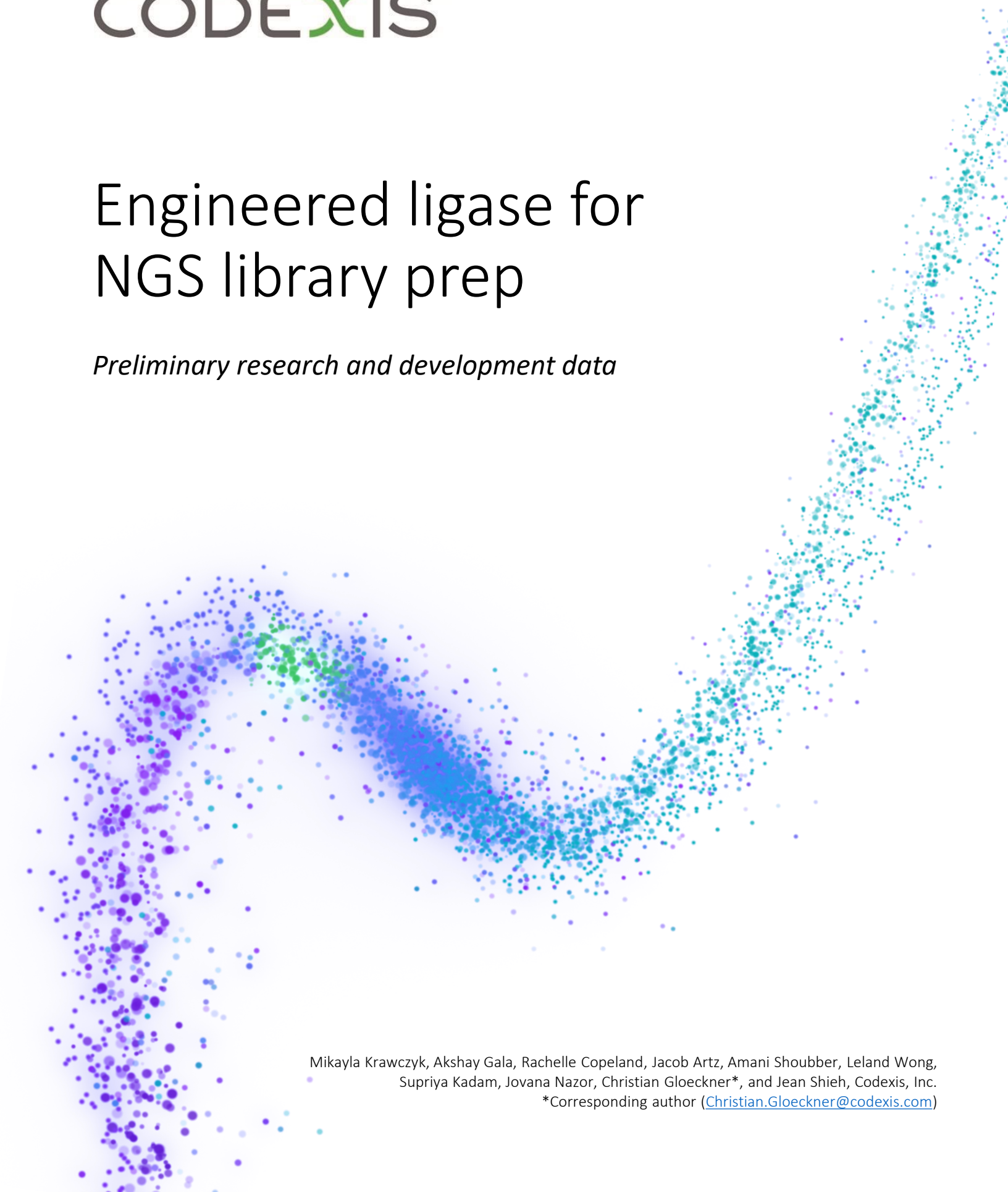
CODEXIS®

Engineered ligase for NGS library prep

Preliminary research and development data



Mikayla Krawczyk, Akshay Gala, Rachelle Copeland, Jacob Artz, Amani Shoubber, Leland Wong, Supriya Kadam, Jovana Nazor, Christian Gloeckner*, and Jean Shieh, Codexis, Inc.
*Corresponding author (Christian.Gloeckner@codexis.com)

Current challenges in NGS library preparation

Ligation is arguably the most critical step in NGS library preparation. Without bilateral adapter ligation, a molecule cannot be amplified during the subsequent PCR, which prevents it from being sequenced and thus detected. Efficient ligation is therefore essential to achieving high sensitivities and detecting low allelic frequency DNA alterations. However, achieving high ligation efficiencies is challenging when the amount of starting material is limiting, as is common with liquid biopsy derived cell-free DNA or fine needle aspirate derived FFPE DNA.

Traditionally, high adapter concentrations and viscous crowding reagents, such as PEG, have been used to enhance ligation efficiency. However, these methods lead to adapter dimer formation, which in turn, necessitates a purification step. This leads to DNA loss, decreasing the library diversity. Furthermore, the reaction viscosity presents a challenge to liquid handling automation.

