

Opinion

Risk-Based Bioengineering Strategies for Reliable Bacterial Vaccine Production

Tjerko Kamminga,^{1,2,5} Simen-Jan Slagman,³ Vitor A.P. Martins dos Santos,^{1,5} Jetta J.E. Bijlsma,⁴ and Peter J. Schaap^{1,5,*}

Design of a reliable process for bacterial antigen production requires understanding of and control over critical process parameters. Current methods for process design use extensive screening experiments for determining ranges of critical process parameters yet fail to give clear insights into how they influence antigen potency. To address this gap, we propose to apply constraint-based, genome-scale metabolic models to reduce the need of experimental screening for strain selection and to optimize strains based on model driven iterative Design–Build–Test–Learn (DBTL) cycles. Application of these systematic methods has not only increased the understanding of how metabolic network properties influence antigen potency, but also allows identification of novel critical process parameters that need to be controlled to achieve high process reliability.

Risk-Based Process Development for Bacterial Antigen Production

Initiatives such as the World Health Organization (WHO) Global Vaccine Action Plan [1] aim to make vaccines more accessible to the human population. Furthermore, there are initiatives to reduce the risks of antibiotic use in livestock. In consequence, there is an increased demand for new, better, and cheaper vaccines. **Antigens** (see [Glossary](#)) in vaccines that confer protection against bacterial infectious diseases are either whole-cell bacteria (inactivated or live-attenuated) or components derived from wild-type bacterial strains ([Table 1](#)). Bacterial antigens are produced in a bioprocess, consisting of an upstream part, where the antigens are produced in large-scale fermenter systems, and a downstream part, where multiple methods are used to purify, concentrate, or formulate antigens ([Figure 1A](#)). The development of a production process ([Figure 1B](#)) for these antigens is a costly (137 million – 1.1 billion US\$ [2]) and time consuming (5–18 years) process, largely because target bacterial production strains and growth media are not directly optimized for use in a bioprocess. Current methods for process development require extensive empirical assessments of strains, growth media, and growth conditions in the feasibility phase. These empirical assessments are needed to understand the relationship between the growth conditions and the **potency** of the antigen produced, while also considering production time, volume, and costs. In addition, control over **critical process parameters (CPPs)** that influence antigen potency is required to ensure **process reliability**. Here we propose a novel **risk-based process development** framework ([Figure 2](#)) that incorporates **systems metabolic engineering** techniques for strain and upstream process development for bacterial antigen production ([Table 2](#)). This novel workflow combines the **Design for Six Sigma (DFSS)** methodology [3] for reliable process design with **Design–Build–Test–Learn (DBTL)** [4] cycles for rational strain improvement ([Box 1](#)). Compared with current methods for process development, this workflow has two main advantages: (i) the duration of the feasibility phase can be reduced because systematic methods can replace time-

Highlights

Risk-based design, aimed at gaining control over critical process parameters, has become the standard in biopharmaceutical process development; however, application for bacterial vaccines is hampered because structurally complex or undefined antigens are needed.

Process analytical technology needs to be applied to actively measure and control novel critical process parameters identified by applying systems metabolic engineering techniques.

Standardized systems metabolic engineering techniques have become accessible to generate the knowledge base that is needed to assign critical process parameters in the feasibility phase of vaccine development.

To develop reliable production processes for bacterial vaccines, risk-based DBTL cycles should be performed to identify novel critical process parameters using systems metabolic engineering techniques.

¹Laboratory of Systems and Synthetic Biology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Wageningen, The Netherlands

²Bioprocess Technology and Support, MSD Animal Health, Boxmeer, The Netherlands

³Manufacturing Science and Technology, Bilthoven Biologicals, The Netherlands

⁴Discovery and Technology, MSD Animal Health, Boxmeer, The Netherlands

⁵<https://www.wur.nl/en/Research-Results/Chair-groups/Agrotechnology-and-Food-Sciences/Laboratory-of-Systems-and-Synthetic-Biology.htm>

*Correspondence: tjerko.kamminga@merck.com (T. Kamminga) and peter.schaap@wur.nl (P.J. Schaap).

consuming trial-and-error methods; and (ii) novel critical process parameters can be identified whose control during antigen production will result in high process reliability.

Basis of the Workflow: Risk-Assessment to Identify Potential CPPs

To guarantee high process reliability, a risk-based process design is required by most regulatory authorities [5] and is standard practice in the biopharmaceutical industry. Risk-based process design requires an effective risk assessment to identify CPPs [6,7]. Examples of current CPPs in the upstream bacterial antigen process are: composition of raw materials, oxygen uptake rate, and concentration of (toxic) byproducts during cultivation (Table 3). These and other parameters are monitored using **process analytical technology (PAT)** and are maintained within defined ranges to guarantee process reliability. Some CPPs can be directly controlled, such as temperature and pH, while others can be indirectly controlled by adapting the environmental conditions for growth or the metabolic capabilities of production strains. An effective risk assessment method for the identification of potential CPPs is **failure-mode and effect analysis (FMEA)** (Box 1), which is widely applied in industry to identify where and how a process step could fail and to estimate the ability to detect and correct the failure once it occurs. Risks are systematically prioritized during FMEA, and high-priority risks are addressed first.

As shown in Figure 2, the DFSS framework consists of five steps: (i) Define, where process specifications are defined; (ii) Measure, where initial system capabilities are determined; (iii) Analyze, where the design space is determined; (iv) Design, where strains are optimised; and (v) Verify, where process reliability is determined using process analytical technology. These five steps are discussed in more detail below. As indicated in Table 2, different systems metabolic engineering techniques may be applied in each step.

