This document is for preliminary drafting only. All text will eventually be transitioned to LaTeX and a git repository. For now we are using the "colwiz" Google Docs add-on for inline citations.

[Freeze 2017-06-04 .. moving to github/latex .. do not edit this doc, your edits will not be reflected. Andrew P.]

Electronic manipulation of polymerase for DNA synthesis

Bryan Bishop <kanzure@gmail.com>
Nathan McCorkle <nmz787@gmail.com>
fenn lipkowitz <fennfoot@gmail.com>
Max ?
Rob ?
Andrew H ? Andrew P ? Mark ? Carl?

Abstract

In this review, we explore the concept of modifying existing DNA polymerase enzymes such that their DNA synthesis activity could be programmed and manipulated in real-time by external control signals including optical stimulation and physical electronic interface. We review a number of concepts and techniques which could be useful for carrying out such a project.

Introduction

All known forms of biological life make use of DNA and related nucleic acid chains to store biological information. In a way similar to how computer programming "controls" computer devices, DNA "controls" the behavior of cells and viruses, playing a role in biology similar to source code for computers. With over 100 trillion years spent "in production", DNA is one of the most widely deployed technologies of all time. From an engineering perspective, better control and synthesis of DNA would allow for many interesting projects such as the development of molecular nanotechnology and projects to defend against disease, memory loss, aging, and death.

The realization of these goals and ambitions will require the invention and development of cheap, large-scale genome synthesis techniques. Perhaps one day it will be possible through DNA synthesis to print billions of base pairs (many genomes worth) of DNA within minutes for pennies. Compared to the current millions of dollars per unique genome

synthesis project, $0.01 USD per 1 billion base pairs would be a welcome, dramatic improvement. With reduced cost, perhaps it would be possible to reduce the design-build-test iteration cycle for bioengineering to more closely match the iteration timelines for software development, where many ideas and variations can be tested within a single day. Without these cost reductions, programming errors and mistakes will continue to be expensive and preclude the possibility of inexpensive innovation.

One possible route to extremely low-cost genome synthesis techniques would use the polymerase enzyme, which is found throughout biology and biology labs. In mammals (including humans), cells are replaced at a rate of about a few million per second throughout every day, and the vast majority of these replacement cells have a copy of the host organism's genome. In the ocean, viruses cause the turnover of about 40% of the ocean's microbes every day, while the overall microbe content stays somewhat stable, indicating daily microbial replication and daily genomic replication. In the molecular biology lab, the polymerase chain reaction (PCR) technique makes use of trillions of units of enzymes for extremely low-cost DNA processing. Demonstrating existing viable DNA processing properties, in a number of different contexts, it is clear that the DNA polymerase enzyme could be an interesting place to start when looking for keys to the development of low-cost genome synthesis techniques.

Polymerase is an enzyme that catalyzes the polymerization of nucleic acids into molecular chains. Different polymerase variants exist throughout the biology kingdoms. Some polymerases only perform replication of existing template strands, while others also perform activities which include proofreading, error checking, error correction, and translesion synthesis. Less commonly, some polymerase enzymes are capable of de novo synthesis in the absence of a template strand molecule. Enzymes that are known to have polymerase activity can be found online by enzyme classification numbers (E.C.) 2.7.7.6, 2.7.7.7, 2.7.7.19, 2.7.7.48, and 2.7.7.49.

(explain more about the natural biology of polymerase enzymes)

(summarizing the introduction section) Controlling an individual DNA polymerase enzyme during every nucleotide incorporation event is currently an unsolved engineering challenge. This type of polymerase technology would be extremely useful to many fields of study including molecular biology, synthetic biology, and molecular nanotechnology. In biology, large quantities of DNA are required to conduct experiments and cure diseases. In molecular nanotechnology, there is a strong requirement for automatic precision positioning of individual atoms and molecules, which can be greatly facilitated through the use of DNA. This controllable polymerase concept tends to capture the imaginations of biologists and engineers alike, but, to date and to the best of our knowledge, only a tiny number of publications have ever discussed this concept. Recent developments in molecular biology, protein engineering, chemistry and other fields might allow for the design, engineering and construction of an electronically-controllable polymerase.

In particular, one DNA sequencing method involves stringing up a polymerase enzyme between two microelectrodes then using voltage, current, conductance, and inductance measurements for determining the identity of a single nucleotide for each incorporation as a style of sequencing-by-synthesis. These and other technologies will be particularly useful for constructing an electronic polymerase.

Besides DNA, there are other synthetic polymers that could theoretically be constructed by controlled polymerases. Molecular nanotech can bring sequenced polymer construction for many other types of materials and modalities. Indeed there are some polymerase enzymes that have already been heavily modified for XNAs, TNAs, LNAs, etc.

Another interesting application is long-term high-density data storage [1]–[3]. Data storage in DNA, needs fast read and write times. Direct electronic control of polymerase is ideal method for achieving these goals. DNA has a long storage time and high density, although read/write will continue to be slow compared to flash memory tech. Large-scale parallelization of DNA polymerase technology could theoretically match flash memory tech through multiplexing of billions of 'different DNA molecules simultaneously.

Biology is an interesting opportunity for short-term progress in molecular nanotechnology. Usually molecular nanotechnology is imagined as diamondoid molecular nanotechnology, such as in the Drexlerian molecular nanotechnology concept [4]. However, there already exist many self-replicating systems in the form of biological self-replicating cells.

(Include here an overview of the remainder of the paper)

Goals

In vitro: The assumption for the most of this paper is that the artificial DNA polymerase will be for in vitro usage. This is expected to be an environment such as a test tube, or some MEMS/CMOS scaffold that interfaces with the polymerase. It is assumed that the polymerase will be meant to function in an aqueous environment. In vitro, polymerase could be attached to a flat surface (such as a CMOS semiconductor microchip) or to the surface of beads, or have no tethering at all and be free floating in droplets, emulsions, or other liquid environments.

In vivo: One alternative is that a polymerase could be made that operates in vivo. There can also be a mix of methods where the polymerase is evolved partially outside of cells in vitro, and then for a second round of selection it is evolved inside of ecoli (or something similar) in cell culture. One proposal is to use a toxin that can be defeated by giving a trivial sequence to the ecoli. Then, use optical stimulation of the ecoli to send signals to the polymerase to construct a certain antitoxin molecule or protein.

In vivo methodology has some other interesting benefits. Usually the purpose of DNA synthesis is to construct DNA and then express the DNA inside of another organism. If we are already going to be handwaving about the construction of a magical polymerase that conforms to our high expectations, then why not also handwave about some magic to make it also work in a cell by purely optical phenomena? This would allow for the construction of DNA at the destination where the DNA is going to be used. This has many advantages such as a reduced reagent cost, a self-renewing system that only requires agarose or other cell culture medium, and can completely skip sequencing by checking for cell survivability when exposed to antimicrobial selection. An interesting example of electronic control over cellular activities is the field of optogenetics, where real-time optical stimulation is used to control the action potential firing of neurons in vivo [5].

Long-term performance of an electronic polymerase .... ultimately our goal should be to construct entire synthetic genomes in a matter of minutes. Using arrays of thousands or millions of polymerases, we can construct many billions of nucleotides per second of novel synthetic DNA. The cost of DNA synthesis should be reduced to be pennies per genome. Making a mistake, like not correctly writing an amino acid sequence for a custom protein, should not be a multi-million dollar mistake for synthetic genome engineering. The design-build iterative cycle needs to be shortened as much as possible.

In vitro methodologies

MEMS and CMOS

Microfluidics

Functionalizing microelectrodes with enzymes

Directed evolution techniques have already been used to modify polymerase function. Examples include modifying the types of substrates incorporated into DNA, increasing the range of temperatures and environmental conditions under which the enzyme can operate, and so on[6]. In directed evolution, multiple rounds of replication and selection are applied against a selective pressure. Directed evolution techniques can be used for aptamers, proteins and other targets.

Compartmentalized self-replication is a technique for the directed evolution of polymerases where each polymerase is isolated in its own individual aqueous compartment, then provided with primers, nucleoside triphosphates, and the genetic material necessary to specify the included polymerase protein. Each compartment starts with one e.coli cell on average, which expresses the polymerase enzyme in each of the isolated compartments. [7] This technique has been used with polymerase to increase Taq polymerase thermostability[7], Taq polymerase resistance to the polymerase inhibitor heparin[7], [insert

some other developments] Additionally, some techniques also couple compartmentalized self-replication with yeast cell surface display [8],[9]

I**n vitro compartmentalization** (IVC), where a system of water-in-oil emulsion compartments, each containing in vitro transcription-translation reagents . This protocol allows for the selection in large libraries up to a size of 10^10 items per mL [10]. For a review, see [11]. IVC serves as an alternative to techniques such as phage display[12], ribosome display and mRNA display[13], covalent DNA display, cell surface display[14], [15], and mRNA-peptide fusion[16]. Bead display plays a role in IVC methods. IVC, in contrast to compartmentalized self-replication (CSR), does not require cell bodies in each emulsified aqueous compartment. Intriguingly, IVC has been used to develop a ribozyme RNA polymerase which could polymerize up to 20 nucleotides [17]. One advantage of IVC techniques using microbeads is that a physical link between genotype and phenotype can be established where the microbead binds to both phenotypes (such as expressed proteins) and the underlying genetic definitions [18]. Recently, IVC techniques have been scaled down to sub-femtoliter droplet size using inkjet printhead technology [19]

Enzymes

We now explore a brief review of a variety of enzymes that could be used as chasses or frameworks for the construction and engineering of an electronic polymerase. The most intriguing prospect may be the use of Taq polymerase, a widely characterized and widely used enzyme from the field of molecular biology.

Other interesting options include TdT (terminal deoxy(ribo) nucleotidyltransferase), poly(A) polymerase (PAPs), and human polymerase mu which shares many similarities with TdT.

Some of the enzymes have an exonuclease and proofreading domain, which has exciting possibilities for the construction of a polymerase. For example, if you assume that there is a template in the aqueous media already, then you could use an existing template of the correct length to bootstrap the construction of another strand of DNA. Then, when the polymerase incorporates any nucleotide, trigger the proofreading or exonuclease functionality to cause the nucleotide to be removed from the strand and replaced with an alternative that matches the programmed synthetic DNA goal string value.

However, other than Taq polymerase, there are a variety of naturally-occurring enzymes that have interesting properties that might be relevant to the construction of an electronic polymerase. Various in vitro evolution methods can be used to take any one of these polymerases and push them towards a mutant form that might be more productively useful for the purpose of an electronic polymerase.

RNA polymerase is an acceptable candidate as well. Whether there is DNA or RNA produced, it does not matter. The goal should be to construct DNA or RNA, since both can be translated almost freely from one to the other using reverse transcriptases and so on.

Here are some worthwhile enzymes to look into:

* existing/known polymerases
* poly(A) polymerase (PAP), such as polynucleotide adenylyltransferase <https://en.wikipedia.org/wiki/Polynucleotide_adenylyltransferase> and polyadenine binding protein I <https://en.wikipedia.org/wiki/PABPI> and polyadenine binding protein II <https://en.wikipedia.org/wiki/PABPII> and PABPN1 (nuclear)
* poly(ADP-ribose) polymerase (PARP) (for protein modification)
* polyribonucleotide nucleotidyltransferase (PNPase) <https://en.wikipedia.org/wiki/Polynucleotide_phosphorylase>; polynucleotide phosphorylase
* GLD-2 (Germ Line Development 2), is a cytoplasmic poly(A) polymerase (cytoPAPs) which adds successive AMP monomers to the 3' end of specific RNAs, forming a poly(A) tail, which is a process known as polyadenylation <https://en.wikipedia.org/wiki/GLD-2>
* RNA-directed RNA polymerase
* DNA-directed RNA polymerase
* RNA-directed DNA polymerase
* DNA-directed DNA polymerase
* template-independent DNA polymerase (there is one that was found in 1997, which is not TdT) (vent polymerase)
* terminal deoxynucleotidyl transferase (TdT) <https://en.wikipedia.org/wiki/Terminal_deoxynucleotidyl_transferase>
* terminal uridylyl transferase (TUT) like <http://www.ncbi.nlm.nih.gov/pubmed/25712096>
* DNA nucleotidylexotransferase (might be same as TdT ??)
* tRNA synthetase
* telomerase (polymerizing activity)
* CCA-adding enzyme (CCAE) <http://enzyme.expasy.org/EC/2.7.7.72>
* kanamycin nucleotidyltransferase
* recombinase
* helicase
* replicase
* DNA ligase
* RNA ligase
* RNA ligase ribozyme
* RNA polymerase ribozyme
* reverse transcriptase (other than RNA-dependent DNA polymerase)
* nucleotidyl transferase <https://www.nlm.nih.gov/cgi/mesh/2011/MB_cgi?mode=&term=Nucleotidyltransferases>
* tRNA nucleotidyltransferase, RNase PH, reversible phosphate-dependent exonuclease <http://enzyme.expasy.org/EC/2.7.7.56>
* see also http://enzyme.expasy.org/EC/2.7.7.
* CRISPR things

DNA polymerase from *Thermococcus litoralis* (commercially available as Vent polymerase) can synthesize DNA in the simultaneous absence of primers and templates, creating DNA

up to lengths of 50,000 bp [20]. Later work found that, in addition to DNA polymerase, adding a thermophilic restriction endonuclease accelerated the creation of de novo DNA by creating primers which would then initiate replication by other Vent polymerase in cycles of digestion and elongation [21]. Vent polymerase, TdT, and other polymerases show that it is possible, in principle, to use these enzymes to synthesize new strands of DNA in a template-independent manner.

Family X DNA polymerases, including TdT and others, show template-independent DNA synthesis activity [22]. Human DNA polymerase θ (POLQ), a family A polymerase, performs translesion synthesis for DNA repair and shows template-independent ssDNA synthesis activity [23]

A family X member polymerase, human DNA polymerase mu, has activity similar to terminal deoxynucleotidyl transferase (TdT). However, polymerase mu also displays a preference for templated synthesis. Loop1 in polymerase mu has been found to conformationally change the shape of the enzyme to provide for both template-independent synthesis behavior and template-dependent synthesis behavior. When human polymerase mu Loop1 was substituted with Loop1 from TdT, the chimeric polymerase mu demonstrated template-independent synthesis activity with strongly reduced template-dependent synthesis activity [24][25]. Loop1 appears in the palm domains for both enzymes.

"... a deletion of the entire Loop1 as in pol λ does confer a limited template-dependent polymerase behavior to Tdt while a chimera bearing an extended pol μ Loop1 reproduces pol μ behavior." [25]

In pol λ, amino acid residue Phe506 is essential for template-independent synthesis activity [26]. Both Pol λ and pol μ belong to the family X DNA polymerases. Pol λ, pol μ, and TdT play roles in DNA repair and DNA recombination.

Polymerization rates of polymerases under physiological conditions can be as high as 10^3 nucleotides per second, with one error per 10^7 nucleotides replicated. Polymerases with proofreading and error correction abilities can reach fidelities of less than one error for every 10^9 nucleotides processed.[27]

Unnatural nucleotides have been proposed for modifying, replacing, and extending the existing natural genetic alphabet.[28][29][30] In addition, unnatural nucleotides have been used to generate aptamers with high binding affinity.[31]

Also.... here is how each of those enzymes seem to work.

Enzyme engineering and rational protein design

There are many techniques available for engineering mutant polymerases, including:

* site-directed mutagenesis
* site-directed chimeragenesis [32]
* chimeric polymerase routines
* compartmentalized self-replication (CSR)
* in vitro compartmentalization (IVC)
* emulsion methods
* directed evolution methods
* in vivo evolution methods
* phage display [33] [34]
* ribosome display [35] [36]
* mRNA display [37] [38]
* DNA display
* cell surface display
* bead display, microbead display, gold nanoparticle display, etc.
* SELEX

Many of the surface display techniques use a cell-free translation system.

Here is a summary of some of the results of enzyme engineering methods.

Wild type Thermus aquaticus DNA polymerase I (Taq pol I) discriminates against the incorporation of dUTP and therefore cannot synthesize RNA. Using site-directed mutagenesis, targeting the dNTP binding site (motif A, amino acids 605-617), mutants of Taq polymerase have been found that show a 10^3 higher ribonucleotide incorporation activity [39]. At least two residues were involved in excluding ribonucleotides from the Taq pol active site, which were demonstrated by mutants that were found to have either a hydrophilic substitution at Ile614, or a substitution at Glu615, the second of which was always found to include mutations in other residues which seem to be necessary to maintain physiological activity of the polymerase [39]. Similarly, in the Klenow fragment of E. coli DNA polymerase I, mutation of the Glu710 residue enables the incorporation of single ribonucleotides, but further RNA synthesis activity is hindered by residue 762 thus demonstrating that multiple residues require mutation to enable RNA synthesis activity in what is otherwise a DNA polymerase [40]

In contrast, modification of only a single amino acid residue in Moloney murine leukemia virus reverse transcriptase is required to endow the reverse transcriptase polymerase with RNA synthesis capability [41]

Using phage display and streptavidin-coated magnetic beads, a selection experiment was able to evolve Taq polymerase to incorporate ribonucleoside triphosphates with efficiency close to wild type Taq polymerase dNTP incorporation rates [42]

Additionally, compartmentalized self-replication (CSR) techniques have been used to evolve Taq DNA polymerase I to simultaneously possess both DNA synthesis activity, RNA synthesis activity, and reverse transcriptase activity [43]

Rational protein design continues to have trouble related to simulation run-times. It seems that the folding problem has yet to be reduced to a set of heuristic known folds and known modifications that can produce predictable protein structures. There are some interesting exceptions to this, such as the development of protein engineering for linear repeat proteins[44]. Alternatively, instead of using protein engineering, it may be prudent to further explore the prospects of DNA origami and structural DNA nanotechnology [45]. Ribozymes, made only of RNA and ribonucleotides, have been used as an RNA polymerase. It seems reasonable to assume that structural RNA nanotechnology might "catch up" to DNA origami at some point [46] [47]. Alternatively, DNAzymes could be used in combination with DNA origami. This direction of research is interesting because it is computationally feasible to predict many DNA origami structures compared to the computational difficulty of protein structure prediction. Polyhedral meshes, like those often stored using the STL file format, can be converted into DNA origami as a set of DNA staples and nucleotide sequences [48]. Also, both DNA and RNA nanotechnology might benefit from graph grammar approaches that focus on graph rewriting rules or shape grammars. DNA origami and structural DNA nanotechnology has also been shown to be compatible with cellular expression in vivo [49]

Control methodologies

Unnatural nucleotides could be useful for a number of scenarios. One is that, for DNA origami, unnatural nucleotides can serve to remotely trigger conformational shape changes. Another possibility is to use unnatural nucleotides during enzymatic DNA synthesis, either for detection of nucleotides or for other reasons, like blocking the entry of a nucleotide into a polymerase enzyme pores.

Unnatural amino acids could be used to remotely cause a conformational shape change in a polymerase enzyme, for a number of reasons. One is to include the same type of unnatural amino acid in multiple locations throughout the same enzyme. If the unnatural amino acid is a photoswitcher, then upon illumination of light you could cause two (or more) simultaneous conformational changes in the polymerase enzyme. Using multiple types of unnatural amino acids, you could have an array of different locations that change in response to specifically tuned optical wavelengths.

Contributions from optogenetics

Contributions from micro electrochemistry

Contributions from DNA nanopore tech, DNA origami in nanopores, etc.

