

ARTICLE

Received 15 Nov 2012 | Accepted 11 Jun 2013 | Published 8 Jul 2013

DOI: 10.1038/ncomms3131

OPEN

# Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion

Johannes Hohlbein<sup>1,†</sup>, Louise Aigrain<sup>1,\*</sup>, Timothy D. Craggs<sup>1,2,\*</sup>, Oya Bermek<sup>2,†</sup>, Olga Potapova<sup>2</sup>, Pouya Shoolizadeh<sup>1</sup>, Nigel D.F. Grindley<sup>2</sup>, Catherine M. Joyce<sup>2</sup> & Achillefs N. Kapanidis<sup>1</sup>

The fidelity of DNA polymerases depends on conformational changes that promote the rejection of incorrect nucleotides before phosphoryl transfer. Here, we combine single-molecule FRET with the use of DNA polymerase I and various fidelity mutants to highlight mechanisms by which active-site side chains influence the conformational transitions and free-energy landscape that underlie fidelity decisions in DNA synthesis. Ternary complexes of high fidelity derivatives with complementary dNTPs adopt mainly a fully closed conformation, whereas a conformation with a FRET value between those of open and closed is sparsely populated. This intermediate-FRET state, which we attribute to a partially closed conformation, is also predominant in ternary complexes with incorrect nucleotides and, strikingly, in most ternary complexes of low-fidelity derivatives for both correct and incorrect nucleotides. The mutator phenotype of the low-fidelity derivatives correlates well with reduced affinity for complementary dNTPs and highlights the partially closed conformation as a primary checkpoint for nucleotide selection.

<sup>1</sup>Biological Physics Research Group, Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK. <sup>2</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. \* These authors contributed equally to this work. † Present addresses: Laboratory of Biophysics, Wageningen UR, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands (J.H.); Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA (O.B.). Correspondence and requests for materials should be addressed to A.N.K. (email: a.kapanidis1@physics.ox.ac.uk).

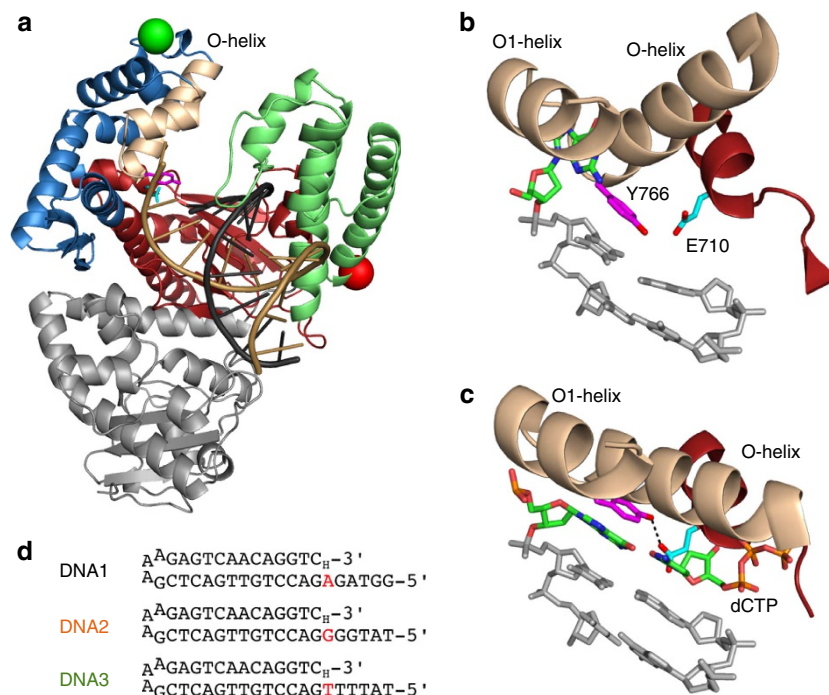
The remarkable fidelity of DNA synthesis is crucial for genome stability<sup>1</sup>. This fidelity depends on non-covalent transitions preceding the chemical step of phosphoryl transfer and serving as checkpoints whereby inappropriate substrates are rejected early during the reaction<sup>2</sup>. Among these transitions, the fingers-closing transition is of particular interest: the binding of the complementary dNTP to a polymerase-DNA (Pol-DNA) binary complex leads to a conformational change in which a part of the fingers subdomain (including the O-helix) moves from an open to a closed conformation (Fig. 1a–c), and forms a tight binding pocket around the nascent base pair<sup>3–6</sup>. Previously, we used single-molecule fluorescence resonance energy transfer (FRET) to examine fingers-closing in the *Escherichia coli* DNA polymerase I (Klenow fragment) (KF; see Santoso *et al.*<sup>7</sup>). Our results identified the open and closed conformations that predominate in the binary Pol-DNA and ternary Pol-DNA-dNTP (complementary) complexes, respectively, and showed that the unliganded polymerase explores both open and closed conformations<sup>7–10</sup>. We also showed that ternary complexes with non-complementary dNTPs or complementary ribonucleotides form distinct FRET species (with FRET signals intermediate between the open and closed states).

The fidelity and speed of DNA replication can be profoundly affected by substitution of amino acids near the polymerase active site<sup>11–14</sup>, implying that specific side chains control

the discrimination between correct and incorrect nucleotides. Two well-studied amino acids contributing to fidelity in KF are E710 and Y766 (Fig. 1a), both universally conserved within A-family polymerases<sup>15</sup>. Residue E710 helps in positioning the incoming nucleotide by interacting with the dNTP base-sugar portion, and acts as a steric gate that blocks incorporation of ribonucleotides<sup>16</sup>. Residue Y766 stacks on the template base of the terminal base pair and blocks the insertion site in the binary complex (Fig. 1b). Upon fingers-closing, Y766 rotates away from the terminal base pair and is replaced by the templating base, which moves into the catalytic pocket (Fig. 1c); the hydrogen-bonded Y766-E710 pair creates part of the binding pocket and controls the nascent base-pair geometry.

Here, we report a conformational analysis of wild-type KF (wt KF) and KF derivatives with single amino-acid substitutions at E710 and Y766, respectively, and obtain conformational energy landscapes for various complexes along the prechemistry pathway. Three of the substitutions examined (E710A, Y766A and E710Q) decrease fidelity, acting as ‘mutators’<sup>11–14</sup>; we also analysed Y766F, a derivative with enzymatic properties similar to wt KF<sup>11,14</sup>.

Since mutators fail to transition efficiently from the partially closed to the closed conformation, our results indicate that the partially closed state is an important fidelity checkpoint and that residues E710 and Y766 are essential for the discrimination between complementary and non-complementary nucleotides.



**Figure 1 | Structural view of DNA polymerase I and key residues in this study.** (a) The open Pol-DNA binary complex (PDB accession code 1L3U) is illustrated using structural data from *Bacillus stearothermophilus* (Bst) DNA polymerase<sup>4</sup>, a close homologue of KF. In the following, the KF numbering scheme is used. The fingers subdomain is coloured blue with the exception of the O- and O1-helices coloured beige, the thumb green, the palm subdomain red and the 3'-5' exonuclease grey. The DNA template strand is coloured brown and the DNA primer strand black. The amino acid Y766 is coloured magenta and E710 cyan. The positions of the residues used for labelling are shown as a red sphere for the acceptor (residue 550) and a green sphere for the donor (residue 744). See also Supplementary Methods and Supplementary Table S1 for details regarding site-specific labelling. (b) Position of E710 and Y766 in the Pol-DNA binary complex (PDB accession code 1L3U). The O-helix is in the open conformation, Y766 is stacked on the template base of the terminal base pair and the next templating base is in a pocket between the O- (mobile) and O1- (immobile) helices. The last two base pairs of the template-primer duplex are in grey. (c) Position of E710 and Y766 in the Pol-DNA-dNTP (correct) ternary complex (PDB accession code 1LV5 (ref. 4)). With the O-helix in the closed position, Y766 has moved away from the terminal base pair, creating space for the templating base. Y766 forms a hydrogen bond with E710, shown as dashed black line. (d) DNA oligonucleotides used in this study. A dideoxy nucleotide (3'-H) prevents extension of the 3' terminus. The templating base is shown in red.

## Results

**wt KF shows an on-pathway intermediate.** The conformational profiles for wt KF alone and in complexes with non-extendable DNA substrates (Fig. 1d) are shown in Fig. 2a. The unliganded polymerase exhibits a broad FRET distribution (Fig. 2a, row 1), slightly biased towards an open conformation (60% occupancy,  $E_{\text{lowFRET}}^* \sim 0.48$ ). Formation of the KF–DNA binary complex (using DNA1, Fig. 1d) further increases the open-state occupancy (81%,  $E_{\text{open}}^* \sim 0.45$ ; Fig. 2a, row 2). The peak positions of the low-FRET species are slightly different in the unliganded polymerase and the binary complex; this is mainly because of unliganded polymerase dynamics<sup>7,8,10</sup>, but may also be due to conformational states other than the open and closed states, as discussed below.

