Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computational Biology and Chemistry 29 (2005) 101–110
Computational Biology and Chemistry
www.elsevier.com/locate/compbiolchem

A macroscopic kinetic model for DNA polymerase elongation and high-fidelity nucleotide selection

Steve Viljoen [a], Mark A. Griep [b], Michael Nelson [c], Hendrik Viljoen [d,*]

[a] *Department of Biochemistry, Nebraska Wesleyan University, Lincoln, NE 68540, USA*
[b] *Department of Chemistry, University of Nebraska, Lincoln, NE 68588, USA*
[c] *Megabase Research Products, Lincoln, NE 68540, USA*
[d] *Department of Chemical Engineering, University of Nebraska, Lincoln, NE 68588, USA*

Received 6 January 2005; received in revised form 16 February 2005; accepted 17 February 2005

Abstract

The enzymatically catalyzed template-directed extension of ssDNA/primer complex is an important reaction of extraordinary complexity. The DNA polymerase does not merely facilitate the insertion of dNMP, but it also performs rapid screening of substrates to ensure a high degree of fidelity. Several kinetic studies have determined rate constants and equilibrium constants for the elementary steps that make up the overall pathway. The information is used to develop a macroscopic kinetic model, using an approach described by Ninio [Ninio J., 1987. Alternative to the steady-state method: derivation of reaction rates from first-passage times and pathway probabilities. Proc. Natl. Acad. Sci. U.S.A. 84, 663–667]. The principle idea of the Ninio approach is to track a single template/primer complex over time and to identify the expected behavior. The average time to insert a single nucleotide is a weighted sum of several terms, including the actual time to insert a nucleotide plus delays due to polymerase detachment from either the ternary (template-primer-polymerase) or quaternary (+nucleotide) complexes and time delays associated with the identification and ultimate rejection of an incorrect nucleotide from the binding site. The passage times of all events and their probability of occurrence are expressed in terms of the rate constants of the elementary steps of the reaction pathway. The model accounts for variations in the average insertion time with different nucleotides as well as the influence of G + C content of the sequence in the vicinity of the insertion site. Furthermore the model provides estimates of error frequencies. If nucleotide extension is recognized as a competition between successful insertions and time delaying events, it can be described as a binomial process with a probability distribution. The distribution gives the probability to extend a primer/template complex with a certain number of base pairs and in general it maps annealed complexes into extension products.

Keywords: Biochemical engineering; Molecular biology; Polymerase chain reaction; Mathematical model

1. Introduction

Template-directed nucleic acid synthesis is one of the greatest biochemical discoveries. Since the activity of the first RNA polymerase and DNA polymerase were studied in 1955 (Grundberg-Manago et al., 1955; Lehman, 2003), template-directed polymerases have contributed to our understanding of many subcellular processes, such as DNA replication and repair, transcription and telomere homeostasis. The accuracy, relative ability to bypass lesions and ability to elongate polymers processively makes these enzymes crucial to understanding such biological processes as infection, cancer, and aging and also identifies them as important drug targets (Ahmed and Tollefsbol, 2003; Miura and Izuta, 2004). Finally, template-directed nucleic acid polymerases are used in many procedures basic to modern biotechnology including PCR and DNA sequencing (Smith, 1980; Saiki et al., 1985).

The microscopic kinetic mechanisms and structure/function relationships of *Escherichia coli* DNA polymerase I, T7 DNA polymerase, KlenTaq DNA polymerase and a *Bacillus* DNA polymerase have been worked out in some detail (Johnson, 1993; Kiefer et al., 1998; Li

* Corresponding author. Tel.: +1 402 4729318; fax: +1 402 4726989.
E-mail address: hviljoen1@unl.edu (H. Viljoen).

1476-9271/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2005.02.003

and Waksman, 2001). Comparative analysis suggests that all DNA polymerases will have the same broadly conserved molecular properties and mechanisms, but that each polymerase will have its unique set of microscopic rates and equilibria (Jager and Pata, 1999; Joyce and Benkovic, 2004; Kunkel, 2004). The mechanisms of nucleotide insertion and incorporation are probably highly conserved although the details of the nucleotide selection process are just beginning to be understood.

