An accelerated protein engineering workflow utilizing linear DNA in a cell-free expression system

INTRODUCTION

Protein engineering and production are essential processes in synthetic biology and biotechnology. Researchers typically use design-build-test cycles to systematically improve the functionality of proteins or even an entire biomolecular pathway. After an initial design phase in which key parameters are identified, the corresponding protein library is built, and prototypes are tested. Thorough data analysis then feeds into the design phase of a second cycle. More cycles can follow for continued improvement until a protein is engineered with the desired characteristics. Traditionally, this cycle is highly dependent on the ability to edit the genetic code for the protein, transforming a desired organism with the genetic material to allow expression, and then growing the organism to generate the newly designed protein, all of which can be extremely time consuming.

In recent times, the design-build-test cycle has been accelerated by using in vitro protein expression systems, also known as cell-free protein expression. In cell-free protein expression systems, gene transcription (TX) and translation (TL) are independent from cell viability and growth, enabling the production of cytotoxic proteins that are typically difficult to produce in cell culture. The lack of cell walls allows on-demand manipulation of reaction conditions and easy access to the produced proteins. Finally, removing the need for transformation from the process and adding the option of using linear DNA fragments as expression templates creates the opportunity to build high-throughput processing platforms for rapid protein engineering.

myTXTL® is a bacterial, cell-free expression system from Daicel Arbor Biosciences utilizing *E. coli* endogenous TXTL machinery while also supporting the use of the common T7 expression system. myTXTL features a Master Mix, which is composed of *E. coli* cell extract and a reaction buffer with co-factors, building blocks, and salts. It is provided in a single tube to facilitate an easy mix-and-go workflow (Figure 1A). After adding a DNA template to the Master Mix and incubating for a few minutes, expressed proteins can be detected directly without purification. Because gene expression is entirely driven by the *E. coli* core RNA polymerase and transcription factor 70 (Figure 1B), myTXTL is compatible with nearly any constitutive or inducible promoter system used for recombinant protein production in *E. coli* cells. Recent advancements have demonstrated cell-free expression technology for an exciting breadth of research applications, including nanotechnologies and disease research [1].

gBlocks™ Gene Fragments are double-stranded DNA fragments available in lengths of 125–3000 bp from Integrated DNA Technologies, Inc (IDT). These synthetic DNA constructs are delivered ready for assembly into longer DNA parts or direct use in a variety of cloning procedures. Traditional gene construction and modification techniques requiring multi-step manipulation with primers, PCR, subcloning, and gene sequencing can be avoided by using gBlocks Gene Fragments. Due to their ease of use, these fragments have already been used in many research fields, such as engineering of enzymes, antibodies, and other proteins; CRISPR-based genome editing; as well as biosensor research and gene circuitry.

IDT gBlocks Gene Fragments combined with the Daicel Arbor Biosciences myTXTL Cell-Free Expression system are technologies for a high-throughput, research workflow that overcomes laborious and time-consuming bottlenecks of traditional recombinant protein production, such as molecular cloning, transformation, and recombinant in vivo gene expression.

In this study, we demonstrate how pairing gBlocks Gene Fragments with myTXTL Cell-Free Expression shortens protein engineering workflows in comparison to those utilizing in vivo protein expression. In addition, common DNA template design considerations are discussed, including template length and sequence features for a fluorescent model protein expressed from two different promoter systems.

