eppendorf

Boost Your (Stem) Cell Culture

We assist you to scale up your bioprocess –
Eppendorf Bioprocess Solutions for Cell & Gene Therapy Development - Flexibe, Scalable, Industrial

The BioFlo® 320 offers flexibility, better control, and maximum functionality while occupying a fraction of the valuable lab space of similar systems. This means greater efficiency and productivity at a lower operating cost for your lab. BioBLU® Single-Use Bioreactors were developed as true replacements for existing reusable vessels.

- > Sterility Assurance Level (SAL): 10^{-6}
- > Simplified handling reduces cross-contamination
- > Reliable scalability from 250 mL - 40 L through industrial design
- > Proven for animal and human cell lines
- > Increased productivity with reduced turnaround time between runs

www.eppendorf.com/BioBLUc

Received: 28 May 2020 | Revised: 26 July 2020 | Accepted: 27 July 2020

DOI: 10.1002/bit.27520

REVIEW

BIOTECHNOLOGY BIOENGINEERING WILEY

Toward in silico CMC: An industrial collaborative approach to model-based process development

David Roush[1] | Dilip Asthagiri[2] | Deenesh K. Babi[3] | Steve Benner[1] | Camille Bilodeau[4] | Giorgio Carta[5] | Philipp Ernst[6] | Mark Fedesco[7] | Sean Fitzgibbon[8] | Matthew Flamm[1] | Jan Griesbach[9] | Tobias Grosskopf[10] | Ernst B. Hansen[3] | Tobias Hahn[11] | Stephen Hunt[8] | Francis Insaidoo[1] | Abraham Lenhoff[12] | Jasper Lin[7] | Henrik Marke[3] | Bruno Marques[13] | Emmanouil Papadakis[3] | Fabrice Schlegel[14] | Arne Staby[3] | Marcel Stenvang[3] | Larry Sun[14] | Peter M. Tessier[15] | Robert Todd[8] | Eric von Lieres[16] | John Welsh[1] | Richard Willson[17] | Gang Wang[18] | Thomas Wucherpfennig[18] | Oleksandr Zavalov[1]

[1]Merck & Co., Inc., Kenilworth, New Jersey

[2]Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas

[3]Novo Nordisk A/S, Bagsvaerd, Denmark

[4]Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, New York

[5]Department of Chemical Engineering, University of Virginia, Charlottesville, Virginia

[6]Bayer AG, Leverkusen, Germany

[7]Genentech, Inc., San Francisco, California

[8]KBI Biopharma, Boulder, Colorado

[9]F. Hoffman-La Roche, Basel, Switzerland

[10]Roche Diagnostics GmbH, Penzberg, Germany

[11]Karlsruhe Institute of Technology, Karlsruhe, Germany

[12]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware

[13]Century Therapeutics, Philadelphia, Pennsylvania

[14]Amgen, Inc., Cambridge, Massachusetts

[15]Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan

[16]Institute of Bio- and Geosciences 1, Research Centre Julich, Julich, Germany

[17]Department of Chemical and Biomolecular Engineering, University of Houston, Houston, Texas

[18]Boehringer Ingelheim, Ingelheim, Germany

Correspondence
David J. Roush, Process Research and Development, Merck & Co., Inc., 2000 Galloping Hill Rd, K15-H206, Kenilworth, NJ 07033.
Email: david_roush@merck.com

Abstract

The Third Modeling Workshop focusing on bioprocess modeling was held in Kenilworth, NJ in May 2019. A summary of these Workshop proceedings is captured in this manuscript. Modeling is an active area of research within the biotechnology community, and there is a critical need to assess the current state and opportunities for continued investment to realize the full potential of models, including resource and time savings. Beyond individual presentations and topics of novel interest, a

substantial portion of the Workshop was devoted toward group discussions of current states and future directions in modeling fields. All scales of modeling, from biophysical models at the molecular level and up through large scale facility and plant modeling, were considered in these discussions and are summarized in the manuscript. Model life cycle management from model development to implementation and sustainment are also considered for different stages of clinical development and commercial production. The manuscript provides a comprehensive overview of bioprocess modeling while suggesting an ideal future state with standardized approaches aligned across the industry.

KEYWORDS

computational fluid dynamics, mechanistic modeling, molecular modeling, plant simulation

1 | INTRODUCTION

Modeling is an active area of research within the biotechnology community. There is a critical need to assess the current state and opportunities for continued investment to realize the full potential of models, including resource and time savings. The Modeling Workshops were initially established as a grass roots movement to share knowledge and experience with mathematical modeling in the industry and to enhance its implementation. Over many conferences and discussions, there was one unifying theme that has been clearly identified. What problems are we trying to solve? This question is juxtaposed with the ultimate challenge to mature modeling to the point where it could potentially replace experimentation.

Hence, the goal of the Third Modeling Workshop, held in May 7 to 9, 2019, was to highlight and assess the current state, challenges, solutions, and opportunities for achieving a future state where a combination of modeling tools can be utilized as a complete simulated plant/digital twin. Sessions were specifically focused on key modeling areas currently in development including molecular modeling, mechanistic modeling, computational fluid dynamics (CFD), and plant simulations. An open challenge session was also included to assist in identification of nascent challenges and tools that are highly interconnected and could be developed to advance the various fields of modeling. One key observation was that models are interconnected (e.g., use of mixed mode isotherms for multiple applications—mechanistic, biophysics, and even plant simulation). Consistent with the evolution of the modeling field, the event has now been converted to a Recovery of Biological Products Workshop and will be distributed to a broader audience for future meetings.

2 | OVERVIEW OF WORKSHOP OUTCOME

The potential exists to utilize and combine a range of modeling tools from molecular design to plant simulation to prospectively evaluate the impact of new technologies, process optimizations, and process changes (see also Figure 1) for a range of products including recombinant proteins and polypeptides. Models can be used to capture these process aspects and insights both quantitatively (e.g., mechanistic models for chromatography) and qualitatively/directionally (e.g., feasibility of utilizing a particular ligand for separation).

One key observation is that models transcend scales of processing (see Figure 2 below). The scale, complexity, potential accuracy, and perceived value of the different bioprocess modeling approaches for design and optimization are shown schematically. Statistical/design of experiments (DOE) based approaches do not require and do not provide mechanistic understanding and can be safely used only to interpolate within experimental conditions (Staby, Ahuja, & Rathore, 2017). At the smallest scale, molecular level models have the potential to provide the greatest possible understanding of molecular interactions as well as true predictions. At the microscopic scale, mechanistic models describe all aspects of the process based on verified physical chemical laws, require in-depth process understanding, and can be used to extrapolate beyond experimental ranges. Hybrid models are intermediate between fully empirical and truly mechanistic approaches. They describe certain well-understood aspects of the process using physical chemical laws and other, less clearly understood aspects through statistical correlations of empirical data. While a decade ago the perceived practical value of modeling for bioprocess design and optimization centered around DOE-based approaches, recent advances have shifted the value curve upwards toward mechanistic approaches and principal component analysis (PCA)/partial least squares (PLS) statistical models or hybrid models (combination of empirical and mechanistic). The Workshop consensus is that, today, hybrid modeling approaches are likely most beneficial in bioprocessing. These approaches combine advances in high-throughput measurements with fundamental understanding of the relevant physical chemical phenomena along with efficient computational technologies. Although these models were developed based on proteins, the applications and approaches can be applied to new biological formats (e.g., conjugates or virus-like particles; Ladd Effio et al., 2016).

A future state may exist where one can leverage databases and predictive models to perform most development efforts in silico. An

	Biophysics	Mechanistic	CFD/Mixing	Plant Model/Facility
Early Stage Process design qualitative models	• Sequence and structure are known • Determine liabilities, including PTM • Rank order candidates to process fit or identify changes required • Downstream perspective: define key unit operations to afford retention and selectivity • Reaction mechanism: detailed understanding to understand impurity profile, time course, assess potential reaction pathways • Evaluate formulation excipients in silico	• Impact of upstream and downstream conditions on CQA (discrete steps) • Define initial control strategy • Primary portion of the filing is modeling • Limited number of experiments to support FIH process (by exception) • Linkage of structure to mechanistic modeling	• Digital twin of equipment exists to assess impact of scale-up/ scale-down and mixing parametric in silico • Modeling solely to support scale-up in filing	• Assess impact of non-platform (different from initial design) to existing facility – linked to process model • Modeling solely to support scale-up and facility fit
Late Stage Characterization validation and quantitative models	• Quantitative predictions of effect process parameter changes on CQAs are feasible including stability based on combination mechanisms and biophysics (file modeling data, supplemental data if different) • Ability to predict macroscopic properties (including viscosity) as a function of concentration	• Upstream model combining all elements of mechanistic understanding (linked, quantitative) integrated with, i.e., stability parameters of the molecule • Unified model (across modeling techniques including linkage to structure) • Downstream: map directly from HTPD to production scale based on first principles (confirmatory runs at scale) • Digital twin of equipment and process exists to assess impact of scale-up/ scale-down and process performance in silico	• Digital twin of equipment exists to assess impact of scale-up/ scale-down and mixing parametric for process characterization (scale-down model qualification)	• Assess impact of connecting various plant components • Assessment of potential impact of range of facility output • Potential for facility transfer amongst
Commercial Scale up, transfer, plant simulation	• Assess impact of RM variability (e.g., resin properties) on process performance; combine with mechanistic prior to implementation or deviation investigation • Allow for more simplified model to support commercial production; more detailed model for investigation	• Unified model (across modeling techniques including linkage to structure) • Late stage + facility • All of the above can be integrated with facility-specific parameters (i.e., UF/DF); different unit operations can be connected	• Digital twin of equipment exists to assess impact of scale-up/ scale-down and mixing parametric for equipment change (PCV) to support risk assessment or facility change	• Plant model is overarching with significant details and interfaces • Connect individual models to plant (in silico assessment end to end) including productivity assessment • Fully integrated facility models with different layers for different products. Allows optimal facility utilization, maybe even across manufacturing networks

