Nucleic-acid-templated synthesis as a model system for ancient translation
Christopher T Calderone and David R Liu*

The translation of nucleic acids into synthetic structures with expanded functional potential has been the subject of considerable research, with applications including small-molecule and polymer evolution, reaction discovery and sensing. Here, we review properties of nucleic-acid-templated synthesis in the context of requirements for prebiotic translation. This analysis highlights the chemical possibilities of ancient translation systems, as well as challenges that these systems may have faced.

Addresses
Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA
E-mail: drliu@fas.harvard.edu

Introduction
For half a century, researchers have appreciated that proteins are translated from information-bearing nucleic acids, which, in turn, are amplified by replication [1]. In the modern world, a network of highly complex macromolecular assemblies mediates these processes of translation and replication. We assume that before the advent of modern translation and replication machinery, information still flowed from replicable carriers to functional molecules to allow the evolution of living systems with diverse chemical capabilities. To better understand the requirements for the prebiotic replication of nucleic acids, researchers began to study the non-enzymatic, template-directed replication of nucleic acids as early as 1966 [2]. More recently, attention has turned to the non-enzymatic translation of nucleic acids into unrelated structures of greater functional potential. Although these recent efforts typically do not operate under reaction conditions chosen to mimic the prebiotic world, they nevertheless provide insights into the potential of nucleic acids to mediate translation without the aid of the modern translational apparatus.

Readout
In the modern world, protein-encoding information is read by the adjacent annealing of three-base tRNA anticodon and a strand of mRNA; second, chemical reaction (acyl transfer from a peptidyl-tRNA to an aminoacyl-tRNA); third, elongation (iteration of template-directed reactions enabled by ribosomal translocation); and fourth, termination (hydrolytic release of the product polypeptide with subsequent dissociation of the mRNA-ribosome complex). Analogues for each of these events have been achieved using nucleic-acid-templated synthesis in the absence of enzymes, and examples of each are presented in this article.

Regardless of its context, translation in the most abstract form involves four events: first, readout (in the biotic world, base pairing between an aminoacyl-tRNA anticodon and a strand of mRNA); second, chemical reaction (acyl transfer from a peptidyl-tRNA to an aminoacyl-tRNA); third, elongation (iteration of template-directed reactions enabled by ribosomal translocation); and fourth, termination (hydrolytic release of the product polypeptide with subsequent dissociation of the mRNA-ribosome complex). Analogues for each of these events have been achieved using nucleic-acid-templated synthesis in the absence of enzymes, and examples of each are presented in this article.
For some rapid reactions such as amine acylation, template–substrate annealing is rate-determining, resulting in reaction rates that are independent of the number of nucleotides separating reactive groups (‘distance independence’). For some slower reactions such as 1,3-dipolar nitro-olefin cycloaddition, the chemical step is rate-determining, yielding overall reaction rates that decrease with increasing separation of reactive groups (‘distance dependence’). The suitability of a nucleic acid template for mediating the successful readout of information therefore depends on the kinetics of the chemical reaction that follows base pairing.

Interestingly, recent studies have shown that readout can take place efficiently even if the coding nucleotides are split into two separate regions of primary sequence (Figure 2) [8,9]. Gartner, Liu and co-workers designed a DNA template in which annealing of both a 10-base region several bases away from the reaction site, as well as a 5-base region adjacent to the reaction site, is required for maximally efficient product formation. This split codon arrangement enabled distance-dependent reactions such as reductive amination or nitro-olefin cycloaddition to proceed efficiently even when one of the coding regions is far from the reactive end of the template. Melting data suggested that annealing of the 5-base region does not increase the overall energetics of duplex formation, but instead barely offsets the entropic cost of ordering the DNA between the two coding regions. These findings collectively imply that coding regions for templated chemistries can lie at a variety of distances along a template yet still direct a wide variety of reactions.

