DECEMBER 2024 Volume 1, Issue 10

NUCLEIC ACID INSIGHTS

SPOTLIGHT ON
Plasmid DNA

Guest Editor
Michelle Berg, JURA Bio, Inc.

NUCLEIC ACID INSIGHTS Volume 1, Issue 10

CONTENTS

SPOTLIGHT Plasmid DNA

NUCLEIC ACID INSIGHTS

PLASMID DNA

SPOTLIGHT

Connecting innovations to great gains for the ultimate stakeholder: the patient

Michelle Berg
JURA Bio, Inc.

"I invite you to join me in welcoming 2025 with great optimism and a devotion to creating meaningful change..."

FOREWORD

Nucleic Acid Insights 2024; 1(10), 379–381

DOI: 10.18609/nai.2024.047

Congratulations and bravo to the creators at BioInsights for completing the inaugural year of *Nucleic Acid Insights* and providing the go-to resource for this constantly evolving space. It has been my joy to have existed at the intersection of the manufacture of nucleic acids (plasmid and mRNA across all scales, quality, and applications) and advocacy for those impacted by rare and unsolved disease. It remains my willful responsibility

to further drive connectivity and mutual understanding from supply chain to patient. Therefore, I'm thrilled to introduce you to the final issue of 2024. The following commentary is through the lens of genetic medicine; however, discoveries and advancements will continue to have clear impact on additional applications.

Reviewing the submissions and interviews reminded me that there is a spectrum to innovation that can lead to broader success for the field and faster access for the ultimate stakeholders—patients. The great gains through leveraged technologies such as machine learning and artificial intelligence are both exciting and wonderful, and yet consistent, incremental improvements to existing art, processes, and reagents also deliver impact. Looking to the rare disease community as forced innovators, we must remain open to creative solutions that aren't readily apparent, become comfortable with the uncomfortable and unanticipated, and realize the benefits of forging into the complex together.

Within this final issue of 2024, we will contemplate these multiple facets of innovation for delivering on the promise that nucleic acid, particularly DNA, holds. Contributors to this issue focus on improving the accuracy, quality, safety, and reproducibility of DNA manufacture. One group is innovating upon how organizations engage and share, with an ultimate goal of establishing a pre-competitive and open-source series of standards and metrics. Another offers insights on their hybrid model utilizing the best of both worlds in outsourcing and do-it-yourself synthesis, while an additional interviewee shares about her team's informed and cost-conscious method for synthesis and multiplexed shots on goal at a scale that is challenging to conceive.

Surely, you will find something of interest and application to take away from this edition. As you navigate the content, I invite you to consider the following questions:

1. What drew you to this field initially and how can you honor this through innovation in the new year?

2. In your role and organization, what is your direct line to making lives better through the value you provide?

3. How can you set subsequent thinkers and doers up for success to iterate and advance?

Thank you for your attention to this issue and support for *Nucleic Acid Insights*. I am grateful for what the subscribers, contributors, and *Nucleic Acid Insights* staff will do to advance this field.

Whether you are a researcher, developer, manufacturer, or resource provider, there is always an opportunity to better understand those we hope to impact. I've found that, when done correctly, bringing patients more proximal to the work not only improves output, but helps people gain a greater sense of purpose in their work. With that, I invite you to join me in welcoming 2025 with great optimism and a devotion to creating meaningful change through the realization of genetic medicine!

BIOGRAPHY

MICHELLE BERG is currently applying her 25-plus years of experience and leadership in the biotechnology sector as an Expert in Residence (XIR) with Digitalis Commons, New York, NY, USA through which she advises ARPA-H, as a member of the ASGCT Patient Outreach Committee, and as a board member with JURA Bio, Emily Whitehead Foundation, Regenerative Medicine Minnesota, and North Dakota State University's Center for Entrepreneurship and Family Business. She served as President, GMP Nucleic Acids for Aldevron, building upon previous roles and expansion cycles within the organization.

During intermission with Aldevron, she held the role of Vice President, Patient Affairs and Community Engagement for Abeona Therapeutics, a company focused on developing novel gene and cell therapy approaches for treatment of people affected by rare diseases. Berg shares her passion as an author, speaker, and advocate for those living with rare disease, enabler of women in science, and driver of change for those with unmet needs through her work in non-profit and start-up organizations in the gene and cell therapy space. She received her BS in Biotechnology at North Dakota State University, Fargo, ND, USA.

AFFILIATIONS

MICHELLE BERG
Board of Directors,
JURA Bio, Inc.,
Boston, MA, USA

AUTHORSHIP & CONFLICT OF INTEREST

Contributions: The named author takes responsibility for the integrity of the work as a whole, and has given their approval for this version to be published.

Acknowledgements: None.

Disclosure and potential conflicts of interest: Michelle Berg reports personal fees from Digitalis Commons, JURA Bio, and Biomed Protection ND.

Funding declaration: The author received no financial support for the research, authorship and/or publication of this article.

ARTICLE & COPYRIGHT INFORMATION

Article source: Invited.

Revised manuscript received: Jan 7, 2025; **Publication date:** Jan 14, 2025.

NUCLEIC ACID INSIGHTS

PLASMID DNA

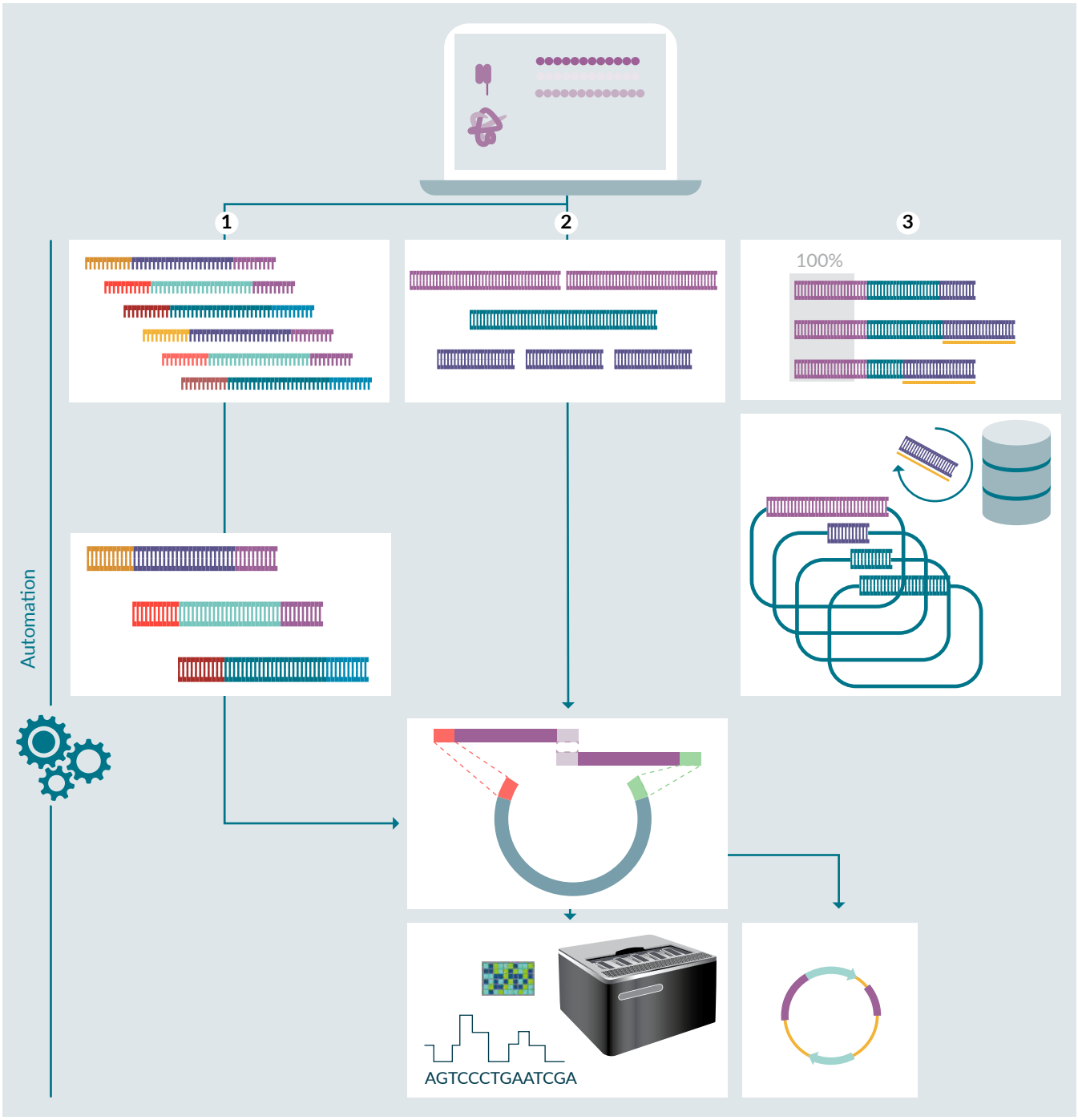
SPOTLIGHT

EXPERT INSIGHT

Automated plasmid assembly: a perspective from early drug discovery

Sanuja Ahammu, Jordi Chi, Tora Edström, and David Öling

Plasmid reagent generation is often both the first and the rate limiting step in early drug discovery projects. At AstraZeneca, we have sought to accelerate the plasmid DNA generation process by integrating laboratory automation and *in silico* processes from digital sequence to clonal plasmid construct. These processes are tracked in a custom-built application, called the Construct Request Portal (CRP), which also hosts various bioinformatic applications such as bespoke codon optimization algorithms and analysis of sequencing data by Oxford Nanopore Technologies (ONT). Drug discovery activities are diverse and may range from requiring a single cell-line engineering construct to variant libraries containing thousands of plasmid variants. Ultimately, the throughput, length, and complexity of plasmid constructs will trigger different DNA fragment formats and assembly pipelines. Herein, we will review the field and discuss advantages and disadvantages of 'do-it-yourself' (DIY) plasmid DNA assembly workflows in the perspective of workflow requirements and rapid sourcing of synthetic DNA.