FIGURE 1 Future state in silico chemistry, manufacturing, and control (Prospective and predictive artificial intelligence) [Color figure can be viewed at wileyonlinelibrary.com]

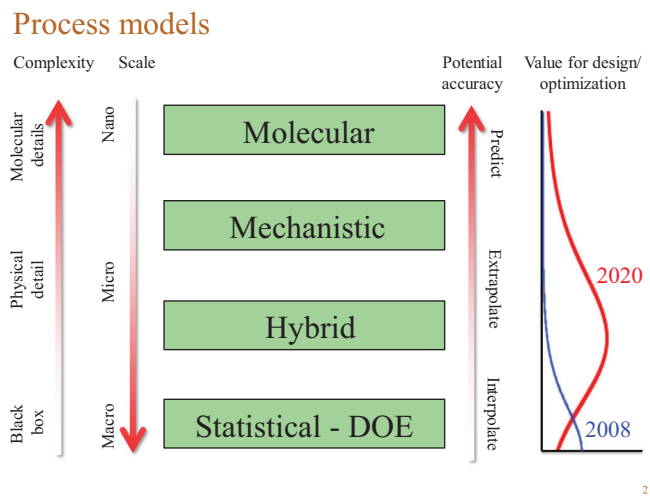

FIGURE 2 Evolution of process models (tradeoffs of applicability and prediction; G. Carta). The scale, complexity, potential accuracy, and perceived value of the different bioprocess modeling approaches for design and optimization are shown schematically [Color figure can be viewed at wileyonlinelibrary.com]

initial Workshop mapping of this functionality to the development state and by modeling activity is presented in Figure 1.

A key element for industry to broadly embark on the future-state modeling approach is the existence of a proper business case for implementation. Although significant investments may be required, including development of more refined, efficient, and quantitative algorithms as well as new tools, utilization of the existing tools in a phase appropriate manner can result in significant resource and time savings; for example, postponement or circumvention of factory investments (Hansen, 2017) and 10+-fold increase in productivity as presented at the Workshop. Scale of capital investment avoidance has significant financial return (10 or 100 s of millions of US dollars). Ideally, models would be fully quantitative and could potentially replace experimentation if that aspirational goal were achieved. Already today models can provide directional guidance to support development (e.g., focus research on areas with higher probability of success) and address regulatory agency questions that cannot be readily explored experimentally.

The impact and return on investment for a model depends on the scope and type of application. For example, empirical models that employ data and statistical approaches for a specific project or

facility will generate value in a linear fashion. Mechanistic models developed on first principles (e.g., chromatographic governing equations, cell growth and metabolism, and mass transfer) can require more initial investments and more time to develop, yet lead to exponential value creation, since they have the potential to be predictive. For example, a combination of models (CFD, mechanistic chromatography modeling, molecular biophysics, and plant simulation) could be performed prospectively to evaluate the impact of proposed equipment or facility changes on productivity or product quality (Figure 1). If the results of these models indicate no impact on critical quality attributes (CQA) then the results of these models could support a decision to implement the changes to equipment, facility, or operating conditions without the cumbersome need to qualify them under protocol.

One of the challenges in model development is mapping the type of model to the specific application. As mentioned in the previous section, simple statistical models can be quite beneficial to support specific applications, projects, or facilities. However, significant value is gained via development of more generalized models, either empirical or predictive. The timing for development and uses of a model may also be dependent on the type of application. For example, one could envision creating a DOE and interrogating process options with a hybrid approach combining modeling and limited number of experiments. This approach could be employed either with directional models to identify key parameters (early stage) or more quantitative models (late stage) depending on the application.

One key area of consensus from the Workshop is that a systematic approach to model development, maintenance, and replacement is required to maximize the return on investment. The key areas of the model lifecycle are captured in Figure 3 (Rolandi, 2019). During the development stage, tools may be directional in nature and used to guide research. Once the tools are deployed and additional data are obtained as well as more clarity obtained on user requirements, the tools are refined. One of the challenges often encountered in the deployment space is the limitation of the amount and quality of data to validate or qualify the tools. This is particularly critical in some areas of modeling such as computational biophysics where access to high quality and diverse data sets is limited. A key area for investment that was identified at the Workshop is the need to perform prospective experiments to validate tools and models with a standard set of conditions. One potential solution proposed would be the formation of consortia to execute these experiments and critically evaluate the performance of the models during the validation expectations (qualitative/directional for early stage and quantitative for late/commercial stage). Translation of this concept to reality has been initiated via the Recovery of Biological Conferences Highland Games (Coffman et al., 2020).

In the following sections, detailed discussions and results of the various Workshop themes are summarized and divided into four sessions: molecular modeling, mechanistic modeling, CFD, and plant simulation, and subsequently the more detailed outcome of the Open Challenge and Future State sessions of modeling are provided. We hope that these discussions will pave the way to achieve the suggested future state for modeling.

2.1 | Molecular modeling

A new but important topic at the Workshop was molecular modeling. Elements of consideration for use and general status of the potential for the field of Biophysics are presented in Figure 1 and discussed in more detail below. Molecular models can be viewed as a mechanistic model at a specific scale built on first principles. A typical approach for therapeutic antibody discovery involved the humanization of a

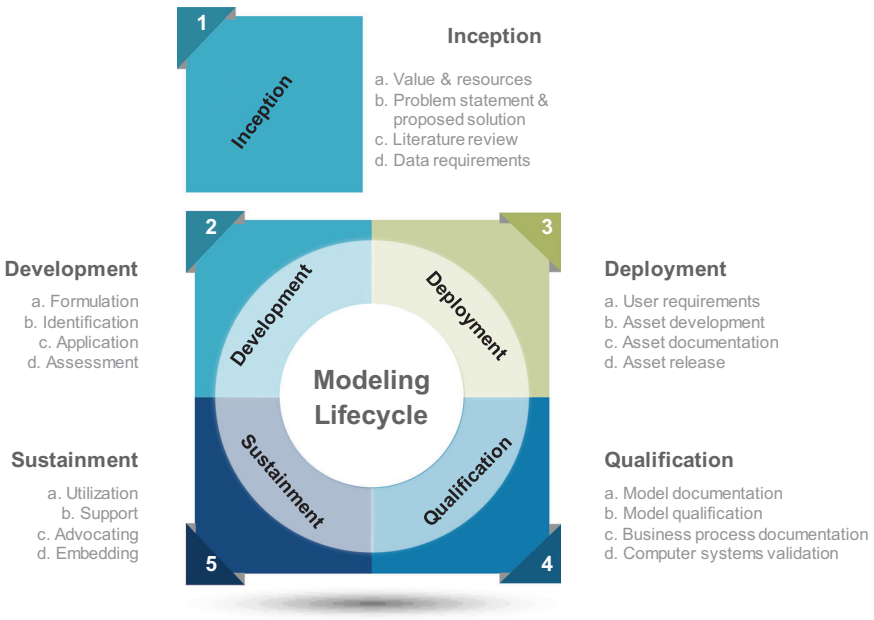

FIGURE 3 Summary of modeling lifecycle (adapted from Rolandi, 2019) [Color figure can be viewed at wileyonlinelibrary.com]

potent murine antibody against a specific antigen or optimization of a lead candidate to improve potency or reduce liabilities. The primary objective at this stage in the discovery process is to improve affinity to a specific antigen and potency. Using sequence alignment and structural information, site specific mutations can be introduced to generate a clinically relevant molecule with the desired therapeutic properties. Yet, the therapeutic potency of an antibody does not directly translate into an ideal process development and manufacturing conditions.

Early development assays provide relevant information during discovery but are limited in predicting the behavior of monoclonal antibodies (mAbs) under relevant process development and manufacturing conditions. Transient transfections typically do not provide the yield and glycosylation profiles as a stable pool cell line. Also, concentrations of the purification pools and purification steps hardly aligns with process development conditions or eventual manufacturing conditions. Thus, if the relevant developability considerations are not accounted for the selection of a lead candidate is subjective and might lead to problems downstream.

