Pipeline for the systematic characterization of heterologous protein secretory load to assess bioproduction efficiency

Sebastián Sosa-Carrillo

▶ To cite this version:

Sebastián Sosa-Carrillo. Pipeline for the systematic characterization of heterologous protein secretory load to assess bioproduction efficiency. Quantitative Methods [q-bio.QM]. Université de Paris, 2021. English. NNT : . tel-03612661

HAL Id: tel-03612661
https://inria.hal.science/tel-03612661

Submitted on 18 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Université de Paris

École doctorale Frontières de l'Innovation en Recherche et Education (474)

Méthodes expérimentales et computationnelles

pour la modélisation des processus cellulaires (InBio)

Pipeline for the systematic characterization of heterologous protein secretory load to assess bioproduction efficiency

Par Sebastián Sosa-Carrillo

Thèse de doctorat de SCIENCES DU VIVANT APPLIQUÉES,
BIOTECHNOLOGIE ET
INGÉNIERIE DES BIOSYSTÈMES MOLÉCULAIRES

Dirigée par Gregory Batt

Présentée et soutenue publiquement le 15 décembre 2021

Devant un jury composé de:

Pr. Megan McClean, University of Wisconsin-Madison	Rapportrice
Pr. Tom Ellis, Imperial College London	Rapporteur
Pr. Nathalie Giglioli-Guivarc'h, Université de Tours	Examinatrice
Pr. Fayza Daboussi, Université de Toulouse	Examinatrice
Pr. Cosmin Saveanu, Institut Pasteur	Examinateur
Pr. Gregory Batt, Inria Paris & Institut Pasteur	Directeur de thèse
Dr. François Bertaux, Institut Pasteur & Inria Paris	Invité

Sebastian Sosa-Carrillo: *Pipeline for the systematic characterization of heterologous protein secretory load to assess bioproduction efficiency*, December 2021. This work is licensed under a Creative Commons Attribution-NonCommercial (CC BY-NC) 4.0 International License.
https://creativecommons.org/licenses/by-nc/4.0/

For more information, contact the author:
Sebastian Sosa-Carrillo (sebassc.bt@gmail.com)

Acknowledgements

I don't like big cities. Even so, I do not regret the last six years in Paris. It has been one of the most intense times I have lived through. My time here, with very good and very bad moments, has been in perspective very positive, and I have to thank those who have been part of it.

First of all, I thank to Gregory Batt for giving me the opportunity to carry out this project. Not only that, over the years I have learned from him every time we interacted. I consider him an important mentor in my career and I will certainly apply in different aspects of my life what I have taken from the experience working with him. Another very important mentor in this experience has been François Bertaux. I feel very fortunate to have worked and shared ambitions with someone so excellent as scientist and human. He has motivated me and boosted my confidence to step out of my comfort zone of knowledge, and that has greatly enhanced my skills and my performance. I also thank Chetan Aditya, as he has recognized, we had "our fair share of conflicts", but still he helped me whenever I needed it.

Within the InBio team, everyone deserves my recognition. Steven Fletcher, Olivier Borkowski, Arthur Carcano, Virgile Andreani and Achille Fraisse have made working days more cheerful. I also thank those members who helped me directly with the PhD project in the early stages such as Mariela Furstenheim, and in the later stages, so that I could devote more time to the writing process, such as Hélène Philippe. Both intern students whose work has been and will surely be remarkable in the science of this project. Moreover, in the scientific side of this project, a very valuable contribution was made by the thesis advisor committee, composed of Gilles Truan and Jean-Marc Nicaud. They have provided an outside, unbiased view of the project when I have needed it, and I have felt very well supported and advised during the annual meetings. I thank them for that. I thank also the members of the thesis defence jury, their comments are very valuable, and I appreciate the feedback very much.

When I arrived in Paris, and still today, I found that living in the city is tough. During this time here, some people have supported me to be able to settle in Paris. Thanks for that to the Centre de Recherches Interdisciplinaires, CRI, especially to Jeanette Nguyen, Chiara Fracassi and Ariel Lindner. I always felt that they looked at the human component of the students. We are more than a CV. The CRI was there in the moments I needed, I can only be grateful to this place.

Beyond the strictly professional sphere, gracias a quien me dio los motivos para venir a Francia y me meció hasta el doctorado, Marie Delage. Gracias también a Stefan Früh, su ayuda ha sido crucial en los tiempos que más perdido he estado. Ellos me han dado lecciones que no olvidaré y me han ayudado a conocerme mejor, me han hecho mejor persona. Además, cuando he tenido que levantarme y tirar adelante, la ayuda de Alicia Calvo, Albin Salazar y Steven Fletcher ha sido muy significativa. Gracias por escucharme y ser mis amigos, gracias por confiar en mí y compartir hogar conmigo. Gracias también a los científicos con quien comparto penas y alegrías, y puedo identificarme: Rebeca Ponce, Roberto Notario, Hugo Magalhães, Roque Pastor, Arantxa Jiménez, Olga Martínez, ...

Agradecido especialmente a mi familia, enormemente, ellos son los que han estado siempre ahí y han hecho que yo pueda estar aquí. He contado con su apoyo incondicional en todos los aspectos: moral, emocional y económico. Tengo la suerte de sentirme parte de una cultura basada en que lo más importante es el amor. Me lo han inculcado mis padres y mis abuelos, y la comparto con mis hermanos y hermana, y con mis amigos más queridos. El amar y ser amado por vosotros es lo que me ha dado la seguridad. No necesito nombraros, sabéis quiénes sois. Habéis sido el eje al que me agarré para no salir volando. Cuando me daba miedo fallar o no ser capaz, y me sentía sin valor, pensaba: algo tengo que valer si ellos me quieren. Eso me daba confianza y ya me daba igual lo demás. Sabed que yo también os quiero, y que me tenéis como yo os tengo. En esto también he de incluir y agradecer a Laura por haber sido mi familia aquí. Esta etapa termina con ella como compañera, dándome el abrazo que necesito, ayudándome, preocupándose y permitiéndome rendir al máximo. En definitiva, compartiendo esa cultura del amor conmigo y meciéndome dulcemente a las siguientes.

Os agradezco a todos vosotros, y a más que no nombro, el haberme ayudado, de una u otra manera, hasta el final de este proyecto. Como sabéis, una tesis no me vale nada sin vuestra confianza y amor.

Ca’ vez que salgo a buscarme
me encuentro más cerca de mí
Como mecido en el aire
deriva me trajo hasta aquí

No quiero ser como nadie
No digas que no lo advertí
No hay manera de encontrarse
si uno no llega hasta el fin

A todos los que me han mecido hasta aquí

Pipeline for the systematic characterization of heterologous protein secretory load to assess bioproduction efficiency

Sebastián Sosa-Carrillo

DECEMBER 2021

Summary

Bioproduction is a field of applied biology that aims to produce goods of interest, mainly proteins or their reaction products, by using biological systems as a factory. It is expected to become one of the most important technologies for sustainable industrial production in the years to come. In the last decades, yeasts have been widely used as a model system for bioproduction. However, despite significant research efforts to elucidate the biology of recombinant protein secretory pathways, strain engineering and bioprocessing optimization to balance production demand with secretory cell capacity largely remains a trial-and-error based process. Important unknowns still remain regarding the bottlenecks that limit secretory function for complex proteins. Such knowledge gaps are often linked to technological limitations in our capacity to tightly control the protein production demand and to perform secretion measurements in real time.

During my PhD, I developed a small collection of yeast strains, producing various proteins of different secretory complexity, and I characterized protein production and secretion, as well as cellular stress, for different levels of demand using optogenetics. I used an automated bioreactor platform that allows to follow the production process by cytometry at high temporal resolution and over long periods of time. I also developed a simple and generic method that enables us to take rapid measurements of secretion levels directly from the cell culture. By combining experimental data and computational methods I developed a mathematical model that is able to predict the contribution of protein active degradation over secretion along the secretory pathway. I quantified the levels of demand beyond which stress response is triggered, growth ceases, and the degradation-secretion balance is lost in favor of degradation. I also found that under high demand conditions, population heterogeneity might limit, and even reduce, secreted protein levels. These results should contribute to a better understanding of the secretory pathway and should help streamline the optimization of protein bioproduction.

Une approche pour caractériser le coût de la sécrétion de protéines et quantifier l'efficacité de la bioproduction

Sebastián Sosa-Carrillo

DÉCEMBRE 2021

Résumé

La biofabrication est un domaine de la biotechnologie qui vise à fabriquer des produits d'intérêt en utilisant des systèmes biologiques comme usine biochimique. Elle devrait devenir l'une des technologies industrielles les plus importantes pour la production durable dans les années à venir. Au cours des dernières décennies, différentes espèces de levures ont été largement utilisées comme systèmes modèles pour la bioproduction, notamment de protéines d'intérêt thérapeutique ou industriel. Cependant, malgré les efforts considérables de la recherche pour élucider la biologie des processus de sécrétion des protéines recombinantes, l'ingénierie des souches et l'optimisation des bioprocédés pour adapter la demande de production à la capacité des cellules reposent encore largement sur une approche par essais et erreurs. D'importantes inconnues subsistent sur les goulots d'étranglement qui limitent la sécrétion de protéines complexes. Ces lacunes sont souvent liées aux limites technologiques de notre capacité à contrôler étroitement la demande de production de protéines et à mesurer les niveaux de sécrétion en temps réel

Au cours de mon doctorat, j'ai développé une collection de souches de levure, qui produisent des protéines de complexité sécrétoire différente, et j'ai caractérisé leur production et leur sécrétion, ainsi que le stress cellulaire produit par différents niveaux de demande en utilisant l'optogénétique. J'ai utilisé une plateforme de bioréacteur automatisée qui permet de suivre le processus de production par cytométrie avec une haute résolution temporelle et sur de longues périodes de temps. J'ai également développé une méthode simple et générique qui permet de mesurer rapidement les niveaux de sécrétion directement à partir d'une culture cellulaire. En combinant des données expérimentales avec des méthodes informatiques, j'ai développé un modèle mathématique capable de prédire la compétition entre la dégradation active des protéines et la sécrétion le long de la voie sécrétoire. J'ai quantifié les niveaux de demande au-delà desquels la réponse au stress est déclenchée, la croissance cesse et la dégradation domine la sécrétion. J'ai également découvert que, dans des conditions de forte demande, l'hétérogénéité de la population peut limiter, voire réduire, les quantités de protéines sécrétées. Ces résultats devraient contribuer à une meilleure compréhension de la voie sécrétoire et aider à rationaliser l'optimisation de la bioproduction de protéines.

Au cours du siècle dernier, l'industrie chimique a contribué au développement de presque tous les aspects de nos sociétés modernes. La production de matériaux, de combustibles, d'aliments et de produits pharmaceutiques, parmi de nombreux autres biens d'intérêt, a été cruciale pour satisfaire la consommation croissante de notre civilisation. Aujourd'hui, l'industrialisation excessive a affecté et dégradé l'équilibre environnemental. Les sociétés doivent développer des alternatives durables capables de faire face aux pires scénarios de changement climatique prévus pour le 21e siècle. Aujourd'hui, à l'aube d'une quatrième révolution industrielle imminente, la bioproduction est appelée à jouer un rôle important dans le développement d'une production industrielle écologiquement durable.

La bioproduction, le bioprocessing ou la biofabrication, est un domaine de la biologie appliquée qui vise à produire des biens d'intérêt en utilisant des systèmes biologiques comme usines ou catalyseurs. Les micro-organismes, les cellules eucaryotes individuelles, les enzymes ou même les systèmes acellulaires basés sur des lysats cellulaires sont des exemples de plateformes de production. La biofabrication présente un énorme potentiel d'amélioration et d'adaptabilité grâce aux progrès réalisés dans des disciplines de recherche telles que la bio-ingénierie, la biologie synthétique, la biologie des systèmes et la biologie computationnelle, entre autres.

Pour produire la plupart des biens d'intérêt, les usines cellulaires synthétisent des protéines, qui sont les produits finis souhaités ou des catalyseurs pour synthétiser le bien d'intérêt. Un exemple courant de produit d'intérêt fabriqué par des usines cellulaires est l'insuline humaine commerciale synthétisée dans *E. coli*. Il a été commercialisé par Eli Lilly en 1982, devenant ainsi le premier produit biopharmaceutique. Depuis lors, l'industrie biotechnologique a mûri et son marché s'est développé. Aujourd'hui, elle vaut des centaines de milliards de dollars. Le secteur vise à devenir l'une des technologies d'énergie renouvelable les plus importantes du 21e siècle.

En ce sens, l'utilisation de bioréacteurs à petite échelle et à faible coût est particulièrement adaptée aux plateformes de recherche. La possibilité de contrôler précisément les conditions de culture, de maintenir des états stables sur de longues périodes ou d'appliquer des perturbations dynamiques, permet de nouvelles approches de recherche pour améliorer le processus de bioproduction. Pour atteindre cet objectif, il est essentiel de détecter et d'interagir avec le système en temps réel, ce qui nécessite l'utilisation de la robotique. En outre, la modélisation et les simulations mathématiques deviennent des facteurs clés dans la compréhension des processus biologiques. Cette compréhension, associée à la fonctionnalité des bioréacteurs, peut être utilisée pour mettre en œuvre un contrôle en temps réel du système biologique et obtenir une performance optimale du processus. Cependant, l'assemblage de tous les dispositifs de mesure et de contrôle avec un ensemble de plusieurs bioréacteurs à culture continue constitue un défi majeur. Ces développements technologiques, ainsi que la normalisation des processus, pourraient contribuer de manière importante à la mise en œuvre et à la démocratisation de la fabrication à petite échelle.

Dans le domaine de la bioproduction, la purification des produits est devenue l'un des facteurs limitatifs les plus importants dans le cadre d'un travail continu. Lorsque des protéines hétérologues d'intérêt sont produites, leur sécrétion facilite généralement le processus de purification ultérieur. La levure est une usine cellulaire eucaryote couramment utilisée pour la bioproduction, notamment en raison de sa capacité à sécréter des protéines d'un certain degré de complexité. Cependant, la biologie sécrétoire des eucaryotes est complexe et régulée de manière multifactorielle. Il a été observé que les niveaux de sécrétion de la protéine d'intérêt ne sont pas toujours en corrélation avec les niveaux de demande de production. Dans le processus de sécrétion, des exigences élevées peuvent conduire à un trafic excessif de protéines par la voie sécrétoire, ce qui entraîne des goulots d'étranglement et, finalement, un stress. Le système adaptatif endogène des cellules résout le problème en dégradant les protéines accumulées, y compris celles qui présentent un intérêt.

En général, la bio-ingénierie a essayé d'éviter les goulets d'étranglement le long de la voie sécrétoire en se basant sur des solutions spécifiques au produit. Malgré son succès, cette approche manque de généralité, car le goulot d'étranglement pour chaque produit peut apparaître à de nombreuses étapes différentes du processus. Des approches génériques basées sur des procédures standard de caractérisation du système et de régulation intelligente sont nécessaires pour réduire le temps et les ressources et ainsi simplifier le processus d'optimisation. La normalisation est considérée comme une caractéristique fondamentale du progrès technologique et a joué un rôle clé dans la dernière révolution industrielle. Le développement de stratégies génériques et systématiques visant à minimiser les processus de caractérisation des composants basés sur les essais et les erreurs sera probablement d'une grande importance pour le développement de la prochaine révolution.

Comme mentionné ci-dessus, dans le domaine de la bioproduction, il est nécessaire de développer des stratégies génériques pour caractériser le système de sécrétion. Les nouvelles approches doivent être normalisées et systématiques pour surmonter l'inefficacité du processus typique d'essais et d'erreurs. Dans cette thèse, je présenterai le processus de développement d'une approche systématique pour caractériser le processus de sécrétion. L'approche intègre trois composantes différentes pour établir un cadre conceptuel permettant de rendre compte des processus de bioproduction impliquant la sécrétion de protéines.

Le premier volet est basé sur une méthode rationnelle de construction des modules biologiques et des souches à caractériser. Ce processus comprend la construction des modules, leur caractérisation individuelle et leur intégration dans l'hôte final. Une fois les souches construites et vérifiées, les différents modules biologiques sont caractérisés individuellement. Il y a trois modules : (i) un système d'expression optogénétique pour le contrôle de la demande de production, (ii) un rapporteur de stress, et (ii) un pool de protéines d'intérêt avec des modifications permettant le suivi en temps réel des concentrations internes et la détection des niveaux de sécrétion. Les trois modules sont ensuite intégrés dans l'hôte final et utilisés pour caractériser le processus de sécrétion.

Le deuxième volet de cette stratégie consiste en des développements technologiques appliqués à une plateforme de bioréacteurs pour des processus continus. Nous avons étendu une plateforme existante de bioréacteurs multiples en mettant en œuvre des capacités optogénétiques et des mesures automatisées à haute résolution temporelle et au niveau de la cellule unique. Ces avancées permettent de réaliser des expériences de contrôle en temps réel. En plus des composants qui font partie de la plateforme, nous avons développé une méthodologie générique pour mesurer les niveaux de sécrétion d'une manière simple et facilement automatisée.

Le troisième volet est basé sur une approche informatique quantitative permettant d'intégrer et d'analyser rationnellement les données générées par les différents dispositifs de mesure. En appliquant un modèle mathématique simple et en dérivant les valeurs des paramètres à partir des ajustements des données expérimentales à la simulation, nous pouvons interpréter les résultats de manière exhaustive. Ce cadre de calcul nous permet de déduire la contribution des composants d'élimination des protéines au cours du processus de sécrétion (sécrétion, dilution due à la croissance cellulaire, ou dégradation active) pour un ensemble de souches qui sécrètent des protéines de complexité sécrétoire différente.

L'intégration de ces trois composantes expérimentales et computationnelles nous permet de développer une approche systématique pour caractériser le processus de sécrétion d'une manière générique et standard. En réalisant des cultures continues des souches hôtes et en appliquant différents niveaux de demande de production, nous avons généré et analysé des données qui nous ont permis de déchiffrer des modèles communs dans le processus de sécrétion. Ces connaissances peuvent être appliquées pour établir des expériences de contrôle de rétroaction sur les plateformes de bioréacteurs afin d'améliorer le processus d'optimisation des souches pour la bioproduction.

Ce projet de thèse présente des contributions technologiques, méthodologiques et théoriques au domaine de la biologie. D'un point de vue technologique, la mise en œuvre de mesures de sécrétion à l'aide d'*immunobads*, présentée au chapitre 3, section 3.2, est, à ma connaissance, la première approche basée sur la cytométrie en flux pour mesurer les niveaux de sécrétion sans avoir besoin de séparer les cellules de l'échantillon. La méthode a le potentiel d'être automatisée, mais pour ce faire, je dois encore surmonter certaines limitations spécifiques à la plateforme. Contrairement à la plupart des méthodes de mesure des niveaux de sécrétion, nos mesures basées sur les *immunobads* sont quantitatives, simples et sensibles, ce qui les rend compatibles avec l'automatisation.

Quant à la contribution méthodologique, je présente dans le chapitre 4 le développement d'une méthode basée sur la biologie systémique et les approches computationnelles pour caractériser le processus de sécrétion selon une stratégie systématique. Nous proposons une approche expérimentale et computationnelle générique pour caractériser le processus de sécrétion sous différentes demandes de production. Les modèles mathématiques présentés montrent les distributions des différents destins, sécrétion, dégradation ou dilution, d'une protéine d'intérêt dans la voie sécrétoire.

L'interprétation de ces résultats contribue également à une meilleure compréhension du processus de sécrétion en fournissant une vue non statique du système. La contribution de la dégradation, de la sécrétion et de la dilution de la croissance est considérablement affectée par le trafic des protéines dans la voie. En ce sens, ces résultats contribuent également au domaine de la bioproduction, car il est probable que des niveaux optimaux de demande de production puissent être trouvés dans l'analyse des modèles de sécrétion.

La stratégie de caractérisation systématique susmentionnée a également révélé l'effet de la population sur le processus de sécrétion (chapitre 4). La détection de populations bimodales définies par les niveaux de protéines internes, ainsi que par les niveaux de stress, est pertinente pour comprendre comment la variabilité de la capacité de sécrétion au niveau de la population peut avoir un impact sur l'efficacité de la bioproduction. A partir de cette observation, nous avons proposé et utilisé les capacités de notre plateforme de bioréacteur pour réaliser des expériences en temps réel afin d'améliorer et de développer des stratégies applicables aux processus de bioproduction.

Les autres contributions de ce travail comprennent la mise en œuvre d'une autorégulation interne, basée sur la biologie synthétique, qui, bien que non achevée, est documentée dans l'annexe A. Dans ce projet, j'ai suivi une approche alternative pour faire correspondre la demande à la capacité des cellules et j'ai mis en œuvre une solution utilisant le contrôle interne au niveau des cellules individuelles. D'autres contributions indirectes de mon projet de doctorat ont été la génération de protocoles, de plasmides et de souches de levure utilisés dans d'autres projets de notre équipe.

Una estrategia para caracterizar el coste de la secreción de proteínas y cuantificar la eficiencia en bioproducción

Sebastián Sosa-Carrillo

DICIEMBRE 2021

Resumen

La bioproduction es un campo de la biotecnología que pretende obtener productos de interés utilizando sistemas biológicos como fábrica bioquímica. Se espera que se convierta en una de las tecnologías industriales más importantes para la producción sostenible en los próximos años. En las últimas décadas, diferentes especies de levaduras se han utilizado ampliamente como sistema modelo para la bioproducción, especialmente de proteínas de interés terapéutico o industrial. Sin embargo, a pesar de los importantes esfuerzos en investigación para dilucidar la biología de los procesos de secreción de las proteínas recombinantes, la ingeniería de las cepas y la optimización de los bioprocesos para adecuar la demanda de producción a la capacidad celular siguen dependiendo en gran medida de un enfoque basado en ensayo y error. Quedan aún importantes incógnitas sobre los cuellos de botella que limitan la secreción de proteínas complejas. Estas deficiencias suelen estar relacionadas con las limitaciones tecnológicas de nuestra capacidad para controlar estrictamente la demanda de producción de proteínas y para realizar mediciones de los niveles de secreción en tiempo real.

Durante mi doctorado, desarrollé una colección de cepas de levadura, que producen proteínas de diferente complejidad secretora, y caractericé la producción y secreción de éstas, así como el estrés celular producido por diferentes niveles de demanda utilizando la optogenética. Utilicé una plataforma de biorreactor automatizada que permite controlar el proceso de producción mediante citometría con una alta resolución temporal y durante largos periodos de tiempo. También he desarrollado un método sencillo y genérico que permite medir rápidamente los niveles de secreción directamente a partir de un cultivo celular. Combinando datos experimentales con métodos computacionales, desarrollé un modelo matemático capaz de predecir la competencia entre la degradación activa de proteínas y la secreción a lo largo de la vía secretora. He cuantificado los niveles de demanda más allá de los cuales se desencadena la respuesta al estrés, cesa el crecimiento y la degradación domina la secreción. También descubrí que, en condiciones de alta demanda, la heterogeneidad de la población puede limitar o incluso reducir las cantidades de proteínas secretadas. Estos resultados deberían contribuir a una mejor comprensión de la vía secretora y ayudar a racionalizar la optimización de la bioproducción de proteínas.

Durante el último siglo, la fabricación de productos químicos ha contribuido al desarrollo de casi todos los aspectos de nuestras sociedades modernas. La producción de materiales, combustibles, alimentos y productos farmacéuticos, entre otros muchos bienes de interés, ha sido crucial para satisfacer el creciente consumo de nuestra civilización. En la actualidad, la excesiva industrialización ha afectado y degradado el equilibrio medioambiental. Las sociedades necesitan desarrollar alternativas sostenibles que puedan hacer frente a los peores escenarios de cambio climático previstos para el siglo XXI. Hoy, en los albores de una inminente cuarta revolución industrial, la bioproducción está destinada a desempeñar un importante papel en el desarrollo de una producción industrial ambientalmente sostenible.

La bioproducción, el bioprocesamiento o la biofabricación, es un campo de la biología aplicada cuyo objetivo es producir bienes de interés utilizando sistemas biológicos como fábricas o catalizadores. Algunos ejemplos de plataformas de producción son los microorganismos, las células eucariotas individuales, las enzimas o incluso los sistemas libres de células basados en lisados celulares. La biofabricación tiene un enorme potencial de mejora y adaptabilidad gracias a los avances en disciplinas

de investigación como la bioingeniería, la biología sintética, la biología de sistemas y la biología computacional, entre otras.

Para producir la mayoría de los bienes de interés, las fábricas celulares sintetizan proteínas, que son los productos finales deseados o los catalizadores para sintetizar el bien de interés. Un ejemplo común de un producto de interés fabricado a través de fábricas celulares es la insulina humana comercial sintetizada en *E. coli*. Fue comercializado por Eli Lilly en 1982, convirtiéndose en el primer producto biofarmacéutico. Desde entonces, la industria biotecnológica ha madurado y su mercado ha crecido. Hoy en día, su valor asciende a cientos de miles de millones de dólares. El sector aspira a convertirse en una de las tecnologías de energías renovables más importantes del siglo XXI.

En este sentido, el uso biorreactores a pequeña escala y de bajo coste es especialmente conveniente para plataformas de investigación. La capacidad de controlar con precisión las condiciones de cultivo, de mantener estados estables durante largos períodos de tiempo o de aplicar perturbaciones dinámicas, permite nuevos enfoques de investigación para mejorar el proceso de bioproducción. Para lograr este objetivo, es esencial detectar e interactuar con el sistema en tiempo real, lo que requiere el uso de la robótica. Además, la modelización matemática y las simulaciones se están convirtiendo en factores clave para comprender los procesos biológicos. Esta comprensión, junto con la funcionalidad de los biorreactores, puede utilizarse para implementar el control en tiempo real del sistema biológico y lograr un rendimiento óptimo del proceso. Sin embargo, ensamblar todos los dispositivos de medición y control con un conjunto de varios biorreactores de cultivo continuo es un reto importante. Estos desarrollos tecnológicos, junto con la normalización de los procesos, podrían contribuir de forma importante a la implantación y democratización de la fabricación a pequeña escala.

En el campo de la biofabricación, la purificación del producto se ha convertido en uno de los factores limitantes más importantes a la hora de trabajar de forma continua. Cuando se producen proteínas heterólogas de interés, su secreción generalmente facilita el proceso de purificación posterior. La levadura es una fábrica celular eucariota de uso común para la bioproducción, especialmente por su capacidad de secretar proteínas con cierto grado de complejidad. Sin embargo, la biología secretora de los eucariotas es compleja y está regulada de forma multifactorial. Se ha observado que los niveles de secreción de la proteína de interés no siempre se correlacionan con los niveles de demanda de producción. En el proceso de secreción, las elevadas demandas pueden conducir a un tráfico excesivo de proteínas a través de la vía secretora, lo que provoca cuellos de botella y, en última instancia estrés. El sistema adaptativo endógeno de las células resuelve el problema degradando las proteínas acumuladas, incluidas las de interés.

Normalmente, la forma en que la bioingeniería ha tratado de evitar los cuellos de botella a lo largo de la vía secretora se basa en soluciones específicas para cada producto. A pesar de su éxito, este enfoque carece de generalidad, ya que el cuello de botella de cada producto puede aparecer en muchas etapas diferentes del proceso. Se necesitan enfoques genéricos basados en procedimientos estándar de caracterización del sistema y regulación inteligente para reducir el tiempo y los recursos, y así, simplificar el proceso de optimización. La estandarización se considera una característica fundamental del progreso tecnológico y ha desempeñado un papel clave en la pasada revolución industrial. El desarrollo de estrategias genéricas y sistemáticas para minimizar procesos basados en ensayo y error para caracterizar componentes será probablemente de gran importancia para desarrollar la próxima revolución.

De acuerdo con lo mencionado anteriormente, en el campo de la bioproducción existe la necesidad de desarrollar estrategias genéricas para caracterizar el sistema de secreción. Las nuevas aproximaciones deben ser estandarizadas y sistemáticas para superar la ineficacia del típico proceso de prueba y error. En esta tesis presentaré el proceso de desarrollo de una aproximación sistemática para caracterizar el proceso de secreción. El enfoque integra tres componentes diferentes para establecer un marco conceptual que dé cuenta de los procesos de bioproducción que implican la secreción de proteínas.

El primer componente está basado en un método racional de construcción de los módulos biológicos y las cepas para realizar la caracterización. Este proceso incluye la construcción de los módulos, su caracterización individual y su integración en el huésped final. Una vez construidas y verificadas las cepas, se caracterizan individualmente los diferentes módulos biológicos. Hay tres módulos: (i) un sistema de expresión optogenética para el control de la demanda de producción, (ii) un reportero de estrés, y (ii) un grupo de proteínas de interés con modificaciones que permiten el seguimiento en tiempo real de las concentraciones internas y la detección de los niveles de secreción. Posteriormente, los tres módulos se integran en el huésped final y se utilizan para caracterizar el proceso de secreción.

El segundo componente de esta estrategia consiste en los desarrollos tecnológicos aplicados a una plataforma de biorreactores para procesos continuos. Hemos ampliado una plataforma existente de biorreactores múltiples implementando capacidades optogenéticas y mediciones automatizadas con alta resolución temporal y al nivel de células individuales. Estos avances permiten realizar experimentos de control en tiempo real. Además de los componentes que forman parte de la plataforma, hemos desarrollado una metodología genérica para medir los niveles de secreción de forma sencilla y fácil de automatizar.

El tercer componente se basa en una aproximación computacional cuantitativa para integrar y analizar racionalmente los datos generados por los distintos dispositivos de medición. Aplicando un modelo matemático sencillo y derivando los valores de los parámetros a partir de los ajustes a los datos experimentales a la simulación, podemos interpretar los resultados de forma exhaustiva. Este marco computacional nos permite deducir la contribución de los componentes de eliminación de proteínas durante el proceso de secreción (secreción, dilución debido a crecimiento celular, o degradación activa) para un conjunto de cepas que secretan proteínas de diferente complejidad secretora.

La integración de estos tres componentes experimentales y computacionales nos permite desarrollar una aproximación sistemática para caracterizar el proceso de secreción de forma genérica y estándar. Realizando cultivos continuos de las cepas del huésped y aplicando diferentes niveles de demanda de producción, generamos y analizamos datos que nos permitieron descifrar patrones comunes en el proceso de secreción. Este conocimiento puede aplicarse para establecer experimentos de control de retroalimentación en plataformas de biorreactores para mejorar el proceso de optimización de las cepas para bioproducción.

Este proyecto de tesis presenta aportaciones tecnológicas, metodológicas y teóricas al campo de la biología. Desde un punto de vista tecnológico, la implementación de las mediciones de secreción utilizando *inmunobeads*, presentada en el capítulo 3, sección 3.2, es, hasta donde yo sé, el primer enfoque basado en la citometría de flujo para medir los niveles de secreción sin necesidad de separar las células de la muestra. El método tiene el potencial de ser automatizado, sin embargo, para hacerlo aún debo superar algunas limitaciones específicas de la plataforma. A diferencia de la mayoría de las metodologías para medir los niveles de secreción, nuestras mediciones basadas en *inmunobeads* son cuantitativas, sencillas y sensibles, lo que las hace compatibles con la automatización.

En cuanto a la contribución metodológica, en el capítulo 4 presento el desarrollo de un método basado en la biología de sistemas y en enfoques computacionales para caracterizar el proceso de secreción siguiendo una estrategia sistemática. Proponemos una aproximación genérica experimental y computacional para caracterizar el proceso de secreción sometido a diferentes demandas de producción. Los modelos matemáticos presentados muestran las distribuciones de los diferentes destinos, ya sea la secreción, la degradación o la dilución, de una proteína de interés en la vía secretora.

La interpretación de estos resultados también contribuye a una mejor comprensión del proceso de secreción al proporcionar una visión no estática del sistema. La contribución de la degradación, la secreción y la dilución del crecimiento se ve afectada significativamente por el tráfico de proteínas en la vía. En este sentido, estos resultados también contribuyen al campo de la bioproducción, ya que es probable que se puedan encontrar niveles óptimos de demanda de producción en el análisis de los patrones de secreción.

La mencionada estrategia sistemática de caracterización también puso de manifiesto el efecto de la población en el proceso de secreción (capítulo 4). La detección de poblaciones bimodales definidas por los niveles internos de proteínas, así como por los niveles de estrés, es relevante para entender cómo la variabilidad en la capacidad de secreción a nivel de población puede impactar en la eficiencia de la bioproducción. A partir de esta observación, propusimos y utilizamos las capacidades de nuestra plataforma de biorreactores para realizar experimentos en tiempo real con el fin de mejorar y desarrollar estrategias aplicables a los procesos de bioproducción.

Otras aportaciones de este trabajo son la puesta en marcha de la autorregulación interna, basada en biología sintética, que, aunque no se ha completado, está documentada en el Apéndice A. En ese proyecto he seguido un enfoque alternativo para adaptar la demanda a la capacidad celular y he aplicado una solución utilizando el control interno a nivel de células individuales. Otras contribuciones indirectas de mi proyecto de doctorado fueron la generación de protocolos, plásmidos y cepas de levadura utilizados en otros proyectos de nuestro equipo.

Contents

---

1 **Chapter 1: Introduction** *1*

1.1 Context and motivation 1
- *1.1.1 Bioproduction and future 1*
- *1.1.2 Small automated bioreactors to democratize bioproduction 2*
- *1.1.3 Generic solutions to simplify bioproduction 2*

1.2 Approach 3
1.3 Contribution 4
1.4 Outline 5

2 **Chapter 2: State of the art** *7*

2.1 Yeasts in bioproduction 8
- *2.1.1 Saccharomyces cerevisiae as a chassis cell 8*
- *2.1.2 Traditional yeast strain optimization approaches for protein secretion 11*
- *2.1.3 Modulation of protein production induction strengths 14*
- *2.1.4 Computational biology to study protein secretion 17*

2.2 Technological Challenge 23
- *2.2.1 Bioreactors setups for systematic data generation and control experiments 23*
- *2.2.2 Measurements of secretion levels for automated setups 24*

2.3 Conclusion 26

3 **Chapter 3: Experimental developments** 29

3.1 Experimental setup: Hardware and Bioware 30
- *3.1.1 Automated multi-bioreactors platform for yeast continuous cultures 30*
- *3.1.2 Optogenetic system to control gene expression 36*
- *3.1.3 Sensor strain to quantify effective induction levels 44*
- *3.1.4 Sensor strain to assess growth rate dynamics 46*
- *3.1.5 Maintenance of culture environmental conditions 48*
- *3.1.6 UPR reporter to quantify stress levels in real-time 49*

3.2 Measuring secretion levels in a systematic manner 52
3.3 Strain construction strategy 58
3.4 Conclusion 62

4 **Chapter 4: Characterizing secretory processes for heterologous proteins** *65*

4.1 Validation of the approach: Characterizing the secretory process for low secretory complexity proteins 67
- *4.1.1 mNeonGreen as model protein 67*

4.2 Leveraging the validated pipeline: Characterizing secretion of different secretory complexity proteins 86
- *4.2.1 endo-1,4-β-xylanase C, XylC, as protein of interest 87*
- *4.2.2 Single chain variable fragment 4M5.3, scFv, as protein of interest 89*
- *4.2.3 ∝-amylase, amylase, as protein of interest 95*
- *4.2.4 Inferring the components of protein removal process 99*
- *4.2.5 Model extension to decompose using XylC as protein of interest 103*

4.2.6 Model extension to decompose using scFv as protein of interest *104*
4.2.7 Model extension to decompose using amylase as protein of interest *105*
4.3 Integrated perspective *106*
4.3.1 Secretory patterns *106*
4.3.2 UPR analysis *108*
4.4 Using the characterization pipeline for bioproduction *112*
4.4.1 The role of accumulators in the overall secretion levels *112*
4.4.2 Feedback control experiment *114*
4.5 Conclusion *117*

5 **Chapter 5: Discussion** *119*
5.1 Thesis summary *120*
5.2 Discussion and conclusions *121*
5.2.1 Technological developments to study secretion in yeast *121*
5.2.2 Development of a systematic pipeline to infer secretory patterns *121*
5.2.3 Population effect in the secretory process *123*
5.2.4 Exploiting the results of this study *124*
5.2.5 General conclusion and take home message *125*
5.2.6 Limitations *126*
5.3 Perspectives and future directions *127*

6 **Chapter 6: Materials and methods** *131*
6.1 *Strains construction* *132*
6.1.1 DNA synthesis and constructions design *132*
6.1.2 Cloning parts and generating strains *133*
6.1.3 Genomic integration and yeast strains construction *136*
6.2 Characterization experiments *139*
6.2.1 Bioreactors experiments *139*
6.2.2 Cytometry acquisition *140*
6.2.3 Analysis of data from bioreactors and model fitting *140*
6.2.4 Secretion levels measurements *141*

7 **Chapter 7: References** *145*

A. **Annex A: Internal feedback control** *155*

B. **Annex B: Supplementary material** *159*

Acronyms

ER: Endoplasmic Reticulum

ERAD: Endoplasmic Reticulum-Associated Degradation

IF: Internal Fluorescence

IF: Internal Fluorescence (Model variable)

POI: Protein Of Interest

SL: Secretion Levels (Model variable)

UPR: Unfolded Protein Response

Chapter 1: Introduction

1.1. Context and motivation

1.1.1. Bioproduction and future

During the last century, manufacturing based on chemical industry has notably contributed to the development of almost every aspect of our modern societies. Production of materials, fuels, food, and pharmaceuticals, among many other goods of interest, has been crucial to fulfill the increasing consumption of our civilization. Nowadays, excessive industrialization has impacted and degraded the environmental balance. Societies need to develop sustainable alternatives that keep the worst-case scenarios of climate change projected for the 21st century (Kim & Worrell, 2002; Kitney, Bell, & Philp, 2021). Today, on the verge of an imminent fourth industrial revolution, bioproduction is playing an important role in the development of environmentally sustainable industrial production (Carvalho, Chaim, Cazarini, & Gerolamo, 2018; Yu, Wu, & Chen, 2019).

Bioproduction, bioprocessing or Biomanufacturing, is a field of applied biology which aims to produce goods of interest using biological systems as a factory or catalyzers. Examples of production platforms include: microorganisms, single eukaryotic cells, enzymes, or even cell free systems based on cell lysates. Bioproduction has huge potential for improvement and adaptability thanks to the advances in research disciplines such as biological engineering, synthetic biology, systems biology and computational biology, among others (Camarasa *et al*, 2018).

To produce the majority of the goods of interest, cell factories synthetize proteins, that are either the desired final products, or catalyzers to synthesize the good of interest (Mattanovich *et al*, 2014). One common example of product of interest produced via cell factories is the commercial human insulin synthesized in *E. coli*. That was marketed by *Eli Lilly* in 1982, becoming the first biopharmaceutical product. Since then, the biotechnology industry has matured and its market has grown and now its size is counted by hundreds billions of dollars (Walsh, 2018). This industry aims to become one of the most important in the renewables technologies for the 21st century.

1.1.2. Small automated bioreactors to facilitate bioproduction

In the biomanufacturing process, cell factories are generally grown in bioreactors. These platforms supply everything necessary for the cells to grow and produce the goods of interest. In the contemporary bioproduction industry, there are two main production platforms. On the one side there are the batch bioreactors. They are currently largely used by industrial firms. In them, whole cell batches are produced at the same time, and must be finished before the next can be started. On the other side one can find the continuous cell culture operations, which are becoming more convenient nowadays. They require smaller equipment, and can be scaled based on time instead of volume (Croughan, Konstantinov, & Cooney, 2015; Kumar, Udugama, Gargalo, & Gernaey, 2020). Continuous cultures have potential for setting small-scale bioproduction plants, and spreading manufacturing around the world. That would bring production closer to consumption agents and contribute to democratize bioproduction processes.

Along these lines, small-scale, low-cost bioreactors are emerging as platforms for research. The capacity to precisely control culture conditions, maintain steady states over long durations, or apply dynamic perturbations, enables new research approaches to improve bioproduction process. To achieve such goal, sensing and interacting with the system in real time is essential, and it requires the use of robotics. In addition, modeling and simulations are becoming key factors in understanding biological processes. This understanding, along with bioreactor functionalities, can be used to implement real-time control in the biological system to obtain optimal performance in bioproduction. However, assembling together all the measuring and controlling devices with a set of continuous culture bioreactors is a major challenge. These technological developments together with standardization of processes could yield major contributions for implementation and democratization of small-scale manufacturing.

1.1.3. Generic solutions to simplify bioproduction

In bioproduction, purification of a product has become one of the most important limiting factors when working in continuous operations (Ramos-de-la-Peña, González-Valdez, & Aguilar, 2019; Beal et al., 2020). When producing heterologous protein of interest, secreting them is convenient to facilitate downstream purification process (Romanos, Scorer, & Clare, 1992; Kahl, 2015). Yeasts are a common eukaryotic cell factory for bioproduction, notably because its capacity to secrete complex proteins (Borodina & Nielsen, 2014; Love et al., 2018). However, the secretory biology of eukaryotes is complex and regulated in a multi-factorial manner.

It has been observed that the secretion levels of the protein of interest (POI) are not always correlated with production demands. In the secretion process, high production demands may lead to excessive protein trafficking through the secretory pathway, resulting in bottlenecks, and eventually secretion-associated stress. The endogenous adaptive system of the cells overcomes the problem by degrading the proteins accumulated, including the proteins of interest (Campbell *et al*, 2017).

Typically, the way bioengineering approaches avoid bottlenecks along the secretory pathway is based on product specific solutions. Despite its success, this approach lacks generality, since the bottleneck for each product may appear at many different stages of the process. Generic approaches relying on standard system characterization procedures and smart regulation are needed to reduce time and resources, simplifying the strain optimization. Standardization is considered a fundamental feature of technological advancement (Beal *et al*, 2020), it played a critical role in the past industrial revolution. The development of generic and systematic approaches to minimize trial-and-error based processes for component characterization will likely be of great importance to dive into the next revolution.

1.2. Approach

According to the previous section, in the field of bioproduction it is necessary to develop generic strategies to characterize the secretory system. Such approaches should be standard and systematic to overcome the inefficiencies of the typical trial-and-error process for strain optimization. In the following chapters I will present the development of a systematic pipeline to characterize the secretory process. The approach integrates three different components to establish a conceptual framework for reporting on bioproduction processes involving protein secretion.

The first component relies on a rational way for constructing the biological modules and strains to carry out the characterization. This process includes the modules construction, their individual characterization, and their integration in the final host (Weber, Engler, Gruetzner, Werner, & Marillonnet, 2011; Lee, DeLoache, Cervantes, & Dueber, 2015). Once the strains have been built and verified, the individual biological modules were characterized individually. There are three modules: (i) an optogenetic expression system for the control of production demand, (ii) a stress reporter, and (ii) a group of proteins of interest allowing real-time monitoring of internal concentrations and detection of secretion levels.

Then, the three modules have been integrated in the final host that is used to characterize the secretory process.

The second component of the approach consist in the technological developments applied to a continuous bioreactors setup. We have expanded a multi-bioreactors platform by implementing optogenetic capabilities and automated measurements with high temporal and single-cell resolution. These developments together enable for real-time control experiments. In addition to the components that are part of the platform, we have developed a generic methodology to measure secretion levels in a simple way and allowing for automation.

The third component is based on a computational framework to integrate and analyze in a rational manner the data generated from different measurements devices. By implementing a simple mathematical model and inferring parameter values from fits to data, we can interpret the results in a comprehensive manner. This computational framework allowed us to infer the contribution of the protein removal components of the pathway (secretion, dilution due to growth, or active degradation) for a set of strains secreting proteins with different secretory complexity.

The integration of these three experimental and computational components together allows to develop a systematic pipeline to characterize the secretory process in a generic and standard manner. By performing continuous cultures of the host strains, and applying different production demands, we generated and analyzed data that allowed us to decipher common patterns in the secretory process. Then, this knowledge can be applied to set up feedback control experiments on the bioreactor platform to improve the optimization process of bioproduction strains.

1.3. Contribution

This thesis project presents technological, methodological and theoretical contributions to the field of biology. On the technological side, the implementation of secretion measurements using immuno-beads, presented in Chapter 3, section 3.2, is to the best of my knowledge, the first flow cytometry-based approach to measure secretion levels without the need to separate cells from sample. The method has the potential to be automated, however, to do so I first have to overcome a few platform-specific limitations. Unlike most methodologies for measuring secretion levels, bead measurements are quantitative, simple and sensible, making them compatible with automation.

Regarding the methodological contribution, I present in Chapter 4 the development of a method based on systems and computational biology approaches to characterize the secretory process in a generic manner. We propose a systematic experimental and computational pipeline to reconstruct the secretory pattern of a strain for different production demands. These patterns show the distributions of different fates, either secretion, degradation or dilution of a protein of interest in the secretory pathway.

The interpretation of these secretory patterns also contributes to the better understanding of the secretory process by proposing a non-static view of the pathway. The contribution of degradation, secretion and dilution due to growth is significantly affected by protein trafficking through the pathway. Along these lines, these results also contribute to the field of bioproduction, since it is likely that optimal levels of production demand can be found in the analysis of the secretion patterns.

The aforementioned pipeline of characterization also shed light on the population effect on the secretion process (Chapter 4). The detection of bimodal populations defined by internal protein levels as well as stress levels is relevant to understand how variability in secretory capacity at the population level can impact bioproduction efficiency. From this observation, we proposed and utilized the capabilities of the bioreactor setup to perform real-time experiments to improve and learn about bioproduction processes.

Other contributions of this work are the implementation of internal feedback, although not completed. It is documented in Annex A. In such project, I followed an alternative approach and partially implemented a solution using internal regulation. Other indirect contributions of my PhD project have been the generation of numerous protocols, plasmids and yeast strains used in other projects of our team. Notably, in the development of *MicroMator* (Fox, *et al*, 2021).

1.4. Outline

The thesis document is organized in six chapters. This chapter is an introductory overview of the general aspects of bioproduction and its potential impact in the coming years. We also presented the strategy followed to address some of the limitations in the field and the main contributions of my work. The goal in Chapter 2 is to pave the ground for the following two chapters. In Chapter 2, we emphasize specific constraints and obstacles for developing generic approaches to characterize the secretion process in the context of bioprocess optimization. In Chapter 3, I present the experimental developments that have been implemented to carry out this project. It covers the technological, methodological and biological implementations necessary to characterize secretion processes. This chapter is rich in technical and methodological details that are important for understanding how the data are produced and analyzed in Chapter 4.

The main results are presented in Chapter 4. These are the systematic pipeline to characterize the secretion process, the results of applying this pipeline to strains that secrete proteins with increasing secretion complexity, the characterization of secretion dynamics at the population level, and real-time feedback control results to potentially improve secretion efficiency. Chapter 5 summarizes the main conclusions of the thesis and I discuss the potential and limitations of the contributions. Finally, Chapter 6 compiles all the materials and methods used in the project, together with the lists of plasmids and strains produced. In Annex A, it is documented the preliminary results for the implementation of an internal feedback based on interference RNA. Annex B is a supplementary material chapter.

Chapter 2: State of the art

As discussed in the previous chapter, protein secretion involves coupled multistep processes that result in complex cellular dynamics. Heterologous protein secretion often challenges the secretion system, triggering stress responses and adaptive mechanisms aimed at restoring cellular physiology. These adaptive processes may ultimately affect the production yield. (Buchberger, Bukau, & Sommer, 2010; P. Xu & Robinson, 2009; de Ruijter, Koskela, Nandania, Frey, & Velagapudi, 2018). In recent decades, impressive efforts and advances have been made to improve the production of secreted heterologous proteins by engineering the secretory pathway (M. Huang, Wang, Qin, Petranovic, & Nielsen, 2018; Thak, Yoo, Moon, & Kang, 2020). However, a more complete understanding of how different biological networks interact to achieve the maintenance of cellular homeostasis would help to develop strategies that increase secretory efficiency in different contexts, such as industrial production processes. In this chapter, I will outline and discuss traditional and state of the art methodologies to improve the performance of heterologous protein secretion, as well as the limitations and needs that should be addressed to develop generic production approaches.

This chapter is divided into two sections. First, the strain engineering approaches for bioproduction and some relevant computational models for studying yeast secretory pathway are presented. In the second section I discuss the technological advances and challenges to perform systems biology studies in the yeast secretory pathway. The first section begins with an introduction to the broad concept of chassis cells and *S. cerevisiae* as model organism in bioproduction. We will continue by reviewing some of the traditional approaches for strain engineering in the context of the secretory pathway. Then, complementary methodologies of a more general scope, such as protein production demand modulation, will be discussed. To introduce the perspective of quantitative biology, the results of three computational models of the secretory pathway are presented and discussed. As for the second section, on technological challenges, I will list some of the latest advances in small-scale bioreactors and their role and shortcomings to study the he secretion process in a systematic manner. In the same vein, the need for development of simpler methodologies to perform measurements of secretion levels *in situ* is also discussed.

2.1. Yeasts in bioproduction

2.1.1. Saccharomyces cerevisiae as a chassis cell

Traditionally, the goods of interest produced by biological systems have been obtained by directly using native host organisms, for instance, the yeast *S. cerevisiae* has been widely used for beer, wine or bread production from the dawn of civilizations to the present time. However, in many cases it turns out that the native organism is not the optimal cell factory for bioproduction in the industrial context (Xu *et al*, 2020). For instance, in the production of commercial human insulin, it is not optimal to synthesize the product in its natural host, the human beta cells, but by engineered *E. coli* with the transgene coding for the human insulin (Baeshen *et al*, 2014). The reasons are often that the native host is not adapted to the conditions and requirements for industrial production, as well as that the lack of manipulation techniques for that specific organism. Due to the aforementioned reasons, the development of so-called chassis organisms has been an important objective in bioproduction.

The concept of a chassis cell refers to a microbial strain that provides a biological background for the expression of heterologous proteins or pathways in industrial applications. An ideal chassis cell would be a unique and versatile cellular platform capable of producing any product by simply introducing the genes encoding the required pathway (S. Y. Lee & Kim, 2015). Instead, such foresight is probably impossible, protein production is a multistep process that requires global modulation of cell metabolism, and its fine-tuning optimization is product-specific (Huang *et al*, 2017). The closest alternative to the ideal chassis cell would be a cellular platform optimized for the production of a wide range of products of interest with similar strain background requirements (Figure 2.1). However, although the interest in this near-ideal chassis cell concept is attracting increasing attention, and recent advances in omics approaches have provided unprecedented insights into host physiology, relatively few strains actually reach the industrial arena. There are still technological challenges to overcome that ultimately translate into economic cost due to the labor and time required to generate such strains (Lee & Kim, 2015).

Currently, the organisms most commonly used as the proxy of the chassis cell are *E. coli*, various yeast species, and mammalian cells. These cell factories produce almost all commercial therapeutic proteins (Walsh, 2018). Compared to *E. coli*, yeasts used in bioproduction possess a more complex and complete post-translational processing machinery. In addition, compared to mammalian cells, yeasts have several advantages, such as the ability to grow in inexpensive media, and a tradition of well-established industrial production technologies. Among the yeasts used for bioproduction, *Saccharomyces cerevisiae* is one of the most studied and developed. In addition, it has several advantages as a host: it is generally recognized as safe (GRAS), there is a wide range of established technologies for engineering its metabolism, and it natively possesses a number of characteristics favorable for industrial fermentation. In fact, it has already been used intensively in a wide range of industrial processes, from pharmaceuticals or bioethanol production to beer, wine or bread. (Nielsen & Jewett, 2008; Borodina & Nielsen, 2014).

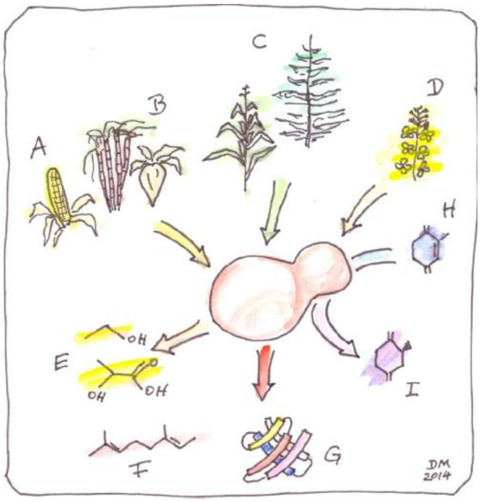


Figure 2.1. ***Yeast as a cell platform for bioproduction***, from Mattanovich, Sauer, & Gasser, (2014). An ideal chassis cell is capable of employing different carbon sources to perform various bioprocesses for industrial purposes. The substrate is converted into products of interest by engineering heterologous metabolic pathways within the endogenous cellular network. A-D: Corn starch, cane sugar, lignocellulose, crude glycerol. E-H: Primary and secondary metabolites, and recombinant proteins.

The secretory biology of *S. cerevisiae* functions similarly to that of other eukaryotic hosts (Papanikou & Glick, 2009; Buchberger, Bukau, & Sommer, 2010). In brief, protein secretion involves the translocation of newly synthesized peptides from the cytosol to the endoplasmic reticulum (ER), followed by the appropriate protein folding and modifications, and finally vesicle-mediated transport to the external milieu (Figure 2.2). These steps are part of a complex system that precisely controls the structural quality of proteins and adapts the cellular proteome to the given needs. Nonetheless, when protein production is very demanding, the secretory machinery becomes overwhelmed and is unable to handle an excessive influx of proteins to process. This situation can result in undesired degradation of the target product instead of its efficient secretion (Buchberger *et al*, 2010), and subsequently a decrease in target protein yield. To overcome these capacity deficiency, many efforts have been made to engineer the secretory pathway with varying successes (Kahl, 2015).

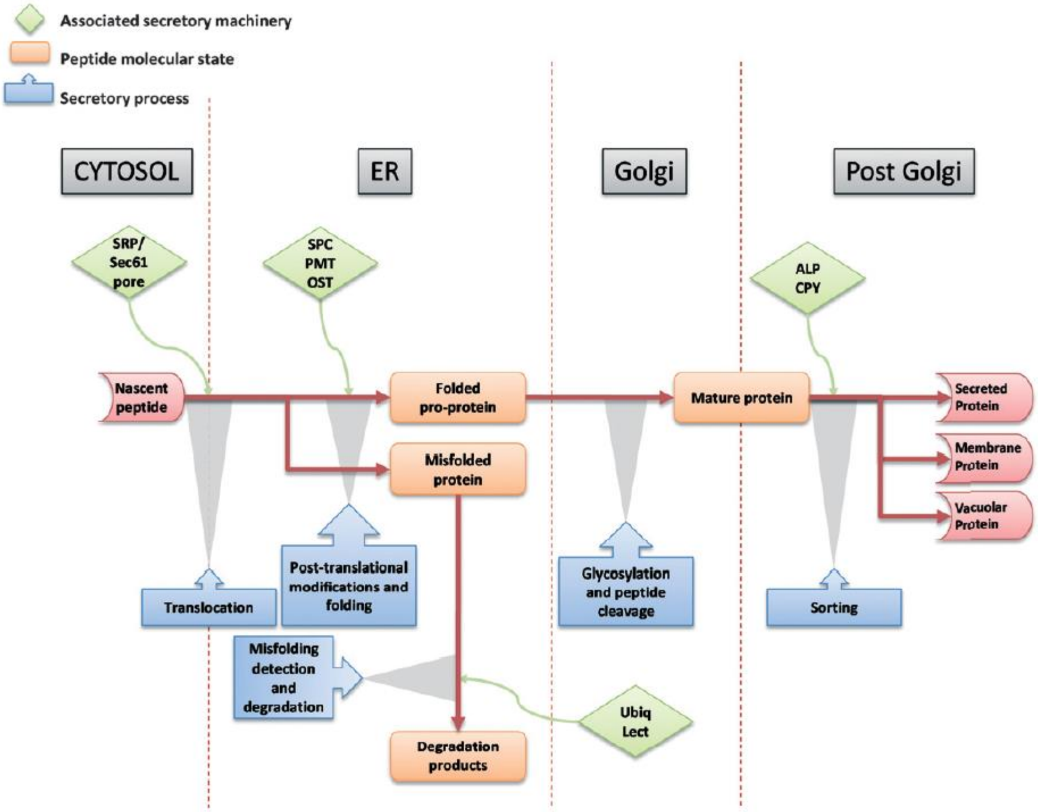

Associated secretory machinery
Peptide molecular state
Secretory process
CYTOSOL
ER
Golgi
Post Golgi
SRP/ Sec61 pore
SPC PMT OST
ALP CPY
Nascent peptide
Folded pro-protein
Misfolded protein
Mature protein
Secreted Protein
Membrane Protein
Vacuolar Protein
Translocation
Post-translational modifications and folding
Glycosylation and peptide cleavage
Sorting
Misfolding detection and degradation
Ubiq Lect
Degradation products

Figure 2.2. ***Overview of the secretory pathway,*** from Hou, Tyo, Liu, Petranovic, & Nielsen (2012). After or during translation the peptide is translocated into the endoplasmic reticulum (ER). There, it is folded to achieve its native conformation required to pass the quality control, otherwise it is identified and targeted for degradation. Once folded and mature, the protein is shorted to its final location by vesicle-mediated transport, either to the membrane, to the vacuole, or to secretion. SRP, signal recognition particle; SPC, signal peptidase complex; PMT, protein O-mannosyl transferase; OST, oligosaccharyl ransferase; Ubiq, ubiquitin; Lect, Lectin; ALP, arginine transporter pathway; CPY, carboxypeptidase Y pathway.

To conclude on what has been mentioned in the previous paragraphs, the implementation of a chassis cell for producing proteins of industrial interest has been a goal for decades in the field of bioproduction. The yeast *Saccharomyces cerevisiae* has shown enormous potential to be exploited as model organism for heterologous protein and pathway expression in eukaryotes (Dymond & Boeke, 2012; Jouhten et al., 2016; Chi et al., 2019). In fact, it has already firmly established itself as a cell factory in many aspects of industrial biotechnology (Darby, Cartwright, Dilworth, & Bill, 2012). However, despite reported the efforts and impressive progress improving heterologous protein production in *S. cerevisiae*, there is still a need to gain more fundamental insight into the key processes of the biological system, and specifically into the protein secretion pathway in order to implement rational engineering more efficiently.

2.1.2. Traditional yeast strain optimization approaches for protein secretion

It is known that high-level demands of heterologous protein expression often cause physiological burden by interfering with homeostatic processes in different microbial hosts (Glick, 1995; Jordà et al., 2014; Klein et al., 2014; Ceroni, Algar, Stan, & Ellis, 2015; Kafri, Metzl-Raz, Jona, & Barkai, 2016; M. Huang et al., 2017). Such burden is often associated to resource reallocation into processes related to heterologous protein production. In addition, when referring specifically to production of secreted heterologous proteins in yeast, even at moderate expression levels, physiological burden is also linked to the presence of bottlenecks on the different stages of the secretory pathway (Romanos *et al*, 1992). Such bottlenecks are often due to the saturation of the secretory machinery, and it negatively affects cell growth and production yields. When the secretory cellular machinery is overwhelmed, protein accumulation occurs and stress responses such as the unfolded protein response (UPR) are triggered (Figure 2.3). UPR induces many different adaptation mechanisms that affect the secretory pathway in a complex manner (Buchberger *et al*, 2010).

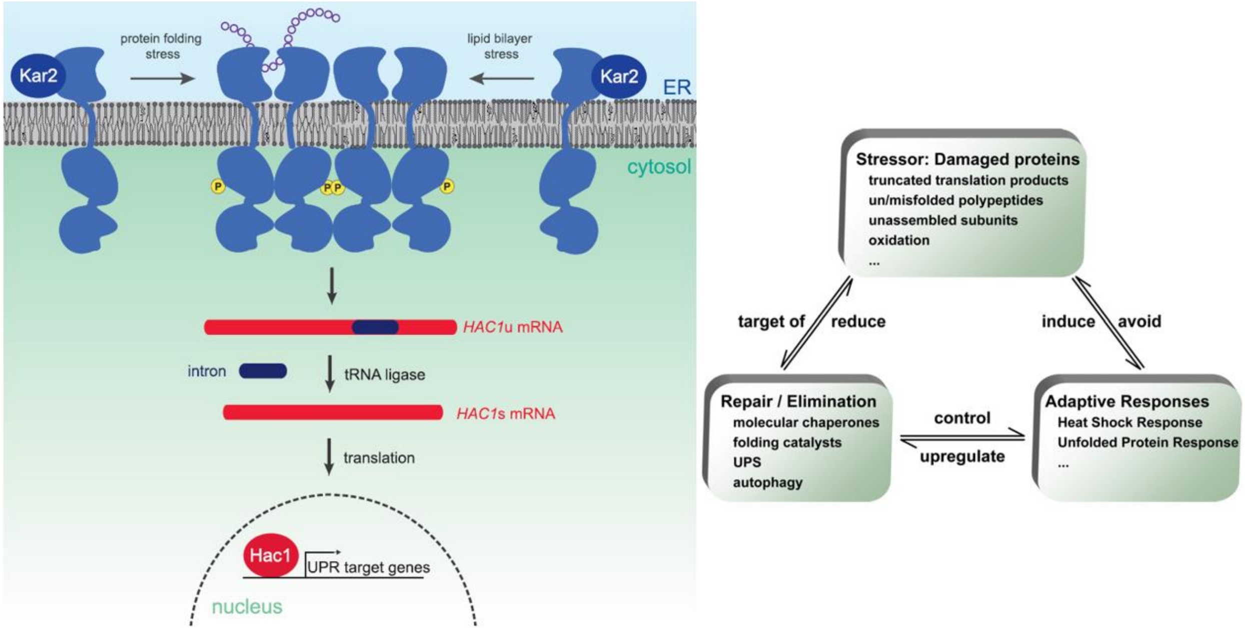

protein folding stress
lipid bilayer stress
Kar2
Kar2
ER
cytosol
P
PP
P
HAC1u mRNA
intron
tRNA ligase
HAC1s mRNA
translation
Hac1
UPR target genes
nucleus
Stressor: Damaged proteins
truncated translation products
un/misfolded polypeptides
unassembled subunits
oxidation
...
target of
reduce
induce
avoid
Repair / Elimination
molecular chaperones
folding catalysts
UPS
autophagy
control
upregulate
Adaptive Responses
Heat Shock Response
Unfolded Protein Response
...