Synthetic plasmid DNA assembly workflows. Amino acids are entered into the Construct request portal (CRP) which codon optimizes and fragments the DNA sequences in sizes suitable for one of three cloning workflows: **(1)** Gene assembly from arrayed or pooled oligos; **(2)** Gene assembly from dsDNA fragments; **(3)** Gene assembly from clonal DNA fragments.

Nucleic Acid Insights 2024; 1(10), 363–370

DOI: 10.18609/nai.2024.045

DNA is an integral tool for research in many different fields such as chemistry, biology, and drug discovery. Since the discovery and understanding of its structure a few decades ago, the field has been rapidly transformed by the development of new technologies that made it possible to read and write DNA. Innovation in next-generation sequencing (NGS) have allowed reading DNA at a much faster speed and greater capacity, revolutionizing the fields of genomics and transcriptomics. On the other hand, developments in DNA synthesis have opened a myriad of new possibilities when it comes to understanding living systems and engineering biology.

DNA SYNTHESIS

Chemical synthesis of DNA relies on the phosphoramidite method, which was developed in the 1980s and has been the cornerstone of DNA synthesis for decades [1]. This chemical synthesis approach involves the sequential addition of 5′-protected dimethoxytril (DMT) nucleotide phosphoramidites to a growing DNA chain, enabling the production of custom DNA sequences with high fidelity. This method was used by Applied Biosystems to develop the first automated DNA synthesizer in the 1980s [2,3]. The technology has matured over the years and current technologies can now synthesize up to a million oligonucleotides in parallel. Despite its widespread use, the method has some limitations. There are currently several oligo synthesis methods which are reviewed elsewhere [4–6]. The main challenge is the generation of >300 nucleotide DNA due to the elongation cycle efficiency. In addition, imperfect sequences are frequently generated containing truncations, indels or substitutions [7,8]. Such imperfections can be corrected enzymatically (mutS) and removed by an additional size selection step [9,10]. Chemical synthesis uses large quantities of organic solvents, and the reaction used to remove the 5′-DMT protecting group can cause depurination, resulting in a decrease in yield and purity of the oligonucleotide.

Chemical synthesis remains the golden standard today where leading companies such as Twist, GeneArt, GenScript, Genewiz, Eurofins, and IDT are competing based on price, speed, and customization. More recently, companies such as Elegen have emerged, specializing in long (<7 kb), high quality dsDNA fragments. Similarly, Ribbon Biolabs has produced >10 kb dsDNA using pools of 8–26:nt oligos (with complements) providing all building blocks required for gene assembly. In general, the process limitations in length and complexity that arise from chemical synthesis can be mitigated by assembling shorter DNA fragments, or codon optimization to adjust GC content and remove repetitions or homopolymers [11].

As the knowledge of the regulatory sequences in our DNA increases, demands for non-coding DNA sequences also grow. These sequences can usually not be altered, and they often have repetitive regions, or regions with very high GC or TA content—sequences that present difficulties to traditional chemical synthesis. Enzymatic synthesis offers an alternative to chemical synthesis using terminal deoxynucleotidyl transferase (TdT), which is a template independent polymerase that attaches deoxynucleotide triphosphates (dNTPs) to the 3′ end of ssDNA. It was described first in 1959 and is now widely explored for commercial template-independent enzymatic oligonucleotide synthesis (TiEOS) methods [12]. This method uses NTPs modified with a protected group at 3′ that interrupts synthesis to ensure that only one single nucleotide is added per reaction step. The protected group is removed at the end of each cycle to allow incorporation of the next desired nucleotide. Since TdT has higher affinity towards natural nucleotides (nt:s), it has been re-engineered to accommodate the protected nt:s in its active site [13]. In addition, TdT in its native form is error prone and biased towards G/C incorporation, requiring mutations of the enzyme or other

workarounds to reach a high-yield, low-error process. An excellent and extensive review of literature describing TdT engineering has recently been published elsewhere [5].

There are now many vendors offering enzymatic DNA synthesis like Ansa Biotech, Molecular Assemblies, NunaBio, and DNAscript, which all use modified versions of TdT to deliver sequences including extreme GC content, secondary structures and repetitive sequences. This technology is still young, and as the development progresses, longer sequences may become available. Recently, Ansa Biotechnologies generated a 1005:nt long oligonucleotide sequence by template independent synthesis. It remains to be seen if other enzymes such as poly-U-polymerase will be able to compete with TdT.

DNA ASSEMBLY

Regardless of DNA source, oligos normally need to be assembled into larger DNA fragments for the downstream application. The most common assemblies are based on DNA polymerase assembly such as PCA, Gibson assembly, or on TypeIIS restriction enzymes such as Golden Gate (GG) assembly. For all cases, fragments up to 100 kb have been assembled. Assembly reactions often rely on cherry picking and automated liquid handling processes, which are amenable to automation and are critical for process improvement and accelerating timelines. Indeed, a few companies are homing in on 'on-site' DNA foundries, or providing reagents, automation, and IT infrastructure to set up an end-to-end process allowing researchers to gain control of the whole supply chain.

DNA Script's SYNTAX system offer a benchtop system that relies on a reagent kit and TdT for oligo synthesis. Currently, the SYNTAX is limited by throughput and oligo lengths up to 120 bp. However, additional automation may be integrated to enable plasmid assembly and 'on-site' production of genes. Similarly, the Gibson SOLA platform from Telesis promise to enhance DNA and mRNA workflows by integrating the technology with their BioXP3250 and 9600 platforms. Camena, a Cambridge, UK-based company, are instead focusing on integrating their reagent kit (gSynth) and bioinformatics pipeline with common liquid handlers to enable complex gene assembly. Another Cambridge-based company in this space, Evonetix, has partnered up with ADI to develop a benchtop semiconductor DNA synthesizer.

TRANSLATION INSIGHT

When working with high-throughput plasmid cloning, sample tracking and automation are key to efficiency. Rather than working with multiple different softwares, we have incorporated our workflow in the Construct Request Portal (CRP). Briefly, a request is submitted as an amino acid or DNA sequence together with information on plasmid backbone and host organism for expression. The CRP provides a unique ID to each requested construct to facilitate tracking. Various bioinformatic processes are triggered such as codon optimization and fragmentation, described in more detail below. The CRP will output DNA sequences in a format that is compatible with most synthetic DNA vendors.

Outsourcing of gene synthesis is a viable option but it is not the most efficient way to accelerate the many iterative cycles of drug discovery projects. While there are some options available for benchtop synthesizers, each has its own drawbacks. For example, the Äkta synthesizer with its seven column reactors is limited in throughput, while the kilobaser produces short oligos unfit for gene assembly. As a workaround, we have adopted a partial DIY workflow where we rely on outsourced DNA but control the assembly reactions internally with bioinformatics and automation.

The type of synthetic DNA product will be determined depending on the downstream application. We have listed pros and cons of the most common synthetic DNA products

TABLE 1

Features pros and cons of the most common synthetic DNA products for gene assembly.

	Oligo pools	DsDNA linear	Cloned DNA
Speed (for 2 kb gene)	Fast (5-10)	Fast (5–10)	Medium (8–10)
Synthesis success rate (for 2 kb gene)	Medium	Medium	High
Synthesis cost (for 2 kb gene)	Low	Medium	High
Sequence complexity	Limited variation	Limited variation	Flexible
Sequence accuracy	~85%	>85%	100%
Sequence length	Short (~350 bp)	3–7 kb*	>7 kb

for gene assembly in Table 1. Codon optimized constructs for protein expression are amenable to synthesis as most complexities such as repetitions, homopolymers and high GC content can be removed by controlling the codon usage. On the other hand of the spectrum are plasmids for long transcripts for IVT or synthesis of other genomic DNA sequences such as homology arms for targeted genome integration. Similarly, enhancer or promoter libraries have little flexibility for sequence alterations. Such long complex DNA constructs have been challenging to acquire in a rapid, reliable, and cost-efficient manner, if even possible to synthesize. While many vendors are improving on production of long and complex genes, delays or cancellations are frequent.

