A novel machine learning-driven tool for oligonucleotide characterisation and manufacturing

cpi intellegens

The challenge

Oligonucleotide characterisation, impurity control, and manufacturing

Oligonucleotides are a major emerging new class of medicines, offering the potential to treat a range of common and rare genetic diseases. However, broader clinical application is limited by challenges in oligonucleotide synthesis and sustainable manufacturing.

With no industry standard nomenclature for oligonucleotides, **characterisation is difficult**. The size and complexity of these molecules makes prediction laborious and complex. **Impurity** prediction tools designed for small molecules struggle with this application [Song, 2020]. Interpretation of mass spectroscopy data to support impurity control requires deep understanding of the molecules and potential impurities [Roberts, 2020].

Current **manufacturing** technology is expensive, has poor sustainability profiles, and is unsuited to large-volume products. It is particularly challenging to understand relationships between impurity profile and manufacturing process with sub-optimal approaches to quality control strategies and improvement of product quality.

The project

Applying machine learning to improve understanding and optimise processes

Intellegens, in collaboration with CPI, is developing a **machine learning (ML)-powered digital tool** that enables rapid, best-in-class prediction of potential impurities and characterisation of oligonucleotides. Industry partners will contribute critical expert input and testing of the digital tool.

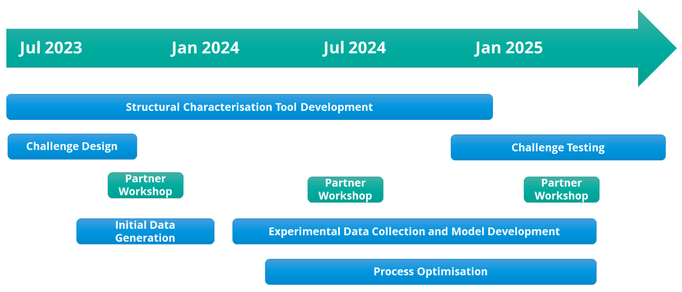
At the heart of the tool is the **Alchemite** method, which has been extensively validated [References below] for difficult experimental and process datasets. The project will build a **database** of structures associated with oligonucleotides. Linked to mass spectra results, this will enable profiling and identification of potential impurities. ML can then be applied to learn from structural, impurity, and process condition data.

Applying a Bayesian experimental design paradigm, the ML will identify **optimal process parameters** for a given oligonucleotide, achieving improved yields while reducing required experimental workloads.

Progress

Funding secured, design phase underway

In August 2023, the project partners announced the award of £1.6m in grant funding to support this project from Innovate UK, part of UK Research and Innovation (UKRI). The project, which will run until mid-2025, is already underway, with progress made on initial experimental planning, database development, and software design and prototyping.



Method

For more on the Alchemite method see, for example:

T.M. Whitehead et al, *Imputation of Assay Bioactivity Data using Deep Learning, Journal of Chemical Information and Modeling* **59**, 1197 (2019)

See also case studies at intellegens.com/casestudies/

Contacts

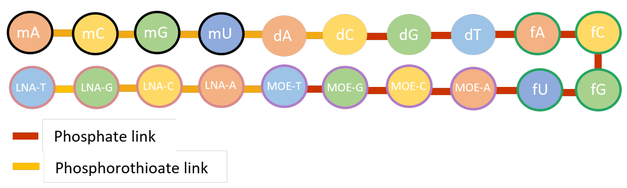
Claire MacLeod
Grand Challenge Lead,
CPI MMIC
claire.macleod@uk-cpi.com

Ben Pellegrini
CEO,
Intellegens
ben@Intellegens.com

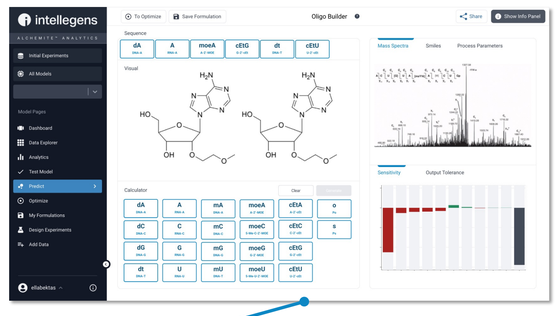
Expressions of interest welcome.