Given the vast space covered during the discovery phase it is not possible to perform all the relevant experiments on potential candidates. A synergy between in silico prediction and experimental studies has the potential to expand the search space computationally and focus experimental efforts on relevant constructs with improved developability parameters. However, there is a lack of relevant experimental data for a diverse set of antibodies to identify specific structural features that predict salient parameters that will inform developability.

Upstream process development is mainly interested in expression and efforts to increase the titer at harvest. Additional consideration during upstream processes is to minimize posttranslational modification (PTM) and impurities (host cell protein [HCP], lipase, etc.) that might be produced as a byproduct. Factors that influence successful downstream processes include metabolite consumption and byproducts, amino acid and nutrient consumption, aeration, agitation, and viability. With regard to computational analysis and optimization of these factors, a critical question that remains unanswered is the extent to which the DNA construct (plasmid) or the protein sequence it encodes influence upstream processes. Currently, enough data exist to optimize protein sequence to minimize a limited set of posttranslational modifications (glycosylation, deamidation, etc.). Other unanswered questions include: is it possible to predict the upstream process outcome based on antibody DNA or protein sequence? Is there a link between cellular consumptions and the type of protein produced? Can HCP be modulated through lead candidate selection or is it a function of production titer and cell culture stress.

If PTMs, lipase levels, and HCPs are controlled upstream, it will reduce the burden on downstream process development. If not, tools need to be developed to address these impurities through downstream processes. Addressing these challenges during downstream process development will require an entirely different set of tools and considerations. For the removal of PTM variants, biophysical differences between the main molecule and PTM variants can be exploited to separate these impurities. Thus, a structural insight into the PTM variants is required. The challenge is that it is not possible to predict all possible variants and, even if possible, can any observed or calculated difference be exploited to optimize a purification process? Presence of lipases in the final drug product has remained active even at sub-parts per billion (ppb) levels. How these molecules remain present through a series of purification steps is not completely understood. One possibility is that some lipases associate/bind to the antibody (or other target proteins) and do not dissociate through the purification process. If this is the case, some level of thermodynamic analysis of the interaction between antibody/target protein and lipase is required. Another challenge is that there is not a complete set of characterized lipases that is present in Chinese hamster ovary (CHO) cells. Knowledge of the sequence and structure of these lipases will significantly transform our understanding of this purification impurity. A similar set of considerations exist for clearance of other HCPs. The diversity of these impurities and lack of comprehensive characterization of the variants presents a challenge for the field.

Additional consideration for downstream process optimization is to define and better understand the principles that govern purification modalities (chromatographic and non-chromatographic). Process development can be accelerated by minimizing the number of experiments through computational modeling. The driving forces that modulate protein retention on a chromatographic column are not easily predicted by looking at the structure or biophysical properties of the antibody. Drawing a correlation between structural properties and its influence on binding and elute or flow-through operating conditions will be a significant advancement of the field. One limitation is our inability to completely characterize aggregation. Efforts to distill the types of aggregates generated and their structural features will open the space for novel tools to be developed to address these challenges. Beyond packed beds, non-chromatographic purification methods have the potential to expand the current tools in this field.

During formulation, the stability of the final drug substance is critical for extending shelf life and potency. The molecular structure can influence self-association and aggregation under higher protein concentration and as a function of excipients, buffer and pH conditions. The influence of these parameters (excipients, buffer, and pH) on protein stability, solubility, and aggregation is determined by the structure and biophysical properties of the biological molecule.

Even though antibodies are structurally very similar, there is significant diversity of the biophysical properties of most of the molecules that have made it to different stages in clinical development. In addition, there is an infinite number of possible permutations of antibodies possible with specificity toward different antigens. The diverse possibility of properties makes predictive assessment of antibody properties challenging. Further, the number of antibody molecules that have undergone some level of clinical development is a significantly limited subset of all the possible variants of antibodies possible in nature. A third limitation to predicting biophysical

properties that influence process development is the diversity in process development platforms and an even lower number of molecules that have been developed by each unit operation. Consolidation of data from different organizations and molecules can improve the odds of a successful predictive too if the data is normalized across platforms.

As more data becomes available through the formation of consortiums and the development of novel molecules, there will be a point where the generated data can be used to predict biologics development. At this current state, the computational cost of building representative homology models, calculation of biophysical descriptors, and analysis of the data to either understand or predict experimental outcome is not a limiting factor. To make meaningful correlation between the in silico biophysical calculation and experimental outcome, both sets of data need to be standardized or normalized. Normalizing experimental data can be challenging when generated under different conditions as required for clinical development. An alternative approach would be to specifically mutate molecules to generate a diverse set of molecules that can be studied under similar experimental conditions. Yet, as new predictive tools are developed, testing the algorithms against a defined set of molecules will be an effective way to benchmark the success of each predictive tool. This will be an effective cross validation and an iterative approach to determine progress within the field of predictive bioprocess development.

A comprehensive process predictive algorithm would start from humanization, through sequence optimization, liability elimination, plasmid design, expression and scale-up, downstream purification modality prediction, and formulation conditions. Experimentally, each of these steps generate specific experimental data which are a function of the molecular biophysical properties. Identifying the molecular biophysical properties that give rise to these experimental outcomes can guide experimental studies. These predictive guidelines can either be used to classify molecules into buckets (good vs. bad) or to predict the exact experimental conditions required to execute each unit operation. Other process attributes (e.g., isotherms, HCP:protein interactions, etc.) are a little more challenging to predict, and these include residual clearance of HCP. Since different cell lines and expression conditions are required for each upstream production, the quantity and types of HCPs can vary. Predictive clearance of the different HCPs will require a complete classification of all the types of HCPs present and their levels under each expression conditions.

From lead selection to final formulation, computational biophysical tools can be deployed to optimize the selection process, optimize lead candidates, and accelerate the development process. Significant time (years) and associated costs are incurred for moving a candidate along to a given stage gate. Should the molecule fail to proceed past that stage, the investment in time and resources is lost. The development of tools that will facilitate the progression of a molecule past each stage gate will save the stated financial cost and will represent the true value of that computational tool. However, the actual cost of developing these computational tools should the data exist will be significantly less than the projected value. The current limitation in producing such a tool is the associated cost of generating relevant data that can be used to develop a predictive platform.

2.2 | Mechanistic modeling—chromatography and other general mechanistic models

Mechanistic models are typically utilized industrially at the unit operation level and across all stages of development including as an early stage activity (Figure 1). Mechanistic models exist for almost all unit operations; however, the current development and implementation level of mechanistic models varies strongly with the specific unit operation. To mention a few, models based on fundamentals for membrane processes in general (Liderfelt & Royce, 2018; van Reis & Zydney, 2007), some chromatographic applications (Benner, Welsh, Rauscher, & Pollard, 2019; Hunt, Larsen, & Todd, 2017), and chemical/enzymatic reactions (Sejergaard et al., 2013) are well established and broadly implemented in industry, while fermentation processes are very (too?) complex to describe solely by mechanistic approaches. Mechanistic modeling focus of the Third Modeling Workshop was on various aspects of chromatographic unit operations with contributions/discussions also covering reactions/synthesis and fermentation.

At the Workshop, applications of mechanistic modeling of chromatography was covered by three presentations (Figure 4). Examples of mechanistic chromatography modeling discussed at the Workshop included incorporating models in typical industry workflows using high throughput scale-down techniques and using models

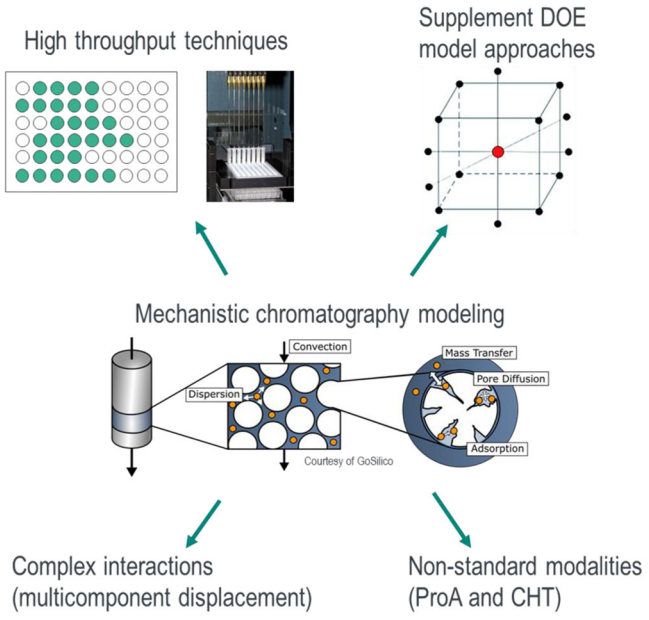

FIGURE 4 Mapping mechanistic modeling to a range of applications (adapted from GoSilico) [Color figure can be viewed at wileyonlinelibrary.com]

to supplement traditional DOE approaches. Other more novel applications included exploring Protein A (ProA) and ceramic hydroxyapatite (CHT) modalities as well as modeling multicomponent displacement effects. One presentation discussed high throughput methods to obtain necessary model parameters through plate based screening or shallow bed chromatography with a focus on ProA affinity chromatography. Another presentation discussed an approach to supplement a typical DOE statistical approach for cation exchange chromatography by applying mechanistic modeling to better understand the significance (or confirm insignificance) of factors not studied in the DOE. A final presentation considered mechanistic models for understanding frontal loading and multicomponent interactions for CHT chromatography and used these models to demonstrate separation of mAb monomer and dimer species.