Table 1. Major Bacterial Antigens Produced for Human Vaccines Using Bacterial Production Strains

Vaccine	Disease	Antigens	Doses administered in 2017 (\times million) ^a
DTP	Laryngeal diphtheria, toxic myocarditis, muscular stiffness or spasm, bronchopneumonia	<i>Corynebacterium diphtheriae</i> and <i>Clostridium tetani</i> toxoids, <i>Bordetella pertussis</i> whole-cell inactivated or acellular	342.2
PCV	Pneumonia, meningitis, and febrile bacteraemia	Polysaccharide conjugate <i>Streptococcus pneumoniae</i>	193.9
BCG	Tuberculosis	Live attenuated <i>Mycobacterium bovis</i>	129.0
Hib	Pneumonia, meningitis	Conjugated polyribosylribitol phosphate capsular polysaccharide of <i>Haemophilus influenzae</i> type b	106.3
TT2+/Td2+	Laryngeal diphtheria, toxic myocarditis, muscular stiffness, or spasm	<i>Corynebacterium diphtheriae</i> and <i>Clostridium tetani</i> toxoids	75.8
MenA	Meningitis	Polysaccharide conjugate <i>Neisseria meningitidis</i> group A	3.6

^aNumber of doses of vaccines used in humans in 2017 for which antigens are produced with nonoptimized bacterial strains. Data have been adapted from information provided by the WHO (https://www.who.int/immunization/monitoring_surveillance/data/en/) on country-reported administered doses per year.

Glossary

Antigen: live or inactivated biological component capable of inducing an immune response. Bacterial antigens are undefined (whole-cell inactivated), defined (conjugates, toxoids, or subunits), or live-attenuated.

Chemically defined culture medium: growth medium in which the exact chemical composition of all components is known.

Conditionally essential genes: genes that are required for growth under specific environmental and nutritional conditions.

Constraint-based genome-scale metabolic models (GEMs): mathematical reconstruction of biochemical (metabolic) reactions in an organism where reactions are associated with genetic/genomic information, bibliomic information, or are automatically assigned based on network structures. The mathematical format allows assessment of phenotypic properties (e.g., growth rate or specific production rate for biochemical) using a variety of computational techniques.

Critical process parameter (CPP): key process parameter that should be maintained within a specified range to ensure that the product reaches critical targets related to quality (e.g., safety or potency criteria).

Design-Build-Test-Learn (DBTL): iterative design methodology that consists of developing a design or prototype, testing of the design, analyzing the performance against design specifications and learning what worked well and what did not. When conditions have been met (design matches specifications) the iteration stops.

Design for Six Sigma (DFSS): methodology for design of a product or process based on the needs of the business and the customer. Uses techniques applied for Six Sigma process improvement where the aim is to reach high process reliability (3.4 failures in 1 million opportunities).

European Pharmacopoeia: official publication that provides common quality standards for the pharmaceutical industry in major regions in Europe to control the quality of medicines, and the

Step 1: Define Product Specifications and Potential CPPs

DFSS starts with a Define phase, where product specifications are set depending on business and customer needs. For bacterial vaccines, the product specifications defined by the customer are related to safety, efficacy, and ease of use. Business specifications are defined by financial constraints (e.g., minimal profit margin for the product) and from legal and regulatory requirements (e.g., in Europe, raw material requirements are described in the **European Pharmacopoeia**). In addition to the definition of product specifications, an initial assessment of potential CPPs is also required by performing an FMEA based on a draft process map. The draft process map should contain information on the preferred cultivation systems, the number of steps needed for scale up, the growth media available at the start of the project, the preferred downstream process technology, and the tests that can be used to determine product quality and yield. It is essential to determine potential CPPs early in the project to prevent delays in future stages caused by missed CPPs, which often occur when applying current methodology.

Step 2: Measure Initial Process Capabilities

After setting the product specifications and defining potential CPPs, the Measure phase in DFSS is started which determines the initial capability of the process. Important parameters such as the potency of the antigen, growth rate of strains, and the total amount of biomass produced in the final fermentation are measured. This phase can be accelerated by applying automatically created draft **constraint-based genome-scale metabolic models (GEMs)** (Table 2 and Box 2) to assess metabolic capabilities of strains *in silico* [8–10], followed by rationally designed *in vitro* experiments to establish strain performance. For example, draft GEMs can be used to provide insight into potential carbon sources and auxotrophic requirements such as amino acids or nucleotides that need to be added for growth. Ideally, a test is available at this stage to determine potency by measuring the concentration of one or more antigenic components (proteins or polysaccharides). Finally, in high-throughput experiments multiple model-derived medium compositions can be tested under **scaled-down industrial conditions** and the strain and conditions which show maximized potency will be selected for the Analyze phase.

Step 3: Analyze Initial Design Space

In the Analyze phase of DFSS, the initial design space for the production process will be determined by performing experiments to test value ranges for potential CPPs. For example, the optimal concentration of specific components in a selected growth medium or the optimal cultivation temperature and pH is analyzed. At this stage, the automatically created draft GEM should be manually curated [11] and constrained based on the experimental data obtained during the Measure phase (Box 1 and Figure 2). This manually curated GEM can then be used to design an improved complex or a **chemically defined culture medium** [12,13]. However, the predictive power of the GEM is limited in designing chemically defined culture media because the exact biomass composition of bacterial strains is not always known, and some components may be taken up from the environment (Table 2). Using GEMs in this stage can also identify potential metabolic bottlenecks or wasteful byproduct formation as targets for optimization during the Design phase. In addition to *in vitro* methods, clinical studies could be considered to show efficacy of the chosen antigen produced under model-optimized conditions.

Step 4: Rational Strain Design to Improve Process Performance

In the design phase of DFSS, the actual bioengineering takes place to improve strain performance and process robustness. Design strategies vary based on the type of antigen that is needed (Figure 1A). GEMs can be applied in this stage to predict the effect of using gene

substances used to manufacture them.

Failure mode and effects analysis (FMEA): structured, systematic technique to analyse how a product or process could fail and to prioritize identified failure risks based on probability, severity, and the chance a failure is detected.

Flux balance analysis (FBA): mathematical method for simulating the flow of metabolites through a metabolic network reconstruction.

High-throughput insertion tracking by deep-sequencing (HITS): mapping of transposon insertion sites by next-generation sequencing after transposon insertional mutagenesis. The method is similar to Tn-seq but uses alternative transposons that require nontargeted DNA fragmentation (Box 3).

Pathogenic bacteria: bacteria that can cause disease.

