Ontology-based Model Composition in Bioprocessing

The hereby described invention discloses a method and a system for integrating software-based models in a Digital Twin Application for simulating and improving a bioprocess.

Technical Field

The invention deals with the technological area of model engineering in the context of digital twin development for bioprocessing.

Background and description of the prior art

A digital twin is a virtual representation of a physical system or process that replicates its characteristics, behaviors, and interactions. A model is an essential component of a digital twin as it provides the foundation for simulating and representing the real-world system accurately. By simulating the model, the digital twin can replicate the system's performance, behavior, and response to different inputs and conditions.

Bioprocessing modeling involves the development and application of mathematical models to simulate and predict the behavior and performance of bioprocesses. It encompasses various stages of biomanufacturing, including upstream processes, e.g., cell culture and production or fermentation, downstream processes like purification, filtration and process integration.

Through the application of mathematical modeling and simulation, bioprocessing modeling contributes to the development of efficient, scalable, and cost-effective processes for the production of biopharmaceutical products.

Biological systems exhibit inherent variability due to factors such as genetic differences, cell line heterogeneity, and variations in physiological states. Thus, a model has to be adapted or newly built to account for the potential variances in biological behavior and response between the biological systems. Moreover, each bioprocess has its own set of operating conditions and equipment. Updating model parameters or model structure to handle variability then requires addressing each time a new biological system or a new bioprocess.

However, model engineering is a tedious task that requires specific modeling skills. Not all companies have modeling experts among their resources and therefore each time a new biological system or bioprocesses is studied the help of external consultants is required.

Therefore designing a digital twin in the industry can be a challenge. Main reasons for this are in particular the following points:

1.) The Model Selection: There might be multiple models available that represent different aspects of the physical asset or system. Choosing the most appropriate model that accurately represents the behavior and characteristics can be challenging. It requires understanding the specific use case, desired outcomes, and aligning the capabilities and limitations of the models with the requirements.

2.) The Model Accuracy: Ensuring the model's accuracy is critical for reliable simulations and analysis. It can be challenging to develop or identify a model that accurately captures the complex dynamics and interactions of the physical asset or system. Model validation against real-world data and thorough testing are essential to verify the accuracy of the chosen model.

3.) Model Complexity: Depending on the complexity of the asset or system, finding a model that strikes a balance between accuracy and computational efficiency can be challenging. Highly detailed models may provide accurate results but require significant computational resources and time, while simplified models may sacrifice accuracy for performance.

4.) Model Availability and Accessibility: Access to suitable and validated models can be limited. Acquiring or developing appropriate models may require specialized expertise, proprietary knowledge, or costly resources. Availability and accessibility of models can pose challenges, especially for specific industries or niche use cases.

Furthermore, in bioprocesses, many models exist but are working standalone so an end-to-end simulation can't leverage interoperable models, they all need some integration.

This leads to another problem when discussing the digital twin and standalone models – the lack of interoperability and the inability to leverage multiple models in an end-to-end simulation. In this context, interoperability refers to the seamless exchange of data and integration of different models within the digital twin ecosystem.

The key challenges associated with standalone models in digital twin applications include:

a) A Limited System Understanding: When models work independently and are not interoperable, the digital twin may lack a comprehensive understanding of the entire system's behavior. This hinders the ability to capture complex interactions and dependencies between different components or subsystems.

b) Incomplete Simulation Capabilities: Standalone models may provide valuable insights within their respective domains but fail to capture the holistic behavior of the entire system. This limits the accuracy and completeness of the simulations performed by the digital twin.

c) Missed Optimization Opportunities: Without interoperability, the digital twin cannot take advantage of the synergies and optimization possibilities that arise from integrating multiple models. This restricts the ability to identify system-wide improvements or optimized operating conditions.

The task of this patent application is now to address these challenges and enable interoperability and leverage multiple models in an end-to-end simulation within a digital twin framework.

