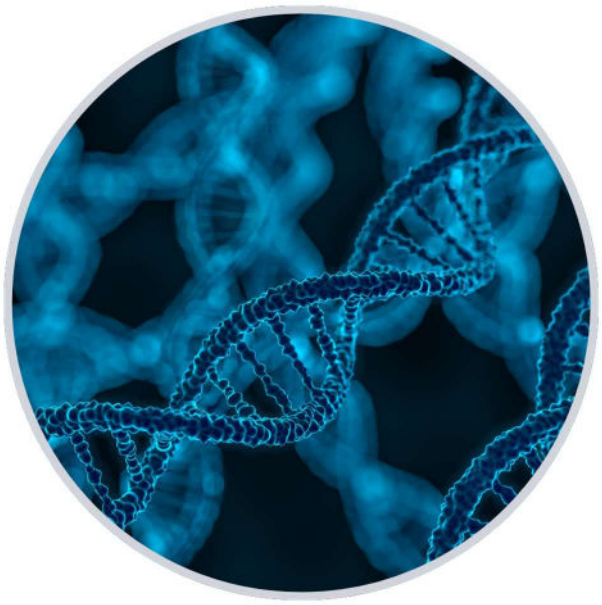
TECHNICAL PAPER

Accelerating data-driven oligonucleotide manufacturing

The Alchemite™ for Oligonucleotide Manufacturing solution



intellegens.com | info@intellegens.com
Intellegens, The Studio, Chesterton Mill, Cambridge, CB4 3NP, UK

intellegens

Executive Summary

Therapeutic oligonucleotides are scaling from discovery to commercial supply, but variability and impurities introduced during solid-phase oligonucleotide synthesis (SPOS) remain a significant barrier to robust, economic manufacturing. When impurities emerge early, downstream cleavage and deprotection (C&D), purification, and analytics become slower and costlier. Intellegens and CPI, funded by Innovate UK, have developed a novel software solution to this problem. Alchemite™ for Oligonucleotide Manufacturing leverages the proven capabilities of the Alchemite™ machine learning (ML) method to train ML models that uncover complex relationships linking production processes and product quality. These models enable accurate performance prediction and development of optimized quality control strategies. By interpreting both structural information of oligonucleotides and key processing parameters, Alchemite™ delivers a comprehensive, system-level view enabling efficient and informed process development. This paper introduces the Alchemite™ solution and reports the results of validation work conducted at CPI. Results include:

- Improved product yields by reducing critical impurities.
- Increased productivity through automated review of process data and identification of impurities.
- A reduction in the number of experiments required to achieve goals.
- Support for sustainability goals, minimizing waste.
- Knowledge capture and sharing to develop and retain corporate memory.

The manufacturing challenge

Solid Phase Oligonucleotide Synthesis (SPOS) is efficient for assembly but difficult to control at commercial scale and quality. Each synthesis cycle has dozens of variables (e.g., reagent equivalents, coupling and detritylation timings, oxidation/thiolation conditions), and outcomes are sequence-specific: the same synthesis parameters can behave differently by nucleotide context. Meanwhile, the deconvoluted mass-spec tables used in analysis of synthesis products rarely map cleanly to theoretical species, obscuring what is full-length product versus truncated, swapped, or chemically modified impurities. Add the heavy solvent usage required for washes and you have elevated PMI, cost, and sustainability risks.

Navigating this complex manufacturing landscape often requires in-depth knowledge and the expertise to apply it. In most organizations, such knowledge lives in bespoke spreadsheets and personal silos; when staff move on, teams must relearn at expense.

Alchemite™ for Oligonucleotide Manufacturing

A two year project, supported by Innovate UK, led by Intellegens, with support from CPI and six leading pharma and biotech partners, has created a software solution to assist with the oligonucleotide manufacturing challenge, accelerate process development, and capture vital knowledge in ways that can be re-used.