Rather than optimizing the reaction for an inefficient ligase, we address the challenges at the root cause through enzyme engineering. Here we present preliminary data of CDX Ligase, a pre-commercial enzyme engineered to maximize the ligation efficiency of DNA fragments and adapters. This is especially beneficial during preparation of libraries for NGS workflows requiring the highest sensitivity, even when the amount of starting material is constrained.

Engineering the next generation DNA Ligase

Utilizing Codexis' proprietary CodeEvolver® technology platform, sequence space was interrogated by iterative rounds of diversity generation through site-saturation mutagenesis. Mutation positions were selected based on the enzyme performance targets (**Fig. 1 - left**). Beneficial diversity was identified in high-throughput screening assays and recombined in subsequent rounds of evolution. All assays were designed to represent challenging reaction conditions specific to each enzyme trait targeted for improvement. Engineering was concluded when the enzyme performance targets of CDX Ligase were achieved, and the NGS library prep performance targets were demonstrated (**Fig. 1 - right**).

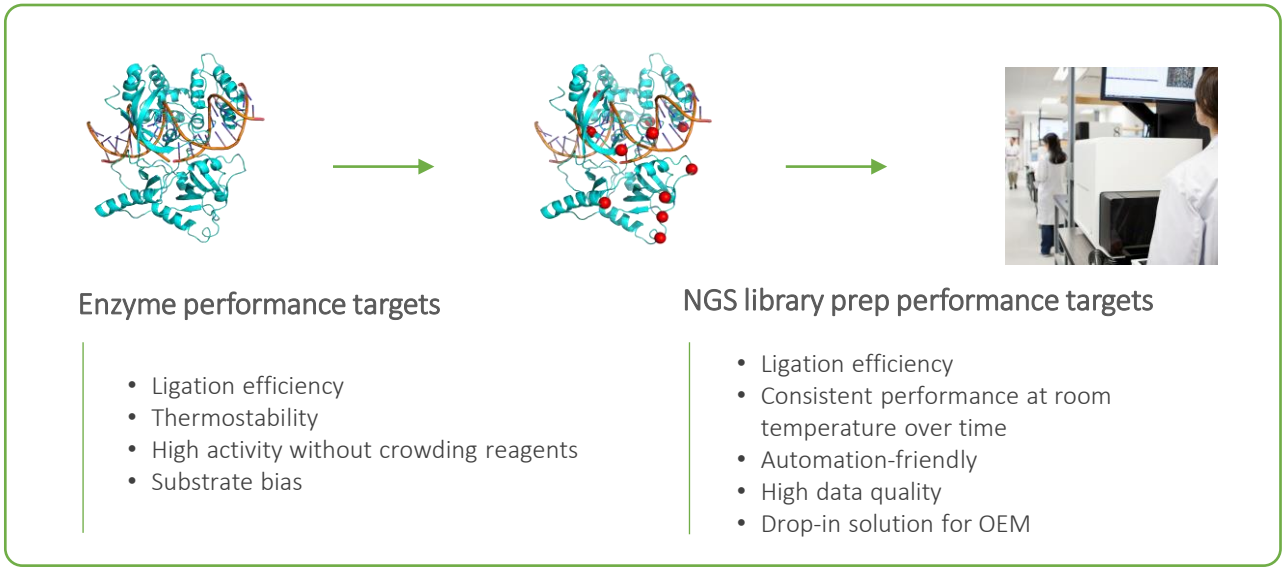


Figure 1. Target enzyme performance and NGS library prep performance improvements

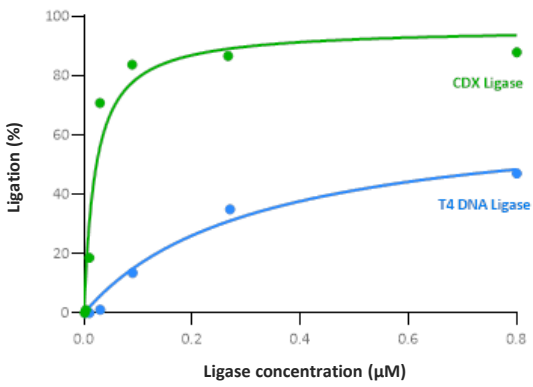
Materials and method

For each enzyme performance test, a 30-minute ligation reaction was performed under the following conditions unless stated otherwise: 1x ligation buffer with 6% PEG, 100 ng of an A-tailed dsDNA substrate, 200 nM NGS adapter and DNA ligase at optimum concentration (0.2 µM for CDX Ligase and 1.7 µM for T4 DNA Ligase). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, USA), T4 DNA Ligase was purchased from Qiagen (Beverly, USA), and ligation conversion was analyzed using Perkin Elmer (Waltham, USA) Labchip GX DNA High Sensitivity Assay.