Figure 2.3. ***Activation of the unfolded protein response (UPR) and associate adaptive mechanisms***, adapted from Buchberger et al., (2010); Halbleib et al., (2017). To the left, the scheme depicts molecular mechanisms for UPR activation. (i) Excessive accumulation of unfolded proteins in the ER lumen, as well as unbalanced composition of the lipid bilayer, induce Ire1 (Blue transmembrane proteins) oligomerization. (ii) The oligomerization of Ire1 triggers the activation of its cytosolic kinase and RNase domain. (iii) IRE1's RNase domain mediates the unconventional splicing of the HAC1 mRNA. (iv) Upon translation of the spliced HAC1 mRNA, the transcription factor Hac1 activates UPR target gene expression. To the right, the diagram relates the causes and consequences of the UPR activation. UPR produces the upregulation of hundreds of genes involved in many secretory functions to overcome stressful conditions. One of the main roles of UPR is counteracting the unfolded protein accumulation by enhancing the ERAD pathway to increase unfolded protein degradation. Kar2: Endoplasmic reticulum resident chaperone. UPS: Ubiquitin proteasome system.

Bottlenecks arise along the secretory pathway in various ways for different proteins, depending on the specific requirements of each protein. In the context of strain engineering for heterologous protein secretion, given that distinct secretory bottlenecks exist even for the same POI (Figure 2.4), alleviating one in the early stages of the pathway may not affect the yield, as others may appear downstream, and *vice versa* (Wang *et al*, 2017). Moreover, the adaptive dynamic mechanisms driven by the UPR may impact the behavior of the system in a complex manner. For instance, giving rise to ambivalent responses, that increase unfolded protein removal by degradation, and at the same time enhancing the folding capacity.

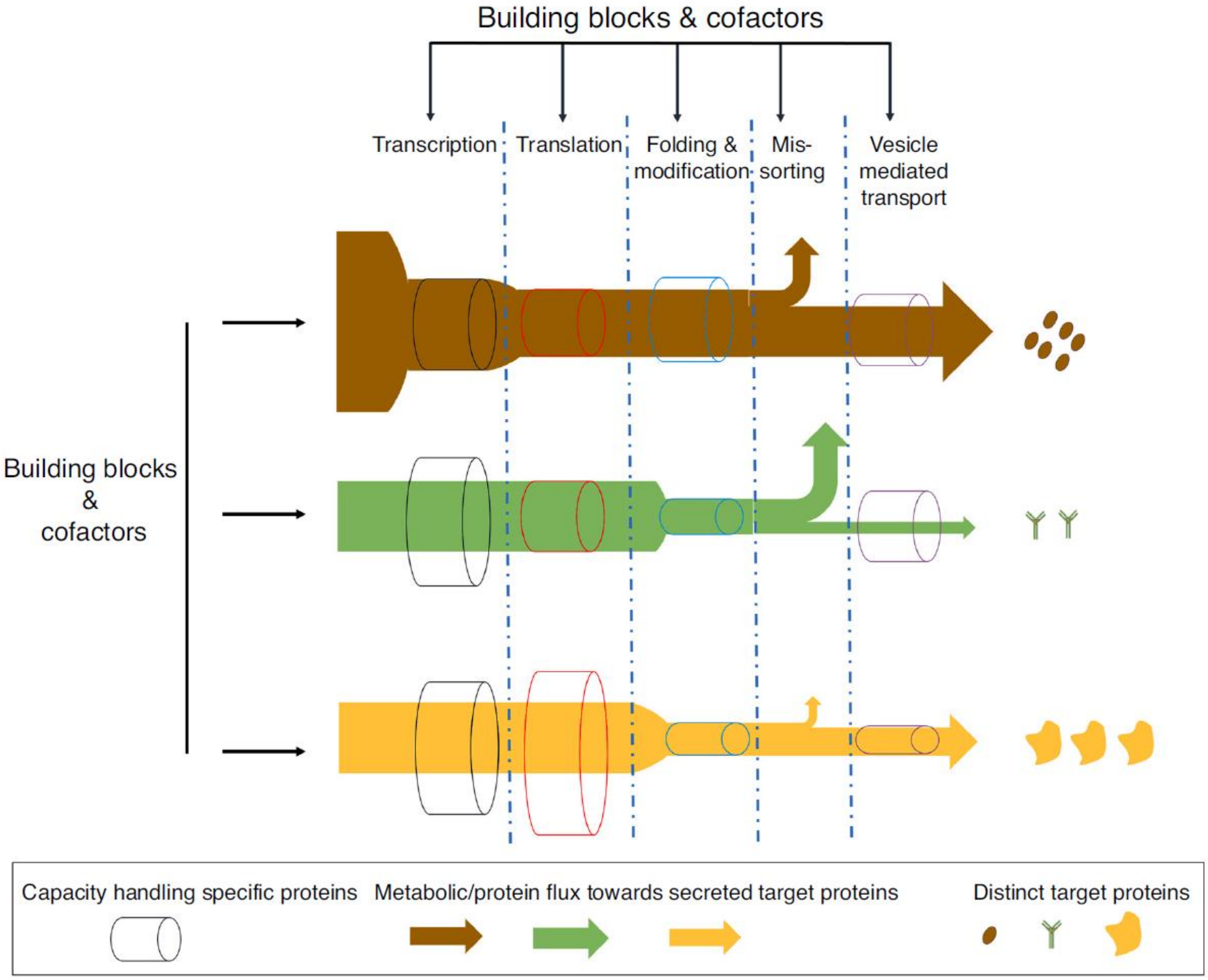

Building blocks & cofactors
Transcription
Translation
Folding & modification
Mis-sorting
Vesicle mediated transport
Building blocks & cofactors
Capacity handling specific proteins
Metabolic/protein flux towards secreted target proteins
Distinct target proteins

Figure 2.4. **Bottlenecks along the secretory pathway for different products**, from Wang et al., (2017). Native constraints limit the production of heterologous proteins at different stages of the secretory pathway. Even if stage-specific bottlenecks are addressed, others may occur downstream and constrain the secretion efficiency.

Bottlenecks in the secretory pathways are often resulting from events related to the folding properties of the POI. Based on this knowledge, the main approaches tackle the problem by addressing specific steps in the folding process. One common solution is overexpressing a particular molecular chaperone that performs a specific posttranslational modification (Delic *et al*, 2014). Unfortunately, these type of *ad hoc* strategies to optimize protein secretion processes are dependent on the protein of interest and they don't work in a generic context. Thus, while engineering a specific secretory process may improve the productivity of a given heterologous protein, this may not work to enhance the production of other proteins. For example, overexpression of the key ER chaperone Kar2 in *S. cerevisiae* improves the extracellular activity of β-glucosidase, but overexpression of the same chaperone in a strain expressing secreted α-amylase shows reduced enzymatic activity (Figure 2.5) (Tang *et al*, 2015). In addition to this one, there are numerous reports of host modification improving secretion for one protein of interest, but those modifications turn out ineffective or even detrimental for others (Harmsen, Bruyne, Raué, & Maat, 1996; Thak, Yoo, Moon, & Kang, 2020).

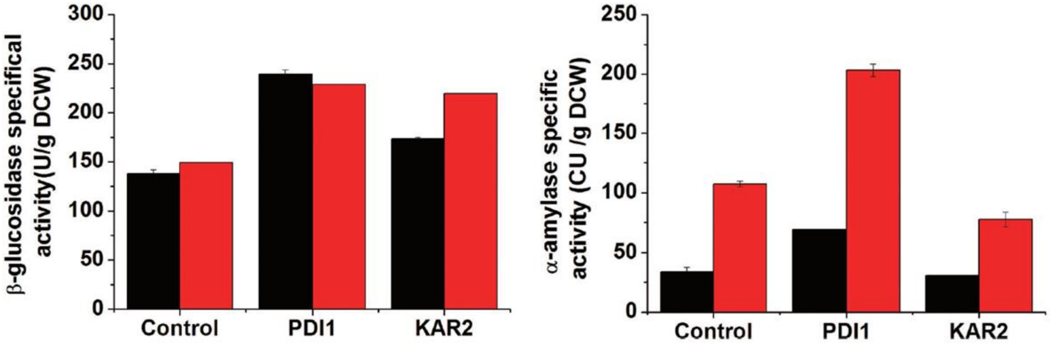


*Figure 2.5. **Effect of*** **Pdi1** ***or*** **Kar2** ***chaperones overexpression in extracellular specific activity of heterologous secreted proteins***, from Tang et al., (2015). Black bars represent the enzymatic activity levels in the supernatant after ER chaperon overexpression 36h and red lines 72h.

In conclusion, protein secretion is a complex process wherein proteins undergo a strict folding-based quality control. Excessive trafficking through the secretory pathway produces an accumulation of proteins, triggering stress responses aimed at activating adaptive mechanisms that interact dynamically with the pathway capacity. Since each protein has its own structural characteristics, the native cellular machinery may saturate when secreting heterologous proteins at high levels, resulting in bottlenecks at different stages of the pathway that are protein dependent. Those obstacles make the typical host optimization framework for bioproduction purposes an *ad hoc* process based on trial-and-error, and therefore, inefficient, slow and costly. Thus, the development of a chassis cell for generic production of secreted heterologous proteins is an enormous challenge. These limitations in turn open the door to explore generic approaches focused on matching the strength of the production demand to the capacity of the pathway.

2.1.3. Modulation of protein production induction strengths

The impact produced by protein overexpression is difficult to predict, and it is known as protein burden or cost effect (Ceroni, Algar, Stan, & Ellis, 2015; Eguchi et al., 2018). One area rarely explored to optimize protein secretion is the quantitative impact of the demand strength in the efficiency of the process. As mentioned previously, overexpression of secreted protein may be detrimental to cellular homeostasis, and it could depend on the specific functional and structural characteristics of the POI. Excessive flux through the secretory pathway overwhelms the system and saturates its capacity. Consequently, proteins accumulate, producing stress and eventually activating adaptive responses that lead to POI degradation, thus decreasing the efficiency of bioproduction (Buchberger *et al*, 2010).

To avoid saturation of the host's secretion capacity by pathway overload and accumulation of unfolded proteins, and consequent bottlenecks, the protein trafficking should be adapted through the protein production demand. In this sense, promoter strength is known to be a central component for efficient production. To modulate gene expression at different levels and activate it at different times there are several options. On the one hand there are constitutive promoters such as pTDH3, pRPL18B, pREV1. These promoters continuously express the protein of interest at a constant level if the physiological state of the cell remains constant as well. On the other hand, inducible promoters trigger expression of the gene of interest upon a specific controlled change in environmental conditions (Figure 2.6).

There are two families of inducible promoters that dominate bioproduction processes in *S. cerevisiae* (Kahl, 2015). The copper-controlled promoters, those that trigger expression upon increasing copper concentration in the medium. And galactose-controlled promoters, which control promoter activity as a function of the galactose/glucose ratio. None of these controllers are ideal for large-scale bioproduction due to their weak activity at maximal levels of induction, and lack of precision in controlling expression (Kahl, 2015; Escalante-Chong et al., 2015; Sheng, Flick, & Feng, 2017). Moreover, their reliance on small chemical agents, such as galactose or copper, limits their precise and rapid capabilities for dynamic control. Chemical inducers, such as copper or sugars, must diffuse into the medium to reach the target, and be diluted for their removal, implying slow activation and deactivation. In addition, other inconvenient is their potential off-target effect on the cellular system, for example, the very presence of galactose, might influence many other metabolic pathways in a complex manner (Ideker *et al*, 2001).

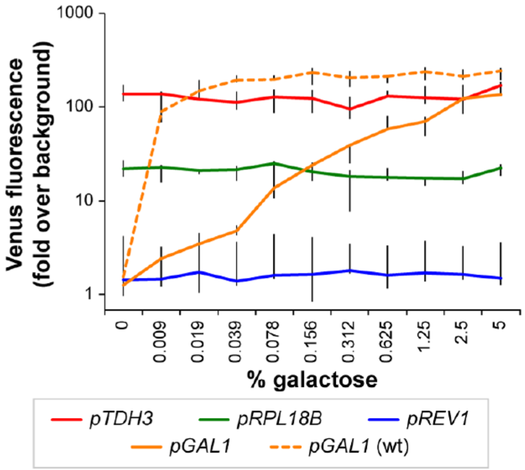


Figure 2.6. ***Characterization of constitutive and inducible promoters***, from M. E. Lee, DeLoache, Cervantes, & Dueber, (2015). Characterization of constitutive and inducible promoters expressing Venus fluorescent reporter protein in *S. cerevisiae*. Constitutive promoters with different expression strength do not change expression levels at different galactose concentration. Galactose induction of the two pGAL1 variants increases from background levels to their highest expression levels with increasing concentration of galactose. The dots represent the mean value of four biological replicates, and the error bars show the range.

Other alternatives for inducible gene expression are light-controlled transcriptional factors, which act as molecular promoter activators driven by light inputs (Figure 2.7). In general, they consist of a variety of photosensitive proteins that have been designed to provide light-driven regulation of gene expression. They are included within the methodologies known as optogenetics (Salinas, Rojas, Delgado, Agosin, & Larrondo, 2017). The term optogenetics has been used to describe methods that use light to modulate molecular interactions in living systems. In this project we will use and refer to optogenetics in the sense restricted to the control of gene expression. Using light at a specific wavelength as an inducer offers many advantages. It is fast and generally harmless, as it is minimally invasive to cells. In addition, in comparison with chemical inducers, light is an orthogonal tool in the case of yeast hosts, and cheap, thus serving as a convenient alternative. The tools most associated with gene expression optogenetic systems are dimers formed or dissociated in the presence of light at a certain wavelength. Then, by fusing them with a DNA-binding domain, they interact with the target promoter sequence and activate transcription of the target gene through the action of an also fused activation domain.

There is a wide repertoire of optogenetic devices for gene expression. These are activated by a broad light spectrum, from red to UV light-inducible photoreceptors (Salinas, Rojas, Delgado, Agosin, & Larrondo, 2017). Although optogenetic tools have had significant use in research, their direct application in industrial processes has been minimal, if used at all. However, it is a tool with potential beyond the laboratory, and has already been used for the optimization of strains and approaches intended for industrial purposes

(Zhao et al., 2018; Lalwani, Kawabe, Mays, Hoffman, & Avalos, 2021; Moreno Morales, Patel, Stewart, Sweeney, & McClean, 2021; Lovelett et al., 2021).

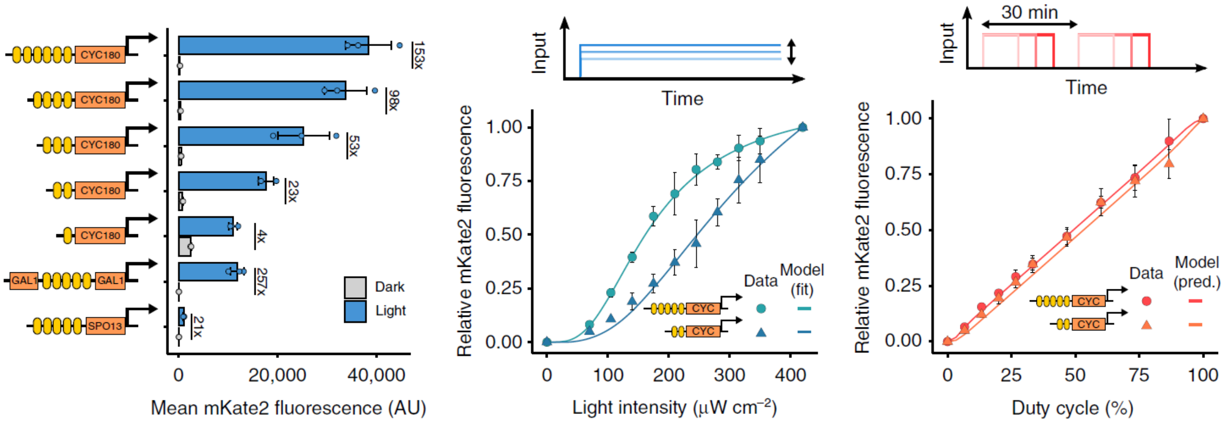


Figure 2.7. ***Characterization of the EL222-based optogenetic expression system in S. cerevisiae***, from Benzinger & Khammash, (2018). Left: Promoter library of different copies of the DNA-binding sequence for the EL222-based transcription factor (yellow boxes), the fluorescence of the reporter mKate2 is measured after 6h in the dark or in the presence of blue light. Center: dose-response relation of two promoters induced by different light intensities. Right: dose-response relation of two promoters induced by different duty cycle, also called pulse-width modulation. Duty cycle represents the percentage of light exposure time in a period of 30 minutes.

Another way to control gene expression levels that is compatible with and additional to those discussed above is by controlling gene copy number. Gene construct cassettes for heterologous proteins are introduced into the yeast genome by two main strategies, either by plasmid-based (episomal) systems or by chromosomal integration. In the first case, plasmid or episomal systems, the cassette is independent of the chromosomes, such process is more prone to produce copy number variability, as plasmid copies might segregate unevenly between daughter cells after division. In addition, plasmid copy number maintenance can produce selection pressure, leading to genetic instability and varying gene copy number. (Li *et al*, 2019). These facts significantly affect the performance in large-scale and long-term industrial bioproduction, therefore, genetic integration into the host chromosome, the second strategy, might be a more convenient for large-scale approach.

Chromosomal integration can be carried out by either non-homologous, or homologous recombination. In the first case, the recombination is not targeted, and might produce undesired effects, such as disruption of a random gene or alteration of its expression. Moreover, the number of gene copies integrated is not known *a priori*. On the other hand, for those integrations based on homologous recombination, the cassette integrates at a known locus, usually in the location of a gene previously deleted. Therefore, it does not affect host physiology, and allows selection marker integration for phenotypic screening of positive clones. The integration can be done at multiple loci in successive steps, allowing to control the gene copy number added to genome. (Thak *et al*, 2020). Depending on the number of copies of the integrated gene, this will result in various levels of gene expression. However, the correlation between gene copy number and protein yield is not necessarily linear.

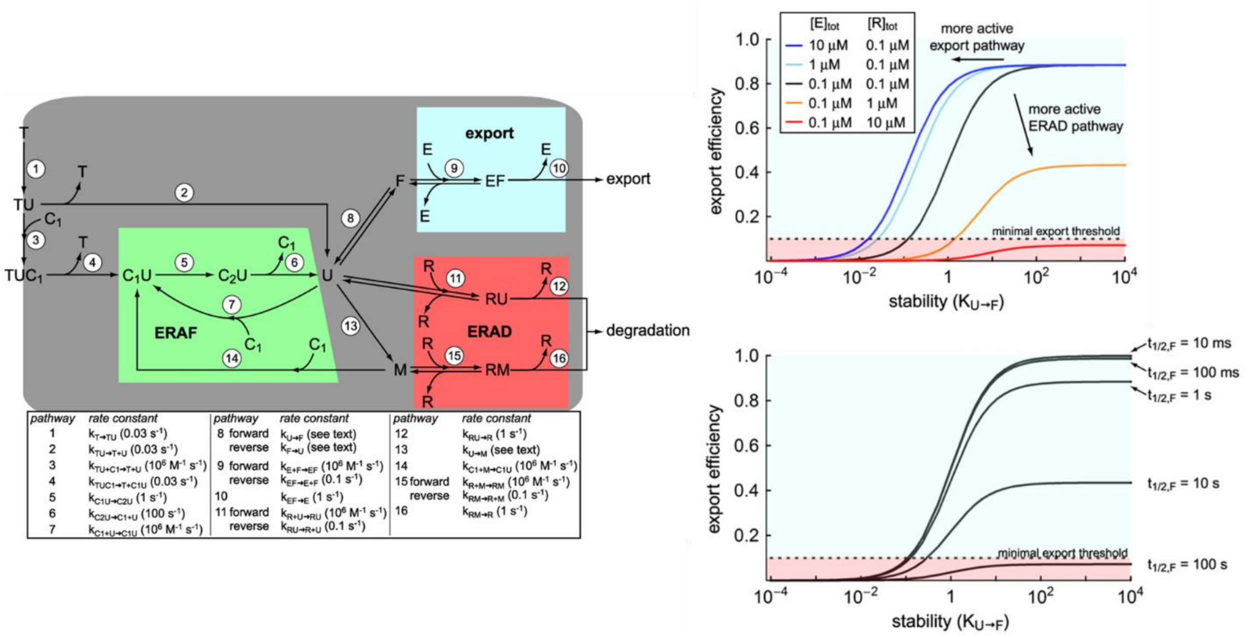
As mentioned, there are several ways to approach the control of gene expression levels. Constitutive promoters provide constant levels, and inducible promoters allow expression levels to be modulated dynamically. On top of that, gene copy number sets a maximum expression level that correlates with copy number to some extent. This is important because gene expression strength is a key parameter to optimize production of a given POI. However, although increasing the expression levels is apparently desirable for optimizing protein production, the occurrence of an optimal level of gene expression seems to be a feature for heterologous proteins (Ceroni *et al*, 2018), especially when referring to secreted ones (Wittrup, Robinson, Parekh, & Forrester, 1994; R. N. Parekh & Wittrup, 1997). Different studies indicate that the optimal level of production demand exists because the secretory pathway becomes saturated with an excessive protein flux. This is probably because the excess of unfolded protein accumulated in the ER leads to degradation by the ERAD. This hypothesis has motivated numerous studies aimed at modulating the regulation of UPR-associated genes, such as chaperones, or even the activation of the pathway itself. (Harmsen et al., 1996; Valkonen, Penttilä, & Saloheimo, 2003; Tang et al., 2015; Thak et al., 2020). In contrast, quantitative studies to characterize the impact of various levels of production demand on secretory efficiency remains surprisingly limited. Controlling gene expression levels to avoid saturation of the secretory pathway may be important for maximizing protein secretion levels in eukaryotic organisms. However, it is necessary to obtain an integrated view of the secretory pathway system to understand the biological reason for such behavior. With the use of mathematical models and systematic characterization capabilities this could be achieved in a quantitative manner.

2.1.4. Computational biology to study protein secretion

Computational and systems biology approaches are helping to uncover the underlying mechanisms involved in the secretory processes of the cell. There are many definitions of systems biology, but in most cases it refers to strategies for carrying out a quantitative description and understanding of a system through the use of global analysis and mathematical modeling. All this under a paradigm claiming that the behavior of the system as a whole is more than the sum of its parts, therefore a holistic point of view could provide new interpretations on the interactions occurring in the system under study. In the context of bioproduction, adapting organisms for the secretion of non-native products is not an easy task. It requires a thorough understanding of cellular processes and the relationship between them. Specifically, in the case of secretory processes, wherein regulations are multifactorial due to the enormous number of interactions involved. (Travers et al., 2000; ; Thibault, Ismail, & Ng, 2011). System-level data integrated into mathematical modeling allows researchers and product developers to identify common regulatory patterns of these complex interactions. This could help to derive general rules for efficient secretion in order to optimize the bioproduction process and rationally design optimal cell factories.

As said, one of the goals of systems and computational biology research is to obtain a quantitative understanding of the biological system under study. Such understanding can be described in the form of a mathematical model, then it can be used as a representation of the system, but also to extract information inaccessible by experimental approaches. For instance, it can predict characteristics of the system different from those used to derive the model, sometimes exceeding the possibilities allowed by experimental procedures. Models can be classified in many ways, one that might be convenient in the scope of this work is that which differentiates between high-level models (or coarse grained models), and low-level models (or fine grained models). High-level models show the system in a relatively abstract form, focusing more on the interactions than on the details of the components. Low-level models are usually used to describe the molecular mechanisms underlying the interactions of individual components (Nielsen & Jewett, 2008).

Surprisingly, there are only a few examples of models describing characteristics and processes of the secretory pathway. In this section, I have selected three examples of models of the secretion process that approach the system from different perspectives. For example, in Wiseman, et al., 2007, the authors modeled the ER functions related to protein removal, either by secretion or degradation. They comprehensively capture features related to cellular proteostasis that would be difficult to describe from experimental approaches. The model is high-level description of the system (Figure 2.8), and suggests a framework for explaining distinct secretory capacities as a function of the protein stability and cellular phenotype. By modeling the path that proteins can follow as a function of their folding state, for example towards folding, degradation, or secretion, the model predicts the distribution of proteins in each of the possible pathway motifs.



pathway	*rate constant*	*pathway*	*rate constant*	*pathway*	*rate constant*
1	$k_{T\to TU}$ (0.03 s^{-1})	8 forward	$k_{U\to F}$ (see text)	12	$k_{RU\to R}$ (1 s^{-1})
2	$k_{TU\to T+U}$ (0.03 s^{-1})	reverse	$k_{F\to U}$ (see text)	13	$k_{U\to M}$ (see text)
3	$k_{TU+C1\to T+U}$ (10^6 M^{-1} s^{-1})	9 forward	$k_{E+F\to EF}$ (10^6 M^{-1} s^{-1})	14	$k_{C1+M\to C1U}$ (10^6 M^{-1} s^{-1})
4	$k_{TUC1\to T+C1U}$ (0.03 s^{-1})	reverse	$k_{EF\to E+F}$ (0.1 s^{-1})	15 forward	$k_{R+M\to RM}$ (10^6 M^{-1} s^{-1})
5	$k_{C1U\to C2U}$ (1 s^{-1})	10	$k_{EF\to E}$ (1 s^{-1})	reverse	$k_{RM\to R+M}$ (0.1 s^{-1})
6	$k_{C2U\to C1+U}$ (100 s^{-1})	11 forward	$k_{R+U\to RU}$ (10^6 M^{-1} s^{-1})	16	$k_{RM\to R}$ (1 s^{-1})
7	$k_{C1+U\to C1U}$ (10^6 M^{-1} s^{-1})	reverse	$k_{RU\to R+U}$ (0.1 s^{-1})		

Figure 2.8. ***High-level model of the secretory pathway and results***, from Wiseman, Powers, Buxbaum, Kelly, & Balch, (2007). Left: schematic of the mathematical model for protein export. ERAF: ER-assisted folding, ERAD: ER-associated degradation, U: unfolded protein, M: misfolded protein, F: folded protein, T: transporter, C: chaperone, E: export machinery, R: retrotranslocation machinery. The numbers indicate the rate constant of each process, specified in the table below. Unfolded proteins enter the secretory pathway and can be folded by the action of chaperones or become misfolded. Unfolded and misfolded proteins can be folded by chaperones or undergo degradation. Chaperones can either successfully fold proteins, or fail. Folded proteins interact with the export machinery to be secreted. Top right: export efficiency versus protein stability. Unstable protein will undergo degradation depending on ERAD activity and secretion capacity, significantly affecting secretion efficiency. Bottom right: export efficiency versus protein stability. Secretory efficiency is affected by folding rate. This shows a slight influence for short folding times, but becomes important when the folding rate values are comparable to the rates of the export and retrotranslocation machinery.

The article draws conclusions with potential applications in the rational design of cellular factories for secreted proteins. Their model shows that secretion efficiency depends on protein's folding complexity, the folding capacity of the secretory pathway, and the adjustable contributions from ERAD. The study also suggests that the so-called quality control of the secretory pathway ensures protein homeostasis by reducing the pernicious accumulation of unfolded proteins. That is achieved by tuning the ERAD/folding balance for a given protein. This can be important in bioproduction, as not only protein stability influences secretion efficiency, but also the phenotypic ERAD/folding balance of the cell factory. When a protein shows short folding time, the cell can manage to translocate it through the pathway. However, when the folding rate values are comparable to the rates of the export and retrotranslocation, the secretory efficiency decreases. Longer folding times lead to protein accumulation, and trigger protein degradation by ERAD action (Figure 2.8). Therefore, maintaining a production demand according to the folding capacity, for a given protein, may reduce protein accumulation and improve secretory efficiency.

Another example of a computational biology approach with relevance to bioproduction, and using in this case a low-level model, is the one reported by Brown & Koslover, 2021. Here the authors modeled the architecture of the quality control subsystem within the secretory pathway at molecular level (Figure 2.9). In this work, the authors investigate how the quality control stage of the secretory pathway decides whether an unfolded protein can still attempt to fold, and how pathway-specific details affect the performance of the process. One of the conclusions of the study is that an optimal quality control system limits the accumulation of unfolded proteins by adapting kinetic parameters to protein production levels.

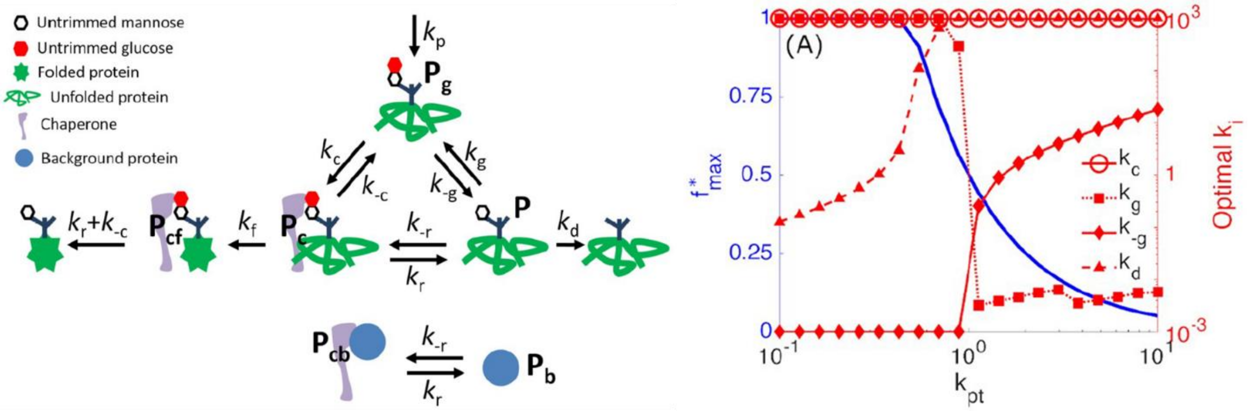


Figure 2.9. ***Low-level model of the secretory pathway quality control and results***, from Brown & Koslover, (2021). Left: Glycoprotein quality control via chaperon binding cycle scheme of the mathematical model. Pg: monoglucosylated proteins, Pc: unfolded chaperone-bound protein, Pcf: folded chaperone-bound proteins, P: proteins lacking a glucose tag, Pb: background proteins, Pcb: chaperone-bound background proteins. Unfolded proteins enter the cycle with a monoglucosylated state and can bind to an available chaperone to form an unfolded protein-chaperone complex. Unfolded protein-chaperone complexes can give rise to a folded protein and trim a glucose from the protein, then the folded protein exits the cycle. Unfolded proteins clipped with glucose can be reglucosylated or enter the degradation pathway. Monoglucosylated proteins that are not bound to a chaperone can have a glucose removed, serving as a mechanism to avoid excess unfolded proteins when no chaperone is available. Right: Optimal performance and parameters values. f*max: maximum folding fraction, kpt: total production rate. At low production demand, almost all proteins entering the quality control cycle are successfully folded. When protein production demand is higher, the unfolded chaperone-bound protein can no longer be formed with the overwhelming flow of incoming proteins. Therefore, the chaperones in the system become saturated and folding efficiency decreases because degradation is favored.

According to the authors, a robust secretory system design requires optimized parameters for low protein production demand levels, and an adaptive process that adjusts its activity to changes in demand. In the scope of the secretory pathway, and in agreement with the model results, the aforementioned adaptive process is the ERAD, whose parameters modulation is regulated by the UPR. This system can maintain maximum folding efficiency and low accumulation of unfolded protein at low and medium regimes by adapting the degradation of unfolded protein as production demand increases (Figure 2.9). This is significant in the field of bioproduction because it seems to point to an optimal level of production demand. Such point would be at low production demands, before degradation becomes significant compared to folding. In such state most proteins that enter the quality control cycle are successfully folded. When protein production demand is higher, an overwhelming flow of incoming proteins saturate free chaperones and over-accumulation of unfolded proteins occurs. That decreases secretory efficiency because ERAD is triggered and degradation is favored.

In another quantitative study of the secretory process combining an experimental and computational approach (Love *et al*, 2012), the authors implemented a high-level model of the steady-state distribution of the population (Figure 2.10). In it, different strains secreting heterologous proteins of distinct secretory complexity are considered. This study indicates that variations in secretion efficiency depend on both protein secretory complexity and production demand levels. Their main results suggest that there is a fixed secretion capacity and it is independent of the folding complexity of the secreted protein. Protein accumulation in the cell depends on the ratio between a fixed secretion rate, and the adaptive nature of the ERAD. The most efficient secreting cells are those that do not exceed a given ratio of ERAD over secretion rates.

In the absence of degradation, for a given production rate, the relation of internal protein levels and secretion rate should be inverse. Since more protein is exported out for the same input rate, the accumulation is lower, and so the internal levels of protein. When the pathway is saturated, due to a low secretion rate relative to production, internal protein levels increase in the cell and the UPR is activated. Consequently, ERAD is upregulated, and it reduces internal protein levels by degradation. Therefore, if only internal protein levels are observed, two different secretion rates may present the same phenotype, depending on the ERAD's contribution (Figure 2.10). Low secretors with high ERAD activation, present low levels of internal proteins, as it does for higher secretors. Cells with low secretion rate and low degradation levels will accumulate more internal protein. Hence, it is the cell secretion capacity what determines the efficiency of the secretory process, according to this study. To achieve optimal secretory efficiency, that is, minimize protein flux to degradation while maximizing secretion, it is necessary to match production demand to secretory capacity.

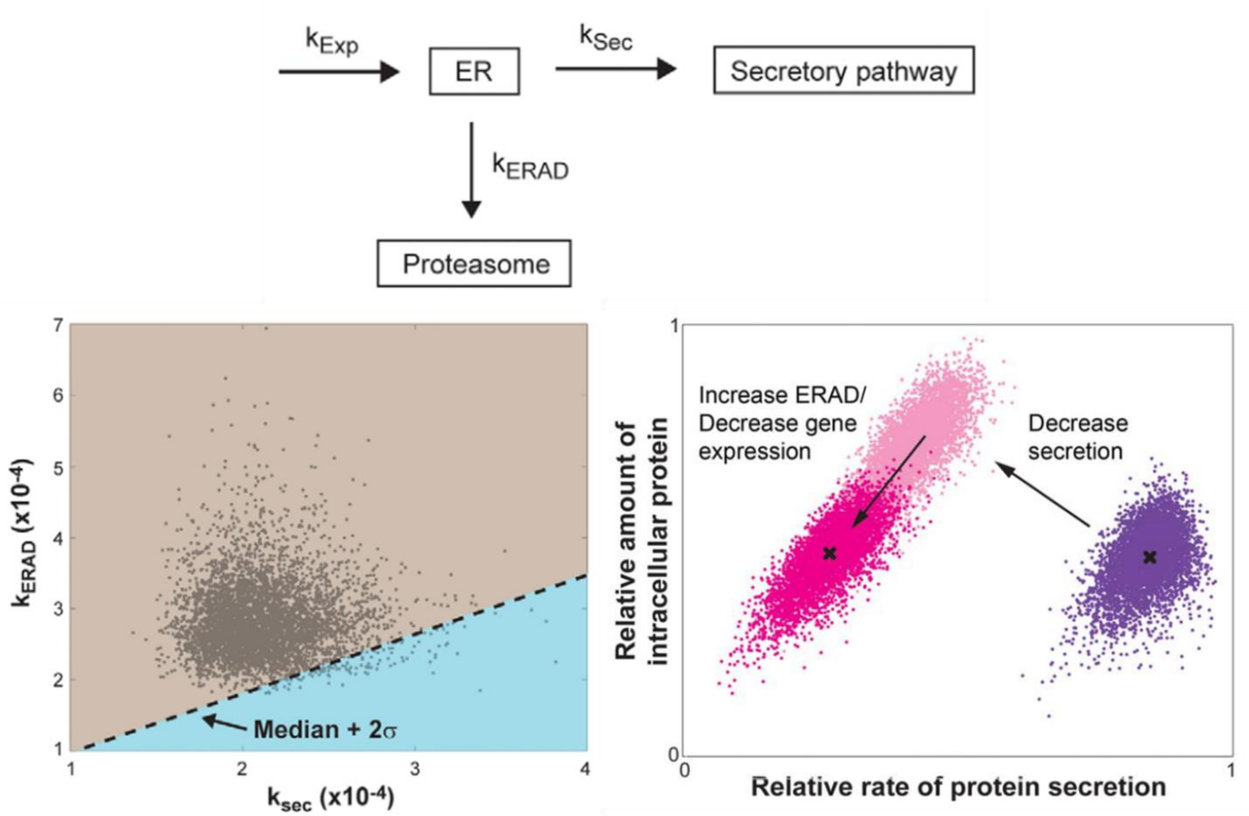


Figure 2.10. ***High-level population model of the secretory pathway and results***, from Love et al., (2012). Top: Model for the steady state distribution of secretory pathway trafficking scheme. Proteins enter the ER and then get out either towards secretion or degradation. Bottom left: Experimental data plotted as function of the inferred secretion and degradation rates. Blue shading indicates efficient secretors cells, those that exceeded a certain threshold ratio ksec/kERAD. Bottom right: Effects of altering secretion and degradation rates on intracellular protein. Purple: initial population, Light pink: secretion-inhibited population, Dark pink: stress-induced population.

The three aforementioned studies agree on at least one of their results: the protein folding complexity and the relative contribution of ERAD determine secretory efficiency. Protein accumulation is a consequence of the relation between the secretion capacity, folding complexity of the POI, and the strength of the production demand. In turn, protein accumulation triggers UPR, and consequently its degradation by ERAD. That said, it follows that efficient secretion of heterologous POI involves optimizing the levels of production demand. However, it requires a comprehensive and quantitative understanding of endogenous cellular processes, including those specific responses that are not even feasible to directly measure in real-time with current experimental platforms, such as ERAD, secretion or folding capacity. In this regard, integrated global analyses and computational modeling are powerful tools for inference, characterization, and interpretation of fundamental questions of the secretion process. However, in some cases the limitations of experimental setups prevent the systematic generation of data with sufficient throughput to achieve a comprehensive characterization at the system level. Further development of experimental platforms is needed to better understand these complex processes with several dynamically changing variables.

2.2. Technological challenge

2.2.1 – Bioreactors setups for systematic data generation and control experiments

Systems and quantitative biology studies are powerful and have shown great potential to benefit industrial biotechnology (Campbell *et al*, 2017). Often, in order to perform global analysis, these approaches require the integration of dynamic data from different biological processes obtained by different measurement devices. For instance, monitoring optical density from a culture of cells in a bioreactor, fluorescence measurements of those cells in a flow cytometer, and secretion levels from the reactor measured by activity assay in a plate reader. Generating and integrating data from different sources in a reproducible and systematic manner is challenging because it necessitates of assembled platforms with high levels of automation in order to yield the sufficient throughput and temporal resolution over long durations.

In the context of bioproduction, cell growth usually takes place in bioreactors, either in batch or continuous cultures. One of the challenges to generate accurate and reproducible data in experimental bioreactors setups is to maintain unchanging environmental conditions for the whole duration of the experiments. Uncontrolled environmental perturbations might change the physiological state of the system under study, and cause reproducibility issues. Continuous cultures are those in which fresh media is introduced at the same rate as it emerges, thus the culture remain at steady state. It makes the culturing modality suitable to maintain constant environmental conditions over long durations (Bull, 2010). For example, a turbidostat dilutes the cell culture during growth to maintain constant optical density and environmental parameters such as pH or nutrient concentration. However, the drawback of continuous cultures has always been their difficult construction process, the requirement of external hardware such as pumps, the size needed to be set up, and the lack of flexibility to easily tune them for different purposes.

In recent times, small-scale bioreactors have emerged as powerful platforms for synthetic and systems biology research (Wong, Mancuso, Kiriakov, Bashor, & Khalil, 2018; Steel, Habgood, Kelly, & Papachristodoulou, 2020). They allow to perform continuous cultures with precise control of culture parameters, such as temperature, cell density or media flow rate, enabling long-term experiments. Moreover, there are some examples of automating high-throughput measurements of in-home bioreactor platforms in batch and continuous cultures (Figure 2.11) (Milias-Argeitis, Rullan, Aoki, Buchmann, & Khammash, 2016; Harrigan, Madhani, & El-Samad, 2018). Precisely, automated cytometry connected to a continuous culture multi-bioreactor setup in the context of microbial systems and synthetic biology enables single-cell resolution data of microbial populations across time. Such technology, in combination with optogenetics, is a powerful tool for monitoring and controlling yeast populations in the research of the secretory process.

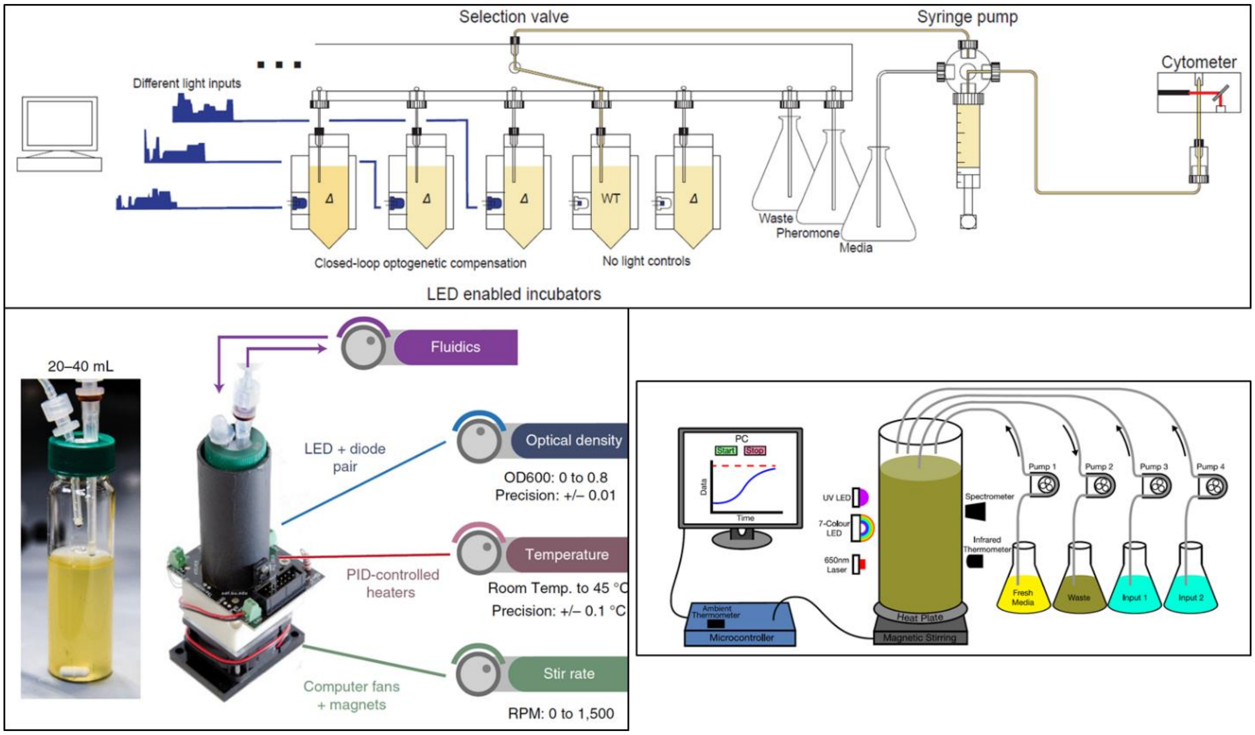

Selection valve
Syringe pump
Different light inputs
Cytometer
Δ
WT
Waste
Pheromone
Media
Closed-loop optogenetic compensation
No light controls
LED enabled incubators
20–40 mL
Fluidics
LED + diode pair
Optical density
OD600: 0 to 0.8
Precision: +/– 0.01
PID-controlled heaters
Temperature
Room Temp. to 45 °C
Precision: +/– 0.1 °C
Computer fans + magnets
Stir rate
RPM: 0 to 1,500
PC
Start
Stop
Data
Time
UV LED
7-Colour LED
650nm Laser
Spectrometer
Infrared Thermometer
Pump 1
Pump 2
Pump 3
Pump 4
Fresh Media
Waste
Input 1
Input 2
Ambient Thermometer
Microcontroller
Heat Plate
Magnetic Stirring

Figure 2.11. ***Example of small-scale bioreactors***. Top: from Harrigan et al., (2018); schematic showing the configuration of the in-house bioreactors setup for yeast continuous culture, LED-enabled, and including a liquid-handling device used for real-time cytometry measurement. Bottom left: Wong et al., (2018); *eVOLVER* open-hardware bioreactors, it is a do-it-yourself (DIY) framework to carry out high-throughput growth automated experiments. Bottom right: Steel et al., (2020); schematic of *Chi.Bio* bioreactors, a parallelized open-source continuous-culturing platform that allows measurement and control actions that can be applied in real-time.

2.2.2. Measurements of secretion levels for automated setups

Measuring secretion levels is of major importance when studying the secretory process, since it is the main product of the pathway. In particular, if the goal is to study secretion efficiency, measuring secretion levels in a quantitative manner is essential. When working in bioreactor setups, wherein cells and secreted proteins are mixed, it is necessary to separate one from the other to avoid the interference of the firsts with measurements, especially when fluorescent reporters are used. This constraint is a major difficulty for automated measurements, as it requires centrifugation, or other methods to discard the cells and debris before obtaining the cell-free medium containing the secretome.

Developing secretion measurements that can be made directly from the cell culture, that are generic for a wide range of proteins, that are relatively simple in its procedure, and yield quantitative data, is a challenging endeavor. Some methods for measuring secretion, such as halo assays, can be made directly from the cell culture, but they are qualitative, and require growth on a solid platform, normally agar plates. (Zhao *et al*, 2007). For quantitative methods to measure secretion, most of them are based on fluorescence measurements or activity assays of the POI in supernatants or purification products (D. Huang, Gore, & Shusta, 2008; Tang et al., 2015). Some are based on immune-affinity detection, such as ELISA-based assays or western-blots (Huang, Gore, & Shusta, 2008; Rogers, McConnell, & Greig, 2012). Alternatively, the supernatants or the purified protein can also be analyzed quantitatively by mass spectrometry, with or without the use of chromatography (Sorgo *et al*, 2010). Despite that these methods are established, they are difficult to automate for different POIs, mainly because they require several preparation steps, or are protein specific.

In addition to the methods discussed above, there are also some approaches that allow quantitative and automated secretion measurements directly from the cell culture. Protein A chromatography has been for years the gold standard for inmunopurification of secreted protein in continuous culture. It is widely used in the purification of monoclonal antibodies, and is already established for large-scale industrial setups. However, it remains a complex method and requires a great deal of expertise and a significant economic investment for research laboratories. (Ramos-de-la-Peña *et al*, 2019).

Another method with great potential for strain improvement and high-throughput screening are flow cytometry-based methods. These approaches have emerged as a potential solution for library screening in combination with yeast surface technologies, wherein the secreted protein binds to high-affinity protein ligands expressed on the cell wall (Boder & Wittrup, 1997). Different studies have attempted to demonstrate the usefulness of the method (Rakestraw, Baskaran, & Wittrup, 2006; Huang & Shusta, 2005), and study the secretory pathway with biotechnological perspectives in *S. cerevisiae* (Wentz & Shusta, 2007). These methodologies function similarly to the autocrine signaling present in some mammalian cell types: the cell secretes a protein that binds to specific receptors in its own surface (Figure 2.12).

Despite its usefulness, this technology imposes some limitations when working with secreted fluorescent reporters, that is very convenient for real-time monitoring in a flow cytometer. It would be difficult, if feasible at all, to differentiate the fluorescence of the secreted target protein bound to the surface from the protein still inside the cell. In addition, secreting the additional protein that is anchored to the cell wall and captures the protein of interest once secreted, may place an additional burden on the secretory pathway, since it follows that pathway to reach the cell wall. However, this technique has huge potential to measure secretion measurements in real time in bioreactors, if it is used in a co-cultured strain just to sense secretion levels.

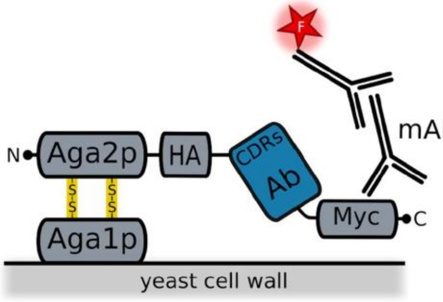


Figure 2.12. **Standard yeast surface display system**, from Uchański et al., (2019). The monoclonal antibody (mAb), is secreted by the cell. The single-domain antibody binding the secreted protein of interest (Blue) is fused to the Aga-2p protein. In this case, the secreted protein in the cell surface is detected by using fluorescently labelled secondary antibodies.

Sampling and performing automated secretion levels measurements from cell cultures growing in a continuous process involves many hurdles. Some of the most notable constraints are that sample processing should be minimized. In addition, the automation integration also implies that the process should be coupled to the robotic platform that works while the continuous culture is performed. Moreover, automation requires standardization in the measurements, and not all proteins of interest are measured in the same way, so it is necessary to develop general methodologies to detect the amounts of secreted proteins of different nature. The development of this technology is both as challenging as potentially useful, and will enable the integration of automated dynamic measurements of secretion in a systematic manner into the studies of the secretory process.

2.3. *Conclusion*

The aim of this chapter is to provide detailed information on approaches for strain engineering, and process optimization in order to improve protein secretion efficiency in the context of bioproduction. I also presented some of the most interesting studies, to my knowledge, in quantitative biology approaches to study the secretion process. I then reviewed state-of-the-art technologies in small-scale bioreactor setup, including some that allow automated monitoring and control of yeast cell populations.

In the first part, I presented the concept of the ideal chassis organism as a cell platform that is able to produce a wide range of POI. Then, by explaining the traditional approaches for strain optimization it has become clear that such endeavor is unreachable. Every POI has its own specific features, thus, engineering a strain to overproduce a given product might be inefficient for production of other. Therefore, the construction of a standard strain for bioproduction is not a realistic objective. The next subsection on the control of POI expression levels and its effect in secretory efficiency indicated that the general solution might lie in matching production demand with cellular capacity. Accordingly, in the subsection on the computational analysis of protein secretion, the three studies concluded that the key to maximizing secretory efficiency is to minimize protein accumulation. From the information presented in this section, I conclude that there is a need to develop generic and systematic approaches to characterize the secretion process based on the relationship between production demand and secretory efficiency. This approach could uncover common regulatory patterns that allow matching the production demand to the capacity of a given strain expressing a specific POI.

In the second section of this chapter, I discussed the needs and possibilities to overcome the technological challenge of performing quantitative dynamic studies of the secretory process. That is, by enabling high temporal resolution, as well as single cell level data, these two features imply sampling automation. First I went through the already existing bioreactors setups to monitor and control in real-time continuous cultures. Among the several examples in the literature, the only one matching the aforementioned requirements is an in-home setup developed by Harrigan et al., 2018. The fact that is not commercial, implies that we have to build our own setup to match all the requirements for the project. The second technological challenge discussed in this section is related to the need for a systematic framework to perform measurements of secretion levels in an automation-compatible, generic for any protein and quantitative manner. The main limitations among the current options are that either the POI-specific nature of the methods makes them not generic and incompatible with automation, or that they are highly complex and costly. However, the last methodology presented, surface display, may accomplish all the requirements. Unfortunately, we did not consider to use it in this project because it didn’t match with the plans in the first term.

Chapter 3: Experimental developments

As seen in previous chapters, the secretory biology of yeasts is a complex multi-stage process for which molecular mechanisms and interactions are well known (Papanikou & Glick, 2009). However, much remains to be understood regarding the relationship between protein production demand and secretory efficiency in bioreactors setups. The study of such complex processes is challenging and requires system-level approaches in which data are obtained with sufficient resolution and frequency to capture dynamic processes at the single-cell level, from various physiological process simultaneously (Nielsen & Jewett, 2008). Recent standardization and spread of small-scale and open-hardware bioreactors have provided powerful tools to perform such a systematic type of experiments (Wong, Mancuso, Kiriakov, Bashor, & Khalil, 2018; Steel, Habgood, Kelly, & Papachristodoulou, 2020). However, this type of platforms has to be further developed to achieve the experimental throughput required to implement dynamic a thorough characterization of the secretory process.

In this chapter I will present the experimental procedures and the platform we developed to perform the characterization of the secretory process. The chapter is divided into three main sections, the first one deals with the automated bioreactor platforms, the biological modules and strains used to work on it, and the experimental conditions required to achieve the objective of the project: developing a systematic characterization pipeline for the impact of heterologous protein secretory process. In the second section, I will review the methodology developed to quantify secretion levels directly from the cell culture, that enables to perform automated measurements of relatively low titers. Then, in the third section, I describe the strain construction process, the methodology employed and the rationale for the process. Finally, I discuss the obstacles faced, the options and decisions made, the contribution and limitations of the methodologies, and the significance of the approach in different contexts.

3.1. Experimental setup: Hardware and Bioware

3.1.1. Automated multi-bioreactors platform for yeast continuous cultures

Note: In the present subsection, I will present the bioreactor platform that allows automated experiments. Here it is important to note that it has been the project of Dr. François Bertaux, who has led the design and the development of the platform. My contribution has been notably in the production of yeast strains for hardware implementation and adjustments, performing calibration experiments in connection with metrological development, and active contributions to design and construction. I have also contributed intensively to the fine tuning and improvement once the platform was up and running.

To address the problem of characterizing the protein secretion process from a quantitative biology perspective, it is necessary to perform dynamic measurements with sufficient temporal and biological resolution allowing to properly observe the physiological changes associated to the secretion pathway on the appropriate time scale. Variables such as intracellular protein levels, stress response activation, growth rate, and secretion levels need to be quantified and integrated to perform a global analysis of the process. In addition, strict control of protein production demand and maintenance of environmental and growth parameters are necessary to yield robust and reproducible results. These parameters notably include temperature, cell density, nutrient composition, pH, aeration and oxygen concentration, and homogenization of the overall culture composition.

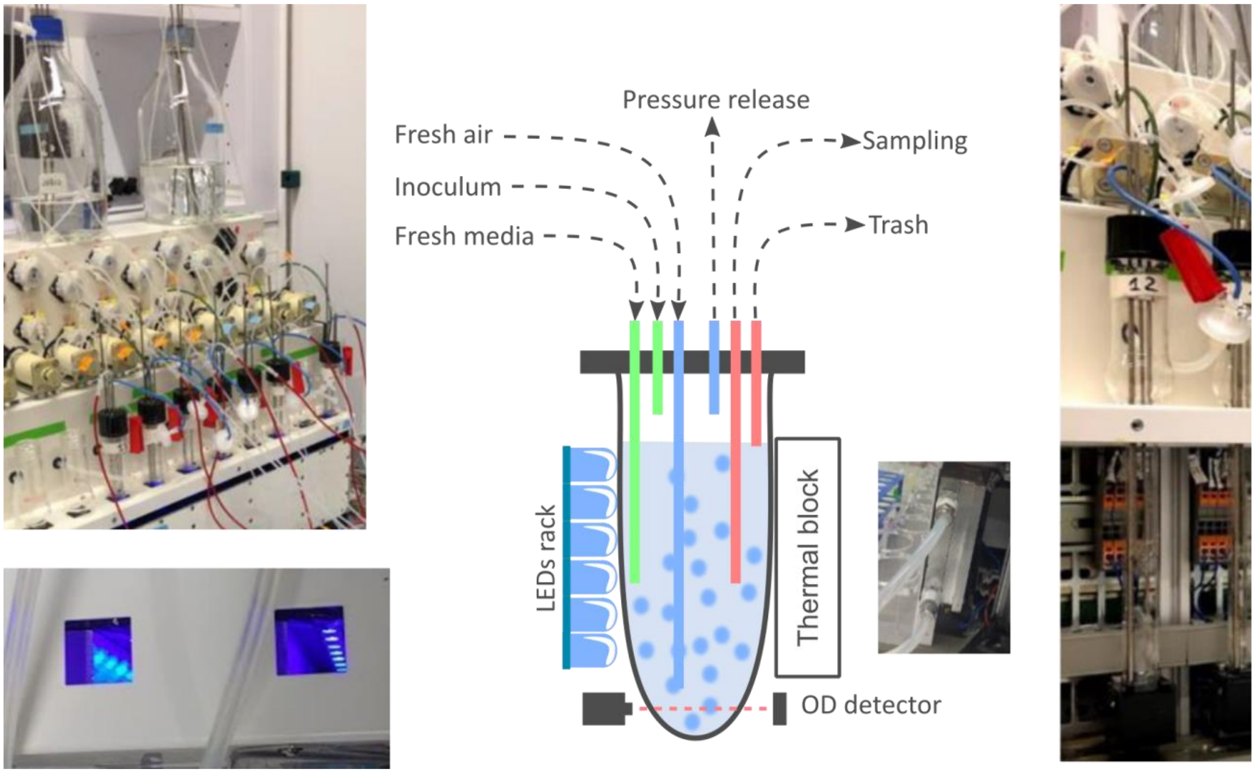


Figure 3.1. ***Multi-bioreactor platform for continuous yeast cell culture experiments with optogenetic control of gene expression.*** In the center is depicted the schematic of the bioreactor configuration, including the light-emitting diodes (LED), the thermal block and the optical density (OD) detector. In the upper left corner, the chassis structure is shown with the pumps, valves, vessels and thermal block in position. The lower left corner shows the LED strips inside the slots, wherein the vessels fit, to illuminate the sample and activate the optogenetic system. On the right side of the figure, the vessel inside its slot is shown, with the thermal block removed, and the OD detector at the bottom of that image. Next to the schematic of the thermal block, in the center, there is an image showing the tube from the water bath entering inside the thermal block, and a second tube that returns the water to be heated in the bath again.

To meet the aforementioned requirements, in terms of containing and maintaining a yeast cell culture under constant environmental conditions, we built a multi-bioreactor platform that allows to perform continuous yeast cell culture experiments with optogenetic control of gene expression and automated cytometry measurements (Figure 3.1). The platform is an extended version of a custom setup with eight reactors that enables independent experiments in each of them, designed and built for in-house use at *Institut Pasteur* (E. Frachon, A. Jacquier and C. Saveanu). Each of the reactors in which the cell cultures are housed is a transparent glass vessel fitting inside a slot in the chassis of the original setup. It is equipped with a thermal block that is in contact with one side of the reactor vessels to keep the temperature controlled by circulating hot water from a heating water bath by the action of an independent peristaltic pump. On the other side of the reactor vessel, attached to the mounting chassis, there is a strip of six light-emitting diodes (LEDs) that illuminate the cell culture, allowing optogenetic control of gene expression. The optical density detector, located at the bottom of each vessel, is already part of the original reactor configuration. In order for the medium to flow through the vessels and establish continuous cultures, two peristaltic pumps, for each reactor, control the inflow and outflow respectively. The outflow is divided into the waste and sampling lines by a pinch valve selector, thus, with the same output pump both flows are controlled independently. Note that the difference between trash and sampling lines is just based on the position where the volume is taken. Whereas the trash tube is located at the higher limit of the liquid culture, to control maximum volume, the sampling tube goes to the lower half of the liquid culture height. The inoculum of cells is done through an additional tube that enters the vessel, and is usually done by using a syringe. Finally, aeration is introduced into the reactors from the general laboratory air supply. It passes through a 0.2 µm membrane filter, and the excess pressure is released through an open tube at the top of the reactor, also protected from contaminations by a membrane filter with the same pore diameter. Mixing of the culture is achieved by bubbling caused by the aforementioned air circulation through the reactor.

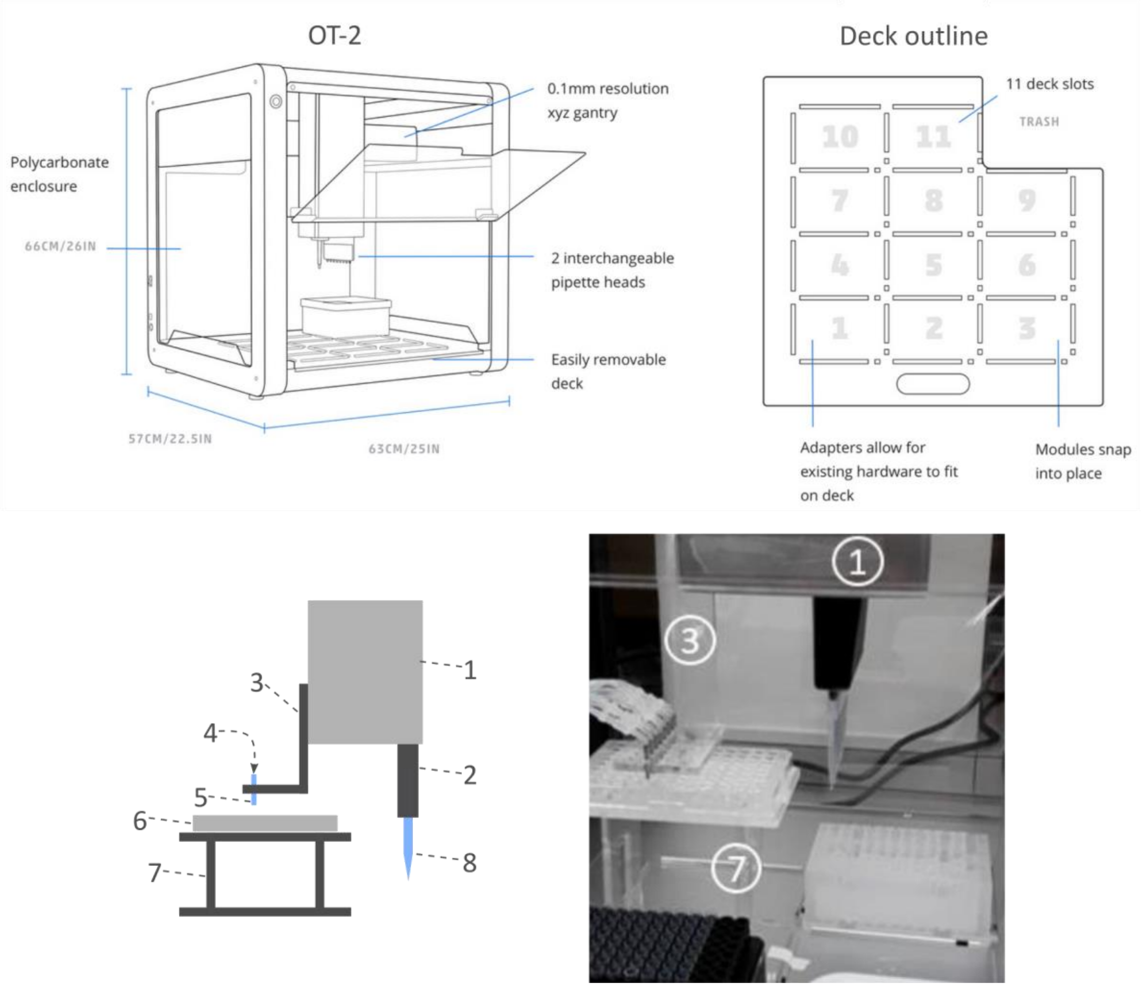

OT-2
0.1mm resolution xyz gantry
Polycarbonate enclosure
66CM/26IN
2 interchangeable pipette heads
Easily removable deck
57CM/22.5IN
63CM/25IN
Deck outline
11 deck slots
TRASH
10
11
7
8
9
4
5
6
1
2
3
Adapters allow for existing hardware to fit on deck
Modules snap into place
1
2
3
4
5
6
7
8
①
③
⑦

Figure 3.2. ***Schematic of the robot to connect the multi-bioreactors platform with the measuring device.*** The top scheme is adapted from the *Opentrons OT-2* website. In it is shown the overall view of the open-source robot, to the left, and the outline of the deck where the different modules, or accessories are located, to the right. In the common routine of the experiments described in this thesis, the deck layout is distributed as follows: deck 1 for tips; deck 3 for a funnel directed to the liquid trash

bin; decks 4 and 7 sampling plates; decks 5 and 8 for water reservoirs used for sample dilution in the sampling plate; decks 6 and 9 for tray of the cytometer measuring device. At the bottom, left and right of the figure, are depicted the sampling in-house made module to automatically sample from the bioreactors setup. 1, robot arm; 2, robot pipette head; 3, 3D-printed arm to hold the reactors sampling output tubes; 4, reactors sampling output tubes; 5, end of the reactors sampling output line; 6, sampling plate placed in robot decks 4 or 7; 7, 3D-printed plate holder in order to reach the required height for the robot configuration; 8, pipette tips for dilutions and liquid handling.