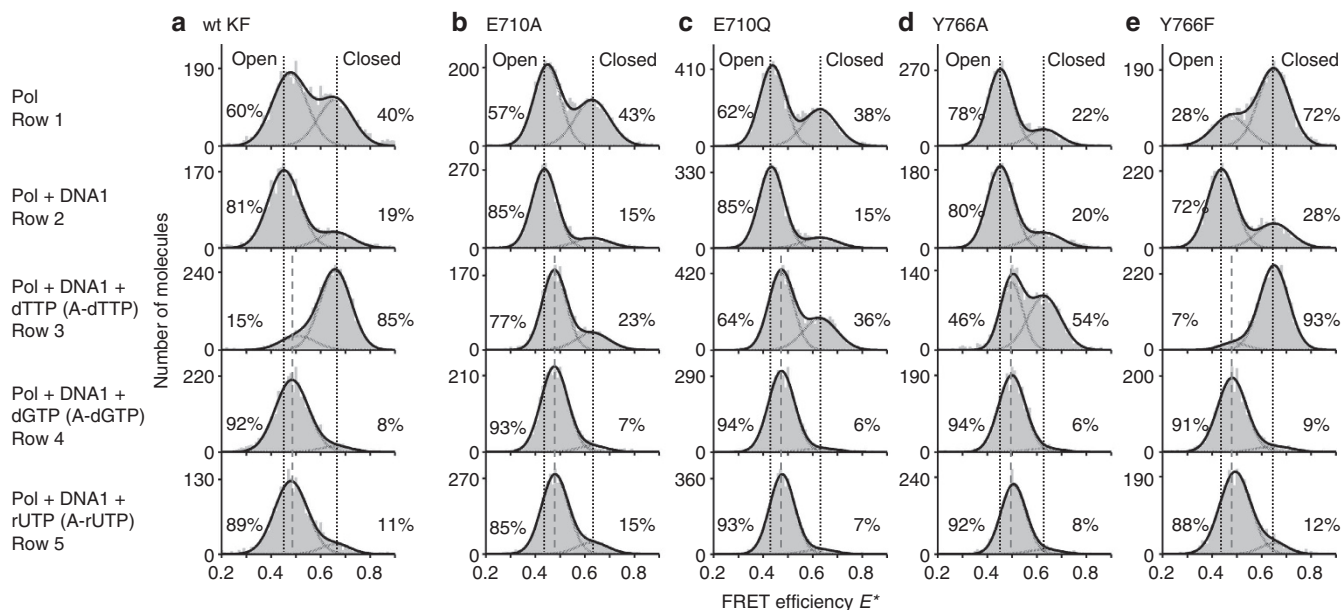
Addition of the complementary nucleotide (A–dTTP) forms a ternary complex, wherein  $\sim 85\%$  of all molecules display a closed conformation ( $E_{\text{closed}}^* \sim 0.66$ ; Fig. 2a, row 3). Notably, the presence of 1 mM dTTP (instead of 10  $\mu\text{M}$  used previously<sup>7</sup>) increases significantly the  $E^*$  of the low-FRET species ( $E_{\text{lowFRET}}^* \sim 0.51$ ) compared with the  $E_{\text{open}}^*$  in the binary complex (see Supplementary Fig. S1). Indeed, the  $E_{\text{lowFRET}}^*$  for 1 mM dTTP is closer to the intermediate-FRET species ( $E^* \sim 0.48$ ) also populated by complexes containing 1 mM of either non-complementary dNTP or complementary rNTP (A–dGTP and A–rUTP, respectively; Fig. 2a, rows 4 and 5). For both A–dGTP and A–rUTP ternary complexes, the occupancy ratio of the  $E^* \sim 0.48$  FRET species and the closed conformation is  $\sim 9:1$ .

We measured nucleotide-binding affinities by analysing FRET histograms generated at different nucleotide concentrations (Fig. 3a). The equilibria between the FRET species are depicted in Fig. 3c: the initial binding of a nucleotide (N) to the open

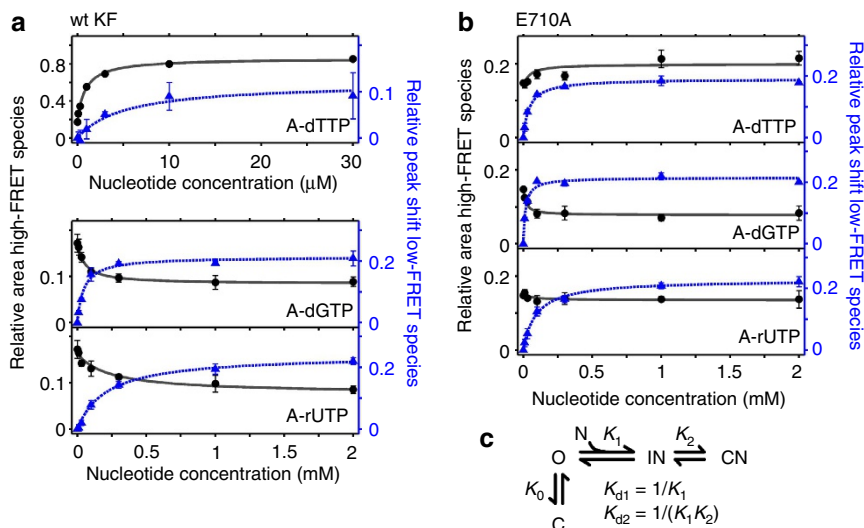
binary complex (O) giving the intermediate-FRET species (IN) is described by the association constant  $K_1$  (or its reciprocal  $K_{d1}$ ), whereas the conformational equilibrium between (IN) and the closed ternary complex (CN) is described by constant  $K_2$ . In addition, the binary complex is in equilibrium between open (O) and closed (C) states, described by constant  $K_0$ . The nucleotide titrations provide two observables (Fig. 3a): the first is the peak shift of the low-FRET Gaussian (comprising the unresolvable O and IN species); the halfway point, corresponding to  $[O] = [IN]$ , is at a nucleotide concentration equal to  $K_{d1}$ . The second observable is the fraction of polymerase complexes in high-FRET conformations. We used global fitting of both observables to obtain  $K_0$ ,  $K_1$  and  $K_2$  (Table 1). Extensions of our model are possible without altering its general conclusions; a five-state model including a potential intermediate-FRET species in the binary complex led only to a small additional scaling factor for  $K_{d1}$  (see Supplementary Methods).

With wt KF, the complementary dNTP (A–dTTP) forms the intermediate-FRET species with a  $K_{d1}$  of 5.1  $\mu\text{M}$  (Table 1); the favourable equilibrium towards the closed complex ( $K_2 = 6.1$ ) augments the binding affinity so that the closed conformation is stabilized (dissociation constant of the closed ternary complex,  $K_{d2} = 1/(K_1K_2) = 0.83 \mu\text{M}$ ). The affinity of incorrect nucleotides or complementary rNTPs for the intermediate state is weaker ( $K_{d1}$ : 46  $\mu\text{M}$  for A–dGTP, 210  $\mu\text{M}$  for A–rUTP and 30  $\mu\text{M}$  for A–rGTP; Table 1) and the low values of  $K_2$  ( $< 0.2$  in all cases) inhibit full fingers-closing.

Titration of wt KF with nucleotides in the absence of DNA (Supplementary Fig. S2e) show formation of an intermediate-FRET species, as reported previously<sup>7</sup>. The initial binding affinity  $K_{d1}$  (3.0  $\mu\text{M}$  for dGTP; 12  $\mu\text{M}$  for rGTP) is similar to that for a



**Figure 2 | Conformational profiles for wt KF and its derivatives.** Histograms of the FRET efficiency  $E^*$  for wt KF (a), E710A (b), E710Q (c), Y766A (d) and Y766F (e). Each histogram (70 bins) was derived from several thousand fluorescence bursts detected from labelled polymerases diffusing through a confocal spot. The histograms were fitted to a double-Gaussian model to determine the relative occupancies (shown as percentages) and mean peak positions of the individual FRET species (black lines, sum of Gaussians; grey lines, individual Gaussians). First row: unliganded polymerase (Pol). Second row: binary complex, formed by adding DNA1 (with A as the templating base, final DNA concentration: 100 nM) to the polymerase (Pol + DNA). Third row: ternary complex with the correct nucleotide, formed by adding dTTP (1 mM) to the binary complex. Fourth row: ternary complexes with an incorrect nucleotide, formed by adding dGTP (1 mM) to the binary complex. Fifth row: ternary complex with the complementary ribonucleotide, formed by adding rUTP (1 mM) to the binary complex. The two dotted vertical lines mark the mean  $E^*$  values of the open and closed FRET species in the binary complex. Histograms for each enzyme were measured on the same day. The dashed line marks the mean  $E^*$  value of the intermediate-FRET species as populated in the ternary complexes (see also Supplementary Fig. S1).



**Figure 3 | Determining equilibrium constants for wt KF and E710A.** (a,b) Nucleotide dependence of FRET observables associated with the formation of ternary complexes of wt KF (a) and E710A (b) in the presence of DNA1. The fraction of the molecules in the high-FRET species (black circles) and the shift of the mean  $E^*$  of the low-FRET species (blue triangles) were plotted as a function of nucleotide concentration. The shift was normalized relative to the  $E^*$  difference between the means of the open and closed conformations. Data points with error bars are represented as mean  $\pm$  s.e.m., derived from three independent measurements (see Supplementary Fig. S2 for examples of titrations). The data were globally fitted using a four-state equilibrium model (c). See Table 1 for results. (c) Four-state model describing the nucleotide-concentration-dependent FRET observables (for details, see main text and Methods).  $K_0$  represents the equilibrium between open and closed forms of the binary complex (O: open, C: closed). Upon nucleotide (N) addition, the nucleotide-bound intermediate-FRET state (IN) and closed state (CN) are formed sequentially. The nucleotide-dependent equilibria are described by  $K_1$  and  $K_2$ , respectively. The dissociation constants  $K_{d1}$  and  $K_{d2}$  are calculated from  $K_1$  and  $K_2$ .

