*Synthetic Biology*, 2025, **10(1)**, ysaf006
DOI: https://doi.org/10.1093/synbio/ysaf006
Advance Access Publication Date: 29 April 2025
**Research Article**

# Synthetic biology design principles enable efficient bioproduction of Heparosan with low molecular weight and low polydispersion index for the biomedical industry

Yadira Boada[1,2], Marcelo Flores[1,3], Martin Stiebritz[4], Marco Córdova[5], Francisco Flores[5,6], Alejandro Vignoni[1,*]

[1]Synthetic Biology and Biosystems Control Lab, Instituto de Automática e Informática Industrial, Universitat Politècnica de València, Camino de Vera s/n, Valencia 46022, Spain
[2]Grado en Ingeniería y Gestión Empresarial, Centro Universitario EDEM, Plaça de L'aigua, Poblados Marítimos, Valencia 46024, Spain
[3]Grupos de Investigación en Cloud Computing Smart Cities and High Performance Computing, Universidad Politécnica Salesiana, Calle Vieja 12-30, Cuenca 010105, Ecuador
[4]Lehrstuhl für Biotechnik, Department für Biologie, Friederich-Alexander-Universität, MVC, Henkestraße 91, Erlangen 91052, Germany
[5]Departamento de Ciencias de la Vida y de la Agricultura, Universidad de las Fuerzas Armadas ESPE, Av. Gral. Rumiñahui s/n, Sangolquí 171103, Ecuador
[6]Centro de Investigación de Alimentos, CIAL, Facultad de Ciencias de la Ingeniería e Industrias, Universidad UTE, C. Rumipamba s/n y Bourgeois, Quito 170147, Ecuador
*Corresponding author. E-mail: vignoni@isa.upv.es

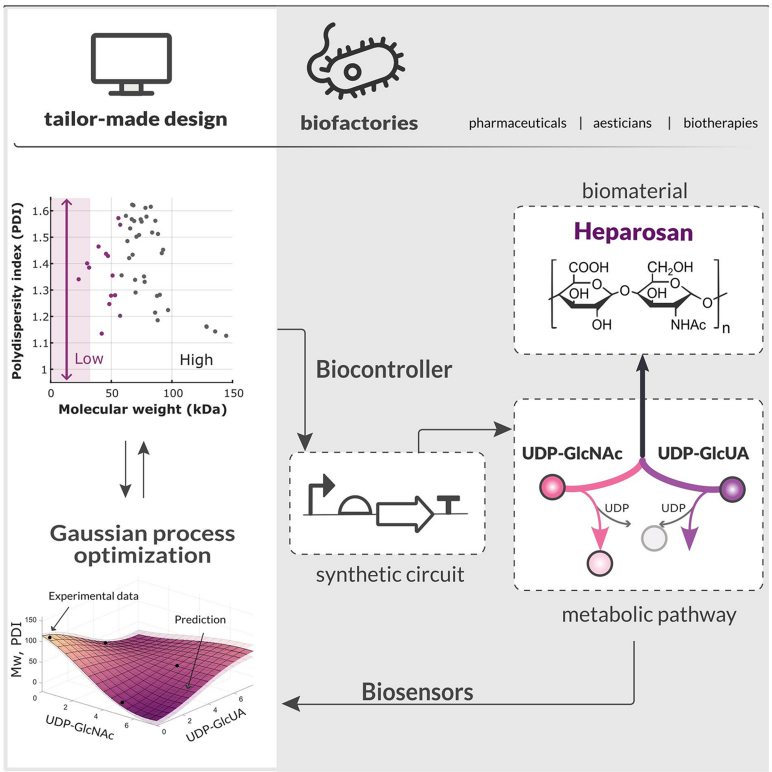
## Abstract

Heparosan is a natural polymer with unique chemical and biological properties, that holds great promise for biomedical applications. The molecular weight (Mw) and polydispersion index (PDI) are critical factors influencing the performance of heparosan-based materials. Achieving precise control over the synthesis process to consistently produce heparosan with low Mw and low PDI can be challenging, as it requires tight regulation of reaction conditions, enzyme activity, and precursor concentrations. We propose a novel approach utilizing synthetic biology principles to precisely control heparosan biosynthesis in bacteria. Our strategy involves designing a biomolecular controller that can regulate the expression of genes involved in heparosan biosynthesis. This controller is activated by biosensors that detect heparosan precursors, allowing for fine-tuned control of the polymerization process. Through this approach, we foresee the implementation of this synthetic device, demonstrating the potential to produce low Mw and low PDI heparosan in the probiotic *E. coli* Nissle 1917 as a biosafe and biosecure biofactory. This study represents a significant advancement in the field of heparosan production, offering new opportunities for the development and manufacturing of biomaterials with tailored properties for diverse biomedical applications.

**Keywords:** Biomanufacturing; synthetic biology; heparosan; polydispersion index

Submitted 7 July 2024; Received 7 July 2024; Revised 11 March 2025; Accepted 26 March 2025

## Graphical Abstract




# 1 Introduction

The global pharmaceutical industry faces significant challenges, including reliance on imported components for drug manufacturing, supply chain disruptions, and the need for sustainable and efficient production methods [56; 4]. Among these, the exploration of alternative sources and innovative manufacturing approaches has become imperative to meet the growing demand for pharmaceutical products worldwide [17; 13].

Heparosan is a natural polymer belonging to the glycosaminoglycan family that exhibits exceptional chemical and biological properties, making it invaluable in various biomedical applications [41]. While being normally utilized as a precursor to heparin [6; 58], recent research has unveiled its potential in diverse fields, including cosmetology, aesthetics, dermatology, and notably, pharmaceuticals [22; 51]. Heparosan's versatility has been underscored by its emergence as a promising alternative to hyaluronic acid, a widely utilized carbohydrate polymer in anticancer therapy [42; 49]. Whether as a precursor for heparin production or for emerging pharmacological uses, the molecular weight (Mw) of heparosan is a critical parameter inherent to its synthesis [15]. High Mw heparosan is a biomaterial with excellent viscoelastic properties, moisture retention, and greater biocompatibility compared to other biomaterials like hyaluronan, as it is more resistant to hyaluronidase [22; 31]. Conversely, low Mw heparosan has been utilised as an antithrombic or anticoagulant drug, as well as in pharmacological delivery applications [47]. Therefore, heparosan's Mw is a key characteristic that significantly influences its bioactivity, including the efficacy and functionality of its various uses.

Heparosan production in bacteria presents a major hurdle due to the competition for precursors essential for bacterial survival. Heparosan is synthesized from UDP-N-acetylglucosamine (UDP-GlcNAc) and UDP-glucuronic acid (UDP-GlcUA) [22; 38]. These UDP-sugars are vital metabolites in bacterial metabolism and growth (Fig. 1a). While UDP-GlcNAc is essential in the production of the cell wall peptidoglycans [5], UDP-GlcUA plays a role in the synthesis of glycoproteins, proteoglycans, and polysaccharide [26]. Consequently, redirecting these metabolic pathways towards heparosan biosynthesis can disrupt essential cellular functions, potentially compromising bacterial viability [14].

Additionally, heparosan synthesis presents inherent high polydispersity or polydispersion index (PDI) [15,37,46]. Polydispersity measures the size distribution within a polymer sample (Fig. 1b), and it significantly influences the quality, functionality, and reproducibility of biopolymers [6]. Heparosan with high PDI results in Mw variability, leading to heterogeneity and unpredictability of its properties, while hindering its suitability for various applications [14]. Therefore, the development of efficient and reproducible production methods capable of producing heparosan with a low PDI (ideally close to unity) remains a challenging task, although some efforts are already ongoing. [42].

Synthetic biology is an interdisciplinary field blending engineering principles with biological systems [1] based on the engineering design-build-test-learn (DBTL) cycle [20], which offers a promising avenue to address the challenges associated with heparosan production [51]. Leveraging synthetic biology approaches, particularly in *E. coli* as a model organism, has shown potential in enhancing the efficiency and quality of heparosan synthesis. By optimizing gene expression, utilizing alternative substrates, and engineering polymerization enzymes, synthetic biology enables precise control over heparosan biosynthesis, paving the way for low PDI heparosan production and different Mws.

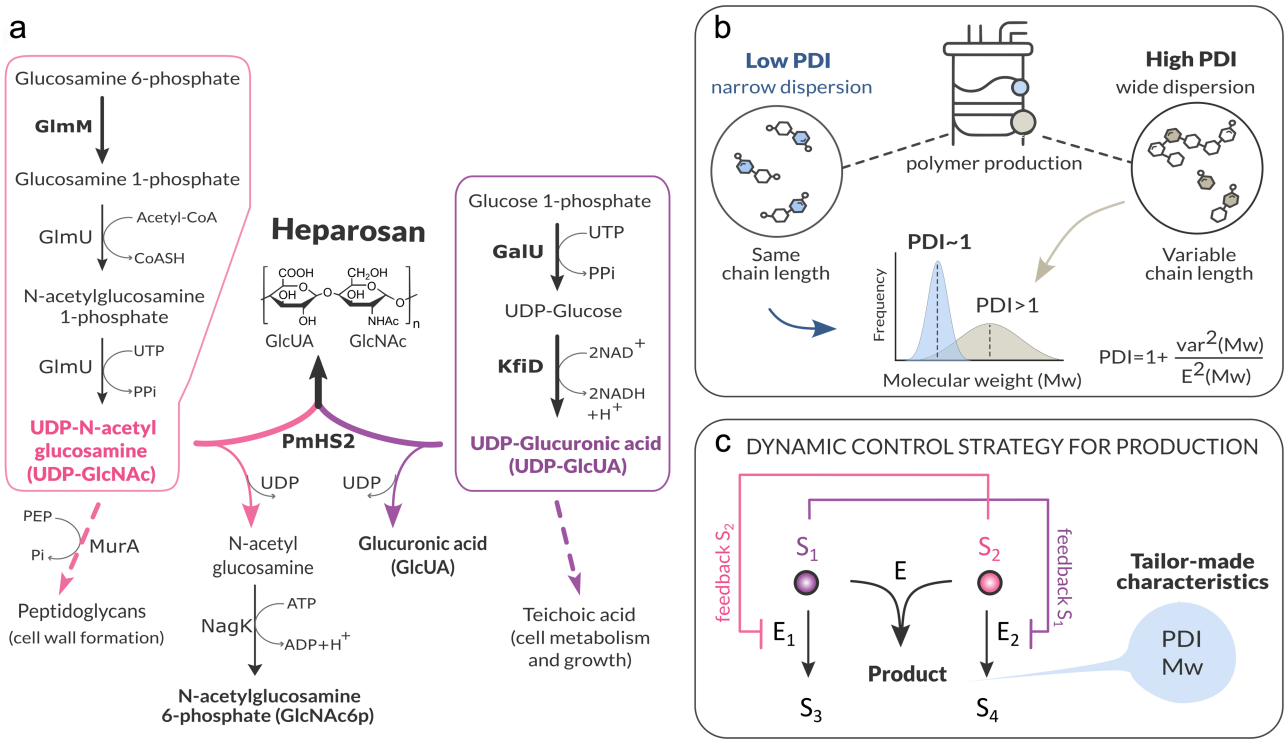



**Figure 1. a**. Heparosan production pathway from Glucosamine 6-phosphate (GlcN6P) and Glucose 1-phosphate (Glc1P). The pathway involves the uptake of UDP-GlcNAc for the production of the cell wall peptidoglycans, whereas UDP-GlcUA is consumed in the synthesis of teichoic acid, glycoproteins, proteoglycans, and polysaccharides. The panel also includes an extended metabolite biosensor (GlcNAc6p) for sensing UDP-GlcNAc. **b.** Heparosan polydispersity, where a beaker symbolizes the produced heparosan, with an explanation of the PDI formula provided. Additionally, two Mw distributions are depicted as examples: one with low PDI (dark blue curve), and another with high PDI (grey curve). The contrast between the distributions highlights the importance of achieving a low PDI for homogenous heparosan production. **c.** Control dynamic strategy applied to a generic metabolic pathway for converting precursors $S_1$ and $S_2$ into a target product through the action of enzyme E. $S_1$ and $S_2$ are also utilized for other cellular processes via enzymes $E_1$ and $E_2$, resulting in the generation of metabolites $S_3$ and $S_4$, respectively. Our strategy involves real-time sensing of precursors $S_1$ and $S_2$ while modulating the production of the opposite enzymes $E_2$ and $E_1$ through negative feedback loops to maintain a predefined balance between the precursors. This can ensure optimal production of a product despite using the precursors in other cellular processes.

In this study, we introduce a novel approach to heparosan production in *E. coli* Nissle 1917 [32] using a dynamic feedback biomolecular controller (Fig. 1c) rooted in synthetic biology principles [10]. In particular, we focus on the Design stage of the DBTL cycle with our *in-silico* experiments, and use the results to plan the subsequent Build stage. Here, we understand 'design' as defining the function, predicting the behaviour, and determining the structure of the genetic circuits through mathematical modelling, which is essential for this stage in synthetic biology. By using this model to set parameters that lead to desired outcomes, we effectively lay the foundation for selecting or creating genetic DNA parts in the Build stage, in line with the DBTL cycle principles. Ideally, the necessary DNA design could be made entirely out by reusing previously designed DNA parts (making new and different combinations of them), which constitutes one of the foundations of synthetic biology as a discipline. In practice and as synthetic biology matures as a discipline, we are still making new DNA designs as we go, but eventually, we will reach closer to the ideal.

