

Proposal for Drug Discovery Coordination Platform for Biopharmaceutical Development

#Healthcare, QoL #IoT/Data Utilization #Industry & Distribution Solutions

Author

Nakamoto Yoichi

Systems Department 2, Pharmaceutical Systems Division, Enterprise Solutions Division, Industrial Digital Business Unit, Hitachi, Ltd.
Current work and research: Development and deployment of solutions as a leader working on drug discovery DX solutions for pharmaceutical manufacturers (Hitachi Digital Solution for Pharma).
Society memberships: Information Processing Society of Japan

Kido Kunihiko Ph.D.

Digital Healthcare Research Department, Healthcare Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research into healthcare AI.
Society memberships: IEEE

Kido Yutaro

Digital Healthcare Research Department, Healthcare Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research into healthcare AI.

Highlight

The pharmaceutical sector has seen a shift over recent years as the primary strategy for drug discovery has moved away from small-molecule drugs and toward biopharmaceuticals. The increasingly competitive field of developing next-generation antibodies such as ADCs and multi-specific antibodies faces a particularly high level of manufacturing difficulty due to the numerous challenges posed by the greater structural complexity of molecules compared to past antibodies, not the least of which is the difficulty of achieving reliable mass production quality. Moreover, how to address and resolve the downstream issues of CMC and manufacturing in the upstream basic research stages of drug discovery is a key challenge when seeking to improve the success rate and speed of the process.

This article describes the architecture of a drug discovery platform that seamlessly coordinates the different processes involved and a numerical analysis technique developed by Hitachi that enhances the value of process coordination in drug discovery.

1. Introduction

The pharmaceutical sector has seen a shift over recent years as the primary strategy for drug discovery has moved away from small-molecule drugs and toward biopharmaceuticals. Whereas small molecule drugs accounted for approximately 80% of drug approvals in Japan in the early 2000s, this has fallen to around 60% in the 2020s. Meanwhile, the share of biopharmaceuticals has grown annually to the extent that, by 2019, biopharmaceuticals made up approximately half of the top 100 drugs by international sales^{1), 2)}. Antibody drugs are a class of biopharmaceuticals that feature heavily in drug company development pipelines, with rising competition between companies to develop drugs that can target the limited number of promising antigens.

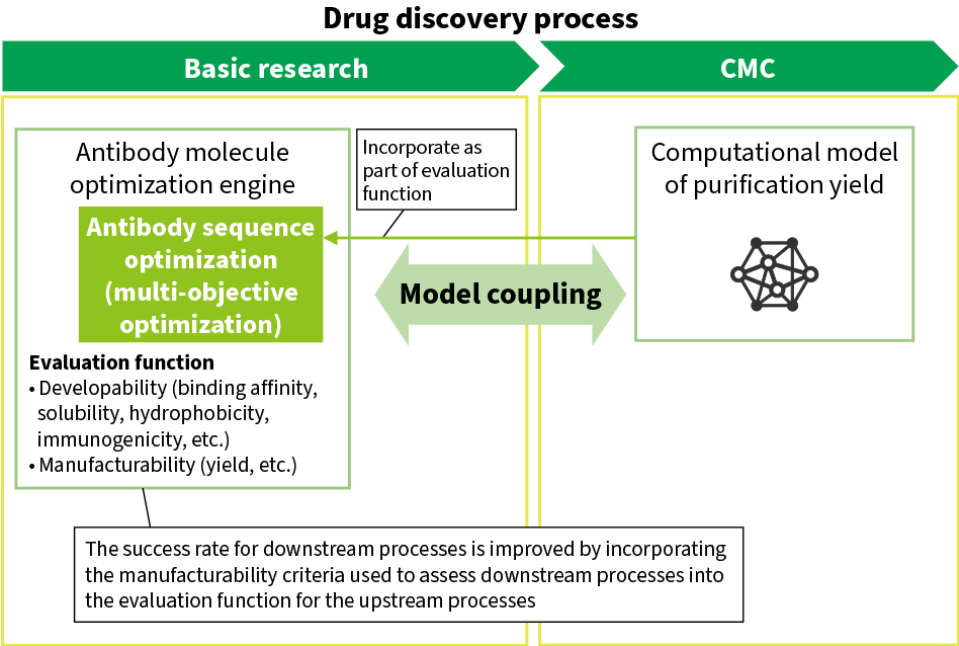
Amid these changes, pharmaceutical manufacturers, in seeking to obtain a competitive advantage, are shifting the focus of their research and development away from the monoclonal antibodies that predominated in the past and toward next-generation antibodies with high added-value such as antibody-drug conjugates (ADCs) and multi-specific antibodies. Meanwhile, one of the challenges of specialized next-generation antibodies is that they combine multiple functional components, making the quality of industrial production less reliable than that of conventional antibodies. This means that special measures are