Electric fields to control molecular machines

* Mechanical (polymerase)
* Mechanical (nucleotides)
* Mechanical (template strands)
* Mechanical (pressure)
* Mechanical (stress)
* Mechanical (contact)
* Electronic (current, voltage, ...)
* Optical
* Confinement
* Ions
* pH
* Inhibitors, both chemical and protein and others

Electric sheep dream about state machines

Assuming the incorporation of different control modalities into a polymerase molecular machine, what sort of external control would be usable? What is the minimum number of signals that could be used to control the machine? In particular, the goal is to minimize the number of input signals and the number of output signals, while also minimizing the amount of engineering work that goes into the protein.

In an ideal world, it would be possible to treat proteins as state machines whose state changes are triggered by external signals. It would be helpful to have a molecular protein-based counter that could increase over time through external stimulation. However, it seems like this is unlikely to engineer for now, especially since it is not a natural component of the polymerase enzyme. A state machine would be useful because then you could conscript the behaviors of the enzyme based on different states-- if this then that, if the current state is X then only the following transitions are allowed; and each interaction with other molecules would first check the internal state before applying an enzymatic catalytic reaction etc... But the truth is that an enzyme works the other way around; it's the current structure that causes behavior, so if you want a specific state then you need to change the current shape of the enzyme to prohibit other behavior.

How do you count a signal? One of the proposed tools is to have an unnatural amino acid that incorporates azobenzene, and is responsive to two separate wavelengths of light. In this scenario, we would consider the unnatural amino acid to have a single trigger, but multiple frequencies... Actually this terminology needs to be fixed. and what if the amino acid is incorporated into multiple places into the same enzyme? Should the total number of signals be increased because the same amino acid type was used in multiple places in the same enzyme? How do you count?

Signals in:

1 signal ---- A single signal could be used in many ways. One is that, after electronic detection of the nucleotide type that the polymerase is attempting to incorporate, the signal could be used for whether to incorporate that particular nucleotide. This one signal could also be used for whether to undo the previous incorporation, perhaps to proofreading mechanisms or an error-correction module. For example, for error correction, the idea would be to continue to signal to correct for an error after each incorporation event, until the polymerase happens to incorporate the correct nucleotide.

2 signals-- two different control signals, two different amino acid types (one can be azobenzene). This could be interesting for ...

Different behavior that could potentially be controlled by DNA polymerase:
* Start incorporation signal, and debug output signal
* Stop incorporation signal, and debug signal
* Ratchet forward signal, and debug signal
* Ratchet backwards signal, and debug signal
* Capture next nucleotide
* Select next nucleotide type
* Incorporate current nucleotide
* Ratchet template strand forward (if any)
* Ratchet template strand backward (if any)

Output signals that would be useful to detect from a polymerase enzyme:
* nucleotide incorporation event
* which nucleotide was incorporated
* which nucleotide is about to be incorporated
* move/ratchet forward
* move/ratchet backwards
* error correction behavior

Concluding remarks

Components

Enzymes
Unnatural nucleic acids
Unnatural amino acids
DNA origami and structural DNA nanotechnology
Ribozymes, DNAzymes and riboswitches
DNAzymes
Nanopores

AFM
TEM, etc.
Arrays
Optical imaging

Debugging molecular machines (fluorophores, etc.)

Evaluation of modalities

We next evaluate various physical modalities available to digital polymerase design.

Mechanical (polymerase)
Mechanical (nucleotides)
Mechanical (template strands)
Mechanical (pressure)
Mechanical (stress)
Mechanical (contact)
Electronic (current, voltage, ...)
Optical
Confinement
Ions
pH
Inhibitors

* short review of chemical phosphoramidite oligonucleotide synthesis
* short review of the source of errors in phosphoramidite oligonucleotide synthesis
* a description of the problem with DNA assembly reactions, why we should prefer to directly write long DNA

* sequencing by synthesis
* nanopore sequencing
* nanopore stuff
* Oxford Nano are using a protein-based nanopore; repeatability, semiconductor nanopore.

* ribozyme polymerase
* DNAzymes
* DNA origami with functional properties

No DNAzyme yet with polymerizing catalytic activity. You are getting less electron density from the DNA than you are from the RNA. So it's less of a pull to pull something over a reaction or something. Less voltage. We can add more unnatural nucleotides if RNA can't do it and if DNA can't do it. Enzymes are more abundant in this catalytic space, it seems like, one they are bigger which makes it easier for us to find and capture and utilize, and more stable. But there are at least 20 amino acids so you have more variability in your solution

space as far as coming up with a configuration that does the electronic reorganization mechanism that you want.-

Someone should compare ribozyme polymerase to protein polymerase mechanism of action. Are they the same?

Instead of gears, it's electrons. You pull some electrons on one side, that causes a void, it wants to tug some electrons from an intermediary, and finally the other end of the enzyme has some electrons being pulled and then it does catalysis on the active site or something.

Azobenzene unnatural nucleotide. Or maybe a poisonous conformational change that blocks its entry into polymerase? Pulsing a laser to heat up an enzyme; well, it seems like it it might not be able to heat up and cool down to really high control over the activity. But potentially yeah being able to do a laser or an LED or whatever. You can get super fast switching times because that's how fiber internet works.

Azobenzene unnatural amino acids conformational shape change in proteins. Throw that into TdT with the idea Nathan proposed. You don't have to worry about diluting your solution for nucleotides like crazy. You just flash your femtosecond laser one pulse, send a single photon or whatever, I don't know. How would you do it? You want to only activate it for only a nucleotide worth of time. Well, we could have two signals, one to stop the polymerase, the other to incorporate and choose a nucleotide. Send one color and it causes the enzyme to become active, and then you want to immediately send deactivate signal and widen your gap until you only got a single addition.

As long as you can activate and deactivate faster than the enzyme can polymerize, then it could work. So you need a really slow polymerase. But your idea Bryan is more of a lock-step with every step being under control. If there was electronic control, then you could read real-time feedback. We're going to have to get this ribozyme or peptide chain enzyme synthesized, with unnatural nucleotides or unnatural amino acids. So if we could come up with the next best thing and then use that to start, basically increasing our iteration speed.

We should minimize the amount of work we have to do with changing the protein design, like for ratcheting or something, or lock-step stuff.

Debugging an enzyme. Maybe fluorescence. FRET. Maybe an amino acid that has a fluorophore or something, and a quencher. Or maybe a set of amino acids that can fluoresce like GFP. When they are together, they quench the fluorescence. That's basically what FRET is. If you have an enzyme grabbing a molecule, you can have a quencher on one side, and they will be fluorescing when the gripper is apart, but when gripped together, the quencher is adjacent to the fluorescent, and then it quenches and that's within the enzyme physically. So you could probably look up multi-channel FRET to see if people are ever using multiple colors within the same enzyme for different active sites or different functional subunits. You would probably also want microwells or nanowells, because otherwise you would have multiple fluorescence enzymes in the same, in the same pixel of your detector. A lot of this is single-enzyme with one enzyme. At the stage of building this debugging

framework or building the feature for debugging, you would need to test it in isolation. Whether you would also apply it in isolation in the functional system well, that's something else. How would you develop this debugging activity before actually using it? But yes in application it's going to be single molecule enzyme stuff. For things that single molecule is hard to do or too expensive or not feasible for practical reasons, then maybe statistics will save us and machine learning and deconvolution of data. OpenCV, machine vision, that's the solution always no matter what.

Double-wide nucleotide polymerase stuff ? What makes them big? Nucleotides in sperm shape vs umbrella shape. Maybe the polymerase could incorporate those nucleotides when they are in the "off" state. Could a chemist do this? Is it reasonable to ask them to take this on? Will it only be a viable option if the device that uses that nucleotide is also a success story? Perhaps click chemistry or something. If the synthesizer we come up with is so amazing that everyone buys it, and they also need the special nucleotides, it might become a viable ecosystem. At that point, is it easier to synthesize your weirdo polymerase, and you just have it sitting in a plasmid whenever you need to synthesize it, versus your synthetic chemistry? Well, bootstrapping version might have to use synthetic nucleotides, but after that, who cares? Cooper Union at iGEM... they were saying, we got this source, from some chemist, is a photocleavable terminator nucleotide, it's a terminator nucleotide until you flash light, and then it's no longer a terminator and you can keep incorporating on to that strand. They were using that as their ability to use TdT to only add one nucleotide. And these nucleotides I think they were around for a while, but I know that commercially, that one sequencing company that went out of business but their sequencers are still being maintained and used, Helix maybe? Not CustomArray. I googled this a week ago. True single molecule sequencing. Helicos. They were using the same type of nucleotides for their sequencing reactions where they were not only photoinducible terminators, photocleavable terminators, but also they had a fluorophore. They were incorporating it, blocking the reaction, the nucleotide with a terminator fluorophore on it would get into, it was sequencing-by-synthesis, so they would take a pic with a good lens camera high quantum yield like cooled CCD camera or whatever, and then they would shine light on the whole array which would cleave off the terminator which had the fluorophore. And then they would flow in A, then they would check for pics, and if A got incorporated, then that microwell with that A incorporation would be fluorescing. Then they would cleave and so on... and at every step, only one nucleotide gets added. But it's fluorescent as well. It sounds like the polymerase is detaching. It's tethered at the bottom of a microwell or nanowell. It's probably less than the wavelength of light that they are using, so they are using a near-field effect. Basically the only light that is making it through the microwell from below to above to the camera, is anything that is happening actually in that microwell, at least as far as my understanding of "zero-mode waveguide" applies, and I might have a wrong understanding. I've been trying to understand that for 7 years, so maybe my understanding is better now than when I first read it, maybe. But yeah, basically they shine light from underneath, there's a tethered polymerase at the bottom of a well, they put a single ssDNA into there, they have terminator nucleotide fluorophores, and they take pictures and cleave in between flowing back in. This big company, and had this double refrigerator sized machine that they were selling for however many millions, they were you know, amazing tech when it comes to people doing sequencing and true single molecule sequencing, and you know, very different

error profile than all the other companies. They went under as far as their company was concerned. And you know, I don't know why that is, if the nucleotides were too expensive, or because it was complementary then perhaps it was only used in certain situations where Ion torrent maybe wasn't working, they had to go to Helicos or something- but only in those circumstances? I don't know why they failed. I don't think they made terminator nucleotide fluorophores any cheaper, I don't think there was a breakthrough there. There are several other DNA sequencing companies, so perhaps being in a different market would have different impact on reducing the price of the cost of the nucleotide, or perhaps prevent the product from taking off. From a really high-end, I think price is probably no object for companies like BP or whatever that want algae that poop diesel as long as it's cheaper than what it is today from Twist Bioscience or whoever the latest $0.10/bp amazing tech company is out there advertizing. Lowest price is probably $0.03/bp. They are not including assembly for those $0.03/bp prices. I think that's for the actual synthetic product, that's not for taking it and making a gene out of it.

Fluorophores are very interesting; maybe some new fluorophores. Probably a lot of people focusing on that. Talk to that guy who used to work for Eureka. Rob? He's working for Google now. Sequencing by synthesis, there's probably lots of fluorophore stuff that would be useful. Josh Perfetto, maybe. Synthego people perhaps. Life Technologies retiree down in Eugene... qPCR machine fluorophores. Perhaps there are nucleotides that are genetically fluorescent, perhaps it's something that has been encoded into polymerase enzymes and nucleotide synthesis enzymes. The proteins would have to do some chemical reactions to add those changes to the nucleotides. Or maybe an enzyme that just adds azobenzene to each nucleotide.

Surface plasmon resonance being used for detection- Dylan Morris/Robb Walters- it seems like his company if it still exists, it's probably the same super stealth webpage that has been since ever online, and he's the only person I know trying to develop SPR. Maybe they pivoted though. It did not seem trivial. We need to keep it simple.-

DNA origami that changes conformation, in response to light or stress or pressure or mechanical movement or electric fields, such that it makes relevant movements that could help with DNA synthesis and polymerization. Novel linkages in the molecule.

* riboswitch stuff?

* azobenzene alternatives
* azobenzene unnatural amino acids
* azobenzene for ribozyme modulation
* azobenzene dNTP modifications and permissive polymerases that can incorporate dNTPs with azobenzene etc. See "Highly accurate synthesis of the fully 2′-fluoro-modified oligonucleotide by Therminator DNA polymerases" http://www.sciencedirect.com/science/article/pii/S0960894X14002510 -- perhaps photoswitching could be used to make a dNTP that is incapable of incorporation by polymerase when switched in one mode, and capable when switched in another mode? Azobenzene incorporation is mentioned in "Modified nucleoside triphosphates for in-vitro

selection techniques" 2016 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4854868/ ...
* short review of chemistries incorporated into unnatural amino acids, other than azobenzene
* click chemistry stuff, re: oligonucleotides, nucleotides, amino acids, proteins and ligands and binding and catalytic activity, and phosphoramidite techniques

Click chemistry. It's a polymerization mechanism that is really well known. Applying that mechanism to a generic molecule that you might want to polymerize. You can add a "click API" to a molecule, then you can click it. It would almost be like how biotinylation with streptavidin, a classical generic tethering solution for molecules; if you can add a biotin site or streptavidin, then you can use biotinylation to bind and fix stuff. If you can put a click site on your molecule, then you can use click chemistry to join them. CMU professor that does polymer stuff. This is atom transfer radical polymerization, not click chemistry.-

* opsins and rhodopsins, for photosensors inside of proteins
* conformational change in proteins
* short review of rational protein design
* graph rewriting rules for automated design of ribozymes and proteins ?

Catalytic site rational design, seems to work by copying catalytic domains from other proteins. This seems to work alright. It seems like you have to express the domains in the right place, depending on the overall shape of the protein. If you can imagine how the electron densities move around the 3d space if you can imagine the polymer as a semiconductor, and the charge doesn't want to move from one place to another but it could be forced to, if given a certain buffer or another molecule, or salt content in the buffer, maybe it does become more obvious that yeah we can put more charge delocalization here by switching from one amino acid residue to another one here, and that will totally lower the activation energy, or increase the catalytic rate. If you pull electrons away, then they are going to be less likely to offend some incoming molecule and bounce it off. Think about electron clouds like magnets. You shove opposite ends of magnets together; it's basically like that with electrons. Positive ions only go to negative ions, and it's a spectrum of how strong those interactions are. So if you can gauge the relative strength, we have lots of words for what the strengths of reactions are, like van der Waals, we can call them ionic or covalent or maybe other words to describe electron pull strength, but yeah maybe at some point this just becomes intuitive and they can I don't know.. Maybe computers are good enough, algorithms are good enough and they can do hybridized approach. And then all the people that played the FoldIt game, who are really good at folding now.

Coiled residues... and spirals.. so if we have to do rational protein design can we just restrict ourselves to the folds that we already know in the literature? Do we really need the full spectrum? Can we still get our enzyme to work? Can we add a catalytic domain from polymerase into a coiled amino acid? Are there coiled amino acid proteins that bind to DNA? Add the catalytic domain and then see what happens. Would we try to use coil structure to approximate the other one? Or would it be totally different? We still need the charge localization to be pretty similar to bind DNA and do polymerization. Perhaps we can't do the same charge localization, but perhaps we can use some other charge localization scheme of

some kind.

Enzymes that undergo large conformational change? Motor proteins. Flagellar motors. Have a large conformational change that causes something to block the nucleotides from going into the polymerase nucleotide pore. This might cause a driving force that causes the polymerase to drive into the wall or away from the DNA or something. Maybe an amino acid residue thing that blocks the pore. Or maybe it just makes it unfavorable for nucleotides to stick around. Just turn off the attractiveness. Stop attracting molecules. Instead of blocking, just pull the active site away or something, pull the nucleotide binding domain, move the fingers further apart so that all the stickiness is not concentrated in one locality. If you move the sticky zone just proximally away from the catalytic zone, then who cares if nucleotides are still binding, they are just so what, they are binding but, it doesn't matter until the enzyme is in the active in terms of the sticky zone moving into the catalytic zone. Well, you would have to make sure the nucleotides don't get stuck and then moved into the catalytic zone if the nucleotides are the wrong ones. If barrels and coiled proteins are the most predictable, what blocking things can we do, can we pull things away, which of those is a coil, and then let's hack it in that direction.

Motor protein review - how many different motor mechanisms, what are their requirements, can any of them be optically actuated? What are their requirements for function? Are there any proteins that incorporate motor components that are not themselves supposed to be motor proteins, for example proteins that use motor components for conformational shape change?

Conformational shape change in response to a hydrophobic response... to cause shape change.

Opsins and rhodopsins. All the optogenetics stuff. Light-gated ion channels. Gated nucleotide channel that happens to be attached to a polymerase instead of the membrane. How small is that switching motif? Is it 10 amino acids? Or is it 100 amino acids to achieve photoswitching? How big of a spatial constraint does adding that into a polymerase? Can we even fit into a polymerase or into the binding site area, or would it be way too big that we couldn't actually get that in there? Azobenzene, that is small enough, it's smaller than an amino acid, or it's on the same scale even or it could be incorporated into an amino acid. Maybe we need a bunch of azobenzene molecules to get the right amount of twist. Do we need multiple opsin motif or something to cause this to twist or achieve enough movement in the polymerase? How are they achieving this? Is it some specific tryptophan and some other thing? Is there some repeating structure that gets twisted and maybe those structures cross-talk from opposite sides of a barrel, and tryptophan binds to guanine or something?

GFP shine light to start, shine another light to stop the fluorescence. Use a quencher in some cases.

* photomodulation of ligands
* photomodulation of free-floating inhibitors/cofactors

Cofactors. One-use co-factors, which then break away and float downstream. Maybe they are one-off, they attach to polymerase, then they break off and get degraded by light.

Dual tethered proteins. Where the cofactor is tethered as well. And you spin two spinners, such that they intersect and get tangled up.

Photo-inducible solvability changes. In one configuration, the molecule is very soluble, and then when you pulse some light, it forms some crystals or something because it becomes, you hide some polar amino acids with a configuration state, and water then doesn't like it as much anymore, so it starts to conglomerate or something and forming bulky things like crystals. That could be useful for purification perhaps. You flush the cofactors, switch some light, crystalize it and then centrifuge them to clean it out or something. Or this might not work; you might be trying to move your binding domain away, but then you're crystallizing and blocking the microchannel or something.

* pH modulation, buffers, salt concentration, small molecule cofactors, altering solubility, water being attached to things that masks or demasks or something, capping and decapping enzymes and such.
* thermal modulation

Being able to modulate the pH rapidly I think is going to suffer thermal modulation... there are photo-acids, those are pretty common in photoresist industry and some of the DNA synthesis papers use DLP stuff with photogenerated acids. So you can do modulated pH control in that way. And then those molecules are going to diffuse and stuff.