To obtain high temporal resolution in the measurements it is necessary to rely on automation of the sampling procedures. The low-cost and open-source *Opentrons OT-2* robot is the device we use to enable automated measurements by connecting the bioreactor setup with the cytometer (Figure 3.2). It enables for continuous and relatively high sampling frequency, and makes the bridge between the bioreactor setup and the measurement device. The link between the bioreactors and the robot consists of connecting the reactor pump-controlled output lines to the robot arm. The output lines are attached to the robot via a custom 3D printed arm extension next to the pipetting head. Since the sample and waste come from the same pump, the position of the robot arm will determine whether the outflow goes to the waste device, a 3D printed funnel for removal of significant volumes, or to the sample plate to undergo premeasurement treatment. Since the robot-driven measurements are performed using a cytometer, the samples must have an adequate cell density. Therefore, the culture sample from the bioreactors is treated by performing a dilution in water prior to measurement. The robot performs a 20-fold dilution of the raw culture volume in the sample plate before pipetting the treated sample into the cytometer. Finally, after the entire cytometer measurement cycle, the robot washes the measurement plate, in the cytometer, so as not to bias the next measurement with residual cells from previous sampling points (Bertaux et al., 2021).

As mentioned, the robot connects the bioreactors with the measurement device, a bench top cytometer (*Guava EasyCyte 14HT, Luminex*). At each time point samples are transferred from the robot to the cytometer, hence it requires that both of them are properly positioned and aligned to load the sample from each of the eight reactor into a 96-wells plate directly placed in the cytometer tray (Figure 3.3). Cytometry measurements allow single-cell characterization of microbial cultures without interfering with cell physiology. It brings several advantages to our setup, such as single-cell resolution of the population, at-line and sufficiently fast measurements, and it allows to measure internal fluorescence unbiased from extracellular components (McKinnon, 2018). These last features make it very convenient when studying secretory processes using optogenetics and fluorescent reporters, since it is not desired to make measurements directly on the culture. First because light inputs to excite the fluorophores could activate the optogenetic expression system, second because the fluorescence of the secreted protein must be differentiated from the intracellular one, and finally to avoid fluorophores bleaching in the culture volume.

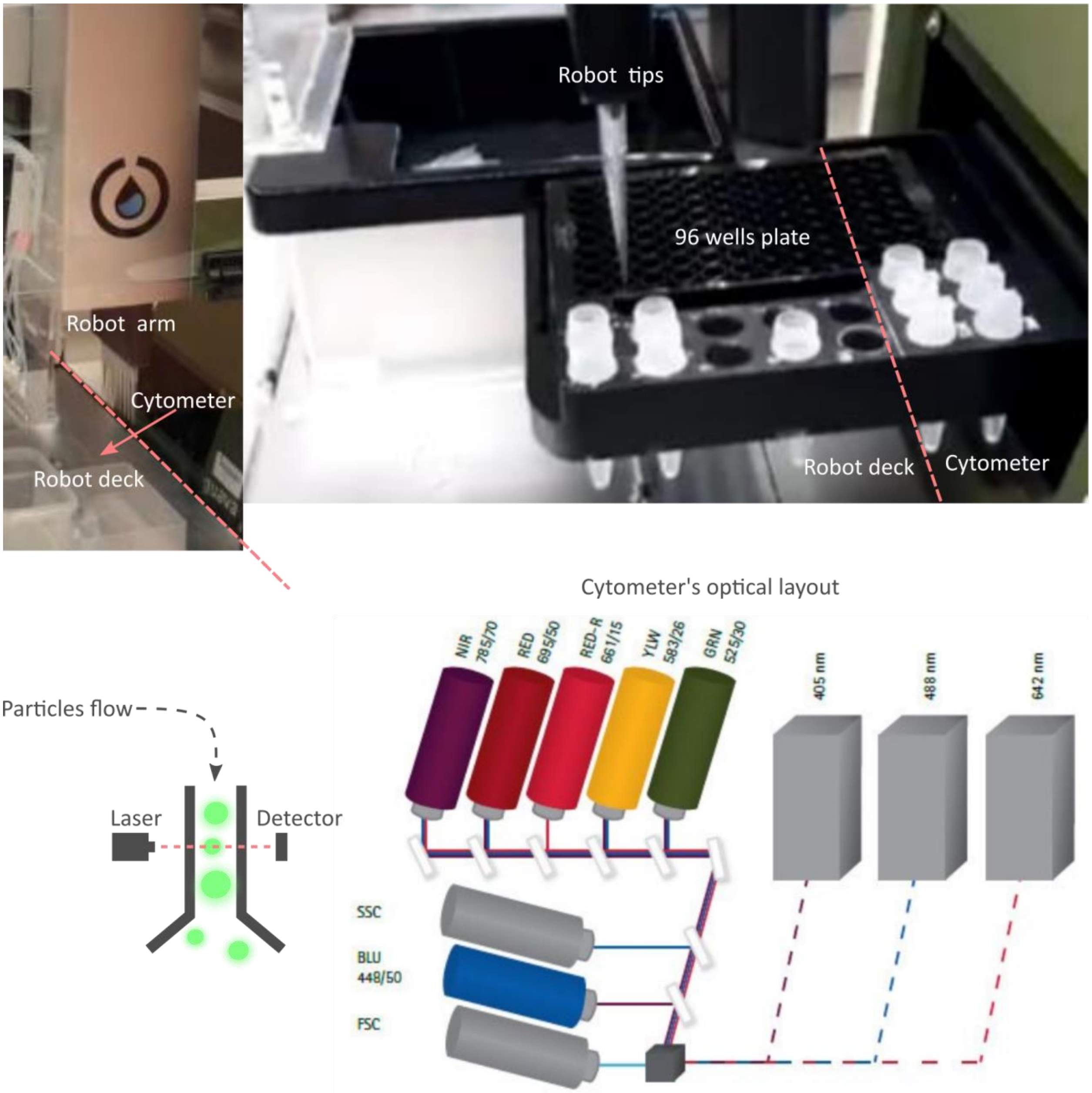

Robot tips
96 wells plate
Robot arm
Cytometer
Robot deck
Robot deck
Cytometer
Cytometer's optical layout
Particles flow
Laser
Detector
NIR 785/70
RED 695/50
RED-R 661/15
YLW 583/26
GRN 525/30
405 nm
488 nm
642 nm
SSC
BLU 448/50
FSC

Figure 3.3. ***Link of the cytometer with the automated multi-bioreactors platform, and schematics of cytometry data generation.*** On the top is shown the hardware link between the robot and the cytometer to perform single-cell measurements. The cytometer tray enters the robot deck in the direction of the red arrow. Then the robot's arm has access to the 96-well plate to deposit the sample from each of the reactors, and after dilution treatment, in eight different wells. At the bottom is represented the principle of the flow cytometer. The schematic at bottom right is taken and adapted from the *Guava EasyCyte* website. The particles, for instance cells, pass one by one between the light source and the different detectors that receive the scattered or emitted light from the individual particles by fluorescence. After measurement is taken the sampling volume is trashed.

Besides having all instruments next to each other in the proper orientation and alignment, the different actions to perform the automated measurements need to be orchestrated according to the experiment needs. To do so, we used the in-house developed *ReacSight* strategy (Bertaux et al., 2021), that enables the integration of the full platform via the *Python* web application micro-framework *Flask*. According to our interests, *Flask* provides the function of managing HTTP request from a unique controller computer to coordinate the different devices connected to the local network (Figure 3.4). This is very convenient specially working with open-source devices such as *Arduino*, or the *Opentrons OT-2* robot, since there are already established libraries for controlling them, or they are even conceived with *Python* application programming interfaces (API). Even in the case of closed-source software instruments, such us the used for controlling the cytometer, the automated processes to perform measurements are carried out by an automatic 'clicking'-based control software based on *Python* (*PyAutoGUI*). That allowed to control a software based graphic user interphase (GUI) by the remote controller computer via *Flask* API and HTTP requests. This strategy allows for coordination of the platform and experimental performance in an automated fashion, controlling all the components simultaneously from the controller computer connected to the local network.

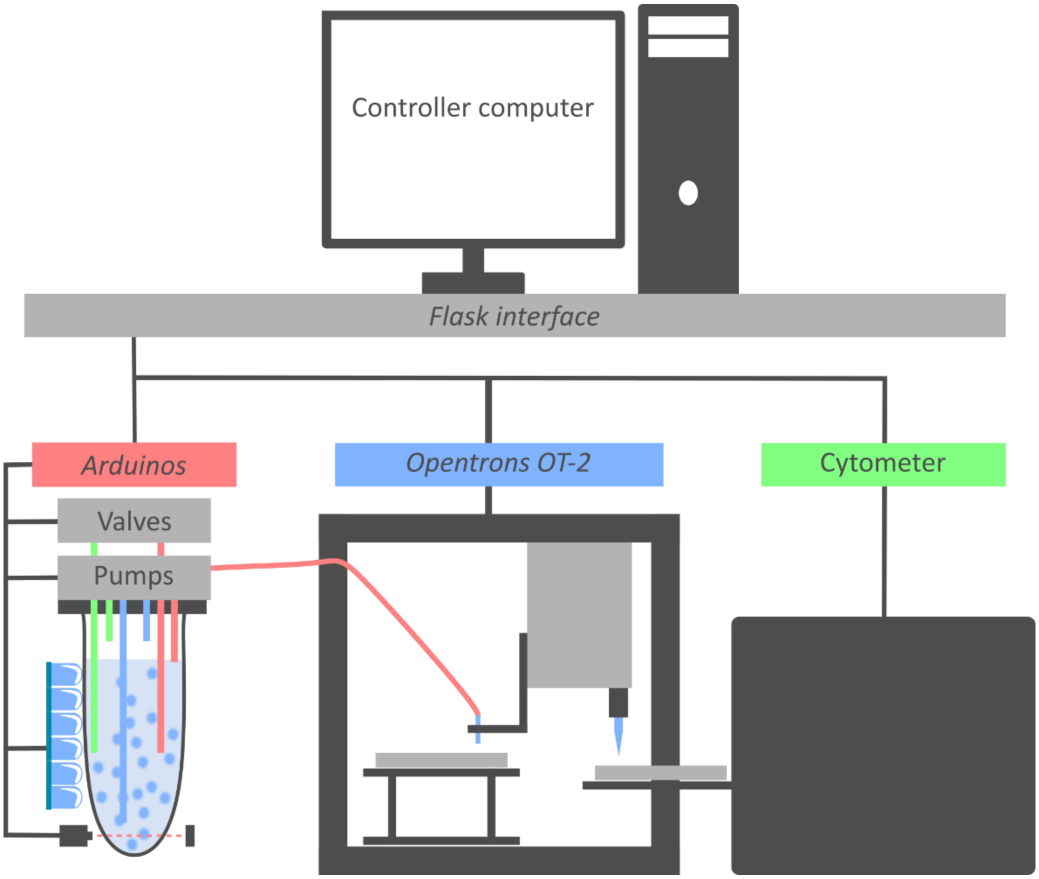


Figure 3.4. ***Multi-bioreactors platform enabled for automated cytometry measurements and optogenetic control of continuous yeast cell cultures.*** A unique controller computer orchestrates the platform function through the *Reacsight* strategy enabling full platform integration *(Bertaux et al., 2021)*. The *Flask API* interfaces the different independent components of the platform and

coordinates the experiments through an event system that enables taking measurements and controlling cell culture parameters simultaneously.

In summary, we have extended an existing bioreactor platform that allows for eight parallel experiments with robust maintenance of growth conditions in continuous culture. The new platform enables real-time optogenetic control of gene expression and automated cytometry measurements. That is achieved by coordinating the multi-bioreactors setup with an open-source pipetting robot and a flow cytometer controlled by close-source software. Altogether, this technology allows us to perform systematic experiments to produce data at the single cell and population level, measuring variables such as growth rate and internal fluorescence levels for several reporters simultaneously at high temporal resolution over a long duration.

3.1.2. Optogenetic system to control gene expression

As mentioned in the previous chapter, the term optogenetics refers to the control of biological processes with light inputs (Salinas, Rojas, Delgado, Agosin, & Larrondo, 2017; Figueroa, Rojas, Romero, Larrondo, & Salinas, 2021). In this document I will narrow down the concept just as the technique to specifically trigger transcriptional activation by exposing cells to light, because it is the exclusive use of optogenetics in this work. Thus, in the context of this project, once the ability to perform automated flow cytometry measurements from continuous cultures is achieved, combining it with optogenetic gene expression systems allows for dynamic characterization of the secretion process with tight control of the production demand.

A large catalog of fully characterized light-controlled optogenetic devices for gene expression control is available for yeast (Salinas, Rojas, Delgado, Agosin, & Larrondo, 2017; Figueroa, Rojas, Romero, Larrondo, & Salinas, 2021). These systems respond to a wide range wavelengths for transcriptional activation, from red to blue light (Shimizu-Sato, Huq, Tepperman, & Quail, 2002; Motta-Mena et al., 2014). Each system offers its own advantages and disadvantages. For example, red light-activated optogenetic switches have the advantage of being activated by low energy light input. It preserves the cell physiology, and minimizes bleaching of those fluorophores that have the absorption peak at higher energies. However, they often require the presence of cofactors that yeasts do not produce by themselves, thus, they must be supplied in the media, or the host be engineered for its production (Shimizu-Sato, Huq, Tepperman, & Quail, 2002). Blue light-activated optogenetic systems, unlike red-activated ones, require more energetic light sources for its activation, which could harm the cells (Robertson *et al*, 2013), and in most cases, they do not need additional cofactors (Motta-Mena et al., 2014). Therefore, the choice of an optogenetic systems that suits the best to a research scheme is not a trivial decision, as there will always be trade-offs, and the experimental design will be strongly influenced according to that decision.

In the context of this project, in which the goal is to characterize the physiological impact of producing heterologous secreted proteins under different production demands, a cofactor-independent optogenetic system is preferred. Since one of the main objectives is to study and quantify the burden generated exclusively by the expression of secreted proteins, it is not desired to produce additional physiological load by expressing the required cofactor. Another objective is to have fast and tight control of expression levels with minimal light intensity of the visible spectrum. This is in order to minimize the cell damage produced by the light, and to fit to the limitations of the light source used in the bioreactor LEDs configuration. After reviewing many options, and according to the mentioned constrains, two different systems were considered as most suitable: (i) *PhiReX*, based on the PhyB-PIF3 system in combination with *synTALE* DNA-binding domain (Hochrein *et al*, 2017), and (ii) the EL222, derived from *Erythrobacter litoralis* native transcription factor (Motta-Mena *et al*, 2014).

The two optogenetic systems mentioned above were elegantly developed and presented for their use in systems and synthetic biology, well characterized for different research groups (Hochrein, Machens, Messerschmidt, & Mueller-Roeber, 2017; Benzinger & Khammash, 2018; Zhao et al., 2018), and suitable for our setup. However, even if both systems showed rapid activation and wide dynamics range for gene expression levels, the best candidate, and the one finally chosen, was the EL222 derived system. The main reason is based on the absence of external cofactors required for its use, as well as the thorough characterization available in the literature. The recent, at the time, preprint version of an article performing a quantitative characterization of optogenetic systems with various promoter configurations and induction strengths, and the available DNA sequence in such preprint version, influenced the decision of selecting this optogenetic system (Benzinger & Khammash, 2018).

Other advantages of the EL222 optogenetic system are that it only requires one light color to control it: upon blue light illumination the transcription is activated, and in absence of light it deactivates. It also shows fast dynamics: the activation occurs in around 10 seconds after light induction, and the deactivation after 50 seconds in darkness (Motta-Mena et al., 2014). It allows high levels and tight control of gene expression, the dynamic range of induction is 150-fold, the steady state levels of expression increase linearly with the duty cycle (also called “pulse wide modulation”, it refers to the fraction of time of light stimulation with respect to a given period). In addition, when the induction is applied as duty cycle with fixed light intensity, the variability of expression levels observed a the level of population is lower than when the induction is based on light intensity (Benzinger & Khammash, 2018).

At the molecular level, the EL222 optogenetic systems consists in a light inducible gene expression device using part of the original *Erythrobacter litoralis* transcription factor of 222 aminoacids modified to only contain the minimal required elements for light-dependent transcriptional activation. These elements are the light-oxygen-voltage (LOV) domain and helix-turn-helix (HTH) DNA binding domain (Motta-Mena *et al*, 2014). In the modified version here used, it is fused with a nuclear localization signal, and the viral activation domain VP16 to recruit RNA polymerase (Sadowski *et al*, 1988). Upon blue light illumination, with maximal absorption at 465nm (Zoltowski *et al*, 2013), the EL222 proteins dimerize and bind the DNA target sequence pC120 to trigger transcription of the gene under the control of such promoter (Figure 3.5).

The same construct used by Benzinger & Khammash, 2018, was used for this study. The only changes applied by us to their published DNA sequence were the codon optimization for *S. cerevisiae* and removal of the restriction target sequence for the enzymes used in our cloning strategy. Regarding the promoter topology, one of the published configurations in Benzinger & Khammash, 2018 was used in our system, taking advantage of the rich characterization presented in the article. The selected promoter configuration contains five copies of the DNA-binding sequence, pC120, followed by the minimal promoter CYC180, and a consensus Kozak sequence (Kozak, 1999). To express the EL222 transcription factor, unlike Benzinger & Khammash, 2018, in this study I used the strong constitutive promoter pTDH3 (Lee *et al*, 2015) (Figure 3.5). In all cases the genetic cassettes were integrated in one copy at different loci, as it will be seen in the following sections of this chapter.

To ensure that the EL222 optogenetic system meets our requirements in this research project, a thorough characterization was performed in our own experimental setup. The main requirements for the system in the context of this project are to achieve a sufficient controllable dynamic range between the uninduced and fully induced state, to maintain low expression levels due to leakage, to ensure maximal activity of the EL222 promoter, to show expression variability similar to other common expression systems, and of course, not to interfere with cellular homeostasis.

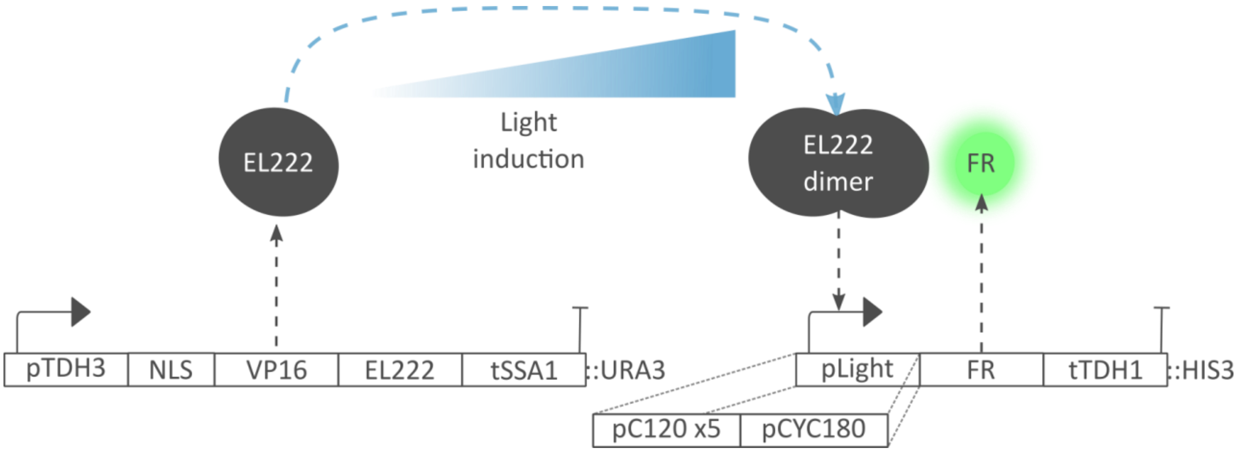


Figure 3.5. ***EL222 optogenetic system functioning and transcriptional units used in this study.*** The transcriptional unit for the expression of the transcription factor is integrated in the URA3 locus (Left). It is composed by the pTDH3 strong constitutive promoter, from *S. cerevisiae*, driving the expression of the modified EL222 protein fused to a nuclear localization signal (NLS) and the VP16 activation domain (VP16), the terminator used is the tSSA1 from *S. cerevisiae*. The integrated cassette includes a transcriptional unit enabling constitutive expression of the URA3 protein to carry out phenotypic selection of the integration positive clones. The inducible transcriptional unit (Right), confers constitutive expression of HIS3 for positive clone selection and it is integrated in the HO locus. The transcriptional unit includes the light inducible promoter (pLight), composed of five copies of the pC120 EL222-binding sequence, and the minimal promoter CYC180. Upon blue light illumination, the EL222 proteins dimerize and activate the expression of a fluorescent reporter (FR) under the control of pLight and the terminator tTDH1 from *S. cerevisiae*.

To characterize the system, I built a yeast strain that expresses the fluorescent reporter mNeonGreen, a monomeric green fluorescent protein (Shaner *et al*, 2013), chromosomally integrated in one copy, and whose expression is controlled by the EL222 optogenetic system (Figure 3.5). The first feature considered for characterization is whether the transcriptional activation should be induced by intensity modulation or by duty cycles. In order to compare, first it is needed to assess which intensity should be used for duty cycle induction approach. Therefore, I first characterized the effect of different blue light intensities in gene expression levels controllability, dynamic range, phenotypic variability, and growth rate. Then, I select the most suitable light intensity for the characterization of duty cycle induction.

I performed an experiment with two replicas in which the expression of mNeonGreen is induced at different intensities, then fluorescence is monitored every hour during 7 hours. In the Figure 3.6, the dynamic evolution of the internal fluorescence, normalized to the forward light scatter is shown for different levels of intensity modulated induction. The median fluorescence signal from mNeonGreen expression driven by the constitutive strong promoter pTDH3 is also shown as reference. To compare the background levels of expression, usually do to promoter leakage, or light contamination inherent to the setup, the levels of autofluorescence of a yeast strain that expresses no fluorophore, but has the EL222 transcription factor integrated in its chromosome, is shown.

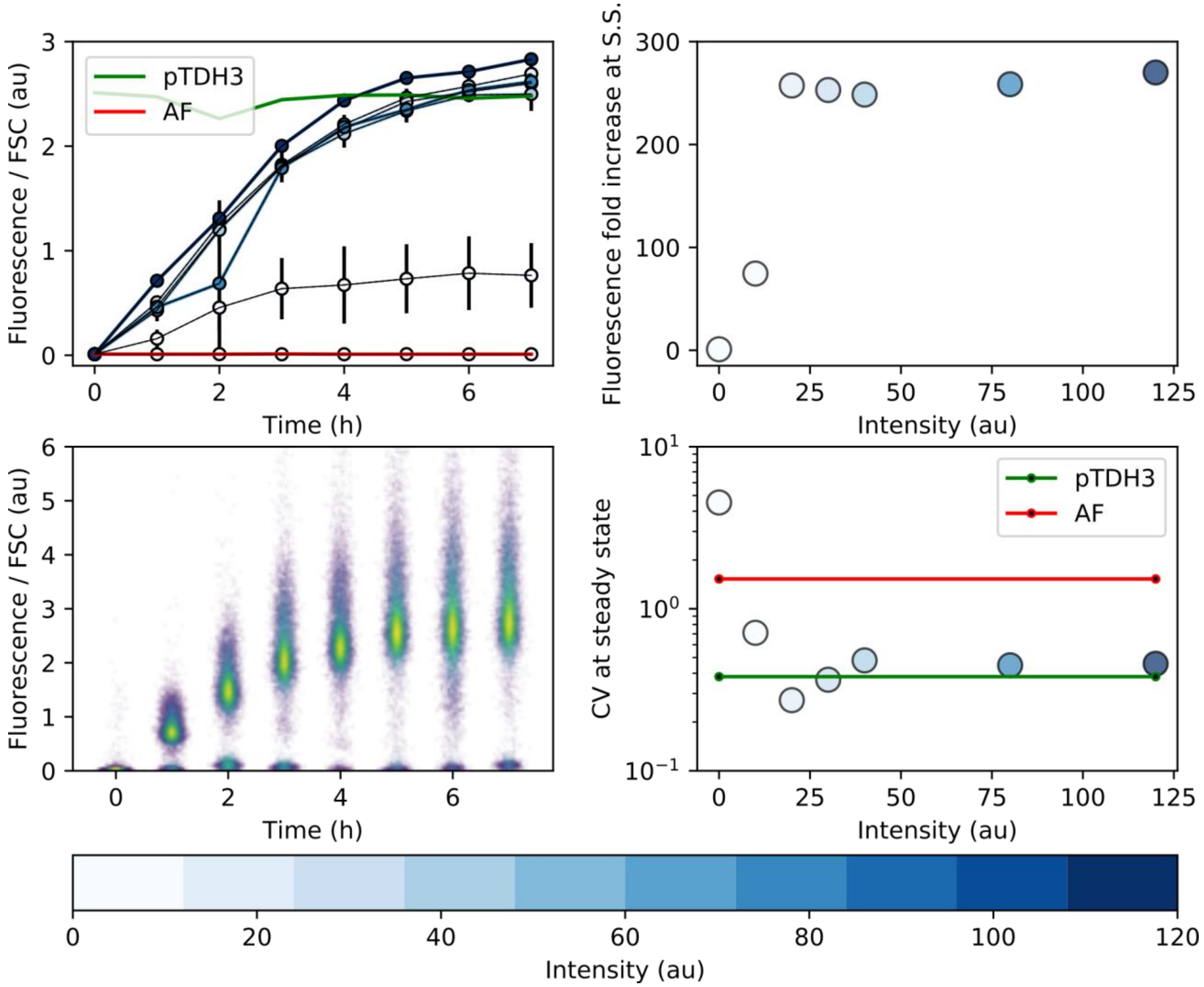


*Figure 3.6. **Light intensity modulation of EL222 optogenetic system transcriptional activation.*** On top representations the median of fluorescence dynamics from the mNeonGreen channel normalized to the forward light scatter (FSC) is shown. The blue color intensity represents the light intensity from 0 to 120 arbitrary units (au), the green line represents constitutive expression of mNeonGreen from the pTDH3 promoter, the red line autofluorescence (AF). At top left, the mean and standard deviation of the two replicas performed. In the top right plot, the mean steady state levels (S.S.) of fluorescence relative to the uninduced state and autofluorescence subtracted are shown for different light inductions for the two replicas. At bottom left, the population of cells for one replica at 120 units of light intensity. In this case the color from purple to yellow represents increasing event density. In the bottom right plot, the mean coefficient of variation (CV = standard deviation / mean) of the two replicas at steady state for each induction levels compared to constitutive expression and autofluorescence.

The system works satisfactory, it reaches high levels of expression, above pTDH3 levels, and the uninduced state has very low leakage, showing fluorescence levels similar to those corresponding to background levels of the autofluorescence reference strain. The steady state is reached around six hours after light induction, except for intensity levels 10, that seems to do it before. As the Figure 3.6 indicates, fluorescence levels stabilize after intensity induction 20. Indicating that levels of the EL222 transcription factor expressed under the control of pTDH3 promoter are sufficient to reach maximal activation the optogenetic system. Regarding phenotypic variability, in general it seems remarkably constant, except for low levels of expression, as seen for intensity 10, the two replicas are far apart in their steady state levels and dynamics.

A major potential limitation of using this optogenetic system, which is activated by blue light, is the possibility of impacting cell physiology. Blue light can be detrimental to cellular homeostasis at high intensities by increasing reactive oxygen species and triggering the oxidative stress response (Robertson *et al*, 2013). In the context of this project, a main requirement for the gene expression system is to fully activate promoter activity, in order to challenge the yeast secretory system, while not generating detrimental effects due to external factors other than those arising from the stress produced exclusively by the secretion process. Therefore, a light induction level allowing full activation of the promoter and not affecting cell physiology is needed.

To identify a light intensity that does not compromise cell physiology, the growth rate of the cultures induced at different light intensities in Figure 3.6 was monitored to assess whether it has a detrimental effect (Figure 3.7). The data indicate that intensities above 80 au decrease the growth rate, probably because it interferes with normal cell homeostasis. The possibility that this effect on growth is due to a physiological burden associated to the reporter expression is ruled out because induction at lower intensities yield same expression levels while no growth defect. According to the results shown in Figure 3.7 the intensities that allows full promoter activity and do not show harmful effects are those below 80.

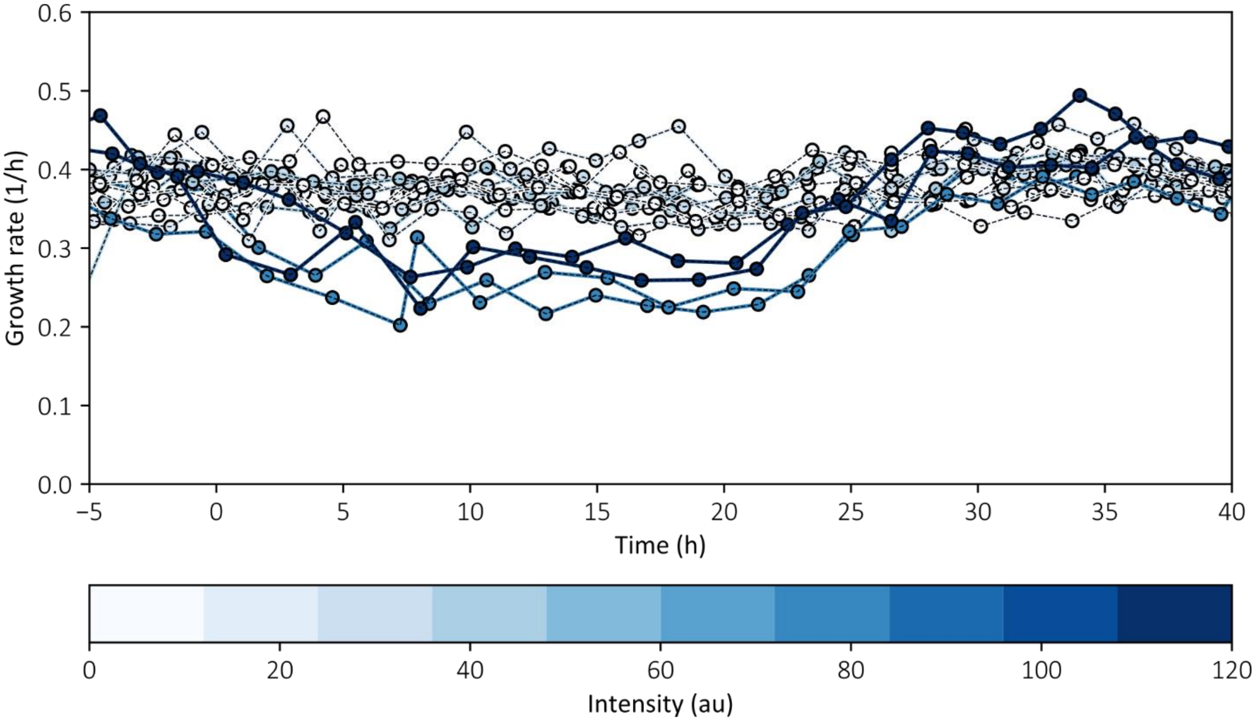


Figure 3.7. **Effect of light intensity exposure in growth rate as a proxy of cell homeostasis.** Growth rate of the cultures performed in Figure 3.6. The light exposure lasts from 0 to 22 hours. The growth rate raw data has been smoothed by applying a Savitzky–Golay filter from the *SciPy Python* library. The blue color intensity represents the light intensity from 0 to 120 arbitrary units (au).

From these results, intensity 20 was selected to carry out the characterization of the system by duty cycle induction (Figure 3.8), as it is the minimal to achieve maximum levels of expression and shows no undesirable physiological effects on the growth rate (Figure 3.6 and Figure 3.7). For such characterization, the cultures were exposed to light for different fractions of time over a 30-minute period, that is to different duty cycles. For example, a 10% duty cycle over a 30-minute period results in three minutes of light induction and 27 minutes of darkness. As in the characterization performed in Figure 3.6, the same background and constitutive expression reference strains are used.

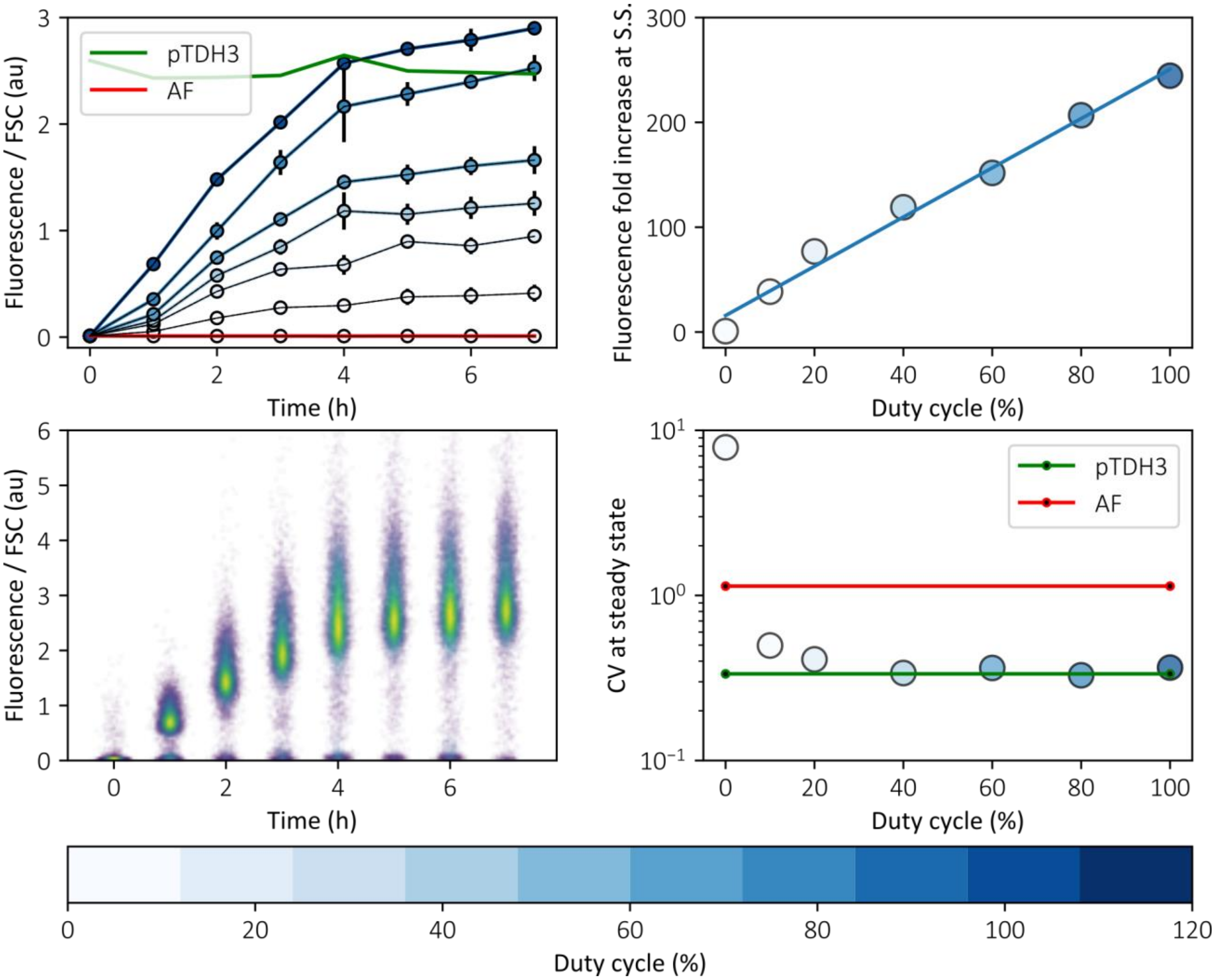


Figure 3.8. ***Light modulation by duty cycle control of EL222 optogenetic system transcriptional activation.*** On top representations the median of fluorescence dynamics from the mNeonGreen channel normalized to the forward light scatter (FSC) is shown. The blue color intensity represents the duty cycle applied from 0 to 100% of a 30-minutes period, the green line represents constitutive expression of mNeonGreen from the pTDH3 promoter, the red line autofluorescence (AF). At top left, the mean of the two replicas performed and shown in the plot at bottom left. In the top right plot, the mean steady state levels of fluorescence relative to the uninduced state and autofluorescence subtracted are shown for different light inductions for the two replicas. At bottom left, the population of cells for one replica at 100% duty cycle. In this case the color from purple to yellow represents increasing event density. In the bottom right plot, the mean coefficient of variation (CV = standard deviation / mean) of the two replicas at steady state for each induction levels compared to constitutive expression and autofluorescence.

In the case of duty cycle modulation, a full linear relationship between light exposure time and expression levels was observed. The maximum activity levels were similar as previously observed in the intensity modulation experiment, between 250 and 300 arbitrary units (au), which means that maximal activation is reached. The variability between replicates seems lower with duty cycle modulation especially for lower levels of induction. Regarding the phenotypic heterogeneity represented by the coefficient of variation (CV) of the population, in both modulation modalities it is similar to the one observed for constitutive expression of pTDH3. It is reduced with increase of mean fluorescence levels. Finally, the time to reach steady state is also about six hours.

According to these results, indicating linear relationship between duty cycle and expression levels, higher reproducibility between replicates, maximal levels of fluorescence reached, and not negative effect observed on cell physiology. The EL222 system seems to meet all the requirements exposed above, and duty cycle induction modulation the most appropriated way to regulate gene expression with this optogenetic system.

In conclusion, here I have selected an optogenetic system for inducible gene expression that overcomes the limitations imposed by our experimental setup, and does not impact cellular physiology by adding an additional burden. The EL222 optogenetic has been selected because its fast activation and deactivation dynamics, and independence of adding additional cofactors. It met all the requirements, such as having a wide dynamic range and high expression levels, tight control of gene expression, the same phenotypic variability as observed for constitutive expression by pTDH3 promoter, and it did not show detrimental effect on cell physiology. After characterization of the system in our bioreactor setup, I tested its usefulness and established the best parameterization for its performance in our future experiments.

3.1.3. Sensor strain to quantify effective induction levels

During the various experiments performed on the bioreactor platform in the scope of this project and others carried out in our setup, we observed variability in induction levels that was not fully understood (Figure 3.9). Such variability could be due to the intensive use of LEDs leading to reduction of intensity, washing-associated changes in the transparency of the bioreactor glass vessel, or simply day-to-day variability attributed to unknown effects in the bioreactors platform.

Variability in the induction system from one vessel to another is certainly a problem for future experiments studying the secretory process. Assessing gene expression levels directly from internal fluorescence signal in cells secreting the reporter protein is not feasible because an unknown fraction of the protein is secreted. To address these experimental limitations in the platform, a sensor strain has been constructed to have unbiased measurements of the induction levels actually perceived by cells.

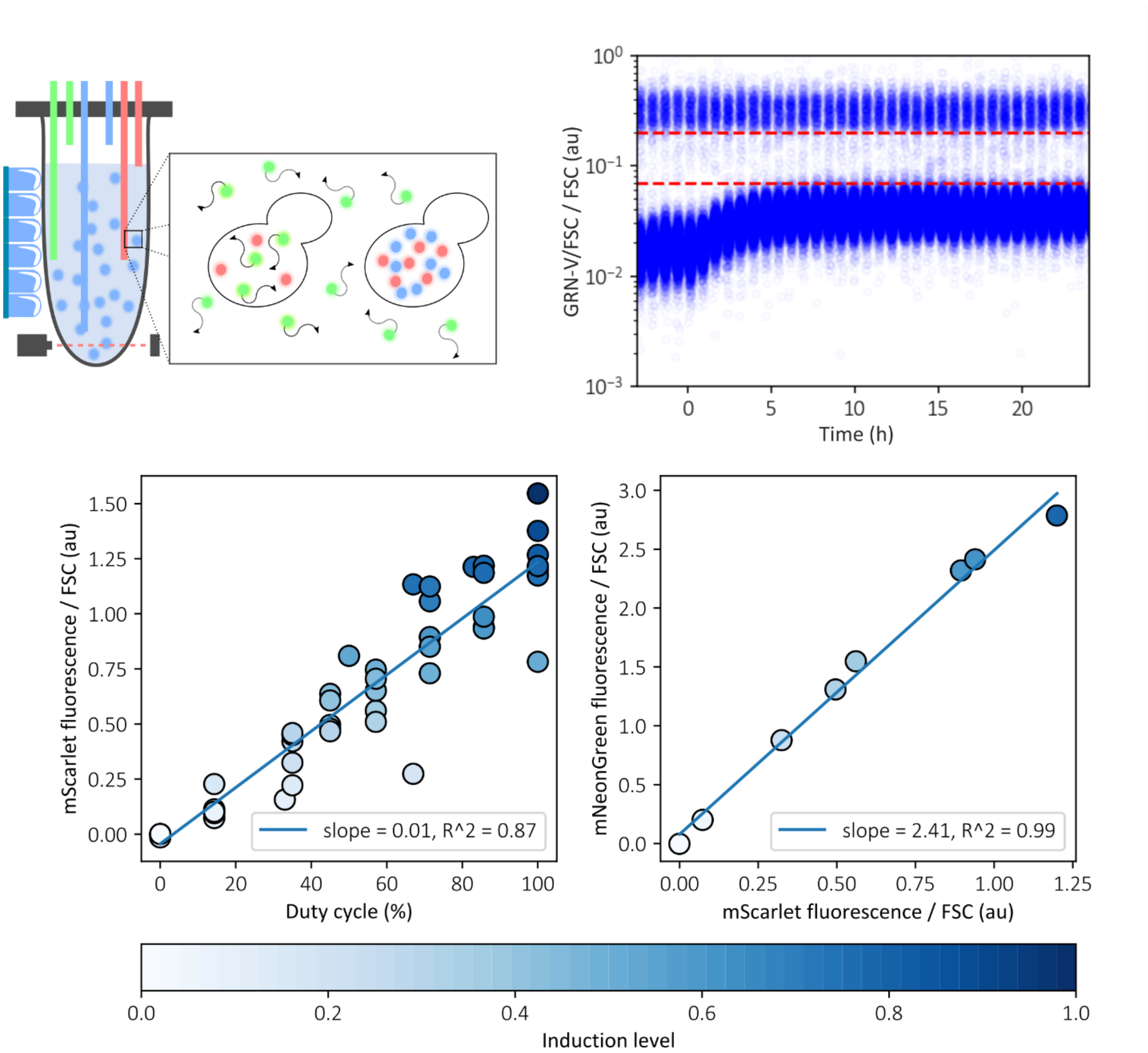


Figure 3.9. ***Sensor strain usage in the bioreactors platform to assess induction levels.*** A scheme of the co-culture of sensor strain and case study strain in the bioreactor is shown in the top left figure. The sensor strain produces mCerulean (blue dots) under the control of a constitutive promoter, and mScarlet-I (red dots) under the control of the EL222 optogenetic system, as it is the case for the secreted protein in the case study strain (green dots with arrow). The case study strain also produces an intracellular red fluorophore which expression is regulated by the stress response (red dots). In the top right plot it is shown the blue fluorescence distributions corresponding to the two strains, and the red lines are the thresholds to differentiate the sensor strain (blue fluorescence above 0.2 arbitrary units (au)) from the case study strain (blue fluorescence below 0.07 arbitrary units (au)). In the bottom left plot is represented the variability between internal fluorescence and duty cycle among different experiments using the sensor strain, as well as the linear regression. The blue light intensity represents the real induction levels assessed using the internal fluorescence of the sensor strain. The bottom right figure, shows the relation between the internal levels of fluorescence at steady state for both, case study, expressing a non-secreted mNeonGreen, and sensor strain in co-culture, together with the linear regression.

Such a yeast sensor strain expresses a red fluorescent protein, mScarlet-I (Bindels *et al*, 2016), whose expression is controlled by the EL222 optogenetic system as for the gene of interest in the strain under study. Therefore, at the steady state, expression levels of mScarlet-I in the sensor strain should match those for the protein of interest in the case study strain. In addition, the sensor strain constitutively expresses the blue fluorescent reporter mCerulean, a monomeric protein that shows rapid maturation and is phylogenetically close to mNeonGreen (Rizzo & Piston, 2005). This feature allows differentiating one yeast strain from the other in co-culture by observing blue fluorescence levels of both.

As shown in Figure 3.9, there is a linear relationship between the expression levels of the sensor strain and that of the case study strain for a non-secreted protein. Therefore, from now on, the internal fluorescence levels at the steady-state in the sensor strain will be used as a proxy for the actual induction levels in the case study strain, instead of using the duty cycles directly.

As a summary, some limitations of the bioreactor platform in terms of light induction variability were detected among several experiments. The origin of such variability is not clearly understood, thus, it is difficult to solve the problem as there are many possible sources causing the discrepancies. This is a major issue because one cannot consider duty cycles as a reference indication of induction levels. But using fluorescence levels of a stable and not secreted fluorescent reporter at steady state is a convenient proxy for induction strength estimation. To overcome this limitation of the platform, a sensor strain was constructed and co-cultured with the strain of interest in all experiments. This strain constitutively expresses a blue fluorophore that differentiates it from the case study strain, and a not secreted red fluorophore expressed under the control of the same optogenetic system as the strain of interest. Hence, the levels of induction of gene expression of the red fluorophore in the sensor strain match those of the case study strain. Accordingly, it makes possible to assess induction levels in both strains. Moreover, other functions of the sensor strain such as growth rate dynamics measurements are also applicable.

3.1.4. Sensor strain to assess growth rate dynamics

To monitor changes in growth rate as a proxy for the physiological state of the cell it is possible to use OD measurements collected from the bioreactor platform. In the case of working in a turbidostat mode, where the OD is kept constant, assessing the growth rate is done through the dilution rate of the media supply pumps. However, in the case of our experimental setup of bioreactors, they contain a co-culture, as mentioned in the prior section. Thus, changes in OD are dependent on the growth of two strains that do not necessarily grow at the same rate.

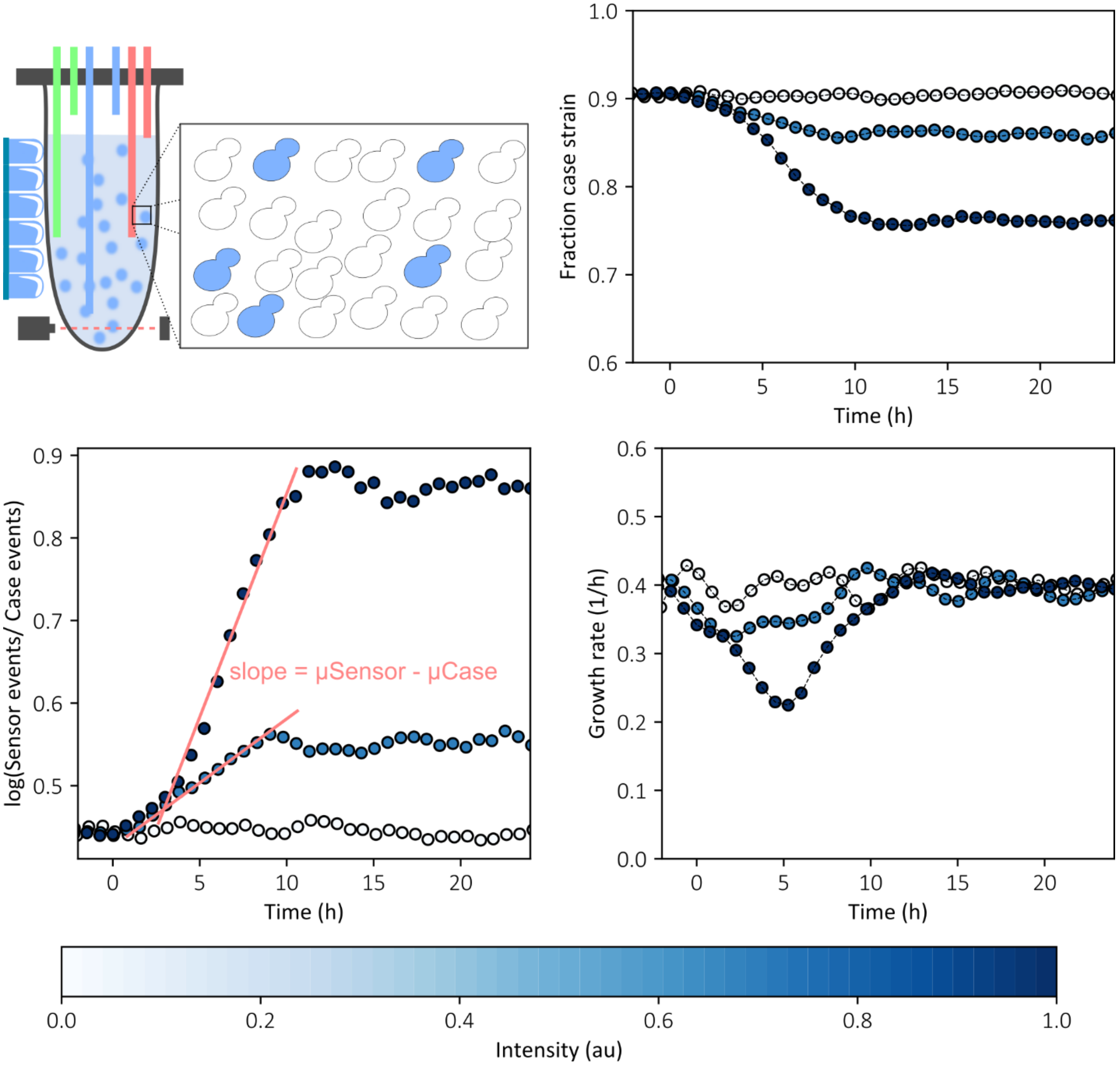


Figure 3.10. ***Sensor strain usage in the bioreactors platform to assess growth rate of the case study strain.*** A scheme of the co-culture of sensor strain and case study strain in the bioreactor is shown in the top left figure. The sensor strain produces mCerulean that is used to differentiate it from the case study strain and monitor the dynamic changes of its fraction (top right plot). Those changes in the fraction are used to compute the strains ratio changes (bottom left) and from them, the changes in the growth rate (bottom right).

To consider changes in growth rate in a more robust way, and those that exclusively relate to the case study strain, it is possible to use changes in the relative abundance of sensor cells as a reference. By starting with a known fraction of sensor strain cells in the population and following dynamically the changes, it is possible to assess growth rate of the strain under study, assuming that the growth rate of the sensor strain as constant. As it has been already shown in the optogenetic system characterization (Figure 3.7). As described in Bertaux et al., 2021, the rate of change of the logarithm of the strains ratio over time is equal to the difference of the growth rates . Thereby, since the sensor strain has a constant growth rate, it is possible to deduce the growth rate of the case study from real-time data (Figure 3.10).

As conclusion, in this project we aim to monitor the growth rate of the case study strain as a proxy for its physiological state. The sensor strain, used to measure induction levels, also serves as a sensor to assess changes in the growth rate of the strain under study. Since the sensor cells have a constant growth rate, changes in the initial fraction of these cells respect to the total population are explained by variations in the growth rate of the strain of interest. These growth rate changes in the case study strain should be attributed solely to the secretion process, as this is the only physiological process that differentiates the two strains.

3.1.5. Maintenance of culture environmental conditions

Continuous cultures allows to maintain steady-state environmental conditions over long periods of time, together with the use of defined media, it favors reproducibility of the experimental procedure (Winder & Lanthaler, 2011). In the case of our experimental design, maintaining constant environmental conditions during the whole length of the experiment is important and challenging. Yeast cultures tend to decrease extracellular pH below 4.5. That has an important impact on the study of protein secretion, since there are extracellular proteases whose activity is pH-dependent (Kang et al., 2000; Görgens, Van Zyl, Knoetze, & Hahn-Hägerdal, 2005). This implies that part of the secreted proteins are degraded if the extracellular pH is not controlled. Such fact is of major importance in the context of this project, given that protein secretion level is one essential variable to measure and to take into account in order to interpret the results. Along these lines, the required media conditions that allow to maintain pH levels high enough along the experiment to preserve secreted protein stability as well as keeping cell homeostasis, are a key factor for our study.

Maintaining redox conditions to avoid degradation of secreted proteins poses a real challenge in the context of a continuous culture in our experimental setup. That is because most of the buffers tested (Potassium phosphate, sodium potassium phosphate, phosphate buffer saline, and even HEPES) produced crystals that clogged the pumping system. We encountered this problem at the concentration provided in the literature to maintain the pH above five, that is the value recommended to keep secreted protein integrity (Kang et al., 2000; Huang, Gore, & Shusta, 2008). The presence of crystals either in the source medium, or in the reactors vessels, is not compatible with our bioreactor configuration because, as said, they clog the tubing and pumps, thus causing major problems in continuous culture mode. It posed a real challenge, because without the possibility of using a buffer the pH-drops below pH 4.5 and secreted protein integrity is really compromised. Such that we could not detect the presence of secreted proteins (by different methodologies) in continuous cultures until the problem was solved.

Finally, we found a solution in the literature. Previous studies (Chung & Park, 1998; Kang et al., 2000) reported that addition of the amino acids, and specifically L-arginine at a concentration of 0.4 M, decreases proteolytic activity. Our results show that L-arginine addition, even at lower concentrations, maintains pH above seven in continuous cultures, and allows detection of the secreted proteins. To obtain this conclusion I had to perform several trial-and-error experiments until I found a concentration of arginine that did not produce crystals and kept pH above seven in continuous culture. The best match is L-arginine at a concentration of 5 mM in the source medium, that is the media used to feed the reactors volume. Above that L-arginine concentration a sediment is produced, which is not detectable in the source media flasks but it is in the vessels, maybe because increase of pressure or concentration in the circulation system. This change in pH does not affect the growth rate of the cells, as we observed from experimental data, and reported in literature (Peña *et al*, 2015).

As conclusion, the pH of the culture must be above a value of five to avoid secreted protein degradation. After dealing with many difficulties through a long and hard trial-and-error process, I managed to reach a constant pH compatible with the hardware constrains of the multi-bioreactors platform. It was solved by thoroughly exploring the literature, and adapting existing knowledge to the specificities of our setup. Precisely, adding L-arginine at a concentration of 5 mM of in the source media was sufficient to maintain protein stability and avoid pernicious crystals formation.

3.1.6. UPR reporter to quantify stress levels in real-time

As seen in the introduction of this thesis, in *S. cerevisiae*, the accumulation of unfolded proteins in the endoplasmic reticulum (ER), as well as other alterations downstream in the secretory pathway, triggers the unfolded protein response (UPR). This is a multi-factorial response that eventually gives rise to the upregulation of hundreds of genes involved in protein folding, unfolded proteins degradation, and secretion, among others (Travers et al., 2000; Buchberger, Bukau, & Sommer, 2010). It makes this adaptive response a good indicator of ER-associated and secretory stress (Merksamer, Trusina, & Papa, 2008; Pincus et al., 2010).

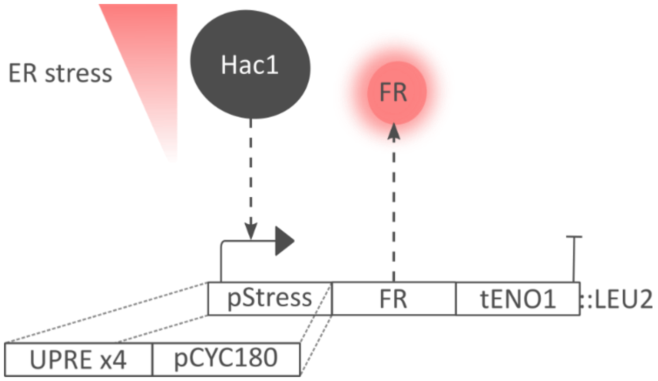


Figure 3.11. ***Stress reporter integration cassette used in this study.*** Scheme of the stress reporter based on the unfolded protein response (UPR) from *S. cerevisiae*. The endoplasmic reticulum (ER) associated stress triggers the mRNA splicing and subsequent translation to synthesize the transcription factor Hac1. This modulates the expression of genes under the control of UPR elements (UPRE) regulated promoters. In the stress reporter shown in the figure, the expression of a red fluorescent reporter (mScarlet-I) is regulated by the pStress promoter, which is composed of four copies of a consensus UPRE and the CYC180 minimal promoter. Therefore, upon ER-associated stress, the expression of the red fluorescent reporter (FR) is triggered. The terminator used in the transcriptional unit is tENO1 from *S. cerevisiae*. The integrated cassette confers constitutive expression of LEU2 for positive clone selection and it is integrated in the LEU2 locus.

Measuring stress levels associated to the secretion process is very convenient because it helps to connect the causes and consequences of the stress. In order to monitor the secretion-associated stress I took advantage of previous designs using the UPR as stress reporter (Pincus *et al*, 2010). Thus, I integrated in the yeast chromosome a transcriptional unit in which the expression of a fluorescent reporter, mScarlet-I (Bindels *et al*, 2016), is under the control of a UPR-regulated promoter (Figure 3.11). Such that, when the UPR is active, the levels of the red intracellular fluorescence account for the levels of the stress response. It acts as a UPR activity sensor that allows to detect a fluorescent signal in real-time.

To characterize the UPR activity sensor, yeast cells carrying the device were treated with dithiothreitol (DTT), a redox reagent that prevents disulfide bonds formation, thereby inducing protein misfolding and UPR activation (Figure 3.12). Two strains were added and considered as reference fluorescence levels. One harboring a cassette with pTDH3 constitutive promoter controlling the expression of mScarlet-I, and other strain without any fluorophore expression, indicating just background autofluorescence. In Figure 3.12 the cells with UPR reporter were incubated during 6 hours in absence or presence of different concentrations of the drug. The results shown that red fluorescence increases with increasing concentration of DTT. It is possible to observe the UPR signal saturation beyond a DTT concentration of 4 mM, these levels are close to the pTDH3 expression levels. However, it is noteworthy that comparison may be misleading, since cells expressing the fluorophore under the control of pTDH3 were not under the influence of DTT. The drug could affect fluorescent signal due to redox effects, or through growth rate and its effects fluorescence dilution, because cells incubated with certain concentrations of DTT decreased their growth rate (Pincus *et al*, 2010). What is also observed from these data is that there is a constant basal UPR signal that is around 20% higher than the autofluorescence background.

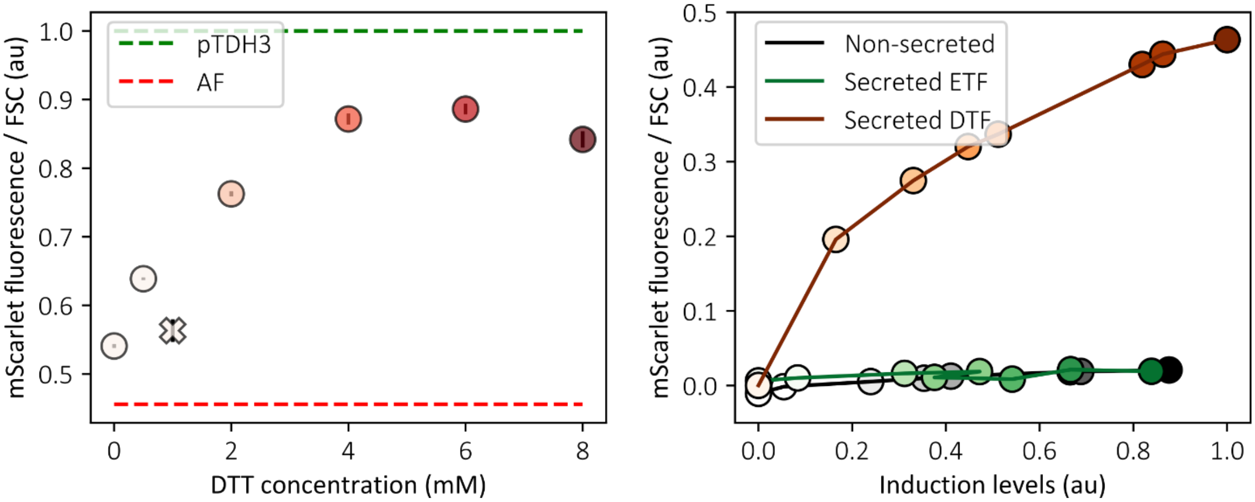


Figure 3.12. ***UPR reporter characterization experiments with DTT and heterologous protein expression induction.*** In the left figure is shows the mean and standard deviation of two replicas experiment to test the activation of the UPR reporter by dithiothreitol (DTT) addition at different concentrations during 6 hours. The red intensity corresponds to the concentration of DTT in the media. As reference levels of fluorescence, constitutive expression of mScarlet-I under the control of the *S. cerevisiae* strong constitutive promoter (green dashed line), and autofluorescence strain without fluorescence expression (red dashed line, not in the DTT concentration scale) are shown. Both reference samples were performed in absence of DTT, and all values are normalized to pTDH3 expression levels. The points represented by a cross are considered as outliers. In the right plot, the UPR reporter is tested for expression of heterologous proteins of different secretory complexity based on whether it is secreted and the presence of posttranslational modifications. The proteins are expressed at different induction levels driven by the EL222 optogenetic expression system. The color intensity represents the real induction levels assessed using the internal fluorescence of the sensor strain in arbitrary units (au). The expressed proteins are mNeonGreen (secreted or not secreted), and secreted α-amylase from *Aspergillus oryzae* (*Genbank* ID: CAA31220), that requires one N-glycosylation and four disulfide bounds to be properly folded and secreted. All three proteins contain three copies of the FLAG tag peptide fused to the C-terminal end, in the case of the α-amylase it is prior a fusion with mNeonGreen. Secreted proteins are driven to secretion by the N-terminal fusion with the alpha-pre-pro leader peptide from the alpha factor from *S. cerevisiae*. ETF: Easy To Fold; DFT: Difficult To Fold.

