ChemBioChem

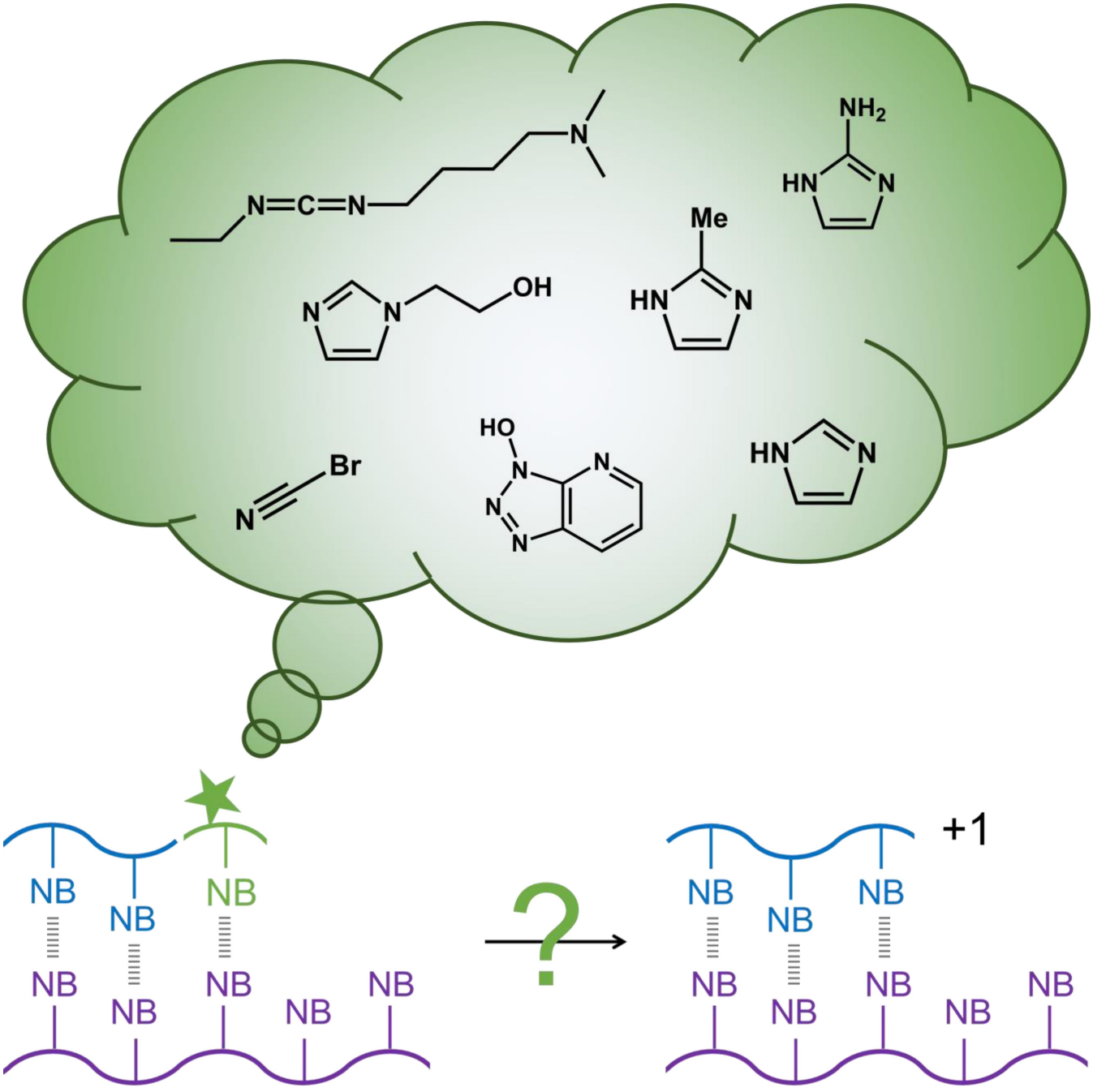
Review
doi.org/10.1002/cbic.202300859

Chemistry Europe
European Chemical Societies Publishing

www.chembiochem.org

The Impact of Activating Agents on Non-Enzymatic Nucleic Acid Extension Reactions

Kimberley L. Callaghan,*[a] Peter C. Sherrell,[a, b] and Amanda V. Ellis*[a]



Non-enzymatic template-directed primer extension is increasingly being studied for the production of RNA and DNA. These reactions benefit from producing RNA or DNA in an aqueous, protecting group free system, without the need for expensive enzymes. However, these primer extension reactions suffer from a lack of fidelity, low reaction rates, low overall yields, and short primer extension lengths. This review outlines a detailed mechanistic pathway for non-enzymatic template-directed primer extension and presents a review of the thermodynamic driving forces involved in entropic templating. Through the lens of entropic templating, the rate and fidelity of a reaction are shown to be intrinsically linked to the reactivity of the activating agent used. Thus, a strategy is discussed for the optimization of non-enzymatic template-directed primer extension, providing a path towards cost-effective in vitro synthesis of RNA and DNA.

1. Introduction

Non-enzymatic template-directed primer extension is an emerging one-pot reaction method of DNA and RNA replication.[1] Performed in an aqueous medium to maintain native RNA and DNA duplexes, non-enzymatic primer extension generates a phosphodiester backbone without the use of protecting groups.[2] This simplicity is an improvement over alternate chemical routes for DNA or RNA synthesis such as oligonucleotide synthesis which uses phosphoramidite coupling chemistry,[3] and requires organic solvents and protecting groups that cannot be tolerated by biological scaffolds.[4] Primarily investigated in the context of prebiotic RNA, non-enzymatic primer extension can also be used to produce DNA, albeit at a slower rate than RNA production.[5]

Using non-enzymatic primer extension offers the potential for synthesis of native DNA and RNA attached to novel scaffolds such as biologicals including proteins or antibodies for hybrid therapeutics,[6] polymers including drug-delivery systems,[7] or other scaffolds such as non-organic materials.[8] These cannot be used as starting points for enzymatic synthesis due to steric effects preventing enzyme access to the reaction site, the enzymes inability to recognize and work with novel substrates which the enzymes have not been evolved to handle,[9] or denaturing effects of the materials.

Non-enzymatic primer extension also offers a simpler route to incorporate unnatural nucleotides or reactive handles into RNA and DNA sequences, which can introduce a wide range of chemical functionalities and the potential for further modification.[10] The use of bio-orthogonal click chemistry to conjugate pre-formed oligonucleotide sequences to these target materials has already been of great benefit.[4c,11] The applications accessed already include research platforms such as DNA-force spectroscopy,[4c] DNA-paint for microscopy[12] and attachment of DNA-tethers to antibodies for labelling or surface tethering.[13] However, these bio-orthogonal reactions introduce non-native chemistries to the DNA or RNA system,[14] which have been shown to be tolerated by cellular machinery only at extremely low levels.[15] By contrast, non-enzymatic primer extension results in native phosphodiester backbone formation, offering the opportunity for *in vitro* replication of native DNA or RNA in a trace-free coupling system.[16] The trace-free material should have superior *in vivo* performance, especially for therapeutics and diagnostics.

Despite promising advances, non-enzymatic template-directed primer extension is still under development, with unresolved issues of insufficient extension length,[17] low rates, low yields of extension,[18] and low fidelity in mixed nucleotide systems.[19] This review explores the fundamental mechanism of non-enzymatic template-directed primer extension and highlights the thermodynamic importance of entropic templating. Using this framework, a pathway towards the optimization of these reactions is discussed and a comparison is offered between the different activating agents routinely used.

2. The Non-enzymatic Primer Extension Reaction

2.1. The reaction system

Non-enzymatic template-directed primer extension is a chemically driven reaction that generates native RNA or DNA sequences. During the reaction the 3′-hydroxyl (OH) group of a primer strand undergoes phosphodiester bond formation with an activated 5′-phosphate group of an incoming nucleotide, to generate an extended primer + 1 nucleotide product.

The primer extension reaction (Figure 1a) has two distinct steps: (1) an initial association of the nucleotide to the primer-template-helper complex (sequence selectivity step) and (2) the phosphodiester bond-forming, extension reaction step between the primer and the activated nucleotide (chemical selectivity step) (Figure 1b),[20] to give the primer + 1 extension product.

2.1.1. Step 1 – Association and Sequence Selectivity

The first step of non-enzymatic template-directed primer extension is nucleotide association (Figure 1a). This process starts with an 'empty complex', consisting of a template

[a] *Dr. K. L. Callaghan, Dr. P. C. Sherrell, Prof. A. V. Ellis*
Department of Chemical Engineering, The University of Melbourne, Parkville, Victoria, 3010, Australia
E-mail: kimberley.callaghan@unimelb.edu.au
amanda.ellis@unimelb.edu.au
[b] *Dr. P. C. Sherrell*
School of Science
RMIT University, Melbourne, Victoria, 3000, Australia

(purple) and primer (blue) which are non-covalently bound through complementary Watson-Crick interactions. A helper strand (pink) can be used, which binds to the template strand downstream of the reaction site but is otherwise not directly involved in the chemical reaction. This binding creates a "gap" on the template bound duplex for the activated nucleotide to bind. The empty complex shown (Figure 1a, i) has an unoccupied one nucleobase gap on the template strand, this gap is the site of association for an activated nucleotide. For sequence selectivity to be achieved, the incorporated nucleotide must be complementary to the template, defined as a high fidelity reaction.

If the incoming activated nucleotide is complementary to the template strand (Figure 1a, ii, NB_c, green), it will associate through Watson–Crick base pairing to the template strand and through non-specific π–π stacking interactions with the bordering primer and helper strands, to form the 'bound complex' (Figure 1a, ii, left). If the incoming nucleotide is mismatched (Figure 1a, ii, NB_m, grey) then the Watson-Crick base pairing rules will not be met, resulting in minimal hydrogen bonding interactions with the template strand (Figure 1a, ii, right). The non-specific π–π stacking interactions with the bordering primer and helper strand will still be present in the mismatched 'bound complex'.

As the primer extension reaction (phosphodiester bond-formation) has not yet occurred, the 'bound complex' exists in an equilibrium with the 'empty complex', with the position of the equilibrium dictated by the association rate (k_{on}) and dissociation rate (k_{off}). The magnitude of these rates is expected to follow the binding strength between nucleobases, with complementary base-pairs (sequence selectivity) favouring the bound complex at equilibrium through a high association rate ($k_{on\text{-}c}$) and a low dissociation rate ($k_{off\text{-}c}$). For mismatched base-pairs, the weaker interactions between the activated nucleotide and the template strand are expected to increase the dissociation rate ($k_{off\text{-}m}$), resulting in dissociation back to the 'empty complex'.

The desired outcome from the association/dissociation equilibration is a high concentration of the complementary bound complex population, with minimal mismatched bound complex present. The rate balance between complementary base pair association ($k_{on\text{-}c}$) and mismatch dissociation ($k_{off\text{-}m}$) is crucial to obtaining high fidelity primer extension reactions. These bound complexes can then undergo the primer extension reaction, step 2 (Figure 1a) of the process.