* electrical current
* ... other physical modalities to potentially use for controlling a polymerase
* physical isolation and encapsulation of nucleotides in attobubbles
* nucleotide dilutions (single nucleotide in a bubble)
* physical placement of nucleotides near a polymerase using AFM tool tips
* moving a bead with DNA and polymerase attached into regions with nucleotides

Dilutions... well, the things I have looked at are single-molecule manipulation or detection or synthesis. We definitely have single molecule detection. Helicos DNA sequencer, they were literally using a fluorophore at the end of the single molecule of DNA and they had a really nice cooled camera and a good optical microscope and they were taking pictures of it. We have other DNA dyes. And maybe even Helicos is using the same dye. You can see DNA being electrophoresed in a single nanochannel that is 100 nm wide. We have cameras that are good enough and fluorophores that are good enough. There have also been electrode detection methods using a signal generator, scanning a frequency across soe electrodes and then see if the frequency response changes in response to floating molecules between the electrodes. A lot of research has gone into that for chemical detection and toxin identification and things like that as well as there has been some single molecule detection of nucleotides using raman spectroscopy. You shine a laser in and look at how a given volume of stuff alters the frequency of that laser in a very very tiny way, like 10ths of hundreds of a nanometer in wavelength alteration. So a very good laser at the front-end and

decent optics to block the laser when looking at what comes out the other side, because you will get swamped by the input laser signal when you are trying to look at just small variations, the dynamic range is just too great, so block out the original laser so you only see things that aren't the original laser. You can detect single nucleotides coming into and out of a reaction zone that has polymerase in it. They were trying to do sequencing by synthesis. Flow in a single nucleotide; or maybe it was some very low concentration nucleotide; it could have been very low concentration and comparing the relative input concentration against relative output concentration, like if you saw 10 nM going in and 9 nM going out, then the polymerase must have absorbed some of it into a growing molecule, and they were trying to do base calling based on that. But my takeaway was the detection scheme, like raman and that sort of thing. It has been in the academic world for a long time, it's been commercialized for what seems like a decent amount of time to where you can get a good raman spectroscopy setup for like $5 to $10k. Probably a cooled CCD and you get into the 90% quantum yield for photons in, electrons-detected out or whatever. And then there's a bunch of enhancement techniques where if you resonate the tip of your sample at some frequency, physically resonate at the tip of a crystal oscillator to synchronize your laser with that, get a lock in amplifier effect and it goes up even higher for things like raman. So raman is an interesting technique in general. That's kind of actually driven my interest in high-performance analog stuff. Analog validation group.

* how in vitro selection works
* how directed evolution works
* selection experiments (like SELEX and in vitro selection) for using directed evolution to get a relevant controllable DNA polymerase
* long DNA synthesis as a precursor to controllable DNA synthesis, e.g. improving template-independent polymerization activity
* improving template-independent polymer initiation
* statistical methods for searching for moderately controllable polymerases, e.g. by using light/optics/illumination and then checking if polymerase output was influenced in any way
* basic principle of first searching for a polymerase that incorporates only one kind of nucleotide (for each type), then taking all 4 polymerases and using rational design to combine them

We should change one of the polymerases to incorporate only a certain nucleotide type, by looking at the catalytic sites of some of the defective mutant polymerases. So we might already have single-nucleotide-type only polymerase incorporation.

* an analysis of the different potential signals to control a digital polymerase
* 4 different nucleotides need to be selected, so 4 different external signals? could be 4 colors?
* alternatively: use a single external signal to switch between all 4 nucleotides. This signal can either modulate dNTPs or it could module inhibitors that bind to dNTPs or it could modulate polymerase to switch between 4 different matching modes (which could be bad because polymerase might not transition during one event, and then you have an "off by one" error during synthesis.....)
* a signal for stopping polymerase, for making polymerase move one step forward, a signal

to prevent polymerase from moving, a signal to make polymerase incorporate one nucleotide, a signal for selecting which nucleotide for polymerase to incorporate

The fastest method would be just switching a light and keeping up with the polymerase enzyme, as a way to drive the enzyme, without stopping it at any step. Keep it running at a constant speed. Calibrate to some fluorophore output or something. And then you would use different light signals for selecting different nucleotides. Also you would want to change the concentration of dNTPs in solution to approximate the final nucleotide usage of the long DNA molecule you are synthesizing.

The concern was the random incorporation of nucleotides while you are not flashing the light.

Start incorporation signal, and debug output signal
Stop incorporation signal, and debug signal
Ratchet forward signal, and debug signal
Ratchet backwards signal, and debug signal
Capture next nucleotide
Select next nucleotide type
Incorporate current nucleotide
Ratchet template strand forward (if any)
Ratchet template strand backward (if any)

Perhaps the polymerase would only capture the nucleotide while the light is shining? Or is it something like, ... if our mouths were the active sites, and our hands were the binding site, your hand is going to find lots of apples in the apple store, but you are only going to take a bite when your elbow ratchets your hand towards the face. So the elbow jerk toward your face would be the result of a photon hitting your elbow. Your elbow is usually extended to grab an apple. Then a photon comes in and twists some azobenzene or other reactive domain if they are small enough to fit, and then you know, brings that hand close to your face hopefully having an apple in it as well.

Dilution and temperature of the nucleotide matters, nucleotides might be hitting the polymerase every second, or maybe not if you have a low concentration. If ratcheting was a 5th color, another signal, could you just keep hitting add nucleotide add nucleotide, and then after a nucleotide gets added... there's an apple in your mouth, your hand has an apple and your teeth grab it, and your teeth are stronger, so you keep the apple. Your hand goes back, but the apple stays in your mouth, so now you are holding the apple to your teeth. If you bring another apple to your mouth, your hands are just going to bounce off. Until you do a swallow command, which is your ratchet command signal, you won’t be able to accept another nucleotide apple. Maybe I should use grapes as an analogy, because you can hold a grape in your mouth and it's not hanging or something, and not swallow it. So yeah maybe trying to reload a nucleotide, after already loaded, maybe that's okay because it bounces back and it's a null, a no effect operation. So then you say okay, now we have pulsed the add operation enough ties that we are convinced we have an add operation. Now ratchet, and we can be pretty convinced that the active site is not empty. It could also be that the

chemistry requires that... imagine a spot welder or something, maybe the active site can't be touching the molecule because if it was then it might detach, or the active site might lower the activation energy in a reversible manner. So you actually need to load a molecule, bring the active site down to the area, and then it might catalyze that polymerization and then it pulls back, it springs back or something. I don't know the specifics enough about polymerase... how many precise movement actions does it have?

Is there an augmented reality for looking at polymerase animations? Some kind of a protein animation thing, with a trackball, and a cardboard interface and have a 3d visualizer that knows about the 10 different conformations that we have crystallography and tomography data for, and almost see a movie of the different conformational states, the different jiggling... almost like, let's watch the movie of polymerase doing its thing naturally. And now once we have a movie of all the crystallography steps, and then we can call the different movements "reload" or "ratchet" or "polymerization" or something.. And then there's ratcheting, which is separate from reloading. We can enumerate all of those movements.

Can we add azobenzene to a certain nucleotide? To a certain amino acid? Yeah I guess we must have that kind of control.

Use an e-beam to charge a surface flow nucleotides over it, write the charge on some surface, maybe on a carbon nanotube that has the same parameter as a DNA helix. And then deposit the charge you want with an e-beam, and have a little lathe inside your STM microscope, and as you rotate the carbon nanotube, you deposit the exact charge and somehow it just gets the nucleotide, that's your false complementary strand or something, and then you know, yeah, just let nature do its work. Then the question is well, okay, how, if we can write charge that finely, then how do you get it to actually stay when you flow aqueous nucleotides over it without the water stealing the charge localization and averaging it all out? Biotinylated on a surface, ... well that's too large. A complementary strand, you need to engineer the, just the piece of DNA that binds with Watson-Crick pairing, just a couple of atoms charge difference away from each other. Even with the best insulators, if you add some electrons on the crystal endpoint of the atom, like they won’t jump over to a nanometer away to another crystal endpoint or whatever. Maybe that idea is totally outlandish, but maybe your insulator instead of being an insulator, maybe it's a material that switches state, where at 30kV e-beam it keeps it, and you change the voltage on the e-beam and maybe you're doing some almost like imaging of molecules but in a controlled way where it's not just a charge floating on a surface. It's a lot of engineering.

* using GFP and other genetically-encoded fluorescent protein domains as a way of communicating information about the internal state of the polymerase enzyme

* evolving a polymerase enzyme where there is a large distance between where the template strand is located and where the new strand is located, such that the polymerase enzyme must change conformational shape internally (or some other internal state other than shape) to control the nucleotide incorporation at the distant site

* search all protein motifs for domains that match any of the catalytic domains from any of

the above proteins; are any of these proteins that are undocumented, are any of them structurally unique compared to the common classification of these proteins?
* * note that this technique was partly used to make a library of mutant chimeric polymerases

* a short review of the mechanisms of action of each of the above polymerase enzymes

* our cells make long DNA all day long, amazing performance of Taq polymerase and other polymerases, cost-effectiveness of a single polymerase enzyme
* estimations regarding length of time to write an entire genome with a single polymerase enzyme
* estimations regarding total cost to write an entire genome with a single polymerase enzyme

* how long can polymerase sit on a DNA molecule and not move?
* how many template-independent polymerases are there?

* nmz787's TdT ideas, filter mesh, etc. (see recent enzymaticsynthesis mailing list posts re: "DIYBIO DNA synthesis" subject line)
* cathal's telomerase ideas (see enzymaticsynthesis mailing list posts from cathal)
* light-directed RNA polymerase ideas (see ref from 2011)
* hessel's electric polymerase ideas (see ref from 2010)
* wittig's beta clamp ideas (he's been contacted)
*-

* microfluidics

What about synthetic chemistry with phosphoramidite chemistry and oligonucleotide synthesis?

Liquid phase DNA synthesis using phosphoramidite synthesis. Liquid phase was abandoned a long time ago because it was harder to keep track of everything. It was harder to purify because you have to change the solubility in a way where you can collect your product aside from the reactant or whatever. You mentioned click chemistry ,which is a synthetic technique, but could phosphoramidite chemistry be replicated with something more sane, more motifs? Rather than the one on the chemistry they chose, what about something switched so that we can get a chemistry not receptive to water contamination? By switching your basic chemicals or something. You could conceivably see that tech as being significantly optimized and scaled down on to a chip-level. We still have to get it from somewhere. Someone has to grind up dirt and do an extraction on the minerals to get some special metal ions.

If this is in vivo, you shine light on it to make it clear state, delete all flash memory, then you manufacture DNA inside of the cell. And it constructs its own dNTPs. You don't even have to electroporate. When it fails, you just delete memory and try again. It minimizes the external material requirements. There's definitely some.. a lot of ways to think about this.

One way to sell this as something other than just DNA synthesis--- you could sell this as a polymerization machine, for sequence-programmed polymers, for other industrial applications. So you could have an engineered polymerase that makes other forms of polymers. You could do polymers for information storage. You could polymerize silk, for example. Spider silk goat milk GMO goat from 20 years ago or something... not sure how much industry was involved in that, maybe it was purely academic idea or something.-

---

.. Drew Barry, videos about cell metabolism

we could do some visualizations of what the polymerase would look like
in... blender stuff
... it could be submitted to nature or something

A thing that looks like nucleotide, then polymerase grabs on it, but it jams up. But when you shoot light at it, you knock it out of the enzyme and it still works again. That was a way to stop the polymerase. You could add, so, proteins are hold together because they are folded up. If you yank on them from both ends, then they unravel (mechanical yanking). One way to yank on them is to just add energy, like in the form of visible light radiation or infrared radiation. If you have like chromophores attached to either end of the protein chain, then you apply a lot of energy, then the protein will denature. Might not be a permanent denaturation, i.e. Taq.

You need a polymerase clock signal. For every nucleotide, it has a tail that prevents the chain from extending. To advance to the next nucleotide, you just add ultraviolet radiation and it cleaves the tail. And then it's just a normal nucleotide and the polymerase could continue advancing. You would need to only cleave it once the nucleotide is inside the polymerase. Fluorescent events to debug information, enzyme fluoresces for a while after incorporating a nucleotide.

You could have a signal to detach from the DNA molecule. How often will we dissociate from the DNA molecule? Can we find a polymerase that stays attached for a very long time? If it has detached, then you need a way to know when the polymerase reattaches. And you need to know what state the polymerase is in.

Is this controlling a single molecule? or a population of molecules?

Can you measure the state of a polymerase enzyme in a nanopore? Or if you do multiple angles; you measure across it in one direction, and across it in another direction. Electrically. Maybe you have a polymerase and you have a nanopore, and around the nanopore you have a circle of 8 different electrodes or something. You would have 4 directions on either

side of the plane, like a circuit board where you have a top layer, like a two-sided PCB with a via, and the via is a nanopore. The polymerase is.. so that gives you how many, .. 8 different paths through the nanopore? So maybe it's all measuring the same thing, because they are all effectively a measurement of the nanopore. I don't know enough about nanopores to know whether you can make a really thin ... membrane as thin as a protein. And then lock a protein in that membrane. Does that work? I don't know.

Nanopore sequencers using sequencing by synthesis? No they are probably not doing that. They are just sending a DNA into a nanopore. Well, it should be looked into more. If any of them are doing polymerases in nanopores.

Make a video of the different polymerases, spinning around in visualizing..

Are people designing proteins manually? or do they have some computer program that looks at every polymerase in the database, and then imagines that they are using... we should do machine learning and recurrent neural networks over a large database, as a way to design an automatic digital polymerase.

Charged ion current, tethered polymerase, there would be charged molecules moving around in solution, blow past the polymerase. Maybe it would oxidize dirt on the protein. Or reduce them. Electronic signals are way too small or slow to interact with the polymerase enzymes. Molecules interact in the terahertz or higher frequency range. So you could have either a really slow polymerase, ... 1000 nt/second incorporation rate, but is that with proof-reading?-

Really slow polymerase. 1 nt/sec incorporation.

DNA strand on a bead. And you can move the bead around with magnets, or an optical tweezer. You have really different regions that only incorporate a certain kind of nucleotide. It's triggered somehow to deliver a nucleotide. Photocleavable tail on a nucleotide. Imagine there is an enzyme that cleaves the tail. It's light-activated. These things have to be roughly in the same place at the same time to have a nucleotide added to the chain. You would ove the bead to the A region, then you would shine ultraviolet light on it, it would cleave the tail, and it would incorporate an A nucleotide. Zero waveguide thing, where, it's a small volume, so there's only one molecule inside of that volume. One nucleotide per attoliter. We could count how many nucleotides are in an attoliter bubble, by looking at fluorescence. We could wait until the right nucleotide enters, and then flash the ultraviolet light to cleave the linker on the nucleotide tail. And then just hope that it gets incorporated by the polymerase.. so you could have a polymerase that fluoresces for an entire second or 2 after incorporation, so if it doesn't fluoresce then you know that the nucleotide hasn't been incorporated, even though you already cleaved the fluorescent tail to the nucleotide. Grid of holes, a honeycomb with tiny holes, the enzyme is pressed up against glass at the bottom of the hole. The light can't get to the bottom of the hole; the fluorescence excitation can penetrate into it it's the attoliter volume, which is where your zero wave guide comes in. It's hard to understand how that works without reading about it.

How do you deliver only one nucleotide over time to a polymerase? A specific nucleotide, in a specific order at a time? You would physically confine the polymerase, and flow the nucleotide to the polymerase where the only way the nucleotide could flow is by moving towards the polymerase. If there's no way to turn polymerase on and off, either with light or electronically, then the bubbles each have 1 nucleotide in them, a random nucleotide, then when you know it's the right nucleotide, you turn the polymerase on and let it use the nucleotide. If you are at a nanopore, then it might get clogged with a bunch of the wrong nucleotides. Turn the polymerase off, by making nucleotides not attracted to the polymerase.

Have two nanopores, if it's the wrong gate you open the one without a nucleotide. So if it's the wrong one, it goes through the wrong channel, not going through the polymerase. So it doesn't interact the polymerase. So the polymerase is a normal polymerase in a nanopore. And it's trying to incorporate nucleotides. You never give it a chance unless it's the correct one. Yes this is Nathan's idea, with TdT in front of a molecular filter. You could have another fluorescent event when the nucleotide is incorporated. And if it doesn't fluoresce, then you know that the next nucleotide you need to send to the enzyme has to be the same nucleotide. A simpler parallelizable version would be to just have dilute nucleotides floating around so that on average there is only one at a time above a nanopore. A detector above the nanopore turns on an electrode below the nanopore when it sees the appropriate nucleotide, pulling it through the nanopore and past the polymerase with a flow of ions.

Everything has noise and uncertainty in it. Your camera is not perfect. Your fluorescence is not perfect. You would have to run the system at really low speeds. You need to have really good control over everything. It has to be really cold. You need to have an error rate of 1 error in 100 bp. That's the current state of the art for chemical synthesis. To be relevant, it has to be better than that. Cooled CCD camera. The polymerase needs to be slow. Water freezes at 0 celsius. So how cold are you talking about?

Need a signal from a polymerase that lasts for multiple seconds to say "hey I did something". Maybe another protein that binds to the polymerase only in its conformation after incorporation of a nucleotide. So there's two states. There's an empty polymerase, and a full polymerase. If you see the nucleotide go in, and you see a full polymerase, and then you see an empty polymerase, then you know it incorporated. You won’t see the nucleotide exit the other side of the nanopore, because it was incorporated. It’s hard to tell which side of a nanopore the nucleotide is on.

Too much fluorescence makes molecules explode. They will react and turn into something else that doesn't fluoresce, and that happens randomly. Imaging single molecules is something I didn't believe in until I was forced to believe it.

Why a single fluorescence event? Why not a giant molecule that has a giant light show attached to it? And it's carrying a nucleotide. Like a tRNA synthetase and it has a whole bunch of different fluorescent proteins and fluorescent groups on it.

There was a polymerase mentioned today in the channel that incorporates larger nucleotides or something. (double wide nucleotides.) So maybe the light show would be attached to that,

and maybe that will work. And this would be for increasing signal to noise ratio of the imaged events.

Binding pocket geometry optimizer/finder, use a SAT solver to define geometry and charge of reactant molecules (i.e. DNA and nucleotides), define state change needed (nucleotide incorporation), ask for solutions (geometry of charges (or molecules with motifs of charges, aka amino acids)) that would bind said reactants.

There was an idea where you have an etched finger that pretend they are a DNA strand. It swaps in different base pairs, or different bases, and then the polymerases uses that as if it was the template strand.

A blank template strand so you can use regular polymerase instead of TDT.

Grid of fluidic interconnects and dead-ends (storage), all electrophoretic transportation, no pressurized fluid stuff (thus no bubbles to contain oligo pools). Use library of short oligos, PCR, ligation, etc.

Physical confinement for assembly assistance, reducing ability for growing DNA to fold on itself and interfere with ligation/extension.

Optimize a polymerase to incorporate only a certain nucleotide type at a specific temperature. This could be easily selected in vitro with SELEX or something. Use multiple different temperatures, one for each nucleotide. Make sure the polymerase can stay attached to the DNA throughout the different temperature changes.

Getting error rate below 1% would be good. Getting real-time error checking with nanopore stuff. That could maybe work. And maybe TdT stuff. It's just a matter of, wanting to try the chaining the 8 bases with ligase - - which should be another method that we do anyway for another time.

molecular engineering: making a rotating belt of tools for the polymerase molecule

laser-induced '"partial"' denaturation of polymerase (?)

Wittig's idea about lithography + DNAaseH?

