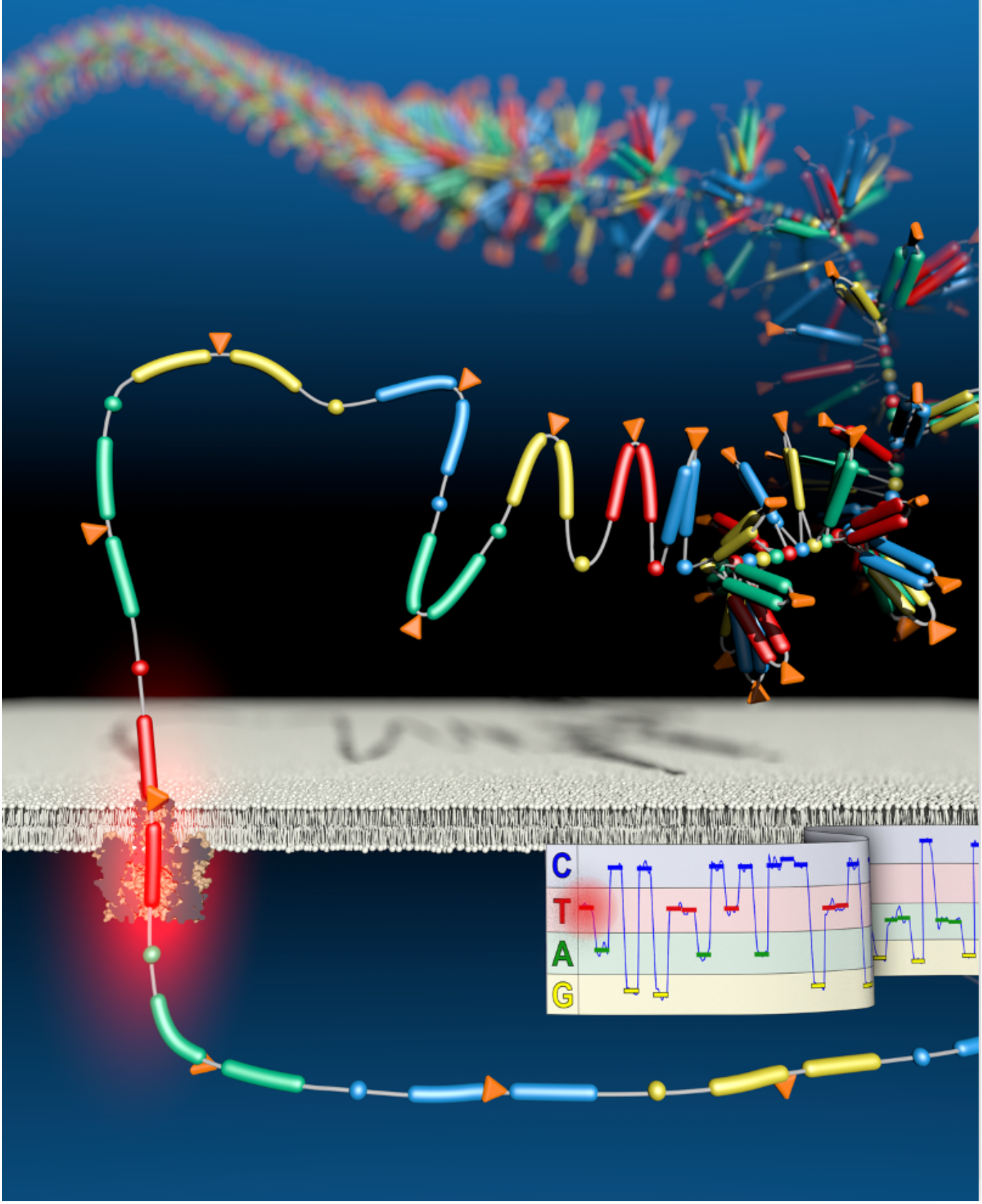
Sequencing by Expansion (SBX)

A novel, high-throughput single-molecule sequencing technology



Dear Readers,

Nucleic acid sequencing is now well established as a core readout for all molecular, cell and tissue biology, with applications from basic discovery to clinical practice. With every increase in scale or drop in cost, the applications and impact of sequencing have increased, as evidenced in the last few years by the rise of population scale efforts like UKBB, scRNA-seq based screens (Perturb-Seq) or the Human Cell Atlas. Despite great advances, high quality sequencing at scale, in terms of both speed and cost, remains a challenge. While costs decreased and scale increased dramatically, it is still an expensive step in every lab experiment and it still limits the routine application of sequencing in molecular diagnostics.

Nanopore sequencing originally offered a promising direction, but the signal-to-noise challenges of direct DNA measurement limited application of the technology. The Sequencing by Expansion (SBX) technology was invented in 2007 at Stratos Genomics, Inc. (now Roche) to address this challenge. Here, we report a key breakthrough step: Sequencing by Expansion on a single pore system. This submission will be the first publication of SBX and will describe molecular designs, syntheses and processes along with sequencing results that were generated prior to the Roche acquisition in 2020. In the coming months, we intend to publish current results for SBX, combined with our high-throughput 8M sensor array, to demonstrate the technology's performance for key application areas like somatic and germline whole genome sequencing as well as read-intensive methods like perturb seq (scRNA) and spatial sequencing, to name a few. We consider the advancements described in this manuscript to be an important demonstration of what is possible with innovative molecular, protein and systems engineering.

This technical achievement was made possible because of the development of the following key innovations:

- XNTPs, nucleoside triphosphates modified with structures engineered to control translocation, expand, provide well-resolved nanopore based reporter codes and enable template-dependent incorporation with a modified polymerase (Xp Synthase).
- PEMs, a class of critical cofactors that enable polymerase extension of XNTPs beyond 10 bases by stabilizing the elongating heteroduplex structure.
- Xp Synthases, highly mutated DNA polymerases that processively and accurately incorporate XNTPs to synthesize Xpandomers (Xp).
- Reaction formulations and processes, optimized for accurate, template-dependent synthesis of Xp using a solid-phase workflow.
- Xp end structures and sequencing conditions, engineered to efficiently and accurately thread Xp molecules though a nanopore.

For a quick overview of SBX, the URL for a concept animation is included on the following webpage: https://sequencing.roche.com/global/en/article-listing/nanopore-sequencing.html

Thank you,

Mark Kokoris (corresponding author - mark.kokoris@roche.com)

Head of SBX Technology, Roche Sequencing Solutions, Seattle WA, USA

Co-founder and CEO of Stratos Genomics, Inc. (2007-2020)

Sequencing by Expansion (SBX) – a novel, high-throughput single-molecule sequencing technology

Mark Kokoris[1,2], Robert McRuer[1,2], Melud Nabavi[1,2], Aaron Jacobs[1,2], Marc Prindle[1,2], Cynthia Cech[1,2], Kendall Berg[1,2], Taylor Lehmann[1,2], Cara Machacek[1,2], John Tabone[1,2], Jagadeeswaran Chandrasekar[1,2], Lacey McGee[1,2], Matthew Lopez[1,2], Tommy Reid[1,2], Cara Williams[1,2], Salka Barrett[1,2], Alex Lehmann[1,2], Michael Kovarik[1,2], Robert Busam[1,2], Scott Miller[1,2], Brent Banasik[1,2], Brittany Kesic[1,2], Anasha Arryman[1,2], Megan Rogers-Peckham[1,2], Alan Kimura[1,2], Megan LeProwse[1,2], Mitchell Wolfin[1,2], Svetlana Kritzer[1,2], Joanne Leadbetter[1,2], Majid Babazadeh[1,2], John Chase[2], Greg Thiessen[1,2], William Lint[1,2], Drew Goodman[1,2], Dylan O'Connell[1,2], Nadya Lumanpauw[1,2], John Hoffman[2], Samantha Vellucci[1,2], Kendra Collins[1,2], Jessica Vellucci[1,2], Amy Taylor[2], Molly Murphy[1,2], Michael Lee[1,2], Matthew Corning[2]

Affiliation

[1]Roche Sequencing Solutions Inc., 645 Elliott Avenue West, Seattle WA, 98119

[2]Formerly Stratos Genomics Inc., 2007-2020, 2211 Elliott Avenue, Seattle WA, 98121

Abstract

Remarkable advances in high-throughput sequencing have enabled major biological discoveries and clinical applications, but achieving wider distribution and use depends critically on further improvements in scale and cost reduction. Nanopore sequencing has long held the promise for such progress, but has had limited market penetration. This is because efficient and accurate nanopore sequencing of nucleic acids has been challenged by fundamental signal-to-noise limitations resulting from the poor spatial resolution and molecular distinction of nucleobases. Here, we describe Sequencing by Expansion (SBX), a single-molecule sequencing technology that overcomes these limitations by using a biochemical conversion process to encode the sequence of a target nucleic acid molecule into an Xpandomer, a highly measurable surrogate polymer. Expanding over 50 times longer than the parent DNA templates, Xpandomers are engineered with high signal-to-noise reporter codes to enable facile, high-accuracy nanopore sequencing. We demonstrate the performance of SBX and present the specialized molecular structures, chemistries, enzymes and methods that enable it. The innovative molecular and systems engineering in SBX create a transformative technology to address the needs of existing and emerging sequencing applications.

Introduction

Sequencing is now a mainstream tool that drives biological and genomics research as well as clinical applications (*1*). Sequencing technologies have been evolving continually with performance improvements and lower costs (*2*), driving broader utilization and development of new applications by offering higher accuracies, higher throughputs, faster turnaround and increasingly simplified workflows.

Nanopore measurement and its potential for sequencing has been in development for over three decades (*3–6*) because of its simple electrical implementation and single-molecule measurement sensitivity. A typical nanopore sequencer has two reservoirs of electrolyte that are fluidically connected through a nanometer-scale pore inserted into an ion barrier membrane. A potential difference between the reservoirs results in a measurable ion current passing through the pore.

If a negative polyelectrolyte such as DNA is introduced to the negative potential reservoir, it can be threaded through a right-sized pore and drawn into the positive potential reservoir. Measurement of the ion current records the level of ion blockage during the polymer translocation through the pore and holds the promise of yielding sequence information.

However, efficient and accurate single-molecule nanopore sequencing of nucleic acids remains challenging because of fundamental signal-to-noise limitations resulting from the poor spatial resolution and molecular distinction of nucleobases. For example, the Oxford Nanopore (ONT) technology (*7*), which directly sequences DNA, is limited by raw read average error rates >3% (sup basecaller) (*8*) and generally poor homopolymer resolution (*9*). While their read lengths lead the industry (*10*, *11*), throughput, costs and accuracy are not competitive with existing platforms (*12*, *13*). As a result, this approach is limited in its application. An alternative to direct nanopore sequencing with new chemistry and measurement innovations is required to meet this difficult challenge.

Here, we introduce Sequencing by Expansion (SBX), a nanopore-based single-molecule approach that overcomes these measurement limitations through a biochemical process that encodes the sequence of a target nucleic acid molecule into a highly measurable surrogate polymer called an Xpandomer (Xp) (*14*). Xpandomers encode the sequence information with high signal-to-noise reporters, enabling high-fidelity, single-molecule nanopore sequencing. We describe the innovations in molecular and protein engineering that enable SBX and demonstrate its performance with data generated using a simple, accessible nanopore set up. When combined with array-based systems, the near real-time high-throughput and workflow flexibility of SBX will allow researchers to rethink how sequencing projects can be implemented.

Results

SBX was conceived as a solution to the fundamental signal-to-noise challenges of direct DNA sequencing. Our data demonstrates that by spatially expanding the sequence information and adding translocation control features, single-molecule sequencing is achieved with high resolution and throughput [Fig. 1E-G]. Using a systems approach, purpose-built molecular structures and workflows were co-optimized to produce the observed results.

SBX Overview

SBX consists of two workflows: synthesis and sequencing. First, during synthesis, the DNA template is enzymatically and chemically converted into the lengthened Xp molecule that contains high signal-to-noise reporters [Fig. 1A-C]. Next, the Xp is sequenced by stepwise translocation through a nanopore, [Fig. 1D] where each step measures a well-defined current specific to one of four reporters that identifies the DNA base [Fig. 1E-G]. Separation of the DNA conversion chemistry from the molecular measurement allows each workflow to be independently optimized to its own unique requirements without compromising the other. This decoupling is a fundamental advantage of the SBX process. These workflows, along with the enabling molecular structures, chemistries and methods are summarized below and are further elaborated in the Supplement.

In the synthesis step, a solid-phase method is used to convert DNA to Xp. A DNA template is hybridized to a specialized, acid-stable primer called an extension oligo (EO) [Fig. 2H], which is anchored to the cyclic olefin walls of a flow channel [Fig. 2A-D]. Next, the primed DNA template is exposed to a reaction mix that comprises a highly-engineered DNA polymerase (Xp Synthase), the four base-specific expandable nucleoside triphosphate substrates (XNTPs) [Fig. 2E, Fig. 3A-B], and a reaction cofactor called a polymerase enhancing moiety (PEM), in a formulation of salts, buffer, solvents and crowding agents optimized for Xp synthesis. The Xp Synthase performs template-dependent primer extension using XNTP substrates under customized reaction conditions.