Our biomolecular controller is tailored to dynamically regulate heparosan biosynthesis by sensing the concentration of the key precursors UDP-GlcNAc and UDP-GlcUA with two biosensors [12]. It allows us real-time monitoring and control of the polymerisation process to obtain the desired Mw heparosan biopolymers with the lowest possible PDI, even in the presence of changes in precursor availability (Fig. 1c). Our approach has advantage over *in-vitro* production systems, where real-time feedback is not possible.

Due to stochasticity underlying the heparosan polymerization reaction, we optimized heparosan Mw and PDI as a function of the concentration levels of its precursors UDP-GlcNAc and UDP-GlcUA following the Gaussian process (GP) regression [59]. Using this approach, we can predict the production of heparosan with desired Mw with a low PDI, while minimally affecting the fitness of the host cell. In addition, we can also obtain heparosan biopolymers with different Mws by interchanging the biological parts of our system, and following the design guidelines using the dynamic production model.

## 2 Methods

### 2.1 Heparosan biosynthesis pathway

Heparosan is a polymer naturally produced in different microorganisms such as *E. coli* as a result of several steps known as polymerization reaction (Fig. 1a). Polymerization chemically combines small molecules (units) of one or more metabolites (precursors), into a very large chain called polymer. Particularly, heparosan is constituted of the repeating unit of both UDP-GlcNAc and UDP-GlcUA synthesized by one or two enzymes depending on the microorganism [16]. The availability of these two UDP-sugars will influence the length, and thus the Mw of the produced heparosan polymers. A balance between the amounts of UDP-GlcNAc and UDP-GlcUA can determine the PDI of heparosan, which is a key aspect for its potential applications in the pharmaceutical industry [32; 51].

*2.1.1 Formation of UDP-GlcNAc* The first precursor involved in the heparosan production is UDP-GlcNAc (Fig. 1a pink panel). D-glucose-6-phosphate is converted to D-Fructose-6-phosphate by Glucose-6-phosphate isomerase, which in turn, is transformed to D-Glucosamine 6-phosphate by Glutamine-fructose-6-phosphate transaminase (GlmS). After being catalyzed by Phosphoglucosamine mutase (GlmM), the enzymes Glucosamine-1-phosphate N-acetyltransferase and UDP-N-acetylglucosamine diphosphorylase (GlmU) convert D-Glucosamine 6-phosphate into UDP-GlcNAc [22]. Interestingly, UDP-GlcNAc is a crucial metabolite involved in the cell wall formation [57].

*2.1.2 Formation of UDP-GlcUA* In Fig. 1a (purple panel), we can see the UDP-GlcUA branch involved in heparosan synthesis. Naturally, its formation starts with the conversion of glucose (carbon source) to D-glucose-6-phosphate by the enzyme hexokinase. Then, phosphoglucomutase transforms D-glucose-6-phosphate to D-glucose-1-phosphate which is subsequently converted to UDP-glucose under the action of the enzyme UTP-glucose-1-phosphate uridylyltransferase (GalU). UDP-glucose is finally converted into UDP-GlcUA by the UDP-glucose 6-dehydrogenase (KfiD) [15; 22]. UDP-GlcUA also plays an important role in the formation of Teichoic acid used for cell growth.

*2.1.3 Heparosan synthesis* Heparosan polymerization can be catalyzed by a range of enzymes from different organisms. For instance, *E. coli* K5 synthesizes heparosan using the glucosyltransferases KfiA and KfiC, which alternately add GlcNAc and GlcUA to extend the heparosan chain [22]. In our study, we selected PmHS2 from *P. multocida* as the primary synthase due to its high expression levels in *E. coli*, substrate flexibility, and independence from an oligosaccharide primer. These together enhance PmHS2 suitability for recombinant production of heparosan [32]. PmHS2 synthase performs the joint action of the KfiA-KfiC complex endogenous to *E. coli* Nissle, due to its two catalytic transferase domains with glucuronyltransferase (GlcUA-T) and N-acetylglucosaminyltransferase (GlcNAc-T). Also, PmHS2 has a high hydrolase activity with UDP-sugars, e.g. UDP-GlcNAc and UDP-GlcUA [14; 15]. In this way, we can reduce the number of reactions, and the number of biochemical species involved in the heparosan synthesis, while decreasing the cell metabolic burden. Moreover, PmHS2 has ~2-fold higher *de novo* initiation activity of heparosan chains, obtaining higher titers (from 104 mg/L/OD-20h, up to 1.29 g/L) in *E. coli* Nissle, with a moderate range of Mws [51].

PmHS2 synthesizes heparosan in two simultaneous phases: initiation and chain elongation. In the initiation phase, heparosan disaccharide is formed by the transfer of UDP-GlcUA (donor molecule) to UDP-GlcNAc as the acceptor molecule. This new compound molecule is known as GlcUA-GlcNAc-UDP. For the elongation phase, the PmHS2 synthase elongates the heparosan disaccharide to its final structure [54]. At equimolar and low UDP-sugars concentrations (~250 $\mu$M), the initiation phase is limited, reducing the heparosan chains to be initiated. Conversely, at high concentrations of each precursor, longer heparosan chains with lower polydispersity (PDI~1) are synthesized [16].

*2.1.4 Limited resources* Heparosan properties such as polydispersity and chain length will depend on the availability of both precursors. On the one hand, UDP-GlcUA production (Fig. 1a purple panel) is a bottleneck, since D-Glucose 1-phosphate is a fundamental metabolite produced by *E. coli* as the sole source of carbon for its own growth. Moreover, UDP-GlcUA plays a fundamental role in the synthesis of glycoproteins, proteoglycans, and polysaccharides [26].

On the other hand, the conversion of D-Glucosamine 6-phosphate into D-Glucosamine 1-phosphate is a critical step prior to UDP-GlcNAc production [15]. It has been shown that overexpression of the GlmM enzyme results in increased heparosan production [32] (Fig. 1a pink panel). UDP-GlcNAc (and GlcNAc) is readily recycled and metabolized by glycolysis at the same time that it is also transformed by the enzyme UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA) into peptidoglycans for the outer cell wall formation [5; 57]. Therefore, this is a competitive biochemical reaction with the one of heparosan polymerization. The peptidoglycans rate limiting can be modulated by the MurA enzyme (Fig. 1a pink-dashed line) to enable the regulation of the UDP-GlcNAc supply.

Heterologous expression of enzymes from the heparosan biosynthesis pathway in *E. coli* Nissle 1917 is an alternative to increase production. Single overexpression of genes bsgalU3 *(B. subtilis 128)*, eckfiD and ecglmM *(E. coli MG1655)* can significantly increase heparosan production (from 0.15 g/L to 0.39 g/L) [32], while modifying the Mw and PDI. Figure 1a depicts how GlmM enzyme speeds up the production of UDP-GlcNAc, and enzymes GalU and KfiD catalyze UDP-GlcUA formation. For purified PmHS2, low concentrations of these UDP-sugars favoured the production of high Mw heparosan chains with low PDI, whereas high precursor concentrations had the opposite effect [14; 51]. Interestingly, studies have shown that over-expressing enzyme GlmU alone led to an unexpectedly reduced heparosan yield [32]. Thus, we need to over-express PmHS2 synthase, as well as enzymes GalU, KfiD, and GlmM. Not only will heparosan titer improve (up to 1.29 g/L in flask cultures), but also the ranges of the Mw obtained and polydispersity can be extended [39; 57].

## 2.2 Modelling the heparosan Mw and PDI

As for any other polymer, polydispersity is usually originated during the heparosan polymerization reaction. The normal slow-chain initiation step (section 2.1.3) produces heparosan with different sizes depending on the ratio between UDP-GlcNAc and UDP-GlcUA. This is far from what is required: a very narrow distribution approaching the ideal polydispersity value of a low PDI~ 1 [52; 22]. Having a PDI ~1 would mean that all the polymers have the same Mw, and hence the same chain length.

In the pharmaceutical industry, formulation of drug nanocarriers requires that the materials used have a uniform distribution, i.e. low PDI to avoid their accumulation in target tissues [21]. Experimentally measuring and characterizing the chain length of a produced polymer, such as heparosan, can be achieved through various analytical techniques. The most common one is Size Exclusion Chromatography. In this method, the polymer sample is passed through a column packed with porous beads. Smaller polymer chains penetrate deeper into the pores, resulting in longer elution times, while larger polymer chains elute more quickly. By comparing the elution times of the polymer sample with standards of known Mws, the Mw distribution of the sample can be determined.

Once the Mw distribution is obtained, the PDI can be calculated from the statistical moments of that distribution as $PDI = 1 + \mathrm{var}^2(\mathrm{Mw})/\mathrm{E}^2(\mathrm{Mw})$. Therefore, the higher the variance relative to the square of the mean, the larger the PDI. This indicates a broader Mw distribution within the polydisperse polymer sample. Conversely, a smaller variance results in a smaller PDI, indicating a narrower distribution (Fig. 1b).

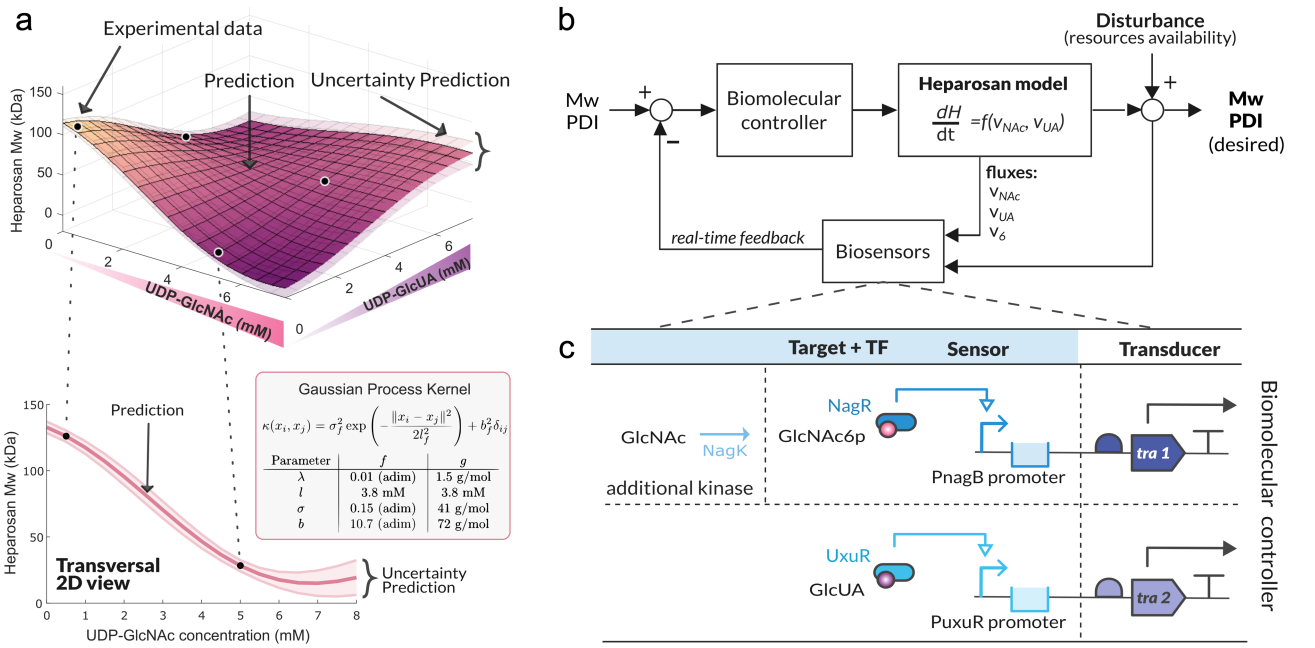



**Figure 2. a**. GP estimation of the Mw of heparosan and its relationship with the precursor's concentration (top). Transparent sheets on top and below the estimation are the 95% confidence intervals of the prediction. Black dots are experimental data from[16]. The kernel of the GP is in the shaded box, together with the parameters of the two estimations: Mw and PDI. The transversal 2D view is an example of the prediction of heparosan Mw for a fixed concentration of UDP-GlcUA (0.5 mM), while varying UDP-GlcNAc concentration (bottom). The central line is the Mw predicted value, and the shaded band is the uncertainty of that prediction. Transversal 2D view for the other precursor can be found in Supplementary section 4, Figure S9. **b.** Feedback control loop for dynamic regulation of heparosan synthesis with tailor-made characteristics. **c.** The biosensor sets are designed to detect GlcNAc6P and GlcUA. GlcNAc6p is used as a proxy for UDP-GlcNAc, and GlcUA approximates UDP-GlcUA. The activity of each biosensor is repressed (OFF) when the TF is bound to the promoter (blue). The biosensor output is activated (ON) when the target metabolite (GlcNAc6P or GlcUA) binds to the TF, releasing the binding region for the polymerase to initiate transcription of the downstream genes. These are the so-called transducers to display an output signal or drive a cellular change. *tra*1 is the transducer of the UDP-GlcNAc biosensor, and *tra*2 corresponds to the UDP-GlcUA biosensor.

PmHS2 can be utilized *in-vitro* to synthesize quasi-monodisperse (i.e. very narrow size distributions approaching the ideal PDI= 1) polymer preparations [15] with homogenous reactive end groups for coupling to biologic targets. The narrow size distribution is achieved by synchronizing the polymerization reaction using a primer, a short heparosan fragment, which allows the normal slow-chain initiation step of biosynthesis to be bypassed. Therefore, all polymers are rapidly extended by PmHS2 in a virtually parallel fashion, where all final chains have a very similar length.

