDC::erm$) and RW922 ($\Delta polA::kan, \Delta polB::spec \Delta dinB::zeo, \Delta umuDC::erm/pCJ102::cat$). Both are derivatives of BL21(DE3) (EMD Biosciences, San Diego, CA, USA). These strains were specifically chosen because they lack three (RW644) or four (RW922) of *E. coli*'s endogenous polymerases, thereby simplifying downstream purification of the recombinant human polymerase. However, because the expression vectors described here do not need any special induction procedure, any desired host could, in principle, be used for expression of recombinant proteins. In general, we found that using cells that had been freshly transformed with the expression plasmid gave the highest yield of soluble protein compared with the yield from cells that had been transformed and maintained for some time.

Antibodies

The following antibodies were used to visualize the recombinant proteins: anti-His (Genscript, cat. no. A00186), anti-GST (GE Healthcare, cat. no. 27-4577-01), anti-FLAG (Abnova, cat. no. MAB2094), and anti-pol η (Abnova, cat. no. H00353497-A01).

Low-copy-number kanamycin-resistant GST, His, and FLAG expression vectors

The unique *Nde*I site located in the vector backbone of pHUC25 [4] was removed by digestion with *Nde*I, and the ends were filled with DNA polymerase (pol)³ I and subsequently religated. Sequences encoding GST, His, and FLAG affinity tags, as well as a short restriction enzyme polylinker, were synthesized by Genscript (Piscataway, NJ, USA) and cloned into the modified pHUC25 vector as *Eco*RI–*Xho*I fragments to generate pJM871 (His-tag), pJM872 (GST-tag), and pJM873 (FLAG-tag), respectively (Fig. 1). Note that if the recombinant gene is cloned into the *Xho*I site without a stop codon, a C-terminal His-tag is added to the protein, such that the N-terminal GST- and FLAG-tagged recombinant proteins are also His-tagged at the C terminus (Fig. 1) and can be purified by tandem affinity chromatography.

Low-copy-number ampicillin-resistant GST, His, and FLAG expression vectors

The β -lactamase (*bla*) gene encoding ampicillin resistance was subcloned from pET22b+ (EMD Biosciences) as an approximately 1-kb *Bsp*HI fragment into the *Bsp*HI backbone fragment of the low-copy kanamycin-resistant vectors, and the orientation of the *bla* gene was confirmed by restriction enzyme analysis. This generated His-tagged expression plasmids pJM974/pJM975 where the *bla* gene in pJM974 is in the same orientation as the *neo* (kanamycin resistance) gene of pJM871 and in the opposite orientation in pJM975. Similarly, GST expression plasmids pJM976 and pJM977 have the *bla* gene in the same or opposite orientation as pJM872. Likewise, the FLAG expression plasmids pJM978 and pJM979 have the *bla* gene in the same or opposite orientation as pJM873.

Expression, purification, and in vitro characterization of human pol η

The human *POLH* gene encoding DNA polymerase η was codon optimized for expression in *E. coli* and chemically synthesized by Genscript. The synthesized gene was subsequently cloned into the *Nde*I–*Xho*I sites of pJM871 to generate pJM879, which expresses N-terminal His-tagged pol η . pJM879 was used to transform the *E. coli* B strain, RW644 [4]. A

³Abbreviations used: pol, DNA polymerase; BME, -mercaptoethanol; EDTA, ethylenediaminetetraacetic acid; BSA, bovine serum albumin.

