Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Engineering

A roadmap to AI-driven *in silico* process development: bioprocessing 4.0 in practice

Moritz von Stosch[1], Rui MC Portela and Christos Varsakelis[2]

In silico process development constitutes a viable option for accelerating CMC development timelines and can be achieved through either a hybrid-modeling-driven or an Artificial Intelligence (AI)-driven avenue. Each pathway has its own pros and cons but the biggest difference is that the former can be developed in-house whereas the latter requires inter-corporation collaboration. Motivated by the precedence of inter-corporation data ecosystems targeting drug discovery, in this paper we bring forward the case of the AI-driven approach to *in silico* process development in terms of both technical feasibility and scientific soundness. Our analysis asserts that methodologically, AI is now mature to be employed for the task at hand, provided that the overall exercise is driven by multi-disciplinary experts. Further, *in silico* process development should be understood as a collection of models, with each process unit being endowed with its own model.

Address
GSK Vaccines, Rixensart, Belgium

Corresponding author: Varsakelis, Christos (cvarsake@its.jnj.com)
[1] Current affiliation: DataHow AG, Zurich, Switzerland.
[2] Current affiliation: The Janssen Pharmaceutical Companies of Johnson and Johnson, Beerse, Belgium.

Current Opinion in Chemical Engineering 2021, **33**:100692

This review comes from a themed issue on **Biotechnology and Bioprocess Eng: modeling and related issues**

Edited by **Cleo Kontoravdi** and **Colin Clarke**

For a complete overview see the Issue

Available online 16th June 2021

https://doi.org/10.1016/j.coche.2021.100692

Introduction

The application of Artificial Intelligence (AI)[3] for drug discovery is projected to speed-up and rationalize drug development. Potentially, this will disrupt the current Ways-of-Working (WoW) though both the decrease of candidates screened in Pre-clinical studies and the increase of the total Probability of Technical Success [1,2••]. Given the continuously disappointing rates of return in the pharmaceutical sector as a whole [3], it is not surprising that investments already sky-rocketed with small start-ups raising over 4.5 bn EUR in 2019 alone [4]. Moreover, net-revenues are estimated to rise at a Compound Annual Growth Rate (CAGR) of around *22%*, reaching approximately *2* bn EUR by *2022* [5].

What is more striking, however, is the modus operandi shared by both big and small players in the field. Many of the involved parties consent that the pathway to harvesting the power of AI in drug discovery goes through data ecosystems, compiling data from multiple corporations and open sources [6••]. This happens due to the low likelihood of one company having a sufficiently diversified data portfolio to proceed independently. Additional benefits of type of joint ventures include risk mitigation, owing to their co-investment nature. Also, they offer protection against potential disruption for it is not only the competitive advantage of participating that may be a game changer but also the competitive disadvantage of being locked out. To date, both open innovation and investment-based approaches have been proposed [7,8].

The disruption appetite celebrated in drug discovery does not seem to carry over to Bioprocessing. To be sure, AI is currently exercised in Bioprocessing throughout the R&D – manufacturing continuum, even though most efforts end up in limbo in the so-called Proof of Concept (POC) purgatory [9]. However, despite the hype that surrounds these efforts [10], they all seem to share the same objective and strategy: the acceleration of CMC development times by integrating AI alongside existing WoW through the Quality by Design (QbD) framework [11–13]. This is somewhat reminiscent of the accommodation of mechanistic process modeling, that shares a similar course [14••]. At this point, there exists no unilateral or collaborative plan for AI to disrupt traditional WoW for processes development by mimicking the activities in the drug discovery sector. In other words, when confronted with the augment versus disrupt dilemma, Biopharma stakeholders have opted for the former which, in turn, begs the question: is process development a good investment?

Traditionally, process development has lived in the shadow of drug discovery for two reasons. First, Return on Investment (ROI) in drug discovery cannot compare to those in process development and a portfolio analysis asserts the minimum allocation of resources. Second, process development has seldom been the bottleneck in terms of success rates, with the vast majority of drugs

[3] The term artificial intelligence is used here in a rather generous way, since we do not aim at detailing the technical approaches for developing an *in silico* platform [52].

and biologics failing outside the CMC development circle [15]. Prima facie, therefore, the answer appears to be negative. This narrative, however, is flawed on at least three fronts.

