



Short Communication

Model-based optimization of enzyme cocktails for sustainable lignocellulosic biomass saccharification

Thomas Waluga^{a,*}, Hossein Nazarpour Najafabadi^a, Michael Zavrel^b, Mirko Skiborowski^a^a Institute of Process Systems Engineering, Hamburg University of Technology, Hamburg, Germany^b Professorship for Bioprocess Engineering, Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Straubing, Germany

ARTICLE INFO

Dataset link: <https://collaborating.tuhh.de/v-4/psi-public/mbo4saccharification.git>

Keywords:

Bayesian optimization

Saccharification

Lignocellulose

ABSTRACT

Lignocellulosic biomass has a high potential as resource for the chemical industry, but its conversion to fermentable sugars requires efficient pre-treatment and a subsequent enzymatic saccharification. Here, we present a model-based workflow for the design of enzyme cocktails for lignocellulosic biomass saccharification. Saccharification performance depends on several factors, such as substrate structure and pre-treatment severity, enzyme synergy, and inhibition effects. Because of this complexity, optimization by experimental trial-and-error is expensive and time-consuming, and enzyme mixtures optimized for one substrate often lack transferability to other feedstocks. To address these challenges, this work couples a stochastic simulator of lignocellulose hydrolysis with multi-objective optimization, explicitly resolving the trade-off between conversion speed and enzyme usage as a proxy for costs. The results demonstrate the feasibility and potential of model-based optimisation for rational enzyme cocktail design.

1. Introduction

Lignocellulosic biomass (LCB) is the most abundant renewable carbon source on Earth, offering a sustainable alternative to fossil fuels for the production of biofuels, biochemicals, and biomaterials (Brethauer et al., 2020). Comprising cellulose, hemicellulose, and lignin, LCB is a cornerstone of the circular bioeconomy, aligning with global sustainability goals such as climate action (SDG 13) and responsible consumption (SDG 12) (Bytyqi et al., 2025; Gohain and Borgohain, 2025). However, the enzymatic saccharification of LCB, converting its carbohydrate fraction into fermentable sugars, remains a critical bottleneck in biorefinery processes due to the structural complexity of plant cell walls and the economic constraints of enzyme usage (Guo et al., 2018).

The efficiency of saccharification depends on three interdependent factors: substrate properties (e.g., crystallinity, lignin content), enzyme synergy (e.g., endoglucanases, lytic polysaccharide monooxygenases, cellobiohydrolases, β -glucosidases), and operating conditions (e.g., temperature, pH) (De et al., 2025). While pre-treatment methods such as steam explosion or alkaline delignification can reduce recalcitrance, they often leave residual barriers, such as dissolved lignin droplets or crystalline cellulose, that limit enzyme accessibility (Bytyqi et al., 2025; Huang et al., 2022). Moreover, product inhibition further complicates

the design of optimal enzyme cocktails (Guo et al., 2018). Traditional empirical approaches like trial-and-error supplementation are resource-intensive and yield solutions that are substrate-specific, lacking generalizability (Lopes et al., 2018).

Mechanistic models, such as the Plant Cell Wall Saccharification Model (PCWSM) (Behle and Ragun, 2021), simulate saccharification kinetics at the mesoscale, capturing several major phenomena like enzyme adsorption, processivity, and inhibition. However, optimizing these models requires efficient algorithms capable of navigating high-dimensional, noisy, and computationally expensive design spaces. The complexity of the simulation model and the lack of derivative information require a sample-efficient derivative-free optimization approach. The stochastic nature of the simulation model makes a Bayesian optimization approach particularly suited (Han and Ouyang, 2022). Based on these requirements and previously reported results for experimental design we considered the Thompson Sampling Efficient Multi-Objective Optimization (TSEMO) (Bradford et al., 2018) as particularly suited for this task. TSEMO combines Bayesian optimization with surrogate models for multi-objective acquisition functions to identify Pareto-optimal trade-offs between conflicting objectives like conversion time vs. costs, here represented by enzyme loading.

As a proof of concept, we therefore decided to couple the PCWSM

* Corresponding author.