well-isolated kanamycin-resistant colony was picked after overnight growth at 37 °C on LB plates containing 30 µg/ml kanamycin and was used to inoculate a 12-ml starter culture. After overnight growth, the starter culture was used to seed 1 L of LB medium (plus 30 µg/ml kanamycin). Cells were harvested by centrifugation after an additional 5 to 6 h of growth at 37 °C. Cell pellets were frozen at –80 °C until required. The frozen cell pellet was thawed on ice, resuspended in approximately 40 ml of lysis buffer (50 mM Tris–HCl [pH 7.5], 0.3 M NaCl, 20 mM imidazole, 10% glycerol, 10 mM β-mercaptoethanol [BME], ethylenediaminetetraacetic acid [EDTA]-free protease inhibitors [Roche Diagnostics, Indianapolis, IN, USA], and 2 mM 4-(2-aminoethyl)benzenesulfonyl fluoride [AEBSF, MP Biochemicals, Solon, OH, USA]) and lysed by sonication. The soluble cell extract was separated from cell debris by centrifugation at 35,000g for 30 min and applied to a 1.75-ml Ni-NTA column (Qiagen, Valencia, CA, USA). The column was washed with 3 column volumes of buffer W1 (20 mM Tris–HCl [pH 7.5], 1 M NaCl, 20 mM imidazole, 10% glycerol, and 10 mM BME), followed by 3 column volumes of buffer W2 (10 mM Na–phosphate [pH 7.7], 0.3 M NaCl, 20 mM imidazole, 10% glycerol, and 10 mM BME), and finally eluted in buffer H (10 mM Na–phosphate [pH 7.7], 0.3 M NaCl, 200 mM imidazole, 10% glycerol, and 10 mM BME). The 4 or 5 peak fractions (~0.8 ml each) containing polη were pooled and directly applied to a 3-ml hydroxyapatite column (Bio-Rad, Hercules, CA, USA). The column was washed with 3 column volumes of buffer H, and polη was eluted in a linear gradient of Na–phosphate (10–200 mM) in buffer H. Peak fractions were pooled and dialyzed overnight against buffer M (20 mM Na–phosphate [pH 7.3], 100 mM NaCl, 10% glycerol, and 10 mM BME) and applied to a MonoS ion exchange column (GE Healthcare, Piscataway, NJ, USA). Polη was eluted in a linear gradient of NaCl (100–500 mM) in buffer M. Fractions containing polη were pooled, glycerol was added to a final concentration of 20%, and fractions were aliquoted prior to storage at –80 °C.

Polη is best characterized for its efficiency to bypass a thymine–thymine cyclobutane pyrimidine dimer (CPD) [5]. In contrast, *E. coli* pol I is unable to bypass a CPD. Therefore, the activity of the purified polη enzyme was compared with that of commercially available *E. coli* pol I (New England Biolabs, Ipswich, MA, USA) on an undamaged template (UTTA48: 5′-TCG ATA CTG GTA CTA ATG ATT AAC GAA **TTA AGC ACG TCC GTA CCA** TCG-3′) or on a template (TTA48) containing a single T–T CPD indicated in bold font. A 5′-³²P-labeled primer (SSP1: 5′-TGG TAC GGA CGT GCT T-3′) was annealed (position underlined) to the UTTA48 or TTA48 template at a 1.5:1 molar ratio by heating the required mixture in an annealing buffer (50 mM Tris–HCl [pH 8.0], 10 mM NaCl, 50 µg/ml bovine serum albumin [BSA], and 1 mM dithiothreitol [DTT]) for 5 min at 100 °C, followed by slow cooling to room temperature. The reaction mixture (10 µl) contained 10 nmol of labeled primer/template, 40 mM Tris–HCl (pH 8.0), 50 mM NaCl, 2.5% glycerol, 10 mM BME, 0.25 mg/ml BSA, 1.5 mM MgCl₂, 100 µM dNTPs, and 3 nmol of polη or 1 U of *E. coli* pol I. Reactions were incubated at 37 °C for 10 min and terminated by adding 10 µl of 2× loading buffer (97% formamide, 10 mM EDTA, 0.1% xylene cyanol, and 0.1% bromophenol blue). The products were heat denatured and immediately cooled on ice. Reaction products were resolved by denaturing polyacrylamide gel electrophoresis (PAGE, 8 M urea, 18% acrylamide, and 1× TBE buffer), followed by visualization and quantification using a Fuji image analyzer (FLA-5100) and ImageGauge software.

Expression, purification, and in vitro characterization of human polη

The human *POLH* gene encoding DNA polymerase η was codon optimized for expression in *E. coli* and chemically synthesized by Genscript. The synthesized gene was subsequently cloned into the *Nde*I–*Xho*I sites of pJM871 to generate pJM868, which expresses N-

terminal His-tagged pol η . Human pol η was expressed and purified using the same protocol described above for pol η .

Pol η is best characterized for its low processivity and ability to readily misincorporate dG opposite template T rather than the normal Watson–Crick base, dA [6]. In contrast, *E. coli* pol I exhibits much higher fidelity and is considerably more processive. Therefore, the activity of the purified pol η enzyme was compared with that of commercially available *E. coli* pol I on the undamaged template, UTTA48. Assay conditions were identical to those described above for pol η except that NaCl was omitted from the reaction mixture and 1.5 mM MgCl₂ was replaced with 500 μ M MnCl₂.