Improved ligation efficiency

We assessed the ability of CDX Ligase and T4 DNA Ligase to convert the A-tailed dsDNA substrate into double-ligated product. As shown in **Figure 2**, CDX Ligase achieved greater than 80% conversion of the DNA insert into double-ligated product at enzyme concentrations from 0.8 µM to 0.1 µM across two DNA input amounts. CDX Ligase outperforms T4 DNA Ligase at all enzyme concentrations tested below 0.8 µM.

A



B

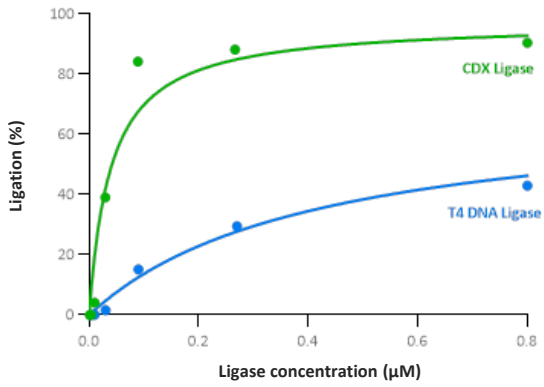


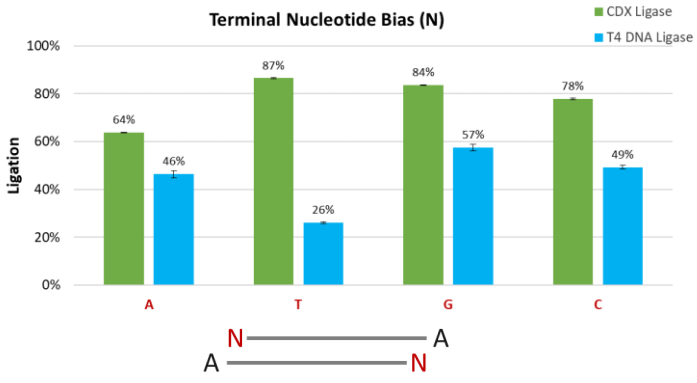
Figure 2. Ligation efficiency at varying enzyme concentrations for two DNA input amounts

In Figure 2A, with 30 ng 160-mer insert, CDX Ligase achieved >80% ligation at or above 0.1 µM ligase, while T4 DNA Ligase never achieved >50% ligation, even at 0.8 µM ligase. Figure 2B shows similar results for 100 ng 160-mer insert.

Decreased substrate bias

To investigate the ligases' sequence bias, we designed individual dsDNA 50-mer oligonucleotides and altered either the terminal nucleotide (**Fig.3, A**) or the penultimate nucleotide (**Fig.3, B**) to all four bases. T4 DNA Ligase demonstrated strong sequence bias for both N and N-1 positions with altered ligation efficiencies depending on the nucleotide present. CDX Ligase showed little or no substrate bias. CDX Ligase's reduced sequence bias is critical to generating complex sequencing libraries and detecting rare fragments independent of their terminal sequence composition.

A



B

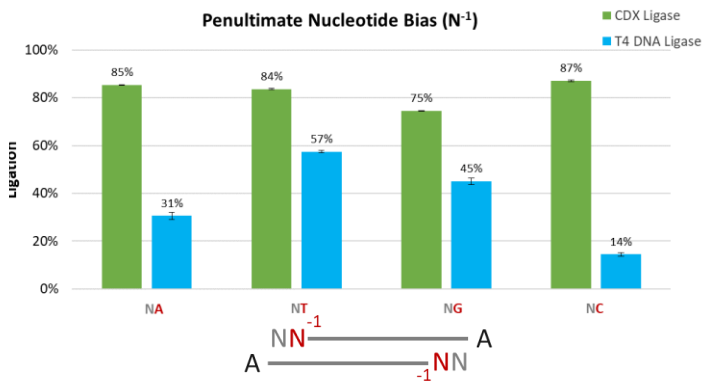


Figure 3. Substrate bias across varying terminal and penultimate nucleotide sequence

T4 DNA Ligase showed a pronounced sequence bias against thymine in the terminal position with a reduction of 31% ligation efficiency. CDX Ligase showed bias against adenine, whereas no bias was observed for the other nucleobases (**A**). Strong influence of the nucleobase identity was also observed with T4 DNA Ligase in the 5' penultimate position (N-1), where bias against adenine, guanine, and cytosine decreased ligation to 31% (NA), 45% (NG), and 14% (NC), while % ligation of CDX Ligase was between 75% to 87% (**B**).

Superior substrate conversion without PEG at low and high DNA inputs

Crowding reagents such as PEG present a challenge for liquid handling equipment. Therefore, we examined ligation efficiency in the presence and absence of PEG using both low and high DNA input amounts with a 20-fold molar excess of NGS adapter. Substrate conversion to double-ligated product remained high for CDX Ligase (**Fig. 4**), even in the absence of PEG. This makes CDX Ligase ideal for automated liquid handling protocols.