E-mail address: thomas.waluga@tuhh.de (T. Waluga).<https://doi.org/10.1016/j.dche.2026.100344>

Received 29 April 2026; Received in revised form 17 August 2026; Accepted 15 September 2026

Available online 16 September 2026

2772-5081/© 2026 The Author(s). Published by Elsevier Ltd on behalf of Institution of Chemical Engineers (IChemE). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Open access funding enabled and organized by Projekt DEAL.

model with TSEMO to systematically optimize enzyme cocktails for substrates of varying pre-treatment severity data. The original data for the pre-treated substrate are published by Bura (Bura et al., 2009). These data are also used for model validation (Behle and Raguin, 2021). While both PCWSM and TSEMO are established methodologies, their combination enables the substrate-dependent optimization of enzyme dosing while explicitly accounting for the trade-off between reaction time and enzyme usage. This approach leverages TSEMO's sample efficiency and robustness to stochastic noise to explore the design space of enzyme dosing units. The results reveal substrate-specific Pareto fronts, highlighting the trade-offs between reaction time and enzyme usage, and provide actionable insights for biorefinery design. Thus, the novelty of the present work lies not in the development of a new modelling or optimization method, but in integrating these complementary approaches namely an external optimization approach with the stochastic simulation model, into a workflow for substrate-specific enzyme cocktail optimization.

2. Materials and methods

2.1. Stochastic saccharification model

The optimization workflow couples an extended version of the PCWSM (Behle and Raguin, 2021) with TSEMO. PCWSM is a mesoscale, stochastic simulator based on the Gillespie algorithm, representing enzymatic hydrolysis of a coarse-grained microfibril structure (Fig. 1). The model includes four enzyme classes endoglucanases (EG), cellobiohydrolases (CBH), β -glucosidases (BGL), and xylanases (XYL) and simulates their interactions with cellulose, hemicellulose, and lignin. This model assumes that there is only one CBH, which hydrolyses from both ends, whereas in reality there are CBH I and CBH II, which hydrolyse from different ends of the polymer and may have different kinetic properties. The single effective CBH is parameterised as Cel7A-like because the objective is total cellobiose release rather than end-type specificity.

Within this study the PCWSM was extended to include an additional product inhibition as well as the processivity of the CBH. The inhibition was implemented as published by (Holtzapple et al., 1984) with

$$f_{inh}(t) = \frac{1}{1 + \frac{G(t)}{K_G} + \frac{B(t)}{K_B}} \quad (1)$$

where $G(t)$ and $B(t)$ are glucose and cellobiose concentrations that result in a reduction of the enzymes activity, with K_G and K_B are inhibition constants. The parameters for the CBH processivity were adjusted to reflect *Trichoderma reesei* kinetics (Brady et al., 2015; Westh et al., 2025). The factor scales enzyme activity when the reaction propensities are calculated. Including it lowers the overall mean squared deviation from measured saccharification (cf. Supplementary Information).

According to the original publication of the PCWSM, three variants of substrate with different severity of pre-treatment were simulated (Behle and Raguin, 2021). Each simulation run used 10 independent replicates: the normalised deviation between the mean conversion curves from 10 and from 100 replicates is 0.93%, 0.29% and 1.5% for high, medium and low severity, that is about 1% at one-tenth of the cost.

2.2. Optimization problem statement

The enzyme cocktail was represented as an integer decision vector $x = [x_{EG}, x_{CBH}, x_{BGL}, x_{XYL}]$, where x_i denotes dosing units (DU) for enzyme i . The bi-objective optimization problem was further defined according to Eq. (2) as

$$\min_x (t_{95}(x), L(x)) \quad (2)$$

where $t_{95}(x)$ is the time to reach 95% glucan conversion (stochastic, evaluated via restricted mean survival time (RMST) (Royston and Parmar, 2013)). $L(x) = \sum x_i$ represents the total enzyme loading (DU), constrained to $L \leq 1036$ DU. This threshold was set as it allows 95% of conversion for all substrates, while lower threshold ended up in incomplete conversion for the substrate with a low severity of pre-treatment. The target of 95% was chosen as an industrially desirable benchmark and as the most demanding target all three substrates reach, which allows the severities to be compared on a common basis. An additional inequality constraint was set as $x_i \geq 1$ DU for all enzymes, ensuring inactivity of all enzymes. To select one cocktail from a front we report its knee, the solution with the smallest Euclidean distance to the utopia point after min-max normalisation of both objectives over the feasible front