I also tested the effect of proteins with different secretory complexity, and at different levels of expression, in the UPR reporter measured signal (Figure 3.12). The non-secreted protein, mNeonGreen, as well as the easy to fold (ETF), secreted mNeonGreen, do not trigger the UPR signal. In the case of a protein with a higher secretory complexity, due to its folding requirements, as for the protein α-amylase, the increase in UPR is obvious, and even saturates at high levels of expression.

To summarize, I have built a device to measure the secretion-associated stress by taking advantage of an existing solution called a UPR reporter. It is based on coupling the expression of a fluorescent reporter to the transcriptional activation of the stress response pathway. Once constructed, the reporter was characterized using a reducing agent that promotes protein misfolding, and expression of heterologous proteins with different secretory complexity and induction levels. The device was able to report stress levels by showing increased signal for a higher secretory complexity protein, with increasing signal at higher induction levels. Conveniently, the protein expressed is not secreted, or with special folding requirements the signal remained at basal levels.

3.2. Measuring secretion levels in a systematic manner

The amounts of protein produced and secreted by cells are key variables to assess the secretory efficiency. Detecting secreted proteins directly from the cell culture, in a quantitative, automated and standard manner for different proteins is challenging, and usually requires several steps. A common and simple approach to quantitatively measure secretion levels is by fusing a fluorescent reporter to the protein of interest (Huang et al., 2008; Besada-Lombana & Da Silva, 2019). However, since proteins targeted for secretion are inside the cells prior to being secreted to the extracellular environment, cells are also fluorescent, thus, quantifying secretion levels requires separating cells from supernatant. Alternatively, secretion levels can be quantitatively measured by methods such as enzyme activity assays, requiring protocols that are protein specific. Those constraints make standardization and automation of the measuring process for secreted protein levels a real challenge.

Our goal is to develop a generic approach to detect secretion levels of a wide range of proteins using a cytometer. Therefore, to implement the methodology that allows quantification of different secreted proteins using a single pipeline, and with a cytometer to perform the measurement, fluorescence seems the best option for protein detection. However, to measure the secreted delocalized proteins in a cytometer, it is first necessary to collect them, so that they reach a minimum size to scatter the beam of the device. To gather secreted proteins, we developed a method in which agarose microbeads bound with antibodies recognize and bind a common epitope in the different secreted POIs (Figure 3.13).

By incubating a sample of the culture with the immuno-beads, only the secreted proteins have access to bind the beads, whereas those inside the cell cannot interact with the antibodies. After incubation, the mixture of beads and cells passes through the cytometer. The distinction between cells and secreted proteins bound to the beads can be done by gating on the forward/side scatter ratio, since this property varies between cells and beads. Moreover, if the beads used are magnetic (Yang, Lien, Huang, Lei, & Lee, 2008; Su, Tang, Feng, Yao, & Chen, 2020), it is possible to separate beads from cells after incubation in an automated manner using the magnetic module of the *opentrons OT-2* robot.

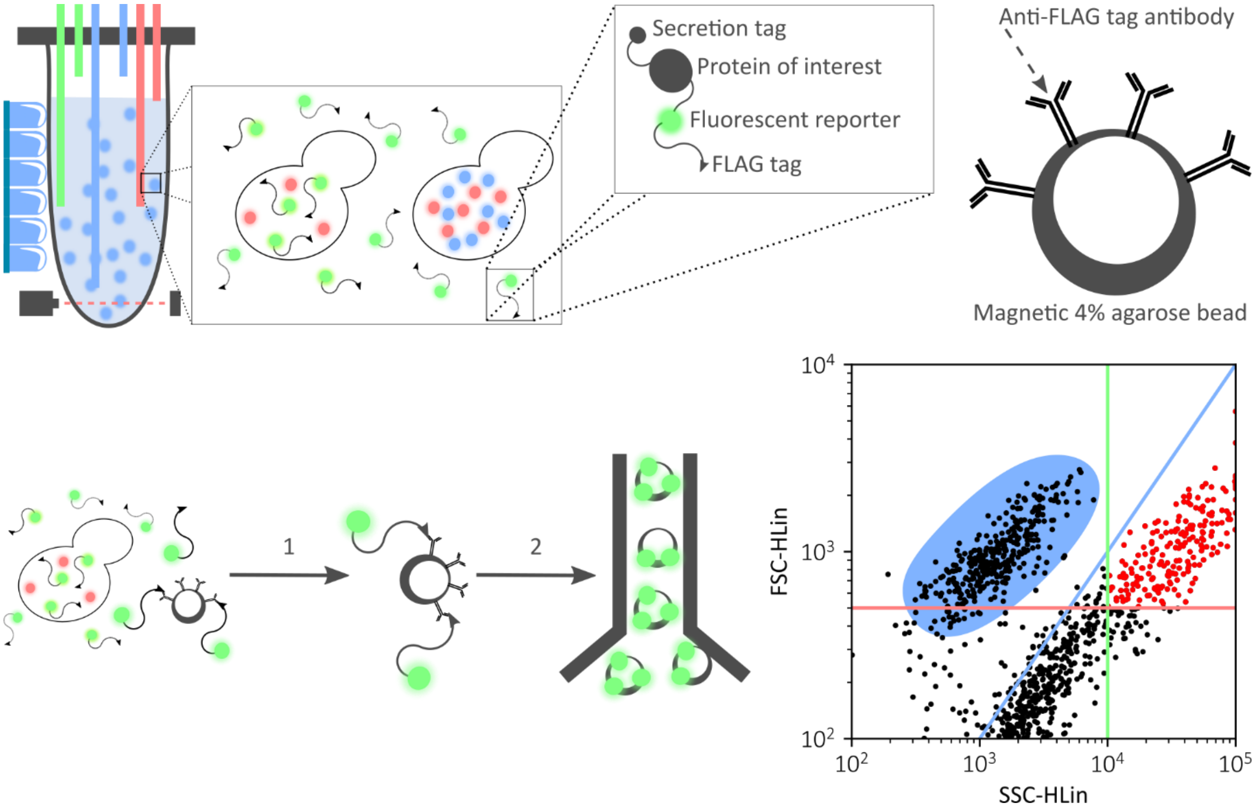


Figure 3.13. ***A generic pipeline to measure secretion levels of a range of proteins directly from the culture media*.** The top left figure represents the protein topology compatible with the method. The secreted protein of interest is fuse to a fluorescent reporter followed by the FLAG tag purification peptide. Such FLAG tag binds the monoclonal anti-FLAG antibodies attached to a 4% agarose magnetic bead (top right) that allows for its detection in a cytometer. As indicated at the bottom left representation, a sample of the culture is incubated with the magnetic beads, wherein only the secreted proteins have access to the anti-FLAG antibodies. After one hour of incubation, beads are partially washed using a magnetic rack, or module if performed in an automated platform (1). Once washed, the beads pass through a cytometer (2) and single bead fluorescence is detected. To distinguish beads from cells, different light scatter properties are taken into account. The plot at bottom right represents the forward light scatter (FCS-HLin) versus the side light scatter (SSC-HLin). The blue line separates the ratio (SSC/FSC > 10) that distinguish beads from cells. The green line restricts the gating to those events higher than the maximal SSC observed for cells (10^4). The red line shows the threshold set for beads binding capacity (500), below which beads show reduced fluorescence. The events within the blue area correspond to the majority of cells.

Following the reasoning stated above, we developed the pipeline to measure secretion levels in a generic manner using a flow cytometer. The secreted proteins of interest were fused to the fluorescent reporter mNeonGreen at the C-terminal end to detect the specific protein by fluorescence measurements. As a common epitope, to bind the target proteins that are secreted into the medium, but not those inside the cells, three consecutive copies of the FLAG tag were fused at the C-terminal end of the fused mNeonGreen (Figure 3.13). The procedure works as follows: (i) a sample of the culture is incubated with the magnetic beads containing the anti-FLAG antibody, (ii) then, the secreted proteins have contact with the antibody, while those still in the secretory pathway, inside the cell, do not. (iii) Using a magnetic grid, the cells are separated from the beads and the beads can be washed from the culture medium to pass through the flow cytometer.

For optimization of the method, different types and concentrations of beads were tested for sufficient binding properties and non-clogging of the flow cytometer capillary. To test the binding of the beads and the feasibility of the method, the first assays were performed with pure commercial GFP fused to the FLAG tag at C-terminal (Figure 3.14). The results show a linear relationship between GFP concentration and fluorescence readings on the cytometer, surprisingly over three orders of magnitude, with the minimum concentration in the linear range being 0.02 nM, and the maximum being 90nM of pure GFP.

Once we established a protocol to produce our own mNeonGreen fused to three copies of the FLAG tag in yeast batch cultures, the binding was assayed directly with the supernatant of the batch culture (Figure 3.14). Then, gating is required to differentiate cells from beads, even when using supernatants and washing beads after incubation. We observed that beads can be differentiated from cells by some light scattering properties. Therefore, three gating conditions are applied: (i) To discard cells, the minimal side scatter to consider events is restricted to 10^4 arbitrary units (au). To avoid beads with low binding properties, (ii) the minimal forward scatter is set to 500 (au), and to differentiate beads from cells (ii) the ratio side/forward scatter must be above ten.

In the case of batch culture supernatants, the concentrations of mNeonGreen are unknown, and the fluorescence signal is compared to the dilution factor of the supernatant in pure water. The linear range covers two orders of magnitude, but the measurements saturate at the non-diluted, and dilution factor four. However, in both cases, pure protein and supernatants, the spread of the events is too high due to systematic variability added by the heterogeneity of the bead features (Figure 3.14).

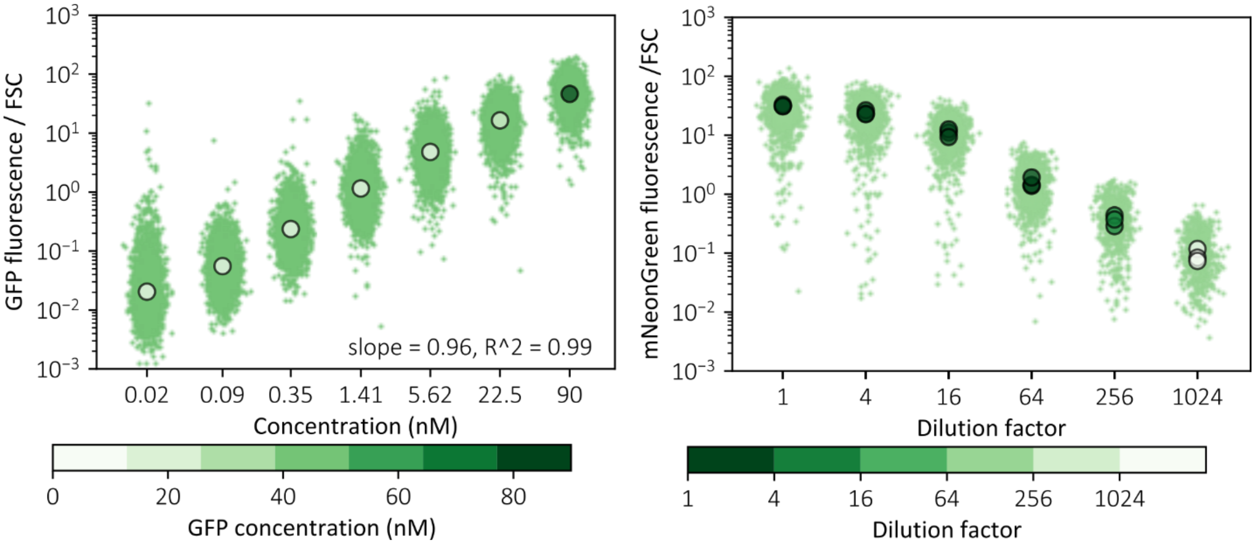


Figure 3.14. ***Relation of protein concentration with beads fluorescent.*** To the left the relation of pure commercial GFP-FLAG concentration and forward light scatter (FSC) normalized fluorescence per bead. The solvent used to dilute the GFP is pure water, with a total volume of 200µL. The linear regression parameters are shown in the lower right side of the plot. To the right the relation of diluted supernatant (200µL) of a batch culture of cells with constant induction of mNeonGreen-(3xFLAG) and forward light scatter normalized fluorescence per bead. The solvent used to dilute the supernatant is pure water. Medians of the population of each of three replicates are represented by the thicker dots; which color code in greens is proportional to the GFP or supernatant dilution, respectively.

The variability observed on the fluorescence of beads populations in each sample is high, two orders of magnitude in some cases. It affects the robustness and reproducibility of the method as seen for the batch supernatants wherein the median values between replicas is around 50% changing. This approach introduces several sources of variability related to the beads nature. The binding capacity may change among single beads in the same sample. It is related with the antibodies per bead, and it will affect the fluorescence per bead regardless the secreted protein levels. The incubation time might be different for different samples, since the cytometer measures one sample at the time, and it lengthens the incubation time of the last samples measured. Moreover, when manipulating beads mixtures, it was observed that there is an obvious heterogeneity in the concentration per sample. That is inherent to the pipetting process as a result of the viscosity of the beads mixture (Figure 3.15).

To account for the variability that is inherent to the method, we implemented ratiometric measurements. In such type of approach, a constant concentration of a second fluorescent reporter with the same binding properties as the protein of interest is added to each sample. That second fluorescent reporter is used as a reference, and it is subjected to the same binding constraints and sources of variability as the protein of interest. Therefore, when accounting for the ratio of the fluorescence of the POI to the fluorescence of the reference reporter, since all samples contain the same amount of reference protein, only the variation in the concentration of the POI makes a difference (Figure 3.15). The reference reporter is the mCerulean protein fused to three copies of the FLAG tag, and obtained from batch cultures. In Figure 3.15 is shown how the sample variability is now reduced significantly, and the linear relation between the dilution factor of the source sample and the ratio as improved and it increased up to four orders of magnitude.

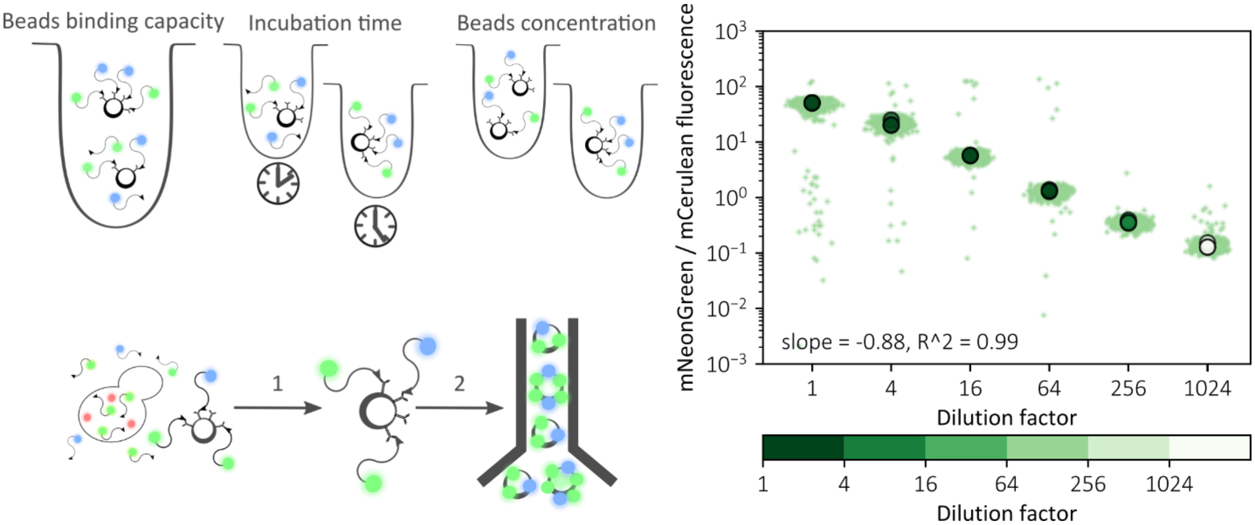


Figure 3.15. **Sources of variability inherent to beads method, and ratiometric measurements to assess secretion levels.** At top left the different sources of variability are explained, and the reference blue fluorescent protein accounting for it. Note that despite that those variability issues remain there, the ratio of green over blue fluorophores per bead is constant. On the left bottom, the protocol explained in *Figure 3.13* is depicted, but now including the reference fluorophore. To the right, the relation of diluted supernatant (200 µL) of a batch culture of cells with constant induction of mNeonGreen-3x(FLAG) normalized to mCerulean-3x(FLAG) fluorescence per bead. Per-cell deconvolution is performed for mNeonGreen and mCerulean fluorescence, and autofluorescence using beads only binding mCerulean is subtracted. The solvent used to dilute the supernatant is pure water. Medians of the population of each of three replicates are represented by the thicker dots; which color code in greens is proportional to the supernatant concentration. The linear regression parameters are shown in the lower right side of the plot.

As indicated, the previous presented data have been obtained from reporter produced in batch cultures, wherein the product accumulate in the media for the whole length of the production phase. The next step is to check whether this methodology is sensitive enough to measure the secretion levels from samples directly taken from the continuous culture bioreactors setup. That is more challenging because the continuous cultures are constantly diluting the culture to maintain OD, and thus the concentration in the vessels might be too low in comparison to the batch culture. To do so, I took a volume of the cell culture directly from the reactors at different levels of light induction, and after 24 hours of starting induction. Then, the exact same protocol and analysis pipeline as for the characterization experiments shown above was followed. As seen in Figure 3.16 the method is good enough to measure secretion levels in samples directly taken from the bioreactors platform. A linear relation between induction and secretion levels of a protein with low secretory complexity, mNeonGreen, is observed.

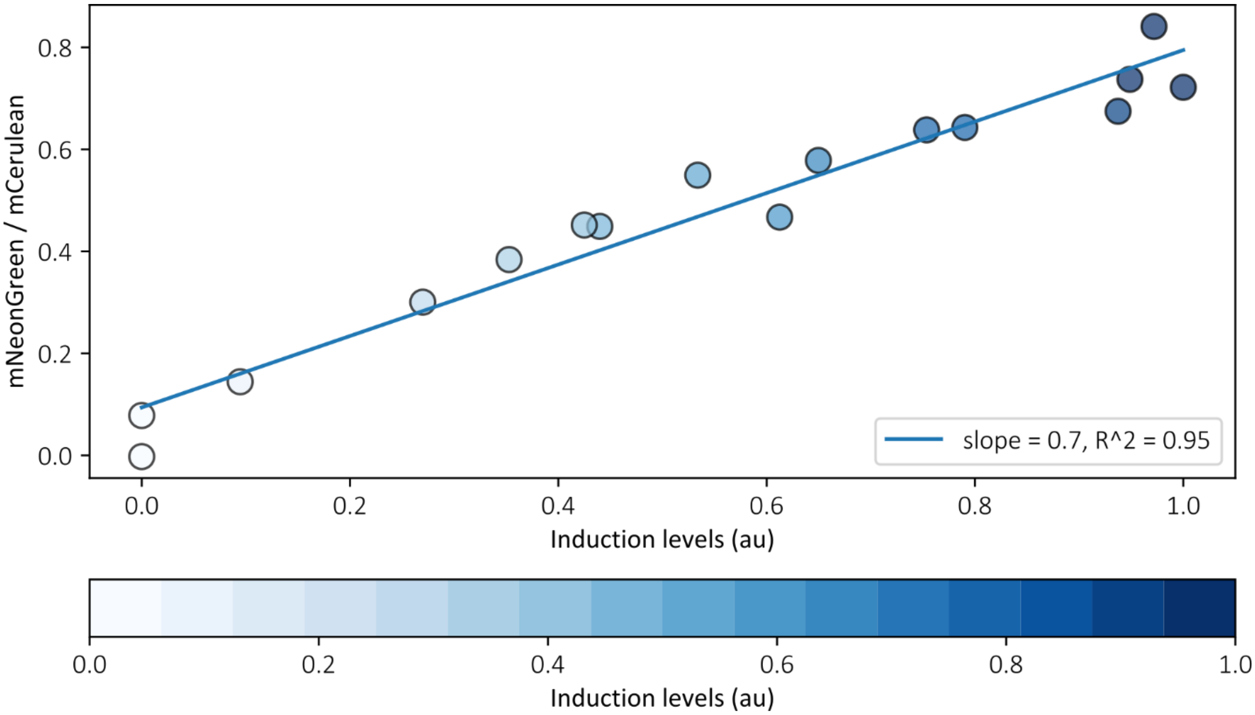


Figure 3.16. **Ratiometric measurements to assess secretion levels from m*ulti-bioreactors platform of continuous yeast cell cultures.*** Sixteen measurements taken in two different experiments of eight parallel bioreactors continuous cultures are shown. The measured samples are 200µL of culture media incubated with beads following the ratiometric measurements approach depicted in Figure 3.13 and Figure 3.15. Per-cell deconvolution is performed for mNeonGreen and mCerulean fluorescence, and autofluorescence using beads only binding mCerulean is subtracted. The secretion levels are normalized to the fraction of strain of interest present in the population with respect to the sensor strain, and to a non-diluted reference sample from a batch supernatant measured in each replica. The blue light intensity represents the real induction levels assessed using the internal fluorescence of the sensor strain in arbitrary units (au). The linear regression parameters are shown in the lower right side of the plot. Secretion levels are presented in arbitrary units.

In addition, to validate the approach using an alternative method to measure protein secretion levels, I produced a set of samples from continuous cultures of strains producing different secreted POIs under the same conditions explained above. Then, in parallel to the beads measurements, a fraction of the supernatants from each of these cultures was also measured by MALDI-TOF, a mass spectrometry technique in which peptides can be measured based on their mass-to-charge ratio. A set of reference proteins of known molecular weight and concentration was used to serve as a spike-in reference to estimate the mass of POIs in each sample. Comparison of the two methods yielded a similar qualitative interpretation. In particular, most of the proteins ranked in the same position according to their amounts when measured by mass spectrometry and beads. The results are not included in the main text.

The last step is integrating the protocol in the automated pipeline. Unfortunately, I didn’t have enough time to carry out such process before the finalization of the thesis. An unexpected constraint was detected and it needed a minor change in the hardware configuration of the platform. Since the process of incubation of beads with cell culture takes one hour, and maintaining OD constant is done every two minutes, it was needed to add an additional tube in the output lines of the reactor vessels to directly lead the trash tube independent of the robot arm. The reason for that is to avoid interference with the sampling process while the washing and sampling of beads is performed. It also required some changes in the general software pipeline of the platform to coordinate the processes. Due to the extensive use of this platform among the different projects carried out in our team, and the time needed to fine tune the protocol, it was very difficult to perform at the time. Therefore, in further experiments, in the next chapter, the samples analyzed and shown are taken by manual procedure at different time points of the experiment. It means that the sampling, the incubation, the wash, and the load of the sample in the cytometer is done by human. It restricts the number of time points that can be taken in 24 hours. The whole process takes around one hour and a half between the sampling and the end of the measurement, in which one hour is just incubation.

To conclude, in this section I described the method developed to measure secretion levels in a quantitative and standard manner for secreted proteins. By using fluorescent reporters, magnetic beads with monoclonal antibodies attached to them, and a cytometer, it is possible to measure secretion levels directly from the cell culture, without separation of supernatant. The method has been characterized and used in a real experimental configuration in the bioreactors platform, and has demonstrated to work robustly. The variability inherent to the handling process and beads nature has been also overcome by performing ratiometric measurements. However, the automation of the process has been incomplete because lack of time, and availability of the platform to carry out the necessary modifications and fine tuning. This method allows rapid and simple measurements of the secretory levels with high sensitivity and following a generic pipeline. In the next chapter it will be tested and used to characterize the secretory process of different proteins with various secretory complexity levels.

3.3. Strain construction strategy

Engineering biological systems is not a trivial task, genetic manipulation to add new cellular functions lead to unexpected results, and troubleshooting requires time and additional experiments. Having a modular cloning strategy, parallelizing processes, and performing thorough characterization of the biological modules is important for the strain construction efficiency (Lee *et al*, 2015). In this project, I had to generate several strains to characterize each of the different biological modules that are used in the final strains to achieve the general goal of the project. As seen in the previous sections of this chapter, those modules are: an optogenetic expression system, a stress reporter, and multi-domain proteins to measure secretion. All these modules have a function that has to be characterized and orchestrated to be merged in a final strain that contains all the functions. To do so I followed a rational strategy to parallelize modules construction, characterization and assembly in an efficient manner (Figure 3.17). In this section I describe the process and the rationale behind of the strain construction process.

One of the first decisions to be made is the host strain used to carry out the project. As mentioned in previous chapters, *S. cerevisiae* is a model organism for bioproduction used in myriad of projects (Nielsen & Jewett, 2008; Borodina & Nielsen, 2014). There are many strains of *S. cerevisiae* with different features and needs, depending on the goal of the study. In this project I decided to work with the strain BY4741, it is a deletion strain derived from the widely-used laboratory strain S288C (Mortimer & Johnston, 1986; Winston, Dollard, & Ricupero-Hovasse, 1995). The strain possesses specific features that are interesting in the context of this project. First, it is a widely used and relatively simple strain with a generic genetic background, and that makes easier to generalize the conclusions of our study. Second, it is a haploid strain and it facilitates the control of gene copy number when introducing genetic modifications, and allows an easier maintenance of isogenic populations under stressful conditions. Finally, it possesses a set of deletions that are compatible with the most spread integration cassettes, and it makes clones selection and genetic engineering approaches simpler and manageable. The strain genotype, in brief, carries deletions in the genes *MET15*, *URA3*, *LEU2*, and *HIS3* (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0)*. It makes the strain an auxotroph for the synthesis of those essential resources (methionine, uracil, leucine, and histidine). Therefore, it is possible to use integration of cassettes in those loci, or others, that re-introduce the deleted gene as genetic markers to complement those auxotrophies and select for clones harboring the integration cassette with the transcriptional unit of interest.

Regarding the cloning approach, I used, and extended with new parts, an already characterized Yeast Toolkit (*YTK*) for modular cloning procedure (MoClo) (Lee *et al*, 2015). It is a library of parts based on Golden Gate assembly and MoClo (Carola Engler, Romy Kandzia, 2008). Golden Gate assembly is based on the use of Type IIs restriction enzymes. They cut outside of the target recognition sequence, providing *a la carte* cohesive ends, which incredibly increases the flexibility of the cloning and allows multiplexing insertion in one-shot reactions. MoClo is based on step-cloning process, it allows interchanging different parts as desired by designating specific overhangs to each part type (Weber *et al*, 2011).

The *YTK* library, as most of restriction-based cloning methods, has some specific restriction enzymes used to perform the cloning and the following chromosomal integration. It means that the presence of restriction recognition sequences out of the cloning target sites generate undesired fragments and overhangs, negatively affecting cloning efficiency. Thus, to avoid those out-of-target restriction sites, and also to optimize codon usage for *S. cerevisiae*, the exogenous DNA used in most genetic manipulations and cloning in this project was designed to avoid that issue and artificially synthesized by external suppliers, as will be explained in Chapter 6. The codon optimization was done by using dedicated free online software (*IDT* codon optimization tool). The undesired restriction sites out of target were those belonging to *BsmBI*, *BsaI*, and *NotI* restriction enzymes. To avoid them, in the case of the coding regions, it was done by changing the codon to the second most frequent in *S. cerevisiae*, coding for the same amino acid as the original sequence.

To summarize, the strategy to carry out the cloning process was based on three main premises: (i) minimizing the cloning reactions, (ii) minimizing integrations, and (iii) characterizing modules in parallel and independently (Figure 3.17). That is possible if we consider three biological modules for the final strain: the optogenetic expression system, the stress reporter, and the heterologous secreted protein of interest. By integrating each of the modules independently in a different single locus, it is possible to minimize the cloning, since each transcriptional unit is cloned, integrated and characterized independently.

Following this rationale, I managed to make a small library of proteins of interest with several copies of each, as well as combinations of different proteins, testing different stress reporters, and all without the need of starting from the wild type strain. During the whole length of my PhD period I made more than 150 strains in a systematic manner following this principle. Moreover, it allowed other members of the lab to easily use these strains for its performance or further development in other projects (Bertaux et al., 2021; Fox et al., 2021; Aditya, Bertaux, Batt, & Ruess, 2021). The Figure 3.17 describes the process followed to generate all the strains used for the final purpose of this project.

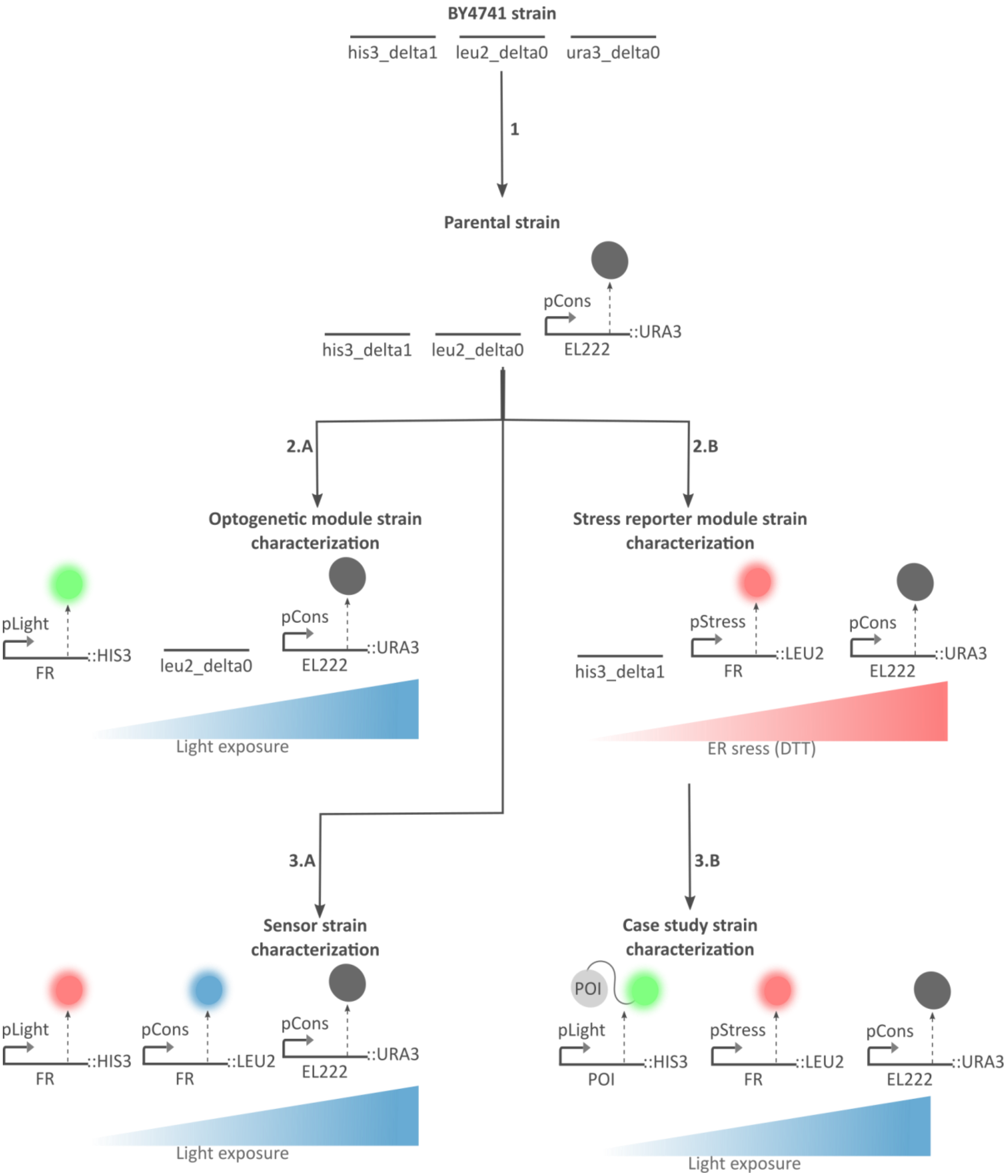

BY4741 strain
his3_delta1
leu2_delta0
ura3_delta0
1
Parental strain
pCons
::URA3
EL222
his3_delta1
leu2_delta0
2.A
2.B
Optogenetic module strain characterization
pLight
::HIS3
FR
leu2_delta0
pCons
::URA3
EL222
Light exposure
Stress reporter module strain characterization
his3_delta1
pStress
::LEU2
FR
pCons
::URA3
EL222
ER sress (DTT)
3.A
3.B
Sensor strain characterization
pLight
::HIS3
FR
pCons
::LEU2
FR
pCons
::URA3
EL222
Light exposure
Case study strain characterization
POI
pLight
::HIS3
POI
pStress
::LEU2
FR
pCons
::URA3
EL222
Light exposure

Figure 3.17. **Strain development and characterization pipeline.** Strain construction and characterization was multiplexed as shown in the workflow. First, (1) the common background was generated (Parental strain). Optogenetic and stress reporter modules were built (2.A, 2.B) and characterized in parallel. Then, the different case study strains were built from the stress reporter module strain (3.B) and in parallel with the sensor (3.A). Characterization of the case study strain required the sensor strain. In theory just three cloning batches would be required to produce all strains, since the sensor strain was not constructed until the full characterization of the optogenetic system.

3.4. Conclusion

In this chapter I have described the experimental methods carried out for the achievement of the main project of this thesis. That is to establish a generic framework to characterize the burden generated by heterologous protein secretion, and its potential applications to improve bioproduction efficiency. The technological developments are divided into two main components, (i) the multi-bioreactors platform enabling automated flow cytometry measurements, and (ii) secretion levels measurements pipeline, not yet automated. In the scope of the biological developments, three modules have been implemented to characterize the secretory process: first, (i) the optogenetic system to control protein production demand, including the use of a co-cultured sensor strain to measure actual induction levels; (ii) the stress reporter to monitor stress levels in real time, that also uses the sensor strain to monitor growth rate defect; and (iii) the engineered protein of interest allowing for quantification of intracellular and secreted levels by using fluorescence enabling systematic measurements in a flow cytometer. Both of them, technological and biological developments, work together and are exploited in the next chapter.

Regarding the multi-bioreactors platform, despite the few limitations exposed, as the variability on the induction levels, and the modifications needed to perform automated secretion measurements, its full potential and versatility is scarcely explored in this thesis. However, its development has been essential to yield the high temporal resolution in the experiments. For example, prior to the implementation of measurement automation, sampling was manual and therefore constrained to laboratory working hours, which limited throughput and required labor time. The development of the platform, together with the *ReacSight* strategy to exploit bioreactors capabilities by facilitating automated measurements and control experiments, is available in Bertaux *et al*, 2021.

The biological modules have successfully worked individually, as well as assembled together to generate Figure 3.12 right, thus they are ready to be used in the next chapter. The optogenetic system has accomplished all the requirements: it showed low variability, high expression levels, tight induction control, and no harm to cell physiology. The role of the stress reporter module in this project adds a lot of information for the interpretation of the secretory load. By analyzing the dynamic relationship of stress with variables such as growth rate, secretion levels or internal fluorescence, and using computational methods, cellular adaptation strategies can be evaluated, as we will see in the next chapter. That motivation led to including the stress reporter in the analysis, however, the UPR-based stress reporter was not the only one I considered.

I have characterized another reporter based on the yeast cell wall integrity (CWI) pathway. This one is also triggered by transcriptional activation, but in this case the transcription factor is Rlm1 instead of Hac1 (Sophie Mokas et al., 2009; García, Sanz, Rodríguez-Peña, Nombela, & Arroyo, 2016). The CWI stress response is triggered by secretory stress but through a different molecular mechanism. I planned to use it as a complementary stress reporter to gain additional understanding of the secretory associated stress. However, for the few heterologous proteins used for its characterization it turned out to behave as the UPR reporter in terms of dynamics. Moreover, due to the additional work load that including such reporter added, doubling the number of experiment and strains required, I finally discarded the idea of using it in this project. Thus, I chose the UPR based reporter, since there are solid and numerous publications using it, and the characterization here performed shows that it accomplishes the requirements for this project.

Finally, one limitation in the UPR reporter, difficult to overcome by using all three modules in the same strain, is the difference in the maturation time of mScarlet-I with respect to mNeonGreen or mCerulean (Guerra *et al*, 2021). It should be noted that red fluorophores have a slow maturation time, despite that mScarlet-I is one of the fastest, it takes around 30 minutes to maturate in *S. cerevisiae* at 30°C (Balleza *et al*, 2018). Therefore, it must be taken into account if comparing dynamics of processes monitored through these to fluorescent proteins.

As for secretion measurements, an enormous amount of time has been invested in their development, and they have been almost abandoned on several occasions. Finally, they have made an important contribution to the possibilities of the project. The method still shows limitations, most of them feasible to overcome, but requiring more time. The most obvious in the context quantitative biology are the units of secretion levels, and how to relate them quantitatively to intracellular fluorescence units. It should be noted that what is measured as secretion levels are actually ratios. These only correspond to the proteins that bind to the antibodies attached to the beads, and pertaining to a small fraction of the culture volume. Therefore, to quantitatively extrapolate such ratios to protein mass or concentration, some type of calibration curve is required to evaluate the absolute protein concentration from ratiometric measurements. I tried to implement such a calibration curve using mass spectrometry, the results were positive, but not extrapolated to the quantitative domain. Therefore, more experiments should be performed, at significant cost, to have an absolute quantitative characterization. Despite these details, the method is reproducible, so it is still worth using it carefully and interpreting its results with its limitations in mind.

In terms of the overall design of the approach and methods, there are things to improve, but overall it has demonstrated robustness and performance to thoroughly characterize the secretory process of heterologous protein from a quantitative perspective. The three fluorescent proteins used are stable in time and their expression do not affect cell growth, as shown in this project and in the project carried out in parallel using the same strains (Bertaux et al., 2021). General aspects can also be criticized, such as the fact that secreting fusion proteins (Figure 3.14) can disrupt cell physiology more than the native one.

Finally, in the next chapter, it is shown the exploitation and applied use of the technological developments together with biological modules developed here to characterize the secretory process from a quantitative perspective, and interpreting the results through the use of computational methods.

Chapter 4: Characterizing secretory processes for heterologous proteins

As seen in Chapter 2 and Chapter 3 yeasts are a common host for protein production in many contexts (Nielsen & Jewett, 2008; Borodina & Nielsen, 2014; Kerry R. Love, Dalvie, & Love, 2018). These organisms offer many advantages as being safe, well known, and many tools have been developed for engineering them. However, when referring to secretion of heterologous proteins in yeast, the process becomes complicated. Secretion is a multifactorial process that produces stress in different manners, to adapt to high demands of secretion, yeast cells trigger feedbacks to balance proteostasis. Moreover, monitoring the variables involved in the process in real-time and at single cell resolution is challenging. Different studies have reported evidence about the existence of an optimal level of demand to optimize secretion, but that level highly depends on the specific features of the protein of interest (Wittrup, Robinson, Parekh, & Forrester, 1994; Parekh & Wittrup, 1997). Therefore, there is a need for developing standard approaches to characterize the process in a generic manner.

Table 4.1. ***Panel of heterologous proteins of interest and their features***

Protein of interest	Group	PTM*	Native organism	Size (aa)**	*GenBank* ID
Non-secreted mNeonGreen[1]	1	N/A	N/A	332	BBB44438
Secreted mNeonGreen[1]	1	N/A	N/A	354	BBB44438
Endo-1,4-beta-xylanase C[2]	2	1x S-S	*Aspergillus niger*	662	EU848304
Single chain variable fragment 4M5.3[3]	2	2x S-S	N/A	609	1X9Q_A
α-amylase[4]	2	4x S-S 2x N-glyc	*Aspergillus oryzae*	832	CAA31220

* **PTM**: posttranslational modifications; **S-S**: disulfide bond; **N-glyc**: N-glycosylation. ** Includes, if present, the secretion tag, the protein of interest, the fluorescent reporter, and the three FLAG tag copies. [1](Shaner *et al*, 2013), [2](Boder *et al*, 2000), [3](Do *et al*, 2013), [4](Liu *et al*, 2012)

In this chapter I am going to present a systematic pipeline to characterize the secretion of different proteins of various folding complexity, a feature that significantly influences the secretory process. First, I will describe and validate the method for the secretion of a group of two proteins with low secretory complexity (Table 4.1, Group 1). Then, once the pipeline is established and quantitatively analyzed, a second group (Table 4.1, Group 2) of proteins will be thoroughly characterized following the same workflow, and extending it to capture features not overserved in the first group of proteins. Then, by integrating the results of the characterization, I obtain an overview of the different processes involved in secretion, I call it secretory pattern. Finally, I will use such knowledge to extend its potential to the biotechnological context, and show preliminary data of an application using the real-time control capabilities of the multi-bioreactors platforms presented in Chapter 3.

4.1. Validation of the approach: Characterizing the secretory process for low secretory complexity proteins

4.1.1. mNeonGreen as model protein

To validate the data generation and analysis pipeline for characterizing the secretion process, I first use two strains expressing the same heterologous protein secreted or non-secreted. Different levels of production demand are applied by controlling the expression levels through optogenetic light induction. Several variables of interest, such as growth rate, stress, internal and secretion levels for the heterologous protein are monitored in real-time. A coarse-grained mathematical model is then applied and its output is fitted to the data to infer the parameter values and simulate the system behavior. The use of such a model is not limited to describing the behavior of the system, but also to analyze and better understand it. Once the experimental pipeline and model are validated for low secretory complexity proteins (Table 4.1, Group 1), both are used to characterize the secretion of heterologous proteins with increasing secretory complexity.

For the experimental validation, eight parallel continuous cultures take place in the multi-bioreactors automated platform presented in Chapter 3. A turbidostat program is run to maintain constant cell density. It works by momentarily opening the input and output lines when the culture passes a certain optical density (OD) to slightly dilute the sample and keep the volume constant. In these experiments, two strains expressing mNeonGreen secreted or non-secreted, respectively, are used (Table 4.1 and Figure 4.1). mNeonGreen is a heterologous protein with rather low secretory complexity, as it is small and has not posttranslational modifications. In both cases the protein of interest, mNeonGreen, is fused to three copies of the FLAG purification peptide in C-terminal. In the case of being secreted, the alpha-pre-pro signal peptide is fused to the N-terminal extreme of the protein. The secretion-associated stress is measured by the growth rate data using the sensor strain, and the fluorescence signal from the UPR reporter, both presented in Chapter 3. Heterologous protein expression is driven be the optogenetic expression system EL222, presented in Chapter 3. Note that mNeonGreen is the fluorescent reporter that is used fused to all the rest of proteins in this study, thus it serves as a good “proof of concept” protein for the validation of the pipeline.

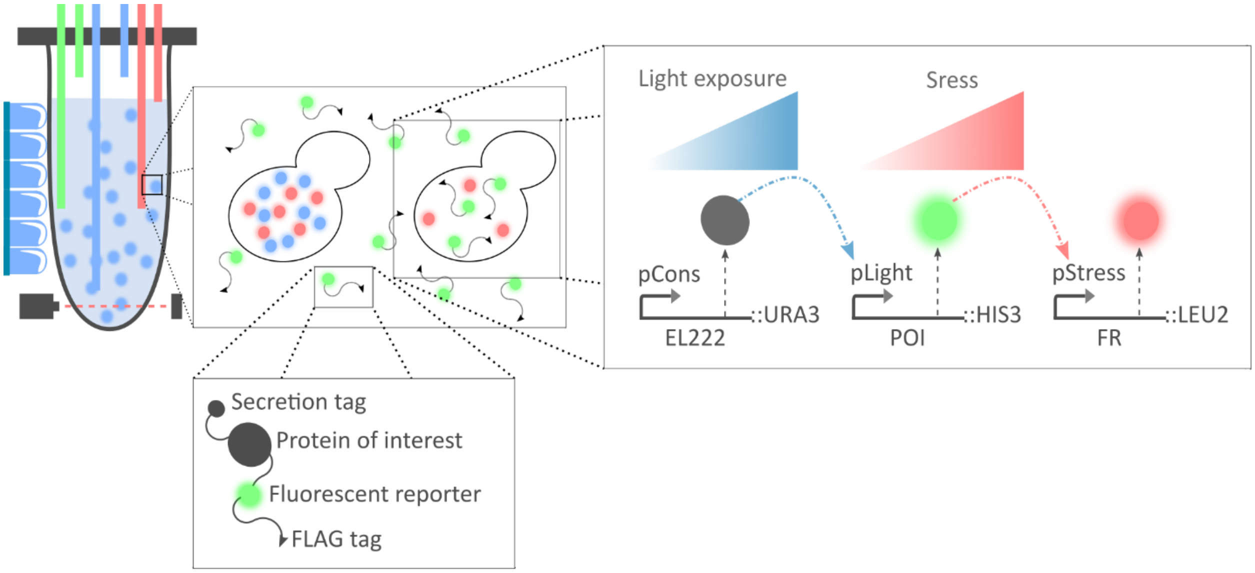


Figure 4.1. ***Experimental platform, strain genome, and heterologous protein of interest schematic.*** A co-culture with the sensor and the case study strains is run in the automated multi-bioreactors platform (top left). The case study strain harbors three transcriptional units chromosomally integrated in different *loci*, URA3, HIS3, and LEU2, respectively (top right). Those integrations confer optogenetic control of transcriptional activation (pLight) of the gene coding for the protein of interest (POI). Upon secretion-associated stress, a red fluorescent reporter (FR) expression is driven by activation of the stress controlled promoter (pStress). The POI (bottom left) is fused to a green fluorescent reporter, mNeonGreen, followed by three copies of the FLAG tag to allow detection and purification, respectively. In case of being secreted, a secretion tag is fused to the C-terminal extreme of the POI. In the case of protein from the group 1 (Table 4.1), the POI is mNeonGreen itself, thus no fluorescent reporter is fused, therefore, the sole module lacking is the Protein of interest depicted in the figure.

Eight parallel experiments are performed per POI, each of them corresponds to a different production demand level. For some POI there are two sets of eight parallel experiment, in total 16, that belong to different day repetition. Note that no real replicas are performed because the levels of induction change from one set of parallel experiments to the other, due to reactor-to-reactor variability exposed in Chapter 3. Light induction is applied for 24 hours and automated cytometry measurements are programmed to perform reads every 45 minutes. The secretion measurements are manually taken at line, as described in Chapter 3, at 24 hours after light induction. In some cases, secretion measurements are also carried out six hours after the start of induction to check whether steady state has been reached (Figure 4.2). Finally, note that the median value is used for the analysis of data produced in the cytometry platform, the internal fluorescence and secretion levels. That decision is based on the fact that this statistic is expected to represent the most common behavior of the population.

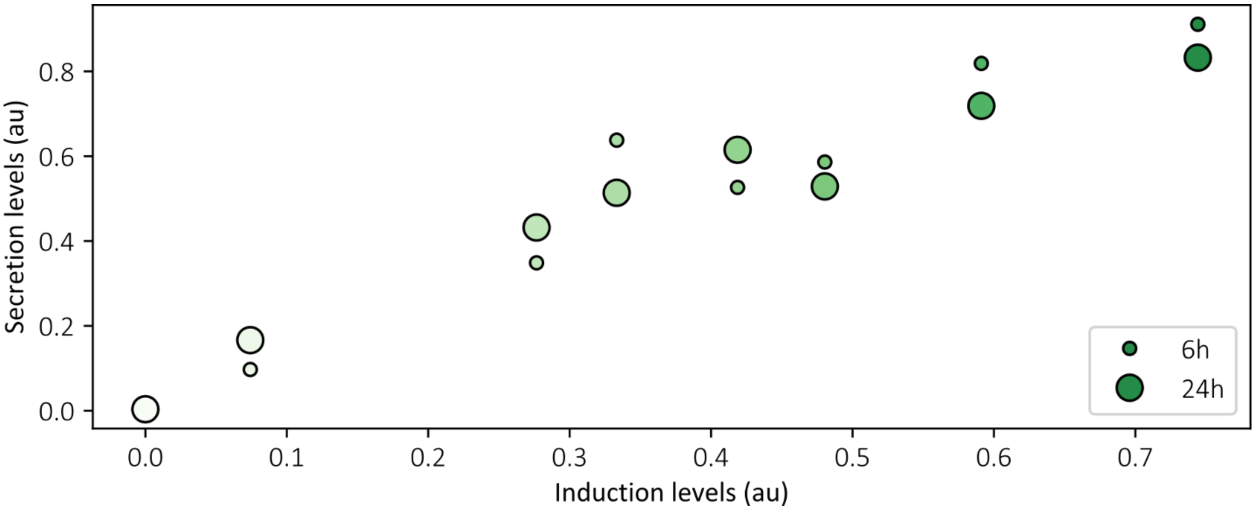


Figure 4.2. ***Comparison of secretion levels measured at different time points for mNeonGreen***. The plot shows the two measurements from height parallel experiments at different induction levels. The points correspond to the median of the beads population measured by the ratiometric approach explained in Chapter 3. The intensity of the color corresponds to the induction levels. The size of the points is used to differentiate between the measurements taken at 6 or 24 hours after induction.

Once the experimental workflow is clear, the data generation is carried out. It is important to mention that for secreted mNeonGreen two sets of experiments were produced. During the second repetition an unexpected error occurred with the platform, therefore the data has a gap between 10 and 24 hours after induction. However, it didn't affect the results in regard to data analysis because the steady state is reached before 8 hours after light induction, and the lasts time points at 24 hours were possible to take.

Figure 4.3 summarizes the steady state data and the dynamics of growth rate obtained from the experiments described above for non-secreted and secreted mNeonGreen. In both cases no growth rate decay is observed, the computed average growth rate values from data are 0.392 ± 0.004 h^{-1} for non-secreted, and 0.394 ± 0.016 h^{-1} for secreted mNeonGreen expression. Those values correspond to a division time of around 107 minutes, which is in agreement with the typical yeast growth rate in synthetic complete media (Waldron & Lacroute, 1975; Orij et al., 2012). This is not surprising in the case of non-secreted mNeonGreen expression, since it is known that not secreted proteins expressed in one copy have minimal impact in yeast growth rate (Kafri *et al*, 2016). In the case of secreted mNeonGreen, the absence of growth decay comes with no surprise because the protein is relatively small, and easy to fold, therefore it shouldn't create bottlenecks in the pathway.

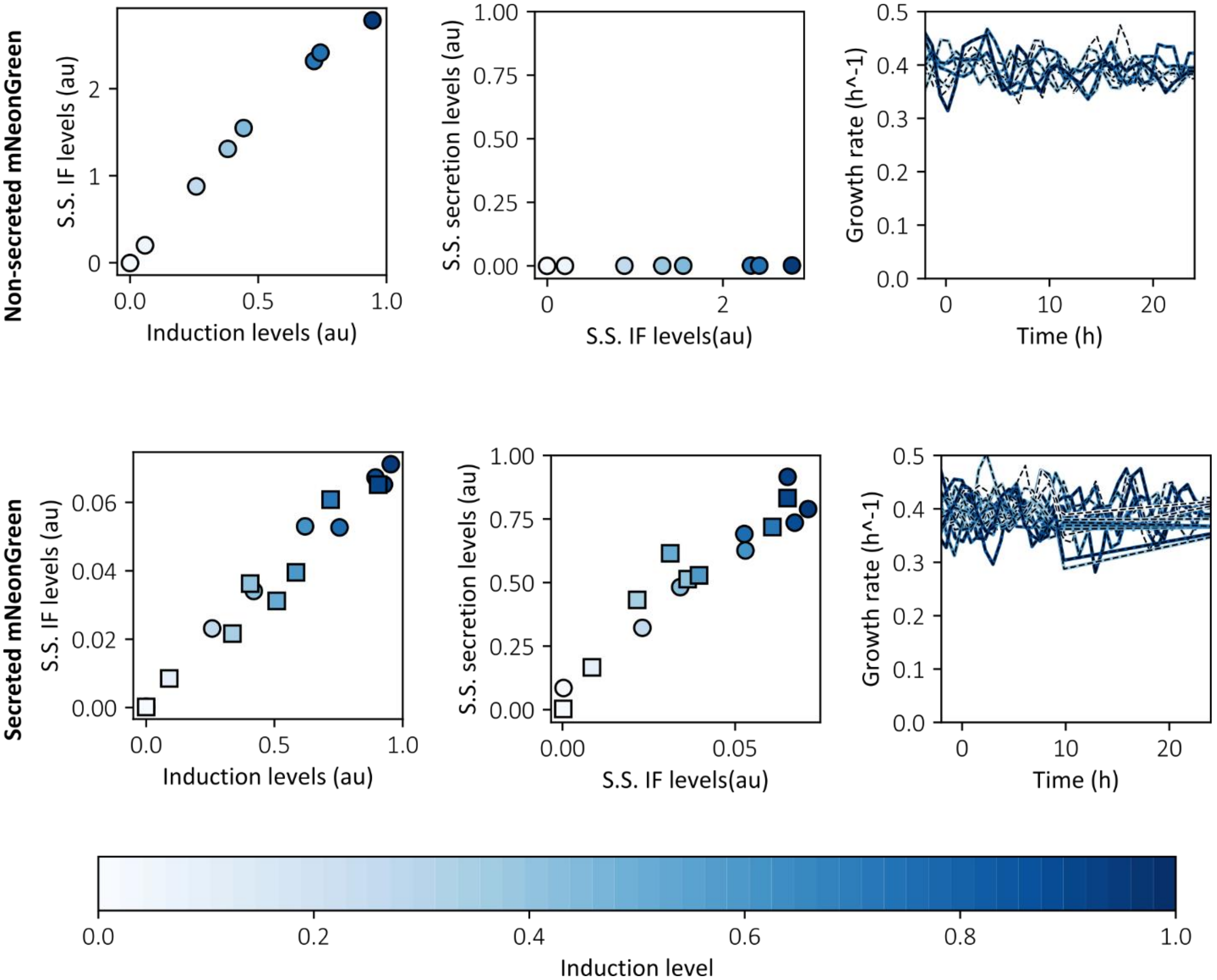


Figure 4.3. ***Data summary for mNeonGreen secreted and non-secreted expression.*** Top row correspond to non-secreted, bottom row to secreted mNeonGreen data. The square data points correspond to a different set of eight parallel experiments. The first column shows the median of cellular internal levels of fluorescence at steady state for different induction levels. The second column shows the median secretion levels at steady state obtained by ratiometric measurements with beads after 24 hours induction. In the third column, the computed growth rate over the length of the experiment using the sensor strain are shown for each experiment. Note the gap on secreted mNeonGreen growth data, between 10 and 24 hours, mentioned in the text due to a technical issue.

Given that secreted proteins must be translocated and folded into the ER upon translation, the machinery performing such functions could saturate, and it would impact the levels of internal fluorescence observed. Thus, the presence of bottlenecks in the production part of the pathway can be deduced from the relationship between production demand and internal levels of heterologous protein. Comparing the normalized internal levels for both POIs at steady state, using the non-secreted protein as a reference for non-saturation, a linear relation is observed (Figure 4.4). That suggests that cellular machinery in charge of producing heterologous protein is not overwhelmed, and presents full capacity for the whole range of production demanded. The same type of rationale can be applied to the secretion part of the pathway; the internal levels of POI should show a linear relation with the secretion levels in case of no bottle necks in the secretion machinery. However, looking at the secretion levels as a function of the internal fluorescence, only for secreted mNeonGreen, the linear relation is not as evident, as it might be a small plateau for high internal levels of protein.

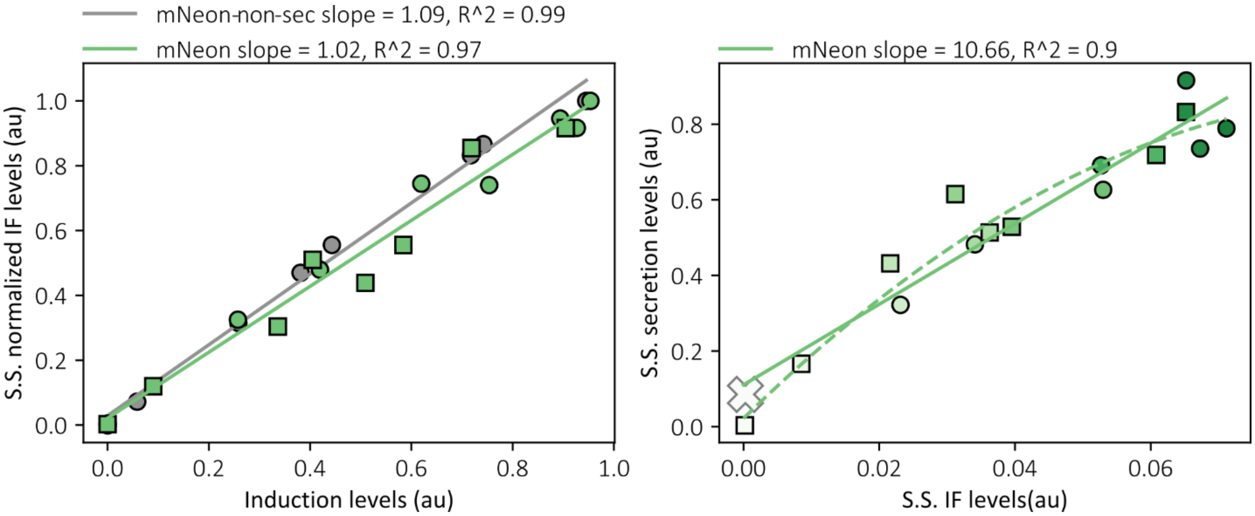


Figure 4.4. ***Relation of variables to check system's saturation.*** To the left the relation of internal levels of fluorescence (IF) at steady state (S.S.) with the induction levels. The IF for secreted and non-secreted mNeonGreen is normalized to their maximums, respectively. The gray dots correspond to non-secreted mNeonGreen, the green points, squares or dots, for the secreted protein. Linear regression parameters are indicated above the chart. To the right, the relation of steady state secretion levels with the IF levels, also at steady state. The point marked with a white "X" indicates a data considered as outlier, because it does not start at zero, and it is not accounted in the linear regression. The dashed line represents a polynomial fit of degree two.

To further analyze the secretion process, taking into account the dynamic behavior of the system, I implemented a simple mathematical model based on ordinary differential equations (ODE) in order quantify the POI production and removal parameters values. As we will see, the parameters values are then useful to discover processes that are not accessible by only analyzing the steady state data. The model described in Figure 4.5 is a high level representation of the secretory process. It is divided in three main dynamic components, (i) the production demand that is applied at the level of transcriptional activation by the optogenetic expression, (ii) the process involving the heterologous protein synthesis stages previous to fluorophore maturation, and the (iii) removal processes including the secretion, and the dilution of the POI due to growth.

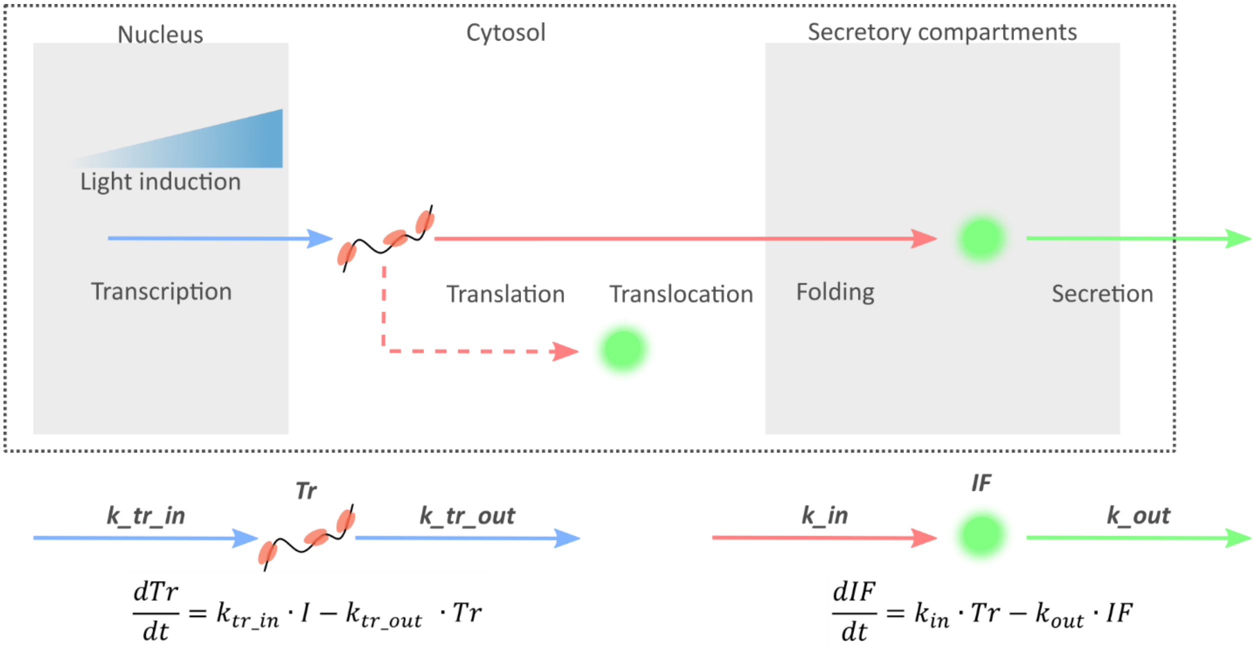

Nucleus
Cytosol
Secretory compartments
Light induction
Transcription
Translation
Translocation
Folding
Secretion
Tr
k_tr_in
k_tr_out
IF
k_in
k_out
$\frac{dTr}{dt} = k_{tr\_in} \cdot I - k_{tr\_out} \cdot Tr$
$\frac{dIF}{dt} = k_{in} \cdot Tr - k_{out} \cdot IF$

Figure 4.5. ***Model of the secretory process.*** The rectangle delimited by the dotted line represents the cell, in gray boxes the cellular compartments as the nucleus, and those related with secretion, for instance, endoplasmic reticulum, Golgi apparatus, or vesicles. The cytosol is the rest of the internal area of the schematic cell. The blue arrows indicate the process representing the demand of production. That includes processes as transcriptional activation driven by light thanks to the EL222 optogenetic system. It generates a specie called *Tr* that represents the transcripts coding for the protein of interest. Then, the *Tr* is used in a gratuitous way to generate the internal fluorescence *IF*. This process is represented by the red lines and includes translation, translocation, folding, and maturation of the fluorophore. In the case of a not secreted protein, dashed red line, the process includes less steps. Finally, the *IF* is removed of the system by processes as secretion and dilution due to growth, that are represented by the green arrows. Note that *IF*, either secreted or not, is also removed proportionally to its quantity by dilution driven by cell growth (not represented in the scheme), such process is also included in k_{out}. The representation of the bottom depicts the kinetics of the system for *Tr* to the left, indicating the kinetic parameters and the differential equation describing its temporal evolution. To the right the same schematic principle for the *IF* dynamics.

