A major hurdle for molecular mechanistic studies of many proteins is the lack of a general method for fluorescence labeling with high efficiency, specificity and speed. By incorporating an aldehyde motif genetically into a protein and improving the labeling kinetics substantially under mild conditions, we achieved fast, site-specific labeling of a protein with ~100% efficiency while maintaining the biological function. We show that an aldehyde-tagged protein can be specifically labeled in cell extracts without protein purification and then can be used in single-molecule pull-down analysis. We also show the unique power of our method in single-molecule studies on the transient interactions and switching between two quantitatively labeled DNA polymerases on their processivity factor.

In recent years, single-molecule techniques have become standard tools for studying complex biological problems. Often a prerequisite for these studies is the fluorescence labeling of a protein at one location with high efficiency. Poor site specificity introduces undesirable heterogeneity, whereas low labeling efficiency limits the data throughput and makes certain experiments challenging. For example, when determining the solution stoichiometry for a hexameric complex, only 1.56% of the molecules will show all six photobleaching steps if the labeling efficiency is 50%. The lack of a generally-applicable protein labeling method has been a major hurdle for many mechanistic ensemble or single-molecule studies. Current methods rely on cysteine-specific chemistry, N-terminal transamination, installation of a peptide such as tetracysteine or polyhistidine and so on; however, these techniques suffer from common problems including poor site specificity, location requirement, byproduct formation, limited commercial availability, unpredictable protein yield, low labeling efficiency or limited fluorophore choice.

Given the limitations of existing techniques, we took a different approach, using the genetically encoded aldehyde tag LCTPSR, which can be fused to a protein. The cysteine in this motif is converted in vivo into formylglycine by coexpressed formylglycine-generating enzyme (FGE). The aldehyde group in this residue then serves as an exclusive target for labeling with commercially available cyanine hydrazides, even in the presence of other cysteines on the protein. However, this seemingly straightforward strategy comes with several practical challenges. First, the condition reported for aldehyde labeling is harsh for many proteins. Second, fast and quantitative labeling is yet to be demonstrated for any protein. Third, the cysteine to formylglycine conversion may not be complete. Despite the importance of these issues, none have been addressed in the literature thus far. Moreover, labeling site specificity and fluorophore linkage stability may present additional challenges.

Here we establish a simple and robust method for rapidly labeling a protein at one location with ~100% efficiency under physiologically relevant conditions. We labeled two different DNA polymerases with different fluorophores, allowing us to detect the transient and dynamic interactions between the two polymerases on their processivity factor. We also show that an aldehyde-tagged protein can be labeled efficiently and specifically in cell extracts and used for single-molecule pull-down analysis.

RESULTS
Quantitative labeling of aldehyde-tagged proteins
We constructed plasmids for expression of archaeal DNA polymerases PolBI and DinB from *Methanosarcina acetivorans* with sequence encoding an aldehyde tag positioned at the N terminus followed by a polyhistidine tag and a thrombin cleavage site, and coexpressed the protein with FGE in *Escherichia coli*. PolBI and DinB have ten and four native cysteines, respectively, making it difficult to engineer a ‘cysteine-light’ mutant rationally in the absence of structural information. We exchanged purified PolBI or DinB into the buffer reported in literature for labeling (pH 5.5 with 1% (wt/vol) SDS) and incubated at 37 °C for 2 h. After this treatment, both polymerases lost the primer-extension activity (Supplementary Note 1). For this reason, we attempted to use a much milder condition: pH 7.0, with no detergent, at 4 °C. However, even with excess Cy3 hydrazide (Cy3HZ), only ~5% of DinB and ~40% of PolBI were labeled after 4.5 d and 9.7 d of incubation, respectively (Supplementary Note 1). By measuring the labeling kinetics of aldehyde-tagged PolBI (Ald6N-PolBI), we determined that the time to reach half of the maximal labeling efficiency, $\tau_{1/2}$, was 6.3 ± 0.6 d for the condition above.

1Howard Hughes Medical Institute, Urbana, Illinois, USA. 2Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. 3Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. 4Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. 5Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. 6Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. Correspondence should be addressed to T.H. (tjha@illinois.edu).

RECEIVED 19 SEPTEMBER 2011; ACCEPTED 8 MARCH 2012; PUBLISHED ONLINE 1 APRIL 2012; DOI:10.1038/NMETH.1954
**ARTICLES**

After characterizing the effects of temperature, [Cy3HZ], pH and the reported catalyst aniline\(^1\) on the labeling kinetics and efficiency (Supplementary Note 1), we focused on increasing [Cy3HZ], while keeping the temperature at 4 °C and pH at 7.0. Owing to the high water solubility of Cy3, a concentration of 75.6 mM can be obtained for Cy3HZ, which allowed us to achieve ~100% labeling efficiency for Ald\(_6\)N-PolBI at 4 °C and pH 7.0 (Fig. 1a). Furthermore, \(\tau_{1/2} \) was shortened markedly, from 6.3 d to 0.4 d. Despite the high concentration of dye, the overall contribution of nonspecific dye association with protein to the labeling efficiency observed was minimal, as shown using a protein without the aldehyde tag (Supplementary Note 2).

