Thermodynamics and kinetics of Lesion Induced DNA Amplification (LIDA)

Jan Engelhardt[1], Bent Andresen[1] and Steen Rasmussen[1,2,*]

[1]Center for Fundamental Living Technology (FLinT), University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark
[2]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[*]steen@sdu.dk

Introduction

Non-enzymatic short-chain XNA replication is often prohibited due to product inhibition, and product inhibition has been one of the major stumbling blocks for implementing prebiotic information replication systems both in bulk[4] and as part of synthetic protocells.[7] Product inhibition implies that a newly formed complementary template is very unlikely to dehybridize from the template. However, Gibbs-Davis and her team[1] have developed a method to circumvent product inhibition based on mismatches and lesions between the complementary templates that destabilize them and thus enables a continued replication process. In this work we explore in simulation the thermodynamic and kinetic properties of lesion induced DNA amplification processes.

The method we use focuses on a characterization of the oligomers by their free energies rather than by their detailed base-pair structure.

Reaction kinetics

A model of the lesion induced DNA amplification (LIDA) kinetics is presented below. The rates of formation of each of the X-containing species are[3]

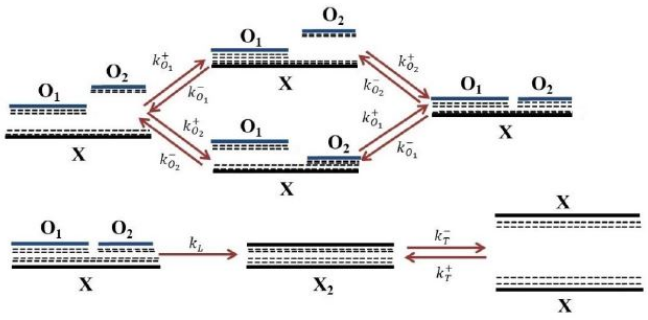

Figure 1: Schematic of idealized the LIDA reaction kinetics.[3] The reaction kinetics is simplified as the plus and minus template are assumed identical. Further, blunt-end ligations are not included.

$$X + O_1 \underset{k^-_{O_1}}{\overset{k^+_{O_1}}{\rightleftarrows}} XO_1;\; XO_1 + O_2 \underset{k^-_{O_2}}{\overset{k^+_{O_2}}{\rightleftarrows}} XO_1O_2 \tag{1}$$

$$X + O_2 \underset{k^-_{O_2}}{\overset{k^+_{O_2}}{\rightleftarrows}} XO_2;\; XO_2 + O_1 \underset{k^-_{O_1}}{\overset{k^+_{O_1}}{\rightleftarrows}} XO_1O_2 \tag{2}$$

$$XO_1O_2 \overset{k_L}{\rightarrow} X_2 \underset{k^+_T}{\overset{k^-_T}{\rightleftarrows}} 2X, \tag{3}$$

where X and O_1, O_2 denote the template strand and oligomers 1 and 2, respectively. Further, $k_i^{+,-}$ with $i \in \{O_1, O_2, T\}$ are the rate constants[2] (on-/off-rates) describing the corresponding forward and backward reactions. The differential equations which are derived from the rates of formation describing the processes in Fig. 1 are given by[3]

$$\begin{aligned}\frac{d}{dt}[X] = &-k^+_{O_1}[X][O_1] - k^+_{O_2}[X][O_2] + k^-_{O_1}[XO_1] \\ &+ k^-_{O_2}[XO_2] + 2k^-_T[X_2] - 2k^+_T[X]^2\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{d}{dt}[XO_1] = &\, k^+_{O_1}[X][O_1] - k^-_{O_1}[XO_1] - k^+_{O_2}[XO_1][O_2] \\ &+ k^-_{O_2}[XO_1O_2]\end{aligned} \tag{5}$$

$$\begin{aligned}\frac{d}{dt}[XO_2] = &\, k^+_{O_2}[X][O_2] - k^-_{O_2}[XO_2] - k^+_{O_1}[XO_2][O_1] \\ &+ k^-_{O_1}[XO_1O_2]\end{aligned} \tag{6}$$

$$\begin{aligned}\frac{d}{dt}[XO_1O_2] = &\, k^+_{O_2}[XO_1][O_2] + k^+_{O_1}[XO_2][O_1] \\ &- \left(k^-_{O_1} + k^-_{O_2} + k_L\right)[XO_1O_2]\end{aligned} \tag{7}$$

$$\frac{d}{dt}[X_2] = k_L[XO_1O_2] - k^-_T[X_2] + k^+_T[X]^2. \tag{8}$$

Furthermore, the total molar template concentration is

$$[X]_{\text{total}} = [X] + [XO_1] + [XO_2] + [XO_1O_2] + 2[X_2]. \tag{9}$$

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2020/32/269/1908548/isal_a_00312.pdf by guest on 03 August 2024

