Time course analysis of enzyme-catalyzed DNA polymerization

Julius Rentergent, Max D. Driscoll, and Sam Hay*

Manchester Institute of Biotechnology, University of Manchester, UK

E-mail: sam.hay@manchester.ac.uk

Phone: +44 (0)161 306 5141

Running header
Abstract

Extracting kinetic parameters from DNA polymerase-catalysed processive polymerisation data using traditional initial-rate analysis has proven problematic for multiple reasons: The first substrate, DNA template, is a heterogeneous polymer and binds tightly to DNA polymerase. Further, the affinity and speed of incorporation of the second substrate, deoxynucleoside triphosphate (dNTP), varies greatly depending on the nature of the templating base and surrounding sequence. Here, we present a mathematical model consisting of the DNA template-binding step and a Michaelis-Menten-type nucleotide incorporation step acting on a DNA template with finite length. The model was numerically integrated and globally fitted to experimental reaction time courses. The time courses were determined by monitoring the processive synthesis of oligonucleotides of lengths between 50 and 120 nucleotides by DNA polymerase I (Klenow fragment exo⁻) using the fluorophore PicoGreen. For processive polymerisation, we were able to estimate an enzyme-template association rate $k_1$ of $7.4 \mu M^{-1} s^{-1}$, a disassociation rate $k_{-1}$ of $0.07 s^{-1}$, a $K_d$ of 10 nM, and the steady-state parameters for correct dNTP incorporation give $k_{cat}$ of 2.5 to 3.3 s⁻¹ and $K_m$ of 0.51 to 0.86 µM. From the analysis of time courses measured at temperatures between 5 and 25 °C, an activation energy for $k_{cat}$ of 82 kJ mol⁻¹ was calculated and it was found that up to 73% of Klenow fragment becomes inactivated or involved in unproductive binding at lower temperatures. Finally, a solvent deuterium kinetic isotope effect (KIE) of 3.0 to 3.2 was observed under processive synthesis conditions, which suggests that either the intrinsic KIE is unusually high, at least 30 to 40, or that previous findings, showing that the phosphoryl transfer step occurs rapidly and is flanked by two slow conformational changes, need to be re-evaluated. We suggest that the numerical integration of rate equations provides a high level of flexibility and generally produces superior results compared to initial-rate analysis in the study of DNA polymerase kinetics and, by extension, of other complex enzyme systems.
The enzyme DNA polymerase is of fundamental importance to the process of DNA replication, which is at the heart of life. The catalysed reaction is a phosphoryl transfer, in which a phosphodiester bond is formed between the 3' hydroxyl group of the primer and the α-phosphate of the 5' end of an incoming dNTP (1). DNA polymerase I is a single polypeptide with 928 amino acid residues from *Escherichia coli*. Its large C-terminal fragment (Klenow fragment, KF) contains the (324-517aa) polymerisation domain and the (521-928aa) 3'-5' exonuclease domain, but not the (1-324aa) 5'-3' exonuclease domain (2). The remaining 3'-5' exonuclease activity is abolished in the D355A/E357A mutant (3), which retains the same structure and kinetic properties as the native protein (4). This simplifies KF to a suitable model enzyme for processive DNA polymerisation without the need to include depolymerisation steps.

A plethora of complexities can be ascribed to the kinetic mechanism of DNA polymerase. Many of these arise from the nature of its first substrate, DNA template. The DNA template is a heterogeneous polymer chain composed of four different types of dNTPs. Due to its length, it offers multiple DNA polymerase binding sites with varying properties, which may further change dynamically dependent on the degree of DNA polymerisation. DNA template also binds tightly to most DNA polymerases, which makes the application of steady-state kinetics difficult (see Discussion). Most DNA polymerases catalyse multiple dNTP incorporations before dissociating, and kinetic models need to account for this processivity. Lastly, the identity of the templating base guides the selection of the second substrate, dNTP. The correct dNTP must be chosen with a very high selectivity over all alternative substrates and the identity of the next correct dNTP may change after each incorporation.

There has been early interest in the construction of template-directed polymerisation models. When the role of DNA template-primer as initiator for synthesis was still unknown, Simha et al. (5) modelled the dNTP monomer adsorption as a first-order process on an
infinite chain, and contrasted kinetic predictions made based on whether the availability of DNA polymerase restricts the number of active growing centers or not. Pipkin and Gibbs (6) extended the second case, in which polymerisation kinetics are assumed to be independent of DNA polymerase concentration, to the treatment of a template with finite length. In another report (7), the polymer species were explicitly modelled by setting up a system of differential equations and these were solved with a Laplace transform. However, this was only possible when enzyme binding was assumed to be slow and irreversible up to the template end, where the rate of release was fast. All three reports show that the molecular weight distribution of DNA polymer species can be approximated by a Poisson distribution in their models (5–7). While such a statistical distribution serves as an excellent way to model the reaction progress in simple conditions, more complex models do not have a similar analytic solution. Most importantly, reactant concentrations need to be approximated to be constant and simplifying assumptions, such as permanent binding of polymerase to DNA template, are required and appropriate experimental conditions need to be enforced when collecting experimental data. Further, any non-uniform behaviour along the DNA template, such as the slow replication at a pausing site, cannot be readily integrated. The application of the statistical template-polymer approach has therefore been limited.

A more traditional approach was taken by McClure and Jovin (8) who developed an early detailed processive kinetic model for DNA polymerase I and used steady-state equations to determine kinetic parameters. They were the first to report a burst phase, ie a faster incorporation of the first nucleotide than the succeeding ones, and later extensions of the model enabled the investigation of polymerase translocation and processivity (9). However, the initial-rate approach does not take finite template length into account and often fails when applied to high-affinity interactions such as the one between DNA polymerase and template.
Most of the mechanistic details and thermodynamic parameters of dNTP incorporation have been determined by combining steady-state and pre-steady-state methods, first for DNA polymerase I \((10)\) and later for KF \((11–13)\). However, the pre-steady-state approach is usually carried out under single-turnover conditions. Although this simplifies the product analysis, the approach neglects post-chemistry reaction steps, translocation of the polymerase to the next template position and processive polymerisation \((14)\).

Here, we present a highly flexible mathematical model for processive DNA polymerisation that explicitly contains the enzyme-template interactions and a finite DNA template length. Numerical integration of this model enables us to extract more information from experimental time course data than would have been possible with an initial-rate approach. The data were collected using our previously published PicoGreen (PG) assay \((15)\).

**Model construction**

KF-catalysed DNA polymerisation is a complex process involving two substrates, DNA template and dNTP, in an ordered-binding mechanism, which is identical for all polymerase families \((10)\). The KF mechanism is illustrated in Scheme 1 (see \((16)\) for a review). First, KF binds to DNA to form a binary complex (the corresponding rate is given by \(k_1\)). Subsequently, dNTP binds to the KF-DNA complex \((k_2)\) and induces a conformational change \((k_3)\) \((17)\). This is followed by catalysis of the phosphoryl transfer \((k_4)\) and a second conformational change \((k_5)\) \((13)\). After release of the product pyrophosphate \((\text{PP}_i)\), the binary KF-DNA complex is restored and may either dissociate or undergo the next round of polymerisation.

A two-step model is presented in the form of a system of ordinary differential equations (ODEs), which has three main features: (i) the enzyme-template interaction is explicitly modelled, (ii) all remaining steps (bracketed part of Scheme 1) are condensed into one
reaction following Michaelis-Menten kinetics, (iii) the polymer nature of DNA template is accounted for by including $2n + 2$ DNA species, where $n$ is the number of nucleotides on the ssDNA portion.