**Table 1 | Equilibrium constants for nucleotide binding.\***

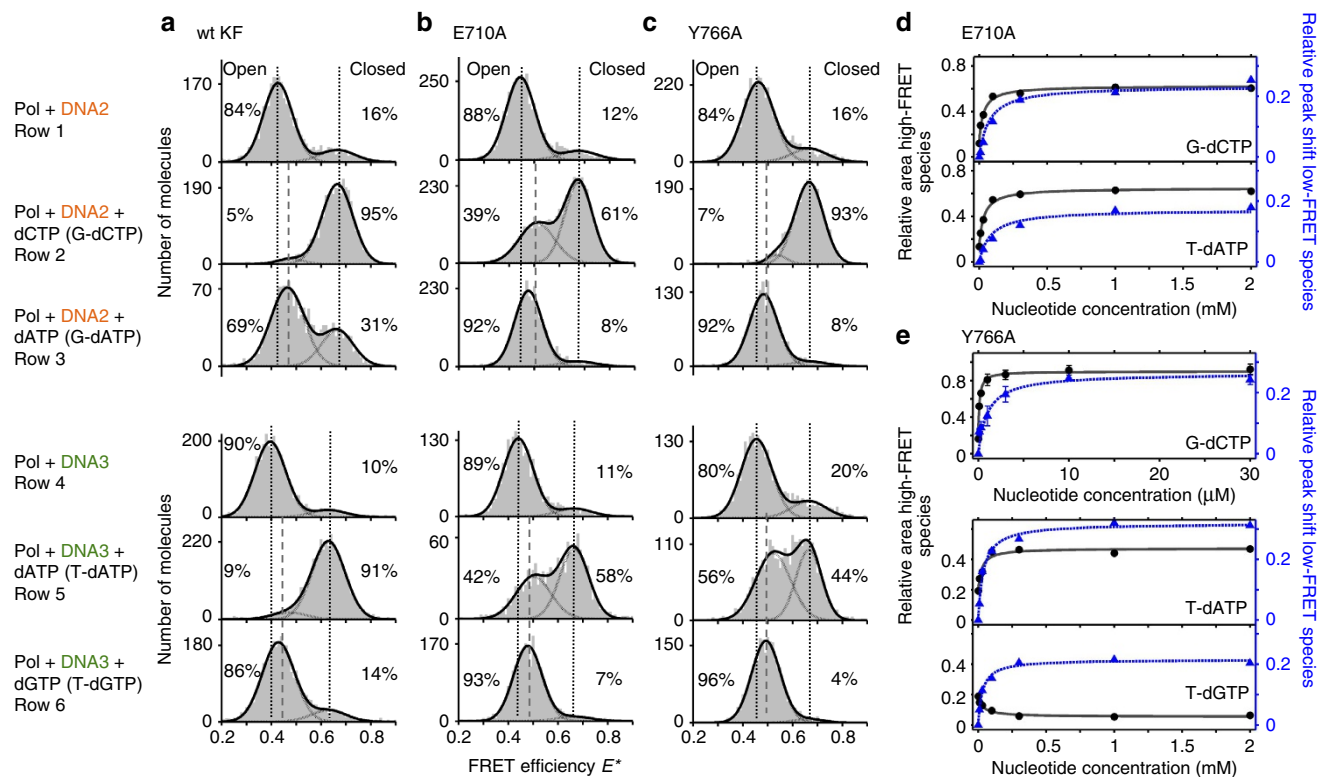
Complex	Constants <sup>†</sup>	wt KF	E710A	E710Q	Y766A	Y766F
A-dTTP <sup>‡</sup>	$K_{d1}$ ( $\mu$ M)	$5.1 \pm 0.6$	$43 \pm 10$	$120 \pm 27$	$80 \pm 30$	$18 \pm 4$
	$K_2$	$6.1 \pm 0.5$	$0.25 \pm 0.01$	$0.52 \pm 0.03$	$1.0 \pm 0.1$	$9.6 \pm 1.7$
	$K_{d2}$ ( $\mu$ M)	$0.83 \pm 0.12$	$170 \pm 40$	$230 \pm 50$	$79 \pm 28$	$1.9 \pm 0.6$
A-dGTP <sup>‡</sup>	$K_{d1}$ ( $\mu$ M)	$46 \pm 5$	$16 \pm 2$	$33 \pm 4$	$26 \pm 4$	$66 \pm 7$
	$K_2$	$0.091 \pm 0.005$	$0.084 \pm 0.06$	$0.061 \pm 0.005$	$0.059 \pm 0.008$	$0.087 \pm 0.007$
A-rUTP <sup>‡</sup>	$K_{d1}$ ( $\mu$ M)	$210 \pm 24$	$92 \pm 8$	$126 \pm 7$	$98 \pm 10$	$280 \pm 36$
	$K_2$	$0.083 \pm 0.007$	$0.16 \pm 0.00$	$0.056 \pm 0.003$	$0.059 \pm 0.007$	$0.073 \pm 0.013$
A-rGTP	$K_{d1}$ ( $\mu$ M)	$30 \pm 7$				
	$K_2$	$0.13 \pm 0.01$				
G-dATP	$K_{d1}$ ( $\mu$ M)	$230 \pm 67$			$90 \pm 54$	
	$K_2$	$0.55 \pm 0.04$			$0.076 \pm 0.030$	
G-dCTP	$K_{d1}$ ( $\mu$ M)	$1.2 \pm 0.4$	$70 \pm 11$		$1.0 \pm 0.2$	
	$K_2$	$6.1 \pm 1.2$	$1.7 \pm 0.1$		$8.9 \pm 1.2$	
	$K_{d2}$ ( $\mu$ M)	$0.20 \pm 0.07$	$42 \pm 7$		$0.11 \pm 0.02$	
G-dTTP	$K_{d1}$ ( $\mu$ M)	$230 \pm 76$			$49 \pm 30$	
	$K_2$	$0.10 \pm 0.02$			$0.22 \pm 0.04$	
T-dATP	$K_{d1}$ ( $\mu$ M)		$82 \pm 10$		$38 \pm 6$	
	$K_2$		$1.8 \pm 0.1$		$0.90 \pm 0.04$	
	$K_{d2}$ ( $\mu$ M)		$45 \pm 6$		$42 \pm 7$	
T-dGTP	$K_{d1}$ ( $\mu$ M)				$30 \pm 3$	
	$K_2$				$0.057 \pm 0.005$	
dGTP (no DNA)	$K_{d1}$ ( $\mu$ M)	$3.0 \pm 0.7$				
	$K_2$	$0.19 \pm 0.03$				
rGTP (no DNA)	$K_{d1}$ ( $\mu$ M)	$12 \pm 4$				
	$K_2$	$0.080 \pm 0.057$				

\*Determined using nucleotide titrations (Figs 3 and 4, Supplementary Figs S2 and S3) as described in Methods (data represented as mean  $\pm$  95% confidence interval representing the fitting error).

<sup>†</sup>See Fig. 3c for a description of the presented constants.

<sup>‡</sup>Calculated data for these base pairs were derived from three independent measurements.





**Figure 4 | Conformational profiles depend on the templating base.** (a–c) FRET histograms for binary and ternary complexes formed using wt KF (a), E710A (b) and Y766A (c). Conditions and style are as in Fig. 2, except that the DNA substrates were DNA2 (template G) or DNA3 (template T); for sequences, see Fig. 1d. (d,e) Nucleotide-concentration dependence of the conformational distribution of the E710A ternary complex with G-dCTP and T-dATP base pairs (d) and of the Y766A ternary complex with G-dCTP, T-dATP and T-dGTP base pairs (e). Data analysis was as in Fig. 3; the results are shown in Table 1. In e, the error bars on the G-dCTP plot represent the s.e. from two independent titrations; all other titrations represent single measurements.

correctly paired A-dTTP ternary complex; however, the subsequent transition to the closed complex is not favoured ( $K_2 < 0.2$ , Table 1).

The conformations of wt KF with complementary and mismatched dNTPs at a template G and a template T (Fig. 4a) were similar to those for template A (Fig. 2a).

#### E710 has a critical role in the fingers-closing transition.

The conformational profiles of unliganded E710A and its binary complex with DNA1 resemble those of wt KF (Fig. 2b, rows 1 and 2). However, addition of the complementary nucleotide (A-dTTP; Fig. 2b, row 3), even at high concentrations ( $\geq 1$  mM), led to only a modest increase in the occupancy of the closed state. Instead, the bulk of the population shifted from the open state ( $E^* \sim 0.44$ ) to an intermediate-FRET species ( $E^* \sim 0.48$ ). The  $K_{d1}$  for the E710A A-dTTP ternary complex was  $43 \mu\text{M}$  (Table 1 and Fig. 3b), similar to that for wt KF with a mismatch (A-dGTP), but several fold higher than that for wt KF with the complementary A-dTTP. Further, the  $K_2$  for the E710A A-dTTP ternary complex was  $< 1$ , similar to mismatches with wt KF (Table 1), preventing efficient formation of the closed conformation. The interactions of E710A with both a non-complementary dNTP (A-dGTP) and a complementary rNTP (A-rUTP) showed the same shift towards the  $E^* \sim 0.48$  state as with wt KF (Fig. 2b). With A as the templating base, E710A binds the incorrect dGTP 3-fold more tightly ( $K_{d1} = 16 \mu\text{M}$ ) than does wt KF ( $K_{d1} = 46 \mu\text{M}$ ); interestingly, this incorrect interaction is also threefold tighter than the binding of the correct dNTP (A-dTTP) by E710A. For the complementary rNTP (A-rUTP), both  $K_{d1}$  and

$K_2$  are  $\sim 2$ -fold more favourable for E710A than for wt KF, probably owing to easier accommodation of the ribose 2'-OH when the glutamate side chain is replaced by alanine.