[[ReAsH-based CALI]] to partially denature polymerase (pause it) (works within 30 seconds)

Some sort of "molecular marker" system for pausing polymerase? (what was this idea?)

hack Arg 283 & Lys 280 re: "ghost template strand" (virtualization)

'The single-stranded template nucleotides do not pass through a cleft formed by the thumb and fingers sub-domains, but rather lie on the surface of the protein.'

metal-assisted mechanism of phosphoryl transfer

polB has no proof-reading activity

* (ddCTP) hPol beta w/ PDB code: 1BPY, acting on 1 nt gap
* (dGTP) hPol beta w/ PDB code: 1ZJN, acting on Primer A:A mismatch
* (dAMP) hPol beta w/ PDB code: 1MQ2, acting on Grapped DNA 8oxodG
* (dUMPNPP) hPol beta w/ PDB code: 2FMQ or 2FMS, both acting on 1 nt gap (Na+/Mg2+)

"The DNA binding channel in Pol B is lined with positively charged lysine and arginine side chains that function to stabilize the negatively charged backbone of the DNA . The Pol B catalytic site contains three key aspartate residues . Asp256 is involved primarily with stabilizing the transition state complex and asp190 and asp192 both aid in positioning an incoming nucleotide. These residues are complexed with two magnesium ions that are essential in binding the incoming nucleotide . In the crystal structure shown, the incoming nucleotide is a dideoxynucleotide (ddCTP), used to "freeze" the polymerase so that co-crystals of template, primer and polymerase could be obtained. The three catalytic asp residues have been modeled to go through several conformational changes during catalysis. ..... The magnesium ions have different functions in catalysis. One Mg2+ ion binds specifically to the beta and gamma phosphates of the ddNTP as a bidentate . The other Mg2+ acts to stabilize the negative a-phosphate of the ddNTP . The metal ions are an essential component of the overall reaction for positioning, stabilization, and activation. .... Pol B also interacts with the sugar moiety of the incoming ddNTP by way of van der Waals contacts between the ribose ring of the incoming nucleotide and three residues of Pol B . Since the only difference between ribonucleotides and deoxyribonucleotides is a hydroxyl at the 3' carbon of the ribose, these van der Waals interactions may participate in dNTP-NTP differentiation. To further position the incoming ddNTP, there are several additional contacts between the ribose of the ddNTP and Asn279 & Asp276 of Pol B . Numerous other contacts between the phosphates of the ddNTP and Pol B residues are not shown. The ddNTP will eventually be added to the growing DNA strand (primer) . The 3' OH group of the last sugar poised on the phosphate sugar backbone nucleophilically attacks the phosphate of the incoming ddNTP . In Pol B, the Mg2+ of site B, acting as a Lewis base, activates the 3' OH of the primer, while Asp256 acts as the proton acceptor . In the process, two phosphates from the ddNTP are removed."

* '"[[pseudotemplate]]s'"
* family X, Rev1 / Rev1p, polA / pol1, etc.

Check electrical resistance of DNA polymerase for each nucleotide incorporation event. Detect which nucleotide is incorporated, either through fluorescence and fluorophores or through electrical detection. Then trigger the polymerase to "fix" the immediately prior nucleotide that was inserted. Continue to "fix" the last insertion until the correct nucleotide is inserted. This requires an "error checking" mechanism. Do the error checking mechanisms

only check 1 nt behind the current insertion position? How many bp away does the error proof reading mechanism move?

Physical limits of polymerase incorporation rate. What's the total maximum polymerization rate of any hypothetical molecular machine? It is somewhat unfortunate that organisms have to use 10,000 separate replication forks on the same DNA molecule. The organisms are replicating faster than the DNA polymerase can work. This means that there's probably no mutations easily accessible on the evolutionary landscape to make DNA polymerase work much more efficiently. So perhaps some theoretical molecular machines would be able to work faster than DNA polymerase? What are the theoretical limits of this machine? The reason why this question is interesting is because we should avoid DNA assembly where two DNA strands have to be connected together. Using a parallel nanoarray of electronic DNA polymerases would mean that each of their outputs would have to be conjugated together in some other separate reaction. The ideal scenario is that one polymerase would operate quickly enough to print out entire genomes or thousands of genomes in a timely manner. However, so far the best polymerases that we have found are only operating at 1000 nt/second.

Long-term, we should set some goals for polymerase performance. For example, once we construct the first electronic polymerase, it would be good to set some goals like Moore's law, we could say something like a 10x speedup every 5 years. The goal should be to make an entire human genome (3 billion bp) in a few minutes at most. To do 3 billion bp in 5 days would require 7 bp/millisecond. To do 3 billion bp in 5 minutes would require 10k bp/millisecond. Since this is probably infeasible given current technologies and current polymerases, it might be necessary to investigate DNA ligases for assembling large genomes from multiple strands of DNA.

Other molecular machines could be imagined for ultra high performance DNA synthesis and ligation. Large-scale nanoscopic molecular factories, made entirely of a single protein, could be fabricated to make a synthesis "assembly line" where DNA is physically transported through nanopores to different reaction sites. Nucleotides could be provided to this molecular factory by way of a giant funnel. Magnesium could also be provided in a more precise way, not left to chance encounter in the aqueous environment.

Is the main catalytic activity mediated by the metal ions? Why not use more metal ions? I suspect the reason is that the polymerase would slow down if it required more metal cations. The concentration of metal cations inside of the cellular environment must be carefully managed. The concentration of metal cations, necessary to support polymerization, might be a level that is highly toxic to the other operations of a cell. Also, mutations for productively using those metal cations, might be inaccessible on the evolutionary landscape. In vitro, metal cation concentration can be much higher as long as the template DNA and synthesized DNA remain unharmed by increasing concentrations of metal atoms.

If a mutant polymerase can operate much more quickly given more energy, then another possibility to consider would be to provide a way to transfer charge to the polymerase from nanoscopic electrodes and transistors. Perhaps the electrical device can be used to

"re-charge" the polymerase after it has used its electrons to catalyze some kind of chemical reaction.

It's interesting that DNA polymerase makes use of physical conformational changes to move nucleotides and to initiate its catalytic activity. What if, instead, the DNA polymerase enzyme was made to operate by chemistry instead? For example, instead of requiring charged ions, perhaps the polymerase would store electron charge provided by nearby electrodes from a CMOS chip. And instead of waiting for a dNTP to bump into the polymerase, what about having a base nucleotide that the polymerase can transform into another dNTP through the use of charge?

What you need is an enzyme that has a long "tail" of not-yet-reacted nucleoside bases. Upon receiving electric charge, or other forms of stimulation, the enzyme should take a nucleoside base and convert it into the correct type of nucleotide. Then it should incorporate that nucleotide into a growing DNA strand. Alternatively, an enzyme could be imagined that takes an existing strand of "DNA" (perhaps with only undifferentiated nucleobases) and then converts each base pair into an actual nucleotide. The advantage of having a "tail" of ready-available chemical parts is that there is no time spent on waiting for a dNTP to fly around and hit the polymerase in the correct way. Instead, the polymerase can run as quickly as possible.

http://gnusha.org/logs/2016-07-19.log
10:29 < kanzure> you could have capactive labels on nucleotides that have a certain electrical signal in response to optical illumination, e.g. to make a really strong signal for electrical detection
10:29 < kanzure> or making a really specific signature i mean

Is there any way to add "internal state" into an enzyme? It looks like conformational shape changes are "stateful" but they are not an internally-queryable state; rather, they are "stateful" in the sense that there are different energy valleys that components of the enzyme can be pushed down. Giving "internal memory" to an enzyme might be helpful. Or we will just be stuck with trying to think of molecular mechanical motions that can be strung together that satisfy all of the reaction conditions and reaction constraints.

Making a really slow polymerase

Add residues that can cover the external pores of the polymerase enzyme so that nucleotides cannot enter. Then make these residue arms or covers in a way such that by using light or electricity or the presence of other molecules, such that the enzyme pores are covered or uncovered in a controlled way.

Then the next step is to make the movement of the polymerase enzyme something controlled. The goal is to only move after incorporation, and only to incorporate when an external signal has been received. The pore for accepting a new nucleotide should be covered before movement. There should be no nucleotide inside of the enzyme during

movement -- otherwise, the pore might be "blocked" but a nucleotide might already be ready to be incorporated into the growing oligo or DNA molecule. So the enzyme should be able to never hold a nucleotide ....

What about one-shot polymerases that have nucleotides stored inside? What if they can't be reused or something? That might be useful, except for getting them to latch on to the growing oligo molecule....

We can use selection experiments to create unique and novel polymerases. For example, making a super large polymerase that can be more easily physically gripped might be an interesting project, because a larger polymerase might be easier to physically handle and also have more room for incorporating strange amino acid residues for strange new functionality.

Functions of polymerase that need to be controlled:
- nucleotide selection
- stepping forward on the DNA molecule, ratcheting
- pausing
- nucleotide incorporation mechanism
- attaching/detaching from a DNA molecule
- speed
- error correction (needs to be removed, really)
- templatelessness, or make it ignore the template

What is the minimum volume of water that polymerase needs to be located within? Can it work without water?

Could there be a selection experiment to make a polymerase enzyme that works outside of water? Are there any enzymes that work in open air environments? Can most of the cellular machinery be made to work in dry atmosphere?

Another option is to analyze the physical mechanisms by which DNA polymerase works, and then creating a non-enzymatic approach that applies the same chemical transformations to a growing DNA molecule.

What about repurposing the ribosome to do DNA synthesis? The ribosome is already made to work with tRNA synthetase. It is already made to accept mRNA fragments. What if the tRNA synthetase stuff is repurposed to program the ribosome to select nucleotides? And also, the ribosome output pore would be changed to create a strand of DNA using the same mechanism from inside the polymerase enzyme. There would be a few problems here; one is that the ribosome needs mRNA at the moment, and it needs tRNA synthetase to bring amino acid residues. What we need is something like--- given a tRNA synthetase enzyme with a nucleotide or some payload, incorporate the payload into the growing molecular strand.

We could have a general purpose protein factory that makes growing strands of anything. The growing strand could be amino acid residue linkages, to make proteins, or it could be DNA, or it could be RNA. The point though is that each of these require control and require selection from a number of options. Perhaps we can do some rational protein engineering to create a design that works for all of these cases? This certainly increases the value of the final protein machine. And it could also create other forms of output too, like other polymerized molecules or something. It's a general purpose one-dimensional programmable nanofactory.

There should be a review of all conformational changes in known enzymes. Also a review of rational protein engineering. What sort of shapes can we definitely create? What are our current rational protein engineering capabilities?

Is it possible that there are other existing enzymes that might have already been sequenced that could provide useful components for a controlled polymerase? We should search over the set of all sequenced proteins, and then look at their shapes and also figure out if any of those shapes would be helpful.

Why is the polymerase enzyme so small? Why not a giant funnel structure, where 20 or 40 bp are inside of the enzyme at the same time?

Would a DNAzyme be a more appropriate target for engineering a controlled polymerase? What about aptamers?

Another method is to use yoctoliter droplets and bubbles to deliver specific reagents to a polymerase enzyme. Then force the polymerase enzyme to incorporate those molecules (perhaps use a surprisingly permissive polymerase for these purposes). Another option is to use yoctoliter droplets to deliver reagents for oligonucleotide synthesis by phosphoramidites. This should in theory fix the yield problems, as long as you can be sure about the contents of a droplet (perhaps through fluorescence measurements of the droplet before sending the droplet to the reaction chamber). The trouble is, is a yoctoliter too big of a droplet? You basically need to manipulate one molecule at a time in a single drop of water. How to do that? Dilutions? With fluorescence you can be somewhat sure that there's only one molecule in the droplet, but at these scales things are much harder to deal with.... and measurements are much harder to get correct.

Physical arm covering the external pores of the polymerase enzyme. Use light control, electricity, or chemical control perhaps ion concentration, to control whether the arm is deployed to cover the polymerase pore or not.

Controlled error correction? When the polymerase enzyme incorporates a nucleotide, perhaps a fluorescent event can be given off. And then if the nucleotide type was wrong, then force the polymerase to go back and try to incorporate a different nucleotide. Then wait until the polymerase incorporates the correct nucleotide. How long can a polymerase enzyme sit there and replace nucleotides in the same oligo? Or is there another step for

removing the previous nucleotide choice? My guess is that error correction should be avoided for a controlled polymerase, it's unlikely to yield a simpler or more achievable result.

Is there any other enzyme that is more important to physically engineer and control? What could possibly be more valuable than DNA polymerase being under more control? is there an enzyme that needs to be rationally engineered that would revolutionize some industries or something? Perhaps it would be financially viable to start there, rather than at polymerase. But DNA is pretty valuable if you can make lots of it at a single time.

A reasonable path forward on designing and creating a controlled polymerase:
- consortium of protein engineering people
- read all the polymerase/ and DNA/ papers
- get someone involved who has done rational protein design
- Financially, you can de-risk this by choosing to work on the controlled ribosome idea. This can be used for protein molecular nanotechnology, and it could inform a DNA polymerase machine, but it won't be directly useful for DNA polymerase. Still, it might provide the revenue necessary to go figure out a DNA polymerase machine.
- read some reviews on protein design, polymerase, DNA synthesis, conformational change in proteins
- conference on controlled polymerase, with a paper that results from it, etc.
- DARPA funding for a controlled polymerase research project
- design competition?? what would be the objectives or judging criteria?

For DNA repair enzymes, perhaps it would be simpler to "proof read" the DNA molecule, and then control the repair enzyme for which nucleotide it chooses to incorporate into a specific position.

How do you debug a living active polymerase enzyme? If you had a really huge polymerase enzyme, and if you were okay with it consuming lots of energy, you could have giant conformational changes to report information, or you could have multiple fluorescent colors to report information. This would be really helpful for live debugging of a polymerase enzyme..... other forms of debugging would also be helpful. This will reduce the amount of time for the iteration cycles from designing and testing and debugging why something isn't working. With more information, you can go much faster. You could call this an "instrumented polymerase" or a "debugger polymerase" or something.

In fact, all enzymes could benefit from debugging instrumentation. There should be much more reporting about their conformational shapes. Shape changes. Whether they are bound or unbound to other objects. Fluorescence to signal different events. Long-stable persistent light. Shape changes to communicate information. Unrelated chemical reactions to indicate to the environment the status of the enzyme. Getting information out of an enzyme is a good first step. We need to both send information but also receive information from these tiny molecular machines.

"Domesticated polymerase"

mutagenesis over polymerase, array
search for viable candidates
statistical distribution of the random synthesis...
use a buffer solution, put lots of polymerases in there, run a current through the solution, check with PCR whether there are any interesting results and signals in the mix

need nanoelectronic electrode, chemical linker, enzyme polymerase molecule bought off the shelf, electrode array and signal generator for entire array
enzymes in a tube, tube is the electrical circuit, use buffer solution and change chemical environment

You could use the "conductance sequencing" polymerase method as a way to debug other enzyme types. By probing for changes in conductance and resistance and inductance, you could determine different details about the current state of the enzyme. Using an array with different mutants, perhaps you could characterize a million different mutant enzymes simultaneously?

By looking at thousands of PDB database images, you could have a program that infers intermediate conformational states (which may not be present in any of the captured crystallography results), based on graph grammar graph rewriting rules and also molecular dynamics simulations and other heuristic rules. Using the inferred results, you could have a much larger catalog of protein conformational changes to work with. These could then be used when designing new proteins and new enzymes.

Are there any enzymes that have an "internal state" counter? Are they able to switch between multiple states? And do they have internal molecular logic based on the current state? If so, can that sort of system be copied into polymerase if polymerase does not also share these characteristics? Would this be useful? How to interrogate the current state of an enzyme?

TODO:

- Include juul things (2016-07-21)
- Include rob things (2016-07-21)

From the G.M.C patent:
"""
The nucleotide type incorporated can be determined by: a) the intersection of a light pulse coincident with a particular dNTP (or rNTP or other monomer class) present at that time point in a cyclic pattern of dNTP solutions. b) ‘caged’ (i.e. photo-activatable or photo-inactivatable) dNTPs, rNTPs or cations. c) base-specific, light-modulated steric or conformational selectivity (see Hoppmann C, Schmieder P, Heinrich N, Beyermann M. (2011) Chembiochem. 12(17):2555-9. doi: 10.1002/cbic.201100578. Epub 2011 October 13. Photoswitchable click amino acids: light control of conformation and bioactivity). Poly(A)

polymerase is particularly useful since its specificity for ATp relative to other rNTPs is due to a conformational change which can be mimicked by a photo-sensitive amino acid linkage (like azobenzene, with or without crosslinking).
"""

From the pacbio patent:
"""
In some cases the substrate is exposed to four types of nucleotide analogs corresponding to A, G, C, T, or A, G, C, U, each of the four types of nucleotide analogs having a different impedance label. In some cases the impedance label is attached to the polyphosphate portion through a linker. In some cases the impedance label comprises either a capacitance label or a conductivity label.
"""

"""
proximate to the nanoscale electrode; exposing the polymerase to a plurality of types of nucleotide analogs, each comprising a different capacitive label attached to the phosphate portion of the nucleotide analog under conditions whereby polymerase mediated nucleic acid synthesis occurs, resulting in cleavage of the capacitive label and the growth of a nascent nucleic acid strand; applying electrical signals comprising alternating current over time to the nanoscale electrode, whereby when a nucleotide analog resides in the active site of the enzyme, the capacitive label on the nucleotide analog produces a measurable change in the capacitance at the nanoscale electrodes; monitoring the electrical signal at the nanoscale electrode over time, whereby the electrical signal indicates an incorporation event for a type of nucleotide analog having a specific capacitive label; and using the monitored electrical signal at the electrode over time to determine a sequence of the template nucleic acid.
"""

"Nucleotide analogs comprising conductivity labels will typically be larger, i.e. have a larger molecular weight than natural nucleotides. These analogs can include, for example, nucleotide analogs describe in U.S. patent application Ser. No. 13/767,619 entitled Polymerase Enzyme Substrates with Protein Shield, filed Feb. 14, 2013, and in U.S. Patent Application 61/862,502, entitled Protected Fluorescent Reagent Compounds, which are incorporated herein by reference for all purposes."

"""
In some cases the conductivity labels comprise beads, for example beads comprising multiple nucleotides attached via their polyphosphate portion. Such analogs are described, for example in U.S. Pat. No. 8,367,813 which is incorporated by reference herein in its entirety for all purposes. The beads can be coated with charged functional groups, anionic, cationic, or a combination of anionic and cationic groups. The amount of charge on the bead can be controlled in order to control the electrical signal at the gate of the nanoFET. The beads can have any usable size range, for example, between about 2 nm and about 50 nm in size. The shapes of the beads can be spherical, elongated, or other effective shape for controlling the current at the gate of the nanoFET.
"""

"""
The characteristics of the nanoparticle can be varied in order to produce different capacitance values. The size of the nanoparticle can influence the capacitance of the particle, as well as the chemical structure. Nanoparticles of metals, semiconductors, glasses, oxides, carbon, silicon, protein, polymers, ionic materials, can be used and can be produced to have widely different impedance magnitude and impedance versus frequency characteristics. The size of the particles can be varied over a wide range, for example from about 2 nanometers to about 50 nanometers in diameter. One large contributor to the impedance change near an electrode is the capacitance characteristics of the material itself. However, it is to be understood that the impedance that is being measured is that of the region around the electrode, and not just that of the label. For example, a nanoparticle label will displace the solution near the electrode, such that the measured impedance will include that change. Thus, a capacitive label near the electrode can result in the impedance either going up or going down as compared to the impedance when the label is not present.
"""

"""
Differentiating nucleotide analogs based on the magnitude of impedance or conductivity change can be carried out, for example, by providing a capacitive or conductivity label having multiple capacitive or conductive moieties on a nucleotide analog. Nucleotide analog structures including those having multivalent scaffolds and nucleotides having multiple moieties can be prepared as described, for example, in US Patent Application 20120058473 Molecular Adaptors for Dye Conjugates, and US Patent Application 20120077189 Scaffold-Based Polymerase Enzyme Substrates, which are incorporated herein by reference for all purposes. While these references generally describe a fluorescent label, it is to be understood in conjunction with the teachings of this application that a suitable capacitive label or conductivity label connected by a suitable linker as described herein can be substituted for the fluorescent label.
"""

"""
Nucleotides or analogs that can thus be identified by the spectrum of the electrical oscillation they produce. In some cases, oscillations looks like noise, but with reproducible and identifiable characteristics including the frequency and the magnitude of the signal. These different types of oscillations can be used like different colored dyes are used to differentiate between different nucleotide analogs in optical systems, thus, we refer herein to a distinguishable type of current oscillation as a current oscillation color.
"""

"Non-limiting exemplary binding moieties for attaching either nucleic acids or polymerases to a solid support include streptavidin or avidin/biotin linkages, carbamate linkages, ester linkages, amide, thiolester, (N)-functionalized thiourea, functionalized maleimide, amino, disulfide, amide, hydrazone linkages, among others. Antibodies that specifically bind to one or more reaction components can also be employed as the binding moieties. In addition, a

silyl moiety can be attached to a nucleic acid directly to a substrate such as glass using methods known in the art."