First, for Biologicals, process development can add up to 2.5 years in the total time to market and for vaccines this elevates to 5 and can get up to 7.[4] Now, Net Present Values (NPVs) of drugs remain notoriously confidential, however, estimates from [16] assert that being just a quarter of a year faster increases the NPV by 5%. Therefore, even though the relation between CMC development timelines and overall time to market is sublinear and there is an inherent stochasticity in clinical trials, faster CMC development timelines can affect time to market. Second, traditional timelines are constantly challenged [17–19] and the on-going COVID-19 pandemic has exerted even more pressure towards accelerating efforts [20]. Third, although Biopharma has just started to digest the on-going discussions about Industry 4.0 [21,22], the voluminous cashflows for AI in drug discovery could lead to an expanded number of simultaneous candidates [23,24], and thus render process development departments a bottleneck due to their intrinsic limited parallelization capacity.

If process development then constitutes an attractive investment, is augmentation of current WoW a viable strategy for seeing it done? The answer is negative. Traditional process development is based on experimental campaigns and statistical modeling and cannot accelerate beyond its parallelization capacity. For instance, scale-up studies, even within the context of High Throughput Screening, are bound to be carried serially. Consequently, augmentations based on the infusion of AI or mechanistic models will only result in incremental gains. Substantial improvements require considerable upfront CAPEX investment in resources, either in house or to Contract Developing and Manufacturing Organizations (CDMOs). However, since both insources and CDMOS use traditional WoW, increases in parallelization capacity will approximately result in proportional increases in the cost thereby casting serious doubts on the business cases.

In silico process development, defined as process development through mathematical models with a minimum, confirmatory set of experiments, emerges as interesting candidate to screen. Recently, [25•] simulated *in silico* process development scenarios based on End-2-End Digital Twins. They found that a modest acceleration of two quarters, living otherwise total product cycle-times unchanged, yields net savings from 40 to 140 million dollars. But the real potential for disruption lies not in cost savings but in reduction of development timelines. Recent examples stemming from the industry give a range from 50% to nearly 75% [26,27].[5] Concerning handling capacity, recent advances in hardware and software dictate that, at least for bioprocessing, the combined CAPEX and OPEX costs for *in silico* handling of multiple candidates is inconsequential when compared against the one required for traditional process development [28].

In silico process development can be achieved through two avenues. First, through a mechanistically oriented approach, predicating on hybrid modelling, and second through a purely AI-driven one. Each approach has its own peculiarities and even conceptually requires different strategic planning with the core being the following: the mechanistically oriented avenue can be achieved within the corporation whereas the AI-driven one requires an inter-corporation data ecosystem. With the mechanistically oriented case being recently explored in [29•] we therefore focus exclusively on the AI-driven pathway. The main question that this paper address is thus:

Is AI-driven *in silico* process development technically feasible and scientifically sound? And if yes, what is a practical way forward?

From product centric to knowledge centric process development

For years, the focus of process development is on maximizing yield, robustness and process understanding for the specific product at hand, which dictated the corresponding experiment design strategy. The Quality by Design (QbD) concept has even increased this focus, as it starts from the attributes of the product to infer the design of the process and identify the process parameters of interest [30••]. Hence, process development at this moment is to the largest degree product centric.

Quality by design can only be cost-effective if knowledge is transferred from one product to the next, iteratively decreasing the number of experiments required to generate process understanding [25•]. In absence of a clear understanding of the best way to achieve this, access to data was identified as a prerequisite. Consequently, a rise in data initiatives could and can be observed across

[4] The exact timeline varies between small and large molecules. Jones *et al.* [53] estimates CMC development times at the order of 18 months for monoclonal Antibody Products and [52] provides estimates for vaccines.

[5] These examples point to specific modules and not towards the End-2-End process development chain so they should be interpreted with the usual caveats. Nevertheless, it is possible to carry out a formal analysis for expected gains. For vaccines such an exercise is presented in Stosch and Varsakelis [54].

pharmaceutical and contract manufacturing companies [31–34], the strategy now being data centric.