Expression, partial purification, and in vitro characterization of human pol ν

The human *POLN* gene encoding DNA polymerase ν was codon optimized for expression in *E. coli* and chemically synthesized by Genscript. The synthesized gene was subsequently cloned into the *NdeI*–*XhoI* sites of pJM975, pJM977, and pJM979 to generate pJM981, pJM983, and pJM985, respectively.

These expression plasmids were used to transform the *E. coli* B strain, RW922. Cultures were grown and soluble cell extracts were prepared as described above for pol η and pol η . His-tagged pol ν was applied to an Ni-NTA column and eluted in buffer H (10 mM Na-phosphate [pH 7.7], 0.3 M NaCl, 200 mM imidazole, 10% glycerol, and 10 mM BME). GST-tagged pol ν was applied to a GST–SpinTrap column (GE Healthcare) and eluted in 50 mM Tris–HCl (pH 8.0), 20 mM glutathione, 10% glycerol, and 10 mM BME. FLAG-pol ν was applied to an anti-FLAG M2 affinity resin (Sigma, St. Louis, MO, USA) and eluted with 0.2 mg/ml FLAG peptide (Sigma) in 10 mM Tris–HCl (pH 7.5), 150 mM NaCl, and 10 mM BME. The activity of partially purified FLAG-pol ν was assayed in the in vitro assays described above using the radiolabeled SSP1 primer annealed (location underlined) to a 29-mer template (polNtemp: 5′-AGC GAA GAG GAA GCA CGT CCG TAC CAT CG-3′). Reaction conditions were identical to those described for pol η except that 1.5 mM MgCl₂ was replaced with 5 mM MgSO₄.

Results

Generation of expression vectors that increase the solubility of recombinant proteins in *E. coli*

We recently developed novel expression vectors for the purification of *E. coli* DNA polymerase V, a heterotrimeric complex consisting of UmuD′₂ and UmuC [4]. When expressed individually, the small acidic UmuD′₂ subunit is readily overproduced and purified. In contrast, the basic UmuC protein is found largely in insoluble inclusion bodies, either when expressed on its own or when coexpressed with UmuD′₂ [7]. This fact notwithstanding, pol V can be purified from *E. coli*, but yields are extremely low, with just a few micrograms obtained from hundreds of liters of cell culture [8,9]. In an attempt to increase the yield of pol V obtained from conventional expression systems, we serendipitously identified vectors that, instead of overproducing UmuC, actually expressed the protein at a low basal level, but the protein was nevertheless highly soluble and allowed us to purify as much as 1 mg of pol V from just 4 L of cell culture [4].

Given our success at increasing the yield of soluble UmuC, which over the years has proven to be difficult to purify, we were interested in determining whether the strategy might be applied to other proteins with a history of poor solubility. The vector backbone was derived from the kanamycin-resistant UmuC expression vector, pHUC25, but the genes encoding His-tagged *umuC* and *umuD′* were replaced with a synthetic DNA encoding N-terminal GST, His, or FLAG affinity tags, followed by a short polylinker (Fig. 1). Recombinant

genes can be cloned in-frame with the N-terminal affinity tag by using the *NdeI* restriction enzyme site in the vector polylinker. We have also developed similar vectors encoding resistance to ampicillin so as to broaden the *E. coli* host range available for protein expression (Fig. 2).

Expression and purification of human DNA polymerases η , ι and ν

To test whether our vectors might be suitable for the expression and purification of recombinant human proteins in *E. coli*, we attempted to purify full-length human DNA polymerases η , ι and ν . These enzymes were chosen as targets for two reasons: first, as far as we are aware, the full-length DNA polymerases have yet to be successfully purified from *E. coli* (see Tables 1–3), and second, each polymerase exhibits biochemical properties in vitro that provides a unique signature for each respective enzyme and thereby confirms that the recombinant protein purified from *E. coli* retains catalytic activity and is not contaminated by an endogenous *E. coli* polymerase. To this end, we first cloned the human genes encoding full-length human DNA polymerase ι (*POLI*) and DNA polymerase η (*POLH*) into the N-terminal His-tagged expression vector. In both cases, we were able to recover significant amounts of soluble protein from whole cell extracts (data not shown). In an attempt to increase yield further, both *POLI* and *POLH* genes were codon optimized, and the recombinant genes were synthesized in vitro (Genscript) and subsequently cloned into the His-tagged expression vector. In both cases, the codon-optimized genes resulted in a higher yield of soluble protein and, as a consequence, were used for further studies. Both human DNA polymerases were purified to near homogeneity using standard chromatographic methods (Fig. 3A and B). The approximate yield of highly purified His-tagged pol ι or His-tagged pol η ranged from approximately 0.5 to 1.5 mg from 3 to 4 L (~3 g wet cell paste) of starting *E. coli* culture.