2.1.2. Step 2 – Reaction and Chemical Selectivity

The second step of non-enzymatic template-directed primer extension is the phosphodiester bond forming reaction. The reaction proceeds from the 'bound complex' (Figure 1a, ii), with phosphodiester bond formation between the primer and the incoming activated nucleotide forming a primer + 1 extension product (Figure 1a, iii). The formation of the phosphodiester bond involves nucleophilic attack of the 3′-hydroxyl group of the primer to the activated 5′-phosphate of the nucleotide (Figure 1b, c), suggested to proceed via an S_N2 mechanism with activation of the 3′-hydroxyl through magnesium ion binding.[20–21]

A range of activating agents have been used to promote the primer extension reaction (Figure 1c), including cyanogen bromide (CNBr),[4b] 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)[22] and the addition of organocatalysts, including 1-hydroxyethylimidazole (HEI)[23] or pyridine[24] which displace the activating agent on the nucleotide before undergoing the extension reaction in an energy

Dr. Kimberley L Callaghan is a Biophysical Chemist with a focus on building a chemical understanding of biological systems to enable their use and modification for fundamental research and applied technologies or therapeutics. She received her Ph.D. from the University of Cambridge in 2021 after completing her BSc (Honours) between the University of Melbourne and the Walter and Eliza Hall Institute in 2016. Kimberley is currently a Research Associate in DNA/polymer hybrid materials within the Department of Chemical Engineering at The University of Melbourne.

Dr. Peter C Sherrell is a Materials Chemist with a focus on using controlling kinetic energy to enable green chemical synthesis. He received his Ph.D. from the University of Wollongong in 2012. He is currently an Elizabeth & Vernon Puzey Research Fellow, and Chemical Engineering Teaching Fellow at The University of Melbourne. He was previously a Marie Sklodowska-Curie Individual Fellow in the Department of Materials and Imperial College London (2015–2018), and a Research Fellow at Linköping University (2013–2015).

Prof. Amanda V Ellis is the Head, School of Chemical and Biomedical Engineering at The University of Melbourne. She graduated with a Ph.D in 2003 from the University of Technology, Sydney. She became a Full Professor (2013), and an Australian Research Council (ARC) Future Fellow (2014). She then was the Head, Department of Chemical Enginnering at The University of Mebourne in 2019–2022. Her research work focussing on the fundamentals and applications of DNA nanotechnology and genotyping, polymer membrane science, and energy storage/ harvesting.

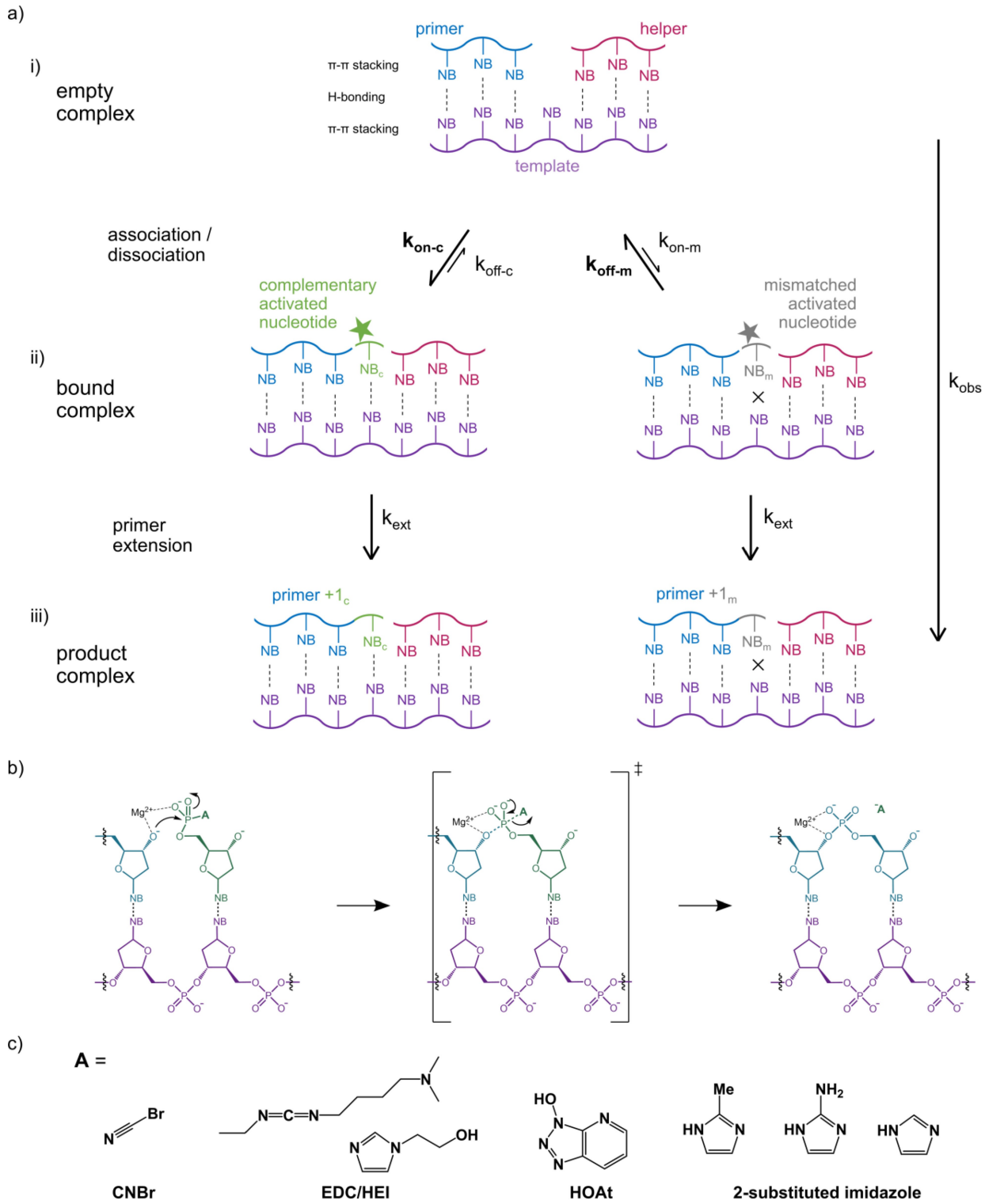


Figure 1. Mechanism of non-enzymatic template-directed primer extension reaction on a primer-template-helper system. a) Pathway of non-enzymatic template-directed primer extension in a mixed nucleotide system. Extension of a primer with complementary base pair (NBc) or mismatching base pair (NBm) incorporation to form the extended primer + 1 product. Wavy line represents the sugar-phosphate backbone of RNA or DNA, with pendant nucleobases (NB). Dotted lines show hydrogen bonding between strands and cross shows mismatching base pair with no canonical hydrogen bonding. b) Schematic of the non-enzymatic primer extension mechanism with a nucleotide activated at the 5′-phosphate group (activating agent (A)) and a 3′-OH of a primer (hydroxyl group as the nucleophile) activated by co-ordinated magnesium ion ($Mg2+$), reacting to form a phosphodiester bond via a pentaphosphate transition state (‡). c) Activating agents (A) used through literature, cyanogen bromide (CNBr), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), 1-hydroethylimidazole (HEI), 1-hydroxy-7-azabenzotriazole (HOAt), and 2-substituted imidazoles (imidazole, 2-methylimidazole, 2-aminoimidazole).

cascade pathway.[25] The use of 1-hydroxy-7-azabenzotriazole (HOAt),[26] has also been widely studied.

Another class of activating agents which have been studied are the 2-substituted imidazole variants,[27] which have been proposed to form a more reactive dimer in solution.[20] For the 2-substituted imidazole nucleotide variants a dinucleotide gap on the template strand (Figure 1a, i) is required for dimer binding.

2.2. The difference between activating agents

Though each activating agent has a different mechanism of nucleotide activation, the mechanism of primer attack on the activated nucleotide (Figure 1b) remains comparable between these different activating agents. As such, the difference in activating agents is their chemical reactivity. This intrinsic reactivity impacts the rate at which the activated nucleotides undergo the primer extension reaction (k_{ext}) (Figure 1a, ii–iii) and the rate at which competing side reactions occur.

2.2.1. Impact on Sequence Selectivity

When the primer extension reaction rate (k_{ext}) is combined with the association (k_{on}) and dissociation (k_{off}) equilibrium, the overall observed reaction rate (k_{obs}) can be determined. The reaction rates reported throughout the literature for primer extension are generally k_{obs} rates. These determine the potential yield of the system and the rate at which this yield can be achieved.[1] However, minimal studies on the reaction rate of the involved steps makes optimization of the primer extension reaction difficult.

These bulk level descriptions are particularly problematic for improving the fidelity of primer extension. To prevent incorporation of mismatches into the primer strand to obtain a high fidelity reaction, mismatch dissociation must occur faster than primer extension can proceed, such that $k_{off\text{-}m}$ must be faster than $k_{ext.}$ This suggests that slowing k_{ext} could be used to increase fidelity, however, if k_{ext} is too slow, complementary nucleotide dissociation, $k_{off\text{-}c}$, will occur prior to any primer extension reaction and thus no extension will be observed. Unfortunately, fine tuning this balance of the relative reaction rates, such that $k_{off\text{-}m} > k_{ext} > k_{off\text{-}c}$, is difficult to achieve.

The lack of a focus on fidelity is likely due to the experimental setups and analysis methods employed throughout literature. Experimental setups which focus on the optimization of reaction conditions may only employ a single nucleotide variant,[16,21c] removing fidelity from the system. Similarly, the predominant method of experimental analysis is gel electrophoresis,[28] which provides no information on the exact identity of the nucleobase incorporated at each extension step. This is likely due to gel electrophoresis being much more accessible, straightforward, and rapid to perform than techniques such as MALDI[19,26] or nanopore sequencing[29] which would give sequence information and therefore information on fidelity. This bias in analysis was highlighted by Duzdevich and co-workers in their work to generate a deep sequencing methodology for primer extension.[29b] In order to develop a viable primer extension technique, the use of mixed nucleotide systems and an analysis of their fidelity will be necessary.

2.2.2. Impact on Chemical Specificity

The rate of non-enzymatic template-directed primer extension can be extremely slow, with reactions taking days[22] to weeks.[1] Long reaction timescales can result in off-target, non-templated side reactions accumulating.