Polysaccharide conjugates: vaccines that contain polysaccharides, often obtained from a bacterial cell capsule, covalently linked to a protein which elicits a T cell-mediated immune response (e.g., tetanus toxoid, diphtheria carrier protein, or meningococcal group B outer membrane protein).

Potency: a term used to indicate the immunogenicity of the antigen, that is, the ability to induce an immune response. For bacterial antigens this correlates with, for example, the concentration of polysaccharides produced with a correct formation or the concentration of toxoids produced.

Process analytical technology (PAT): mechanism to design, analyse, and control a pharmaceutical manufacturing process through timely or continuous measurement of critical process parameters.

Process reliability: probability that a process performs adequately and matches predefined specifications for quality in a defined period under specified process conditions.

Scaled-down industrial conditions: simulation of large-scale industrial conditions in small-scale systems where critical parameters of the large-scale system (e.g., mixing time and gas transfer) will be mimicked on a small scale.

knockout or knock-in mutant strains with respect to optimal growth rates or antigen production [14]. Although not all bacterial pathogenic strains are considered to be genetically amenable, recent advances in gene editing techniques have considerably expanded the toolbox enabling researchers to make targeted gene deletions even in bacteria that were previously thought to be inaccessible [15,16]. At this stage in the project, aside from process risks, the regulatory aspect should also be considered since the use of genetically modified organisms (GMOs) is often tightly regulated (e.g., Directive 2009/41/EC) and permission to use an engineered strain is based on information regarding the safety of the strain.

Strain design strategies depend on the type of antigen needed in a vaccine and differ between whole-cell inactivated vaccines with undefined antigens, vaccines with defined antigens, and live-attenuated vaccines. Design strategies for each type of antigen are discussed below in more detail.

Rational Strain Design for Whole-Cell Inactivated Vaccines with Undefined Antigens

Robust strains used to produce whole-cell inactivated vaccines should have an optimized and well-controlled *in vitro* growth rate and consistent antigen potency. Strain robustness could be improved in a DBTL cycle through genome streamlining by knockout of nonessential genes predicted by the GEM and verified with gene-essentiality screens using random mutagenesis, such as recently performed for the bacterial pathogens *Mycoplasma pneumoniae* [17], *Pseudomonas aeruginosa* [18], and *Streptococcus pneumoniae* [19]. Alternatively, metabolism can be re-engineered to remove production of toxic compounds and to remove energy sinks when competing pathways are present [4]. Strain performances must be verified under scaled-down industrial conditions and ultimately in an efficacy study after each design cycle as there is a risk that the potency of biomass is diminished.

Rational Strain Design for Bacterial Vaccines with Defined Antigens

When antigens are defined (specific virulence factors on inactivated whole cells, **toxoids**, or **polysaccharide conjugates**), the relationship between the formation of antigens and metabolic fluxes should be understood, resulting in more predictive metabolic models. This aspect is challenging because the influence of metabolic networks on the production of virulence factors in **pathogenic bacteria** has only recently been investigated [20–24]. GEMs have been successfully applied to optimize production strains for production of biochemicals [25,26]. To optimize antigen potency, process engineers may attempt to increase metabolic flux toward the formation of virulence factors or their precursors. Random transposon mutant libraries can help to assess with high throughput detection methods [e.g., **transposon sequencing (Tn-seq)** or **high-throughput insertion tracking by deep-sequencing (HITS)**] [27] whether genes are needed for growth, antigen potency, or both [20]. If growth and antigen potency are coupled, algorithms such as Optknock [28] can be used to suggest genetic manipulations to rationally engineer strains with increased antigen potency. In addition to these targeted approaches, genomes could be optimized to obtain more robust phenotypes as for undefined antigens. Furthermore, pathogenic bacteria can use alternative metabolic pathways during growth in the host [24,29]. Mimicking these growth conditions during fermentation could result in increased expression of virulence factors [18,30] and possibly increased antigen yield.

Rational Strain Design for Live-Attenuated Vaccines

By simulating a range of growth conditions, GEMs can be applied to predict genes that are either essential or **conditionally essential** [31], which allows researchers to predict metabolic reactions that are essential for growth in a host but nonessential for growth in the production

Systems metabolic engineering:

integration of computational modelling with rational engineering (including synthetic biology), evolutionary engineering, and systems biology to develop industrially competitive microorganisms.

Toxoid vaccines: vaccines that contain a toxin that has been made harmless but still elicits a neutralizing immune response against the toxin.

Transposon sequencing (Tn-SEQ): mapping of transposon insertion sites by next-generation sequencing after transposon insertional mutagenesis using the Himar I *Mariner* transposon (Box 3).