We next turned our attention to human pol ν . Full-length pol ν is somewhat larger than pol ι and pol η (100.3 kDa vs. 80.3 and 78.4 kDa, respectively) and, to our knowledge, has not been purified (from any expression system/host), possibly due to its proline rich C-terminal tail. The human gene encoding *POLN* was codon optimized for expression in *E. coli*, synthesized, and cloned into the GST-, His-, and FLAG-tagged expression vectors. In this instance, the ampicillin-resistant vectors were chosen because the host strain used for expression of the *POLN* vectors carries a deletion allele of *E. coli polA* that is marked by kanamycin resistance. The three expression vectors all produced soluble full-length pol ν that could be partially purified using the appropriate affinity resin (Fig. 4). Western blot analysis with antibodies to the specific affinity tag, or to the C terminus of pol ν , confirmed the soluble expression of full-length pol ν from *E. coli* and its partial purification. The degree of purity varied depending on the N-terminal affinity tag. In the case of pol ν , the greatest purity obtained in a single step was with the FLAG-tagged construct, which resulted in a preparation of pol ν that was estimated to be 30% pure based on visual inspection of the Coomassie-stained gel (Fig. 4C).

Characterization of purified recombinant DNA polymerases η , ι and ν

A key requirement for any purification protocol is that the final purified product be highly active and not contaminated with interfering *E. coli* proteins. In our situation, that would include endogenous polymerases or exonucleases. Of particular concern is DNA polymerase I, which is a robust and processive enzyme with both 5′–3′ and 3′–5′ exonucleolytic activities. Fortunately, both DNA polymerases ι and η lack intrinsic exonucleolytic activity, and each possesses unique catalytic activities that easily distinguishes it from *E. coli* pol I. For example, pol ι is a distributive enzyme that is known for its propensity to misinsert dG opposite template T [6]. In contrast, pol η is best characterized by its ability to bypass a thymine–thymine dimer in DNA [10]. As seen in Fig. 3C and D, both activities are

observed, and the patterns of the respective replication products are very different from that observed in reactions containing commercially available *E. coli* pol I. As a consequence, we confirmed that the purified full-length human DNA polymerases α and η are highly active and free from any *E. coli* contaminants.

The activity of human pol ν was harder to determine because it belongs to the same “A-family” of DNA polymerases as *E. coli* pol I and its activities are less specialized, although it does misincorporate T opposite template G with high frequency [11] and, like pol ι and pol η , it lacks intrinsic exonucleolytic activity [12]. To simplify downstream purification, we expressed and purified pol ν from an *E. coli* strain, RW922, lacking pols I, II, IV, and V. The cell’s sole endogenous polymerase in RW922 is the high-fidelity replicase, pol III. Therefore, the activity of human pol ν can be distinguished from pol III by its ability to readily misincorporate T opposite template dG. As seen in Fig. 4D, the partially purified preparation of full-length FLAG-pol ν has polymerase activity with a propensity to misincorporate T opposite G, indicating that the enzyme is highly active when expressed in *E. coli*. We should note that DNA synthesis promoted by the partially purified full-length pol ν on the 48-mer oligonucleotide template appears to be somewhat less processive than previously reported for longer DNA templates [12]. At the current time, however, it is unclear whether these properties are intrinsic to full-length pol ν or are simply due to differences in the replication assay conditions. Therefore, further characterization of more highly purified preparations of full-length pol ν is clearly warranted.