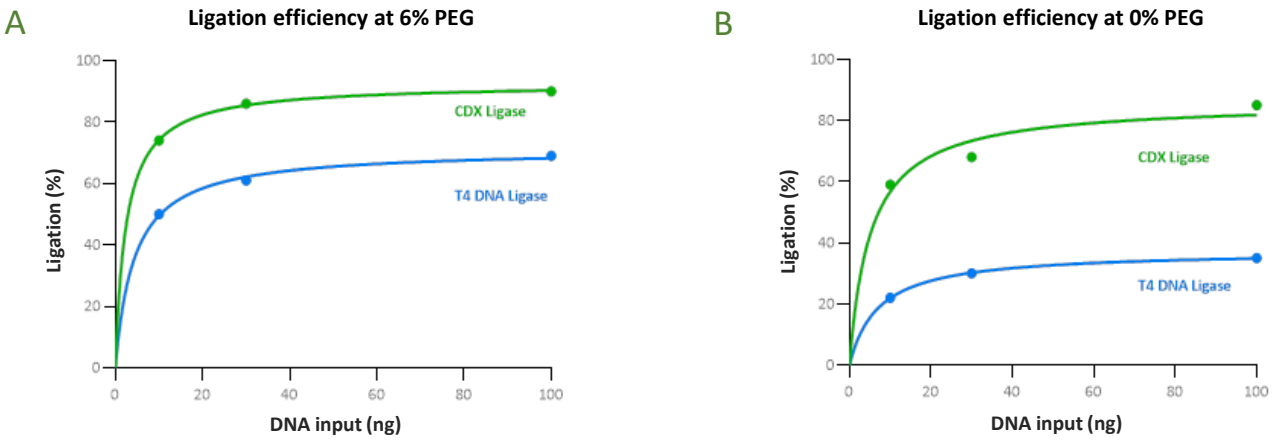


Figure 4. Substrate conversion to double ligated product in presence and absence of PEG

CDX Ligase demonstrated higher ligation efficiency than T4 DNA Ligase with DNA input as low as 10 ng in presence of 6% PEG (**A**). In the absence of PEG (**B**), ligation efficiency remained high for CDX Ligase, whereas a dramatic reduction was observed for T4 DNA Ligase at all DNA input amounts. CDX Ligase converted more than 2x the amount of double-ligated product compared to T4 DNA Ligase.

Stable at room temperature for up to 14 days

High-throughput library preparations using automated liquid handling often require reagents to be stable at room temperature for extended periods of time. Engineering a ligase towards improved thermostability resulted in an enzyme with ambient temperature stability even in absence of glycerol (**Fig 5**). Consistently high enzymatic activity was observed for CDX Ligase after incubation at 22°C for up to 14 days in both buffers tested. With T4 DNA Ligase, double-ligated product formation decreased after only 16 hours at 22°C, and a continuous decrease over time was observed for both buffers tested.

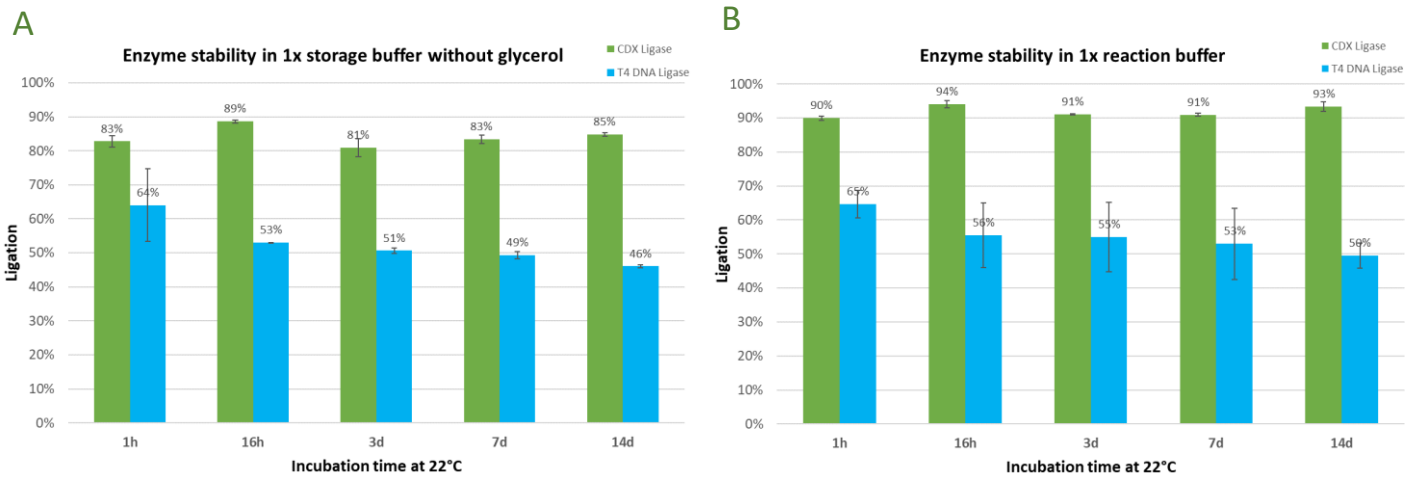


Figure 5. Enzyme stability at 22°C in storage buffer and in reaction buffer

Both ligases were incubated at 22°C for up to 14 days and sampled at the indicated timepoints. (**A**) Results after incubation of enzymes in 1x storage buffer in the absence of glycerol. (**B**) Results after incubation in 1x reaction buffer, mimicking a reaction set-up on a non-cooled liquid handler. Ligases were tested using 10 ng of a 160-mer dsDNA oligonucleotide.