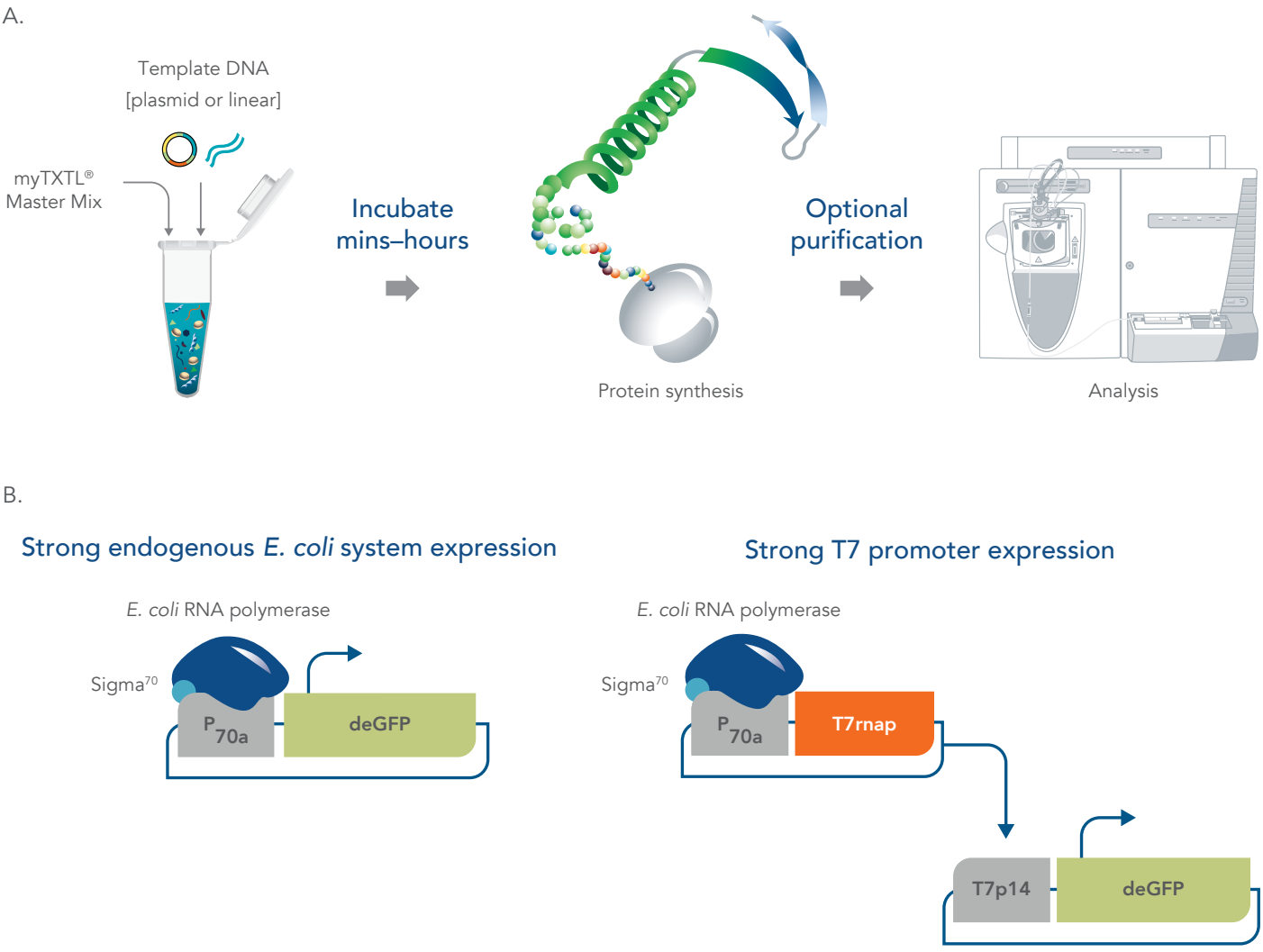


Figure 1. myTXTL Cell-Free Expression. (**A**) Upon the addition of circular or linear template DNA to the myTXTL Master Mix, cell-free gene expression starts immediately. Produced target proteins are readily accessible for rapid downstream processing due to the lack of cell membrane and compartmentalization. (**B**) The myTXTL system supports gene expression regulated by different promoters including inducible and constitutive endogenous *E. coli* promoters as well as phage promoters, such as the T7 system.

MATERIALS AND METHODS

Preparation of IDT gBlocks Gene Fragments for cell-free gene expression

A panel of gBlocks Gene Fragments (IDT) was designed to encode a fluorescent reporter gene called deGFP, an engineered version of eGFP [2], under transcriptional control of the Sigma70-specific promoter (P70a) or a T7 phage promoter (T7p14) (Figure 2). The impact of varying lengths of the 5′- and 3′-flanking regions, as well as the presence of a terminator sequence were explored. Lyophilized gBlocks Gene Fragments were reconstituted in molecular-biology-grade water and subjected to a heating step for 20 minutes at 50°C to fully resuspend and dissolve the DNA. At this point, the gBlocks Gene Fragments were used for cell-free gene expression. The DNA concentrations of individual gBlocks Gene Fragments were confirmed by absorbance measurements on a Nanodrop™ spectrophotometer (Thermo Fisher Scientific). For a subset of gBlocks Gene Fragments, cell-free gene expression was also investigated after amplification with PCR. For this, 10 ng of reconstituted gBlocks Gene Fragment DNA was amplified and the PCR amplicons were subjected to a clean-up procedure (PureLink Kit, Thermo Fisher Scientific) with a final elution in molecular-biology-grade water prior to cell-free gene expression.