The side reactions observed within non-enzymatic primer extension include hydrolysis of the activated nucleotides[23] (Figure 2a), for example of HOAt-activated monomers (Figure 2a, monomer hydrolysis) or 2-aminoimidazole[20] or 2-methylimidazole[21c] activated dimers (Figure 3a, dimer hydrolysis). These reactions deplete the nucleotide feedstock and introduce hydrolyzed nucleotides that can bind to the template strand and block primer extension (Figure 2c, template bound hydrolyzed nucleotide).[21c,30] Hydrolysis of the template, primer and helper strand can also occur. This is more common for RNA

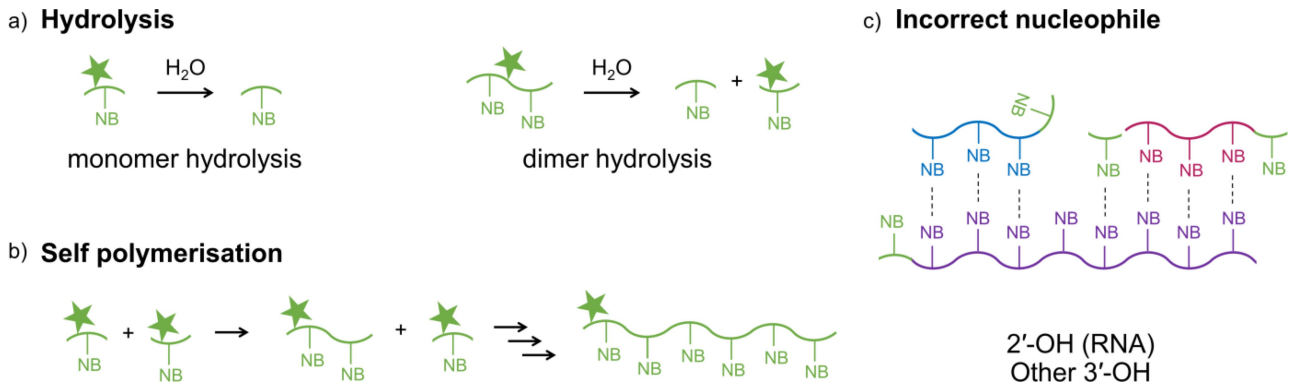


Figure 2. Chemically off-target reactions of the activated nucleotide which can occur during non-enzymatic template-directed primer extension, including a) hydrolysis of the activated monomer or dimer, b) self-polymerization of the activated nucleotide, and c) off-target reactions with an incorrect nucleophile in the system, for a lack of regioselectivity involving reaction with the 2′-OH of RNA or other 3′-OH groups, rather than the desired primer 3′-OH group. Also shown in c) is binding inhibition from the hydrolyzed monomers.

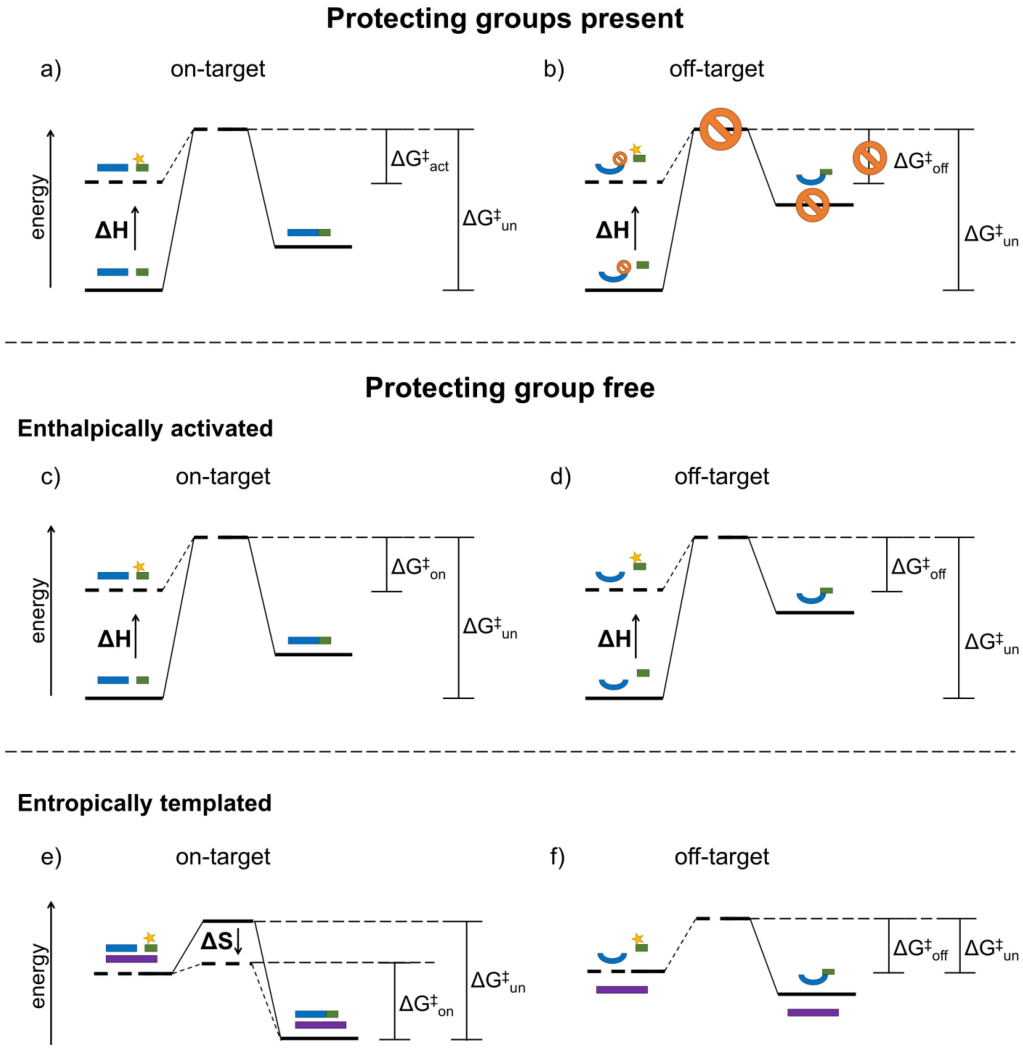


Figure 3. Alternate energetic activation mechanisms for achieving chemical specificity in systems with or without protecting groups present. a) The activation energy of an activated ($\Delta G^{\ddagger}_{act}$) reaction pathway (dashed line) relative to an unactivated ($\Delta G^{\ddagger}_{un}$) reaction pathway (solid line) of non-specific enthalpic (ΔH) activation. b) The off-target ($\Delta G^{\ddagger}_{off}$) reaction pathway is also activated enthalpically, but the reaction pathway is blocked by a protecting group (orange), preventing off-target reactions. In protecting group free reactions, the activation energy of on-target ($\Delta G^{\ddagger}_{on}$) and off-target ($\Delta G^{\ddagger}_{off}$) reaction pathways (dashed-line) relative to an unactivated ($\Delta G^{\ddagger}_{un}$) reaction pathway (solid line) of non-specific enthalpic (ΔH) activation on c) on-target and d) off-target pathways, compared to templated entropic (ΔS) activation of e) on-target and f) off-target pathways. Reaction components: reagent (green square), on-target reactant (blue rectangle), off-target reactant (curved blue rectangle), activating agent (star), template (long purple rectangle), and protecting group (orange block).

non-enzymatic template-directed primer extension, especially at elevated pH levels (pH 8–9).[1,31]

Self-polymerisation of activated nucleotides is another off-target reaction. Here, the free 3′-OH of one activated nucleotide monomer reacts with the activated 5′-phosphate of another nucleotide monomer resulting in oligomer formation in solution and depletion of the monomer feedstock (Figure 2b).[22] These oligomers have a large template binding capacity, so can be potent inhibitors of homopolymeric templates.[22]

Reactions with the incorrect nucleophile can also occur when the DNA or RNA strands are not complexed in the double-stranded helix complex. This includes nucleophilic attack from the terminal 3′-OH of the template or helper strand (Figure 2c), or non-templated extension reactions between strands and nucleotides not bound to the template strand but instead existing free in solution. Additionally, RNA can undergo non-regioselective primer extension involving nucleophilic attack of the 2′-OH of RNA to the activated nucleotide.

Decreasing these side reactions is a target of significant research,[32] but is unfortunately complicated by the reaction system conditions, including the aqueous reaction environment, lack of protecting groups, and long reaction times from low intrinsic reactivities of the activated nucleotides.

3. Chemical Specificity from Entropic Templating

3.1. Promoting a chemical reaction

The process of promoting a desired reaction pathway involves improving its Gibbs free energy (ΔG). The Gibbs free energy (ΔG) has an enthalpy (ΔH) and an entropy (ΔS) component (Eq. 1). The more negative the ΔG of a reaction, the larger the energy difference between the reagents and products and therefore the more favourable a specific reaction is over an alternate pathway.

$$\Delta G = \Delta H - T\Delta S \quad [1]$$

The process of obtaining chemical specificity is arguably better described through the relative rates of alternate reaction pathways. The activation energy ($\Delta G^{\ddagger}$) gives the energy difference between the starting material and the transition state ($\ddagger$) and dictates the number of successful reactions which will occur over a given timeframe. This determines how fast the reaction will proceed, and the overall yield obtained. Contrary to the ΔG, a smaller $\Delta G^{\ddagger}$ will result in a more favourable reaction pathway.

To decrease the $\Delta G^{\ddagger}$ of a reaction, the enthalpy (ΔH) of the reagent can be increased with the use of an activating agent (Figure 3a, yellow star). The introduction of these groups raises the intrinsic energy of the activated material, resulting in a smaller activation energy ($\Delta G^{\ddagger}_{act}$) (Figure 3a, dashed line), compared to the larger activation energy for the unactivated pathway ($\Delta G^{\ddagger}_{un}$) (Figure 3a, solid line).

For chemical reaction systems, increasing the ΔH as much as is feasible is favoured for increasing the reaction rate and yield, often with the use of halogen groups under anhydrous and anoxic conditions. The resulting decrease in $\Delta G^{\ddagger}$ occurs regardless of whether that reaction is on-target ($\Delta G\ddagger$on, Figure 3a) or off-target ($\Delta G\ddagger$off, Figure 3b), resulting in activation of all reaction pathways the reagent is involved in.

Introducing specificity within such highly reactive systems is a process of excluding alternate reaction pathways (Figure 3b). This can be achieved through the introduction of protecting groups to block alternate pathways in the system (Figure 3b, orange).

3.2. Promoting a chemical reaction in a biological system

For biological systems such as DNA, there are two issues which must be overcome to promote the desired reaction pathway. The first is the water-based reaction environment and the second is the lack of protecting groups.

3.2.1. Protecting group free chemistry in a biological system

One of the distinct advantages of non-enzymatic primer extension is the lack of protecting groups. This simplifies the chemical reaction pathway by enabling a one-pot reaction system without the cyclic deprotection steps found in solid-phase synthesis. This is necessary for using non-enzymatic primer extension in biological systems where the use of protecting groups is not a feasible strategy, due to the complexity of the systems and the difficulty in purification of the products due to the difference in scale of biomacromolecules (kDa–mDa size range) and chemical groups (Da–kDa size range).

The lack of protecting groups is problematic, as a high intrinsic reactivity (ΔH) of the activated nucleotide will equally promote the on-target primer extension reaction ($\Delta G^{\ddagger}_{on}$, Figure 3c) and any off-target reaction pathways ($\Delta G^{\ddagger}_{off}$, Figure 3d) such as hydrolysis.[30]

3.2.2. Aqueous phase chemistry in a biological system

For DNA and RNA an aqueous environment must be used to maintain the thermodynamic favourability of the double helix to ensure base pairing is present for fidelity. However, the presence of water is a major obstacle to chemical specificity, as water can hydrolyze the activated nucleotide, resulting in a depletion of the activated nucleotide concentration and an increase in the hydrolyzed monomer inhibitor concentration.

There are two options to overcome the issue of hydrolysis. One option is to use bio-orthogonal reaction systems, such as click chemistry reactions.[4c,14] These reaction pathways are less liable to hydrolysis and are unique within biological systems, effectively removing the off-target reactions ($\Delta G^{\ddagger}_{off}$, Figure 3d) and obtaining specificity in this way. Whilst click chemistry has led to many advances, the resulting chemical linkages are generally not found in nature. As such, the resulting click products may have different biophysical properties and behaviours, may not be recognised by enzymes as a substrate for replication or transcription, and may not be degraded appropriately by the body which can result in toxic accumulation.

