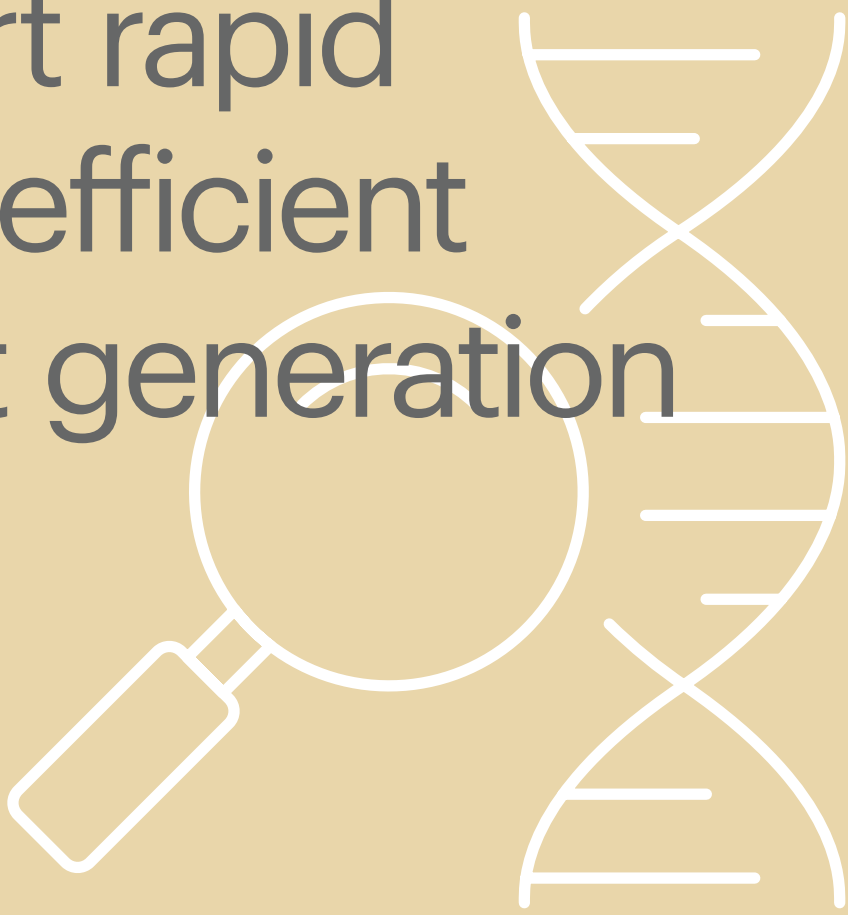
White Paper

How AstraZeneca created an automated DNA assembly framework to support rapid and cost-efficient construct generation



Benchling AstraZeneca

Introduction

What you'll learn when you read this article:

FRAGment recycLER (FRAGLER) is AstraZeneca's software solution used to increase efficiency and throughput of DNA synthesis on a wide variety of targets	**50%-90% reduction in costs to complete DNA synthesis and sequencing as a result of the fragment recycling and reuse and associated time savings afforded by FRAGLER automation in Benchling**	**Decreased end-to-end DNA construction process by ~1-4 weeks from initial request to sequence-validated plasmid**

It takes around 10 years and millions to sometimes billions of dollars for most new drugs to make it to market. Unsurprisingly, this timeline and cost profile doesn't cut it for drug development patients, companies, and clinical trial sponsors. One of the earliest steps that can be optimized is DNA construct development for early drug discovery, as it affects all downstream processes.

An essential first step in drug discovery is recombinant protein production. Expression and purification of the intended target protein enables candidate drug screening as well as structural and mechanistic interrogations that guide the design and evaluation of candidate drugs. Later, cell lines are engineered and used in *in vitro* screening to evaluate toxicity and confirm a candidate drug's mode of action. These early tests are critical for eliminating ineffective and toxic drugs early, before billions of dollars are spent on clinical trials doomed to fail.