We have addressed the current challenges in synthetic DNA manufacturing by adopting different workarounds. The key part of the process is to generate and outsource production of multiple shorter fragments, single-stranded (ss) or double-stranded (ds) DNA and assemble in our own lab. We have geared the pipeline towards high reliability, cost-efficiency, and sustainability. At the core of this process is a DNA fragmentation and recycling application, which relies on multiple sequence alignment and generation of 0.3–1 kb fragments. We have observed that this is a fragment range where most vendors are highly successful at delivering DNA in a timely manner. Identical amino acid (AA) sequences such as domains are identified, and an assembly matrix is generated with the purpose to only synthesize identical sequences once. Prior to fragmentation, an algorithmic database search is performed on the AA level to maximize the value of already synthesized DNA by recycling it [14]. This process is facilitated by species-specific meta data tags since identical AA sequences may differ on the DNA level due to alternative codon usage. The pipeline benefits from clonal DNA fragments that are stored in an automated freezer (Biostore) and made available via acoustic tubes and ECHO plating upon request. The pipeline can also be run with non-clonal dsDNA fragments readily available from most vendors. Clonal fragments are more expensive but can on the other hand be stored stably enabling recycling and facilitate re-supply. In fact, we have seen a 30% reduction in cost by using FRAGLER and an additional 15% reduction by reusing commonly used affinity tags, promoters, and fluorescent or luminescent reporters. For single site mutagenesis library type campaigns with highly similar AA sequences, up to 90% of nucleotides can be synthesized once and recycled, enabling significantly higher cost reduction.

To accelerate and scale our process, we implemented another workflow based on gene assembly from oligo pools. This workflow has been inspired by a recent publication from where the team addressed the challenge of costly gene assembly at scale. They used a combination of barcoding, long 300-mer oligo pools, and a robust GG assembly approach to generate hundreds of constructs from one oligo pool [15]. This is a powerful de-risking approach as multiple designs of oligo pools can be generated to maximize

success of synthesis and ensure full coverage of ordered oligos. We have built on this workflow by developing a graphing fragmentation solution to choose optimal 4 bp overhangs for fragmentation and GG assembly. In addition, we have implemented an AI-powered sequence generator that will add optimal buffering sequences to create oligos of equal length, facilitating the downstream size selection process. The sequences are optimized for lack of restriction enzyme recognition sequences and GC content, and are devoid of repeat regions.

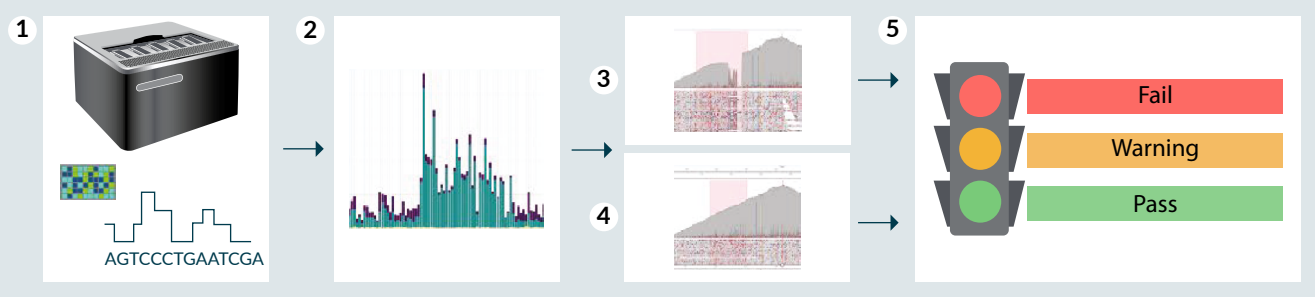
Sequence imperfect dsDNA fragments as well as oligo pools increase the screening and sequencing burden of DIY workflows. Sanger sequencing has long been the gold standard, providing well over 1 kb quality reads. However, constructs required for cell-line engineering and gene editing, such as Cas enzymes, are often >4 kb sequences that require multiple sequencing reactions to cover the whole insert. Long sequences in combination with the need to screen more clones may lead to sample sizes that rapidly get out of hand. To this end, we have enhanced our sequencing capacity by implementing an Oxford Nanopore Technology (ONT) platform similar to Circuit-seq [16]. On-site sequencing reduced logistical issues with lost samples and enabled direct analysis following plasmid extraction. Importantly, it increased quality by providing long reads to cover the full plasmid. The rapid barcoding kit by ONT (SQK-RBK114.96) allowed us to demultiplex 96 plasmids in a single experiment and we could generate more than 100x sequencing depth on a Flongle flow cell. Apart from the advantages of real-time data generation, long read length, and throughput with ONT, we also acknowledge its error rate and susceptibility to artefacts. While NGS or Sanger may be a better choice when it comes to higher accuracy and reliability, their shorter read length, pre-processing steps, and low throughput (Sanger) make them less desirable for longer inserts, or when sequencing of the full plasmid is required. Moreover, ONT is amenable to automation and enables primer-free sample preparation in a cost-effective manner. Importantly, the seamless integration with our bioinformatic pipeline and real-time analysis reduces the turnaround time.

Adapting the workflow for inexpensive Flongle flow cells made the process highly economical. The bespoke bioinformatic pipeline (Figure 1) performs (1) sequence analysis, (1–2) outputs read quality, (3–4) coverage images, and (5) aligns the generated reads with the reference summarized in an intuitive read-level report. This eliminates the need for programming skills when analyzing ONT sequenced data.

Both chemically or enzymatically generated oligos and genes will likely prevail as

FIGURE 1

A bespoke bioinformatics pipeline for whole plasmid QC.



(1) Local basecalling and generation of raw data on Oxford GridIon. **(2)** Processing of raw data and mapping of reads to plasmid and host DNA to estimate the proportion of host contamination. (3) Coverage plots from Integrated Genome Viewer (IGV) of a failed **(3)** or passed alignment **(4)**. The red region highlights the insert on the plasmid **(5)**. Based on this reference-based read assembly, the pipeline gives a PASS/FAIL output in the automated QC report.

they have different advantages and disadvantages. Currently, the main hurdles seem to be related to logistics and process scheduling related issues. Moving DNA synthesis capability into the hands of scientists will at the minimum enable quick access to oligos for individual labs to optimize their own gene assembly processes. However, automated workflows require lengthy optimization cycles such as bacterial plating and colony picking using liquid handling robots. Thus, a major limitation of DIY workflows is that they rely heavily on costly lab automation and IT investments so a certain throughput will have to be met for DIY workflows to be economically sound.

REFERENCES

1. Beaucage SL, Caruthers MH. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* 1981; 22, 1859–1862.

2. Caruthers MH. The chemical synthesis of DNA/RNA: our gift to science. *J. Biol. Chem.* 2013; 288, 1420–1427.

3. Matteucci MD, Caruthers MH. Synthesis of deoxyoligonucleotides on a polymer support 1981. *Biotechnology* 1992; 24, 92–98.

4. Ma Y, Zhang Z, Jia B, *et al.* Automated high-throughput DNA synthesis and assembly. *Heliyon* 2024; 10, e26967.

5. Hoose A, Vellacott R, Storch M, *et al.* DNA synthesis technologies to close the gene writing gap. *Nat. Rev. Chem.* 2023; 7, 144–161.

6. Hughes RA, Ellington AD. Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* 2017; 9, a023812.

7. Kosuri S, Eroshenko N, Leproust EM, *et al.* Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* 2010; 28, 1295–1299.

8. Ma S, Saaem I, Tian J. Error correction in gene synthesis technology. *Trends Biotechnol.* 2012; 30, 147–154.

9. Smith J, Modrich P. Removal of polymerase-produced mutant sequences from PCR products. *Proc. Natl Acad. Sci. USA* 1997; 94, 6847–6850.

10. Binkowski BF, Richmond KE, Kaysen J, *et al.* Correcting errors in synthetic DNA through consensus shuffling. *Nucleic Acids Res.* 2005; 33, e55.

11. Tang NC, Chilkoti A. Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nat. Mater.* 2016; 15, 419–424.

12. Jensen MA, Davis RW. Template-independent enzymatic oligonucleotide synthesis (TiEOS): its history, prospects, and challenges. *Biochemistry* 2018; 57, 1821–1832.

13. Sarac I, Hollenstein M. Terminal deoxynucleotidyl transferase in the synthesis and modification of nucleic acids. *Chembiochem* 2019; 20, 860–871.

14. Oling D, Lan-Chow-Wing O, Martella A, *et al.* FRAGLER: a fragment recycler application enabling rapid and scalable module DNA assembly. *ACS Synth. Biol.* 2022; 11, 2229–2237.

15. Lund S, Potapov V, Johnson SR, *et al.* Highly parallelized construction of DNA from low-cost oligonucleotide mixtures using data-optimized assembly design and golden gate. *ACS Synth. Biol.* 2024; 13, 745–751.

16. Emiliani FE, Hsu I, McKenna A. Multiplexed assembly and annotation of synthetic biology constructs using long-read nanopore sequencing. *ACS Synth. Biol.* 2022; 11, 2238–2246.