2.3. Bayesian optimization with TSEMO

To efficiently navigate the discrete design space of enzyme cocktails, we implemented a multi-objective Bayesian optimization approach by

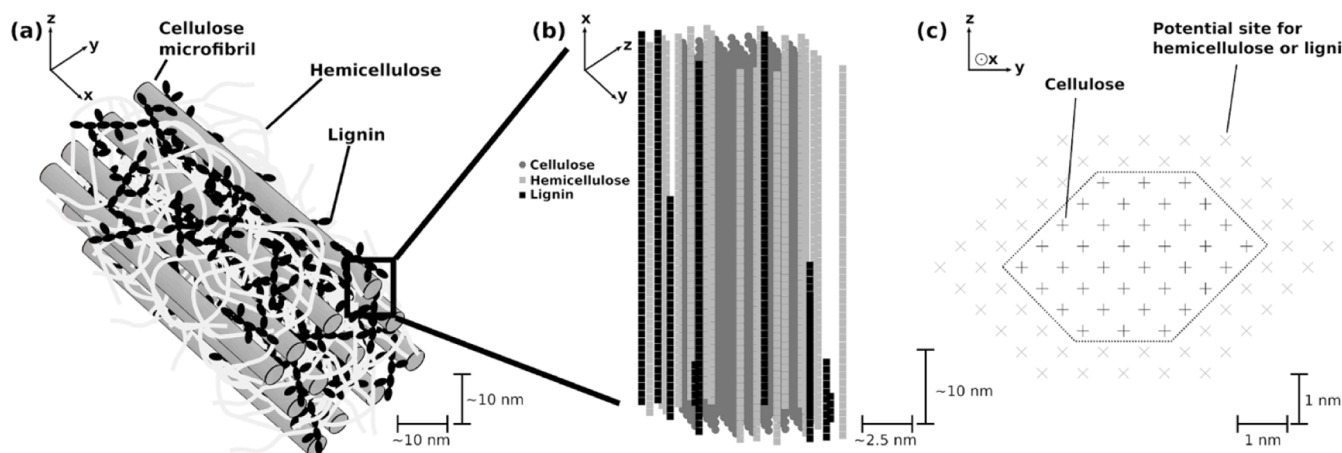


Fig. 1. Schematic representation of the substrate structure used in the model. (a) Multiple cellulose microfibrils are embedded in a hemicellulose matrix reinforced with lignin. (b) Side view of a single cellulose microfibril, together with part of the surrounding matrix, illustrating the relative arrangement of hemicellulose and lignin as well as the presence of gaps. In both (a) and (b), the different polymer types are indicated as cellulose: dark grey; hemicellulose: light grey; lignin: black. (c) Top view of the structure depicted in (b). The core region, marked by the dotted line, consists of cellulose chains (black crosses) arranged according to the structure reported by (Ding et al., 2014). The positions in the two outer layers (grey crosses) may be occupied by either hemicellulose or lignin polymers. Reproduced from (Behle and Raguin, 2021) under the Creative Commons Attribution 4.0 License (CC BY 4.0).