Summary of the invention

This task has been solved by a method for integrating at least two software-based models in a Target Digital Twin simulating a bioprocess via a computer, comprising the following steps of a Standardized Model Integration Step wherein a common ontology is defined by a user for the Target Digital Twin including the at least two software-based models; A Data Integration Step wherein the computer uses the ontology to facilitates the integration of heterogeneous data from various sources and the at least two software-based models; A Data Mapping Step wherein the computer maps the ontology-integrated at least two software-based models to the designated bioprocess; An Inter-model Communication Step wherein the computer uses the ontology to exchange data about the Bioprocess between the at least two software-based models; A Simulation and Analysis Step, wherein the computer uses the ontology to align the at least two software-based models regarding their computations, inputs, and outputs to facilitate the integration of the at least two software-based models into an end-to-end simulation of the bioprocess in form of the Digital Twin; and an

Application Step of using the Digital Twin to simulate the bioprocess and its respective hardware instruments and optimize the hardware instruments and/or process settings with the respective simulation results. The ontology acts thereby as a mediator, enabling the at least two models to understand and interpret data shared between them, ensuring a coherent and consistent flow of information. It enables the creation of a comprehensive knowledge representation of the entire system. It captures the relationships, dependencies, and interactions between different entities and components, leading to a more holistic understanding of the system's behavior. Using ontology, the models can better align their computations, inputs, and outputs, facilitating the integration of models into an end-to-end simulation. This enables more accurate and comprehensive simulations, allowing for system-wide analysis and optimization. The ontology also provides a structured and extensible framework that can accommodate evolving models and system requirements. It enables the integration of new models or the modification of existing models without disrupting the overall interoperability of the digital twin ecosystem.

The solution provides furthermore several key features: Model Integration for establishing mechanisms for sharing, exchanging, and integrating data and information between different models. This enables a more comprehensive and interconnected representation of the entire system. Data Harmonization: to ensure that data formats, standards, and protocols are compatible across different models. This enables seamless data exchange and connectivity between models. And Synchronization and Coordination for coordinating the execution of multiple models to ensure consistent and synchronized simulations. This allows for a more accurate representation of system behavior and dynamic interactions. By promoting interoperability and leveraging multiple models in an integrated manner, the Digital Twin provides a more comprehensive understanding of the system, enable accurate simulations, and facilitate optimization opportunities for improved performance and decision-making. Furthermore by leveraging

ontology, the Digital Twin implementation can overcome the challenge of interoperability and establish a common foundation for integrating diverse models into the bioprocess. This promotes seamless communication, integration, and collaboration among models, enhancing the effectiveness and utility of the digital twin in simulating and analyzing complex systems.Ontology in this context means to provide a formal representation of knowledge within a specific domain, defining concepts, relationships, and properties.

Advantageous and therefore preferred further developments of this invention emerge from the associated subclaims and from the description and the associated drawings.

One of those preferred further developments of the disclosed method comprise that for the Data Integration Step a shared vocabulary and semantic framework is provided that enables semantic interoperability between different software-based models by aligning the meaning and interpretation of occurring data elements. The data elements usually include, but are not limited to, process parameters of the respective bioprocess and the used hardware instruments.

Another one of those preferred further developments of the disclosed method comprise that the output data of one of the at least two software-based models is used as input data to the next software-based model, enabling a user to choose different models for different bioprocess steps, hardware instruments or consumables and chain the models to create a simulation of connected hardware instruments and/or at least a part of the bioprocess for an end-to-end simulation. By generalizing the use of a common bioprocessing ontology for all used models the output of a model can be the input of the next. Customer can choose models for the different steps, equipment or consumable and they can be chained to create a simulation for at least a part of the bioprocess to an end-to-end simulation.

Another one of those preferred further developments of the disclosed method comprise that the bioprocess is performed using several different hardware instruments, like a bioreactor and a pump, wherein the at least two software-based models simulate the behavior of the different hardware instruments with the hardware instrument settings and parameters and the bioprocess parameters being the input and output data of the at least two software-based models. The approach of chaining or mapping different models to create an end-to-end simulation for the bioprocess is not limited to the components being specifically hardware instruments, but those instruments, like a bioreactor and peripheral devices like a pump, are the most common required components.

Another one of those preferred further developments of the disclosed method comprise that the at least two software-based model's input and/or output data comprises of parameters like the physical quantity, in particular temperature or flow rate, the units of measurement, in particular °C or ml/min, the model process parameters, the model quality attributes or relevant chemical elements. These are the most common parameters, but all suitable process or hardware instrument parameters which can be used in a Digital Twin are part of the disclosed method.