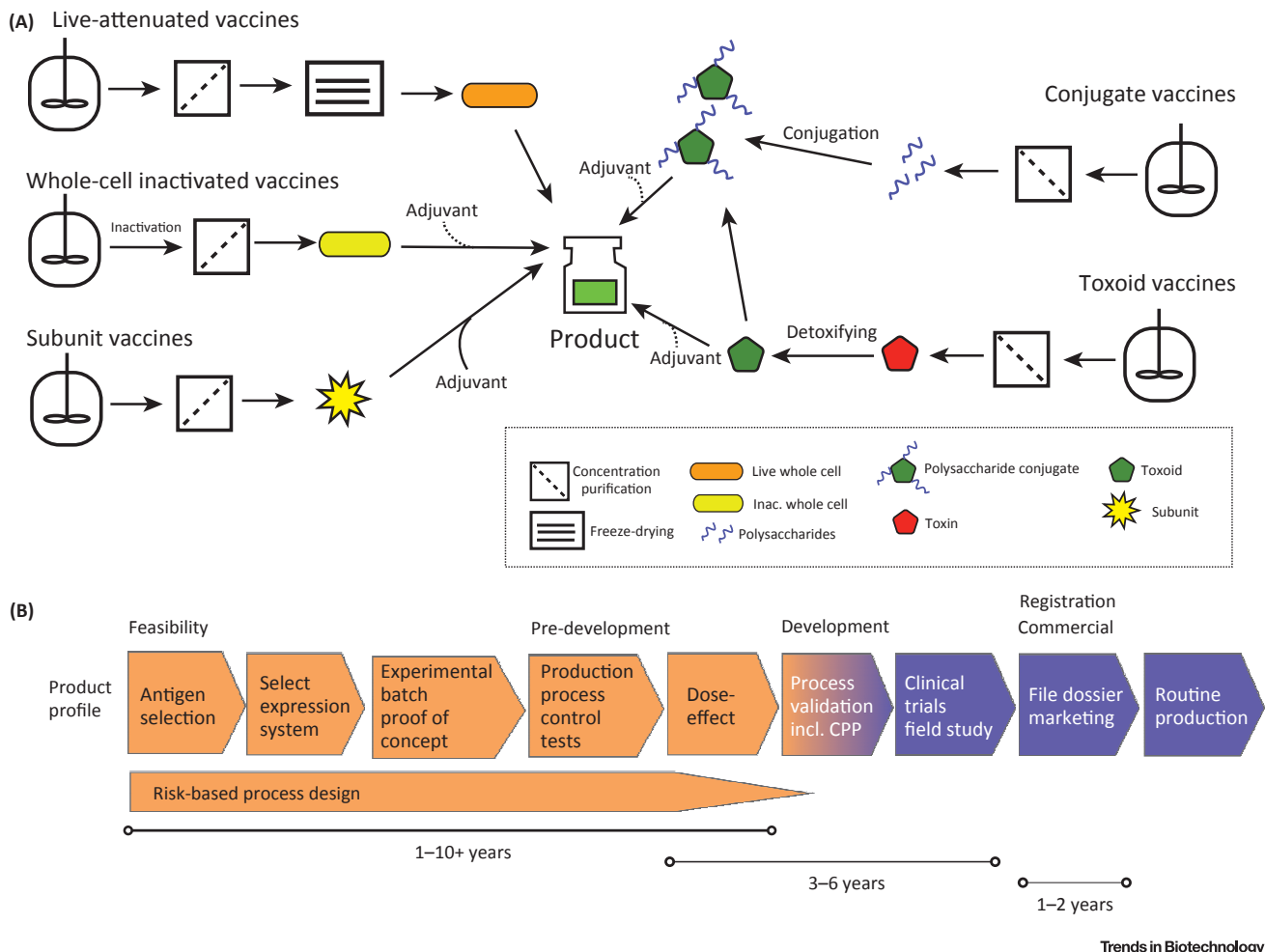
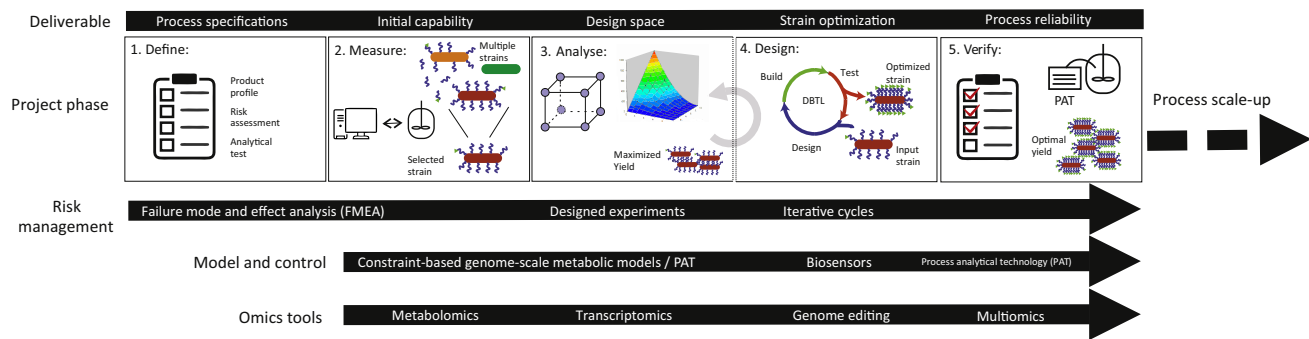


Figure 1. Production Platforms for Bacterial Vaccine Production and the Process Development Pipeline. (A) Simplified platforms for production of bacterial antigens. Five types of antigen are used in bacterial vaccines: live-attenuated and inactivated whole-cell bacteria, subunits, toxoids, and polysaccharide conjugates. Each process starts with cultivation of bacteria in controlled fermentation systems to reach high antigen potency. Next, cultures are concentrated and purified using a variety of methods [53]. After purification an adjuvant is usually added to increase the immune response. (B) Antigen process development timeline. The feasibility phase (first three blocks) can be most time consuming. Application of risk-based process design and systems metabolic engineering techniques until early development (orange blocks), can decrease the duration of the feasibility phase and increase process reliability by uncovering novel critical process parameters. Abbreviations: CPP, critical process parameter; inac., inactivated; incl., including.

process. Disrupting such genes could result in live-attenuated strains with robust fermenter growth and improved safety in the host. Once a suitable live-attenuated strain is obtained, a genome-scale metabolic model of this strain can be used to further optimize growth conditions to reach higher cell counts in live culture. However, a key challenge is to balance attenuation with immunogenicity [32] and therefore engineered strains should always be tested in animal or clinical studies.

Testing Performance of Engineered Strains

Once a production strain has been built, its performance should be tested under scaled-down industrial conditions by performing the test phase in DBTL, which is in principle a



Trends in Biotechnology

Figure 2. Risk Management Workflow for Bacterial Antigen Process Development. The novel workflow combines the DFSS methodology for robust process design with systems metabolic engineering techniques (Table 2) for rational strain improvement applying DBTL cycles (Box 1). DFSS consists of five steps: (1) Define, definition of process specifications; (2) Measure, determination of initial system capabilities; (3) Analyse, where the design space is determined; (4) Design, strain optimization; and (5) Verify, determination of process reliability using process analytical technology. The Analyse phase (Step 3) can be repeated during DBTL cycles in the Design phase (Step 4). In Step 2–Measure, automatically created constraint-based metabolic models are applied to evaluate the potential of different strains (Box 2). In Step 4–Design, strain optimization is performed using iterative DBTL cycles. In the workflow, process risk is decreased by implementing risk management techniques and development of process analytical technology to control (Box 2) or analyse the process (Box 3). Abbreviations: DBTL, Design–Build–Test–Learn; DFSS, Design for Six Sigma.