Although the vast majority of research in nucleic-acid-templated synthesis has exploited canonical Watson–Crick base pairing as the basis for readout, other approaches are also possible. Poulin-Kerstien, Dervan and co-workers have used sequence-specific binding of polyamides to the minor groove of DNA to template the cycloaddition between alkynes and azides in a highly distance-dependent manner [10]. Others have used Hoogsteen pairing to the major groove of a DNA duplex to template the formation of a native DNA phosphodiesteer bond [11,12]. Although triplex formation has not yet been reported to translate DNA sequences into structures unrelated to the nucleic acid backbone, the generality of DNA-templated synthesis as summarized above and below strongly imply this possibility.

**Chemical reactions**

Early work in nucleic-acid-templated synthesis assumed that the transition state of a DNA-templated reaction should closely mimic the structure of the phosphoribose backbone to proceed efficiently [13]. Researchers have recently shown that structural mimicry of a nucleic acid backbone is not necessary for efficient and sequence-specific DNA-templated synthesis. DNA is now known to support template-directed formation of phosphoramidates [14], phosphoroselenates [15] and phosphorothioates [16]; [2+2] photodimerization [17]; nucleophilic substitution and conjugate addition [6]; amine acylation, reductive amination, nitro-aldol and nitro-Michael addition, Wittig olefination, and transition metal-catalyzed cross-coupling [7*]; heterocycle formation [9]; cycloadditions [7*,8,10]; and metal coordination complex formation (Figure 3) [18,19,20]. Reductive amination in particular has proven to have intriguing kinetic properties, displaying strong distance dependence but rapid overall rates of reaction, an observation that has not been fully explained thus far.

Other researchers have used DNA-templated synthesis to direct functional group transformations such as ester
Many reactions can proceed sequence-specifically in a DNA-templated format [6,7,8–10,14–18].

hydrolysis rather than to couple DNA-linked substrates. Ma and Taylor linked an imidazole moiety to DNA to catalyze the hydrolysis of a para-nitrophenyl ester sequence-specifically [21,22]. Brunner, Kraemer and co-workers appended a ligand for copper to DNA and demonstrated metal-catalyzed hydrolysis of picolinates in a sequence-specific manner [23].

The surprisingly wide range of chemical reactions supported by DNA-templated synthesis can be understood...
by considering the role of the oligonucleotides as simply elevating the effective molarity of reactants. DNA’s role as a passive tether was demonstrated by a series of DNA-templated conjugate additions in which the 10 nucleotides separating reactive groups were systematically replaced with backbone analogs lacking nucleobases, riboses or phosphates with little effect on reaction efficiency [6]. In addition, Li and Liu observed that template-induced stereoselectivity in DNA-templated synthesis vanishes when even a modest number of flexible bonds are present in the linkage between reactants and oligonucleotides [24]. These findings are consistent with the conclusion that simply linking reactive groups to complementary oligonucleotides upon DNA hybridization can elevate the effective molarity of reactive groups by many orders of magnitude over their nanomolar absolute concentrations [6]. The presence of macromolecules with precise tertiary structures is therefore not necessary for nucleic-acid-templated chemistries to proceed efficiently and sequence-specifically, although such structures can dramatically accelerate reaction rates, especially between relatively unreactive substrates [25].

**Elongation**

Elongation of a translation product requires that a single template direct multiple successive chemical reactions. At least three approaches have achieved multistep templated syntheses on a single strand of DNA in the absence of enzymes (Figure 4). In one approach, the template architecture is designed to allow multiple oligonucleotide-linked reactants to independently approach and react with a common template-linked substrate. Alternatively, reactive groups can be appended to oligonucleotides through cleavable linkers, allowing the substrate-delivering DNA to be removed after chemical reaction. In a third approach to multistep nucleic-acid-templated syntheses, template-annealed building blocks are consecutively coupled to one another to form a polymer chain that grows along a DNA template.

