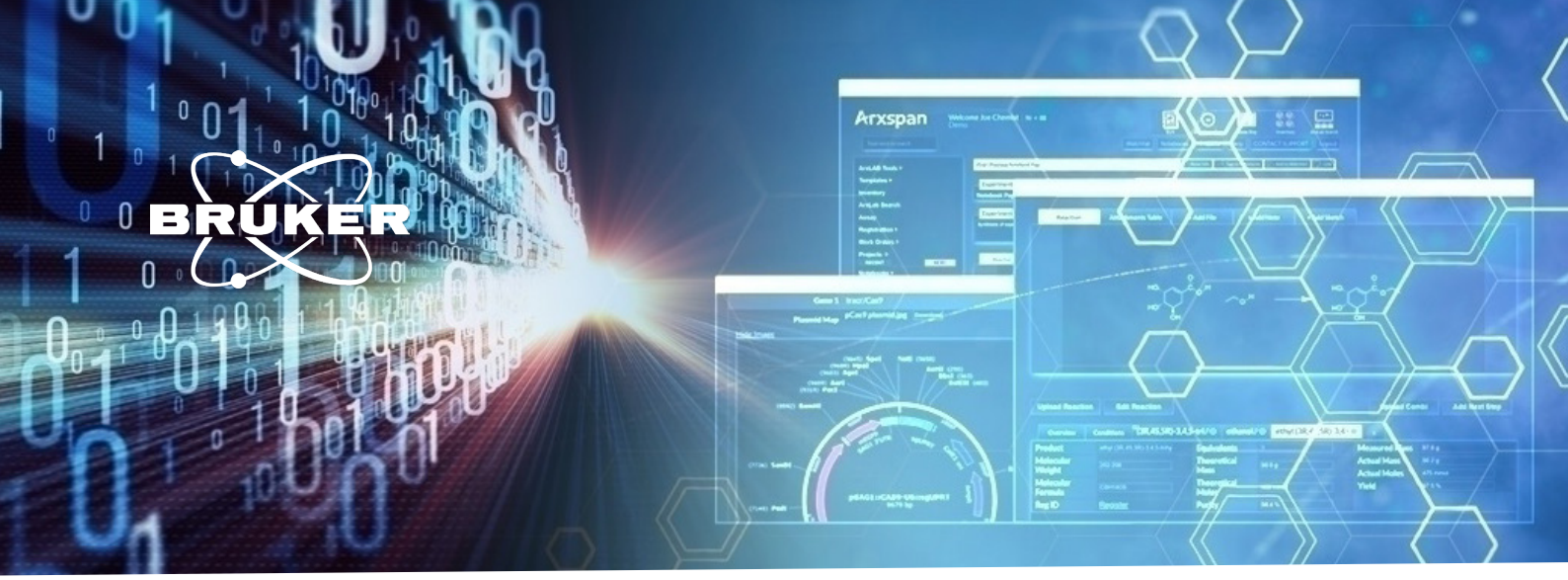

LABORATORY DATA MANAGEMENT

Addressing the Data Challenges of Biologics Research

Innovation with Integrity

The last two decades has seen a dramatic rise in the number of biological therapies, such as monoclonal antibodies, recombinant proteins and vaccines, being brought to market.

In 2020, the US Food and Drug Administration's (FDA) Center for Drug Evaluation and Research (CDER) approved 53 new molecular entities (NMEs), of which 13, or 25%, were biologics[1]. Biologics also account for an increasing proportion of the top 100 drugs and, of the top 10 selling drugs projected for 2021, five are biologics, including the top two: AbbVie's Humira (adalimumab) and Merck & Co.'s Keytruda (pembrolizumab). Since 2014, biologics have accounted for 93% of the growth in net drug spending[2].

Percentage of Biologics Approved as New Molecular Entities (NMEs) by FDA's CDER, 2016–2020.

2020	2019	2018	2017	2016
25%	21%	29%	26%	32%
53 NMEs approved: 13 were biologics	48 NMEs approved: 10 were biologics	59 NMEs approved: 17 were biologics	46 NMEs approved: 12 were biologics	22 NMEs approved: 7 were biologics

Biologic approvals include those approved as new molecular entities via an original biologics license application (BLA) by the US Food and Drug Administration's Center for Drug Evaluation and Research.

Source: US Food and Drug Administration's Center for Drug Evaluation and Research

There are several drivers for this. Compared to small molecules, biologics have a relatively low rate of attrition - 24.4% of preclinical stage biologics survived up to the market stage compared to 7.1% for small molecules[4]. Biologics are also harder to 'copy', such that companies developing them face less severe competition from biosimilars, extending their revenue streams even post-patent expiry. It is also often easier to extend the number of indications to which biologics can be applied compared to small molecules, as exemplified by Humira, which has already been registered for use in more than 10 indications in the US alone.