Table 2. Applications and Limitations of Systems Metabolic Engineering Techniques for Optimization of Bacterial Vaccine Production Processes

Project stage ^a	Technique ^b	Applications	Limitations	Refs
Step 2: initial capability	Genome-scale constraint-based metabolic modelling (GEM: Box 2)	Comparison of different metabolic landscapes, rational decision making when selecting a production strain	Automatically created models are incomplete but are still useful as a knowledge base, capturing all available biochemical knowledge in an actionable format	[8–10]
Step 4: strain optimization DBTL	Curated predictive GEM constrained with knowledge obtained in Step 3	Growth medium development, improvement of yield and robustness	Model functions as an improved knowledge base but may not be predictive for all antigen design specifications defined in Step 1	[12,13]
	Model-based genome editing/streamlining (Box 2)	Gene knock-in/knockout for strain improvement and flux redirection to production of precursors for virulence factors	Requires development of genome engineering tools, strict regulations for engineered strains	[4,14,20]
	Transcriptomics	Model validation, process optimization, biosensor discovery	Transcript levels do not always represent protein levels or correlate with flux distributions	[33,35]
	Metabolomics + ¹³ C flux analysis	Model validation, process optimization, biosensor discovery	No complete metabolome coverage, analysis and sampling techniques are complex, intracellular biosensors require strain engineering	[36]
	Model validation and lead discovery for further improvement	Restart the DBTL cycle for iterative improvement	Multiple iterations may be required which can be time consuming	[54]

Table 2. (continued)

Project stage ^a	Technique ^b	Applications	Limitations	Refs
Step 5: process reliability	Predictive GEM	Model predicted control of CPPs	GEM assumes a steady-state system, highly dynamic systems require alternative models	[38,39]
	Transcriptomics	PAT – process consistency	Currently offline sample analysis required. Online analysis methods need to be developed.	[55,56]
	Metabolomics	PAT to monitor process consistency	For complex/intracellular metabolites, online analysis is not yet possible	[57]

^aProject stage in DFSS (Figure 2).

^bSee Boxes 2 and 3 for details.

targeted repetition of the Analyze phase (Step 3, Figure 2). In this phase, GEMs can be applied to design the experiments to test the engineered strains, and transcriptomic and metabolomic analyses can be performed to verify model predictions (Table 2). Although transcriptomic analysis cannot be easily correlated with metabolic flux distributions [33,34], applying model-driven design and -omics analysis in this stage has the advantage that novel CPPs can be identified and translated to a PAT. For example, an expression-level analysis of virulence factors determined the optimal harvest point during antigen production for an inactivated *Bordetella pertussis* vaccine [35]. In this study, a novel CPP was determined as a function of glutamate and lactate concentrations following analyses of the transcriptional landscape at maximal antigen potency. Novel CPPs can also be identified by applying a focused metabolomics analysis strategy. Toxin production in *Clostridium tetani* was recently shown to be induced by a metabolic switch from consumption of free amino acids to consumption of peptides from complex growth medium [36]. This observation not only provided leads to improve the production medium, but also enabled the development of an additional PAT by measuring concentrations of amino acids or peptides using either online analysis methods or biosensors. In summary, GEMs and omics-based insights into the regulation of expression of virulence factors enable the study of novel gene clusters, the discovery of novel CPPs, as well as the implementation of novel PATs.

Step 5: Verifying Process Consistency

After successfully finishing the DBTL phase for rational strain design, the last phase in DFSS is the verification of strain performance by showing process consistency. At this stage, it is

Box 1. Risk Management Methods

DFSS [3] is a project management method related to Six Sigma that uses statistical tools in predefined phases to develop a product or process such that it matches business and customer needs. DBTL [4] defines the four phases in a metabolic engineering project, which are followed consecutively and often in an iterative cycle to rationally improve strain performance. Both methods rely on effective risk assessment. An important tool for risk assessment is FMEA [40]; a systematic method to analyse how and where a process or design might fail. To identify failure modes, FMEA requires participation of all experts needed to successfully develop and produce a vaccine on a large scale. A process flowchart shows all steps in a process; for each step, possible failure modes are assessed by scoring for occurrence, level of impact, and for the likelihood of detection. Using overall scores, risks are prioritized and should be addressed before moving to the next project phase.

Table 3. Advantages and Disadvantages of Current and Novel PAT to Control Potential Upstream CPPs during Bacterial Antigen Production

Process step	CPP	PAT and control measures	Advantages	Disadvantages
Raw materials	Nonconforming Chemical composition	RAMAN spectroscopy	Fast, early in the process which allows time for corrective action	Accuracy not high enough to measure individual components
		Near-infrared spectroscopy		
Growth medium production	Temperature during heat-sterilization	Online temperature probe measurement used for temperature control	Fast measurement, no further testing needed for sterility of medium if temperature profile is correct	N.A. ^a
	Bioburden load or filter integrity for filter sterilization	Offline filter integrity test and offline determination of bioburden levels prior to sterile filtration	Filter integrity test is fast and standardized	Tests are performed after the medium is produced risking batch rejection
	Medium pH	Online measurement with probes	Fast, standardized measurement	N.A.
Cultivation	Environmental conditions (temperature, pH, dissolved oxygen, dissolved CO ₂)	Online measurement of critical parameter with probes and direct control using control mechanisms such as heating/cooling or addition of base/acid	Fast measurement, standardized systems for control	Probes could fail, risking batch rejection, could be prevented by using multiple probes
	Cell growth and metabolic rates measured by oxygen uptake rate or carbon dioxide production rate	Off gas analysis, direct comparison with historical expectations, allows setting of feed strategies to control culture growth	Fast measurement, directly correlated to growth phase	Requires relatively high concentration of component of interest in off gas, not sensitive enough when using a high airflow rate on fermenter systems
	Concentration of medium components needed for growth or byproducts that limit growth	Online measurement with probes or sensors. Offline measurement with (automated enzyme assays) or chromatography (HPLC, LC-MS/MS). Allows control of feed strategy or removal of byproducts using filtration or dialysis.	Online measurements are fast, optimized growth with fed-batch strategies	Calibration needed before each production run, LC-MS/MS analysis is labour intensive and time consuming
	Mixing time and aeration rate	Measurement and control of the stirrer speed and aeration rate used to agitate and aerate the vessel. Mixing time can be calculated for different vessels based on physical dimensions of the vessel and impeller. For dynamic simulation of fluid motion and air in the vessel, computational fluid dynamics can be applied.	Standard control system on fermenter systems, mixing time can be controlled consistently between vessels	Parameters such as power input to the liquid phase or impeller tip speed will vary between vessels at different scales when there is only focus on controlling mixing time, computational fluid dynamics studies when applied to determine stirrer speed are labour intensive
	Optimal harvest moment	Online measurement of biomass with a biomass probe or a parameter that consistently changes when cultures reach stationary phase or maximum potency	Fast measurement, could utilize PAT currently in the process	Sensitivity of biomass probes insufficient to measure low biomass concentrations, harvest moment needs to fit with the production schedules