Scheme 1. Kinetic model for processive DNA polymerisation on a DNA template with finite template length $n$. DNA$_0$ designates the initial template-primer species, while the fully polymerised template species is DNA$_n$. The binding of KF to DNA template is described by mass-action parameters $k_1$ and $k_{-1}$. Binding and disassociation can occur at any point, independent of the degree of polymerisation of the DNA template. The dNTP incorporation process is condensed to one step modelled by steady-state parameters $k_{\text{cat}}$ and $K_m$, which consumes the substrate dNTP and releases pyrophosphate PP$_i$. The reverse reaction is assumed to be negligible. While the bracketed part is not included in the mathematical model, it illustrates the steps contributing to the steady-state parameters $k_{\text{cat}}$ and $K_m$ based on the model by Dahlberg and Benkovic (13). The steps show dNTP binding ($k_2$), pre-chemistry conformational change ($k_3$), phosphoryl transfer ($k_4$), post-chemistry conformational change ($k_5$) and release of PP$_i$ ($k_6$).

The integration of template-binding dynamics is necessary, because dissociation from template ($k_{-1}$) is the slowest step in the overall reaction scheme (11). Taking this into account, explicit modelling of the binding reaction enables the analysis of the remaining
reaction scheme comprising the dNTP incorporation using Michaelis-Menten kinetics involving just two parameters, $k_{\text{cat}}$ and $K_m$. In contrast, when applying the Michaelis-Menten equation to processive polymerisation data directly, the resulting parameters $k_{\text{cat}}$ and $K_m$ are strongly dependent on the particular DNA template and experimental conditions chosen, and give little information on the DNA polymerisation process. For example, under single-nucleotide-incorporation conditions, in which just a single nucleotide is to be incorporated per template strand, the reported $k_{\text{cat}}$ would be equal to the dissociation rate ($k_{-1}$) (11). Such analyses have therefore been criticised (18). Even the extension of the Michaelis-Menten equation to a two-substrate model is not adequate when combined with an initial-rate approach for reasons outlined in the Discussion. The solution presented here therefore does not rely on the estimation of initial rates, but instead focuses on the analysis of full reaction time courses.

Even though no initial-rate approach is used here, the Michaelis-Menten equation is still useful to construct a model with identifiable parameters. In Equation 1, $v$ is the dNTP incorporation velocity, which depends on the steady-state parameters $k_{\text{cat}}$ and $K_m$, the concentration of the KF-DNA complex and of the complementary dNTP as defined by template identity.

$$v = \frac{k_{\text{cat}} \times \text{KF-DNA} \times \text{dNTP}_{\text{complementary}}}{K_m + \text{dNTP}_{\text{complementary}}}$$ (1)

In the experiments presented in this study, all nucleotides are added in equal starting concentrations. To simplify the analysis further, the sequences of the templates employed have an even ratio of all four dNTPs, which leads approximately to a simultaneous equal reduction of all dNTP types during DNA synthesis. This holds true, as long as the distribution of nucleotides in the sequence is not heavily biased, and the DNA templates were designed accordingly. The model can be easily adapted to work with varying nucleotide concentrations, or with different kinetic parameters at each individual DNA template.
position. Instead of fitting Equation 1 to initial rates, it is integrated together with the rest of the model to produce the full time course. This is a suitable approximation as long as \( \frac{K_F}{(dNTP + K_m)} \ll 1 \) (19), where \( K_F \) is the initial enzyme concentration. The typical \( K_F \) concentration here is 50 nM, which is only a tenth of the estimated \( K_m \), so the approximation is assumed to hold even in those cases where the dNTP pool is completely depleted.

Generally, previous kinetic models have not accounted for the polymer nature of DNA template. However, if the enzyme is able to dissociate from DNA template before the template end, the decline of available polymer substrate during a reaction is fundamentally different from that of a pooled substrate: In a plot of substrate concentration vs time, the first follows a sigmoidal curve whereas the second declines exponentially. Thus, we included all polymer species in the model and compared it to the results of reduced model approximations.

In the following, the model parameters will first be estimated in a global fit to experimental time course data. The data were collected by varying the amounts of the components \( K_F \), DNA and dNTP. Subsequently, experimental data produced at variable oligonucleotide lengths, temperatures and in \( D_2O \) will be analysed.

**Methods**

Steady-state assays were performed essentially as described previously using a stopped PG assay (15). All experiments were performed in 10 mM Tris/HCl, 50 mM NaCl, 10 mM MgCl\(_2\), 1 mM dithiothreitol, pH 8.0 or pH* 8.0 (in the case of solvent KIE measurements performed in \( D_2O \)) and the reaction temperature was controlled using a circulating water bath.

**Reagents.** \( K_F \) exo\(^-\) was obtained from New England Biolabs (Hitchin, Hertfordshire, UK), was monitored between batches by SDS/PAGE (single band at 68 kDa) and its
concentration was determined spectrophotometrically (\(\varepsilon_{280} = 55450 \text{ M}^{-1} \text{cm}^{-1}\)) (20). dNTPs were obtained from Peqlab Ltd (Sarisbury Green, Hampshire, UK). PicoGreen was obtained from Invitrogen (Life Technologies, Paisley, UK) and its concentration was determined by \(\varepsilon_{500} = 70000 \text{ M}^{-1} \text{cm}^{-1}\) (21). D_{2}O was purchased from Goss Scientific Equipment Ltd (Crewe, Cheshire, UK). HPLC-purified DNA oligonucleotides were obtained from Eurofins MWG Operon (London, UK) and were designed to anneal to the T7 primer, leaving a 30 to 100 nucleotide long 5’ overhang. The sequence of the ssDNA region is randomised and has approximately equal ratios of A : T : G : C. No secondary structure is predicted in any of the sequences. All oligonucleotides were annealed by heating a 1:1:2 mixture of oligonucleotide:T7 primer to 95°C for 10-15 min, before allowing it to cool to room temperature over at least 45 min. The sequences are given below, with || indicating the start of the T7 complement:

(i) T7pR30 5’-TACGGATCCATGCTAGTCCATTAGCAGGTG||CCCTATAGTGAGTCTATAGTTA-3’
(ii) T7pR50 5’-GTGGACAGTCTGGTATGTAGTCAGGCTCCAGGAGTCGCCTATGC\nCAACCT||CCCTATAGTGAGTCTATAGTTA-3’
(iii) T7pR80 5’-GTAGCTGACTTCTTCACCACATCTACCAAAGTGGGCAGTCTGGTATGCAGTCAGGCTCCAGGAGTCGCCTATGCCAACCT||CCCTATAGTGAGTCTATAGTTA-3’
(iv) T7pR100 5’-GACTCGAATGTGACTCAGTGATAGCTGACTTCTTCACCACATCTACCAAAGTGGGCAGTCTGGTATGCAGTCAGGCTCCAGGAGTCGCCTATGCCAACCT\nCCT||CCCTATAGTGAGTCTATAGTTA-3’

Model setup. The model shown in Scheme 1 was assembled in Copasi (22). In its largest form, 205 species are involved in the reactions: 101 free DNA and 101 enzyme-bound KF-DNA species, free KF, dNTP and PP_{i}. Although the model was fitted to fluorescence
data, it was not necessary to include the fluorescent agent PG in the model, because it quenches the reaction immediately when added \((15)\), and the concentration is sufficient to fully saturate the produced dsDNA. Therefore, the amount of dsDNA formed is taken to be directly proportional to the increase in fluorescence signal, \(\Delta f\), which measures the difference between the fluorescence emission from PG bound to dsDNA and ssDNA (Eq 2). Further, the amount of dsDNA formed, or dNTP incorporated, is identical to the amount of PP\(_i\) formed, which is a convenient species to track in the computational model. Thus, the experimental signal \(F\) is represented by Eq 3. While it is necessary to include an offset proportional to the initial amount of DNA template-primer present, DNA\(_0\), the signal from all other species and instrument noise were found to be insignificant \((15)\).