"Where desired, an enzyme or other protein reaction component to be immobilized may be modified to contain one or more epitopes for which specific antibodies are commercially available. In addition, proteins can be modified to contain heterologous domains such as glutathione S-transferase (GST), maltose-binding protein (MBP), specific binding peptide regions (see e.g., U.S. Pat. Nos. 5,723,584, 5,874,239 and 5,932,433), or the Fc portion of an immunoglobulin. The respective binding agents for these domains, namely glutathione, maltose, and antibodies directed to the Fe portion of an immunoglobulin, are available and can be used to coat the surface of a capacitive device of the present invention. The binding moieties or agents of the reaction components they immobilize can be applied to a support by conventional chemical techniques which are well known in the art."

"The various components of the surface of the capacitive devices can be selectively treated in order to bind the polymerase-template complex to a specific portion of the substrate. Selective treatment and immobilization is described, for example, in U.S. Pat. No. 5,624,711; U.S. Pat. No. 5,919,523; Hong et al., (2003) Langmuir 2357-2365; U.S. Pat. No. 5,143,854; U.S. Pat. No. 5,424,186; U.S. Pat. No. 8,137,942; U.S. Pat. No. 7,993,891 Reactive surfaces, substrates and methods of producing and using same; U.S. Pat. No. 7,935,310; U.S. Pat. No. 7,932,035 U.S. Pat. No. 7,931,867 Uniform surfaces for hybrid material substrates and methods of making and using same; and U.S. Pat. No. 8,193,123 Articles having localized molecules disposed thereon and methods of producing same, all of which are incorporated herein by reference for all purposes."

"One aspect of carrying out capacitance measurements in solution is controlling the ionic strength of the medium. It is know that polymerase enzymes can effectively operate over a range of ionic strengths, and that the ionic strength can be varied by changing the levels of monovalent ions such as Li+, Na+, K+, Rb+, or Cs+. As has been shown the amount of one or more of these cations can have an effect on the kinetics of the polymerase, and that the kinetic behavior can be tuned by varying the relative amounts of these ions. Using combinations of these ions, conditions can be chosen where both the kinetic parameters of the enzyme, and the ionic strength for capacitive detection can be useful for the instant methods. See, e.g. U.S. Patent Application 20120009567 which is incorporated herein by reference for all purposes."

"Enzymatic reactions are often run in the presence of a buffer, which is used, in part, to control the pH of the reaction mixture. Buffers suitable for the invention include, for example, TAPS (3-{[tris(hydroxymethyl)methyl]amino}propanesulfonic acid), Bicine (N,N-bis(2-hydroxyethyl)glycine), TRIS (tris(hydroxymethyl)methylamine), ACES (N-(2-Acetamido)-2-aminoethanesulfonic acid), Tricine (N-tris(hydroxymethyl)methylglycine), HEPES 4-2-hydroxyethyl-1-piperazineethanesulfonic acid), TES (2-{[tris(hydroxymethyl)methyl]amino}ethanesulfonic acid), MOPS (3-(N-morpholino)propanesulfonic acid), PIPES (piperazine-N,N′-bis(2-ethanesulfonic acid)), and MES (2-(N-morpholino)ethanesulfonic acid)."

"Exemplary modifications that can be detected by the methods of the invention include, but are not limited to methylated bases (e.g., 5-methylcytosine, N6-methyladenosine, etc.), pseudouridine bases, 7,8-dihydro-8-oxoguanine bases, 2′-O-methyl derivative bases, nicks, apurinic sites, apyrimidic sites, pyrimidine dimers, a cis-platen crosslinking products, oxidation damage, hydrolysis damage, bulky base adducts, thymine dimers, photochemistry reaction products, interstrand crosslinking products, mismatched bases, secondary structures, and bound agents. In preferred embodiments, nucleotides or analogs thereof that are incorporated into a nascent strand synthesized by the enzyme are distinctly labeled to allow identification of a sequence of specific nucleotides or nucleotide analogs so incorporated. Labels are linked to nucleotides or nucleotide analogs through a phosphate group, e.g., a phosphate group other than the alpha phosphate group. As such, the capacitive labels are removed from the nucleotide or nucleotide analog upon incorporation into the nascent strand. Techniques for kinetically identifying modified bases are described, for example in U.S. Patent Application 20110183320 Classification of Nucleic Acid Templates which is incorporated herein by reference for all purposes."

"Although certain specific embodiments of the invention are described in terms of 5-methylcytosine detection, detection of other types of modified nucleotides (e.g., N^6-methyladenosine, N^3-methyladenosine, N^7-methylguanosine, 5-hydroxymethylcytosine, other methylated nucleotides, pseudouridine, thiouridine, isoguanosine, isocytosine, dihydrouridine, queuosine, wyosine, inosine, triazole, diaminopurine, β-D-glucopyranosyloxymethyluracil (a.k.a., β-D-glucosyl-HOMedU, β-glucosyl-hydroxymethyluracil, “dJ,” or “base J”), 8-oxoguanosine, and 2′-O-methyl derivatives of adenosine, cytidine, guanosine, and uridine) are also contemplated. Further, although described primarily in terms of DNA templates, such modified bases can be modified RNA bases and can be detected in RNA (or primarily RNA) templates. These and other modifications are known to those of ordinary skill in the art and are further described, e.g., in Narayan P, et al. (1987) Mol Cell Biol 7(4):1572-5; Horowitz S, et al. (1984) Proc Natl Acad Sci U.S.A. 81(18):5667-71; “RNA's Outfits: The nucleic acid has dozens of chemical costumes,” (2009) C&EN; 87(36):65-68; Kriaucionis, et al. (2009) Science 324 (5929): 929-30; and Tahiliani, et al. (2009) Science 324 (5929): 930-35; Matray, et al. (1999) Nature 399(6737):704-8; Ooi, et al. (2008) Cell 133: 1145-8; Petersson, et al. (2005) J Am Chem Soc. 127(5):1424-30; Johnson, et al. (2004) 32(6):1937-41; Kimoto, et al. (2007) Nucleic Acids Res. 35(16):5360-9; Ahle, et al. (2005) Nucleic Acids Res 33(10):3176; Krueger, et al., Curr Opinions in Chem Biology 2007, 11(6):588); Krueger, et al. (2009) Chemistry & Biology 16(3):242; McCullough, et al. (1999) Annual Rev of Biochem 68:255; Liu, et al. (2003) Science 302(5646):868-71; Limbach, et al. (1994) Nucl. Acids Res. 22(12):2183-2196; Wyatt, et al. (1953) Biochem. J. 55:774-782; Josse, et al. (1962) J. Biol. Chem. 237:1968-1976; Lariviere, et al. (2004) J. Biol. Chem. 279:34715-34720; and in International Application Publication No. WO/2009/037473, the disclosures of which are incorporated herein by reference in their entireties for all purposes. Modifications further include the presence of non-natural base pairs in the template nucleic acid, including but not limited to hydroxypyridone and pyridopurine homo- and hetero-base pairs, pyridine-2,6-dicarboxylate and pyridine metallo-base pairs, pyridine-2,6-dicarboxamide and a pyridine metallo-base pairs, metal-mediated pyrimidine base pairs T-Hg(II)-T and C-Ag(I)-C, and metallo-homo-basepairs of 2,6-bis(ethylthiomethyl)pyridine nucleobases Spy, and alkyne-,

enamine-, alcohol-, imidazole-, guanidine-, and pyridyl-substitutions to the purine or pyridimine base (Wettig, et al. (2003) J Inorg Biochem 94:94-99; Clever, et al. (2005) Angew Chem Int Ed 117:7370-7374; Schlegel, et al. (2009) Org Biomol Chem 7(3):476-82; Zimmerman, et al. (2004) Bioorg Chem 32(1):13-25; Yanagida, et al. (2007) Nucleic Acids Symp Ser (Oxf) 51:179-80; Zimmerman (2002) J Am Chem Soc 124(46):13684-5; Buncel, et al. (1985) Inorg Biochem 25:61-73; Ono, et al. (2004) Angew Chem 43:4300-4302; Lee, et al. (1993) Biochem Cell Biol 71:162-168; Loakes, et al. (2009), Chem Commun 4619-4631; and Seo, et al. (2009) J Am Chem Soc 131:3246-3252, all incorporated herein by reference in their entireties for all purposes). Other types of modifications include, e.g., a nick, a missing base (e.g., apurinic or apyridinic sites), a ribonucleoside (or modified ribonucleoside) within a deoxyribonucleoside-based nucleic acid, a deoxyribonucleoside (or modified deoxyribonucleoside) within a ribonucleoside-based nucleic acid, a pyrimidine dimer (e.g., thymine dimer or cyclobutane pyrimidine dimer), a cis-platin crosslinking, oxidation damage, hydrolysis damage, other methylated bases, bulky DNA or RNA base adducts, photochemistry reaction products, interstrand crosslinking products, mismatched bases, and other types of “damage” to the nucleic acid. As such, certain embodiments described herein refer to “damage” and such damage is also considered a modification of the nucleic acid in accordance with the present invention. Modified nucleotides can be caused by exposure of the DNA to radiation (e.g., UV), carcinogenic chemicals, crosslinking agents (e.g., formaldehyde), certain enzymes (e.g., nickases, glycosylases, exonucleases, methylases, other nucleases, glucosyltransferases, etc.), viruses, toxins and other chemicals, thermal disruptions, and the like. In vivo, DNA damage is a major source of mutations leading to various diseases including cancer, cardiovascular disease, and nervous system diseases (see, e.g., Lindahl, T. (1993) Nature 362(6422): 709-15, which is incorporated herein by reference in its entirety for all purposes). The methods and systems provided herein can also be used to detect various conformations of DNA, in particular, secondary structure forms such as hairpin loops, stem-loops, internal loops, bulges, pseudoknots, base-triples, supercoiling, internal hybridization, and the like; and are also useful for detection of agents interacting with the nucleic acid, e.g., bound proteins or other moieties."

Acknowledgements

Many of the ideas presented in this paper were publicly discussed by the Enzymatic Synthesis Group mailing list <https://groups.google.com/group/enzymaticsynthesis> and the ##hplusroadmap channel on Freenode IRC.

References

[21], [24], [27]

Scratch notes not yet added

http://diyhpl.us/~bryan/papers2/polymerase/
http://diyhpl.us/~bryan/papers2/DNA/
http://diyhpl.us/~bryan/papers2/bio/
http://diyhpl.us/~bryan/papers2/neuro/
https://groups.google.com/group/enzymaticsynthesis
https://groups.google.com/group/diybio
http://diyhpl.us/wiki/polymerase/notes/

~~Strikethrough~~ indicates that the ref has already been incorporated and cited into the document.

[Hargreaves2011] Light-dependent RNA polymerase
http://diyhpl.us/~bryan/papers2/polymerase/Light-dependent%20RNA%20polymerase%20-%20proposal%20-%20Tom%20Hargreaves.pdf

~~[XXYZXUU] Genetic information 'created' by archaebacterial DNA polymerase~~
http://diyhpl.us/~bryan/papers2/polymerase/Genetic%20information%20'created'%20by%20archaebacterial%20DNA%20polymerase%20-%201997.pdf

[Hessel2010] Exploring template-independent polymerases for automated DNA synthesis
http://diyhpl.us/~bryan/papers2/polymerase/Exploring%20template-independent%20polymerases%20for%20automated%20DNA%20synthesis%20-%202010.pdf

[Cho2006] Reengineering CCA-adding enzymes to function as (U,G)- or dCdCdA-adding enzymes or poly(C,A) and poly(U,G) polymerases
http://www.pnas.org/content/104/1/54.short

[NSNT2014rev] DNA polymerases engineered by directed evolution to incorporate non-standard nucleotides --- Laos
http://diyhpl.us/~bryan/papers2/polymerase/DNA%20polymerases%20engineered%20by%20directed%20evolution%20to%20incorporate%20non-standard%20nucleotides%20-%20review%20-%202014.pdf

[Brak2006] Directed evolution as a tool for understanding and optimizing nucleic acid polymerase function
http://diyhpl.us/~bryan/papers2/polymerase/Directed%20evolution%20as%20a%20tool%20for%20understanding%20and%20optimizing%20nucleic%20acid%20polymerase%20function%20-%20Brakmann%20-%20review%20-%202005.pdf

[NGS2010] Directed evolution of DNA polymerases for next generation sequencing

http://diyhpl.us/~bryan/papers2/polymerase/Directed%20evolution%20of%20DNA%20polymerases%20for%20next%20generation%20sequencing%20-%202010.pdf

~~[IVC2006] Directed evolution by in vitro compartmentalization~~
http://diyhpl.us/~bryan/papers2/polymerase/Directed%20evolution%20by%20in%20vitro%20compartmentalization%20-%202006.pdf

~~A bulk sub-femtoliter in vitro compartmentalization system using super-fine electrosprays~~
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4873800/

Development of a multi-copy bead display approach and its application in in vitro evolution of proteins and nucleic acids
http://www-brs.ub.ruhr-uni-bochum.de/netahtml/HSS/Diss/PaulSiddhartha/diss.pdf

[SzoSyn] TNA synthesis by DNA polymerase
http://diyhpl.us/~bryan/papers2/DNA/TNA%20synthesis%20by%20DNA%20polymerase%20-%20Szostak%20-%202003.pdf

[EDNA910914508581] The evolution of DNA polymerases with novel activities
http://diyhpl.us/~bryan/papers2/polymerase/The%20evolution%20of%20DNA%20polymerases%20with%20novel%20activities.pdf
http://diyhpl.us/~bryan/papers2/polymerase/The%20evolution%20of%20DNA%20polymerases%20with%20novel%20activities.pdf.evolved-or-rationally-designed-polymerases.png
TODO: incorporate above table (or similar)

[2016MIVS] Modified nucleoside triphosphates for in-vitro selection techniques
http://diyhpl.us/~bryan/papers2/polymerase/Modified%20nucleoside%20triphosphates%20for%20in-vitro%20selection%20techniques%20-%202016.pdf

~~[DMFPME] A general strategy for expanding polymerase function by droplet microfluidics~~
http://diyhpl.us/~bryan/papers2/polymerase/A%20general%20strategy%20for%20expanding%20polymerase%20function%20by%20droplet%20microfluidics%20-%202016.pdf

~~[MFDE2010] Ultrahigh-throughput screening in drop-based microfluidics for directed evolution~~
http://diyhpl.us/~bryan/papers2/polymerase/Ultrahigh-throughput%20screening%20in%20drop-based%20microfluidics%20for%20directed%20evolution%20-%202010.pdf

~~[GhadDEpoly] Directed evolution of polymerase function by compartmentalized self-replication~~
http://diyhpl.us/~bryan/papers2/polymerase/Directed%20evolution%20of%20polymerase%20function%20by%20compartmentalized%20self-replication%20-%20Ghadessy.pdf

~~[2013DPE] Directed polymerase evolution~~
http://diyhpl.us/~bryan/papers2/polymerase/Directed%20polymerase%20evolution%20-%20review%20-%202013.pdf

[13] Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T., Kodama, K., Yokogawa, T., Nishikawa, K. and Yokoyama, S. (2002) An unnatural base pair for incorporating amino acid analogs into proteins. Nat. Biotechnol. 20, 177–182.
[14] Li, Z., Lavergne, T., Malyshev, D.A., Zimmermann, J., Adhikary, R., Dhami, K., Ordoukhanian, P., Sun, Z., Xiang, J. and Romesberg, F.E. (2013) Sitespecifically arraying small molecules or proteins on DNA using an expanded genetic alphabet. Chem. Eur. J. 19, 14205–14209.
~~[15] Seo, Y.J., Malyshev, D.A., Lavergne, T., Ordoukhanian, P. and Romesberg, F.E. (2011) Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs. J. Am. Chem. Soc. 133, 19878–19888.~~
~~[16] Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S. and Hirao, I. (2013) Generation of high-affinity DNA aptamers using an expanded genetic alphabet. Nat. Biotechnol. 31, 453–457.~~

[El2014] Mutant Taq DNA polymerases with improved elongation ability as a useful reagent for genetic engineering
http://diyhpl.us/~bryan/papers2/polymerase/Mutant%20Taq%20DNA%20polymerases%20with%20improved%20elongation%20ability%20as%20a%20useful%20reagent%20for%20genetic%20engineering%20-%202014.pdf
chimeric polymerase paper (above)

Recombinant polymerases with increased phototolerance
https://patents.google.com/patent/US8906660B2

US 20120034602 Recombinant Polymerases for Improved Single Molecule Sequencing
US 20100093555 Enzymes Resistant to Photodamage
US 20110189659 Generation of Modified Polymerases for Improved Accuracy in Single Molecule Sequencing
US 20100112645 Generation of Modified Polymerases for Improved Accuracy in Single Molecule Sequencing
US 2008/0108082 Polymerase enzymes and reagents for enhanced nucleic acid sequencing
US 20110059505 Polymerases for Nucleotide Analogue Incorporation
WO 2007/075987 Active surface coupled polymerases
WO 2007/076057 Protein engineering strategies to optimize activity of surface attached proteins
U.S. patent application Ser. No. 12/414,191 Two slow-step polymerase enzyme systems and methods

Hiraga and Arnold (2003) “General method for sequence-independent site-directed chimeragenesis: J. Mol. Biol. 330:287-296

U.S. patent application Ser. No. 12/584,481 filed Sep. 4, 2009, by Pranav Patel et al. entitled "Engineering polymerases and reaction conditions for modified incorporation properties"

Sequencing reactions with alkali metal cations for pulse width control
https://patents.google.com/patent/US20120009567

Methods and apparatus for synthesizing nucleic acids
https://www.google.com/patents/US9279149

"Another embodiment for using non-template dependent polymerase/transferase enzymes would be to using protein engineering or protein evolution to modify the enzyme to remain tightly bound and inactive to the nascent strand after each single nucleotide incorporation, thus preventing any subsequent incorporation until such time as the polymerase/transferase is released from the strand by use of a releasing reagent/condition. Such modifications would be selected to allow the use of natural unmodified dNTPs instead of reversible terminator dNTPs. Releasing reagents could be high salt buffers, denaturants, etc. Releasing conditions could be high temperature, agitation, etc. For instance, mutations to the Loop1 and SD1 regions of TdT have been shown to dramatically alter the activity from a template-independent activity to more of a template dependent activity. Specific mutations of interest include but are not limited to Δ3384/391/392, del loop1 (386→398), L398A, D339A, F401A, and Q402K403C404→E402R403S404. Other means of accomplishing the goal of a post-incorporation tight binding (i.e., single turnover) TdT enzyme could include mutations to the residues responsible for binding the three phosphates of the initiator strand including but not limited to K261, R432, and R454."