The exposure of the pharmaceutical companies to the 'data industry' lead pharma to increasingly understand the data business, where data is the asset.[6] They have become aware that data gathering is occurring and that by sharing their data with partners, service providers and so on, their data could though not be gathered (likely contractually prohibited) potentially be used by the other parties to further their tools. Consequently, more refined clauses to protect their data and derived IP have been and are appearing in contracts trying to remediate this risk. Also, the use of online tools and services, for example, for sequence design, translation, and so on. or even google translate, has been seen more critically and was restricted. Logically, the data gathering, and querying capabilities were and are reshaped, contextualizing data such that they can share specific parts of it with partners in the future, would they desire to. Their first (investment) focus is (or should be), of course, on drug discovery and clinical trial data, as they are expensive to generate and might contain insights for the discovery of new drugs. Though starting to appear, investment in the same manner is not seen on the process side, process development being a necessary evil to them. For CDMOs this seems a completely different story, process development typically being important for attracting clients for manufacturing contracts. CDMOs have taken a similar position in data sharing, yet as process development is an essential part of their business, they realized that the kind of variations in drugs and processes they see and the number of processes they deal with permits them to create valuable data assets. Hence, they could try to explore them by themselves or with the help of a technology partner to streamline process development. An even different position is taken by (small) biotech companies, which either outsource process development or pursue it with minimal effort in a proof-of-concept modus operandi. These companies are willing to share their data for the sake of accelerated and facilitated process development as long as their major IP remains intact, a possible gateway for technology partners. Logically, all players might reconsider their position to gather a chair on the data business table, for example, using 'gated open innovation' ecosystems, as elaborated on later in this contribution. Process development such will become a service, that is, the strategy being process knowledge centric.

Knowledge-centric process development is an exponential technology, because (1) it marginalizes the cost of data, thus prospectively the cost of experimentation and process development; (2) it democratizes the access to data, that is, optimal process development capabilities; and (3) it is ubiquitous, living in the cloud.

Methods and technical capabilities – state of the art

There are two prerequisites for the application of AI for *in silico* process development: (1) high volumes of qualitative data; and (2) methods for the analysis and modeling of multi-product spanning process data.

Why data need to be contextualized

The application of AI for *in silico* process development requires process data of a number of different drug variants, beside variations in process design, materials and process parameters. At present, few companies have a data portfolio (in particular easy access to these data) that is (i) diversified enough and (ii) dense enough to unilaterally develop an adequately trained AI-process design tool. Therefore, cross-company data gathering approaches would be sought as well as incentives for companies to promptly participate.

However, each company (and even departments with the same company) likely has a different approach to data contextualization, that is, differences in data recording, analytic methods, and so on.

Harmonization by agreeing on a common standard and ontology to data storage and exchange could be one solution to address this issue. However, analytical methods would have to be harmonized as well. Harmonization requires a huge effort involving long times to develop and comply. For instance, in light of industry 4.0 (analytic) equipment providers have been discussing several standards for years, such as SILA2 [35,36] and others. Despite that big providers of automation, lab or device solutions (such as Siemens, ABB, Applied Materials, etc.) could implicitly drive harmonization. Experiment protocols/recipes exist in the automation world already for decades, their extension to lab experimentation with the increased use of robots and automated high-throughput platforms seems consequential. Companies providing analytic devices could also further implicit harmonization, as analytic methods could potentially be developed faster and more robust using insights from various laboratories. Providers of workflow and data capturing solutions such as Riffyn, IDBS, Lab-forward, and so on certainly drive the realization of a certain data standard or at least decrease data recording entropy.

[6] Big biopharma companies are owning the entire value stream, starting from a market analysis, via target identification, drug design, pre-clinical and clinical evaluation, process development, manufacturing, supply and marketing, with almost all of the activities being carried out by themselves. Each of these activities in the value stream might be dominated in the future by companies that are specialized working cross-industry, that is, transversal, and are hence much faster and more cost-effective than one big biopharmaceutical company. However, biopharma companies have unique knowledge on the overall managing of the value stream, so this could potentially be the fallback option.