Table 3. (continued)

Process step	CPP	PAT and control measures	Advantages	Disadvantages
		of biomass is reached (e.g., oxygen consumption, base consumption, depletion of a component in growth medium)		
	Expression of genes related to antigen potency (future CPP)	Measurement of gene expression levels using offline methods (microarray, quantitative RT-PCR, RNA sequencing), active control using biosensors	Adds a novel level of control, data obtained can be integrated with GEMs to build a knowledge base needed to improve process performance	Transcriptomics methods require extensive validation and are time consuming, biosensors are novel in the field, and robustness needs to be optimized for the specific application (stimulus strength, response time, crossreactivity [58]), requires genetic modifications
	Intracellular concentration of critical metabolites (future CPP)	Online measurement of intracellular metabolite levels using spectroscopy, possibility to control metabolite levels using feed strategies	Adds a novel level of control, data obtained can be integrated with GEMs to build a knowledge base needed to improve process performance	Requires genetic modification, biosensors are novel in the field, and robustness needs to be optimised for the specific application

^aAbbreviations: HPLC, high performance liquid chromatography; N.A. not applicable.

Box 2. Tools for Process Modelling and Process Control

GEMs are created in a bottom-up approach by translating genome information into metabolic capabilities, from which a complete metabolic map of the cell is reconstructed, capturing the stoichiometry, directionality, and gene–protein relationships for all known metabolic and transport reactions. A GEM thus represents a species-specific knowledge base that can be used as a platform for hypothesis-driven investigations, interpreting multiomics data, and rational strain and process design [11]. Based on the genome sequences of candidate bacterial strains, descriptive draft GEMs can be created automatically at low cost [41–44], allowing researchers to start a vaccine development project with a set of GEMs [45]. Insight into the metabolic capabilities is further increased by **flux balance analysis (FBA)**, which utilizes reaction stoichiometry to model metabolic flux at steady state using linear problem solving [46–48]. To obtain a physiological meaningful solution space, the network is constrained for uptake of nutrients (e.g., maximal glucose uptake rate that is determined by transporter capacity). Solving an FBA problem means that the minimal or maximal flux through a target reaction is calculated. Relevant target reactions to optimize for production of bacterial antigens could be biomass formation, that is, calculation of the growth rate or production rate of a compound needed for virulence. Data needed to constrain the model are obtained from metabolomics analyses that can be rapid online analysis, using for instance, automated enzyme assays or enzyme probes, or more complex offline analysis such as liquid chromatography mass spectrometry (LC-MS/MS). For model validation, especially flux analysis, using ¹³C-labeled metabolites is widely applied [49]. Metabolite profiling using MS analysis and model-based interpretation of these data could result in detection of novel metabolite concentrations that need to be measured and controlled during a production process. Real-time monitoring of CPPs can be reached using PAT (Table 3 in main text). Introduction of PAT to the production process for bacterial vaccines is crucial, because insufficient control over critical process parameters could result in production batch failure. Examples of PAT are: temperature probes, pH probes, or online metabolite analyzers. Advanced PAT can also be developed using molecular biology techniques allowing for instance the online detection of maltose or glutamate levels [50,51] in the cell using biosensors.

Box 3. Omics Tools Applied during the Design and Verify Phases

There is a multitude of tools that can be used for targeted modification of a bacterial genome using sequence-specific nucleases [15,16]. Zinc finger nucleases and transcription activator-like effector (TALE) nucleases are widely applied techniques. Recently, genome editing using CRISPR-Cas9 has become popular since its specificity is directed by guide RNA, and protein engineering of the DNA binding domain is no longer needed. Nontargeted modification of a bacterial genome can be applied when a process is optimized using evolutionary engineering or when gene essentiality is studied. In the latter case, transposon mutagenesis [27] (Tn-seq or HITS) is often applied. Tn-seq or HITS examine gene essentiality by randomly inserting transposon sequences in the genome of bacterial strains. Insertion of a transposon sequence in most cases inactivates the gene function. The location of insertions in nonessential genes can be determined using second-generation DNA-sequencing techniques. Tn-seq uses specifically a Himar I *Mariner* transposon that introduces *MmeI* restriction sites in the inverted repeat sections of the minitransposon. When DNA is fragmented using a type IIS restriction endonuclease, terminal ends are obtained with a length of 16 base pairs, which can be amplified, sequenced, and used to map the location of insertions in the bacterial genome. HITS uses a similar method but with nonspecific DNA fragmentation, resulting in a variety of terminal end lengths that need to be mapped to the bacterial genome after DNA sequencing. DNA or RNA sequencing techniques allow the rapid elucidation of bacterial strain genomes or transcriptomes. Currently, third-generation sequencing techniques are in development that allow the rapid detection of long fragments of DNA or RNA without chemical modification [52]. Second-generation sequencing techniques, especially Illumina sequencing, have been widely applied to the sequencing of bacterial genomes and rely on fragmentation of DNA or RNA, amplification of fragments, and massively parallel sequencing of short fragments generating small reads. These reads can be mapped to the genome of a strain in the case of RNA sequencing or combined to generate a genome in the case of DNA sequencing. The combined use of short-read, second-generation sequencing, and long-read, third-generation sequencing, should result in high quality genome or transcriptome data.