Therefore, the second option for overcoming hydrolysis is more favourable, which is to minimise the ΔH activation and instead utilise entropic (ΔS) activation to promote the desired reaction pathway.

3.3. Entropically promoting a desired reaction pathway

Activating a reaction entropically (ΔS) is more complicated and less well understood than enthalpic activation (ΔH) but is a crucial mechanism in both chemical and sequence specificity in non-enzymatic primer extension.

The entropy of a reaction (S) is described by the accumulation of reactive microstates (W), proportional to Boltzmann's constant (k_B) (Eq. 2).

$$S = k_B \bullet \ln W \quad [2]$$

For non-enzymatic primer extension, this microstate accumulation comes in the form of entropic templating. Entropic templating occurs when a reaction system is held in the appropriate conformation (microstate) for a reaction to proceed, over prolonged time periods. This presents several opportunities for the reagents to interact (potential microstates), until an interaction in which the desired reaction proceeds (productive microstate). Importantly, the entropic templating mechanism both increases the local concentration of the activated nucleotide through the template binding interaction and introduces a pre-organisation effect by holding the geometry of the reaction site in the correct orientation for the primer extension reaction to proceed. This effect is similar to that of enzyme active sites, which entropically activate their catalysed reaction. The presence of the entropic activation mechanism works cumulatively with the ΔH boost of the activated nucleotide to decrease the $\Delta G^{\ddagger}$ of the templated reaction ($\Delta G^{\ddagger}_{on}$) (Figure 3e, dashed line), relative to the pathway which is only enthalpically activated ($\Delta G^{\ddagger}_{un}$) (Figure 3e, solid line). Importantly, entropic templating is only active in a templated reaction, thus this mechanism will only enhance the rate of the $\Delta G^{\ddagger}_{on}$ (Figure 3e), with no enhancement of the $\Delta G^{\ddagger}_{off}$ (Figure 3f). This distinction introduces a mechanism for chemical specificity in the primer extension reactions, favouring the desired primer extension reaction over any other off-target non-templated reactions (Figure 3e). However, entropic templating is limited by the time required for adequate microstate accumulation of the bound complex state to occur, given the low intrinsic reactivity (ΔH) of the activated nucleotides to minimise nucleotide hydrolysis. When combined with the balance of k_{ext} and k_{off}, the rate of non-enzymatic primer extension is slow, leading to the reaction timescales of days to weeks.[39]

4. Effects of Entropic Templating

Despite not being identified in previous studies there is extensive evidence of entropic templating promoting non-enzymatic template-directed primer extension. The intricacies and impact of this mechanism have thus far not been delineated or investigated. The following section highlights some of the reaction idiosyncrasies which arise from the entropic templating mechanism. The impacts on rate, yield, and fidelity are also commented upon, alongside suggestions of factors worth considering in future optimization efforts. These effects are summarized in Figure 4, with discussion on each concept found within the text.

It is important to note that the reaction values presented here are from isolated studies. Whilst these values can be compared within a study, they cannot be directly compared between studies. This is due to extensive differences between studies, including reaction conditions (buffer systems including Mg^{2+} and pH, temperature, time), activating agents (EDC/HEI, HOAt, Me–Im, NH_2–Im), reaction systems (DNA, RNA, 3′-NH_2-DNA, LNA, DNA origami), cofactors (helper present or absent, catalysts), nucleotide systems (single monomer system, mixed nucleotide solution with single monomer addition, dimers, activated oligomers) and parameters measured (yield over what time period and measured by gel or MALDI, rate including k_m or k_{max}, fidelity). As such, whilst trends can be observed across the literature, which are commented on here, a systematic study and comparison of rates between works remains difficult.

4.1. Nucleotide incorporation correlates with template occupancy

A tighter nucleotide binding to the template complex improves the reaction outcome, through potential microstate accumulation.

The rate and yield of primer extension varies dramatically with nucleotide identity, particularly for 2-aminoimidazole activated dimers, with a reaction rate of 0.35 h^{-1} for U*U dimers up to 84 h^{-1} for C*G dimers.[33] This variation is despite the nucleophilic attack of the primer 3′-OH being the rate limiting step for all activating agents.[34] The higher incorporation rate observed with the C*G dimers is a result of higher template occupancy from a higher k_{on} rate and a lower k_{off} rate. The greater k_{on} versus k_{off} results in more microstates of the 'bound complex' existing and therefore more opportunities for extension to occur.[35]

Higher template occupancy can result from several factors. The first factor is the strength of the hydrogen bonding interaction between the incoming activated nucleotide and the template. Over two times higher incorporation of C and G nucleotides,[20] which form 3-hydrogen bonds, is observed during template-directed primer extension compared to the lower incorporation observed for A, U and T nucleotides which only form 2-hydrogen bonds.[33,36] The stronger template binding of C and G results in a strong sequence bias towards C and G, with mismatches involving incorporation of these two nucleobases being more common when sequence control is lower.[29b,37] The second factor is the π–π stacking interaction between the incoming activated nucleotide and adjacent strands. Purines (G, A) have a larger surface area than pyrimidines (C, T, U), resulting in greater π–π stacking and therefore higher rates of incorporation.[33] This results in a higher rate of incorporation for A than T and particularly U which has very poor π-stacking.[28]

Template binding effects also impact the prevalence of wobble-base pairs being incorporated. Wobble-base pairs involve hydrogen bonding interactions between two nucleotides that do not follow Watson–Crick base pair rules. The main wobble-base pairs present in non-enzymatic template-directed primer extension are A:C and either G:U in RNA or G:T in DNA. The presence of a hydrogen bonding interaction between the nucleotide and the template strand results in a lower k_{off} rate for wobble-base pairs than complete mismatches. As such, these pairs are incorporated at a higher rate than background mismatches, impacting fidelity.[37] The G:U and G:T wobble-

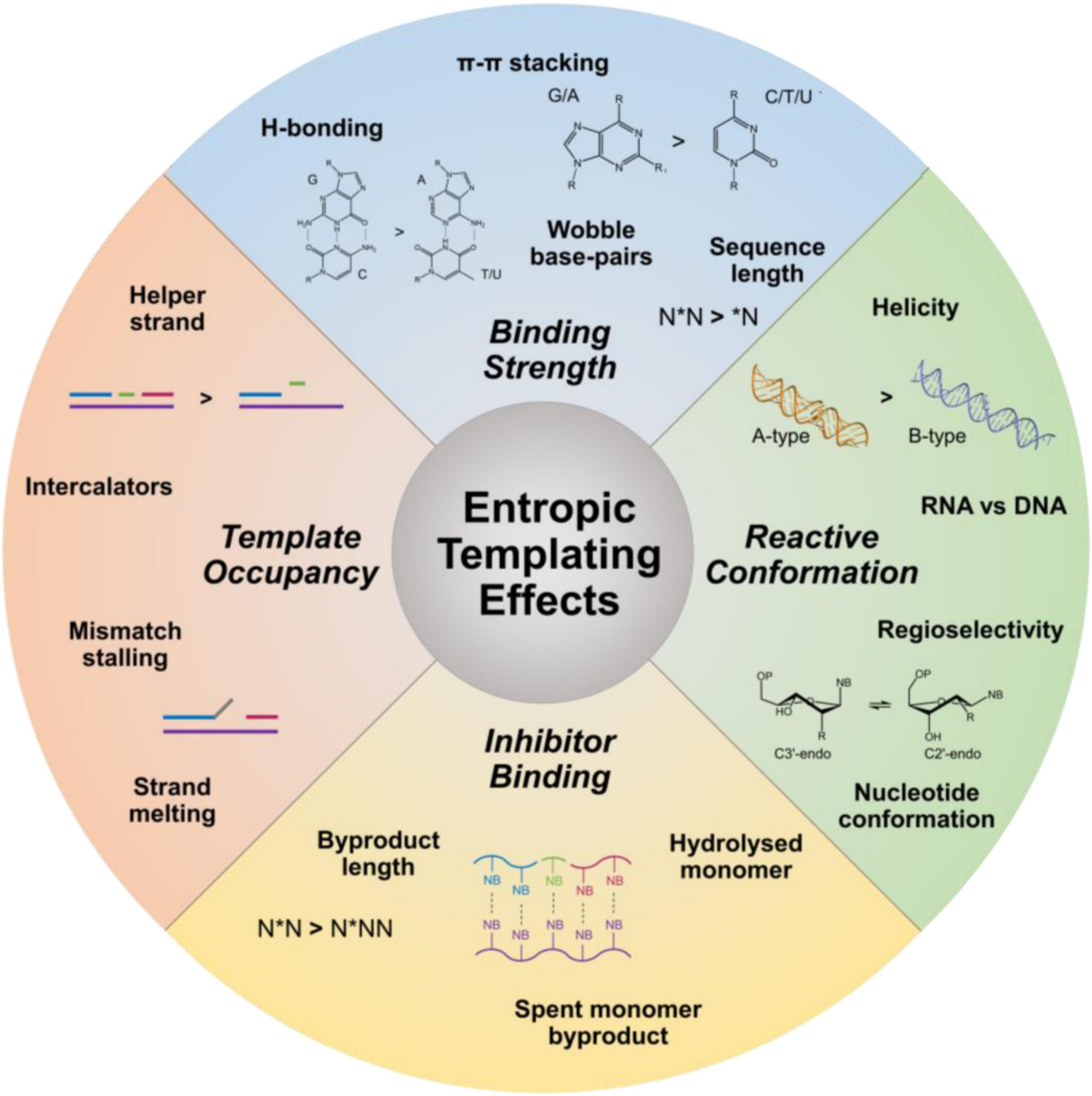


Figure 4. A summary of the effects of entropic templating found throughout the literature on non-enzymatic template-directed primer extension. Effects include 'Binding Strength', in which monomers with a higher binding strength show a greater reaction rate, shown through a higher π–π stacking surface, a greater number of hydrogen bonds including the dominance of specific wobble base-pairs, and the higher reactivity of dimers over monomers. Components which increase 'Template Occupancy' increase the number of potential microstates, including the introduction of the Helper strand or intercalators to increase π-stacking, and the impact of complex stability with an increased rate at lower temperatures from minimal strand melting and a decreased rate from mismatch stalling. 'Inhibitor Binding' from spent monomer or oligomeric by-products or from hydrolyzed monomers also decreases the number of potential microstates. The 'Reactive Conformation' of the system refers to the structural requirement of the S_N2 reaction, including the nucleotide conformation and helicity which result in the higher observed reactivity of RNA over DNA and the resulting regioselectivity of the primer extension given the structural templating effects present.

base pairs have been reported to have the highest impact on fidelity, due to their strong binding interaction.[30] The G:T wobble pair displays a more favourable binding affinity (K_d) than the canonical A:T pairing and has a binding force only 10 times weaker than the canonical G:C base pair, making avoiding mismatch incorporation very difficult.[19,38] One suggestion for overcoming the G:U and G:T wobble-base pairs is to use non-native nucleotides, specifically 2-thiouridine (sU) and 2-thiothymine (sT).[38b,39] These have been reported to have weaker binding affinity to G,[27b] suppressing G:U and G:T mismatch formation by increasing the k_{off} rate. Additionally, sU and sT have stronger reported binding to A,[38b] this decreases the sU:A and sT:A k_{off} rate, increasing the amount of bound complex present. As such, there is less opportunity for C:A binding to occur, supressing C:A mismatch formation. Although a non-natural nucleotide, the use of sU and sT within non-

enzymatic primer extension reaction may still be interesting for applications where the biological degradation is not an issue.[33]

4.2. An organised primer-nucleotide-template complex

4.2.1. A stable complex

Reaction conditions or components which stabilize the primer-nucleotide-template complex show an increase in reaction rate and yield, due to the entropic templating mechanism.