Contextualization/Harmonization is a 'motivation' problem, in the sense that it requires significant effort without providing direct benefit to the person putting the effort in. Hence, harmonization is likely to only get us one step forward but it will not solve everything. Intelligently designed solutions that provide direct benefit to the person contextualizing the data in a systematic and harmonized way are required and need to be sought. Bioinformatics NGS sequencers and cloud solutions could provide an inspiration. Here different sequencing technologies generate different file types (Oxford Nanopore generates fast5 while Illumina sequencers generate bcl files). These files have to be pre-processed and converted in a common standard file (fastq). In some cases it can be performed by the device (e.g. Illumina MiSeq can generate fastq files). Such files are submitted to a webpage (amenable to include contextualized data) to predict some properties (∼immediate benefit).

In essence recording of the workflow is a 'perception problem' and likely there always is circumstance in which a parameter that was not considered important and hence not recorded renders impossible the replication (or even reproduction) of the experiment results. However, it should be eminent that granularity when it comes to where, who and how data where generated is of utmost importance.

Multi-product spanning process data analysis and modeling

Pilot studies for *in silico* process development have appeared in the literature focusing on one particular part of process development, namely drug formulation [37•]. Inspired by quantitative structure-activity relationship (QSAR) models, they used molecular descriptors to represent different drug compounds, which dominate the release behavior, yet disregarded differences that might arise from differences in laboratories and analytic methods [38] elaborated how such a molecular descriptor concept could be used for a directed Quality by Design guided monoclonal Antibody process development. However, apart from formulation and capture steps in downstream purification the process behavior can be regarded largely independent of the drug compound, that is, it will not be possible to predict the behavior of the bioprocess based purely on the aminoacid sequence.

Given the heterogeneity of the product, process, analytic methods and data context, modeling and analysis methods are required that can deal with this heterogeneity or at least account for it. Bioinformatics, in particular, -omics data analysis has, though to a lesser degree, faced a conceptually alike problem, normalization having become an inherent and critical component of the -omics data ingestions pipelines [39]. Similarly, in a desire to make Raman monitoring independent of one particular application, methods of wavelength alignment and scaling have been proposed for transversal Raman modeling across probes, scales and processes. Yet scaling exhibits a number of challenges [40–44] and alone will not suffice to address heterogeneity.

Using categorical variables to represent the data context is another possibility that can be used in addition to normalization. For analysis purposes, categorical variables are classically encoded as dummy variables (statistics)/ one-hot encoding (data-science). Other encodings are possible, the translation from categorical feature to numeric representation typically being chosen by the user [45]. The general problem with this type of encoding is that for process development of a new drug compound a new feature (binary representation) would be added as input, requiring the execution of several experiments for the model to 'understand' how the behavior of the new process relates to that of the old processes. Recently, vector embeddings that are identified from data, were proposed as an automatic approach to alleviate this problem [26], yet the execution of a reduced, potentially specific, number of experiments is required. Hence, knowledge transfer approaches, as the vector embedding approach, are of particular interest for cases where the process behavior is not governed by the product molecule, such as in upstream processes. Though not completely *in silico*, knowledge transfer approaches could in the first line help to reduce the amount of required experimentation significantly. Prospectively, taking cell-lines, media and other factors into consideration it might even become possible to predict the behavior of the most suitable cell lines going one step further towards full *in silico* process design. Hence, investigation in this direction is needed.

On the sort to mid run, a combination of sequence-activity and knowledge transfer approaches will likely prove successful in practical applications, choosing the most adequate way for each process unit. A process unit-by-unit guided development of an *in silico* platform would also benefit from the granularity of data and separate parameter spaces, combatting the curse of dimensionality. This also constitutes a substantial advantage because the viability of a divide-and-conquer and agile-based approach allows for the systematic leverage of pre-existing process knowledge in form of hybrid modeling [46,47••,48,49]. This furthers the possibility of reducing experiments yet faithfully representing the 'dominant' behavior of the system. It is interesting to note that with an increase in the quantity of data it will be possible to see and model nuances in process behavior, something that to date is not possible due to the 'low sample size'.

Conclusion

In silico process development comes with an attractive business case and quantifiable benefits in terms of accelerating timelines and parallelization capacity. In this

paper, we have examined the case of the AI-driven approach to *in silico* process development in terms of both technical feasibility and scientific soundness.

The need to shift mindset from to product-centric to knowledge-centric has been identified as a prerequisite. Within the classical QbD paradigm, acceleration and cost effectiveness can only happen if Baysianity and knowledge transfer are embraced.