The behaviour of E710A at template G and template T (Fig. 4b) was as described for template A, except in the presence of complementary dNTPs. Despite weak initial binding of the complementary dNTPs, the more favourable  $K_2$  values of  $\sim 2$  resulted in greater occupancy of the closed conformation (61% for G-dCTP and 58% for T-dATP at 1 mM dNTP); however, the overall affinity of E710A for all three complementary base pairs is low (Fig. 4d, Table 1), indicating that E710A discriminates poorly between correct and mismatched dNTPs regardless of templating base. The conformational profile for E710Q at template A (Fig. 2d) was similar to that of E710A (Table 1).

#### Y766A perturbs fingers-closing for complementary dNTPs.

With template bases A and T, the Y766A mutator followed the pattern seen for E710A and E710Q (Figs 2d and 4c). Addition of complementary nucleotides (A-dTTP or T-dATP) led to equal occupancy of the intermediate-FRET and high-FRET states at high nucleotide concentrations ( $\geq 1$  mM). Addition of incorrect nucleotides (A-dGTP, G-dATP, or T-dGTP) formed predominantly the intermediate-FRET species. At template A and T, all dNTPs showed weak binding regardless of their complementarity to the templating base (Table 1).

However, at template G, Y766A bound the complementary dCTP as strongly as wt KF ( $K_{d1} = 1.0 \mu\text{M}$  for Y766A versus  $1.2 \mu\text{M}$  for wt KF, Fig. 4e) and formed the closed complex almost

exclusively, reflecting a very favourable  $K_2$  (Table 1). Moreover, the single-turnover kinetic parameters for G-dCTP incorporation by Y766A and wt KF (Supplementary Table S2) are similar, in contrast to those for A-dTTP incorporation. Thus, at template G, Y766A distinguishes well between correct and incorrect dNTPs.

The complexes of Y766F, which has similar fidelity to wt KF<sup>11,14</sup>, resembled those of wt KF (Fig. 2e, Table 1), the only difference being a  $\sim 3$ -fold decreased binding affinity ( $K_{d1}$ ) for the complementary nucleotide (A-dTTP) and an increased  $K_2$  (9.6) so that the overall  $K_{d2}$  is  $\sim 2$ -fold weaker for Y766F.

### Y766 modulates the equilibrium in unliganded polymerases.

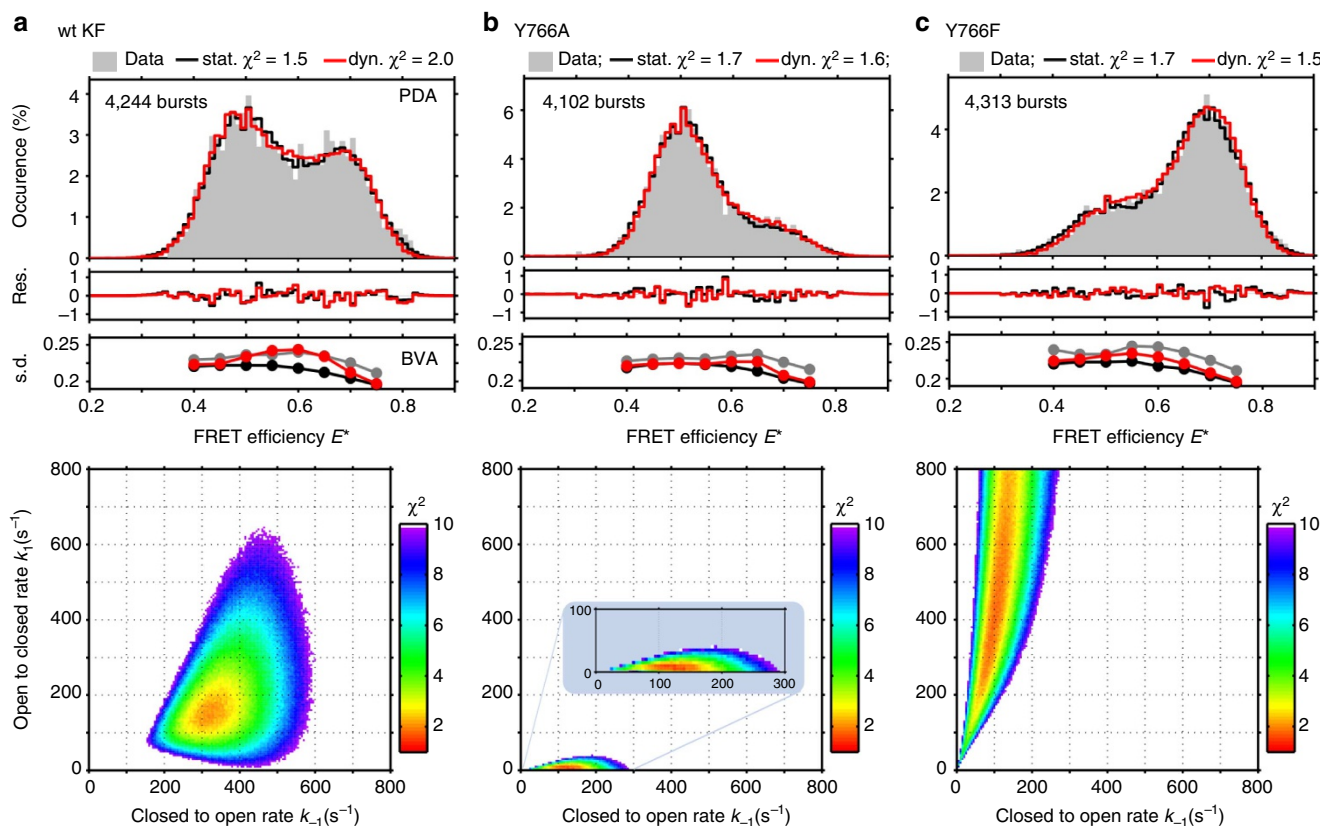
Whereas the unliganded wt KF, E710A and E710Q polymerases exhibit similar conformational equilibria, Y766 substitutions shift the equilibrium (Fig. 2d,e). Loss of the bulky phenyl ring from the Y766 side chain increases the open-state occupancy to 78% (Y766A), whereas the more conservative Y766F substitution, which removes a potential hydrogen-bond donor, decreases the open-state occupancy to 28%.

The large differences between wt KF, Y766A and Y766F indicate that the interconversion rates between open and closed states depend on residue 766. To characterize the equilibria, we used probability distribution analysis (PDA)<sup>8,17,18</sup> and burst variance analysis (BVA)<sup>7,10</sup>, both described in Methods. PDA

revealed that the FRET distribution fits well either to a ‘static’ model of two independent molecular entities or a ‘dynamic’ model<sup>19,20</sup> that invokes dynamic interconversions of a single molecular entity between two distinct FRET states (Fig. 5a, top).

We used BVA to show that the dynamic model prediction fitted the data significantly better than the static model (Fig. 5a), suggesting the presence of dynamic heterogeneity for all three KF derivatives. To further analyse the conformational dynamics, we performed PDA using the FRET efficiencies  $E^*_i$  obtained from the initial PDA fit (dynamic model), and plotted the  $\chi^2$ -values as a function of  $k_1$  and  $k_{-1}$  (Fig. 5a, bottom). For wt KF, the minimization surface forms a distinct well at  $k_1 = 160 \text{ s}^{-1}$  and  $k_{-1} = 320 \text{ s}^{-1}$  ( $\chi^2 < 2$ ), consistent with our previous reports<sup>7,8,10</sup>. In contrast, both Y766A and Y766F display different landscapes (Figs. 5b,c), whereas Y766A shows a distinct well at  $k_1 = 15 \text{ s}^{-1}$  and  $k_{-1} = 120 \text{ s}^{-1}$ , Y766F displays a broad range of possible rate combinations with a shallow minimum at  $k_1 \sim 450 \text{ s}^{-1}$  and  $k_{-1} \sim 120 \text{ s}^{-1}$ .