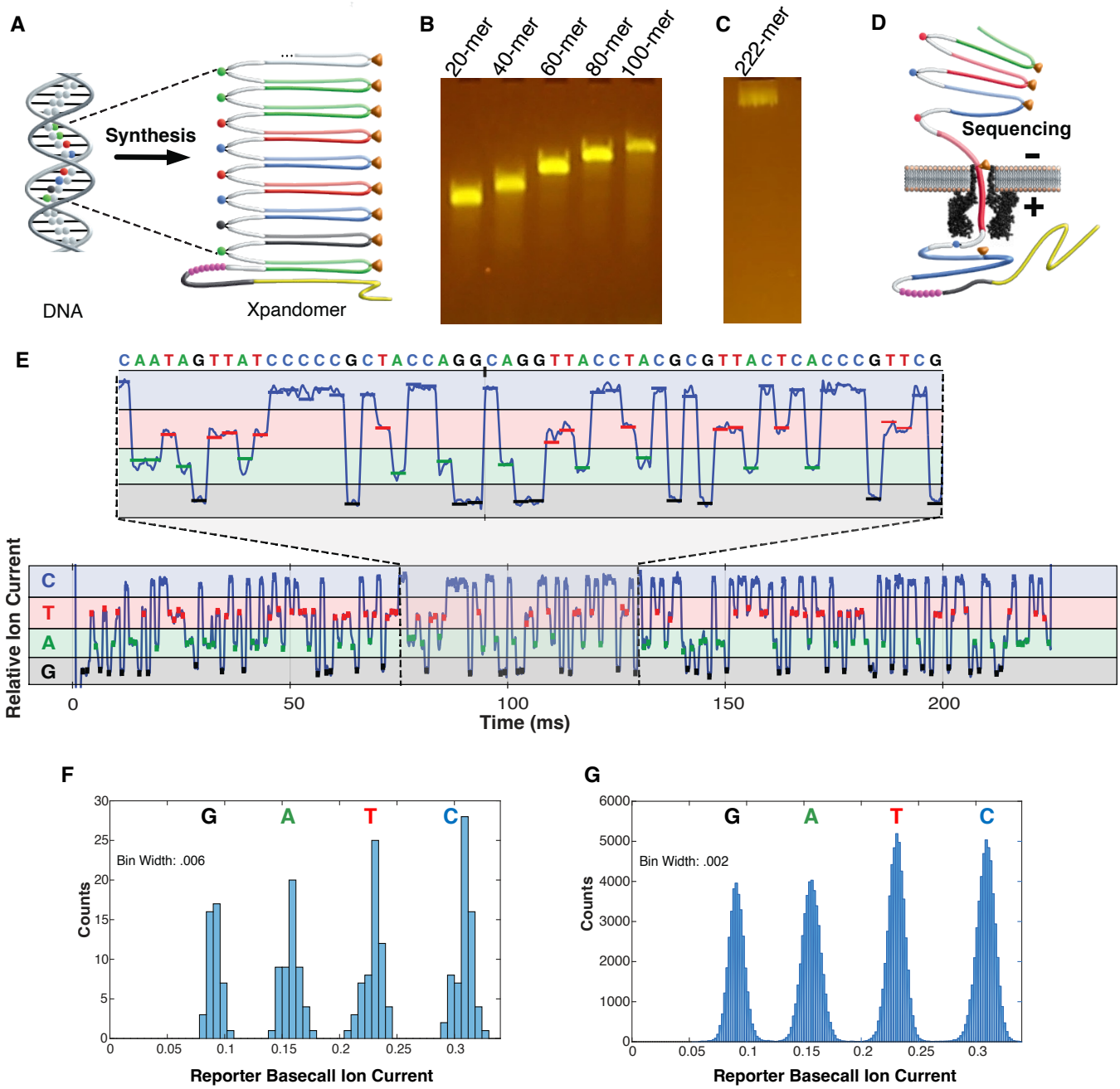


Fig. 1. SBX separates chemistry from measurement. (A) SBX chemistry synthesizes a surrogate of the DNA called the Xpandomer (Xp). **(B)** Unexpanded Xp synthesized from templates of varying length using a fluorescent primer, analyzed by electrophoresis in a 2.5% agarose gel. **(C)** Expanded Xp synthesized from a 222-mer template using a sequencing-compatible extension oligo (EO). Xp products were analyzed by electrophoresis in a 2.5% agarose gel after hybridizing a fluorescent reporter to the EO. **(D)** A biological nanopore inserted into a lipid bilayer constrains the Xp to sense individual reporters enabling high signal-to-noise, single-molecule measurement and simple sequence readout. **(E)** Example ion current trace recorded during translocation of an Xp synthesized from a 222 base *Streptococcus* DNA template. During translocation, each Xp reporter was paused in the nanopore for a 1 ms period, during which the ion current was sampled. The median ion current derived from samples captured in the last half of the period was normalized by the open pore current (72 pA) to determine the Xp reporter basecall ion current shown as 1 ms colored bars along the above trace. **(F)** Histogram of the reporter basecall ion currents from the above molecular trace. **(G)** Histogram of the reporter basecall ion currents for the set of 966 aligned Xp reads analyzed in this paper.

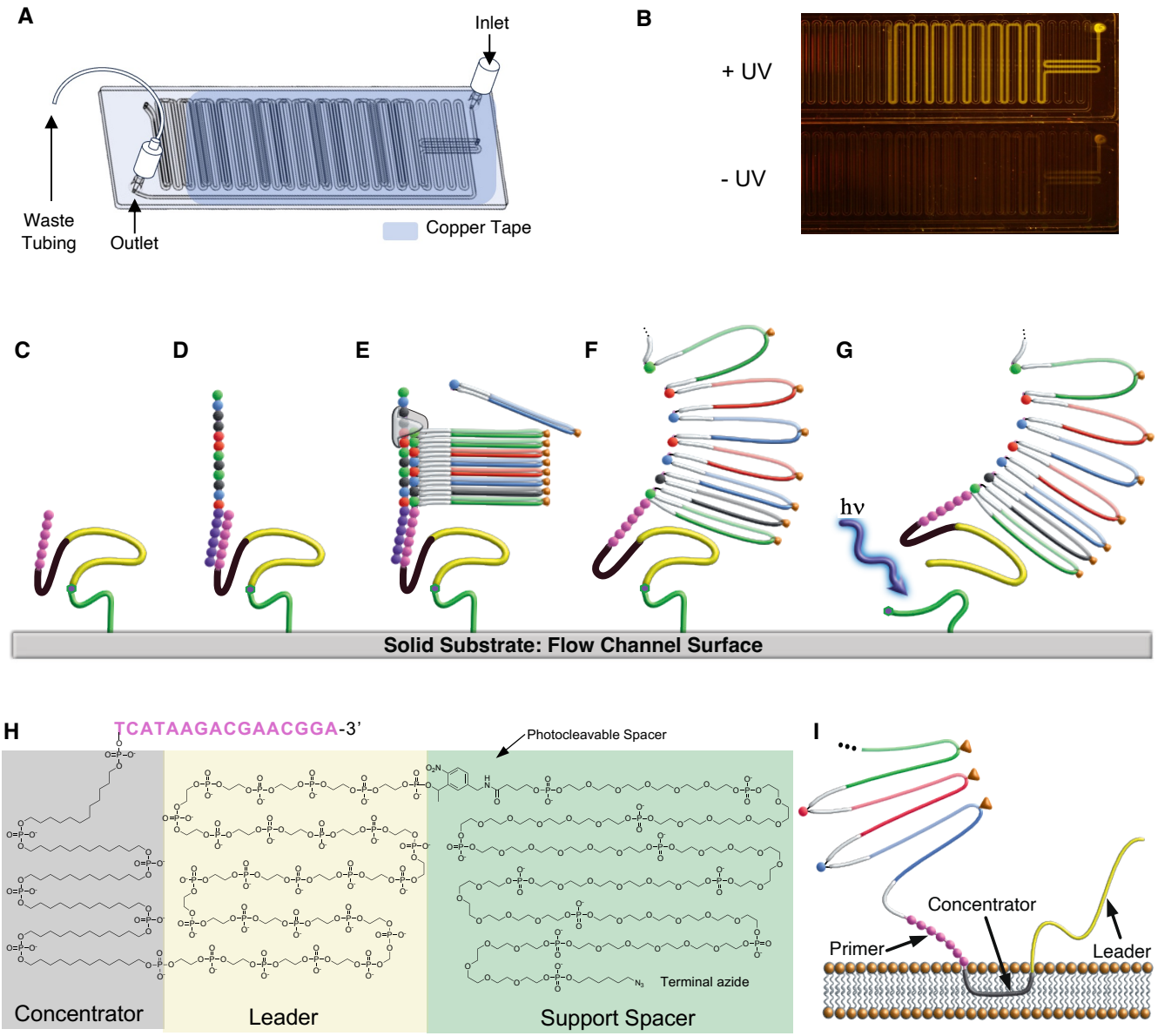


Fig. 2. Synthesis of Xp from DNA on solid support. (A) Cyclic olefin microfluidic chip (Chipshop, Fluidic 243). 0.6 mm x 0.3 mm x 1,637 mm channel dimension. ~500 mm^2 was functionalized to support Xp synthesis. **(B)** Functionalization was carried out in the first 50 µL of the channel in 80% DMSO, 1% SDS, 2 mM propargyl PEG-4 maleimide, which was irradiated (365 nm, 4.2 mW/cm^2) for 20 minutes followed by a washing step (see Supplemental). For testing purposes, a fluorescent 5'-azide oligonucleotide was clicked to the surface and unbound material was washed away. In the absence of irradiation, minimal background was observed. **(C)** For Xp synthesis, the functionalized solid substrate is conjugated to the EO through click chemistry. **(D)** Target DNA hybridizes to the support-bound EO. **(E)** Xp Synthase performs template-dependent extension with XNTP substrates. **(F)** The Xp is expanded using acid cleavage. **(G)** Release of expanded Xp from solid substrate by photocleavage. **(H)** EO features regions of hydrophobicity (concentrator), high negative charge density (leader), a photocleavable spacer (PC) and a terminal azide for conjugation to the solid support. The 3' terminus contains acid-stable 2'-OMe nucleotides to hybridize the template and support extension. After photocleavage from support, the leader terminates in a 5'-phosphate. **(I)** When used in sequencing, the concentrator and leader domains carry out specific functions, localizing the Xp to the lipid membrane and promoting capture by the pore.

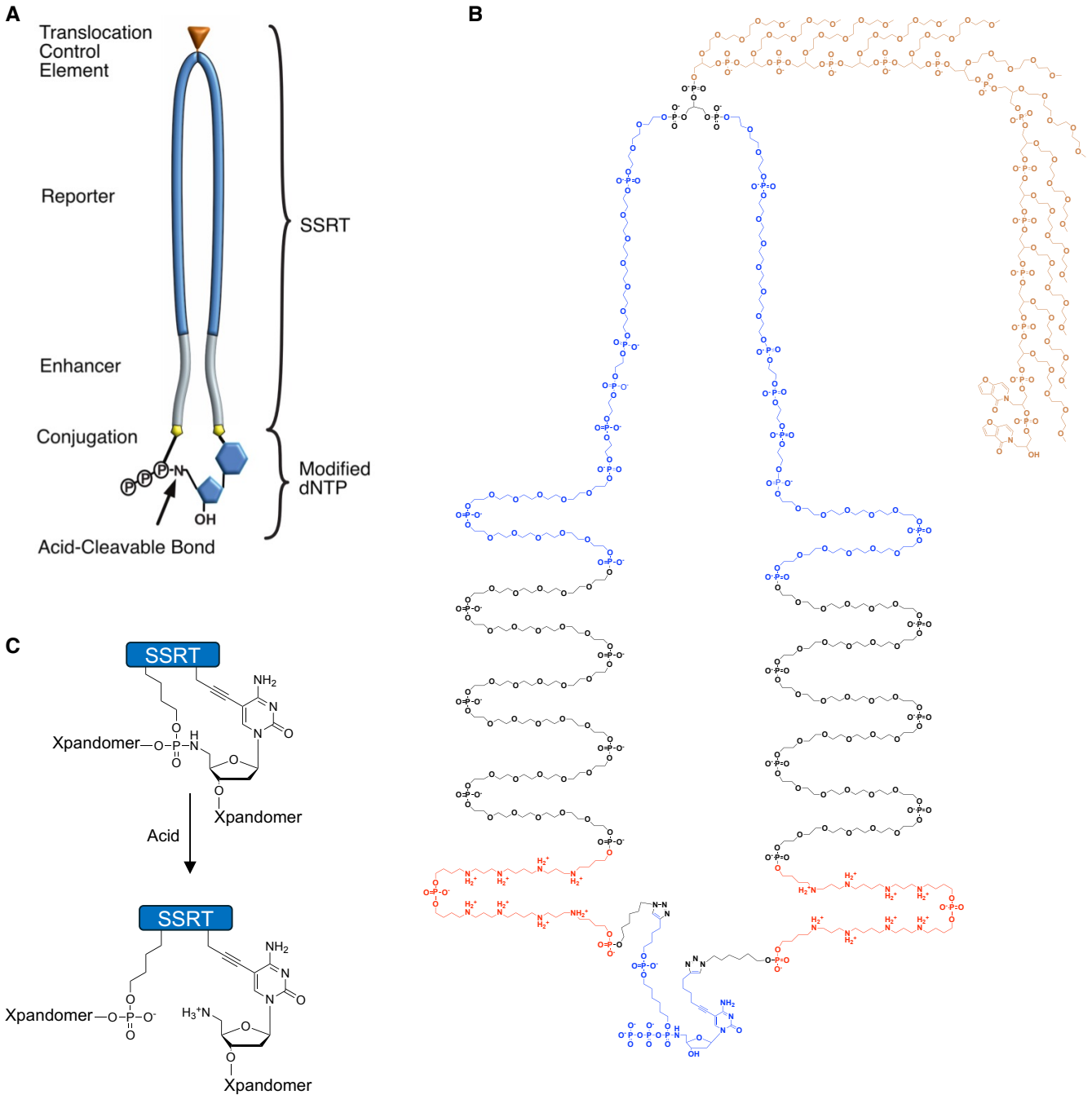


Fig. 3. Expandable nucleotide structure. (A) Functional schematic of the XNTP. The XNTP is synthesized by double conjugation of the SSRT (symmetrically-synthesized reporter tether) and the modified dNTP. Each SSRT contains an identifying reporter that can be read in the nanopore when held in place by the translocation control element (TCE). Each XNTP also contains an enhancer region consisting of polyamines, which is necessary for efficient incorporation by the Xp Synthase. **(B)** Chemdraw image of XCTP. The SSRT containing the enhancer (red), level 1 (blue) and TCE (gold) is clicked to the bis-alkyne dCTP. The synthesis of the SSRT follows modified oligonucleotide synthesis (See Supplement), with many incorporations of specific spacer phosphoramidites. The average XNTP has a molecular weight of 16 kDa, charge of -38 and can be handled and purified much like a 50-mer oligonucleotide. **(C)** When the XNTPs are incorporated into an Xp, the resulting molecule can be expanded with an acid treatment. The P–N bond is broken, reforming the backbone to include the entire length of the tether.