Thermodynamics

The equilibrium constants corresponding to the rates of formation in eqs. (1)-(3) are given by

$$K_i = \frac{k_i^+}{k_i^-} = \exp\left(-\frac{\Delta G_i}{RT}\right), \quad i \in \{O_1, O_2, T\}, \quad (10)$$

where ΔG_i is the change in free energy per mol for the reaction, R is the gas constant (1.987 cal/mol·K) and T represents the absolute temperature in kelvin. The off-rates are calculated by fixing the on-rate to a specific value and using the equilibrium condition from equation (10), as we can assume[2] $k_i^+ = 2.0 \cdot 10^7$/(mol/l s) because the on-rate is rather independent of the sequence details. Thus the ΔG_i's for the three hybridization processes determine the overall kinetics.

In this investigation O_1 consists of 8 base pairs and O_2 of 9 base pairs with a one base dangling end. The strategy for achieving the most efficient DNA replication is by dialling in destabilization.[1] Already with these short oligomers the number of possibilities to order base pairs as well as to dial in destabilization is rather high. Therefore we focus on the resulting free energy for the single oligomers as well as for the ligated template and not on the detailed base sequences.

The energy change for two neighboring base pairs at standard conditions ($T = 310.15 \sim 37°\text{C}$) ranges from -2.24 kcal/mol to 1.33 kcal/mol with a initiation cost per oligomer of 1.96 kcal/mol.[8] After ligation we assume the dangling end causes a bulge loop of size one with a corresponding energy cost of 4 kcal/mol.[8] Since the total template concentration in eq. 9, depends on the on- and off-rates, which again are linked to the free energy change (eq. (10)), the template replication efficiency solely depends on the values of ΔG_i.

The difference between the free energy changes of the two oligomers and the ligated total template ΔG_T mainly relies on the cap contribution[8] of O_1, the intervening nearest neighbor base pair[8] of $O_1 + O_2$ and especially on the bulge loop.[8] This relation shall be defined by $\Delta G_T - (\Delta G_{O_1} + \Delta G_{O_2}) \equiv \gamma$. In particular, for $T = 310.15$ K $\gamma > 0$ is always true.

There are usually multiple solutions in terms of the base sequence resulting in the same free energy changes, which means it does not matter whether there e.g. are more AT/TA base pairs or GC/CG pairs combined with internal mismatches.[8] However, what matters is the intervening nearest neighbour base pair of the $O_1 + O_2$ ligation, as it, besides the bulge contribution, decisively influences the energy difference between the single oligomers and the total template. Consequently, the task is to find the optimal values for ΔG_i depending solely on γ.

The equations eq. (4-9) are implemented and simulated by matlab for different on-/off-rate combinations calculated from the respective ΔG_is.

Simulation results

Before investigating our system for various on-/off- and ligation rates, we first present in Fig. 2 a comparison of our simulation approach with the experimental data from Gibbs-Davis et al (2015).[1] Therefore, we use a identical initial template and oligomer concentrations as in the experiment and fix all on-rates i to $k_i^+ = 2.0 \times 10^7/(\text{mol/ls})$ and the ligation rate to $k_L = 0.02/\text{s}$. These same values for the rates are used throughout our simulations unless stated differently. Since the reported experiment was conducted at