~~Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase~~
http://www.pnas.org/content/99/10/6597.short

~~Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide~~
http://www.sciencedirect.com/science/article/pii/S0022283606007911

Very efficient template/primer-independent DNA synthesis by thermophilic DNA polymerase in the presence of a thermophilic restriction endonuclease
https://www.bu.edu/cab/CAB%20PDF/Liang%20et%20al%20Biochemistry%20'04.pdf
(cited once, but not completely read yet)

Mutator phenotypes caused by substitution at a conserved motif A residue in eukaryotic DNA polymerase δ

"Here, we used site-directed mutagenesis to create all 19 amino acid substitutions for the conserved Leu612 in Motif A of Saccharomyces cerevisiae Pol δ. We show that substitutions at Leu612 differentially affect viability, sensitivity to genotoxic agents, cell cycle progression, and replication fidelity. The eight viable mutants contained Ile, Val, Thr, Met, Phe, Lys, Asn, or Gly substitutions. Individual substitutions varied greatly in the nature and extent of attendant phenotypic deficiencies, exhibiting mutation rates that ranged from near wild type to a 37-fold increase. The L612M mutant exhibited a 7-fold elevation of mutation rate but essentially no detectable effects on other phenotypes monitored; the L612T mutant showed

a nearly wild type mutation rate together with marked hypersensitivity to genotoxic agents; and the L612G and L612N strains exhibited relatively high mutation rates and severe deficits overall."

Evolving a polymerase for hydrophobic base analogues
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2762193/

Interactions of non-polar and “Click-able” nucleotides in the confines of a DNA polymerase active site
http://kops.uni-konstanz.de/bitstream/handle/123456789/20494/Obeid_204941.pdf?sequence=2

Structure and Function of an RNA-Reading Thermostable DNA Polymerase

Highly tolerated amino acid substitutions increase the fidelity of Escherichia coli DNA polymerase I

In Vitro Production and Screening of DNA Polymerase η Mutants for Catalytic Diversity
http://www.biotechniques.com/multimedia/archive/00010/02335dd08_10101a.pdf

ϕ29 DNA polymerase residue Leu384, highly conserved in motif B of eukaryotic type DNA replicases, is involved in nucleotide insertion fidelity

Random mutagenesis by error-prone Pol I plasmid replication in Escherichia coli
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4112372/

"The error-prone DNA polymerase I that we use (low fidelity-Pol I or “LF-Pol I”), bears three mutations, namely I1709N (in motif A), A759R (in motif B) and D424A (in the proofreading domain) (14,15). LF-Pol I is expressed in an Escherichia coli strain, JS200, which has a temperature-sensitive allele of Pol I (polA12) (16) so that LF-Pol I becomes the predominant Pol I activity at 37 °C. ..... Compared to in vitro mutagenesis methods, the main disadvantages of this approach are lack of ability to restrict mutagenesis to a target gene (with the consequent concern about mutations modulating expression rather than activity) or to a specific area within a target gene[.]"

TdT-related references

High priority

A theoretical model for template-free synthesis of long DNA sequence
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2735642/
"This theoretical scheme is intended to formulate a potential method for high fidelity synthesis of Nucleic Acid molecules towards a few thousand bases using an enzyme system. Terminal Deoxyribonucleotidyl Transferase, which adds a nucleotide to the 3′OH

end of a Nucleic Acid molecule, may be used in combination with a controlled method for nucleotide addition and degradation, to synthesize a predefined Nucleic Acid sequence. A pH control system is suggested to regulate the sequential activity switching of different enzymes in the synthetic scheme. Current practice of synthetic biology is cumbersome, expensive and often error prone owing to the dependence on the ligation of short oligonucleotides to fabricate functional genetic parts. The projected scheme is likely to render synthetic genomics appreciably convenient and economic by providing longer DNA molecules to start with."

"For example TdT will add a homopolymeric tract of nucleotides when provided with a primer and a single type of NTP (Bollum 1978; Ratliff 1981). When provided with a mixture of nucleotides and a primer, TdT act as a random-sequence generator (Bollum 1978; Ratliff 1981)."

"TdT has already been utilized in genomics. It has been used for the production of synthetic homo and heteropolymers (Bollum 1974), homopolymeric tailing of linear duplex DNA (Deng and Wu 1983; Eschenfeldt et al. 1987), oligodeoxyribonucleotide and DNA labeling (Deng and Wu 1983; Tu and Cohen 1980; Vincent et al. 1982; Kumar et al. 1988; Igloi and Schiefermayr 1993), rapid amplification of cDNA ends (Frohman et al. 1988) and in situ localization of apoptosis (Gorczyca et al. 1993). Scheele and Fukuoka (1992, 1997) used TdT to add homopolymeric oligo dC tract to 3′ end of ss linear DNA in order to facilitate synthesis of ds linear DNA using oligo dG primer. However, no attempt to use TdT to synthesize a defined DNA sequence has been reported. Perhaps the random sequence generation by TdT in an uncontrolled system has prevented this."

Loop 1 modulates the fidelity of DNA polymerase λ

Conferring a template-dependent polymerase activity to terminal deoxynucleotidyltransferase by mutations in the Loop1 region
http://nar.oxfordjournals.org/content/early/2009/06/05/nar.gkp460.short
[25] (cited, not completely read yet)

~~"... a deletion of the entire Loop1 as in pol λ does confer a limited template-dependent polymerase behavior to Tdt while a chimera bearing an extended pol μ Loop1 reproduces pol μ behavior."~~

Sustained active site rigidity during synthesis by human DNA polymerase μ
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4164209/

Structures of intermediates along the catalytic cycle of terminal deoxynucleotidyltransferase: dynamical aspects of the two-metal ion mechanism
http://www.sciencedirect.com/science/article/pii/S0022283613004415

Functional analyses of polymorphic variants of human terminal deoxynucleotidyl transferase
http://www.nature.com/gene/journal/v16/n6/abs/gene201519a.html

" ... evaluate the activity and the N-addition levels of six natural (SNP) variants of hTdT. In vitro, the variants differed from wild-type hTdT in polymerization ability with four having significantly lower activity"

~~De novo DNA synthesis by human DNA polymerase λ, DNA polymerase μ and terminal deoxyribonucleotidyl transferase~~
http://www.sciencedirect.com/science/article/pii/S0022283604003705

~~"... Furthermore we found that the amino acid Phe506 of pol λ is essential for the de novo synthesis."~~

~~Promiscuous DNA synthesis by human DNA polymerase θ~~
http://nar.oxfordjournals.org/content/early/2011/11/30/nar.gkr1102.short

~~DNA polymerases that perform template-independent DNA synthesis~~
http://link.springer.com/chapter/10.1007/978-3-642-39796-7_5

??? The Effects of Magnesium Ions on the Enzymatic Synthesis of Ligand-Bearing Artificial DNA by Template-Independent Polymerase
http://www.mdpi.com/1422-0067/17/6/906/htm

Limited terminal transferase in human DNA polymerase defines the required balance between accuracy and efficiency in NHEJ
http://www.pnas.org/content/106/38/16203.full.pdf

"By site-directed mutagenesis in human Pol, we have identified a specific DNA ligand residue (Arg387) that is responsible for its limited terminal transferase activity compared to that of human TdT, its closest homologue (42% amino acid identity). Pol mutant R387K (mimicking TdT) displayed a large increase in terminal transferase activity, but a weakened interaction with ssDNA. That paradox can be explained by the regulatory role of Arg387 in the translocation of the primer from a nonproductive E:DNA complex to a productive E:DNA:dNTP complex in the absence of a templating base, which is probably the rate limiting step during template-independent synthesis. Further, we show that the Pol switch from terminal transferase to templated insertions in NHEJ reactions is triggered by recognition of a 5-P at a second DNA end, whose 3-protrusion could provide a templating base to facilitate such a special ‘‘pre-catalytic translocation step.’"

Terminal Deoxynucleotidyl Transferase: The Story of a Misguided DNA Polymerase
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2846215/

"Attempts to overexpress TdT in bacteria systems have generally failed due to three prominent factors: (1) mismatches in the codon frequencies and tRNA pools between E. coli and eukaryotes, (2) low solubility of the protein in these systems, and (3) lower levels of enzyme activity [59, 74, 75]. These complications can be alleviated by overexpressing a rare argU tRNA in the E. coli system and growing cultures at 15°C [64, 76] to boost the production of soluble and active forms of the enzyme. Purification of TdT to apparent

homogeneity is accomplished by column chromatography which is facilitated by the presence of a hexa-histidine tag attached to the N-terminus that does not affect its activity [76]. Recombinant human TdT has also been overexpressed in the baculovirus expression system and purified to homogeneity in one step from Trichoplusia ni larvae using a monoclonal antibody affinity column [59, 74]. This last advancement has played an instrumental role in generating significant amounts of protein required for structural characterization of TdT."

"The unique activity of TdT to incorporate nucleotides in a template-independent manner was reported simultaneously by Karkow and Kamen [77] and Bollum [78] using calf thymus extracts. The template-independent activity of TdT was distinguished from that of other template-dependent DNA polymerases by comparing levels of the incorporation of radiolabeled nucleotides using single-stranded versus double-stranded DNA [1]. Activity measured using single-stranded DNA is attributed only to that of TdT [2]. Subsequent biochemical studies confirmed that TdT requires a single-stranded initiator that is at least three nucleotides long with a free 5′-phosphate end and a free 3′-hydroxyl end for extension [78]. The replication of homopolymers by TdT requires an initiator chain of more than six nucleotides for poly(dA) and more than five nucleotides for poly(dT) [79]. The presence of a ribonucleotide 5′-monophosphate (rNMP) at the 3′-end of the primer does not inhibit the enzyme-primer complex formation [80]. However, further elongation occurs at a slower rate and the addition of more than two ribonucleotides to the single-stranded initiator does inhibit activity [80]. TdT is also unique for its ability to perform de novo synthesis of polynucleotides ranging in size from 2- to 15-mers when provided with dNTPs in absence of a primer [81]. These DNA fragments are hypothesized to act as signals for DNA repair or recombination machinery [81]. There may indeed be credence to this provocative hypothesis since small RNAs, designated as microRNA, influence the activity of many biological pathways [82]. It will prove interesting to firmly test this hypothesis to evaluate if de novo DNA synthesis does indeed occur in vivo and that small DNA fragments can regulate biological processes such as DNA repair and recombination."

"All DNA polymerases require the presence of a divalent metal ion to catalyze the phosphoryl transfer reaction associated with nucleotide incorporation. The general catalytic mechanism for phosphoryl transfer utilized by DNA polymerases is provided in Figure 5. In this mechanism, an aspartate residue near the deoxyribose sugar of the incoming dNTP serves as the general base to abstract the proton from the 3′-OH to generate a more reactive nucleophile. The electron rich 3′-oxygen then attacks the α-phosphate creating a trigonal-bipyramidal penta-coordinated transition state. This step results in the inversion of the α-phosphate stereochemistry [83, 84] and the concerted release of the pyrophosphate leaving group coordinated to another divalent metal ion [85–88]."

"TdT, like all DNA polymerases, also requires divalent metal ions for catalysis [73, 89]. However, TdT is unique in its ability to use a variety of divalent cations such as Co^{2+}, Mn^{2+}, Zn^{2+} and Mg^{2+}. In general, the extension rate of the primer p(dA)n (where n is the chain length from 4 through 50) with dATP in the presence of divalent metal ions is ranked in the following order: $Mg^{2+} > Zn^{2+} > Co^{2+} > Mn^{2+}$ [73]. In addition, each metal ion has different effects on the kinetics of nucleotide incorporation. For example, Mg^{2+} facilitates the

preferential utilization of dGTP and dATP whereas Co2+ increases the catalytic polymerization efficiency of the pyrimidines, dCTP and dTTP [90]. Zn2+ behaves as a unique positive effector for TdT since reaction rates with Mg2+ are stimulated by the addition of micromolar quantities of Zn2+ [90]. This enhancement may reflect the ability of Zn2+ to induce conformational changes in TdT that yields higher catalytic efficiencies [90]. Polymerization rates are lower in the presence of Mn2+ compared to Mg2+, suggesting that Mn2+ does not support the reaction as efficiently as Mg2+ [73]."

"A DNA polymerase is classically defined as an enzyme that catalyzes template-dependent addition of mononucleotides into a growing primer in the presence of four deoxynucleotides (dNTPs) [91]. Nevertheless, many archaeal, bacterial, viral, and eukaryotic DNA polymerases also catalyze non-templated nucleotide additions to the 3′-termini of blunt-ended DNA [92–96]. In addition, numerous DNA polymerases have been demonstrated to incorporate nucleotides opposite non-instructional DNA lesions such as abasic sites via translesion DNA synthesis [96–103]. In both of these cases, however, these polymerases require the use of duplex DNA as a substrate for efficient catalysis to occur. In fact, the sole polymerase that can add nucleotides using only a single-stranded DNA initiator as a substrate is the template-independent DNA polymerase, TdT [1, 2]. Terminal deoxynucleotidyl transferase activity has also been observed for pol μ [38] and pol λ [104, 105], although DNA synthesis is primarily restricted to the use of a DNA template [38, 104, 105]. In the sections provided below, we compare and contrast the mechanism of nucleotide incorporation catalyzed by TdT with template-dependent DNA polymerases. Particular emphasis is placed on the replication of non-instructional DNA lesions as this represents the most frequent mode of template-independent DNA synthesis catalyzed by these enzymes."

"A naïve mechanism to explain the utilization of these non-natural nucleotides is that the primary molecular determinant for nucleotide binding resides simply with interactions of the negatively-charged triphosphate moieties with positively charged amino acids in the polymerase’s active site. This is unlikely to be the sole source for binding as it cannot account for the significant differences observed for the incorporation of various natural and non-natural nucleotide substrates. For example, 5-NIMP is incorporated and extended 10-fold more efficiently than the closely related analog, 5-AIMP, which is poorly incorporated even at concentrations greater than 500 μM [113]. In addition, the replacement of an active site arginine residue (R336) involved in binding the triphosphate moiety of a dNTP with either glutamine (R336Q) or alanine (R336A) does not completely abolish nucleotide binding. Instead, these amino acid substitutions reduce the binding affinities of dGTP and dATP by only 10-fold [59]. These data indicate that ionic interactions between the active site arginine and the triphosphate group are important but not essential for nucleotide binding."

"A final point to discuss is with respect to the kinetics of elongation. Our studies with the various 5-substituted indolyl deoxynucleotides provided in Figure 6 reveal an inverse correlation in the kinetics of primer elongation as a function of nucleobase size [113]. For example, large analogs such as 5-PhIMP, 5-CEIMP, and 5-NapIMP are incorporated but not efficiently elongated whereas their smaller counterparts (IndMP, 5-FIMP, and 5-NIMP) are elongated with similar efficiency as natural nucleotides [113]. The inability of TdT to elongate bulky non-natural nucleotides likely results from the steric constraints imposed by the

16-amino acid loop present within TdT (see below) that appears to function as a steric gate to provide selectivity for single-stranded versus duplex DNA."

"All template-dependent DNA polymerases characterized to date by structural methods show an overall molecular architecture that is analogous to a right hand containing thumb, fingers, and palm subdomains (Figure 8, inset 1) (reviewed in [86]). Although these subdomains work synergistically during the polymerization process, it is easier to describe their functions independently. The palm subdomain is considered to be the heart of the polymerase as this is where the phosphoryl transfer reaction takes place. This subdomain contains at least two carboxylate residues that are highly conserved amongst all DNA polymerases [138] which function to coordinate metal ions that act as Lewis acids. These metal ions lower the activation energy barrier for the phosphoryl transfer reaction by stabilizing the build-up of the negative charge that accumulates on the α-phosphate and the oxyanion of the β- and γ-phosphate leaving group of the incoming dNTP substrate [88]. The function of the fingers subdomain is to coordinate interactions between the templating base and the incoming dNTP. These interactions are designed to align the nucleobases properly prior to catalysis. With template-dependent DNA polymerases such as pol β [87, 139, 140], the thumb subdomain serves a dual role by positioning duplex DNA for accepting the incoming dNTP as well as for translocation of the polymerase to the next templating base position after catalysis."

"Delarue and colleagues originally pointed out that TdT appears to be sealed in a closed conformation that is similar to pol β [127], suggesting that TdT can add nucleotides without undergoing an obligatory conformational change [127]."

"One group of proteins reported to regulate the activity of TdT are referred to as TdT interacting factors (TdiFs). TdiF1 is a protein that binds to the C-terminus of TdT and increases its polymerase activity by ~4-fold [108]. TdiF2 is another protein that binds TdT through its C-terminus, possibly through interactions with the proline-rich and pol β-like domains. However, the interaction of TdiF2 with TdT decreases the polymerase activity of TdT by ~2-fold [108]."

"One example is the nucleoside analog cordycepin (3′-deoxy adenosine) (Figure 11A). This nucleoside analog is cytotoxic against TdT-positive leukemias, especially when used in combination with the adenosine deaminase inhibitor deoxycoformycin [158, 159]. The cytotoxic activity of cordycepin correlates with the ability of the triphosphate form to inhibit single-stranded DNA synthesis by TdT in vitro [160]. Since cordycepin lacks a 3′-OH moiety, incorporation of this analog terminates primer extension and generates abortive intermediates along the recombination pathway that may induce apoptosis."

Evolving DNA repair polymerases from double strand break repair to base excision repair and VDJ recombination
http://cdn.intechopen.com/pdfs/42471/InTech-Evolving_dna_repair_polymerases_from_double_strand_break_repair_to_base_excision_repair_and_vdj_recombination.pdf

"In the case of TdT, residue Phe401 (corresponding to Phe385 in Polμ), is involved in maintaining the fixed position of Loop 1 via a strong stacking interaction between its aromatic ring and His475 (His459 in Polμ), located in a mini-loop at the thumb subdomain (Fig. 7). Mutant F401A in TdT had a striking phenotype, turning a completely template independent enzyme into a DNA-instructed DNA polymerase [90]. This mutation clearly disrupted the network of interactions needed to maintain a fixed orientation of TdT Loop 1, that is now endowed with a greater degree of flexibility, as in Polμ, thus allowing TdT to accept a template strand. Phe389 is again conserved among Polμs and TdTs (Phe405) of different species, and in both cases it seems to be involved in maintaining the shape and orientation of this motif. Mutation of this residue to alanine in TdT abolishes terminal transferase activity and allows templated insertion of only one nucleotide on a template/primer substrate [90]"

"We produced mutants in the implicated residues of Polμ and all of them lacked terminal transferase activity, indicating that the network of interactions maintaining the conformation of Loop 1 in TdT is conserved in Polμ [92]. Also, in TdT Loop 1 is interacting with another very small loop located in the thumb through His475 (Fig. 7), that is conserved in Polμ (His459). This mini-loop is also present in the other members of the X family, but its function is different: residues from this loop directly interact with the template strand. In Polμ this mini-loop has both roles: depending on the substrate used and the desired conformation of Loop 1, the mini-loop may interact either with the template strand (through Asn457) or with Loop 1 itself (through His459). Accordingly, the asparagine is only needed during templated additions, and dispensable for terminal transferase activity of Polμ, while the histidine had the opposite effect [92]. We propose a regulatory function for the NSH motif in the thumb mini-loop, helping to accommodate either the template strand (as in Polß of Polλ) or Loop 1 (as in TdT) as suits best for each individual situation."