Alchemite™ for Oligonucleotide Manufacturing is an end-to-end machine learning (ML) solution that converts heterogeneous lab records into actionable, sequence-specific guidance. The solution is built on the proven **Alchemite™** platform for ML in R&D, adding domain-specific extensions that support the following oligonucleotide workflow.

1. Automated data extraction

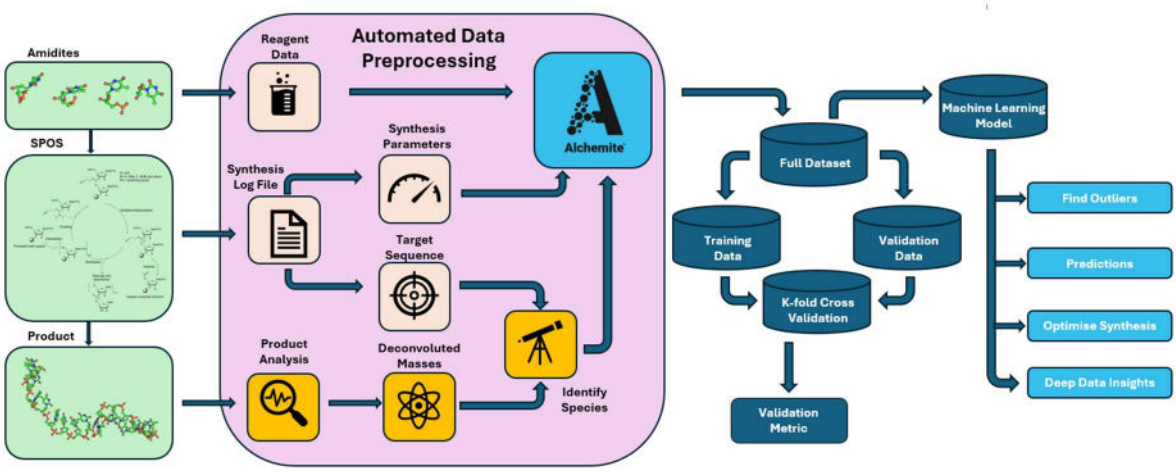


Fig 1. Mapping the workflow enabled by Alchemite™ for Oligonucleotide Manufacturing.

Inputs: the system makes it easy to ingest and process key oligonucleotide processing data:

- Synthesis logs: Parses synthesis parameters directly from oligonucleotide synthesizer equipment (initial implementation on the widely deployed Cytiva Oligosynt machine), reconstructing sequence order and per-step conditions.
- Mass spectrometry: Aligns deconvoluted peaks to theoretical species to generate an impurity profile (full-length, gain/loss-mers, nucleotide swaps, chemical modifications and adduct species).

- Reagents & yield: Normalizes equivalents, integrates OD260 measurements for crude yield, and captures reagent age and activator selection.
- Sequence features: Encodes sequences as SMILES and computes chemical/physical descriptors to generalize learning to unseen sequences.

Output: a tidy, machine learning-ready dataset created without manual wrangling.

2. Uncertainty-quantified machine learning

The Alchemite™ ML method can now be used to analyze your data, automatically building a mathematical model that captures key relationships between inputs to the system and outputs from it. This process is called **training an ML model**. The ML model can be built from scratch using your data or you can use your data to retrain existing models, such as those provided by Intellegens, leveraging the knowledge gained during previous projects. Once trained, an Alchemite™ model can be used to predict expected outcomes such as yield, purity, and impurity profiles for new sets of synthesis parameters, or to propose the synthesis conditions most likely to achieve specific target performance criteria. In the context of oligonucleotide process development, this enables more efficient process optimization by guiding experimental design toward conditions that deliver the desired quality attributes.

Key features of the Alchemite™ method are:

- Provides a highly accurate method for estimating the **uncertainty** of its predictions. This feature is vital for good decision-making. It enables, for example, quantitative assessment of which experiments are most likely to succeed.
- Handles the **sparse, noisy, high-dimensional datasets** common in development – these often cause problems for training models using generic ML methods.
- **Predicts synthesis outcomes** and enables analysis by flagging outliers in real time and ranking synthesis parameters by influence on purity, yield, impurity classes.
- **Proposes high-confidence experiments** to meet predefined thresholds (purity/yield) so teams can perform high value studies instead of spending months varying single synthesis parameters across experimental iterations.