employing TSEMO (Bradford et al., 2018). Three properties of the simulator shape this coupling: a right-censored objective, aggregated as a restricted mean survival time truncated at the horizon rather than as a mean over replicates; an integer, budget-constrained feasible set, handled directly on a discrete candidate pool instead of by relaxation and rounding; and a second objective that is deterministic and linear in the decision vector, so one Gaussian process suffices rather than one per objective. Candidates are proposed in each iteration by a Gaussian process (GP) regressor for the stochastic objective $t_{95}(x)$, while the total enzyme loading follows directly from the dosages. The GP's predictive mean and uncertainty were used to generate Thompson samples (8 per iteration), which scored candidates based on their hypervolume improvement relative to the current Pareto front. Candidate selection was restricted to a discrete pool of 25,000 feasible cocktails, generated via random sampling under the constraints $x_i \geq 1$ DU and $L(x) \leq 1036$ DU. The pool is redrawn in every iteration, so the union examined over a run is far larger. The solutions nevertheless remain optimal within the sampled pool rather than over the full integer space, and the fronts should be read as approximations. The selected cocktail was then evaluated using the saccharification model, with t_{95} estimated as the RMST from 10 independent simulation replicates. This approach accounted for stochastic variability while avoiding bias from censored runs (i.e., trajectories failing to reach 95% conversion within a 149 h horizon a 25% margin above the longest time observed in preliminary runs). The optimization began with a precomputed initialization dataset of 300 evaluated cocktails from a space-filling design. Convergence was monitored using the dominated hypervolume (HV) of the Pareto front, with the algorithm terminating when the HV improvement fell below 1% over 10 consecutive iterations. This criterion balanced solution quality with computational efficiency, particularly critical for the stochastic, expensive evaluations. Alternative stopping rules were not tested. Every optimization was instead repeated with the criterion deactivated and the full 1200-evaluation budget exhausted (cf. Supplementary Information).

2.4. Software and tools

The framework for the simulation-based optimization was implemented in Python 3.8, leveraging the scikit-learn 1.5.1 library for Gaussian process regression with NumPy 1.26 and SciPy 1.11 and custom scripts for TSEMO and subsequent analysis of the resulting Pareto front. The stochastic saccharification model was executed as a command-line tool in C++, with parameter files defining substrate properties, enzyme kinetics, and simulation controls. The simulator was compiled with GCC 11.2 and all computations were performed on the TUHH high-performance computing cluster. Optimizer streams were seeded from a fixed base seed. The replicate generator inside the simulator is randomly seeded, so repeated runs reproduce the reported means and intervals but not individual trajectories. The code and data are available in our repository (cf. Data availability statement) to ensure reproducibility within the range of the stochastic saccharification model.

3. Results and discussion

3.1. Model-based optimization reveals trade-offs in enzyme cocktail design

The simulation-based optimization coupling PCWSM with TSEMO enabled systematic exploration of enzyme cocktail formulations across substrates of varying pre-treatment severity. The resulting Pareto fronts are illustrated in Fig. 2 for the different pre-treatments and quantify the inherent trade-off between processing time (time to 95% glucan conversion) and enzyme costs (here total loading in dosing units), providing actionable insights for biorefinery design. For the high-severity substrate, TSEMO identified solutions achieving 95% conversion in under 2

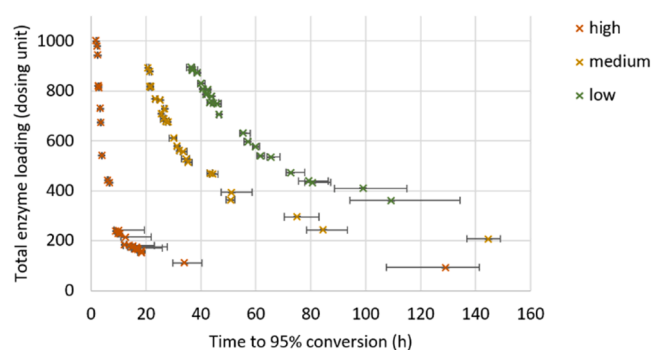


Fig. 2. Pareto fronts under relaxed enzyme loading in the objective space of time to 95% conversion and total enzyme loading for substrates with different severity of pre-treatment. Error bars represent the 95% confidence intervals for 10 simulations.

h at loadings of 1000 DU, while the low-severity substrate required significant reaction times (>20 h). This shift reflects the increasing recalcitrance of less pre-treated biomass, where residual lignin and cellulose crystallinity limit enzyme accessibility (Bisaria, 2024).