"Why and how the terminal transferase activity of TdT is much higher than that of Polμ? Combined structural and functional evidences for both Polμ and TdT indicate that there is one residue modulating the terminal transferase activity of both enzymes. That residue (Arg387 in Polμ and Lys403 in TdT) tunes the catalytic efficiency of the terminal transferase reaction, by regulating the rate-limiting step. Judging by the structural data available, this residue could be establishing dual and alternative interactions during the catalytic cycle of both Polμ and TdT: when the primer is bound at the unproductive position (TdT crystal 1KDH), the residue is interacting with the primer strand, while in the Polμ crystal in which the primer strand is correctly positioned in a productive complex (2IHM), the arginine is interacting with the -3 position of the template strand (Fig. 8B). In the case of Polμ, and assuming an alternative interaction as that seen in TdT, Arg387 acts as a brake for the necessary movement of the primer, to limit nucleotide additions before end bridging. In fact, the single change of this residue for the TdT counterpart (Polμ mutant R387K) showed an increase in untemplated additions that ranged from 10- to 100-fold, reaching levels comparable to those of TdT itself [93]."

Ribozyme-related references

High priority

[IIRZYM] In-ice evolution of RNA polymerase ribozyme activity
http://diyhpl.us/~bryan/papers2/bio/In-ice%20evolution%20of%20RNA%20polymerase%20ribozyme%20activity%20-%202013.pdf

[RNAWS2015] RNA synthesis by in vitro selected ribozymes for recreating an RNA world
http://diyhpl.us/~bryan/papers2/bio/RNA%20synthesis%20by%20in%20vitro%20selected%20ribozymes%20for%20recreating%20an%20RNA%20world%20-%202015.pdf

[RCT2011] Ribozyme-catalyzed transcription of an active ribozyme - polymerase ribozyme
http://diyhpl.us/~bryan/papers2/bio/Ribozyme-catalyzed%20transcription%20of%20an%20active%20ribozyme%20-%20polymerase%20ribozyme%20-%202011.pdf

[CD2013] Chemical fidelity of an RNA polymerase ribozyme
http://diyhpl.us/~bryan/papers2/bio/Chemical%20fidelity%20of%20an%20RNA%20polymerase%20ribozyme%20-%202013.pdf

[RCRP2001] RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension
http://diyhpl.us/~bryan/papers2/bio/RNA-catalyzed%20RNA%20polymerization:%20Accurate%20and%20general%20RNA-templated%20primer%20extension%20-%20Johnston%20-%202001.pdf

[TFYRV] Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today?
http://diyhpl.us/~bryan/papers2/bio/Thirty-five%20years%20of%20research%20into%20ribozymes%20and%20nucleic%20acid%20catalysis:%20where%20do%20we%20stand%20today%3f%20-%20review%20-%202016.pdf

Comparison of transition state models of ribozyme and protein RNA polymerase - Schechner
http://diyhpl.us/~bryan/papers2/bio/Comparison%20of%20transition%20state%20models%20of%20ribozyme%20and%20protein%20RNA%20polymerase%20-%20Schechner.gif

Polymerase structure and mechanism

High priority

[309408fafq] A specific loop in human DNA polymerase mu allows switching between creative and DNA-instructed synthesis

http://diyhpl.us/~bryan/papers2/polymerase/A%20specific%20loop%20in%20human%20DNA%20polymerase%20mu%20allows%20switching%20between%20creative%20and%20DNA-instructed%20synthesis.pdf

[Mu2015] Creative template-dependent synthesis by human polymerase mu
http://diyhpl.us/~bryan/papers2/polymerase/Creative%20template-dependent%20synthesis%20by%20human%20polymerase%20mu%20-%202015.pdf

[201609r0321] Computational evaluation of nucleotide insertion opposite expanded and widened DNA by the translesion synthesis polymerase Dpo4
http://diyhpl.us/~bryan/papers2/polymerase/Computational%20evaluation%20of%20nucleotide%20insertion%20opposite%20expanded%20and%20widened%20DNA%20by%20the%20translesion%20synthesis%20polymerase%20Dpo4%20-%202016.pdf

[FRET2010] Conformational transitions in DNA polymerase I revealed by single-molecule FRET
http://diyhpl.us/~bryan/papers2/polymerase/Conformational%20transitions%20in%20DNA%20polymerase%20I%20revealed%20by%20single-molecule%20FRET%20-%202010.pdf

DNA Polymerases: Structural Diversity and Common Mechanisms

Rev1 Employs a Novel Mechanism of DNA Synthesis Using a Protein Template

Mutagenesis of polB dNTP binding pocket to probe fidelity

A Single Highly Mutable Catalytic Site Amino Acid Is Critical for DNA Polymerase Fidelity

Burgers et al. (2001) “Eukaryotic DNA polymerases: proposal for a revised nomenclature”

Hübscher et al. (2002) “Eukaryotic DNA Polymerases” Annual Review of Biochemistry Vol. 71: 133-163

Alba (2001) “Protein Family Review: Replicative DNA Polymerases” Genome Biology 2(1):reviews 3002.1-3002.4

and Steitz (1999) “DNA polymerases: structural diversity and common mechanisms” J Biol Chem 274:17395-17398

Crameri et al. (1998) “DNA shuffling of a family of genes from diverse species accelerates directed evolution”

Techniques used to study the DNA polymerase reaction pathway
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2846202/

"""

In separate studies of Klentaq and Pol I(KF), measurement of the rate of change of the FRET signal by stopped-flow (Fig. 6b) indicated that the fingers-closing transition is faster than the dNTP incorporation rate measured by chemical quench [29, 35], thus placing the fingers-closing ahead of the step that is rate-limiting for chemical incorporation. A third study, also of Pol I(KF), obtained a slower rate for fingers-closing, almost certainly due to interference in the reaction by the donor fluorophore which was attached close to the DNA primer terminus [36]. A more recent study of fingers subdomain movement in Klentaq used a different strategy for placement of the FRET probes, with both fluorophores attached to the protein [37]. The results obtained were similar to those described above, but the labeling strategy, by avoiding the need to place a FRET probe on the DNA, allows greater flexibility for future studies."

"Despite the focus of fluorescence studies on the prechemistry steps of the DNA polymerase mechanism, the nature of the slow step that is rate-limiting for dNTP incorporation remains a mystery. In some cases, as described earlier for T7 DNA polymerase, it is possible that a more detailed analysis of the reaction pathway will reveal that the chemical step itself is rate-limiting [18]. In others, exemplified by Klenow fragment, the rate-limiting step may involve relatively subtle structural changes and consequently be fluorescently silent with the probes that have been used. A stopped-flow fluorescence experiment, in which Ca2+ was used instead of Mg2+, suggests that the rate-limiting prechemistry step of Klenow fragment could involve the entry into the active site of the metal ion that activates the primer 3′OH [29]. Likewise, it has been suggested that entry of the second catalytic metal ion could take place immediately before chemistry in the DNA polymerase β mechanism [22, 23]."

Figure 8 shows the chemical reaction pathway provided by DNA polymerase (in the above paper).

Nucleotide insertion initiated by van der Waals interaction during polymerase beta DNA replication
http://arxiv.org/pdf/1109.1259.pdf

Role of conformational motions in enzyme function - Selected methodologies and case studies
http://www.mdpi.com/2073-4344/6/6/81/pdf

DNA polymerase conformational dynamics and the role of fidelity-conferring residues: Insights from computational simulations
http://journal.frontiersin.org/article/10.3389/fmolb.2016.00020/full

The Closing Mechanism of DNA Polymerase I at Atomic Resolution
http://www.sciencedirect.com/science/article/pii/S0969212615002695
"... Identify a conserved histidine, to initiate catalysis, as a potential proton acceptor from the primer 3′-hydroxyl"

Molecular events during translocation and proofreading extracted from 200 static structures of DNA polymerase

http://nar.oxfordjournals.org/content/early/2016/06/20/nar.gkw555.full

"DNA polymerases in family B are workhorses of DNA replication that carry out the bulk of the job at a high speed with high accuracy. A polymerase in this family relies on a built-in exonuclease for proofreading. It has not been observed at the atomic resolution how the polymerase advances one nucleotide space on the DNA template strand after a correct nucleotide is incorporated, that is, a process known as translocation. It is even more puzzling how translocation is avoided after the primer strand is excised by the exonuclease and returned back to the polymerase active site once an error occurs. The structural events along the bifurcate pathways of translocation and proofreading have been unwittingly captured by hundreds of structures in Protein Data Bank. This study analyzes all available structures of a representative member in family B and reveals the orchestrated event sequence during translocation and proofreading."

"This enzyme catalyzes incorporation of hundreds of nucleotides each second in host cells. That is to say, its fingers domain is wagging between the open and closed positions at a frequency of several hundred Hz during normal elongation (300 Hz is used in calculations throughout this paper). The fingers domain acts as an oscillator just like one in a mechanical clock and is responsible for power management of both translocation and processive active site switching. Once this oscillation is established, each leaving pyrophosphate (PPi) bound to the fingers in the closed conformation replenishes energy to maintain the oscillation. With each push from a newly cleaved PPi resulting from dNTP hydrolysis, the closed fingers accelerate to adopt an open conformation. At the end of a large swing, the fingers slam into the N-terminal domain. When the motion of the fingers is completely halted at the open position, their momentum is transferred to power the translocation. The motion of the thumb drives a back-and-forth displacement of the product duplex during translocation. The structural events during translocation and active site switching share much of the same conformational pathway until they branch out at an advanced stage."

"Only the closed state is catalytically capable of the phosphoryl transfer reaction. It is postulated that the highly discrete states of the closed and open conformations are associated with an energy minimum and an entropy maximum, respectively (27). Interestingly, no intermediate structure between them has ever been observed by static crystallography (Figure 2C and D; Supplementary Figure S1f). The discreteness of states has also been detected in ionic current traces when a single molecule of ϕ29 DNA polymerase, another member in family B, performs translocation on an α-hemolysin nanopore in an electric field (28,29). Concerted and continuous motions of the exo and thumb are observed only when the polymerase is in the open state as many open structures scatter along c3′ dimension, suggesting a rather flat, low energy valley. The two end conformations along this dimension are denoted as near and far conformations, respectively (Figure 2D). In contrast, similar motions are absent in the closed state. That is to say, the large swing of the fingers is not in concert with motions of the exo and thumb as indicated by two orthogonal arrows closed→open and near→far (Figure 2C and D). The open conformation of the fingers is apparently a prerequisite for the concerted motions of the exo and thumb domains."

"""
[28] Dahl J.M., Mai A.H., Cherf G.M., Jetha N.N., Garalde D.R., Marziali A., Akeson M., Wang H., Lieberman K.R. Direct observation of translocation in individual DNA polymerase complexes. J. Biol. Chem. 2012;287:13407-13421.
[29] Lieberman K.R., Dahl J.M., Mai A.H., Cox A., Akeson M., Wang H. Kinetic mechanism of translocation and dNTP binding in individual DNA polymerase complexes. J. Am. Chem. Soc. 2013;135:9149-9155.
"""

"Binding of any type of triphosphate ligand, such as dNTP or non-hydrolyzable analogs (30–32), at the polymerase active site seems to be sufficient to induce the closed conformation. On the other hand, all 69 structures adopt the open state in the absence of bound triphosphate. The only exceptions are the chimeric structures 3KD1 and 3KD5 with extensively altered amino acid sequences in the fingers. A key substitute V478W among many other changes introduces a clash in the open state, which significantly shifts the equilibrium toward the closed state regardless whether a triphosphate is bound (33). Both chimeric structures with and without phosphonoformic acid bound in the active site are located at the extreme of the closed conformation where c′2 ≈ −1.5 (Figure 2F)."

Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution (2009)
http://www.pnas.org/content/106/50/21109.full.pdf

~~Multiple amino acid substitutions allow DNA polymerases to synthesize RNA~~
http://www.jbc.org/content/275/51/40266.full

~~A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides~~
http://www.pnas.org/content/95/7/3402.short
Distinct function of conserved amino acids in the fingers of Saccharomyces cerevisiae DNA polymerase α

Poly(A) tail length and Poly(A) polymerase references

Medium priority

Poly (A) tail length control is caused by termination of processive synthesis
http://www.jbc.org/content/270/6/2800.short

"Poly(A) polymerase synthesizes poly(A) tails rapidly and processively only when the substrate RNA is bound simultaneously by two stimulatory proteins, the cleavage and polyadenylation specificity factor (CPSF) and poly(A)-binding protein II (PAB II). A burst of synthesis terminates after the addition of about 250 nucleotides, a length corresponding to that of newly synthesized poly(A) tails in vivo. Further elongation is slow. Length control can be reproduced with premade poly(A) tails of different lengths and is insensitive to large changes in the elongation rate. Thus, the control mechanism truly measures the length of

the poly(A) tail. The stimulatory action of PAB II is similar on long and short tails. Coating of poly(A) with one PAB II molecule for approximately 30 nucleotides is required, such that the number of PAB II molecules in the polyadenylation complex is a direct measure of poly(A) tail length. CPSF also stimulates poly(A) polymerase on long and short tails. Long tails differ from short ones only in that they do not permit the simultaneous stimulation of poly(A) polymerase by CPSF and PAB II. Consequently, elongation of long tails is distributive. Thus, length control is brought about by an interruption of the interactions responsible for rapid and processive elongation of short tails. The 3′-end of the poly(A) tail is not sequestered in the protein-RNA complex when the correct length has been reached. Neither ATP hydrolysis nor turnover of the polymerized AMP is involved in length control."

Poly (A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation

Yeast Pab1 interacts with Rna15 and participates in the control of the poly (A) tail length in vitro.

Stimulation of poly (A) polymerase through a direct interaction with the nuclear poly (A) binding protein allosterically regulated by RNA

Poly (A) tail length is controlled by the nuclear poly (A)-binding protein regulating the interaction between poly (A) polymerase and the cleavage and polyadenylation …

Control of poly (A) tail length
http://onlinelibrary.wiley.com/doi/10.1002/wrna.56/abstract

Poly (A)-tail profiling reveals an embryonic switch in translational control

Polyadenylation-dependent control of long noncoding RNA expression by the poly (A)-binding protein nuclear 1

Rrp6p controls mRNA poly (A) tail length and its decoration with poly (A) binding proteins

The long and the short of it: the role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly (A) tail length
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3345082/

Crystal structure of human poly (A) polymerase gamma reveals a conserved catalytic core for canonical poly (A) polymerases
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3878066/

"Rigid body domain movement is a component of the induced-fit catalytic mechanism proposed for PAP 28; 39; 45. As previously reported for DNA polymerases, PAPs undergo a conformation change leading to domain closure and assembly of active site residues upon binding of the RNA primer and correct incoming nucleotide, Mg-ATP 46. It should be noted that the conformational change is less pronounced than that reported for replicative DNA

polymerases 42; 47. hPAPγ was superimposed with bPAPα based on the central domain (residues 17-58 & 173-351)(Fig. 4)39 to illustrate the movements of the catalytic and RRM domains. The rms deviation between the two models is 1.198 Å (comparing Cαs in PDB ID 4LT6 (this study) and 1Q78 24), and 1.465 Å comparing 4LT6 and 1Q79 24; 31, indicating that the structures are overall quite similar but that there are regions that differ between the two models. A closer inspection reveals that PAPγ is in a slightly more closed conformation as compared to the bPAPα structures (Fig. 4A). The catalytic domain moves perpendicular whereas the RRM moves roughly parallel to the substrate binding cleft, as previously reported for yPap1 (Fig. 4B)39. Upon domain closure, the nucleotide triphosphate tail is contacted by three conserved residues, Ser101, Lys227, and Tyr236 (Fig. 2B). The latter two residues correspond to Lys215 and Tyr224 in yPap1, whose interactions with the triphosphate moiety are only observed in the closed state 28. One difference with the yeast enzyme is that Asn238 in hPAPγ is located 6 Å from the adenine base, compared to ~3 Å for the analogous Asn226 in the nucleotide-bound closed form of yPap1 (Similarly Asn239 in bPAPα lies far from the nucleotide, about 8 Å). In the ternary complex of yPap1 with oligo(A) and MgATP Asn226 hydrogen bonds with the phosphate of the RNA primer terminal base and is within van der Waals distance of the adenine base 28. hPAPγ would be predicted to further close in the presence of both nucleotide and oligo(A). A detailed analysis of the interaction of the conserved aspargine with the RNA primer and nucleotide will have to await a structure of the human enzyme with both substrates."

Structure of mitochondrial poly(A) RNA polymerase reveals the structural basis for dimerization, ATP selectivity and the SPAX4 disease phenotype
http://nar.oxfordjournals.org/content/early/2015/08/28/nar.gkv861.short

"Polyadenylation, performed by poly(A) polymerases (PAPs), is a ubiquitous post-transcriptional modification that plays key roles in multiple aspects of RNA metabolism. Although cytoplasmic and nuclear PAPs have been studied extensively, the mechanism by which mitochondrial PAP (mtPAP) selects adenosine triphosphate over other nucleotides is unknown. Furthermore, mtPAP is unique because it acts as a dimer. However, mtPAP's dimerization requirement remains enigmatic. Here, we show the structural basis for mtPAP's nucleotide selectivity, dimerization and catalysis. Our structures reveal an intricate dimerization interface that features an RNA-recognition module formed through strand complementation. Further, we propose the structural basis for the N478D mutation that drastically reduces the length of poly(A) tails on mitochondrial mRNAs in patients with spastic ataxia 4 (SPAX4), a severe and progressive neurodegenerative disease."