3. Adaptive DOE for efficient process design

Alchemite™ is tuned to explore and exploit high-dimensional design spaces and to help researchers design experimental programs that achieve their targets, faster. It achieves these objectives through the ability of the ML model to capture and apply complex, non-linear relationships in data and by supporting *adaptive* Design of Experiments. The adaptive approach is illustrated in Figure 2 and discussed in more detail in a separate Intellegens white paper[1]. Once trained, the machine learning model is used to advise on which new experiments

[1] Intellegens White Paper: Machine learning for adaptive experimental design (2025). Available at intellegens.com/whitepapers.

are most likely to increase its accuracy as it aims to predict the ideal synthesis parameters. Those experiments are conducted and their results used to retrain the model. The process repeats until targets are achieved.

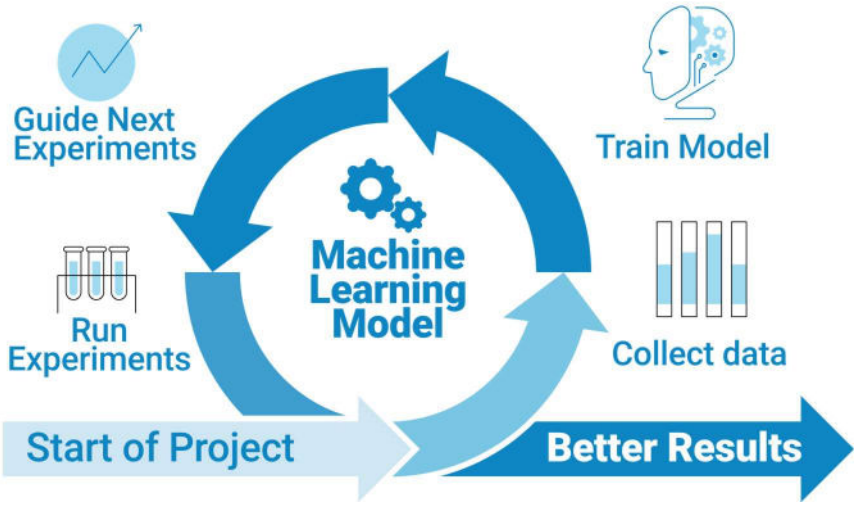


Figure 2. A schematic of the adaptive experimental design approach.

Compared to traditional DOE, this iterative approach hits project goals in fewer experiments (typically a 50-80% reduction). Conventional DOE approaches can also be limited in their ability to explore complex experimental design spaces, requiring only a small number of parameters at a time to be varied and missing multi-parameter optima. ML overcomes this restriction. In the case of oligonucleotides this means that Alchemite™ can, for example, recommend conditions that improve crude yield and purity while reducing sequence-specific impurities without defaulting to costly suggestions of more equivalents, more activator. The result is **fewer runs, cleaner crudes, and easier purification**.