However to achieve the desired Mw and PDI in the *in-vivo* production of heparosan, it is crucial to have a precise control over the precursor concentrations [15]. It has been shown experimentally that maintaining a lower precursor concentration (either UDP-GlcNAc or UDP-GlcUA) promotes the formation of high Mw heparosan with a low PDI. However, increasing precursor concentrations in a specific way can promote the formation of lower Mw with varying PDI (usually >> 1) [16]. Thus, achieving the appropriate balance to obtain low-Mw heparosan with the lowest possible PDI is a great challenge.

*2.2.1 Gaussian processes* To establish a relationship between precursor concentrations and their impact on both the PDI and Mw of heparosan, we used experimental data from [16]. We leverage a GP as part of our methodology, akin to the principles of ordinary kriging, to estimate and understand the underlying trends and uncertainties in these data. A GP is a machine learning tool employed for modelling and predicting relationships between variables (Fig. 2a). It is particularly useful when dealing with intricate or uncertain patterns in data [59]. However, while primarily based on data, GPs also allow the inclusion of parametric components, incorporating classical regression techniques [60].

The formalization of any 'random function' with input $X = [x_0, x_1, \ldots, x_N]$ and output $Y = [y(x_0), y(x_1) \ldots, y(x_N)]$ where the relationship between $X$ and $Y$ is given by the function $y(x) = f(x) + w$, $w = \mathcal{N}(0, \lambda^2)$, where $\lambda$ is the standard deviation of the noise measurement that we assume to be associated with the observations of $f$. It is commonly done through a GP, characterized by a mean $\mu(x)$ and a covariance generator or kernel ($\kappa(x_i, x_j)$) as seen in [59]. In the particular case of *ordinary krigging*, the process has zero mean ($\mu(x) = 0$ ), and as a parametric part, it only has a bias term $b$. The GP has an exponential kernel characterized by the hyperparameters $l$ (length scale) and $\sigma_f$ (signal variance). In this way, the GP can be used such as for any set of input points $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$, the joint distribution of the function values at these points $f(\mathbf{X})$ follows a multivariate Gaussian distribution with mean $\mathbf{0}$ and a covariance matrix given by the exponential kernel function:

$$\kappa(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{\|x_i - x_j\|^2}{2l_f^2}\right) + b_f^2 \delta_{ij}, \quad (1)$$

where $\delta_{ij}$ is the Kronecker delta. The GP is then fully characterized by its mean function $m_f(\mathbf{X}) = \mathbf{0}$, the covariance matrix $\kappa_f$ defined by the kernel function, and the bias $b_f$.

In our case, we used one GP ($f$) to model the relationship between the concentrations of precursors UDP-GlcUA, UDP-GlcNAc (the $X$ of the GP) and the PDI (Y), and a second GP ($g$) to model the relationship between the concentrations of these precursors and the Mw. Figure 2a-top shows the GP ($g$), and the parameters associated with both GPs are listed in the shaded

box. Figure 2a-bottom depicts a transversal cut of the GP computed for the Mw with a fixed concentration of UDP-GlcUA= 0.5 mM. As we can see, the predicted Mw has low uncertainty in the estimation for any UDP-GlcNAc concentration. On the one hand, when the concentration of UDP-GlcNAc gets closer to the available experimental value, we obtain a better estimation (with a smaller uncertainty). On the other hand, when the value of the concentration gets further away from the experimental data point, the estimation uncertainty increases. Transversal 2D view for the other precursor can be found in Supplementary section 4, Figure S9, and further details in Supplementary section 4.

## 2.3 Optimization process

For optimizing heparosan production, we used the framework where a single-objective problem (SOP) can be expressed as follows:

$$\min_{\theta} J(\theta) = J(\theta^*) \tag{2}$$

subject to:

$$\begin{cases} \mathbf{K}(\theta) \leq 0 \\ \underline{\theta_i} \leq \theta_i \leq \overline{\theta_i}, \ \forall i = [1, \ldots, n] \end{cases} \tag{3}$$

where $J(\theta)$ represents the objective function, $\theta = [\theta_1, \theta_2, \ldots, \theta_n]$ denotes the decision vector containing the decision variables for single-objective optimization, and $\mathbf{K}(\theta)$ denotes the inequality and equality constraint vectors. The bounds $\underline{\theta_i}$ and $\overline{\theta_i}$ represent the lower and upper bounds, respectively, within the decision variable space $\theta$. The SOP (2) seeks to find a single optimal solution $\theta^*$ that minimises the objective function $J(\theta)$.

## 2.4 Biosensors design

Figure 2b illustrates the role of genetic biosensors as powerful tools for real-time monitoring of natural product biosynthesis in microbial cell factories [34]. However, there are still several challenges underlying the design and application of genetic biosensors, i.e. characterising the dynamic range, sensitivity, and substrate specificity, and developing multigene-controlling genetic biosensors [30].

A biosensor usually consists of a *i)* sensor part, which detects an input signal, and *ii)* a transducer part to generate a measurable output signal (Fig. 2c). Specifically, the sensor part can include interactions among the promoter (DNA binding region), proteins (transcription factors, enzymes, etc.), and the target compound (ligand) we want to detect. This is known as *biosensor set*. The transducer translates these interactions into an output signal (e.g. fluorescence, colourimetry, and resistance). We used two biosensor sets of UDP-GlcNAc and UDP-GlcUA to provide feedback on the status of precursors during heparosan synthesis, while a biomolecular controller regulates the heparosan Mw and PDI (Fig. 2b).

In the biosensor design, sometimes there is no possibility of sensing the metabolite of interest directly, i.e. with a transcription factor (TF) that binds to the desired metabolite. One valid approach is the use of extended biosensors to measure the desired compound through an intermediate molecule, derived from the first one [12]. This is useful to avoid crosstalk and improve the specificity of the biosensor device. We used Sensipath [36], which is web-server assisting tool to design Sensing-Enabling Metabolic Pathways (SEMPs). A SEMP creates a link between a nondetectable target compound and another one known through enzymatic reactions. The result is a biosensor that can detect the presence/quantity of the target compound, while it is transferred to another reliable compound easy to detect by a biosensor. When the transcriptional activity of the promoter is mediated directly by a TF, the biosensor is known as a TF-based biosensor [30].

We selected one TF-based extended biosensor for the precursor UDP-GlcNAc, and one TF-based biosensor for UDP-GlcUA based on: *i)* target compounds should not be ubiquitous in the cell to reduce background noise (crosstalk) and increase the orthogonality of biosensors or other engineered genetic devices, *ii)* the enzymatic products in the detection pathway (including the SEMPs) must not be too toxic to the cell, *iii)* each biosensor is specific to its detectable target compound, *iv)* the promoter transcriptional activity is the response to one TF, and the promoter sequence has not been found in other databases, and *v)* each biosensor avoids reactions with enzymes capable of processing several structurally similar substrates (enzyme promiscuity).

*2.4.1 Biosensor of UDP-GlcNAc* We identified several detectable metabolites from the SEMP analysis of UDP-GlcNAc (Supplementary section 6.1). N-acetylglucosamine-6-phosphate (GlcNAc6P) was selected as one of the metabolites produced after the N-acetylglucosamine phosphorylation from UDP-GlcNAc branch in Fig. 1a. This is one of the faster reactions in biological systems. Therefore, we can rapidly convert GlcNAc into GlcNAc6P to use it as a proxy for sensing UDP-GlcNAc concentration as they have the same dynamics [7]. Our genetic device produces GlcNAc6P from the interactions between the N-acetyl-D-glucosamine kinase *NagK* (native from *Plesiomonas shigelloides*) and GlcNAc as a substrate (Fig. 3). The resulting metabolite will slightly change the amount of UDP-GlcNAc available in the cell. Therefore, to minimise the use of GlcNAc6P and its flux $\nu_6^q$ as much as possible, we quantified its effect with the model (5) as a rate equation (7), in every $q-th$ cell.

Additionally, the SEMP results showed this is a TF-based extended biosensor that also requires a TF, for which there were two options: *NagC* or *NagR*. In the native GlcNAc catabolic pathway, first *NagR* or *NagC* inhibits the activity of promoter *PnagB*, and then when there is enough concentration of GlcNAc6P, the TF dissociates from the *PnagB* promoter region allowing for the transcription of the GlcNAc catabolic genes [45]. We used RegulonDB, a database of *E. coli* K-12 regulatory network, to confirm how *NagR* and *NagC* bind to the promoter *PnagB*, and how it is released upon interaction with GlcNAc6P. We also analysed the interactions with *PnagB* as a TF regulon, by using the genomic-scale RegPrecise database (http://regprecise.lbl.gov).

Furthermore, there is experimental evidence from the group of one of the authors, where they obtained a high-precision crystallographic position and orientation of *NagR* protein bound to both the ligand GlcNAc6P, and the DNA binding region of *PnagR* promoter [24]. Therefore, we considered *NagR* as the TF, and the promoter region of *PnagB* as the transducer of the UDP-GlcNAc biosensor. When *NagR* is bound to promoter *PnagB*, the expression of the downstream genes is inhibited. De-repression of *PnagB* occurs when a sufficient amount of GlcNAc6P binds to NagR and the PnagB promoter. In that way, the amount of GlcNAc6P corresponding to the amount of GlcNAc produced in a cell can be measured (Fig. 2c).

*2.4.2 Biosensor of UDP-GlcUA* In the case of UDP-GlcUA, not only did we find few candidates to detect derivate compounds from the SEMP (Supplementary section 6.2), but we also detected some direct GlcUA TF-based biosensors. As we can see in Fig. 1a, UDP-GlcUA can be directly measured as trace levels of glucuronic acid (GlcUA) for two reasons: *i)* it is an intermediate in the UDP-GlcUA degradation pathway, and *ii)* this compound is not involved

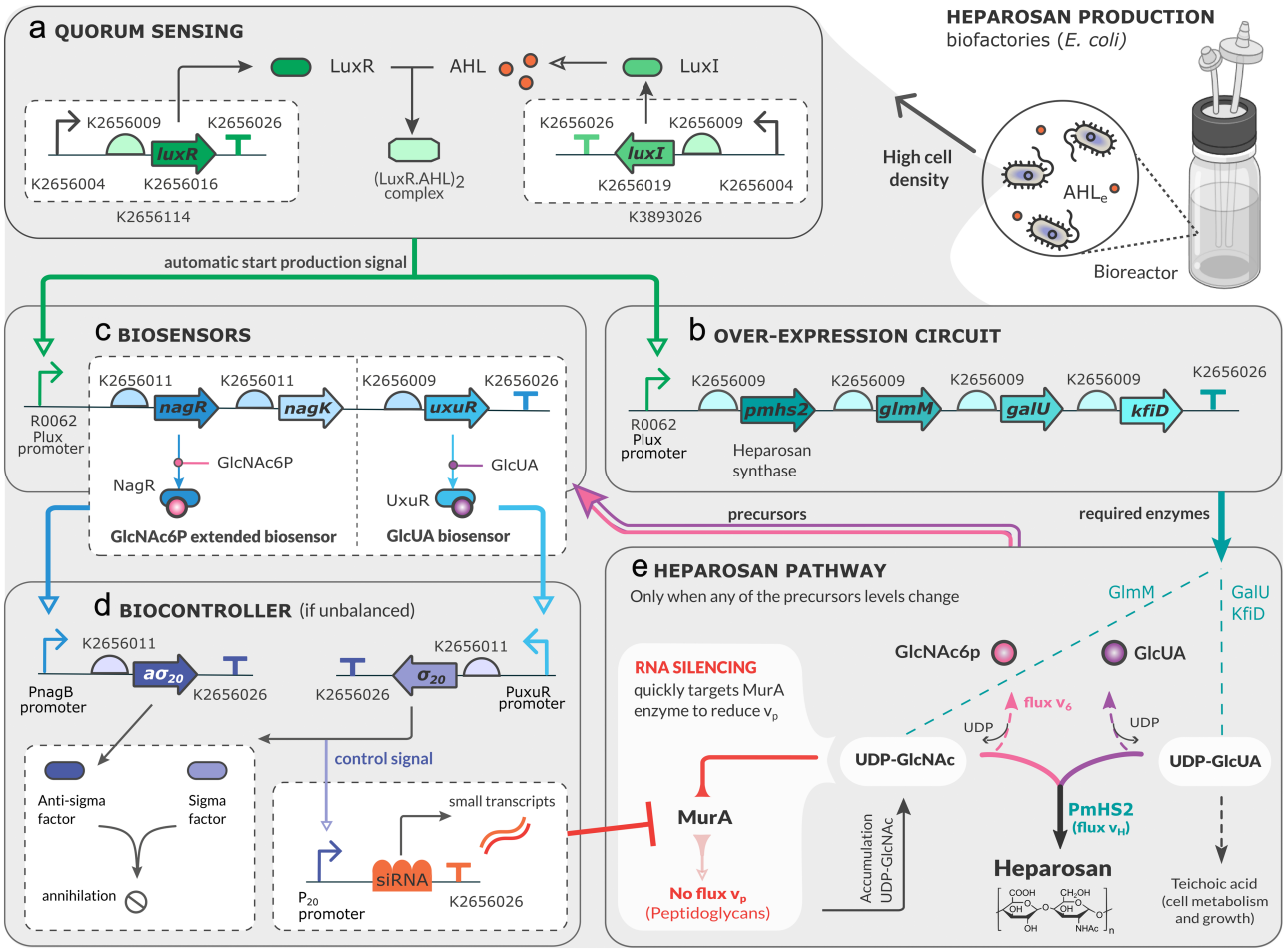



**Figure 3.** Schematic representation of the genetic circuits involved in the regulation of heparosan production. **a.** The QS circuit is responsible for initiating the production process. **b.** depicts the overexpression circuit, which produces the necessary enzymes for heparosan biosynthesis upon activation. **c.** illustrates the biosensors that measure the levels of precursors in the heparosan pathway. These signals are used by the biocontroller **d**, which generates siRNA to silence the MurA enzyme, reducing utilization of UDP-GlcNAc heparosan precursor to restore the balance between the UDP-sugars. **e.** shows the heparosan biosynthesis pathway, demonstrating how siRNA-mediated regulation affects precursor utilization. Together, these circuits enable precise control over heparosan production in response to cellular signals.

in the central metabolism of *E. coli* Nissle, proposed as a chassis for our circuit. This was verified by performing a comparative analysis of genome regulation and metabolism of the *E. coli* Nissle chassis in the EcoCyc (https://ecocyc.org/) database. Interestingly, the GlcUA biosensor senses it as a result of the hydrolysis reaction of this ligand [8]. Hence, not only is this reaction a naturally occurring process in *E. coli*, but it is also a proxy of UDP-GlcUA heparosan precursor (Fig. 1a).

