# Automated CRISPR-Based Iterative Genome Engineering for Rapid Strain Improvement

## Introduction

Proteins represent an important class of bioproducts with many applications in biotechnology, biopharma, cellular agriculture, materials, and more. A growing demand for these products is encouraging scientists to develop efficient methods for their production in microbial cell factories, as well as to facilitate engineering of new protein variants with improved properties. Advances in synthetic biology and genome engineering have enabled significant progress in these areas [1,2]; however, challenges remain when it comes to strain generation time and throughput.

The Design–Generate–Test–Learn (DGTL) cycle is an increasingly adopted engineering framework that represents an efficient approach to biological engineering. Significant improvement can be achieved in a short amount of time by recombining beneficial edits through successive rounds of genome engineering. The effectiveness of this method relies on the ability to rapidly move through iterations of the DGTL cycle. Until recently, the Design and Generate stages have been the rate-limiting; however, novel genome editing technologies have significantly shortened these steps [3]. The Onyx® Digital Genome Engineering Platform further accelerates DGTL by automating the design and construction of thousands of edited strains in parallel. In addition, the recently introduced Onyx Iterative Genome Engineering workflow streamlines the plasmid curing and preparation of strains for the next round of editing, allowing users to shorten the time between cycles (**Figure 1**).

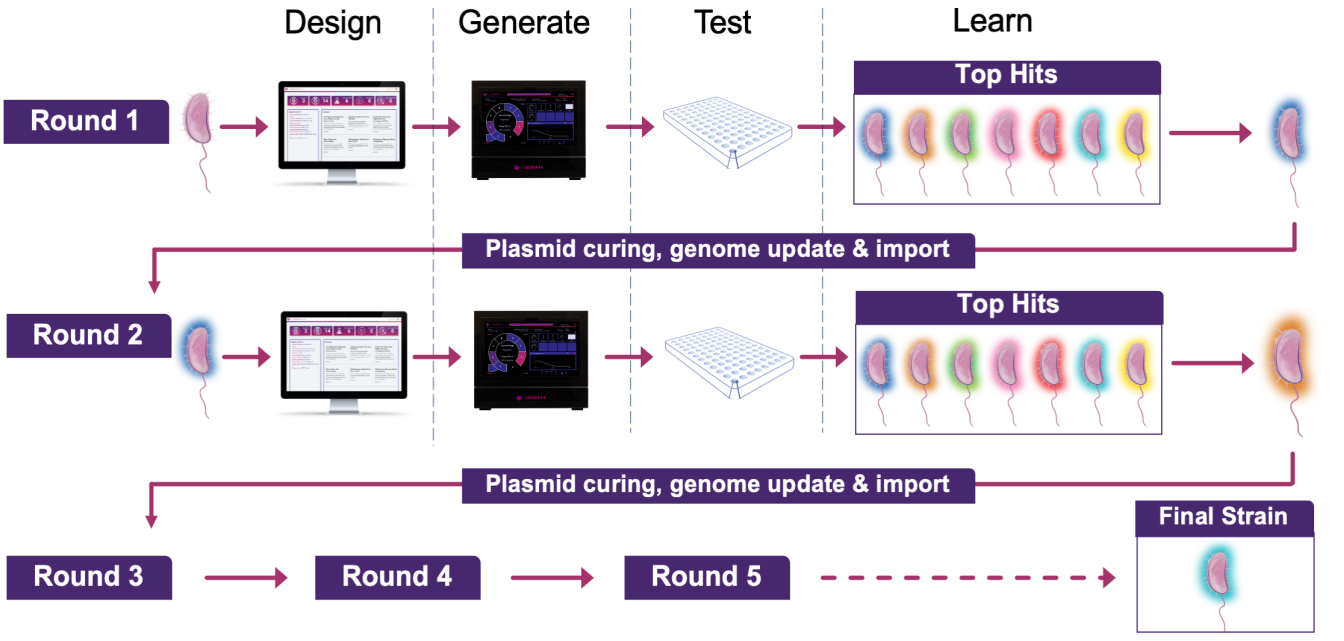


**Figure 1:** The Design–Generate–Test–Learn (DGTL) cycle using the Onyx Iterative Genome Engineering workflow.

The DGTL framework can be successfully applied to both protein and strain engineering. While protein engineering focuses on editing the sequence of a single gene, improving protein production strains requires a genome-wide approach. However, both objectives can be conceptualized as a similar problem of optimizing a large combinatorial search space [4]. Here, we showcase how the Onyx Iterative Genome Engineering workflow can be successfully applied to both problems to accelerate the development of improved protein variants and optimize strains for heterologous protein production.