The first part of the model introduces the transcript (*Tr*) variable, it has its own dynamics that are independent of the internal fluorescence observed. The production of transcripts, simulates the delay observed between light induction and fluorescence increase. The *Tr* is produced at a rate that is modulated by the induction levels (*I*), and degraded at a rate that is proportional to the transcript levels:

$$\frac{dTr}{dt} = k_{tr\_in} \cdot I - k_{tr\_out} \cdot Tr$$

For simplicity, when inferring parameters because the lack of experimental data accounting for *Tr*, I assume that $k_{tr\_in} = k_{tr\_out}$, from now on both called k_{tr}. Note that k_{tr_in} and k_{tr_out}, have same value but different units. When *Tr* is at steady state:

$$\frac{dTr}{dt} = 0 \rightarrow$$

$$\rightarrow 0 = k_{tr} \cdot I - k_{tr} \cdot Tr_{ss} \rightarrow Tr_{ss} = \frac{k_{tr} \cdot I}{k_{tr}}$$

$$\rightarrow Tr_{ss} = I$$

The second differential equation of the model represents the dynamics of the internal fluorescence (*IF*):

$$\frac{dIF}{dt} = k_{in} \cdot Tr - k_{out} \cdot IF$$

The process has two contributions, first the production of *IF* proportional to the quantity of transcripts. Second, the removal of *IF* from the system, which is proportional to *IF*. In the first, the parameter k_{in} accounts for processes such as translation, since this part of the model describes the production of internal POI from transcripts in a gratuitous manner. In the case of secreted proteins, k_{in} also accounts for translocation into the ER, folding, fluorophore maturation, and probably other steps that are part of the secretory processes previous to the *IF* increase. For the *IF* removal process, the parameter k_{out} accounts for dilution of cell content lead by cell growth and division, as well as for secretion in the case of the secreted POI. This is the case because I assume that secreted mNeonGreen is not actively degraded, since the UPR signal, driving the ERAD, is minimal. As a matter of fact, to justify the assumption, Figure 4.6 shows how the secreted mNeonGreen triggers the UPR at the same levels as the not secreted one.

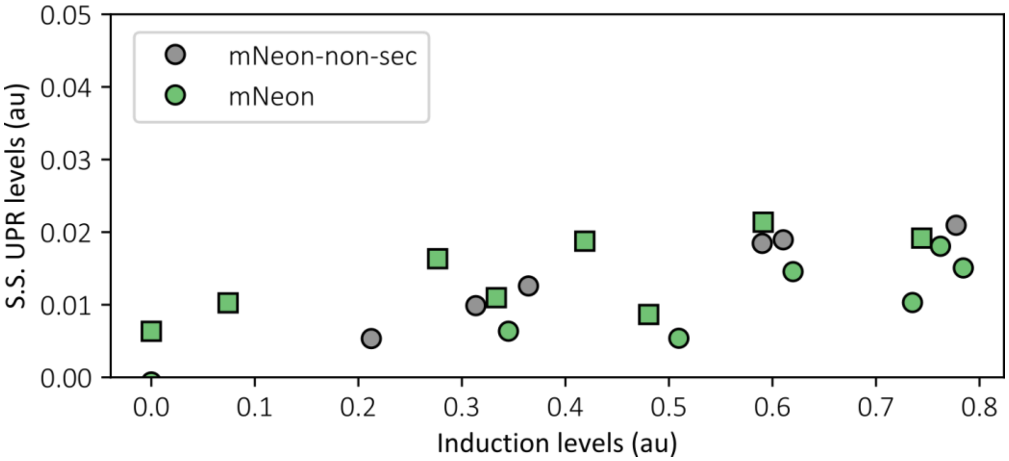


Figure 4.6. ***UPR signal triggered by secreted and non-secreted mNeonGreen.*** Steady state levels of the unfolded protein response (UPR) reporter signal at steady state for different levels of induction of a protein with low secretory complexity, mNeonGreen, and the not secreted version of the same protein. Square markers represent data collected in a different set of experiments.

Inferring parameters values is informative to understand the system because they can be used as proxy for biological processes. Its analysis can bring information about capacity of the secretory system under different conditions, the presence of the bottlenecks, and the adaptation mechanisms. Thanks to the high temporal resolution dynamic data here presented, allowed by the automation of the platform, it is possible to extract more information than when restricted to the steady states. For example, at the moment just after induction is triggered, *IF* tends to zero, and k_{in} is the parameter determining the system's evolution together with *Tr*. Therefore:

$$\frac{dIF}{dt} = k_{in} \cdot Tr - k_{out} \cdot \lim_{t \to 0} IF(t) \rightarrow$$

$$\rightarrow \frac{dIF}{dt} = k_{in} \cdot Tr$$

Then, as the time increases it is k_{out} the parameter determining the systems behavior. To understand how the model proposed describes such dynamics I need to integrated the differential equation:

$$\frac{dIF}{dt} = k_{in} \cdot Tr - k_{out} \cdot IF \rightarrow$$

$$\rightarrow IF(t) = \ IF_{ss} \cdot \left(1 - e^{-k_{out} \cdot t}\right); \ IF_{ss} = \frac{k_{in}}{k_{out}} \cdot I$$

So, as the equation above shows, it is k_{out} the parameter determining the needed time to reach steady state. And the model also assumes that IF_{ss} is proportional to I:

$$\frac{dIF}{dt} = 0 \rightarrow 0 = k_{in} \cdot I - k_{out} \cdot IF_{ss} \rightarrow$$

$$\rightarrow IF_{ss} = \frac{k_{in}}{k_{out}} \cdot I$$

Using this information one can precisely identify the value of both parameters, whereas it is not possible by just looking at steady state experimental data. To infer parameters value from dynamic data, I use the global optimization tool CMA-ES, also named covariance matrix adaptation evolution strategy (Hansen & Ostermeier, 2001). It has been already applied to perform parameters inference through model fitting in other research projects (Uhlendorf et al., 2012; Lugagne et al., 2017). The rationale behind the approach is to use the optimization algorithm to infer those parameters values that best fit the model simulation with the experimental data. The software searches for parameters to minimize the value of a cost function that computes the residuals between the model output and the real data. Regarding the cost function, it should be noted that, as the residual increases with the *IF* value, to give the same weight to the cost of all experiments, it is computed from the *IF* values normalized to the steady state levels for each experiment.

One aspect to take into account is the model fitting strategy for each of the different POIs data sets. It can be done either in a joint manner for all induction levels of a given POI, so each set of data has same parameters value, or inferring parameters individually for each experiment. This is an important decision because on the one hand, inferring parameters in an experiment-specific manner could introduce discrepancies among samples if there are unidentifiable parameters. On the other hand, such approach is very useful to identify how different parameters values evolve under different production demands. Regarding the joint-fitting, it would work well, given the model proposed, only if the relation between *IF* and induction levels actually evolves linearly. Otherwise it may omit interesting features of the system such as the presence of bottlenecks.

Given the data and conclusion extracted from Figure 4.3 and Figure 4.4, suggesting that the secretory system maintains its full capacity across all demand levels, because the linear relation between induction and internal fluorescence. The fitting is first done for all experiments together, in a joint manner, for each POI. Then, I will perform experiment-specific fitting, and compare both approaches to choose the most precise based on quality of the simulation respect to the data.

To test for the suitability of the model, we first fit the data of the non-secreted mNeonGreen. Given that the sole source of *IF* removal in that case is assumed to be dilution due to growth, the inferred k_{out} is expected to be equal, or very close, to the growth rate observed experimentally. As it is indeed the case (Table 4.2), k_{out} parameter is fixed to 0.4 h^{-1}. Note that in the case of k_{tr}, the inferred value is consistent with that corresponding to a half-life of 20 minutes for *Tr* (Table 4.2). That value has been reported in different studies as the mean half-life of mRNA in *S. cerevisiae* (Wang et al., 2002; Geisberg, Moqtaderi, Fan, Ozsolak, & Struhl, 2014). Thereby, from now on the value for k_{tr} is fixed to the corresponding to a half-life of 20 minutes. Thus, I assume no saturation at the level of transcription, as it has been observed in the characterization of the EL222 optogenetic system in Chapter 3. In addition, since to the best of my knowledge, general transcription regulation is independent of the expressed heterologous protein and the secretory pathway stress levels in *S. cerevisiae* (Travers et al., 2000; Buchberger, Bukau, & Sommer, 2010), the value of k_{tr} is going to be fixed for the fittings of the different POIs in this study.

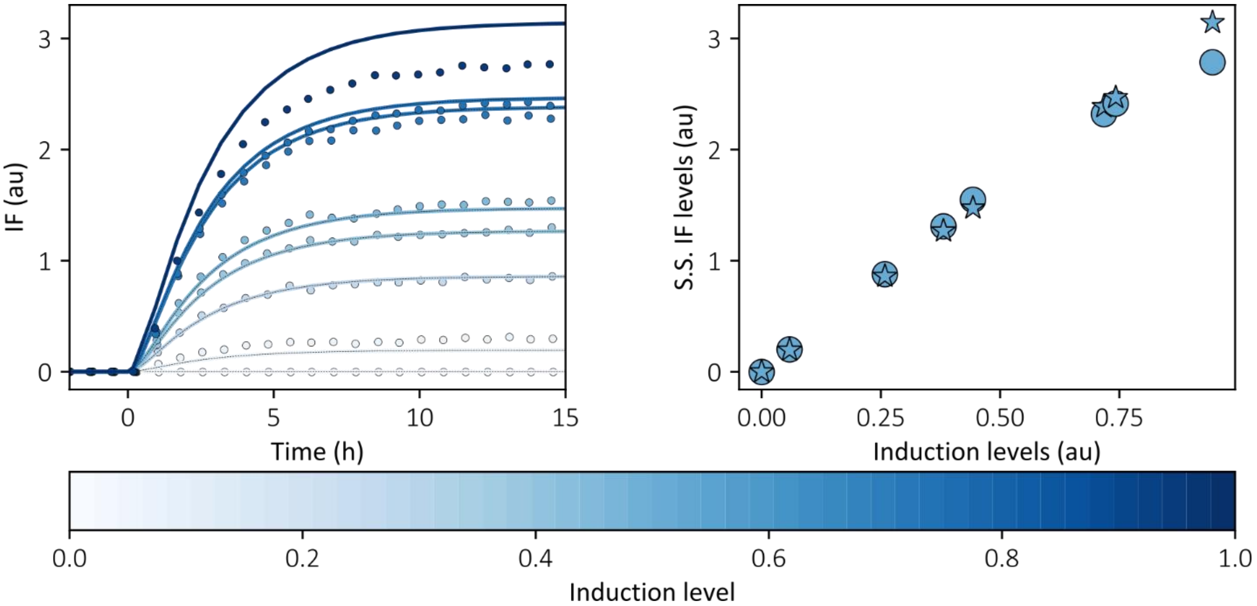


Figure 4.7. ***Comparison between data and simulation of non-secreted mNeonGreen expression.*** Parameters values have been obtained by joint-fitting strategy. To the left the dynamics of internal levels of fluorescence represented by dots, and the simulation results for *IF* using parameters shown in Table 4.2, Fit 2. To the right the steady state levels for data, represented as circles, and simulation as stars. The fitting is done for the time points between -2 and 15 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes, k_{out} is fixed to 0.4 h^{-1}.

In Figure 4.7 it is shown the fitting of non-secreted mNeonGreen, with k_{tr} and k_{out} fixed. The steady state levels fit well for all experiments except for the highest demands. The prediction for the dynamics is precise for intermediate levels of induction but it fails at the lowest and highest. For low levels of demand, the model gets short in *IF* as compare to the data, and the opposite is observed for higher levels. It might indicate that there is some over capacity to produce at low levels of demand that the model cannot capture. In contrast, the capacity to produce might be shortened for higher levels of demand. As mentioned above, if that is the case, the joint-fitting approach cannot capture those changes in parameters value across experiments. This could be attributed to a myriad of different biological processes, and it is difficult to estimate just by looking at the model fitting. But as an example, this observation could suggest that it due to ribosomal availability, being more than enough at low levels of expression, and scarce at higher levels.

The same strategy of fitting the model to all experiments jointly with the same set of parameters for a given POI is applied for the secreted mNeonGreen data. As indicated above, the parameter k_{tr} is fixed to the value corresponding to a 20 minutes half-life, as in the non-secreted protein case. The Figure 4.8 shows that the fitting is good for most of the cases. As observed in the case of the non-secreted protein, at least regarding the dynamic behavior, preceding the steady state, the model results are lower than data in the case of lower demands, and *vice versa* for higher levels. For the steady state levels, the simulation yields a linear relation that fits very well the data.

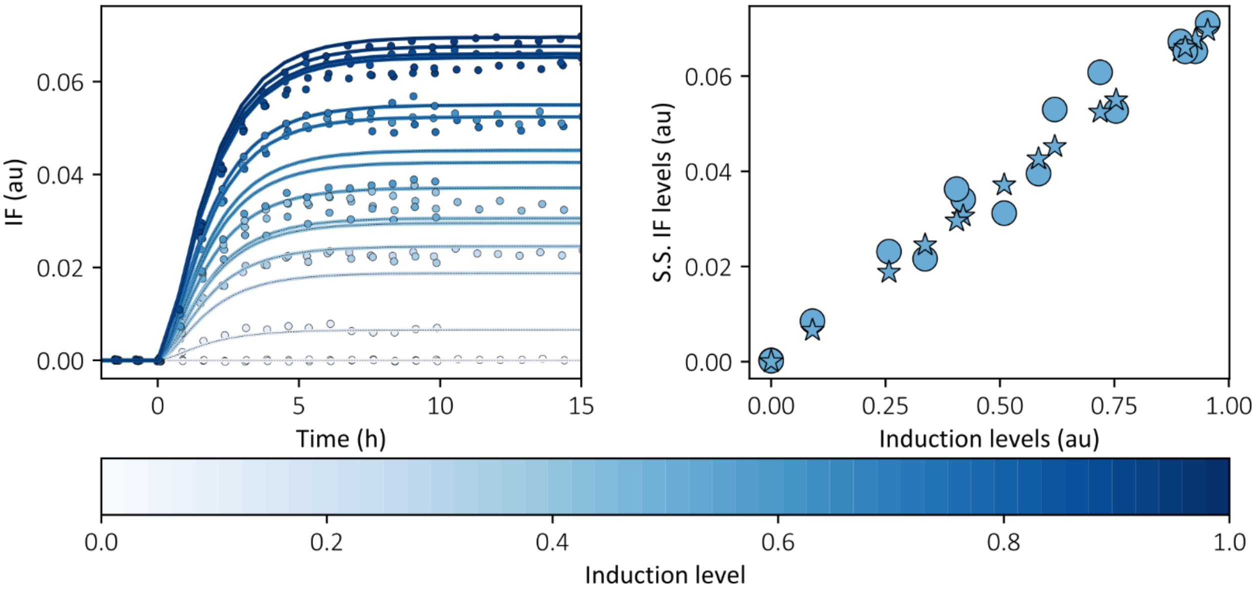
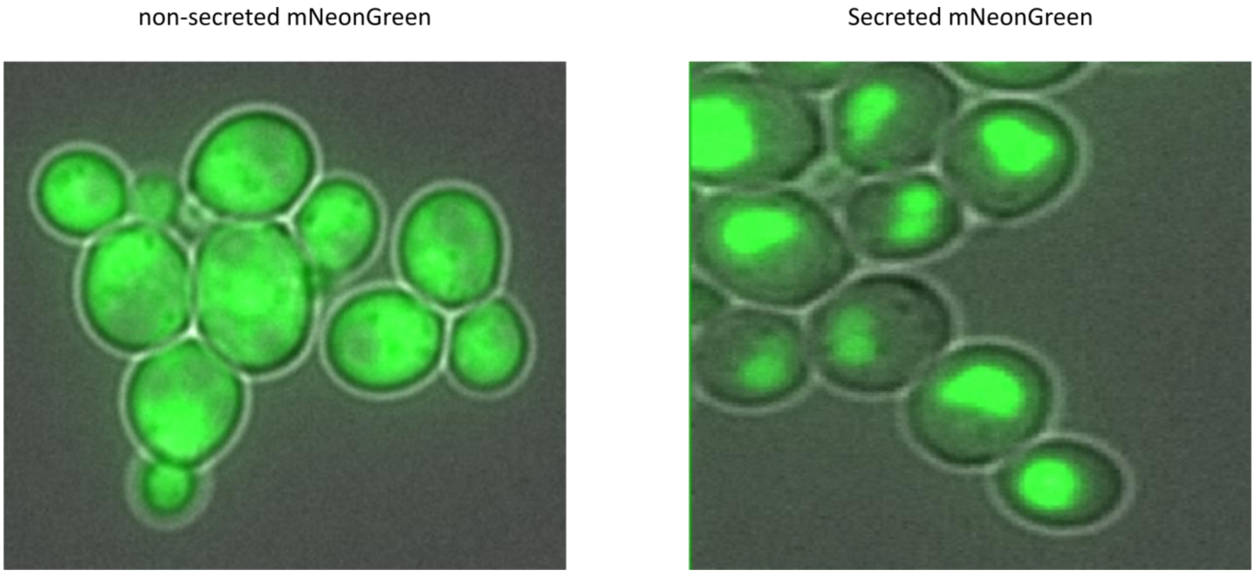


Figure 4.8. ***Comparison between data and simulation of secreted mNeonGreen expression.*** Parameters values have been obtained by joint-fitting strategy. To the left the dynamics of internal levels of fluorescence represented by dots, and the simulation results for *IF* using parameters shown in Table 4.2, Fit 1. To the right the steady state levels for data, represented as circles, and simulation as stars. The fitting is done for the time points between -2 and 15 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

Importantly, in the case of secreted mNeonGreen, since the growth rate is known, it is possible to assess the secretion rate by subtracting it, 0.4 h^{-1}, to k_{out}. The resulting value is 0.25 h^{-1}, a rather low secretion rate, being even slower than growth rate. Thus, for the *IF* decay observed in a given period, around 40% corresponds to secretion, the rest is dilution due to cell growth. The other interesting parameter to analyze is k_{in} that is almost 27 times lower in comparison with a non-secreted protein. The difference should rely on the process of translocation to the endoplasmic reticulum, and maturation of the fluorophore in such compartment. It is known from the literature that the secretion signal here used, the alpha-pre-pro leader peptide, is a posttranslational signal (Rapoport, Matlack, Plath, Misselwitz, & Staeck, 1999; Johnson, Powis, & High, 2013). Meaning that the synthesized peptide is translocated to the ER after full translation. However, such process requires the protein to be unfolded, and therefore non-fluorescent. Consistently with this, we have observed by fluorescence microscopy (Figure 4.9) how the non-secreted mNeonGreen is spread across the cytoplasm, as opposed to the secreted one, which is located exclusively in what should be the ER.


non-secreted mNeonGreen

Secreted mNeonGreen

Figure 4.9. ***Fluorescence microscopy images of yeast cells producing mNeonGreen secreted and not secreted.*** To the left cells expressing non-secreted mNeonGreen, to the right cells expressing the secreted protein. In both cases the expression was induced by constant light.

Table 4.2. ***Parameters values obtained by joint-fitting strategy***

Protein of interest	**Fit**	k_{tr}^{*}	$60 \cdot \frac{\log(2)}{k_{tr\_out}}$**(min)**	$k_{in}\left(\frac{IF_{units}}{h \cdot Tr_{units}}\right)$	$k_{out}(h^{-1})$	$60 \cdot \frac{\log(2)}{k_{out}}$**(min)**
Non-secreted mNeonGreen	1	1.78	23	1.42	0.43	97
	2	2.08^{F}	20^{F}	1.33	0.4^{F}	104
Secreted mNeonGreen	1	2.08^{F}	20^{F}	0.05	0.65	N/A

***** k_{tr} has different units depending if refers to $k_{tr_{in}}\left(\frac{Tr_{units}}{h}\right)$, or to $k_{tr_{out}}(h^{-1})$

^{F} Refers to fixed parameter in a given fitting process

To conclude on the joint-fitting strategy, the overall quality of the fitting has been rather good. However, those aforementioned mentioned mismatches with the lowest and highest levels of induction, raise the question of whether there could be bottlenecks even when secreting heterologous proteins with low secretory complexity. It motivates the idea of trying to uncover subtle features of the system in a more data-driven manner. Thus, I decided to pursue the experiment-specific fitting strategy, since it does not impose a relation between rates and experiments.

For the experiment-specific fitting process, first using the non-secreted mNeonGreen experiments individually I infer k_{out} to check that the value is consistent with the actual growth rate (Figure 4.10). Indeed, the k_{out} values inferred are very close to the mean of the measured growth rate, the mean of the inferred k_{out} is 0.398 ± 0.020 h^{-1}. In the second fit, fixing k_{out} to 0.4 h^{-1}, the obtained values for k_{tr} are also close to those corresponding to a half-life of 20 minutes (Figure 4.10). Those have a mean of 2.125 ± 0.298, that corresponds to a 19.6 minutes half-life. As explained above, this value is also fixed for future fittings, assuming that the optogenetic system is not saturated and the global transcription not regulated by stress.

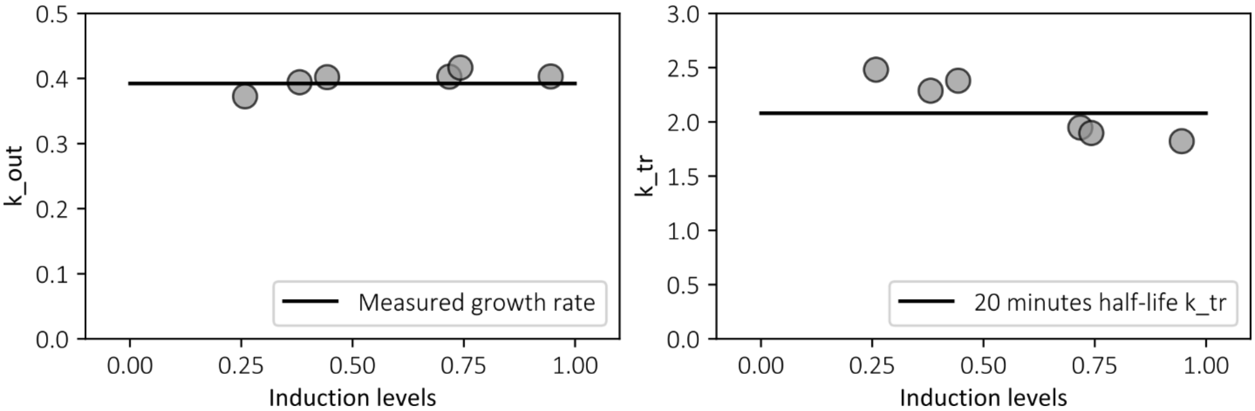


Figure 4.10. ***Parameters values obtained by experiment-specific fitting strategy***. To the left the k_{out} values for different levels of expression. The black line represents the mean measured growth rate, 0.39 h^{-1}. To the right, the k_{tr} values for different levels of expression obtained from a fitting with k_{out} fixed at 0.4 h^{-1}. The black line represents the k_{tr_out} value corresponding to 20 minutes half-life. The fitting is done for the time points between -2 and 24 hours. Note that experiments wherein no light exposure is applied are not used for fitting.

With k_{out} and k_{tr} fixed, only k_{in} is inferred for model fitting to non-secreted mNeonGreen data (Figure 4.11). Unsurprisingly, the simulation results have notably improved for all experiments with respect to the joint-fitting strategy, in both dynamics and steady states, using the independent-fitting. It is clear that by construction the experiment-specific fitting will yield fits at least as good as the joint-fitting approach.

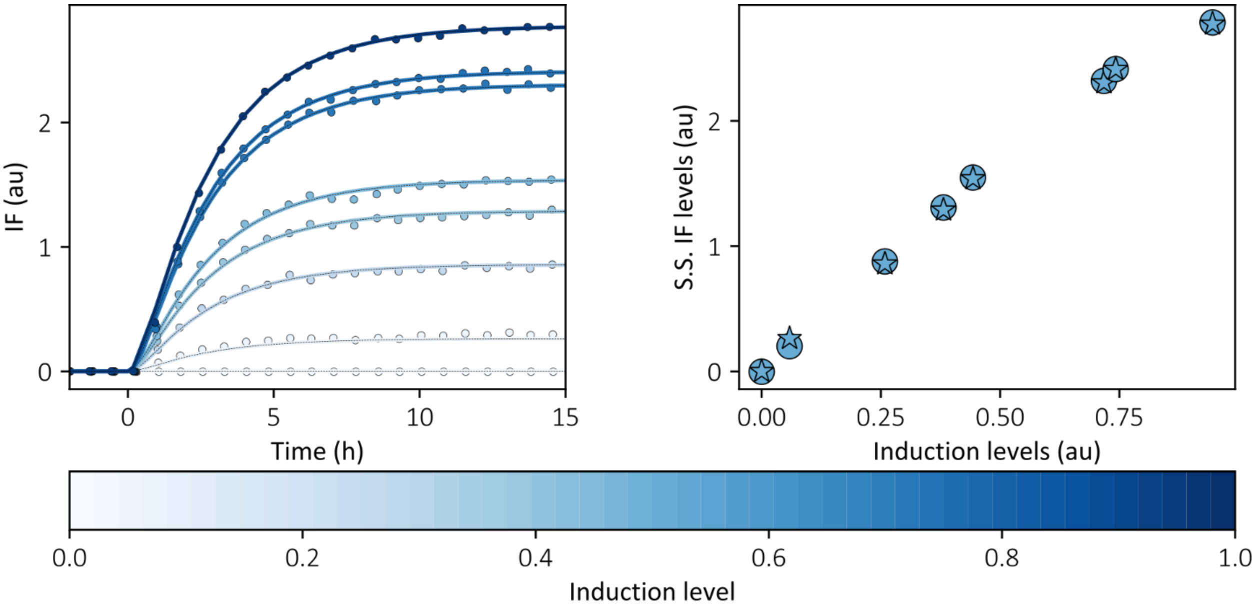


Figure 4.11. ***Comparison between data and simulation of non-secreted mNeonGreen expression.*** Parameters values have been obtained by experiment-specific fitting strategy. To the left the dynamics of internal levels of fluorescence represented by dots, and the simulation results for *IF*. To the right the steady state levels for data, represented as circles, and simulation as stars. The fitting is done for the time points between -2 and 24 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes, k_{out} is fixed to 0.4.

In Figure 4.12 the values of k_{in} for different experiments are shown. The parameter values slightly change at extreme induction levels, suggesting that the capacity of the cell to produce protein may change among the different demands levels applied. Being higher at the lowest levels of expression, and reduced for highest levels. Although difficult to justify, that could be due to an excess, or limitation of synthesis resources, for low and high levels of demand, respectively.

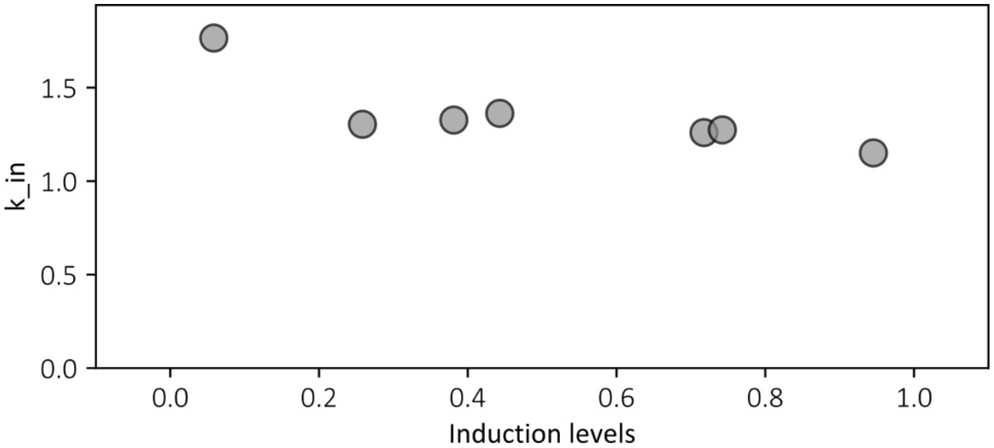


Figure 4.12. ***Parameters values obtained by experiment-independent fitting strategy***. k_{in} values for different levels of expression. The fitting is done for the time points between -2 and 24 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes, k_{out} is fixed to 0.4. Note that experiments wherein no light exposure is applied are not used for fitting.

For secreted mNeonGreen experiments, the same strategy of fitting the model independently for each experiment is followed. The simulation is precisely matching the data in all cases, for dynamic behavior as for steady state levels (Figure 4.13).

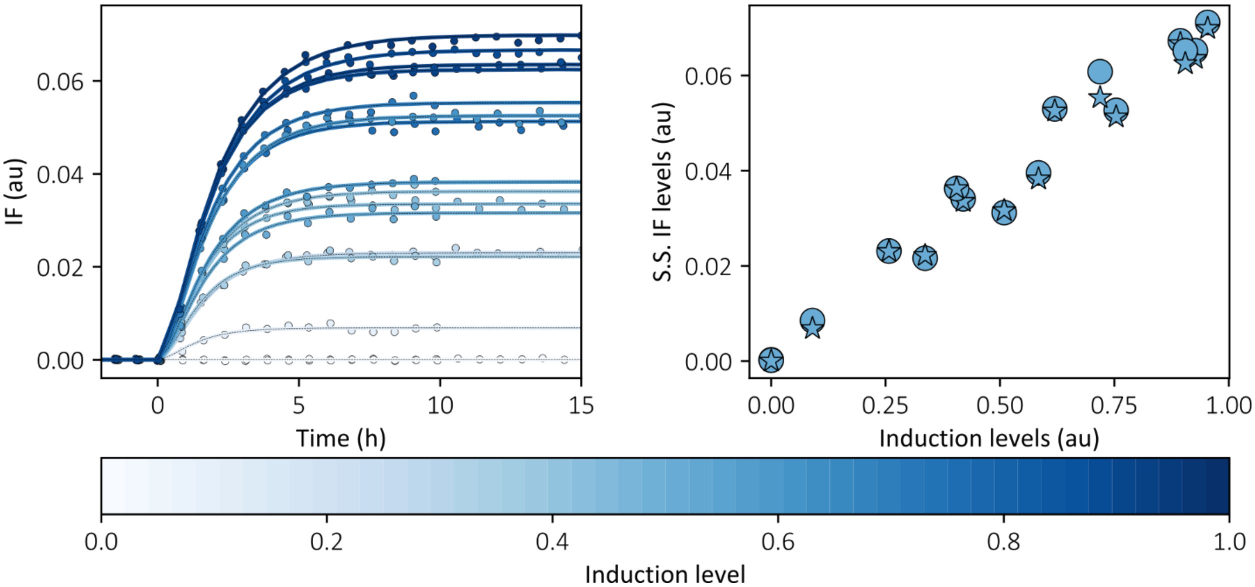


Figure 4.13. ***Comparison between data and simulation of secreted mNeonGreen expression.*** Parameters values have been obtained by experiment-specific fitting strategy. To the left the dynamics of internal levels of fluorescence represented by dots, and the simulation results for *IF*. To the right the steady state levels for data, represented as circles, and simulation as stars. The fitting is done for the time points between -2 and 24 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

In Figure 4.14, the parameters values inferred suggest an alternative interpretation of the data contrasting to that argued previously to motivate the joint-fitting approach. The decreasing trend of k_{in} indicates a loss of capacity to produce POI as induction levels increase. In the same way, k_{out} decreases too, suggesting a loss in capacity to secrete as the internal levels of fluorescence increases. Indeed, both parameters show very similar decrease in their slopes with respect to the induction levels. It makes necessary to revise the arguments stated previously in reference to conclusions obtained from Figure 4.3 and Figure 4.4. There, I said first that the secretory system could maintain full capacity because the relation between internal fluorescence and induction levels at steady state is linear. Secondly, I also stated that the relation between secretion levels and internal fluorescence could be also linear, but not evident.

Regarding the first hypothesis, and despite the real linear relation between internal fluorescence and induction levels. Figure 4.14 suggest that the linear relation aforementioned is due to the decrease in both, the production rate by k_{in}, and removal rate by the changes in k_{out}. As both parameters decrease with similar slope, one compensates the other, given that the steady state level is determined by the ratio $\frac{k_{in}}{k_{out}}$. Therefore, the internal levels observed change linearly with the induction, however, the cellular capacity is saturated. These results suggest that even for mNeonGreen, a heterologous protein of low secretory complexity, several bottlenecks are encountered. Subtly, these bottlenecks have an impact at both parts of the pathway, the production and the removal, and that makes them non-detectable by just looking at the internal levels of POI. Additionally, note that the hypothetical bottlenecks, produce a decay of around 60% of the parameter value for each of both processes.

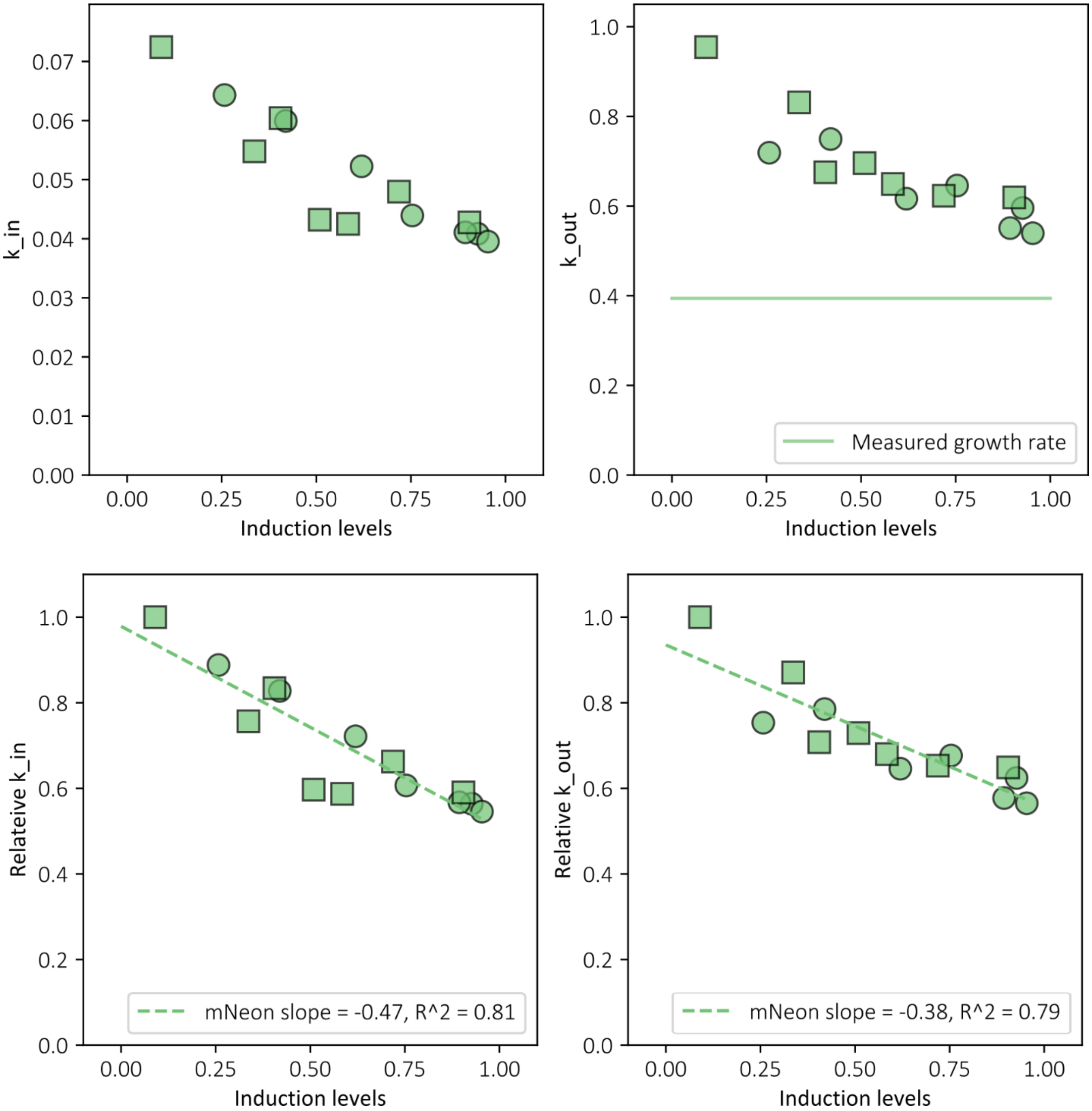


Figure 4.14. ***Parameters values obtained by experiment-specific fitting strategy and relation between them***. At top left the k_{in} values for different levels of expression. At top right, the k_{out} values for different levels of expression. The green line represents the mean measured growth rate, 0.39 h^{-1}. At bottom left, the relative k_{in} values with respect to its maximum value for different levels of expression. The dashed line indicates the linear regression to infer the slope. At bottom right, the relative k_{out} values with respect to its maximum value for different levels of expression. The dashed line indicates the linear regression to infer the slope.

Regarding the relation between internal POI and secretion levels, as seen in Figure 4.15 there is a decreasing k_{out} with increasing levels of internal fluorescence. Therefore, assuming that all the POI removal is due to constant dilution by growth and secretion, one would expect a non-linear relation between internal levels of fluorescence and secretion levels. Indeed, when looking to data more in detail than in Figure 4.4, a decrease in slope with increasing internal fluorescence at steady state levels is observed.

To estimate the difference between the two behaviors, full capacity or saturation of the secretory machinery, I fit the data to a linear and a polynomial relation. In fact, the difference corresponding to a hypothetical maximal secretion level in case of linear relation, and the predicted by polynomial fit of degree 2, that fits the data better, is close to 60%. Which is the same value as the decrease in relative k_{out} observed in Figure 4.15 top right.

All these arguments suggest that the secretory system loses efficiency when secreting mNeonGreen, a relatively easy to fold protein, at increasing levels of demand. Such loss of capacity occurs at both, production and secretion processes. Therefore, it is not detectable by only looking to the internal levels of fluorescence at steady state, since a similar decrease in the kinetic constants driving the processes do the steady state levels proportional to the induction. This shows the potential benefit of a detailed analysis of the system's dynamics to characterize the secretory process.

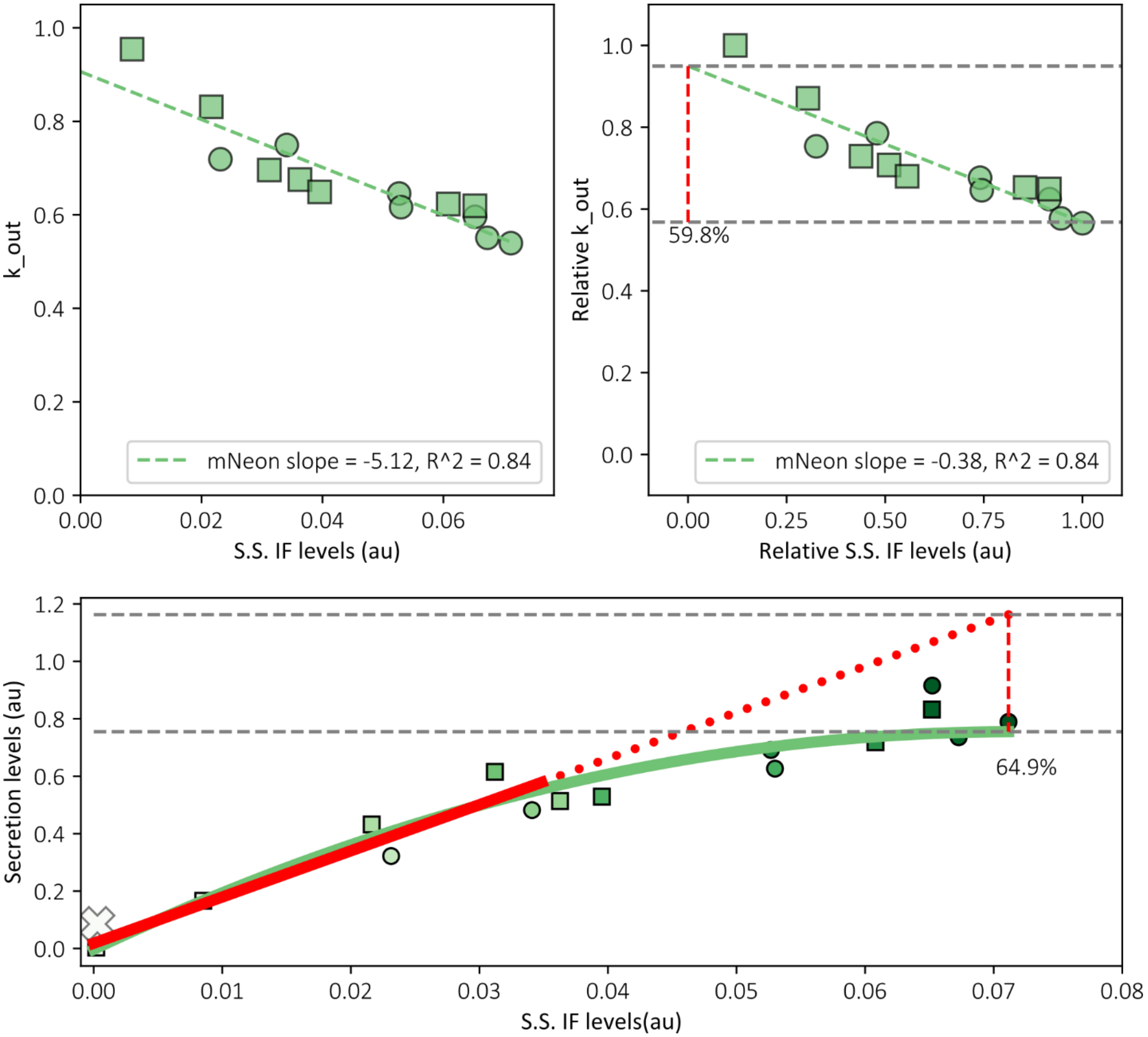


Figure 4.15. ***Analysis of parameters values obtained by experiment-specific fitting strategy and comparison with data***. At top left the k_{out} values for different levels of internal fluorescence (IF) at steady state (S.S.). The dashed line indicates the linear regression to infer the slope. At top right, the relative k_{out} values respect to its maximum for different levels of IF at S.S. The dashed line indicates the linear regression to infer the slope. The gray lines indicate the limits of the linear regression used to compute the relative decrease in k_{out}, 59.8%, red dashed line. At the bottom, the relation of secretion levels with the IF at steady state. The upper gray line indicates the limit of the linear regression used to compute the expected secretion levels in case of linear relation. The solid red line indicates the data points to perform the linear fit, the dotted line shows the rest of the inferred linear relation. The lower gray line indicates the limit of the polynomial fit of degree 2, solid green line, and the predicted actual maximal secretion level. The difference between hypothetically linear and actual predicted is indicated by the dashed red line, and it is 64.9% respect to the linear. The point marked with a white "X" indicates a data considered as outlier, because it does not start at zero, and it is not accounted for the linear regression. The intensity of the green is proportional to the induction levels. Square markers represent data collected in a different set of experiments.

To conclude, in this section is described the experimental procedure used to produce data in order to characterize the secretory process. Continues cultures of cells secreting a heterologous protein of interest at different production demand levels were carried out in parallel. Variables as growth rate, internal levels of fluorescence and stress, were monitored in real time with high temporal resolution, secretion levels were also measured at steady state.

First data has been produced using two strain expressing a low complexity protein that serves as fluorescent tag for characterization of the rest of POIs in this study, mNeonGreen (Figure 4.1). The protein is secreted, or not secreted, in each strain respectively. The data that has been produced is used to validate the relevance of a high-level model of the secretory process. Using an optimization software to fit model simulations with experimental data, parameters values can be inferred to extract.

In fact, interesting features of the system have been learned from the analysis of a protein with relatively simple secretory complexity. The secretory system decreases its capacity when the levels of production demand increase. Such a bottlenecks occurs both at the level of production as for secretion. That feature cannot be captured just by looking at the growth rate, or the internal levels of POI at steady state. In the next section, the pipeline implemented here, based on data generation and model fitting to interpret parameters values, will be used with a panel of protein with different secretory complexity.

4.2. Leveraging the validated pipeline: Characterizing secretion of different secretory complexity proteins

In the previous section, I presented and validated the pipeline for characterizing the secretory process from a quantitative perspective. In this section, a second group of heterologous proteins with different secretory complexity are studied using the same pipeline (Table 4.1, Group 2). The criteria to classify the secretory complexity of this set of proteins is based on the presence of posttranslational modifications needed to achieve folding. As mentioned in Chapter 2, such folding requirements make that proteins are more prone to accumulate in ER to pass the quality control. Thus, possibly triggering stress, activating the UPR pathway, and consequently undergoing active degradation by the ERAD machinery.

Given the different types and number of posttranslational modification in the second group of heterologous proteins analyzed here, different behaviors are expected for each POI. Therefore, in this section the data generated from each POI production is presented individually, and then the model fitting is applied to infer parameters values. It is important to note that fluorescent reporters fused to the C-terminal region of the POI are used in this study (Figure 4.1). Here we assume that folding and maturation of the fluorescent reporter is fast and independent of the POI fused to it. Therefore, the internal levels of fluorescence are considered proportional to the total amount of POI inside the cell, either unfolded or already folded.

4.2.1. endo-1,4-β-xylanase C, XylC, as protein of interest

We first consider the endo-1,4-β-xylanase C, from now on called XylC (Figure 4.16). This protein does not induce growth rate decay. The relation between internal fluorescence at steady state and induction levels suggests saturation at the production level. A linear relation for secretion levels and internal fluorescence at steady state is observed. This could be explained by the requirement of a disulfide bond to proper folding of the protein. That might produce a bottleneck in the production, increasing the ratio unfolded/folded protein. Thus, reducing the subtract for secretion, only allowed for folded proteins, and preventing the saturation of the secretion process. If comparing the internal levels of fluorescence between mNeonGreen and XylC, the second shows 70% the levels of the first. This can be explained by the action of ERAD, given that UPR is triggered, decreasing the quantity of internal unfolded protein. For the secretion levels, XylC reaches 30% of the mNeonGreen levels at steady state, this could be caused by a reduced fraction of folded proteins able to pass the quality control and be secreted.

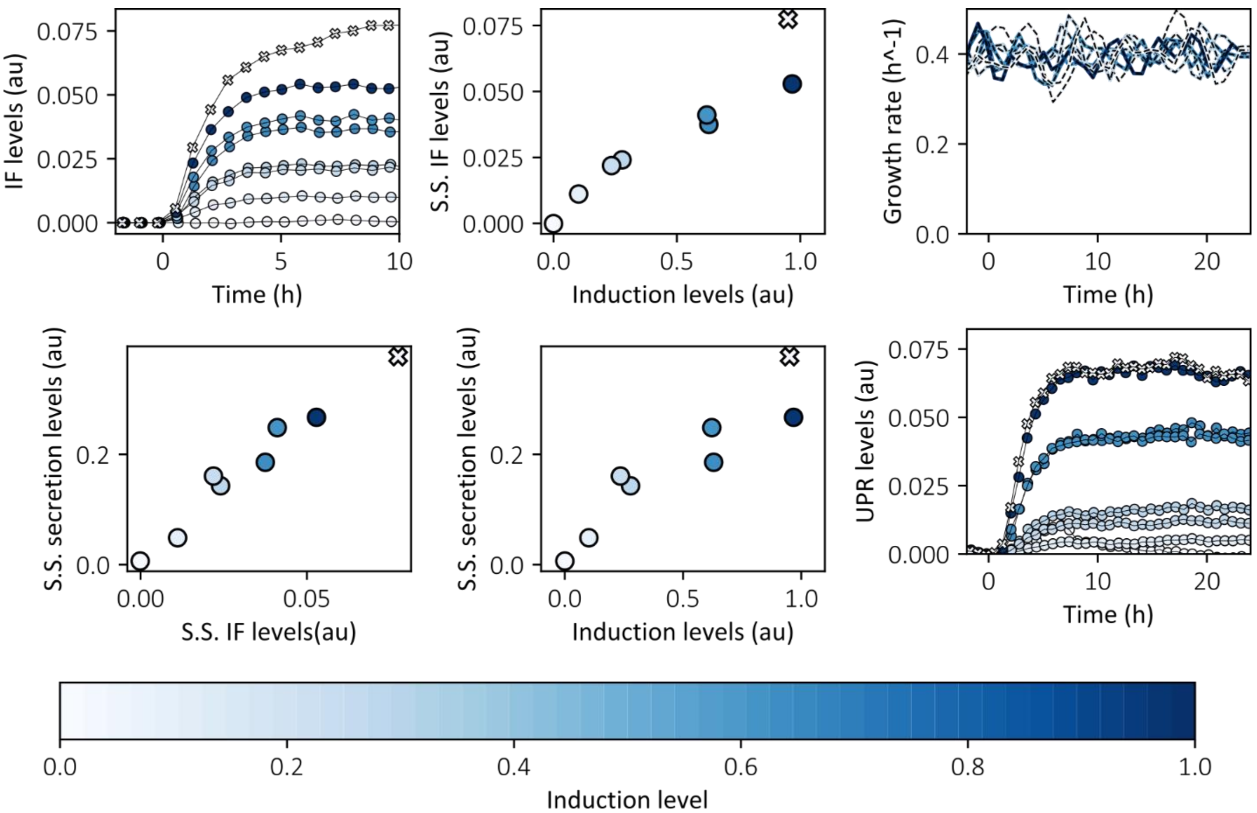


Figure 4.16. ***Data summary for XylC secretion.*** Top left plot shows the internal fluorescence (IF) levels over the time respect to the light induction starting point, from -2 to 10 hours. Top center shows the relation between IF at steady state (S.S.) and induction levels. Top right shows the growth rate over the length of the whole experiment. Bottom left shows the relation between IF levels at S.S. and S.S. secretion levels. Bottom center shows the relation between induction levels and the secretion levels at S.S. Bottom right shows the UPR signal over the length of the whole experiment. The data points marked with a white "X" indicates data considered as outlier.

The model fitting is performed for each individual experiment as shown in the previous section of this chapter. As for all the rest of proteins, k_{tr} is fixed to the value corresponding to a 20 minutes half-life for mRNA. No fit is done for induction levels corresponding to experiment with duty cycle equal zero, and neither for the outlier shown in Figure 4.16. In agreement to what has been deduced from the experimental data, we observe decrease in k_{in} with increasing demand levels, and no change in k_{out}, corroborating a saturation of the production part of the pathway but not in the secretion (Figure 4.17). Moreover, it is noteworthy to point that k_{in} evolves in the same way and values as for secreted mNeonGreen with respect to the induction levels. In the case of k_{out}, although it shows different relation with the induction levels, it presents same values as maximal k_{out} in secreted mNeonGreen. Suggesting that no bottleneck is present in the secretion part of the pathway.

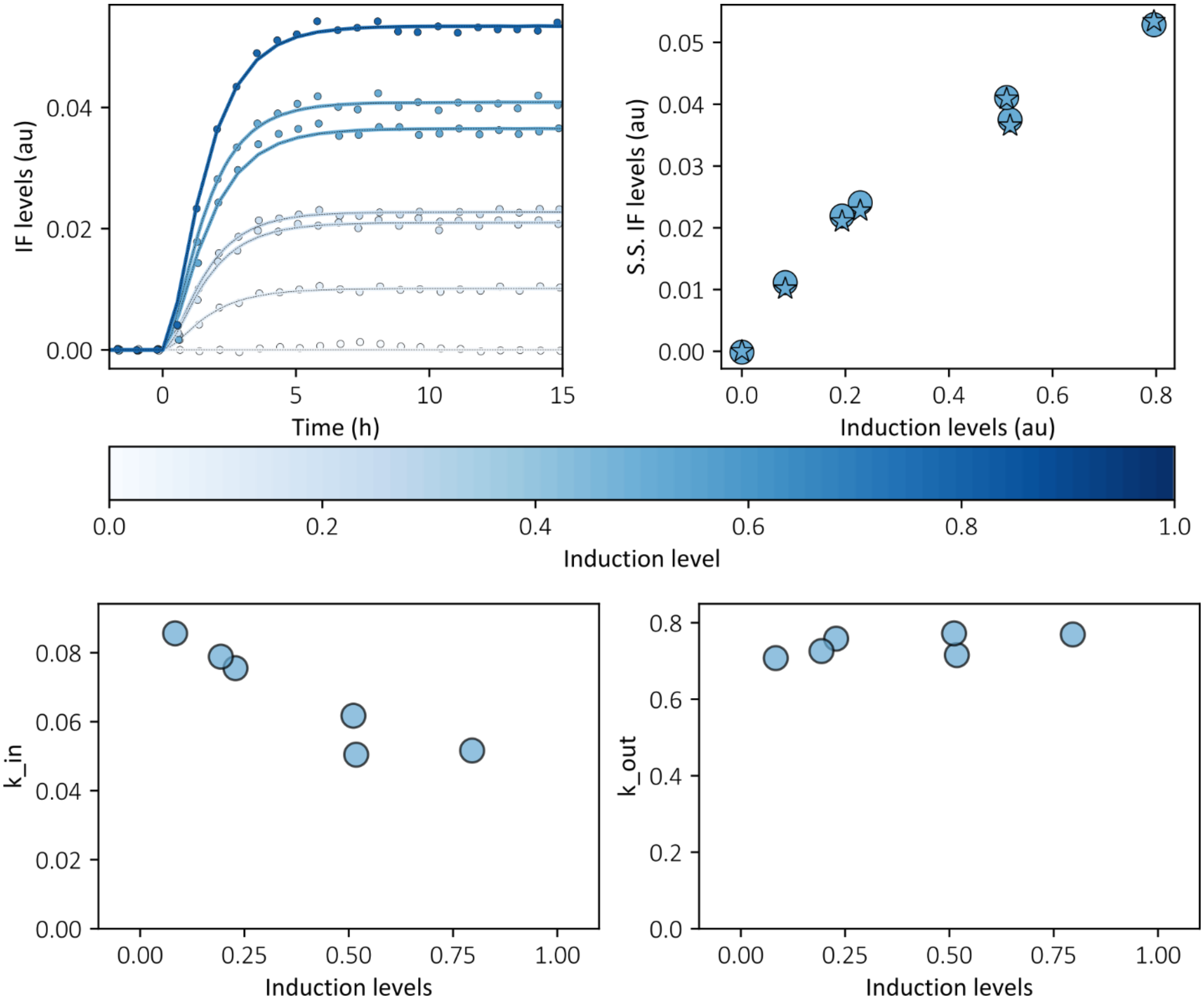


Figure 4.17 ***Comparison between data and simulation for XylC secretion.*** Parameters values have been obtained by experiment-specific fitting strategy. At top left the dynamics of internal levels of fluorescence represented by dots, and the simulation results for *IF*. At top right the steady state levels for data, represented as circles, and simulation as stars. At bottom left the k_{in} values for different levels of expression. At bottom right, the k_{out} values for different levels of expression. The fitting is done for the time points between -2 and 15 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

4.2.2. Single chain variable fragment 4M5.3, scFv, as protein of interest

The next protein of interest to analyze is the single chain variable fragment 4M5.3, from now on called scFv (Figure 4.18). This protein shows several features of interest that complicates its analysis. First, the internal fluorescence dynamics are not monotonic in time from high levels of demand. The steady state levels of internal fluorescence saturate for high demands, and present even higher values than for secreted mNeonGreen. Second, it shows a transient growth decay of approximately 50% that eventually recovers to the initial value of 0.4 h^{-1}. Third, the steady state levels of UPR are around 3-4 times higher than for XylC. The fourth feature of interest is that there is non-monotonic relation between demand and secretion levels. Interestingly, it suggests that may be an optimal level of production demand to maximize the secreted protein obtained.

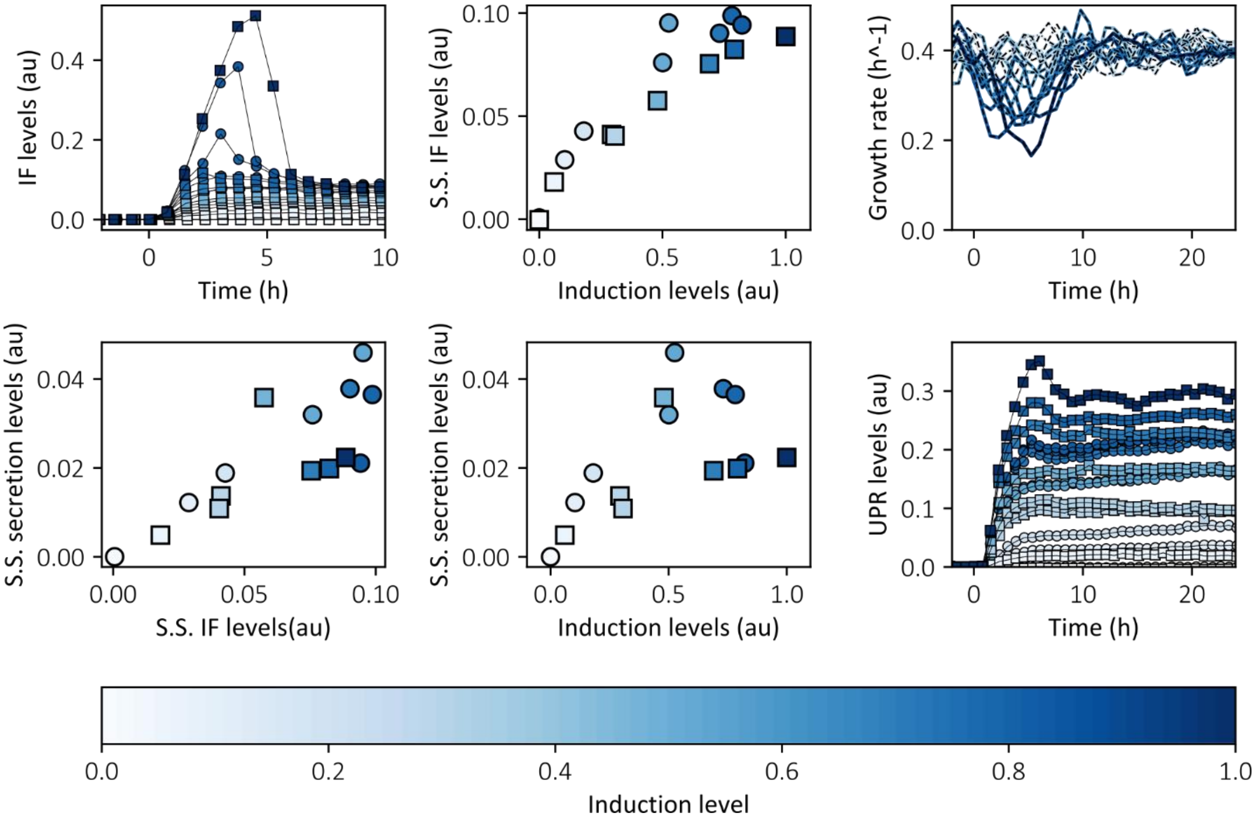


Figure 4.18. ***Data summary for scFv secretion.*** Top left plot shows the internal fluorescence (IF) levels over the time respect to the light induction starting point, from -2 to 10 hours. Top center shows the relation between IF at steady state (S.S.) and induction levels. Top right shows the growth rate over the length of the whole experiment. Bottom left shows the relation between IF levels at S.S. and S.S. secretion levels. Bottom center shows the relation between induction levels and the secretion levels at S.S. Bottom right shows the UPR signal over the length of the whole experiment. Square markers represent data collected in a different set of experiments.

Obviously, the proposed mathematical model is unable to fit these dynamics, since they are non-monotonic. To better understand which processes are at play, I use the single cell data obtained by the cytometry. Figure 4.19 shows that for scFv, upon high levels of induction a transient bimodal distribution emerges. The overall population is composed by a high internal fluorescence phenotype subpopulation, and other subpopulation that quickly reaches the steady state internal fluorescence at lower levels. Since the high internal fluorescence subpopulation is characterized by presenting an exacerbated accumulation of POI, from now on I will refer to such phenotype as the accumulator phenotype.

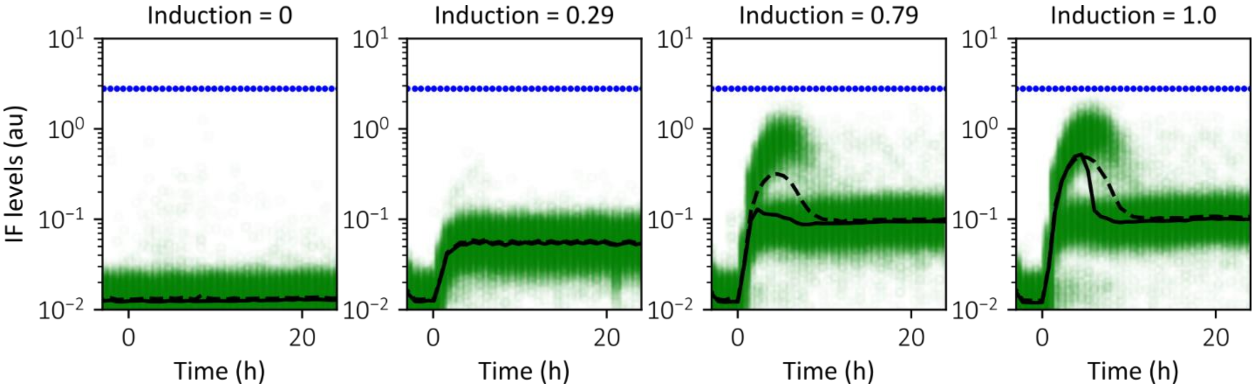


Figure 4.19. ***Distributions of cells expressing scFv at different expression levels over the length of the experiment.*** In each of the plots the temporal evolution of internal fluorescence (IF) is represented in log scale over time. Note that in this case the autofluorescence is not subtracted. Every single cell captured by the cytometry is shown as a green dot. The mean is represented as a dashed line, the median as solid black line. The dotted blue line represents the median steady state IF levels for the non-secreted mNeonGreen from Figure 4.3.

From the experimental data, one can suggest some of the causes that may cause emergence and disappearance of the accumulator phenotype. The fact that this subpopulation reaches internal fluorescence levels relatively close to the non-secreted mNeonGreen at steady state suggests that secretion rate might be low for cells presenting the accumulator phenotype (Figure 4.19). Accordingly, Figure 4.20 shows that the secretion levels do not increase, or even decrease for experiments in which such phenotype arises. The presence of bimodal populations respect to the internal levels of protein in yeast has been already observed and reported by Love et al., 2012, as commented in the Chapter 2. The authors reproduced computationally those population distributions by adding variability to the secretion rate, and the following reduction of internal POI levels by increasing degradation of internal protein.