With respect to technical feasibility, the fundamental observation is that no company has a sufficiently diversified data or product portfolio to proceed on its own. In direct analogy with the drug discovery counterpart, the road to AI-driven process development goes through data ecosystems and thus joint ventures. This is expected to create compatibility issues which is why contextualization of data is identified as a pivotal competence.

We have identified studies that can be classified as preludes to AI-driven process development. Their synthesis advocates two things: (a) methodologically, AI is now mature to be employed for the task at hand, provided that the overall exercise is driven by multi-disciplinary experts and (b) *in silico* process development should be understood as a collection of models, with each process unit being endowed with its own model.

Herein, we have limited ourselves to *in silico* process development. Nevertheless, the potential of AI in process development extends further to process maintenance and control in manufacturing. Although questions concerning both technical aspects (e.g. appropriate PAT for online or at-line measurements, [50]) and/or regulatory requirements do exist [10], discussions are already intensifying even at the level of CDMOs [51].

Author's contributions

M.v.S. and C.V. conceived of the presented idea. All authors discussed the results and contributed to the final manuscript

Conflict of interest statement

M.v.S., R.M.C.P. and C.V. are, or were at the time of the study, employees of the GSK group of companies. M.v.S. is now an employee of DataHow AG, Zurich, Switzerland, and C.V. is now an employee of Janssen, part of the pharmaceutical companies of Johnson & Johnson.

Sponsorship

This work was sponsored by GlaxoSmithKline Biologicals SA.

Acknowledgements

The authors acknowledge Julia Pence and Sandrine Dessoy for the scientific review of the paper. The help of Ulrike Krause and Lucy Hitchcock is gratefully aknowledged. We also thank the anonymous Reviewers whose comments improved the quality of the manuscript.

References

1. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T: **The rise of deep learning in drug discovery**. *Drug Discovery Today* 2018, **23**:1241-1250 http://dx.doi.org/10.1016/j.drudis.2018.01.039.

2. •• Fleming N: **How artificial intelligence is changing drug discovery**. *Nature* 2018, **557**:S55-S57 http://dx.doi.org/10.1038/d41586-018-05267-x
A review article that summarizes the state-of-the-art of AI in drug discovery and offers a critical and compelling commentary.

3. Stott K: *Pharma's Broken Business Model: An Industry on the Brink of Terminal Decline*. Endpoints News; 2017 https://endpts.com/pharmas-broken-business-model-an-industry-on-the-brink-of-terminal-decline/.

4. Kristensen U: *VC Funding for AI in Drug Development & Clinical Trials Hits $5.2B*. 2019 https://www.signifyresearch.net/healthcare-it/vc-funding-ai-drug-development-clinical-trials-hits-5-2b/.

5. Research and Markets: *Growth Insight—Role of AI in the Pharmaceutical Industry, Global, 2018-2022*. 2019 https://www.researchandmarkets.com/reports/4846380/growth-insight-role-of-ai-in-the-pharmaceutical.

6. •• Hinkson IV, Madej B, Stahlberg EA: **Accelerating therapeutics for opportunities in medicine: a paradigm shift in drug discovery**. *Front Pharmacol* 2020, **11**:770 http://dx.doi.org/10.3389/fphar.2020.00770
The article presents the ATOM consortium: a government/academia/industry ecosystem aiming in disrupting drug discovery. This is an example of a paradigm shift.

7. Arshad Z, Smith J, Roberts M, Lee WH, Davies B, Bure K, Hollander GA, Dopson S, Bountra C, Brindley D: **Open access could transform drug discovery: a case study of JQ1**. *Expert Opin Drug Discov* 2016, **11**:321-332 http://dx.doi.org/10.1517/17460441.2016.1144587.

8. MELLODDY (2020, November 18): *Melloddy*. https://www.melloddy.eu/.

9. Digital McKinsey: *Digital Manufacturing – escaping pilot purgatory*. 2018 In: *https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Operations/Our%20Insights/How%20digital%20manufacturing%20can%20escape%20pilot%20purgatory/Digital-Manufacturing-escaping-pilot-purgatory.pdf*.

10. Demesmaeker M, Kopec D, Arsenio A: **Bioprocessing 4.0 – where are we with smart manufacturing in 2020?** *Pharm Outsourcing* 2020, **21**:28-31.