Higher ligation efficiency in less than 5 minutes

Getting from sample to answer quickly is a key challenge for many NGS applications; therefore, improving the efficiency of library preparation workflows is a continuous goal. As shown in **Figure 6**, with 30 ng input DNA, CDX Ligase demonstrated high conversion rates of 84% and 85% ligation at 2.5 min and 5 min incubation, respectively. This could be increased to 88% after 30 min. T4 DNA Ligase generated only 39% double-ligated product at 2.5 min and 54% at 30 min. CDX Ligase is a highly efficient and active enzyme, facilitating reduction of turnaround times in NGS library preparation protocols.

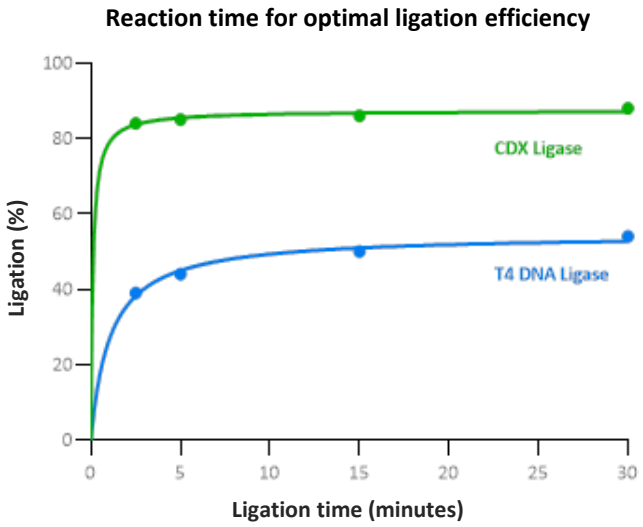


Figure 6. Formation of double-ligated product in a time course

CDX Ligase quickly generated double-ligated product, reaching 84% at the shortest time-point tested (2.5 min), with a minimal increase to 88% observed after 30 min. T4 DNA Ligase demonstrated a greater increase in ligation over time, but still only achieved 54% ligation after 30 min.

Lower adapter dimer formation

Formation of adapter dimers is problematic in NGS library preparation protocols and requires purification prior to PCR amplification. CDX Ligase and T4 DNA Ligase were both incubated for 16 hours with 1 µM of adapter in a 1x reaction buffer, and the formation of adapter dimers was subsequently assessed via Labchip GX DNA High Sensitivity Assay. As shown in **Figure 7**, CDX Ligase demonstrated low inherent dimerization activity of NGS adapters relative to T4 DNA Ligase.

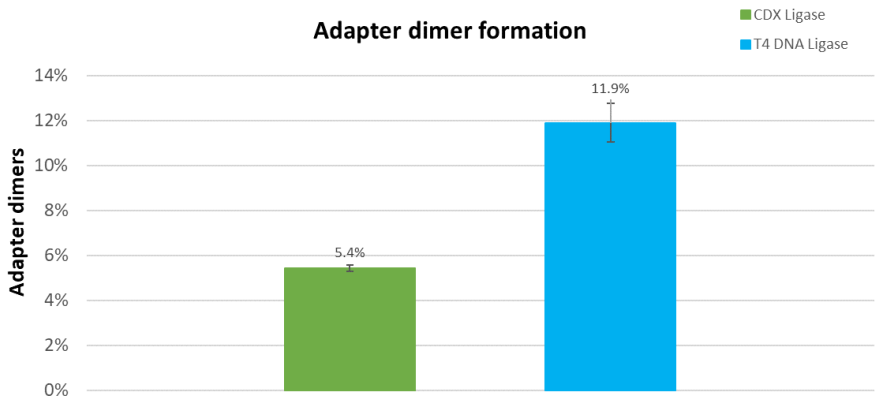


Figure 7. Formation of adapter dimers

Adapters incubated in the presence of CDX Ligase resulted in 5.4% adapter dimers, while T4 DNA Ligase generated more than double the adapter dimers at 11.9% under the same testing conditions.

Confirming CDX Ligase performances in NGS workflows

CDX Ligase was engineered to improve NGS library preparation workflows that require high sensitivity and fast turnaround times. Some applications are limited to a minimal amount of sample material, such as liquid biopsies, fine needle aspirates, or small tissue and FFPE samples. To achieve the highest sensitivity, every molecule needs to be converted into a sequenceable NGS library. Here, we demonstrate the drop-in performance of CDX Ligase by using it in place of the original ligase included in three market-leading library preparation kits.