Along these lines, another interesting feature is that the accumulator phenotype disappears after ten hours of induction, indicating either adaptation or dilution due to the continuous culture. As shown in Figure 4.20 there is a clear correlation between the fraction of overall cells presenting the accumulator phenotype and the fraction of overall population growth decay. Such relation suggests that the accumulators do not growth. Another argument to support that statement, is that those two variables are also correlated in time. Moreover, regarding the growth arrest argument, there is also literature demonstrating that upon ER-associated stress, the ER stress surveillance (ERSU) pathway, avoids ER inheritance and cytokinesis, therefore cell division (Babour, Bicknell, Tourtellotte, & Niwa, 2010; Jonas, Royle, Aw, Stan, & Polizzi, 2018). This also supports the hypothesis of adaptation as a possible cause of the disappearance of the subpopulation.

According to the arguments exposed above, the accumulator phenotype implies accumulation of internal protein, growth arrest, and reduced secretion. As opposed to the accumulator phenotype, the subpopulation showing lower levels of internal fluorescence is likely to secrete, thereby from now on I will refer to that as secretor phenotype.

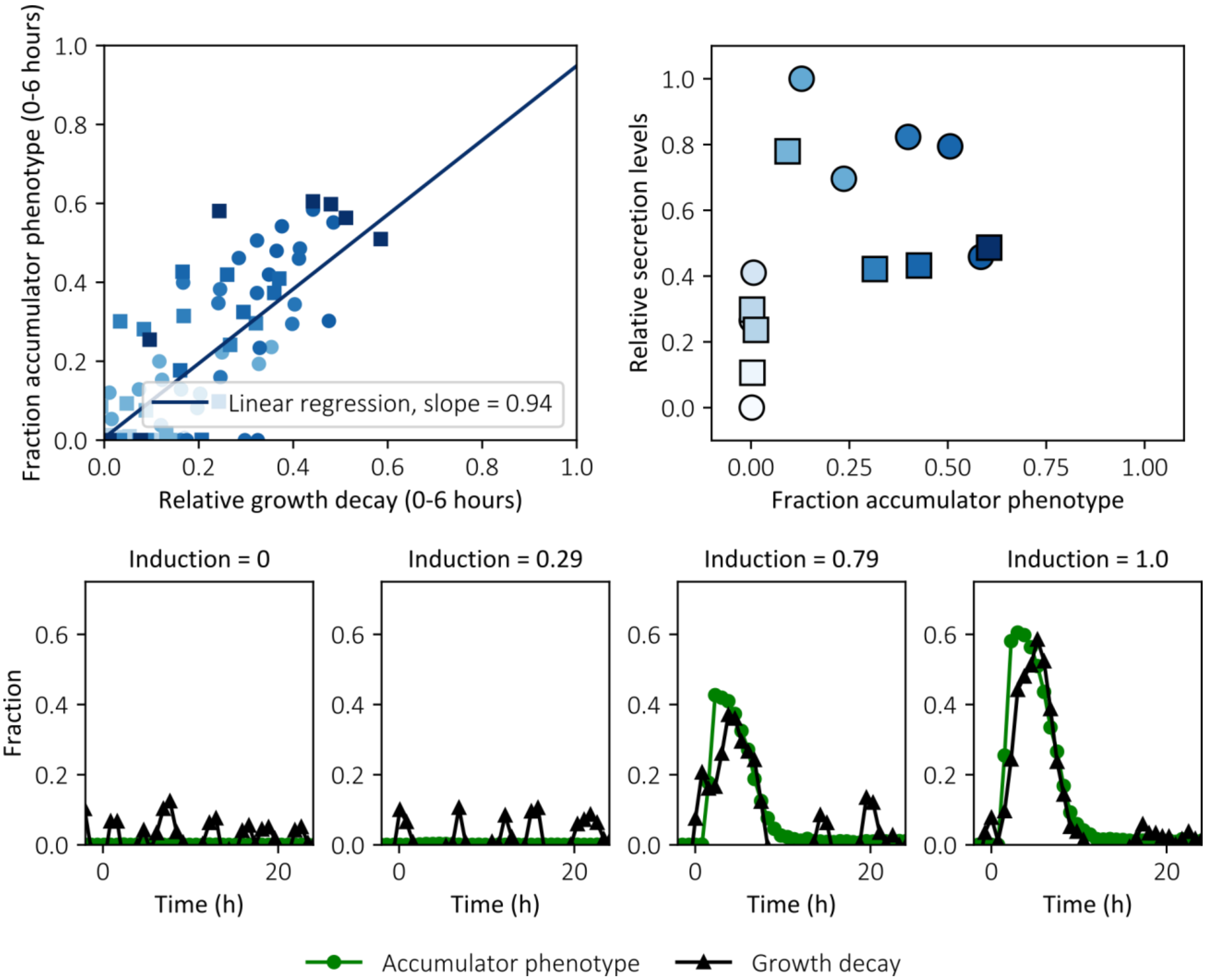


Figure 4.20. ***Accumulator phenotype features.*** The top left plot shows the correlation between overall population growth decay and fraction of cells presenting the accumulator phenotype over all the cells measured in the cytometer (5000 events per data point), during the first five hours after induction starts. At top right, the plot shows the relation between fraction of accumulators and the relative secretion levels for each experimental set at steady state (S.S.). Square markers represent data collected in a different set of experiments. The lower row shows the dynamics of the relative growth decay, in black, and the fraction of accumulator phenotype, in green.

Taking into account what is explained above, I continue the characterization of the secretory process for scFv, just taking into account the population showing the secretor phenotype. To differentiate either of the two populations I use a Gaussian mixture model (Figure 4.21). It is a probabilistic model that assumes that data is generated from normal distributions. Based on an optimization strategy it finds normally distributed subpopulation within an overall population. Specifically, I use the *Python* package *sklearn.mixture.GaussianMixture* that enables for fitting of mixture-of-Gaussian models. The algorithm is only applied to the data points corresponding to induction levels above 0.4, which is the threshold level to observe the phenotype, and the period between 0 to 15 hours after induction is triggered.

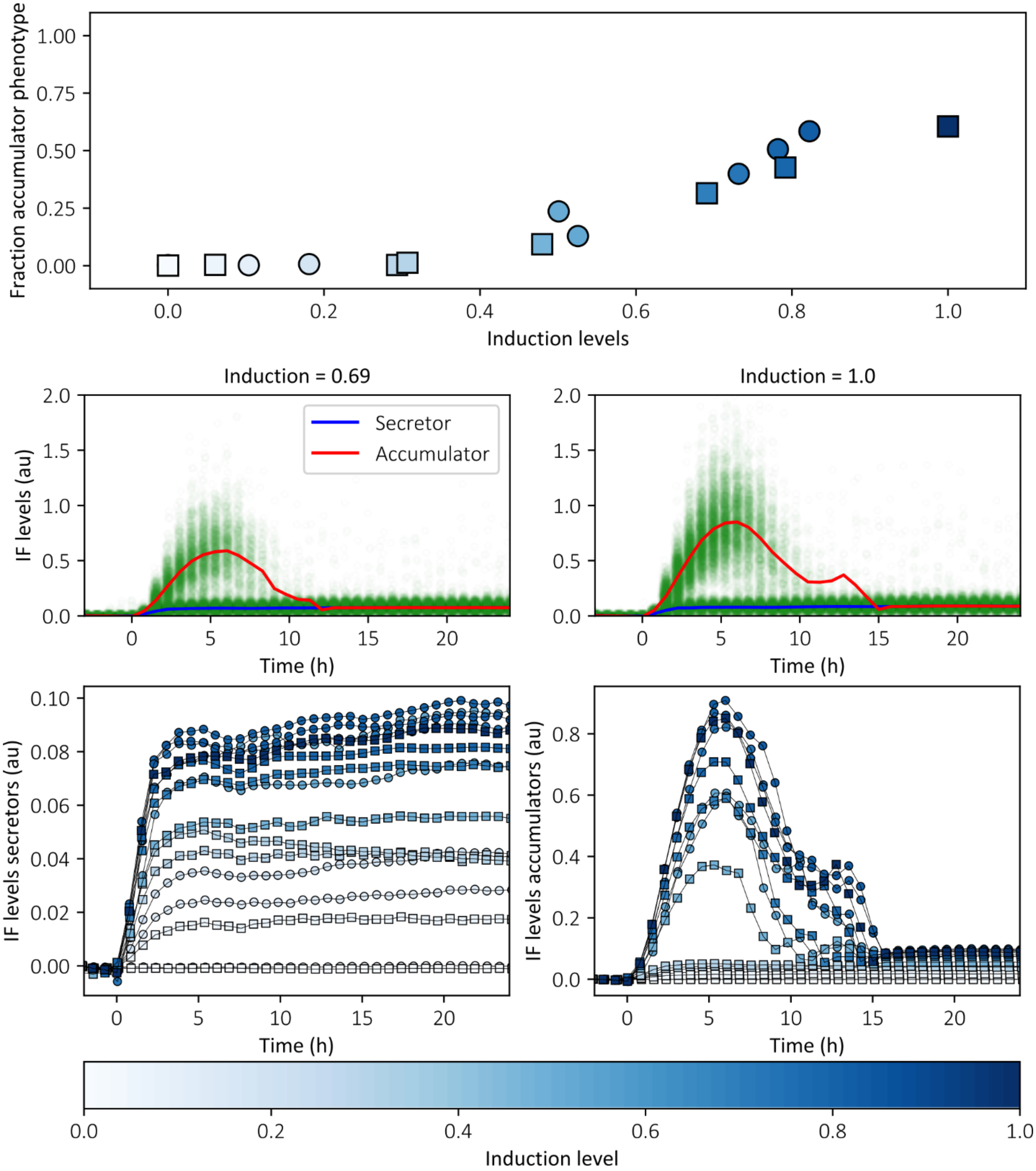


Figure 4.21. ***Differentiation of secretor and accumulator phenotype from the overall population.*** On top most, the plot shows the relation between induction levels and accumulator phenotype fraction, at the time point with maximum proportion of accumulators. The middle plots show two examples of Gaussian mixture model used to discriminate both phenotypes. In each of them the temporal evolution of internal fluorescence (IF) in log scale over time. The median for accumulators is represented as

a red line, the median for secretors as a blue line. The bottom plot shows the IF median of the two phenotypes after populations filtering, to the left the secretors, to the right the accumulators. Square markers represent data collected in a different set of experiments.

It is important to note that from now on, the variable *IF* accounts only for the median of internal fluorescence of the secretor phenotype. Once the accumulator phenotype is filtered out from the overall population, the model fitting can be performed for the secretor population. As before, only k_{in} and k_{out} are searched by experiment-specific fitting. Now, the obtained fitting of the model to the data is overall very good (Figure 4.22).

Regarding the parameters value, as opposed to secreted mNeonGreen, here k_{out} evolves by increasing with induction levels. However, as seen in Figure 4.19 it is not the case for secretion, suggesting that it is likely degradation what contributes to the increase of k_{out}. In the case of k_{in}, it decreases with increasing induction levels, this has been seen for all the previous cases. Such pattern can be interpreted as saturation of the production machinery, probably due to excessive trafficking of POI at the level of translocation or folding. Notably, in this case the value of k_{in} is twice the value than for secreted mNeonGreen. It could be due to cytoplasmic fluorescence of scFv queuing to be translocated into the ER. Since we use a posttranslational translocation secretion tag, overwhelming of the translocation machinery could led to cytoplasmic folding of the fluorescent reporter (Rapoport, Matlack, Plath, Misselwitz, & Staeck, 1999; Johnson, Powis, & High, 2013). That should be checked by microscopy as done for mNeonGreen in Figure 4.9, but unfortunately it has not been done yet.

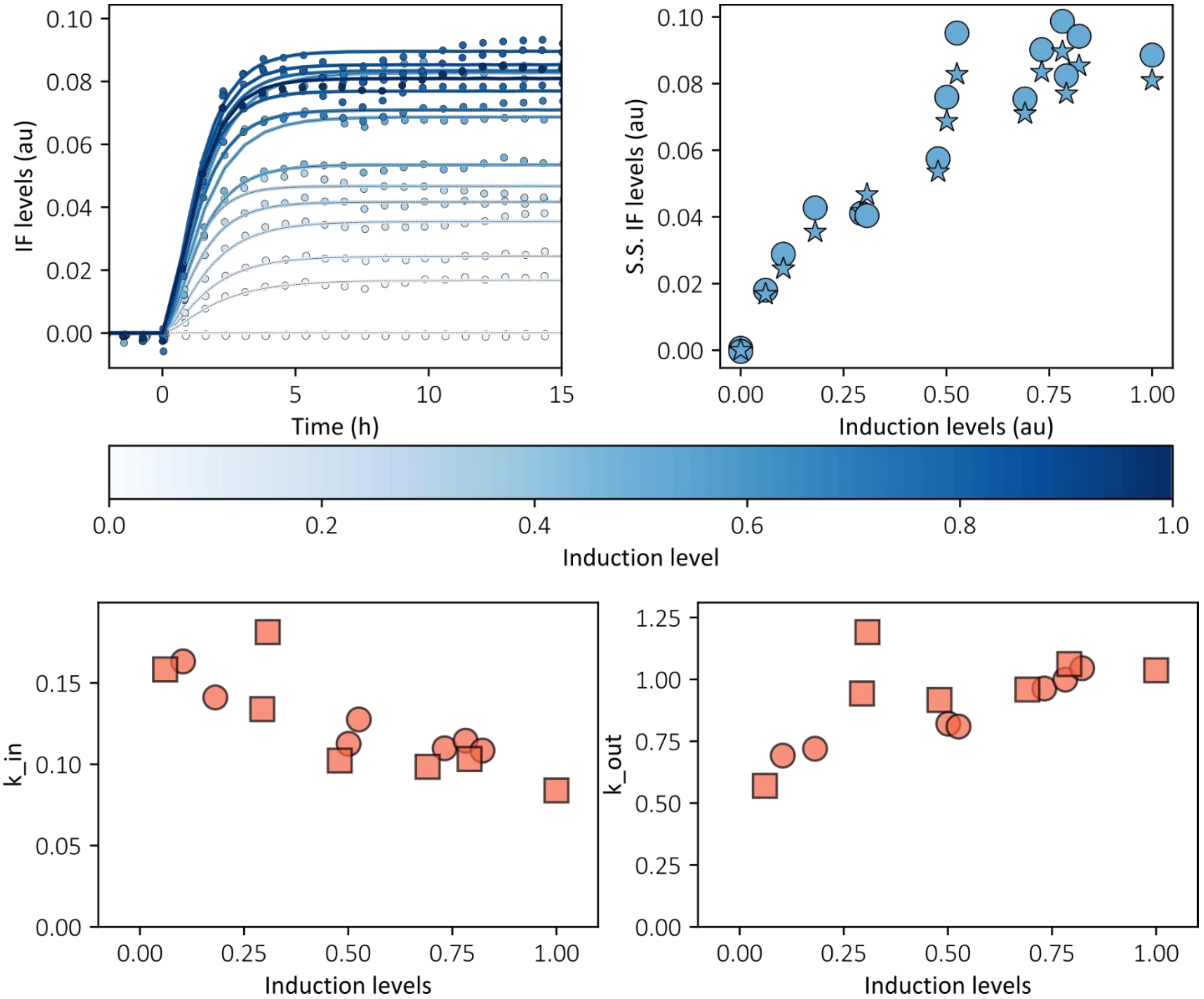


Figure 4.22. ***Comparison between data and simulation for scFv secretion.*** Parameters values have been obtained by experiment-specific fitting strategy. At top left the dynamics of internal levels of fluorescence represented by dots, and the simulation results for *IF*. At top right the steady state levels for data, represented as circles, and simulation as stars. At bottom left the k_{in} values for different levels of expression. At bottom right, the k_{out} values for different levels of expression. The fitting is done for the time points between -2 and 15 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

4.2.3. ∝-amylase, amylase, as protein of interest

The last POI of the second group is the α-amylase, from now on called just amylase. It is assumed to be the protein with higher secretory complexity, since it harbors the highest number of disulfide bonds, four in total, and one N-glycosylation. As for scFv, this protein shows non-monotonic dynamics for the internal levels of fluorescence over time (Figure 4.23), a plateau for the internal fluorescence at steady state with respect to the induction levels, and growth decay. However, for this protein the levels of secretion reach a saturation with the induction levels, but are monotonic and 5 times lower than for scFv. The maximal UPR signal at steady state is 1.5 times higher than for scFv, but also the internal levels of fluorescence, which at steady state are 1.25 times higher with respect to scFv.

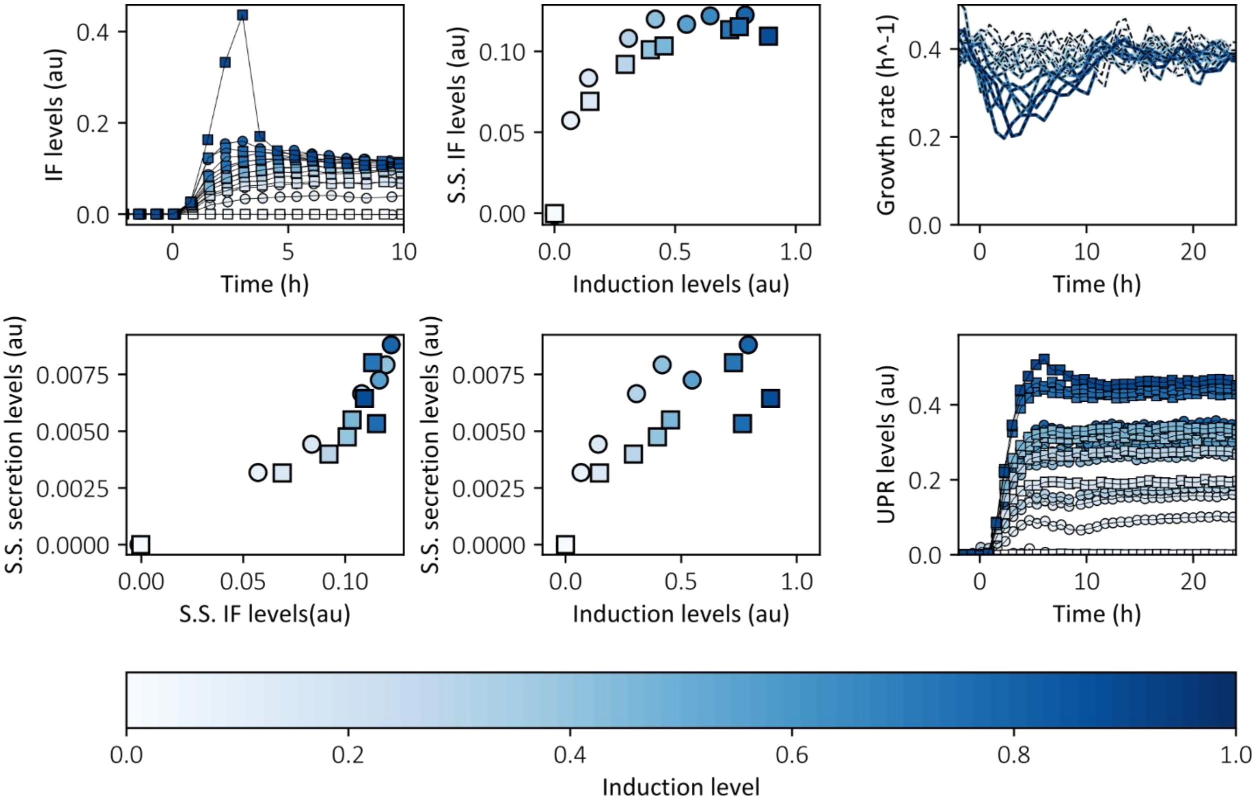


Figure 4.23. ***Data summary for amylase secretion*** Top left plot shows the internal fluorescence (IF) levels over the time respect to the light induction starting point, from -2 to 10 hours. Top center shows the relation between IF at steady state (S.S.) and induction levels. Top right shows the growth rate over the length of the whole experiment. Bottom left shows the relation between IF levels at S.S. and S.S. secretion levels. Bottom center shows the relation between induction levels and the secretion levels at S.S. Bottom right shows the UPR signal over the length of the whole experiment. Square markers represent data collected in a different set of experiments.

Given the non-monotonic dynamics of internal fluorescence, as seen for scFv, the single cell data is analyzed for detecting the rise of the accumulator phenotype (Figure 4.24). In fact, a bimodal population is observed, in which its subpopulations show the same features as those observed when scFv is expressed. The transient accumulator phenotype fraction in the overall population is correlated with a growth decay and decrease in a secretion levels. To distinguish the accumulators population from the secretors, I applied the same pipeline as previously, a Gaussian-mixture model with the same specificities as for the scFv protein. The results of the subpopulations differentiation can be found in the Annex B.

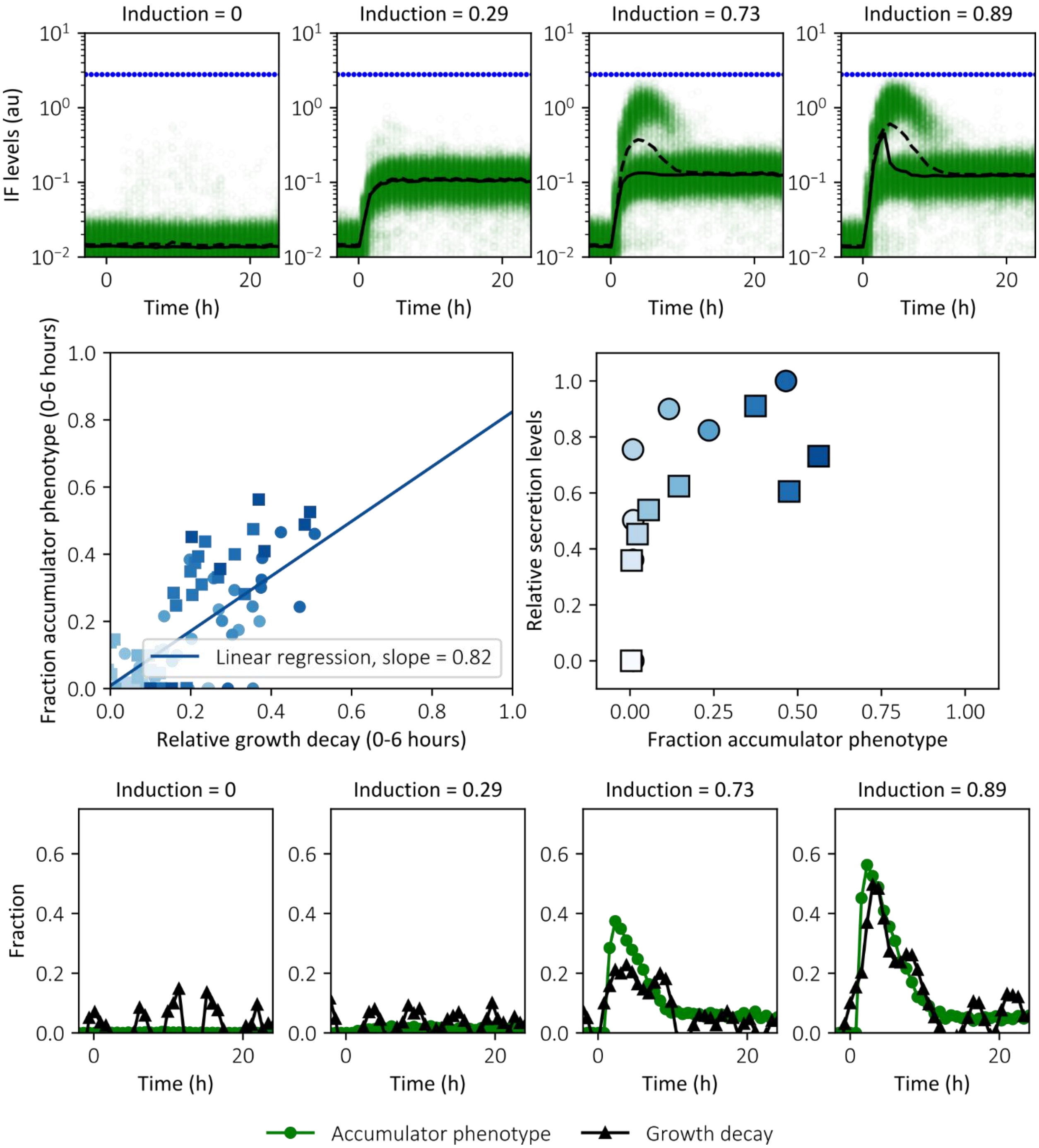


Figure 4.24. ***Distributions of cells expressing α-amylase at different expression levels over the length of the experiment and accumulator phenotype features.*** On the top row, data corresponding to four different experiments from the same data set. In each of them the temporal evolution of internal fluorescence (IF) in log scale over time. The mean is represented as a dashed line, the median as solid black line. The dotted blue line represents the median steady state IF levels for the non-secreted mNeonGreen from Figure 4.3. Middle row to the left plot shows the correlation between overall population growth decay and fraction of cells presenting the accumulator phenotype over all the cells measured in the cytometer (5000 events per data point), during the first six hours after induction starts. Middle row to the right, the plot shows the relation between fraction of accumulators and secretion levels at steady state (S.S.). Square markers represent data collected in a different set of experiments. The third row shows the dynamics of the relative growth decay, in black, and the fraction of accumulator phenotype, in green.

To continue the analysis, the model fitting is performed to the secretor subpopulation data. The simulation has a good match with the data dynamics (Figure 4.25), and also for steady state levels. For this heterologous protein, as for all the previous ones, k_{in} decreases with the production demand. As for scFv, k_{in} starts at very high levels. In the case of k_{out}, a more interesting behavior is observed. It remains close to the same value for most of the experiments. Such behavior is difficult to explain with the information that this model yields.

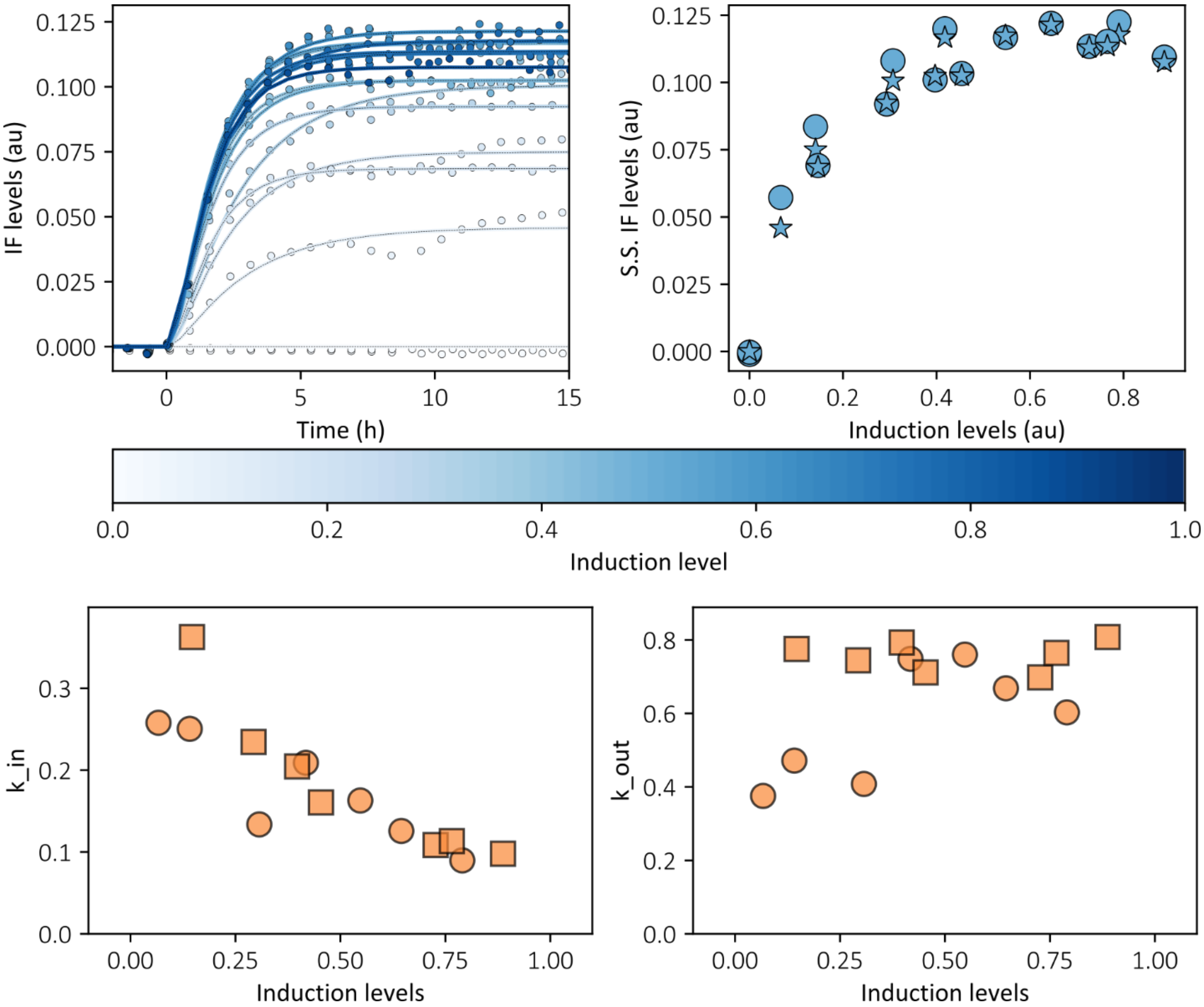


Figure 4.25. ***Comparison between data and simulation for amylase secretion.*** Parameters values have been obtained by experiment-specific fitting strategy. At top left the dynamics of internal levels of fluorescence represented by dots, and the simulation results for *IF*. At top right the steady state levels for data, represented as circles, and simulation as stars. At bottom left the k_{in} values for different levels of expression. At bottom right, the k_{out} values for different levels of expression. The fitting is done for the time points between -2 and 15 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

Indeed, k_{out} accounts for the effect of several protein removal processes including secretion, degradation, and dilution due to growth (constant for secretors). Therefore, the evolution of the parameter as a function of the induction levels could be explained in different ways depending on the relation between secretion and degradation.

4.2.4. Inferring the components of protein removal process

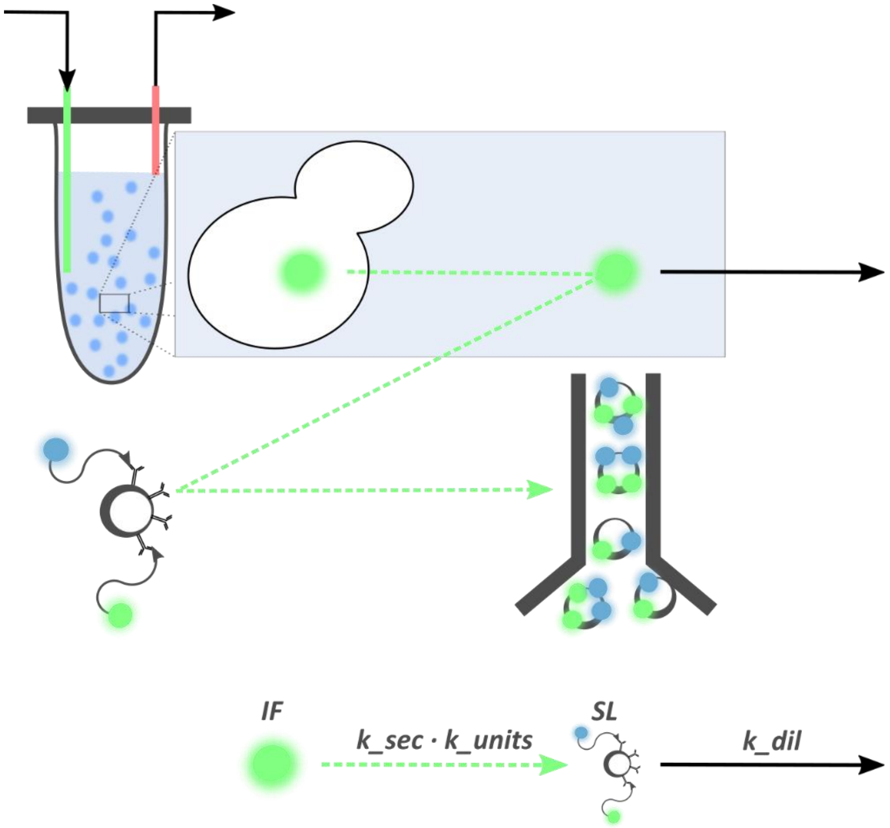
We use the secretion data to obtain more information from the model fitting process by parameters inference. To achieve that, a simple extension of the model is implemented to account for secretion levels. Given that the growth rate is known and constant for the secretor phenotype, and the secretion levels at steady state are also known. We can use such data to decompose k_{out} by considering that k_{growth} accounts dilution due to constant growth, k_{sec} accounts for secretion rate, and $k_{out} = k_{growth} + k_{sec} + k_{deg}$, where k_{deg} is the degradation rate. Thus, by fitting a new version model (Figure 4.26) using the secretion and growth information, it allows to infer the degradation components from k_{out}, even if no degradation measurements are performed. The new model maintains the same mathematical description for the dynamics of *Tr*, and *IF*. However, in this new version $k_{out}$is decomposed in other three parameters:

$$\frac{dIF}{dt} = k_{in} \cdot Tr - \left(k_{growth} + k_{sec} + k_{deg}\right) \cdot IF$$

For describing the dynamics of the secretion obtained by the ratiometric measurements (Chapter 3). We consider the following differential equation:

$$\frac{dSL}{dt} = k_{sec} \cdot k_{units} \cdot IF - k_{dil} \cdot SL$$

Where k_{units} is included in the model to account for the units change from single cell internal fluorescence to secretion levels (*SL*) of the population. And k_{dil} accounts for dilution rate of the reactor, that in a continuous culture using a turbidostat program is equal to the growth of the overall culture.



$$\frac{dSL}{dt} = k_{sec} \cdot k_{units} \cdot IF - k_{dil} \cdot SL$$

Figure 4.26. ***Schematic of the model extension accounting for population secretion levels.*** The population of cells secretes *IF* to produce secreted proteins. Since the later cannot be directly measured, in this study it is monitored by the ratiometric measurements explained in Chapter 3, represented with a green dashed arrow in the scheme. Those measurements quantify the secretion levels at steady state (*SL*). The parameter (k_{units}) is introduced to account for the change in units from secreted protein to secretion levels. The secreted protein is constantly diluted in the bioreactor at a rate k_{dil}, the process is represented by the black arrow. The representation in the bottom depicts the kinetics of the system for *SL*, indicating the kinetic parameters and the differential equation describing its temporal evolution.

Note that here I consider $k_{dil} = k_{growth}$, however, it is not actually the case when the accumulator phenotype arises because it presents growth arrest. In such situation we have:

$$k_{dil} = k_{growth} \cdot (1 - F_{ac})$$

Where F_{ac} is the fraction of the accumulators. Therefore, the dilution rate, or the overall growth rate, decreases proportional to the fraction of accumulators. If it is taken into account that only the secretor phenotype fraction of the population contributes to the secretion, the actual equation to describe secretion levels dynamics looks like this:

$$\frac{dSL}{dt} = k_{sec} \cdot k_{units} \cdot F_{sc} \cdot IF - k_{growth} \cdot (1 - F_{ac}) \cdot SL$$

Where F_{sc} is the fraction of secretors. However, it can be transformed into the first differential equation proposed to describe SL dynamics. Taking the information from experimental data in Figure 4.19 and Figure 4.24, we see that during the last 10 hours of experiment F_{sc} and IF are constant, and F_{ac} is equal to zero. Then, it is possible to consider that the right term of the equation is constant during such period of the experiment, and then the equation can be written as:

$$\frac{dSL}{dt} = constant\ term - k_{growth} \cdot SL \rightarrow$$

$$\rightarrow SL(t) = \frac{constant\ term}{k_{growth}} \cdot \left(1 - e^{-k_{growth} \cdot t}\right)$$

As seen before, in the case of *IF*, the analytical solution of this equation demonstrates that the time to reach steady state is dependent only on k_{growth}, that has a time scale of approximately of 2 hours. If performing a simulation of the equation $1 - e^{-k_{growth} \cdot t}$, 90% of the steady state levels is reached in 6 hours. Moreover, experimental data for secretion levels at induction below 0.4, that is with no accumulators, shows that at 6 hours after induction the secretion levels are already close to the measured values at 24 hours (Figure 4.2 and Figure 4.27). Meaning close to reach steady state levels.

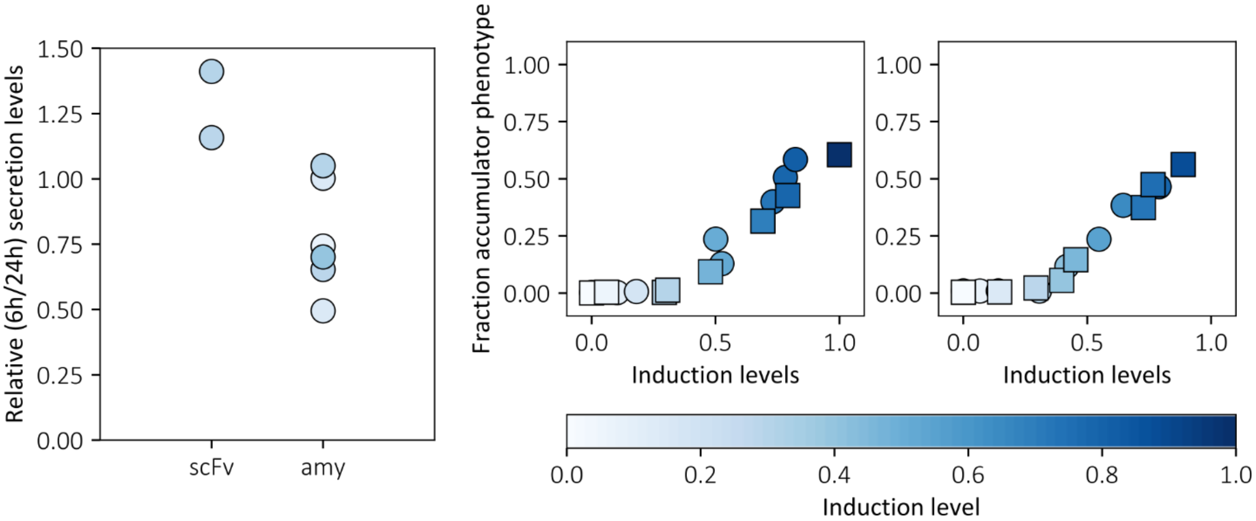


Figure 4.27. ***Comparison of secretion levels measured at 6 and 24 hours after induction***. To the left the ratio of the secretion level measurement after 6 hours induction over the one taken at 24 hours for scFv and amylase. At center and right, the plots show the relation between induction levels and accumulator phenotype fraction, at the time point with maximum proportion of accumulators. The color code matches the right chart data. Square markers represent data collected in a different set of experiments.

Taking these arguments, I assume that the dynamics of secretion levels are sufficiently fast with respect to the changes in internal fluorescence and populations fractions during the last 10 hours of experiment, to apply the idea of time-scales separation for the dynamics of these variables. Then, a *quasi*-steady state approximation is used (Bond *et al*, 1998). Thus, assuming that steady state is reached: $F_{ac}(t_{ss}) = 0$, and $F_{sc}(t_{ss}) = \frac{N_{sc}}{(N_{sc}+N_{ss})}$, where N is the number of secretors (*sc*), or sensor strain cells (*ss*), and t_{ss} is time at which steady state is reached. In the case of *IF*, the model simulations as the experimental data already reaches steady state at 10 hours. So, the differential equation looks as follows:

$$\frac{dSL}{dt} = k_{sec} \cdot k_{units} \cdot F_{sc}(t_{ss}) \cdot IF - k_{growth} \cdot SL$$

To understand how the $F_{sc}(t_{ss})$ is removed from the equation it is important to note that the sole data point that I fit for secretion levels is the taken at 24 hours after induction, at the end of the experiment. As indicated in Chapter 3, in all cases the measured secretion levels are normalized to the fraction of case study strain, $(1 - F_{ss})$, that at steady state corresponds to $F_{sc}(t_{ss})$, because no accumulators are observed. Therefore, since I fit secretion data already corrected for the fraction of secretors, the final equation to use in the model is:

$$\frac{dSL}{dt} = k_{sec} \cdot k_{units} \cdot IF - k_{dil} \cdot SL$$

For the k_{units} parameter value, it is fixed for all the simulations. It is done assuming that all POI fused the mNeonGreen followed by three copies of the FLAG purification tag in C-terminal have similar affinity for beads, and fluorescent properties, as the secreted mNeonGreen alone. Therefore, the experiments performed to characterize the secretory process of secreted mNeonGreen are used to assess k_{units}. From fitting such data, the median k_{units} value is then used to fit the simulations of XylC, scFv and amylase secretion. The obtained median value is 22.13 ± 7.99, and it is fixed to 22.13.

I perform model fitting again for mNeonGreen secreted only searching for k_{in} and k_{sec} (Figure 4.28). In the case of k_{deg} it is fixed to be zero, since no degradation is assumed, as explained in the section 4.1, and k_{growth} is known and constant. For mNeonGreen the simulation fits very precisely the data, for dynamics as well as steady states levels for *IF* and *SL*. The parameters values match pretty well with those predicted just fitting k_{in} and k_{out} with the previous model. However, here it seems that one of the replicas has in general lower values. It might be due to differences in secretion levels measurements for both set of experiments. Overall the fits are good, and the conclusions are in agreement with those exposed for the previous version of the model.

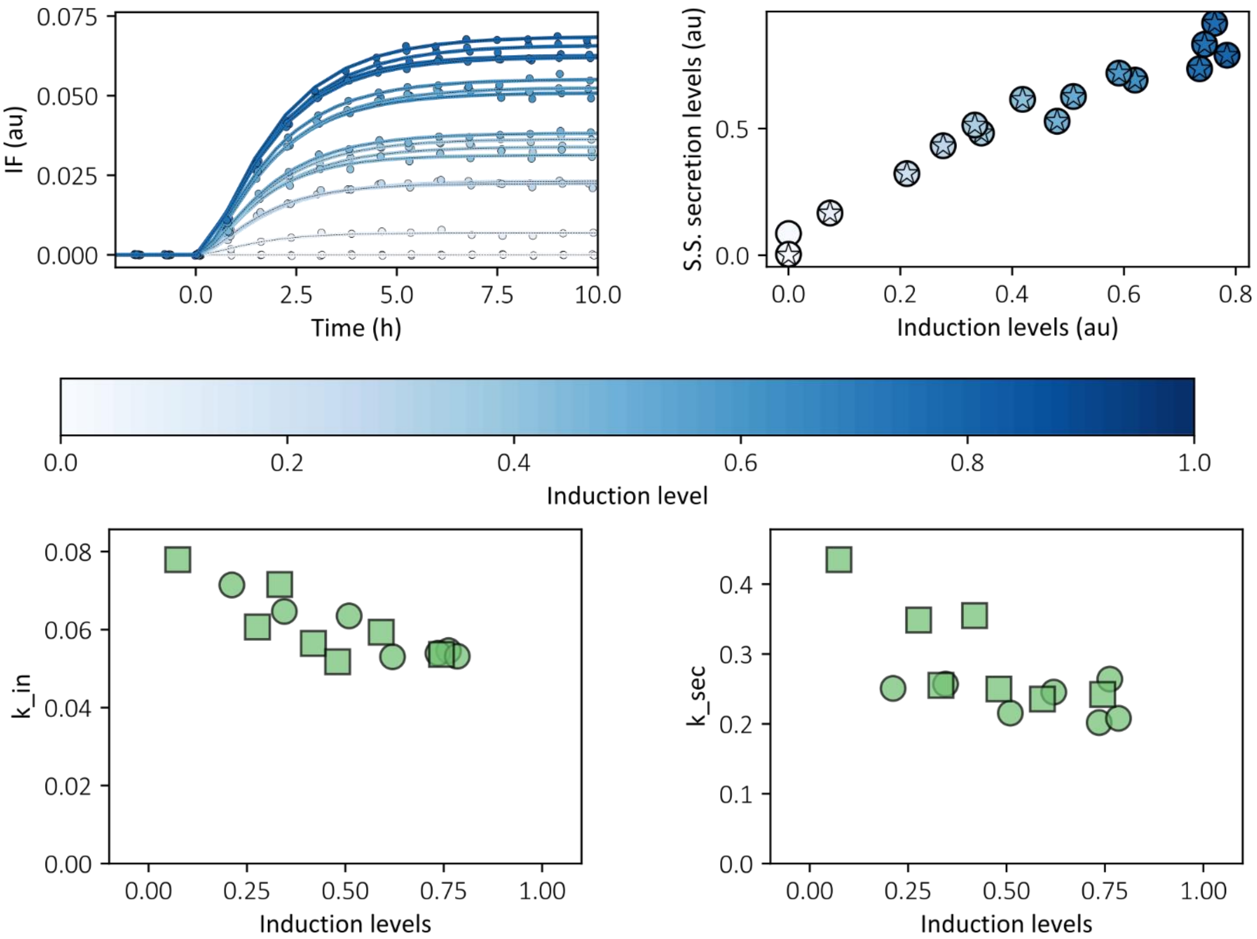


Figure 4.28. ***Comparison between data and simulation for mNeonGreen secretion.*** Parameters values have been obtained by independent-fitting strategy. At top left the dynamics of internal levels of fluorescence represented by dots, and the simulation results for *IF*. At top right the steady state levels for secretion data, represented as circles, and simulation as stars. At bottom left the k_{in} values for different levels of expression. At bottom right, the k_{sec} values for different levels of expression. The fitting is done for the time points between -2 and 24 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

4.2.5. Model extension to decompose using XylC as protein of interest

For second group of proteins k_{in}, k_{sec} and k_{deg} are inferred, since the rest of parameters are known either from data of from the previous version of the model. Figure 4.29 shows the inferred parameters for each experiment of XylC. The fitting (not shown) is in general good for all variables at dynamics and steady state. Interestingly, what was a constant k_{out} in the previous model, turns out to be the sum of a decreasing secretion efficiency and an increasing degradation activity.

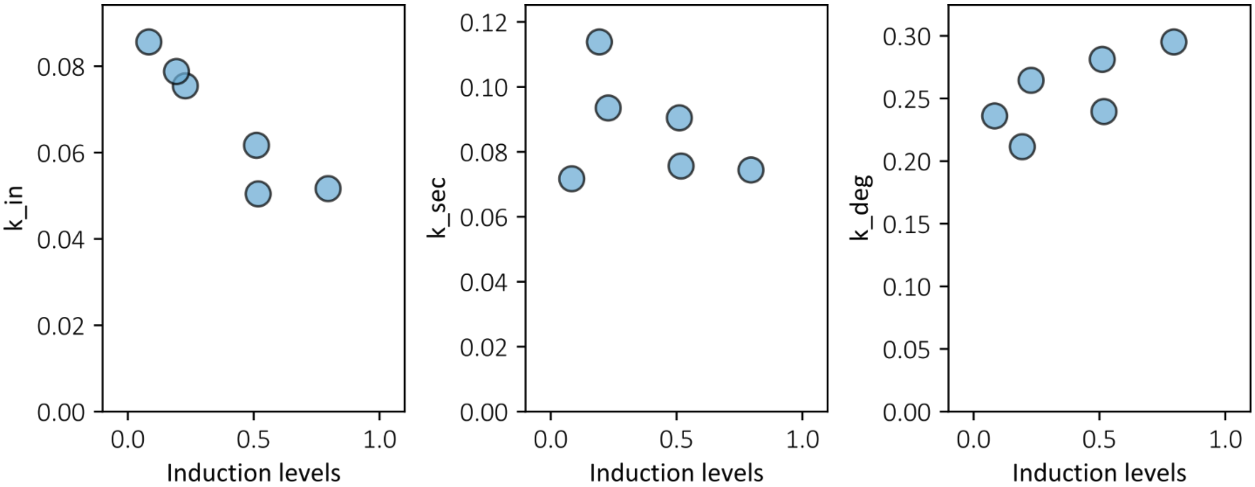


Figure 4.29. ***Comparison between data and simulation for XylC secretion.*** Parameters values have been obtained by independent-fitting strategy. The fitting is done for the time points between -2 and 24 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

4.2.6. Model extension to decompose using scFv as protein of interest

In the case of the scFv, the model is able to capture the dynamics of *IF* and the steady states for both *IF* and SL (Figure 4.30). The parameter k_{sec} is clearly at two levels, and that is probably a bias by the secretion measurements. Surprisingly, there is a peak in both set of data, as for model parameters as experimental data, at induction levels close to 0.5. Degradation increases with the demand, the relation seems linear. With all this features together, the secretion of this heterologous protein looks prone to have an optimal production demand point below the maximal levels of induction.

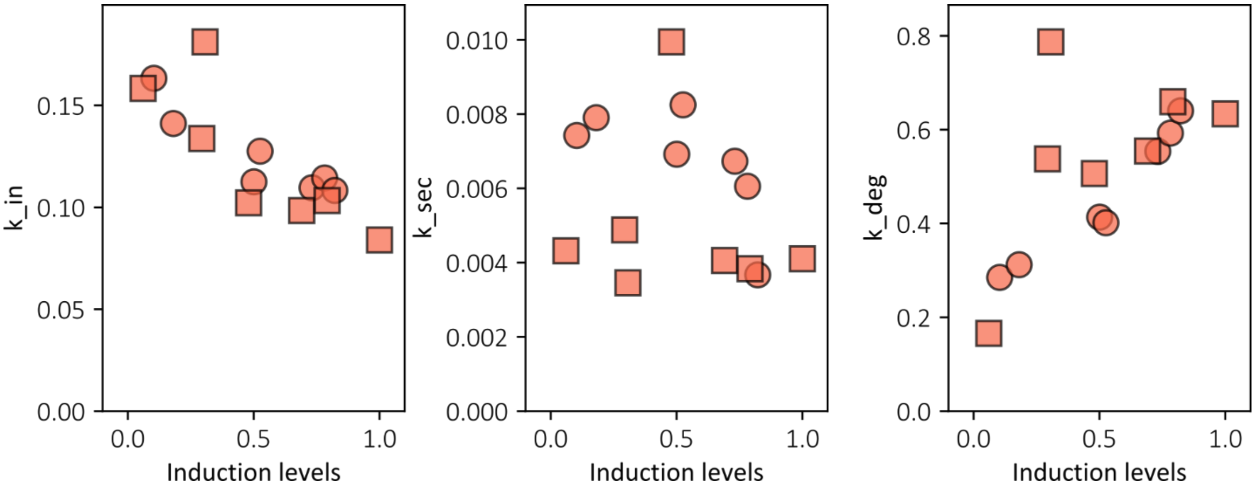


Figure 4.30. ***Comparison between data and simulation for scFv secretion.*** Parameters values have been obtained by independent-fitting strategy. The fitting is done for the time points between -2 and 24 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

4.2.7. Model extension to decompose using amylase as protein of interest

Finally, for the amylase, the model seems again very good fit with the experimental data (not shown). The parameters values for k_{sec} remain constant, and interestingly, k_{deg} seems to be activated even at very low demand levels (Figure 4.31). That is consistent with the results, since amylase is the secreted protein with the lowest secretion levels.

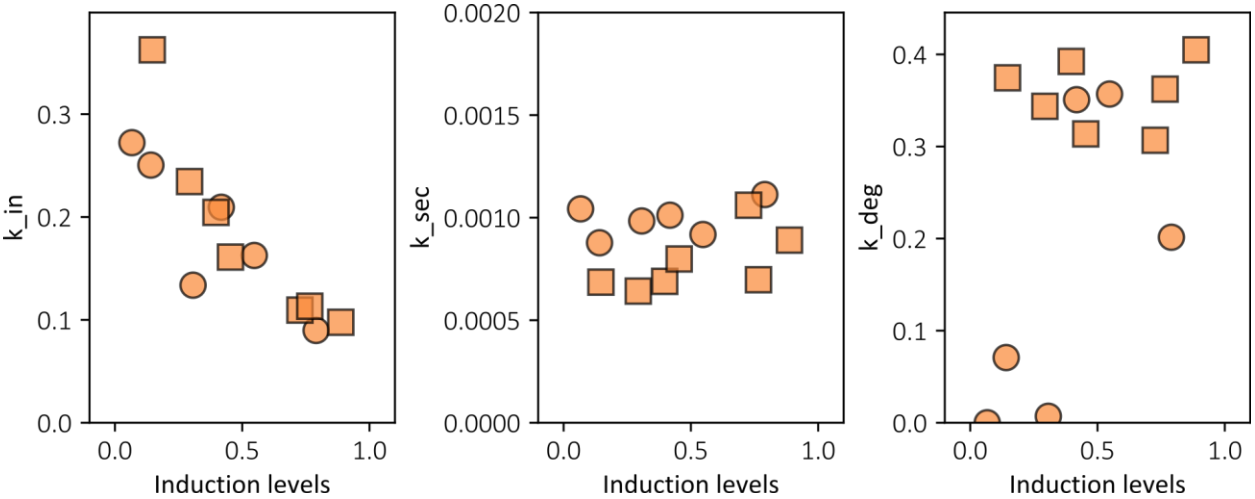


Figure 4.31. ***Comparison between data and simulation for α-amylase secretion.*** Parameters values have been obtained by independent-fitting strategy. The fitting is done for the time points between -2 and 24 hours, k_{tr} is fixed to a half-life corresponding to 20 minutes.

To summarize, in this section the characterization pipeline validated in section 4.1 has been applied for proteins of higher complexity that are of interest in biotechnological processes. During the characterization, non-monotonic dynamics of the median internal heterologous protein levels have been observed for two of the three proteins tested. That resulted to be the effect of bimodal populations generated by the emergence of a growth arrested non-secreting subpopulation, whose individuals are here called accumulators. This subpopulation ended up by either being diluted due to the continuous culture effect, or merging with the second subpopulation by adaptation. In contrast, the second population apparently maintained proteostasis, reaching faster the steady state levels of internal heterologous protein. Thereby, I focused the subsequent study in that balanced phenotype, here referred as secretor.

Once both population were identified and separated one from the other, the internal levels of fluorescence of secretor phenotype are used to continue with the characterization pipeline and analysis. The conclusion of the analysis is that secretion of heterologous proteins, even at low demand and relatively low complexity, produces excessive protein trafficking and compromises the capacity of the secretory system, giving rise to bottlenecks. This justified by the evolution of parameter values associated to the protein production, that decrease with increasing demand levels. As well as by parameter values describing protein removal that increase even at demands in which secretion levels do not, indicating that degradation is likely to be playing a role.

To understand the contribution of degradation to the secretory process, secretion data is used together with an extension of the model implemented in section 4.2. Then, the protein removal parameter can be decomposed into the growth, secretion and degradation components. Since there are data to identify the two first, the third can be also assessed by assuming that there are no other sources of protein removal. Using model fitting approaches the parameters values were inferred, and the contribution of each one for the expression of different proteins under increasing production demands has been represented and analyzed.

4.3. Integrated perspective

4.3.1. Secretory patterns

With this information of the system it is possible to compute the contribution of each of the processes described in the model, either secretion, degradation or dilution for each protein, and production demand level. Figure 4.32 shows the secretory patterns for the different POI characterized in the previous sections. It represents the estimation of the distribution of protein fates across the secretory process, according to the parameters inferred in the last version of the model. It has been assessed by doing polynomial fit of degree two in the parameters values as a function of the induction levels.

The analysis shows that most of the produced protein is accumulated and eventually diluted at low levels of demand in all cases. For the proteins of group 2, as production demand increases the fraction of protein targeted to degradation also increases, until some level of plateau in the case of scFv and amylase. Specifically, for these two POI degradation is massive in compare to secretion. As seen in the parameters values inferred, the secreted protein fraction also decreases with the increase of induction levels. However, when scaling with the production term $k_{in} \cdot I$, secretion also increases, as in the case of mNeonGreen. Otherwise, for the rest of proteins, the protein secreted saturates or even decreases as for scFv.

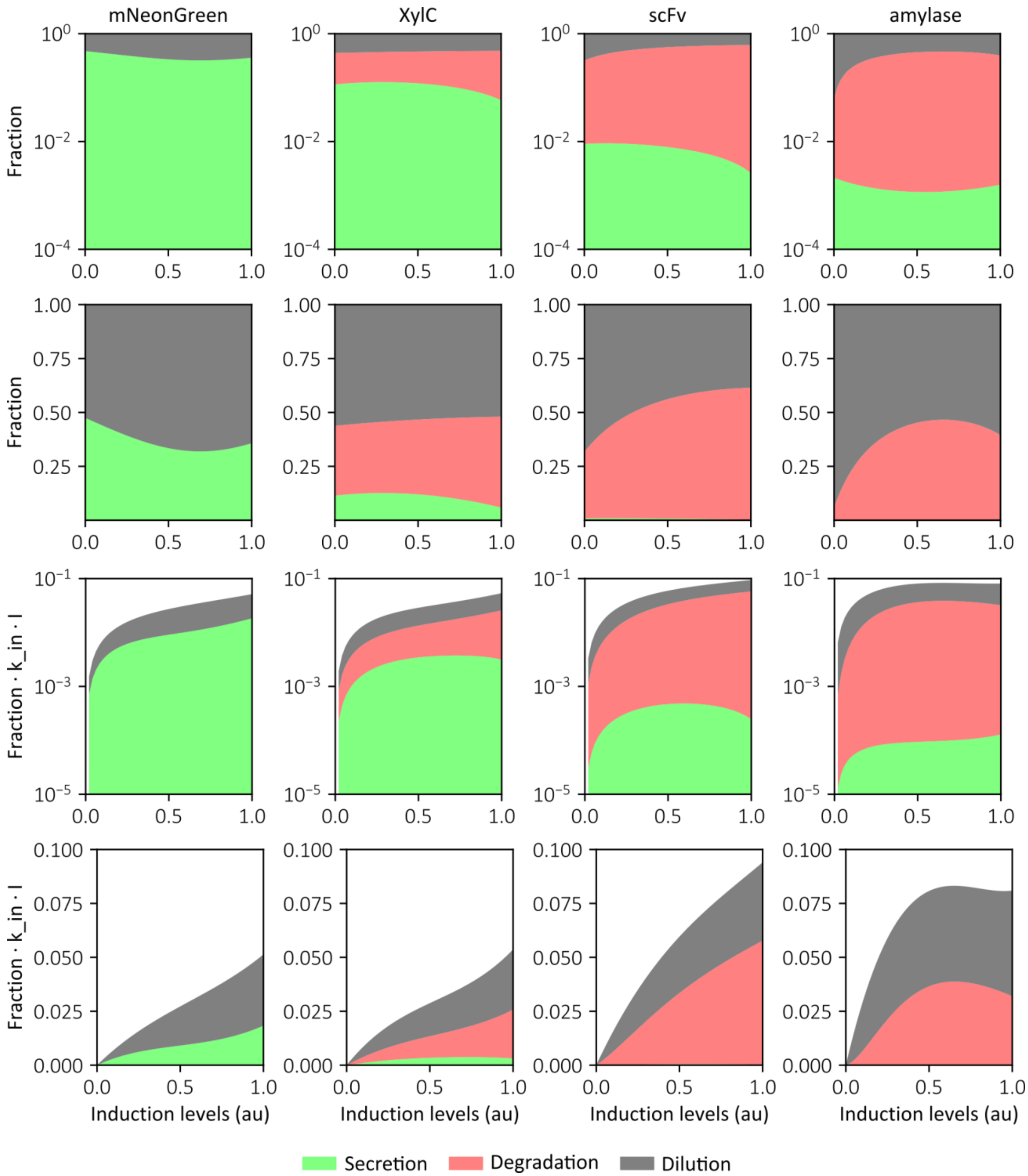


Figure 4.32. ***Secretory patters describing the protein's fate distribution at the steady state in the secretory process.*** Each column corresponds to a protein of interest, specified on the top of each column. The first and the second row shows the fraction of contribution of each of the three processes defined in the mathematical model for protein removal, secretion in green, degradation in red, and dilution in gray. In logarithmic scale the first row, and linear scale the second to facilitate interpretation of the results. In the third and fourth rows, the fraction is multiplied by the production term in the differential equation describing *IF* dynamics ($k_{in} \cdot I$). In logarithmic scale the third row, and linear scale the fourth.

4.3.2. UPR analysis

Until now, the relation between production demand levels and the contribution of degradation and secretion for protein removal has been studied. The adaptive response UPR is known as a key regulator for these processes. Given that when secreting either scFv or amylase transient bimodal populations are observed, it is worth to check such behavior also in the single cell data for UPR.

When looking at the single cell data, two populations are also observed for the UPR signal. Moreover, when applying the same procedure as that used to differentiate the accumulator phenotype with internal levels of POI, two populations with different UPR levels are also identified (Figure 4.33). Those populations are clearly correlated with each of the phenotypes identified previously. Indicating that UPR levels and internal levels of POI define similar populations. Thus, accumulator cells have an acute stress and exacerbated UPR.

Note that as previously done for the internal levels of fluorescence, the study is going to focus in the population of secretor phenotype. So, from now on every time that UPR is mentioned, it will refer to the UPR corresponding to the secretor population. If at some point there are references to the overall UPR, or the accumulators UPR, it will be specified.