Other other references

Low priority

Efficient labelling of enzymatically synthesized vinyl-modified DNA by an inverse-electron-demand Diels–Alder reaction
https://kops.uni-konstanz.de/bitstream/handle/123456789/28692/Busskamp_286929.pdf?sequence=1

Polymerase enzyme substrates with protein shield
https://patents.google.com/patent/US20130316912A1

Molecular adaptors for dye conjugates
https://patents.google.com/patent/US20120058473A1/

Scaffold-based polymerase enzyme substrates
https://patents.google.com/patent/US20120077189A1/

Alpha-morpholino ribonucleoside derivatives and polymers thereof
https://patents.google.com/patent/US5235033A/en

Uncharged morpholino-based polymers having achiral intersubunit linkages
https://patents.google.com/patent/US5034506A

Template immobilization
U.S. patent application Ser. No. 12/562,690

Mechanical tension

Medium priority

Tuning and switching a DNA polymerase motor with mechanical tension

Tension-induced attenuation of polymerase activity

Pausing of DNA Polymerases on Duplex DNA Templates due to Ligand Binding in Vitro
http://linkinghub.elsevier.com/retrieve/pii/S0022283603000445

Towards mechanical characterization of biomolecules by MNEMS tools
http://doi.wiley.com/10.1002/tee.20154

Mechanical and chemical unfolding of a single protein: A comparison
http://www.pnas.org/cgi/content/abstract/96/7/3694

Raman scattering sequencing

Low priority

Nucleic acid sequencing by Raman monitoring of uptake of nucleotides during molecular replication
https://www.google.com/patents/US20050147979

Methods and device for DNA sequencing using surface enhanced Raman scattering (SERS)

http://www.google.com/patents/US20040110208

Denaturation

Low priority

Selective destruction of protein function by chromophore-assisted laser inactivation.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC281775/

Genetically targeted chromophore-assisted light inactivation
http://www.ncbi.nlm.nih.gov/pubmed/14625562

Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores

Electronics

High priority

[EBS2015] Polymerase-DNA interactions and enzymatic activity: multi-parameter analysis with electro-switchable biosurfaces
http://diyhpl.us/~bryan/papers2/polymerase/Polymerase-DNA%20interactions%20and%20enzymatic%20activity:%20multi-parameter%20analysis%20with%20electro-switchable%20biosurfaces%20-%202015.pdf

[BND481085] Electronic control of DNA polymerase binding and unbinding to single DNA molecules
http://diyhpl.us/~bryan/papers2/polymerase/Electronic%20control%20of%20DNA%20polymerase%20binding%20and%20unbinding%20to%20single%20DNA%20molecules%20-%202009.pdf

[ARRDNP] Automated forward and reverse ratcheting of DNA in a nanopore
http://diyhpl.us/~bryan/papers2/DNA/Automated%20forward%20and%20reverse%20ratcheting%20of%20DNA%20in%20a%20nanopore%20-%20Cherf%20-%202012.pdf

[DNANPTRNS] Challenges in DNA motion control and sequence readout using nanopore devices
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4710574/

[TDNSQ] Three decades of nanopore sequencing
http://www.nature.com/nbt/journal/v34/n5/abs/nbt.3423.html

DNA sequencing technology based on nanopore sensors by theoretical calculations and simulations
http://link.springer.com/article/10.1007/s11434-014-0622-x

[PNPHI] Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3076064/

[SMNDSNR] A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2453067/

Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3757088/

DNA sequencing using electrical conductance measurements of a DNA polymerase
https://ir.nctu.edu.tw/bitstream/11536/22351/1/000319979400018.pdf
https://www.google.com/patents/US20150017655
(similar) https://www.google.com/patents/US20150065353

Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment)
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3738269/

"Consistent with these nanocircuit-based observations, the enzyme closed complex forms a phosphodiester bond in a highly efficient process >99.8% of the time with a mean duration of only 0.3 ms for all four dNTPs. The rate-limiting step for catalysis occurs during the enzyme open state, but with a nearly two-fold longer duration for dATP or dTTP incorporation than for dCTP or dGTP into complementary, homopolymeric DNA templates."

A protein transistor made of an antibody molecule and two gold nanoparticles

Electrochemical bioelectronic device consisting of metalloprotein for analog decision making
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4585857/

Recent progress in atomistic simulation of electrical current DNA sequencing
http://arxiv.org/pdf/1411.3437

Proteins as electronic materials: electron transport through solid-state protein monolayer junctions
http://arxiv.org/pdf/0903.4255

Bacteriorhodopsin as an electronic conduction medium for biomolecular electronics
https://www.researchgate.net/profile/David_Cahen/publication/23412882_ChemInform_Abstract_Bacteriorhodopsin_as_an_Electronic_Conduction_Medium_for_Biomolecular_Electronics/links/0fcfd5107f4fa3bb91000000.pdf

Molecular electronics: effect of external electric field

http://arxiv.org/pdf/0807.3505

Effects of high external electric fields on protein conformation
http://www.academia.edu/download/44424937/Proceedings_of_SPIE_2005.pdf

Patolsky et al., “Electrical Detection of Viruses”, PNAS, 101(39), 14017, 2004

Surface reaction limited model for the evaluation of immobilized enzyme on planar surfaces
https://ir.nctu.edu.tw/bitstream/11536/7386/1/000264759400043.pdf

Topological and electron-transfer properties of glucose oxidase adsorbed on highly oriented pyrolytic graphite electrodes
http://pubs.acs.org/doi/abs/10.1021/jp071122h

Conformational Behavior of Genetically-Engineered Dodecapeptides as a Determinant of Binding Affinity for Gold.
http://pubs.acs.org/doi/abs/10.1021/jp404057h

Modeling and simulation of protein–surface interactions: achievements and challenges
http://journals.cambridge.org/article_S0033583515000256

Photoswitching AAs

Highest priority

[CAA2015] Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells
http://diyhpl.us/~bryan/papers2/bio/Genetically%20encoding%20photoswitchable%20click%20amino%20acids%20in%20Escherichia%20coli%20and%20mammalian%20cells%20-%202015.pdf

In situ formation of an azo bridge on proteins controllable by visible light
http://pubs.acs.org/doi/abs/10.1021/jacs.5b06234

Photosensitive GFP mutants containing an azobenzene unnatural amino acid
http://www.sciencedirect.com/science/article/pii/S0960894X14013432

Optical Control of a Neuronal Protein Using a Genetically Encoded Unnatural Amino Acid in Neurons
http://www.ncbi.nlm.nih.gov/pubmed/27078635

Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4331201/

Synthesis and Site-Specific Incorporation of Red-Shifted Azobenzene Amino Acids into Proteins
http://pubs.acs.org/doi/abs/10.1021/acs.orglett.5b03268?journalCode=orlef7

Synthesis and Evaluation of Photoreactive Tetrazole Amino Acids
http://pubs.acs.org/doi/abs/10.1021/ol901300h

Selective Functionalization of a Genetically Encoded Alkene-Containing Protein via "Photoclick Chemistry" in Bacterial Cells
http://pubs.acs.org/doi/abs/10.1021/ja803598e

Fluorogenic, Two-Photon-Triggered Photoclick Chemistry in Live Mammalian Cells

A Biosynthetic Route to Photoclick Chemistry on Proteins

Protein Surface Recognition by Synthetic Receptors: A Route to Novel Submicromolar Inhibitors for α-Chymotrypsin
http://pubs.acs.org/doi/abs/10.1021/ja981504o

A Visible-Light-Triggered Conformational Diastereomer Photoswitch in a Bridged Azobenzene
http://onlinelibrary.wiley.com/doi/10.1002/chem.201601400/full

Light-driven electron injection from a biotinylated triarylamine donor to [Ru(diimine)3]2+-labeled streptavidin
http://pubs.rsc.org/en/content/articlelanding/2016/ob/c6ob01273f/unauth#!divAbstract

DNA origami

High priority

Prediction and design of DNA and RNA structures

Mechanical design of DNA nanostructures
http://diyhpl.us/~bryan/papers2/bio/dna-origami/Mechanical%20design%20of%20DNA%20nanostructures%20-%202015.pdf

Investigating the dynamics of surface-immobilized DNA nanomachines
http://diyhpl.us/~bryan/papers2/bio/dna-origami/Investigating%20the%20dynamics%20of%20surface-immobilized%20DNA%20nanomachines%20-%202016.pdf

Genetic encoding of DNA nanostructures and their self-assembly in living bacteria
http://diyhpl.us/~bryan/papers2/bio/dna-origami/Genetic%20encoding%20of%20DNA%20nanostructures%20and%20their%20self-assembly%20in%20living%20bacteria%20-%20Voigt%20-%202016.pdf

DNA nanotechnology: Understanding and optimisation through simulation
http://diyhpl.us/~bryan/papers2/bio/dna-origami/DNA%20nanotechnology:%20understanding%20and%20optimisation%20through%20simulation%20-%202015.pdf

Rational protein design and protein engineering

High priority

Advances in the directed evolution of proteins (2014)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Advances%20in%20the%20directed%20evolution%20of%20proteins%20-%202014.pdf

Computational tools for designing and engineering enzymes (2014)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Computational%20tools%20for%20designing%20and%20engineering%20enzymes.pdf

A general computational approach for repeat protein design (2014)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/A%20general%20computational%20approach%20for%20repeat%20protein%20design%20-%202014.pdf

Control of repeat protein curvature by computational protein design (2015)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Control%20of%20repeat%20protein%20curvature%20by%20computational%20protein%20design%20-%202015.pdf

Exploring the repeat protein universe through computational protein design (2016)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Exploring%20the%20repeat%20protein%20universe%20through%20computational%20protein%20design%20-%202016.pdf

Algorithms for protein design (2016)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Algorithms%20for%20protein%20design%20-%202016.pdf

Fast search algorithms for computational protein design (2016)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Fast%20search%20algorithms%20for%20computational%20protein%20design%20-%202016.pdf

The protein folding problem, 50 years on (2012)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/The%20protein%20folding%20problem,%2050%20years%20on.pdf

Why reinvent the wheel? Building new proteins based on ready-made parts (2016)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Why%20reinvent%20the%20wheel%3f%20Building%20new%20proteins%20based%20on%20ready-made%20parts%20-%202016.pdf

De novo protein design: how do we expand into the universe of possible protein structures? (2016)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/De%20novo%20protein%20design:%20how%20do%20we%20expand%20into%20the%20universe%20of%20possible%20protein%20structures%3f%20-%202016.pdf

Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments (2013)
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Design%20of%20a%20single-chain%20polypeptide%20tetrahedron%20assembled%20from%20coiled-coil%20segments%20-%202013.pdf

Design of protein function leaps by directed domain interface evolution
http://diyhpl.us/~bryan/papers2/bio/protein-engineering/Design%20of%20protein%20function%20leaps%20by%20directed%20domain%20interface%20evolution.pdf

References:

[1] G. M. Church, Y. Gao, and S. Kosuri, “Next-generation digital information storage in DNA,” *Science*, vol. 337, pp. 1628–1628, 2010.
[2] J. P. Cox, “Long-term data storage in DNA,” *Trends Biotechnol.*, vol. 19, no. 7, pp. 247–250.
[3] N. Goldman *et al.*, “Towards practical, high-capacity, low-maintenance information storage in synthesized DNA,” *Nature*, vol. 494, no. 7435, pp. 77–80, 2011.
[4] K. E. Drexler, *Nanosystems: molecular machinery, manufacturing, and computation*. John Wiley & Sons, Inc., 1990.
[5] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth, “Millisecond-timescale, genetically targeted optical control of neural activity,” *Nature Neuroscience*, vol. 8, no. 9, pp. 1263–1268, 2003.
[6] T. Chen and F. E. Romesberg, “Directed polymerase evolution.,” *FEBS Lett.*, vol. 588, pp. 219–229, Jan. 2014.
[7] F. J. Ghadessy, J. L. Ong, and P. Holliger, “Directed evolution of polymerase function by compartmentalized self-replication.,” *Proc. Natl. Acad. Sci. United States Am.*, vol. 98, pp. 4552–7, 1999.
[8] J. J. Agresti *et al.*, “Ultrahigh-throughput screening in drop-based microfluidics for directed evolution.,” *Proc. Natl. Acad. Sci. United States Am.*, vol. 107, pp. 4004–4009, Mar. 2010.
[9] A. C. Larsen, M. R. Dunn, A. Hatch, S. P. Sau, C. Youngbull, and J. C. Chaput, “A general strategy for expanding polymerase function by droplet microfluidics.,” *Nat. Commun.*, vol. 7, Apr. 2016.
[10] O. J. Miller *et al.*, “Directed evolution by in vitro compartmentalization.,” *Nat. methods*, vol. 3, pp. 561–570, Jul. 2006.
[11] W.-C. Lu and A. D. Ellington, “In vitro selection of proteins via emulsion compartments.,” *Methods*, vol. 60, pp. 75–80, Mar. 2013.
[12] J. McCafferty, A. D. Griffiths, G. Winter, and D. J. Chiswell, “Phage antibodies: filamentous phage displaying antibody variable domains,” pp. 552–4, 1988.

[13] D. Lipovsek and A. Plückthun, “In-vitro protein evolution by ribosome display and mRNA display.,” *J. Immunol. methods*, vol. 290, pp. 51–67, Jul. 2004.
[14] G. Georgiou, C. Stathopoulos, P. S. Daugherty, A. R. Nayak, B. L. Iverson, and R. Curtiss, “Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines.,” *Nat. Biotechnol.*, vol. 15, pp. 29–34, Jan. 1997.
[15] E. T. Boder and K. D. Wittrup, “Yeast surface display for screening combinatorial polypeptide libraries.,” *Nat. Biotechnol.*, vol. 15, pp. 553–557, Jun. 1997.
[16] R. W. Roberts and J. W. Szostak, “RNA-peptide fusions for the in vitro selection of peptides and proteins.,” *Proc. Natl. Acad. Sci. United States Am.*, vol. 94, pp. 12297–12302, Nov. 1997.
[17] H. S. Zaher and P. J. Unrau, “Selection of an improved RNA polymerase ribozyme with superior extension and fidelity.,” *RNA*, vol. 13, pp. 1017–1026, Jul. 2007.
[18] S. Paul, A. Stang, K. Lennartz, M. Tenbusch, and K. Überla, “Selection of a T7 promoter mutant with enhanced in vitro activity by a novel multi-copy bead display approach for in vitro evolution.,” *Nucleic acids Res.*, vol. 41, Jan. 2013.
[19] B. Sharma, Y. Takamura, T. Shimoda, and M. Biyani, “A bulk sub-femtoliter in vitro compartmentalization system using super-fine electrosprays.,” *Sci. reports*, vol. 6, May 2016.
[20] N. Ogata and T. Miura, “Genetic information ‘created’ by archaebacterial DNA polymerase.,” *Biochem. J.*, vol. 324 (Pt 2), pp. 667–671, Jun. 1997.
[21] X. Liang, K. Jensen, and M. D. Frank-Kamenetskii, “Very efficient template/primer-independent DNA synthesis by thermophilic DNA polymerase in the presence of a thermophilic restriction endonuclease.,” *Biochemistry*, vol. 43, pp. 13459–13466, Oct. 2004.
[22] A. J. Berdis, “DNA Polymerases That Perform Template-Independent DNA Synthesis,” 2012, pp. 109–137.
[23] M. Hogg, A. E. Sauer-Eriksson, and E. Johansson, “Promiscuous DNA synthesis by human DNA polymerase θ.,” *Nucleic acids Res.*, vol. 40, pp. 2611–2622, Mar. 2012.
[24] R. Juarez, J. F. Ruiz, S. A. N. McElhinny, D. Ramsden, and L. Blanco, “A specific loop in human DNA polymerase mu allows switching between creative and DNA-instructed synthesis,” *Nucleic Acids Res.*, vol. 34, Aug. 2006.
[25] F. Romain, I. Barbosa, J. Gouge, F. Rougeon, and M. Delarue, “Conferring a template-dependent polymerase activity to terminal deoxynucleotidyltransferase by mutations in the Loop1 region.,” *Nucleic acids Res.*, vol. 37, pp. 4642–4656, Aug. 2009.
[26] K. Ramadan, I. V. Shevelev, G. Maga, and U. Hübscher, “De novo DNA synthesis by human DNA polymerase lambda, DNA polymerase mu and terminal deoxyribonucleotidyl transferase.,” *J. Mol. Biol.*, vol. 339, pp. 395–404, May 2004.
[27] S. D. McCulloch and T. A. Kunkel, “The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases.,” *Cell Res.*, vol. 18, pp. 148–161, Jan. 2008.
[28] A. Henry, “Beyond A, C, G and T: augmenting nature’s alphabet,” *Curr. Opin. Chem. Biol.*, pp. 727–733, 2001.
[29] I. Hirao and M. Kimoto, “Unnatural base pair systems toward the expansion of the genetic alphabet in the central dogma.,” *Proc. Jpn. Acad. Ser. B, Phys. Biol. Sci.*, vol. 88, pp. 345–367, 2010.
[30] Y. J. Seo, D. A. Malyshev, T. Lavergne, P. Ordoukhanian, and F. E. Romesberg, “Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs.,” *J. Am. Chem. Soc.*, vol. 133, pp. 19878–19888, Dec. 2011.
[31] M. Kimoto, R. Yamashige, K. Matsunaga, S. Yokoyama, and I. Hirao, “Generation of

high-affinity DNA aptamers using an expanded genetic alphabet.," *Nat. Biotechnol.*, vol. 31, pp. 453–457, May 2013.
[32] K. Hiraga and F. H. Arnold, "General method for sequence-independent site-directed chimeragenesis.," *J. Mol. Biol.*, vol. 330, no. 2, pp. 287–296, Jul. 2003.
[33] A. Fernandez-Gacio, M. Uguen, and J. Fastrez, "Phage display as a tool for the directed evolution of enzymes.," *Trends Biotechnol.*, vol. 21, pp. 408–414, Sep. 2003.
[34] K. T. O'Neil and R. H. Hoess, "Phage display: protein engineering by directed evolution.," *Curr. Opin. Struct. Biol.*, vol. 5, pp. 443–449, Aug. 1995.
[35] J. Hanes and A. Plückthun, "In vitro selection and evolution of functional proteins by using ribosome display.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 94, pp. 4937–4942, May 1997.
[36] P. Amstutz *et al.*, "In vitro selection for catalytic activity with ribosome display.," *J. Am. Chem. Soc.*, vol. 124, pp. 9396–9403, Aug. 2002.
[37] B. Seelig, "mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries.," *Nat. Protoc.*, vol. 6, pp. 540–552, May 2011.
[38] P. A. Barendt, D. T. W. Ng, C. N. McQuade, and C. A. Sarkar, "Streamlined protocol for mRNA display.," *ACS Comb. Sci.*, vol. 15, pp. 77–81, Feb. 2013.
[39] P. H. Patel and L. A. Loeb, "Multiple amino acid substitutions allow DNA polymerases to synthesize RNA.," *J. Biol. Chem.*, vol. 275, pp. 40266–40272, Dec. 2000.
[40] M. Astatke, K. Ng, N. D. Grindley, and C. M. Joyce, "A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 95, pp. 3402–3407, Mar. 1998.
[41] G. Gao, M. Orlova, M. M. Georgiadis, W. A. Hendrickson, and S. P. Goff, "Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 94, pp. 407–411, Jan. 1997.
[42] G. Xia, L. Chen, T. Sera, M. Fa, P. G. Schultz, and F. E. Romesberg, "Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase," *Proc Natl Acad Sci U S*, vol. 99, pp. 6597–602, 2000.
[43] J. L. Ong, D. Loakes, S. Jaroslawski, K. Too, and P. Holliger, "Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide.," *J. Mol. Biol.*, vol. 361, pp. 537–550, Aug. 2006.
[44] T. J. Brunette *et al.*, "Exploring the repeat protein universe through computational protein design.," *Nature*, vol. 528, no. 7583, pp. 580–584, Dec. 2015.
[45] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," *Nature*, vol. 440, pp. 297–302, 2004.
[46] J. M. Stewart and E. Franco, "Self-assembly of large RNA structures: learning from DNA nanotechnology," *DNA and RNA Nanotechnology*, vol. 2, no. 1.
[47] H. Kim *et al.*, "Nucleic Acid Engineering: RNA Following the Trail of DNA.," *ACS Comb. Sci.*, vol. 18, no. 2, pp. 87–99, Feb. 2016.
[48] E. Benson *et al.*, "DNA rendering of polyhedral meshes at the nanoscale.," *Nature*, vol. 523, no. 7561, pp. 441–444, Jul. 2015.
[49] J. Elbaz, P. Yin, and C. A. Voigt, "Genetic encoding of DNA nanostructures and their self-assembly in living bacteria.," *Nat. Commun.*, vol. 7, p. 11179, Apr. 2016.