Discussion

We have developed plasmid vectors that enable the recovery and purification of recombinant proteins that have previously proven to be difficult to express in soluble form. The vectors described here have a pSC101 origin of replication [13] and, therefore, are maintained in *E. coli* at approximately 5 copies per cell [14]. DNA for these low-copy-number vectors can be obtained by conventional plasmid purification methods (e.g., QIAprep spin column, Qiagen) by simply increasing the recommended volume of the *E. coli* culture used for the preparation of plasmid DNAs from 1–2 ml to 10–20 ml. Unlike many expression vectors where the goal is to produce high levels of the target protein, the recombinant proteins expressed from the vectors described here are expressed at low basal levels, most likely by transcriptional read-through from the *lacIq* promoter in the vector backbone (Fig. 2), which is approximately 1 kb upstream of the *NdeI* cloning site and the initiator codon of the recombinant protein. Cells are grown in LB medium at 37 °C, and expression of the target protein occurs without any external induction. As a consequence, cell cultures can be harvested at any time that is convenient for the researcher. In the examples described above, cells were generally harvested anywhere between 3 and 6 h of growth after dilution of the initial overnight culture and at an OD₆₀₀ of 1 to 2, with only minimal differences in the amounts of soluble protein obtained. However, it is likely that the cell density resulting in the greatest solubility of each individual target protein needs to be determined empirically.

We have used these vectors to express and purify three full-length human DNA polymerases in *E. coli* and confirmed that the respective purified polymerase is highly active and free from *E. coli* contaminants. We assume that a similar approach can be readily applied to recombinant proteins from a number of sources and with different catalytic activities. In addition, the expression vectors described here have a pSC101 origin of replication and, therefore, are fully compatible with the replicons of conventional high-expression systems (e.g., pET vectors, EMD Biosciences). This compatibility may be especially useful for the expression of multi-meric protein complexes, where many of the components can be highly

expressed in soluble form and only one subunit of the complex is problematic and needs to be expressed at low levels.

Acknowledgments

This work was supported in part by the National Institute of Child Health and Human Development/National Institutes of Health Intramural Research Program (to R.W., E.G.F., D.H., and J.P.M.) and the Special Coordination Funds for Promoting Science and Technology of the Japanese Ministry of Education, Culture, Sports, Science, and Technology (to K.K.).

References

1. Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* 2006; 24:179–185. [PubMed: 16503059]
2. Armstrong N, de Lencastre A, Gouaux E. A new protein folding screen: application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase. *Protein Sci.* 1999; 8:1475–1483. [PubMed: 10422836]
3. Vasina JA, Baneyx F. Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli* *cspA* and *tac* promoter systems. *Protein Expr Purif.* 1997; 9:211–218. [PubMed: 9056486]
4. Karata K, Vaisman A, Goodman MF, Woodgate R. Simple and efficient purification of *E. coli* DNA polymerase V: cofactor requirements for optimal activity and processivity in vitro. *DNA Repair.* 2012; 11:431–440. [PubMed: 22341652]
5. Masutani C, Araki M, Yamada A, Kusumoto R, Nogimori T, Maekawa T, Iwai S, Hanaoka F. Xeroderma pigmentosum variant (XP-V) correcting protein from HeLa cells has a thymine dimer bypass DNA polymerase activity. *EMBO J.* 1999; 18:3491–3501. [PubMed: 10369688]
6. Tissier A, McDonald JP, Frank EG, Woodgate R. Pol η , a remarkably error-prone human DNA polymerase. *Genes Dev.* 2000; 14:1642–1650. [PubMed: 10887158]
7. Woodgate R, Rajagopalan M, Lu C, Echols H. UmuC mutagenesis protein of *Escherichia coli*: purification and interaction with UmuD and UmuD'. *Proc Natl Acad Sci USA.* 1989; 86:7301–7305. [PubMed: 2552436]
8. Schlacher K, Jiang Q, Woodgate R, Goodman MF. Purification and characterization of *Escherichia coli* DNA polymerase V. *Methods Enzymol.* 2006; 408:378–390. [PubMed: 16793381]
9. Fujii S, Gasser V, Fuchs RP. The biochemical requirements of DNA polymerase V-mediated translesion synthesis revisited. *J Mol Biol.* 2004; 341:405–417. [PubMed: 15276832]
10. Masutani C, Kusumoto R, Iwai S, Hanaoka F. Mechanisms of accurate translesion synthesis by human DNA polymerase η . *EMBO J.* 2000; 19:3100–3109. [PubMed: 10856253]
11. Takata K, Shimizu T, Iwai S, Wood RD. Human DNA polymerase ν (*POLN*) is a low fidelity enzyme capable of error-free bypass of 5S-thymine glycol. *J Biol Chem.* 2006; 281:23445–23455. [PubMed: 16787914]
12. Arana ME, Takata K, Garcia-Diaz M, Wood RD, Kunkel TA. A unique error signature for human DNA polymerase. *DNA Repair.* 2007; 6:213–223. [PubMed: 17118716]
13. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci USA.* 1973; 70:3240–3244. [PubMed: 4594039]
14. Hasunuma K, Sekiguchi M. Replication of plasmid pSC101 in *Escherichia coli* K12: requirement for *dnaA* function. *Mol Gen Genet.* 1977; 154:225–230. [PubMed: 337104]
15. Johnson RE, Washington MT, Prakash S, Prakash L. Fidelity of human DNA polymerase η . *J Biol Chem.* 2000; 275:7447–7450. [PubMed: 10713043]
16. Biertümpfel C, Zhao Y, Kondo Y, Ramon-Maiques S, Gregory M, Lee JY, Masutani C, Lehmann AR, Hanaoka F, Yang W. Structure and mechanism of human DNA polymerase η . *Nature.* 2010; 465:1044–1048. [PubMed: 20577208]
17. Johnson RE, Washington MT, Haracska L, Prakash S, Prakash L. Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. *Nature.* 2000; 406:1015–1019. [PubMed: 10984059]