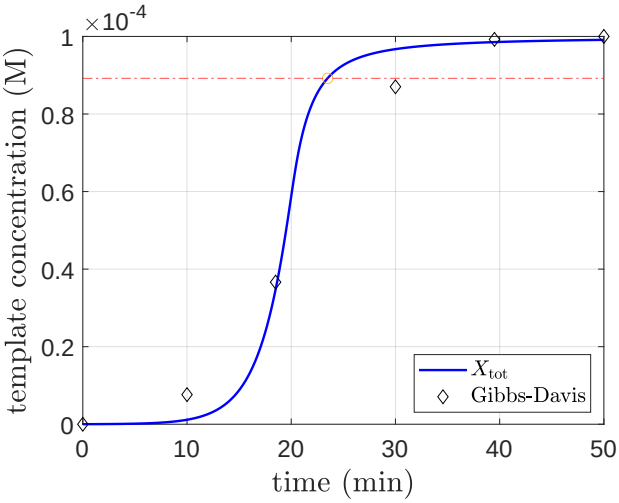

Figure 2: Total template concentration as a function of time for corresponding experimental[1] (black diamonds) and simulated dynamics (blue solid curve). The blue curve is generated using the parameters from the experiment approximating $\Delta G_{O_{1,2}}(26°\text{C}) \approx \Delta G_{O_{1,2}}(37°\text{C}) - 1$ kcal/mol and $\gamma_{26°\text{C}} \approx \gamma_{37°\text{C}}$, see text for details. The dotted red line intercepting the solid red curve characterizes the ratio $[X_{\text{tot}}]/\max\{[X_{\text{tot}}]\} = 0.9$, which is used through out the paper to compare the simulation results. The values used are: $k_L = 0.02$/s, $\Delta G_{O_1} = -6.44$ kcal/mol, $\Delta G_{O_2} = -9.85$ kcal/mol and $\gamma = 3.84$ kcal/mol, which corresponds to $k_{O_1}^- = 394.25$/s, $k_{O_2}^- = 1.27$/s and $k_T^- = 0.016$/s. All on-rates $k_{O_{1,2},T}^+ = 2 \times 10^7$/(mol/l s).

$T = 299.15$ K and the free energies of the base pairs are given at $T = 310.15$ K,[8] we need to translate the free energies between these two temperatures. We can either use an approximation of ΔG_{O_i} by consulting Table 1 in Wu et al[9] or consulting the www.nupack.org DNA/RNA complex free energy calculator.[6] In both cases we find that $\Delta G_{O_{1,2}}(26°\text{C}) \approx \Delta G_{O_{1,2}}(37°\text{C}) - 1$ kcal/mol. Further, we can approximate $\gamma_{26°\text{C}} \approx \gamma_{37°\text{C}}$, as it is defined as a difference between the involved free energies where we assume the oligomers each decrease with $\sim$ - 1 kcal/mol and the double length template decrease with $\sim$ - 2 kcal/mol. Even though these approximations are rather rough, the compari-

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2020/32/269/1908548/isal_a_00312.pdf by guest on 03 August 2024

son in Fig. 2 shows that our approach is compatible with the experimental data.

We now turn to unspecified base pair sequences and therefore a characterization of the oligomers solely by their free energies at $T = 310.15$ K. To have an unambiguous measure distinguishing the total template replication efficiencies for different sets of rates in different simulations, we use the time t where $[X_{\text{tot}}]/\max\{[X_{\text{tot}}]\} = 0.9$ (red dotted line in Fig. 2).

In Fig. 3 the results for $\gamma = 3.89$ kcal/mol are computed for a range of free energy values for ΔG_{O_1} and ΔG_{O_2} in the interval $[-6; -9]$ kcal/mol, which corresponds to off-rates in the interval $\left[1.18 \cdot 10^3; 9.09\right]$/s. The used value for the energy difference γ of ligated template and single oligomers is estimated to be the largest possible with a bulge (loop of size one) arising from the ligation process. Each of the graphs correspond to a given value of ΔG_{O_1}, and its minima indicate the highest replication efficiency and therefore the optimal value not only for ΔG_{O_1} but also for ΔG_{O_2}. The optimal replication efficiency is obtained for $\Delta G_{O_1} \approx -8$ kcal/mol while $\Delta G_{O_1}/\Delta G_{O_2} \approx 1$.

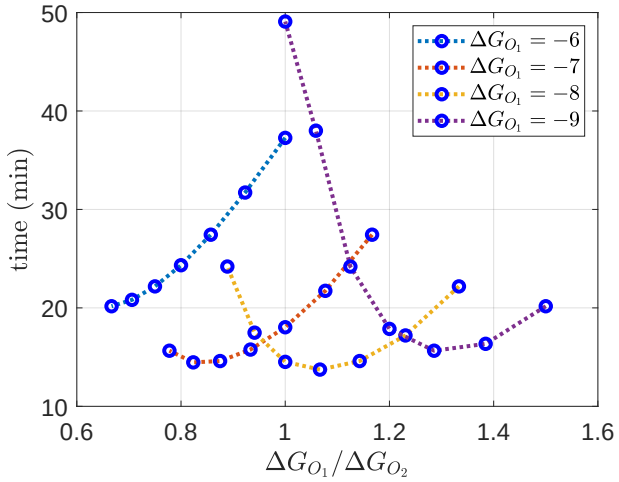

Figure 3: Replication efficiency of the total template concentration for $\gamma = 3.89$ kcal/mol, $k_L = 0.02$/s and all on-rates i $k_i^+ = 2.0 \times 10^7$/(mol/l s).