Usually, a decrease in reaction temperature results in a slower reaction rate as components are moving slower with less energy. The opposite is true in non-enzymatic primer extension, with an increase from 24% to 49% incorporation being observed when decreasing the temperature from 22 °C to 4 °C.[40] Similarly, higher reaction rates have been observed at 0 °C rather than at higher temperatures.[22] This result only makes sense if k_{ext} is rate limiting due to the need to accumulate microstates of the bound complex. The concentration of the bound complex would be increased at lower temperatures due to the lack of strand melting and dissociation which occurs at higher temperatures.

Mismatch incorporation has also been shown to result in a strong stalling effect.[41] This is likely due to a weaker association between the strand termini at the mismatch location, resulting in strand melting. This decrease in the concentration of the primer-template complex reduces the number of empty complex microstates for the nucleotide to bind,[19] resulting in a lower number of bound complex microstates and a lower overall reaction rate.[18]

The addition of small molecules, such as the intercalator ethidium bromide, can also increase the stability of the primer-nucleotide complex, enhancing the rate of reaction 25-fold.[21a] An interesting avenue for investigation in increasing the effective concentration or strength of the primer-nucleotide complex is the introduction of confinement effects.[17,42] This could be achieved with the addition of crowding agents or through confinement of the reaction within liposomes to increase the local concentration of reagents.[43] Freeze-thaw cycling has also been shown to effectively increase conversion through increasing the effective concentration of activated nucleotides at the reaction site and may be an interesting avenue for further investigation.[44]

4.2.2. The Helper strand

The helper strand has no covalent involvement in the non-enzymatic primer extension reaction, but it can greatly accelerate the rate of reaction, increasing reaction rates by over an order of magnitude from 0.016 to 0.11/h with a preformed 2-aminoimidazole activated G*G dimer.[45] This rate acceleration allows incorporation of dimers such as the 2-aminoimidazole activated U*U dimer which are nearly undetectable without the helper strand, but with the helper strand rates up to 5.7 h^{-1} can be observed.[33]

The helper strand can accelerate the rate of primer extension by 3 orders of magnitude,[46] through two main mechanistic effects. The first, is through the provision of an additional π-stacking surface for the incoming nucleotides, such that helpers with terminal purine bases, particularly G, offer the greatest rate acceleration effect through a greater increase in template occupancy.[33,46] The second, is through a conformational preorganisation effect, as the helper strand holds the nucleotide in an optimal position for attack of its activated 5′-phosphate from the primer 3′-OH group. This is shown through the higher degree of crystal ordering in the nucleotide dimer when a helper strand is present.[47]

A similar templating effect also assists in directing the desired native phosphodiester bond formation. When template bound, a catalytic Mg^{2+} ion is complexed by the DNA or RNA backbone, promoting reaction between the 3′-OH and 5′-phosphate, increasing the regioselectivity over alternate linkages such as the 2′-OH of RNA which occur in non-templated reactions.[21b]

However, the helper strand presence blocks primer extension beyond single nucleotide addition. This is particularly problematic for systems in which very limited options exist to accelerate the reaction. For example, the use of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and 1-ethylimidazole (HEI) with C or G on an RNA template only results in a single extension event when the helper strand is not present.[36] To gain the benefit of a helper strand, a suggestion has been to use strand displacement to remove the helper strand after each round of nucleotide addition,[48] thus preventing it from blocking extension. The template strand can potentially be bead-bound, to facilitate removal and washing away of the unbound helper strand.[24] The logistics of such a systems remains to be proven.

4.3. Role of template and nucleotide conformation and their impact on reactivity

4.3.1. Reactivity difference of DNA and RNA

One of the more interesting quirks in non-enzymatic template-directed primer extension in the reactivity difference between RNA and DNA.

RNA has been shown to react more rapidly than DNA in non-enzymatic template-directed primer extension.[26,28] The difference in reactivity was observed for extension off a DNA primer occurring at only 5.3 h^{-1} compared to 60 h^{-1} for an RNA primer.[34] The difference in reaction rates was also observed for RNA or DNA nucleotides, with incorporation of a 2-aminoimidazole bridged G*G dimer proceeding at 44 h^{-1} with an RNA G*G dimer compared to only 24 h^{-1} with a DNA G*G dimer.[33] The effect is also observed when DNA is used as a template, with slower primer extension rates for DNA observed than for RNA, for the same primer and nucleotide system.[28]

The presence of an additional hydroxyl group at the 2′-position in RNA offers an additional nucleophile for primer extension. However, there is the strong directing effect of the

primer-template complex which promotes extension from the 3′-position and so would dilute the benefit of the second nucleophile of the 2′-OH.[21b] The presence of the 2′-OH group has been shown to impact the electronegativity of the 3′-OH of RNA, rendering it more reactive than the 3′-OH of DNA.[34] Yet the higher electronegativity is not sufficient to explain the scale of the observed reactivity differences.

As such, the chemical reactivity differences between DNA and RNA are not sufficient to explain the observed reactivity differences.

4.3.2. Impact of helix and nucleotide conformation on reactivity

There is increasing evidence to support the reactivity difference between RNA and DNA being due to a structural difference in the helix and nucleotide conformation adopted by each system.[19,49] rather than the chemical differences between the deoxyribose and ribose sugars.[28] The (deoxy)ribose sugars of DNA and RNA are pentose sugars and so do not sit in a planar conformation in solution. Instead, a puckered conformation is adopted, with the C2′-endo and C3′-endo puckers being most commonly adopted.[50] The conformation of the sugar ring impacts the structure of the DNA or RNA helix formed, due to changes in positioning of the 3′-hydroxyl and 5′-phosphate groups which form the phosphodiester backbone. In the C2′-endo conformation the 3′-OH and 5′-phosphate groups are axially positioned, resulting in a thin elongated helix known as the 'B-type' helix. In the alternate C3′-endo conformation the orientation of the 3′-OH and 5′-phosphate groups is switched to equatorial. As a result, a wider and shorter helix is formed, labelled as the 'A-type' helix. In the A-type helix conformation the 3′-OH is positioned close to the activated 5′-phosphate, facilitating a rapid primer extension reaction, whilst in the B-type helix the orientation and distance of these groups is suboptimal for the extension reaction to proceed. This reactivity difference is important as DNA preferentially adopts a B-type helix but can adopt an A-type helix under the influence of solvent conditions[51] or binding factors.[49] However, steric occlusion of the 2′-OH in RNA results in the A-type helix being preferentially adopted.

The observed rate of non-enzymatic template-directed primer extension has been shown to scale with the degree of A-type helicity.[52] Reactions were fastest on a Locked-DNA (LNA) template which is covalently locked in an A-type helix, taking 10 min to reach 98% conversion from the primer to the primer + 1 product, compared to an RNA template which took 3 h to reach 98% conversion, with the much slower DNA template taking 3 h to reach only 68% conversion.[28]

The conformational effect also extends to the nucleotide component. The nucleotide conformation must be compatible with the primer-template-helper complex for binding to occur.[35,53] The sugar-ring conformation can flip between C2′-endo and C3′-endo to match the template,[35] however a preference for the C3′-endo conformation facilitates rapid binding. As such, the conformational preference of nucleotides has a measurable effect on the observed extension rate.[33,35] The sugar-ring conformation adopted in solution varies with nucleotide identity, with cytosine existing in the C3′-endo conformation more than 50% of the time. The preference for the C3′-endo conformation is strongest for C, then U, G and then A. This trend accounts for the higher rate of incorporation of C over A or G, despite C being a pyrimidine base which should have a weaker template association than the purine bases.[33] The increased binding of C also explains the strong sequence bias that exists for the misincorporation of C.[41] The preorganization effect also contributes to the high performance of the modified nucleotide 2-thiouridine, as the unactivated nucleotide 2-thio-UMP has been shown to adopt a C3′-endo conformation in solution, which combines with the superior template binding effect to make sU a superior substrate to U for primer extension.[35]

The reactivity of DNA could be improved by pushing DNA into the more reactive A-type conformation, either through solvent effects[51] or binding partners.[49] The use of 3′-NH_2-DNA in which the 3′-hydroxyl has been replaced with a more nucleophilic amine group has been suggested as a more reactive alternative to DNA.[17,21a,24,54] The 3′-NH_2-DNA is particularly efficient because in addition to the increased nucleophilicity of the amine group they adopt an A-type helix with a strong monomer preference for the C3′-endo conformation.[55] However, the resulting oligomers do not have a native phosphodiester backbone due to the presence of the nitrogen atom.[56] Whilst enzymes can be evolved to recognize this alternate backbone, as has been achieved with a DNA polymerase,[55–56] this solution is only viable *in vitro* and is not applicable for *in vivo* applications.

4.4. Other avenues for optimizing non-enzymatic template-directed primer extension reactions

One of the biggest issues impacting non-enzymatic primer extension is its slow rate, especially for reactions involving DNA. The slow reaction rate results in the accumulation of side-reactions discussed, particularly monomer hydrolysis. These hydrolyzed monomers contribute the major inhibition pathway of non-enzymatic primer extension and must be overcome.[32]

The hydrolysis of activated nucleotides is problematic for two reasons.[40] The first is the depletion of the activated nucleotide feedstock. The second issue is the inhibition caused by the hydrolyzed nucleotides, when these bind the template strand but are not able to undergo the primer extension reaction. This blocks the binding of activated nucleotides and slows down the overall reaction rate.[32] The depletion of activated nucleotides upon hydrolysis can be overcome with higher initial nucleotide concentrations, however this causes the inhibition by hydrolyzed nucleotides to be even larger in magnitude.

To overcome the issue of hydrolysis, three main avenues have been investigated.[57] The first is replacement of the hydrolyzed nucleotide with fresh substrate.[17] This approach requires the ability to separate the template-primer-helper complex from the solution phase, either through attachment of

the template to a bead as in solid-phase synthesis[24,26,58] or sequestration in a liposome.[17] Whilst an additional step, the replenishment of nucleotides has led to far superior reaction outcomes and so may be worth further investigation. The second option to overcome hydrolyzed nucleotide inhibition is to reactivate the hydrolyzed nucleotide. This is difficult to achieve directly with the 2-aminoimidazole system as the activation occurs under very different solvent conditions to the primer extension reactions.[22] There have, however, been promising results showing the *in situ* activation with isocyanide chemistry, which will be discussed later.[5] The third option to overcome hydrolysis is to run the reactions in such a manner that hydrolysis is not an issue. This would involve very rapid primer extension reactions with very low hydrolysis levels over the length of the reaction. This has been suggested as a valid option for the 3′-NH_2-DNA system, as these systems react rapidly, with timescales in the order of minutes.[24,32] However, this regime remains to be achieved for 2-aminoimidazole activated nucleotides given their low reactivity rates and the hydrolysis half-life of approximately 2 h[33] to 4 h.[20]

5. Choice of Activating Agent for Different Rate Regimes

Thus far, the mechanistic pathway of non-enzymatic template-directed primer extension has been outlined, the thermodynamic mechanism of entropic templating has been described, and the implications of this mechanism have been discussed, specifically with regards to how differences in nucleotide incorporation arise during the primer extension reaction. This next section will focus on how this understanding can be used to deconvolute the fidelity and rate of different activating agents. From here the ideal application of each activating agent will be discussed, based on the reaction rates of each system. Finally, the concepts discussed will be brought together to present a route towards optimizing the non-enzymatic template-directed primer extension reaction system.