\[
\Delta f = f_{ds} - f_{ss} \tag{2}
\]

\[
F = PP_i \times \Delta f + DNA_0 \times f_0 \tag{3}
\]

The initial concentrations of all species, other than the ones declared in the experimental conditions, are zero. To account for a reduced amount of active enzyme in the temperature dependence experiments, the concentration of active KF (KF\(_{\text{active}}\)) in the model is determined by scaling the expected enzyme concentration (KF\(_0\)) with a dimensionless parameter (KF\(_{\text{mod}}\)): KF\(_{\text{active}}\) = KF\(_0\) × KF\(_{\text{mod}}\). The same approach was used to modify the concentration of available template-primer DNA with a dimensionless parameter (DNA\(_{\text{mod}}\)) for all data sets: DNA\(_{\text{active}}\) = DNA\(_0\) × DNA\(_{\text{mod}}\). Seven of the data sets showed a marked improvement in quality of fit when including this modification. Because the fitted parameter was consistently estimated to be lower than 1 (DNA\(_{\text{mod}}\) = 0.80 ± 0.06, \(n = 7\)), we concluded that the annealing process of primer to oligo was not fully efficient. It is expected this is also true for all other data sets, which do not contain sufficient information to determine DNA\(_{\text{mod}}\). Therefore, we set the parameter to 0.8 when it could not be unambiguously fitted for.
The model was also set up independently in Mathematica 10.0 (Wolfram Research, Inc., Champaign, IL) employing the NDSolve framework. Because this approach requires entering the differential equations directly, it lends itself to a cross-validation platform for large Copasi models, which take a reaction-centric approach. It was also used to simulate the data analysed in the reduced model approximation and to produce all figures.

*Global regression analysis.* The model from Scheme 1 was numerically integrated with the LSODA algorithm and was fit to experimental data in Copasi using the standard parameters for each optimisation algorithm. First, an Evolutionary Strategy algorithm was applied (SRES), followed by local optimisation with the gradient-based Levenberg-Marquardt algorithm. This procedure was shown to be successful for a number of biological optimisation problems by Rodriguez-Fernandez et al. (23). It should be noted that the same minima were found when applying the faster Hooke-Jeeves direct-search algorithm in tandem with Levenberg-Marquardt.

In the initial calibration, the kinetic parameters $k_1$, $k_{-1}$, $k_{\text{cat}}$ and $K_m$ were globally fitted to the experimental data sets of KF, DNA and dNTP dependencies. The global optimisation was weighted in order to ensure similar contributions of each data set to the overall objective value. The weights were 1x for KF, 3x for DNA, and 1x for dNTP. For all other regressions, kinetic parameters were fixed to the values from this calibration (Table 1), unless they were estimated. The calculated sum of squares and auxiliary parameters $\Delta f$, $f_0$, $\text{KF}_{\text{mod}}$, $\text{DNA}_{\text{mod}}$ and $k_{\text{end}}$ are included in Table S1. While $\Delta f$ and $f_0$ are expected to be constants under a given set of conditions, the values change depending on the DNA template employed, fluorometer settings and lamp output. However, the specific values can be readily estimated from the analysis of multiple time courses in a global regression. Note that, if multiple time courses are not available, a standard curve can be constructed to estimate the parameters (15).
The standard errors of fitted parameters were taken from the Copasi output. The standard errors of parameters which are ratios of fitted parameters were calculated using two different methods: for the KIE, which is a ratio of uncorrelated parameters, Equation 4 was used.

\[
\text{KIE}_{SE} = \frac{X_{Value}}{Y_{Value}} \times \sqrt{\left( \frac{X_{SE}}{X_{Value}} \right)^2 + \left( \frac{Y_{SE}}{Y_{Value}} \right)^2}
\]  \tag{4}

For parameters made from a combination of parameters that have some degree of correlation, such as \( k_{\text{cat}}/K_m \) and \( K_d \), the standard errors were again estimated in Copasi. This was done by making one of the single parameters dependent on the other two and subsequently fitting for the combined parameter.

**Analysis of published data.** The model in Scheme V from Dahlberg and Benkovic (13) was implemented in Copasi including all relevant parameters and initial conditions. This model was first used to accurately calculate their implicit steady-state rate and compare the result to our \( k_{\text{cat}} \) values. Next, a number of steady-state rate calculations were made for different values of KIE\(_{\text{int}}\) after substituting \( k_4 \) and \( k_{-4} \) for \( 150/\text{KIE}_{\text{int}} \) s\(^{-1}\) and \( 37.5/\text{KIE}_{\text{int}} \) s\(^{-1}\). The values of KIE\(_{\text{int}}\) which corresponded to a 3 to 3.2-fold drop in the steady-state rate are taken to reflect the KIE\(_{\text{int}}\) of phosphoryl transfer, assuming the model is correct. Finally, the model was numerically integrated and the full time courses were compared and fitted to the pulse-chase/pulse-quench data from their Figure 2 (Figure S2).

**Results**

*Model Calibration.* The presented model in Scheme 1 is designed to produce the time courses of DNA polymerase-catalysed synthesis of dsDNA of arbitrary finite length. Three species are necessarily involved in this process: DNA polymerase, dNTPs and a template-primer strand of DNA forming the initiation locus. The most straightforward test for the model are sets of
experiments, in which the concentration of one species is varied, while the other two are held constant, and the time course of the polymerisation progress is recorded. Subsequently, the time courses can be analysed with regression methods: if a single time course is analysed, this is referred to as local nonlinear regression. Global nonlinear regression, on the other hand, is performed when multiple time courses, or a set of time courses, at various concentrations of species are analysed simultaneously and model parameters are shared. The sharing of parameters often leads to an increase in confidence of the estimated parameters, which can, in some cases, enable the estimation of novel parameters that can not be determined when fitting the data sets independently. Sharing parameters across the fit of multiple datasets can further reveal inconsistencies in the data or false assumptions in the model, which may have gone unnoticed when analysing data locally (24).

For a simple reaction involving just two species, enzyme and substrate, it would be sufficient to record the time courses at different dNTP concentrations and fit an appropriate model to all of the data points in a global regression analysis. While this is an appropriate method to analyse a one-substrate reaction model, it is insufficient for determining all parameters of a more complex kinetic model such as the one depicted in Scheme 1. For example, if this model is fitted to only one experimental dNTP dependency data set (in which dNTP is the variable component), it would be difficult to determine the polymerase-DNA association rate constant $k_1$ and $k_{\text{cat}}$ at the same time, because their behaviour is correlated in the search of the optimal parameter: if $k_1$ increases and therefore more enzyme-bound DNA species are present at any one time, much of the effect on the polymerisation rate can be countered by decreasing $k_{\text{cat}}$, and vice versa. One popular workaround for this particular problem has been to use an excess of DNA polymerase, thus assuming that the concentration of DNA template is identical to the concentration of the active polymerase-DNA complex (14). However, this approach has a number of disadvantages: (i) it may give inaccurate results, because the active polymerase-DNA complex may exist
in significantly lower quantities than expected throughout the time course; (ii) including necessary assumptions about the polymerase-DNA interaction makes it impossible to gain information on this interaction, such as the calculation of a $K_d$; (iii) it is possible that the system can only be probed in concentration regimes that are not biologically relevant; (iv) if the polymerase-DNA interaction behaves in a more complicated way than assumed, e.g. due to enzyme dimerisation (25), the model fitting may not necessarily fail, but the computed parameters could be unreliable.