**Intact protein function after quantitative labeling**

To confirm the biological function of labeled protein, we used standard ensemble biochemical assays for DNA polymerases and observed identical primer-extension and strand-displacement activities of unlabeled and Cy3HZ-labeled PolBI (Supplementary Note 3). We also examined DNA binding and replication using single-molecule fluorescence resonance energy transfer (FRET)\(^1\),\(^\text{18,19}\), Cy3HZ-labeled PolBI showed robust binding to a Cy5-labeled DNA immobilized on the surface (Fig. 1c,d). Adding dATP led to a substantial decrease in the FRET efficiency, \(E_{\text{FRET}} = I_{\text{Cy5}}/(I_{\text{Cy5}} + I_{\text{Cy3}})\), where \(I\) is the fluorescence intensity, as a result of template-dependent replication (Fig. 1c,d). Real-time single-molecule FRET trajectories showed a gradual FRET decrease during the reaction, and fitting the replication time histogram to a Gaussian distribution gave a replication rate of 34 ± 13 nt s\(^{-1}\) (Fig. 1e). This value agrees well with the value of ~23 nt s\(^{-1}\) estimated from bulk primer-extension experiments\(^\text{15} \) and is close to the value of ~17 nt s\(^{-1}\) reported for a closely related member of the PolB family from *Thermococcus litoralis*\(^2\). As a control, adding an incorrect dNTP gave no detectable change in \(E_{\text{FRET}}\) (Fig. 1d and Supplementary Fig. 1). These single-molecule data show that Ald\(_6\)N-PolBI was functional after labeling.

**Site specificity, linker stability and general applicability**

Through incubation with restriction-grade thrombin, Cy3HZ-labeled Ald\(_6\)N-PolBI was cleaved at the thrombin site between the aldehyde tag and the main protein (see Supplementary Note 4 for amino acid maps), producing a small N-terminal fragment that was fluorescent and a large C-terminal fragment that was nonfluorescent (Fig. 2a and see Supplementary Fig. 2 for the case of aldehyde-tagged DinB (Ald\(_6\)N-DinB)). These results show that only the N-terminal region of PolBI or DinB containing the aldehyde tag was labeled, despite the high concentration of dye used. In comparison, a widely used method that targets the N-terminal ε-amine at pH 7.0 (Supplementary Fig. 3)\(^\text{21} \) by using the difference in pK\(_a\) and, thus, reactivity between this group and the ε-amine of lysines, did not result in N terminus–specific labeling, as evaluated by the same restrictive proteolysis assay (Supplementary Note 5).

---

**Figure 1** Development and validation of the quantitative fluorescence labeling of an aldehyde-tagged protein. (a) Normalized UV-vis spectra (top) and labeling kinetics (bottom) of Ald\(_6\)N-PolBI at 4 °C incubated with 75.6 mM Cy3HZ in a pH 7.0 buffer containing 250 mM potassium phosphate, 500 mM KCl and 5 mM DTT (green). The labeling result before optimization is shown for comparison (black). (b) Schematic for the single-molecule polymerase assay. The DNA is labeled near the primer-template junction with a FRET acceptor (Cy5), whereas PolBI is labeled with a FRET donor (Cy3). (c) Total internal reflection microscopy images of the Cy3 and Cy5 channels taken before (left, −dNTP) and after (top, +dATP) addition of dATP to PolBI and DNA immobilized as shown in b. (d) FRET efficiency histograms before (bottom) and after (top) addition of dATP (left) or dGTP (right) to PolBI and DNA. (e) Donor (green) and acceptor (magenta) intensity and FRET efficiency (blue) traces of a molecule during replication (top, dATP added at ~2.5 s), and replication time (Δt) histogram (bottom).

**Figure 2** General applicability of labeling an aldehyde-tagged protein. (a) Fluorescence (left) and white-light (right) images of a gel for Cy3HZ-labeled Ald\(_6\)N-PolBI after incubation with thrombin (+ thrombin) or no thrombin (control). The nonfluorescent band between 25 kDa and 37 kDa in thrombin-treated sample is from thrombin. (b) Normalized UV-vis spectra (left) and the labeling kinetics (right) of Ald\(_6\)N-PolBI incubated with 40.5 mM Cy3HZ under the condition shown in Figure 1a. (c) Detection of the fragment ions from an Ald\(_6\)N-PolBI’s N-terminal peptide using high-resolution mass spectrometry. (d) Separation of Cy3-labeled DinB from unlabeled protein in a partially labeled sample of Ald\(_6\)N-DinB using hydrophobic interaction chromatography.
We tested the hydrolytic stability of the hydrazine linker formed between the fluorophore and protein and observed only a negligible loss of the fluorophore from the protein after several days of incubation under standard polymerase storage conditions. This result indicates good stability of the hydrazine linker under these conditions (Supplementary Fig. 4).

Aside from Cy3, we found that the labeling of Ald<sub>6N</sub>-PolBI with Cy5 hydrazide (Cy5<sub>HZ</sub>) was also quantitative under similar conditions (Fig. 2b), allowing us to label two DNA polymerases with different fluorophores and to observe their transient interactions. Such quantitative labeling can probably be extended to other fluorophores such as Alexa Fluor 555 and Alexa Fluor 647.

The procedure optimized for PolBI also improved the labeling efficiency of DinB from ~5% to 60%, but was not quantitative (Supplementary Note 6). Unlike Ald<sub>6N</sub>-PolBI, which showed a complete cysteine to formylglycine conversion in vivo (Fig. 2c), Ald<sub>6N</sub>-DinB showed only incomplete aldehyde biosynthesis, limiting the maximal labeling efficiency achievable. Using hydrophobic interaction chromatography<sup>21</sup>, we separated Cy3<sub>HZ</sub>-labeled Ald<sub>6N</sub>-DinB from unlabeled protein (Fig. 2d), thereby obtaining ~100% labeled protein.