5.1. Different Activating Regimes

Optimizing non-enzymatic primer extension has mainly focused on the issues of rate and yield. The issue of fidelity must be addressed by the choice of an activating agent with a high intrinsic fidelity. Throughout literature several different activating agents have been used, however, the choice of activating agent results in very different reaction outcomes. It is therefore important that a balance of reaction conditions to produce a high rate and yield should come once an apparent fidelity has been achieved. This balance has not been recognized throughout literature, with papers primarily focusing on obtaining high rates and yields, often working on single nucleotide systems to optimize reaction conditions and therefore not considering fidelity.

The different rates and fidelities of reactions arise as each activating agent has a different primer extension rate (k_{ext}), based on the intrinsic reactivity (ΔH) of the activated nucleotide. The intrinsic reactivity (ΔH) of the activated nucleotide has been discussed in the context of chemical selectivity in Section 3.

The discussion thus far has assumed that activated nucleotides are present at excess concentrations in the reaction system from the start of the reaction. These nucleotides either react with the primer in the non-enzymatic template-directed primer extension reaction, or undergo side reactions over time, such as hydrolysis. However, there are in fact three different activation regimes, based on the activating agent used. The three regimes are *in situ* activation, pre-activation (*ex situ*), and pre-activation (*ex situ*) with further *in situ* activation. These regimes arise from the activation method used to produce the activated nucleotide, however, they are also indicative of the intrinsic reactivity of each activating agent. This reactivity level impacts the k_{ext}, and how rapidly a template-primer bound activated nucleotide will react with the primer strand. As mentioned in Section 2.2.1, the relative rate of dissociation and extension determines fidelity, with $k_{off\text{-}m} > k_{ext} > k_{off\text{-}c}$ required for a high fidelity, high yield system. As such, the choice of activating agent has important implications on the potential fidelity that can be obtained.

The rate of the nucleotide activation (k_{act}) can be rate limiting based on the activation regime. As this has not been systematically evaluated, the observed rate k_{obs} could be a measure of k_{act} instead of k_{ext}, becoming an inaccurate measurement of the activated nucleotide reactivity in the primer extension reaction. From the relative rates of k_{act} and k_{ext}, and the mechanism of activation, each activating agent is best suited to different applications, which are summaried in Table 1.

5.2. In situ activation – EDC/HEI and catalytic additives

In the case of *in situ* activation, the unactivated nucleotide and activating agent are added to the reaction and the nucleotide activation occurs during the primer extension reaction. *In situ* activation is the mechanism present for the EDC and HEI systems with nucleophilic attack of the nucleotide to the EDC, followed by substitution of the of the EDC with HEI in an energy cascade reaction pathway, before the primer extension reaction itself.[27] CNBr also undergoes *in situ* activation.[4] This type of activation is ideal for truly one-pot reactions, as no pre-activation is required and conditions for re-activation of hydrolyzed monomers *in situ* can be achieved.[22]

In situ activation with EDC and HEI, or CNBr is required, as the resulting activated nucleotides are very reactive, and cannot be purified.[22] Unfortunately, this high reactivity renders these activating agents inappropriate for high fidelity non-enzymatic template-directed primer extension. Oligomerization of nucleotides in solution has been observed,[22,29b] suggesting a high level of non-templated reactivity. This is confirmed by the lack of sequence control observed when the EDC and HEI system

Table 1. Reaction properties of each activating agent investigated for non-enzymatic primer extension and ideal applications for each. Letters (a,b,c) refer to Figure 5, abbreviations: monomer (mon), dimer (dim), med (medium).

	activation regime		fidelity	reactivity	rate balance	application
CNBr EDC/HEI		*In situ* (mon)	Very Low	High	$k_{ext} >> k_{diss}$	DNA origami, nick ligation[(a)]
HOAt		*In situ* (mon)	Low	Good	$k_{ext} > k_{diss}$	3′-NH2-DNA[(b)]
Me-Im	*Ex situ* (mon)	*In situ* (dim)	Med	Med	$k_{ext} \geq k_{diss}$	
NH_2-Im	*Ex situ* (mon)	*In situ* (dim)	Good	Low	$k_{ext} \approx k_{diss}$	Primer extension[(c)]
	Ex situ (mon, dim)					
H-Im	*Ex situ* (mon)	*In situ* (dim)	unknown	Very low	$k_{ext} < k_{diss}$	
	Ex situ (mon, dim)					

was used for primer extension, suggesting that mismatch dissociation does not have the opportunity to occur prior to primer extension proceeding and a lack of microstate accumulation is required for entropic activation for primer extension to occur.[13] The rapid k_{ext} reaction rate in consistent with a high intrinsic reactivity of the EDC, HEI, or CNBr.

In EDC reactions the nucleotide activation (k_{act}) is rate limiting, with a very fast k_{ext}. Thus, EDC and other activating agents in this regime are not suited to non-enzymatic template-directed primer extension, but are ideally suited to nick ligation reactions where a single phosphodiester bond is formed between two strands bound to a template (Figure 5a).[18] Nick ligation has been shown in the assembly of oligonucleotide fragments into longer gene strands.[59] Applications also include the construction of DNA origami, where oligomers are preassembled on a scaffold and only following this pre-assembly is the EDC/HEI or CNBr system introduced to ligate the fragments.[4b,21a,60] In these reactions, the rapid primer extension rate is ideal, as the system is pre-positioned and the desired outcome is a fast reaction with high yields.

Similarly, the use of catalytic additives such as HEI have been shown to accelerate reaction rates more than 10-fold via a catalytic energy cascade pathway involving replacement of the activating agent followed by almost immediate primer extension.[21a] This is too reactive, with poor fidelity observed and high incorporation of G:T and A:C mismatches,[17] suggesting that mismatch dissociation did not occur prior to the primer extension proceeding with minimal sequence control observed. Similar results were observed with pyridine used as an organ catalyst.[24] This suggests that similar to the EDC/HEI systems, where $k_{ext} \gg k_{diss\text{-}m}$, energy cascade pathways are rendered too reactive for high fidelity primer extension reactions.

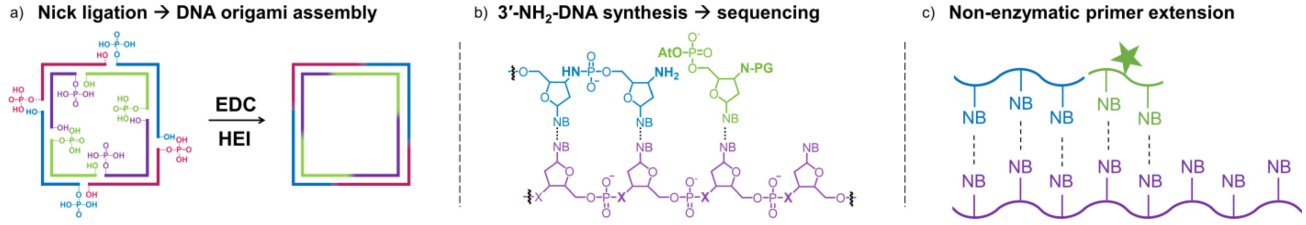


Figure 5. Ideal applications for each of the activating agents investigated. a) Application of EDC and HEI, nick ligation between 3′-OH and 5′-phosphate for forming DNA origami structures. b) Application of HOAt in 3′-NH2-DNA synthesis, with amine groups replacing the 3′-OH of DNA, note that in the template strand X can be either O or N. c) 2-aminoimidazole is the ideal activating agent for non-enzymatic primer extension, of those investigated.

The CNBr system is too active for sequence control as discussed, but a modification of this system with isocyanide chemistry has been shown to reactivate nucleotides *in situ*.[5] This could be a promising way to cycle hydrolyzed nucleotides back into the active 2-aminoimidazole form and then reintroduce it back to the template-primer-nucleotide system, for example in different chambers in a flow reactor for a one-pot flow setup, which will be discussed further.

5.3. Pre-activation – HOAt

Pre-activation is possible with less reactive activating groups including HOAt and the 2-x-imidazole agents. This pre-activation is necessary, as the activation reaction conditions are not conducive to *in situ* synthesis.[16,32] The HOAt and 2-x-imidazole activated nucleotides are stable enough to survive the purification process following synthesis, with a long enough half-life to be added to the primer extension reaction.[21c,37] This higher stability suggests the HOAt and 2-x-imidazole activated nucleotides are less reactive than the EDC activated nucleotides, with a lower k_{ext}.

The HOAt activated nucleotide monomers react directly with the primer on the template. The addition of activated nucleotides to the primer extension reaction removes the activation step and the k_{act} as a variable during the extension reaction itself and so the k_{ext} is rate limiting. Despite having a lower intrinsic reactivity than EDC, shown by the ability of HOAt-activated nucleotides to be purified, HOAt is too reactive for high fidelity, with only 56–66% sequence selectivity observed in a mixed nucleotide system after 20 h.[1] This suggests that $k_{ext} > k_{diss\text{-}m}$ and that sequence control with HOAt is too low to be a good activating agent for RNA or DNA based primer extension.

HOAt is ideally suited to 3′-NH_2-DNA systems (Figure 5b), as it is a very efficient activating agent for these reactions in which an amine is the nucleophile, with fast rates and high yields recorded.[24] HOAt activated nucleotides could be particularly efficient in mononucleotide systems where fidelity is not a concern, as a simpler method of 3′-NH_2-DNA production,[61] or used for a novel and rapid sequencing setup if coupled to a reversible termination protocol.[61] It should be noted that when the amine terminated primer is used, the primer extension reactions are performed at pH 8.9 in the absence of magnesium ions.[21a]

5.4. Pre-activation with in-situ activation

5.4.1. 2-x-imidazole

The 2-x-imidazole family fall into the third regime, pre-activation with further *in situ* activation. Similarly to HOAt, the 2-x-imidazole group is coupled to monophosphate nucleotides in a pre-activation regime, originally identified via imidazole.[54] The activated nucleotide monomers are added to the primer extension reaction system where these monomers undergo a dimerization reaction.[21b]

Walton et al.[27a] have shown that the 2-x-imidazole activated nucleotides undergo dimerization in solution. Dimer formation was confirmed via NMR,[20] with the dimer being much more active than the monomer and acting as the main substrate for the primer extension reaction to give the primer + 1 extension product.[27a]

As such the concentration of the dimer is rate limiting, with negligible contributions to k_{obs} from the activated monomer.[20] The presence of a rate limiting dimerization reaction k_{dim} reduces the overall reaction rate k_{obs} and the yield that can be obtained before the activated nucleotides hydrolyze.[20] This additional reaction step also complicates the reaction system and interpretation of the results obtained.