A subsequent acid treatment [Fig. 2F, Fig. 3C] degrades the DNA template and cleaves each incorporated XNTP at the sugar-phosphate linkage, expanding the Xp backbone to include the length of the tethers located between adjacent nucleobases. Each tether contains a reporter segment corresponding to the identity of the subsequent DNA base. The fully formed Xp is released from the solid substrate by photocleavage [Fig. 2G], and the eluted sample is ready for measurement in the nanopore sequencer.

In the sequencing step, the Xp is measured in the nanopore. Here, we use an alpha-hemolysin (αHL) nanopore inserted into a lipid bilayer membrane which fluidically connects the electrolyte of the *cis* and *trans* reservoirs, and functions as a conduit for the Xp [Fig. 4A-B]. A periodic electrical field is imposed between the *cis* and *trans* reservoirs through Ag/AgCl electrodes [Fig. 4D, Fig. S1A-C]. When Xp is added to the *cis* reservoir, the electrical potential drives it to thread the nanopore [Fig. 4E]. Threading of the Xp occurs at the 5′ end due to two capture enhancement features of the EO: the concentrator and the leader [Fig. 2H-I].

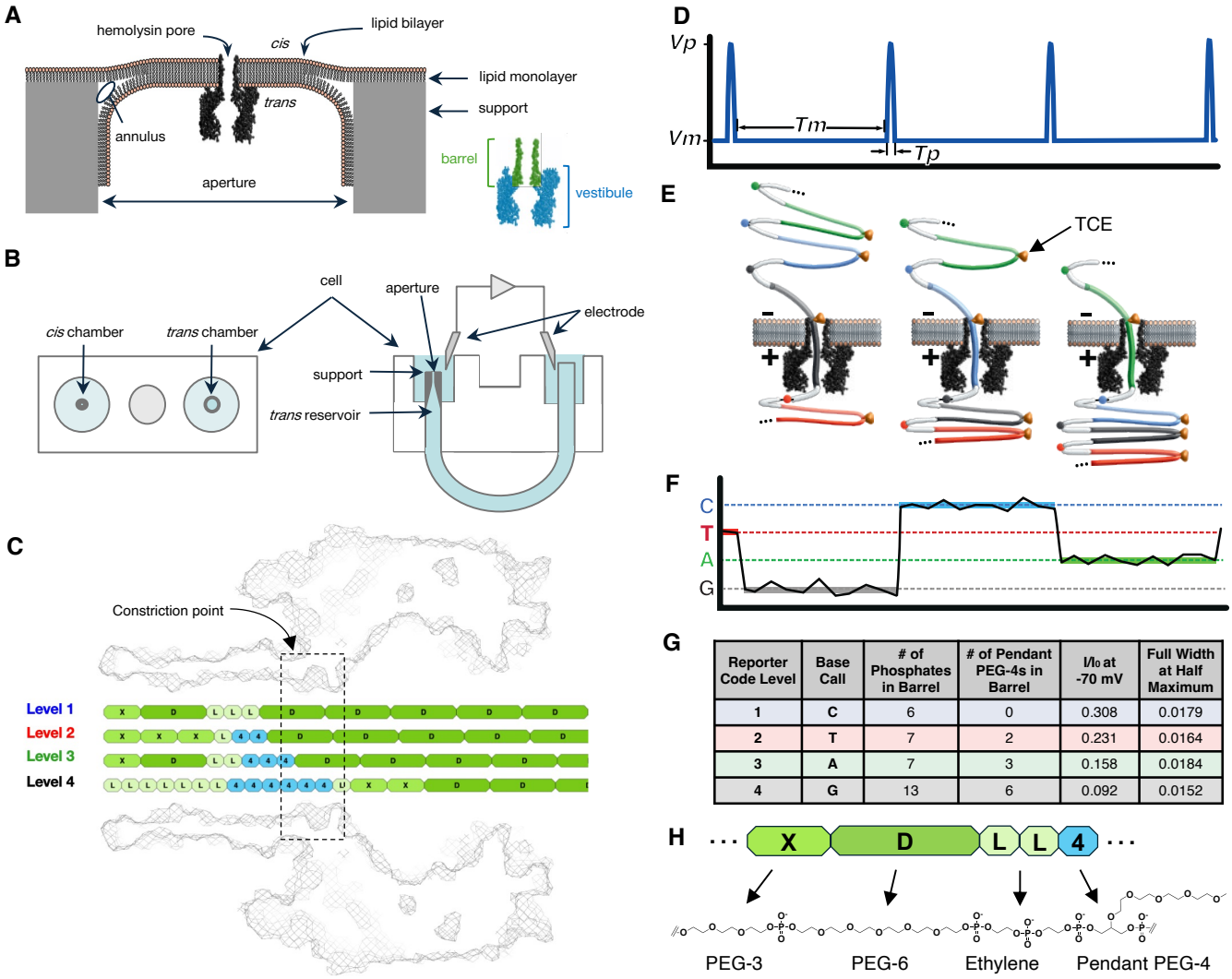
After threading, the Xp continues to translocate under the applied potential until paused by a molecular structure called a Translocation Control Element (TCE) [Fig. 4E]. The TCE is positioned to hold its adjacent base reporter in the 5 nm long barrel of the nanopore while the ion current is measured. Four different reporter types [Fig. 4C] yield four distinct ion current levels that correspond to each XNTP base-type [Fig. 4F-G]. The applied waveform sets the period of measurement (typically 1 ms @ -70 mV) and is followed by a brief, high-voltage release pulse (typically 6 µs @ -600 mV) before repeating [Fig. 4D]. The pulse controls the translocation of the TCE, precisely advancing the Xp to the next TCE and subsequent reporter measurement [Fig. 4E]. This cycle continues until the Xp is fully translocated and the nanopore becomes open again to capture another Xp. The resulting ion current trace from Xp translocation shows a series of periodic steps between the four possible levels that are easily translated into the DNA sequence [Fig. 4F].

Molecular Structures, Chemistries and Methods that Enable SBX

Novel molecules, synthetic routes and enzymes, along with specialized solid-phase synthesis workflows, reagent formulations and sequencing methods were developed for this technology. The molecular design required hundreds of iterations of the XNTP and the EO, a process facilitated by the strategic decision to leverage phosphoramidite chemistries. With in-house capability and access to >200 phosphoramidites, designs could be routinely conceived, synthesized and tested within days. The modular design of molecular systems meant that separate functional parts of these molecules, such as the base reporters and enhancers, could be optimized with relative independence from each other. Similar systems approaches were used in developing PEMs [Fig. S2] and the Xp Synthase. With few exceptions, SBX optimization required the nanopore measurement to resolve molecular level effects [Fig. 5A-C]. Here, these designed elements and workflows underlying SBX chemistry and sequencing are described in detail, followed by a demonstration of the method.

Expandable Nucleoside Triphosphate (XNTP)

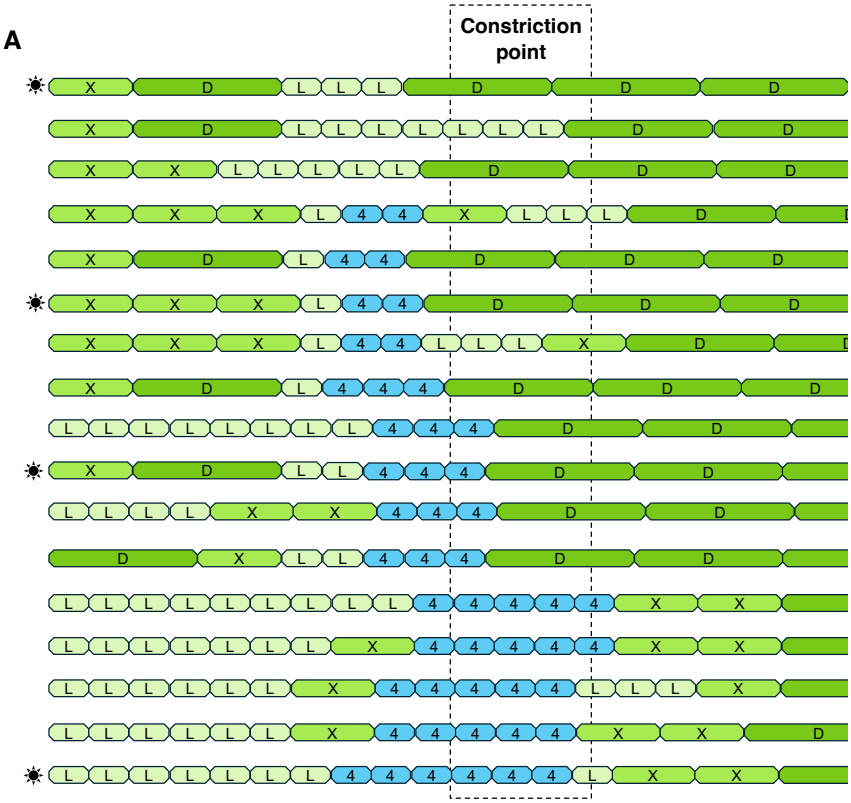
The key building blocks of Xp synthesis are the four XNTPs [Fig. S3-S6], one for each nucleotide. Each of the XNTPs uniquely encodes the base identity of its cognate nucleotide and serves as the substrate for Xp Synthase. The XNTP consists of a modified deoxynucleoside triphosphate (dNTP) that is doubly-conjugated to a symmetrically-synthesized reporter tether (SSRT) to form a cycle using copper-catalyzed azide-alkyne "click" chemistry (*15*). The SSRT is synthesized using phosphoramidite chemistry which enables precise design and rapid optimization of the XNTP's molecular features, including high signal-to-noise reporters and non-stochastic translocation control [Fig. 6A-B]. A glossary of all phosphoramidite structures used is shown in Fig. S7.



Reporter Code Level	Base Call	# of Phosphates in Barrel	# of Pendant PEG-4s in Barrel	I/I₀ at -70 mV	Full Width at Half Maximum
1	C	6	0	0.308	0.0179
2	T	7	2	0.231	0.0164
3	A	7	3	0.158	0.0184
4	G	13	6	0.092	0.0152

Fig 4. Stepped translocation and measurement of Xp in nanopore. (A) Cross-section of the sequencing environment. An assembled heptameric alpha-hemolysin pore is oriented with the barrel facing the *cis* side in a lipid bilayer. The lipid bilayer is formed across the aperture of a PTFE/FEP support. **(B)** Top-down view (left image) and cross-section side view (right image) of the sequencing cell. A PTFE/FEP tube bridges *cis* and *trans* chambers which are each filled with electrolyte solution. A voltage potential is imposed by positioning AgCl electrodes connected to an Axopatch amplifier into the *cis* and *trans* chambers (*42*). After forming the lipid-pore complex, Xp sample is added to the *cis* chamber and driven through the pore via an imposed electric field. **(C)** Base reporters were designed to have four distinct ion current level responses, with similar voltage-pulse translocation control. **(D)** A periodic waveform applied across the nanopore (*cis* to *trans*) consists of a baseline measurement voltage *Vm* and duration *Tm* followed by a short translocation pulse with peak voltage *Vp* and duration *Tp*. Typical times and voltages for the measurement and pulse periods are 1.0 ms / -70 mV and 6.0 μs / -600 mV respectively. **(E)** An Xp in three consecutive measurement states of three consecutive reporters. In each state, the TCE holds the reporter in the nanopore until the pulse advances the TCE through the nanopore to stop at the next TCE along the Xp. **(F)** Depiction of a molecular trace showing three measurement states. **(G)** Characteristics of the reporters producing the best signal levels. The reporter signal values and the FWHM distribution widths for their ion current levels (I) are normalized to the open pore ion current level (I_0), typically ~70 pA @ -70 mV. **(H)** Components of the reporter are units of **D** (PEG-6), **X** (PEG-3), **L** (ethylene diol) and **4** (pendant PEG-4) assembled by phosphoramidite chemistry.

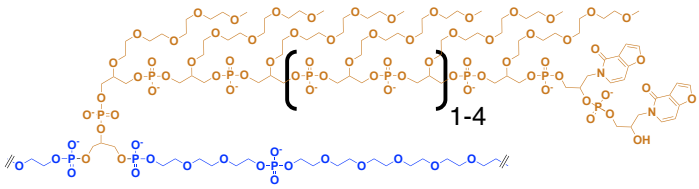
A



Level	# of 4's in TCE	I/I_0 at -60 mV	% Ins	% Del
L1.35	12	0.326	0.06	0.03
L1.37	12	0.324	0.05	0.23
L1.36	12	0.299	0.04	0.29
L2.49	12	0.266	0.04	0.22
L2.44	12	0.262	0.11	0.34
L2.46	12	0.248	0.09	0.23
L2.50	12	0.229	0.04	0.20
L3.48	12	0.191	0.23	0.40
L3.53	12	0.186	0.05	0.40
L3.49	12	0.174	0.27	0.28
L3.52	12	0.172	0.05	0.62
L3.50	12	0.156	0.19	0.27
L4.68	12	0.148	0.11	0.23
L4.69	12	0.132	0.04	0.31
L4.70	12	0.118	0.13	0.30
L4.67	12	0.106	0.14	0.50
L4.66	12	0.100	0.18	0.29

B

Level	# of 4's in TCE	% Ins	% Del	% Sub
L1.35	12	0.06	0.03	0.48
L1.35	10	0.06	0.21	0.38
L1.35	8	0.09	0.62	0.32
L1.35	6	0.05	16.68	0.63