For a fully mechanistic approach to modeling of chromatographic unit operations, proper mathematical descriptions of fluid flow, mass transport, and where appropriate adsorption isotherms are required. This is generally established for ion-exchange (Briskot et al., 2019; Hahn et al., 2016; Rischawy et al., 2019), size-exclusion chromatography/gel filtration (Hagel, 2011), and affinity chromatography (Benner et al., 2019). However, a molecular level understanding of the chromatographic surface and protein/surface interactions is not yet available. Appropriate adsorption isotherms for reversed-phase (RP), hydrophobic interaction (HIC), and mixed-mode (MM) chromatography are more challenging for proteins, for example, among others due to lack of suitable mathematical description of protein unfolding or conformational changes during operation with hydrophobic surfaces and in the presence of organic solvents. A possible future opportunity for developing mechanistic isotherm models is to combine these models with molecular models (see Figure 2) to obtain a better description of protein/solute behavior and avidity effects, which could extend applicability to new modalities, both new protein formats (e.g., bispecifics, antibody-drug conjugates, antibody fragments, cytokines, etc.) as well as new-therapeutic targets (e.g., cell therapies, viruses, virus-like particles, messenger RNA therapies, and other new modalities). However, it was also suggested at the Workshop to take a hybrid approach and combine machine learning approaches for adsorption with mechanistic understanding/models to obtain better models now for all modes of chromatography, especially RP, HIC and MM.

Scale-up is another example of well-established mechanistic models for chromatography (Benner et al., 2019; Carta & Jungbauer, 2020). This late stage activity is applied either through constant linear or constant volumetric flow rate models. Workshop consensus was that current approaches are well understood for modeling purposes but additional challenges for implementation still exist (e.g., imperfect mixing, spatial inhomogeneity, varying flow-paths, increased/decreased hold-up volumes). This late stage activity is applied either through constant linear or constant volumetric flow rate models. The Workshop consensus was that current approaches work satisfactorily.

Chemical and enzymatic reactions for protein modifications (conjugations of PEG, fatty acids, Fc regions etc., amino acid extensions/substitutions and others) are areas that have proven to be well-described and established by mechanistic models using various rate models (Sejergaard et al., 2013). A future extension of this approach would be to develop general models for example, shelf life and in-use time for protein formulations as a predictive tool to supplement or replace real-time and accelerated stability studies.

Fermentation (possibly including large-scale mixing) and other complex systems are examples of unit operations where current mechanistic models are insufficient to describe and predict the behavior. For these systems, machine learning approaches and/or statistical approaches involving PCA/PLS techniques may currently be optimal, but attempts to provide more mechanistic understanding to these systems are on-going. A Workshop consensus was that the future of modeling of complex systems as well as systems where sufficient mechanistic description is missing would be to aim for hybrid modeling approaches to increase process understanding. Regardless of the modeling approach taken, everybody agreed that comprehensive and good data is required, and the possibility of establishing common databases to the benefit of all and to improve modeling work was discussed.

2.3 | Computational fluid dynamics

CFD is a specific type of mechanistic modeling and branch of fluid mechanics that uses numerical analysis and data structures to analyze and solve problems involving fluid flows. Similar to mechanistic models, CFD is commonly applied at the unit operation level as a digital twin for process equipment (Figure 1).

CFD models are now ubiquitous in aerodynamic design. The last two decades has seen an increased interest in the application of CFD models to other industries including biotechnology processing. This is partly in response to the development of sophisticated commercial software packages that have lowered the barrier of entry for this type of analysis, and most major pharmaceutical companies have now established internal CFD capability and are broadly applying the technology.

The session successfully motivated the value proposition for CFD. The contributors provided numerous examples of how the industry is currently leveraging CFD and the close agreement with experimental results that can be achieved. Agitated vessels continue to be the primary application of CFD to biotechnology processing with bioreactors and fermenters being an extension of these models that include multiphase flow and potentially chemical reactions (Dhanasekharan, Sanyal, Jain, & Haidari, 2005; Haringa et al., 2018; LaRoche, 2005; Wutz et al., 2016). The success of CFD in predicting mixing times, power input, and mass transfer coefficients (k_La) has also led to its use for other unit operations including: centrifugation, chromatography, ultrafiltration, membrane systems, microfiltration, spray drying, and freeze drying.

Despite the ability to provide high resolution results, CFD is still generally used to provide directional guidance for early stage activities or as supporting evidence. A lack of published industry-accepted

approaches for model development and validation often prevents CFD from being used as primary data for later stage activities (e.g., mesh density, domain decomposition, turbulence models, population balance, multiphase flow, non-Newtonian liquids, coupling with metabolic cell models, etc.). This is not a problem specific to CFD but impedes development of strategies for improving current work processes through modeling.

The FDA is actively investing in CFD and its Center for Devices and Radiological Health (CDRH) has a Fluid Dynamics Laboratory focused on problems involving fluid flow and the fluid interactions with medical devices and the human body. There is also an FDA-wide working group on Modeling and Simulation, sponsored by the Office of the Chief Scientist, which launched in 2017. There was general agreement at the Workshop that we need a strategy to engage the agency in our efforts and develop common standards for building and utilizing CFD models. A small step in this direction proposed by the group was to have industry practitioners from the Workshop group develop and document a standard approach CFD mixing and model validation.

2.4 | Plant simulation

Plant simulation was set as an overarching goal for modeling at the Workshop. A number of considerations are required for implementation of plant simulation with respect to timing, applications, underlying tools, and restrictions. In transitioning from early stage design to manufacturing (see Figure 1), an inherent multiscale approach is applied that requires the application of multiscale modeling. Multiscale modeling is defined as the symbiosis between different modeling complexities and disciplines to represent elements of a system and thereby, the full system using in silico methods and tools. The problems to be solved, whether in early stage design or manufacturing are the same, for example, process design, process understanding, process control, process monitoring, process optimization, and process intensification to name a few. However, the constraints associated with these problems differ based on the stage of the design, for example, where CQAs are specified and CPPs operating windows (lower and upper bounds) are estimated, justified and fixed for manufacturing, CQAs are set as process specifications and CPPs are improved/optimized within their predefined operating window. In the pharmaceutical industry (new) products are ultimately anchored in a facility. A facility is defined as a system where multiple processes are operating in parallel to support one or more core manufacturing processes making a value-added product. In contrast to a factory (or plant), a single core process exists to make a value-added product. Pharmaceutical processes are typically batch and therefore, inherently dynamic. A challenge that can be transformed into an opportunity is how to probe the design space for new as well as existing facilities to quantify the impact of design/retrofit decisions. To navigate the complexity for modeling of a facility to feasibly probe, analyze, evaluate and generate the best investment portfolio based on the multiscale, the facility can be decomposed into different scales: the facility (Scale 1), the individual processes (Scale 2) and the unit operations in each process (Scale 3).

Figure 5 shows the symbiosis of the different models utilized for facility (and supply chain) simulation. At the higher end of the scale, Scale 1, production planning-scheduling (PPS) is used to feasibly probe the design/operating space and estimate expected outcomes (reactive approach) and test ideas to evaluate beneficial outcomes/synergies (proactive approach). However, the extent of the expectations that can be obtained are at the lower end of the scale,