**Labeling a protein in cell extracts and pull-down**

Because cellular proteins rarely have aldehyde functionality, it should be possible to label an aldehyde-tagged protein specifically even without purification. We therefore labeled Ald<sub>6N</sub>-DinB in E. coli extracts using the same procedure. As expected, crude lysate shows the presence of many proteins; however, only DinB showed substantial labeling in this large pool of proteins (Fig. 3a), demonstrating high target specificity. As a control, we examined the labeling of E. coli extract with FGE or DinB expressed alone and observed no appreciable labeling of any protein in either case (Fig. 3b). These results confirm the specificity of our method, and show that the background, FGE-like activity in E. coli<sup>10</sup> was much weaker than that of coexpressed FGE.

We also examined the effects of adding protease inhibitors and removing small molecules. Adding a protease inhibitor cocktail to the extract limited proteolysis, with no substantial impact on the labeling efficiency of target protein (Fig. 3a). Removing small, exchangeable molecules led to a marked improvement in the labeling efficiency, probably owing to the exclusion of molecules containing a carbonyl group such as pyruvate (Fig. 3a). We estimated the labeling efficiency to be 70–75% after 1 d of incubation at 4 °C (Supplementary Fig. 5).

After labeling, we subjected the cell extract containing labeled DinB to single-molecule pull-down assay. We immobilized a Cy5-labeled DNA molecule on the surface and then loaded onto this DNA a polymerase processivity factor PCNA<sup>1</sup> (Fig. 3c). After adding an extract containing 1 nM Cy3<sub>HZ</sub>-labeled DinB followed by a brief incubation, we used single-molecule FRET to detect binding of DinB from the extract to PCNA-DNA; in contrast, without PCNA we detected only nonspecific binding (Fig. 3d). These data closely resemble those obtained with purified DinB, and this similarity is also seen in the FRET traces of individual molecules (Fig. 3e,f). The capacity to label a protein in cell extracts should allow examination of protein function under near-native conditions, in which proteins maintain their native post-translational modifications and physiological partners<sup>11</sup>.

---

**Figure 3** Specific and efficient labeling of unpurified Ald<sub>6N</sub>-DinB in cell extracts and single-molecule pull-down. (a) White-light (left) and fluorescence (right) images of a gel for extracts of E. coli coexpressing Ald<sub>6N</sub>-DinB and FGE, labeled at 4 °C with 100 µg Cy3<sub>HZ</sub> in 3 µl for 1 d, with no other treatment (lane 2), with the addition of a protease inhibitor cocktail (lane 3), or with both the addition of protease inhibitors and the removal of small, exchangeable molecules (lane 4). (b) White-light (left) and fluorescence (right) images of a gel for extracts of E. coli expressing FGE (lane 2) or DinB (lane 3) alone, or both simultaneously (lane 4), all labeled as in (a) (lane 4), and ~92% labeled purified DinB at various concentrations used for calibration (lanes 5–8). (c) Schematic for the labeling and pull-down of Ald<sub>6N</sub>-DinB (blue) from a cell extract, using a surface-immobilized, Cy5-labeled DNA in complex with PCNA (orange). (d) Total internal reflection microscopy image showing the binding of DinB from cell extract to DNA in the presence of PCNA (left), and the control in the absence of PCNA (right). (e) Representative donor (green) and acceptor (magenta) intensity and FRET efficiency (blue) traces of individual molecules obtained by direct pull-down. (f) Representative data obtained with purified DinB.
Observation of transient interactions between polymerases

Using quantitatively labeled PolBI and DinB, we carried out a single-molecule study of transient interactions and switching between these two polymerases on PCNA. This is relevant to the process of translesion DNA synthesis, in which the actions of a high-fidelity replicative polymerase (PolBI) and an error-prone translesion polymerase (DinB) need to be coordinated\(^\text{23–25}\). First, we examined the dynamic interaction between PolBI and DNA alone, without DinB and PCNA (Fig. 4a). Using Cy\(^3\)HZ-labeled PolBI and a DNA molecule labeled with Cy5 at a four-way junction away from the 3’ end of the primer, we observed frequent FRET fluctuations between two well-defined states at 0.5 and 0.8 (Fig. 4b,c). This observation, along with the results obtained using several related DNA constructs and with multicolor FRET\(^\text{26}\) (data not shown) indicated that PolBI can shuttle repetitively on a single-stranded DNA between the proximal and distal positions (Fig. 4d).

Next, we examined the ‘toolbelt’ model, in which more than one polymerase can bind to a PCNA molecule simultaneously\(^\text{25}\). This proposal has been difficult to validate owing to the involvement of highly transient and dynamic interactions. Using quantitatively labeled PolBI and DinB, we captured relevant events in real time. When we added free Cy3-DinB to Cy5-PolBI already on PCNA-DNA, there was a brief and highly dynamic interaction between the two polymerases, observed as FRET fluctuations (Fig. 4e); this interaction requires PCNA. The interaction between DNA-bound Cy5-PolBI and free Cy3-DinB also requires PCNA, but we observed for many molecules a long delay between PolBI binding and the appearance of appreciable FRET between PolBI and DinB, suggesting long-lived colocalization of both polymerases on the same DNA (Fig. 4f). Combined with PolBI shuttling, our data suggest that the two proteins can bind to the same DNA, one on PCNA and the other off PCNA, until making a direct contact. A more in-depth examination of this and other possible models (Supplementary Fig. 6) will be reported elsewhere.

These data are the first single-molecule observation of two DNA polymerases interacting and switching in real time. This was made possible by our labeling method, which ensures minimal sample heterogeneity, adequate data throughput and preservation of biological function.

DISCUSSION

Although we report labeling only at the N terminus of a protein, previous studies suggest our method should be applicable to other locations. The formylglycine motif in sulfatases, from which the aldehyde tag was originally identified, is always located 50–80 amino acids away from the N terminus, suggesting that coexpressed FGE can modify the aldehyde tag in an internal location. Indeed, we made a preliminary observation of specific labeling of a helicase with the aldehyde tag inserted into an internal loop. In addition, the aldehyde formation at the C terminus had 99% and 45–69% efficiency for maltose-binding protein expressed in E. coli\(^\text{9}\) and the Fc region of immunoglobulin G expressed in CHO cells\(^\text{10}\), respectively.