**Accessing the equilibrium dynamics of wt KF complexes.** To characterize the timescales of interconversions within the sequential four-state model (Fig. 3c), we used PDA and a novel analysis of FRET distribution widths. Using PDA, we characterized the interconversions between O and C for the wt KF binary complex, and



**Figure 5 | Conformational dynamics of unliganded wt KF and Y766 derivatives.** (a) Analysis of wt KF using PDA (upper panel), BVA (middle panel) and a rate-dependent PDA  $\chi^2$ -minimization surface (lower panel). The  $E^*$  histogram was fitted using PDA, assuming either a static (black line) or a dynamic (red line) two-species model, wherein the polymerase occupies either the closed or open state yielding  $k_1 = 150 \text{ s}^{-1}$ , and  $k_{-1} = 320 \text{ s}^{-1}$  for the dynamic model (see Supplementary Methods for fitting parameters). Whereas PDA achieved satisfactory fits to the data in either scenario, as shown by the residuals plot ( $\chi^2 \leq 2$ ), the BVA data (grey line) favoured the dynamic model (red line). The PDA  $\chi^2$ -minimization surface uses the mean FRET positions and excess widths from the dynamic fit, plotting the  $\chi^2 < 10$  values as a function of the rates. Best fits are obtained for  $k_1 \sim 160 \text{ s}^{-1}$  and  $k_{-1} \sim 320 \text{ s}^{-1}$ . (b,c) Analysis of conformational dynamics for Y766A and Y766F, as in a, yielding interconversion rates of  $k_1 = 9 \text{ s}^{-1}$ ,  $k_{-1} = 128 \text{ s}^{-1}$  for Y766A and  $k_1 = 435 \text{ s}^{-1}$ , and  $k_{-1} = 108 \text{ s}^{-1}$  for Y766F. For fitting parameters, see Supplementary Methods; for PDA of the wt KF binary complex and the wt KF ternary complex with an incorrect nucleotide, see Supplementary Fig. S4.

between IN and CN for the wt KF ternary complex with an A-dGTP mispair at saturating nucleotide conditions (Supplementary Fig. S4); this analysis yielded rate constants of  $k_{O \rightarrow C} = 15 \text{ s}^{-1}$  and  $k_{C \rightarrow O} = 85 \text{ s}^{-1}$  for the binary complex and  $k_{IN \rightarrow CN} = 3 \text{ s}^{-1}$  and  $k_{CN \rightarrow IN} = 45 \text{ s}^{-1}$  for the mispaired ternary complex.

Assessing the remaining rates (corresponding to rate constants  $k_{O \rightarrow IN}$  and  $k_{IN \rightarrow O}$ ) is challenging, as the analysis of the low-FRET species is hampered by our inability to resolve the FRET distributions of the O and IN states. Width analysis (see Methods) of our data (Supplementary Fig. S5c) revealed broadening of the low-FRET species around  $K_{d1}$  for several nucleotide titrations (for wt KF: A-dGTP and A-rUTP; for E710A: A-dTTP, A-dGTP and A-rUTP), indicating that the interconversions are at the millisecond timescale or slower. For the E710A A-rUTP titrations, for example, there is a maximum broadening of  $\sim 6\%$  at  $100 \mu\text{M}$  rUTP, close to the corresponding  $K_{d1}$  of  $92 \mu\text{M}$ . For a better estimate of the timescale, we simulated the four-state model using four fixed rates (by fixing  $k_{O \rightarrow C}$ ,  $k_{C \rightarrow O}$ ,  $k_{IN \rightarrow CN}$  and  $k_{CN \rightarrow IN}$ ) and varying  $k_{O \rightarrow IN}$  and  $k_{IN \rightarrow O}$  (Supplementary Fig. S5d; as in Santoso *et al.*<sup>7</sup>). The width of the low-FRET species at a nucleotide concentration equal to  $K_{d1}$  remains constant for high IN-to-O rates ( $\geq 300 \text{ s}^{-1}$ ), whereas for rates of  $< 300 \text{ s}^{-1}$  there is a significant width increase around  $K_{d1}$ . We thus propose a rate of  $300 \text{ s}^{-1}$  as an upper limit for the IN-to-O transition.

## Discussion

Previously, we described a FRET state between the open and closed states for wt KF ternary complexes with incorrect nucleotides<sup>7</sup>. Data presented here strongly suggest that wt KF accesses this intermediate state even with complementary dNTPs, consistent with this state being on the path to nucleotide incorporation. Whereas the intermediate-FRET state is sparsely populated in the ternary complex of wt KF with a complementary dNTP, the three mutators (Y766A, E710A and E710Q) behave differently; their ternary complexes with complementary dNTPs occupy mainly the intermediate-FRET state, with Y766A and the G-dCTP base pair being the only exception.

Initially, we attempted to relate the intermediate-FRET species to available cocrystal structures, speculating that the ternary complex with a complementary rNTP undergoes partial fingers closure, whereas the mispaired ternary complex resembles a fingers-open conformation<sup>7</sup>. Since then, a better structural candidate for the intermediate-FRET species has been described<sup>21</sup>. Using *Bacillus stearothermophilus* DNA polymerase (a KF homologue), Wu and Beese<sup>21</sup> captured a partially closed ('ajar') fingers conformation (Fig. 6a) in a ternary complex with a G-dTTP mismatch, using both a mutator (Y714S, equivalent to Y766S in KF) and a non-mutator (F710Y, equivalent to F762Y in KF) polymerase; the partially closed state corresponds to fingers-closing along the open-to-closed transition, leading to a distance change of  $\sim 2 \text{ \AA}$  between the sites carrying the FRET probes, in excellent agreement with our observations.

Evidence for a partially closed state was also reported in a recent FRET study on immobilized KF complexes<sup>22</sup>. In contrast to our assay, which reports directly on conformational states by monitoring an intra-protein distance, Berezhna *et al.*<sup>22</sup> used donor-labelled DNA and acceptor-labelled polymerases to infer a protein conformational change by measuring a DNA-protein distance; in the latter assay, any protein conformational changes may be superimposed on concurrent DNA conformational changes, as well as on possible polymerase repositioning relative to the DNA.

The partially closed cocrystal structures may also relate to other intermediate-FRET species we have observed. By introducing a mutation that prevents full closing of the O-helix, Wu and Beese<sup>21</sup>

captured a partially closed *B. stearothermophilus* ternary complex with a correctly paired dNTP, a candidate for the sparsely populated species of intermediate-FRET value in the ternary complexes of wt KF with correct dNTPs. More speculatively, we note that the Pol-dNTP binary complex (without DNA) has a similar intermediate-FRET signature (Supplementary Fig. S2e; see also Santoso *et al.*<sup>7</sup>), suggesting that O-helix movement might be initiated by dNTP binding in the KF cleft without direct DNA involvement. Our  $K_d$  measurements (Table 1) indicate that nucleotide binding is destabilized by the presence of DNA with a non-complementary template base, just as DNA binding is destabilized by a mismatched dNTP (O.B., N.D.F.G., C.M.J., unpublished)<sup>23</sup>; both cases promote dissociation of mispaired ternary complexes.

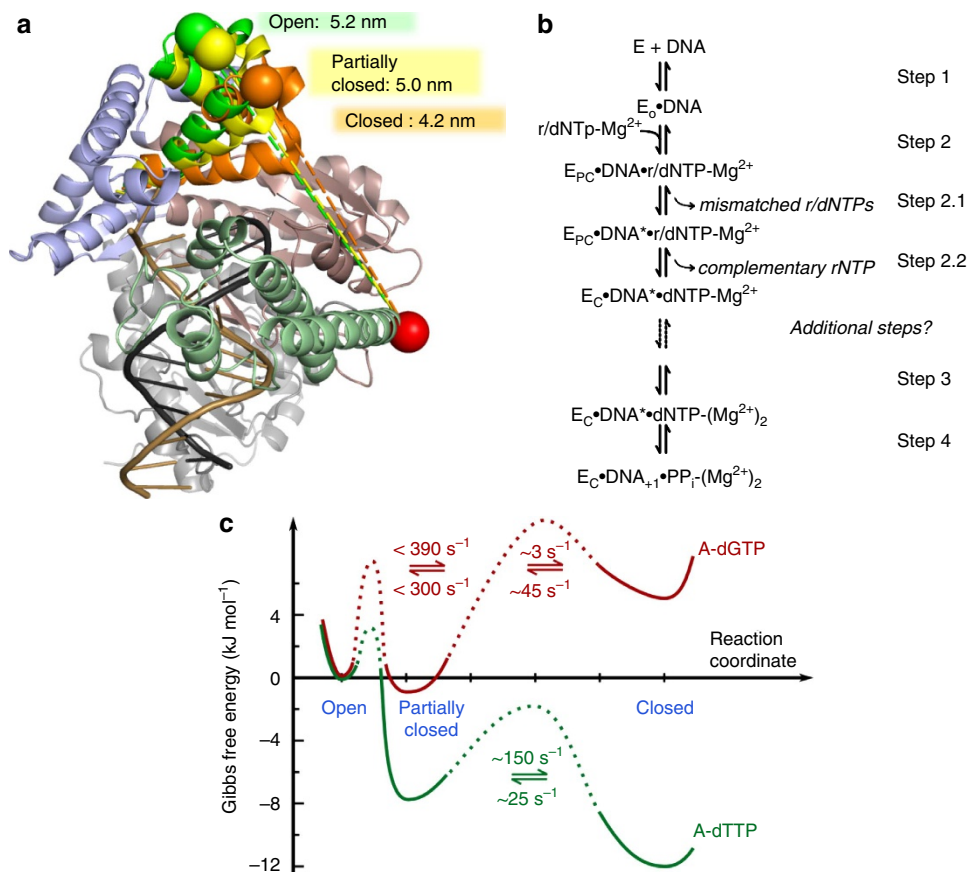
It seems reasonable to assume that the partially closed complex is on the pathway to the chemical step, given that the partially closed state is structurally part way between open and closed states. We cannot directly rule out the alternative scenario whereby the partially closed conformation is an 'off-pathway' state in equilibrium with the open but not the closed state as, in either case, the equilibria follow similar relations (Supplementary Methods).