needed, both in the chemistry, manufacturing, and control (CMC) steps at the downstream end of the drug discovery process and in commercial production. Immunoglobulin G (IgG) biospecific antibodies are one example of a commercially available multi-specific antibody medication. Here, commercial production was made possible by modifications to part of the antibody sequence. By altering the electrical charge characteristics of the antibody, these modifications made it easier to separate out the byproducts of the manufacturing process³).

To achieve greater competitiveness in the research and development of biopharmaceuticals such as next-generation antibody medications, coordinating the different steps in the drug discovery process will likely become more important than ever. This involves addressing the challenges posed by the downstream phases of CMC and manufacturing during early-phase basic research (see Figure 1). Whereas past practice was for the functional evaluation of manufacturability to be addressed at the downstream stages of drug discovery, it is anticipated that rework can be avoided by instead including this work in the functional evaluations conducted during antibody design (at the upstream end of the process). Other potential benefits include a faster overall drug discovery process and lower costs. Unfortunately, the different processes involved in drug discovery are not predicated on seamless integration, tending instead to be siloed between different departments or business units. That is, this work tends to be organized in such a way that most of the data and models generated at each step are only used within their own stage of the process, with no standardization of the metadata and interfaces for these data and models across the different activities. What is needed, then, is a platform that can manage the data and models across the entire drug discovery process, thereby enabling the different steps to be coordinated. Moreover, while such a platform may enable the exchange of data between the different process steps, this will only deliver adequate value if the platform also provides analysis techniques that can be applied to this data. One example is the need for numerical analysis to offer more sophisticated techniques such as multi-objective optimization so that the indicators used for downstream evaluation can be incorporated into the upstream evaluation function as described above.

This article presents an architecture for Hitachi’s drug discovery platform that is intended to overcome the challenges posed by process coordination. As an example of one of the numerical analysis techniques required for this coordination, the article also describes a search technique developed by Hitachi for antibody sequences that incorporates a method for performing multi-objective optimization in antibody sequence design.

Figure 1—Example of Process Coordination in Drug Discovery



CMC: chemistry, manufacturing, and control

The diagram shows an example use case involving the coupling of models for the upstream and downstream processes. The success rate for downstream processes is improved by incorporating the manufacturability parameters used to assess downstream processes into the evaluation function used for optimization of the upstream processes.

2. Drug Discovery Platform Architecture

Figure 2 shows the architecture proposed for Hitachi’s drug discovery platform. The purpose of the platform is to integrate data so that the same models can be used across different steps in the process as each step works through its own design-make-test-analyze (DMTA) cycle. Crucial to coordination of the drug discovery process is the coordinated management of data and models across all steps. The following three functions are provided for this purpose.

(1) Experiment planning and execution

Researchers and technicians conduct experiments on the basis of the experiment plans (experiment protocols) they devise. Then, they analyze the resulting data and plan the next experiment based on the results. If the experiment plan and methodology are poorly specified or documented, experimental repeatability may be diminished, resulting in greater variability in the data obtained. For example, if someone working at the same or a different step in the

process analyzes the data obtained from experiments, the reliability of the analysis results will be compromised if the analysis mixes data obtained by different experimental methods. This means that functions for experiment standardization are required to eliminate any uncertainties in the plan and procedure, and to ensure repeatability and that all experimental data is of consistent quality. Accordingly, Hitachi has adopted the “lab-as-code” concept and is developing standardization functions that are being built on a LabOps*1 base.