Cell-free gene expression

Each cell-free expression reaction was set up in a 2 mL tube and contained 9 µL of myTXTL Linear DNA Master Mix (Linear DNA Expression Kit, Daicel Arbor Biosciences) and 20 nM of a gBlocks Gene Fragment construct (IDT) in a final volume of 12 µL. In the case of T7 promoter constructs, T7 RNA polymerase was expressed from the helper plasmid, pTXTL-P70a-T7rnap (Daicel Arbor Biosciences), at a final concentration of 0.1 nM in the myTXTL reaction. In an incubator set to 29°C, cell-free expression was conducted in triplicate for 16 hours.

Protein quantification

Model protein deGFP yield was used to evaluate the influence of several DNA template designs on cell-free gene expression. For deGFP quantification, the fluorescence signal from a myTXTL reaction expressing deGFP was analyzed by linear regression of a standard curve of purified recombinant deGFP. After completion of the myTXTL reaction, samples were placed on ice for 5 minutes, followed by centrifugation at 16,300 x g for 3 minutes. Then, each sample was diluted with PBS to fit into the linear range of the standard curve, and aliquots of each sample and standard protein dilution were loaded onto a black Nunc™ 384-well plate (Thermo Fisher Scientific) in triplicate. deGFP fluorescence was excited at a wavelength of 485 nm and detected at a wavelength of 535 nm using a GENios FL fluorescence plate reader and appropriate filter sets (Tecan).

RESULTS

Rationale of gBlocks Gene Fragment design

The ability of cell-free systems to accept linear DNA templates encoding the target gene is one of the major advantages over traditional in vivo gene expression, as it reduces the necessary genetic information to only a few central elements: promoter/ribosome binding site (RBS), open reading frame (ORF), and terminator. Using Circular DNA templates (plasmids) as vehicles for recombinant protein production are well understood for in vivo expression systems, and many of the same plasmids are suitable for in vitro cell-free expression. However, design rules for linear DNA templates have not been fully explored. In this study, a reference sequence for the P70a and the T7 promoter was built from a 5′ non-coding region, a promoter, the deGFP gene, a terminator, and a 3′ non-coding region (Figure 2A). From this, the influence of the non-coding 5′- and 3′-flanking region length on protein yield was investigated. Specifically, the 5′-flanking region could have an influence on gene expression because it is the binding area of the transcriptional apparatus. Furthermore, gene expression from constructs missing the terminator sequence was examined. Understanding the requirements of a minimal construct minimizing DNA synthesis cost and for maximizing ORF size while minimally compromising sequence fidelity.

A.

B.

gBlocks construct	5′ region (bp)	Promoter	Open reading frame	Terminator	3′ region (bp)	Length (kbp)	deGFP protein	
							Concentration (μM)	Yield (μg)
1A	250	P70a	deGFP	T500	100	1.211	27.8±2.2	9.2
2A	100				100	1.062	26.2±2.8	8.7
3A	15				100	0.977	18.3±1.2	6.1
4A	15				15	0.892	17.5±1.0	5.8
1B	250			–	100	1.182	25.7±2.1	8.5
2B	100			–	100	1.032	21.9±1.4	7.2
3B	15			–	100	0.947	14.7±0.1	4.8
4B	15			–	15	0.862	11.4±0.2	3.8
5A	250	T7p14	deGFP	T7 term	100	1.304	29.5±1.1	9.8
6A	100				100	1.154	28.4±1.9	9.4
7A	15				15	0.984	21.6±1.8	7.1
5B	250			–	100	1.214	12.7±0.2	4.2
6B	100			–	100	1.064	12.1±0.8	4.0
7B	15			–	15	0.894	8.0±0.1	2.7