The partially closed conformation has the hallmarks of an early checkpoint, being populated in all wt KF ternary complex species, although these species progress differently along the reaction pathway. This conformation must be formed at an early step, most probably immediately after the Pol-DNA binary complex binds a nucleotide and its  $\text{Mg}^{2+}$  ion (step 2, Fig. 6b). Although our FRET pair does not distinguish between mispaired and rNTP-containing ternary complexes, experiments using 2-aminopurine on DNA substrates show that wt KF complexes with a complementary rNTP proceed further than mispairs (O.B., N.D.F.G., C.M.J., unpublished)<sup>23</sup>. A nanopore-based study of single KF-DNA complexes also supports this conclusion<sup>24</sup>. The partially closed conformation, therefore, encompasses at least two related ternary complexes (Fig. 6b) whose equilibria determine the progression of a particular ternary complex towards product formation. First, the partially closed ternary complex undergoes a base-checking step (step 2.1) in which the equilibrium is unfavourable for all non-complementary nucleotides (dNTPs and rNTPs), thus rejecting six out of the eight nucleoside triphosphates. Next, the sugar structure is tested against the active-site pocket (step 2.2), rejecting the bulkier rNTPs and allowing complexes containing complementary dNTPs to proceed to the closed conformation, from which they undergo subsequent steps, including D705 reorientation and binding of the second  $\text{Mg}^{2+}$  ion<sup>25</sup>.

The calculated equilibrium constants (Table 1) enabled us to propose a free-energy landscape for the prechemistry nucleotide selection steps (Fig. 6c). Although a simplified landscape has been proposed before (on the basis of fluorescence enhancement of a dye-labelled DNA upon T7 DNA polymerase binding<sup>26</sup>), our ability to monitor the conformational coordinate directly allowed us to add the important partially closed conformation checkpoint.

The dissociation constant for a nucleotide in the partially closed ternary complex,  $K_{d1}$ , reports on interactions in the wt KF active site. The binding of the complementary dNTP is favoured, leading to a stabilized partially closed state ( $-8 \text{ kJ mol}^{-1}$  for A-dTTP versus  $-0.7 \text{ kJ mol}^{-1}$  for A-dGTP, using the nucleotide concentrations in *E. coli*<sup>27</sup>); this may result from tighter binding within either or both of  $\text{E}_{\text{PC}} \bullet \text{DNA} \bullet \text{r/dNTP} \bullet \text{Mg}^{2+}$  and  $\text{E}_{\text{PC}} \bullet \text{DNA} \bullet \text{r/dNTP} \bullet \text{Mg}^{2+}$  (see Fig. 6b). Moreover, the similar  $K_{d1}$  values for A-dGTP and A-rGTP complexes (Table 1) indicate that the ribose 2'-OH is not detected in the initial  $\text{E}_{\text{PC}} \bullet \text{DNA} \bullet \text{r/dNTP} \bullet \text{Mg}^{2+}$  complex.





**Figure 6 | Roles of the partially closed state in nucleotide selection.** (a) Proposed structure of the polymerase conformation corresponding to the intermediate-FRET state. The crystal structure of the Pol-DNA binary complex (PDB accession code 1L3U, mobile section green)<sup>4</sup> is superimposed on the crystal structures of a ternary complex where the fingers adopt a partially closed ('ajar') conformation (PDB accession code 3HPO, mobile section yellow)<sup>21</sup> and a ternary complex where the fingers adopt the closed conformation (PDB accession code 1LV5, mobile section orange)<sup>4</sup>. The distances between the C<sub>β</sub> positions of residue 550 (red sphere on the thumb) and 744 (sphere on the fingers) are 5.2 nm in the open, 5.0 nm in the partially closed and 4.2 nm in the closed conformation. The predicted distance change of ~2 Å on going from the open to the partially closed conformation is consistent with the observed shift of  $\Delta E^* \sim 0.04$ , as we operate in the linear part of the FRET dynamic range. (b) Proposed reaction pathway for KF from DNA binding to phosphoryl transfer, based on references<sup>7,21,23</sup> and our current results. E<sub>o</sub> and E<sub>C</sub> represent the open and the closed conformations, respectively, whereas E<sub>PC</sub> represents the intermediate-FRET state interpreted as a partially closed conformation. (c) Proposed free-energy landscapes for the prechemistry nucleotide selection steps prepared using the equilibrium constants in Table 1 and *in vivo* dNTP concentrations<sup>27</sup> (green curve: complementary nucleotide A-dTTP; red curve: non-complementary nucleotide A-dGTP; see also Methods). Reaction coordinate: degree of fingers-closing (shown to scale). The free energy of the open binary state is set to zero. Binding of a nucleotide stabilizes the complex with the partially closed conformation, either weakly (incorrect base pair A-dGTP) or strongly (correct base pair A-dTTP). Equilibration to the fully closed conformation is favoured only for the correct nucleotide. Although the exact activation energy barriers (dotted lines) between the conformations are unknown, we included the rates for the transition of the A-dGTP ternary complex between the open and partially closed states (estimated using FRET-width analysis; Supplementary Fig. S5) and the rates for the transitions between the partially closed and the closed states for correct and incorrect nucleotides (determined using PDA; Supplementary Figs S1c and S4, respectively).

$K_2$ , describing the equilibrium between partially closed and closed complexes, defines the free energy of the closed ternary complex. When  $K_2$  is  $>1$ , as for correctly paired ternary complexes of wt KF, the closed complex is the lowest-energy species ( $-4 \text{ kJ mol}^{-1}$  relative to the A-dTTP partially closed state) on the prechemistry pathway, and its enhanced stability relative to the partially closed ternary complex ( $K_{d2} < K_{d1}$ ) drives the forward reaction. With an inappropriate substrate, such as a mispair,  $K_2$  is  $<1$  so that the closed complex is destabilized relative to the partially closed ternary complex ( $+6 \text{ kJ mol}^{-1}$  for A-dGTP) and the equilibrium favours species formed early in the pathway. The extremely slow rates for misinsertion ( $\geq 10^3$ -fold slower than complementary dNTP additions) probably result from low occupancy of the closed conformation combined with active-site distortions<sup>13,14</sup>.

The fidelity mutants in this study show a reduced ability to recognize complementary dNTPs as 'correct' (Table 1), having lower affinity for the initial binding of the complementary dNTP to the partially closed complex and an unfavourable  $K_2$ . Therefore, the partially closed conformation is highly populated in the resulting ternary complexes so that, in the absence of the E710 or Y766 side chains, the complex with the complementary dNTP resembles a mispaired complex. Our data show a continuum of behaviours, with both the templating base and the particular mutator allele influencing how far the conformational equilibria shift towards the closed state; at one extreme is the E710A A-dTTP complex, which has little of the closed conformation ( $K_2 = 0.25$ ), at the other is the Y766A G-dCTP complex, which resembles wt KF ( $K_2 = 8.9$ ); in between are several complexes with  $K_2 \sim 1$ , having substantial amounts of both closed and partially closed complexes.



Why should alteration of E710 or Y766 destabilize the closed ternary complex? Structurally, E710 and Y766 define much of the surface of the active site-binding pocket so that they interact more closely with the nascent base pair in the closed complex than in the partially closed complex, where the base pair is more solvent exposed. Surprisingly, the hydrogen bond between E710 and Y766, despite being invariant in A-family DNA polymerases, has a minor role in the closed complex stability, as shown by the similar FRET profile for complexes of Y766F and wt KF. More important may be the overall size and shape of the 766 side chain and its ability to interact with and position the nearby F762 side chain, which forms part of the binding pocket.

Structural studies argue that the transition from partially closed to fully closed conformation is required for catalysis of the polymerase reaction. In the partially closed conformation, the primer 3'-OH is too far from the  $\alpha$ -phosphate of the bound nucleotide; moreover, the rest of the active site, including the catalytic metal ions, is not properly assembled for a chemical attack<sup>21</sup>. Our data support this reasoning, showing a rough correlation between the rate of nucleotide incorporation ( $k_{\text{pol}}$ , Supplementary Table S2) and closed-complex occupancy; the most notable example is G-dCTP incorporation by Y766A, where high occupancy of the closed complex correlates with near-wild-type kinetics. Moreover, for the mutator proteins, the single-turnover  $K_d$  for A-dTTP incorporation ( $K_{d(\text{dNTP})}$ , Supplementary Table S2) approximates to  $K_{d2}$  (Table 1), consistent with closed-complex formation being required for the chemical step. (This assumes that the chemical step is rate limiting for these mutant proteins. For wt KF (and probably Y766F), the same relationship does not hold, as prechemistry step 3 is rate limiting)<sup>28</sup>. Nevertheless, as the low incorporation rate for E710Q shows (Supplementary Table S2), achieving the closed state with moderate efficiency is, of itself, no guarantee of effective catalysis if the mutation distorts the final active-site geometry.

Our data support the generalization that alterations in polymerase fidelity correlate with the kinetics of handling correct, rather than incorrect, nucleotides<sup>29</sup>. Compared with wt KF, the mutator variants show at most a threefold enhancement in the binding ( $K_{d1}$ ) of the non-complementary A-dGTP and no improvement in  $K_2$ . This contrasts with the much larger deficits in both constants for the mutators when binding the correct A-dTTP. As the closed complex selects for the correct template-dNTP pairing, any change to the energetics of this complex will exclusively affect correct dNTP incorporation. Moreover, even in the partially closed conformation, changes in active-site side chains diminish specificity for the correct pairing more than they enhance the interaction with mismatches.