Researchers have used the first approach to perform up to three reactions on a single DNA template (Figure 4a). For example, Gartner, Liu and co-workers used a DNA template displaying a primary amine group on the non-Watson–Crick face of an internal thymidine to direct contemporaneous amine acylation and Huisgen cycloaddition. Separately, a similar template architecture was used to mediate nucleophilic substitution followed by Wittig olefination [8]. Eckardt and von Kiedrowski used a branched DNA template to mediate tris-hydrazone formation between three DNA-linked hydrazines and a non-DNA-linked trialdehyde [26].
The second approach to multistep DNA-templated synthesis uses DNA-linked reagents in which the reactive functional group is connected to the delivering oligonucleotide by a cleavable linker (Figure 4b). A set of such linkers has been devised for this purpose, including esters [27,28*]; thioesters [9]; diols, ethyl sulfones and phosphoranes [28*]; and disulfides [29]. In some cases, DNA-templated product formation induces spontaneous linker cleavage as a mechanistic consequence of the reaction [28*,30]. Combinations of cleavable linkers have been used to perform three-step DNA-templated syntheses of peptides, branched thioethers [28*], macroyclic fumaricamides [30], and N-acyloxazolidinones [9]. Taking inspiration from ribosomal peptide synthesis, Tamura and Schimmel recently reported a system in which a series of acyl transfer reactions on an RNA template led to tripeptide synthesis in a single vessel [27,31*].

The third elongation strategy most closely resembles modern translation. In ribosomal protein biosynthesis, consecutive acyl transfers from peptidyl-tRNA to aminoacyl-tRNA generate proteins containing many amino acids (Figure 4c). A simplified version of this approach replaces aminoacyl-tRNAs with monomers that directly recognize the encoding template. Such DNA-templated polymerizations are perhaps the most chemically challenging form of multistep templated synthesis due to their stringent requirements for efficiency, sequence-specificity, and site-selectivity to generate even a small fraction of full-length polymeric products of correct sequence (Figure 4d).

The use of reductive amination has proven to be a key feature of successful DNA-templated polymerizations. The strong distance dependence of this reaction leads to high selectivity for adjacent monomer couplings. In addition, reductive amination proceeds through a reversible imine intermediate. Lynn proposes a model in which the added stability of the imine–template complex relative to the ternary substrate–template complex biases the reaction towards polymerization as well as providing an additional measure of sequence selectivity (Figure 5) [32]. By exploiting these features of imine formation, Li, Lynn and co-workers quantitatively and chain-length-specifically translated octathymidine into a homoadenosine DNA analog with step-growth kinetics [33*,34]. Rosenbaum and Liu were able to extend these findings by translating mixed-sequence DNA into 40-mer peptide nucleic acid (PNA) analogs in one solution through 10 consecutive DNA-templated reductive amination coupling reactions [35]. The resulting sequence-defined DNA-templated polymers are of length similar to that of functional proteins and nucleic acids, but are made entirely of synthetic monomers without the assistance of macromolecular machinery. Given the much higher degree of design and engineering needed to implement the first two elongation strategies, it is tempting to speculate that this third approach to non-enzymatic multistep templated synthesis may be the most relevant to prebiotic elongation.

**Termination**

Compared with ribosomal termination in which the recognition of stop codons by protein release factors triggers ester hydrolysis of peptidyl-tRNAs, the chemical events following the last step in a typical non-enzymatic templated synthesis are relatively simple. Nevertheless, these events have important ramifications for prebiotic translation. Molecular evolution requires that the functional properties of translated products affect the ability of their encoding templates to survive and replicate. For nucleic-acid-templated products to impart a survival advantage during a selection step requires a stable association between the template and the corresponding translated structure. Many of the laboratory examples described above use a stable template–product covalent bond to establish this linkage and do not require an explicit termination step. For example, the template–product linkage within each member of a small DNA-templated macrocycle library enabled Gartner, Liu and co-workers to subject the library to an *in vitro* selection and enrich for a single template sequence based on the ability of its corresponding translated macrocycle to bind to a target protein [30].