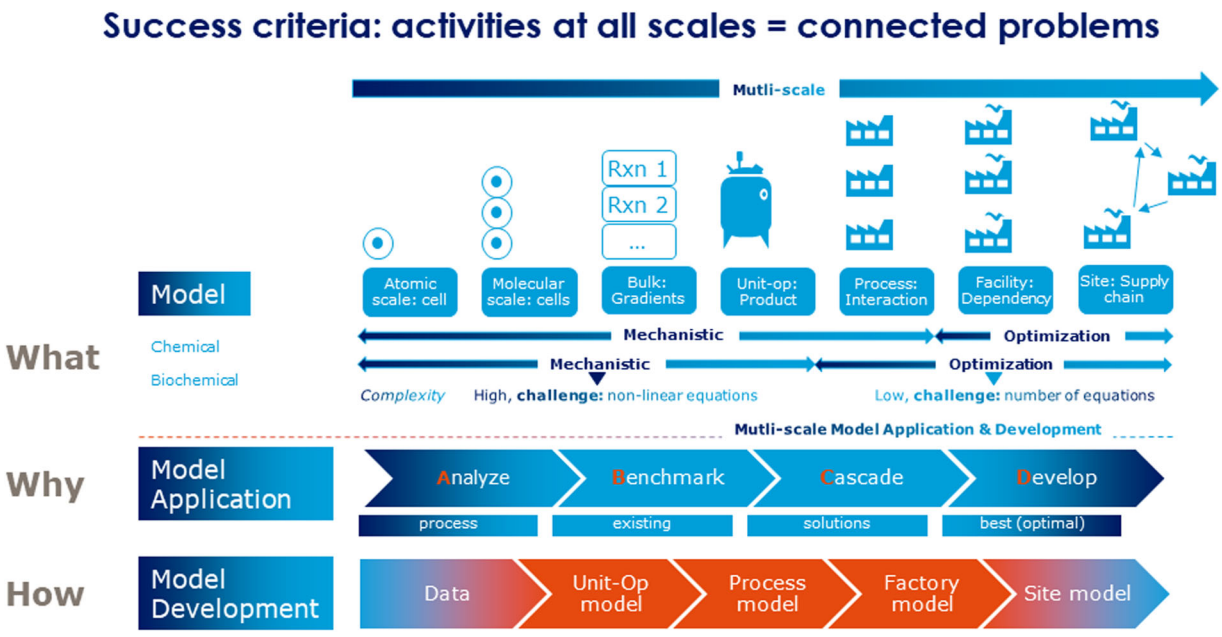

FIGURE 5 Multiscale modeling overview and applicability [Color figure can be viewed at wileyonlinelibrary.com]

Scales 2–3, where a combination of experiments and mechanistic/hybrid models are used. For PPS, the models are of low fidelity (mass balance models), however, solution complexity of the flowsheet model requires the solution of a high number of equations subject to constraints (superstructure optimization). Such problems are formulated as a mixed integer nonlinear programming (MI[N]LP) problem because of the decision making (e.g., equipment selection) uses integer variables and the (non-) linear unit operations models (Bertran et al., 2017). This problem can either be solved using optimization or discrete event algorithms or a combination of both (Harjunkoski et al., 2014).

Using PPS the following analyses can be performed either a priori (early stage design) or posteriori (manufacturing) or a combination of both to evaluate technology transfer and synergies/symbiosis across the value chain. PPS allows the solution of both the planning problem, that is, when to produce what, how much to produce, and the sequence of what to produce and the scheduling problem, that is, how to produce what in a favorable time subject to manufacturing constraints. Examples of analyses are faster product-process designs across a given governance structure, minimization of variables/full manufacturing costs, tact (cycle) time, raw material consumption, and supply chain design/optimization. It should be noted that these minimizations can be defined as reactive because they are classical optimization problems and normally existing experience and data are available. However, the identification of process limitations and bottlenecks, and identification of how to mitigate/remove them while minimizing reactive improvements are also of importance. These minimizations/maximizations (formulated as optimization problems) can be defined as proactive improvements. In exploring proactive improvements using in silico methods for screening, ranking, and selecting feasible, implementable ideas are of importance. As an example, for *n* feasible identified improvement ideas, a total of *n*! can be explored for order of implementation, that is, 1 idea = 1 (1!), 2 = 2 (!), 3 = 6 (3!), 5 = 120 (5!) and 9 = 362,880 (9!). How to select which ideas for investment and what should be the screening and ranking criteria? Or consider an additional constraint where funding is available now but not later in a given project life cycle: which additional ideas should be selected out of the total set? Here, risk-based approaches need to also be incorporated into PPS and mechanistic/hybrid models to account for both uncertainties and sensitivities.

The exploration of synergistic effects, whether positive, negative, or neutral, between the different processes within a facility must be explored for quantification of requirements of auxiliary processes (utility system flowsheet design, size/volume/capacity, location etc.), topological selection (centralized/distributed/de-centralized) and supply/demand of raw materials (e.g., bulk material to produce elution solvents that then become raw materials for the core manufacturing API process). The auxiliary processes are of importance because of the approach to their design and influence on the core manufacturing API process. In terms of design, consideration of simultaneous or hierarchal process design should be performed. Specifically, should facility design-simulation be performed to simultaneously capture synergistic effects on, for example, capacity, to design the auxiliary process or should the core process be designed first, and then auxiliary processes designed around it. A recommendation would be a hybrid approach where hierarchal process design is performed first followed by simultaneous design for refinement. New concepts for flexible process design can be explored, for example, modular design (Baldea, Edgar, Stanley, & Kiss, 2017). The PPS approach allows the exploration of design space discretely for modular design as a function of, for example, numbering up (Bieringer et al., 2016). Discrete sizes are considered because unit operations typically have finite sizes that can be converted into autonomous modules that can be utilized as plug and play. Using an optimization-based approach for PPS (e.g., MILP formulation) m number of modules (autonomous fermenters, tanks, etc.) can be estimated for discrete step changes in ensuring that market demand for a given product can always be met.

From an industrial perspective another problem of equal consideration besides design is retrofit, since API manufacturing processes already exists, that is, whether to produce at pilot or an existing facility scale. The question to consider is how a new process can exist with minimal challenges in an existing facility and if so, what process changes are required? Again, a hybrid approach is recommended using the existing facility simulation model, and the retrofit problem for a new product can be studied and evaluated keeping the auxiliary processes fixed to estimate current change requirements. These requirements that are then tested as ideas for anchoring the project portfolio (both short and long-term) for transforming the current, existing process into the future process via retrofit. Here Scales 2–3 and therefore, mechanistic modeling is of importance to confirm what is expected to be realized.

2.5 | Open Challenges session

Additional aspects, challenges, and opportunities of the four major topics and others were addressed at the Open Challenge session. Many opportunities in modeling exist at the interface of the traditional modeling areas or are a hybrid of various modeling techniques. The Open Challenges session (another new session at the Workshop) was specifically designed to identify areas where tools need to be developed to address complex modeling challenges that transcend multiple areas. The majority of the session focused on flexible approaches and the development of hybrid models, including Artificial Neural Networks (ANN) and the extension of models to optimization of molecular design via development of rules to avoid key primary liabilities.

One specific application was a combination of process modeling and equipment characterization for a lyophilization process. Although one could consider lyophilization as a relatively mature unit operation, it was noted that there is often an extended lag in time from publication of key results and development of a process model. One potential origin of the lag in time is lack of standardization of equipment and experimental approaches in the research field.

Additional opportunities were identified to combine the existing model with other modeling efforts (e.g., CFD for material/heat transfer). Although the model required some lab scale verification of heat transfer coefficients, it was quite useful since it contained a web-based interface to allow for utilization by a researcher who was not a subject matter expert (SME) in the specific field of lyophilization. This case study (as well as other papers in the session on first principles chromatography modeling detailed below) are also excellent examples of some of the challenges associated with deploying models to a wider user community, where there is the potential to further realize the return on investment for the model development. Specifically, the end-user should be aware of the limitations on the interpretation of the simulation results. For example, if a hybrid model is developed with a specific range of parameters (e.g., cake height for the lyophilization system), then the inputs and results may need to be constrained within the model algorithm to specific ranges for both the inputs and outputs of the model. This approach could reduce the risk for misinterpretation or misapplication of the model.

Examples were provided for the current state of chromatography modeling where the technology and algorithms exist (e.g., Chromatographic Analysis and Design Toolkit [CADET]; Leweke & von Lieres, 2018; GoSilico [Karlsruhe, Germany]) and offer a range and combination of models (e.g., multistate steric mass action [SMA], various transport models, multiple unit-operations, including columns, tanks, and other parts such as tubes, valves, detectors, etc.) and allow the user to select the most appropriate tool for the specific application. These tools also exercise the option to model a range of outcomes for chromatography based on Bayesian or Markov Chain Monte Carlo (MCMC) approaches. One of the challenges in deploying these tools is that a SME is required to gather the inputs, interpret the results, and ensure the appropriate selection of the specific tools, see Figure 2.

Opportunities certainly exist for the development of hybrid models, which can exist in multiple forms including a combination of empirical and mechanistic approaches. The combination of mechanistic and statistical models is currently challenging since this approach requires large amount of data for parameters (potentially 18 for SMA and pH in an empirical model) which are not readily available to the model developers. Once developed, these approaches could be used to predict key parameters for chromatography breakthrough similar to CFD power number estimation and be applied with limited experimental data from breakthrough and isotherms to predict elution. Other forms of hybrid models include combinations of various forms of mechanistic models, mixture of atomistic and coarse grain molecular biophysics, combination of CFD flow patterns, and SMA. It is important to note that empirical models can also exist in multiple forms (e.g., lumped kinetic model, isotherm model, etc.) Hence, it is critical when referencing hybrid models to be specific in the approach and goal for the model to ensure lack of ambiguity. One key outcome of the Workshop was that a gap exists in terminology for modeling including the definition of scale and types of hybrid models—a gap that needs to be addressed by the community.