The effects of template base and mutator allele on closed-complex energetics provide a partial explanation of the observed error specificities of the KF mutators (Supplementary Fig. S6)<sup>11–14</sup>. Closed complexes with the correct dNTP are least populated when E710A encounters template A or when Y766A encounters template A or T, making these the situations where recognition of the correct base pair is least effective. One would therefore expect E710A to be less accurate at template A than that at G or T, and Y766A to be less accurate at templates A and T, relative to G. This is consistent with the high frequency of A-dCTP errors made by E710A and T-dGTP made by Y766A but does not explain the lower frequency of other errors at these particular template bases. The substantial frequency of Y766A G-dTTP errors is noteworthy, given that Y766A handles the G-dCTP pair with almost wild-type efficiency; this could be a case where a mutator actually facilitates the interaction with the mismatch, consistent with earlier kinetic data<sup>14</sup>. As the experimentally observed polymerase errors in

Supplementary Fig. S6 require not only covalent addition of mismatched bases but also extension of the resulting mismatched primer termini, they will be influenced by steric or geometric constraints in intermediates for both reactions so that the overall error specificity reflects the combination of several factors.

Finally, our results on the unliganded Y766 polymerase derivatives reinforce our previous observations that the conformational equilibrium of the unliganded polymerase is delicate; thus, the removal of even a single oxygen atom (as in Y766F) increased substantially the occupancy of the closed state for the unliganded polymerase, whereas the more drastic Y766A mutation had the opposite effect. By contrast, changes to the 710 side chain did not influence the conformational equilibrium, suggesting that the behaviour of Y766 derivatives is not directly related to the missing 710–766 hydrogen bond in the closed conformation. An alternative explanation considers, instead, the open conformation, where the 766 side chain is extremely solvent exposed; one might expect the side chains to favour the open conformation in the reverse order of their hydrophobicity: A > Y > F, as is observed. Unlike the unliganded polymerase, the FRET distributions of binary complexes of the mutant and wt polymerases were similar, indicating that the conformational landscape is dominated by Pol-DNA interactions, and that the apparent conformational equilibrium of unliganded polymerases is not a reliable predictor of the conformational landscape of the corresponding binary complexes.

## Methods

**Single-molecule FRET measurements.** Single-molecule FRET experiments were performed at room temperature ( $\approx 22^\circ\text{C}$ ) using a home-built confocal microscope equipped with 20 kHz alternating laser excitation between a 532-nm (operated at 240  $\mu\text{W}$ ) and a 638-nm laser (operated at 60  $\mu\text{W}$ ), coupled to a  $\times 60$ , 1.35 NA, UPLSAPO 60XO (Olympus) objective. The alternating laser excitation allows filtering for correctly labelled species bearing an active donor and an active acceptor<sup>30</sup>. After filtering each fluorescent burst for the correct labelling stoichiometry, we calculated the apparent FRET efficiency  $E^*$  for each burst as  $E^* = \text{DA}/(\text{DD} + \text{DA})$ , where DA is the number of photons in the red detection channel after green excitation and DD the number of photons in green detection channel after green excitation. Doubly labelled KF derivatives were present at  $< 100 \text{ pM}$  in single-molecule observation buffer (40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)-NaOH, pH 7.3, 10 mM  $\text{MgCl}_2$ , 1 mM dithiothreitol, 100  $\mu\text{g ml}^{-1}$  bovine serum albumin, 5% (vol/vol) glycerol, 1 mM mercaptoethylamine).

**Analysis of FRET histograms.** Single-molecule FRET experiments are sensitive to subtle changes in conditions such as room temperature and set-up alignment<sup>30,31</sup>. To minimize changes, every set of experiments contains reference measurements of a DNA FRET standard with a reference FRET value of  $E^* \sim 0.5$  (ref. 32), the unliganded polymerase, the binary complex formed with DNA and the ternary complex with DNA and nucleotides. Each titration experiment was completed in a single day.

The stoichiometry-filtered  $E^*$  FRET histograms were fitted to double-Gaussian functions that were constrained by a maximum width of  $\sigma_{\text{max}}(E^*) \leq 0.07$ . The double Gaussians report on peak positions and relative occupancies. For titration experiments, we plotted the relative peak shift  $(E^*_{\text{obs}} - E^*_{\text{open}})/(E^*_{\text{closed}} - E^*_{\text{open}})$  of the low-FRET Gaussian and the fractional occupancy of the high-FRET Gaussian as a function of nucleotide concentration. We fixed the position of the high-FRET Gaussians to the value obtained from a free fit of the binary complex or a saturated ternary complex with complementary dNTPs. We did not fix the position of the low-FRET Gaussian.

**Derivation of the multistate equilibrium model.** To determine the parameters governing the equilibria between the species observed in the single-molecule FRET histograms (Fig. 3c), we derived equations for the two concentration-dependent FRET observables: the peak position of the low-FRET Gaussian and the fraction of the population in the high-FRET Gaussian. The binary complex is in dynamic equilibrium between open (O, low-FRET) and closed (C, high-FRET) conformations, described by the equilibrium constant  $K_0 = [\text{C}]/[\text{O}]$ . Binding of a nucleotide (N) to the open conformation, forming the intermediate-FRET ternary complex (IN), is described by the association constant  $K_1 = [\text{IN}]/[\text{O}][\text{N}]$ . The intermediate-FRET ternary complex can undergo an additional conformational

transition to the closed ternary complex (CN, high-FRET), governed by the equilibrium constant  $K_2 = [\text{CN}]/[\text{IN}]$ . The corresponding dissociation constants are defined as  $K_{d1} = 1/K_1$  for the formation of the intermediate-FRET state and  $K_{d2} = 1/(K_1K_2)$  for the formation of the high-FRET (fully closed) state. The total polymerase concentration is

$$\begin{aligned} [\text{Total}] &= [\text{O}] + [\text{C}] + [\text{IN}] + [\text{CN}] \\ &= [\text{O}] + K_0[\text{O}] + K_1[\text{O}][\text{N}] + K_1K_2[\text{O}][\text{N}]. \end{aligned} \quad (1)$$

The fractions of each species as a function of nucleotide concentration are:

$$[\text{O}]/[\text{Total}] = 1/(1 + K_0 + K_1[\text{N}] + K_1K_2[\text{N}]), \quad (2)$$

$$[\text{C}]/[\text{Total}] = K_0/(1 + K_0 + K_1[\text{N}] + K_1K_2[\text{N}]), \quad (3)$$

$$[\text{IN}]/[\text{Total}] = K_1[\text{N}]/(1 + K_0 + K_1[\text{N}] + K_1K_2[\text{N}]), \quad \text{and} \quad (4)$$

$$[\text{CN}]/[\text{Total}] = K_1K_2[\text{N}]/(1 + K_0 + K_1[\text{N}] + K_1K_2[\text{N}]). \quad (5)$$

The two FRET observables are given (a) by the normalized peak shift of the low-FRET Gaussian

$$\begin{aligned} (E_{\text{obs}}^* - E_{\text{open}}^*)/(E_{\text{closed}}^* - E_{\text{open}}^*) &= y_{\text{max}}[\text{IN}]/([\text{O}] + [\text{IN}]) \\ &= y_{\text{max}}K_1[\text{N}]/(1 + K_1[\text{N}]) \\ &= y_{\text{max}}[\text{N}]/(K_{d1} + [\text{N}]), \end{aligned} \quad (6)$$

where  $y_{\text{max}}$  corresponds to the normalized peak shift at saturating nucleotide concentration and (b) by the fraction of high-FRET species

$$([\text{C}] + [\text{CN}])/[\text{Total}] = (K_0 + K_1K_2[\text{N}])/(1 + K_0 + K_1[\text{N}] + K_1K_2[\text{N}]). \quad (7)$$

We used a global fitting approach to fit both FRET observables simultaneously as a function of nucleotide concentration, and to determine the equilibrium constants  $K_0$ ,  $K_1$  and  $K_2$ , and the parameter  $y_{\text{max}}$ . From the equilibrium constants and their s.e., we calculated the dissociation constants. Identical equations can be derived to describe nucleotide binding to the polymerase in the absence of DNA; in this case, O and C refer to the unliganded protein.

**Probability distribution analysis and burst variance analysis.** For PDA<sup>8</sup>, the static model is fully described by two peak positions  $E_i^*$ , two excess widths  $\sigma_i$  (accounting for additional broadening of  $\langle E_i^* \rangle$  due to acceptor photophysics<sup>17,19</sup>) and two relative occupancies  $N_i$ . In comparison, the dynamic model replaces the relative occupancies with two rate constants:  $k_1$  for the open-to-closed transition and  $k_{-1}$  for the closed-to-open transition. We optimized parameters for both models using a least mean square algorithm implemented in MATLAB. BVA directly evaluates FRET distributions for the presence of dynamic heterogeneity<sup>10</sup>. In BVA, the experimentally measured s.d. of the FRET efficiency for individual transits through the confocal spot is compared with the s.d. predicted by the static and dynamic models.