## Engineering novel protein functions using Iterative Genome Engineering

The typical workflow for protein engineering includes cloning a gene of interest into a plasmid and performing site-saturation or random mutagenesis, which can be accomplished using a variety of plasmid editing techniques. However, this approach has several potential drawbacks. First, plasmids present a metabolic burden for the host cell and require antibiotics for maintenance. Second, the level of protein expression may be difficult to control due to plasmid copy-number variations. This presents a problem for functional screening and ability to quantitatively assess improvement in engineered variants. Finally, if the engineered protein is to be expressed for large-scale commercial production, the gene will often be integrated into the host chromosome for stable genomic expression, creating a lag in technology transfer and scale-up.

Highly efficient multiplex CRISPR editing offers a way to circumvent these potential problems by performing protein engineering directly on the genome. To demonstrate this workflow in *Escherichia coli*, we integrated a codon-optimized super-folder green fluorescent protein (sfGFP) variant from *Aequorea victoria* [5] into the host chromosome using Inscripta's open-sourced MAD7™ CRISPR nuclease. The integration cassette contained a refactored sfGFP gene under a strong constitutive J23102 promoter, a strong synthetic RBS, and a T0 terminator (**Figure 2A**). We then used the Onyx platform to generate 723 targeted edits against mutational hotspots (**Figure 2B**) in the sfGFP sequence identified through literature and database searches [6].

After generating the edited cell library using the Onyx workflow, we selected 192 individual colonies to screen for altered spectral characteristics. Fluorescence intensity was measured at three excitation and emission wavelengths based on blue, green, and yellow fluorescent protein (YFP) peaks. One variant showed a 1.5-fold increase in green fluorescence intensity compared to wild-type sfGFP (**Figure 2C**). The edit responsible for this variant was identified to be F145M. This position had been previously implicated in improving the folding kinetics that contribute to faster maturation of sfGFP [7].

To see whether the spectral properties could be further improved or diversified with additional rounds of editing, the $sfGFP_{F145M}$ variant was selected for the Iterative Genome Engineering workflow. First, the variant was taken through the plasmid curing workflow (see **Onyx Plasmid Curing Handbook - E. coli – Doc. # 1002402**) to prepare the isolate for the next Genome Engineering Run. The designs for the second editing cycle used the same site-saturation library, resulting in 722 edits stacked on top of the F145M edit. Out of those, 96 colonies were screened and two novel variants identified with significantly improved yellow fluorescence intensity (**Figure 2D**). The variants contained edits T203F and T203I in addition to the previous amino acid change. This example shows how you can use the