Another key opportunity for modeling is the potential application of machine learning. However, this is currently limited by the size of high-quality data sets and the ability to directly compare the various modeling approaches. To overcome these challenges, one key recommendation at the Workshop is the definition of Good Modeling Practice (such as utilization of consistent approaches for Uncertainty Quantification) via definition of what is an acceptable variance, characterization of output from Bayesian assessments, checks for correlated variables, degrees of freedom, under/over-specified systems, and so forth. In addition to consistency of metrics in Good Modeling Practice, incorporation of a standard set of conditions or systems (e.g., NIST mAb) in the various modeling approaches would be extremely beneficial to advance the modeling field since it can provide objective measures of algorithm efficiency and assess gaps amongst the various modeling approaches. One other area of discussion was the significant value of access to high quality/diverse data sets. One approach could be to develop a structure to amalgamate existing data sets into structured data to support modeling (e.g., research project through a consortium) including black box models. Some examples could include biophysical properties of proteins and protein/protein complexes, protein/ligand complexes, chromatographic retention data for a wide range of molecular types and ligands and simulations data acquired for key model systems over a range of parameterization). This approach could lead to a significant advancement in molecular-scale perspective, provide a reference point and a basis benchmarking for future modeling activities.

To implement this structure an approach to address Intellectual Property (IP) concerns and data integrity would need to be developed. Input and acceptance by regulatory bodies (within companies and at health authorities) in the context of International Conference on Harmonization Guidelines, either existing or in development (e.g., ICH Guidelines Q8–11) (ICH Q10: Pharmaceutical quality system, 2008; ICH Q11: Development & manufacture of drug substances (chemical entities & biotechnological/biological entities), 2012; ICH Q8(R2): Pharmaceutical development, 2009; ICH Q9: Quality risk management, 2005) is essential.

3 | FUTURE STATE

As a result of presentations and discussions, the group defined scenarios and approaches for the desired future state of modeling.

3.1 | Interconnectivity

In an ideal future state of modeling in Chemistry, Manufacturing, and Controls (CMC), the various areas of modeling that have been illustrated in the above sections of this report will be interconnected. This is most likely achieved by an approach that is called multiscale modeling (MSM). MSM describes an approach where physics-based models describe a process at different scales from molecular models to meso- and macro-scale models with decreasing levels of

complexity, but on increasing length and time scales. At the macroscale, many of the more detailed models should be combined as simplified versions. Figure 6 aims to describe the concept of multiscale modeling. Models can be categorized along a scale of decreasing level of detail from atomistic, molecular, micro–meso–macro scales, CFD, and plant simulation/scheduling. However, depending on the task at hand whether it is process characterization or batch scheduling, different levels of detail may be required. The concept of multiscale modeling stipulates that parameters defined at high level of detail will be passed on to the next lower level of detail model, so that a high-detail level model feeds into a lower level of detail. This approach will ensure that models are connected and prevent models be established from scratch again, just because a single parameter needed to be adjusted. Furthermore, changes to the model at a higher-detail level could easily be transferred to the model at the larger scale. Obviously, this may as well work in the opposite direction.

As an example, the development of a chromatography process may require detailed understanding of how a particular antibody interacts under different conditions with a particular chromatography resin. For this purpose, the establishment of a molecular model may be appropriate, but to characterize various other parameters, like flow rate and length of wash, some detailed molecular parameters may be lumped together into a charge parameter that is utilized in an SMA-type of model. Thus, only the lumped parameter will be passed on, but ensure that the connection between the models is kept. Another example is the utilization of quantitative structure–activity relationship models to select operating conditions (e.g., resin, pH) to evaluate in high-throughput screening for the entire process (orthogonal selectivity). These models can be developed using computational biophysics and applied to facilitate mechanistic modeling and effectively span both modeling areas synergistically (see also Opportunities in Figure 7). For the plant simulation model, the only parameters that will be passed on from the SMA-type of model are for each time dependent activity, the duration and the material flow that are used. Lastly, the top right corner of Figure 6 displays that while the level of detail mostly transcends unit operations, the different unit operation models will be connected together to create a unified model for a whole process, and in a facility model different process models are operated to schedule production.

It is crucial for this approach, though, that information gained and parameters calculated are and can be passed on to the next model. As an example, one may consider a molecular assessment probing a variety of things such as charge distribution, hydrophobicity, or patches thereof, chemical stability of charged residues. The parameters that describe this model should then be passed on not only to inform the developability of the molecule but also to be used as initial starting points for informing models on the next scale. For example, parameters for charge may be lumped together and utilized as charge parameters in mechanistic chromatography models. Likewise, a detailed CFD model of the flow-through in a chromatography column will be reduced to a handful of one-dimensional flow distributions that can in turn be used in mechanistic chromatography models.

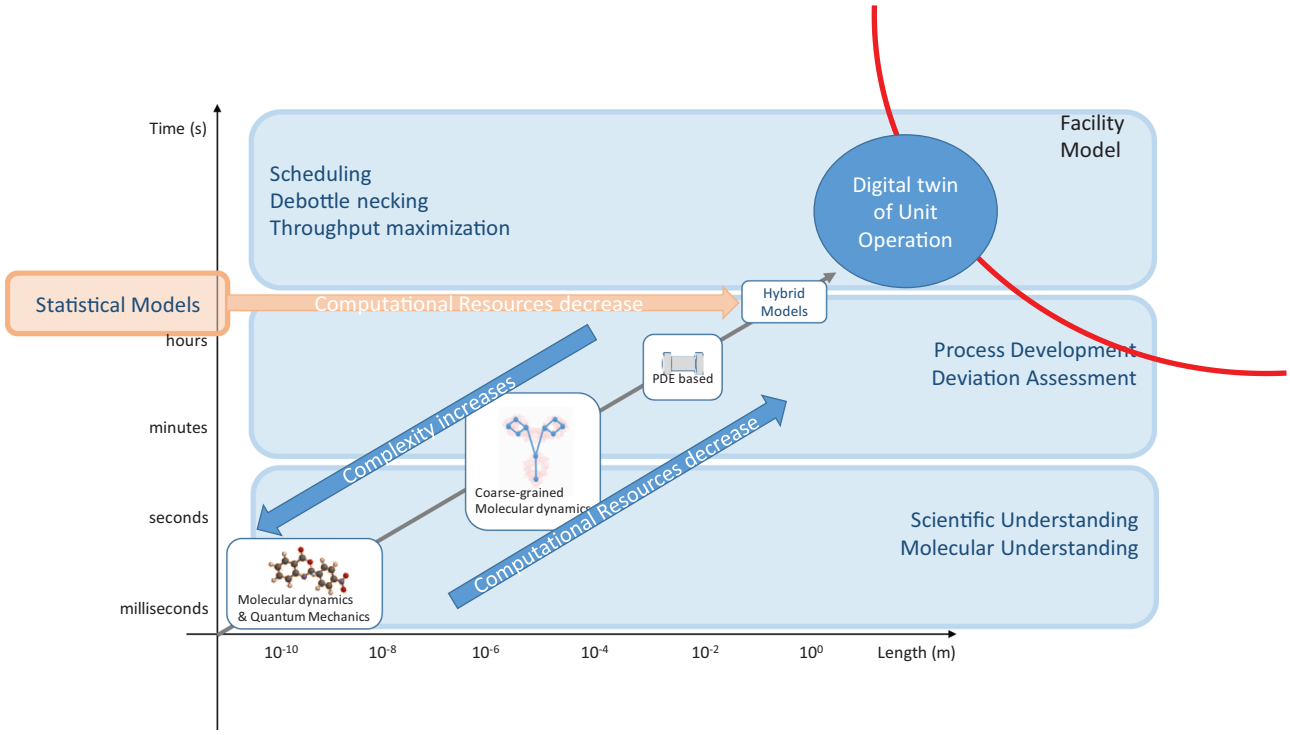

FIGURE 6 Overview of interconnected multiscale models (adapted from Cavallotti & Salvalaglio, 2012) [Color figure can be viewed at wileyonlinelibrary.com]