Similarly to the GlcNAc6P biosensor, we identified potential TFs for the UDP-GlcUA biosensor set using Sensipath, RegPrecise, and RegulonDB databases. Protein *UxuR* is the TF that binds to the *PuxuR* promoter and repressing its activity (Fig. 2c) until the ligand is bound. This biosensor works in the same way as the UDP-GlcNAc biosensor set. *UxuR* inhibits the expression of genes necessary for gluconate transport and glucuronoside catabolic processes. Then, GlcUA prevents binding of *UxuR* with the operator, so promoter *PuxuR* is de-repressed and generating the biosensor output. The UDP-GlcUA biosensor response was also verified by using the RegPrecise database.

Furthermore, *UxuR* (belonging to the GntR family of bacterial repressors) has a peculiar $Zn^{2+}$-binding motif, consisting of four metal-coordinating amino acids (three histidines and one aspartate) and one auxiliary arginine [2]. The free coordination sites of the central $Zn^{2+}$ provide an anchor point for the effector molecule's carboxylate and neighbouring hydroxyl groups, while the former is additionally held in place by the auxiliary arginine via a salt bridge. This intricate coordination arrangement allows *UxuR* to specifically recognize glucuronic acid and to discriminate against nonoxidized sugar molecules, such as GlcNAc or GlcNAc6P, reducing the possibility of cross-talk. We also monitor in real time the UDP-GlcUA while quantifying its effect in the heparosan synthesis model (5).

## 2.5 Dynamic model of heparosan synthesis

Mathematical dynamic modelling of gene synthetic circuits can be based on the knowledge of biochemical interactions, protein mechanisms, and experimental data. We focused on the first stage of the DBTL cycle. We designed and characterised a model based on ordinary differential equations (ODEs) to describe changes in the heparosan production over time. The model also can predict the temporal evolution of both UDP-GlcNAc and UDP-GlcUA, which eventually determine the structural properties and quality of the produced heparosan.

For a metabolite $M_i$, the changes in its concentration satisfy the equation $\frac{d[M_i]}{dt} = \sum_j S_{ij}\nu_j$, where $S_{ij}$ is the stoichiometric coefficient of metabolite $M_i$ in reaction $j$ and $\nu_j$ is the flux of reaction $j$. If we are

producing heparosan in a living microorganism, the dilution effect due to cell growth reduces the concentration of the metabolite $M_i$. Now, the previous equation is slightly different:

$$\frac{d[M_i]^q}{dt} = \sum_j S_{ij}\nu_j - \mu[M_i] \tag{4}$$

where $\mu$ is the specific growth rate of the $q-th$ cell. Considering eq. (4), the changes in the heparosan concentration (H) depend on both precursors UDP-GlcUA (UA) and UDP-GlcNAc (NAc), and the metabolic fluxes $\nu_j^q$ resulting from their interactions with other cell pathways. Therefore, they are limiting resources that can be represented as the dynamic model (5).

$$\begin{cases} \frac{d[H]^q}{dt} & = \nu_H^q - \mu[H] \\ \frac{d[UA]^q}{dt} & = \nu_{UA}^0 - \nu_H^q - \mu[UA] \\ \frac{d[NAc]^q}{dt} & = \nu_{NAc}^0 - \nu_H^q - \nu_6^q - \nu_P^q - \mu[NAc] \end{cases} \tag{5}$$

where $\nu_{UA}^0$ and $\nu_{NAc}^0$ are the estimated initial fluxes of UDP-GlcUA and UDP-GlcNAc in the $q-th$ cell, respectively; $\nu_H^q$ is the flux coming from the heparosan synthesis reaction, $\nu_6^q$ is the flux from GlcNAc6P phosphorylation (sections 2.4, and 3.2), $\nu_P^q$ is the peptidoglycans flux going to the cell wall formation (section 2.1.4), and $\mu$ is the specific growth rate of the cell (we assumed a doubling time of 60 min in a typical bioreactor).

We consider $\nu_{UA}$ and $\nu_{NAc}$ as limiting fluxes for catalyzing the downstream reactions, including heparosan biosynthesis. However, the initial fluxes $\nu_{UA}^0$ and $\nu_{NA}^0$ were estimated through flux balance analysis (FBA) of the *E. coli* Nissle 1917 genome-scale metabolism [35] network using COBRA Toolbox [43; 27]. Further details are in Supplementary section 5, and the FBA-COBRA simulation code is our Github repository supp_data. The remaining $\nu_j$ fluxes in model (5) obey Michaelis–Menten kinetics [25]:

$$\nu_H^q = \frac{\text{cat}_P\,[PmHS2][UA][NAc]}{\frac{1}{\text{Km}_U\text{Km}_N} + \frac{[UA]}{\text{Km}_N} + \frac{[NAc]}{\text{Km}_U} + [UA][NAc]} \tag{6}$$

$$\nu_6^q = \frac{\text{cat}_K[NagK][NAc]}{\text{Km}_6 + [NAc]} \tag{7}$$

$$\nu_P^q = \frac{\text{cat}_M[MurA][NAc]}{\text{Km}_N + [NAc]} \tag{8}$$

where $\text{cat}_P$, $\text{cat}_K$, and $\text{cat}_M$ are the catalytic rates of enzymes *PmHS2*, *NagK* and *MurA*, respectively, $\text{Km}_U$ and $\text{Km}_N$ are the Michaelis–Menten constants of both UDP-GlcUA and UDP-GlcNAc, respectively; and $\text{Km}_6$ is the Michaelis–Menten constant of *NagK* and N-acetylglucosamine. As *PmHS2* catalyses the heparosan synthesis reaction, and the enzymes *NagK* and *MurA* are involved in the heparosan pathway, their concentration levels need to be regulated for optimal production. These dynamics are described in section 3.1.2 for *PmHS2* synthase, section 2.4 for *NagK*, and 3.3 for *MurA* enzyme. Table 1 shows the enzyme kinetic parameters of the fluxes (6-8) obtained from BRENDA [50].

## 3 Results

We focus on *i)* increasing heparosan production of a desired Mw, and *ii)* reducing the PDI to optimize its homogeneity and quality by balancing the precursor supplies UDP-GlcNAc and UDP-GlcUA. Figure 3 illustrates the synthetic device implemented following the design guidelines of the dynamic heparosan production model.

**Table 1.** Parameter values of the metabolic pathway

| | Description | Value |
|---|---|---|
| $\mu$ | specific growth rate (*E. coli*) | 0.01 $\text{min}^{-1}$ |
| $\nu_{UA}^0$ | Initial UDP-GlcUA flux | 9e3 molec/min |
| $\nu_{NAc}^0$ | Initial UDP-GlcNAc flux | 4.1e4 molec/min |
| $\text{cat}_P$ | PmHS2 catalytic constant | 240 $\text{min}^{-1}$ |
| $\text{cat}_K$ | NagK catalytic constant | 0.8 $\text{min}^{-1}$ |
| $\text{cat}_M$ | MurA catalytic constant | 228 $\text{min}^{-1}$ |
| $\text{Km}_U$ | Michaelis constant PmHS2-UA | 6e6 molec |
| $\text{Km}_N$ | Michaelis constant PmHS2-NAc | 1.5e5 molec |
| $\text{Km}_6$ | Michaelis constant NagK-NAc | 1.5e5 molec |

### 3.1 Automatic initiation of heparosan production

Usually, a genetic circuit is triggered by an external inducer. However, this implies a thorough understanding of the biological process, and full characterization of its biological parts is not always possible. We aim to automate the production of heparosan without any external induction. Quorum sensing (QS) is a cell-to-cell communication mechanism in which bacteria exchange chemical signalling molecules, called autoinducers, to generate a jointly coordinated response to a stimulus [11]. We used QS and this cell's ability to synchronise the operation of our four genetic circuits to produce heparosan.

*3.1.1 Quorum sensing-based circuit* Figure 3a shows the first genetic circuit based on QS [9], where the Lux system (native from *Vibrio fischeri* [11]) produces both LuxI homoserine lactone synthase, and the LuxR TF, respectively. In turn, LuxI synthesises N-acyl homoserine lactone molecules (AHL) as autoinducer signals across the population. The dynamic model (9) captures this cell-to-cell communication among every $q-th$ cell in a population of $N$ cells. At low cell densities, intracellular AHL molecules diffuse passively out of the cell down a concentration gradient across its membrane, according to Fick's law [23]. Then, the extracellular AHLe level increases in the culture medium.

$$\begin{aligned} \frac{dN(t)}{dt} &= \mu N(t)\left(1 - \frac{N(t)}{N_{max}}\right) \\ \frac{d[AHLe]}{dt} &= -DV[AHLe]N(t) + D[AHL]N(t) - d_{Ae}[AHLe] \end{aligned} \tag{9}$$

where $\mu$ is the specific cell growth rate, $N_{max}$ is the maximum number of cells in the population, D is the diffusion rate of AHL molecules across the cell membrane, $V = V_{cell}/V_{bioreactor}$ is the ratio between the cell volume and the bioreactor volume, and $d_{Ae}$ is the AHLe degradation rate. Model (9) was obtained from the biochemical reactions involving the biochemical species LuxR, LuxI, AHL, and AHLe (Supplementary section 1.1). The model parameters values are listed in Table 2.

In contrast, at high cell densities AHL is accumulated and the intracellular levels are equivalent to the extracellular ones (AHLe). The higher the biomass $N$, the larger the AHL level and the faster QS starts. The extracellular environment interplays with every $q-th$ cell through the expressions in (10). With this dynamic model we can quantify the required conditions of LuxR, LuxI, and AHL to successfully activate QS, only when we have enough biomass. We carried out several *in-silico* experiments considering 48 hours for heparosan fermentation period, in a 7L bioreactor. [32; 15]. Figure 4a depicts that the optimal biomass of OD=50 will automatically trigger heparosan biosynthesis 10 hours after the fermentation has started.

**Table 2. Model parameters of the genetic circuits.** Summary of parameters used in the genetic circuit models, including QS, overexpression, biosensor, and biocontroller circuits. Parameters are categorized by function, with values obtained from experimental data, estimates, or literature references. Parameters include rates of transcription, translation, degradation, diffusion, dissociation constants, and cell volume, providing a comprehensive framework for the model's predictive behaviour and performance under simulated conditions.