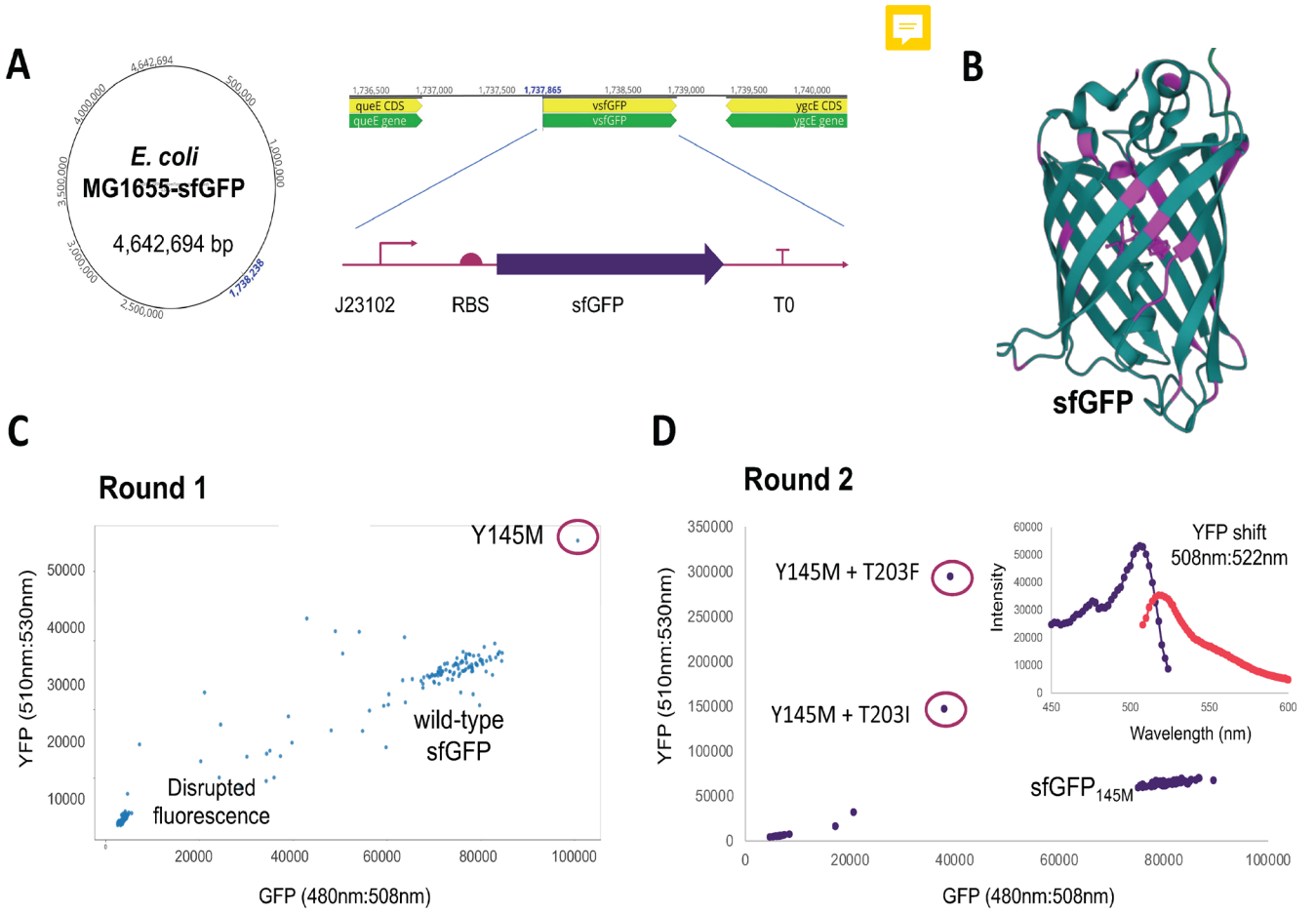


**Figure 2:** Iterative Genome Engineering approach to engineering new spectral characteristics in the Green Fluorescent Protein (GFP). **A)** Chromosomal integration construct, with sfGFP, J23102 promoter, RBS and T0 termination elements inserted into the genome as one cassette. **B)** sfGFP structural model: mutational hotspot residues indicated in magenta. **C)** Fluorescence intensity of screened library variants from the first round of editing along the GFP and YFP excitation-emission spectra compared to the wild-type sfGFP and YFP. The highlighted Y145M variant was taken to the next round of editing. **D)** Identification of two novel YFP variants with significantly improved intensity (Y145M+T203F and Y145M+T203I) obtained in the second round of editing by stacking the original edit library on top of the sfGFPF145M variant.

same library designs to combine different edits and identify synergistic interactions. With significant reduction in the library design and construction steps, Onyx Iterative Genome Engineering workflow cuts down typical protein engineering timeline from months to just a few weeks.

## Iterative strain engineering for improved heterologous protein production in yeast

Protein production in industrial hosts like *Saccharomyces cerevisiae* is an essential tool in biotechnology, with diverse applications from therapeutics to cellular agriculture [8,9]. Strain development for optimal product yield typically involves both rational and empirical methods such as laboratory evolution. The later can result in accumulating mutational load that can weaken the strain over time [10]. Optimizing production strains requires a comprehensive approach that includes codon usage optimization, promoter and ribosomal binding site (RBS) substitution, as well as genome-wide

knockouts and mutagenesis [8,11]. The Onyx platform enables introducing diverse edit types across a wide range of targets to create rich libraries. Subsequent cycles of Iterative Genome Engineering can be used to combine the beneficial variants identified in the initial screen or add new diversity to the strain.