	Biophysics	Mechanistic	CFD/Mixing	Plant Model/Facility
Early Stage Process design qualitative models	**Upstream** • Limited ability to predict epitope binding/molecular design via in silico screen (ability for relative rank ordering) **Downstream** • Models exist for candidate selection/ liability ID and stability (aka in silico developability) • Potential to screen modalities for purification **Gap** • Standard system to predict stability, expression, and purification modality (e.g., impact of subtle changes) • Some limitations on prediction of specific PTM • Large curated high-quality data sets (potentially via consortium) • Opportunity to use QSAR models to select operating conditions to evaluate in HTS	**Upstream** • Mathematical or hybrid models primarily employed (e.g., scale-up/scale-down) and systems biology **Chromatography/Conjugation** • Current State: Simplified/basic models and commercial tools exist to support PD via fully empirical isotherms • Surrogates for impurities (lumped) • Understanding limited to specific modalities (IEX) as opposed to mixed mode (RP and Protein A) **Gap** • Limited isotherms, mechanistic understanding would be beneficial, primarily applied to but not limited to Mab (applicable to other modalities) • Opportunity to use QSAR models to select operating conditions (e.g., resin, pH) to evaluate in HTS for entire process (orthogonal selectivity)	• Ability to predict power numbers computationally (more accurate than experiment) • Facility fit: Mapping required mixing conditions and vessels to process requirements (including engineering factors) • Quantification of mixing: Which model to apply? • Confirmation of modeling results (feasible for pH) via multiple point sampling • Multiphase modeling exists **Gap** • Bubble coalescence	• Models are dependent on the complexity of the mechanistic models of unit operations since this understanding is translated into states and tasks per unit time **Gap** • If mechanistic understanding is limited, the worst case model can only solve a purely planning-scheduling problem. However, this does not limit evaluation because the objective here is to minimize investment risk (= probability x consequence). Here, probability is high but consequence is low
Late Stage Characterization validation and quantitative models	**Upstream** • Limited ability to predict epitope binding/molecular design via in silico screen (ability for relative rank ordering) **Downstream** • Utilize biophysics to address PC-related question (e.g., parametric understanding) • Potential to utilize models to support process parameter classification • Potential to predict aggregation/ deamidation (limited examples) **Gap** • Missing link from biophysics parameters to process and stability (in lieu of experiments) • Biophysics to supplement mechanistic model • Mapping/characterization of process- and product-related impurities on chromatographic performance	**Upstream** • TBD **Chromatography** • Current State: Simplified/basic models exist to support PD • Surrogates for impurities (lumped) • Understanding limited to specific modalities (IEX) as opposed to mixed mode • Initial models for scale-down/HTS **Gap** • Mechanistic understanding of all modes of chromatography • Limited isotherms • Scale-down for HTS systems • Standardization (including reference data sets) for modeling tools does not exist	**Upstream** • Initial models established (including multiphase) e.g., ambr scale-down • Predictive models do not necessarily exist (confirmation of experimental results feasible) • Modeling not yet fully quantitative (e.g., aggregate formation, discrete particles, bubble size distribution/ coalescence) • Directional/semi-quantitative effects of shear are feasible (experimental confirmation is challenging) **Chromatography** • Heterogeneity of packing, flow distribution • Experimental confirmation of distribution determined	**Upstream** • Same as Early Stage **Gap** • Additionally, having material flow mapped, sustainability among selected alternatives using process mass intensity (PMI) can be evaluated for selection of the best process
Commercial Scale-up, transfer, plant simulation	**Upstream** • Same as Late Stage **Mixed Mode Isotherms** **QSPR-Tools**	**Upstream** • Limited understanding of process variability, scale-up/scale-down **Chromatography** • Current State: Simplified/basic models exist to support PD surrogates for impurities (lumped) • Understanding limited to specific modalities (IEX) as opposed to mixed mode • Initial models for scale-down/HTS **Gap** • Mechanistic understanding of all modes of chromatography • Limited isotherms • Scale-down for HTS systems • Standardization (including reference data sets) for modeling tools does not exist	• Quantitative assessment of facility fit to potentially support PPQ • Utilized to support deviation management	• Same as Early Stage • Additionally, this is the best approach for understanding the process to be built. Model frameworks and solution strategies independently exist for solving the plant simulation problem **Gap** • As with the previous two stages, complexity of the models used here are dependent on the mechanistic models, and as they improve, the plant simulation improves. Ideally, the plant model also helps with commissioning; planning IQ, OQ, PPQ, etc.; and characterization of short (mid-/long) term changes

FIGURE 7 Current state of modeling areas and gaps/investments required to achieve future state (Figure 1). Green—models are sufficiently developed to support implementation now. Yellow—gaps identified that require modest investments to address before implementation (2–5 years). In some specific cases, computational limitations may currently exist which impact deployment. Gray— opportunities for exploration or significant gaps required to achieve realization, for example, mechanistic modeling of fermentation/cell culture (5–10 years) or may not be scientifically feasible (initial assessment required). Early stage indicates models appropriate for supporting first-in-human studies or developmental studies. Late stage indicates quantitative models potentially applicable for biologics licensing applications [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | Usability

Apart from the ability to pass on physical parameters, in the future state there will be platforms available that facilitate this interconnection. A future computer tool will integrate different modeling approaches and facilitate the integration of the various modeling techniques described. Currently many tools are custom made, require in depth expert modeling knowledge, and are focused on only one aspect of modeling. As mentioned in the opening section, although these models were developed based on proteins, the applications and approaches can be applied to new biological formats (e.g., conjugates, virus-like particles or virus production for gene therapy) since the same principles and parametric approach is applicable. In the future state the computational tools will be simple enough to use that a competent process engineer is capable of using and applying those tools without expert knowledge in programming or mathematics, but with proper understanding of model limitations, and so forth. To achieve this future state, additional investments in curricula at universities and industry may be required to ensure appropriate knowledge of tool development and implementation including contextualization of results.

3.3 | Standardization

Lastly, the modeling approaches will be standardized. Currently, there are multiple approaches at the mechanistic or meso-scale level, and they are using different but similar parameters to describe similar phenomena. This makes it difficult to move between modeling approaches as each of them may require a different set of calibration experiments. A similar problem occurs with raw materials. Ionic capacities, binding capacities, and/or porosities as examples are specified on certificates of analysis (CoAs) in many ways by chromatographic resin vendors and determined by as many methods, resulting in the need to prepare cumbersome calibration experiments before a model may be formulated. In the future state, the vendors of raw materials will supply the parameters required for their components and similarly the suppliers of equipment will supply preformulated models that can be used by the process engineer, including a digital twin (e.g., CFD model development). Not having to spend their efforts on characterizing the equipment and raw materials, they can focus on developing the process. To achieve this state, standardization of requirements, reference conditions (for evaluation of performance) and versioning will need to be deployed. This approach is essential to address the ability to utilize this information for regulatory filings and to advance the modeling fields.

To achieve the Future State of In Silico Development, it is important to reflect on the current state and existing gaps. This is done in Figure 7 with color coding of the different stages of development of the various modeling tools and timing from Figure 1. Prospective investments to address the opportunities identified (gray shading) are required to transition from the current to the envisioned future state.

ACKNOWLEDGMENT

The authors would like to acknowledge Pablo Rolandi and Ben Smith (Amgen, Inc.) for their input and detailed reviews of the manuscript.

ORCID

Francis Insaidoo http://orcid.org/0000-0002-0800-4634
Abraham Lenhoff http://orcid.org/0000-0002-7831-219X
John Welsh http://orcid.org/0000-0002-2422-037X
Richard Willson http://orcid.org/0000-0002-0900-6148

REFERENCES

Baldea, M., Edgar, T. F., Stanley, B. L., & Kiss, A. A. (2017). Modular manufacturing processes: Status, challenges, and opportunities. *AIChE Journal*, *63*(10), 4262–4272.

Benner, S. W., Welsh, J. P., Rauscher, M. A., & Pollard, J. M. (2019). Prediction of lab and manufacturing scale chromatography performance using mini-columns and mechanistic modeling. *Journal of Chromatography A*, *1593*, 54–62. https://doi.org/10.1016/j.chroma.2019.01.063

Bertran, M.-O., Frauzem, R., Sanchez-Arcilla, A.-S., Zhang, L., Woodley, J. M., & Gani, R. (2017). A generic methodology for processing route synthesis and design based on superstructure optimization. *Computers & Chemical Engineering*, *106*, 892–910.

Bieringer, T., Bramslepe, C., Brand, S., Brodhagen, A., Dreiser, C., Fleischer-Trebes, C., & Kockmann, N. (2016). *Modular Plants: Flexible Chemical Production by Modularization and Standardization-Status Quo and Future Trends*. DECHEMA Gesellschaft für Chemische Technik und Biotechnologie eV.

Briskot, T., Stückler, F., Wittkopp, F., Williams, C., Yang, J., Konrad, S., … Hepbildikler, S. (2019). Prediction uncertainty assessment of chromatography models using Bayesian inference. *Journal of Chromatography A*, *1587*, 101–110.

Carta, G., & Jungbauer, A. (2020). *Protein chromatography: Process development and scale-up*. Newark, NJ: John Wiley & Sons.

Cavallotti, C., & Salvalaglio, M. (2012). Molecular modelling of the affinity chromatography of proteins. *Chimica oggi/Chemistry Today*, *30*(5).

Coffman, J., Marques, B., Orozco, R., Aswath, M., Mohammad, H., Zimmermann, E., … Willson, R. C. (2020). Highland games: A benchmarking exercise in predicting biophysical and drug properties of monoclonal antibodies from amino acid sequences. *Biotechnology and Bioengineering*, *117*(7), 2100–2115. https://doi.org/10.1002/bit.27349

Dhanasekharan, K. M., Sanyal, J., Jain, A., & Haidari, A. (2005). A generalized approach to model oxygen transfer in bioreactors using population balances and computational fluid dynamics. *Chemical Engineering Science*, *60*(1), 213–218.