DNA constructs — artificially designed vectors or plasmids that deliver target DNA sequences into cells, where the gene product is produced — underpin recombinant protein production and cell line engineering. Rapid and flexible generation of these constructs at scale is one of the most limiting first steps in drug discovery. A large time sink associated with this process is the *de novo* synthesis of long coding sequences, a challenge that modular cloning kits haven't been able to solve effectively.

To address this, AstraZeneca developed a comprehensive DNA assembly framework, which they integrated into Benchling to support rapid, cost-efficient, and scalable construct generation. Their unique solution takes an unheard-of 3 weeks and led to cost savings from 50 to 90%.

Fragment recycling significantly reduces DNA synthesis time and costs

AstraZeneca's solution is an elegant integration of multiple softwares and automation, spearheaded by Associate Principal Scientist David Öling and his group in Sweden. This group provides most of the DNA constructs used by AstraZeneca's teams across Sweden and the U.K.

Öling's team began by optimizing a golden gate assembly-based approach to building DNA constructs. They created and validated 19 different modules that are sufficient for most complex constructs. But the key to AstraZeneca's success is in how the fragments that are plugged into those modules to create the complete construct are identified, generated, and pieced together.

Traditionally, a significant amount of synthesis time and cost is spent on DNA sequences that are present multiple times across a construct, as those sequences have to be ordered individually as many times as they are present. To address this problem, Öling and his team worked with the computational biology team at AstraZeneca to create what he calls "a workaround for the whole DNA synthesis industry" — an algorithm for fragment recycling called FRAGLER (FRAGment recycLER).

Fragment recycling identifies shared coding sequence regions across the construct so that those sequences only need to be ordered once, reducing both the cost of DNA synthesis and the time it takes to generate DNA fragments. FRAGLER not only performs amino acid sequence alignment and codon optimization of desired sequences, but it also fragments those sequences to increase the success of DNA production. Fragmentation of long sequences into shorter fragments of 300-900 base pairs reduces synthesis

time and ensures production success, says Öling. Moreover, fragmentation eliminates sequence complexities and minimizes cancellations, which are very common for longer sequences.

To illustrate how FRAGLER works, AstraZeneca used its software solution to rapidly generate complex SARS-CoV-2 spike protein constructs for expression optimization **(Figure 1)**. The spike protein coding sequence was fragmented into four submodules, which could then be combined with various signal peptides and trimerization domains. The resulting constructs were expressed in HEK293 cells. FRAGLER was also used to fragment 30 full-length SARS-CoV-2 spike protein variants, generating 55 unique fragments and recycling 11 fragments 65 times — a base pair recycle rate of 55.3%.