Another protein-labeling technique with which high labeling efficiency and specificity can be achieved is based on the ybbR tag\(^\text{27}\). This method\(^\text{27,28}\) is based on an enzymatic reaction catalyzed by Sfp or AcpS phosphopantetheinyl transferase used to attach a small molecule such as a fluorescent dye in a synthetic CoA conjugate to a specific serine in a peptide tag\(^\text{27,28}\). In comparison to the aldehyde tag method, which can use most cyanine, Alexa and ATTO dyes, the ybbR tag method is currently limited to only Dy-547, Dy-647 and ATTO488. In addition, the ybbR tag method involves the expression and purification of Sfp or AcpS and necessitates purification of the target protein to remove the enzyme after labeling. The aldehyde tag is slightly shorter than the ybbR tag (6 residues versus 11–12 residues), and the ybbR tag method had yielded only 17% labeling efficiency in an earlier study\(^\text{29}\). With additional optimization, the ybbR tag method may complement the aldehyde tag method, for example, in applications requiring orthogonal labeling chemistries.

Our aldehyde-based labeling method should not be limited to proteins produced in bacteria. If a host-compatible FGE is
used, proteins expressed in eukaryotic cells could also be labeled. Coexpression of a target protein with human FGE in mammalian cells has been demonstrated. In principle, this strategy is applicable to proteins expressed in insect cells using human or fruitfly Drosophila melanogaster FGE and the MultiBac vector. Besides purified ligands, an unpurified protein can be labeled in cell extracts directly and then pulled down from the mixture in a physiological form with appropriate post-translational modifications or in a complex with cellular partners. The use of a short aldehyde tag and small cyanine dye makes this labeling scheme more advantageous than existing methods using a fluorescent protein or labeled antibody for imaging, both of which are much larger in size and may not accurately reflect the physiological function, complex stoichiometry, in vivo modification and interaction with cellular partners associated with a protein of interest. Note that it may be difficult to apply this method to systems in which genetic manipulation is cumbersome, such as Xenopus laevis extracts.

By using the aldehyde-based labeling method, many proteins can now be examined directly. One good example is the five-subunit, clamp-loader complex RFC, which has been difficult to label using conventional methods. This 222-kDa complex consists of three components named RFCL, RFCS1 and RFCS2 in a 1:3:1 ratio, containing 16 cysteines and 132 lysines in total. We labeled this protein complex specifically at the N terminus of RFC1 with ≥78% efficiency (X.S., Y.J., I.K.O.C. and T.H., unpublished data). The range of proteins we labeled implies that even larger entities, such as the ribosome, may be tackled the same way.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank C. Bertozzi (University of California, Berkeley) for providing the plasmid DNA for FGE through Addgene, J. Fei for the suggestion of using hydrophobic interaction chromatography, and K. Raganathan for a critical reading of the manuscript. This work was supported by the US National Institutes of Health grants GM065367 and AI083025 (to T.H.) and the US National Science Foundation grants MCB-0238451 (to I.C.) and PHY-0822613 (to T.H.).