Realising the insufficiency of single dependency experiments, we first fitted three sets of time courses to the model in Scheme 1 in a global regression (Figure 1). In each of these sets, either KF, DNA or dNTP concentration was the variable component, while the other two were held constant. The resulting data set was therefore 5-dimensional; three for species, one for the time point and one for the resulting fluorescence. All of the parameters $k_1$, $k_{-1}$, $k_{\text{cat}}$ and $K_m$ (Table 1) were uniquely determined. The combined parameters $k_{\text{cat}}/K_m$ and $K_d = k_{-1}/k_1$ were subsequently derived from the results.
Figure 1. Global calibration fit. The varied component was either (A) KF, (B) T7pR80 DNA template, or (C) dNTP, and the reaction was started by addition of KF and stopped with 3 µM PG. The experiment was conducted in 10 mM Tris/HCl, 50 mM NaCl, 10 mM MgCl₂, 1 mM dithiothreitol, pH 8.0 at 25°C. KF dependence: 0 to 200 nM KF, 50 nM T7pR80, 500 µM dNTP. T7pR80 dependence: 5 nM KF, 0 to 50 nM T7pR80, 500 µM dNTP. dNTP dependence: 50 nM KF, 100 nM T7pR80, 0 to 100 µM dNTP.

Table 1. Results of global regression of Scheme 1 to the three data sets shown in Figure 1. $k_1$, $k_{-1}$ and $K_d$ refer to the enzyme-template binding parameters, while $k_{cat}$ and $K_m$ refer to the steady-state parameters of dNTP substrate.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value ± Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$ / (µM⁻¹ s⁻¹)</td>
<td>7.4 ± 1.0</td>
</tr>
<tr>
<td>$k_{-1}$ / s⁻¹</td>
<td>0.07 ± 0.01</td>
</tr>
<tr>
<td>$K_d$ / nM</td>
<td>10.0 ± 1.8</td>
</tr>
<tr>
<td>$k_{cat}$ / s⁻¹</td>
<td>3.3 ± 0.1</td>
</tr>
<tr>
<td>$K_m$ / µM</td>
<td>0.51 ± 0.04</td>
</tr>
<tr>
<td>$k_{cat}/K_m$ / (µM⁻¹ s⁻¹)</td>
<td>6.5 ± 0.6</td>
</tr>
</tbody>
</table>
Reduced model approximation. In the regression analysis above, the model explicitly contains every single polymer species: For example, simulating a 100-mer requires 202 species of DNA and polymerase-DNA species. While this is the most flexible form, because all kinetic parameters at each polymer step can be directly accessed and changed intuitively, some applications may require a computationally less expensive approximation. Examples of such applications include the analysis of longer DNA templates, the addition of multiple enzyme-template states or the calculation of Monte Carlo confidence intervals. Therefore, we constructed a number of reduced models with fewer species. Such models, of polymer length $l_{\text{model}}$, were then compared to the model of the desired oligo length, $l_{\text{target}}$. In the short models, the rate $v_{\text{short}}$ directs the rate of catalysis and the production of $\text{PPi}$. It is calculated as $v_{\text{short}} = v \times l_{\text{model}}/l_{\text{target}}$. Next, multiple time courses were generated assuming a model with $l_{\text{target}} = 100$ and a generic parameter set of $k_1 = 10 \text{s}^{-1}$, $K_d = 10 \text{nM}$, $k_{\text{cat}} = 5 \text{s}^{-1}$, $K_m = 0.5 \mu\text{M}$. The initial simulation conditions were similar to those used in the experimental calibration experiment (Figure 1). The time course data were fitted using global regression (Figure S1).

It is apparent that fitting a polymer reaction with a monomer model is unsatisfactory, especially if the amount of DNA polymerase is higher than the amount of DNA template. The main time course features can be reproduced by models above a length of 3. Figure 2 illustrates how the estimated parameters differ from its known value. It is interesting to note that it only requires the simulation of a length 5 polymer to estimate $k_{\text{cat}}$ and $K_m$ within 10% of their true value. Since many applications involve the incorporation of dNTPs using well-studied enzymes, this could be a useful approximation. In order to determine $k_1$ and $K_d$ with a similar accuracy, a model polymer length of 10 to 15 is necessary. As expected, the parameters only result in the true value at a model length of 100, but in many cases half the model length produces indistinguishable results. Nevertheless, for the following experiments, it was possible to perform each parameter
regression using the full model in hours on a desktop computer, thus no approximation has been made. The model was used to determine a selection of the kinetic parameters under different conditions, while the remaining parameters were fixed to the values estimated in the calibration above. The novel conditions tested include variation of (i) DNA template length, (ii) temperature and (iii) solvent isotopic content.

![Figure 2](image.png)

**Figure 2.** Reduced model approximation of a 100-mer. Global regression estimates depend on the number of explicit DNA species in the model. All parameter results were divided by the corresponding simulation values.

**DNA template length dependence.** We next investigated whether the model can reproduce the time courses of experiments in which the oligonucleotide length is varied. The oligonucleotides comprised a single-stranded portion of 30, 50, 80 or 100 nucleotides. The data were again analysed using global nonlinear regression (Figure 3). Interestingly, a $k_{\text{cat}}$ of 2.5 s$^{-1}$ can be employed to reproduce all of the time courses (Table S2), which indicates that the increase in template length does not significantly increase unproductive binding by interaction with longer single-stranded and completed oligonucleotides. The fit was also significantly better if the dissociation of KF ($k_{-1}$) from the final template position (DNA$_n$) was allowed to differ from its expected value of 0.07 s$^{-1}$ (Table 1).

The value was estimated to be $1.7 \pm 2.3$ s$^{-1}$ (Table S2), so it is likely that the dissociation of DNA polymerase from completely polymerised DNA is somewhat faster than dissociation from uncompleted template-primer. This effect becomes more important the shorter the
oligonucleotide is, and may therefore have been hidden in the analysis of the previous experiments with the longer 80-mer template. We thus included a separate $k_{-1}$ for the final dissociation in a later experiment in which T7pR30 template is used again.

Figure 3. DNA template length dependence fit with a shared $k_{\text{cat}}$ parameter for all data sets. The reaction was started by adding KF and stopped with 3 µM PG. The experiment was conducted with 50 nM KF, 100 nM T7pR30 or T7pR50 or T7pR80 or T7pR100, 500 µM dNTP in 10 mM Tris/HCl, 50 mM NaCl, 10 mM MgCl$_2$, 1 mM dithiothreitol, pH 8.0 at 25°C.

Temperature dependence. Brown and LiCata (26) have suggested that KF becomes inactivated at temperatures below approximately 20°C, while still tightly bound to DNA. We tested whether the Scheme 1 model is able to account for these proposed differences and fitted it to a series of experimental time course data measured under saturating dNTP conditions, in which both KF concentration and temperature were varied. Global regression analyses were performed for data sets at each of the temperatures 5, 10, 15, 20 and 25°C (Figure 4). The results show $k_{\text{cat}}$ is a well-determined parameter and decreases from 5.6 s$^{-1}$ at 25°C to 0.43 s$^{-1}$ at 5°C. All $k_{\text{cat}}$ values were fitted to the Arrhenius equation, which revealed an activation energy of 82 ± 10 kJ mol$^{-1}$. This value is in agreement with a previous estimation of 71 kJ mol$^{-1}$ (8).

The factor KF$_{\text{active}}$ reduces the initial KF concentration and is an essential part of the model below 20°C. It is important to note that this reduction in active enzyme concentration
can not simply be substituted by a reduction in the rate of KF-DNA association ($k_1$) or a reduction in $k_{cat}$ (see Discussion).