Figure 2. Design of gBlocks Gene Fragment constructs and deGFP production in myTXTL cell-free expression. (**A**) Non-coding sequences flanking the gene cassette were generated with an online random sequence generator tool choosing 50% GC content. gBlocks Gene Fragment constructs exhibited variations in the flanking region length as well as presence of the terminator sequence. (**B**) Expression of deGFP was quantified after 16 hours of cell-free expression.

myTXTL Cell-Free Expression from linear gBlocks Gene Fragments

Upon delivery, the designed gBlocks Gene Fragment constructs were rehydrated in sterile water and subsequently used as template DNA material—without any further preparation steps—for cell-free expression with myTXTL Linear DNA Master Mix. The amount of deGFP expression from each linear DNA construct was assayed the next day by measuring the amount of fluorescence produced. Remarkably, all gBlocks Gene Fragment constructs produced significant amounts of protein with a concentration of up to 0.8 mg/mL, equivalent to 10 µg per myTXTL reaction (Figure 2B and 3). Overall, gBlocks Gene Fragments constructs encoding a P70a promoter (constructs 1–4) produced deGFP even when there was modification of both flanking regions as well as the removal of the T500 terminator sequence (Figure 3A). Truncating the 5′-flanking region in P70a reference construct 1A by 150 bp (construct 2A) or absence of the entire terminator sequence (construct 1B) had only minimal impact on the production of deGFP. Further length reduction in the non-coding region on either end (with or without terminator present) resulted in a decrease of protein yield of up to 60% for the shortest construct, 4B, which lacks a terminator. This indicates that the optimum 5′-flanking region length on P70a promoter constructs is between 15 and 100 bp, and the 3′-flanking region can be as short as 15 bp.

Gene expression controlled by a T7 promoter in myTXTL reactions requires co-expression of T7 RNA polymerase. Typically, a helper plasmid, pTXTL-P70a-T7rnap, is added to the myTXTL Master Mix in addition to DNA template harboring the T7 promoter (constructs 5–7). Protein production from the reference sequence (construct 5A) was very similar to the P70a reference construct indicating a comparable strength of those two promoter systems (Figure 3B). Shortening the T7 promoter reference construct (constructs 6A and 7A) resulted in a slight to moderate decrease in protein yield, therefore, a 5′- and 3′-non-coding region length between 15 and 100 bp produced the highest amount of deGFP protein. In contrast, removal of the T7 terminator sequence had a detrimental effect on protein production with a greater than 50% decrease for the reference-derived construct 5B.