18. Zhang Y, Yuan F, Wu X, Wang Z. Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase ϵ . *Mol Cell Biol*. 2000; 20:7099–7108. [PubMed: 10982826]
19. Choi JY, Guengerich FP. Kinetic evidence for inefficient and error-prone bypass across bulky N2-guanine DNA adducts by human DNA polymerase ϵ . *J Biol Chem*. 2006; 281:12315–12324. [PubMed: 16527824]
20. Nair DT, Johnson RE, Prakash S, Prakash L, Aggarwal AK. Replication by human DNA polymerase ϵ occurs by Hoogsteen base-pairing. *Nature*. 2004; 430:377–380. [PubMed: 15254543]
21. Pence MG, Blans P, Zink CN, Hollis T, Fishbein JC, Perrino FW. Lesion bypass of N2-ethylguanine by human DNA polymerase ϵ . *J Biol Chem*. 2009; 284:1732–1740. [PubMed: 18984581]
22. Kirouac KN, Ling H. Structural basis of error-prone replication and stalling at a thymine base by human DNA polymerase ϵ . *EMBO J*. 2009; 28:1644–1654. [PubMed: 19440206]
23. Zietlow L, Smith LA, Bessho M, Bessho T. Evidence for the involvement of human DNA polymerase ν in the repair of DNA interstrand cross-links. *Biochemistry*. 2009; 48:11817–11824. [PubMed: 19908865]
24. Arana ME, Powell GK, Edwards LL, Kunkel TA, Petrovich RM. Refolding active human DNA polymerase ν from inclusion bodies. *Protein Expr Purif*. 2010; 70:163–171. [PubMed: 19853037]

Fig. 2.
Plasmid maps of kanamycin- and ampicillin-resistant expression vectors: (A) pJM871; (B) pJM872; (C) pJM873; (D) pJM974; (E) pJM976; (F) pJM978; (G) pJM975; (H) pJM977; (I) pJM979.

Fig. 3. Purification and characterization of human pol ι and pol η . (A) purification of pol ι (lane 1: whole cell extract; lane 2: fraction from the Ni-NTA column; lane 3: fraction from the hydroxyapatite column; lane 4: fraction from the MonoS column). The arrow on the right-hand side of the panel indicates the position of pol ι . (B) Purification of pol η (lane 1: whole cell extract; lane 2: fraction from the Ni-NTA column; lane 3: fraction from the hydroxyapatite column; lane 4: fraction from the MonoS column). The arrow on the right-hand side of the panel indicates the position of pol η . (C) Replication assays comparing purified pol ι with *E. coli* pol I in vitro. Standing start primer extension reactions were performed in the absence of dNTPs (lanes -), dGTP (lanes G), or dATP (lanes A) or in the presence of all four dNTPs (lane 4). The position of the primer is indicated on the left of the image, and the local sequence of the template downstream of the primer is indicated on the right-hand side of the image. (D) Replication assays comparing purified pol η with *E. coli* pol I in vitro. Standing start primer extension reactions on an undamaged template (U) or a template with a T–T CPD (TT) were performed in the absence of dNTPs (lanes -) or in the

presence of all four dNTPs (lanes 4). The location of the T–T CPD is indicated on the right-hand side of the image.