AFFILIATIONS

Sanuja Ahammu
Discovery Sciences,
Biopharmaceuticals R&D,
AstraZeneca,
Gothenburg, Sweden

Jordi Chi
Discovery Sciences,
Biopharmaceuticals R&D,
AstraZeneca,
Gothenburg, Sweden

Tora Edström
Discovery Sciences,
Biopharmaceuticals R&D,
AstraZeneca,
Gothenburg, Sweden

David Öling
Discovery Sciences,
Biopharmaceuticals R&D,
AstraZeneca,
Gothenburg, Sweden
(Author for correspondence)
david.oling@astrazeneca.com

AUTHORSHIP & CONFLICT OF INTEREST

Contributions: Sanuja Ahammu, Jordi Chi, and Tora Edström contributed equally to this article. The authors take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.
Acknowledgements: None.
Disclosure and potential conflicts of interest: The authors have no conflicts of interest.
Funding declaration: The authors received no financial support for the research, authorship and/or publication of this article.

ARTICLE & COPYRIGHT INFORMATION

Article source: Invited; externally peer reviewed.
Submitted for peer review: Oct 6, 2024; **Revised manuscript received:** Nov 24, 2024;
Publication date: Dec 19, 2024.

NUCLEIC ACID INSIGHTS

PLASMID DNA

SPOTLIGHT

INTERVIEW

What removing the DNA synthesis bottleneck means for the future of drug discovery and development

Elizabeth Wood, CEO of Jura Bio, tells **David McCall**, Senior Editor, BioInsights, about the sheer scale of the opportunity for biopharma presented by combined innovations in DNA synthesis and machine learning.

Nucleic Acid Insights 2024; 1(10), 383–389

DOI: 10.18609/nai.2024.048

Q What are you working on right now?

EW: I am a biophysicist and a theoretical computational biologist by training. I spent years and years chasing the problem of how we could use computers to design medicines that cut the cost and duration of the R&D process as well as the potential for dangerous off-target effects or other safety concerns.

I continue to work on these challenges today. I'm very lucky to be in the position as CEO of JURA to be able to rewrite and rebuild biologics drug discovery and development as I would want it to be done.

Q Tell us more about JURA Bio's platform and approach.

EW: JURA got its start in 2017 as a cell therapy company focused on autoimmune diseases. I call this our 'misspent youth'! We were fresh out of the lab, very eager and optimistic that it was just a matter of building a company and everything else would follow. Then, of course, we met all the issues that one confronts as one tries to take a drug through the development phases—in our case, just getting to Phase 1 trials was a huge challenge. There were regulatory questions to address and difficult developmental choices along the path as to which therapeutic candidate to prioritize. It was a wonderful and important experience for us. I now think of it as the lived and worked example that allowed us to figure out what the pain points really were in the drug development process. Without having had that experience of trying to take a new drug to clinic, we would have been working on entirely the wrong problems...But by 2019, armed with that insight, we had noticed a few juicy problems along the discovery and development path that would really lend themselves well to modern machine learning.

We talk a lot about the biotech renaissance that has happened recently with the likes of CRISPR genome editing and with gene therapy in general—things that make this a really special time to be in biotechnology. However, modern machine learning has also seen a renaissance in biotech application, albeit one that has gone somewhat under the radar, comparatively speaking. It extends well beyond large language models and other machine learning models—we have seen profound advances in technical mathematics, for example. It is really modern inference writ large.

Going back to JURA in 2019, we knew where the problems were where we could make order of magnitude gains that would help us in developing a next wave of therapeutic candidates. We started knocking off those problems one by one, and the first thing we focused on was the phenomenon of ChatGPT and other generative AI tools basically redefining what was possible in real-time.

One could be forgiven for wondering why we are not witnessing the same degree of hype in the field of machine learning-driven biology, where it really counts in terms of getting drugs to the clinic. The answer is that modern biology and biotechnology suffer from a fundamental lack of data at scale. We think we have a lot of data—we as biologists know how much effort has gone into each experiment. We think of tens of thousands of data points generated by each high-throughput experiment, and we know that is a huge amount of data compared to what was available to previous generations. But at the same time, it is still a truly minuscule amount—perhaps 0.000001%—of the data that is available to the likes of Google, Facebook, and OpenAI every single day. Just by clicking on the internet, you get a huge amount of data. And it is data of a special type: it's distributed, but it still provides the likes of ChatGPT with

"...we recognized the need to rigorously test our model at scale, which required overcoming the synthesis bottleneck. Our solution was variational synthesis..."

instantaneous, large-scale feedback on their predictions. For example, every time ChatGPT spits out something out, a user reads it and gives it a thumbs-up or a thumbs-down review. They can therefore test and refine their models in this way. However, when we think about large-scale data in biology—perhaps we have done a cell atlas—we have this large, fixed and static data set. And when we go to train our models, we find ourselves deeply bottlenecked by our inability to test our predictions. We don't have the chance to get in front of 6 billion users!

Now, bringing DNA synthesis into the conversation, our therapeutic candidate pipeline is fundamentally bottlenecked by the synthesis bottleneck. Consider our ability to build and test the therapeutic candidates we have predicted, let's say in a functional screen. We are spending cents per base pair, and maybe we are looking at 10,000 samples from our machine learning model that we are going to test. Unfortunately, most antibody screening problems don't yield a single hit in 10,000. So, maybe we will spend all this money, we will test everything, and we will get one or two predictions at most with which to try to train our models. That is a very hard way to train a model—it certainly isn't 6 billion data points a second. It will slow everyone down.

At JURA, we recognized the need to rigorously test our model at scale, which required overcoming the synthesis bottleneck. Our solution was variational synthesis—a novel process of physical computation that shifts computational workloads from the silicon chips in traditional computers to the silicon chips used in gene synthesis. This breakthrough enables us to construct trillions of designs for the cost of just a few hundred using conventional gene synthesis technologies.

The Bayesian-optimized Design of Experiments approach we have introduced at JURA is all about keeping scientists in the loop or, to put it another way, allowing AI to hold the hand of our scientists and help them make informed decisions. For instance, to help them identify which 500 leads from 100,000 hits should be progressed to the next round of experiments, and then once that next round of experiments has been conducted, to help them decide whether to move forward solely with a few of those 500 leads, or to dip back into the original pool of 100,000 and pick some new ones.

Q Stepping back for a moment, can you reflect on the journey that culminated in your current work in applying generative AI to DNA synthesis, specifically?

EW: We have been through a lot. We have been through some very unique times for biotech, whether it was the heyday of 2017/18 when we got our start, or the COVID-19 pandemic period, or the present reality, which is very asset-focused and traditional. Even as

"...we are building the infrastructure for a future where some of the business models that are available to our tech colleagues become available to us in biotech."

machine learning is exploding in our industry, there is predominantly a very asset-focused response to it, and the unknowns around what it can do for us. Additionally, I am a first-time founder and a woman founder, and I believe that currently, only 2% of venture capital reaches women-founded and women-led companies.

We have never had a tremendous amount of capital available to us, even in the boom times of a few years ago. And to be honest, that has been our strength: throughout these periods, we have continued to slowly and steadily grow our team. We never had to react to market conditions because we were solely concerned with hiring exactly who we needed when we needed them, and building exactly as we needed to at the time. I honestly think that in the absence of that profound lack of access to capital, we would never have invented variational synthesis. Had we had the capital, the other option would have been to just spend a lot of money making the constructs and testing them. If we had had hundreds of millions of dollars, maybe we wouldn't have invested all of the people time, all of the theory time in this wild and profound goal of bringing down the cost of gene synthesis a trillion-fold. That is really a question you only ask out of desperation! If you have access to lots of capital and all the pressures that go with it, you don't tend to step back and ask questions like 'how do we profoundly change the way that we sample from a generative model?'

We are very optimistic for the future. We have adopted an asset-focused pathway forward where we're going after hard-to-drug intracellular targets with antibodies. We have generated some beautiful results, as one should when one is able to build, test, and screen hundreds of millions of AI-designed examples. It is natural that it should work, but it still makes you feel good when it does.

I do feel like we are building the infrastructure for a future where some of the business models that are available to our tech colleagues become available to us in biotech. So, right now we are an asset-focused biotech, but I would love to think that in 5–10 years' time we will be an enablement company—a ScaleAI for biotech, where we provide the data, the inference, and all the tools to support others who are great at getting therapies through Phase 3 clinical trials.

Q What is your vision for the harnessing of AI in the DNA synthesis space—firstly, in terms of the specific applications in which you expect to see its initial impact?

EW: In terms of specific applications, we have turned to designing antibodies against extremely hard targets. A lot of people think about antibody screens as being a solved problem,

but of course, there are a number of targets for which it is not solved. These are the G-protein-coupled receptors (GPCRs) of this world, or any time that agonism is involved. There are questions about mechanism of action, relating to things like ion channels, or intracellular targets. However, undoing the synthesis bottleneck means that you can directly build and test anything you want in both the DNA and protein spaces. You basically move your limitations downstream to your assay capacity.