To demonstrate the utility of this approach on yeast strain optimization for protein production, we chose the fungal cellobiohydrolase I (CBH1) enzyme for heterologous expression. A single copy of the cbh1 gene from *Taloromyces emersonii* was introduced into the chromosome of *S. cerevisiae* CEN.PK strain. The expression cassette contained a native pENO2 constitutive promoter, a native *T. emersonii* signal sequence, and a DIT1 terminator. The heterologous protein expression was confirmed by the CBH1 activity assay with chromophoric substrate pNP-beta-lactopyranoside and baseline activity was established.

The edit libraries were designed to include both known and new targets to mix rational and exploratory approaches [12-14]. These included knockouts of genes previously shown to improve heterologous protein production in yeast; knockouts of known glycosylation and protein degradation genes to improve protein stability; promoter diversification through transcription factor binding site (TFBS) shuffling; a synonymous codon library for the *cbh1* gene; untargeted genome-wide knockouts (GW KO); finally, short, deletions and terminator libraries were included as a genomic background diversification strategy. The total number of edits designed was >14,000 (**Table 1**). Edited cell libraries were constructed over 7 runs on the Onyx instrument and assessed using the Onyx genotyping workflow.

| Library Concept | Library size | Isolates Screened | Est Library coverage | # of retest hits | # of unique hits | Hit rate |
|---|---|---|---|---|---|---|
| GW KO | 5133 | 2024 | 18% | 40 | 26 | 2.0% |
| Deletions across genome (1-59bp) | 1142 | 1104 | 26% | 11 | 6 | 1.0% |
| GW terminator | 5145 | 1104 | 4% | 13 | 8 | 1.2% |
| Targeted KO | 201 | 920 | 68% | 21 | 7 | 2.3% |
| PENO2 TFBS ins/del | 280 | 1104 | 48% | 22 | 4 | 2.0% |
| PENO2 TFBS swap | 1588 | 1104 | 9% | 23 | 19 | 2.1% |
| CBH1 Alternate codon | 1370 | 1104 | 32% | 6 | 4 | 0.5% |
| **Total** | **14859** | **8464** | | **136** | **74** | |

**Table 1:** Summary of edit design libraries used in the first round of genome engineering and number of hits.

Over 8,400 total variants from the first engineering cycle were screened using the CBH1 activity assay and 136 hits selected for retesting. Out of those, 74 unique hits were identified all 7 library concepts (**Table 1**). Isolate genotyping revealed that top hits belonged to diverse functional categories, such as protein degradation, secretion and glycosylation, stress response, transcription and translation, nuclear transport, as well as ENO2 promoter and alternate-codon variants.

The top 2 variants from the first were taken through the plasmid curing and quality control (QC) workflow following the procedures describe in the **Onyx Plasmid Curing and Iterative Genome Engineering Handbook - S. cerevisiae – Doc. # 1002403**. The genomes of these two strains were updated using the Caber tool and imported into InscriptaDesigner™ software to create new design libraries with the goal of further increasing genomic diversity in subsequent engineering cycles.

To design the next round of edits, two different strategies were undertaken. In the first, the top hit identified in the first round of screening was combined with a 5133-edit genome-wide knockout library to generate additional diversity using an unbiased approach. In the second library concept, the two top performing strains from the first round of screening were combined with the other top 103 hits identified through screening. The second-round libraries were constructed in three separate Genome Engineering runs and assessed using the Onyx genotyping assays. Isolates from all three libraries were taken thought the phenotyping workflow to select for top performers. Finally, the top hit identified in the second round of screening was combined with a library of 103 hits from the first round of screening. The third-round library was constructed and phenotyped using the same workflow as described above.

The results from screening of the first, second, and third round libraries showed a stepwise improvement in CBH1 activity after each round of engineering (Figure 3A). In the single-edit libraries, a maximum ~1.6-fold increase over baseline activity was observed; with two-edits libraries, the activity of isolates screened was between 1.6 and 2.3-fold above baseline; the triple-edit isolates showed up to a 3-fold improvement. Genotyping of the libraries reveled a diverse distribution of edits among several functional classes, such as protein degradation and trafficking, stress response, transcription and translation, nuclear transport, glycosylation, and other, including unknown function, genes (**Figure 3B**).