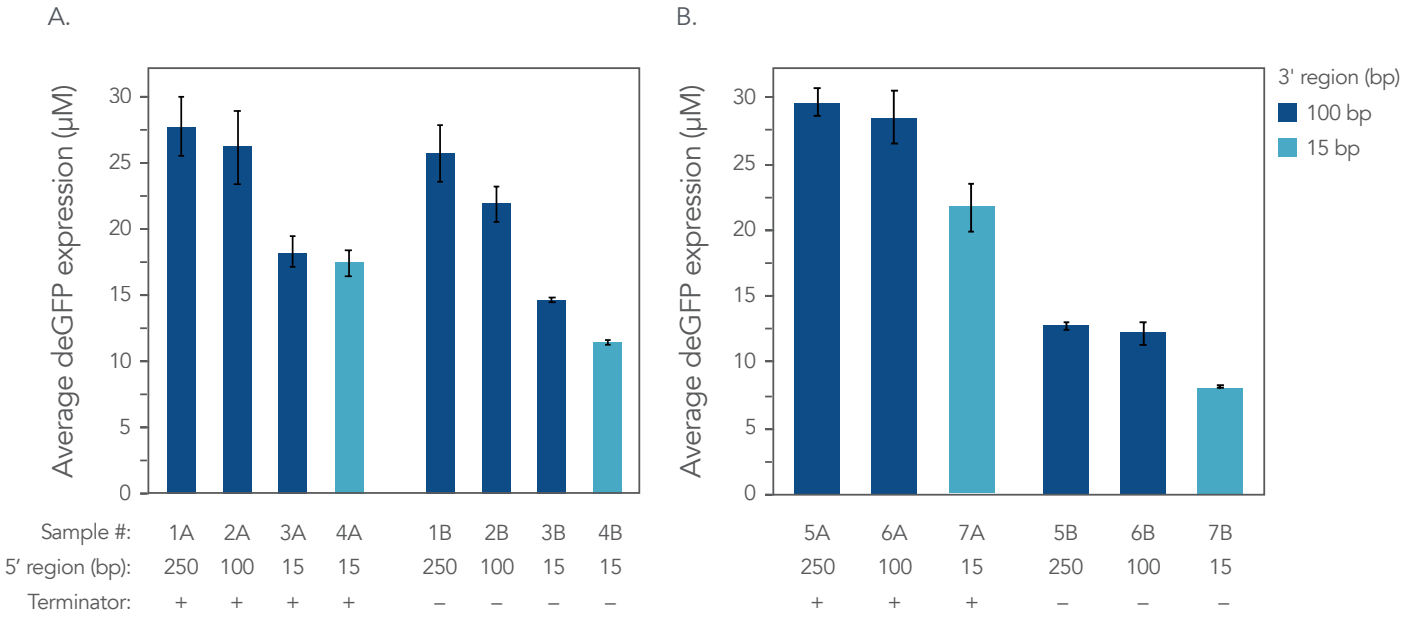


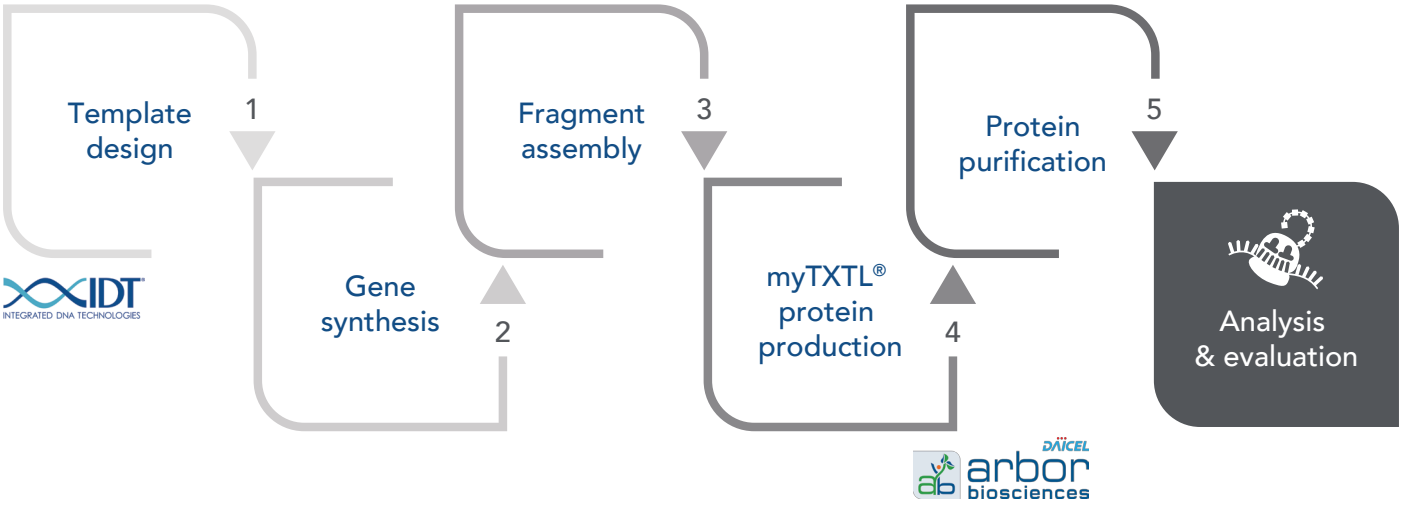
Figure 3. Impact of linear DNA template design on cell-free protein production. gBlocks Gene Fragment constructs harboring the (**A**) strong Sigma^{70}-specific P70a or (**B**) phage T7 promoter were incubated with myTXTL Linear DNA Master Mix, and deGFP fluorescence was quantified. Standard deviation was calculated from three individual myTXTL reactions. While gBlocks Gene Fragments may be amenable to variations in length of 5′/3′ non-coding regions, presence of a terminator sequence seems to be essential for deGFP expression.