**Figure 4.** Temperature dependence fits of the reaction at (A) 5°C, (B) 10°C, (C) 15°C, (D) 20°C, (E) 25°C. The reaction was started by adding KF, as given in legend, and stopped with 3 µM PG. The experiment was conducted with 50 nM T7pR100 DNA template, 500 µM dNTP in 10 mM Tris/Cl, 50 mM NaCl, 10 mM MgCl$_2$, 1 mM dithiothreitol, pH 8.0. (F) Arrhenius plot of $k_{cat}$ with $E_a$ of 82 ± 10 kJ mol$^{-1}$ and pre-exponential factor $A$ of $1.3 ± 5.6 \times 10^{16}$ s$^{-1}$. 
Initially, the fitted KF binding parameters determined from these enzyme dependencies were underdetermined. At lower temperatures, the inclusion of $k_{-1}$ in the model introduced excessive parameter correlation. The DNA dissociation appears to be sufficiently slow at these lower temperatures as to be apparently irreversible over the experimental timescale. Consequently, DNA binding was made irreversible at temperatures below 20°C. At all temperatures, the errors in $k_1$ and $K_d$ were large, because enzyme dependencies alone do not sufficiently restrain those parameters of the model. At all temperatures, the results for $k_{cat}$ and KF$_{active}$ were nearly identical in both magnitude and trend, irrespective of whether the reversible or irreversible model was fitted.

Table 2. Temperature dependence parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Temperature / °C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td>$k_1$ /$\mu$M$^{-1}$s$^{-1}$</td>
<td>5.5 ± 2.6</td>
</tr>
<tr>
<td>$K_d$ /nm</td>
<td>20 ± 8</td>
</tr>
<tr>
<td>$k_{cat}$ /s$^{-1}$</td>
<td>5.6 ± 0.4</td>
</tr>
<tr>
<td>KF$_{active}$/KF$_0$ b</td>
<td>0.95 ± 0.17</td>
</tr>
</tbody>
</table>

$a$ not existent in irreversible model.

$b$ KF$_{active}$/KF$_0$ is the fraction of active enzyme.

Solvent isotope effects. The question of how DNA polymerase achieves its extraordinary fidelity has been the subject of intense debate over the past thirty years (16, 17, 27, 28). Fidelity is a concept defined by the kinetic mechanism and parameters involved. For KF, it is assumed that a slow conformational change precedes a fast chemical step, which is then followed again by a slow conformational change step (bracketed part of Scheme 1) (11, 13). In this model, the first conformational change is rate-limiting in single-nucleotide incorporation, the phosphoryl transfer reaction (which is the isotopically-sensitive step (29)) is rapid, and the second conformational change is rate-limiting in processive nucleotide incorporation. From this, we would expect the $k_{cat}$ measured in our experiments to be dominated by the second conformational change. Thus, a change of the solvent from H$_2$O to
D$_2$O is expected to have little effect on $k_{\text{cat}}$ and the observed KIE should be close to unity, if the pH of the D$_2$O solution is adjusted to offset the change in p$K_a$ of relevant ionisable groups upon deuteration (30, 31).

We tested this prediction by comparing the polymerisation time courses using T7pR80 template with varying dNTP concentrations in D$_2$O at pH* 8.0, and compared it to the dNTP dependency in H$_2$O (pH 8.0) presented before (Table 1). Global regression was performed as before (Figure 5 A) and the results are shown in Table 3. Contrary to expectations, the $k_{\text{cat}}$ reduces from 3.3 ± 0.1 s$^{-1}$ in H$_2$O to 1.1 ± 0.1 s$^{-1}$ in D$_2$O, resulting in a KIE$_{\text{obs}}$ of 3.0 ± 0.2. The KIE$_{\text{obs}}$ on $K_m$ is of similar magnitude, while the KIE$_{\text{obs}}$ on $k_{\text{cat}}/K_m$ is nearly unity. These results and the interpretation of the steady-state parameters will be examined in detail in the Discussion.

Next, we investigated if these intriguing results can be replicated with a shorter template, T7pR30. We recorded time courses from dNTP dependencies in H$_2$O and D$_2$O as before and fitted the Scheme 1 model to the data (Figure 5 B/C). The observed KIEs are not significantly different from those measured with the longer template (Table 3). These experiments clearly demonstrate that there is a KIE$_{\text{obs}}$ greater than 1 on $k_{\text{cat}}$ and $K_m$ in processive nucleotide incorporation for KF, while the KIE on the specificity constant $k_{\text{cat}}/K_m$ is close to 1.
Figure 5. dNTP dependence fits for (A) T7pR80 in D$_2$O, and for (B/C) T7pR30 in H$_2$O and D$_2$O. The reaction was started by adding KF and stopped with 3 µM PG. The experiment was conducted with 50 nM KF, 100 nM T7pR80 or T7pR30 DNA template, 0 to 100 µM dNTP in 10 mM Tris/HCl, 50 mM NaCl, 10 mM MgCl$_2$, 1 mM dithiothreitol, pH/pH* 8.0 at 25°C

Table 3. Results of dNTP dependence data regression including solvent KIE$_{obs}$ values.

<table>
<thead>
<tr>
<th>T7pR80 DNA</th>
<th>H$_2$O$^a$</th>
<th>D$_2$O</th>
<th>KIE$_{obs}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{cat}$ /s$^{-1}$</td>
<td>3.3 ± 0.1</td>
<td>1.13 ± 0.06</td>
<td>3.0 ± 0.2</td>
</tr>
<tr>
<td>$K_m$ /µM</td>
<td>0.51 ± 0.04</td>
<td>0.22 ± 0.03</td>
<td>2.3 ± 0.3</td>
</tr>
<tr>
<td>$k_{cat}/K_m$ /($µM^{-1}s^{-1}$)</td>
<td>6.5 ± 0.6</td>
<td>5.1 ± 0.7</td>
<td>1.3 ± 0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T7pR30 DNA</th>
<th>H$_2$O</th>
<th>D$_2$O</th>
<th>KIE$_{obs}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{cat}$ /s$^{-1}$</td>
<td>2.6 ± 0.1</td>
<td>0.83 ± 0.02</td>
<td>3.2 ± 0.2</td>
</tr>
<tr>
<td>$K_m$ /µM</td>
<td>0.86 ± 0.07</td>
<td>0.40 ± 0.03</td>
<td>2.2 ± 0.3</td>
</tr>
<tr>
<td>$k_{cat}/K_m$ /($µM^{-1}s^{-1}$)</td>
<td>3.1 ± 0.2</td>
<td>2.1 ± 0.2</td>
<td>1.5 ± 0.1</td>
</tr>
</tbody>
</table>

$^a$from Table 1
Discussion

In this paper, we have demonstrated how numerical integration of an enzyme-catalysed polymer model, as shown in Scheme 1, can be used to conveniently analyse time course data of KF-catalysed synthesis of dsDNA of arbitrary template length. In the following, the biochemical insight we have gained will be discussed, and some advantages and limitations of the methods will be highlighted.

Polymerase-DNA interaction. The experimental conditions typically used in DNA polymerase assays are chosen in such a way as to allow simplifying assumptions, such as \([\text{polymerase}] \gg [\text{DNA}]\) (11, 14). Of course, when such conditions are used, this can prove useful for the analysis of dNTP incorporation, but such conditions prohibit the analysis of the interactions between polymerase and DNA. For the analysis of these polymerase-DNA interactions, it is mandatory to create conditions in which the concentration of polymerase and DNA are similar, and, under these conditions, assumptions such as permanent saturation of template with enzyme are invalid. One of our key objectives therefore was to account for the enzyme-template interaction by explicitly including it in the model in the form of a simple mass-action binding step.

The DNA-template interaction often systematically distorts the steady-state dNTP incorporation parameters estimated by initial-rate analysis for a number of reasons. Firstly, the DNA binding is 'tight', with previous reports of the \(K_d\) to be in the low nm range (11). While a Michaelis-Menten analysis can be performed using a modified quadratic form of the Michaelis-Menten equation (32) to account for tight binding and conditions where \([\text{polymerase}] \simeq [\text{DNA}]\), using this equation often violates the basic assumption that the observed system achieves steady-state rapidly. For example, the maximum amount of KF-DNA intermediate in the Scheme 1 model is only reached after 3 to 10 s under the conditions employed here. Even if these complications arising from tight binding were removed, one of
the key parameters estimated from initial-rate analysis, the DNA $K_m$, is not defined (33) for a processive DNA polymerisation model such as the one depicted in Scheme 1 due to the cancellation of terms arising from the non-replenishment of free KF. Consequently, where we (15) and others (8) have previously used a DNA $K_m$ value to give a phenomenological description of experimental data, these must be interpreted within a model-free framework. Therefore, the analysis of whole time courses is mandatory to extract mechanistic details from processive polymerisation data produced with similar concentrations of KF and DNA template.