11. Benedetti A, Khoo J, Sharma S, Facco P, Barolo M, Zomer S: **Data analytics on raw material properties to accelerate pharmaceutical drug development**. *Int J Pharm* 2019, **563**:122-134 http://dx.doi.org/10.1016/j.ijpharm.2019.04.002.

12. Rogers A, Ierapetritou M: **Challenges and opportunities in modeling pharmaceutical manufacturing processes**. *Comput Chem Eng* 2015, **81**:32-39 http://dx.doi.org/10.1016/j.compchemeng.2015.03.018.

13. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S: **Applications of machine learning in drug discovery and development**. *Nat Rev Drug Discov* 2019, **18**:463-477 http://dx.doi.org/10.1038/s41573-019-0024-5.

14. •• Chatterjee S, Moore CMV, Nasr MM: **An overview of the role of mathematical models in implementation of quality by design paradigm for drug development and manufacture**. In *Comprehensive Quality by Design for Pharmaceutical Product Development and Manufacture.* Edited by Reklaitis GV, Seymour C, García-Munoz S. John Wiley & Sons, Inc.; 2017:9-24 http://dx.doi.org/10.1002/9781119356189.ch2
A very important publication that underlies the role of modeling in QbD and what that this imply in the regulatory front.

15. Parasrampuria DA, Benet LZ, Sharma A: **Why drugs fail in late stages of development: case study analyses from the last decade and recommendations**. *AAPS J* 2018, **20**:46 http://dx.doi.org/10.1208/s12248-018-0204-y.

16. Glass H, Muniz E, Kolassa E: **Drug development through modeling and simulation—the business case**. *Applied Clinical Trials Online*. 2016 In: *https://www.appliedclinicaltrialsonline.com/view/drug-development-through-modeling-and-simulation-business-case*.

17. Appl C, Moser A, Baganz F, Hass VC: *Digital Twins for Bioprocess Control Strategy Development and Realisation*. Berlin, Heidelberg: Springer; 2020 http://dx.doi.org/10.1007/10_2020_151.

18. Gottschalk U, Brorson K, Shukla AA: **Innovation in biomanufacturing: the only way forward**. *Pharm Bioprocess* 2013, **1**:141-157 http://dx.doi.org/10.4155/pbp.13.17.

19. Narayanan H, Luna MF, Stosch M, Cruz Bournazou MN, Polotti G, Morbidelli M, Butté A, Sokolov M: **Bioprocessing in the digital age: the role of process models**. *Biotechnol J* 2020, **15**:1900172 http://dx.doi.org/10.1002/biot.201900172.

20. Kis Z, Kontoravdi C, Dey AK, Shattock R, Shah N: **Rapid development and deployment of high-volume vaccines for pandemic response**. *J Adv Manuf Process* 2020, **2**:3 http://dx.doi.org/10.1002/amp2.10060.

21. Macdonald GJ: **Biomanufacturing makes sense of the industry 4.0 concept: the digital transformation of biomanufacturing poses short-term difficulties but promises long-term benefits through real-time monitoring and dynamic process optimization**. *Genet Eng Biotechnol News* 2020, **40**:S7-S10 http://dx.doi.org/10.1089/gen.40.S3.03.

22. Varsakelis C: **Manufacturing in the industry 4.0 era: balancing innovation with performance while not being disrupted**. *Biopharma Asia* 2020, **9**.

23. Mak K-K, Pichika MR: **Artificial intelligence in drug development: present status and future prospects**. *Drug Discov Today* 2019, **24**:773-780 http://dx.doi.org/10.1016/j.drudis.2018.11.014.

24. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK: **Artificial intelligence in drug discovery and development**. *Drug Discov Today* 2021, **26**:80-93 http://dx.doi.org/10.1016/j.drudis.2020.10.010 PubMed.

25. • Portela RMC, Varsakelis C, Richelle A, Giannelos N, Pence J, Dessoy S, von Stosch M: *When is an In Silico Representation a Digital Twin? A Biopharmaceutical Industry Approach to the Digital Twin Concept*. Berlin Heidelberg: Springer; 2020 http://dx.doi.org/10.1007/10_2020_138
Provides cost-reduction opportunities related to Digital Twins applications as well as numbers for potential cost savings by shortening development times using Digital Twins.