Another one of those preferred further developments of the disclosed method comprise that for the Standardized Model Integration Step the at least two software-based models are contextualized by applying the ontology, so that the ontology domain contains all the necessary data to represent the bioprocess and the respective models. This step allows to standardize and harmonize the used description. It also promotes consistency and compatibility, allowing models and process to interoperate seamlessly by sharing a common understanding of concepts and relationships.

Another one of those preferred further developments of the disclosed method comprise that the necessary data comprises information of for which stages and/or operations of the bioprocess a model can be applied, the type of required hardware like bioreactor or pump, hardware characteristics like size, volume, or instrument family and characteristics of the final product of the simulated bioprocess that the respective models to addresses like model effeciency on a specific cell line or clone. Again this data are only the most common used information. For the disclosed method every suitable process data can be used as long as it still allows for linking between the different models and process elements.

Another one of those preferred further developments of the disclosed method comprise that the respective simulation results from the Application Step are additionally used for predictive maintenance of the hardware instruments. In particular the bioreactor used for applying the bioprocess would be suitable for predictive maintenance, since its maintenance procedures are usually the most expensive ones and its maintenance schedule, considering the fact that the grown cell cultures in the bio reactor must under no circumstances perish and therefore a malfunction of the bioreactor has to be avoided at all costs, the most tight from all the hardware instruments. But of course a predictive maintenance schedule can be created or improved using the simulation results for all hardware instruments which parameters are calculatedby the at least two models.

Another solution of the given task is a System for integrating at least two software-based models in a Target Digital Twin of a Bioreactor System performing a bioprocess comprising a computer, a database connected to the computer with several software-based models to select the at least two software-based models from, an ontology database connected to the computer, input and output means connected to the computer for entering and outputting data, wherein the system is configured to perform the previously described method steps.

A further solution includes a computer program product and a computer-readable storage medium and/or data carrier signal having stored thereon the computer program product, which comprises instructions which cause the involved computers to perform the method steps of the previously described methods.

Detailed description of the invention

The method and system according to the invention and functionally advantageous developments of those are described in more detail below with reference to the associated drawings using at least one preferred exemplary embodiment. In the drawings, elements that correspond to one another are provided with the same reference numerals.

The drawings show:

Figure 1: a schematic organigramm showing nexessary steps to result in a final product starting with a CHO cell line

Figure 2: a schematic showing a representation of a knowledge graph of the bioprocess using the ontology

Figure 3: a schematic showing a representation of the knowledge graph for the chromatography model

Figure 4: a schematic showing a representation of the knowledge graph for the mixer model

Figure 5: a schematic showing the final mapped models enabling the end-to-end simulation of the respective CHO cell line bioprocess

One exemplary preferred embodiment of the invented method will be described in the following. The steps itself are performed divergent in every exemplary embodiment dependent on the different conditions. In the

chosen preferred embodiment the target is to produce mAB therapeutics. Figure 1 shows the different steps to produce the final product starting with a Chinese Hamster Ovary (CHO) cell line. The shown hardware infrastructure used for the embodiment is exemplary for this preferred embodiment. It can change for a other embodiments. Especially the kind of involved computers can differ greatly, depending on how much of the steps is performed by human users with the help of computers and application software or done automatically by specific computers using for instance AI based software.

The goal is now to demonstrate the feasibility of temporal evolution, in continuous mode, of the concentration of the target product. This will be done by leveraging macroscopic models.

The representation in the knowledge graph of the process using the ontology for this preferred embodiment when focusing on the chromatography stage is shown in Figure 2.

For this specific preferred embodiment now two different models are required. One for the multi-chromaography column: It will act as an orchestrator for the various pumps injecting either the solution of interest coming from the bioreactor step or a buffer solution, and represent the execution of successive steps of a chromatography cycle within a two-column multi-column chromatography system. Another model is required for the mixer which will homogenize the solution in order to adjust the pH.

The respective representation in the knowledge graph of the chromatography model is shown in Figure 3, while the used model for the mixer is disclosed by Figure 4.