Selecting one cocktail from a front requires an explicit rule. The knee is listed with the extreme solutions in Table 1. It separates steep from diminishing returns: on the high-severity substrate the step from minimum loading to the knee cuts the conversion time from 142.8 h to 11.2 h for 2.4 times the loading, whereas the step from the knee to minimum time buys a further factor of six at 5.4 times the loading, added almost entirely as CBH, so the preferable region follows from the ratio of reactor-time to enzyme cost. Composition is largely conserved across severities but dose is not: 95% conversion within 50 h needs 111 DU at high severity against 394 DU and 706 DU at medium and low severity, and conversion within 20 h is unattainable at any admissible loading on the two milder pre-treatments.

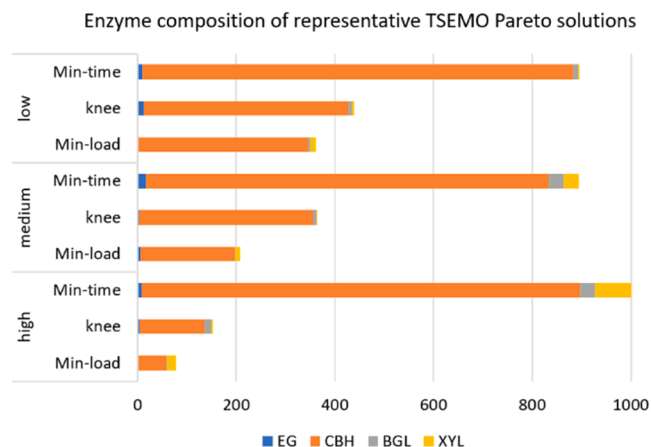
3.2. Enzyme allocation patterns reflect substrate-specific bottlenecks

Analysis of representative Pareto-optimal solutions revealed distinct enzyme allocation strategies tailored to substrate properties (Fig. 3). Across all substrates, cellobiohydrolases (CBH) dominated fast-conversion solutions, comprising 89–98% of the total loading in Min-time cocktails. This aligns with the critical role of CBH in processively degrading cellulose (Sinha et al., 2023). For example, the Min-time solution for the high-severity substrate (1.81 h) allocated 89% CBH of total loading, underscoring its efficiency in accessible substrates. In contrast the Min-loading solution for the high-severity substrate (142.8 h) allocated 73% CBH. A similar trend can be seen for the low-severity pre-treated substrate, where the share of CBH is 95% (min-load solution) increasing to 98% (min-time solution). In contrast, the share of XYL is high for the min-load solutions (22% and 3% for high and low severity of pre-treatment, respectively) while it is low for min-time solutions (7% and 0.2% for high and low severity of pre-treatment, respectively). The high CBH fractions in Min-time solutions align with its processive mechanism, which is critical for degrading crystalline cellulose. CBH binds to chain ends and sequentially releases cellobiose, making it the most efficient enzyme for rapid depolymerization of accessible substrates. This dominance is particularly pronounced in high-severity substrates, where pre-treatment has already reduced recalcitrance, allowing CBH to act unhindered. CBH binds to chain ends and sequentially releases cellobiose, which makes it most efficient for rapid depolymerization and explains why its dominance is greatest where pre-treatment has already reduced recalcitrance. CBH-rich optima are likewise reported in experimental mixture-optimization studies on pre-treated lignocellulose (Machado et al. 2020). XYLs remove hemicellulose barriers, exposing cellulose to cellulases. In Min-loading solutions, where enzyme budgets are constrained, XYL's synergistic effect to

Table 1

Representative TSEMO Pareto optimal enzyme cocktails. For each substrate, the minimum-loading feasible solution (Min-load), knee-point solution (Knee), and minimum-time solution (Min-time) are reported as enzyme units, total loading $\sum e_i$ (dosing unit), and time to reach 95% conversion (hour).