In other contexts — for example, if the nucleic acid is compartmentalized together with its translation product within a vesicle — it may be advantageous for the translated structure to become disconnected from the encoding template. When termination induces breakage of the template–product bond or when the template never becomes covalently attached to products, a single template can be reused to direct the synthesis of multiple copies of a translation product. This form of multiple turnover requires dissociation of the product–template complex to release the unmodified template to direct subsequent syntheses. Dissociation rates compatible with efficient turnover in some cases may be incompatible with the template–substrate affinities that are required to achieve large increases in effective molarity. To overcome this conflict, researchers including Xu and Kool [16] and Albargi and co-workers [36] have used thermocycling to enable rapid product–template dissociation at high temperatures, and stable substrate–template association at lower temperatures. Zhan and Lynn developed an alternative approach, in which a DNA-templated reductive amination directs the coupling of two oligonucleotide-linked amine or aldehyde substrates [37]. The product amine–template duplex is destabilized relative to the intermediate imine–template duplex, enabling efficient product release (Figure 5) [38].

Taken together, these examples demonstrate that a delicate balance of substrate–template affinity, product–template affinity, and reaction conditions enables translation to
terminate in a manner that allows a single template molecule to generate multiple copies of a translation product.

**Reaction modes enabled by nucleic-acid-templated synthesis**

In contrast with conventional laboratory synthesis, DNA-templated reactions are controlled by modulating effective molarities of reactive groups. As a result, templated synthesis enables some modes of controlling reactions that are not possible when random intermolecular collisions direct reactivity. For example, Calderone, Liu and co-workers have shown that in a DNA-templated format it is possible to perform multiple otherwise incompatible reactions such as conjugate additions, reductive amination, amine...
acylation and Wittig olefination simultaneously in a single solution [39]. In a conventional synthesis format, it is not possible to simultaneously control these reactions because many of the substrates are cross-reactive. DNA-templated synthesis has also been shown to enable heterocoupling reactions to take place between substrates that preferentially homocouple under conventional reaction conditions because the effective molarity of heterocoupling substrates can be much higher than the absolute concentration of any single substrate (Figure 6) [29]. A prebiotic world in which chemical reactivity was directed by modulation of effective molarities could have similarly accessed these expanded synthetic capabilities, enabling access to structures that could not have easily arisen from random intermolecular reactions.

A modulated effective molarity system for chemical synthesis can also dramatically decrease non-productive side reactions compared with its random intermolecular counterpart. Due to the high mutual affinity of cDNA oligonucleotides, functional groups can be present at much lower concentrations than those required to drive unassisted intermolecular reactivity. This lower concentration requirement can enable highly reactive groups to remain intact for longer durations without being quenched by unproductive, non-templated reactions.
with other functional groups. The use of nucleic acid templates to direct early translation events therefore implies expanded chemical capabilities beyond simply serving as a program for translation.

Conclusions
Recent studies in DNA-templated synthesis have established systems that chemically read out the information in a DNA template, transduce this sequence information into the synthesis of diverse reactions products over multiple reaction cycles, and release the translated synthetic structure, all without the assistance of protein or nucleic acid enzymes. Although most of these studies were not undertaken with the intent to recapitulate ancient translation, the properties of nucleic-acid-templated synthesis illuminate key obstacles and their possible solutions encountered during the fundamental process of converting information into functional molecules.

Acknowledgements
We are grateful to Professor Gerald Joyce for helpful discussions. CTC is supported by the National Institutes of Health (NIGMS R01GM065865) and the Office of Naval Research (N00014-03-1-0749).

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest


A comprehensive review of developments in nucleic-acid-templated synthesis through early 2004, detailing advances in the physical organic understanding of nucleic-acid-templated synthesis, the scope of the method, and several recent applications.


This report presents the first examples of small molecules synthesized through multiple DNA-templated steps. By using a biotin affinity purification strategy with chemically cleavable linkers, the authors demonstrate the translation of DNA sequence information into a non-natural tripeptide and a branched thioether.


30. Gartner ZJ, Grubina R, Tse BN, Doyon JB, Snyder TM, Liu DR: DNA-templated organic synthesis and selection...


The authors describe sequential RNA-templated acyl transfers to synthesize a tripeptide in a single pot.


This paper is the first report of nearly quantitative, sequence-specific translation of a nucleic acid into a synthetic polymer. Using reductive amination, the authors demonstrate efficient chain length-specific translation of a DNA octamer into an octathymidine amine-linked analog. Polymerization is shown to obey step growth kinetics, and in the case of amide-linked dimer substrates, polymerization occurs in an antiparallel sense.