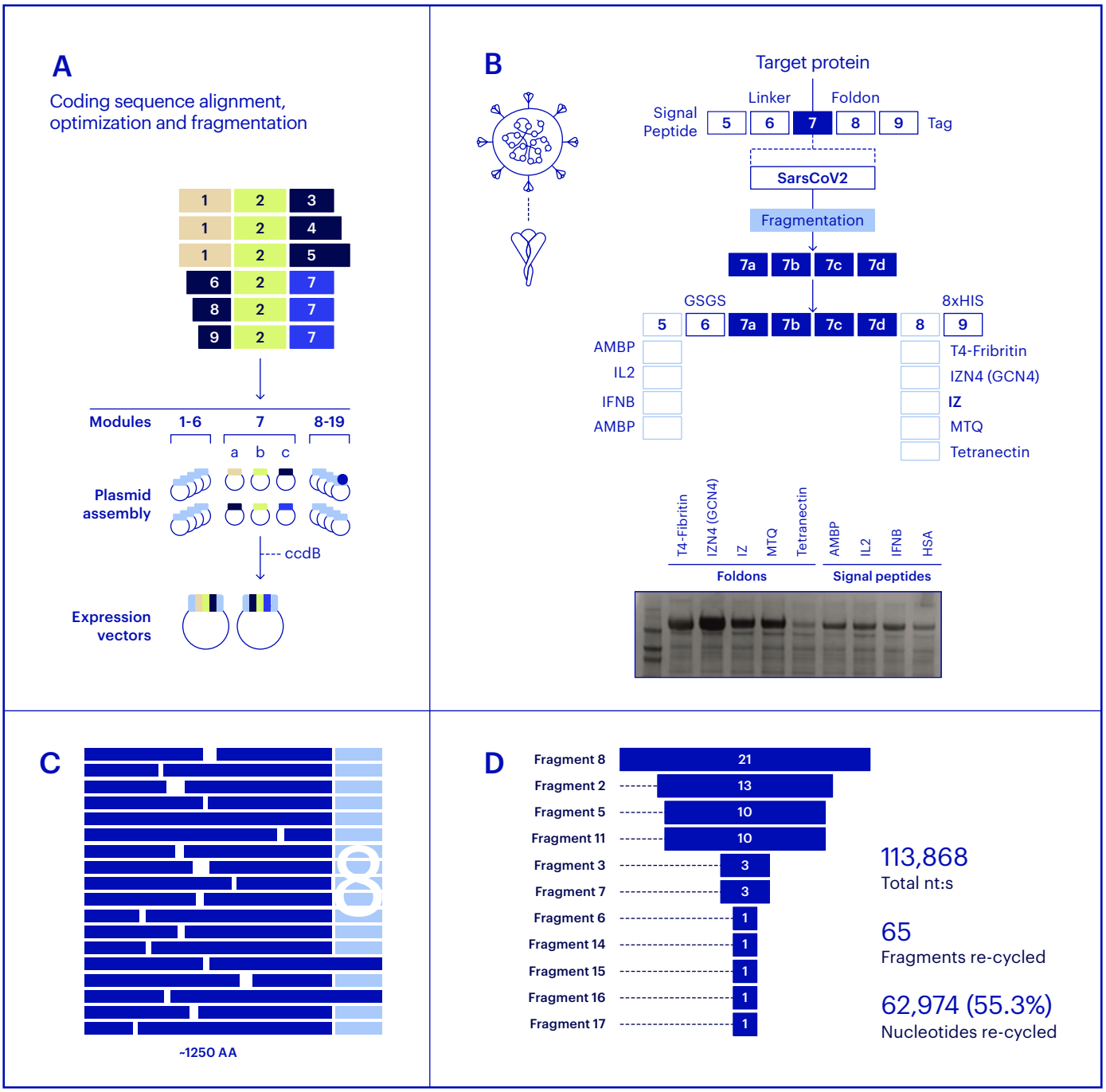


Figure 1. FRAGLER enables rapid generation of SarsCoV2 expression constructs.

A) Sequence alignment, codon optimization, and fragmentation performed by FRAGLER which runs in Benchling to algorithmically search pre-existing DNA sequences.

B) Top: Schematic of the coronavirus spike protein, with module 7 fragmented into four submodules (a, b, c and d) to reduce production timelines. Submodules are combined with various signal peptides (module 5) and trimerization domains (module 8) (top). Bottom: Coomassie-stained gel from a small-scale purification of HIS-tagged spike protein constructs and expression in Hek293 cells.

C) Schematic of SARS-CoV-2 spike protein (PODTC2) alignment and fragmentation with the top 30 hits from Uniprot (90% identity) listed.

D) Top: Graphical representation of recycled fragments. Bottom: Total number of recycled nucleotides and fragments.

According to Öling, the savings afforded by this approach equates to **at least a 50% reduction in costs, and sometimes even more, depending on the constructs.** For example, most of the DNA can be reused when performing single point mutations over a hundred constructs, which affords cost reductions upwards of 80-90%. These long complex constructs were generated rapidly by the fragmentation approach.

Integration with Benchling unleashes the scale-up potential of a fully automated pipeline

While FRAGLER can and has been run offline, offline implementation doesn't permit the type of scale that AstraZeneca needed. Although sequence fragmentation and fragment recycling can significantly drive down costs, that approach realizes its full potential only if you can find the fragments that you're going to reuse. Öling explains, "There needs to be an algorithmic search to properly identify pre-existing fragments for FRAGLER to align. No human can keep track of everything if thousands of constructs are generated annually. Benchling allows us to achieve speed and scale by transforming previously manual processes relying on Excel and various in-house tools into fully automated steps."