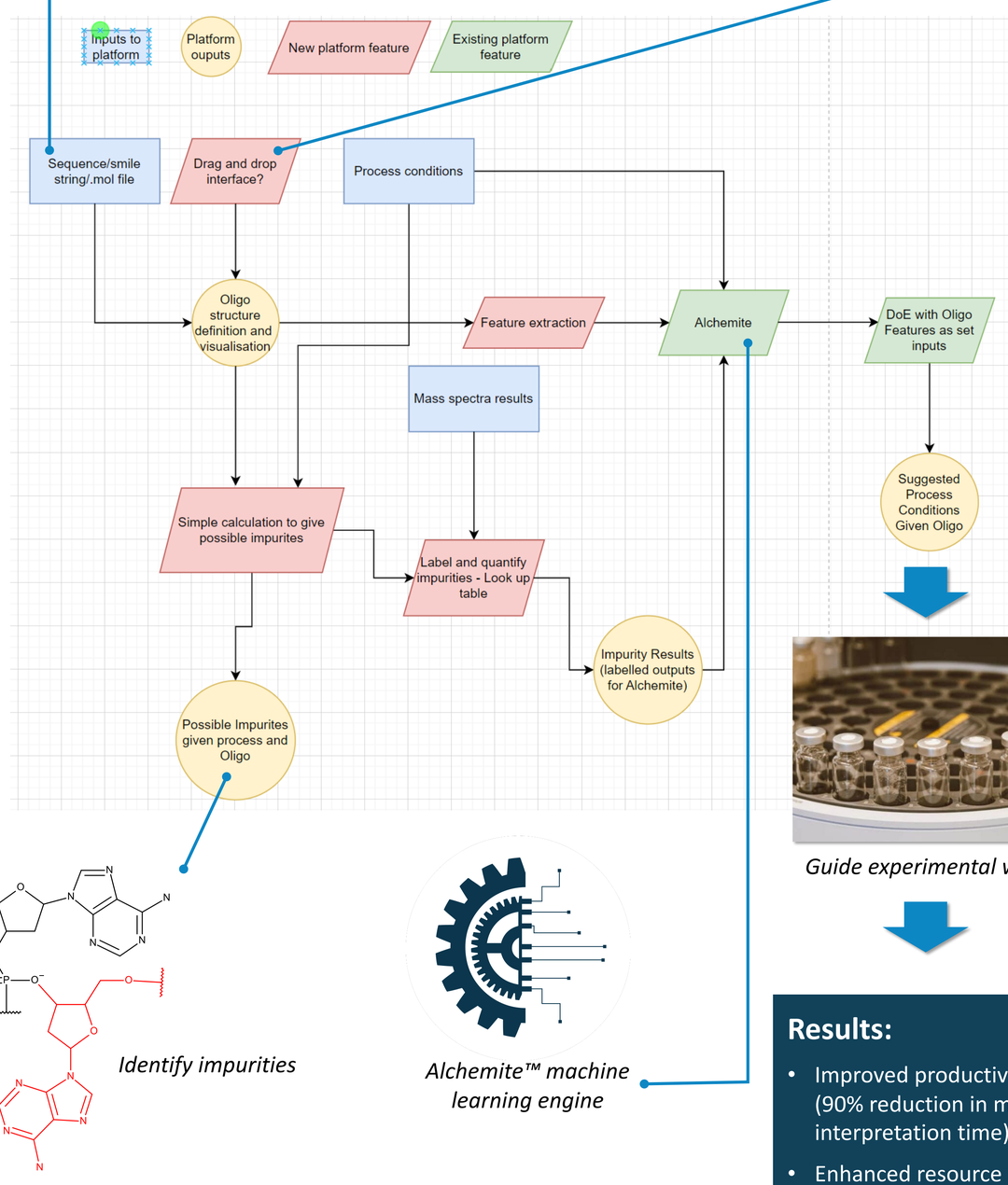
The proposed tool

Define structure from 'monomers'



Prototype user interface





Guide experimental work

Identify impurities

Alchemite™ machine learning engine

Results:

- Improved productivity (90% reduction in manual interpretation time)
- Enhanced resource efficiency (5% yield improvement)

Objectives

- Build and validate a machine learning-based tool to characterise oligonucleotides and guide manufacturing achieving TRL level 7
- Deliver a validated digital tool and high quality oligo synthesis / purification database
- Enable automated identification of impurities
- Achieve targets for productivity (90% less interpretation time) and yield (5% increase)
- Reduce development costs and facilitate manufacturing scale-up

Technical readiness level (TRL)