**Width analysis for obtaining rate constants.** At nucleotide concentrations equal to the dissociation constant  $K_{d1}$  (where  $[\text{O}]$  equals  $[\text{IN}]$ ), the interconversion rates between O and IN,  $k_{\text{O} \rightarrow \text{IN}}$  and  $k_{\text{IN} \rightarrow \text{O}}$ , are equal. These rates also control the width of the low-FRET species distribution: if  $k_{\text{O} \rightarrow \text{IN}}$  and  $k_{\text{IN} \rightarrow \text{O}}$  are high and lead to multiple interconversions during the transit of the complex through the detection volume ( $\sim 3$  ms (ref. 7)), the FRET averaging between O and IN will yield a FRET width for the low-FRET species that matches that for the binary or partially closed ternary complex for all nucleotide concentrations. In contrast, slower interconversions will lead to a FRET-width increase; this increase should peak around  $K_{d1}$  (Supplementary Fig. S5a,b).

**Calculating the free-energy landscape of ternary complexes.** To construct the free-energy landscape for the prechemistry conformations of wild-type DNA polymerase in the presence of nucleotides, we calculated the relative Gibbs free energies  $\Delta_r G$  for the partially closed conformation and the fully closed conformation in ternary complexes by using the open conformation in the binary complex as a reference. To estimate  $\Delta_r G_{(\text{O}/\text{IN})}$ , the free-energy difference between the open conformation and the partially closed conformation, for a complementary dNTP (A-dTTP), we used:  $\Delta_r G = \Delta_r G^0 + RT \ln(Q_r)$ . The first term,  $\Delta_r G^0$ , is the free energy corresponding an O-to-IN equilibrium with 1 M concentrations for all species involved and is calculated using  $\Delta_r G^0 = -RT \ln(K_1)$ , where  $R$  is the ideal gas constant,  $T$  the temperature and  $K_1$  is the equilibrium association constant measured using our nucleotide titration experiments (Table 1). The second term requires the reaction quotient  $Q_r$ , which is a function of the actual concentrations of all species involved in the equilibrium:  $Q_r = \{[\text{IN}]/[\text{O}][\text{N}]\}$ . To estimate  $Q_r$ , we set  $[\text{N}]$  to the *E. coli* intracellular concentration for dTTP (120  $\mu\text{M}$ )<sup>27</sup>, and assumed that  $[\text{IN}] = [\text{O}]$ . To estimate  $\Delta_r G_{(\text{IN}/\text{CN})}$ , the free-energy difference between the partially closed conformation and the closed conformation with bound nucleotide, we used  $\Delta_r G_{(\text{IN}/\text{CN})} = -RT \ln(K_2)$ . Similar expressions were used for the free energies of the ternary complexes formed upon binding to an incorrect dNTP (A-dGTP); for dGTP, we used  $[\text{N}] = 60 \mu\text{M}$ <sup>27</sup>. Using the proposed rate of  $\sim 300 \text{ s}^{-1}$  as an upper limit for the rate of the IN-to-O transition for A-dGTP, we

calculate the rate of the O-to-IN transition at 60  $\mu\text{M}$  dGTP in *E. coli* as follows: rate =  $K_1 \bullet k_{\text{IN} \rightarrow \text{O}} \bullet 60 \mu\text{M} = 21.7 \bullet 10^3 \text{ M}^{-1} \bullet 300 \text{ s}^{-1} \bullet 60 \bullet 10^{-6} \text{ M} = 390 \text{ s}^{-1}$ .

The following procedures are further described in the Supplementary Methods: DNA preparation, protein expression and purification, and site-specific labelling of KF.

## References

- Kunkel, T. A. & Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **69**, 497–529 (2000).
- Joyce, C. M. & Benkovic, S. J. DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* **43**, 14317–14324 (2004).
- Doublé, S., Sawaya, M. R. & Ellenberger, T. An open and closed case for all polymerases. *Structure* **7**, R31–R35 (1999).
- Johnson, S. J., Taylor, J. S. & Beese, L. S. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc. Natl Acad. Sci. USA* **100**, 3895–3900 (2003).
- Kool, E. T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71**, 191–219 (2002).
- Li, Y., Korolev, S. & Waksman, G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514–7525 (1998).
- Santoso, Y. *et al.* Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl Acad. Sci. USA* **107**, 715–720 (2010).
- Santoso, Y., Torella, J. P. & Kapanidis, A. N. Characterizing single-molecule FRET dynamics with probability distribution analysis. *Chem. Phys. Chem.* **11**, 2209–2219 (2010).
- Santoso, Y. & Kapanidis, A. N. Probing biomolecular structures and dynamics of single molecules using in-gel alternating-laser excitation. *Anal. Chem.* **81**, 9561–9570 (2009).
- Torella, J. P., Holden, S. J., Santoso, Y., Hohlbein, J. & Kapanidis, A. N. Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophys. J.* **100**, 1568–1577 (2011).
- Bell, J. B., Eckert, K. A., Joyce, C. M. & Kunkel, T. A. Base miscoding and strand misalignment errors by mutator Klenow polymerases with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. *J. Biol. Chem.* **272**, 7345–7351 (1997).
- Minnick, D. T. *et al.* Side chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **274**, 3067–3075 (1999).
- Minnick, D. T., Liu, L., Grindley, N. D. F., Kunkel, T. A. & Joyce, C. M. Discrimination against purine-pyrimidine mismatches in the polymerase active site of DNA polymerase I: a structural explanation. *Proc. Natl Acad. Sci. USA* **99**, 1194–1199 (2002).
- Carroll, S. S., Cowart, M. & Benkovic, S. J. A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity. *Biochemistry* **30**, 804–813 (1991).
- Astatke, M., Grindley, N. D. F. & Joyce, C. M. How *E. coli* DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides. *J. Mol. Biol.* **278**, 147–165 (1998).
- Astatke, M., Ng, K. M., Grindley, N. D. F. & Joyce, C. M. A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl Acad. Sci. USA* **95**, 3402–3407 (1998).
- Kalinin, S., Felekyan, S., Antonik, M. & Seidel, C. A. M. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *J. Phys. Chem. B* **111**, 10253–10262 (2007).
- Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C. A. M. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970–6978 (2006).
- Nir, E. *et al.* Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B* **110**, 22103–22124 (2006).
- Kalinin, S., Valeri, A., Antonik, M., Felekyan, S. & Seidel, C. A. M. Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J. Phys. Chem. B* **114**, 7983–7995 (2010).
- Wu, E. Y. & Beese, L. S. The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an ‘ajar’ intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* **286**, 19758–19767 (2011).
- Berezhna, S. Y., Gill, J. P., Lamichhane, R. & Millar, D. P. Single-molecule Förster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *J. Am. Chem. Soc.* **134**, 11261–11268 (2012).
- Joyce, C. M. *et al.* Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry* **47**, 6103–6116 (2008).
- Garalde, D. R. *et al.* Distinct complexes of DNA polymerase I (Klenow fragment) for base and sugar discrimination during nucleotide substrate selection. *J. Biol. Chem.* **286**, 14480–14492 (2011).
- Bermek, O., Grindley, N. D. F. & Joyce, C. M. Distinct roles of the active-site Mg<sup>2+</sup> ligands, Asp882 and Asp705, of DNA polymerase I (Klenow fragment)

- during the prechemistry conformational transitions. *J. Biol. Chem.* **286**, 3755–3766 (2011).
26. Luo, G., Wang, M., Konigsberg, W. H. & Xie, X. S. Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. *Proc. Natl Acad. Sci. USA* **104**, 12610–12615 (2007).
  27. Wheeler, L. J., Rajagopal, I. & Mathews, C. K. Stimulation of mutagenesis by proportional deoxyribonucleoside triphosphate accumulation in *Escherichia coli*. *DNA Repair. (Amst.)* **4**, 1450–1456 (2005).
  28. Dahlberg, M. E. & Benkovic, S. J. Kinetic mechanism of DNA polymerase I (Klenow fragment): Identification of a second conformational change and evaluation of the internal equilibrium constant. *Biochemistry* **30**, 4835–4843 (1991).
  29. Beard, W. A. & Wilson, S. H. Structural insights into the origins of DNA polymerase fidelity. *Structure* **11**, 489–496 (2003).
  30. Kapanidis, A. N. *et al.* Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl Acad. Sci. USA* **101**, 8936–8941 (2004).
  31. Lee, N. K. *et al.* Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.* **88**, 2939–2953 (2005).
  32. Holden, S. J. *et al.* Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Biophys. J.* **99**, 3102–3111 (2010).

### Acknowledgements

A.N.K. was supported by grants from the European Commission Seventh Framework Program (FP7/2007-2013 HEALTH-F4-2008-201418), the UK Biotechnology and Biological Research Council (BB/H01795X/1) and the European Research Council (261227). C.M.J.

and N.D.F.G. were supported by the US National Institutes of Health (NIH GM-28550). T.D.C. was supported by a Lindemann Trust Fellowship from the English-Speaking Union.

### Author contributions

J.H., C.M.J., N.D.F.G. and A.N.K. designed the research plan. J.H., L.A. and P.S. performed the confocal experiments. O.B., O.P. and C.M.J. contributed, tested and optimized novel reagents. J.H. and T.D.C. developed new tools for data analysis and analysed the data. J.H., T.D.C., C.M.J., N.D.F.G. and A.N.K. wrote the manuscript.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Hohlbein, J. *et al.* Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nat. Commun.* **4**:2131 doi: 10.1038/ncomms3131 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>