At JURA, we have been seeking partners for a project to design and build CRISPR guides to every single nucleotide in the human genome. This can be built, but it has always been something that would have been so arduous and expensive that nobody really thought seriously about doing it. Now, thanks to variational synthesis reducing costs so dramatically, this suddenly becomes a tens of thousands of dollars question instead of a billions of dollars question. The same can be said for any number of large search space problems, such as peptides. The potential is just so exciting to me, and I hope it will be to others, too.

Q Looking further ahead, what can the convergence of AI and DNA synthesis ultimately deliver for drug discovery and development?

EW: In a few words, modern-day machine learning-scale data from biology. Right now, some of the biggest gains from and most impressive examples of machine learning in application are things that have been trained on literally quadrillions of examples.

I think the biotech industry is in for a rude awakening. Looking at the drug discovery process and the size of the data sets we can currently produce, even when we talk about examples of federated learning where multiple companies pool their data, the difference in scale is staggering. If we were to say that over the past 20 years worldwide, perhaps 30,000 therapeutic programs have reached any sort of clinical phase, and for each one of those, let's say (extremely optimistically) that 10,000 candidates were built and tested. That would total 3 trillion constructs built and tested across all of pharma, biotech, and academia over the course of the past two decades—a number that is very likely far in excess of reality. Even so, the cutting edge of modern machine learning is capable of processing that total data set literally thousands of times per second. It just isn't the scale of data that modern machine learning is built on.

The central motivating vision that I have for variational synthesis is to unlock biological data at a scale that is at least within touching distance of modern machine learning data. There are quadrillions of data points around which adverts we look at, where we click, and what we search for online. Let us allow ourselves to deliver that scale of data for biology, and specifically, for the drug development process. Now that we have overcome the synthesis bottleneck, we as a field collectively face another challenge, which is the high-throughput, high-quality assay development process. This is because we have never been in a position before where our high-throughput assays needed to go beyond one in a million hits, but now we will need to screen against a quadrillion-scale library where each and every one is a hit. We should be asking a lot more of our high-throughput assays: not just 'is it a hit' but does it have off-target hits?

What is its solubility? Any further downstream developability questions that we want to ask, that we previously didn't dare ask of our hits because they were so rare and so precious. Now, in this deeply hit-rich regime, we can start asking more and demanding more, and that will serve to make our machine learning so much smarter. That is what will enable us to one day achieve modern-day machine learning-scale data from biology.

The repercussions for healthcare will be profound. You might be faced with a cancer that has specific expression profile. Let us enable ourselves to come up with an intervention that is going to be safe and effective the first time. Maybe it's a personalized or precision medicine, but whatever it is, we can get it to you in a timescale that matters to you and your family.

Q Can you sum up one or two key goals or milestones that you have for your work over the foreseeable future?

EW: Even though there's a renewed focus on and interest in machine learning, we know that biopharma tends to be cautious and slow to adopt new processes. The pathway for JURA is to get to Phase 1 with an antibody or another biologic against a high-value intracellular target for oncology. I think that by the time we reach that stage, collectively, biopharma is going to be realizing that there really is something to all of this. We will have succeeded along the path that really differentiates great ideas from executable ones.

That will be our most meaningful milestone moving forward. I certainly hope to be working on business development, too, either creating development candidates for partners, or eventually, even datasets. But it might be that this absolute proof is needed in order to create a real shift in mindset in terms of what is possible.

BIOGRAPHY

ELIZABETH WOOD is the founder and CEO of JURA Bio, Inc., Boston, MA, USA, a biotechnology company focused on developing innovative therapies, primarily in the field of immune modulation and cell therapies. Their approach combines cutting-edge advancements in synthetic biology, bioinformatics, and immunotherapy to address unmet medical needs, especially in oncology and autoimmune diseases. JURA Bio's mission is to develop next-generation therapies that push the boundaries of personalized medicine, using cutting-edge biotechnological tools to improve patient outcomes across multiple disease areas. Before starting JURA, Wood was a post-doc in the lab of Adam Cohen at Harvard's Department of Chemistry and Physics. She began her PhD with Angela Belcher and Markus Buehler at MIT, Cambridge, MA, USA and finished it under the supervision of Claus Helix-Nielsen in the Departments of Physics and Civil and Environmental Engineering at The Technical University of Denmark, Copenhagen, Denmark.

AFFILIATION

Elizabeth Wood
Founder and CEO,
JURA Bio, Inc.,
Boston, MA, USA

AUTHORSHIP & CONFLICT OF INTEREST

Contributions: The named author takes responsibility for the integrity of the work as a whole, and has given their approval for this version to be published.

Acknowledgements: None.

Disclosure and potential conflicts of interest: The author has no conflicts of interest.

Funding declaration: The author received no financial support for the research, authorship and/or publication of this article.

Article source: Invited.

Interview conducted: Nov 21, 2024; **Revised manuscript received:** Jan 13, 2025;

Publication date: Jan 14, 2025.

NUCLEIC ACID INSIGHTS

PLASMID DNA

SPOTLIGHT

INTERVIEW

Bringing the power of DNA synthesis to bear in the field of synthetic biology

Synthetic biology offers transformative potential in applications ranging from biosensor design to novel therapeutic R&D. **David McCall**, Senior Editor, BioInsights, speaks to **Paul Freemont**, Head of Structural and Synthetic Biology, Department of Infectious Disease, Imperial College London, about the recent advancements and future prospects in synthetic biology, particularly concerning the development of biofoundries and related work in scaling and automating DNA synthesis.

Nucleic Acid Insights 2024; 1(10), 371–378

DOI: 10.18609/nai.2024.046

Q What are you working on right now?

PF: Currently, my main interest is synthetic biology, also known as engineering biology in the UK. In essence, synthetic biology is a platform technology with nucleic acids at its core. A significant part of this concept involves building DNA constructs, pathways, circuits, and even genomes.

Specifically, I am focused on establishing biofoundries—fully automated, integrated infrastructures that supports the design-build-test-learn cycle in the context of synthetic biology.

In order to accelerate this process, we are developing automated protocols that leverage liquid handling automation and other advanced systems. These protocols enable the efficient screening of different DNA structures, which can be utilized in various spaces such as manufacturing, drug discovery, and biosensor design. For example, the entire manufacturing workflow of synthetic biologics can be streamlined using biofoundry facilities, allowing researchers, startups, and small companies to access the complete infrastructure required to take a project from DNA to its final application.

My laboratory at Imperial College London focuses on cell-free systems, which play a critical role in prototyping within synthetic biology. These systems are non-living extracts of cells that allow developers to execute or run DNA and other nucleic acid designs in a cell-free environment. Whether developing new enzymes, constructing pathways, or creating biosensors, the process begins by designing these components at the nucleic acid level and testing whether or not they work. The quickest method to test the functionality of nucleic acid constructs is to introduce them into a cell-free lysate system, which supports transcription and translation, leading to the production of proteins that hopefully behave as intended for a specific application. We use cell-free systems extensively for prototyping designs, discovering enzymes and natural products, and using them for 'bottom-up' approaches that involve the construction of synthetic cells.

Q You wrote about advances in DNA synthesis technologies some 2 years ago [1]. How has the field continued to evolve since then? Where do you see the cutting edge currently?

PF: Most DNA synthesis technologies rely on phosphoramidite chemistry, a method developed in the 1980s. Today, optimized versions of this chemistry allow for scalable manufacturing as well as parallel processing. However, regarding the enzymatic side of DNA synthesis, while progress is being made, the technology is still not capable of delivering long sequences.

Looking at the bigger picture, the companies producing nucleic acids are performing reasonably well, offering diverse product lines such as libraries of protein or CRISPR-Cas9 designs, large nucleic acid constructs, and gene-type constructs useful for both experimentation and discovery. The field has evolved from relatively simple genes and oligonucleotides to more complex products that align with the needs of high-throughput biology. These products support more effective exploration of both the design and mutational spaces in proteins.

Despite all these advancements, the field seems to have reached a somewhat static phase, though. It is unclear what the next major innovation in this space will be, beyond optimized methods to produce longer DNA sequences at a lower cost.

This said, there are some promising developments on the horizon, such as optimized chip-based approaches and novel techniques for oligonucleotide synthesis. These innovations might improve system efficiency and reduce errors. Currently, errors remain a significant challenge,

"...transitioning from lab-scale innovation to a commercially viable marketed product is an important challenge to address."

and quality control and assurance processes for large oligonucleotides are still time-consuming and costly.

Additionally, enzymatic DNA synthesis was a revolutionary development, and it remains an exciting area with substantial potential. Although it has not fully worked out as yet, there are significant opportunities in this space. However, transitioning from lab-scale innovation to a commercially viable marketed product is an important challenge to address.

From a financial perspective, the development of DNA synthesis technology has not yet led to the sharp price reductions we had hoped for. There are issues tied to the distributed nature of the industry, as well as reagent lock-in practices. These factors, while making sense commercially, do not contribute to reducing costs, although that will hopefully change in the future.

In summary, the contribution of nucleic acid developers to the synthetic biology field has been transformative. However, the potential for truly disruptive, innovative methodologies still lies ahead. Such breakthroughs could significantly lower costs and enable the synthesis of longer gene constructs.