C

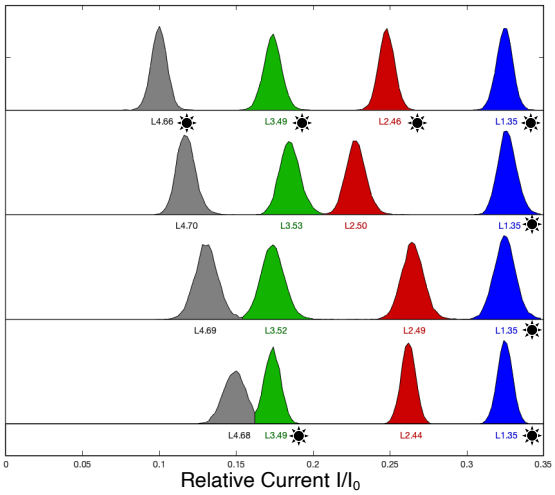
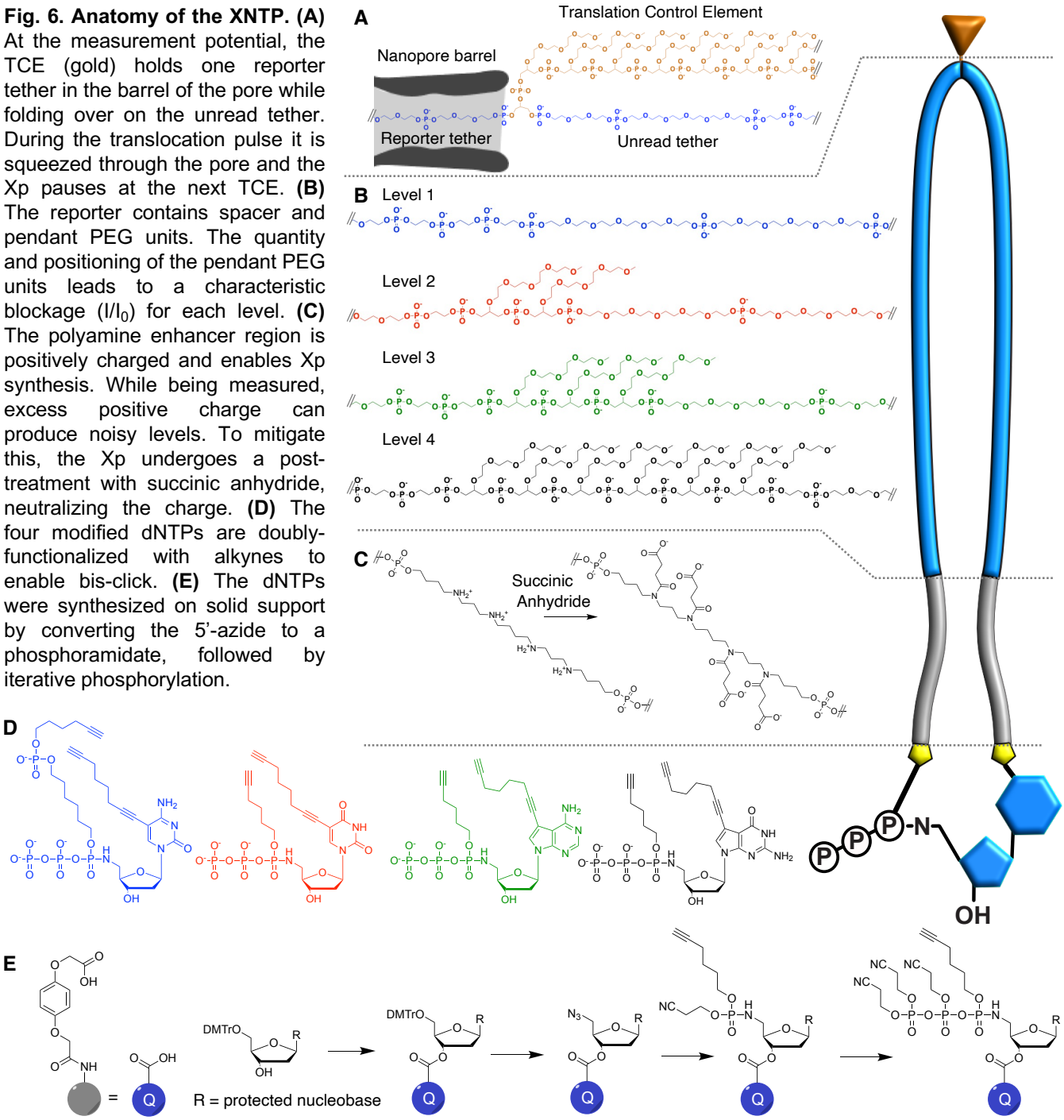


Fig. 5. SSRT composition determines control over levels, translocation and noise. (A) Various reporter code designs with corresponding performance data. When bulky pendant PEG-4 units (light blue) are added to the reporter code, the observed I/I_0 decreases. For a given number of pendant PEG-4 residues, additional tuning of their placement can change the I/I_0. The nanopore constriction point has the most sensitivity for ion current adjustment. Structures used for the sequencing demonstration are highlighted with stars and were chosen for their separation and signal quality. **(B)** Changing the length of the TCE for a given level had little impact on the I/I_0, but a profound impact on the deletion profile. In this example of a Level 1, a TCE of 6 pendant PEG-4 units loses its ability to stall the reporter code in the barrel of the nanopore. For the sequencing data presented herein, all TCEs had 12 pendant PEG units. **(C)** As I/I_0 varies with reporter code design, the levels can become less resolved as the signals begin to overlap, resulting in non-enzymatic substitution errors.

Fig. 6. Anatomy of the XNTP. (A) At the measurement potential, the TCE (gold) holds one reporter tether in the barrel of the pore while folding over on the unread tether. During the translocation pulse it is squeezed through the pore and the Xp pauses at the next TCE. **(B)** The reporter contains spacer and pendant PEG units. The quantity and positioning of the pendant PEG units leads to a characteristic blockage (I/I_0) for each level. **(C)** The polyamine enhancer region is positively charged and enables Xp synthesis. While being measured, excess positive charge can produce noisy levels. To mitigate this, the Xp undergoes a post-treatment with succinic anhydride, neutralizing the charge. **(D)** The four modified dNTPs are doubly-functionalized with alkynes to enable bis-click. **(E)** The dNTPs were synthesized on solid support by converting the 5'-azide to a phosphoramidate, followed by iterative phosphorylation.



Where we could not achieve the desired performance with commercially available products, we synthesized new phosphoramidites [Fig. S8A-C], developing a diverse repertoire of these building blocks to optimize SSRT structure and function. Solubility, length, spacing of charge and positioning along the reporter code were engineered to enhance performance. We synthesized the SSRT as a symmetric molecule by utilizing a branching phosphoramidite [Fig. S9]. This simplifies the final double conjugation of dNTP to the SSRT. XNTP and Xp function is not affected by the orientation of the SSRT-dNTP conjugation due to the symmetry of the SSRT. The XNTP macrocycle products showed an expected lower gel mobility and were easily separated from the linear SSRT by HPLC [Fig. S10A]. Through control of the substrate concentration, SSRT structure, and click reaction conditions we isolated pure cyclized product in excess of 50% yield [Fig. S10C]. The click reaction is known to produce reactive oxygen species that can be detrimental to nucleotides, so XNTP click reactions were performed under argon using minimal copper to avoid these products (*16*). During XNTP isolation, uncyclized and concatemer click products were also removed [Fig. S10B-C]. Any side product other than cyclized XNTP, with a single acid-cleavable bond, could lead to a strand break in the final Xp.

Symmetrically Synthesized Reporter Tether (SSRT)

Translocation Control Element (TCE): The TCE regulates Xp translocation through the nanopore as described above. The structure is composed of a series of pendant polyethylene glycol (PEG) residues (*17*), followed by a symmetric brancher to which the remaining features of the SSRT are constructed through stepwise synthesis. Fig. 6A illustrates our hypothesis of how the TCE pendant PEGs fold back at the nanopore entrance and sterically prevent translocation under the low force of the applied measurement potential (*18*, *19*). In contrast, the TCE will pull through the nanopore barrel under the brief high force delivered by the voltage pulse that follows measurement. This periodic measurement process provides well-controlled, reliable performance, with few incidents of stochastic TCE translocation, which simplifies analysis and allows for highly efficient measurement [Fig. 1E]. Translocation fidelity is important, as failure to advance the Xp by a single reporter per pulse results in either a deletion or insertion sequencing error.

Base Reporter: During Xp nanopore measurement, the TCE precisely positions the reporter in the barrel of the nanopore, impeding the ion flow to one of four current measurement levels associated with each of the four XNTP base types [Fig. 6B]. Base reporters are composed of a series of spacer and pendant PEG phosphoramidites, experimentally optimized to reproducibly generate distinct, well-separated current levels, with minimal noise during measurement. Design considerations included size, charge density, chemical properties and precise positioning in the nanopore. Due to the interdependent function of the reporters and TCE, molecular design optimization occurred simultaneously, to balance both the DC measurement voltage and the release pulse for each reporter [Fig. 5A-C]. Although the SSRT is synthesized symmetrically, only the leading arm occupies the nanopore to generate a base reporter ion current measurement.

Enhancer: This SSRT functional group is essential for efficient Xp synthesis [Fig. 7A] and is composed of polyamine phosphoramidites [Fig. 3B] that provide positive charge density. While it is not fully understood how the enhancer region promotes Xp synthesis, this part of the structure likely interacts with the Xp Synthase, where it could encourage better XNTP binding or reduce the Xp-enzyme dissociation rate. It is also possible that the charge creates organization of the SSRTs in the growing Xp which stabilizes synthesis. The enhancer's positive charge density, while beneficial for Xp synthesis, requires chemical modification prior to Xp sequencing in the nanopore to mitigate measurement interference [Fig. 6C].

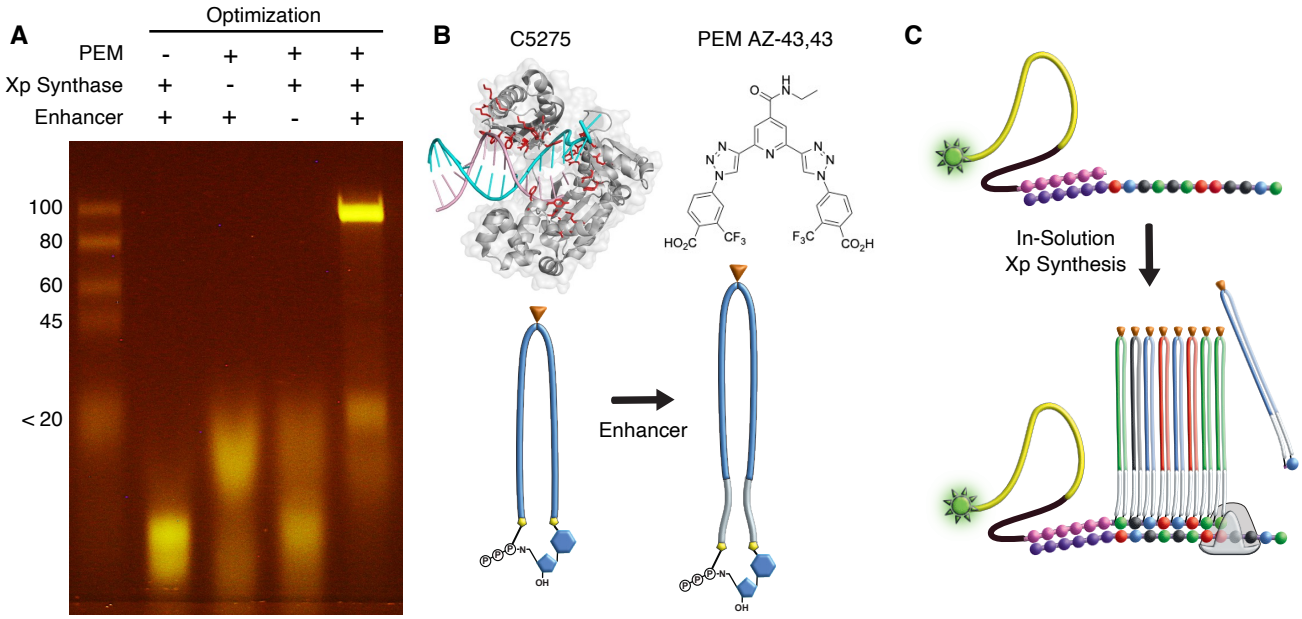


Fig. 7. Critical elements in the synthesis of Xp. (A) PAGE analysis of uncleaved Xp products synthesized from a 100-mer template in solution. Reversions that eliminate any critical element prevent formation of full-length product. **(B)** Reversions in **A** included WT DPO4 replacing Xp Synthase C5275, removal of PEMs and XNTPs missing the enhancer region. **(C)** In-solution synthesis using a 5'-SIMA HEX labeled EO permitted rapid testing of components for optimization of Xp synthesis. In-solution extensions were performed similarly to on-support extensions but were quenched with EDTA and proteinase K. Products from templates shorter than 100-mer can be analyzed by 4-12% polyacrylamide TBE gel, and longer Xp can be analyzed by 2.5% agarose TBE gel.

Azide Terminal Ends: Following the enhancer, the final functional element added to each SSRT arm is a 6-carbon terminal azide linker [Fig. S11] that is used for macrocyclization of the SSRT and modified dNTP via click conjugation (*20*, *21*).

Modified dNTP

Acid-Cleavable Phosphoramidate Bond: The alpha phosphoramidate modification (5′ P–N–C 3') of the dNTP permits template-dependent extension of XNTPs and enables selective cleavage downstream (*22*). After Xp synthesis, the P–N bond is cleaved by treatment with strong acid, which opens the macrocycle to form the expanded Xp backbone to include the length of the SSRTs [Fig. 3C]. Since all other components of the Xp were engineered to survive acid, the expanded molecule provides well-separated, signal-distinct reporters that can be measured independently [Fig. S12A-C].

Alkyne Linker Arms: To enable macrocyclization via a bis-click strategy (*15*, *23–25*) the modified dNTPs have alkyne-functionalized arms [Fig. 6D]. Octadiynyl arms (*26*) were used for the heterocycle linkage (C-5 for pyrimidines, deaza-7 for purines) while hexynyl arms were attached to the alpha phosphoramidate of A/T/G. With this configuration, cleavage of the P–N bond allows the Xp molecule to expand. The structural characteristics of the alkyne arms have a large influence on the incorporation efficiency of the XNTP [Fig. S13A-C]. For example, to increase accuracy and prevent misincorporation of XCTP, a compound arm consisting of hexyl-phosphate-hexyne was used to normalize incorporation efficiency relative to XTTP. Without this adjustment, T to C substitution errors were increased [Fig. S13D].