Outstanding Questions

What are the best strategies to design chemically defined media for bacterial antigen production using metabolic modelling?

Is control over metabolic fluxes sufficient to optimize antigen potency for bacterial pathogens?

Which novel process analytical tools can be applied on large-scale to better control antigen potency based on metabolic flux analysis?

How much faster can novel bacterial vaccines reach the market if systems metabolic engineering techniques are routinely applied for strain improvement and to define and control critical process parameters in large-scale fermentation processes?

important to develop a robust control strategy and demonstrate in at least three consecutive test runs that process performance is consistent and matches the criteria defined in the first stage of DFSS. In addition to the standard CPPs identified in the Analyze phase, the novel CPPs identified in the Design phase should also be controlled using process analytical technology. Further development of biosensors (Table 3) could allow process engineers to continuously monitor gene expression and key metabolite levels [37] in bacterial cultures. Feeding this information into a control model could enable real-time interventions in the process to optimize antigen expression. Multivariate data analysis, together with data about the quality of the medium components and process conditions, can be used to correlate process data with biomass yields or antigen potency [38,39].

Although we focused on the upstream process development, the downstream process can be developed in parallel using a similar risk-based approach and DFSS methodology. The development of highly efficient purification or concentration techniques in the downstream process or strong adjuvants for the vaccine could help lower the minimal requirement for antigen potency in the upstream process and further reduce the risk of batch failure due to low antigen potency. After DFSS is successfully finished, the complete process can be scaled up and validated at final manufacturing scale (Figure 1B). Compared with empirical trial-and-error approaches, the rational approaches discussed here can save time and costs during scale-up, as the process has been designed based on the full-scale process parameters (mixing times and transfer times). Furthermore, controls will be in place for CPPs resulting overall in a reliable, high-yield process with a low chance of failure during scale-up.

Concluding Remarks

Risk-based process design delivers a reliable production process by gaining a more thorough understanding of and control over critical process parameters. A complete understanding of critical process parameters for bacterial antigen production remains difficult because protective structures are often undefined, and strains and media are usually not well characterized.

However, for these antigens, novel critical process parameters may be identified by iterative application of systems metabolic engineering strategies following a DBTL approach within the DFSS framework. Genome-scale metabolic models, in particular, will allow process developers to analyze the influence of metabolic flux distributions on the growth and antigen potency. We strongly encourage further investigation on the coupling of metabolic fluxes and antigen potency (see Outstanding Questions) as this plays a key role in the future systems metabolic engineering framework. Because systematic methods are used to choose strains and optimize performance, fewer experiments will be needed to design the final production process. Within the MycoSynVac project, the aim of which is to design a universal chassis for animal vaccination, we have begun to apply these methods with promising results [see MycoSynVac, Engineering *Mycoplasma pneumoniae* as a broad-spectrum animal vaccine (<http://www.mycosynvac.eu/>)].

Disclaimer Statement

T.K. and J.B. are employed by MSD Animal Health, a pharmaceutical company producing veterinary vaccines. S.J.S. is employed by Biltoven Biologicals, a biopharmaceutical company producing vaccines for use in humans. V.M.dS. and P.S. do not have any competing interests.

Acknowledgments

This work was financially supported by MSD Animal Health, Bioprocess Technology and Support, Boxmeer, The Netherlands. V.M.dS. and P.S. also received funding from the European Union's Horizon 2020 research and innovation programs under grant agreement numbers 634942 and 730976.

References

1. WHO . Global Vaccine Action Plan 2011–2020 2013;
2. Gouglas, D. *et al.* (2018) Estimating the cost of vaccine development against epidemic infectious diseases: a cost minimisation study. *Lancet Glob. Health*, 6, 1386–1396
3. Junker, B. *et al.* (2011) Design-for-Six-Sigma for Development of a bioprocess quality-by-design framework. *PDA J. Pharm. Sci. Technol.* 65, 254–286
4. Nielsen, J. and Keasling, J.D. (2016) Engineering cellular metabolism. *Cell*, 164, 1185–1197
5. US Food and Drug Administration (2004) *Pharmaceutical CGMPs for the 21st Century – a Risk-Based Approach*.
6. Rathore, A.S. (2016) Quality by Design (QbD)-based process development for purification of a biotherapeutic. *Trends Biotechnol.* 34, 358–370
7. Haas, J. *et al.* (2014) Implementation of QbD for the development of a vaccine candidate. *Vaccine*, 32, 2927–2930
8. Bosi, E. *et al.* (2016) Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci. U. S. A.* 113, E3801–E3809
9. Seif, Y. *et al.* (2018) Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. *Nat. Commun.* 9, 1–12
10. Fouts, D.E. *et al.* (2016) What makes a bacterial species pathogenic?: comparative genomic analysis of the genus *Leptospira*. *PLoS Negl. Trop. Dis.* 10, e0004403
11. Thiele, I. and Palsson, B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121
12. Cortés, M.P. *et al.* (2017) Analysis of *Piscirickettsia salmonis* metabolism using genome-scale reconstruction, modeling, and testing. *Front. Microbiol.* 8, 1–15
13. Branco dos Santos, F. *et al.* (2017) Probing the genome-scale metabolic landscape of *Bordetella pertussis*, the causative agent of whooping cough. *Appl. Environ. Microbiol.* 83, 1–19
14. Kamminga, T. *et al.* (2017) Metabolic modeling of energy balances in *Mycoplasma hyopneumoniae* shows that pyruvate addition increases growth rate. *Biotechnol. Bioeng.* 114, 2339–2347
15. Cheng, J.K. and Alper, H.S. (2014) The genome editing toolbox: a spectrum of approaches for targeted modification. *Curr. Opin. Biotechnol.* 30, 87–94
16. Esvelt, K.M. and Wang, H.H. (2013) Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol.* 9, 641
17. Lluch-Senar, M. *et al.* (2015) Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol. Syst. Biol.* 11, 780
18. Turner, K.H. *et al.* (2015) Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proc. Natl. Acad. Sci. U. S. A.* 112, 4110–4115
19. Liu, X. *et al.* (2017) High-throughput CRISPRi phenotyping in *Streptococcus pneumoniae* identifies new essential genes involved in cell wall synthesis and competence development. *Mol. Syst. Biol.* 13, 1–18
20. Bartell, J.A. *et al.* (2017) Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis. *Nat. Commun.* 8, 14631
21. Minato, Y. *et al.* (2013) Central metabolism controls transcription of a virulence gene regulator in *Vibrio cholerae*. *Microbiology*, 159, 792–802
22. Bouillaut, L. *et al.* (2015) Integration of metabolism and virulence in *Clostridium difficile*. *Res. Microbiol.* 166, 375–383
23. Brown, A.J.P. *et al.* (2014) Metabolism impacts upon candida immunogenicity and pathogenicity at multiple levels. *Trends Microbiol.* 22, 614–622
24. Eisenreich, W. *et al.* (2010) Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nat. Rev. Microbiol.* 8, 401–412
25. Harder, B.J. *et al.* (2016) Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*. *Metab. Eng.* 38, 29–37