AUTHOR CONTRIBUTIONS

X.S. and T.H. conceived the study; X.S. designed the experiments; X.S., Y.J., L.-J.L., C.L. and C.W. performed the experiments and analyzed the data; X.S., I.K.O.C. and T.H. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. & Ha, T. Advances in single-molecule fluorescence methods for molecular biology. Annu. Rev. Biochem. 77, 51–76 (2008).
2. Li, G.W. & Xie, X.S. Central dogma at the single-molecule level in living cells. Nature 475, 308–315 (2011).
3. Ha, T. & Tinnefeld, P. Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. Annu. Rev. Phys. Chem. 63, 26.1–26.23 (2012).
4. Gilmore, J.M., Scheck, R.A., Esser-Kahn, A.P., Joshi, N.S. & Francis, M.B. N-terminal protein modification through a biomimetic transamination reaction. Angew. Chem. Int. Edn Engl. 45, 5307–5311 (2006).
5. Algire, M.A., Maag, D. & Lorsch, J.R.P. release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. Mol. Cell. 20, 251–262 (2005).
6. Sletten, E.M. & Bertozzi, C.R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. Angew. Chem. Int. Edn Engl. 48, 6974–6989 (2009).
7. Adams, S.R. et al. New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. J. Am. Chem. Soc. 124, 6063–6076 (2002).
8. Lata, S., Gavutis, M., Tampe, R. & Pielker, J. Specific and stable fluorescence labeling of histidine-tagged proteins for dissecting multi-protein complex formation. J. Am. Chem. Soc. 128, 2365–2372 (2006).
9. Carrico, I.S., Carlson, B.L. & Bertozzi, C.R. Introducing genetically encoded aldehydes into proteins. Nat. Chem. Biol. 3, 321–322 (2007).
10. Wu, P. et al. Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. Proc. Natl. Acad. Sci. USA 106, 3000–3005 (2009).
11. Jain, A. et al. Probing cellular protein complexes using single-molecule pull-down. Nature 473, 484–488 (2011).
12. Yeom, K.H. et al. Single-molecule approach to immunoprecipitated protein complexes; insights into miRNA uridylation. EMBO Rep. 12, 690–696 (2011).
13. Yardimci, H., Loveland, A.B., Habuchi, S., van Oijen, A.M. & Walter, J.C. Uncoupling of sister replisomes during eukaryotic DNA replication. Mol. Cell 40, 834–840 (2010).
14. Hoskins, A.A. et al. Ordered and dynamic assembly of single spliceosomes. Science 331, 1289–1295 (2011).
15. Chen, Y.H. et al. Biochemical and mutational analyses of a unique clamp loader complex in the archaeon Methanosarcina acetivorans. J. Biol. Chem. 280, 41852–41863 (2005).
16. Lin, L.J. et al. Molecular analyses of an unusual translesion DNA polymerase from Methanosarcina acetivorans C2A. J. Mol. Biol. 397, 13–30 (2010).
17. Dirksen, A., Hackeng, T.M. & Dawson, P.E. Nucleophilic catalysis of oxime ligation. Angew. Chem. Int. Edn Engl. 45, 7581–7584 (2006).
18. Ha, T. et al. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. Proc. Natl. Acad. Sci. USA 93, 6264–6268 (1996).
19. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. Nat. Methods 5, 507–516 (2008).
20. Kong, H., Kucera, R.B. & Jack, W.E. Characterization of a DNA polymerase from the hyperthermophilic archaea Thermococcus litoralis. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. J. Biol. Chem. 268, 1965–1975 (1993).
21. Galletto, R., Amatini, I., Baskin, R.J. & Kowalczykowski, S.C. Direct observation of individual RecA filaments assembling on single DNA molecules. Nature 443, 875–878 (2006).
22. Steinberg, S.H., Fei, J., Prywes, N., McGath, K.A. & Gonzalez, R.L. Jr. Translation factors drive intrinsic ribosome dynamics during translation termination and ribosome recycling. Nat. Struct. Mol. Biol. 16, 861–868 (2009).
23. Moldovan, G.L., Pfander, B. & Jentsch, S. PCNA, the maestro of the replication fork. Cell 129, 665–679 (2007).
24. Yang, W. & Woodgate, R. What a difference a decade makes: insights into translesion DNA synthesis. Proc. Natl. Acad. Sci. USA 104, 15591–15598 (2007).
25. Diadani, C., McInerney, P., Georgescu, R., Goodman, M.F. & O’Donnell, M. A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously. Mol. Cell 19, 805–815 (2005).
26. Lee, J. et al. Single-molecule four-color FRET. Angew. Chem. Int. Edn Engl. 49, 9922–9925 (2010).
27. Yin, J. et al. Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinytransferase. ACS Chem. Biol. 2, 337–346 (2007).
28. Lee, G., Yoo, J., Leslie, B.J. & Ha, T. Single-molecule analysis reveals three phases of DNA degradation by an exonuclease. Nat. Chem. Biol. 7, 367–374 (2011).
29. Fitzgerald, D.J. et al. Protein complex expression by using multigene baculoviral vectors. Nat. Methods 3, 1021–1032 (2006).
**ONLINE METHODS**

**Plasmid preparation.** All oligonucleotides used in this study (Supplementary Table 1) were purchased from Integrated DNA Technologies (IDT). For cloning, the two NcoI sites found in the gene encoding PolBI were removed using a QuickChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies).

The sequence of N-terminal aldehyde tag was designed into the primer Ald6N-forward with an NcoI site at the 5′ end. This primer was used in a PCR along with a reverse primer containing a NotI or SalI site at the 5′ end. Purified PCR product was tailed at both 3′ ends with a single adenine using Taq DNA polymerase (New England Biolabs), and then ligated into the pGEM-T Easy vector (Promega) using T4 DNA ligase (Promega). After transformation with the ligation product, JM109 cells (Promega) were cultured at 37 °C for 1 h, spread on an LB-ampicillin plate, and grown at 37 °C overnight. Colonies containing the gene construct encoding Ald6N-PolBI or Ald6N-DinB were cultured in LB medium containing 100 µg ml−1 ampicillin at 37 °C overnight. Plasmid DNA was extracted using the QIAprep Spin Miniprep kit (Qiagen). Restriction digestion and sequencing confirmed the presence of the gene construct for Ald6N-PolBI or Ald6N-DinB in the plasmid obtained. This plasmid, along with the plasmid containing the pET-28a(+) vector, were digested with restriction enzymes at 37 °C for 3 h. Purified Ald6N-PolBI or Ald6N-DinB insert and pET-28a(+) vector were ligated using T4 DNA ligase at 16 °C overnight. The final plasmid was obtained and validated by the same procedure.

DH5α cells transformed with the FGE plasmid in the pBAD/Myc-His A vector were obtained from Addgene (plasmid 16132). Purified plasmid was prepared as described in this section.

**Protein expression and purification.** After co-transformation with the FGE and Ald6N-PolBI or Ald6N-DinB plasmids, BL21(DE3) cells (Agilent Technologies) were cultured at 37 °C for 1 h, spread on an LB-ampicillin-kanamycin plate, and grown at 37 °C overnight. Single colonies were cultured in 10 ml of LB medium with 100 µg ml−1 ampicillin and 30 µg ml−1 kanamycin at 37 °C for 6 h, followed by dilution into 1 liter of LB medium containing the same antibiotics. When the optical density at 600 nm (OD600) reached 0.3, expression of FGE was induced by addition of l-(+)-arabinose to 0.2% (wt/vol). Expression of Ald6N-PolBI or -DinB was induced 30 min later by addition of isopropyl-β-D-thiogalactoside (IPTG) to 0.1 mM, and then the temperature was lowered to 16 °C. Cells were collected after 16 h by centrifugation at 7,000g and then resuspended in lysis buffer (50 mM sodium phosphate (pH 7.0) and 300 mM NaCl), followed by homogenization using a French press. After centrifugation at 10,000g the supernatant was collected, mixed with Talon metal-affinity resin (Clontech), and incubated at 4 °C for 1 h. After washing with lysis buffer plus 10 mM imidazole, proteins were eluted off the resin using lysis buffer plus 150 mM imidazole.