Substrate	Point	EG	CBH	BGL	XYL	Loading (DU)	Time to 95% (h)	CBH share
high	Min-time	8	889	30	74	1001	1.81	88.8%
high	Knee	1	167	7	10	185	11.22	90.3%
high	Min-load	2	56	2	17	77	142.83	72.7%
medium	Min-time	16	817	31	30	894	20.89	91.4%
medium	Knee	3	353	7	1	364	50.79	97.0%
medium	Min-load	6	191	1	9	207	142.84	92.3%
low	Min-time	10	873	10	2	895	35.44	97.5%
low	Knee	5	521	8	6	540	61.12	96.5%
low	Min-load	2	344	4	11	361	105.16	95.3%

**Fig. 3.** Enzyme composition of representative TSEMO Pareto solutions.

enhance CBH access becomes critical, especially for low-severity substrates with residual xylan. The model thus prioritizes XYL to compensate for limited CBH by improving substrate accessibility.

3.3. Sampling efficiency of TSEMO and robustness in noisy environments

TSEMO demonstrated high evaluation efficiency for this computationally expensive optimization problem, as quantified by the dominated hypervolume (HV) metric. After the initialization phase (300 evaluations), TSEMO achieved rapid convergence, reaching 90% of the final HV within 500 evaluations for all substrates. TSEMO met its stopping criterion after 399, 407 and 408 evaluations; since 300 of these are the shared initialization design, only 99 to 108 were model-guided. That design is itself an optimizer-independent space-filling reference and already supplies 84%, 80% and 69% of the final HV reached by TSEMO. The HV is reported in the Supplementary Information. This efficiency stems from TSEMO's Thompson sampling strategy, which balances exploration and exploitation by scoring candidates based on surrogate-model uncertainty (Bradford et al., 2018). The Gaussian process (GP) surrogate, trained on the stochastic time-to-target evaluations, enabled TSEMO to prioritize promising regions of the design space while avoiding redundant simulations. For comparison, a genetic algorithm (GA) failed to converge within the 1200-evaluation budget for medium- and low-severity substrate and met the criterion after 961 evaluations on the high-severity substrate, exhibiting significantly lower HV values across all substrates (see Supplementary Information). While GA achieved comparable results to TSEMO for the low-severity substrate at

high enzyme loadings, TSEMO demonstrated superior performance for high and medium-severity substrates by identifying more efficient solutions with lower loadings and faster conversion times. TSEMO nevertheless dominated 15.0%, 22.0% and 40.3% more of the objective space, with a HV per evaluation higher by factors of 2.8, 3.6 and 4.1, and improved on the space-filling reference by 19.5%, 25.3% and 45.2% where the GA gained only 4.0%, 2.7% and 3.5% in 661 to 900 further evaluations (Supplementary Information). This underscores TSEMO's enhanced capability to explore complex, stochastic design spaces effectively. The stochastic nature of the saccharification model introduced variability in the time-to-target estimates, particularly for solutions with higher reaction times (cf. Fig. 2). TSEMO's surrogate model mitigated this noise by incorporating uncertainty into the acquisition function, resulting in a median interval width of 0.76 h, 1.95 h and 2.58 h for TSEMO and 0.21 h, 1.92 h and 3.30 h for the GA on the high-, medium- and low-severity substrate. For the high-severity substrate, TSEMO explored more extreme low-loading solutions, where stochastic effects were amplified due to near-plateau conversion kinetics. This highlights TSEMO's robustness in noisy environments, a critical advantage for optimization problems involving stochastic simulations. The workflow is also modular in the enzyme set: adding another enzyme function extends the decision vector and the candidate pool without changing the surrogate or the acquisition step, although the pool must grow to preserve its density.

Three findings emerge. First, the workflow resolves the trade-off between conversion time and enzyme loading quantitatively for three pre-treatment severities, and the knee gives an explicit basis for choosing one cocktail. Second, allocation is CBH-dominated throughout, with XYL gaining importance where the budget is tight, so composition is largely conserved while the required dose is not. Third, the surrogate-based search attains these fronts with roughly one hundred model-guided evaluations, two to three times fewer than the evolutionary baseline.