| Parameter | Description | Value | Reference |
|---|---|---|---|
| **QS circuit** | | | |
| $c_n$ | Plasmid copy number | 10 copies | experimental |
| $N_{max}$ | maximum cells capacity | 120 OD | estimated |
| $k_1$ | *luxR* transcription rate | 1 $min^{-1}$ | [12] |
| $p_1$ | $mRNA_{LuxR}$ translation rate | 0.05 $min^{-1}$ | estimated |
| $dm_{1,\dots,11}$ | mRNA degradation rate | 0.23 $min^{-1}$ | [10] |
| $d_{1,\dots,9}$ | protein degradation rate | 0.002 $min^{-1}$ | [10] |
| $k_2$ | *luxI* transcription rate | 2.5 $min^{-1}$ | [12] |
| $p_2$ | $mRNA_{LuxI}$ translation rate | 0.075 $min^{-1}$ | estimated |
| $k_A$ | AHL synthesis rate | 0.04 $min^{-1}$ | [11] |
| D | AHL diffusion rate across cell membrane | 2 $min^{-1}$ | [11] |
| $d_A$ | Intracellular degradation rate | $4 \times 10^{-4}$ $min^{-1}$ | [40] |
| $d_{Ae}$ | Extra cellular degradation rate | $5 \times 10^{-5}$ $min^{-1}$ | [11] |
| $V_{cell}$ | *E. coli* volume | $1.1 \times 10^{-15}$ L | [40] |
| $V_{ext}$ | Culture medium volume | 7 L | experimental |
| **Overexpression circuit** | | | |
| $\alpha$ | $P_{lux}$ basal expression | 0.05 | estimated |
| n | Cooperative Hill coefficient | 2 | [11] |
| $k_{lux}$ | dissociation constant to $P_{lux}$ | 300 molec | [12] |
| kd | dissociation constant AHL to LuxR | 14 molec | estimated |
| $k_{3,\dots,6}$ | *pmhs2, glmM, galU, kfiD* transcription rate | 0.4 $min^{-1}$ | estimated |
| $p_3, p_4$ | $mRNA_{pmhs2}$, $mRNA_{glmM}$ translation rate | 0.02 $min^{-1}$ | estimated |
| $p_5$ | $mRNA_{galU}$ translation rate | 0.05 $min^{-1}$ | estimated |
| $p_6$ | $mRNA_{kfiD}$ translation rate | 0.04 $min^{-1}$ | estimated |
| **Biosensors circuit** | | | |
| $k_{7,8}$ | *uxuR, nagR* transcription rate | 0.6 $min^{-1}$ | [10] |
| $k_9$ | *nagK* transcription rate | 0.1 $min^{-1}$ | estimated |
| $p_7$ | $mRNA_{uxuR}$ translation rate | 0.06 $min^{-1}$ | estimated |
| $p_8$ | $mRNA_{nagR}$ translation rate | 0.03 $min^{-1}$ | estimated |
| $p_9$ | $mRNA_{nagK}$ translation rate | 0.08 $min^{-1}$ | estimated |
| **Biocontroller circuit** | | | |
| $k_{uxu}$ | dissociation constant to $P_{uxuR}$ | $4.3 \times 10^5$ molec | estimated |
| $k_{nag}$ | dissociation constant to $P_{nagB}$ | 1 molec | estimated |
| $kd_{UA}$ | dissociation constant GlcUA to UxuR | 2400 molec | estimated |
| $kd_{NAc}$ | dissociation constant GlcNAc6P to nagR | 1200 molec | estimated |
| $\beta_N, \beta_U$ | $P_{uxuR}$, $P_{nagB}$ leakage | 0.01 | [11] |
| $k_{10}$ | $\sigma_{20}$ transcription rate | 5.3 $min^{-1}$ | estimated |
| $k_{11}$ | $a\sigma_{20}$ transcription rate | 3 $min^{-1}$ | estimated |
| $k_{-c}$ | $(\sigma \cdot a\sigma)$ complex unbinding rate | $1.8 \times 10^{-3}$ $min^{-1}$ | [3] |
| $kd_c$ | $(\sigma \cdot a\sigma)$ complex dissociation constant | 0.01 molec | [3] |
| $p_{10}$ | $mRNA_{\sigma 20}$ translation rate | 0.9 $min^{-1}$ | estimated |
| $p_{11}$ | $mRNA_{a\sigma 20}$ translation rate | 0.18 $min^{-1}$ | estimated |
| $d_{10,11}$ | $\sigma$ and $a\sigma$ degradation rate | $3 \times 10^{-4}$ $min^{-1}$ | [10] |
| $d_{12}$ | $(\sigma \cdot a\sigma)$ complex degradation rate | 0.001 $min^{-1}$ | [3] |
| $k_{mT}$ | siRNA binding rate to target mMurA | $1.3 \times 10^{-4}$ $molec^{-1}$ $min^{-1}$ | [19] |

$$\frac{d[LuxR]^q}{dt} = \frac{p_1 k_1 c_n}{dm_1 + \mu} - (d_1 + \mu)[LuxR]$$
$$\frac{d[LuxI]^q}{dt} = \frac{p_2 k_2 c_n}{dm_2 + \mu} - (d_2 + \mu)[LuxI]$$
$$\frac{d[AHL]^q}{dt} = DV[AHLe] - D[AHL] + k_A[LuxI] - (d_A + \mu)[AHL] \quad (10)$$

where $p_i$ is the translation rate and $k_i$ is the transcription rate of the $i = [1, \dots, 3]$ biochemical species, $c_n$ is the plasmid copy number; $d_i$ is the protein degradation rate, and $dm_i$ is the mRNA degradation rate of the $i$ biochemical species; $d_A$ is the AHL degradation rate.

Only after the 'start production' signal at 10 hours, the overexpression circuit (Fig. 3b) begins the production of UDP-sugars, together with the necessary resources for heparosan biosynthesis (section 2.1). This signal comes from the binding of complex $(LuxR{\cdot}AHL)_2$ to promoter $P_{lux}$ for activating downstream gene expression (Fig. 3a, b and c). Equation (11) is a Hill-like function, where the $P_{lux}$ nonlinear dynamics depend on LuxR and AHL levels [11; 10]. The computational results (Supplementary section 1.1, Figure S1) also demonstrated that to obtain the start signal at 10 hours (when the culture is in exponential phase and around one-third of the final density), we need to produce more

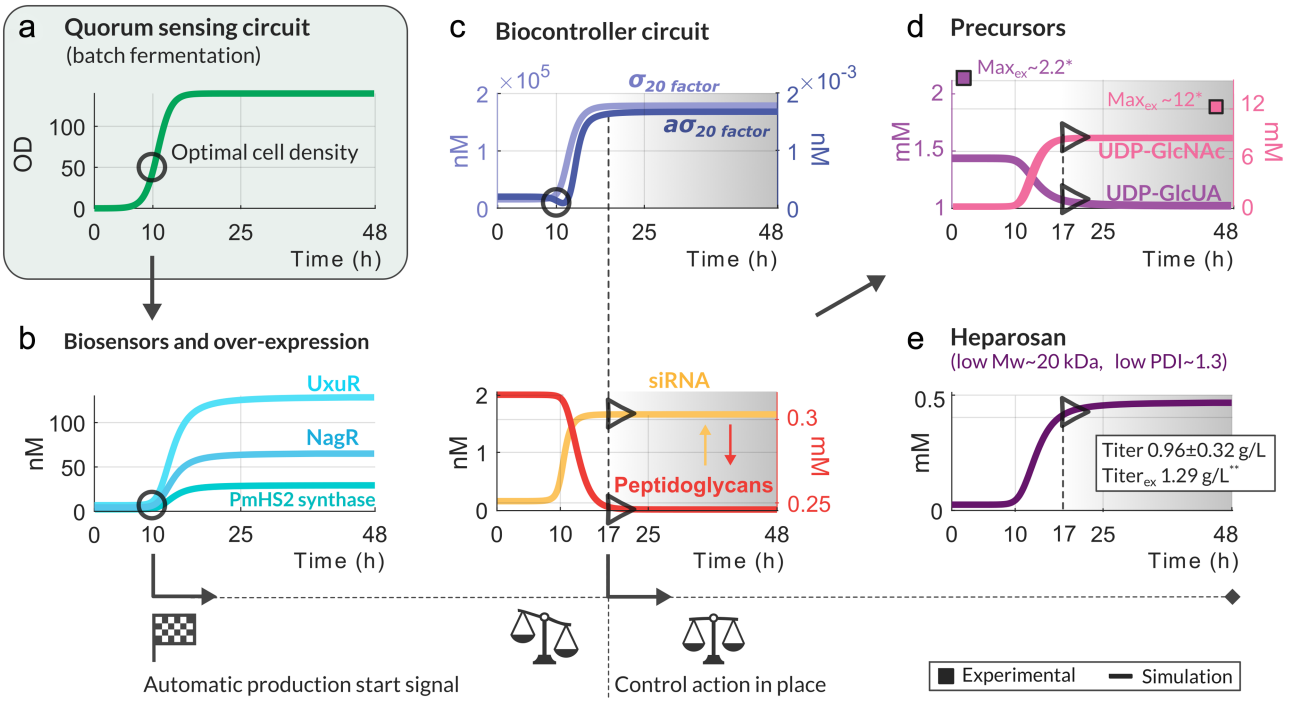



**Figure 4.** Dynamic behaviour of the synthetic device. **a.** Evolution of culture optical density (OD), indicating the initiation of production triggered by the QS circuit at 10 hours. **b.** Dynamics of biosensors and overexpression devices (UxuR, NagR, and PmHS2), showing increased expression postproduction initiation. **c.** Signals from the biocontroller components ($\sigma_{20}$, $a\sigma_{20}$, siRNA transcripts, and peptidoglycans) are fully operating at 17 hours of production (triangle). The biocontroller adjusts siRNA levels and reduces peptidoglycans production to maintain a specified balance between precursors, and to optimize heparosan synthesis. **d-e.** Levels of heparosan precursors. As heparosan production rises, UDP-GlcNAc levels increase, while UDP-GlcUA levels decrease after the biocontroller activation (triangle). The initial unbalance of the precursors concentrations is corrected to produce heparosan with low Mw, and low PDI. Experimental data (purple and pink squares) from *[44], **[31].

*LuxI* (~130 nM) than the LuxR TF (88 nM). We selected the biological parts for the QS-based circuit to specifically meet this behaviour. The DNA sequences corresponding to these parts are detailed in Supplementary section 7.

$$P_{lux} = \alpha + \frac{(1-\alpha)[AHL]^n}{k_{lux}\left(\frac{k_d c_n}{[LuxR]}\right)^n + [AHL]^n} \tag{11}$$

where $\alpha$ is the basal expression of $P_{lux}$ promoter, $n$ is the Hill coefficient (number of activators that need to bind to the promoter and starting the gene expression), $k_{lux}$ is the dissociation constant of $(LuxR{\cdot}AHL)_2$ to the promoter, and $k_d$ is the dissociation constant of LuxR to AHL.

### 3.1.2 *Overexpression of the necessary resources*

As in section 2.1.4, insufficiency of synthase PmHS2, or enzymes GlmM, GalU, and KfiD can directly constrain the synthesis of heparosan and modify its Mw and PDI. The overexpression gene circuit (Fig. 3b) is the second gene circuit built with the biological parts suggested by the results from the dynamic model (12). We designed this dynamic model to predict the expression levels needed for the PmHS2 synthase, and the enzymes GalU, KfiD, and GlmM, in every $q-th$ cell. The biochemical reactions happening in this circuit are described in Supplementary section 1.2. The biological parts of this circuit are listed in Supplementary section 7.

$$\begin{aligned}
\frac{d[PmHS2]^q}{dt} &= \frac{p_3 k_3 c_n}{dm_3+\mu} P_{lux} - (d_3+\mu)[PmHS2] \\
\frac{d[GalU]^q}{dt} &= \frac{p_4 k_4 c_n}{dm_4+\mu} P_{lux} - (d_4+\mu)[GalU] \\
\frac{d[KfiD]^q}{dt} &= \frac{p_5 k_5 c_n}{dm_5+\mu} P_{lux} - (d_5+\mu)[KfiD] \\
\frac{d[GlmM]^q}{dt} &= \frac{p_6 k_6 c_n}{dm_6+\mu} P_{lux} - (d_6+\mu)[GlmM]
\end{aligned} \tag{12}$$

where $p_i$ is the translation rate and $k_i$ is the transcription rate of the $i = [3,\ldots,6]$ biochemical species; $d_i$ is the protein degradation rate and $dm_i$ is the mRNA degradation rate of the $i$ biochemical species, for every $q-th$ cell in the population.

Later on, 10 hours after fermentation has started and only when the cells have grown sufficiently (Fig. 4a), the QS circuit enables the overexpression of these enzymes (activation of promoter $P_{lux}$). The threshold to activate the overexpression circuit is computed as the Hill-like function (11). Then, GlmM catalyzes UDP-GlcNAc at the same time that GalU and KfiD produce UDP-GlcUA (Fig. 4d). PmHS2 synthase reaches a final concentration of ~ 30 nM after nearly 16 hours. This will lead to relatively high heparosan concentrations (~ 0.5 mM), and titers of $0.925 \pm 0.32$ g/L (Fig. 4e, and Supplementary section 3.3). These are comparable to those obtained experimentally on *E. coli* Nissle [18; 51], but with the additional property of narrow polydispersion (Fig. 4e). With these results, we also can compute how much heparosan flux $\nu_H^q$

is the cell synthesizing (eq. 6). The overexpression circuit has a dual function as it increases the concentration of the precursors, as well as dealing with the degradation of the heparosan pathway intermediates. The remaining temporal dynamics of enzymes GlmM, GalU, and KfiD are depicted in Supplementary section 1.2.

## 3.2 Biosensors circuit

After the start of heparosan production, we have to monitor the precursors UDP-GlcNAc and UDP-GlcUA during the whole process. We designed a third gene circuit containing two different biosensors for reporting on both UDP-sugars (Fig. 3c, and Supplementary section 6). Both biosensors activate the biocontroller when it is needed, in order to maintain the balance between these metabolites for proper regulation of the heparosan PDI (see section 2.4).

On one hand, the GlcNAc6P biosensor produces an output signal when it senses this ligand as trace concentrations of UDP-GlcNAc (Fig. 3c- *nagR*, *nagK* genes). This is an extended biosensor that also expresses the kinase NagK to convert UDP-GlcNAc into GlcNAc6P (see section 2.4). The mathematical model (13) characterizes this biosensor set. A full description of the biochemical reactions taking place in this circuit is described in Supplementary section 1.3.

The first two equations of (13) represent the *NagR* and *NagK* levels expressed after the activation of $P_{lux}$ promoter, according to the dynamics in eq. (11). Similarly to the overexpression gene circuit, only at high cell density level the GlcNAc6P biosensor is activated and begins to measure UDP-GlcNAc. Notice that the last equation of (13) is the Hill-like function of $P_{nagB}$ promoter. $P_{nagB}$ responds to changes in the concentration levels of GlcNAc6P (or NAc) bound to the TF *NagR*.

$$\begin{aligned}\frac{d[NagR]^q}{dt} &= \frac{p_7 k_7 c_n}{dm_7+\mu} P_{lux} - (d_7+\mu)[NagR]\\ \frac{d[NagK]^q}{dt} &= \frac{p_8 k_8 c_n}{dm_8+\mu} P_{lux} - (d_8+\mu)[NagK]\end{aligned} \tag{13}$$

$$P_{nagB} = \beta_N + \frac{(1-\beta_N)[NAc]^{n_N}}{k_{nag}\left(\frac{kd_{NAc}\,cn}{[nagR]}\right)^{n_N} + [NAc]^{n_N}}$$

where $p_i$ is the translation rate and $k_i$ is the transcription rate of the $i = [7, 8]$ biochemical species; $d_i$ is the protein degradation rate and $dm_i$ is the mRNA degradation rate of the $i$ biochemical species, $n_N$ is the Hill coefficient, $\beta_N$ is the $P_{nagB}$ promoter leakage, $k_{nag}$ is the dissociation constant of *nagR*-GlcNAc6P to $P_{nagB}$ promoter, and $kd_{NAc}$ is the dissociation constant of *nagR* to GlcNAc6P.