26. Meadows, A.L. *et al.* (2016) Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature*, 537, 694–697
27. van Opijnen, T. and Camilli, A. (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.* 11, 435–442
28. Burgard, A.P. *et al.* (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657
29. Liu, Y. *et al.* (2015) Proteomic analyses of intracellular *Salmonella enterica* serovar Typhimurium reveal extensive bacterial adaptations to infected host epithelial cells. *Infect. Immun.* 83, 2897–2906
30. Schoen, C. *et al.* (2014) Metabolism and virulence in *Neisseria meningitidis*. *Front. Cell. Infect. Microbiol.* 4, 1–16
31. Koehorst, J.J. *et al.* (2016) Comparison of 432 *Pseudomonas* strains through integration of genomic, functional, metabolic and expression data. *Sci. Rep.* 6, 38699
32. Galen, J.E. and Curtiss, R. (2014) The delicate balance in genetically engineering live vaccines. *Vaccine*, 32, 4376–4385
33. Machado, D. and Herrgård, M. (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* 10, e1003580
34. Kim, M.K. *et al.* (2016) E-Flux2 and sPOT: validated methods for inferring intracellular metabolic flux distributions from transcriptomic data. *PLoS One*, 11, 1–22
35. Van De Waterbeemd, B. *et al.* (2009) Gene-expression-based quality scores indicate optimal harvest point in *Bordetella pertussis* cultivation for vaccine production. *Biotechnol. Bioeng.* 103, 900–908
36. Licona-Cassani, C. *et al.* (2016) Tetanus toxin production is triggered by the transition from amino acid consumption to peptides. *Anaerobe*, 41, 113–124
37. Rogers, J.K. and Church, G.M. (2016) Genetically encoded sensors enable real-time observation of metabolite production. *Proc. Natl. Acad. Sci. U. S. A.* 113, 2388–2393
38. Mercier, S.M. *et al.* (2014) Multivariate PAT solutions for biopharmaceutical cultivation: current progress and limitations. *Trends Biotechnol.* 32, 329–336
39. Glassey, J. *et al.* (2011) Process analytical technology (PAT) for biopharmaceuticals. *Biotechnol. J.* 6, 369–377
40. Teng, S.-H. and Ho, S.-Y. (2006) Failure mode and effects analysis: An integrated approach for product design and process control. *Int. J. Qual. Reliab. Manag.* 13, 8–26
41. Henry, C.S. *et al.* (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28, 977–982
42. Faria, J.P. *et al.* (2018) Methods for automated genome-scale metabolic model reconstruction. *Biochem. Soc. Trans.* 46, 931–936
43. Machado, D. *et al.* (2018) Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* 46, 7542–7553
44. Karlsen, E. *et al.* (2018) Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinform.* 19, 467
45. Bartell, J.A. *et al.* (2014) Comparative metabolic systems analysis of pathogenic Burkholderia. *J. Bacteriol.* 196, 210–226
46. Orth, J.D. *et al.* (2010) What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248
47. O'Brien, E.J. *et al.* (2015) Using genome-scale models to predict biological capabilities. *Cell*, 161, 971–987
48. King, Z.A. *et al.* (2015) Next-generation genome-scale models for metabolic engineering. *Curr. Opin. Biotechnol.* 35, 23–29
49. Heux, S. *et al.* (2017) Recent advances in high-throughput 13C-fluxomics. *Curr. Opin. Biotechnol.* 43, 104–109
50. Marvin, J.S. *et al.* (2011) A genetically encoded, high-signal-to-noise maltose sensor. *Proteins Struct. Funct. Bioinform.* 79, 3025–3036
51. Marvin, J.S. *et al.* (2013) An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nat. Methods*, 10, 162–170
52. Heather, J.M. and Chain, B. (2016) The sequence of sequencers: the history of sequencing DNA. *Genomics*, 107, 1–8
53. Josefsberg, J.O. and Buckland, B. (2012) Vaccine process technology. *Biotechnol. Bioeng.* 109, 1443–1460
54. Campbell, K. *et al.* (2017) The impact of systems biology on bioprocessing. *Trends Biotechnol.* 35, 1156–1168
55. Gupta, A. *et al.* (2017) Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat. Biotechnol.* 35, 273–279
56. Ramon, C. *et al.* (2018) Integrating –omics data into genome-scale metabolic network models: principles and challenges. *Essays Biochem.* 62, 563–574
57. Villaverde, A.F. *et al.* (2016) Metabolic engineering with multi-objective optimization of kinetic models. *J. Biotechnol.* 222, 1–8
58. Rogers, J.K. *et al.* (2015) Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res.* 43, 7648–7660