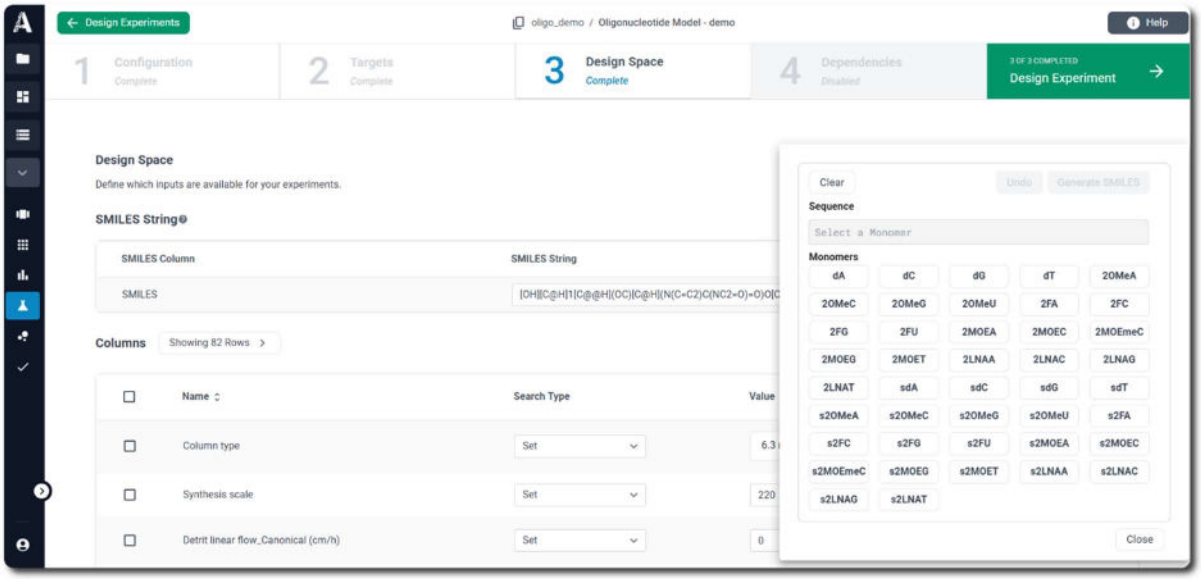


Figure 3. Design of experiments showing sequence data within Alchemite™.

4. Beyond synthesis

Because sequences are stored as SMILES with rich descriptors, providing the software with an understanding of the chemistry, the same transferred learning between sequences framework extends to **C&D, purification, and analytical-method development**, with models that evolve in real time, refining their predictions as new data becomes available.

Validation project: demonstrating commercial value

Validation scenarios and results

Working with CPI, Alchemite™ for Oligonucleotide Manufacturing was applied to real process development. The project iteratively built up a comprehensive dataset comprising 60 experiments across three distinct oligonucleotide sequences, training and optimizing an ML model using these experiments. The model was then applied to two previously unseen cases.

Known sequence optimization: In a single iteration comprising just two Alchemite™ designed experiments, target results were achieved for a characterized siRNA-like sequence with historical purity and yield data. Alchemite™ recommended non-obvious, multi-parameter adjustments – such as maintaining equivalents near ~2 and lowering activator levels – that led to simultaneous improvements in both yield and purity. A 7% increase in crude purity was achieved, while n-2 impurities were significantly reduced to as low as 1.3%, well below the target threshold of <3.5%. The n-1 impurity level was also reduced and identified as a focus for further optimization through the inclusion of finer-grained, position-aware data. This outcome demonstrates a substantial acceleration of process optimization compared to conventional experimental programs in this domain.