Figure 4b (blue-line colour) depicts the temporal dynamics of TF *NagR*. At t=10 hours, QS triggers NagR production up to ~60 nM. This is the concentration needed for activating GlcNAc6P biosensor. Additionally, the sequence encoding the kinase NagK was specifically designed to extract the least possible UDP-GlcNAc needed for the GlcNAc6P biosensor. We obtained that the required concentration of NagK is ~28.5 nM. This way, we ensure optimal UDP-GlcNAc detection with minimum use of resources for the metabolic flux $\nu_6^q$ (eq. 7) in every $q-th$ cell.

Similarly to the GlcNAc6P biosensor, the GlcUA biosensor begins to work after the 'start signal' triggered by the QS-based circuit. This means that, once the $P_{lux}$ promoter (eq. 11) is activated, the GlcUA biosensor is monitoring any change in the UDP-GlcUA level. Model (14) characterises the *UxuR* dynamics and the concentrations required to activate the $P_{uxuR}$ promoter, which efficiently responds to changes in GlcUA while producing heparosan.

$$\begin{aligned}\frac{d[UxuR]^q}{dt} &= \frac{p_9 k_9 c_n}{dm_9+\mu} P_{lux} - (d_9+\mu)[UxuR]\\ P_{uxuR} &= \beta_U + \frac{(1-\beta_U)[UA]^{n_U}}{k_{uxu}\left(\frac{kd_{UA}\,cn}{[UxuR]}\right)^{n_U} + [UA]^{n_U}}\end{aligned} \tag{14}$$

where $p_9$ is the translation rate and $k_9$ is the transcription rate of UxuR; $dm_9$ and $d_9$ are the degradation rates of $mRNA_{UxuR}$ and UxuR protein, respectively; $n_U$ is the Hill coefficient; $\beta_U$ is the $P_{uxuR}$ promoter leakage; $k_{uxu}$ is the dissociation constant of UxuR-GlcUA to $P_{uxuR}$ promoter; and $kd_{UA}$ is the dissociation constant of UxuR to GlcUA. Further details on the biochemical reactions are in Supplementary section 1.3.

Figure 4b (cyan-line) depicts how *UxuR* reaches up to ~120 nM to turn on this biosensor at t =10 hours, when heparosan synthesis has also started. Notice that the GlcUA biosensor uses twice TF than GlcNAc6P biosensor does (UxuR>NagR). Therefore for the biosensors circuit implementation, we selected a stronger RBS (larger translation rate $p_9$) for UxuR expression in comparison with the RBS of the NagR TF ($p_7$), as in Fig. 3c.

## 3.3 Biomolecular controller circuit

Maintaining a balance between UDP-GlcNAc and UDP-GlcUA is a key factor for optimal heparosan synthesis *in-vivo*. Figure 3d illustrates the fourth gene circuit implemented as a biomolecular controller circuit. This biocontroller combines *i*) extracytoplasmic function (ECF) sigma factors, $\sigma$ and $anti-\sigma$ [28; 48], with *ii*) RNA silencing [33; 53] of enzyme MurA for a fast coordination between UDP-GlcNAc and UDP-GlcUA supply. The biocontroller's DNA sequences and parts are detailed in Supplementary section 7.

*3.3.1 Antithetic-based system* The antithetic-based system depicted in Fig. 3d is made up of ECF $\sigma_{20}$ factor, the antifactor $a\sigma_{20}$, and the inducible promoter $Pecf_{20-992}$ ($P_{20}$) [48; 10]. In general, $\sigma$ is a cofactor of RNA polymerase, so it promotes gene expression of the downstream genes. When the factor $a\sigma$ is expressed, it rapidly sequesters $\sigma$ to form the antithetic complex ($\sigma \cdot a\sigma$) that reciprocally annihilates both factors. As a result, this annihilation decreases the available amount of $\sigma$ factor in the cell. This way, we can harness the coordination of transcriptional responses to variations in precursor levels by using ECF factors [55].

The biocontroller supervises in real-time the levels of both UDP-GlcNAc and UDP-GlcUA via the biosensors gene circuit. $P_{nagB}$ promoter senses the GlcNAc6P biosensor input, and generates $a\sigma_{20}$ factor as the output signal (Fig. 3d). Conversely, $P_{uxuR}$ promoter detects the GlcUA biosensor input and produces $\sigma_{20}$ factor as the output that triggers the RNA silencing mechanism as the control action (Fig. 3d). Model (15) characterises how the biosensor's output signals are translated to the biocontroller, and computes the required expression levels of both $\sigma_{20}$ and $a\sigma_{20}$. The silencing mechanism follows a dynamic characterized by the promoter $P_{20}$ (last eq. in 15).

The results are shown in Fig. 4c-d. On one hand, when the established balance between UDP-GlcNAc and UDP-GlcUA is broken in favour of UDP-GlcUA, $\sigma_{20}$ increases more than its anti-$\sigma_{20}$ factor to activate the RNA silencing after 17 hours of fermentation (Fig. 4c top-bottom). Consequently, translation of MurA enzyme is quickly prevented, and the peptidoglycans flux $\nu_P^q$ is decreased (Fig. 4c red line). This control action accumulates UDP-GlcNAc and restores the balance between both precursors (Fig. 4d). The low concentration of $a\sigma_{20}$ factor ($\sim < dispIDdisp13 > 2 \times 10^5$ nM) than its

antifactor, ensuring that the silencing will occur inside the cell.

$$\begin{aligned}\frac{d[\sigma]^q}{dt} &= \frac{p_{10}k_{10}c_n}{dm_{10}+\mu}P_{uxuR} - \frac{k_{-c}}{kd_c}[\sigma][a\sigma] + k_{-c}[C] - (d_{10}+\mu)[\sigma]\\ \frac{d[a\sigma]^q}{dt} &= \frac{p_{11}k_{11}c_n}{dm_{11}+\mu}P_{nagB} - \frac{k_{-c}}{kd_c}[\sigma][a\sigma] + k_{-c}[C] - (d_{11}+\mu)[a\sigma]\\ \frac{d[C]^q}{dt} &= \frac{k_{-c}}{kd_c}[\sigma][a\sigma] - k_{-c}[C] - (d_{12}+\mu)[C]\\ P_{20} &= \gamma + \frac{(1-\gamma)[\sigma]^{n_s}}{k_{20}(kd_s\,cn)^{n_s}+[\sigma]^{n_s}}\end{aligned} \quad (15)$$

where $p_i$ is the translation rate and $k_i$ is the transcription rate of the $i = [10, 11]$ biochemical species, $dm_i$ is the mRNA degradation rate and $d_i$ is the protein degradation rate of the $i = [10, \ldots, 12]$ biochemical species, $k_{-c}$ is the unbinding rate of ($\sigma$ to $a\sigma$), $kd_c$ is the dissociation constant of complex ($\sigma \cdot a\sigma$) $n_s$ is the Hill coefficient of $\sigma$ TF, $\gamma$ is the basal expression of $P_{20}$, $k_{20}$ is the dissociation constant of *nagR*-GlcNAc6P to $P_{nagB}$ promoter, and $kd_s$ is the dissociation constant among $\sigma_{20}$ molecules. The biochemical reactions happening in this circuit are shown in Supplementary section 1.4.

On the other hand, when the balance between both precursors is broken in favour of UPD-GlcNAc, expression of $a\sigma_{20}$ factor is activated by the GlcNAc6P biosensor. $a\sigma_{20}$ rises and starts to annihilate $\sigma_{20}$ factor (Fig. 4c purple-line). This also prevents MurA degradation while increasing the flux $\nu_p^q$ towards the cell wall formation. Thus, the resulting unbalance caused by the excess of UPD-GlcNAc is compensated by the biocontroller. $a\sigma_{20}$ triggers this compensation via $P_{nagB}$ (eq. 13). Again, annihilation is a fast dynamic process that can be characterized by complex C in model (15), ensuring a rapid control action.

*3.3.2 Silencing of the target enzyme MurA* RNA silencing is widespread in nature to reversibly repress translation through RNAs or antisense RNAs [53]. Apart from the RNA-binding proteins found in living organisms, we used synthetic biology tools to design small noncoding regulatory transcripts (siRNAs) that bind to a target mRNA, while inhibiting its subsequent translation and facilitating the degradation of the mRNA itself [29].

We designed the dynamic model (16) to measure the performance of this control action via RNA silencing of MurA. The first two ODEs of (16) represent the amount of siRNAs (*si*) that target the messenger RNA of enzyme MurA (*mMurA*). This translational repression and its effect on reducing/increasing the peptidoglycans flux $\nu_p^q$ is predicted by the availability level of MurA in the $q-th$ cell (last eq. in 16).

$$\begin{aligned}\frac{d[si]^q}{dt} &= k_{13}c_n P_{20} - km_T[si][mMurA] - (d_{13}+\mu)\\ \frac{d[mMurA]^q}{dt} &= k_{14}c_n - km_T[si][mMurA] - (dm_{14}+\mu)[mMurA]\\ \frac{d[MurA]^q}{dt} &= p_{14}[mMurA] - (d_{14}+\mu)\end{aligned} \quad (16)$$

where $p_i$ is the translation rate and $k_i$ is the transcription rate of the $i = [13, 14]$ biochemical species, $dm_i$ is the mRNA degradation rate and $d_i$ is the protein degradation rate of the $i = [13, 14]$ biochemical species, $km_T$ is the binding rate to target RNA *mMurA*. This model was designed from the biochemical reactions happening in Supplementary section 1.4.

Figure 4c bottom depicts how effectively the flux $\nu_p^q$ increases/reduces depending on the *siRNA* control action. Provided an unbalance between the precursors, e.g. in favour of UDP-GlcUA, the biocontroller transcribes siRNA that rapidly silences enzyme MurA. The low level of free siRNA transcripts (~2 nM) suggests that most of them target the messenger RNA *mMurA* for effective degradation (Fig. 4c yellow line). This also reduces the flux $\nu_p^q$ towards the cell wall formation, and the UDP-GlcNAc is increased. Figure 4e depicts how our synthetic device regulated UDP-sugar's supply to synthesize heparosan of low-Mw (20kDa) with a narrow polydispersity (PDI~ 1.3).

We designed the target mRNA sequence from the *murA* gene (*E. coli* K12, NCBI NC-000913.3). The *siRNA* transcripts were engineered using the software tool Invitrogen BLOCK-iT$^{TM}$ RNAi Designer from Thermofisher. We obtained 10 sequences as possible candidates of siRNAs. We selected the sequence (starting in position 639) GCAGUAGCUUCCACACCUUGCAAAU due to: *i)* it has GC content between 36-52%, *ii)* around 25-29 nucleotides (nt), *iii)* 50 to 100 nt downstream of the start codon, and *iv)* five stars rank Invitrogen BLOCK-iT meaning more than 70% probability of generating a knockdown in the desired gene *murA*. In addition, a BLAST was performed to verify the sequence homology with any other region within the genome of *E. coli* Nissle to avoid unwanted effects. Only one site with 100% homology was recognized, corresponding to the MurA enzyme.

## 3.4 Biocontroller optimization and robustness

Following the establishment of mathematical models and subsequent simulation runs, we embarked on an optimization process to determine the optimal conditions for heparosan bioproduction with desired properties. Particularly, we targeted a Mw <20 kDa and minimised the PDI. First, the optimization process involved adjusting the fluxes of UDP-GlcNAc and UDP-GlcUA ($\nu_{NAc}$ and $\nu_{UA}$, respectively) in terms of their corresponding initial ones ($\nu^0_{NAc}$ and $\nu^0_{UA}$). Then, we also optimized the promoter strength of $P_{20}$ ($k_{13}$) to transcribe *siRNAs* for effective silencing of *MurA* enzyme. Thus, the ratios of the fluxes $\nu_{NAc}/\nu^0_{NAc}$, $\nu_{UA}/\nu^0_{UA}$, and the transcription rate $k_{13}$ were considered as decision variables of our optimization problem ($\theta$). This approach is useful when one must desire outcomes while adhering to constraints imposed by the dynamics of the system, i.e. heparosan production with low Mw and low PDI. The statement of the optimization problem is the following:

$$\min_{\theta\in\Re^3} J(\theta) = J(\theta^*),\ \text{subject to}: \begin{cases}\text{equations 5 to 16}\\ \mathbf{K}(\theta)\le 0\\ \underline{\theta_i} \le \theta_i \le \overline{\theta_i},\ \forall i = [1,\ldots,3]\end{cases}$$

where $\theta = [\nu_{NAc}/\nu^0_{NAc}, \nu_{UA}/\nu^0_{UA}, k_{13}]$, $\underline{\theta_{1,2}} = 0.1$, $\overline{\theta_{1,2}} = 10$, $\underline{\theta_3} = 0.1$, $\overline{\theta_3} = 5$, and the constrains $\mathbf{K}(\theta)$ are $Mw \le \overline{20}$, $[NAc] < 8mM$, $[UA] < 8mM$.