Within the 2-x-imidazole family, the 2-methylimidazole variant was identified to be superior to imidazole, with a 5–10 fold faster primer extension and superior 3′ to 2′ regioselectivity observed, and has been extensively studied for non-enzymatic primer extension.[52,54] The 2-methylimidazole activated nucleotides are more stable than HOAt activated nucleotides. HOAt activated DNA monomers have a half-life of 6–16 h, with RNA monomers showing a 6–8 h half-life. By comparison, the 2-methylimidazole activated nucleotides are much more stable with a half-life of 19–29 h for the activated DNA monomers and 53–63 h for the RNA monomers.[21c] As such, non-enzymatic template-directed primer extension reactions using 2-methylimidazole activated nucleotides display superior fidelity of extension compared to HOAt,[19] but inferior reaction rates, with 74–84% fidelity in a four nucleotide reaction and a rate of 38–84 h^{-1} for 2-methylimidazole, compared to 64–93% fidelity in a four nucleotide reaction and a rate of 94–259 h^{-1} for HOAt.[16]

5.4.2. 2-aminoimidazole

In a screen of 2-x-imidazole agents, 2-aminoimidazole was identified to be a superior activating agent, with a primer extension rate of 2.14 h^{-1}, compared to 0.31 h^{-1} for 2-methylimidazole, 0.143 h^{-1} for imidazole and only 0.018 h^{-1} for 2-chloroimidazole activated nucleotides.[27b] This screen showed that the higher the electron donating capabilities of the substituent at the 2-position of the imidazole ring, the greater the observed rate of non-enzymatic primer extension. This result contradicts the reactivity of the activated dimers, as the 2-methylimidazole nucleotide dimers were shown to be more reactive than 2-aminoimidazole dimers.[27b]

Here, the higher reactivity of the 2-methylimidazole nucleotide dimers is disadvantageous, as the activated dimers are much more prone to hydrolysis.[27b] By contrast, 2-aminoimidazole is less reactive, allowing the active dimers to accumulate in solution to higher concentrations than the 2-methylimidazole dimers,[21b] with a dimer concentration of 2.5 μM measured in a 500 μM monomer solution of 2-aminoimidazole activated G.[25] Additionally, the 2-aminomidazole activated nucleotide dimers can be produced synthetically, purified and used in reactions

with a half-life to hydrolysis of 135 min for A*C dimers and 120 min for A*U dimers,[30] and 245 min for C*C dimers.[20]

The lower reactivity of 2-aminomidazole activated nucleotide dimers also allows more chances for mismatch dissociation to occur prior to the extension reaction, increasing the potential fidelity. Given the dimeric nature of the activated nucleotides, if both nucleobases are complementary they should facilitate a larger entropic effect with a greater microstate accumulation, relative to the HOAt or EDC activated monomer.[20]

Additionally, 2-aminoimidazole activated dimers benefit from the same enhanced template interactions observed for HOAt, where the activating agent can form hydrogen bonding interactions to the template strand and enhance the nucleophilicity of the 3′-OH group.[21c] 2-methylimidazole is not able to interact in this manner and therefore does not have the same advantageous effect in the bound complex state.[32]

The low k_{ext} of the 2-aminoimidazole activated nucleotides suggest that they are the ideal candidate for non-enzymatic primer extension reactions (Figure 5c).

5.5. Optimizing non-enzymatic template-directed primer extension

5.5.1. Optimizing the 2-aminoimidazole reaction system

The 2-aminoimidazole system is the current best candidate for the primer extension reaction. The 2-aminoimidazole activated dimers show a low intrinsic reactivity, necessary to prevent off-target reactions, minimize hydrolysis, and balance k_{ext} with $k_{off\text{-}c}$ to provide a high fidelity outcome. While the relatively slow extension rate k_{ext} is required for fidelity, there is still the capacity to optimize the overall reaction rate k_{obs}.

The main difference between the general reaction pathway shown previously (Figure 1a) and the 2-aminoimidazole system is the dimeric nature of the activate nucleotides. The dimeric nature of the activated nucleotides is a distinct disadvantage. It requires a 2-nucleotide gap on the template strand between the primer and helper to accommodate the dimer. However, the larger issue is the dimerization reaction itself. As discussed, the dimerization reaction is rate limiting and slow,[21b,45] resulting in primer extension reactions that take days to run.[26] The reaction is even self-inhibiting, as the dimerization reaction releases one unit of 2-aminoimidazole, which inhibits the dimer formation through equilibrium effects as it accumulates.[37]

The rate limiting k_{dim} step can be overcome if the activated nucleotide dimers are pre-synthesised, purified and then added directly to the primer extension reaction.[33] This shifts the 2-aminoidazole system to the second pre-activation regime, in which the k_{ext} is rate limiting, which is ideal for a high fidelity and high efficiency system. Whilst this would optimize the primer extension reaction itself, performing the pre-dimerization step as routine is restrictive, as the reaction system requires the preparation of two separate activated nucleotides that are coupled together to give the 2-aminoimidazole bridged dimer.[33] In addition, the monomeric by-product which is released upon primer extension can at as an inhibitor to the primer extension reaction,[20,30] in a similar manner to that of hydrolyzed nucleotide monomer.[57] One suggestion to overcome this issue is tethering of the template complex to a solid phase,[46] followed by frequent complete replacement of the supernatant with fresh substrate.[26] Whilst effective, this approach is very resource intensive and does not optimize the fundamental primer extension reaction.

Recent work from Zhang and coworkers offers a unique one-pot reaction setup for the 2-aminoidazole system. Through the combination of eutectic phase freeze-thaw cycling and isocyanamide chemistry, nucleotides are activated *in situ*.[5,44] This system overcomes the differing chemical requirements of the nucleotide activation reaction, dimerization reaction, and primer extension reaction, which have required their separation into different processes and can facilitate reactivation of spent nucleotide.[62] This one-pot system showed very promising high conversion levels above 98%, with mixed nucleotide systems also examined and is well worth further investigation.[44]

5.5.2. Optimizing the activating agent

Even with the advances proposed, the need for reactive 2-aminoimdazole bridges dimers increases the complexity of the nucleotide feedstock dramatically, from 4 activated monomers to 10 activated dimers when pre-synthesised.[33] Even with the isocyanamide formation allowing for *in situ* oligomerization,[62] the concentration of each nucleotide and their depletion rate in solution also needs to be accounted for, further adding to the complexity.[44] These issues are particularly problematic given the need for increased T/U concentrations due to their low binding affinity[33] and propensity for G-based nucleotides to aggregate at higher concentrations, effectively decreasing their concentration dramatically.[44]

As such, development of a novel activating agent for a reactive nucleotide monomer is a potential area for future development. The activated nucleotide would require a reactivity similar to that of the 2-aminoimidazole activated nucleotides, to ensure that $k_{off\text{-}m} > k_{ext} > k_{off\text{-}c}$. The use of a N–P bond between a nitrogen on the activating agent and the phosphate on the nucleotide appears to facilitate this reactivity balance.

For the chemical structure of the activating agent, an arrangement which facilitates hydrogen bonding with the DNA or RNA to position the nucleotide and interact with the catalytic Mg^{2+} ion would be ideal.[21b] This would allow reactions to benefit from the directing and activating effects observed for HOAt and 2-aminoimidazole.[21c]

Similarly to the HOAt activated nucleotides, the activated nucleotides would be required to undergo primer extension in a monomeric form, to simplify the activated nucleotide feedstock complexity. *In situ* activation of the nucleotide, similarly to the EDC/HEI[40] and isocyanamide chemistry,[62] would also facilitate re-activation of spent nucleotide, decreasing the issue of inhibition.

A group such as a guanidinium group may facilitate activation of the phosphate and the desired hydrogen-bonding pattern. This structure mimics the hydrogen bonding capabil-

ities of HOAt and 2-aminoimidazole, which are ablated by the methyl group on the 2-methylimidazole activated nucleotides.[21c,32] However, development of the structure would be required to block dimerization (or trimerization) involving activation of multiple nucleotides from the one guanidinium group. The introduction of a chemical moiety which blocks the reactivity of the nitrogen whilst maintaining hydrogen bonding capacity, in a similar manner as the hydroxyethyl group in HEI, may facilitate this.

If all of these parameters were met, an optimal reagent could potentially be designed which is driven by entropic activation and facilitates a balance in reaction rates and equilibrium positions for a high rate, fidelity and yield reaction outcome.

6. Conclusions

Here, the non-enzymatic template-directed primer reaction has been discussed, with a focus on achieving a balance of high rate, fidelity, and reaction yield. The role of the competing equilibrium reactions was highlighted, with traditional approaches using enthalpic activation shown to result in a low level of fidelity. An alternative approach, using entropic activation was proposed as the dominant mechanism of improving the Gibbs Free Energy of non-enzymatic template-directed primer extension. This shift from considering enthalpic to entropic activation can better enable the design and control of optimized non-enzymatic template-directed primer extension reaction systems for both DNA and RNA into the future.

In order to optimize the reaction system, the different activating agents used throughout literature were compared. Based on their reactivity and mechanism of activation relative each of the six activating agents discussed was suggested to suit different reaction regimes. The EDC/HEI and CNBr systems show a high reaction rate with *in situ* nucleotide activation so are most suited to being introduced to pre-templated systems to facilitate phosphodiester bond formation. The HOAt activated nucleotides are less reactive than ED/HEI or CNBr. They are pre-activated, but display poor fidelity, showing good results for amine linked DNA (3′-NH_2-DNA). For the non-enzymatic template-directed primer extension reactions the 2-X-imidazole systems were the prime candidates, with the 2-aminoimidazole system being proposed as the ideal candidate for a high fidelity primer-extension reaction system. Strategies for optimizing these reactions involve ensuring the primer extension reaction is rate limiting through pre-activation of the 2-aminoimidazole linked dimers.

It is anticipated that the ideas presented here will enable a more informed understanding of the non-enzymatic template-directed primer extension reaction. This will enable the development of primer extension reactions into a powerful and versatile technique for generating long strands of DNA and RNA for a diverse range of applications. These could include DNA or RNA synthesis, with the incorporation of non-natural nucleotides to facilitate healthcare applications such as mRNA vaccine delivery, the development of integrated research tools such as DNA-paint or the use of the material properties of DNA for responsive surface tethering. Further, there is huge potential for the synthesis of DNA or RNA grafting off a diverse range of materials for advanced nanomaterials with DNA origami or hybrid structures with metallic nanoparticles, energy harvesting polymers, or sensors.

Author Contributions

Conceptualization (Kimberley L. Callaghan, Peter C. Sherrell), Data curation (Kimberley L. Callaghan), Funding acquisition (Amanda V. Ellis), Investigation (Kimberley L. Callaghan), Project administration (Amanda V. Ellis), Resources (Amanda V. Ellis), Supervision (Peter C. Sherrell, Amanda V. Ellis), Visualization (Kimberley L. Callaghan, Peter C. Sherrell), Writing – original draft (Kimberley L. Callaghan), Writing – review & editing (Kimberley L. Callaghan, Peter C. Sherrell, Amanda V. Ellis)

Acknowledgements

P. C. S. acknowledges support from RMIT University via the Vice-Chancellor's Research Fellowship Scheme (2023). A. V. E. and K. L. C acknowledges funding from the Australian Research Council (ARC) Discovery program, DP190100055. Open Access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

Conflict of Interests

There are no conflicts to declare.