4. Conclusions

This work couples an extended stochastic plant cell wall saccharification model with multi-objective Bayesian optimization into a workflow that designs enzyme cocktails for a given substrate instead of by experimental trial-and-error. For three pre-treatment severities it delivers Pareto fronts quantifying the trade-off between the time to 95% glucan conversion and the total enzyme loading, and identifies knee cocktails that capture most of the attainable time reduction at a fraction of the maximum loading. Pareto-optimal cocktails are consistently cellobiohydrolase-dominated, while the dose required for a given conversion time rises several-fold as pre-treatment severity decreases. The integration proved evaluation-efficient, reaching these fronts with of the order of one hundred model-guided simulations and with a clear margin

over both an evolutionary baseline and a space-filling reference. For model-based bioprocess design this shows that mechanistic simulation and surrogate-based optimization can be combined into a practical formulation tool for sustainable biorefineries. The results are computational; experimental confirmation of the representative solutions, and the model extensions below, remain to be done.

5. Future work

The model captures key mechanisms (e.g., inhibition, enzyme synergy). It also assumes a dilute, well-mixed system and currently omits relevant aspects in real applications. At very high enzyme loadings, crowding effects may reduce the predicted benefits of CBH-dominated cocktails. The model also does not account for diffusion effects in high-solids systems. Also, only the four mentioned enzymes are considered for the saccharification only. The extensions below are ordered by expected impact: Other enzymes like Lytic polysaccharide monooxygenases are known to enhance cellulose accessibility (Martins et al., 2024; Nevalainen et al., 2024), but are not included in the current model. A high priority is resolving CBH I and II and adding lytic polysaccharide monooxygenases. Then experimental confirmation of the representative solutions and a better representation of accessibility and recalcitrance barriers, where the model is demonstrably weakest. Finally extending validity to high loadings. Methodologically, a second multi-objective Bayesian baseline, a recombining evolutionary comparison, a designed cross-substrate transfer study and a heteroscedastic noise model remain open, each requiring several hundred to several thousand further simulator evaluation. The total sugar yield should also be taken into account. Additionally, coupling TSEMO with high-throughput experimental platforms (Clayton et al., 2020) could accelerate the discovery of optimal enzyme cocktails for diverse feedstocks. Because pre-treatment severity changes the required dose several-fold, pre-treatment intensity and enzyme dosing are economically coupled, which makes joint optimization with upstream pre-treatment and downstream fermentation a natural extension of the same bi-objective formulation for the application in a biorefinery.

CRedit authorship contribution statement

Thomas Waluga: Writing – review & editing, Writing – original draft, Supervision, Project administration, Formal analysis, Conceptualization. **Hossein Nazarpour Najafabadi:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Michael Zavrel:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Mirko Skiborowski:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.dche.2026.100344](https://doi.org/10.1016/j.dche.2026.100344).

Data availability

<https://collaborating.tuhh.de/v-4/psi-public/mbo4saccharification.git> (All code, the archived simulation results, the initialization designs and the scripts are available at).