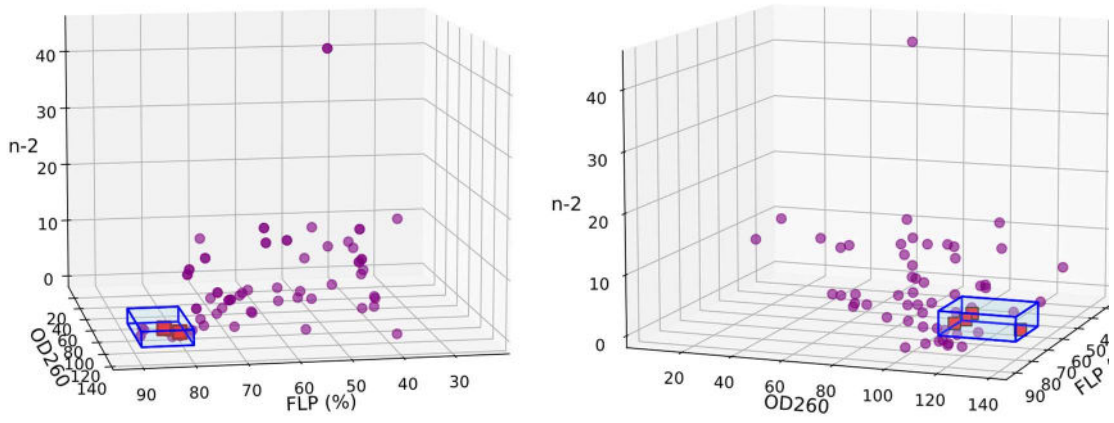


Figure 4. Validation study data. Purple dots show experimental data used to train the ML model. The box indicates the target region for key properties of synthesized oligonucleotides. Red squares show the results of using synthesis parameters proposed by the ML model for known sequence (left) and unseen sequence (right) scenarios.

Unseen sequence: For a development-scale sequence with no prior runs, all targets were met in the initial iteration of four experiments. Crude yields reached up to 138 OD260/µmol and FLP purity up to 78%, with n-2 impurity of 1.9%, confirming Alchemite's ability to identify high-yield and purity conditions for new oligonucleotide sequences.

In both of these scenarios, the resulting ML model captured the key relationships that define the synthesis, and thus vital knowledge that can be re-used in future projects. For example, graphical analytics could be applied to help understand which inputs drive which outputs and the model could be used to predict synthesis outcomes for new sequences.

Fewer runs and less re-work also mean lower solvent use, energy, and waste, directly improving PMI and enabling more sustainable routes to scale.

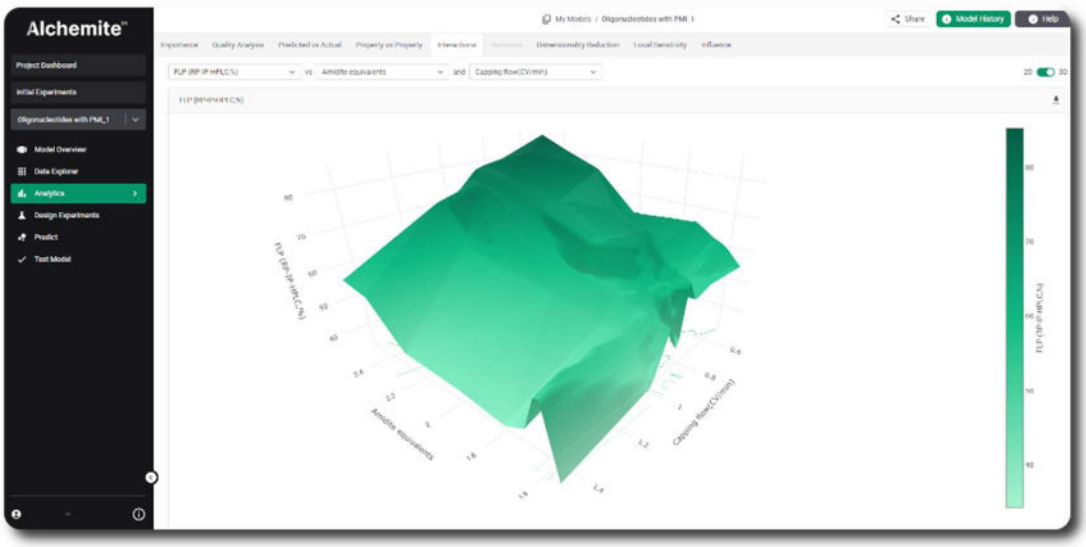


Fig 6. Using the Alchemite ™ software to explore relationships between process parameters and properties uncovered by the machine learning model.

Validation project conclusions

Alchemite™ demonstrated at least a **5% improvement in theoretical product yield** by reducing critical impurities, fulfilling a key acceptance criterion.

Automated Process Review & Productivity: The system achieved a 90% reduction in manual interpretation time for impurity analysis, significantly improving productivity.

Reduced Expert Dependency: The software reduces the need for expert input through automation and guided decision-making.