Proteins were exchanged into buffer A (50 mM Tris (pH 8.5)) using an Amicon Ultra-15 centrifugal filter unit (Millipore), and separated on a Mono Q 5/50 GL column using ÄKTAFPLC (GE Healthcare) and a 50-ml gradient from 0 to 50% buffer B (buffer A plus 1 M NaCl) at a flow rate of 1 ml min−1. Desired fractions were exchanged into the gel-filtration buffer (50 mM Tris (pH 8.5), 150 mM NaCl and 0.5 mM DTT), and further separated on a HiLoad 16/60 Superdex 200 column (GE Healthcare).

Purified protein was exchanged into the storage buffer (gel-filtration buffer plus 10% (vol/vol) glycerol). *M. acetivorans* RFC and PCNA were expressed in *E. coli* and purified as described previously15.

**Labeling aldehyde-tagged proteins.** Ald6N-PolBI or Ald6N-DinB was exchanged into labeling buffer (pH 7.0) using an Amicon Ultra-4 or Ultra-0.5 centrifugal filter unit (Millipore). The labeling buffer contained 50 mM potassium phosphate, 100 mM KCl and 1 mM DTT (pre-optimization), or identical ingredients but each with a concentration five times higher (post-optimization). After buffer exchange, the protein was mixed with dried Cy3HZ or Cy5HZ (GE Healthcare). For the quantitative labeling established in this work, we used 100 µg of dye in 3-µl reactions or 1 mg dye in 30-µl reactions.

To characterize the labeling kinetics, we withdrew at each time point an aliquot of 0.5–1 µl from the 3-µl mixture. Unincorporated free dye was removed by passing the sample through two Micro Bio-Spin columns (Bio-Rad). A UV-vis spectrum was taken for each sample using a Cary 50 spectrophotometer (Varian), from which absorbance measurements were obtained for the label and protein. A correction factor of 0.08 or 0.017 was used for Cy3 or Cy5, respectively, to account for their absorption at 280 nm. We used the published molar extinction coefficients of Cy3 and Cy5: 150,000 M−1 cm−1 and 30,850 M−1 cm−1, respectively (ref. 31). Note that covalent conjugation of these dyes to protein led to a small but detectable red shift in their absorption maxima and, possibly, a minor change in extinction coefficient; however, this was not considered in our calculation. For Ald6N-PolBI and Ald6N-DinB, we used the extinction coefficients calculated by the ProtParam program (ExPASy): 83,500 M−1 cm−1 and 28,500 M−1 cm−1, respectively. In principle, the values obtained experimentally (Supplementary Fig. 7), 81,600 M−1 cm−1 and 30,850 M−1 cm−1, should be used instead. That is, all values of $P_{\text{label}}$ (l) reported in this work should be corrected by −2.3% and 8.2% for PolBI and DinB, respectively.

To label Ald6N-PolBI at the N-terminal α-amine21, we exchanged the protein into the labeling buffer above, mixed it with 2 µg of dried Cy3 NHS-ester (GE Healthcare), and incubated the mixture at room temperature (22–23 °C) for 30 min, followed by 4 °C overnight.

To characterize the stability of the hydrazone linker between the label and protein, we exchanged Cy3HZ-labeled PolBI into the storage buffer (pH 8.5) and kept it in the dark at 4 °C. After 1 d, 2 d and 4 d, we drew aliquots from the sample and removed the free dye using one Micro Bio-Spin column.

**Analysis of the aldehyde-tagged protein’s labeling kinetics.** The labeling of aldehyde-tagged protein with cyanine hydrazide and the reverse reaction of hydrolysis can be described by

$$A + B \rightarrow AB,$$

where $A$, $B$ and $AB$ are the unlabeled protein, dye and labeled protein, respectively. The rate is given by

$$\frac{d[A]_0}{dt} = k_2[A]_0[B] - k_{-1}(AB),$$

where $k_2$ and $k_{-1}$ are rate constants for the forward and reverse reaction, respectively. When $B$ is in large excess, $[B]$,
is essentially a constant, making the forward reaction pseudo-first order. Then
\[
\frac{d[A]_0}{dt} = k_2[A]_0[B]_0 - k_{-1}([A]_0 - [A]_f)
\]
\[= k'_1[A]_0 - k_{-1}([A]_0 - [A]_f)\]
\[= (k'_1 + k_{-1})[A]_0 - \frac{k_{-1}}{k'_1 + k_{-1}}(A)_{0}\]
\hspace{1cm}(3)

where \(k'_1 = k_2[B]_0\) is the rate constant for the pseudo-first-order reaction. Solving equation (3) gives
\[
[A]_t = [A]_0 \times \left[\frac{k_{-1}}{k'_1 + k_{-1}} + \frac{k'_1}{k'_1 + k_{-1}} e^{-(k'_1 + k_{-1})t}\right].
\hspace{1cm}(4)

The labeling efficiency at time \(t\) is given by
\[
P_{\text{label}}(t) = \frac{[AB]_t}{[AB]_0} = \frac{k'_1}{k'_1 + k_{-1}}(1 - e^{-(k'_1 + k_{-1})t})).
\hspace{1cm}(5)

The labeling kinetics is governed by the apparent rate constant \(k_{\text{app}}\), which equals \(k'_1 + k_{-1}\). The time to reach half of the maximal labeling efficiency, \(\tau_{1/2}\), is given by \(\ln(2)/k_{\text{app}}\), whereas the maximal labeling efficiency, \(P_{\text{label}}(t \to \infty)\), is given by \(k'_1/k_{\text{app}}\). When \(k'_1 >> k_{-1}\), \(P_{\text{label}}(t)\) becomes \(1 - \exp(-k'_1t)\) and \(P_{\text{label}}(t \to \infty)\) should reach 100%.