Q Much of the wider field's attention continues to focus on RNA, but where specifically could DNA-based approaches and innovations continue to make headway and open up new opportunities for the nucleic acids space?

PF: My work remains deeply rooted in synthetic biology, where DNA acts as a programmable structural material. Instead of focusing on delivering nucleic acids directly, we aim to leverage the functionality they encode. This approach holds promise in areas such as enzyme replacement, sensing, and interfering. The field of self-replicating RNA vaccines is also particularly interesting. Although it is RNA-based, much of the design work still needs to be done at the DNA level.

In the context of DNA-based therapeutic applications, gene therapy stands out as a key area of interest. The methods used to deliver therapeutics are also evolving, and the real challenge for the field as a whole lies in developing systems that achieve highly efficient uptake of nucleic acids. Whether we aim to replace genes or exogenously introduce new activity or functionality for specific disease contexts, the nucleic acid payload itself can be very effective. However, practical implementation, including delivery mechanisms and specific targeting, remains the most interesting and challenging aspect.

Additionally, utilizing nucleic acids as scaffolds has proven to be incredibly powerful. Their versatility comes from our growing ability to design diverse shapes, sizes, and complex

"..there have been some interesting developments in the field of toehold methodologies—specifically, RNA toeholds."

3D tertiary structures. The DNA origami space, for example, remains an exciting field with significant potential, whether in the context of therapeutic, biosensor, or delivery system development.

The area of modified nucleic acids is emerging as an exciting field of research. However, it is critical to consider the long-term stability of nucleic acid samples.

Lastly, CRISPR/Cas9 technologies also stand out as particularly powerful tools not only for altering cellular functionality, but also for their significant potential in therapeutic applications. The latter includes the ability to deliver specific inhibitory mechanisms or even to replace nucleic acids in order to address certain indications.

Q How are you leveraging the advancements in DNA design and synthesis in your own work?

PF: Most of our work focuses on the functions encoded within nucleic acids, which can be leveraged in various applications. In the drug discovery and development space, the applications include cell-based therapies, regenerative medicine, and pluripotent stem cell models for disease studies, as well as organoids for drug screening. Furthermore, delivery systems for targeting specific cells and delivering payloads, such as conjugated antibodies for cancer treatment, are key areas of development.

As I mentioned earlier, my laboratory is particularly focused on prototyping, developing high-throughput assays, and designing small biosensors. However, we are also focused on building various DNA designs and developing methods to rapidly assemble them. We then explore automated cell transformations and assess how the phenotypes are influenced by these DNA designs. On the 'bottom-up' side, we are delving into synthetic cells by constructing modular DNA components that can be run in cell extracts and provide functionality that mimics living cells. These three domains—DNA design and assembly, synthetic cell development, and biosensors—are at the core of our research, and we work to translate these technologies into practical applications spanning multiple sectors.

Beyond these core activities, there have been some interesting developments in the field of toehold methodologies—specifically, RNA toeholds. In these systems, the nucleic acid regulates the translation of components of a reporter gene based on the presence of a specific target nucleic acid. For example, one can use RNA toehold methods to detect pathogen nucleic acids by targeting specific sequences.

There have been other advancements in this realm such as CRISPR/Cas13a, which was pioneered by James Collins' laboratory [2]. This method involves detecting specific pathogen

nucleic acid sequences and activating Cas13a, which allows for non-discriminating cleavage of other nucleic acid molecules, revealing fluorescent FRET pairs that emit a signal.

These technologies are now being translated into real-world applications. Nucleic acid detection using technologies such as CRISPR/Cas is a rapidly evolving area and is leading the way towards the more routine use of pathogen detection. We have been exploring this area extensively ourselves, particularly in the context of the COVID-19 pandemic.

Q Your interests in synthetic biology extend far beyond applications in the field of life sciences—what other areas do you find particularly exciting at the moment and why?

PF: Synthetic biology is a platform technology with applications across various sectors. These include biomaterials, specialty chemicals (particularly replacing fossil-based chemicals), natural product and enzyme discovery, and alternative food systems such as cellular agriculture.

We co-founded a spin-out company with two colleagues from Imperial College London that focuses on computationally designed protein fibers. We have essentially opened up an entirely new design space by synthesizing proteins that do not necessarily exist in nature. By computationally designing these fibers, we can introduce specific functions or properties tailored to particular applications, such as textiles. With advancements such as AlphaFold 3 and large language models, protein design has become one of the most exciting areas in biotechnology today. The opportunities to design completely novel proteins with highly defined and desired functionalities are just beginning to open up.

Furthermore, the use of AI and machine learning (ML) in this space is truly exciting because of the long list of potential applications, including ligand and antibody design, drug discovery, and predicting protein-ligand and protein-protein interactions. The entire field is revolutionizing how we understand and design proteins for particular functions. For the first time in history, we are entering an entirely new design space where we can create functionality in proteins and unlock insights into very complex biological mechanisms as never before.

However, the success of this technology will depend on the data sets we capture, how we capture them, and the specific questions we aim to address. One key challenge is that these AI- and ML-driven technologies are dual-use, which introduces biosecurity concerns. It is crucial to figure out how to manage the security risks of such powerful technologies, especially when applied to redesigning biological systems.

Q What are your key goals or milestones for the foreseeable future?

PF: I am very keen to see how we can translate biotechnological advances into the real world. The more we can introduce biotechnology, bio-based products, and solutions into

practical applications, the more people will understand the power of the technology and the benefits it can bring. On a related note, I have been actively working in two areas—developing open-source standards and metrology, and scaling up and scaling out the biomanufacturing processes.

Regarding technical standards and metrics, I have been looking at these through the lens of translation and commercialization. For example, I recently completed an 18-month project, which resulted in a major report for a global initiative funded by Schmidt Sciences, Eric Schmidt's philanthropic organization in the USA [3]. As we delved deeper into this project, we discovered that within the biotechnology industry, there are only about 36 or 37 ISO-published standards. Other multi-billion-dollar industries, such as the aviation, car, or even food sectors, have hundreds or even thousands of standards. The question that concerned us is why there are so few standards associated with biotechnology? Although there are plenty of standards related to product safety or regulatory testing, such as those from the US FDA, there is a significant gap when it comes to the actual processes of biotechnology.

This made me realize that the biotechnology industry, especially in biomanufacturing, has many trade secrets in terms of capabilities and know-how, which is not ideal for encouraging expansion. I am working to push the development of open-source standards and metrics for biotechnology innovation and translation, which could be valuable for small companies and startups, especially those bringing new products to the market.

Ultimately, there are two types of biotech companies today—those that possess unique capabilities that no one else has, making them attractive to investors, and those that have unique products they want to take to market, whether it's a fragrance, a food supplement, or a novel therapeutic. The challenge for both types of company is scaling, which is particularly difficult for small startups.

Our focus is on opening up the scaling space and developing metrics and standards around it. While this work might seem unexciting to some, it is crucial nonetheless. This work also ties into the biofoundry space where we have the Global Biofoundry Alliance, comprising roughly 36 publicly funded biofoundries worldwide. I was lucky to serve as the founding Chair of this organization. The goal of these biofoundries is to build public, open, pre-competitive protocols that accelerate the entire process of developing DNA, constructing cells, and conducting design work in the upstream R&D context. This includes developing protocols, standards, open metrology, and software tools, as well as exploring how AI and ML can be integrated into the biofoundry context. I believe it will significantly accelerate the translation and development of biotechnology, particularly in the synthetic biology space. For biomanufacturing to scale globally and operate efficiently, new companies entering the field need more such opportunities.

REFERENCES

1. Hoose, A, Vellacott, R, Storch, M, *et al.* DNA synthesis technologies to close the gene writing gap. *Nat. Rev. Chem.* 2023; 7, 144–161.

2. Gootenberg JS, Abudayyeh OO, Lee JW, *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 2017; 356(6336), 438–442.

3. New report recommends standards, metrics for multi-trillion-dollar global bioeconomy. *Schmidt Sciences* May 7, 2024. https://www.schmidtsciences.org/new-report-recommends-standards-metrics-for-multi-trillion-dollar-global-bioeconomy.

BIOGRAPHY

PAUL FREEMONT is the co-founder of the Imperial College Centre for Synthetic Biology and Innovation and co-founder and co-director of the National UK Innovation and Knowledge Centre for Synthetic Biology (SynbiCITE; since 2013) and co-director of the London BioFoundry (since 2016) at Imperial College London, London, UK. He is also currently the Head of the Section of Structural and Synthetic Biology in the Department of Infectious Diseases at Imperial College London. He was previously the Head of the Division of Molecular Biosciences and Centre for Structural Biology having joined Imperial from Cancer Research UK London Research Institute (now known as the Crick Research Institute) where he was a Principle Investigator and Head of Group. His research interests are focused on developing synthetic biology foundational tools: automation, biofoundries, and cell-free systems for specific applications, including biosensing and metabolic engineering. He is author of over 300 scientific publications and is an elected member of European Molecular Biology Organisation and Fellow of the Royal Society of Biology, Royal Society of Chemistry and Royal Society of Medicine and is an Honorary Fellow of the Royal College of Art. He was a co-author of the British Government's UK Synthetic Biology Roadmap and was a recent member of the Ad Hoc Technical Expert Group (AHTEG) on synthetic biology for the United Nations Convention for Biological Diversity (UN-CBD). He is currently a council member of the US Engineering Biology Research Consortium and member of the UK Government's Office for Science and Technology Engineering Biology sub-group. He is also currently leading a US-funded task force on Engineering Biology Metrics and Technical Standards for Global Bioeconomy.