(2) Ontology management

Data integration requires the standardization not only of experiments, but also of the data itself. Given the different experiment objectives and instruments used, the experimental data obtained at each process step cannot be integrated in its raw form. Data integration involves pre-processing each item of experimental data in accordance with its metadata (information about experimental data items). Unfortunately, the metadata itself is likely to suffer from greater variability if people are free to specify it however they want, without restrictions. If this happens, this pre-processing will take a lot of time and effort. This makes it essential to design and manage metadata using a standard ontology framework that is suited to how data is used in the drug discovery process. This is an area where global standardization efforts are already underway and an ontology is being created using existing de-facto global standards as a base, along with additional elements that are needed for process coordination. These de-facto standards include the work done by the Open Biological and Biomedical Ontologies (OBO) Foundry in the research and development field and by the National Institute for Innovation in Manufacturing Biopharmaceuticals (NIIMBL) in the areas of CMC and manufacturing.

(3) Experiment life cycle management

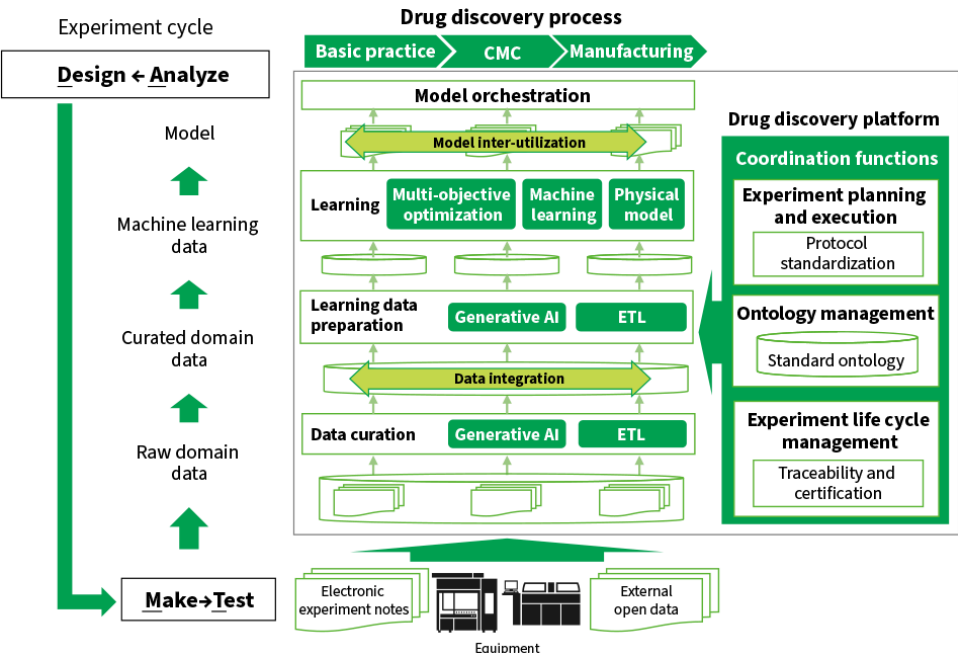
Researchers and technicians working through the DMTA cycle in drug discovery require extensive knowledge of information technologies such as machine learning in addition to their domain knowledge in fields like biology, pharmacology, and biochemical engineering. As projects require specialists from a range of different disciplines to work together, it is essential that an environment be created that improves project traceability and allows for knowledge sharing and the verification of work done. In practice, this involves providing the means for project team members to see what is going on and to enable detailed management of the experimental data and analytical models on the drug discovery platform. This latter includes keeping track of the experimental protocols used to acquire data, how the data was pre-processed, and what model tuning was performed. As work is already being done on the standardization of data processing and machine learning models by organizations such as the W3C*2 PROV and W3C ML Schema groups, Hitachi is developing functions for managing and tracking this information across the entire experiment life cycle by using this work as a base and extending it to cover the DMTA cycle.

As noted in the above function explanations, the drug discovery platform being proposed by Hitachi will use these functions to support seamless coordination both within and between processes as users work through the DMTA cycle for each step.

*1 A general term for systems and tools used to improve the efficiency of laboratory operation. These are used for purposes such as laboratory equipment management, data collection, process automation, and real-time monitoring.

*2 W3C is a registered trademark of the Massachusetts Institute of Technology.

Figure 2—Architecture of Platform for Coordinating the Drug Discovery Process



AI: artificial intelligence, ETL: extract, transform, load

The platform enables data and models to be shared across different steps in the drug discovery process as users work through the design-make-test-analyze (DMTA) cycle for each step.