Since the screening was performed in a high-throughput format with isolates grown and assayed in deep-well plates, we wanted to verify that the improvement seen in the engineered strains would translate to a larger scale. We selected top 3 isolates from each round of engineering and cultured them in 25mL shake flasks in YPD medium supplemented with 20 g/L glucose. We analyzed both the growth rate of the engineered strains (**Figure 3C**) and CBH1 activity over 72 hours (**Figure 3D**). Isolates identified in each of 3 different engineering rounds clustered together in terms of the end-point measurement of CBH1 activity, with two triple-edit isolates demonstrating nearly a 4-fold improvement over base strain. Interestingly, these isolates also demonstrated a reduced growth rate, indicating that the increase in CBH1 activity was due to improved protein production rather than increase in growth rate of the strain.

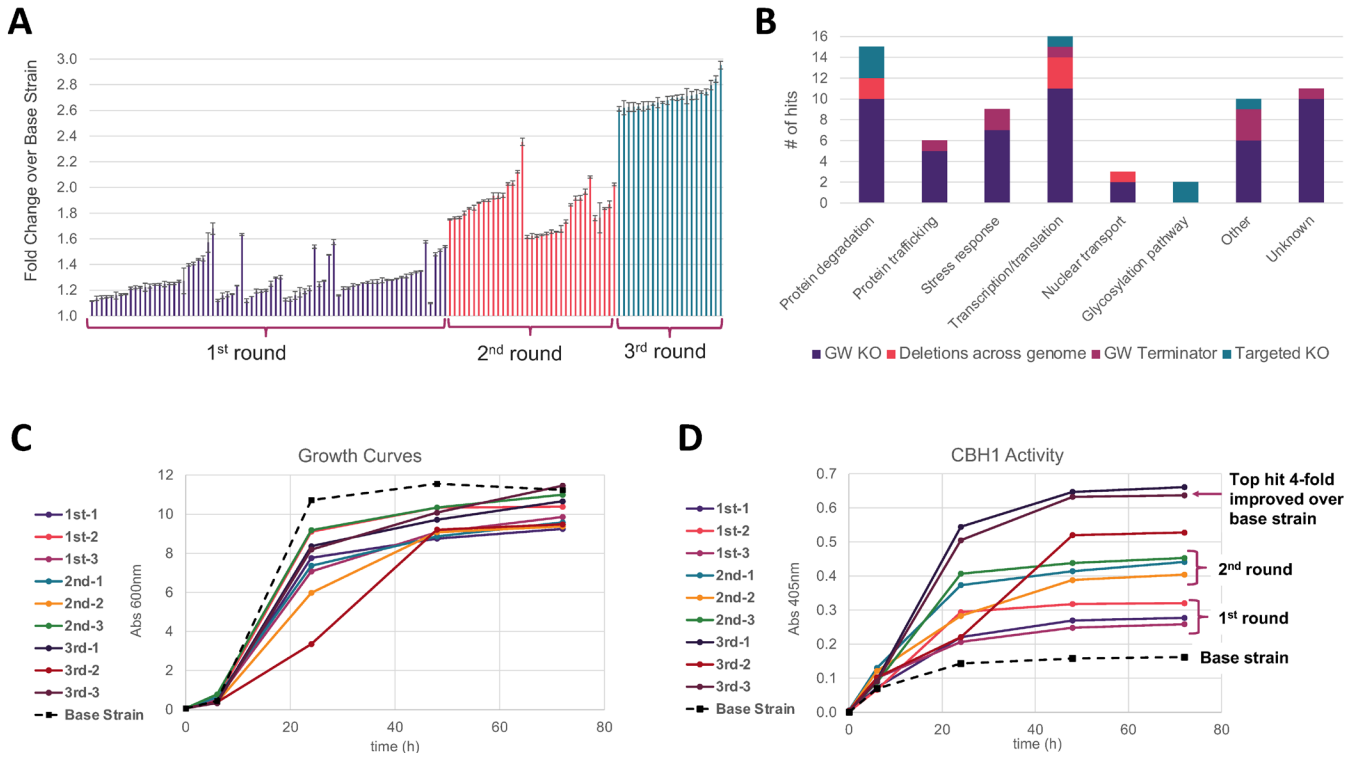


**Figure 3:** Iterative Genome Engineering approach to strain engineering for improved protein production of CBH1. **A)** Fold-change improvement in CBH1 activity in screened variants from the first, second, and third round of editing. Bars represent average values and standard deviation from quadruplicate measurements. **B)** Number of unique variants with improved CBH1 activity sorted by gene function and specific gene targets and their corresponding fold improvement. C) Growth curves of the top 3 variants from each round of editing measured over 72 hours. **C)** CBH1 activity measured in the of the top 3 variants from each round of editing measured over 72 hours. Both growth rate and CBH1 activity were measured for cultures grown in 25mL shake flasks in YPD medium supplemented with 20 g/L glucose.

## Conclusions

As demonstrated by the two applications described above, the Onyx Iterative Genome Engineering workflow significantly accelerates the strain improvement cycle, reducing the time needed to meet your research goals. A four-fold improvement in protein production was seen after 3 cycles of strain engineering, which took just 6 months to complete. Additionally, protein engineering can be greatly sped up by introducing genomic edits into the DNA sequence of a native or genomically integrated heterologous gene, facilitating screening and discovery of new and improved protein variants. With automated library design, massively-multiplexed hands-off strain construction, convenient genotyping assays and seamless workflow integration to facilitate Iterative Genome Engineering cycles, the strain times are reduced from multiple months or even years down to several weeks or months. The Onyx platform is an unparalleled technology for microbial strain optimization, enabling forward engineering applications from developing novel proteins to improving production of biomolecules.