AFFILIATION

Paul Freemont
Head of Structural and Synthetic Biology,
Department of Infectious Disease,
Imperial College London,
London, UK

AUTHORSHIP & CONFLICT OF INTEREST

Contributions: The named author takes responsibility for the integrity of the work as a whole, and has given their approval for this version to be published.

Acknowledgements: None.

Disclosure and potential conflicts of interest: The author has no conflicts of interest.

Funding declaration: The author received no financial support for the research, authorship and/or publication of this article.

ARTICLE & COPYRIGHT INFORMATION

Article source: Invited.

Interview conducted: Nov 27, 2024; **Revised manuscript received:** Jan 3, 2025;

Publication date: Jan 9, 2025.

NUCLEIC ACID INSIGHTS

PLASMID DNA

SPOTLIGHT

INTERVIEW

Exploring alternatives to plasmid DNA for safer, higher quality gene therapy

David McCall, Senior Editor, BioInsights, speaks with **Nafiseh Nafissi**, Chief Technology Officer, Mediphage Bioceuticals, about the applications and advantages of ministring DNA, including as the basis for a novel gene therapy modality.

Nucleic Acid Insights 2024; 1(10), 357–362

DOI: 10.18609/nai.2024.044

Q What are you working on right now?

NN: At Mediphage, we are advancing a proprietary platform called ministring DNA (msDNA™), with the goal of creating safer, re-dosable, scalable gene therapies. We are applying msDNA across multiple domains, including as a starting material to enhance the quality and manufacturing yield of viral vectors such as AAV, and in mRNA production. Our main priority is using msDNA as an active pharmaceutical ingredient for gene supplementation in

gene therapy applications, or as donor DNA in gene editing applications—in particular, for both *ex vivo* and *in vivo* cell therapy modalities.

My role is focused on establishing and expanding strategic partnerships with pharmaceutical companies and CDMOs while driving our internal R&D pipelines and preclinical programs. We are mainly focused on preclinical development for monogenic disorders such as hemophilia and lysosomal storage disorders. In collaboration with partners, we are also working on genetic epilepsy—particularly Dravet syndrome—and ocular conditions such as Stargardt disease.

All of our efforts aim to tackle key gene therapy challenges, including manufacturing scalability, quality, immune tolerability issues, durability of treatment, and the development of long-term therapeutic solutions.

Q Tell us more about msDNA—what differentiates it from pDNA and other DNA vectors in its key applications?

NN: MsDNA is a double-stranded linear mini-DNA vector with closed ends. It is free of bacterial and viral sequences, and is designed to enhance safety, tolerability, and flexibility for gene and cell therapy applications. Our approach offers key advantages over conventional pDNA and synthetic linear DNA vectors, which are produced via polymerase chain reaction (PCR). By removing plasmid backbone elements, we eliminate the safety concerns associated with unwanted immunogenicity and antibiotic-resistant genes potentially present in traditional plasmids.

Compact design is another advantage—msDNA has a streamlined structure to allow enhanced nuclear uptake, transfection efficiency, and targeted gene expression in the target tissues while minimizing off-target effects. The closed ends dramatically reduce the potential of random integration.

We add proprietary sequences to the structure to improve intracellular kinetics and transfection efficiency to hard-to-transfect cells, such as induced pluripotent stem cells (iPSC) and immune cells such as CAR-Ts and CAR-NKs. Additionally, we designed msDNA to have an enhanced safety profile, making it an ideal DNA vector for applications that require high quality and tolerability profiles.

Another benefit is that msDNA is very flexible in terms of potential size of payload. We can accommodate multiple gene expression cassettes, or very large complementary DNA genes and large promoters. We have successfully manufactured msDNA from 2 kb up to 16 kb.

The linear, covalently closed structure makes msDNA more resilient—it is torsion-free and remains intact through mechanical and chemical stress and shearing forces. We plan to publish the data that we have collected on resilience to physical force and low pH in the near future. This is particularly relevant for cell therapy applications that require physical forces for *ex vivo* transfection, and for formulation aspects that require a highly acidic pH.

Last but not least, there is fidelity. Compared to many other synthetic or PCR-based linear DNA vectors, an *E. coli*-based manufacturing approach provides DNA that benefits from the

"...*E. coli*-based products exhibit mutation rates 1,000–10,000 times lower than synthetic DNA and PCR-based vectors..."

mismatch repair (MMR) mechanisms embedded into *E. coli* cells, and therefore achieves a higher level of fidelity, which is crucial for clinical applications.

Q Can you expand on the specific advantages an *E. coli*-based manufacturing approach provides?

NN: *E. coli*-based manufacturing is a platform that has seen wide adoption in the pharmaceutical industry and offers significant advantages, including the intrinsic MMR mechanisms I mentioned. Essentially, the cells' natural proofreading mechanisms ensure high fidelity during DNA replication, thereby minimizing mutation risks compared to synthetic or PCR-based products.

We have developed precise analytical techniques to quantify loss-of-function mutations, demonstrating that msDNA has superior fidelity compared to synthetic DNA vectors. We showed that *E. coli*-based products exhibit mutation rates 1,000–10,000 times lower than synthetic DNA and PCR-based vectors, which again, is crucial to success in clinical applications. This production process ensures superior fidelity when compared to other linear DNA constructs.

A second benefit is the ability to leverage established infrastructure. Our own extensive experience in fermentation process development and scale up, plus the existing knowledge pool in the industry, allows for straightforward scalability from bench to GMP-grade manufacture. The optimized conditions also result in a more cost-effective approach, as we produce high yields per batch.

Q What are the major challenges in msDNA production and analysis, and what are some key technological solutions that you employ to help address them?

NN: As anyone working on the manufacturing of nucleic acids can tell you, there are always challenges! Achieving high-yield conversion of the precursor plasmid to msDNA during the upstream process without compromising quality was a key issue. To address this, we did extensive process development. We developed proprietary protocols and SOPs, tested different *E. coli* strains, and we continue to work diligently on enhancing our manufacturing yield.

"...we envision msDNA-driven therapeutics benefiting multiple areas of gene and cell therapy."

The other major obstacles were around analytical precision: ensuring that we had the right analytical tools for sequence evaluation in order to ensure high fidelity, purity, and functionality of our final product. We are currently utilizing a number of advanced analytical tools. We developed a process for HPLC and qPCR, together with employing next-generation sequencing, and we are running these techniques to validate every batch.

Finally, there is the industry-wide challenge of achieving safe and targeted delivery of your DNA vector. To address this, we are collaborating with delivery platform companies, and internally, we are advancing our nanoparticle and formulation capabilities to enhance the clinical potential of msDNA.

Q How and where do you hope to see msDNA-driven therapeutics having an impact—both initially, and longer term?

NN: As I mentioned, we are currently focused on preclinical development for non-viral gene therapy applications. Ministring DNA enables re-dosable and titratable therapies which are particularly promising for monogenic disorders where traditional treatments or delivery systems face challenges, such as a pre-existing immune response, the size of the transgene, or long-term expression. For conditions such as lysosomal storage disorders, msDNA can deliver larger transgene or multiple gene expression cassettes, overcoming the payload and pre-existing immunity constraints of AAV.

Our R&D pipeline is focused on bringing these applications to clinical reality and addressing unmet needs across a range of diseases. Ultimately, we envision msDNA-driven therapeutics benefiting multiple areas of gene and cell therapy.

Q Finally, can you sum up one or two key goals/priorities/milestones, both for yourself in your own role and for Mediphage as a whole, over the foreseeable future?

NN: For myself, the priority is to expand Mediphage's R&D capabilities through strategic partnerships and to advance our preclinical program towards IND-enabling studies. I am also focused on scaling GMP-grade msDNA production to meet clinical demands.

For the company, our immediate goal and priority is fundraising and securing resources to advance our lead msDNA therapeutic to clinical trials. We would like to secure resources for

scaling GMP manufacturing, either by building the infrastructure to support commercialization or by engaging with existing facilities that can help us to outsource our msDNA and expand its application.

By collaborating with industry leaders, drug developers, and strong academic partners, and with support from the government, we hope we can further explore the utility of msDNA and translate its potential into transformative solutions for unmet medical needs. Manufacturing high-quality nucleic acids is foundational to the clinical success of cell, gene, and nucleic acid-based therapies, and I believe Mediphage is positioned to play a critical role in this new area of genetic medicine.