We also imposed constraints to ensure the concentrations remained for UDP-GlcNAc <12mM and UDP-GlcUA <2.2mM to maintain physiological relevance and considering experimental data (Fig. 4d). Thus, the optimization sought to strike a balance between minimizing PDI, maintaining desired Mw, and respecting the natural constraints of the system (Supplementary section 3.1).

The results revealed that the best combination of initial fluxes and siRNA strength was $\theta^* = [\nu_{NAc}/\nu^0_{NAc} = 7.734, \nu_{UA}/\nu^0_{UA} = 1.582, k_{12} = 1.757]$. This produces heparosan with an Mw of $19.86 \pm 5.91$ kDa, well within the target range, compared to random combinations which often resulted in much higher Mw variants with higher PDIs. We also obtained a remarkable reduction in polydispersity, with a minimum PDI= $1.27 \pm 0.05$ achieved under optimal conditions. In these conditions, including a 48 hr fermentation in a 7 Liter bioreactor reaching 140 OD (Fig. 4a), the yielded titer was $\sim 0.965 \pm 0.32$ g/L (Supplementary section 3.3). These values are compatible with those experimentally obtained in [32] (further results in Supplementary Table S1).

To assess the robustness of the optimized conditions, we introduced disturbances to the system and evaluated its performance

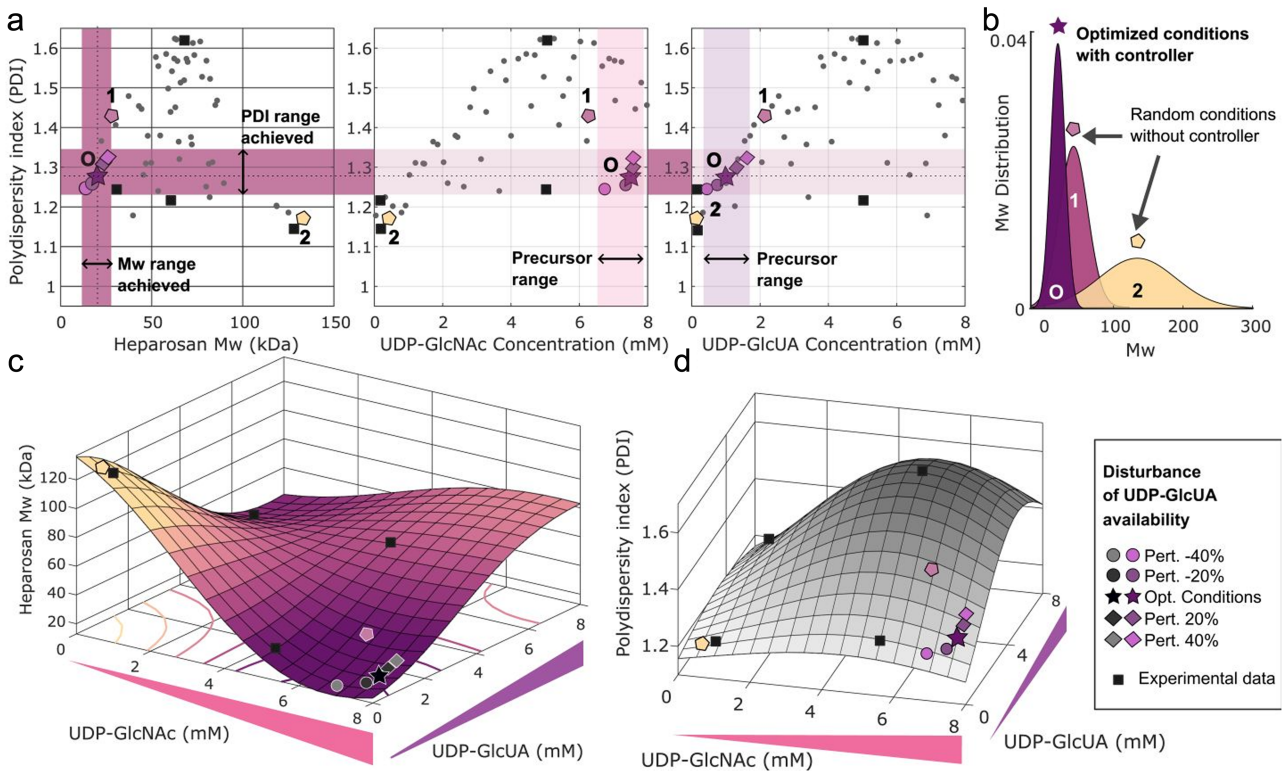



**Figure 5.** Performance analysis of the synthetic device. **a.** We computed the trade-off between the PDI and the Mw of produced heparosan for combinations of different concentrations of precursors. Grey dots represent random combinations, whereas purple shapes (O) are the optimized values. Points 1 and 2 are representatives of low-Mw/high-PDI (1), and high-Mw/low-PDI easily obtained with random combinations of precursors. Black squares are experimental data from [16]. Purple star is the optimized production, within the desired low-Mw zone with the lowest PDI. Circles represent disturbance where the availability of UDP-GlcUA is reduced and Diamonds where the availability is increased. Coloured bands show the PDI and Mw achievable ranges with our optimized conditions and dynamical regulation strategy, and the corresponding precursor range needed. **b.** Mw distributions of the selected points. Optimized conditions give a very narrow, low-Mw distribution of heparosan in comparison with the random points (1-2). **c-d.** GP estimation of the Mw (left) and PDI (right) of heparosan and its relationship with the precursor's concentration (UDP-GlcNAc in pink, and UDP-GlcUA in purple). Markers represent the same data as in A and B.

under these altered conditions (Fig. 2b). Remarkably, even under perturbation, the system maintained its capability to synthesize heparosan with consistent properties. Specifically, the PDI was observed to fluctuate within a modest 4% range, showcasing the resilience of the system to external perturbations. Figure 5 illustrates the performance analysis on the PDI and the Mw of the produced heparosan, under different combinations of precursor concentrations. Notably, points 1 and 2 serve as representatives of the outcomes obtained with random combinations (5a grey dots), where there is a clear trade-off between low-Mw/high-PDI (point 1) and high-Mw/low-PDI (point 2). Conversely, our biocontroller and the resulting production conditions achieved the desired low-Mw zone with the lowest possible PDI (5A,B purple star), in accordance with experimental data as shown by the black squares in Fig. 5. In fact, our results are consistent and robust even when disturbances such as the cell is demanding more UDP-GlcUA (5 circles) or when it is accumulated (5 diamonds). In Fig. 5b, the Mw distributions of points 1 and 2, together with optimal ones (purple star) further elucidate the impact of optimization. We obtained a significantly narrower, low-Mw distribution of heparosan compared to random points 1 and 2.

Figure 5c and d presents the GP estimations of the Mw and PDI of heparosan. The results revealed the relationship among the levels of both UDP-GlcNAc (pink) and UDP-GlcUA (purple) and these attributes. As we can see, the markers from Fig. 5a,b provide a comprehensive visualization on how precursor concentrations influence the heparosan's quality, even when one of the precursors remains at constant levels. Despite the increase/reduction of 20-40% of the available UDP-GlcUA (Fig. 5c circles and diamonds), our system exhibited remarkable stability, producing heparosan with low-Mw (differing by less than 3 kDa and 6 kDa, respectively) compared to the unperturbed condition (Fig. 5c purple start). Overall, Fig. 5d underscores the effectiveness of the optimization process minimal polydispersity (PDI< 1.3) for any scenario.

Figure 6 plots the temporal simulations of heparosan production following optimization and considering disturbances in the UDP-GlcUA availability. At $t = 24$ hours, a deliberate reduction (-40%, -20% ) or increment (20%, 40% ) of the UDP-GlcUA level was introduced. The biocontroller promptly acts to maintain the optimized balance between UDP-GlcNAc and UDP-GlcUA, in less than 10 hours (Fig. 6-bottom). This dynamic response is achieved by dynamically modifying the flux $\nu_p^q$ towards peptidoglycans, resulting in an increase of this flux and the subsequent reduction in the UDP-GlcNAc availability (Fig. 6-top). This simulation provides valuable insights into the efficacy of the biocontroller in maintaining the desired production conditions despite external disturbances. Figure 6 highlights the robustness and adaptive nature of the optimized production system. The computational simulations of the key biochemical species of our device are shown in Supplementary section 2.

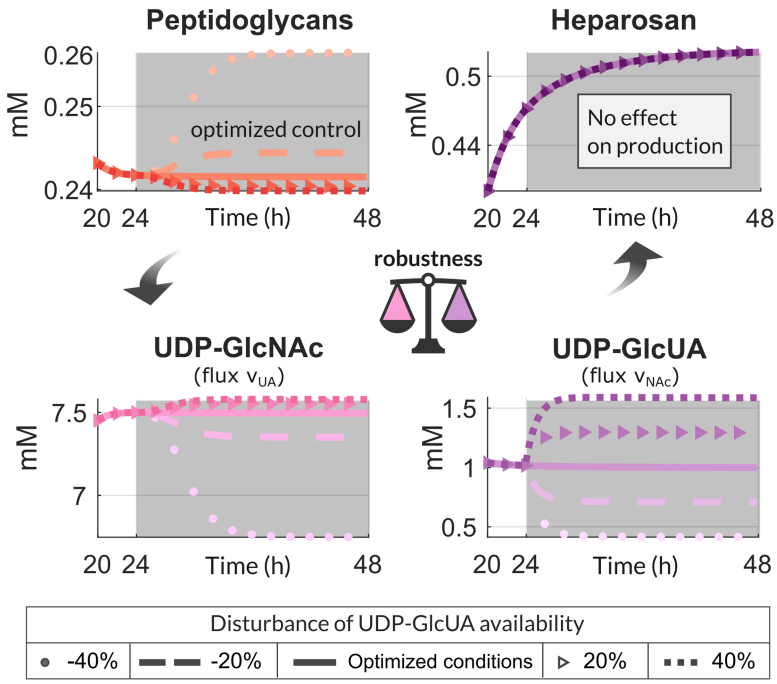



**Figure 6.** Temporal simulations of the heparosan production after optimization. To test the robustness of the biocontroller, in $t$ =24 hours, a change in the UDP-GlcUA availability is introduced (reduction in dots and dashed lines). The biocontroller acts to maintain the optimized balance between UDP-GlcNAc (bottom-left) and UDP-GlcUA (bottom-right). The flux to peptidoglycans $\nu_P^q$ (top-left) increases dynamically to reduce the availability of UDP-GlcNAc, as well as UDP-GlcUA, while the level of heparosan production remains constant (top-right). Conversely, if more UDP-GlcUA is available, the biocontroller will also increase UDP-GlcNAc.

## 4 Discussion

The production of low-Mw heparosan with minimal PDI is of paramount importance in various biomedical applications. In this study, we proposed a novel synthetic biology approach with dynamic regulation to achieve this goal, which represents a significant advancement compared to existing methodologies.

Our results demonstrate the effectiveness of the proposed strategy in producing heparosan within the desired low-Mw range (approximately 19.86 $\pm$ 5.91 kDa) while minimizing PDI to a remarkable extent (as low as 1.27 $\pm$ 0.05), with the achievable ranges with perturbation: PDI $\in (1.24, 1.32)$, Mw $\in (11, 32)$kDa. Previous approaches, in particular the work in [32] where the production of heparosan, also in *E. coli* Nissle 1917, resulted in a lowest Mw of 100.16 $\pm$ 23.87 kDa, with a PDI of 1.27 $\pm$ 0.19 and a titer in batch fermentation of 0.8 *g/L*, shows us that our approach is more than promising to improve those results.

Unlike previous approaches that rely on conventional methods, our synthetic biology-guided dynamic model offers precise control over the whole biosynthesis process, from the optimal cell density to start production, until regulating the heparosan Mw and PDI as a function of the precursors levels UDP-GlcNAc and UDP-GlcUA. Our dynamic model of production allows fine-tuning of the key parameters representing the experimental biological parts, and removes the need to physically assemble and test each potential design. Furthermore, the principle of dynamic regulation enables the optimization of heparosan production under varying conditions, ensuring consistency and reproducibility in the final product.

Another singular strength of our approach is its robustness to perturbations, particularly in UDP-GlcUA availability, a crucial precursor for glycoproteins, proteoglycans, and polysaccharides. Even under these conditions, our system demonstrated resilience, maintaining heparosan production with minimal deviation in Mw compared to unperturbed conditions. This robustness underscores the reliability and adaptability of our synthetic biology-based approach, offering a promising solution for industrial-scale heparosan production.

Furthermore, it is worth noting that our study represents a departure from traditional methodologies employed in heparosan production. While previous works have primarily focused on chemical, enzymatic synthesis, or metabolic engineering in different chassis [51] our utilization of synthetic biology principles with dynamic regulation and extended biosensors introduces a novel paradigm in heparosan production. By leveraging the inherent capabilities of biological systems, we unlock new possibilities for tailored and sustainable production methods, with potential implications for various biomedical and pharmaceutical applications.

Moreover, this study presents a rational, model-based design approach, which integrates the function, behaviour, and structure of the genetic circuit required to produce heparosan with specific characteristics. Additionally, it provides the optimized parameter values for the model, achieved through mathematical optimization and *in silico* analysis, which inform the construction of the genetic circuits.

In conclusion, our study highlights the efficacy and versatility of synthetic biology approaches in achieving the desired low-Mw heparosan with minimal PDI. The dynamic regulation employed in our strategy, coupled with robustness to perturbations, positions it as a promising avenue for advancing heparosan production methodologies, with implications for a wide range of biomedical and industrial applications.