Despite these advances on the molecular level, we still lack an understanding of the relationship of these parameters to macroscopic properties such as chromosomal elongation rates. This lack of theoretical understanding has practical consequences. For instance, it is one of the obstacles that prevent users from selecting the best DNA polymerase for a given PCR application. Currently researchers choose a polymerase, work out the PCR conditions semi-empirically and then adjust if necessary. Our poor theoretical understanding is also an obstacle for those researchers trying to solve the many PCR problem areas such as the amplification of GC-rich sequences, amplification of long sequences, maintaining low error frequencies, etc. There is a commercial impact in that many thermophilic DNA polymerases are available (Vieille and Zeikus, 2001) with various known biochemical properties (elongation rate, processivity, nucleotide selectivity, thermal half-life at 95 °C etc.) but it is not clear which properties are best for which PCR protocol. To bridge the gap between the microscopic and macroscopic in the realm of DNA polymerases, we have developed a model for DNA synthesis that accounts for the template nucleotide sequence and dNTP pool to predict DNA synthesis rates and error frequencies.

2. Mathematical model

The objective is to derive an expression for the extension rate of the polymerase. An outcome of the model is that it becomes possible to predict extension rate, fidelity, processivity and product yield in polymerase chain reactions. Whereas a microscopic model would focus on a complex sequence of steps and the details of the mechanism, the macroscopic model condenses the information into a few key steps. Although a reduction in complexity always accompanies a reduction in detail knowledge, it is also true that if the rate-limiting step can be identified, then further data become largely superfluous. In an effort to formulate a macroscopic model that describes the average dynamics of a polymerase/template complex, it is necessary to conjecture some sequence of steps, even though the exact mechanism of the polymerase is an ongoing study. Johnson (1993) proposed a kinetic scheme for DNA propagation by T7 polymerase and it is schematically presented in Fig. 1. The binary complex primer/template is denoted as ***A***1. This complex binds a polymerase molecule to form a ternary complex ***A***0. The dNTP pool comprises of dATP, dCTP, dGTP, dTTP

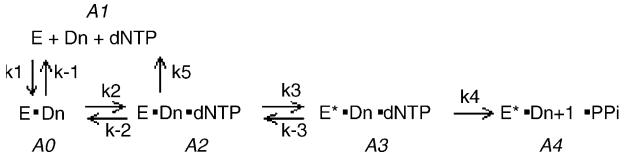

Fig. 1. Schematic of polymerase catalysed DNA elongation. ***A***1 presents the primer/template complex, ***A***0 is the ternary complex that includes the polymerase, ***A***2 is the quaternary complex that forms after addition of a dNTP. The model accounts for polymerase detechment from states ***A***0 and ***A***2. The quaternary complex adopts an activated state ***A***3 followed by phosphodiester bond synthesis, the release of pyrophosphate, and translocation of the enzyme. The schematic also identifies the notation to label the different states and the respective rate constants.

and dUTP. Uracil is formed by thermal deamination of dCTP, thus it will be present. Another component that is not usually considered part of the dNTP pool, but it competes for binding to the ternary complex is PP_i. Pyrophosphate is a product of the extension reaction and it is responsible for the pyrophosphorolysis reaction that reverses extension and produces dNTP. After one of these components have bound to the ternary complex to form the quaternary complex ***A***2, the polymerase may undergo a conformational change denoted as ***A***3. The final step to ***A***4 involves the release of PP_i. Patel et al. (1991) noted that the rate limiting step is the formation of the activated complex ***A***3 and the reaction following that is in equilibrium, because it is so much faster. The state ***A***4 is a ternary complex similar to ***A***0, but the number of base pairs has increased by one. The scheme differs from the one proposed by Johnson because a reaction has been added to account for the detachment of the polymerase from the quaternary state, i.e. ***A***2→***A***1. Due to the paucity of kinetic data, it is reasonable to assume that k_5 and k_{-1} would be similar.

The derivation of the kinetic rate expression that is presented in the following section differs from the conventional approach. There is a striking similarity between these two approaches and the two approaches that are used to derive the equilibrium energy distribution of an ensemble by statistical thermodynamics. It is well known that the same energy distribution is obtained either by sampling a large number of molecules at the same time or tracking a single molecule along a time-line to determine occupation periods of different energy levels. In kinetics, the classical approach to determine the rate of an elementary reaction is to sample the ensemble of reacting molecules over a short interval δt and find the number that has reacted. In the spirit of the ergodic principle, the alternative approach is to track a molecule over time and average the conversion times.

2.1. Reaction rates based on pathway probabilities

Ninio (1987) proposed an alternative approach to derive kinetic expressions for enzymatic reactions. For conventional reaction schemes, there is no benefit in this alternative approach. However, it is much better suited to handle the complexity of the nucleotide selection and editing.