Leveraging Benchling's developer platform, the Benchling team built an integration to fully automate the process of generating de-novo fragments. It also allows for rapid searching of the Benchling database across thousands of pre-existing fragments in order to generate fully assembled in-silico constructs.

By integrating FRAGLER into Benchling, pre-existing fragments are identified and collected for the multiple sequence alignment and fragmentation steps. Benchling also creates pooling instructions for the resulting fragments that are used for construct assembly. Files containing information about the pools and which construct(s) they correspond to can then be entered back into the platform. In addition, Benchling facilitates QC steps, such as monitoring plasmid purity and concentration. The result for AstraZeneca is increased throughput coupled with high confidence in data quality since Benchling is the single source of truth for each step.

AstraZeneca developed an in-house design and request portal to enable global teams across Sweden and the U.K. to request constructs. Like FRAGLER, this portal was easily integrated with Benchling via the developer platform and open APIs. The entire process of DNA construct production, from initial request to sequence-validated plasmid, is automated and can be completed in about three weeks compared to 4-8 weeks needed before leveraging the synergy union between FRAGLER and Benchling **(Figure 2)**. This is a timeline no DNA synthesis vendor can equal, says Öling, particularly for long, complex sequences such as SarsCoV or Cas9 expression constructs.

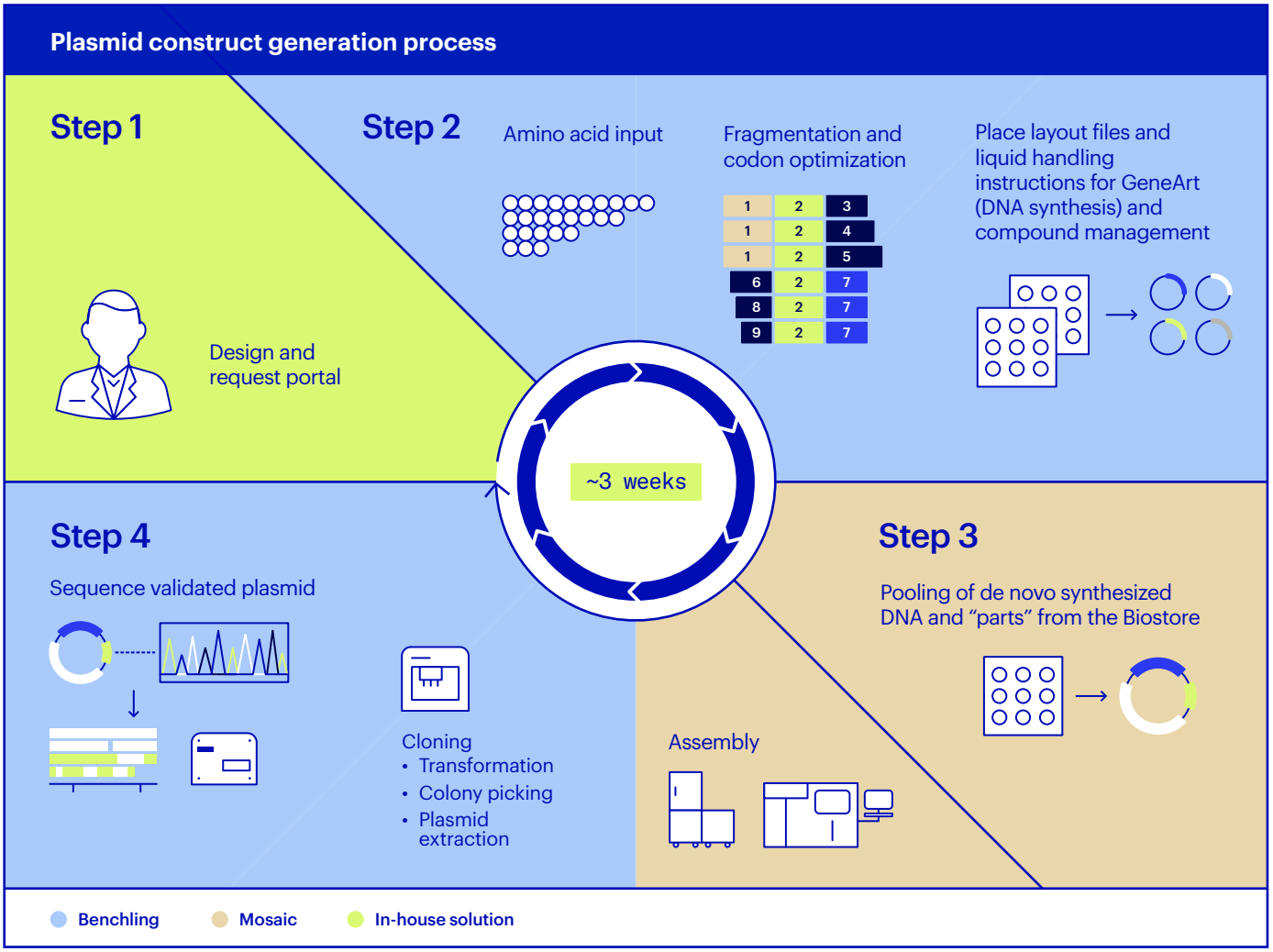


Figure 2. Automated plasmid generation workflow.

Step 1. Multiple Amino acid sequences are requested via an in-house design and request web-portal.

Step 2. Sequences are bioinformatically processed (FRAGLER) and synthesized.

Step 3. In-house DNA fragments from the Biostore are combined with de novo synthesized fragments and assembled using an ECHO 655T integrated on an Access system.

Step 4. The assembly mixture is cloned and single colony-derived plasmids are extracted using Biomek i7 with an integrated colony picker. Plasmids are validated by Sanger sequencing.