For a subset of constructs, the above standard cell-free expression protocol was modified to include steps that are potentially useful during high-throughput processing. gBlocks Gene Fragment constructs 1A and 5B were amplified by PCR and subjected to a clean-up procedure before applying them as template DNA in myTXTL reactions. With a moderately reduced deGFP yield (~25%) compared to the gBlocks Gene Fragment samples used without prior amplification, amplicon generation can serve as a valuable option to enable higher template input or a greater reaction number. Also, myTXTL cell-free reactions with gBlocks Gene Fragments incubated in a 0.2 mL PCR tube in a thermocycler instead of a 2.0 mL reaction vessel resulted in up to 50% diminished protein output.

CONCLUSION

Combining ready-to-use, high-fidelity IDT gBlocks Gene Fragments with high-yield Daicel Arbor Biosciences myTXTL Linear DNA Master Mix enables rapid iterations of design-build-test cycles used in biotechnology and synthetic biology to quickly screen protein libraries, analyze metabolic pathways, test gene circuits, and more. Since gBlocks Gene Fragments do not require any additional processing before setting up a cell-free expression reaction, target proteins are readily available within hours after receipt of DNA material for analysis and characterization. With a yield of up to 10 µg functionally active target protein from a single reaction, myTXTL Linear DNA Master Mix is a cell-free system for linear templates providing sufficient material for downstream analysis and characterization. Both technologies can enable next-generation, high-throughput protein screening workflows, which eliminates traditional molecular cloning or DNA sequencing procedures (Figure 4). Importantly, gBlocks Gene Fragments and myTXTL Master Mixes can be processed with automated liquid handling systems and are available in multi-well format or large aliquot size for ease of use. For protein expression projects at any scale, gBlocks Gene Fragments with myTXTL Master Mixes offer a convenient, highly controllable, and scalable solution to accelerate protein expression and synthetic biology workflows.

High-throughput screening pipeline with myTXTL®



Liquid-handling robots optional for pipeline process

Figure 4. Next-generation, high-throughput protein screening with linear, double-stranded DNA fragments and cell-free expression technology. (Step 1) Various DNA template design tools are available online to help create gene sequences and regulatory elements for gene expression. (Step 2) Desired fragment sequences are submitted for synthesis and are delivered within 1–2 weeks, ready to use for cell-free expression or an optional assembly (Step 3) using PCR if needed. (Step 4) Cell-free expression only requires mixing gBlocks Gene Fragments with myTXTL Linear DNA Master Mix followed by (optional Step 5) protein purification or immediate characterization of the produced target protein. Once a promising protein variant is identified, the DNA template can be subcloned into a plasmid. Then, mid- to large-scale protein production can be performed in vitro or in vivo for further in-depth analysis.

REFERENCES

1. Silverman AD, Karim AS, et al. (2020) Cell-free gene expression: an expanded repertoire of applications. Nat Rev Genet 21(3):151–170.
2. Garamella J, Marshall R, et al. (2016) The all E. coli TX-TL toolbox 2.0: A platform for cell-free synthetic biology. ACS Synth Biol 5(4):344–355.

Technical support: genes@idtdna.com

Integrated DNA Technologies

For more than 30 years, IDT's innovative tools and solutions for genomics applications have been driving advances that inspire scientists to dream big and achieve their next breakthroughs. IDT develops, manufactures, and markets nucleic acid products that support the life sciences industry in the areas of academic and commercial research, agriculture, medical diagnostics, and pharmaceutical development. We have a global reach with personalized customer service.

> SEE WHAT MORE WE CAN DO FOR **YOU** AT **WWW.IDTDNA.COM**.

Daicel Arbor Biosciences

Daicel Arbor Biosciences, a subsidiary of Chiral Technologies, Inc. (part of Daicel Corporation), is a development and manufacturing company founded by scientists to serve our peers in molecular biology applications. We are a passionate organization of scientists determined to deliver cost-effective, user-friendly products to researchers of genetics and synthetic biology. The team at Daicel Arbor Biosciences prides themselves on providing exceptional customer service and timely technical support to new or advanced users on our array of products. We routinely collaborate with our customers and research partners to develop innovative solutions to address their unique applications.

> CONNECT WITH US AT **WWW.ARBORBIOSCI.COM**.

For research use only. Not for use in diagnostic procedures. Unless otherwise agreed to in writing, IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Purchaser is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations.

IDT INTEGRATED DNA TECHNOLOGIES

DAICEL arbor biosciences