However, the cost of developing a new biological therapeutic is high. A major contributor to this stems from the complexity of their structure and the accompanying high degree of characterization required. Underpinning this is the need for effective data management throughout the development lifecycle, from early research through to manufacturing.

Molecular Biology plays a pivotal role in the development of any novel biologic, such as understanding the molecular basis for biological activity and optimizing the sequence of the gene to increase expression levels, etc.

Here we explore how a modern, collaborative approach to the management of molecular biology data can, and ultimately, expedites the development of novel biological therapeutics.

Meeting the Needs of Molecular Biologists

In contrast to their chemistry counterparts, scientific teams responsible for the development of biologics have historically been underserved by the laboratory informatics market. For example, molecular biologists are often highly reliant on a disconnected set of standalone sequence analysis tools, Microsoft Office applications, file shares and even paper, to manage their data.

Sequence data is pivotal to contextualizing and understanding the results of biological experiments. For example, a single point mutation within a gene can have a major impact on the expression levels and function of a protein. To be able to make any kind of correlation between gene variants and the associated results, it's critical to be able to associate experimental data with the specific variant of the gene or protein being studied. Such differences can only be truly encapsulated in the sequence itself.

Rather than simply attaching plasmid map images, incorporating sequence data directly within an ELN experiment is necessary to make this link unambiguous. Attaching a sequence file, such as a GenBank file, to an experimental record is a step in the right direction. However, the ability to centrally manage molecular entities, such as primers, restriction enzymes and sequences directly within the ELN platform not only provides a single point of truth with an associated audit trail and version control, but also enables users to easily incorporate live, interactive sequence data within their experimental documentation.

Moreover, if an ELN also supports common molecular biology workflows, such as designing PCR primers, restriction cloning and transfection experiments, then the incorporation of sequence data will inherently maintain a genealogy between biological entities, enabling traceability from a cell line back to the plasmid and PCR primers used to create it.

Eliminating Data Silos

An ELN that meets the needs of Molecular Biologists is clearly a very significant step forward. However, for companies engaged in both small and large molecule R&D, the proposition of the additional cost and resources required to implement and maintain separate ELNs for chemistry and biology is not particularly appealing.

But there are wider implications which could compromise continuity and integrity in R&D. For example, chemistry and biology teams often interact with the same analytical and central services departments. If the analytical data is in one ELN and the molecular biology data is in another, it becomes challenging to gain a holistic view. Similarly, if the preparation of a batch of media by a central services team is documented in a different ELN to the one used to write up the molecular biology experiment where that media is used, then the traceability of both materials and data is compromised.

Furthermore, many novel therapeutics are not simply chemicals or biologics. Let's take Antibody Drug Conjugates (ADCs), or immunoconjugates, for example. ADCs are an increasingly important sub-class of antibody-related therapeutics, the market for which is predicted to grow to over $13 Billion by 2026[5]. ADCs exploit the specificity of monoclonal antibodies (mAbs) to target drug delivery to tumor cells while minimizing damage to healthy tissue, thus reducing side effects associated with traditional chemotherapy drugs. They can also overcome issues associated with small molecule therapies such as clearance and cellular internalization.

The development of ADC-based therapies clearly requires input from both biology and chemistry teams who are involved in a diverse range of activities. While this results in an increasingly complex picture, collaboration both within and between teams can be significantly simplified when everyone involved uses the same platform to manage their data. This eliminates manual hand-offs of data between groups, reduces ambiguity and streamlines communication between them.

Ultimately, when all research data is managed in one place, scientists don't need to waste time looking in multiple systems to find the data they need. Having a single point of truth enables a 'self-service' approach to data - scientists can access information in a timely manner, without having to work out who in the department or organization has what they need. A common research ELN not only facilitates cross-pollination of ideas between otherwise disconnected departments, but also provides a comprehensive system of tracing interdepartmental collaborations.

Introducing Arxspan BioDrive

To meet the challenges faced by Biologics research, Bruker has launched Arxspan BioDrive which has been designed to support common molecular biology workflows through a modern, easy to use interface. BioDrive is available both as a standalone desktop application and as an enterprise solution embedded within the Arxspan Electronic Laboratory Notebook platform. Data can be easily transferred between the desktop and the cloud, giving molecular biologists the flexibility to work where they want and how they want.