## REFERENCES


1. Sinha, Rajeshwari, and Pratyoosh Shukla. "Current trends in protein engineering: updates and progress." Current Protein and Peptide Science 20.5 (2019): 398-407.
2. Ow, Dave Siak-Wei, et al. "Technological Advances Improving Recombinant Protein Production in Bacteria." Frontiers in Microbiology 12 (2021).
3. Garst, Andrew D., et al. "Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering." Nature biotechnology 35.1 (2017): 48-55.
4. Munro, Lachlan J., and Douglas B. Kell. "Intelligent host engineering for metabolic flux optimisation in biotechnology." Biochemical Journal 478.20 (2021): 3685-3721.
5. Li, C., Zhang, R., Wang, J., Wilson, L. M., & Yan, Y. (2020). Protein engineering for improving and diversifying natural product biosynthesis. Trends in biotechnology 36.7, 729-744.
6. Kashimoto, R., et al. (2021). Expansion and Diversification of Fluorescent Protein Genes in Fifteen Acropora Species during the Evolution of Acroporid Corals. Genes 12.3, 397.
7. Pédelacq, J.-D., et al. (2006). Engineering and characterization of a superfolder green fluorescent protein." Nature biotechnology 24.1, 79-88.
8. Wang, G., Huang, M., and J. Nielsen. "Exploring the potential of *Saccharomyces cerevisiae* for biopharmaceutical protein production." Current opinion in
9. biotechnology 48 (2017): 77-84.
10. Waltz, E. Appetite grows for biotech foods with health benefits. Nature biotechnology 37, 573–480, 2019.
11. Winkler, James, Luis H. Reyes, and Katy C. Kao. "Adaptive laboratory evolution for strain engineering." Systems Metabolic Engineering. Humana Press, Totowa, NJ, 2013. 211-222.
12. Tripathi, N. K., & Shrivastava, A. (2019). Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. Frontiers in bioengineering and biotechnology, 7, 420.
13. Ilmén, M., et al. "High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*." Biotechnology for biofuels 4.1 (2011): 1-15.
14. Penttilä, M. E., et al. "Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*." Gene 63.1 (1988): 103-112.
15. Boer, H., Teeri, T. T., and A. Koivula. "Characterization of Trichoderma reesei cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters." Biotechnology and bioengineering 69.5 (2000): 486-494.


Learn more at **INSCRIPTA.COM/ITERATIVE**

INSCRIPTA®
The Digital Genome Engineering Company

INSCRIPTA.COM



Document 100xxxx RevA/02012022