26. Hutter C, von Stosch M, Bournazou MNC, Butté A: **Knowledge transfer across cell lines using Hybrid Gaussian process models with entity embedding vectors**. *ArXiv:2011.13863 [Cs, q-Bio, Stat]* 2020 In: *http://arxiv.org/abs/2011.13863*.

27. Whitford W, Horbelt D: *Process development in the digital biomanufacturing age*. 2017:65-68 In: *https://cdn.cytivalifesciences.com/dmm3bwsv3/AssetStream.aspx?mediaformatid=10061&destinationid=10016&assetid=26707*.

28. Talib MA, Majzoub S, Nasir Q, Jamal D: **A systematic literature review on hardware implementation of artificial intelligence algorithms**. *J Supercomput* 2021, **77**:1897-1938 http://dx.doi.org/10.1007/s11227-020-03325-8.

29. • Cardillo A, Castellanos MM, Dessailly B, Dessoy S, Mariti M, Portela RMC, Scutellà B, von Stosch M, Tomba E, Varsakelis C: **Towards in silico process modeling for vaccines**. *Trends Biotechnol* 2021 http://dx.doi.org/10.1016/j.tibtech.2021.02.004
This is the first effort that systematically addresses the mechanistically focused *in silico* process development for vaccines.

30. •• Rathore AS, Winkle H: **Quality by design for biopharmaceuticals**. *Nat Biotechnol* 2009, **27**:26-34 http://dx.doi.org/10.1038/nbt0109-26
Details and explains the quality by design paradigm for biopharmaceuticals in a clear and concise manner and provides an easy yet comprehensive entry into this topic.

31. Bean R: *Pharmaceutical Industry as the Next Frontier for Data-driven Leadership*. Forbes; 2019 In: *https://www.forbes.com/sites/ciocentral/2019/12/09/pharmaceutical-industry-as-the-next-frontier-for-data-driven-leadership/*.

32. Cattell J, Chilukuri S, Levy M: *How Big Data Can Revolutionize Pharmaceutical R&D*. 2013 In: *https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d*.

33. Facco P, Zomer S, Rowland-Jones RC, Marsh D, Diaz-Fernandez P, Finka G, Bezzo F, Barolo M: **Using data analytics to accelerate biopharmaceutical process scale-up**. *Biochem Eng J* 2020, **164**:107791 http://dx.doi.org/10.1016/j.bej.2020.107791.

34. Process Excellence Network: *6 Ways Pharmaceutical Companies are Using Data Analytics to Drive Innovation & Value*. 2015 In: *https://www.mckinsey.com/business-functions/operations/our-insights/how-digital-manufacturing-can-escape-pilot-purgatory#*.

35. Allotrope Foundation: *Allotropefoundation*. 2021 In: *https://www.allotrope.org*.

36. Porr M, Schwarz S, Lange F, Niemeyer L, Hentrop T, Marquard D, Lindner P, Scheper T, Beutel S: **Bringing IoT to the lab: SiLA2 and open-source-powered gateway module for integrating legacy devices into the digital laboratory**. *HardwareX* 2020, **8**:e00118 http://dx.doi.org/10.1016/j.ohx.2020.e00118.

37. • Rodrigues de Azevedo C, von Stosch M, Costa MS, Ramos A, Cardoso MM, Danhier F, Préat V, Oliveira R: **Modeling of the burst release from PLGA micro- and nanoparticles as function of physicochemical parameters and formulation characteristics**. *Int J Pharm* 2017, **532**:229-240 http://dx.doi.org/10.1016/j.ijpharm.2017.08.118 PubMed
Modeling framework that renders possible *in silico* process development to obtain desired drug release rates.

38. Karlberg M, von Stosch M, Glassey J: **Exploiting mAb structure characteristics for a directed QbD implementation in early process development**. *Crit Rev Biotechnol* 2018, **38**:957-970 http://dx.doi.org/10.1080/07388551.2017.1421899.

39. Ebrahim A, Brunk E, Tan J, O'Brien EJ, Kim D, Szubin R, Lerman JA, Lechner A, Sastry A, Bordbar A *et al.*: **Multi-omic data integration enables discovery of hidden biological regularities**. *Nat Commun* 2016, **7**:13091 http://dx.doi.org/10.1038/ncomms13091.