We can define an envelope curve that connects the minima of the ratio $\Delta G_{O_1}/\Delta G_{O_2}$ with the variable γ. In Fig. 4 the minima of the graphs for variable ΔG_{O_1} (in Fig. 3) are connected, calculated in half integer steps between [-6;-9] kcal/mol for different $\gamma \in \{2.7, 3.89, 5\}$ kcal/mol with ligation rate $k_L = 0.02$/s and all on-rates i $k_i^+ = 2.0 \times 10^7$/(mol/l s).

In any of the three graphs the data points from left to right stand for the minimal ratio of $\Delta G_{O_1}/\Delta G_{O_2}$ with a fixed value for ΔG_{O_1} (varying ΔG_{O_2}) starting with the highest

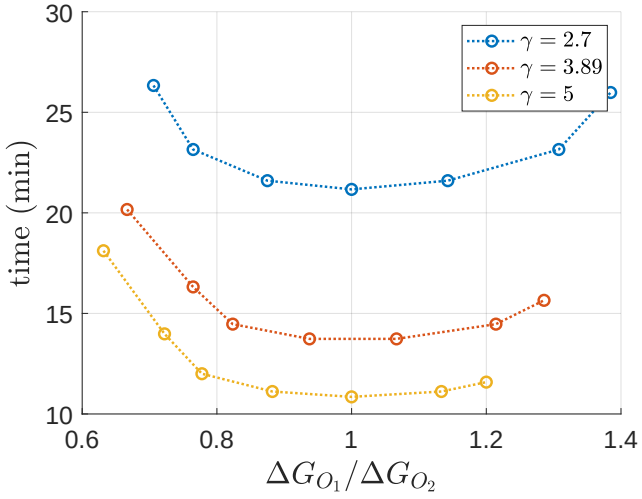

Figure 4: Envelop curves of minina as seen in Fig 3 for different γ values. In any of the three graphs the data points from left to right stand for a minimal ratio of $\Delta G_{O_1}/\Delta G_{O_2}$ with a fixed value for ΔG_{O_1} starting with the highest value $\Delta G_{O_1} = -6$ kcal/mol and decreasing in half integer steps to $\Delta G_{O_1} = -9$ kcal/mol. The two values for γ = 2.7 and 3.89 are physically accessible in the system, while 5.0 is only of theoretical interest. Note that $\gamma = 3.89$ kcal/mol (red curve) with $\Delta G_{O_1} = \Delta G_{O_2} = -8$ kcal/mol predicts significantly faster replication dynamics than achieved experimentally with about 30 minutes versus 14 minutes for reaching 90% completion.

value $\Delta G_{O_1} = -6$ kcal/mol and decreasing in half integer steps to $\Delta G_{O_1} = -9$ kcal/mol. As mentioned before $\gamma = 3.89$ kcal/mol is the largest obtainable for systems with low loop size (< 9 bases) and no hairpins, meaning $\gamma = 5$ kcal/mol is not realistic but of theoretical interest.

First it should be noted that increasing γ means decreasing product inhibition. So for a given ratio $\Delta G_{O_1}/\Delta G_{O_2}$ the replication efficiency increases, indicated by a shift to shorter replication times.

Secondly, it should be noted that the minimum replication time (and thus highest replication efficiency) is indeed maximal for $\Delta G_{O_1}/\Delta G_{O_2} \approx 1$, regardless of γ.

Lastly, for larger γ the optimum in replication efficiency corresponds to lower values in free energy. This is also suggested by a shift of the data points to the left in Fig 4 for increasing γ.

These results, especially that the template replication is most efficient for $\Delta G_{O_1}/\Delta G_{O_2} \approx 1$, imply that the base length of the oligomers is in fact of minor importance as only the free energies and their ratio have an impact. Longer oligomers can be adjusted by the use of AT/TA and internal mismatches instead of GC/CG base-pairs, to achieve the highest replication efficiency for a given γ-value. This

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2020/32/269/1908548/isal_a_00312.pdf by guest on 03 August 2024

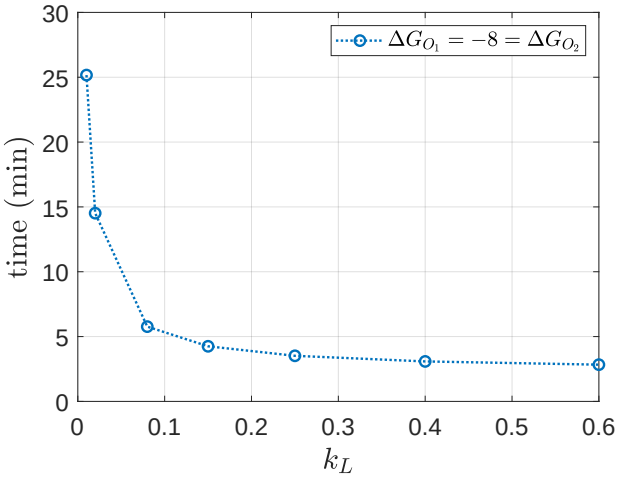

Figure 5: Replication efficiency for $\Delta G_{O_2} = \Delta G_{O_1} = -8$ kcal/mol corresponding to $k_{O_i}^- = 46.06$/(mol/l s) as a function of the ligation rate k_L. All on-rates i are $k_i^+ = 2.0 \times 10^7$/(mol/l s).