References

- Behle, E., Raguin, A., 2021. Stochastic model of lignocellulosic material saccharification. *PLoS Comput. Biol.* 17 (9), e1009262.
- Bisaria, V. (Ed.), 2024. Handbook of Biorefinery Research and Technology: Biomass Logistics to Saccharification. 1st ed. Springer, Dordrecht.
- Bradford, E., Schweidtmann, A.M., Lapkin, A., 2018. Efficient multiobjective optimization employing gaussian processes, spectral sampling and a genetic algorithm. *J. Glob. Optim.* 71 (2), 407–438.
- Brady, S.K., Sreelatha, S., Feng, Y., Chundawat, S.P.S., Lang, M.J., 2015. Cellobiohydrolase 1 from *Trichoderma reesei* degrades cellulose in single cellobiose steps. *Nat. Commun.* 6, 10149.
- Brethauer, S., Shahab, R.L., Studer, M.H., 2020. Impacts of biofilms on the conversion of cellulose. *Appl. Microbiol. Biotechnol.* 104 (12), 5201–5212.
- Bura, R., Chandra, R., Saddler, J., 2009. Influence of xylan on the enzymatic hydrolysis of steam-pretreated corn stover and hybrid poplar. *Biotechnol. Prog.* 25 (2), 315–322.
- Bytyqi, H., Mujdaci, G.N., Ekici, E., Ozturk, A.B., Celik Madenli, E., Kumar, G., 2025. Review of lignocellulosic bio-chemical production: current challenges, advances, and future perspectives. *Energy Nexus* 20, 100543.
- Clayton, A.D., Schweidtmann, A.M., Clemens, G., Manson, J.A., Taylor, C.J., Niño, C.G., et al., 2020. Automated self-optimisation of multi-step reaction and separation processes using machine learning. *Chem. Eng. J.* 384, 123340.
- Sakha De, P., Grande, P.M., Heise, H., Klose, H., Raguin, A., 2025. Stochastic model highlights the impact of crystallinity on saccharification dynamics depending on plant chemotype and pre-treatment. *PLoS One* 20 (12), e0322367.
- Ding, S.-Y., Zhao, S., Zeng, Y., 2014. Size, shape, and arrangement of native cellulose fibrils in maize cell walls. *Cellulose* 21 (2), 863–871.
- Gohain, P.B., Borgohain, D., 2025. Lignocellulosic biomass (LCB) for a sustainable circular economy. *Agric. Res. J.* 62 (3), 246–255.
- Guo, H., Chang, Y., Lee, D.-J., 2018. Enzymatic saccharification of lignocellulosic biorefinery: research focuses. *Bioresour. Technol.* 252, 198–215.
- Han, M., Ouyang, L., 2022. A novel Bayesian approach for multi-objective stochastic simulation optimization. *Swarm Evol. Comput.* 75, 101192.
- Holtzapfel, M.T., Caram, H.S., Humphrey, A.E., 1984. Determining the inhibition constants in the HCH-1 model of cellulose hydrolysis. *Biotechnol. Bioeng.* 26 (7), 753–757.
- Huang, C., Jiang, X., Shen, X., Hu, J., Tang, W., Wu, X., et al., 2022. Lignin-enzyme interaction: a roadblock for efficient enzymatic hydrolysis of lignocellulosics. *Renew. Sustain. Energy Rev.* 154, 111822.
- Lopes, A.M., Ferreira Filho, E.X., Moreira, L.R.S., 2018. An update on enzymatic cocktails for lignocellulose breakdown. *J. Appl. Microbiol.* 125 (3), 632–645.
- Machado, A.S., Valadares, F., Silva, T.F., Milagres, A.M.F., Segato, F., Ferraz, A., 2020. The secretome of *Phanerochaete chrysosporium* and *Trametes versicolor* grown in microcrystalline cellulose and use of the enzymes for hydrolysis of lignocellulosic materials. *Front. Bioeng. Biotechnol.* 8, 826.
- Martins, M., da Silva, M.F., de Melo, A.H.F., Ramalho, E.X., Goldbeck, R., 2024. Enzymatic bioconversion trends for lignocellulosic biorefineries. In: Bisaria, V. (Ed.), Handbook of Biorefinery Research and Technology: Biomass Logistics to Saccharification. Springer, Dordrecht, pp. 841–860.
- Nevalainen, H., Aggarwal, S., Adlakha, N., 2024. Sources, properties, and modification of lignocellulolytic enzymes for biomass degradation. In: Bisaria, V. (Ed.), Handbook of Biorefinery Research and Technology: Biomass Logistics to Saccharification. Springer, Dordrecht, pp. 567–605.
- Royston, P., Parmar, M.K.B., 2013. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med. Res. Methodol.* 13, 152.
- Sinha, T., Sharma, K., Yazdani, S.S., 2023. Cellobiohydrolases. In: Goyal, A., Sharma, K. (Eds.), Glycoside Hydrolases: Biochemistry, Biophysics, and Biotechnology. Foundations and Frontiers in Enzymology. Academic Press, pp. 77–95.
- Westh, P., Kari, J., Badino, S., Sørensen, T., Christensen, S., Røjel, N., et al., 2025. Are cellulases slow? Kinetic and thermodynamic limitations for enzymatic breakdown of cellulose. *BBA Adv.* 7, 100128.