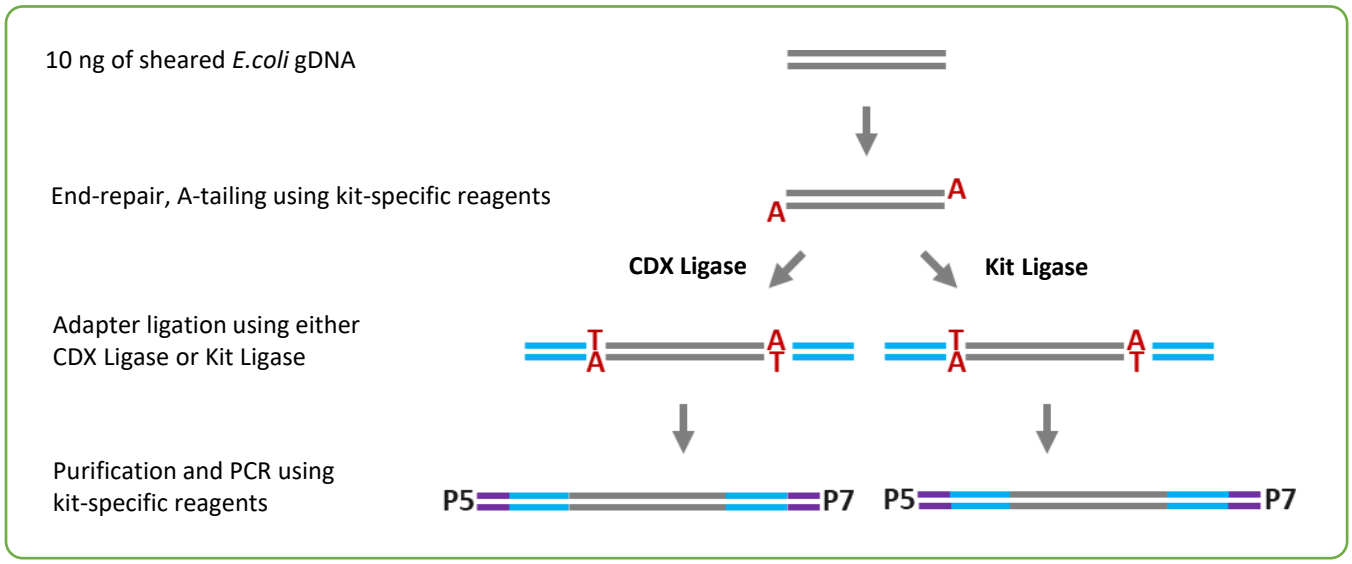


Figure 8. NGS library preparation workflow

E.coli genomic DNA was mechanically sheared and 10 ng was used for each replicate of the library preparation workflow. End-repair and A-tailing was performed as a pool using the manufacturer's provided reagents and protocol. The pool was then split into specific adapter ligation reactions, using either the CDX Ligase or the kit ligase included in each respective library prep kit. Kit specific buffers were used for both ligases. Ligation reactions were subsequently purified and submitted for indexing PCR following each manufacturer's recommendation.

Materials and method

Three commercially available library preparation kits were used to test the performance of CDX Ligase against the kit-provided ligase: Integrated DNA Technologies ("IDT") xGen™ DNA Library Prep MC (10009861), Roche KAPA HyperPrep (7962347001) and Agilent SureSelect® XT HS2 (G9982A). *E.coli* genomic DNA was mechanically sheared using a Covaris S220 Focused-ultrasonicator SN001460. to approximately 400 bp, followed by end-repair and A-tailing reactions using 10 ng of sheared DNA and the recommended reagents and protocols for each kit. Ligation reactions were set-up per manufacturers recommendations using the provided buffer for CDX Ligase and for the kit-provided ligase. All ligation reactions employed the same adapter from IDT (xGEN stubby adaptors #10005924). Libraries were purified and submitted for PCR using the kit-provided PCR reagents and IDT's UDI indexing primers (# 10005922). Quantification of libraries was performed by using Qubit High sensitivity dsDNA kit (Q32851) and measurements were read on Qubit fluorometer 2.0 (Q32866). Libraries were visualized in 2100 bioanalyzer instrument (G2939BA) using Agilent High sensitivity DNA kit (5067-4626). Sequencing was performed using Miseq V3 600 cycle (MS-102-3003) in a 2x250 sequencing on an Illumina-based platform. Data was downsampled for every replicate to 2.3M reads, and analysis was performed using CLC Genomics Workbench 22.0.2 and Picard version 2.27.5.

Higher library yields

Adapter-ligated libraries were amplified using IDT UDI indexing primers in a limited cycle PCR; PCR cycle numbers were adjusted to reflect each manufacturer's recommendations and experimental constraints: IDT kit-generated libraries were amplified using 7 cycles, KAPA libraries were amplified using 8 cycles and Agilent libraries were amplified using 9 cycles of PCR. Both ligated samples were treated identically. Library yield with CDX Ligase was demonstrated to be 1.3-fold to 2.7-fold greater than libraries generated with each respective kit ligase (**Fig. 9**).