Keywords: DNA · RNA · non-enzymatic primer extension · thermodynamics · entropy

[1] S. R. Vogel, C. Deck, C. Richert, *Chem. Commun. (Camb.)* **2005**, *39*, 4922–4924.
[2] S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobe, D. A. Treco, J. W. Szostak, *Nature* **2008**, *454*, 122–125.
[3] a) N. D. Sinha, J. Biernat, J. McManus, H. Koster, *Nucleic Acids Res.* **1984**, *12*, 4539–4557; b) Y. Hayakawa, S. Wakabayashi, H. Kato, R. Noyori, *J. Am. Chem. Soc.* **1990**, *112*, 1691–1696.
[4] a) A. Radakovic, S. DasGupta, T. H. Wright, H. R. M. Aitken, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2021**, *119*, e2116840119; b) M. Kramer, C. Richert, *Chem Biodivers* **2017**, *14*, e1700315; c) M. Synakewicz, D. Bauer, M. Rief, L. S. Itzhaki, *Sci. Rep.* **2019**, *9*, 13820.
[5] S. J. Zhang, D. Duzdevich, J. W. Szostak, *J. Am. Chem. Soc.* **2020**, *142*, 14810–14813.
[6] O. Zavoiura, B. Brunner, P. Casteels, L. Zimmermann, M. Ozog, C. Boutton, M. W. Helms, T. Wagenaar, V. Adam, J. Peterka, C. Metz-Weidmann, P. Deschaght, S. Scheidler, K. Jahn-Hofmann, *Mol. Pharm.* **2021**, *18*, 1048–1060.
[7] Y. Shao, H. Jia, T. Cao, D. Liu, *Acc. Chem. Res.* **2017**, *50*, 659–668.
[8] J. Zhu, Y. Kim, H. Lin, S. Wang, C. A. Mirkin, *J. Am. Chem. Soc.* **2018**, *140*, 5061–5064.
[9] T. Walton, S. DasGupta, D. Duzdevich, S. S. Oh, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 5741–5748.
[10] A. Santorelli, K. V. Gothelf, *Nucleic Acids Res.* **2022**, *50*, 7235–7246.
[11] R. van Wee, M. Filius, C. Joo, *Trends Biochem. Sci.* **2021**, *46*, 918–930.

[12] J. Schnitzbauer, M. T. Strauss, T. Schlichthaerle, F. Schueder, R. Jungmann, *Nat. Protoc.* **2017**, *12*, 1198–1228.
[13] T. Schlichthaerle, A. S. Eklund, F. Schueder, M. T. Strauss, C. Tiede, A. Curd, J. Ries, M. Peckham, D. C. Tomlinson, R. Jungmann, *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 11060–11063.
[14] E. M. Sletten, C. R. Bertozzi, *Acc. Chem. Res.* **2011**, *44*, 666–676.
[15] A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli, T. Brown, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11338–11343.
[16] P. Hagenbuch, E. Kervio, A. Hochgesand, U. Plutowski, C. Richert, *Angew. Chem.* **2005**, *117*, 6746–6750.
[17] D. K. O'Flaherty, L. Zhou, J. W. Szostak, *J. Am. Chem. Soc.* **2019**, *141*, 10481–10488.
[18] C. Obianyor, G. Newnam, B. E. Clifton, M. A. Grover, N. V. Hud, *ChemBioChem* **2020**, *21*, 3359–3370.
[19] K. Leu, E. Kervio, B. Obermayer, R. M. Turk-MacLeod, C. Yuan, J. M. Luevano Jr., E. Chen, U. Gerland, C. Richert, I. A. Chen, *J. Am. Chem. Soc.* **2013**, *135*, 354–366.
[20] T. Walton, J. W. Szostak, *Biochemistry* **2017**, *56*, 5739–5747.
[21] a) L. Zhou, D. K. O'Flaherty, J. W. Szostak, *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 15682–15687; b) T. Walton, W. Zhang, L. Li, C. P. Tam, J. W. Szostak, *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 10812–10819; c) E. Kervio, M. Sosson, C. Richert, *Nucleic Acids Res.* **2016**, *44*, 5504–5514.
[22] M. Jauker, H. Griesser, C. Richert, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 14559–14563.
[23] M. Rothlingshofer, C. Richert, *J. Org. Chem.* **2010**, *75*, 3945–3952.
[24] M. Rothlingshofer, E. Kervio, T. Lommel, U. Plutowski, A. Hochgesand, C. Richert, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 6065–6068.
[25] P. Tremmel, H. Griesser, U. E. Steiner, C. Richert, *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 13087–13092.
[26] C. Deck, M. Jauker, C. Richert, *Nat. Chem.* **2011**, *3*, 603–608.
[27] a) T. Walton, J. W. Szostak, *J. Am. Chem. Soc.* **2016**, *138*, 11996–12002; b) L. Li, N. Prywes, C. P. Tam, D. K. O'Flaherty, V. S. Lelyveld, E. C. Izgu, A. Pal, J. W. Szostak, *J. Am. Chem. Soc.* **2017**, *139*, 1810–1813.
[28] J. P. Schrum, A. Ricardo, M. Krishnamurthy, J. C. Blain, J. W. Szostak, *J. Am. Chem. Soc.* **2009**, *131*, 14560–14570.
[29] a) T. Z. Jia, S. Nishikawa, K. Fujishima, *BBA Adv* **2022**, *2*, 100049; b) D. Duzdevich, C. E. Carr, J. W. Szostak, *Nucleic Acids Res.* **2020**, *48*, e70.
[30] E. Kervio, B. Claasen, U. E. Steiner, C. Richert, *Nucleic Acids Res.* **2014**, *42*, 7409–7420.
[31] M. G. AbouHaidar, I. G. Ivanov, *Zeitschrift für Naturforschung C* **1999**, *54*, 542–548.
[32] M. Sosson, C. Richert, *Beilstein J. Org. Chem.* **2018**, *14*, 603–617.
[33] D. Ding, L. Zhou, C. Giurgiu, J. W. Szostak, *Nucleic Acids Res.* **2022**, *50*, 35–45.
[34] C. Giurgiu, Z. Fang, H. R. M. Aitken, S. C. Kim, L. Pazienza, S. Mittal, J. W. Szostak, *Angew. Chem. Int. Ed. Engl.* **2021**, *60*, 22925–22932.
[35] N. Zhang, S. Zhang, J. W. Szostak, *J. Am. Chem. Soc.* **2012**, *134*, 3691–3694.
[36] G. Leveau, D. Pfeffer, B. Altaner, E. Kervio, F. Welsch, U. Gerland, C. Richert, *Angew. Chem. Int. Ed. Engl.* **2022**, *61*, e202203067.
[37] D. Duzdevich, C. E. Carr, D. Ding, S. J. Zhang, T. S. Walton, J. W. Szostak, *Nucleic Acids Res.* **2021**, *49*, 3681–3691.
[38] a) K. Leu, B. Obermayer, S. Rajamani, U. Gerland, I. A. Chen, *Nucleic Acids Res.* **2011**, *39*, 8135–8147; b) S. Zhang, J. C. Blain, D. Zielinska, S. M. Gryaznov, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17732–17737.
[39] B. D. Heuberger, A. Pal, F. Del Frate, V. V. Topkar, J. W. Szostak, *J. Am. Chem. Soc.* **2015**, *137*, 2769–2775.
[40] M. Sosson, D. Pfeffer, C. Richert, *Nucleic Acids Res.* **2019**, *47*, 3836–3845.
[41] S. Rajamani, J. K. Ichida, T. Antal, D. A. Treco, K. Leu, M. A. Nowak, J. W. Szostak, I. A. Chen, *J. Am. Chem. Soc.* **2010**, *132*, 5880–5885.
[42] D. K. O'Flaherty, N. P. Kamat, F. N. Mirza, L. Li, N. Prywes, J. W. Szostak, *J. Am. Chem. Soc.* **2018**, *140*, 5171–5178.
[43] a) K. Adamala, A. E. Engelhart, N. P. Kamat, L. Jin, J. W. Szostak, *Nat. Protoc.* **2015**, *10*, 927–938; b) K. P. Adamala, A. E. Engelhart, J. W. Szostak, *Nat. Commun.* **2016**, *7*, 11041.
[44] S. J. Zhang, D. Duzdevich, D. Ding, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116429119.
[45] T. Walton, L. Pazienza, J. W. Szostak, *Biochemistry* **2019**, *58*, 755–762.
[46] N. Prywes, J. C. Blain, F. Del Frate, J. W. Szostak, *eLife* **2016**, *5*, e17756.
[47] W. Zhang, C. P. Tam, L. Zhou, S. S. Oh, J. Wang, J. W. Szostak, *J. Am. Chem. Soc.* **2018**, *140*, 2829–2840.
[48] L. Zhou, S. C. Kim, K. H. Ho, D. K. O'Flaherty, C. Giurgiu, T. H. Wright, J. W. Szostak, *eLife* **2019**, *8*, e51888 .
[49] S. J. Park, K. L. Callaghan, A. V. Ellis, *Org. Biomol. Chem.* **2023**, *21*, 6702–6706.
[50] A. Rich, *Nat. Struct. Mol. Biol.* **2003**, *10*, 247–249.
[51] a) V. I. Ivanov, L. E. Minchenkova, E. E. Minyat, M. D. Frank-Kamenetskii, A. K. Schyolkina, *J. Mol. Biol.* **1973**, *87*, 817–833; b) C. H. Lee, H. Mizusawa, T. Kakefuda, *Biochemistry* **1981**, *78*, 2838–2842.
[52] S. Zhang, N. Zhang, J. C. Blain, J. W. Szostak, *J. Am. Chem. Soc.* **2013**, *135*, 924–932.
[53] M. Egli, *The Excitement of Discovery: Selected Papers of Alexander Rich: A Tribute to Alexander Rich* **2019**, *73*, 309–315.
[54] W. S. Zielinski, L. E. Orgel, *Nucleic Acids Res.* **1985**, *13*, 2469–2484.
[55] V. S. Lelyveld, D. K. O'Flaherty, L. Zhou, E. C. Izgu, J. W. Szostak, *Nucleic Acids Res.* **2019**, *47*, 8941–8949.
[56] D. K. O'Flaherty, L. Zhou, J. W. Szostak, *Bio Protoc* **2020**, *10*, e3734.
[57] A. Kaiser, C. Richert, *J. Org. Chem.* **2013**, *78*, 793–799.
[58] A. Kaiser, S. Spies, T. Lommel, C. Richert, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 8299–8303.
[59] R. A. Hughes, A. D. Ellington, *Cold Spring Harbor Perspect. Biol.* **2017**, *9*, a023812.
[60] L. Zhou, D. K. O'Flaherty, J. W. Szostak, *J. Am. Chem. Soc.* **2020**, *142*, 15961–15965.
[61] N. Griesang, K. Giessler, T. Lommel, C. Richert, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 6144–6148.
[62] D. Ding, S. J. Zhang, J. W. Szostak, *Nucleic Acids Res.* **2023**, *51*, 6528–6539.

Manuscript received: December 20, 2023
Revised manuscript received: January 21, 2024
Accepted manuscript online: January 28, 2024
Version of record online: February 12, 2024