## 5 Conclusions

By integrating synthetic biology tools and focusing on the Design stage of the DBTL cycle, we aim to transform heparosan production, offering a sustainable and efficient solution with vast implications for biomedical applications worldwide. This research represents a pivotal step towards enhancing global pharmaceutical capabilities, ensuring the availability of high- quality heparosan for various industrial uses.

## Acknowledgement

YB was supported by Acceso al Sistema Español de Ciencia e Innovación-Universitat Politècnica de València (Grant PAID-10-21), and by Secretaría de Educación Superior, Ciencia, Tecnología e Innovación of Ecuador (Scholarship Convocatoria Abierta 2011).

MF thanks to Universidad Politécnica Salesiana of Ecuador (Doctoral Fellowship CB-805-2017). The authors would like to thank the team iGEM Biotech EC 2022, for valuable insight in the conceptualization of the study and the design of the genetic circuits.

## Supplementary data

Supplementary Data are available at SYNBIO Online. All relevant codes used for the study are available in the Github repository (https://zenodo.org/doi/10.5281/zenodo.11081266)

Competing interests: No competing interest is declared.

## Funding

This work was partially supported by the following funding sources: MCIN/AEI/10.13039/501100011033 project numbers PID2023-151077OB-I00, PID2020-117271RB-C21 and AYUDA PAID-06-24 (20250092) from Universitat Politècnica de Valéncia.

## Material availability statement

Not applicable.

## Data availability statement

Not applicable.

## Author contributions statement

YB and AV conceptualized the study. YB, MF, and MS performed the *in-silico* experiments. AV, FF, and MC designed the synthetic gene circuits. All authors interpreted the data and results. YB and MF wrote the first draft of the manuscript, which was then edited by YB, MS, AV, and FF. All authors reviewed the manuscript.

## References

1. Aldulijan I, Beal J, Billerbeck S *et al*. Functional synthetic biology. *Synthetic Biology* 2023;**8**:ysad006.
2. Almeida BC, Kaczmarek JA, Figueiredo PR *et al*. Transcription factor allosteric regulation through substrate coordination to zinc. *NAR Genomics and Bioinformatics* 2021;**3**:lqab033.
3. Annunziata F, Matyjaszkiewicz A, Fiore G *et al*. An orthogonal multi-input integration system to control gene expression in Escherichia coli. *ACS Synthetic biology* 2017;**6**:1816–1824.
4. Badri A, Williams A, Linhardt RJ *et al*. The road to animal-free glycosaminoglycan production: current efforts and bottlenecks. *Current Opinion in Biotechnology* 2018;**53**:85–92.
5. Barreteau H, Kovač A, Boniface A *et al*. Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiology reviews* 2008;**32**:168–207.
6. Baytas SN, Linhardt RJ. Advances in the preparation and synthesis of heparin and related products. *Drug Discovery Today* 2020;**25**:2095–2109.
7. Berger M, Chen H, Reutter W *et al*. Structure and function of n-acetylglucosamine kinase: Identification of two active site cysteines. *European Journal of biochemistry* 2002;**269**:4212–4218.
8. Bhagavan N.V. and Chung-Eun H Chapter 14 - carbohydrate metabolism ii: Gluconeogenesis, glycogen synthesis and breakdown, and alternative pathways. In N.V. Bhagavan and Chung-Eun Ha, editors, *Essentials of Medical Biochemistry (Second Edition)*, 205–225. Academic Press, San Diego, second edition 2015.
9. Boada Y, Reynoso-Meza G, Picó J *et al*. Multi-objective optimization framework to obtain model-based guidelines for tuning biological synthetic devices: an adaptive network case. *BMC Systems biology* 2016;**10**:1–19.
10. Boada Y, Santos-Navarro FN, Picó J *et al*. Modeling and optimization of a molecular biocontroller for the regulation of complex metabolic pathways. *Frontiers in Molecular Biosciences* 2022;**9**.
11. Boada Y, Vignoni A, Picó J. Engineered control of genetic variability reveals interplay among quorum sensing, feedback regulation, and biochemical noise. *ACS Synthetic Biology* 2017;**6**:1903–1912.
12. Boada Y, Vignoni A, Picó J *et al*. Extended metabolic biosensor design for dynamic pathway regulation of cell factories. *iScience* 2020;**23**:101305.
13. Chaudhary R, Nawaz A, Fouillaud M *et al*. Microbial cell factories: Biodiversity, pathway construction, robustness, and industrial applicability. *Microbiology Research* 2024;**15**:247–272.
14. Chavaroche AAE, Springer J, Kooy F *et al*. In vitro synthesis of heparosan using recombinant Pasteurella multocida heparosan synthase PmHS2. *Applied Microbiology and biotechnology* 2010;**85**:1881–1891.
15. Chavaroche AAE, van den Broek LAM, Eggink G. Production methods for heparosan, a precursor of heparin and heparan sulfate. *Carbohydrate polymers* 2013;**93**:38–47.
16. Chavaroche AAE, van den Broek LAM, Springer J *et al*. Analysis of the polymerization initiation and activity of Pasteurella multocida heparosan synthase PmHS2, an enzyme with glycosyltransferase and UDP-sugar hydrolase activity. *Journal of Biological Chemistry* 2011;**286**:1777–1785.
17. Sung Cho J, Bae Kim G, Eun H. Cheon Woo Moon, and Sang Yup Lee. Designing microbial cell factories for the production of chemicals. *Jacs Au* 2022;**2**:1781–1799.
18. Rok Choi K, Ho Shin J, Sung Cho J *et al*. Systems metabolic engineering of Escherichia coli. *EcoSal Plus* 2016;**7**:10–1128.
19. Cuccato G, Polynikis A, Siciliano V *et al*. Mario di Bernardo, and Diego di Bernardo. Modeling RNA interference in mammalian cells. *BMC Systems biology* 2011;**5**:1–12.
20. Cummins B, Vrana J, Moseley RC *et al*. Robustness and reproducibility of simple and complex synthetic logic circuit designs using a DBTL loop. *Synthetic Biology* 2023;**8**:ysad005.
21. Danaei M, Dehghankhold M, Ataei S *et al*. Impact of particle size and polydispersity index on the clinical applications of lipidic nanocarrier systems. *Pharmaceutics* 2018;**10**:57.
22. DeAngelis PL *Heparosan-multimolecular assembly drug delivery compositions and methods of making and using same, May 21 2015*. US Patent App. 14/607,893.
23. Fick A. The early history of membrane science selected papers celebrating. *Journal of Membrane Science* 1995;**100**:33–38.
24. Fillenberg SB, Grau FC, Seidel G *et al*. Structural insight into operator dre-sites recognition and effector binding in the

GntR/HutC transcription regulator NagR. *Nucleic Acids Research* 2015;**43**:1283–1296.
25. Gatto Jr GJ, Stryer L, Tymoczko JL *et al*. *Biochemistry*. WH Freeman, 2019.
26. Norman Haworth W. Bakerian lecture-the structure, function and synthesis of polysaccharides. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 1946;**186**:1–19.
27. Heirendt L, Arreckx S, Pfau T *et al*. Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0. *Nature protocols* 2019;**14**:639–702.
28. Helmann JD. The extracytoplasmic function (ECF) sigma factors. 2002.
29. Holmqvist E, Vogel J. Rna-binding proteins in bacteria. *Nature Reviews Microbiology* 2018;**16**:601–615.
30. Sakir Hossain G, Saini M, Miyake R *et al*. Genetic biosensor design for natural product biosynthesis in microorganisms. *Trends in biotechnology* 2020;**38**:797–810.
31. Litao H, Wang Y, Yunxiao H *et al*. Biosynthesis of non-sulfated high-molecular-weight glycosaminoglycans and specific-sized oligosaccharides. *Carbohydrate Polymers* 2022;**295**: 119829.
32. Shan H, Zhao L, Litao H *et al*. Engineering the probiotic bacterium Escherichia coli nissle 1917 as an efficient cell factory for heparosan biosynthesis. *Enzyme and Microbial Technology* 2022;**158**:110038.
33. Huang H-H, Qian Y, Del Vecchio D. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nature communications* 2018;**9**:5415.
34. Jones KA, Zinkus-Boltz J, Dickinson BC. Recent advances in developing and applying biosensors for synthetic biology. *Nano Futures* 2019;**3**:042002.
35. Kim D, Kim Y, Ho Yoon S. Development of a genome-scale metabolic model and phenome analysis of the probiotic Escherichia coli strain Nissle 1917. *International Journal of Molecular Sciences* 2021;**22**:2122.
36. Libis V, Delépine B, Faulon J-L. Expanding biosensing abilities through computer-aided design of metabolic pathways. *ACS Synthetic Biology* 2016;**5**:1076–1085.
37. Mellisa L, Wang Z, Laremore TN *et al*. Analysis of e. coli k5 capsular polysaccharide heparosan. *Analytical and Bioanalytical chemistry* 2011;**399**:737–745.
38. Mayer C, Maria Kluj R, Muehleck M *et al*. Bacteria's different ways to recycle their own cell wall. *International Journal of Medical Microbiology* 2019;**309**:151326.
39. Mengin-Lecreulx D, van Heijenoort J. Characterization of the essential gene glmm encoding phosphoglucosamine mutase in Escherichia coli. *Journal of Biological Chemistry* 1996;**271**: 32–39.
40. Milo R, Phillips R *Cell Biology by the numbers*. Garland Science, 2015.
41. Nakato H, Kimata K. Heparan sulfate fine structure and specificity of proteoglycan functions. *Biochimica Et Biophysica Acta (BBA)-General Subjects* 2002;**1573**:312–318.
42. Nehru G, Rami Reddy Tadi S, Mukund Limaye A *et al*. Production and characterization of low molecular weight heparosan in bacillus megaterium using Escherichia coli k5 glycosyltransferases. *International Journal of Biological Macromolecules* 2020;**160**:69–76.
43. Palsson B *Systems biology*. Cambridge university press, 2015.
44. Park JO, Rubin SA, Yi-Fan X *et al*. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nature Chemical biology* 2016;**12**:482–489.
45. Plumbridge J. Co-ordinated regulation of amino sugar biosynthesis and degradation: the NagC repressor acts as both an activator and a repressor for the transcription of the glmUS operon and requires two separated NagC binding sites. *The EMBO journal* 1995;**14**:3958–3965.
46. Raman K, Arungundram S *Chemical Approaches to Prepare Modified Heparins and Heparosan Polymers for Biological Studies*. New York, NY: Springer US, 2022, 289–296.
47. Francesca Restaino O, D'ambrosio S, Cassese E *et al*. Molecular weight determination of heparosan- and chondroitin-like capsular polysaccharides: figuring out differences between wild-type and engineered Escherichia coli strains. *Applied Microbiology and biotechnology* 2019;**103**:6771–6782.
48. Rhodius VA, Segall-Shapiro TH, Sharon BD *et al*. Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Molecular Systems biology* 2013;**9**:702.
49. Rippe M, Stefanello TF, Kaplum V *et al*. Heparosan as a potential alternative to hyaluronic acid for the design of biopolymer-based nanovectors for anticancer therapy. *Biomaterials science* 2019;**7**:2850–2860.
50. Schomburg I, Jeske L, Ulbrich M *et al*. The BRENDA enzyme information system–from a database to an expert system. *Journal of biotechnology* 2017;**261**:194–206.
51. Li Sheng L, Min Cai Y, Li Y *et al*. Advancements in heparosan production through metabolic engineering and improved fermentation. *Carbohydrate Polymers* 2024;**331**.
52. Anshuman Shrivastava. 2 - polymerization. In Anshuman Shrivastava, Eds, *Introduction to Plastics Engineering, Plastics Design Library*, 17–48. William Andrew Publishing, 2018.
53. Sioud M. Rna interference: story and mechanisms. *Design and Delivery of SiRNA Therapeutics* 2021;2282:1–15.
54. Sismey-Ragatz AE, Green DE, Otto NJ *et al*. Chemoenzymatic synthesis with distinct Pasteurella heparosan synthases: Monodisperse polymers and unnatural structures*. *Journal of Biological Chemistry* 2007;**282**:28321–28327.
55. Todor H, Osadnik H, Campbell EA *et al*. Rewiring the specificity of extracytoplasmic function sigma factors. *Proceedings of the National Academy of Sciences* 2020;**117**:33496–33506.
56. Tucker EL, Daskin MS. Pharmaceutical supply chain reliability and effects on drug shortages. *Computers & Industrial Engineering* 2022;**169**:108258.
57. Uehara T, Park JT. The n-acetyl-d-glucosamine kinase of Escherichia coli and its role in murein recycling. *Journal of bacteriology* 2004;**186**:7273–7279.
58. Wang Z, Mellisa L, Zhang F *et al*. k5 fermentation and the preparation of heparosan, a bioengineered heparin precursor. *Biotechnology and bioengineering* 2010;**107**:964—973.
59. Williams CKI, Edward Rasmussen C *Gaussian Processes for Machine learning*. Vol. 2, Cambridge, MA: MIT Press, 2006.
60. Tingfan W, Movellan J Semi-parametric gaussian process for robot system identification. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 725–731, 2012.

Synthetic Biology, 2025, 10(1), ysaf006 , DOI: https://doi.org/10.1093/synbio/ysaf006, Advance Access Publication Date: 29 April 2025, Research Article
Submitted 7 July 2024; Received 7 July 2024; Revised 11 March 2025; Accepted 26 March 2025