Synthesis of Modified dNTPs: Synthesis of such a highly modified dNTP requires a specialized synthetic route [Fig. 6E, Fig. S14A]. The electronics of the 5'-aminonucleoside make classical dNTP synthesis routes such as the Yoshikawa (*27*) and Ludwig-Eckstein (*28*) methods extremely low yielding. Other routes employing trimetaphosphate are incapable of placing the linker arm on the alpha phosphate (*29*). Ultimately, we developed a method based off Engel's chemistry, running through the azide intermediate (*15*, *30*). When incubated with the phosphite of an alkyne arm, the resulting Staudinger reduction can undergo a Michaelis-Arbuzov rearrangement (*31*) to produce the functionalized monophosphate. This chemistry was adapted to a solid support (*32*) to continue building the triphosphate through iterative phosphorylation with phosphoramidites (*15*, *33*).

While modified nucleotides have seen significant utility in many applications (*29*), this is the first report of dNTPs containing a 5'-phosphoramidate diester. These modified dNTPs are stable in neutral to alkaline buffers but the phosphoramidate breaks down under acidic conditions. As expected from the geometry of the Dpo4 active site, only one isomer at the alpha phosphoramidate is a substrate for the Xp Synthase. Therefore, the diastereomers were separated by HPLC prior to click conjugation [Fig. S14B].

Extension Oligo (EO)

The EO [Fig. 2H-I] comprises three functional components: the leader, the concentrator and the Xp primer. The EO has additional functional groups on the leader end that support solid phase synthesis, including an azide for click conjugation to an alkyne-functionalized cyclic olefin substrate and a photocleavable linker to release the Xp after synthesis [Fig. 2G, Fig. S15]. The Xp primer is synthesized with acid-stable 2´-O-methyl A, C and G modified bases, and unmodified acid-stable 2'-deoxy T bases to prevent breakdown during the cleavage step. The primer portion of the EO hybridizes to the DNA template and initiates Xp extension from the 3´end.

The leader and concentrator components promote capture of the Xp by the nanopore [Fig. S16A-B]. The leader, on the 5´-terminus of every Xp, is a linear polymer element with high negative charge density designed to maximize capture by the nanopore biased with a positive potential on the trans reservoir. Following the leader is the concentrator, a hydrophobic linear polymer element designed to associate with, and diffuse along, the surface of the lipid bilayer. This acts to confine the leader near the lipid surface, which increases Xp capture rates by greater than 1000x [Fig. 2I, Fig. S16B].

Polymerase Enhancing Moiety (PEM)

In an effort to increase the average length of Xp synthesis to >10 bases, we discovered that some classes of DNA minor groove binders improved Xp extension length dramatically [Fig. S17A-B]. Based on this finding, we developed a novel family of molecules using a bis-alkyne core that could be double-conjugated to a variety of azide arms (*34*). With a systematic synthesis and screening effort, we further discovered higher-performing combinations, using a large matrix of cores and arms [Fig. S2A-B]. These molecules, which are termed Polymerase Enhancing Moieties (PEMs), have a profound effect on Xp Synthase processivity, enabling Xp synthesis lengths that exceed 400 bases [Fig. 7A, Fig. S2E].

Several molecules in our initial screening demonstrated a PEM effect on a 45-mer template. Both the angle of the arms and the presence of the salicylic acid correlated with improved extension [Fig. S2C-D]. We performed a second round of testing using a 100-mer template and varying the concentration of PEM [Fig. S2D]. While the PEM effect did increase in a concentration dependent fashion, when the concentration of the additives reached 1.5 mM there was a noticeable

drop off in Xp extension length. The lead PEM after these screens (D-8,8) contained a 2,6-pyridine core with *p*-salicylic acid arms.

We next targeted select regions of the PEM for structural improvements [Fig. S18]. When the 4 position on the pyridine core was modified to the nitro, nitrile or *N*-ethylamide the PEM effect increased [Fig. S2C]. Additionally, the solubility improved which allowed use of up to 15 mM PEM in the extension reaction. The arm regions were made more acidic by replacement of the salicylate alcohol with the electron-withdrawing trifluoromethyl substituent. This PEM, called AZ-43,43 [Fig. 7B] resulted in full length Xp of a 411-mer template [Fig. S2E], and was used in the SBX demonstration described below. A potential mechanism for the PEM effect is the inherent instability of the DNA-Xp heteroduplex. Without the charged backbone of DNA, the DNA-Xp heteroduplex could fray once it has passed beyond the influence of the Xp Synthase, at about 9-10 bases in length, resulting in slippage and stalling further synthesis. X-ray crystallography or cryogenic electron microscopy studies of the Xp Synthase tertiary complex with PEMs could provide more insights into their function.

Xp Synthase

The size of the SSRT and the uncharged phosphoramidate diester of the modified dNTP create significant barriers to primer extension with XNTPs by most DNA polymerases. Dpo4 (*Sulfolobus solfataricus*), a Y-family DNA polymerase, has an open DNA binding cleft (*35*, *36*) which provides a continuous passage for the large SSRT structures [Fig. 8A]. Wild-type Dpo4 was capable of only 1-2 XNTP incorporations under the original extension conditions with early SSRT designs [Fig. S19A]. Through a mutagenesis process, we evolved an Xp Synthase with 36 amino acid substitutions [Fig. 8B, Table S1] from wild-type (*37*, *38*), which is capable of generating Xp that are hundreds of bases in length under optimized conditions [Fig. 7A-C].

For the initial phase of mutagenesis, we used large libraries in which 1-8 amino acids were targeted simultaneously for saturation mutagenesis to survey the polymerase, with a focus on regions proximal to the catalytic site and nucleic acid interaction surfaces [Fig. S19B]. Next, we used small libraries and designs focused on regions that generated hits in the initial surveys [Fig. S19E]. We evaluated variant polymerases for extension length using XNTP precursor substrates in a high-throughput functional screen [Fig. S19 C-D]. Most of the adopted amino acid substitutions occurred at nucleic acid binding surfaces, and probably reflect adaptations to the significant difference in charge between the DNA and Xp backbones. These modifications improved Xp extension length and yield, while fidelity was primarily affected by mutations near the catalytic site [Fig. 8D, Table S1]. Most base substitutions in SBX sequencing derive from Xp Synthase base selection errors and can be addressed by mutagenesis. For example, the A57S mutation led to a 1% net reduction in G(A) substitutions, and the V42A reversion improved all base substitutions [Fig. 8D, Table S1].

Without a structure of the Xp Synthase, it can be difficult to assign a mechanism for how individual mutations contribute to the observed improvements in SBX extension, accuracy, soluble yield or thermostability, but there are two observable trends of note: 1) Substitution for aromatic side chains is preferred at many positions that interact with the extending Xp strand [Fig. 8C]. 2) Reducing negative charge on the open side of the polymerase improves average extension lengths [Fig. 8E]. The phosphoramidate diester backbone of the Xp molecule is uncharged, so the preference for aromatic side chains on the primer interacting surfaces may indicate a new way of engaging with that strand. The reduction in negative charge on the open side of the polymerase was engineered to prevent undesired electrostatic interactions between the polyamine enhancer moieties on the SSRTs and the polymerase surface. The mechanistic basis of how the remaining mutations improve polymerase performance will be a topic of future study.

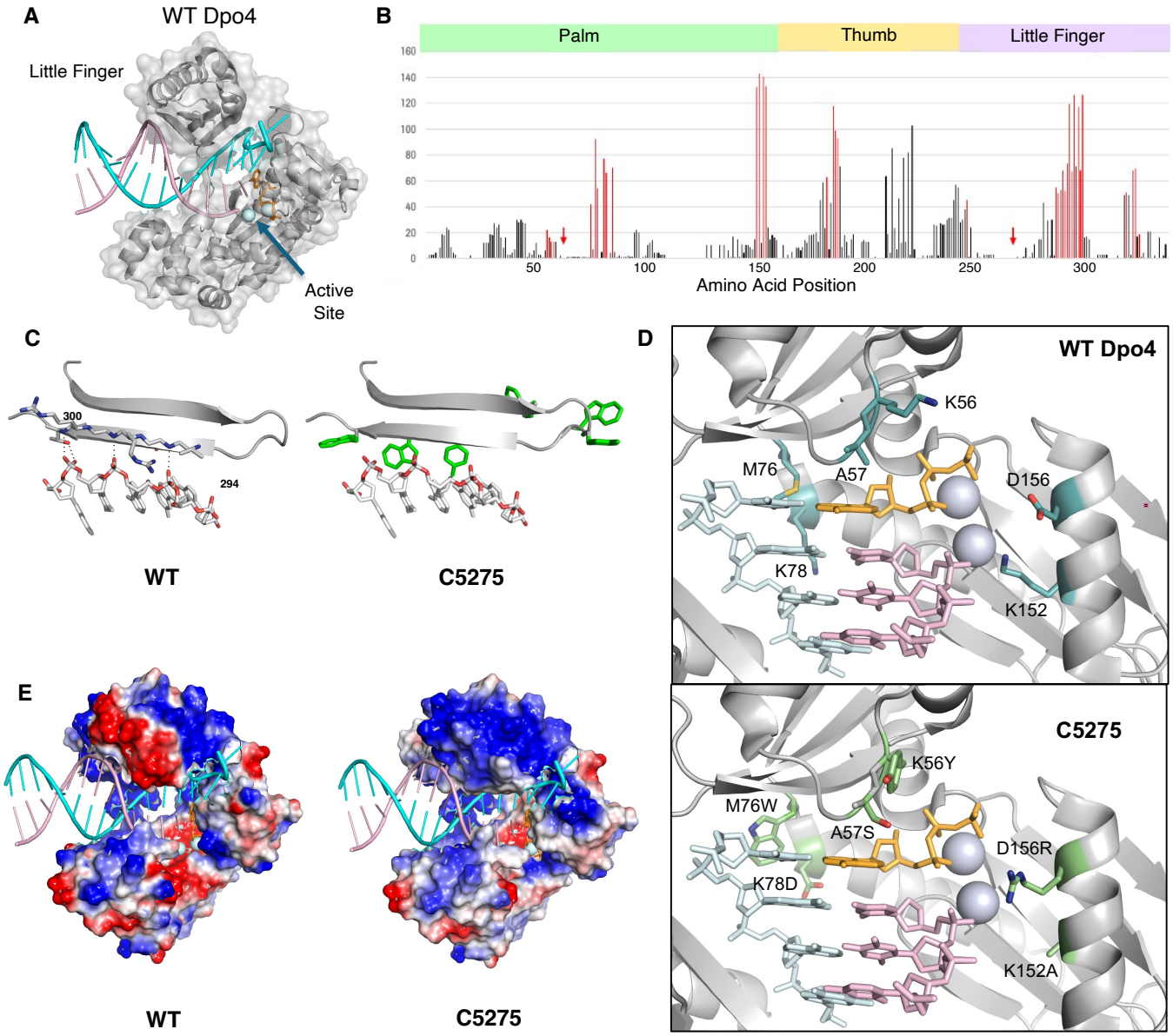


Fig. 8. Evolution of Dpo4 polymerase into an Xp Synthase. (A) Dpo4 was selected as the parent DNA polymerase because of its open DNA-binding cleft that provides continuous passage for the XNTP reporter tethers, about 9-10 XNTP incorporations. PDB accession number 2AGQ with template DNA (cyan) and product strand (rose) was used for all structures. **(B)** Histogram of the number of times an amino acid position was targeted by library mutagenesis. The count includes the number of times a position appeared in a unique library, and the number of times that library was screened. Red indicates positions where mutations were adopted. The red arrows indicate positions 63 and 272, which are mutated in the Xp Synthase but were not targeted by library screening. **(C)** Little finger primer rail positions 293-301 showing WT Dpo4 positions and Xp Synthase mutations modeled in PyMol using lowest strain rotomer orientation. There are key hydrogen bonds (dotted lines) between the DNA backbone and peptide backbone at positions 295, 297, and 301 in the WT structure. The Xp Synthase has six aromatic positions added in this region (V289W, D292Y, L293W, S297Y, R299W, T301W) (green). **(D)** Dpo4 active site comparison showing key mutations that reshape the XNTP binding site. **(E)** Changes in surface charge of Dpo4 from WT to C5275 generated using the PyMol vacuum electrostatics tool. Red indicates higher average negative charge and blue indicates higher average positive charge. Mutations on the open face of the Xp Synthase that change the surface charge: K56Y, E63R, K152A, D156R, E291S, D292Y, D294N, R300S, K321Q, E324K, E325K, E327K.

Xp Synthesis Considerations

We optimized the extension master mix to balance yield and accuracy of Xp synthesis and to reduce template secondary structure. Adjustments to concentrations of PEMs, XNTPs, Mn^{2+}, PEG, solvent and salts as well as buffer pH resulted in accuracy, extension length and yield outcomes that are consistent with the high interactivity of these reaction components. In particular, accounting for the broad interactivity of divalent Mn^{2+} was critical to balancing the reaction. Mn^{2+} is an essential cofactor for catalyzing the template-directed addition of XNTPs to the primer strand, and it cannot be substituted by Mg^{2+} without a significant loss of yield [Fig. S20A, D]. As in PCR, Mn^{2+} contributes to base substitution errors at elevated concentrations, but also enhances the catalytic efficiency of the enzyme [Fig. S20F] (*39*) (*40*) so understanding what factors can alter the effective Mn^{2+} concentration is key to optimizing the reaction. Oxidation of Mn^{2+} during Xp synthesis, which reduces the concentration of the divalent form, leads to slower reaction kinetics and reduced full length yield [Fig. S20B], requiring consideration of pH (*41*), reaction byproducts and reagent purity. It also led us to prepare the master mix immediately before use, with $MnCl_2$ added last. PEMs and PEG both influence the effective concentration of Mn^{2+}. PEMs appear to weakly chelate Mn^{2+}, while the crowding effect of PEG may make Mn^{2+} more available [Fig. S20G, H]. Thus, the development of the SBX extension reaction often revolves around balancing desired primary effects of reaction components with how they might alter the effective Mn^{2+} concentration.