We were able to determine an association rate $k_1$ of KF-DNA binding of 7.4 $\mu$M$^{-1}$ s$^{-1}$ and a dissociation rate $k_{-1}$ of 0.07 s$^{-1}$, resulting in a $K_d$ of 10 nM (Table 1). The dissociation rate is consistent with previously published values of 0.06 s$^{-1}$ (11) or 0.1 s$^{-1}$ (10). Interestingly, Kuchta et al. (11) and others (34) observed an increase in $k_{-1}$ to 0.67 s$^{-1}$ as the polymerase acted close to the end of the template, an effect that we also had to include when investigating shorter oligomers in the template length dependency (Figure 3) and dNTP dependency (Figure 5 B/C). However, the large standard errors associated with the estimated values of 0.43 s$^{-1}$ and 1.67 s$^{-1}$ (Tables S2 and S4) prevent the assignment of a reliable rate constant from the available data.

Previous reports of the $K_d$ estimated using pre-steady-state methods are 4.5 to 5 nM (11) and 5.3 nM (13), which are comparable to the value estimated here. Similarly, Datta and LiCata (35) report a $K_d$ of 8.1 nM. Others have estimated, using DNase I footprinting, a $K_d$ of 2.1 nM (34) and 8 nM (36). The authors note that a significant amount of unpaired oligonucleotides present in the reaction mixture could lead to an increase in the estimated $K_d$. This is plausible seeing that the $K_d$ for the binding to unpaired oligos has been reported to be 68 nM (37), 40 nM (38) or as low as 4.9 nM (39). Thus, some degree of unproductive binding of KF to unpaired oligos, which were present at an amount equal to at least a fifth
of paired oligos in this study, seems likely. However, more mechanistic details are required in order to include this interaction in a computational model, eg the number of binding sites offered by unpaired oligos. The same is true for unproductive binding on dsDNA; the KF-DNA species in Scheme 1 comprises only the enzyme bound at the end of the template-primer portion and ready to bind incoming dNTPs, while other dsDNA-bound states along the ssDNA and dsDNA portions certainly exist. The model should thus be improved in the future with mechanistic details regarding the sliding of polymerase along DNA template and unproductive binding.

We were able to demonstrate the importance of unproductive binding when investigating the temperature dependence of the reaction (Figure 4). The best fit to the data was obtained when allowing the concentration of active KF to decrease with decreasing temperature, down to just 27% at 5°C (Table 2). This agrees with previous findings (26), which indicated that KF becomes inactivated at temperatures around 15°C, although the binding affinity for template-primer DNA is largely unchanged (40). However, at least for Klentaq polymerase, the $K_d$ for polymerase binding to unpaired oligos decreases from 34 nM at 25°C to 22 nM at 5°C (39). If this trend applies equally to KF, the decrease in active enzyme species we observe is probably a combination of tighter binding to unpaired oligos, and a shift in enzyme population to inactive states on the ssDNA- and dsDNA portions of template-primer DNA. It was not possible to recognise this feature using initial-rate analysis. As discussed before, the reaction achieves steady-state conditions slowly and one of the key assumptions of initial-rate analysis does therefore not hold. However, even if initial-rate analysis was applicable, it would have yielded inferior results: the parameters $k_1$, $k_{cat}$ and the concentration of KF would be highly correlated, and a reduction of active KF would therefore have gone unnoticed and unknowingly distorted the kinetic parameters. Contrary to initial-rate analysis, the global analysis of full time courses enabled us to reduce the parameter correlations and distinguish the effects of decreased enzyme concentration from the effects on $k_1$ and $k_{cat}$.
**dNTP incorporation parameters** $k_{\text{cat}}$ **and** $K_m$. In the proposed Scheme 1 model of processive DNA polymerisation, the steps involving dNTP binding, orientation, conformational changes, catalysis of phosphoryl transfer, release of products and translocation are convoluted in one step described by steady-state parameters $k_{\text{cat}}$ and $K_m$. In the following, we will first compare the parameters estimated here to previously published values and then discuss their mechanistic interpretation.

The $k_{\text{cat}}$ values determined for different templates were broadly similar at 2.5 to 3.3 s$^{-1}$ when measured at 25°C. While this could indicate that the polymerisation rate across the full template length is approximately constant, it could also indicate that the ratio of slow to fast polymerisation regions is similar for the DNA templates investigated. We believe the latter explanation is more likely, which implies that the steady-state parameters values estimated here are averaged values across the entire DNA template. While this is a limitation of the current study, the mathematical model can be easily adapted to treat one, multiple, or all of the 101 $k_{\text{cat}}$ and $K_m$ values (for a 100-mer) individually. This level of control cannot be obtained using an initial-rate or analytical integration approach, but obviously more data would be required in order to obtain well-defined parameters. For example, previous polymerase studies have employed sequencing gels to estimate sequential incorporation rates for up to 17 nucleotide incorporations ($41$–$43$).

Numerous evidence shows that this flexibility would be required for a complete description of processive polymerisation in order to account for the complex dependence of the nucleotide incorporation rate on nucleotide identity (A, C, G or T) and sequence environment. Early experiments on homopolymers found that $k_{\text{cat}}$ and $K_m$ can differ up to 30-fold for different dNTPs ($44$). This has, in part, been attributed to the unsuitability of homopolymer templates, because of undesired interactions such as the stacking of T-primers on A-
templates (41). Additional results have, however, confirmed the importance of sequence on polymerisation rate: Bertram et al. (45) found that both the templating base and neighbouring base greatly affect the $k_{\text{cat}}$ and $K_m$ of dNTP incorporation, while Mytelka and Chamberlin (46) found that T7 DNA polymerase and KF can pause significantly, depending on the environment, during incorporation of dCTP opposite a pyrimidine-G-C site. Maier et al. (47) also found that KF pauses for up to 1000s of seconds during replication of such a sequence. Since the DNA templates we used in this study contain two or more pyrimidine-G-C sites, this may explain why the estimated $k_{\text{cat}}$ values are slightly lower than previously estimated rates. Previous measurements of $k_{\text{cat}}$ include 8.3 s$^{-1}$, 5.7 s$^{-1}$, 7.5 s$^{-1}$, 7 s$^{-1}$, and 5.3 s$^{-1}$ (8, 11, 37, 44, 47). Schwartz and Quake (48) calculated a rate as high as 14 s$^{-1}$ when subtracting pauses. However, there are also $k_{\text{cat}}$ values near or below the ones reported here from studies which did not involve pyrimidine-G-C replication sequences such as 3.8 s$^{-1}$, 1.0 s$^{-1}$ and 2.0 s$^{-1}$ (10, 49, 50). The estimated $K_m$ values of 0.51 µM (Table 1) and 0.86 µM (Table 3) reported here are in line with previous estimates of 1 to 2 µM (8), 0.4 to 1.7 µM (49), 1.4 µM (37) and 0.58 µM (50).