BioDrive enables researchers to:

- Create, import, and export sequences, plasmids, and vectors (with support for all common file formats)
- Display, analyze, and annotate sequences, enzymes, and primers
- Automatically detect restriction enzyme digestion sites
- Automatically annotate sequences with shared annotation libraries
- Automatically calculate biochemical properties for any sequence
- Design and plan cloning projects
- Maintain relationships between molecules, such as those between primers, PCR products and the resulting clone
- Align the designed sequence with one or more sequencing files (e.g..abi/.ab1 files)
- Use sophisticated search capabilities, including sequence or keyword searches
- Translate DNA sequence directly to an amino acid sequence
- Generate publication quality graphics

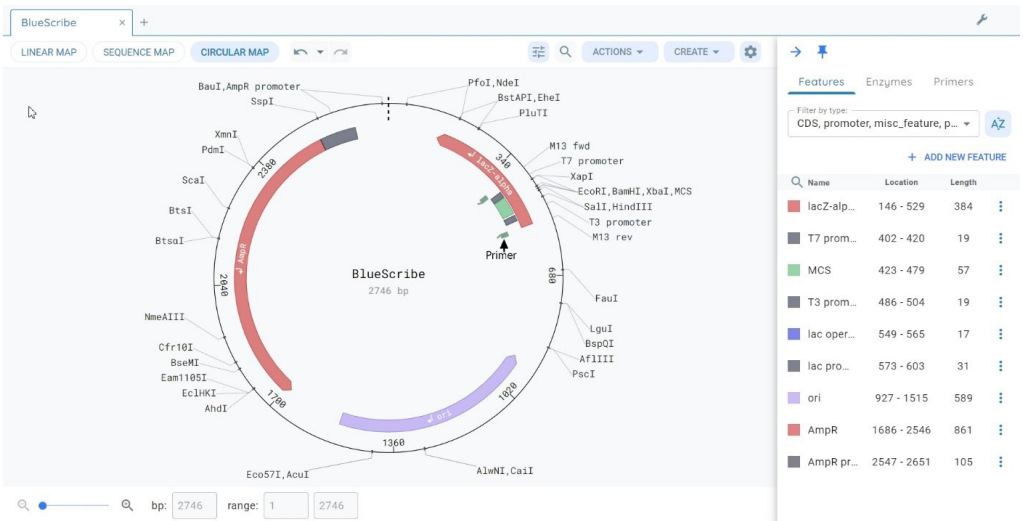

Figure 1: The Display of a circular DNA sequence with annotation features, restriction enzyme sites and primers.

BioDrive can be accessed via a web-browser when deployed as part of the secure, cloud-based Arxspan Electronic Laboratory Notebook platform. This eliminates the need for organizations to maintain a separate sequence repository, while streamlining common molecular biology workflows. For example, before starting work in the lab, users can search the ELN with either the name of a molecule or its sequence to see if other researchers within their organization have tried something similar before and learn from the experience of others. They can also easily check if any necessary reagents are available onsite and where they are stored, such as PCR primers for a gene of interest that have previously been ordered by a colleague in a different department. This helps eliminate unnecessary duplication of effort, ultimately saving time and reducing costs.

BioDrive enables scientists to create their recombinant molecule in silico before starting work in the lab and simulate whether their molecular design will produce the desired product. For example, when designing a plasmid to express a fusion protein, the ability to translate the DNA sequence and confirm that the fusion is in-frame can avoid costly mistakes that would have otherwise only come to light after days, or even weeks, of lab work.

One of the drawbacks of legacy molecular biology tools is that they rarely enable users to record the context of their work, making it more difficult for others to re-use that information in the future. The inability to reproduce experiments due to improper documentation is a huge problem. While using BioDrive as an integral part of the Arxspan ELN platform users can record details of their experiments such as the recipes and conditions used for a ligation or PCR step, thus increasing both the utility, the reusability of the information, and most importantly, the reproducibility of the experiments.

Summary

Biologics present an exciting opportunity for the pharmaceutical industry, but the development of novel biological therapeutics presents a data management challenge.

BioDrive brings biological sequence awareness to the Arxspan platform. With the addition of BioDrive, Bruker delivers a single platform for managing molecular biology, biology, and chemistry data.

To find out more about how BioDrive can help expedite your biologics R&D program, contact us.

References:

1. https://www.dcatvci.org/features/tracking-biologics-what-s-in-store-for-2021/
2. https://www.forbes.com/sites/theapothecary/2019/03/08/biologic-medicines-the-biggest-driver-of-rising-drug-prices/
3. https://www.biopharmatrend.com/post/67-will-small-molecules-sustain-pharmaceutical-race-with-biologics/
4. https://www.drugs.com/history/humira.html
5. Global Cancer Antibody Drug Conjugate Market, Price, Dosage & Clinical Trials Insight 2026

Bruker BioSpin
info@bruker.com

Customer Support
https://www.bruker.com/en/services/support.html

Online information
bruker.com/