The solvent environment is also important to reaction formulation. Molecular crowding by inclusion of PEG 8000 at 20% wt./vol. was essential to achieve the reaction kinetics observed for Xp synthesis. Decreasing concentration to 16% wt./vol., for example, dramatically reduced Xp yield [Fig. S20C]. The co-solvent N-Methyl-2-pyrrolidone (NMP) improves accuracy at the expense of extension length [Fig. S20I]. The amount of solvent also affects the solubility of PEMs, with increased concentrations of solvent enabling the use of higher concentrations of PEMs [Fig. S20E]. Given that PEMs can affect accuracy as well as length, in a concentration-dependent manner [Fig. S20H], the interplay between solvent and PEMs is critical for balancing the reaction conditions.

Sequencing Considerations

The choice of materials, design and protocols evolved to improve the quality and throughput of Xp sequencing in the nanopore. Using 1,2-di-O-phytanoyl-*sn*-glycero-3-phosphoethanolamine (DPhPE) lipid in the nanopore bilayer resulted in a >6-fold increase in Xp throughput over diphytanoylphosphatidyl choline (DPhPC), which could be attributed to better concentrator-lipid interaction [Fig. S16C]. We also evaluated the αHL orientation in the bilayer [Fig. S21A-D]. The barrel-to-vestibule orientation demonstrated higher throughput and much lower measurement noise relative to sequencing in the vestibule-to-barrel orientation. An essential feature of Xp sequencing is the inclusion of a negatively-charged leader and a hydrophobic concentrator on the EO [Fig. S16]. We extensively optimized both features for length and composition. The concentration and choice of sequencing salts also had a significant impact on signal quality and Xp measurement behavior. 1 M KCl is traditionally used for nanopore measurement (*42*), however, NH_4Cl demonstrated the highest signal-to-noise and best overall sequence quality of all the salts evaluated. LiCl in particular had poor ion current signal, possibly due to lower relative ion mobility, and also negatively affected Xp translocation, demonstrating slower, less consistent translocation when compared to the NH_4^+ salt. Chaotropic salts, like GuHCl or Urea, were found to improve Xp yield and length when added to the *cis* reservoir at a concentration of 1 M.

SBX Sequencing of 222 Base DNA Template

To demonstrate the single-molecule signal resolution of SBX, we combined the chemistries and methods above to synthesize and sequence an Xp derived from a 222-base *Streptococcus* DNA template. Additional supporting details for the protocol and reagent mixes are provided in the Supplement.

To generate the highest quality sequencing data, we performed solid-phase Xp synthesis, cleavage and modification steps using an off-the-shelf cyclic olefin chip (Chipshop, Fluidic 243) [Fig. 2A-B] with a functionalized flow channel that allowed for tight control of reaction timing, temperature and introduction of reagents. This approach facilitates removal of reaction components that would affect the stability of the lipid bilayer and overall sequence quality, such as solvents and detergents. All liquid handling steps were controlled by manual pipetting, and utilized 500 mm^2 of the flow channel surface area (~278 mm channel length) for synthesis. The flow channel was functionalized by propargyl-PEG4-maleimide addition, followed by irradiation (*43*). The template was selected for its sequence composition, which includes hairpins, homopolymer regions and dinucleotide repeats [Fig. 9C-D]. To enable efficient hybridization of the template to the EO, we strand-enriched the DNA template using established biotin-streptavidin methods (*44*). After flushing the click reagents and excess EO from the chip, we hybridized the template over 5 minutes on a temperature ramp from 72 C to 38 C.

We initiated Xp synthesis by the addition of SBX Extension Master Mix (see Supplement) and incubated the reaction for 1 h at 37 C, producing a heteroduplex Xp-DNA product that remained anchored on the flow channel. Extension was followed by acid cleavage of the P–N bonds on the Xp strand, allowing the Xp to expand into its sequence-ready structure while minimizing ancillary acid damage. After acid neutralization, the SSRT enhancer moieties were modified by succinic anhydride (SA), which partially neutralizes the positive charge of the polyamine enhancers that flank the nucleotide [Fig. 6C]. Finally, we photocleaved the Xp products from solid support and diluted them for sequencing.

For single pore Xp sequencing (and the associated nanopore instrument), we adapted prior approaches [Fig. S22] (*42*, *45*). Briefly, we inserted a single wild-type αHL nanopore in a DPhPE lipid bilayer, oriented with the barrel-side facing the *cis* reservoir using Insertion/Trans Buffer in both reservoirs. We qualified the correct pore orientation by conductance analysis and then replaced the buffer in the *cis* reservoir by perfusion with 3 mL of Sequencing Buffer, leaving 80 µL in the reservoir. Salt content in the *cis* and *trans* reservoirs during measurement were 0.4 M NH_4Cl, 0.6 M GuHCl and 2 M NH_4Cl respectively, all buffered to pH 7.4. The sample was sequenced for 10 minutes at 20 C.

We captured ion current measurements and applied potentials with patch-clamp instrumentation using a digital interface and a LabVIEW program for control [Fig. S22] (*46*). The potential applied to the *cis* reservoir was -70 mV DC baseline with -600 mV pulses (6 µs duration) applied at 1 kHz [Fig. 4D]. Amplifier voltage and current channel outputs were digitally recorded with 16-bit sampling from separate outputs of the Axopatch amplifier at 100 kSps and captured in a raw data file. The current channel was filtered by the amplifier at 10 kHz.

A MATLAB script was used to parse and translate the basecall in the raw data file from the 10-minute sequencing run [Fig. S23A], beginning with removal of the high voltage pulses from the trace [Fig. S23B-C]. Next, we computationally demarcated ion current ranges for each reporter using the four distinct peaks that emerged from analyzing the ion current data from the aggregate translocation events [Fig. S23D]. Each 1.0 ms reporter measurement was assigned a base identity according to the range in which its ion current fell. An example ion current trace of the full length 222 base *Streptococcus* DNA sequence is shown in Fig. 1E, S24 and the complete 10-minute run trace is in Fig. 10A-C.

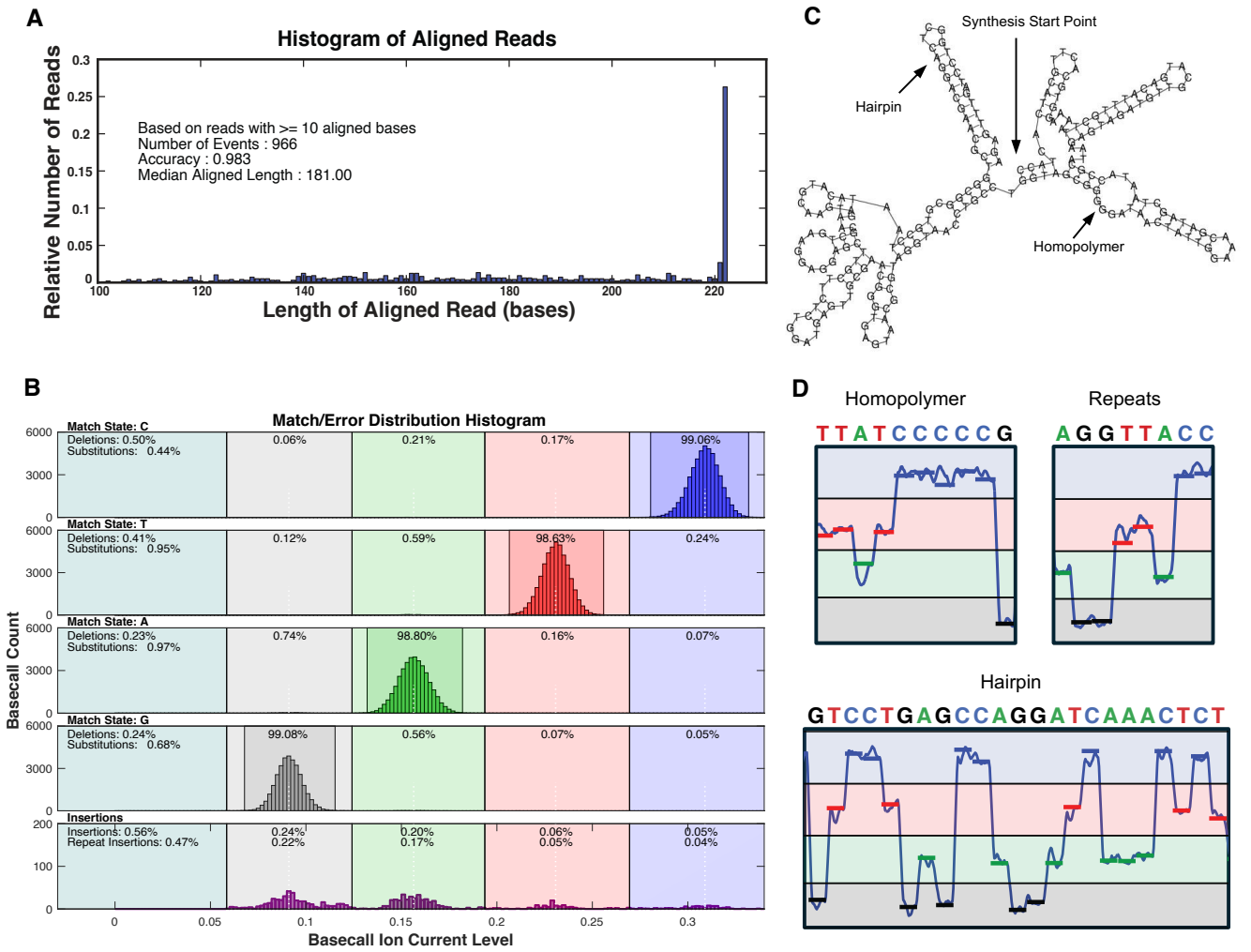
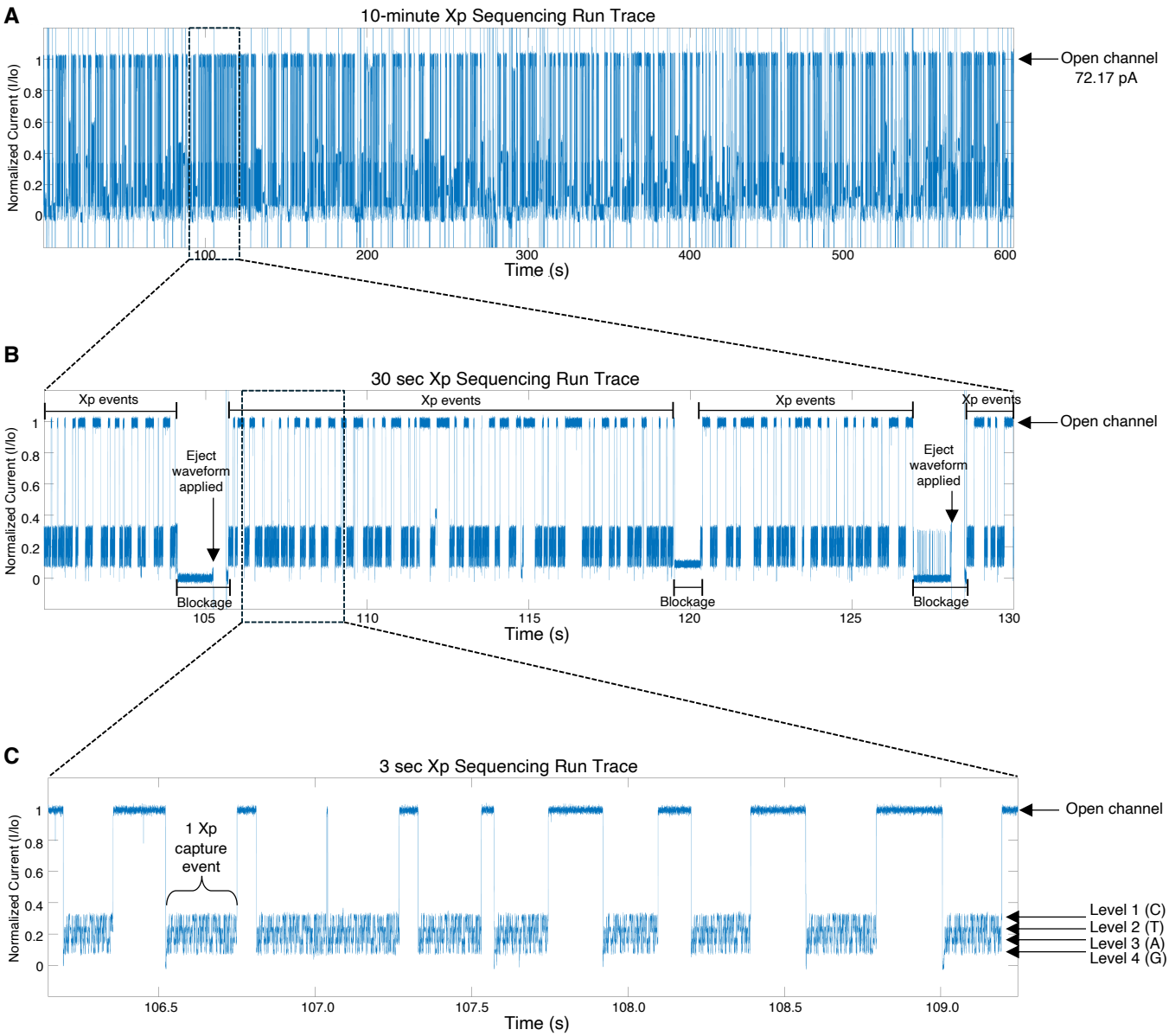


Fig. 9. Xp measurements in a nanopore from 10-minute run. Ion current measurements, captured during each 1.0 ms period that a reporter is in the nanopore, were normalized to the open channel ion current and reduced to produce the basecall ion current level. Basecalls were assigned to these levels according to the ion current range they fall into. The basecalled reads were then aligned to the 222 base DNA template using the Smith-Waterman algorithm. **(A)** The distribution of 966 aligned Xp read lengths longer than 10 bases showed 26% of the reads are full length and have a median aligned read length of 181 bases. **(B)** These histograms present the basecall matches and error distributions of the 966 read set. Basecall ion current level is plotted along the abscissa with basecall call bin counts shown on the ordinate using bin widths of 0.002. The vertical demarcation lines define the ion current ranges, that are used to assign basecalls for Cytosine, Thymine, Adenine, and Guanine ion current levels (in descending value order). The upper four histograms show sequence match and substitution distributions for each of these. Count distributions of matched basecalls for each basetype appear in the range containing the dominant peaks. Base substitutions appear in the three adjacent basecall ranges. Deletions are summarized with total base substitutions in the leftmost column. For each basetype, the percentage matches, substitutions and deletions are normalized to the total base count of that basetype found in the template portions of the aligned reads. As an example, in the top histogram for the cytosine basetype, 99.06% were matched, 0.44% were substitutions (G: 0.06%, A: 0.21%, T: 0.17%) and the remaining 0.50% were deletions. To illustrate the well differentiated signal-to-noise quality of this data, the four subranges shown with darker contrast bound 99.7% of the matches for their respective basetype. The insertion distribution (rescaled by 30x to visualize) is located in the bottom histogram. Insertion percentages (example: C insertions = 0.05%) are normalized to the total number of bases in the template portions of the aligned reads. Listed beneath the insertion percentages are the repeat insertions, a subset of insertions that have an adjacent base of the same basetype as discussed in Sequencing Fidelity. **(C)** Secondary structure of the 222-mer template, generated by the DNA Secondary Structure prediction tool on vectorbuilder.com. **(D)** Magnified regions of ion current traces demonstrating SBX sequencing of homopolymer, repeat, and hairpin regions.