In all conducted experiments, the concentration of non-complementary dNTP is approximately three times higher than the concentration of the complementary dNTP throughout the reaction. KF shows little discrimination in ground-state binding of nucleotides compared to other DNA polymerases (51), so the other nucleotides may act as inhibitors of the reaction. If so, the reported steady-state values become apparent parameters. However, as McClure and Jovin (8) measured a $K_i$ as high as 1 mM for the inhibitors dGTP or ATP, when dATP is expected, it is not clear whether this translates into a significant change in reaction velocity. Others (13) have also found that ATP does not act as a significant inhibitor. We also found that an increase in non-complementary dATP or dGTP from a 2-fold excess to 12-fold excess did not have any inhibitory effect on the reaction rate (not shown).
The meaning of \( k_{\text{cat}}, k_{\text{cat}}/K_m \) and \( K_m \). The interpretation and value of steady-state kinetic parameters for nucleotide incorporation in the study of DNA polymerase has been subject to much debate (18, 52). The difficulty arises because DNA polymerase does not catalyse dNTP incorporation in isolation, but only in conjunction with a DNA template and this interaction may mask the kinetic properties of dNTP incorporation. This was shown by Kuchta et al. (11), who investigated the polymerisation of DNA templates with single-nucleotide overhangs by KF. Using initial-rate analysis, they found that \( k_{\text{cat}} \) equals the dissociation rate from DNA template, \( k_{-1} \). Hence the \( k_{\text{cat}} \) did not report on any step involving dNTP incorporation because a far slower step, DNA dissociation, was necessary for the continuous polymerisation process. This criticism of steady-state parameters does not apply to the estimated values in this study because the Scheme 1 model includes the DNA binding steps and thus does not rely on simplifying binding assumptions.

While the presented model is able to separate the DNA binding process, thereby ensuring that the steady-state parameters report on the dNTP incorporation, it is not obvious which microscopic rates of the complex process (bracketed part of Scheme 1) determine the steady-state parameters. In isolation, the presented processive incorporation data do not allow conclusions on this matter. While it is desirable to devise experiments which allow the calculation of some of the microscopic rate constants during observation of the macroscopic replication process under biologically relevant conditions, this may require the use of multiple simultaneous physical probes and large amounts of data. The method of full time course analysis would remain essential for the analysis of more complex experiments. It is worth noting that previously calculated microscopic rates were also derived from full (pre-steady-state) timecourses (14). For further interpretation of the presented data, it is instructive to draw on the rich literature of Klenow fragment. Previous findings have shown that dNTP binding (\( k_2 \)) is at rapid equilibrium (4, 10), although this equilibrium may be coupled and
result from multiple sequential binding steps, as suggested for T4 DNA polymerase (53). The first conformational change \((k_3)\) occurs at 50 s\(^{-1}\), the phosphoryl transfer \((k_4)\) at a minimum of 150 s\(^{-1}\), the second conformational change \((k_5)\) at 15 s\(^{-1}\), and the release of PP\(_i\) \((k_6)\) is essentially instantaneous \((11, 13)\). The slowest of these rates, the second conformational change at 15 s\(^{-1}\) is thus assumed to dominate the \(k_{\text{cat}}\) we estimate. While others \((47)\) have approximated these two rates to be identical, an improved rough calculation of the steady-state rate is:

\[
k_{\text{cat}} = \frac{1}{K_4 k_5} + \frac{1}{k_5} + \frac{1}{k_3} = 9.7 \text{s}^{-1}
\]  

(5)

A thorough steady-state calculation including all steps and conditions used in the study returns 9.3 s\(^{-1}\). As described above, sequence-dependent pausing could account for the 3-fold difference to our estimated \(k_{\text{cat}}\) values.

The contributions of microscopic parameters to the \(K_m\) value are even more difficult to disentangle. Due to the separation of dNTP binding and rate-limiting step, the \(K_m\) does not have an obvious relation to the binding affinity of the polymerase-DNA complex for dNTP and is therefore of little use \((14)\). This problematic gap between substrate binding and slow step has been treated more generally by Northrop \((54)\), who advises against the use of \(K_m\) for such mechanisms. Although we report the \(K_m\) values for comparison purposes, we follow his recommendation and instead focus on \(k_{\text{cat}}/K_m\). This is the rate constant for the capture of a productive complex committed to turnover and can, in contrast to \(K_m\), be intuitively applied to the mechanism of KF. In Scheme 1, \(k_{\text{cat}}/K_m\) refers to the rate of formation of the fraction of KF-DNA\(_n\)-dNTP, which will eventually form the product KF-DNA\(_{n+1}\) and free PP\(_i\). \(k_{\text{cat}}/K_m\) further possesses special significance for DNA polymerase, because the value of \(k_{\text{cat}}/K_m\) for correct incorporations divided by its value for misincorporations defines the fidelity of the enzyme, one of the key characteristics of DNA polymerases. This idea was put forward by Fersht \((55)\), and was later validated by Bertram et al. \((45)\) in both steady-state and pre-steady-state experiments.
Interpretation of observed KIEs. The observed solvent deuterium KIEs on $k_{\text{cat}}$ using either T7pR80 or T7pR30 template were 3.0 and 3.2, respectively. These results are similar to previous findings (29), where a KIE$_{\text{obs}}$ of 2 to 5 was found for poliovirus RNA polymerase, RB69 DNA polymerase, T7 RNA polymerase and HIV reverse transcriptase. The reported measurements were made in the pre-steady-state and the authors surmise that phosphoryl transfer is partially rate-limiting for those polymerases. It is therefore surprising that a similar KIE$_{\text{obs}}$ is found for KF. For KF, the chemical step ($k_4$) is assumed to be much faster than both conformational change rates ($k_3$ and $k_5$), which are assumed to be insensitive to isotopic substitution. It is perhaps even more surprising that the KIE$_{\text{obs}}$ is measured in the steady-state, because the second conformational change is even slower than the first conformational change (13).

The published model of Dahlberg and Benkovic (13) (bracketed part of Scheme 1) can be used to give an estimate of the intrinsic KIE on the phosphoryl transfer step ($k_4$). In order to achieve a 3-fold reduction in the steady-state rate $k_{\text{cat}}$ (ie a KIE$_{\text{obs}}$ of 3), the KIE$_{\text{int}}$ would have to be 30 to 40 or more. KIEs of this magnitude have been reported for reactions involving a large degree of hydrogen tunnelling (56). Alternatively, our data may suggest that the rate of phosphoryl transfer step is of the same order as the steady-state rate. This would contradict previous findings (13), so it seems appropriate to review the evidence arguing for a fast chemical step flanked by two slow conformational changes.

The initial proposal of a fast phosphoryl transfer step was based on a small elemental effect of 4 to 7 on the polymerisation rate when a nonbridging oxygen on the $\alpha$-phosphate was replaced with a sulfur atom (11). However, this line of evidence was since shown to be unreliable (57), which led others (27) to argue that the highest energy barrier is posed during the chemical step. This mechanism was proposed for DNA polymerase $\beta$
based on fluorescence studies (58), and is also supported by the theoretical reasoning that such a mechanism would account for a stronger discrimination between correct and incorrect dNTPs than the previous model (11, 12). However, for some polymerases, the strongest evidence comes from pulse-chase/pulse-quench experiments (16). In this type of experiment, the polymerisation reaction was either chased with an excess of unlabeled substrate or quenched with acid after a certain amount of time. Subsequently, the yield of labeled dNTP incorporation was compared between the experiments. The idea is that the yield of pulse-chase experiments is only higher if a pre-chemistry species has locked in the substrate and then proceeds with its incorporation despite an excess of unlabeled competition in the surroundings. When adding acid, on the other hand, all reactions simply stop due to denaturation of the polymerase. For KF, the pulse-chase yield was approximately 20% higher than the pulse-quench yield, which was interpreted as requiring two slow conformational changes, both before and after the chemical step (13). Further evidence for a slow conformational pre-chemistry step comes from studies employing fluorescent probes (59–61). For example, Purohit et al. (59) found that the signal referring to the putative rate-limiting step still occurs when using dideoxy-terminated primer, a non-potent substrate. This is interpreted as evidence for a rate-limiting conformational change step before phosphoryl transfer. It is not obvious what the nature of this conformational change is, although the extensive domain movement of fingers-closing, which predominantly occurs after dNTP binding, was ruled out. FRET studies have shown that it occurs at a rate of >1000 s\(^{-1}\) or 100 to 300 s\(^{-1}\), much faster than the rate of incorporation measured at 30 to 40 s\(^{-1}\) (59, 60). While Bakhtina et al. (62) confirm the rapid rate for the subdomain closure and the identity of a slow post-chemistry conformational change, their results from fluorescence studies contradict the idea of a slow conformational change preceding chemistry. When using dideoxy-terminated primer, they do not record a fluorescence transient which supposedly reports on the rate-limiting step. Further, they argue that the aforementioned pulse-chase/pulse-quench data only indicate the presence of a closed KF*-DNA-dNTP species,
but the chemistry step could still be rate-limiting. This situation was shown to be true for T7 DNA polymerase (63), for which the forward conformational change step is faster than chemistry, but the slow reverse conformational change rate is sufficiently slow so that the yield in a pulse-chase/pulse-quench experiment differs markedly.