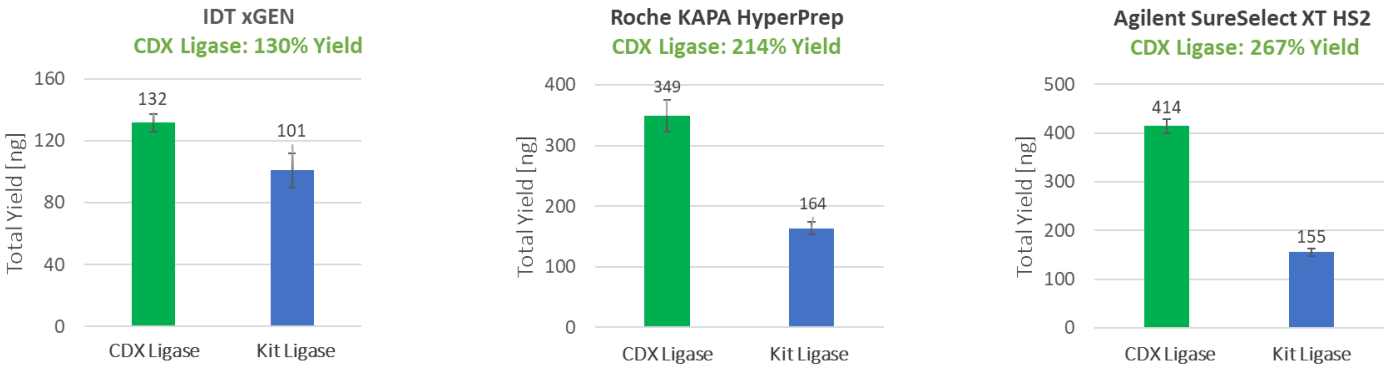


Figure 9. Sequencing library yield

Three library preparation kits were tested using either CDX Ligase or the supplied kit ligase. Yields were assessed after final PCR amplification using indexing primers.

Decreased substrate bias

All libraries were analyzed for sequencing bias of the first 20 bases (**Fig. 10**). The *E. coli* genome is composed of ~25% of each base. Therefore, in the absence of bias, each base should have an even composition of ~25% at every cycle of sequencing. The kit ligases, however, exhibited strong bias against thymine and cytosine in the first cycle of sequencing. This bias continued into subsequent cycles and remained prominent for the KAPA HyperPrep kit ligase and the SureSelect XT HS2 kit ligase up to cycle 10. CDX Ligase demonstrated a much lower sequencing bias compared to its counterparts, which was in line with the reported results of the ligation assay presented in Figure 3.