D

% Open Channel	25.7%
% Xp Capture	28.3%
% Blockage	45.9%

E

Overall Throughput (bases / minute)	16638
Throughput with blockage time excluded (bases / minute)	30754

Fig. 10. Run overview and current trace data. **(A)** Example of a full Xp sequencing run. Current trace represented as normalized current (I/I_0). **(B)** The 30 s window shows clean Xp translocation events, occasional blockages and a few short-lived dysfunctional states. An eject waveform application was often required for removing blockages from the pore to resume an active capturing state. The eject waveform reversed the polarity of the voltage to push out material that was blocking the pore. **(C)** The 3 s window showcases the high capture efficiency of Xp molecules. **(D)** Relative time spent in various nanopore states. **(E)** Blockage states dominate nanopore measurement time, offering an opportunity to improve throughput.

The read rate of 1 kbase/s during an Xp translocation corresponded to the 1 kHz pulse rate. During the run, the nanopore was translocating Xp 28% of the time, and was waiting for Xp capture (open channel) 26% of the time [Fig. 10D]. The nanopore was in a blocked state 46% of the time (207 blocking events), which limited the possible throughput and yield of the run [Fig. 10E]. This occurred when the nanopore failed to return to the open channel condition. Blockages were resolved by temporarily reversing the voltage across the nanopore to eject a stuck Xp molecule from the nanopore before proceeding with the run. This 10-minute run recorded the measurement of 166,383 base reporters, for an average throughput of 280 bases/s.

We assessed the errors by aligning the reads with the DNA target using a Smith-Waterman algorithm (*47*), which provided aligned read lengths as well as insertion, deletion and substitution error counts. The run recorded 966 reads equal to or longer than 10 base calls that aligned to the target DNA sequence. Fig. 9A shows a histogram of this aligned read length distribution. Of these, 26% aligned to the full 222 base template length. Median read length was 181 bases. Smith-Waterman alignment trimmed an average of two unaligned bases from each end of the reads (see Supplement).

Mean accuracy of all reads longer than 10 bases in the run was 98.3% with errors arising from insertions, deletions and substitutions at 0.56%, 0.36% and 0.76%, respectively. Base call and error profile histograms are shown in Fig. 9B. Accuracy is defined here by normalizing the alignment errors with the number of template bases to which reads were aligned.

Discussion

During the development of SBX, important observations and considerations were made that warrant further discussion. Foremost, the interdependence among the novel molecules in the context of the synthesis and measurement environments made SBX development a co-evolutionary process. For example, the Xp Synthase, PEM and Enhancer are all critical to produce high yields of full-length Xp, and removal of any of these three components results in synthesis failure [Fig. 7A-B]. Second, the modularity of the molecular components enabled a systems engineering approach, particularly for the PEMs, SSRT and EO, that accelerated understanding of these components and the technology in whole.

Sequencing Fidelity

SBX sequencing errors occur during both Xp synthesis and nanopore measurement. Errors in Xp synthesis are influenced by many aspects of the reaction. Reaction temperature, order of addition, salt concentration, effective Mn^{2+} concentration, PEM composition, solvents and other reaction additives can all modulate Xp Synthase accuracy [Fig. S20F-I]. The reaction mixture that offers the highest synthesis accuracy likely balances the optimal reaction kinetics with control over solvation near the nascent base-pair, encouraging Watson-Crick geometry over the alternate configurations that are permitted by the combination of Mn^{2+} as the divalent cofactor and the solvent-exposed catalytic site of Dpo4 (*35*, *39*). Xp Synthase mutagenesis and PEMs design have led to significant improvements, and they continue to be a development focus. While the Xp Synthase errors are primarily base substitutions, indel errors also occur. The parent polymerase, Dpo4, is known to introduce deletions via a template slippage mechanism, so it is anticipated that the enzyme will contribute to the deletion rate (*48*).

Sequencing errors during nanopore measurement have multiple sources. The TCE reliably translocates the Xp by a single reporter during each pulse, but some errors do occur. In this experiment, a majority of insertions (0.47% of the total 0.56%) were repeats of their previous basecall. This insertion repeat is characteristic of a TCE insertion error. Experiments exploring pulse optimization show how small pulse voltage adjustments from optimum will drive the deletion and insertion rates in opposite directions. Small pulse force differences between the four base reporters exhibit distinctive deletion-insertion tradeoffs. For example, C and T have the highest deletions [Fig. 9B], but exhibit the lowest insertions in contrast with those of A and G. In an ongoing optimization effort, a matrix of TCE designs with associated reporters have been tested by nanopore sequencing using different reagent and pulse conditions to minimize these errors.

Substitution errors due to a reporter ion current level falling outside its correct range have been reduced through reporter design optimization by altering their size, composition, charge and position [Fig. 5A]. By avoiding the broad or bistable ion current responses exhibited by many TCE-reporter combinations and selecting for a balanced separation across the dynamic range of the ion current measurement, the selected optimized set of reporters exhibits robust, high signal-to-noise responses [Fig. 1G].

The polyamine portion of the enhancer region imparts positive charge to both arms of the tether. This localized charge density affects smooth translocation of the Xp that may cause measurement errors. To mitigate this impact, the more reactive amines of the enhancer region are modified with SA after Xp synthesis to neutralize some of the positive charge [Fig. 6C]. This modification chemistry is tunable, and control over the degree of succinylation can be used to optimize measurement accuracy. Insufficient SA modification results in the polyamine portion retaining too much positive charge, leading to an increase of insertion errors or, in some cases, erratic bistable states and blockages. Over-modification of the amines with SA causes overly fast translocation that results in an increase of deletion errors.

Other sources of noisy Xp nanopore measurement have been identified and addressed during development with minor error contributions relative to those discussed above. Noise and error contributions from steric blockages, pore gating (*49*), environmental noise and co-translocation of multiple Xp have been minimized by effective identification and filtering. Blockages are a common occurrence which are resolved during measurement by briefly reversing voltage to back the Xp out of the pore [Fig. 10B]. Examples of these blocked state ion current traces are shown in Fig. S25A-E. Barrel-first entry of the hemolysin nanopore avoids the congestion of the Xp in the nanopore vestibule that causes erratic deep blockage of the ion current [Fig. S21C].

Near Real-Time Sequencing

A key advantage of SBX is the potential for massively parallel measurements with high accuracy and speed. Such highly parallelized measurements could produce independent streams of data with events that are highly localized in time. Under such a scenario, sufficiently fast computational hardware could execute inference routines at low latency, allowing each read to be called from its corresponding time series data, which can then be further processed shortly thereafter. The short duration and independence of Xp translocation events could allow for such processing to occur even while data acquisition continues over the course of a run. With such immediate processing after measurement, near real-time results can be generated to provide the user with flexibility to end the run when sequence requirements are satisfied. Additionally, run times can be adjusted to match output needs, enabling, for example, small batch multiplexing. By contrast, sequencing-by-synthesis (SBS) based technologies employ long standard cycle times that run to completion before information is available to call a complete read. Access to early result trends and data quality assessment are ancillary user benefits that increase efficiency and flexibility.

Scalability

The high signal-to-noise performance of SBX reduces the need for sophisticated amplifiers and shrinks the footprint requirement of nanopore cells for CMOS-based nanopore array design (*50*). This anticipates scaling to array sizes of 8 million nanopores (*51*, *52*). In these designs, CMOS amplifier arrays are post-processed to add an electrode in a small volume cell above each array amplifier. Prior to operation, the cells can be filled with electrolyte and coated with a lipid bilayer that is a barrier to ion transport. Next, a biological nanopore can be inserted into each cell's bilayer. All of the array cells are fluidically connected through their respective nanopores to a shared flow channel reservoir containing a shared counter-electrode. During operation, the CMOS is designed to rapidly and periodically sample the nanopore impedance of each cell. Leveraging efficient inference routines implemented on accelerated compute hardware, the raw signals produced by Xp measurement events can be translated into base calls (primary analysis). The resulting reads may then immediately be routed for post-primary processing. By adapting Xp measurement to these arrays, it is reasonable to consider that the single pore Xp sequencing rate of 280 bases/s demonstrated here can be scaled to a throughput exceeding 500 million bases/s.

Comparison to On-Market Technology

A natural comparator for SBX would be ONT. It is challenging to find a study to compare this SBX experiment with state-of-the-art ONT sequencing given the scope and the data filtering that is applied in most ONT studies. Reports of average 'raw' accuracies cover a broad range (*8*, *53*, *54*) that depend heavily on the trained basecaller version (e.g. Guppy fast, high accurate (hac) and super accurate (sup) versions) with little detail of the quality cutoffs used. A recent study evaluated the MinION device using the R10.4 flow cell and SQK-LSK112 sequencing kit to sequence HCC78 genomic DNA (*8*). It compared the hac and sup basecallers aligned to the source genome using data filtering that included eliminating reads <200 bases, applying minimal MapQ score 60, and removal of indels with length <100 bases. The reported observed accuracies were 95.6% and 96.8% for the hac and sup basecallers, respectively. By comparison, this SBX experiment used a simple four-range classifier to make millisecond synchronized reporter basecalls followed by alignment and elimination of reads <10 bases to yield an average accuracy of 98.3%. This accuracy is even more significant given the high measurement throughput. In the same study, the reported MinION yield was 12.85 Gb in a 72-hour run or about 50,000 bases/s. The example of SBX scaled to an 8 M pore array would reasonably be >10,000 times that throughput and would also exceed the yields observed for state-of-the-art sequencing by synthesis (SBS) technologies (*55*).

This improved performance, simplicity of measurement and computational reduction clearly demonstrates the signal-to-noise benefits of SBX. The data presented here were captured in January of 2020 using an early chemistry that demonstrates all the critical functionality of the SBX technology and is a prelude to newer chemistry now being applied to large CMOS arrays (*45*, *52*, *56*).

Conclusions

Specialized molecular constructs, chemistries and methods developed for SBX were utilized to demonstrate this novel single-molecule sequencing technology. This approach solves the signal-to-noise challenges of direct sequencing by converting DNA into an expanded, more measurable structure. Key innovations were required to make this technical achievement possible including the development of:

- XNTPs, nucleoside triphosphates modified with structures engineered to control translocation, expand, provide well-resolved nanopore based reporter codes and enable template-dependent incorporation with a modified polymerase (Xp Synthase).
- PEMs, a class of critical cofactors that enable polymerase extension of XNTPs beyond 10 bases by stabilizing the elongating heteroduplex structure.
- Xp Synthases, highly mutated DNA polymerases that processively and accurately incorporate XNTPs to synthesize Xpandomers.
- Reaction formulations and processes, optimized for accurate, template-dependent synthesis of Xp using a solid-phase workflow.
- Xp end structures and sequencing conditions, engineered to efficiently and accurately thread Xp molecules though a nanopore.

The single-molecule results shown provide an opportunity to contemplate a scaled implementation of SBX. As introduced here, single-nanopore SBX was demonstrated with accurate, well-resolved reporter sequencing that can be adapted to high-throughput parallel CMOS-based architectures. Based on Xp sequencing rates of 280 bases/s shown for single pores, throughputs of greater than 500 million bases per second can be reasonably considered. Furthermore, separated Xp synthesis and sequencing workflows allow each process to be independently optimized and efficiently performed. In contrast to current SBS methods (*57*), near real-time SBX data output at high-throughput would allow for flexible operation across a range of experimental scales. Such flexibility would reduce batch requirements, enable faster time to result and ultimately allow users to rethink how sequencing workflows can be implemented for existing and emerging applications.

References

1. C. A. Austin-Tse, V. Jobanputra, D. L. Perry, D. Bick, R. J. Taft, E. Venner, R. A. Gibbs, T. Young, S. Barnett, J. W. Belmont, N. Boczek, S. Chowdhury, K. A. Ellsworth, S. Guha, S. Kulkarni, C. Marcou, L. Meng, D. R. Murdock, A. U. Rehman, E. Spiteri, A. Thomas-Wilson, H. M. Kearney, H. L. Rehm, Best practices for the interpretation and reporting of clinical whole genome sequencing. *Npj Genomic Med.* **7**, 1–13 (2022).

2. J. A. Schloss, How to get genomes at one ten-thousandth the cost. *Nat. Biotechnol.* **26**, 1113–1115 (2008).

3. S. M. Bezrukov, I. Vodyanoy, V. A. Parsegian, Counting polymers moving through a single ion channel. *Nature* **370**, 279–281 (1994).

4. D. Deamer, M. Akeson, D. Branton, Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).

5. Y. Wang, J. Liu, Z. Wang, M. Zhang, Y. Zhang, The Research Progress of Single-Molecule Sequencing and Its Significance in Nucleic Acid Metrology. *Biosensors* **15**, 4 (2024).