**Hydrophobic interaction chromatography.** Partially labeled Ald6N-DinB was diluted into buffer C (50 mM sodium phosphate (pH 8.0) and 1 M (NH4)2SO4) and separated on an TSKgel Phenyl-5PW column (Tosoh Bioscience) at 4 °C using an ÄKTApurifier (GE Healthcare) and a 20-column-volume gradient from 0 to 100% (vol/vol) formamide and 1 mM EDTA. After heat denaturation by dilution in a 1:1 volume ratio into a solution containing 98% (vol/vol) formamide and 1 mM EDTA, the sample was examined by SDS-PAGE at a flow rate of 1 ml min\(^{-1}\).

**DNA preparation.** We mixed 27-nt DNA D27-5Am with an amino-C6-dT modification at the fifth base from the 3′ end and a biotin moiety at the 5′ end, with 200 µg of dried Cy5 NHS-ester (GE Healthcare) in 100 mM sodium tetraborate (pH 8.5) and incubated the mixture at room temperature overnight. Free dye was removed by using two Micro Bio-Spin columns. We used the same method to label the 22-nt strand J7b+4Am with an amino-C6-dT modification at the fourth base from the 5′ end. The labeling efficiency was almost 100% for both strands.

Cy5-labeled D27-5Am was mixed in a 1.5:1:2:2:2 molar ratio with 69-nt strand HJ47A-T20 and the three 22-nt strands J7b, J7h and J7r, respectively. Unlabeled D27-5Am was mixed in the same ratio with these strands. Cy5-labeled J7b+4Am was mixed in a 2:1.5:1:2:2 ratio with D27-5Am, HJ47A-T20, J7h and J7r. After heat denaturation at 90 °C for 1 min and slow cooling to room temperature, annealed DNA was separated on an 8% native polyacrylamide gel. The 27-nt strand D27-5Cy5, which was labeled internally with Cy5, was mixed in a 1:2 ratio with 47-nt or 69-nt strand D27-T20 or D27-T42, respectively, and the partial duplex DNA desired was obtained without gel purification.

**Ensemble polymerase assay.** A mixture of DNA, polymerase and dNTPs was prepared at a concentration of 0.1 µM, 0.5 µM and 250 µM, respectively, in a buffer containing 20 mM Tris (pH 8.8), 2 mM β-mercaptoethanol, 5 mM MgCl₂ and 0.1 mg ml\(^{-1}\) BSA. After 10 min of incubation at 37 °C, the reaction was stopped by dilution in a 1:1 volume ratio into a solution containing 98% (vol/vol) formamide and 1 mM EDTA. After heat denaturation at 95 °C for 5 min, the reaction product was examined using a 10% TBE-Urea gel (Bio-Rad). The gel was then imaged using an ImageQuant LAS 4000 system (GE Healthcare).

**Single-molecule polymerase assay.** Quartz slides and cover glasses were subjected to PEGylation before assembly into an imaging chamber. Cy5-labeled DNA containing a (dT)\(_{20}\) template and a four-way junction was immobilized on the surface through biotin-NeutrAvidin (Thermo) linkage. Cy3\(_{1HZ}\)-labeled PolBI was then added at a concentration of 0.5–1 nM in imaging buffer containing 25 mM Tris (pH 8.0), 5 mM MgCl₂, 0.8% (wt/vol) dextrose, 2 mM Trolox (Aldrich), 0.04 mg ml\(^{-1}\) catalase (Sigma) and 1.0 mg ml\(^{-1}\) glucose oxidase (Sigma). After 5 min of incubation at room temperature, unbound protein was removed. We then added 1 mM dATP or dGTP at pH 8.8 into the chamber and incubated it for 1 min before imaging. In the case of DinB, Cy3\(_{1HZ}\)-labeled polymerase was added at 1–2 nM after incubating the DNA with 20 nM RFC, 1 mM ATP and 5 mM PCNA.

Imaging was performed with the prism-type total internal reflection microscopy. A 532-nm Excelsior laser (Spectra Physics) was used to excite Cy5 on the labeled polymerase. Emission from both Cy3 and Cy5 was collected through a 60x, 1.2 NA, water immersion objective (Olympus), separated with a 630dcxr dichroic beam splitter (Chroma), and detected using an iXon DU-897 EMCCD camera (Andor). Individual molecules were identified using a custom program.

**Restrictive proteolysis.** We incubated 3.1 µl of labeled polymerase with 0.5 U restriction-grade thrombin (EMD Chemicals) at room temperature overnight. After a 1:1 dilution into Laemmli sample buffer (Bio-Rad) and heat denaturation at 95 °C for 5 min, the sample was examined by SDS-PAGE on a 4–20% Tris-HCl gel (Bio-Rad). Unless noted, Precision Plus Protein Kaleidoscope Standards (Bio-Rad) were used as molecular weight markers. The gel was imaged for fluorescence using the imager mentioned before. After staining with Bio-Safe Coomassie (Bio-Rad), this gel was imaged again under white-light illumination.

