Roundtable Session 1 – Table 13 - Practical Applications of AI / ML in Discovery, Development and Platform Establishment Including CQA Controls

Facilitator: Keneshia Haenssen, *Arcus Biosciences*
Scribe: Xiaoyue Jiang, *Vir Biotechnology*

Abstract

Almost unquestionably, there is no more valid taste of the future for so many branches of science and technology than artificial intelligence and machine learning. The domain of drug discovery, development, and associated functions will present a particular abundance of such tools. More and more high-profile laboratories and institutions are rapidly, or at least tentatively, embracing technology and coming to terms with the immense utility and efficiency. AI/ML algorithms serve as innovative technologies that can be used to expedite the analysis of large and complex data sets to propose novel therapeutic targets, and to facilitate high-throughput screening of drug candidates. These algorithms may also be used to facilitate the establishment of continuously monitored platform processes that analyze and interpret data to provide real-time feedback for CQAs of interest.

AI based platforms can use historical and real-time data to predict chemistry-based outcomes including protein folding, stability, and degradation. In this roundtable, we will discuss insights, experiences and key opinions on potential workflows and nuances involved in the incorporation of automated process controls in the context of traditional predictive approaches.

Discussions and notes:

1. **Can someone share any experience with AI/ML used in Biopharmaceuticals? (Note all the discussions below are based on the assumption that AI/ML is set up for company's internal use, which is a closed system and not actually open source AI. The information that is input into the AI/ML is within the company only, and query responses are based on internally uploaded data from within the company.)**

- Some participants thought that the companies do not have sufficient data in CMC for machines to learn yet.
- In contrast, other participants working in research/discovery/early-stage development programs mentioned that they do have a lot of data ready for use. With 50,000 samples analyzed yearly, all the data can be fed into a database, where

AI/ML helps to integrate, score, process and organize into reports, which saves tremendous human power.
- There are vendors working on developing algorithms to put data into data management systems and users can get standardized workflows directly.

2. **In what area is AI/ML being most helpful for drug development?**
- AI/ML will be most useful in the area where prior knowledge is plentiful, eg. candidate screening and identification.
- AI/ML can also be used in CQA controls and post-translational modification monitoring. Eg. AI can monitor the deamidation levels from peptide mapping easily. The same concept can be applied to other structural elucidation.
- The algorithms can detect correlations between specific modifications or species and potency.
- The information on other molecules within the company database can prove very useful on predictions about structural or efficacy effects on the candidate molecule. This level of deep diving is harder to achieve manually.
- Training AI/ML does not need that many datasets. Using 13 datasets could offer 95% accuracy already.
- AI can also help to write INDs by pulling the results from the available report templates.
- AI can help to streamline the internal system, convert multi-group into a more streamed module for submission.

3. **How can we translate AI/ML's use in non-GMP labs to GMP environment?**
- Will need to start with non-GMP environment, which is also a process for analysts to get experience as prompt engineers.
- We can start by implementing simple AI/ML work into GMP environment followed by human verification first. Eg. AI gets the data captured, processed, and built into a report, all of which will still be reviewed by analysts.
- With the wide use of AI and the equivalency slowly built up between human's work and AI/ML's work, trust will be established, and the use of AI/ML will be more accepted.

4. **While large companies could leverage prior knowledge to feed into ML database to train the model, how could small companies without a lot of pipelines and experiences benefit from AI/ML?**
- Small companies can rely on CDMOs to leverage their knowledge from a similar molecule.

- Small companies might adopt AI at a slower pace at the beginning, considering the challenge of relying on the IT to set up the capability.
- Overall, if small companies only have a few modalities, the chance to implement AI/ML is low, as the immediate benefit cannot justify the cost to set up the IT infrastructure.

5. What are the challenges of implementing AI/ML?

- We don't know if the regulatory agency will accept it.
- It should be noted that AI is just a tool, and it is ultimately the human analyst's responsibility to review everything generated by AI is accurate before regulatory submission
- We need to make sure the correct model is used to train AI.
- Even though there is a lot of data generated, they are not quite accessible, especially across organizations.
- If methods get updated, the results might change potentially. This will require new training on the AI or establishing the correlation between old and new data to bridge.
- Data infrastructure can be a challenge. Eg. need to work with Amazon Web Services (AWS).

In summary, AI/ML will be able to process large quantities of data in a broader and more efficient way compared to what we have achieved manually. It has started to be used in different areas, especially in large companies, which included

1) Data integration, processing, report writing
2) Developability/molecule screening
3) Data organization

We expect further expansion of the AI/ML use rapidly. However, the acceptance of AI/ML in drug discovery and development can still be a challenge. The regulatory agencies will be cautious at the beginning and the sponsors need to ensure the data has been verified for accuracy. The acceptance of the AI/ML by regulatory agencies will be partially driven by how widely the tool will be used by the sponsors and how the data is presented.