In light of this important controversy, it seemed appropriate to revisit the original pulse-chase/pulse-quench experiment for KF (13). The published model was numerically integrated and fitted to the published data. The results show that a difference in pulse-chase/pulse-quench yield can occur if the rate of the chemical step is rate-limiting and the reverse rate of the first conformational change is not fast, as suggested by Bakhtina et al. (62). Further, while it is not possible to uniquely determine all parameters, we found that it is at least possible to describe the data (Figure S2) if a pre-chemistry step is of the same magnitude as the phosphoryl transfer step (150 s\(^{-1}\)). Putting together these different strands of evidence, it is currently not clear whether the chemical step is rapid or partially rate-limiting. By extension, further investigation is required to confirm whether the KIE\(_{\text{int}}\) on the phosphoryl transfer step is at least 30 to 40 or lower.

The observed solvent deuterium KIEs on \(k_{\text{cat}}/K_m\) using either T7pR80 or T7pR30 were 1.3 and 1.5, respectively. This finding can be rationalised by taking into account a theoretical model for the specificity constant for T7 DNA polymerase. Johnson (18) suggests that \(k_3/K_{d,d\text{NTP}}\) alone defines the specificity constant if dNTP binding is at rapid equilibrium and the reverse rate of the pre-chemistry conformational change is much slower than the chemical rate, \(k_{-3} \ll k_4\). Notably, the definition does not include the rate of the chemical step. We propose that this approximation can be made for KF too, which explains why substitution of solvent for D\(_2\)O does not have a strong effect on \(k_{\text{cat}}/K_m\). The small KIEs we observe are thus either due to statistical error, or because the reduced phosphoryl transfer rate in D\(_2\)O starts to approach the magnitude of the reverse rate of the conformational
change, thus breaking one of the two assumptions discussed above (18). Further, taking into account our previous definition of $k_{cat}/K_m$ as the rate constant for the capture of productive complex generation, the result is intuitive. As long as the reverse rate of the first conformational change is slow with regards to the following steps, none of these are relevant for the enzyme’s specificity. The enzyme locks in a substrate and rarely returns it, independent of the time it takes for subsequent steps. This kinetic model should be applicable to both KF and T7 DNA polymerase alike, even though the latter has a much higher forward rate for the first conformational change ($660 \text{ s}^{-1}$) and no rate-limiting post-chemistry step (18). Notably, KIEs of around 3 on $k_{cat}$ and unity for $k_{cat}/K_m$ have been observed before in studies of the human enzyme carbonic anhydrase II (64). The authors proposed a rate-limiting proton transfer step after the first irreversible step to explain the result, analogous to our own findings presented here.

Applications. The presented model emphasises the complete process of dynamic enzyme binding and repeated dNTP incorporations on a polymer template. This enables the analysis of complex processes, which are difficult to investigate under single-turnover conditions (65). This includes the translocation step (14), for which models have been developed and could be tested (66). Other applications include the analysis of replication at pausing sites or blocks such as pyrimidine dimers. Further, PCR is a key technology that relies on the processive polymerisation of a wide variety of DNA templates by engineered DNA polymerases. In order to compare and improve these polymerases, it is necessary to characterise them with models that account for the processive nature of the target task. In addition, the cycle run times could be more easily optimised if the process can be simulated using a bottom-up approach. Our finding that a long polymer, or a portion of the polymer, can be approximated by a simulated polymer of reduced length could prove useful in this regard. However, such approximations should be made with care if the enzyme system is not very well-known.
A further challenge in polymerase research arises from the fact that the process is in reality a highly stochastic process, whereas it is not usually treated as such. For example, it was found in a KF single-molecule study that the incorporation rate of the first nucleotide replication rate varies between 1 and 50 s$^{-1}$ (48). While this topic was not covered here, this behaviour could be readily produced using the mathematical model presented here. Instead of calculating the deterministic solution to the system of ODEs, as performed here, a stochastic solution could be calculated by employing Gillespie’s exact method (67) or an approximate stochastic method.

Conclusions

DNA polymerase-catalysed DNA polymerisation is a complex process, which is often inadequately described by initial-rate analysis using steady-state assumptions (18, 52). In this study, we have described the analysis of full reaction time courses by numerical integration of rate equations. The rate equations are in the form of a system of ODEs, which combines mass-action kinetics of enzyme-template binding, Michaelis-Menten-type incorporation of dNTPs, and a finite DNA template length. Global regression analysis of this model allowed the estimation of $k_1$, $k_{-1}$, $K_d$, $k_{\text{cat}}$, $K_m$ and $k_{\text{cat}}/K_m$ from experimental fluorescence data in good agreement with previously published values (Table 1). The KF-template binding interactions could not have been investigated with a steady-state approach due to its tight binding.

It was shown that the model can be used to describe the reaction at a range of temperatures, and that KF increasingly adopts inactive states when the temperature is lowered (Table 2), as suggested previously (26). However, it is not clear whether the shift to these states occurs at the template-primer junction or not. An alternative explanation for
our observations is the dimerisation of Klenow fragment, with changing dimer properties at lower temperatures. There is both evidence for (25) and against (38, 39) KF dimerisation occurring on DNA templates. It would be interesting to investigate whether this putative interaction induces changes that can be observed in a suitable kinetic experiment.

Unexpectedly, a KIE of around 3 was observed on both $k_{\text{cat}}$ and $K_m$ under processive polymerisation conditions in D$_2$O. It was concluded that either proton tunnelling makes a large contribution to the reaction rate, or that phosphoryl transfer is at least partially rate-limiting for KF, contrary to previous findings (13, 16). More investigations are required. For example, it could be instructive to determine the KIE$_{\text{obs}}$ on the first dNTP incorporation using pre-steady-state methods in order to remove the complexities of a second conformational change.

Systems of rate equations have always had the advantage of being simple to set up, intuitive to modify and reliant on fewer assumptions than the corresponding initial-rate approximation. With recent advances in computer hardware and software, most enzyme systems can now be numerically integrated at a speed that allows the estimation of parameters from time course data. This ability should be harnessed, especially for complex enzymes such as DNA polymerase.

**Supporting Information Available**

Assay procedure and initial conditions; full results tables; analysis of pulse-chase/pulse-quench data; Mathematica code for model generation. This material is available free of charge via the Internet at [http://pubs.acs.org/](http://pubs.acs.org/).
Funding

This work was funded by the UK British Biological Sciences Research Council (BB/H021523/1).

References


Polymerase I (Klenow Fragment) and Their Role in Nucleotide Selectivity. *Biochemistry* **47**, 6103–6116.


Julius Rentergent, Max D. Driscoll, Sam Hay
Time course analysis of enzyme-catalyzed DNA polymerization