Hagel, L. (2011). Gel filtration: Size exclusion chromatography. In J.-C. Janson (Ed.), *Protein purification: Principles, high resolution methods, and applications* (pp. 51–91). Hoboken, NJ: John Wiley & Sons.

Hahn, T., Huuk, T., Osberghaus, A., Doninger, K., Nath, S., Hepbildikler, S., … Hubbuch, J. (2016). Calibration-free inverse modeling of ion-exchange chromatography in industrial antibody purification. *Engineering in Life Sciences*, *16*(2), 107–113.

Hansen, E. B. (2017). Chromatographic scale-up on a volume basis. In A. Staby, A. S. Rathore & S. Ahuja (Eds.), *Preparative chromatography for separation of proteins* (p. 227). Hoboken, NJ: John Wiley & Sons.

Haringa, C., Tang, W., Wang, G., Deshmukh, A. T., van Winden, W. A., Chu, J., … Noorman, H. J. (2018). Computational fluid dynamics simulation of an industrial *P. chrysogenum* fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization. *Chemical Engineering Science*, *175*, 12–24.

Harjunkoski, I., Maravelias, C. T., Bongers, P., Castro, P. M., Engell, S., Grossmann, I. E., … Wassick, J. (2014). Scope for industrial

applications of production scheduling models and solution methods. *Computers & Chemical Engineering*, *62*, 161–193.

Hunt, S., Larsen, T., & Todd, R. J. (2017). Modeling preparative cation exchange chromatography of monoclonal antibodies. In A. Staby, A. S. Rathore & S. Ahuja (Eds.), *Preparative chromatography for separation of proteins* (p. 399). Hoboken, NJ: John Wiley & Sons.

ICH Q8(R2): Pharmaceutical development. (2009).

ICH Q9: Quality risk management. (2005).

ICH Q10: Pharmaceutical quality system. (2008).

ICH Q11: Development and manufacture of drug substances (chemical entities and biotechnological/biological entities). (2012).

Ladd Effio, C., Hahn, T., Seiler, J., Oelmeier, S. A., Asen, I., Silberer, C., … Hubbuch, J. (2016). Modeling and simulation of anion-exchange membrane chromatography for purification of Sf9 insect cell-derived virus-like particles. *Journal of Chromatography A*, *1429*, 142–154. https://doi.org/10.1016/j.chroma.2015.12.006

LaRoche, R. D. (2005). *Biopharmaceutical scale-up and scale-down using fluid mixing analysis and computational fluid dynamics (CFD)*. Paper presented at the 2005 AIChE Annual Meeting, Cincinatti, OH.

Leweke, S., & von Lieres, E. (2018). Chromatography analysis and design toolkit (CADET). *Computers & Chemical Engineering*, *113*, 274–294.

Liderfelt, J., & Royce, J. (2018). Filtration principles. In G. Jagschies, E. Lindskog, K. Lacki & P. Galliher (Eds.), *Biopharmaceutical processing: Development, design, and implementation of manufacturing processes* (pp. 279–293). Cambridge, MA: Elsevier.

Rischawy, F., Saleh, D., Hahn, T., Oelmeier, S., Spitz, J., & Kluters, S. (2019). Good modeling practice for industrial chromatography: Mechanistic modeling of ion exchange chromatography of a bispecific antibody. *Computers & Chemical Engineering*, *130*, 106532.

Rolandi, P. A. (2019). *The unreasonable effectiveness of equations: Advanced modeling for biopharmaceutical process development*. Proceedings of computer aided chemical engineering (Vol. 47, pp. 137–150). Elsevier.

Sejergaard, L., Hansen, M. H., Olsen, A. A., Ostergaard, P., Strom-Hansen, T., Ostfeldt, P., … Staby, A. (2013). *Implementation of unit operation design spaces: Experiences with regulatory authorities*. Paper presented at the 245th ACS National Meeting, New Orleans, LA.

Staby, A., Ahuja, S., & Rathore, A. S. (2017). Model-based preparative chromatography process development in the QbD paradigm. In A. Staby, A. S. Rathore & S. Ahuja (Eds.), *Preparative chromatography for separation of proteins* (p. 1). Hoboken, NJ: John Wiley & Sons.

van Reis, R., & Zydney, A. (2007). Bioprocess membrane technology. *Journal of Membrane Science*, *297*(1–2), 16–50.

Wutz, J., Lapin, A., Siebler, F., Schäfer, J. E., Wucherpfennig, T., Berger, M., & Takors, R. (2016). Predictability of kLa in stirred tank reactors under multiple operating conditions using an Euler–Lagrange approach. *Engineering in Life Sciences*, *16*(7), 633–642.

How to cite this article: Roush D, Asthagiri D, Babi DK, et al. Toward in silico CMC: An industrial collaborative approach to model-based process development. *Biotechnology and Bioengineering*. 2020;117:3986–4000. https://doi.org/10.1002/bit.27520

APPENDIX: AGENDA FOR THIRD MODELING WORKSHOP

Third Modeling Workshop Agenda, Day 1 (May 07, 2019)

Topic	Presenter
Mechanistic modeling	Arne Staby/John Welsh
MAGPIE—Algorithmic analyzation of bioprocess data	Tobias Grosskopf (Roche)
Computer-aided nondistributed modeling of industrial fermentation process for process understanding, optimization, monitoring, and control	Deenesh Babi (Novo Nordisk)
Model-based development of a conjugation step	Ernst Broberg Hansen (Novo Nordisk)
Digitalizing biopharma: Application of good modeling practice for industrial chromatography	Gang Wang (Boehringer Ingelheim)
Understanding the impact of scale and resin characteristics on Protein A chromatography elution profiles using mechanistic modeling	Steve Benner (Merck & Co., Inc., Kenilworth, NJ)
Developing a mechanistic chromatography model to perform a simulated design of experiments (DOE) study	Mark Fedesco (Genentech)
Modeling displacement effects in purification of monoclonal antibodies by frontal analysis and overloaded gradient elution chromatography	Giorgio Carta (Univ. of Virginia)
Mechanistic modeling roundtable discussion	Arne Staby/John Welsh
Computational fluid dynamics	Stephen Hunt and Bob Todd
Insight from CFD snapshots inform experimental design	Henrik Marke (Novo Nordisk)
CFD challenges and opportunities in pharmaceutical industry	Matt Flamm (Merck & Co., Inc., Kenilworth, NJ)
Making large scale processes transparent—the application of cfd and classical engineering approaches to mitigate risk during cell culture process transfer	Thomas Wucherpfennig (Boehringer Ingelheim)
CFD: Total cost of ownership	Stephen Hunt (KBI Biopharma)
CFD roundtable discussion	Stephen Hunt and Bob Todd

Third Modeling Workshop Agenda, Day 2 (May 8, 2019)

Topic	Presenter
Plant simulation	Ernst Broberg Hansen
Run-rate and capacity with VirtECS plant simulation software	Ben Smith (Amgen)
Transforming biopharmaceutical manufacturing with plant simulation and scheduling	Larry Sun (Amgen)
Multiscale in silico driven product and process development	Philipp Ernst (Bayer AG)
Computer-aided flowsheet simulation of industrial pharmaceutical processes for identification of improvement-optimization	Marcel Stenvang (Novo Nordisk)
Advanced models and the modeling lifecycle in biopharma process	Pablo Rolandi (Amgen)
Plant simulation roundtable	Ernst Broberg Hansen
Open challenge	David Roush and Jan Griesbach
A new approach to lyophilization process development and transfer enabled by equipment characterization and process modeling	Fabrice Schlegel (Amgen)
Current capabilities and future development of the CADET platform for chromatography modeling	Eric von Lieres (Research Centre Jülich)
Modeling of ion exchange chromatography: from mechanistic to empirical and back	Tobias Hahn (KIT)
Leveraging mechanistic modeling with a black box: neural network estimation of chromatographic parameters	Abraham Lenhoff (Univ. of Delaware)
Chemical and physical determinants of drug-like monoclonal antibodies	Pete Tessier (Univ. of Michigan)
Open challenge roundtable	David Roush and Jan Griesbach

Third Modeling Workshop Agenda, Day 3 (May 9, 2019)

Topic	Presenter
Opening remarks	Organizing Committee
Molecular modeling	Francis Insaidoo
Molecular modeling overview of session topics/themes	Francis Insaidoo
A quasichemical perspective of protein solution thermodynamics	Dilip Asthagiri (Rice University)
Pattern formation on resin surfaces in multimodal chromatography: Quantification of ligand aggregation and development of QSAR descriptions	Steve Cramer (RPI)
Mesoscale Model for the self-assembly and cross-linking dynamics of HPV virus-like particles	Oleksandr Zavalov (Merck & Co., Inc., Kenilworth, NJ)
Using data for *in silico* prediction of monoclonal antibody characteristics	Jasper Lin (Genentech)
Roundtable discussion molecular modeling	Francis Insaidoo

Abbreviation: QSAR, quantitative structure–activity relationship.