BIOGRAPHY

NAFISEH NAFISSI is the CTO of Mediphage Bioceuticals an innovative pharmaceutical company based in Toronto, Ontario, Canada. She leads the company's research and development efforts, steering scientific innovation and overseeing both internal projects and strategic partnerships with major pharmaceutical companies. Dr Nafissi brings over 20 years of hands-on expertise in biotechnology, genome engineering, and gene therapy, alongside over 5 years of executive management experience in both academic and industrial settings. As co-inventor of Mediphage's technology platform and its novel Ministring DNA product, Dr Nafissi has contributed significantly to the field, with over 20 recognized publications and four patents to her name. In her current role, she manages the R&D team, oversees IP generation around Ministring DNA applications, directs technology transfer, and spearheads the commercialization of Mediphage's cutting-edge nucleic acid therapeutics aimed at delivering redosable and accessible treatments for patients globally. Dr Nafissi is also an Adjunct Professor at the University of Waterloo, Waterloo, Ontario, Canada where she co-supervises graduate students, guiding their work toward meaningful industrial applications. Previously, she collaborated with the Pasteur Institute on national vaccine development projects for *Helicobacter pylori* and on genotyping hemophilia patients, contributing to advancements in personalized drug development and prenatal screening. Dr Nafissi holds an MSc (Molecular Biology) and PhD (Pharmaceutical Science) from the University of Waterloo and a BSc (Microbiology) from the University of Tehran, Iran. She has received certification in drug development and effectiveness (CIHR) and specialized training in negotiation, licensing, and commercialization.

AFFILIATION

Nafiseh Nafissi
Chief Technology Officer,
Mediphage Bioceuticals,
Toronto, Ontario, Canada

AUTHORSHIP & CONFLICT OF INTEREST

Contributions: The named author takes responsibility for the integrity of the work as a whole, and has given their approval for this version to be published.

Acknowledgements: None.

Disclosure and potential conflicts of interest: The author is a Mediphage Bioceuticals employee and stock holder, and this manuscript is about Mediphage's ministering DNA technology.

Funding declaration: The author received no financial support for the research, authorship and/or publication of this article.

ARTICLE & COPYRIGHT INFORMATION

Article source: Invited.

Interview conducted: Nov 20, 2024; **Revised manuscript received:** Dec 12 2024;

Publication date: Dec 20, 2024.

BIOINSIGHTS

Thank you to all our peer reviewers in 2024

At BioInsights, we use double-blinded peer review to minimize bias and focus the evaluation solely on the quality of the research rather than personal identities. However, despite their invaluable expertise and thoughtful feedback, peer reviewers often go unrecognized and are not formally acknowledged in publications. Their contributions are crucial to our editorial process, helping us assess articles and advance fields such as cell and gene therapy, nucleic acids, vaccines, and immuno-oncology, and we deeply appreciate the time and effort our reviewers dedicate to this process, enabling us to publish content of the highest quality. To honor their contributions, the editors of BioInsights are proud to publicly acknowledge our top reviewers of 2024 and look forward to collaborating with them in the future.

Sincerely,
BioInsights

Dr Asher Williams, Columbia University
Dr Joseph Zaia, Department of Biochemistry, Boston University School of Medicine
Dr Hemant Dhamne, Autolus Ltd
Dr Eduard Ayuso, Dinaqor AG
Dr Hassan Rashidi, University College London
Dr Davide Danovi, bit.bio
Dr Brian Philip, National Institute for Bioprocessing Research and Training (NIBRT)
Dr Saba Ghassemi, University of Pennsylvania School of Medicine
Dr Austin Thiel, Elevate Bio
Dr Wouter Van't Hof, Cleveland Cord Blood Center
Dr Roger Horton, Anthony Nolan
Dr Jason Acker, University of Alberta
Dr Kavitha Siva, Xintela AB
Dr Victoria Day, NHS Blood and Transplant
Dr Roberto Gramignoli, IRCCS Ospedale Pediatrico Giannina Gaslini
Dr Nick Chen, Fate Therapeutics
Dr Behnam Ahmadian Baghbaderani, Cell Therapy Process Department, Lonza Houston
Dr Daniel Paull, The New York Stem Cell Foundation Research Institute
Dr Dhruv Sareen, Cedars-Sinai
Dr Marina Zenkova, Institute of Chemical Biology and Fundamental Medicine SB RAS
Dr Tonya Villafana, AstraZeneca
Dr Jean Haensler, Sanofi R&D
Dr Eleni Samaridou, Merck
Dr Michaela Sharpe, Moare Solutions Ltd
Mr Paul Heal, Orchard Therapeutics
Dr James McBlane, Licensing Division, Medicines and Healthcare Products Regulatory Agency

Dr Keith Sutton, Resolution Therapeutics
Dr Michela Palmisano, Fondazione Telethon
Dr Sol Ruiz Antunez, Spanish Medicines Agency (AEMPS)
Dr Pasi Virta, University of Turku Finland
Dr Alyssa Hill, Oxford Department of Pediatrics
Dr Ramon Eritja, Department of Surfactants and Nanobiotechnology, Institute for Advanced Chemistry of Catalonia (IQAC), Spanish National Research Council (CSIC)
Dr Akash Bhattacharya, Beckman Coulter Life Sciences
Dr Bryan Troxell, Atsena Therapeutics
Dr Cristiana Boi, North Carolina State University
Dr Vijesh Kumar, Spark Therapeutics
Dr Lefkothea Papadopoulou, Laboratory of Pharmacology, School of Pharmacy, Faculty of Health Sciences, Aristotle University of Thessaloniki
Dr Benjamin Davis, University of Pennsylvania
Dr Jose-Antonio Daros, Instituto de Biología Molecular y Celular de Plantas
Dr Simon Daniel, Imperial College London
Dr Kostadinka Lilova, Autolus Therapeutics
Dr David Shaw, Genentech
Dr Sudhanshu Shekhar, Bristol Myers Squibb
Dr Sameer Kalghatgi, FUJIFILM Diosynth Biotechnologies
Mr David Ede, Sartorius
Dr Stuart Beattie, Biogen
Dr Chantelle Gaskin, Thermo Fisher Scientific
Dr Alexis Cockroft, Lex Regulatory Ltd
Dr Ilona Reischl, Austrian Agency for Health and Food Safety—Austrian Medicines and Medical Devices Agency
Dr Nermin Ibreljic, Sarepta Therapeutics
Dr Alex Chatel, Donaldson
Dr Paul Young, Pharmaron Gene Therapy
Dr Francesca Vitelli, Intellia Therapeutics
Dr Rowan Flynn, RosalinCT
Dr Emily Titus, Notch Therapeutics
Dr Pete Tonge, bit.bio
Dr David Wellis, Excellos
Dr Isabelle Riviere, Takeda
Dr Jennifer Solomon, STEMCELL Technologies Inc.
Dr Saif Raisheed, Biocina
Dr Rana Chattopadhyay, Sanofi
Mr Benjamin Hall, Tessera Therapeutics
Dr Monica Dommel, Curevac
Dr Jane Luo, sciex
Dr Wei He, Lawrence Livermore National Laboratory
Dr Renaud Heine, Erasmus School of Health Policy and Management, Erasmus University
Mr Vikram Shanbhag, SG Analytics
Dr Misganaw Asmamaw Mengstie, Debre Tabor University
Dr Riccardo Privolizzi, Sania Therapeutics
Dr Anjulika Chawla, University College London
Dr Sourav Choudhury, Sanofi
Dr Adrien Lemoine, Bloomsbury Genetic Therapies
Dr Luca Marchetti, University of Trento Centre for Computational and Systems Biology (COSBI)
Dr Atul Goyal, Pfizer
Dr Sita Awasthi, Perelman School of Medicine, University of Pennsylvania
Dr Matthew Smart, Cell and Gene Therapy Catapult
Dr Kevin D'Amour, Stemson Therapeutics
Dr Tessy Hick, Umeå University
Mr Ashish Saksule, Vertex Pharmaceuticals
Dr Sharada Mokkapati, MD Anderson Cancer Center
Ms Min Zong, Cellics Therapeutics
Mr Daniel Koback, Sartorius
Mr Yannick Borkens, Charité—Universitätsmedizin Berlin
Dr Kazutoyo Miura, National Institute of Allergy and Infectious Diseases (NIAID)
Dr Rajender Jena, Serum Institute of India Pvt. Ltd
Dr Piotr Kowalski, School of Pharmacy, University College Cork
Dr Natalie Prow, The Hull York Medical School, University of York
Dr Ruben Esse, Cell and Gene Therapy Catapult
Ms Konstantina Malengou, Cellics Therapeutics
Dr Ferdos Hashi, King's College London
Mr Paul Greback-Clarke, AskBio
Mr Kyle Stead, Virocell Biologics
Dr Gabrielle Humphrey, VECTAPLUS
Dr Catherine Jomary, IPS-Integrated Project Services
Professor Barry Fuller, University College London
Dr Julie Meneghel, Cytiva
Dr Kathryn Murray, Takara Bio Europe
Mr Alexander Kerr, King's College London
Dr Maria Rende, Sartorius
Dr Drew Hope, Exmoor Pharma
Dr Garima Thakur, Regeneron
Dr Adrien Savy, Coave Therapeutics
Dr John F Tisdale, Cellular and Molecular Therapeutics, NIH