After all required data has been made available in the knowledge graph, the Digital Twin of the chromatography stage is designed by determining the

suitable models that are used in this stage. As figure 5 discloses, the output of the multi-column-chromatography model is then mapped to the input of the mixer-model by using the same concept of physical quantity. Here that includes the mass flowrate and the pH values. For the input of the multi-column-chromatography model, the output of the previous step, which in this case would be a bioreactor modeling step, is mapped by using again the common concept of physical quantity, which includes here again the mass flowrate and the pH values. The same accounts for the output of the mixer-model whereby here the output of the mixer model is limited to the mass flowrate of the product of interest.

If a full end-to-end simulation is intended in a further preferred embodiment the output of the mixer model could to be connected to another model until the full chain is created. Any last model in this chain would then provide final simulation result.

By applying this principle a user can chain different models to create the end-to-end simulation of the respective simulated bioprocess.

Patent claims

1. Method for integrating at least two software-based models in a Target Digital Twin simulating a bioprocess via a computer, the following steps comprising:
   - A Standardized Model Integration Step wherein a common ontology is defined by a user for the Target Digital Twin including the at least two software-based models;
   - A Data Integration Step wherein the computer uses the ontology to facilitates the integration of heterogeneous data from various sources and the at least two software-based models;
   - After these two steps a Data Mapping Step wherein the computer maps the ontology-integrated at least two software-based models to the designated bioprocess;
   - An Inter-model Communication Step wherein the computer uses the ontology to exchange data about the Bioprocess between the at least two software-based models;
   - A Simulation and Analysis Step, wherein the computer uses the ontology to align the at least two software-based models regarding their computations, inputs, and outputs to facilitate the integration of the at least two software-based models into an end-to-end simulation of the bioprocess in form of the Digital Twin; and
   - An Application Step of using the Digital Twin to simulate the bioprocess and its respective hardware instruments and optimize the hardware instruments and/or process settings with the respective simulation results.

2. Method according to claim 1, wherein
for the Data Integration Step a shared vocabulary and semantic framework is provided that enables semantic interoperability between

different software-based models by aligning the meaning and interpretation of occurring data elements.

3. Method according to any of the previous claims, wherein the output data of one of the at least two software-based models is used as input data to the next software-based model, enabling a user to choose different models for different bioprocess steps, hardware instruments or consumables and chain the models to create a simulation of connected hardware instruments and/or at least a part of the bioprocess for an end-to-end simulation.

4. Method according to claim 3, wherein the bioprocess is performed using several different hardware instruments, like a bioreactor and a pump, wherein the at least two software-based models simulate the behavior of the different hardware instruments with the hardware instrument settings and parameters and the bioprocess parameters being the input and output data of the at least two software-based models.

5. Method according to claim 4, wherein the at least two software-based model's input and/or output data comprises of parameters like the physical quantity, in particular temperature or flow rate, the units of measurement, in particular °C or ml/min, the model process parameters, the model quality attributes or relevant chemical elements.

6. Method according to any of the previous claims, wherein for the Standardized Model Integration Step the at least two software-based models are contextualized by applying the ontology, so that the ontology domain contains all the necessary data to represent the bioprocess and the respective models.

7. Method according to claim 6, wherein
the necessary data comprises information of for which stages and/or operations of the bioprocess a model can be applied, the type of required hardware like bioreactor or pump, hardware characteristics like size, volume, or instrument family and characteristics of the final product of the simulated bioprocess that the respective models to addresses like model effeciency on a specific cell line or clone.

8. Method according to to any of the previous claims, wherein
the respective simulation results from the Application Step are additionally used for predictive maintenance of the hardware instruments.

9. System for integrating at least two software-based models in a Target Digital Twin of a Bioreactor System performing a bioprocess comprising a computer, a database connected to the computer with several software-based models to select the at least two software-based models from, an ontology database connected to the computer, input and output means connected to the computer for entering and outputting data, wherein the system is configured to perform the method steps of claims 1 to 8.

10. Computer program product comprising instructions which cause the involved computers to perform the method steps of claims 1 to 8.

11. Computer-readable storage medium and/or data carrier signal having stored thereon the computer program product of claim 10 which cause the involved computers to carry out the method steps of claims 1 to 8.