6. M. Wanunu, Nanopores: A journey towards DNA sequencing. *Phys. Life Rev.* **9**, 125–158 (2012).

7. Y. Wang, Y. Zhao, A. Bollas, Y. Wang, K. F. Au, Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).

8. Y. Ni, X. Liu, Z. M. Simeneh, M. Yang, R. Li, Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Comput. Struct. Biotechnol. J.* **21**, 2352–2364 (2023).

9. C. Delahaye, J. Nicolas, Sequencing DNA with nanopores: Troubles and biases. *PLOS ONE* **16**, e0257521 (2021).

10. S. L. Amarasinghe, S. Su, X. Dong, L. Zappia, M. E. Ritchie, Q. Gouil, Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 (2020).

11. A. Payne, N. Holmes, V. Rakyan, M. Loose, Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files. Genomics [Preprint] (2018). https://doi.org/10.1101/312256.

12. Northwestern Medicine CGM Pricing, *NuSeq Core Pricing* (2025). https://www.cgm.northwestern.edu/cores/nuseq/pricing.html.

13. L. Froenicke, UC Rate -- UC Davis and other UC campuses | DNA Technologies Core (2019). https://dnatech.ucdavis.edu/uc-rate-uc-davis-and-other-uc-campuses.

14. Mark Stamatios Kokoris, Robert McRuer, High Throughput Nucleic Acid Sequencing by Expansion (2011).

15. Mark Stamatios Kokoris, John Tabone, Melud Nabavi, Aaron Jacobs, Phosphoroamidate esters, and use and synthesis thereof (2019).

16. G. R. Abel, Z. A. Calabrese, J. Ayco, J. E. Hein, T. Ye, Measuring and Suppressing the Oxidative Damage to DNA During Cu(I)-Catalyzed Azide-Alkyne Cycloaddition. *Bioconjug. Chem.* **27**, 698–704 (2016).

17. Dylan O'Connell, Aaron Jacobs, Jessica Vellucci, Brent Banasik, Matthew Lopez, Drew Goodman, Melud Nabavi, John Tabone, Mark Kokoris, Cara Machacek, Lacey Merrill, Jagadeeswaran Chandrasekar, Svetlana Kritzer, Cynthia Cech, Translocation control elements, reporter codes, and further means for translocation control for use in nanopore sequencing (2020).

18. J. Comer, V. Dimitrov, Q. Zhao, G. Timp, A. Aksimentiev, Microscopic Mechanics of Hairpin DNA Translocation through Synthetic Nanopores. *Biophys. J.* **96**, 593–608 (2009).

19. D. K. Lathrop, E. N. Ervin, G. A. Barrall, M. G. Keehan, R. Kawano, M. A. Krupka, H. S. White, A. H. Hibbs, Monitoring the Escape of DNA from a Nanopore Using an Alternating Current Signal. *J. Am. Chem. Soc.* **132**, 1878–1885 (2010).

20. J. Lietard, A. Meyer, J.-J. Vasseur, F. Morvan, An efficient reagent for 5′-azido oligonucleotide synthesis. *Tetrahedron Lett.* **48**, 8795–8798 (2007).

21. A. H. El-Sagheer, T. Brown, Click chemistry with DNA. *Chem. Soc. Rev.* **39**, 1388 (2010).

22. M. Mag, J. W. Engels, Synthesis and selective cleavage of oligodeoxyribonucleotides containing non-chiral internucleotide phosphoramidate linkages. *Nucleic Acids Res.* **17**, 5973–5988 (1989).

23. D. Pasini, The Click Reaction as an Efficient Tool for the Construction of Macrocyclic Structures. *Molecules* **18**, 9512–9530 (2013).

24. H. Yang, F. Seela, Circular DNA by "Bis-Click" Ligation: Template-Independent Intramolecular Circularization of Oligonucleotides with Terminal Alkynyl Groups Utilizing Bifunctional Azides. *Chem. – Eur. J.* **22**, 1435–1444 (2016).

25. M. B. Koo, S. W. Lee, J. M. Lee, K. T. Kim, Iterative Convergent Synthesis of Large Cyclic Polymers and Block Copolymers with Discrete Molecular Weights. *J. Am. Chem. Soc.* **142**, 14028–14032 (2020).

26. F. Seela, V. R. Sirivolu, P. Chittepu, Modification of DNA with octadiynyl side chains: synthesis, base pairing, and formation of fluorescent coumarin dye conjugates of four nucleobases by the alkyne--azide "click" reaction. *Bioconjug. Chem.* **19**, 211–224 (2008).

27. M. Yoshikawa, T. Kato, T. Takenishi, A novel method for phosphorylation of nucleosides to 5′-nucleotides. *Tetrahedron Lett.* **8**, 5065–5068 (1967).

28. J. Ludwig, F. Eckstein, Rapid and efficient synthesis of nucleoside 5'-0-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **54**, 631–635 (1989).

29. R. L. Letsinger, J. S. Wilkes, L. B. Dumas, Nucleotide chemistry. XVII. Enzymic synthesis of polydeoxyribonucleotides possessing internucleotide phosphoramidate bonds. *J. Am. Chem. Soc.* **94**, 292–293 (1972).

30. J. P. H. Verheyden, J. G. Moffatt, Halo sugar nucleosides. I. Iodination of the primary hydroxyl groups of nucleosides with methyltriphenoxyphosphonium iodide. *J. Org. Chem.* **35**, 2319–2326 (1970).

31. A. K. Bhattacharya, G. Thyagarajan, Michaelis-Arbuzov rearrangement. *Chem. Rev.* **81**, 415–430 (1981).

32. R. T. Pon, S. Yu, Hydroquinone-O,O'-diacetic acid ('Q-linker') as a replacement for succinyl and oxalyl linker arms in solid phase oligonucleotide synthesis. *Nucleic Acids Res.* **25**, 3629–3635 (1997).

33. G. S. Cremosnik, A. Hofer, H. J. Jessen, Iterative Synthesis of Nucleoside Oligophosphates with Phosphoramidites. *Angew. Chem. Int. Ed.* **53**, 286–289 (2014).

34. Mark Stamatios Kokoris, John Tabone, Melud Nabavi, Aaron Jacobs, Dylan O'Connell, Drew Goodman, Lacey Merrill, Jagadeeswaran Chandrasekar, Kendall Berg, Samantha Vellucci, Jessica Vellucci, Enhancement of nucleic acid polymerization by aromatic compounds (2024).

35. H. Ling, F. Boudsocq, R. Woodgate, W. Yang, Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91–102 (2001).

36. F. Boudsocq, S. Iwai, F. Hanaoka, R. Woodgate, Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic polη. *Nucleic Acids Res.* **29**, 4607–4616 (2001).

37. Mark Stamatios Kokoris, Marc Prindle, Melud Nabavi, Craig Ostrander, Taylor Lehmann, Samantha Vellucci, Michael Kovarik, John Chase, Robert Busam, Miranda Lahman, Dpo4 Polymerase Variants (2017).

38. Mark Stamatios Kokoris, Marc Prindle, Michael Kovarik, Alex Lehmann, Dpo4 Polymerase Variants with Improved Properties (PCT application) (2024).

39. A. Vaisman, H. Ling, R. Woodgate, W. Yang, Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis. *EMBO J.* **24**, 2957–2967 (2005).

40. T. A. Kunkel, L. A. Loeb, On the fidelity of DNA replication. Effect of divalent metal ion activators and deoxyrionucleoside triphosphate pools on in vitro mutagenesis. *J. Biol. Chem.* **254**, 5718–5725 (1979).

41. John D. Hem, Chemical equilibria affecting the behavior of manganese in natural water. *Hydrol. Sci. J.* **8**, 30–37 (1963).

42. N. N. Jetha, M. Wiggin, A. Marziali, "Forming an α-Hemolysin Nanopore for Single-Molecule Analysis" in *Micro and Nano Technologies in Bioanalysis*, R. S. Foote, J. W. Lee, Eds. (Humana Press, Totowa, NJ, 2009; http://www.springerlink.com/index/10.1007/978-1-59745-483-4_9)vol. 544, pp. 113–127.

43. R. Silva, E. C. Muniz, A. F. Rubira, Maleimide Immobilized on a PE Surface: Preparation, Characterization and Application as a Free-Radical Photoinitiator. *Langmuir* **25**, 873–880 (2009).

44. R. Wilson, Preparation of single-stranded DNA from PCR products with streptavidin magnetic beads. *Nucleic Acid Ther.* **21**, 437–440 (2011).

45. The work described herein was done at Stratos Genomics to demonstrate the capabilities of the SBX chemistry prior to its acquisition by Roche Molecular Systems in May 2020. The SBX chemistry has been incorporated into a novel and proprietary instrument and workflow currently in development.

46. C. J. Kalkman, LabVIEW: a software system for data acquisition, data analysis, and instrument control. *J. Clin. Monit.* **11**, 51–58 (1995).

47. S. F. Altschul, B. W. Erickson, Optimal sequence alignment using affine gap costs. *Bull. Math. Biol.* **48**, 603–616 (1986).

48. R. J. Kokoska, K. Bebenek, F. Boudsocq, R. Woodgate, T. A. Kunkel, Low fidelity DNA synthesis by a y family DNA polymerase due to misalignment in the active site. *J. Biol. Chem.* **277**, 19633–19638 (2002).

49. J. Ivica, P. T. F. Williamson, M. R. R. De Planque, Salt Gradient Modulation of MicroRNA Translocation through a Biological Nanopore. *Anal. Chem.* **89**, 8822–8829 (2017).

50. C. W. Fuller, S. Kumar, M. Porel, M. Chien, A. Bibillo, P. B. Stranges, M. Dorwart, C. Tao, Z. Li, W. Guo, S. Shi, D. Korenblum, A. Trans, A. Aguirre, E. Liu, E. T. Harada, J. Pollard, A. Bhat, C. Cech, A. Yang, C. Arnold, M. Palla, J. Hovis, R. Chen, I. Morozova, S. Kalachikov, J. J. Russo, J. J. Kasianowicz, R. Davis, S. Roever, G. M. Church, J. Ju, Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *Proc. Natl. Acad. Sci.* **113**, 5233–5238 (2016).

51. R. J. A. Chen, H. Tian, B. Maney, United States Patent: 9557294 - Nanopore-based sequencing with varying voltage stimulus (2017). http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=17&f=G&l=50&co1=AND&d=PTXT&s1=genia.ASNM.&OS=AN/genia&RS=AN/genia.

52. Roche Touts “Good Progress” on Nanopore Sequencer, Stays Mum on Details, Timeline, *GenomeWeb* (2024). https://www.genomeweb.com/sequencing/roche-touts-good-progress-nanopore-sequencer-stays-mum-details-timeline.

53. N. Lerminiaux, K. Fakharuddin, M. R. Mulvey, L. Mataseje, Do we still need Illumina sequencing data? Evaluating Oxford Nanopore Technologies R10.4.1 flow cells and the Rapid v14 library prep kit for Gram negative bacteria whole genome assemblies. *Can. J. Microbiol.* **70**, 178–189 (2024).

54. N. El-Lagta, L. Truong, F. Ayora, F. Mobegi, S. Bruce, P. Martinez, L. D’Orsogna, D. De Santis, Revolutionising High Resolution HLA Genotyping for Transplant Assessment: Validation, Implementation and Challenges of Oxford Nanopore Technologies’ Q20^{+} Sequencing. *HLA* **104**, e15725 (2024).

55. M. Eisenstein, Illumina faces short-read rivals. *Nat. Biotechnol.* **41**, 3–5 (2023).

56. Roche Group Media, Roche acquires Genia Technologies to strengthen next generation sequencing pipeline (2014). https://assets.roche.com/imported/en/med-cor-2014-06-02-e.pdf.

57. C. W. Fuller, L. R. Middendorf, S. A. Benner, G. M. Church, T. Harris, X. Huang, S. B. Jovanovich, J. R. Nelson, J. A. Schloss, D. C. Schwartz, D. V. Vezenov, The challenges of sequencing by synthesis. *Nat. Biotechnol.* **27**, 1013–1023 (2009).

Acknowledgements

We thank A. Stephan, H.F. Johnson, J. Steel, S. Rose, L. Huntsman, H Kawaguchi and H. Dreismann for program support; L. Loeb, A. Marziali, V. Makhijani, D. Heindl, H. Kuchelmeister, J. Mannion, D. McAndrew, A. Wallace, M. St. Louis, M. Lahman, A. Asbe, A. Holt, E. Quig, E. Hermanns, J. Loo, R. Walsh, K. McElligott, C. Berrios, A. Lee, M. Carroll, J. Horsman, E. Williams, N. Westergreen, T. Frese, C. Rosevear, Y. Umezawa, K. Fosnacht, K. Smith and J. Chang for technical support; K. Walker for patent support; B. Poppell for IT support; D. Jones for finance support; B. Anderson for 3D animation; A. Regev for publishing support. Funding for this work was provided by Stratos Genomics Inc., grants from the National Human Genome Research Institute (R21 HG006304) and Department of Defense Small Business Innovation Research (DARPA-BAA-09-22 Strategic Technologies). Funding was acquired by A. Stephan, M. Kokoris and R. McRuer. M. Kokoris and R. McRuer conceptualized the program. M. Kokoris is the first author and PI. All authors participated in the research and investigation. RM developed the software. MK, RM, MN, AJ, MP, CC, TL, CM, JT, SV, KC, supervised the research. MK, RM, MN, AJ, MP, CC, KB, TL wrote the original draft. MK, RM, MN, AJ, MP, CC, KB, LM, CW, TL reviewed and edited. Competing interests: Various authors are inventors on the following patents 7939259, 8349565, 8324360, 9920386, 9670526, 10851405, 11920184, 10457979, 10676782, 10301345, 10774105, 10745685, 11299725, 12037577, 12116570, 11530392, 11970731, 9771614, 8586301 and 10996213 held by Roche Sequencing Solutions, Inc. that cover SBX Technology. MK, RM, MN, AJ, MP, CC, KB, TL, JC, LM, ML, TR, CW, SB, AL, MK, RB, SM, BB, BK, ML, MW, AK, AA, MRP, SK, JL, MB are employees of Roche Sequencing Solutions, Inc. SBX, Xpandomer and XNTP are registered trademarks of Roche Sequencing Solutions, Inc.