means that the template replication presumably could be made significantly more efficient by expanding the oligomer and template length to include larger loops or hairpins.,[8] corresponding to larger γ values. Hairpins are used exactly in this manner by Lincoln and Joyce(2009)[5] in their self-replicating RNA system.

What is left to examine is the influence the ligation rate has on the template replication efficiency. For this purpose the replication time for $\Delta G_{O_2} = \Delta G_{O_1} = -8$ kcal/mol is plotted over various values of the ligation rate k_L as illustrated in Fig. 5. Clearly, higher ligation rates lead to faster replications although the efficiency seems to converge towards a threshold. It should be noted that k_L larger then $\sim 10^{-2}$ are mainly of theoretical interest.

Conclusion

Our investigations of the free energies of oligomer-template interactions and their impact on the reaction kinetics have show us what mainly determines replication efficiency in non-enzymatic, short-chain, nucleotide systems. The replication efficiency does not depend on the details of the base-pair sequences. Product inhibition decreases and can eventually be eliminated when the value of the difference between the free energy of template hybridization and the free energies of the oligomer hybridizations increases as defined by $\gamma \equiv \Delta G_T - (\Delta G_{O_1} + \Delta G_{O_2})$. Desired γ values be obtained by designing the replication systems with bulges and an appropriate balance between AT and CG base pairs as well as mismatches. The bigger the γ the faster the replication. Further, the template replication efficiency has a maximum when the hybridization energies of the two oligomers are identical $\Delta G_{O_1}/\Delta G_{O_2} \approx 1$. Finally, increasing the ligation rate can boost replication efficiency independently of the free energy relationships between the oligomers and the template.

Based on our analysis we predict that a significantly faster experimental kinetics should be achievable for parameters around: $\gamma = 3.89$ kcal/mol with $\Delta G_{O_1} = \Delta G_{O_2} = -8$ kcal/mol, $k_L = 0.02$/s and assuming all on-rates i $k_i^+ = 2.0 \times 10^7$/(mol/l s).

As our results are independent of the amount of base pairs per oligomer, the concept should still be valid for larger system sizes (longer templates and oligomers). Therefore it should be possible to make the replication process even faster by increasing the system size so that larger loops and/or hairpins become accessible.

References

[1] B.Safeenaz Alladin-Mustan, C. J. Mitran and J. M. Gibbs-Davis, Achieving room temperature DNA amplification by dialling in destabilization, *Chem. Commun. (2015)* 51, 9101.

[2] U. Christensen, N. Jacobsen, V. K. Rajwanshi, J. Wengel, and T. Koch (2001) Stopped-flow kinetics of locked nucleic acid (LNA)oligonucleotide duplex formation: studies of LNA-DNA and DNA-DNA interactions, *Biochem. J.* 354, 481-484.

[3] Adi Constantinescu, Carsten Svaneborg and Steen Rasmussen, *preprint*, 2020.

[4] Sievers, D. Von Kiedrowski, G. (1994) Self-replication of complementary nucleotide-based oligomers. *Nature* 369, 221–224.

[5] Tracey A. Lincoln and Gerald F. Joyce (2009) Self-sustained Replication of an RNA Enzyme *Science*. 2009 Feb 27; 323(5918): 1229–1232.

[6] http://www.nupack.org, accessed June 2020.

[7] Steen Rasmussen, Adi Constantinescu and Carsten Svaneborg (2016) Generating minimal living systems from nonliving materials and increasing their evolutionary abilities, *Phil. Trans. Royal Soc. B. Biological Sciences.* 1701, 371.

[8] John SantaLucia, Jr. and Donald Hicks, The Thermodynamics of DNA Structural Motifs (2004) *Annu. Rev. Biophys. Biomol. Struct.* 33, 415-440.

[9] P. Wu, S. Nakano and N. Sugimoto (2002), Temperature dependence of thermodynamic properties for DNA/DNA and RNA/RNA duplex formation.

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2020/32/269/1908548/isal_a_00312.pdf by guest on 03 August 2024