"The first time you run a campaign, FRAGLER will identify duplicate DNA sequences and help you save up to 50% of the cost of synthetic DNA synthesis by enabling fragment recycling," says Öling. "But for each campaign after that," he continues, "there is an estimated 10-30% additional cost reduction per iterative cycle, as more and more fragments are recycled and reused with each campaign. Eventually, most DNA fragments will be available in-house, meaning that you'll reach nearly zero cost for generation of new synthetic DNA constructs. This, of course, can only be fully realized by the automated sequence search in Benchling, as manual searches for duplicate sequences at this scale are impossible."

Original processes	With FRAGLER	With FRAGLER + Benchling automated sequence search:
Cycle 1: $1000	Cycle 1: $500	Cycle 1: $500
Cycle 2: $1000	Cycle 2: $500	Cycle 2: $400 (additional 20% cost reduction)
Cycle 3: $1000	Cycle 3: $500	Cycle 3: $320
Cycle 4: $1000	Cycle 4: $500	Cycle 4: $256 (additional 20% cost reduction)
		...Cycle N: ~$0 (additional 20% cost reduction)
Total: $4000	**Total: $2000** [50% total cost reduction from original]	**Total: $1476** [63% total cost reduction from original, 26% total cost reduction from FRAGLER-only]

Delivering a fully automated laboratory ecosystem

“Plasmids are critical for so many processes in early drug discovery,” says Öling. “They’re critical for protein production, cell line engineering, CAR-T cell therapy, viral vector generation, therapeutic genome editing, mRNA/Vaccine production and more.” The time it takes and the cost of producing them is a critical limiting factor. When it comes to automation combined with a fragmentation algorithm, AstraZeneca is the only team doing it at this scale. Benchling helped make that possible by streamlining processes, eliminating time-consuming manual interrogations, adding structure to sample management, and bringing standardization to their entire workflow.

While Öling’s team has used Benchling for DNA construct assembly, he says it’s easy to imagine using Benchling for other design, discovery, or workflow handoffs. The most time-consuming part was configuring the whole workflow, says Öling, adding that the learning curve for Benchling was very quick. “There are several appealing things about Benchling — it’s cloud based, intuitive, has good DNA assembly tools, and is super easy to get into,” he says. Other teams at AstraZeneca have expressed interest in leveraging Benchling for their own processes, too. Öling is working with Benchling to implement a way to track timelines and cost savings from their automated DNA construct assembly platform directly within Benchling.