**Measurement of the protein extinction coefficient.** An aliquot of aldehyde-tagged polymerase was diluted into either the native storage buffer, or a denaturing buffer containing 20 mM sodium phosphate (pH 6.5) and 6.0 M guanidine hydrochloride\(^{33}\). The extinction coefficient of a protein in the storage buffer, \(\varepsilon_{\text{nat}}(\lambda_{\text{max}})\), is
\[
\varepsilon_{\text{nat}}(\lambda_{\text{max}}) = \frac{A_{\text{nat}}(\lambda_{\text{max}})}{A_{\text{denat}(280)}} \varepsilon_{\text{denat}(280)},
\hspace{1cm}(6)
\]
where \(A_{\text{nat}}(\lambda_{\text{max}})\) and \(A_{\text{denat}(280)}\) are the peak absorbance under native conditions and the 280-nm absorbance under denaturing conditions, respectively, and \(\varepsilon_{\text{denat}(280)}\) is the protein’s extinction coefficient at 280 nm under denaturing conditions (Supplementary Fig. 7). \(\varepsilon_{\text{denat}(280)}\) is obtained by
\[
\varepsilon_{\text{denat}(280)} = \varepsilon_{\text{W}} e_{\text{Y}} + \varepsilon_{\text{C}} e_{\text{C}},
\hspace{1cm}(7)
\]
where \(\varepsilon_{\text{W}}, \varepsilon_{\text{Y}}, \text{ and } \varepsilon_{\text{C}}\) are the number of tryptophans, tyrosines and cysteines in the protein, and \(\varepsilon_{\text{W}}, \varepsilon_{\text{Y}}, \text{ and } \varepsilon_{\text{C}}\) are the extinction coefficients of the corresponding model compounds at 280 nm under denaturing conditions\(^{33}\). The extinction coefficient of Ald6N-PoBI and Ald6N-DinB under native conditions was thus determined to be 81,300–81,900 M\(^{-1}\) cm\(^{-1}\) and 30,700–31,000 M\(^{-1}\) cm\(^{-1}\), respectively.
High-resolution mass spectrometry. Aldehyde-tagged polymerase was mixed in a 50:1 weight ratio with sequencing-grade modified trypsin (Promega) in a buffer containing 100 mM ammonium bicarbonate (pH 7.8) and incubated at 37 °C overnight. The product was separated on a nanoLC-2D system (Eksigent) at a flow rate of 300 nL min\(^{-1}\), using a gradient from 95% buffer E containing 5% (vol/vol) acetonitrile and 0.2% (vol/vol) formic acid, and 5% (vol/vol) buffer F containing 95% (vol/vol) acetonitrile and 0.2% (vol/vol) formic acid, to 55% buffer E and 45% buffer F in 50 min, and then to 15% buffer E and 85% buffer F in 10 min.

Sample eluted off the reverse-phase liquid chromatography (RPLC) was electrosprayed into an 11-tesla LTQ-FT Ultra hybrid mass spectrometer (Thermo). For tandem mass spectrometry (MS/MS) analysis, we used a data-dependent, top-3 strategy with a 10 m/z isolation window. MS1 was performed in the 350–1,300 m/z scan range at 50,000 resolving power with an FT-ICR cell. MS2 was performed at 25,000 resolving power after collision-induced dissociation. Data were analyzed with the ProSightPC 2.0 SP1 software (Thermo)\(^3\).

Labeling Ald\(_{6N}\)-DinB in cell extracts and pull-down. BL21(DE3) cells co-transformed with the Ald\(_{6N}\)-DinB and FGE plasmids were cultured in 1 liter LB medium as described earlier. The same procedure was carried out for the controls, except that cells were transformed with either plasmid alone and then cultured in the presence of a single antibiotic. Cells were collected after 16 h and resuspended in a 1:4.2 weight-to-volume ratio in labeling buffer, containing ~1 nM Cy3 HZ-labeled DinB was added to surface-immobilized DNA in complex with PCNA (see above). After a 5-min incubation at room temperature, DinB was pulled down from the mixture of cellular proteins and other macromolecules, and unbound molecules were removed. As a control, the imaging buffer was added in place of the RFC-ATP-PCNA mixture before pulling down DinB.

Imaging transient interactions and switching between polymerases. The DNA mentioned above containing a (dT)\(_{20}\) template and Cy5-labeled four-way junction was immobilized in the imaging chamber. Cy3 HZ-labeled PolBI was added at 1 nM. After 5 min, unbound protein was removed using imaging buffer (pH 8.0). To observe the interactions and switching between different DNA polymerases, the DNA described earlier containing a (dT)\(_{20}\) template and an unlabeled four-way junction was immobilized on the surface and incubated for 5 min with a mixture of 20 nM RFC, 1 mM ATP and 5 nM PCNA. Cy5 HZ-labeled DinB was then added at a concentration of 5 nM, followed by Cy3 HZ-labeled PolBI to a concentration of 1 nM. A movie was recorded during the flow of PolBI. Alternatively, Cy5 HZ-labeled PolBI was added first at a concentration of 2 nM, and then Cy3 HZ-labeled DinB was added to a concentration of 1 nM. A similar procedure was carried out for the controls, except that the imaging buffer was added in place of the RFC-ATP-PCNA mixture, Cy3-labeled polymerase or Cy3-labeled polymerase, respectively.

\(31\). Berlier, J.E. et al. Quantitative comparison of long-wavelength Alexa Fluor dyes to Cy dyes: fluorescence of the dyes and their bioconjugates. J. Histochem. Cytochem. 51, 1699–1712 (2003).

\(32\). Shi, X., Lim, J. & Ha, T. Acidification of the oxygen scavenging system in single-molecule fluorescence studies: in situ sensing with a ratiometric dual-emission probe. Anal. Chem. 82, 6132–6138 (2010).

\(33\). Gill, S.C. & von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 182, 319–326 (1989).

\(34\). Lee, J.E. et al. A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. J. Am. Soc. Mass Spectrom. 20, 2183–2191 (2009).