Streamlines data capture and analysis: the ML model offers unique insights and helps to overcome current process limitations.

Optimizes synthesis: the solution identifies sequence-specific parameters to predict optimal process parameters in virtual experiments.

Alchemite™ unified SPOS data, learned the complex mapping from synthesis conditions to outcomes, and delivered executable recommendations, closing the loop between experiment and optimization. Each run strengthens the model and reduces waste, enabling faster, cleaner, and more economical routes to commercial-grade oligonucleotides.

Implementing the solution

Alchemite™ for Oligonucleotide Manufacturing is available via a user-friendly web browser-based app or programmatic API access (using the Alchemite™ Architect technology) for integration with ELN, data lakes, and automation.

Built-in governance and auditability: the system enables standardized ingestion, versioned models, and reusable feature/descriptor libraries.

Deployable to single sites or multi-site networks to enable cross-site reproducibility.

Any pharma or biotech R&D team can deploy the solution as commercial-off-the-shelf (COTS) software, with an integrated package of software and partner services to help them maximize oligonucleotide development productivity. A typical implementation process is:

1. **Discovery** – understanding your data by building ML models using Alchemite™ and applying its graphical analytics to explore the relationships in your system can prioritise high-value synthesis workflows, helping you to meet defined synthesis quality and PMI targets.

2. **Proof of concept** – run a representative route, using Alchemite™ to propose the experiments and process parameters that are likeliest to succeed. Benchmark improvements to impurity profile, crude yield, and runtime.

3. **Scale-Up** – roll out to all oligo programs; extend to purification and analytics; integrate with digital systems.

4. **Knowledge Hub** – establish a living repository of sequence–process–outcome relationships and model artefacts to secure institutional memory.

Conclusion: accelerating process development

Alchemite™ Suite offers a new software extension for oligonucleotide manufacturing which has been customized to handle oligonucleotide manufacturing data, providing a single front door for data visualization and comparison across multiple processing data types while using machine learning to enable more robust and faster decision making.

Alchemite for Oligonucleotide Manufacturing is a solution that can be used across the full manufacturing process including synthesis, cleavage and deprotection, purification, and analytical method development. In this paper, we have summarized its validation for sequence-specific optimization in solid-phase oligonucleotide synthesis (SPOS). It has demonstrated improved yields by predicting critical impurities and automating impurity identification. Key features include ML-guided design of experiments, intelligent impurity analysis, and automated reporting from chromatography and MS workflows—reducing experimental load and accelerating development.

This results in improved product yield, improved productivity, and reduced time and cost through reduction in the number of experiments required for process development.

To explore how Alchemite™ can accelerate your manufacturing programmes, contact Intellegens to arrange a technical demonstration or discovery call.

About CPI

CPI runs the Medicines Manufacturing Innovation Centre (MMIC), a collaborative hub near Glasgow that helps the pharmaceutical industry adopt better, faster, and greener ways to make medicines. MMIC brings together leading companies, universities, and government partners – including AstraZeneca, GSK, and the University of Strathclyde – to turn promising ideas into proven, scalable solutions. Through joint projects and demonstration facilities, MMIC helps organizations de-risk new technologies, improve productivity, and move from development to commercial manufacture more smoothly. The goal is simple: enable the UK and global life sciences community to deliver high-quality treatments to patients sooner.

About Intellegens

Our vision is that machine learning will drive innovation and deliver value wherever data is used in R&D. We aim for best-in-class easy-to-use machine learning software for data analysis in chemicals, materials, life science, and manufacturing. Our Alchemite™ technology originated at the University of Cambridge and development is on-going at Intellegens, in close collaboration with our growing community of customer organizations. These represent sectors including additive manufacturing, aerospace, alloys, batteries, biopharmaceuticals, ceramics, chemical processes, composites, consumer products, cosmetics, drug discovery, energy, food and beverage, formulated products, paints, plastics, and printing technology.

www.intellegens.com | info@intellegens.com