40. Berry B, Moretto J, Matthews T, Smelko J, Wiltberger K: **Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis**. *Biotechnol Prog* 2015, **31**:566-577 http://dx.doi.org/10.1002/btpr.2035.

41. Chen Z-P, Zhong L-J, Nordon A, Littlejohn D, Holden M, Fazenda M, Harvey L, McNeil B, Faulkner J, Morris J: **Calibration of multiplexed fiber-optic spectroscopy**. *Anal Chem* 2011, **83**:2655-2659 http://dx.doi.org/10.1021/ac103145a.

42. Mehdizadeh H, Lauri D, Karry KM, Moshgbar M, Procopio-Melino R, Drapeau D: **Generic Raman-based calibration models enabling real-time monitoring of cell culture bioreactors**. *Biotechnol Prog* 2015, **31**:1004-1013 http://dx.doi.org/10.1002/btpr.2079.

43. Steinbach D, Anderson CA, McGeorge G, Igne B, Bondi RW, Drennen JK: **Calibration transfer of a quantitative transmission Raman PLS model: direct transfer vs. global modeling**. *J Pharm Innov* 2017, **12**:347-356 http://dx.doi.org/10.1007/s12247-017-9299-4.

44. Tulsyan A, Schorner G, Khodabandehlou H, Wang T, Coufal M, Undey C: **A machine-learning approach to calibrate generic Raman models for real-time monitoring of cell culture processes**. *Biotechnol Bioeng* 2019, **116**:2575-2586 http://dx.doi.org/10.1002/bit.27100.

45. Portela RMC, von Stosch M, Oliveira R: **Hybrid semiparametric systems for quantitative sequence-activity modeling of synthetic biological parts**. *Synth Biol* 2018, **3**:ysy010 http://dx.doi.org/10.1093/synbio/ysy010.

46. Destro F, Facco P, García Muñoz S, Bezzo F, Barolo M: **A hybrid framework for process monitoring: enhancing data-driven methodologies with state and parameter estimation**. *J Process Control* 2020, **92**:333-351 http://dx.doi.org/10.1016/j.jprocont.2020.06.002.

47. •• Smiatek J, Jung A, Bluhmki E: **Towards a digital bioprocess replica: computational approaches in biopharmaceutical development and manufacturing**. *Trends Biotechnol* 2020, **38**:1141-1153 http://dx.doi.org/10.1016/j.tibtech.2020.05.008
A timely article that discusses the way forward towards Bioprocessing 4.0 and the different modeling capabilities.

48. von Stosch M, Oliveira R, Peres J, Feyo de Azevedo S: **Hybrid semi-parametric modeling in process systems engineering: past, present and future**. *Comput Chem Eng* 2014, **60**:86-101 http://dx.doi.org/10.1016/j.compchemeng.2013.08.008.

49. Yang S, Navarathna P, Ghosh S, Bequette BW: **Hybrid modeling in the era of smart manufacturing**. *Comput Chem Eng* 2020, **140**:106874 http://dx.doi.org/10.1016/j.compchemeng.2020.106874.

50. Guerra A, von Stosch M, Glassey J: **Toward biotherapeutic product real-time quality monitoring**. *Crit Rev Biotechnol* 2019, **39**:289-305 http://dx.doi.org/10.1080/07388551.2018.1524362.

51. Macdonald GJ: **CDMOs embrace industry 4.0 to satisfy customer demand: the digitalization of bioprocessing technology isn't happening for its own sake, but to ensure that outsourced manufacturing tasks, however diverse, will be part of highly integrated processes**. *Gen Eng Biotechnol News* 2021, **41**:43-45 http://dx.doi.org/10.1089/gen.41.04.14.

52. Dinges R: *Risk-based Speed: Approaches for Fast to FIH and Accelerated Process Development*. 2017.

53. Jones S, Seymour P, Levin H: *CMC Activities for Development of MAbs*. Contract Pharma; 2010 In: *https://www.contractpharma.com/issues/2010-04/view_features/cmc-activities-for-development-of-mabs/*.

54. Stosch M, Varsakelis C: *AI Driven in Silico Process Development: The Business Case*. Sub Judice; 2021.