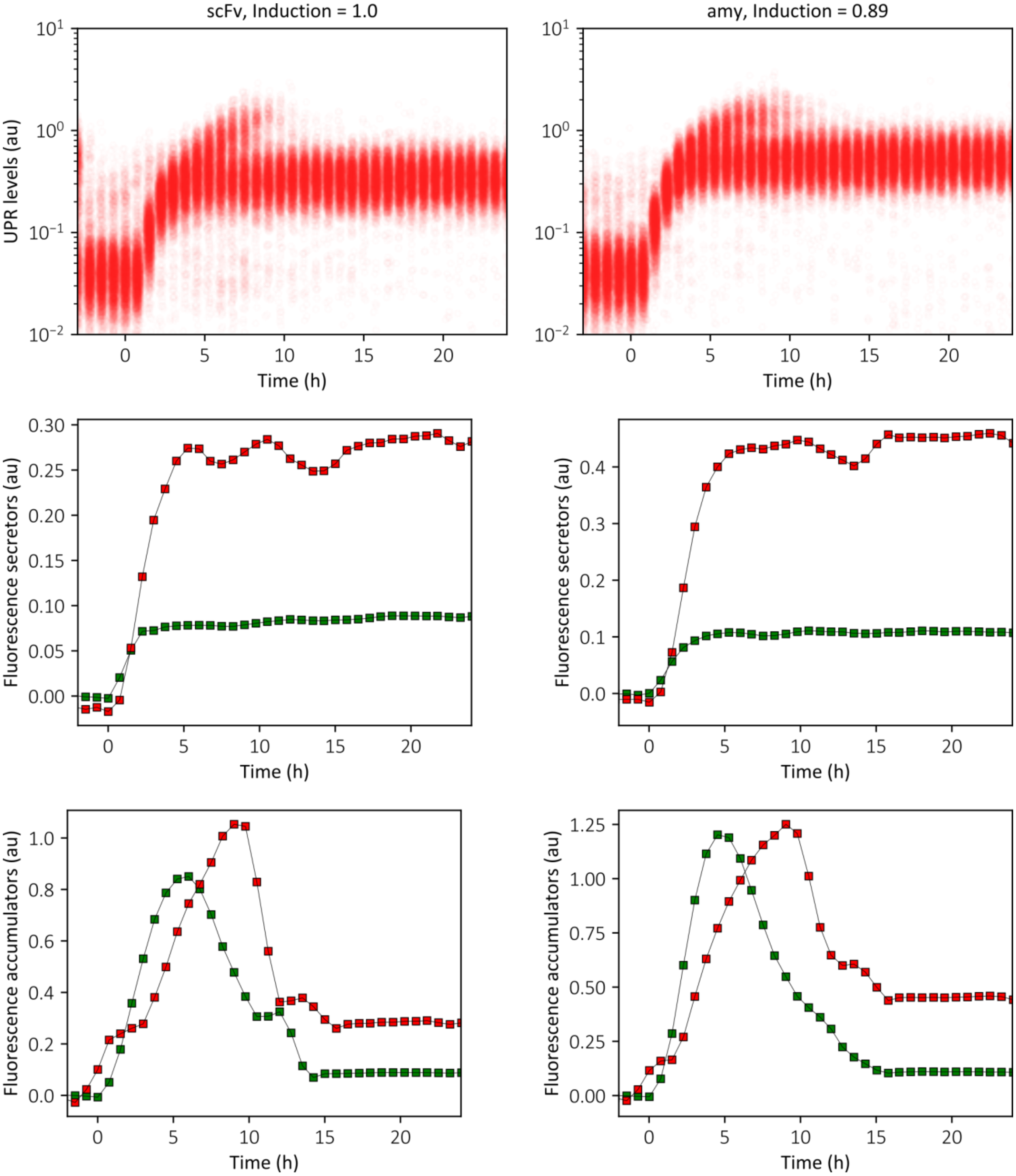


Figure 4.33. ***Differentiation of secretor and accumulator subpopulations from the overall population, and comparison of UPR with the internal levels of POI.*** The left column corresponds to scFv data, the right column to amylase data. In the top row, the overall populations at the single cell level, over the time of the experiment, is shown. In the middle row the median for the secretor phenotype is shown, for UPR and POI levels. On the lowest row the same data but for accumulators. In red the data associated to the UPR signal, and the POI data in green.

Interestingly, a delay is observed in the UPR dynamics with respect to internal fluorescence in the accumulators data. To quantitatively analyze the delay among the two signals, I computed the cross-correlation between the two time-series to assess the lag value for both phenotypes during the first 15 hours after induction. To do it I use the *Python* package *scipy.signal.correlate* with the default settings, such that it returns the correlation of the two time-series with lags from -15 to 15 hours. Then, the lag that yields the best correlation is considered as the time delay between the two variables. It is important to note that both fluorophores have a difference of around 15-30 minutes in the time for fluorophore maturation (Balleza, Kim, & Cluzel, 2018; Guerra et al., 2021).

In the case of the secretors, for both scFv and amylase secretion data, no delay is detected, both signals increase synchronously. However, in the case of the accumulator phenotype, the lag that best correlates both signals during the first 15 hours of induction depends on the demand levels. For scFv all samples with induction levels above 0.4 show a lag of around 1.5 hours, except the maximal induction level, wherein the lag is approximately 2.5 hours. In the case of amylase, the lag is close to 1.5 hours for inductions between 0.4 and 0.5, all the rest show the best correlation at a lag around 2.5 hours. That suggests that such delay may depend on the nature of the protein expressed and the production levels. It could be due to the time that the UPR needs to first detect and respond, and then deal with the over-accumulation of protein.

Nevertheless, for both phenotypes data shows that UPR is correlated with internal levels of POI, thus, not surprisingly it is triggered by protein accumulation. Moreover, it starts to decrease when the internal levels of POI are low, once the proteostasis is restored, probably by degradation of the accumulated protein, according to Figure 4.33. As matter of fact, when comparing UPR signal with internal levels of fluorescence, both at steady state (Figure 4.34 top), it is observed how at high levels of demand, the internal level of POI reaches a plateau, but the UPR still increases. It suggests that the UPR is inducing the degradation of POI, thus keeping the POI levels constant despite the increase in production demand. That supports the idea arguing that accumulator subpopulation adapts by degradation of internal proteins, giving rise to a phenotype that has the same internal levels of POI than the secretor. Finally, note that given that the steady state levels of UPR are correlated with the induction levels (Figure 4.34 bottom), the relation of demand levels with the secretory system patterns can be extrapolated also to the UPR with such patterns.

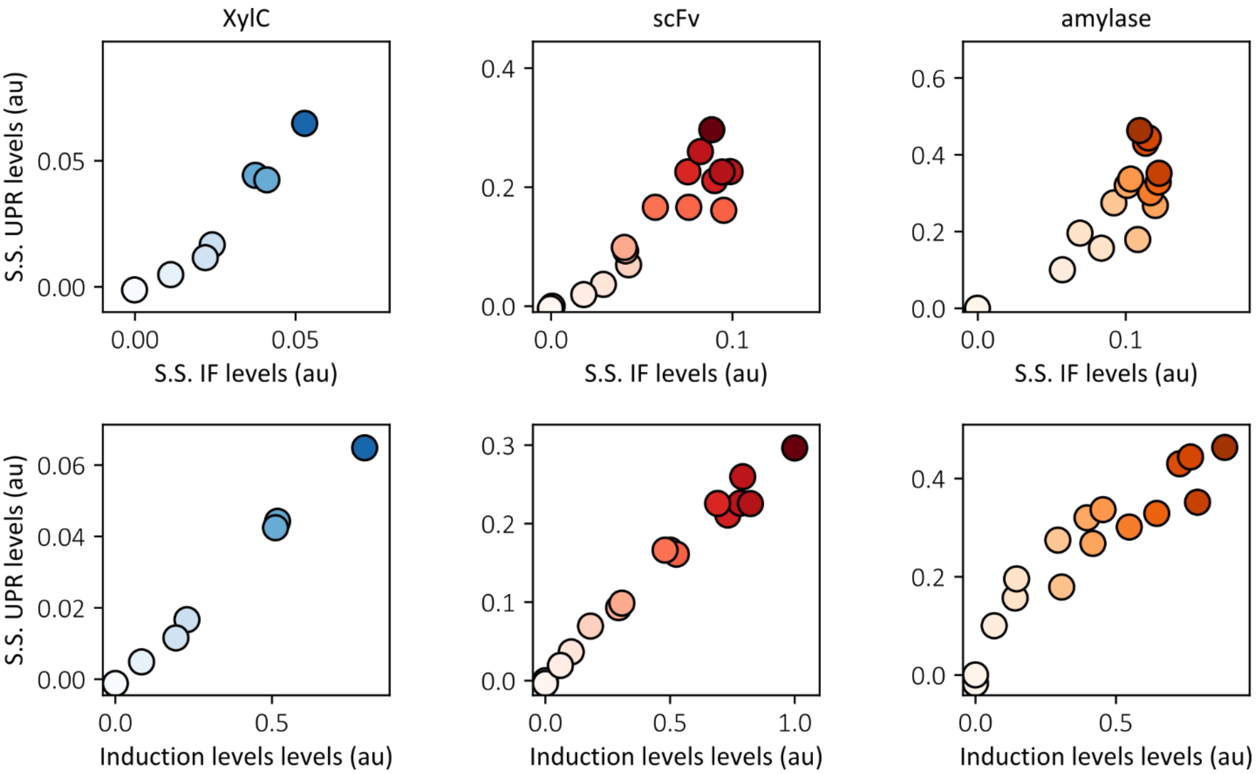


Figure 4.34. ***Relation of UPR and internal fluorescence levels at steady state.*** Top row represents the unfolded protein response (UPR) levels in relation to the internal fluorescence (IF) levels, both at steady state (S.S.). The bottom row shows relation between UPR at S.S. and the induction levels. The intensity of the color is proportional to the induction levels.

To conclude with this section, here we first have integrated the results from the section 4.2 to build the secretory patters. These representations of the system show the contribution of each internal protein removal process described in the model to the overall protein traffic. Interestingly, it turned out that for the proteins of the second group studied here, secretion is the least likely fate in the secretion process. Specifically, for the secretory process of scFv and amylase, those are massively degraded if compared with fraction of POI undergoing secretion. The contribution of the different removal components changes with the induction demands, suggesting that might be an induction levels different from the maximal that optimizes the process.

In the second par to the section, the UPR data has been analyzed for the presence of bimodal populations, and so related the accumulator phenotype with an increased UPR activation. That fact has significant implications in this project because it suggests that adaptation is likely reached through degradation. Therefore, an adaptation of accumulator should not be interpreted as a secretor, but rather as a “degradator”. To support that arguments, the impact of UPR on internal levels of POI fluorescence at steady state was analyzed. It showed that high UPR levels are correlated to the saturation of internal levels of fluorescence. That is probably caused by the ERAD activation, as secretion levels do not increase at those levels of induction. As UPR at steady state and induction levels are very correlated, an increase of UPR is also associated with an increase of k_{deg}. Therefore, maintaining high levels of UPR could be detrimental in the context of obtaining high efficiency for POI secretion.

Along the lines of what has been presented in this section, the impact of induction levels and UPR in secretory efficiency, as well as the possible adaptation process of the accumulator subpopulation, in the next section I am going to comment and suggest a potential biotechnological application by combining the results obtained and the technological developments presented in Chapter 3.

4.4. Using the characterization pipeline for bioproduction

4.4.1. The role of accumulators in the overall secretion levels

In the previous sections I characterized the secretory process for a panel of heterologous proteins with different complexities. To do that I used a systematic pipeline that generates information about the capacity to secrete the POI, and the sources of its removal from within the cell, including the distinction between secretion and degradation processes. Interesting biological phenomena have been observed and analyzed, such as the high contribution of degradation to the process, the presence of different phenotypes, and some implications of the UPR activation in each of them. In this section I am going to propose potential applications for such knowledge in the context of biotechnological processes. I will also take advantage of the technological contributions described in Chapter 3, specifically the capacity to perform real time control experiments in the multi-bioreactors platform.

Here I propose to carry out feedback control based on the UPR reporter signal. I could also propose feedback on the internal levels of POI, but it requires to use fused fluorescent reporters, and that is not feasible in all biotechnological applications. Whereas using a UPR reporter clearly proposes a general biosensor for secretory-associated stress, independently of the POI secreted. Along these lines, I suggest a way in which applying real-time control to maintain the UPR levels may improve biotechnological processes by increasing the proportion of cells secreting the POI with respect to those degrading it.

From the analysis performed in previous sections, it clearly appears that the presence of the accumulator phenotype is detrimental for secretion efficiency. It is very probable that these cells adapt by reducing the internal levels of POI activating degradation, until they reach proteostasis at the same levels as the secretor phenotype. Therefore, the new subpopulation of adapted accumulator is not a secretor, but a high degradator. It implies that the population in the reactor is enriched of these non-secreting cells, thus the levels of secretion do not increase even if the induction levels do. As a matter of fact, Figure 4.18, Figure 4.20, Figure 4.23, and Figure 4.24 show that when the fraction of accumulators rise, the relation between secretion and induction is not linear anymore.

We observed that in a population of cells secreting proteins with lower secretory complexity at high production levels, the accumulator phenotype arises all along the experiment but does not lead to a significant subpopulation (Figure 4.35). The reason might be related to an insufficient UPR response associated to the POI at a given production demand. Indeed, if the UPR does not reach a threshold, there is not adaptation and cells presenting accumulator phenotype are diluted because the growth arrest triggered by an accumulation of POI. Above certain levels of UPR, the cells upregulate active degradation and recover proteostasis, thus they can restore growth and merge with the secretors population, but maintaining high degradation capacity.

It is important to note that the fraction of the population that becomes accumulator has an impact on adaptive capacity as well. When the fraction of accumulators is sufficiently high, it affects significantly the overall growth rate, with the subsequent increase in the residence time of the accumulator in the reactor. Thus, it gives extra time to the accumulator cells to adapt before being diluted out of the reactor.

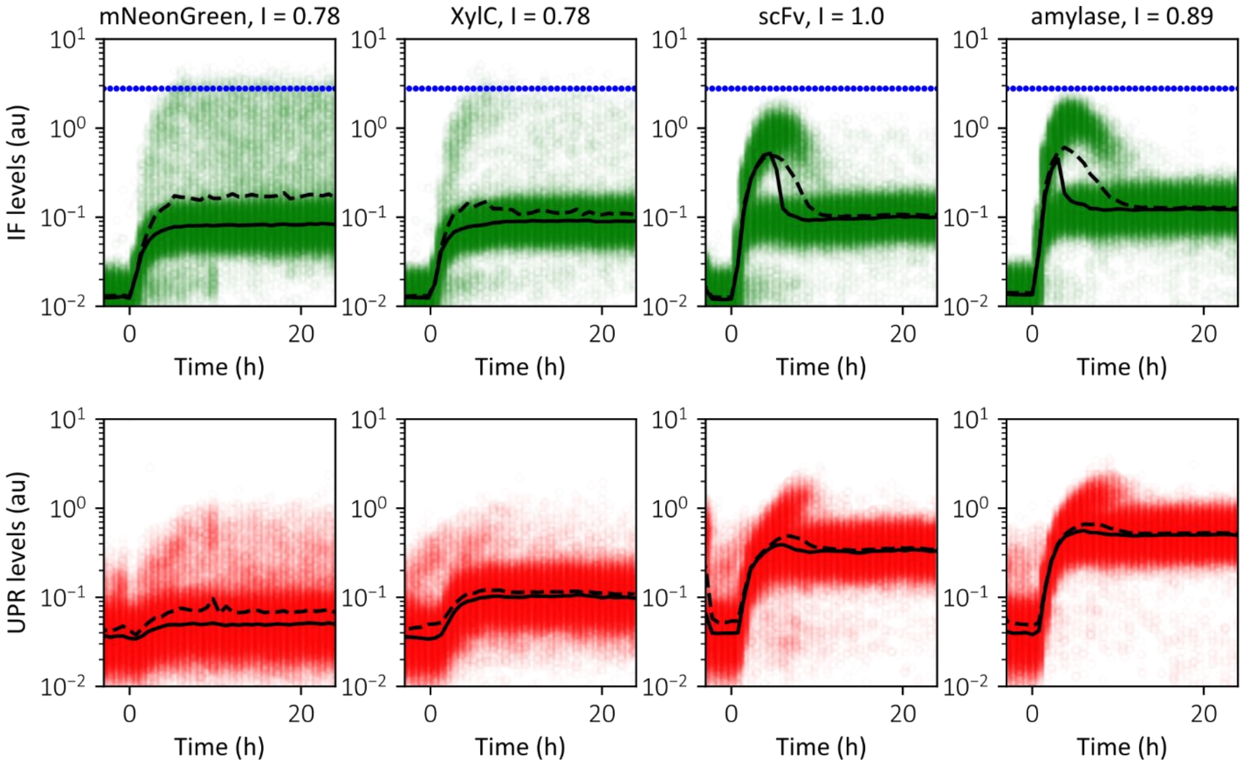


Figure 4.35. ***Internal levels of fluorescence and UPR distributions of cells expressing different POIs over the length of the experiment.*** Data corresponding to the expression of four different proteins of interest (POI), one per column and indicated on top. Top row corresponds to internal levels fluorescence of the POI (IF); bottom row corresponds to signal from UPR reporter. The mean is represented as a dashed line, the median as solid black line. The dotted blue line on the top row represents the median steady state IF levels for the non-secreted mNeonGreen from Figure 4.3.

4.4.2. Feedback control experiment

Performing feedback control to maintain the UPR levels below the adaptation threshold seems a good strategy to maintain a homogeneous population of secretors. For the real-time feedback control, the mean of UPR levels might be a good variable to target for control, because it is biased by bimodal populations if the subpopulation is sufficiently dense. To set the threshold value, I propose to use the steady state UPR level, as it is the stress corresponding to a balanced proteostasis. As the optimal UPR target value is probably POI dependent, a first characterization is necessary. Once that point has been identified, feedback control can be implemented.

At the time of writing this document I have not yet performed such control experiment. However, I have produced preliminary data showing the capacity of the platform to carry out close-loop feedback control on the UPR levels on cells expressing XylC (Figure 4.36). I implemented a BANG-BANG controller on the UPR median value to avoid the population to stay above the threshold. Such controller stops induction when the data is above the threshold, and *vice versa*. The experiment target value has been set in an *ad* *hoc* manner. In Figure 4.36 the UPR median value is controlled and maintained below the UPR target value. There is a subpopulation observed in the internal POI levels, but not in the UPR signal. Oscillations around the target levels are observed because the controller used is not optimal. The light induction is triggered when the last measurement time point is below the threshold, so when demand is triggered the UPR levels have decreased significantly. Unfortunately, the secretion measurements at that time were not prepared to be done in a systematic way, so there is not results of secretion levels associated to this experiment.

To conclude, in this section I have comprehensively analyzed results to implement a feedback control strategy. First, I have described a hypothesis on the impact of the accumulator phenotype on the overall population, and its significance in the final levels of secretion. The presence of accumulators is inevitable; it is probably due to random variability in the secretion rate. The accumulators do not secrete, and adapt by degradation, so they regulate their internal levels of protein and restore growth. Hence, they manage to consolidate their presence in the population. The consequence of accumulator adaptation in secretion levels is that the fraction of secreting cells in the population is decreased. Therefore, the overall quantity of secreted protein does not increase despite the increase in demand, since higher induction levels produces higher quantities of accumulators.

It has been observed that if UPR levels are not sufficiently high, single cells presenting accumulator phenotype are likely to not adapt. The presence can be detected through and exacerbated UPR signal with respect to the secretors, due to and acute stress lead by over accumulation of internal POI. That makes the UPR levels a good variable to control. One advantage of identifying the presence of accumulators by monitoring the UPR signal, is that it is a biological marker orthogonal to the POI. Meaning that it can be used for a wide range of secreted POI in a generic manner. Hence, performing feedback control on the UPR levels is the next step to carry out in this research project to ascertain the proposed hypothesis.

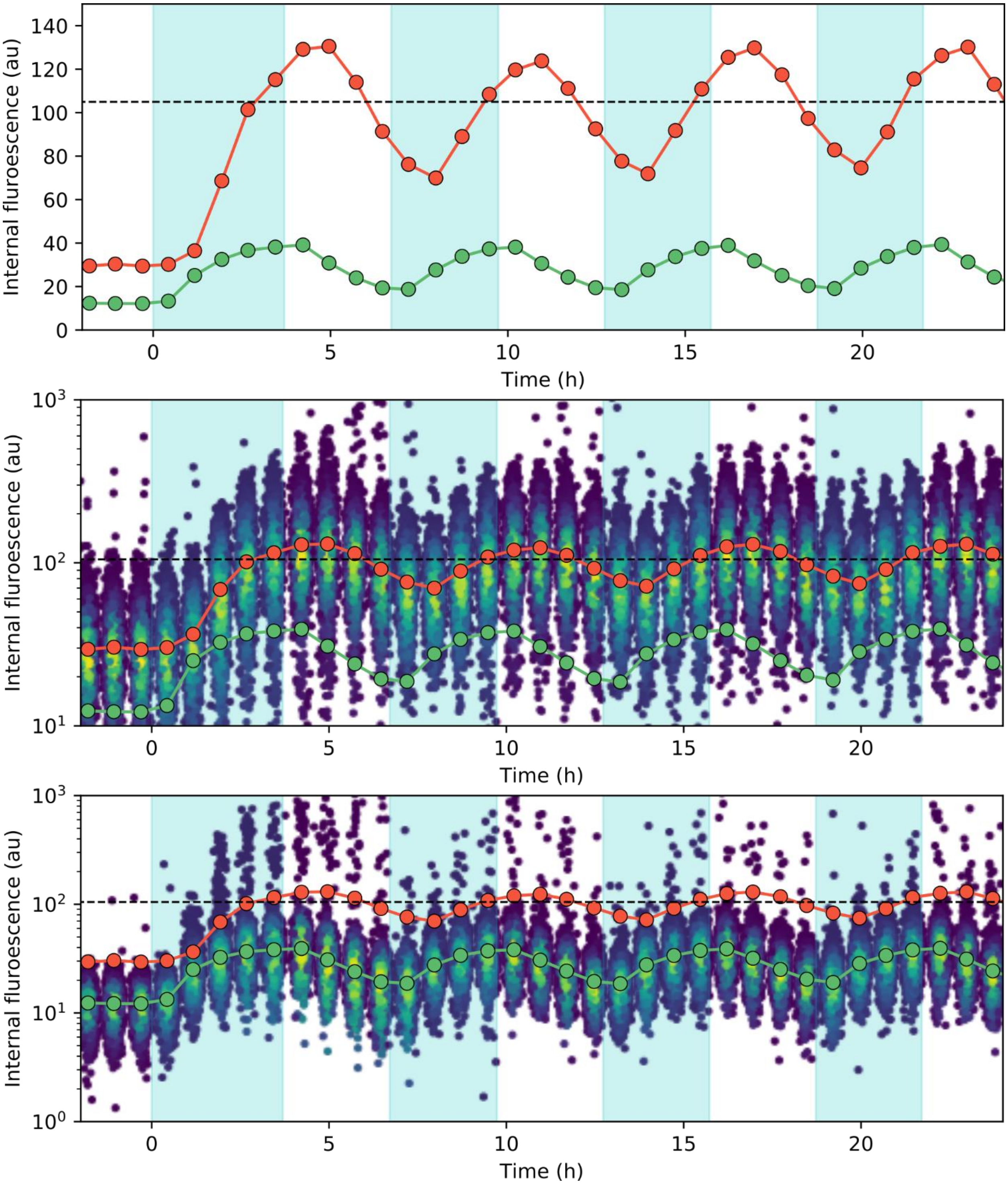


Figure 4.36. ***Feedback experiment to maintain the median UPR value below a threshold in a population of cells secreting XylC.*** On the top plot the medians for internal protein of interest (POI) levels, in green, and for the UPR, in red, are shown. The light induction periods are highlighted in blue, and the threshold (105 (au)) value with a black dashed line. In the middle plot the same information is shown but adding the population of UPR measurements. In the bottom plot the same information is shown but with the population of POI measurements. The color of the distributions goes from purple to yellow for increasing cell density. The fluorescence is not normalized to forward scatter. The light is turned on when the last data point (median) was below the threshold. Then turned off when the median passes above the threshold.

4.5. Conclusion

In this chapter I have presented, validated and applied a pipeline to characterize the secretion process of proteins with different secretory complexity. In the first section the experimental and modelling procedure have been described and validated. Using proteins of low secretory complexity, I have quantitatively assessed the quality of the methods and selected the most suitable. In the second section, three different heterologous proteins with different posttranslational features that increase their secretory complexity have been characterized. That has allowed to propose an overview of the secretory processes by decomposing the contribution of the different sources of protein removal, and the presence of bottlenecks at the level of trafficking. During the characterization process, heterogeneity of the population has been observed, and identified as possible cause of inefficiency in the secretion levels. By studying the relations of the UPR and the population heterogeneity, a feedback strategy has been proposed to increase secretion levels. It led to the hypothesis that controlling the UPR might reduce heterogeneity, hence the fraction of cells secreting the heterologous protein is maintained as high as possible. Finally, a demonstration of a feedback experiment controlling the UPR levels has been presented.

Despite that there many arguments supporting the idea proposed above, the feedback experiments are not performed and the hypothesis is not validated. However, there is literature describing the heterogeneity in populations of yeast secreting heterologous proteins, and the adaptation of low secreting phenotypes by degradation (Love, Panagiotou, Jiang, Stadheim, & Love, 2010; Love et al., 2012). Even if the feedback experiments are not a real solution, it is very likely that the presence and consolidation of the accumulator phenotype is an issue in the context of bioproduction.

The presence of a population of non-secreting cells that growths and competes for resources, limits the fraction of those cells secreting the POI. In batch cultures wherein growth and production take place at different times, preventing the presence of accumulators would not be possible with the feedback here proposed. That is because with the turbidostat approach, the accumulators are discarded by growth arrest and dilution, whereas in a batch culture there is not dilution. However, if the fraction of accumulators is significantly high, the residence time might increase, thus favoring adaptation.

Regarding the experimental limitations of the data shown in this chapter, one clear thing to improve is the reproducibility of secretion measurements for low levels of secretion. It is clear that there are discrepancies in the different set of experiments. Such differences are in the quantitative but do not affect the qualitative behavior of the secretion levels for a given POI. Something also needed for the quality of the data is to generate more data points for those POIs where only one set of 8 experiments is available, such as for non-secreted mNeonGreen, and XylC.

Finally, in reference to the contribution of this work, the set of data generated regarding dynamics of the secretion process as it is produced in this project, has not precedent to the best of my knowledge. That will be openly available once published and it is expected to be significant and contribute to extend the knowledge in the field of protein secretion in eukaryotes. More models can be proposed to explain the data, and alternative ways of measuring secretion can be also performed. The presented pipeline to characterize the pathway can be applied to other types of stress and processes. It has shown potential to contribute to applied research in protein production in general and secretion in particular, and can be extended to an endless number of proteins.

In terms of industrial applicability, scaling up is challenging, and just few approaches really manage to be applied. Despite that, in the last section I used the knowledge obtained in the previous ones to propose a general approach to characterize and improve the secretory process. It is general because it uses the UPR reporter that is a common biomarker for secretion-associated stress for any protein. In theory, the approach does not need to deal with the POI, besides cloning it. Therefore, in the best case scenario, just adding a UPR reporter to any strain of interest and finding the steady state level of UPR to perform the feedback would be sufficient. Still dealing with population heterogeneity by applying inputs at the population level is probably a limited approach in many contexts. The UPR threshold level might be difficult to assess, since the rise of accumulators is a single cell process and that cannot be monitored and controlled in bioreactors setups, especially in the industrial scope. In the Annex A I propose an internal feedback based on interference RNA that controls the production at the single cell level in order to maintain homogeneous populations.

Chapter 5: Discussion

In this section I will first summarize the thesis, giving an overview of what have been seen throughout the chapters, and what have been the most relevant points of each in a synthetic manner. Then, I will discuss more in detail the main contributions of this work, one by one, and finally I deliver what I consider to be the take home message of the main results of the project from the scientific perspective. To end with the most technical aspects of this chapter I will go through limitations of the work, and how I would tackle them. Finally, I will discuss the next steps I hope to take to improve the quality and extent the results I have presented, followed by a description of what I envision myself doing in the long-term in the context of the project.

5.1. Thesis summary

This thesis started with an introductory overview of the general aspects of bioproduction and its potential impact in the coming decades. At the same time, I also commented on some limitations for the engineering of biological systems and related processes in the context of bioproduction. Then, I presented our approach for developing a generic way to characterize secretory processes. Finally, I listed the most significant contributions of this PhD project.

In Chapter 2, we went in more detail through the state of the art approaches, the concept of chassis cell and its limitations in the field of bioproduction, and the quantitative approaches to study secretory processes. Then, I also introduced the technological challenge for studying secretory processes in a systematic manner using continuous bioreactors while emphasizing the need for automation of measurements, single-cell resolution data, and simple, quantitative measurements of secretion levels.

Along these lines, in Chapter 3, I presented the technological developments that we have carried out for the characterization pipeline of the secretory process. A multi-bioreactor platform with automated cytometry measurements, that enables feedback control of eight cell cultures in parallel is presented together with a method to measure secretion levels in a simple and rapid manner. In this chapter I have also described the strategy for constructing a set of strains for real-time monitoring of different secretion-related characteristics, including internal POI and stress.

Taking advantage of the platform, the methods, and the strains mentioned, I describe in Chapter 4 how we implemented a systematic pipeline to characterize secretory processes. First, using experimental and computational methods I validated the framework with a group of proteins. Then, a panel of three other different proteins of increasing secretory complexity was characterized to obtain an overview of the secretion process and the patterns of protein fates. During the study of the second group of proteins, bimodal behaviors defined by internal levels of protein have been observed. The knowledge acquired on these populations was used to propose a generic approach based on feedback control aiming at increasing secretion efficiency in bioproduction processes.

5.2. Discussion and conclusions

5.2.1. Technological developments to study secretion in yeast

In this project we have developed a systematic characterization pipeline that offers an integrated view of the quantitative aspects of the secretory process. The characterization workflow relies in a customized multi-bioreactor connected to an open source robot that allows automated cytometry measurements providing high temporal resolution, described in Chapter 3, section 3.1. Furthermore, beyond the sampling automation function, the integrated nature of the platform also enables to perform real-time control experiments, that can be used to apply feedbacks at the population level. The manuscript of the article relative to the bioreactors platform and the strategy to carry out reactive experiments is available in Bertaux et al., 2021. It is important to note that I have contributed to the development of this platform as part of the team of developers, fulfilling the role of main tester. My specific contributions have consisted in providing specifications to define the necessary functionalities based on experiments, as well as performing troubleshooting actions when necessary. I have also actively collaborated with Dr. François Bertaux in the improvement of robustness by performing metrological studies.

To monitor secretion levels in a simple and systematic manner, we have also developed a method based on magnetic immuno-beads that enables for measurements in an automated fashion using a flow cytometer. The development of this methodology was extremely challenging, and we had to face numerous problems along the process. Some of the major causes of issues have been related with maintaining proper redox conditions in the culture media in our setup for continuous cultures, the features related to the beads size and the constrains of the cytometer circuitry, and the fine tuning of the protocol allowing *in situ* measurements directly from the cell culture. Finally, automation has not been completed at the present time, nonetheless we have exploited the potential of the method as it is for now, and we will continue developing it.

5.2.2. Development of a systematic pipeline to infer secretory patterns

High production demands for the POI leads to excessive trafficking through the secretory pathway, leading to the emergence of bottlenecks and accumulation of unfolded protein. This produces stress and eventually the degradation of the POI. Thereby, expression strength is not always correlated with secretion levels. By combining the technological developments mentioned above, and an engineered set of yeast strains for controlling production demand, and stress monitoring in real-time, we developed a generic pipeline for the characterization of the secretory process. Such characterization pipeline (Figure 5.1) aims to establish a standard framework able to capture the dynamic behavior of the system under different conditions, and so provide a picture of the secretory pattern of the system for such conditions. Following this systematic process, we have been able to generate a unique set of data on the dynamics of the secretory process. The quantitative analysis of these data also helped to characterize in depth the behavior of the cellular system by studying the secretion process of three proteins with different secretory complexity under different production demands.

The results indicated that most of the protein that enters the secretory process is either degraded or diluted due to growth. In the best case scenario, for a protein with low secretory complexity at mild induction strength, and no UPR triggered, only 40% of the produced protein is secreted. We also quantified that the secretory capacity for this protein decreases to 60% of its maximum when production demands are high. Interestingly, the presence of the bottlenecks causing such decrease is not detectable just by looking at steady state levels of internal POI.

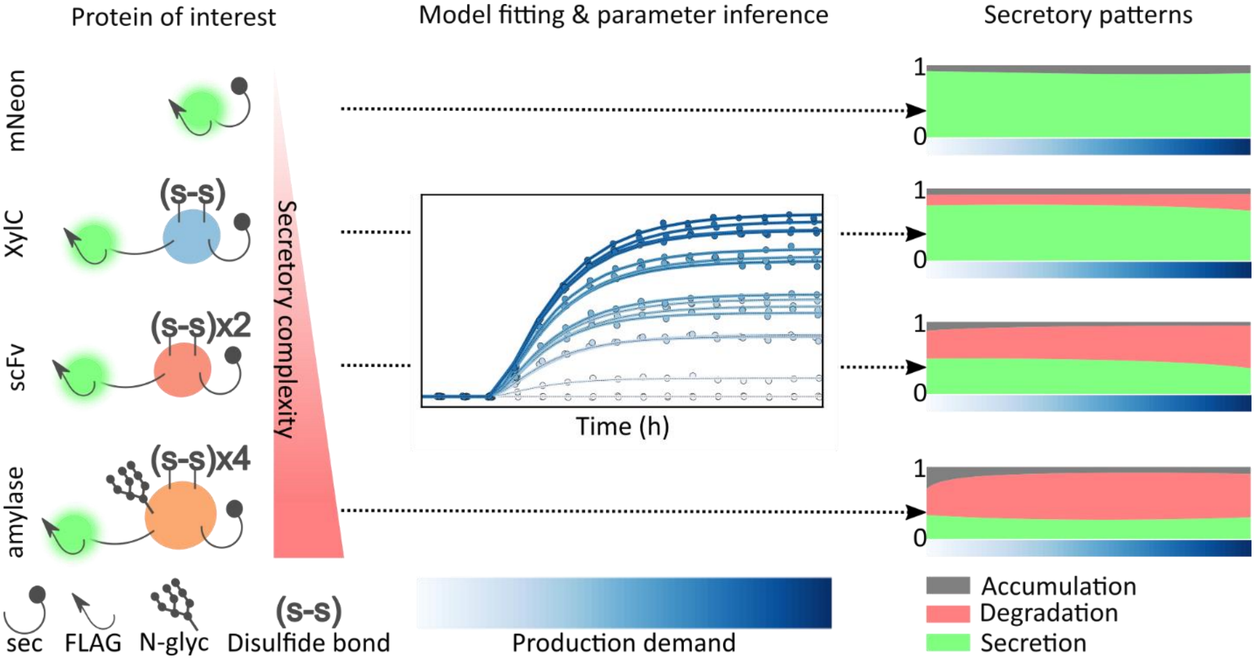


Figure 5.1. ***Schematic description of the pipeline to characterize protein secretion.*** From left to right, a panel of secreted proteins with different features of interest fused to a fluorescent reporter to quantify them, and a purification tag to measure secretion levels is used for its characterization. In parallel, a mathematical model describing the dynamics of the system is implemented to be able to infer processes that are not monitored experimentally, most notably degradation rate. Once the dynamic data for all proteins and different induction levels is generated, the model is fitted to the data in order to infer parameters value. Then parameters values are used to assess the secretory pattern of the system under different conditions. Sec: secretion tag, FLAG: purification tag, N-glyc: N-glycosylation.

5.2.3. Population effect in the secretory process

We found that for secretion of proteins with high secretory complexity, the populations defined by internal levels of POI were bimodal during the first 10-15 hours after induction. With one subpopulation of cells presenting high accumulation of internal protein, followed by acute levels of stress, and subsequent analysis showed that those cells are likely to be growth arrested. After adaptation by triggering a high UPR, and consequently degrading the POI, these cells restored their normal phenotype for internal levels of POI and UPR. That was an exciting and interesting observation, and our interpretation is consistent with the adaptation scheme proposed by the study discussed in Chapter 2 (Love *et al*, 2012).

In that study, by using computational approaches, the authors obtained three phenotypes based on the ratio of secretion rate over degradation rate. Out of these three phenotypes, only two are distinguishable by looking at the level of internal protein at steady state. The population with high secretion rate, and the one with lower, show the same internal levels of protein when the second has adapted by increasing degradation. I considered that as a highly probable explanation of what we have observed in the transient bimodal behavior mentioned above. Thus, I called accumulators to the cells with low secretion rate, low degradation rate, growth arrest and high internal POI, adapted accumulators to cells with low secretion rate but high degradation rate and restored growth, and secretors to the ones with high secretion rate and normal growth (Figure 5.2).

From that rationale, I hypothesize that the presence and adaptation of accumulators is probably limiting the efficiency of secretion at the population level. Accordingly, the secretion levels lose the linear relation with the internal levels of protein when this phenotype arises. In the worst scenario, the fraction of accumulators reached 50%, meaning that if they adapt and restore growth, only half of the population is actually contributing to the POI secretion levels.

We found a common regulatory pattern for these cells that differentiate them from the secretor phenotype. By looking at the dynamics of the accumulator behavior for all the data sets, it is observed that the rise of the accumulator phenotype is always present regardless the POI secreted, and the moment of the experiment. This phenotype seems to adapt and remain in the overall population only if they pass and maintain certain level of UPR signal before being diluted. Otherwise, they probably disappear because the low stress response is not enough to overcome the growth arrest.

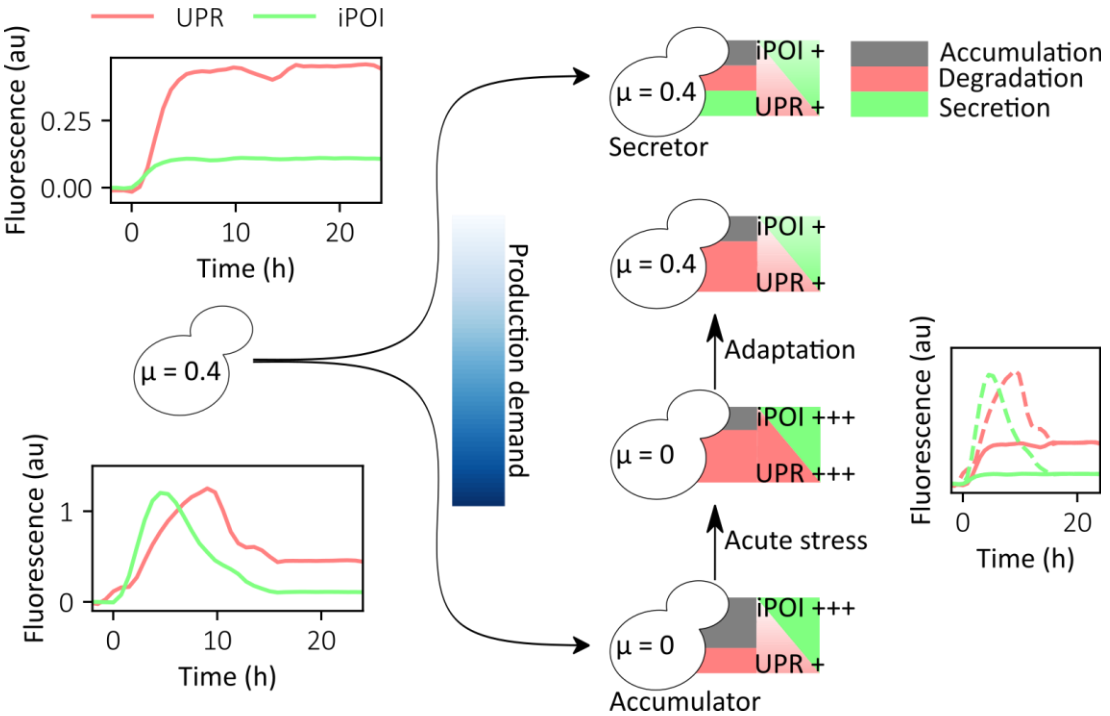

UPR
iPOI
Fluorescence (au)
0.25
0.00
0
10
20
Time (h)
μ = 0.4
Fluorescence (au)
1
0
0
10
20
Time (h)
Production demand
μ = 0.4
iPOI +
UPR +
Secretor
Accumulation
Degradation
Secretion
μ = 0.4
iPOI +
UPR +
Adaptation
μ = 0
iPOI +++
UPR +++
Acute stress
μ = 0
iPOI +++
UPR +
Accumulator
Fluorescence (au)
0
20
Time (h)

Figure 5.2. ***Rise and consolidation of the accumulator cells in the overall population.*** Depending on the strength of the production demand, a fraction of the population becomes an accumulator cell (lower arrow and plot). Accumulators are characterized by having high levels of internal levels of protein of interest (iPOI), and growth arrest (μ = 0). Due to that physiological state, the UPR is activated in order increase degradation and restore iPOI levels. Once the iPOI levels a decreased to reach proteostasis, the UPR, the iPOI and the growth rate, show similar values as the secretor cells (upper arrow and plot).

5.2.4. Exploiting the results of this study

Taking into account the results of the characterization pipeline and the observation of the accumulator subpopulation of cells, I proposed a feedback control strategy based on UPR levels. By controlling the induction levels, we can maintain the average UPR below a threshold, such that the accumulator cells do not adapt. Or alternatively, they are not numerous enough to slowdown dilution rate of the continuous culture.

As opposed to the typical strain optimization framework in which the goal is to engineer a specific stage of the secretory pathway. Here the goal is to match the production demand with the cell secretory capacities. This approach is in principle generic, since UPR is a common indicator for secretion-associated stress regardless the protein secreted. Thus, by just integrating the UPR reporter in the system, production could be optimized with this method in any strain.

5.2.5. General conclusion and take home message

It is known that secretion is a complex and highly regulated process in eukaryotic systems. As exposed in Chapter 2, bottlenecks arise along the secretory pathway at different stages depending on the specific requirements of each protein. Some studies also indicate that an optimal level of production demand is a common feature for the secretion of any heterologous protein (Wittrup, Robinson, Parekh, & Forrester, 1994; Parekh & Wittrup, 1997).

Here we provide tools to quantitatively characterize the secretion processes under different levels of production demand, and show through the secretory patterns representation the non-static nature of the secretory process. For the same given strain, there will be a different secretory pattern depending on the overall trafficking through the pathway (Figure 5.3). By extension of this rationale, every strain expressing a POI should show a specific secretory pattern as a function of the production demand and the nature of the POI.

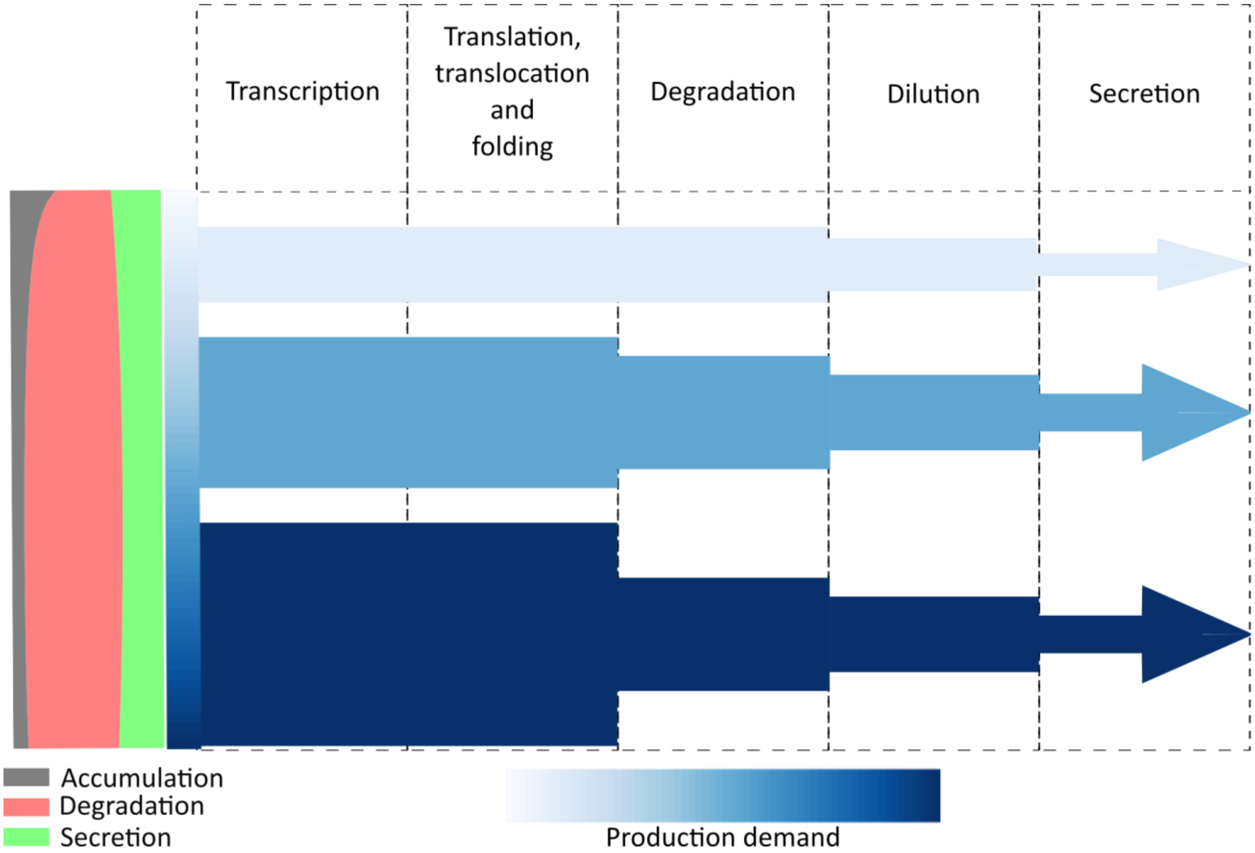


Figure 5.3. ***Trafficking in secretory pathway for different production demand levels.*** (Inspired from Figure 2.4 in Chapter 2). The relative contribution of the different components in the secretion process changes for different production demands. The bottlenecks arise in the pathway when the trafficking is excessive. To the left it is shown the secretory pattern for a given protein across all induction levels. For increasing induction levels, degradation adapts to keep internal levels of protein and proteostasis.

Therefore, the quantity of secreted protein obtained can be non-monotonic and even decreased with increasing demand. The thickness of an arrow is representative of the trafficking magnitude through a given stage of the pathway.

5.2.6. Limitations

There are many limitations in this project that I would like to approach if starting the PhD project again. Nevertheless, I still might have some time to approach them after the completion of this manuscript. It is evident that the characterization pipeline is dependent on a platform that can yield the data with enough temporal throughput to perform the analysis of the system's dynamics. In Chapter 4, the parameter inference relied in the dynamic data, as well as the detection of bimodal transient populations relied in single cell resolution capabilities. In contrast, flow cytometry does not provide temporal data for individual cells, which could be very interesting to study the dynamics of accumulator cells. Microfluidics is a possible alternative solution for that, however, microscopy data can be modest in terms of cell numbers.

A constrain of the bioreactors setup is the limited for OD measurements range (≅ 1.5 units of OD at 600nm). Here I pretend to generate knowledge with potential applications in industrial bioproduction, whereas experiments are done at an OD that is orders of magnitude lower than those used in actual industrial processes. This does not mean that the knowledge generated here is not applicable, but rather that it should be used as a contributing piece to the whole puzzle of industrial applications.

The optogenetic system has met our expectations: it showed low variability, high expression levels, tight induction control, and no harm to cell physiology. However, there are limitations if extrapolating to the context of industrial bioproduction, wherein optical densities are orders of magnitude higher than the ones used here. Working at such cell densities, implies that high light intensity is needed to reach the deepest volumes of the reactor. However, at very high light intensities, cells closer to the light source might die. Hence, the aim of using an optogenetic induction system in this project is not to simulate actual industrial setups, but having tight, fast, homogeneous and precise control of the external conditions. This helps to establish a solid knowledge that could eventually be incorporated to industrial applications, even if less flexible induction systems are used. Note that there have been recent developments in optogenetic systems that allow for slow-deactivation dynamics, that may be compatible with the industrial scope, since cells need to be less often under light illumination (Guerra *et al*, 2021).

Regarding reproducibility aspects, improvement of secretion measurements, and increasing the data sets would be important to demonstrate robustness on the results. For the secretion measurements, other alternative method has been published at the time the first version of this manuscript has been finished

(Boone *et al*, 2021). In such method, the authors use yeast surface display, where the POI is a fusion to the Sag1 protein on the external surface of the yeast. They use this approach to study the secretability of thousands of human proteins fragments fused to the external surface protein Sag1. Their approach to study secretion is really appropriated for library screening, and our is limited in this sense. However, I think the goal of each of the two methodologies is complementary, and its efficiency depends on the goal of the experiment. Probably our approach is more suitable for systems biology projects, where several variables are aimed to be measured simultaneously. For instance, internal protein level is an important variable in our research, that could not be measured with surface display.

As alternative, trying to find the best compromise between the two methods, other options could be considered for future projects in this line of research, as for instance using surface display in the sensor strain (Chapter 3) instead of beads. By expressing an antibody anti-FLAG in a yeast display fashion in the external surface of the sensor strain. Then use the sensor cells as beads, since the secreted protein would bind the external surface of the sensor cells. It would simplify the process, and increase the temporal throughput in the secretion levels measurements, since the measurement would be taken at every time point in the cytometer from the surface of the sensor cell. This approach was never used in our project because the lack of time and the challenge that it implies at the late stages of the project.

Regarding the feedback strategy, it must be tested to check whether it works. That is the next thing to do in my 'to do' list. In parallel I want to implement a model describing the behavior of the cells undergoing the accumulator phenotype. Such a model should include a feedback driven by the UPR. Modelling the UPR dynamics is something that I have already started with success, but the relation of degradation with the UPR dynamics to simulate the feedback will be more challenging.

5.3. Perspectives and future directions

In the sense of extending the contributions of this project, in the short term I would work mainly in overcoming some of the limitations mentioned above. Then, implementing a model that is able to reproduce and explain the behavior of the accumulators and their adaptation would be one of my first objectives. I think that key elements of the secretory efficiency can be uncovered by counting for populations. The development of an internal feedback (Annex 1) that can match production levels with cell capacity at the single cell level is also one of the projects that I would like to pursue in the midterm. The idea is to implement a feedback control established by an interference RNA system which expression is driven by the UPR activation. Such interference RNA is complementary to the mNeonGreen sequence present in all POI 3' region. To summarize, looking at the short/midterm directions, I would focus in obtaining a better understanding of the yeast secretory biology at the single cell and population level.

Regarding the development of technological platforms, I would like to adapt the multi-bioreactors setup enabling conditions closer to the industrial scope in terms of cell density. As said previously, it would affect the controllability of gene expression by optogenetic devices, but the use of optogenetic systems as the low-deactivation EL222 mentioned above would help to overcome such issue.

One interesting aspect to investigate in the scope of bioproduction in continuous cultures is the continuous purification process. I would like to propose immuno-beads usage to carry out the purification of POI from the culture media in a continuous manner. It would consist in using magnetic beads either bound to antibodies or other binding agents such as nickel ions. These beads would be introduced in the reactors by the culture media, thus constantly accessible to secreted protein with the purification tag. When the output flow carrying beads and cells leaves the reactors, it is temporarily collected in a magnetic rack, where the beads with the protein of interest are separated from the culture that is directed to the trash. Then, once the quantity of beads has reached a threshold in the magnetic rack, a protein purification process starts driven by a pipetting robot. The elution volume containing the protein would be constantly stored and the beads recycled. The main limitations of the process are the need of maintaining sterile conditions, the effect of beads in the optical density of the reactor, and the intensive use of the pipetting robot that is constantly purifying protein.

Then, in the very long term, once the pipeline for the characterization of the secretory process for various strains is standardized, and the bioreactor platform along with the automated purification system is set up. One can imagine its commercialization as a standard small-scale facility for protein production. Thus, making it possible to acquire an automated platform for the production of a wide range of goods of interest. The target market would be small local companies, or "garage" and community laboratories. In addition to providing the platform as the product itself, *a la carte* strains perfectly suited to this specific production platform can be offered. It would be done by using systematic characterization pipelines such as that presented here. That would provide the strain with a very detailed set of instructions, including a mathematical model, for the optimal production of the desired product in the platform, together with an efficient troubleshooting service.

The business model based on small-scale platform could cover the needs of small customers, for example, producing a molecule for a rare disease therapy that is not economically efficient for big pharma, or the production of specific compounds based on personalized medicine that have a very small and specific market. That model would also allow for continuous development by contributions of the community to extend the platform in a collaborative manner, democratize the use of biotechnologies, and certainly revolutionize several markets in a non-oligarchic structure. However, one of the main challenges to overcome, even more difficult than the technological development, would be the change in paradigm for the acceptance of genetically modified organisms. And the risks of having them out of contained environments and accessible for open and ethics-free development.

Chapter 6: Materials and methods

In this chapter I will present in detail the methods, procedures and protocols that I followed to build strains, perform experiments and analyze data. The objective is to provide all the necessary details to understand the data acquisition and treatment that was not discussed in the main text. I will also show the list of plasmids and strains used in the main text of the thesis. In some specific cases I will refer to sections of other chapter of the thesis, especially Chapter 3. Some of the details for the data acquisition are found in Bertaux et al., 2021, and I will refer to it when necessary.

6.1. Strains construction

6.1.1. DNA synthesis and constructions design

Some of the genetic construction used in this project were synthesized by external suppliers. The design of the parts was carried out to make the parts compatible with the Modular Cloning strategy that was used for generating the plasmids collection. Such collection started with the one provided in the MoClo Yeast Tool Kit (YTK), and from that was extended (Lee *et al*, 2015). The DNA synthesis was purchased from Twist Biosciences and GENEWIZ, depending on the specific needs, such as number of repeated sequence, length of the fragment, turnaround time.

Table 6.1. ***Synthesized DNA parts obtained from external suppliers***

Protein of interest	**Native organism**	***GenBank* ID**
mNeonGreen[1]	N/A	BBB44438
mCerulean[5]	N/A	AJN91096
Endo-1,4-beta-xylanase C[2]	*Aspergillus niger*	EU848304
Single chain variable fragment 4M5.3[3]	N/A	1X9Q_A
α-amylase[4]	*Aspergillus oryzae*	CAA31220
5xC120-CYC180pr-Kozak (pLight, Chapter 3, 3.1.2)[6]	N/A	N/A
4x UPRE-CYC1 (pStress, Chapter 3, 3.1.6)[7]	N/A	N/A

[1](Shaner *et al*, 2013), [2](Boder *et al*, 2000), [3](Do *et al*, 2013), [4](Liu *et al*, 2012), [5] (Rizzo & Piston, 2005), [6] (Benzinger & Khammash, 2018), [7] (Pincus *et al*, 2010)

The sequences synthesized corresponded in all cases to coding sequences or promoters. More sequences in the context of this project were synthesized, and even used in strains and characterized. However, the results are not included in the thesis, because the final characterization was not completed as for those in Chapter 4, and the results were out of context. First the sequence of the proteins is obtained from literature and verified in *GeneBank* (Table 6.1). The coding sequences were codon optimized for *S.* cerevisiae by using the IDT codon optimizer tool. The subsequent changes in the sequence were done by using the software *SnapGene*. Once optimized, the recognition target sequence for the restriction enzymes *BsmBI*, *BsaI*, and *NotI* was avoided by changing manually, not changing codon code in the case of CDS, but using the second more frequent. The synthesis of genes also took into account the overhangs compatible for a level 0 part in the cloning strategy (Figure 6.1), such that the received plasmid from the vendor can be either directly used as a level 0 part, that is as a part in the YTK context, or cloned into the entry vector to become a level 0 part with the YTK backbone.

6.1.2. Cloning parts and generating strains

For cloning I used, and extended with new parts, an already characterized Yeast Toolkit for modular cloning procedure (MoClo) (Lee *et al*, 2015). It is a library of parts based on Golden Gate assembly and MoClo (Carola Engler, Romy Kandzia, 2008). Golden Gate assembly in based on the use of Type IIs restriction enzymes, they cut outside of the target recognition sequence, providing *a la carte* cohesive ends, which incredibly increases the flexibility of the cloning and allows multiplexing insertion in one-shot reaction. Modular cloning is based on step-cloning process (Figure 6.1), it allows interchanging different parts as desired by designating specific overhangs to each part type (Weber *et al*, 2011). This library, as most of restriction-based cloning methods, has some specific restriction enzymes used to perform the cloning and the following chromosomal integration.

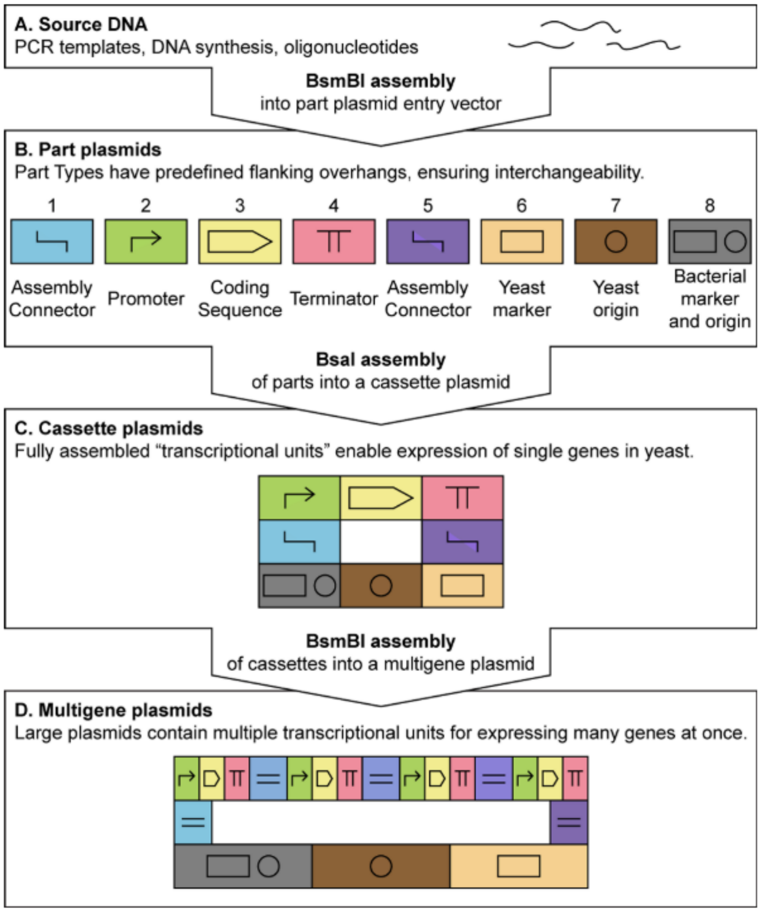

A. Source DNA
PCR templates, DNA synthesis, oligonucleotides
BsmBI assembly
into part plasmid entry vector
B. Part plasmids
Part Types have predefined flanking overhangs, ensuring interchangeability.
1
2
3
4
5
6
7
8
Assembly Connector
Promoter
Coding Sequence
Terminator
Assembly Connector
Yeast marker
Yeast origin
Bacterial marker and origin
BsaI assembly
of parts into a cassette plasmid
C. Cassette plasmids
Fully assembled "transcriptional units" enable expression of single genes in yeast.
BsmBI assembly
of cassettes into a multigene plasmid
D. Multigene plasmids
Large plasmids contain multiple transcriptional units for expressing many genes at once.

Figure 6.1. ***Cloning strategy based on Golden Gate and Modular Cloning.*** From Lee et al., 2015. Schematic description of the cloning strategy used in this project.

The cloning reaction is performed by adding the 20fmol (if possible in 1 µL) of each plasmid to the reaction volume described in the table below.

Reactant	**Volume/sample (µL)**
Water	Up to 10 µL
T4 ligase buffer	1
Restriction enzyme	0.8
Bovine serum Albumin (100 µg/ml)	0.8
T4 ligase	0.4
Total	10

Finally, the reaction takes place in a cycle of restriction and ligation steps that lasts around 6 hours. Its efficacy is close to 100% of positive clones, though decreasing with increasing number of parts to clone.

The reaction mixture was introduced in the thermocycler with the following set up:

1. 37°C for 15 minutes
2. 37°C for 2 minutes
3. 16°C for 5 minutes
4. Back to step 2 x50 times
5. 37°C for 15 minutes
6. 50°C for 5 minutes
7. 80°C for 5 minutes

The reaction mixture is directly cloned in CaCl2 competent to *E.coli DH5-alpha*. The procedure to make competent cells is the following:

1. The night before, inoculate a 5 ml culture and grow overnight with selection.
2. Dilute cells ~ 1:200 into selective media
3. Grow the cells to an OD600 of 0.5 – 0.6 (This takes about 3 hours)
4. Spin down the cells at 4 ºC, 4000 rpm, 15 minutes.

Note: Keep the cells at 4 ºC from now on.

5. Resuspend cells in 15 ml, ice-cold 100 mM $CaCl_2$
6. Leave on ice 4 hours to overnight
7. Spin down the cells at 4 ºC, 4000 rpm, 15 minutes.
8. Resuspend cells in 4 ml, ice-cold 100 mM $CaCl_2$ + 15% glycerol (final concetration).

9. Aliquot 20µL of cells into pre-chilled Eppendorf tubes. Use immediately or store at -80ºC.

Note: Frozen cells are only good once.

The reaction mixture is directly cloned in CaCl2 competent to *E.coli DH5-alpha*. Following the heat-shock transformation procedure:

1. Thaw CaCl2 competent cells on ice
2. Place 20 µL of cells in a pre-chilled Eppendorf tube
3. Add 3µL of golden gate product
4. Mix gently by flicking the tube
5. Chill on ice for 10 minutes
6. Heat shock at 42 °C for 30 seconds
7. Return to ice for 2 minutes
8. Add 200µl LB medium and recover 60 minutes the cells by shaking at 37 °C
9. Plate out the cells on selective LB, one plate with 20µL, the other one with the rest of volume, up to 200µL
10. Incubate at 37 °C overnight

Then, the positive clones are selected by antibiotic markers in selective media, and in some cases also fluorescence selection. They are checked by colony PCR with low fidelity polymerase reaction (GreenTaq from ThermoFisher), directly using the colonies. Those positive clones are grown over night in LB at 37 °C during 24 hours. Then a miniprep is done to obtain and sequence the plasmid, note that for sequencing, all the target sequence is read from two different primers. If the sequence is correct, the strains are stored at -80C in 25% glycerol.

In many occasions I generate new parts from others already existing. For instance, the mNeonGreen protein used as reporter, fused or not to the POI, was used as part 3, 3a, 3b and 4b. All came from one template mNeonGreen coding region. The interchange between parts was just simply doing high fidelity PCR (PhusionTaq from Thermofisher) from the template par with the right primers for the transformation into the new part.

Using this methodology I have produced around 200 plasmids, counting for levels 0, 1, and 2. All of them have been constructed in the context of the Yeast Tool Kit library, thus, the extension for such library has been significant. And the integrative plasmids used for publications will be available in the public nonprofit repository *Addgene* data base when the manuscript is available in *bioRxive*. In the following list I will show the level 2, so integrative plasmids used in the results presented in this thesis.

Table 6.2. ***Level 2 plasmids used in the experiments presented in this thesis***

Plasmid	Locus	Level	Database name
Integrative_EL222_pTDH3	URA3	2	pIB0120
Integrative_3x(pC120_sec-mCerulean_FLAG)	LEU2	2	pIB0381
Integrative_3x(pC120_sec-mNeonGreen_FLAG)	LEU2	2	pIB0369
Integrative_pTDH3_mCerulean	LEU2	2	pIB0273
Integrative_pTDH3_mNeonGreen	LEU2	2	pIB0274
Integrative_pTDH3_mScarlet	LEU2	2	pIB0035
Integrative_pC120_mScarlet	LEU2	2	pIB0113
Integrative_UPR_reporter	LEU2	2	pIB0115
Integrative_1x(pC120_mNeonGreen-FLAG)	HO	2	pIB0682
Integrative_1x(pC120_sec-mNeonGreen_FLAG)	HO	2	pIB0399
Integrative_1x(pC120_sec-4_4_20_scFv-mNeonGreen_FLAG)	HO	2	pIB0401
Integrative_1x(pC120_sec-alpha_amylase-mNeonGreen_FLAG)	HO	2	pIB0677
Integrative_1x(pC120_sec-xylanase_C-mNeonGreen_FLAG)	HO	2	pIB0679

6.1.3. Genomic integration and yeast strains construction

In all the strains shown in this thesis, the expression cassettes have been integrated in the chromosomes of the yeast using homologous recombination, in which the construct is integrated at a specific locus. The transformation is based on an adapted version of the lithium acetate/single-stranded carrier DNA/PEG method of transformation that I modified from Gietz, 2014.