3. Development of Hitachi’s Own Search Technique for Antibody Sequences

This section describes Hitachi’s search technique for antibody sequences using antibody sequence design as an example use case for process coordination.

Progress is currently being made on using in-silico drug discovery for the computer design of antibody sequences. This involves using a generative artificial intelligence (AI) model to produce antibody sequences in large numbers and then using machine learning and physical models to predict antibody functions from their sequences. These predictions are used to narrow down and optimize the candidate sequences. This use of computers to identify the most promising antibody candidates prior to performing wet-lab experiments can significantly reduce the time and cost of experimental work. Two key factors to consider in this sequence design process are multi-objective optimization and the generation of a diverse range of candidate sequences.

Antibody development involves progressively narrowing down the initial candidate antibodies identified by computer based on wet-lab experiments that assess them in terms of antibody functionality criteria such as binding affinity, solubility, and viscosity. This proceeds until only those candidates that satisfy all criteria remain. In doing so, it helps if the initial candidate antibodies have sequences that are predicted to satisfy as many of the criteria as possible. That is, it involves solving a multi-objective optimization problem that combines multiple evaluation functions. In the future, it is hoped that models for predicting antibody functionality can be further enhanced to achieve high accuracy by accumulating data from experiments that evaluate the antibodies based on a wide range of factors across different steps in the process, including basic research and CMC. This in turn will require multi-objective optimization that can handle even larger numbers of evaluation functions.

As it is not possible to accurately predict every physical property on a computer, the step of using wet-lab experimental testing to narrow down the candidate antibodies will always be needed. In doing so, as antibodies with similar sequences tend to have similar functionality, if all candidate antibodies have similar sequences, there is a heightened risk that all of the candidates will be eliminated from consideration if just one sample fails a particular wet-lab experiment. To avoid this and improve the odds of candidate survival, it is desirable to select candidate antibodies with as diverse a range of sequences as possible provided that they meet certain criteria.

Hitachi has developed its own search technique for antibody sequences that is based on GFlowNets⁴⁾, with additional improvements for use with multi-objective optimization problems⁵⁾. GflowNets has the ability to generate a diverse a range of sequences. When candidate antibody sequences generated using this technique were assessed on five indicators, including binding affinity and solubility, the pass rate of candidates that satisfied all of the evaluation criteria was 10 times higher than the method used previously. Hitachi is currently making further improvements to the technique prior to commercialization, including boosting learning efficiency and enhancing the search algorithm to identify better solutions.

4. Conclusions

This article has presented work being done by Hitachi to enhance competitiveness in future biopharmaceutical development, describing the concepts behind a platform for coordinating the drug discovery process and a search technique for antibody sequences that Hitachi developed itself. Product development for the platform is currently underway through collaborative creation with pharmaceutical manufacturers and other corporate partners, with the platform to be made available in the near future as part of the Hitachi Digital Solution for Pharma⁶⁾.

REFERENCES
1) Ministry of Economy, Trade and Industry, “Strengthening Bio CMO/CDMO” (Nov. 2020) in Japanese.
2) Office of Pharmaceutical Industry Research, “Trends in Drug Discovery Modalities for New Drugs” in Japanese.
3) Z. Sampei et al., Identification and Multidimensional Optimization of an Asymmetric Bispecific IgG Antibody Mimicking the Function of Factor VIII Cofactor Activity, PLOS ONE (2013.2)
4) M. Jain et al., Biological Sequence Design with GFlowNets, Proceedings of Machine Learning Research (2022)
5) T. Toyomura, et al., “Application of Multi-mode Sampling Generative Model to Antibody Sequence Generation and Improvement of Sequence Diversity,” 22nd Forum on Information Technology (FIT2023) (Sept. 2023) in Japanese.
6) Hitachi, Ltd., Hitachi Digital Solution for Pharma in Japanese.

Hitachi Review

Hitachi Review is a technical medium that reports on Hitachi's use of innovation to address the challenges facing society.

The *Hitachi Review* website contains technical papers written by Hitachi engineers and researchers, special articles such as discussions or interviews, and back numbers.

Hitachi Hyoron
(Japanese) website

<https://www.hitachihyoron.com/jp/>



Hitachi Review
(English) website

<https://www.hitachihyoron.com/rev/>



Hitachi Review Newsletter

Hitachi Review newsletter delivers the latest information about Hitachi Review when new articles are released.