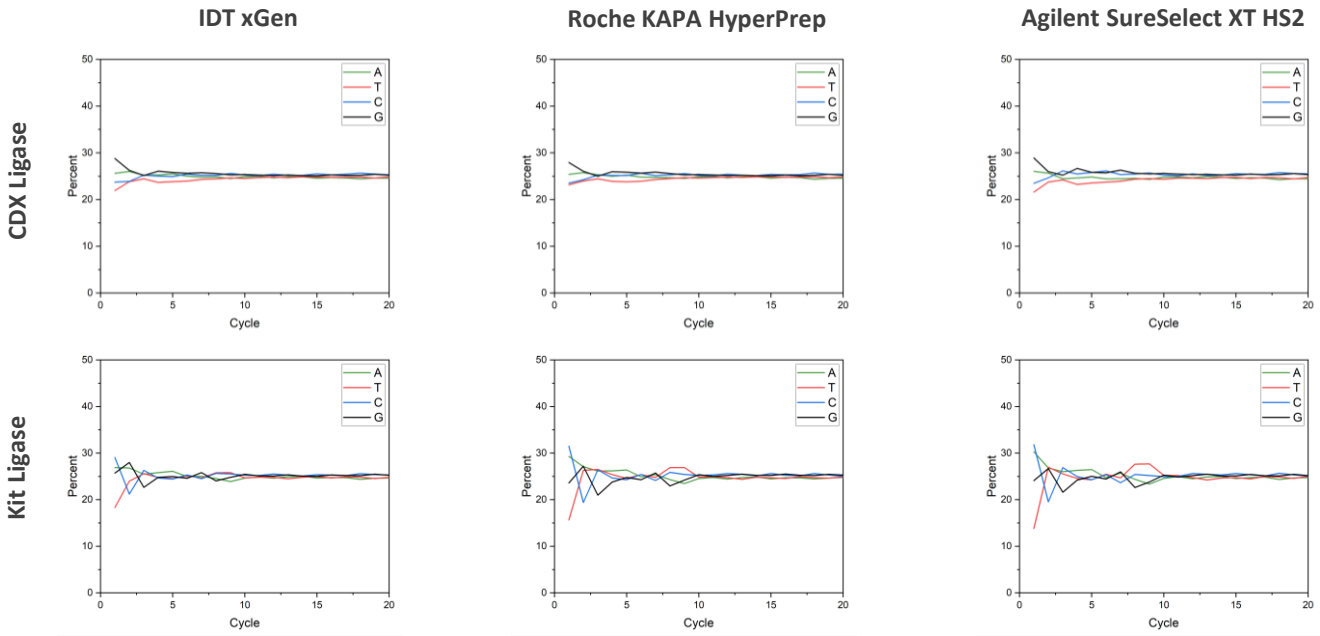


Figure 10. Ligation bias in sequencing reads

Sequencing bias was examined for all samples by observing the first 20 cycles of sequencing. CDX Ligase data (top row) demonstrated less bias than data from each individual kit ligase (bottom row).

Higher library complexity

Library complexity is a measure of the sequencing library diversity. Thus, it represents the ability of a library preparation kit to transform unique sample molecules into a sequenceable library. This metric was examined for CDX Ligase and the three kit ligases (**Fig. 11**). CDX Ligase generated higher library complexity, with observed increases ranging from 106% to 167% relative to the kit ligases. Of note, these gains were generated without any prior optimization of buffer composition or other protocol variables. These data provide additional support of CDX Ligase's reduced sequencing bias and its overall higher activity and efficiency in forming double-ligated products.

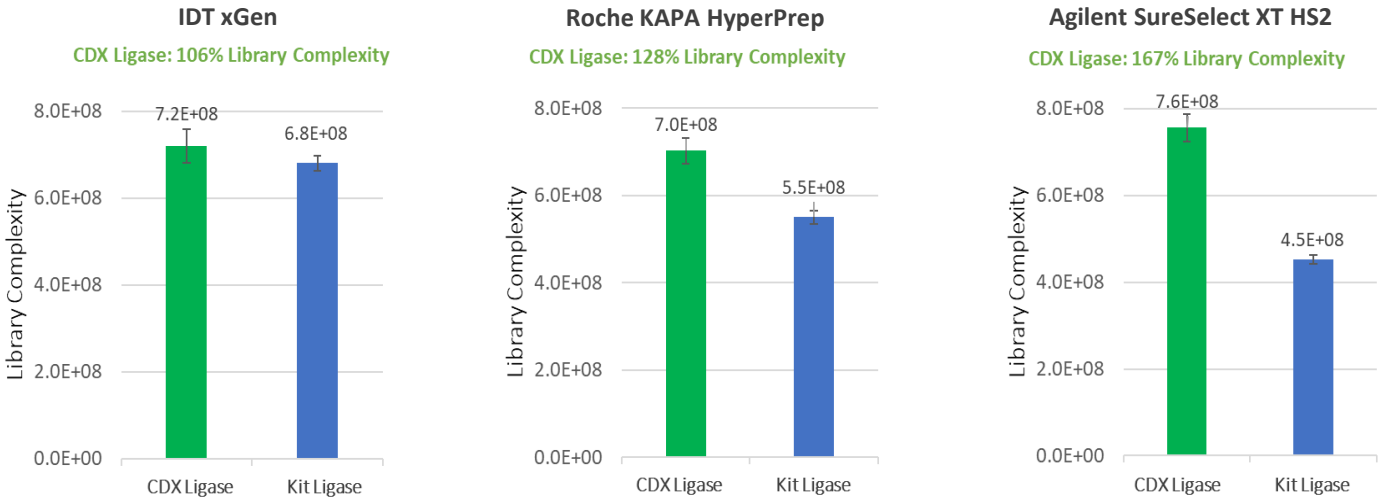


Figure 11. Library complexity

CDX Ligase demonstrated increased library complexity than that observed for the kit supplied ligase. Library complexity was assessed after downsampling all libraries to 2.3 million reads.

Conclusion

Data from ligation assays verified CDX Ligase's performance in ambient temperature over time and its independence from crowding reagents. Furthermore, CDX Ligase delivered higher ligation efficiency in shorter reaction time, offering automation process optimization opportunities to increase overall NGS library sample prep efficiency. CDX Ligase, engineered to address sequencing challenges, can boost assay sensitivity and improve the number of successfully processed samples. CDX Ligase was tested across various assays to demonstrate usability and adaptability in existing sample prep workflows. It demonstrates efficient ligation, even in the absence of PEG (**Fig 4**), as well ambient temperature stability (**Fig. 5**). This, combined with a fast reaction time of 2.5 min (**Fig. 6**), makes CDX Ligase well suited for high throughput automation. DNA sequencing results confirmed that the CDX Ligase successfully increased both library yield (**Fig. 9**) and complexity (**Fig. 11**) when used as a drop-in replacement for the ligase provided in three commercially available library prep kits. Taken together, the CDX Ligase will allow its users to significantly improve the critical step of adapter ligation, which, in turn, will improve library complexity, detection and discovery.

Codexis, Inc.
200 Penobscot Drive, Redwood City, CA 94063 USA
www.codexis.com | lifesci@codexis.com