To do so I first digested the level 2 plasmids to generate linear DNA. It is recommended to transform 34µL with 1-4 µg of DNA. I digest with *NotI* 40µL per reaction. The reaction was performed during 2 hours at 37°C, followed by 5 minutes at 80°C to inactivate the enzyme.

The night before I grew the yeast strains that I want to transform in 5mL of YPD at 30 °C. From the overnight culture, I dilute 100 times into 50mL of YPD fresh media and let it grow for 4 hours at 30 °C, shaking. (The OD600 should be between 0.3-0.4). Then I follow this procedure:

1. Harvest the cells by pelleting at 3000 rpm in a centrifuge for 10 minutes at room temperature and discard the supernatant
2. In parallel, place a thawed tube of carrier DNA reagent in a boiling water bath for 5 minutes and chill immediately in ice.
3. Resuspend the cells in 10 mL of pure water
4. Harvest the cells by pelleting at 3000rpm in a centrifuge for 10 minutes at room temperature and discard the supernatant
5. Resuspend the cells in 1 mL of 100mM LiAc and transfer to a 1.5 mL tube
6. Pellet the cells at 3000 rpm for 10 minute, remove the supernatant, and resuspend the pellet in 400 µL of 100 mM LiAc
7. Aliquot 50 µL of cell suspension into 1.5 mL tubes (One per transformation)
8. Pellet the cells at 3000 rpm for 1 minute and remove the supernatant
9. Add to the pellet 240 µL PEG 3350 (50% w/v), 36 µL LiAc (1.0M), 50µL single-stranded carrier DNA (2.0 mg/mL) and 34 µL of sterile water plus any plasmid DNA (up to 1 µg if it is episomal plasmid, and 1-4µg if it is linear DNA for integration), in the order given, and mix the pellet by vortex mixing briskly until resuspended, no clumps

Usually I transform directly the restriction product.

10. Incubate for 30 minutes at 30 °C
11. Incubate the tube in a water bath at 42 °C for at least one or two hours. Temperature-sensitive strains can be left on the bench overnight, then carried on to the next step
12. Centrifuge the transformation tube at 3000rpm for 1 minute at room temperature and remove the transformation mix with a pipette
13. Resuspend by pipetting gently in 1 mL of YPD and incubate two hours at 30°C shaking.
14. Harvest the cells by centrifuging at 3000 rpm 10 minutes
15. Resuspend the cells in 200 µL of sterile water by pipetting gently
16. Plate the cells suspension on selectable plates (20 µL in one plate, 180 µL in the other)

Then I select the positives based on prototrophic selection. I check the positives by colony PCR for yeast, that is a bit subtler than for bacteria. Basically, the lysis is done with NaOH during 20 minutes at 90C, and the maximal recommended length for the amplicon is of around 1kb.

Once I have selected positive clones by colony PCR, I perform genome extraction using a commercial kit, MasterPure™ Yeast DNA Purification Kit. To the extracted genome and I make a high fidelity PCR, usually Phusion PCR from Thermofisher. Then, I purify the amplicons by using a commercial kit and send them to sequence, note that for sequencing, all the target sequence is read from two different primers. If the sequence is correct, the colony selected is grown overnight at 30 °C in YPD and then are stored at -80 °C in 25% glycerol.

Using this methodology I have produced around 100 yeast strains. However, I did not perform 100 genome extractions, I only sequenced final strains that were used in experiments. In the following list I will show the ones used in the results presented in the main text this thesis. All these strains share the parental BY4741, it is a deletion strain derived from the widely-used laboratory strain S288C (Mortimer & Johnston, 1986). More detail about the use of this yeast strain can be found in Chapter 3, section 3.3.

Table 6.3. ***Yeast strains used in the experiments presented in this thesis***

Strain	Genome	Database name
BY4741	MATa his3Δ1 met15Δ0 ura3Δ0 leu2Δ0	IB20019
UPR reporter	IB20019 + Leu2Δ0::UPRE-mScarlet-I-tENO1-LEU2	IB20031
Constitutive expression of EL222	IB20019 + ura3Δ0::pTDH3-V40 NLS-VP16AD-EL222-tSSA1-URA3	IB20032
UPR reporter & EL222 (Parental case study strains)	IB20032 + leu2Δ0::UPREx4-pCYC-mScarlet-I-tENO1-LEU2	IB20090
mCerulean batch	IB20090 + HO::3x(pC120- alpha-pre-pro-mCerulean-3xFLAG-tTDH1)-HIS3	IB20162
mNeonGreen batch	IB20090 + HO::3x(pC120- alpha-pre-pro-mNeonGreen-3xFLAG-tTDH1)-HIS3	IB20148
pTDH3 reference	IB20032 + leu2Δ0::pTDH3-mScarlet-I-tTDH1-LEU2	IB20043
EL222 characterization	IB20032 + leu2Δ0::pC120-mScarlet-I-tENO1-LEU2	IB20044
Sensor strain	IB20032 + leu2Δ0::pC120-mScarlet-I-tENO1-LEU2 + HO::2x(pTDH3-mCerulean-tTDH1) HIS3	IB20337
Non-secreted mNeonGreen	IB20090 + HO::pC120-mNeonGreen-3xFLAG-tTDH1-HIS3	IB20330
Secreted mNeonGreen	IB20090 + HO::pC120- alpha-pre-pro-mNeonGreen-3xFLAG-tTDH1-HIS3	IB20169
XylC	IB20090 + HO::pC120- alpha-pre-pro-xylanase_C-mNeonGreen-3xFLAG-tTDH1-HIS3	IB20315
scFv	IB20090 + HO::pC120- alpha-pre-pro-4_4_20_scFv-mNeonGreen-3xFLAG-tTDH1-HIS3	IB20171
Amylase	IB20090 + HO::pC120- alpha-pre-pro-alpha_amylase-mNeonGreen-3xFLAG-tTDH1-HIS3	IB20313

6.2. Characterization experiments

6.2.1. Bioreactor experiments

For bioreactors experiments, the yeast strains are first revived by plating in a agar + YPD plate and incubating at 30 °C during 48 hours. Note that from now, all manipulation must be protected from direct light if working with strains with optogenetic controlled expression. Then, a single colony is picked and gown in a preculture of 5 mL of YPD overnight at 30 °C. The next day, the preculture is repeated by inoculating 500 µL of the overnight YPD culture in 25 mL of synthetic complete medium (ForMedium LoFlo yeast nitrogen base CYN6510 and Formedium complete supplement mixture DCS0019) 2% glucose (w/v), at 30 °C. After 6 hours the culture is inoculated in the reactors by adding 2 mL approximately to each of the eight vessels by using a syringe.

All experiments were performed in 20 mL culture volume bioreactors at 30 °C and in turbidostat mode (OD 0.5, typically corresponding to 107 cells/mL according to cytometry data) with synthetic complete medium (ForMedium LoFlo yeast nitrogen base CYN6510 and Formedium complete supplement mixture DCS0019) 2% glucose (w/v) and 0.5mL of L-arginine. As said in Chapter 3, section 3.1.5, arginine is used to maintain pH 7, thus keeping secreted protein integrity along the experiment duration to allow secretion measurements.

After an 8 hours of the incubation, the automated cytometry measurements start and are programed to take samples every 45 minutes. Approximately after 3-4 hours of measurements in dark conditions, the LEDs of the platform are switched on, at intensity 20, and different duty cycles in each reactor to induce protein expression. The induction lasts 24 hours, then, the experiment is over and the secretion measurements are taken. Sometimes the secretion measurement is also taken after 6 hours induction. In either case the pH is always measured when taken a sample to check that it is at pH 7.

6.2.2. Cytometry acquisition

The cytometer (Guava EasyCyte 14HT, 362 Luminex) settings and reproducibility studies, as well as the robot calibration were done once for all in parallel of this project and are specified in (Bertaux *et al*, 2021b). For the cytometry data when measuring cells, I don’t use deconvolution or automatic gating. For the gating I use my own pipeline of code, implemented in *Python 3*.

6.2.3. Analysis of data from bioreactors and model fitting

For gating and be sure of what I work with are cells and not doublets (not common in my experiments) or cell debris. I gate based on the forward and side scatter. For the forward scatter I just consider events between 600 and 1500 (au), whereas for the side scatter the limits are 1200 and 3000. Then, the signal from the channels for mNeonGreen (GRN-B), mCerulean (GRN-V), and mScarlet-I (ORG-G), is normalized for each event by its forward scatter. That can be interpreted as a way to normalize by the volume of the cell. In some cases, it is discussed if using forward or side scatter. Nevertheless, here I use the forward scatter as a representative of the cell volume. The decision was motivated by empirical experience while working with beads, when I observed a correlation between forward scatter and fluorescence of the bead.

The second gating is done based on the GRN-V channel to separate the sensor strain from the case study strain. In this case, as mentioned in Chapter 3, section 3.13, the events with blue fluorescence (GRN-V/forward scatter) above 0.2 (au) are consider sensor cells. Whereas those with blue fluorescence below 0.07 (au) are considered case study strain. After the separation of the two strains, the levels of internal fluorescence are measure independently for each type through the red fluorescence for the sensor strain, and the green fluorescence for the case study strain. In the case of the second the red fluorescence accounts for the UPR reporter signal. Note that for green and red fluorescence, the autofluorescence is subtracted. It is done by computing autofluorescence as the average fluorescence for each color independently during the 2 hours previous to induction in each reactor respectively.

The growth rates are also computed in the same pipeline as the gating as explained in Chapter 3, section 3.1.4. All the analysis for data parsing is done in using *Python 3* with code implemented by me. The packages used are *pandas* (McKinney, 2010)*, numpy* (Harris *et al*, 2020)*,* and *SciPy* (Virtanen *et al*, 2020). The data representations are done with *matplotlib* library, and the linear and polynomial fitting, as well as the data smoothing is done by using the packages *numpy, sklearn,* and *SciPy*. The model fitting has been done using the CMA-ES algorithm by using the *pycma* package from Hansen et al., 2021. The model ODEs integration has been carried out by the numerical solver *scipy.integrate.solve_ivp*. More specific analysis as the Gaussian Mixture Model to identify bimodal populations used in section 4.2 has been implemented by using the package *sklearn.mixture.GaussianMixture* (Pedregosa FABIANPEDREGOSA *et al*, 2011). All analysis pipelines are done using Jupyter notebook interface. All scripts are notebooks are available upon request. The pipeline used for raw data managing previous to analysis is included in Annex B.

6.2.4. Secretion levels measurements

The secretion measurements using immuno-beads have been thoroughly described in the Chapter 3, section 3.2. Here I am going to mainly describe the protocol and the details omitted in the mentioned reference to the main text.

For the secretion levels measurements, I used the Anti-FLAG® M2 Magnetic Beads from Sigma (M8823), the bead size is ranged in 20-75 µm, which allows us to measure them in the cytometer. These are 4% agarose micro-beads bound to the Anti-FLAG M2 monoclonal antibody. Such antibody recognizes the FLAG sequence at the C-terminus of the mNeonGreen, fused or not to the POI. Regarding the details of the protocol, it is described in the following lines:

Samples taken directly from the reactors cultures, 200 µL of it are mixed with 26 µL of phosphate buffer 10x, and 26 µL of stock dilution of the supernatant from a batch culture of cells producing mCerulean, that is used as the reference protein for ratiometric measurements. Such batch culture is previously done and last for many experiment, stored at – 80 °C.

Three reference samples are done following the same procedure as for bioreactors samples. The first reference contains only 200 µL of supernatant form batch culture of cells producing mNeonGreen. This is used as the highest signal reference, can be also seen as a positive control for the method. The second contains 200 µL of water instead of sample, and serves as a blank value, since it only contains the 26 µL of stock dilution of the supernatant from a batch culture of cells producing mCerulean, and the same volume of buffer. The third reference does not contain fluorescent protein at all, just 226µL of water counting for mCerulean and sample. This reference accounts only for the autofluorescence of the beads.

Note that all the manipulations with samples containing cells should be done in the dark. After incubation during 1 hour at room temperature, the samples are washed before measurements by removing the supernatant and resuspending in 200 µL of TBS buffer 1x, then repeat 3 times.

Regarding the beads preparation, the protocol needs 0.4 µL of packed gel volume of beads in a reaction, for instance, for 10 samples (volume of 252 µL each before the beads) it is 4 µL of packed gel volume. It is very important to add the beads the last, because otherwise the incubation time for the POI and the reference reporter mCerulean won't be similar. Therefore, the ratiometric measurements will not work well. After the beads equilibration (Described below), collect the beads using the magnetic rack and resuspend them in 120 µL of water. Then, once the preparation of the samples is complete 10 µL are added to each sample.

Protocol for beads equilibration:

1. Thoroughly resuspend the resin by gentle inversion. Make sure the bottle of ANTI-FLAG M2 Magnetic Beads is a uniform suspension. Remove an appropriate volume for use.
2. Transfer resin to an appropriate sized tube. Equilibrate beads by resuspending with 5 packed gel volumes of TBS. Mix thoroughly. Using the magnetic rack collect the beads. Remove and discard the storage buffer/TBS mixture.
3. Equilibrate beads by resuspending with another 5 packed gel volumes of TBS. Mix thoroughly.
4. Using the magnetic rack, the beads. Remove and discard the storage buffer/TBS mixture.

5. Repeat steps 3 and 4 once. Allow a small amount of buffer to remain on the top of the beads.

For the data treatment of beads, as explained in Chapter 3, section 3.2. The measured samples of 200µL of washed beads follow the ratiometric measurements approach. Per-cell deconvolution between overlapping fluorophores is performed for mNeonGreen and mCerulean fluorescence, using as autofluorescence the reference sample three, containing only beads. The details of the deconvolution algorithm can be found in Bertaux et al., 2021.

The beads can be normalized by the forward scatter for each event, however, for ratio metric measurements that is not significant, since the ration discards the bead size effect. However, the forward and side scatter properties are used to gate applying the following gating conditions: (i) To discard cells, the minimal side scatter to consider events is restricted to 10^4 arbitrary units (au). To avoid beads with low binding properties, (ii) the minimal forward scatter is set to 500 (au), and to differentiate beads from cells (ii) the ratio side/forward scatter must be above ten. See Figure 3.13, in Chapter.

Then, the fluorescence of each event obtained by the channels GRN-V and GRN-B, after deconvolution, are used to make the ratio mNeonGreen/mCerulean. Blank fluorescence is subtracted using beads from the second reference sample only binding mCerulean subtracted. Finally, the resulting number is divided by the ratiometric measurement from the first reference sample, the one containing the supernatant of mNeonGreen, discard day-to-day variability effect. Once this is done the secretion levels are normalized to the fraction of strain of interest present in the population with respect to the sensor strain, and to a non-diluted reference sample from a batch supernatant measured in each replica.

Chapter 7: References

Aditya C, Bertaux F, Batt G & Ruess J (2021) A light tunable differentiation system for the creation and control of consortia in yeast. *Nat Commun 2021 121* 12: 1–10

Babour A, Bicknell AA, Tourtellotte J & Niwa M (2010) A Surveillance Pathway Monitors the Fitness of the Endoplasmic Reticulum to Control Its Inheritance. *Cell* 142: 256–269

Baeshen NA, Baeshen MN, Sheikh A, Bora RS, Ahmed MMM, Ramadan HAI, Saini KS & Redwan EM (2014) Cell factories for insulin production. *Microb Cell Fact* 13

Balleza E, Kim JM & Cluzel P (2018) Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* 15: 47–51

Beal J, Goñi-Moreno A, Myers C, Hecht A, Vicente M del C, Parco M, Schmidt M, Timmis K, Baldwin G, Friedrichs S, *et al* (2020) The long journey towards standards for engineering biosystems. *EMBO Rep* 21: 1–5

Benzinger D & Khammash M (2018) Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat Commun*: 1–38

Bertaux F, Sosa-Carrillo S, Fraisse A, Aditya C, Furstenheim M & Batt G (2021a) Enhancing bioreactor arrays for automated measurements and reactive control with ReacSight. *bioRxiv*: 2020.12.27.424467

Bertaux F, Sosa-Carrillo S, Fraisse A, Aditya C, Furstenheim M & Batt G (2021b) Enhancing bioreactor arrays for automated measurements and reactive control with ReacSight. *bioRxiv*: 2020.12.27.424467

Besada-Lombana PB & Da Silva NA (2019) Engineering the early secretory pathway for increased protein secretion in Saccharomyces cerevisiae. *Metab Eng* 55: 142–151

Bindels DS, Haarbosch L, Van Weeren L, Postma M, Wiese KE, Mastop M, Aumonier S, Gotthard G, Royant A, Hink MA, *et al* (2016) MScarlet: A bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* 14: 53–56

Boder ET, Midelfort KS & Wittrup KD (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci U S A* 97: 10701–10705

Boder ET & Wittrup KD (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol 1997 156* 15: 553–557

Bond RAB, Martincigh BS, Mika JR & Simoyi RH (1998) The Quasi-Steady-State Approximation: Numerical Validation. *J Chem Educ •* 75: 56

Boone M, Ramasamy P, Zuallaert J, Bouwmeester R, Van Moer B, Maddelein D, Turan D, Hulstaert N, Eeckhaut H, Vandermarliere E, *et al* (2021) Massively parallel interrogation of protein fragment secretability using SECRiFY reveals features influencing secretory system transit. *Nat Commun 2021 121* 12: 1–16

Borodina I & Nielsen J (2014) Advances in metabolic engineering of yeast Saccharomyces cerevisiae for production of chemicals. *Biotechnol J* 9: 609–620

Brown AI & Koslover EF (2021) Design principles for the glycoprotein quality control pathway. *PLoS Comput Biol* 17: 1–24

Buchberger A, Bukau B & Sommer T (2010) Protein Quality Control in the Cytosol and the Endoplasmic Reticulum: Brothers in Arms. *Mol Cell* 40: 238–252

Bull AT (2010) The renaissance of continuous culture in the post-genomics age. 993–1021

Camarasa C, Chiron H, Daboussi F, Valle G Della, Dumas C, Farines V, Floury J, Gagnaire V, Gorret N, Léonil J, *et al* (2018) INRA's research in industrial biotechnology: For food, chemicals, materials and fuels. *Innov Food Sci Emerg Technol* 46: 140–152

Campbell K, Xia J & Nielsen J (2017) The Impact of Systems Biology on Bioprocessing. *Trends Biotechnol* 35: 1156–1168

Carola Engler, Romy Kandzia SM (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3

Carvalho N, Chaim O, Cazarini E & Gerolamo M (2018) Manufacturing in the fourth industrial revolution: A positive prospect in Sustainable Manufacturing. *Procedia Manuf* 21: 671–678

Ceroni F, Algar R, Stan GB & Ellis T (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat Methods* 12: 415–418

Ceroni F, Boo A, Furini S, Gorochowski TE, Borkowski O, Ladak YN, Awan AR, Gilbert C, Stan G-B & Ellis T (2018) Burden-driven feedback control of gene expression. *Nat Methods 2018 155* 15: 387–393

Chi H, Wang X, Shao Y, Qin Y, Deng Z, Wang L & Chen S (2019) Engineering and modification of microbial chassis for systems and synthetic biology. *Synth Syst Biotechnol* 4: 25–33

Chung BH & Park KS (1998) Simple Approach to Reducing Proteolysis During Secretary Production of Human Parathyroid Hormone in Saccharomyces cerevisiae.

Crook N, Sun J, Morse N, Schmitz A & Alper HS (2016) Identification of gene knockdown targets conferring enhanced isobutanol and 1-butanol tolerance to Saccharomyces cerevisiae using a tunable RNAi screening approach. *Appl Microbiol Biotechnol* 100: 10005–10018

Crook NC, Schmitz AC & Alper HS (2014) Optimization of a yeast RNA interference system for controlling gene expression and enabling rapid metabolic engineering. *ACS Synth Biol* 3: 307–313

Croughan MS, Konstantinov KB & Cooney C (2015) The future of industrial bioprocessing: Batch or continuous? *Biotechnol Bioeng* 112: 648–651

Darby RAJ, Cartwright SP, Dilworth M V & Bill RM Chapter 2 Which Yeast Species Shall I Choose? Saccharomyces cerevisiae Versus Pichia pastoris (Review). *Methods Mol Biol* 866

Delic M, Göngrich R, Mattanovich D & Gasser B (2014) Engineering of protein folding and secretion - Strategies to overcome bottlenecks for efficient production of recombinant proteins. *Antioxidants Redox Signal* 21: 414–437

Do TT, Quyen DT, Nguyen TN & Nguyen VT (2013) Molecular characterization of a glycosyl hydrolase family 10 xylanase from Aspergillus niger. *Protein Expr Purif* 92: 196–202

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR & Bartel DP (2009) RNAi in budding yeast. TL - 326. *Science* 326 VN-: 544–550

Dymond J & Boeke J (2012) The Saccharomyces cerevisiae SCRaMbLE system and genome minimization. *Bioeng Bugs* 3: 168

Eguchi Y, Makanae K, Hasunuma T, Ishibashi Y, Kito K & Moriya H (2018) Estimating the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins. *Elife* 7: 1–23

Escalante-Chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ & Springer M (2015) Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proc Natl Acad Sci U S A* 112: 1636–1641

Figueroa D, Rojas V, Romero A, Larrondo LF & Salinas F (2021) The rise and shine of yeast optogenetics. *Yeast* 38: 131–146

Fox ZR, Fletcher S, Fraisse A, Aditya C & Sosa S (2021) MicroMator: Open and Flexible Software for Reactive Microscopy. *bioRxiv*: 1–9

García R, Sanz AB, Rodríguez-Peña JM, Nombela C & Arroyo J (2016) Rlm1 mediates positive autoregulatory transcriptional feedback that is essential for Slt2-dependent gene expression. *J Cell Sci* 129: 1649–1660

Geisberg J V., Moqtaderi Z, Fan X, Ozsolak F & Struhl K (2014) Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. *Cell* 156: 812–824

Gietz RD (2014) Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol Biol* 1205: 1–12

Glick BR (1995) Metabolic load and heterologous gene expression. *Biotechnol Adv* 13: 247–261

Görgens JF, Van Zyl WH, Knoetze JH & Hahn-Hägerdal B (2005) Amino acid supplementation improves heterologous protein production by Saccharomyces cerevisiae in defined medium. *Appl Microbiol Biotechnol* 67: 684–691

Guerra P;, Guerra P, Vuillemenot L-A, Rae B, Ladyhina V & Milias-Argeitis A (2021) Systematic in vivo characterization of fluorescent protein maturation in budding yeast.

Halbleib K, Pesek K, Covino R, Hofbauer HF, Wunnicke D, Hänelt I, Hummer G & Ernst R (2017) Activation of the Unfolded Protein Response by Lipid Bilayer Stress. *Mol Cell* 67: 673-684.e8

Hansen N & Ostermeier A (2001) Completely derandomized self-adaptation in evolution strategies. *Evol Comput* 9: 159–195

Hansen N, yoshihikoueno, ARF1, Nozawa K, Chan M, Akimoto Y & Brockhoff D (2021) CMA-ES/pycma: r3.1.0.

Harmsen MM, Bruyne MI, Raué HA & Maat J (1996) Overexpression of binding protein and disruption of the PMR1 gene synergistically stimulate secretion of bovine prochymosin but not plant Thaumatin in yeast. *Appl Microbiol Biotechnol* 46: 365–370

Harrigan P, Madhani HD & El-Samad H (2018) Real-Time Genetic Compensation Defines the Dynamic Demands of Feedback Control. *Cell* 175: 877-886.e10

Harris CR, Jarrod Millman K, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, *et al* (2020) Array programming with NumPy. *Nature* 585: 357

Hochrein L, Machens F, Messerschmidt K & Mueller-Roeber B (2017) PhiReX: A programmable and red light-regulated protein expression switch for yeast. *Nucleic Acids Res* 45: 9193–9205

Hou J, Tyo KEJ, Liu Z, Petranovic D & Nielsen J (2012) Metabolic engineering of recombinant protein secretion by Saccharomyces cerevisiae. *FEMS Yeast Res* 12: 491–510

Huang D, Gore PR & Shusta E V. (2008a) Increasing yeast secretion of heterologous proteins by regulating expression rates and post-secretory loss. *Biotechnol Bioeng* 101: 1264–1275

Huang D, Gore PR & Shusta E V. (2008b) Increasing yeast secretion of heterologous proteins by regulating expression rates and post-secretory loss. *Biotechnol Bioeng* 101: 1264–1275

Huang D & Shusta E V. (2005) Secretion and Surface Display of Green Fluorescent Protein Using the Yeast Saccharomyces cerevisiae. *Biotechnol Prog* 21: 349–357

Huang M, Bao J, Hallström BM, Petranovic D & Nielsen J (2017) Efficient protein production by yeast requires global tuning of metabolism. *Nat Commun* 8

Huang M, Wang G, Qin J, Petranovic D & Nielsen J (2018) Engineering the protein secretory pathway of Saccharomyces cerevisiae enables improved protein production. *Proc Natl Acad Sci U S A* 115: E11025–E11032

Ideker T, Thorsson V, Ranish J, Christmas R, Buhler J, Eng J, Bumgarner R, Goodlett D, Aebersold R & Hood L (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science (80-)* 292: 929

Johnson N, Powis K & High S (2013) Post-translational translocation into the endoplasmic reticulum. *Biochim Biophys Acta - Mol Cell Res* 1833: 2403–2409

Jonas FRH, Royle KE, Aw R, Stan GB V. & Polizzi KM (2018) Investigating the consequences of asymmetric endoplasmic reticulum inheritance in Saccharomyces cerevisiae under stress using a combination of single cell measurements and mathematical modelling. *Synth Syst Biotechnol* 3: 64–75

Jordà J, Rojas HC, Carnicer M, Wahl A, Ferrer P & Albiol J (2014) Quantitative Metabolomics and Instationary 13C-Metabolic Flux Analysis Reveals Impact of Recombinant Protein Production on Trehalose and Energy Metabolism in Pichia pastoris. *Metab 2014, Vol 4, Pages 281-299* 4: 281–299

Jouhten P, Boruta T, Andrejev S, Pereira F, Rocha I, Kiran & & Patil R (2016) Yeast metabolic chassis designs for diverse biotechnological products OPEN. *Nat Publ Gr*

Kafri M, Metzl-Raz E, Jona G & Barkai N (2016) The Cost of Protein Production. *Cell Rep* 14: 22–31

Kahl G (2015) Recombinant protein production

Kang HA, Choi ES, Hong WK, Kim JY, Ko SM, Sohn JH & Rhee SK (2000) Proteolytic stability of recombinant human serum albumin secreted in the yeast Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* 53: 575–582

Kim Y & Worrell E (2002) CO2 emission trends in the cement industry: An international comparison. *Mitig Adapt Strateg Glob Chang* 7: 115–133

Kitney RI, Bell J & Philp J (2021) Build a Sustainable Vaccines Industry with Synthetic Biology. *Trends Biotechnol* 39: 866–874

Klein T, Lange S, Wilhelm N, Bureik M, Yang TH, Heinzle E & Schneider K (2014) Overcoming the metabolic burden of protein secretion in Schizosaccharomyces pombe – A quantitative approach using 13C-based metabolic flux analysis. *Metab Eng* 21: 34–45

Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene* 234: 187–208

Kumar A, Udugama IA, Gargalo CL & Gernaey K V. (2020) Why is batch processing still dominating the biologics landscape? Towards an integrated continuous bioprocessing alternative. *Processes* 8: 1–19

Lalwani MA, Kawabe H, Mays RL, Hoffman SM & Avalos JL (2021) Optogenetic Control of Microbial Consortia Populations for Chemical Production. *ACS Synth Biol* 10: 2015–2029

Lee ME, DeLoache WC, Cervantes B & Dueber JE (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth Biol* 4: 975–986

Lee SY & Kim HU (2015) Systems strategies for developing industrial microbial strains. *Nat Biotechnol* 33: 1061–1072

Li L, Liu X, Wei K, Lu Y & Jiang W (2019) Synthetic biology approaches for chromosomal integration of genes and pathways in industrial microbial systems. *Biotechnol Adv* 37: 730–745

Liu Z, Tyo KEJ, Martínez JL, Petranovic D & Nielsen J (2012) Different expression systems for production of recombinant proteins in Saccharomyces cerevisiae. *Biotechnol Bioeng* 109: 1259–1268

Love KR, Dalvie NC & Love JC (2018) The yeast stands alone: the future of protein biologic production. *Curr Opin Biotechnol* 53: 50–58

Love KR, Panagiotou V, Jiang B, Stadheim TA & Love JC (2010) Integrated single-cell analysis shows Pichia pastoris secretes protein stochastically. *Biotechnol Bioeng* 106: 319–325

Love KR, Politano TJ, Panagiotou V, Jiang B, Stadheim TA & Love JC (2012) Systematic single-cell analysis of pichia pastoris reveals secretory capacity limits productivity. *PLoS One* 7: 1–11

Lovelett RJ, Zhao EM, Lalwani MA, Toettcher JE, Kevrekidis IG & Avalos JL (2021) Dynamical Modeling of Optogenetic Circuits in Yeast for Metabolic Engineering Applications. *ACS Synth Biol* 10: 219–227

Lugagne JB, Sosa Carrillo S, Kirch M, Köhler A, Batt G & Hersen P (2017) Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nat Commun* 8

Mattanovich D, Sauer M & Gasser B (2014) Yeast biotechnology: teaching the old dog new tricks.

McKinney W (2010) Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference* pp 56–61.

McKinnon KM (2018) Flow cytometry: An overview. *Curr Protoc Immunol* 2018: 5.1.1-5.1.11

Merksamer PI, Trusina A & Papa FR (2008) Real-Time Redox Measurements during Endoplasmic Reticulum Stress Reveal Interlinked Protein Folding Functions. *Cell* 135: 933–947

Milias-Argeitis A, Rullan M, Aoki SK, Buchmann P & Khammash M (2016) Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat Commun* 7: 1–11

Moreno Morales N, Patel MT, Stewart CJ, Sweeney K & McClean MN (2021) Optogenetic Tools for

Control of Public Goods in Saccharomyces cerevisiae. *mSphere* 6

Mortimer RK & Johnston JR (1986) Genealogy of Principal Strains of the Yeast Genetic Stock Center. *Genetics* 113: 35

Motta-Mena LB, Reade A, Mallory MJ, Glantz S, Weiner OD, Lynch KW & Gardner KH (2014) An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nat Chem Biol* 10: 196–202

Nielsen J & Jewett MC (2008) Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. *FEMS Yeast Res* 8: 122–131

Orij R, Urbanus ML, Vizeacoumar FJ, Giaever G, Boone C, Nislow C, Brul S & Smits GJ (2012) Genome-wide analysis of intracellular pH reveals quantitative control of cell division rate by pHc in Saccharomyces cerevisiae. *Genome Biol 2012 139* 13: 1–15

Papanikou E & Glick BS (2009) The yeast Golgi apparatus: Insights and mysteries. *FEBS Lett* 583: 3746–3751

Parekh RN & Wittrup KD (1997) Expression level tuning for optimal heterologous protein secretion in Saccharomyces cerevisiae. *Biotechnol Prog* 13: 117–122

Pedregosa FABIANPEDREGOSA F, Michel V, Grisel OLIVIERGRISEL O, Blondel M, Prettenhofer P, Weiss R, Vanderplas J, Cournapeau D, Pedregosa F, Varoquaux G, *et al* (2011) Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12: 2825–2830

Peña A, Sánchez NS, Álvarez H, Calahorra M & Ramírez J (2015) Effects of high medium pH on growth, metabolism and transport in Saccharomyces cerevisiae. *FEMS Yeast Res* 15: 5

Pincus D, Chevalier MW, Aragón T, van Anken E, Vidal SE, El-Samad H & Walter P (2010) BiP binding to the ER-stress sensor Ire1 tunes the homeostatic behavior of the unfolded protein response. *PLoS Biol* 8

Rakestraw JA, Baskaran AR & Wittrup KD (2006) A Flow Cytometric Assay for Screening Improved Heterologous Protein Secretion in Yeast. *Biotechnol Prog* 22: 1200–1208

Ramos-de-la-Peña AM, González-Valdez J & Aguilar O (2019) Protein A chromatography: Challenges and progress in the purification of monoclonal antibodies. *J Sep Sci* 42: 1816–1827

Rapoport TA, Matlack KES, Plath K, Misselwitz B & Staeck O (1999) Posttranslational Protein Translocation Across the Membrane of the Endoplasmic Reticulum. 380: 1143–1150

Rizzo MA & Piston DW (2005) High-Contrast Imaging of Fluorescent Protein FRET by Fluorescence Polarization Microscopy. *Biophys J* 88: L14–L16

Robertson JB, Davis CR & Johnson CH (2013) Visible light alters yeast metabolic rhythms by inhibiting respiration. *Proc Natl Acad Sci* 110: 21130–21135

Rogers DW, McConnell E & Greig D (2012) Molecular quantification of Saccharomyces cerevisiae α-pheromone secretion. *FEMS Yeast Res* 12: 668–674

Romanos MA, Scorer CA & Clare JJ (1992) Foreign gene expression in yeast: a review. *Yeast* 8: 423–488

de Ruijter JC, Koskela E V., Nandania J, Frey AD & Velagapudi V (2018) Understanding the metabolic burden of recombinant antibody production in Saccharomyces cerevisiae using a quantitative metabolomics approach. *Yeast* 35: 331–341

Sadowski I, Ma J, Triezenberg S & Ptashne M (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nat 1988 3356190* 335: 563–564

Salinas F, Rojas V, Delgado V, Agosin E & Larrondo LF (2017a) Optogenetic switches for light-controlled gene expression in yeast. *Appl Microbiol Biotechnol* 101: 2629–2640

Salinas F, Rojas V, Delgado V, Agosin E & Larrondo LF (2017b) Optogenetic switches for light-controlled gene expression in yeast. *Appl Microbiol Biotechnol* 101: 2629–2640

Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, Sell BR, Allen JR, Day RN, Israelsson M, *et al* (2013) A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat Methods 2013 105* 10: 407–409

Sheng J, Flick H & Feng X (2017) Systematic optimization of protein secretory pathways in Saccharomyces cerevisiae to increase expression of hepatitis B Small Antigen. *Front Microbiol* 8: 1–10

Shimizu-Sato S, Huq E, Tepperman JM & Quail PH (2002) A light-switchable gene promoter system. *Nat Biotechnol 2002 2010* 20: 1041–1044

Sophie Mokas JRM, Cristina Garreau M-J, Fournier ´e, Robert F, Arya P, Kaufman RJ, Pelletier J & Mazroui* and R (2009) Uncoupling Stress Granule Assembly and Translation Initiation Inhibition. *Mol Biol Cell* 20: 2673–2683

Sorgo AG, Heilmann CJ, Dekker HL, Brul S, Koster CG de & Klis FM (2010) Mass spectrometric analysis of the secretome of Candida albicans. *Yeast* 27: 661–672

Steel H, Habgood R, Kelly CL & Papachristodoulou A (2020) In situ characterisation and manipulation of biological systems with Chi.Bio. *PLOS Biol* 18: e3000794

Su R, Tang X, Feng L, Yao G long & Chen J (2020) Development of quantitative magnetic beads-based flow cytometry fluorescence immunoassay for aflatoxin B1. *Microchem J* 155: 104715

Tang H, Bao X, Shen Y, Song M, Wang S, Wang C & Hou J (2015) Engineering protein folding and translocation improves heterologous protein secretion in Saccharomyces cerevisiae. *Biotechnol Bioeng* 112: 1872–1882

Thak EJ, Yoo SJ, Moon HY & Kang HA (2020) Yeast synthetic biology for designed cell factories producing secretory recombinant proteins. *FEMS Yeast Res* 20: 1–17

Thibault G, Ismail N & Ng DTW (2011) The unfolded protein response supports cellular robustness as a broad-spectrum compensatory pathway. *Proc Natl Acad Sci* 108: 20597–20602

Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS & Walter P (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* 101: 249–258

Uchański T, Zögg T, Yin J, Yuan D, Wohlkönig A, Fischer B, Rosenbaum DM, Kobilka BK, Pardon E & Steyaert J (2019) An improved yeast surface display platform for the screening of nanobody immune libraries. *Sci Reports 2019 91* 9: 1–12

Uhlendorf J, Miermont A, Delaveau T, Charvin G, Fages F, Bottani S, Batt G & Hersen P (2012) Long-term model predictive control of gene expression at the population and single-cell levels. *Proc Natl Acad Sci U S A* 109: 14271–6

Valkonen M, Penttilä M & Saloheimo M (2003) Effects of inactivation and constitutive expression of the unfolded- protein response pathway on protein production in the yeast Saccharomyces cerevisiae. *Appl Environ Microbiol* 69: 2065–72

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, *et al* (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods 2020 173* 17: 261–272

Waldron C & Lacroute F (1975) Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *J Bacteriol* 122: 855–865

Walsh G (2018) Biopharmaceutical benchmarks 2018. *Nat Biotechnol* 36: 1136–1145

Wang G, Huang M & Nielsen J (2017) Exploring the potential of Saccharomyces cerevisiae for biopharmaceutical protein production. *Curr Opin Biotechnol* 48: 77–84

Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D & Brown PO (2002) Precision and functional specificity in mRNA decay. *PNAS April* 30: 5860–5865

Weber E, Engler C, Gruetzner R, Werner S & Marillonnet S (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS One* 6: e16765

Wentz AE & Shusta E V. (2007) A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl Environ Microbiol* 73: 1189–1198

Williams TC, Averesch NJH, Winter G, Plan MR, Vickers CE, Nielsen LK & Krömer JO (2015) Quorum-sensing linked RNA interference for dynamic metabolic pathway control in Saccharomyces cerevisiae. *Metab Eng* 29: 124–134

Winder CL & Lanthaler K (2011) The Use of Continuous Culture in Systems Biology Investigations. *Methods Enzymol* 500: 261–275

Winston F, Dollard C & Ricupero-Hovasse SL (1995) Construction of a set of convenient saccharomyces cerevisiae strains that are isogenic to S288C. *Yeast* 11: 53–55

Wiseman RL, Powers ET, Buxbaum JN, Kelly JW & Balch WE (2007) An Adaptable Standard for Protein Export from the Endoplasmic Reticulum. *Cell* 131: 809–821

Wittrup KD, Robinson AS, Parekh RN & Forrester KJ (1994) Existence of an optimum expression level for secretion of foreign proteins in yeast. *Ann N Y Acad Sci* 745: 321–330

Wong BG, Mancuso CP, Kiriakov S, Bashor CJ & Khalil AS (2018) Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nat Biotechnol* 36: 614–623

Xu P & Robinson AS (2009) Decreased secretion and unfolded protein response up-regulation are correlated with intracellular retention for single-chain antibody variants produced in yeast. *Biotechnol Bioeng* 104: 20–29

Xu X, Liu Y, Du G, Ledesma-Amaro R & Liu L (2020) Microbial Chassis Development for Natural Product Biosynthesis. *Trends Biotechnol* 38: 779–796

Yang SY, Lien KY, Huang KJ, Lei HY & Lee G Bin (2008) Micro flow cytometry utilizing a magnetic bead-based immunoassay for rapid virus detection. *Biosens Bioelectron* 24: 855–862

Yu LP, Wu FQ & Chen GQ (2019) Next-Generation Industrial Biotechnology-Transforming the Current Industrial Biotechnology into Competitive Processes. *Biotechnol J* 14: 1–9

Zhao EM, Zhang Y, Mehl J, Park H, Lalwani MA, Toettcher JE & Avalos JL (2018) Optogenetic regulation of engineered cellular metabolism for microbial chemical production. *Nature* 555: 683–687

Zhao X, Chang AY, Toh-e A & Arvan P (2007) A Role For Lte1p (a Low Temperature Essential Protein Involved in Mitosis) in Proprotein Processing in the Yeast Secretory Pathway. *J Biol Chem* 282: 1670–1678

Zoltowski BD, Motta-Mena LB & Gardner KH (2013) Blue light-induced dimerization of a bacterial LOV-HTH DNA-binding protein. *Biochemistry* 52: 6653–6661

Annex A: Internal feedback control

A.1. *Introductory remarks*

Based on the observation in Chapter 4, showing the rise of a transient subpopulation defined by internal levels of protein and stress levels, here I proposed an internal feedback controller that aims to reduce population heterogeneity. The approach aims to implement a controller that auto-regulates the expression of the protein of interest, at the level of RNA interference, based on the stress levels. The interference RNA is complementary to the mNeonGreen sequence that is common in all proteins of interest, because the fusion of the reporter. Thus, the accumulator cells, presented in Chapter 4, would minimize its presence in the overall population, because they do not produce protein of interest upon stress. The idea of an internal feedback came inspired by the study carried out in Ceroni et al., 2018. These results are documented in an annex because I did have time just to start the strain construction, but I considered that the characterization of the secretory process, and the external feedback control should be implemented and analyzed before carrying this approach.

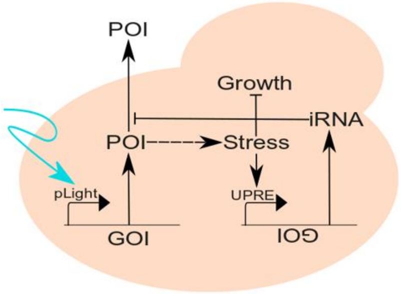


Figure A.0. ***Schematic of the internal feedback approach.*** GOI: gene of interest, iRNA: interference RNA, UPRE: UPR driven promoter, pLight: optogenetic promoter. Blue arrow: Light induction.

Despite being conserved in most fungal kingdoms; RNAi pathway is not functional in *S. cerevisiae*. However, in previous works (Drinnenberg *et al*, 2009) the RNAi pathway has been reconstructed in *S. cerevisiae* by introducing two genes (AGO1 and DCR1) from *S. castellii*. This, together with a transcriptional unit that expresses an interference RNA (iRNA) complementary mNeonGreen part of the to the target transcript, is sufficient to produce the small interference RNA (siRNA), which reduces the expression of a complementary target gene at the level of translation. The interference intensity of such target gene can be modulated through the length and expression levels of the iRNA. Therefore, the first step is to characterize the interference intensity on the target gene for different expression levels of the iRNA.

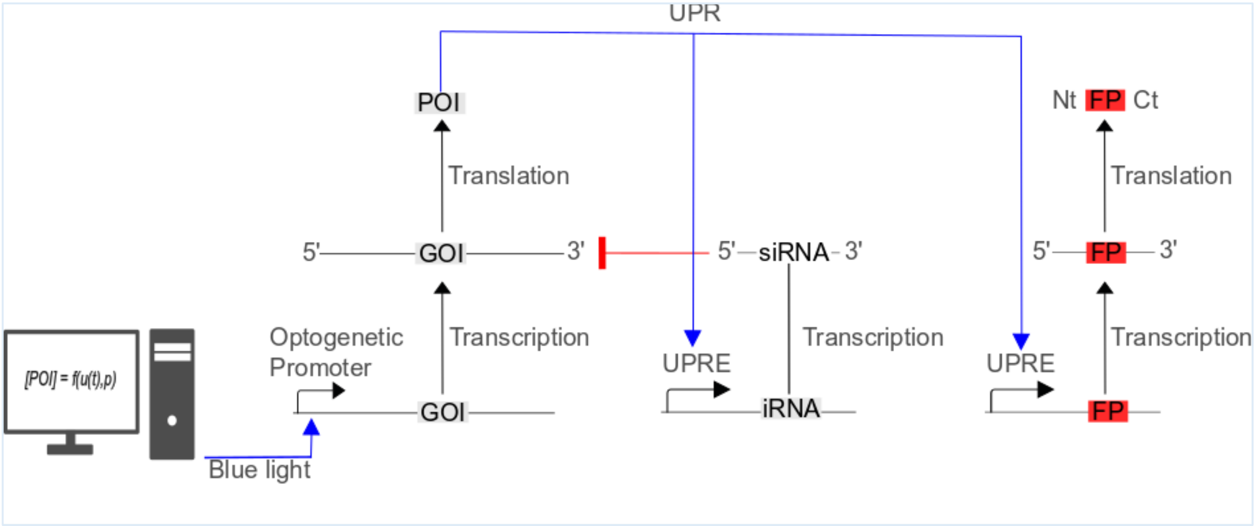


Figure A.14. ***General controller circuit.*** *From left to right; the transcriptional activity of the POI is controlled by the bioreactor platform that illuminates the sample with blue light to activate the optogenetic expression system. Then the GOI will be expressed and eventually the POI will generate secretory pathway stress. This activates the UPR response by the UPRE upregulation. In this controller system there are two exogenous genes controlled by the UPRE, first a iRNA, that is the complementary sequence to the GOI CDS, hence forming a double stranded RNA upon transcription. That double stranded RNA is processed to generate siRNA that targets the POI mRNA to degradation, by the action of the RNAi complementary proteins. The second gene controlled by the UPRE codes for a fluorescent reporter (FP), to monitor the UPR activation, built in Objective 1, that should match with the iRNA levels.*

In order to find the appropriate features that are required for the RNAi system I started from existing ones (N. C. Crook, Schmitz, & Alper, 2014; N. Crook, Sun, Morse, Schmitz, & Alper, 2016). According to previous works, the highest interference is reached when the hairpin formed by the iRNA is complementary to 200bp of the target mRNA. Therefore, I built a hairpin with such length, and half of it, to be expressed under the control of three different constitutive promoters (REV1, RPL18B, and TDH3) that show low, medium and high expression levels, respectively, and also another under the control of the optogenetic expression system. The topology of this design consisted in a sequence of the length previously commented, followed by an intron from *S. pombe*, and the reverse complementary of the first region. Such regions should be the same as the CDS sequence that is pretended to target. Unfortunately, I could not use such designs because I couldn't clone them. Such constructions have palindromic regions, essential to form the hairpin structure, that probably makes the cloning process challenging in *E. coli*, because interferes with the replication machinery. Therefore, I abandoned these designs and opted for something simpler. I still took advantage of existing designs (Williams *et al*, 2015), which consisted just in using the reverse complementary of the targeted sequence in its full length. In such case, once transcribed, the resulting RNA will form a double stranded RNA with the target messenger, since they are complementary. In this case the cloning worked well and I could make the constructions and integrate them in the yeast chromosome without problem. The yeast strain incorporated three new transcriptional units in the same locus, in addition to the AGO1, and DCR1, necessary genes for the RNAi pathway. One transcriptional unit carrying the CDS that is transcribed into the iRNA (the antisense) under the control of the optogenetic induction system. The second, coding for the target POI (the sense), in this case mNeonGreen. And the last, coding for a reporter that serves as reference for expression levels, mScarlet-I. Both, the target POI and the reference reporter are expressed under the control of pTDH3, a strong promoter.

A.2. *Results*

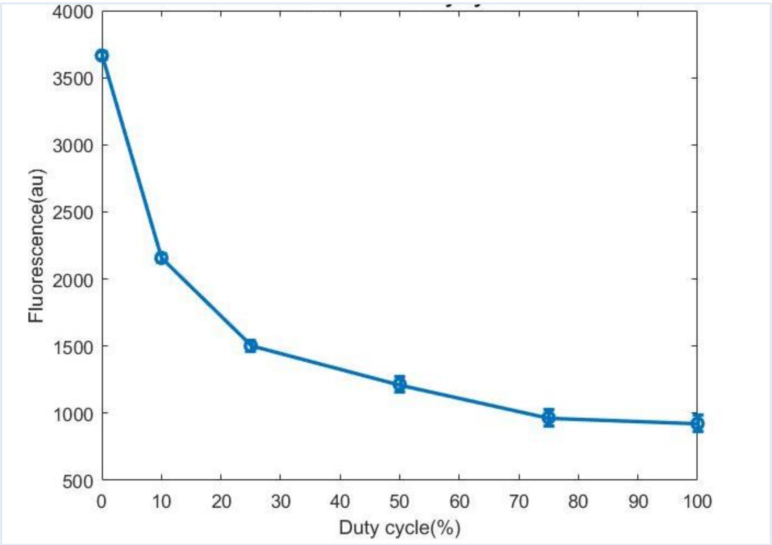


*Figure A.25. **Fluorescence reduction level by the expression of a targeted iRNA.** The figure shows the median fluorescence levels at steady state, after 18 hours' induction of cell cultures expressing the iRNA system, which expression is driven by the EL222 optogenetic expression system and induced at different duty cycles. The data was obtained from cells growing in synthetic complete low fluorescence media, 2% glucose, at 30C, in the InBioreactors platform.*

In Figure A.2 is shown the performance at steady state of the interference RNA system. It reduces the fluorescence levels upon induction of the optogenetic system driven the transcription of the antisense RNA, anti mNeonGreen, full length. The interference levels seem exponential with the duty cycle, and it is able to reduce expression up to four fold.

In Figure A.3 I just show the dynamics over the time from a microscope experiment, to have a notion of the time needed to reach the steady state. Notice that the results from microscopy cannot be compared quantitatively with the ones obtained in the bioreactors, since the light intensity for induction, as well as the growth conditions are not comparable. According to the results shown in Figure A.3, it takes around 8 to 9 hours to reach steady state of interference. However, it could be less, due to the fact that in our microscopy setup, the cells grow in a microfluidic chip that at some point saturates of cells, and those accumulate protein because the lack of growth. Therefore, it can show an increased fluorescence at the end of the experiment

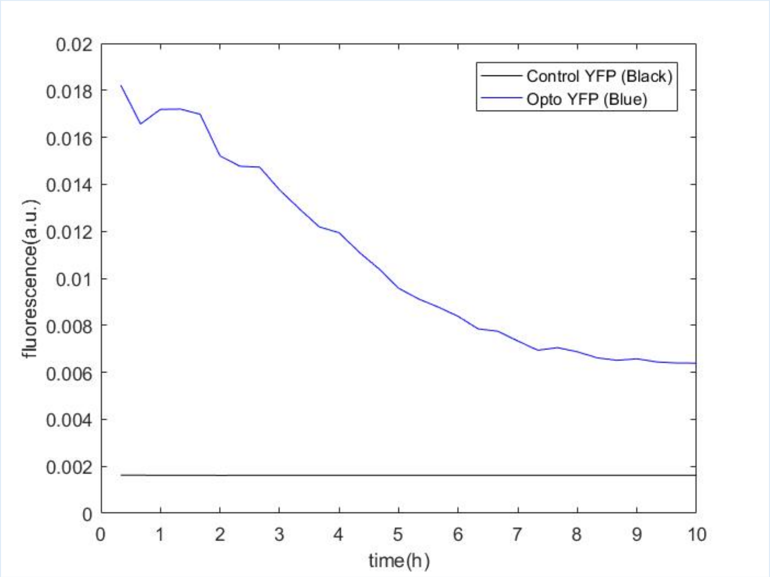


Figure A.36. ***RNA interference on time.*** *Interference levels of iRNA over time obtained from single cells by fluorescence microscopy. The cells grew at 30°C under blue light in a microfluidics chip in the microscope from time = 0. Media was constantly supplied. The mean fluorescence computed from 100 individual cells in each condition, during 10 hours.*

A.3. *Conclusion*

The results from Annex A have been generated by using the *InBioreactor* platform, and the microscopy platform. I have managed to build the RNA interference system, and start its characterization. If continue I will build the strain that expresses the iRNA under the control of the UPRE, to perform the negative internal feedback that regulates the expression of the POI as a function of the stress levels.

Annex B: Supplementary material

B.1. *Results Chapter 4*

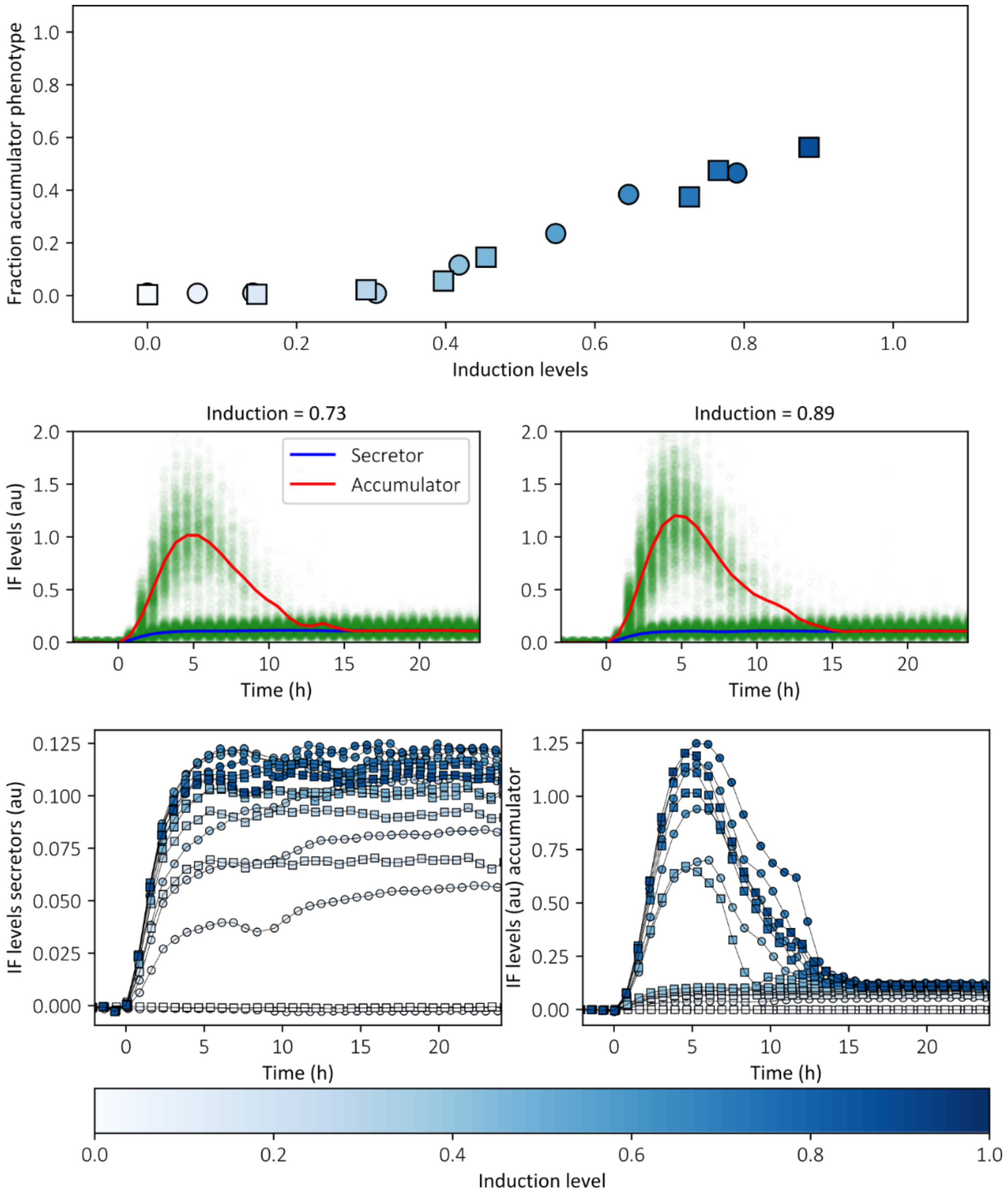


Figure B.1. ***Differentiation of secretor and accumulator phenotype from the overall population.*** On most top, the plot shows the relation between induction levels and accumulator phenotype fraction, at the time point with maximum proportion of accumulators. The middle plots show two examples of Gaussian mixture model used to discriminate both phenotypes. In each of them the temporal evolution of internal fluorescence (IF) in log scale over time. Note that in this case the autofluorescence is not subtracted. Every single cell captured by the cytometry is shown as green dots. The different induction levels are indicated on top of each plot. The median for accumulators is represented as a red line, the median for secretors as a blue line. The bottom plot shows the IF median of the two phenotypes after populations filtering, to the left the secretors, to the right the accumulators. The intensity of the blue is proportional to the induction levels. Square markers represent data collected in a different set of experiments.

B.2. *Code example Chapter 6*

Example of code for processing the raw data from the reactors platform

Note: It requires a .csv file with the metadata

```
def process_data_dynamics(experiment_date):
    import pandas as pd
    import numpy as np
    import os
    from scipy.signal import savgol_filter

    folder = 'C:\\...'

    metadata =  pd.read_csv(folder + '\\' + str(experiment_date) + '_metadata.csv')
    reactors = metadata['reactor']
    cells_raw = {}
    cells = {}
    sensor = {}
    POI = {}
    sensor_median = {}
    sensor_mean = {}
    POI_median = {}
    POI_mean = {}
    LED = {}
    induction_starts = {}

    for reactor in reactors:
        path = folder + '\\reactor-data\\reactor-' + str(reactor)

        # Load cytometry data:
        with open(path + '\\cells.csv', 'r') as infile:
            cells_raw[reactor] = pd.read_csv(infile)

        # Load LED data:
        with open(path + '\\LEDs.csv', 'r') as infile:
            LED[reactor] = pd.read_csv(infile)

        # Process LED data
        if metadata.iloc[reactor - 1]['DC'] == 0:
            path_LED = folder + '\\reactor-data\\reactor-' + str(reactor + 1)
            with open(path_LED + '\\LEDs.csv', 'r') as infile:
                LED_TEMP = pd.read_csv(infile)
            LED[reactor]['time_s'] = LED_TEMP['time_s']
            LED[reactor]['intensity'] = np.zeros(len(LED[reactor]['time_s']))
            # Computing time in hours units
            LED[reactor]['time_h'] = (LED[reactor]['time_s'] - \
            cells_raw[reactor]['time_s'].iloc[0])/60/60
    # Computing when induction starts
    induction_starts_idx = min(LED_TEMP[LED_TEMP['intensity'] > 0].index.values)
    induction_starts[reactor] = LED[reactor]['time_h'].iloc[induction_starts_idx]
if metadata.iloc[reactor - 1]['DC'] > 0:
    # Computing time in hours units
    LED[reactor]['time_h'] = (LED[reactor]['time_s'] - \
    cells_raw[reactor]['time_s'].iloc[0])/60/60
    # Computing when induction starts
    induction_starts_idx = min(LED[reactor][LED[reactor]['intensity'] > 0].index.values)
    induction_starts[reactor] = LED[reactor]['time_h'].iloc[induction_starts_idx]

# Normalize time to induction
LED[reactor]['time_norm'] = LED[reactor]['time_h'] - induction_starts[reactor]

# Gatting based on FSC and SSC
cells[reactor] = cells_raw[reactor][(cells_raw[reactor]['FSC-HLin'] > 6*10**2) & \
                              (cells_raw[reactor]['FSC-HLin'] < 1.5*10**3) & \
                               (cells_raw[reactor]['SSC-HLin'] > 1.2*10**3) & \
                               (cells_raw[reactor]['SSC-HLin'] < 3*10**3)].copy()

# Normalize fluorescence by FSC
cells[reactor]['GRN-V/FSC'] = cells[reactor]['GRN-V-HLin']/cells[reactor]['FSC-HLin']
cells[reactor]['GRN-B/FSC'] = cells[reactor]['GRN-B-HLin']/cells[reactor]['FSC-HLin']
cells[reactor]['ORG-G/FSC'] = cells[reactor]['ORG-G-HLin']/cells[reactor]['FSC-HLin']

# Computing time in hours units
cells[reactor]['time_h'] = (cells[reactor]['time_s'] - \
cells[reactor]['time_s'].iloc[0])/60/60

# Normalize time to induction
cells[reactor]['time_norm'] = cells[reactor]['time_h'] - induction_starts[reactor]

# Separate the sensor strain (['GRN-V/FSC'] > 0.2) from the POI strain
sensor[reactor] = cells[reactor][cells[reactor]['GRN-V/FSC'] > 0.2].reset_index()
POI[reactor] = cells[reactor][cells[reactor]['GRN-V/FSC'] < 0.07].reset_index()

# Compute medians
sensor_median[reactor] = sensor[reactor].groupby(['time_h']).median().reset_index()
POI_median[reactor] = POI[reactor].groupby(['time_h']).median().reset_index()

# Compute and substract autofluorescence (from -3 hours before induction)
AF = np.median(POI_median[reactor]['GRN-B/FSC'][(POI_median[reactor]['time_norm'] < 0) & \
                                (POI_median[reactor]['time_norm'] >= -1.5)])
POI_median[reactor]['IF'] = POI_median[reactor]['GRN-B/FSC'] - AF
AF = np.median(POI_median[reactor]['ORG-G/FSC'][(POI_median[reactor]['time_norm'] < 0) & \
                                (POI_median[reactor]['time_norm'] >= -1.5)])
POI_median[reactor]['UPR'] = POI_median[reactor]['ORG-G/FSC'] - AF
```

```
    AF = np.median(sensor_median[reactor]['ORG-G/FSC'][(sensor_median[reactor]['time_norm'] < 0) &
\
                     (sensor_median[reactor]['time_norm'] >= -1.5)])
    sensor_median[reactor]['mScarlet'] = sensor_median[reactor]['ORG-G/FSC'] - AF

    # Compute growth rates
    # Assume growth rate of sensor strain is 0.4h^-1
    s = []
    p = []
    points_slope_gr = 3
    for i in range(0, len(sensor_median[reactor]['time_h'])):
      s.append(len(sensor[reactor][sensor[reactor]['time_h'] == \
                       sensor_median[reactor]['time_h'].iloc[i]]))
      p.append(len(POI[reactor][POI[reactor]['time_h'] == \
                     POI_median[reactor]['time_h'].iloc[i]]))
    sensor_median[reactor]['events'] = s
    POI_median[reactor]['events'] = p
    POI_median[reactor]['ratio'] = POI_median[reactor]['events']\
    /sensor_median[reactor]['events']
    POI_median[reactor]['log_ratio'] = np.log(POI_median[reactor]['ratio'])
    a = []
    for i in range(1, len(sensor_median[reactor]['time_h'])-points_slope_gr):
      slope = np.polyfit(POI_median[reactor]['time_norm'][i:i+points_slope_gr], \
                  POI_median[reactor]['log_ratio'][i:i+points_slope_gr],1)
      a.append(slope[0]+0.4)
    a = a + [np.mean(a[-10:-1])]*points_slope_gr
    a = [a[0]] + a

    # Filter to reduce noise
    gr_filtered = savgol_filter(a, points_slope_gr, 1)
    POI_median[reactor]['gr'] = gr_filtered
```