Fig. 4.

Purification and characterization of human polv. (A) Purification of His-polv (lane 1: whole cell extract; lane 2: soluble cell extract; lane 3: fraction from the Ni-NTA column). Proteins were visualized by staining with Coomassie blue (left-hand panel), Western blotting with His antibodies (center), or polv antibodies (right-hand panel). The arrow on the right-hand side of the image indicates the position of polv. (B) Purification of GST-polv (lane 1: whole cell extract; lane 2: soluble cell extract; lane 3: fraction from the glutathione column). (C) Purification of FLAG-polv (lane 1: whole cell extract; lane 2: soluble cell extract; lane 3: fraction from anti-FLAG column). (D) Activity of purified FLAG-polv. Standing start primer extension reactions were performed in the absence of dNTPs (lanes -) or the

presence of dTTP (T), dCTP (C), or all four dNTPs (lane 4). The position of the primer is indicated on the left-hand side of the image, and the local sequence of the template downstream of the primer is indicated on the right-hand side of the image.

Table 1Selected published purification protocols for human pol η .

Affinity tag	Expression organism	Purification strategy	References
No tag	HeLa cells	Nuclear extract Phosphocellulose P11 Hydroxyapatite Ammonium sulfate precipitation HiTrap phenyl-Sepharose HP TSK gel phenyl 5PW TSK gel HA1000 MonoQ PC MonoS PC	[5]
C-term His-1-713	Sf9 insect cells	Hi-Trap Q Ni-NTA MonoS PC	[10]
N-term GST-1-713	Yeast	Glutathione Sepharose 4B Phenyl Sepharose PC	[15]
N-term His-1-432	<i>E. coli</i>	Ni-NTA MonoS Superdex 75 His-tag removed by PreScission protease	[16]

Note: Full-length human pol η consists of 713 amino acids.

Table 2Selected published purification protocols for human pol α .

Affinity tag	Expression organism	Purification strategy	References
N-term	Sf9 insect cells	Glutathione agarose	[6]
GST-1-715		Hydroxyapatite	
		Superdex 200	
N-term	Yeast	Glutathione Sepharose 4B	[17]
GST-1-715		Thrombin GST cleavage	
		Mini Q PC	
N-term	Yeast	Ni-NTA	[18]
His-1-715		Polyethylene glycol concentration	
		Sephadex G-25	
		MonoS HR	
		Polyethylene glycol concentration	
		Superdex 200	
N-term	Sf9 insect cells	His-Trap column	[19]
His-1-715		MonoQ	
		Amicon concentration	
		Superdex 200	[20]
GST-1-420	Yeast	Ammonium sulfate precipitation	
		Glutathione Sepharose	
		PreScission protease removal of GST	
		MonoQ	
N-term	<i>E. coli</i> BL21(DE3)	Amylose affinity resin	[21]
MBP-1-420	Rosetta 2 grown at 17 LC	PreScission-MBP cleavage	
		Phosphocellulose P11	[22]
N-term	<i>E. coli</i> BL21(DE3)	Glutathione agarose	
GST-1-420		GST-tag cleaved by TEV protease	
		Ni-NTA, HiTrap SP Sepharose	

Note: Full-length human pol α consists of 715 amino acids.

Table 3

Selected published purification protocols for human polv.

Affinity Tag	Expression organism	Purification strategy	References
N-term-His	<i>E. coli</i>	FLAG affinity resin	[11]
C-term-FLAG 1-863	BL21(DE3)/ pRARE	TALON affinity resin	
N-term-His	<i>E. coli</i> BL21(DE3)	DEAE ion exchange	[23]
C-term-FLAG 1-863	Rosetta 2/pRARE	Phosphocellulose P11 TALON affinity resin	
N-term His-(175-863)	<i>E. coli</i> BL21(DE3)	Inclusion body isolation	[24]
	Rosetta 2 pLacI	Hydrostatic pressure refolding Superdex 200 HiTrap heparin	
N-term His-NusA (175-863)	<i>E. coli</i> BL21(DE3)	Ni-HiTrap	[24]
	Rosetta 2 pLacI	His-NusA tag cleaved by AcTEV protease	
	Grown at 30 °C	HiTrap heparin	

Note: Full-length human polv consists of 900 amino acids.