



Optimisation de la conception de bioprocédés : vers une approche intégrée biologie de synthèse et conduite du procédé

Guillaume Jeanne

► To cite this version:

Guillaume Jeanne. Optimisation de la conception de bioprocédés : vers une approche intégrée biologie de synthèse et conduite du procédé. Autre [cs.OH]. Université Paris Saclay (COMUE), 2018. Français. NNT : 2018SACLC064 . tel-02015607

HAL Id: tel-02015607

<https://theses.hal.science/tel-02015607>

Submitted on 12 Feb 2019

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.

Optimisation de la conception de bioprocédés : vers une approche intégrée biologie de synthèse et conduite du procédé

Thèse de doctorat de l'Université Paris-Saclay
préparée à CentraleSupélec

École doctorale n° 580 Sciences et technologies de
l'information et de la communication (STIC)
Spécialité de doctorat : Automatique

Thèse présentée et soutenue à Gif-sur-Yvette, le 27 septembre 2018,

par **Guillaume JEANNE**

Composition du jury :

Paul-Henry COURNÈDE	Président
Professeur, CentraleSupélec	
Olivier BERNARD	Rapporteur
Directeur de recherche, INRIA Sophia-Antipolis	
Denis DOCHAIN	Rapporteur
Professeur, Université Catholique de Louvain	
Stéphanie HEUX	Examinatrice
Chargée de recherche, INRA Toulouse	
Didier DUMUR	Directeur de thèse
Professeur, CentraleSupélec	
Anne GOELZER	Co-encadrante
Ingénieur de recherche, INRA Jouy-en-Josas	
Vincent FROMION	Co-encadrant, invité
Directeur de recherche, INRA Jouy-en-Josas	
Sihem TEBBANI	Co-encadrante, invitée
Professeur, CentraleSupélec	
Pierre MARTINON	Invité
Chercheur, INRIA Paris-Saclay	

Remerciements

En premier lieu, je voudrais remercier très chaleureusement mes directeurs et encadrants de thèse, MMme Sihem Tebbani et Anne Goelzer et MM. Didier Dumur et Vincent Fromion, pour leur accompagnement sans faille du premier au dernier jour. Je suis conscient d’avoir eu le privilège de travailler avec eux et que ces travaux n’auraient su aboutir sans leur concours. Merci à Sihem pour toutes ces « cinq minutes » de discussion qui se sont souvent transformées en deux heures d’échanges passionnants. Merci à Vincent d’être venu tous les jours m’abreuver de sa culture scientifique. Cette heure quotidienne, que je me suis forcé à suivre au début, a progressivement fait son effet, en provoquant des déclics et déverrouillant en douceur des portes que je n’aurais su forcer. Merci à Anne de m’avoir poussé à m’engager pleinement dans des échéances à court terme bénéfiques à l’avancée des travaux et à mon enrichissement personnel. Merci à Didier d’avoir tenu la barre et su me ramener sur le droit chemin lorsque j’étais tenté de m’en éloigner. À tous les quatre, merci pour ces journées, ces soirées (parfois prolongées) à tenter d’être à la hauteur du sujet et du cadre de travail idéaux dont j’ai eu la chance de profiter.

Je tiens également à remercier MM. Olivier Bernard et Denis Dochain pour l’honneur qu’ils m’ont fait d’accepter de rapporter cette thèse. Merci à eux d’avoir consacré leur temps précieux à la relecture de mon manuscrit et pour leurs commentaires très enrichissants qui m’ont permis d’envisager mes travaux sous un angle neuf. Je remercie aussi Mme Stéphanie Heux, M. Paul-Henry Cournède et M. Pierre Martinon d’avoir accepté de participer à mon jury de thèse et pour l’intérêt qu’ils ont témoigné à ces travaux. En particulier, merci à M. Martinon pour son aide et sa disponibilité concernant l’utilisation du logiciel Bocop. Merci à M. Cournède d’avoir présidé le jury et participé à mon comité de mi-thèse avec M. Bernard, par ailleurs.

Merci aux membres de l’équipe Maïage de l’INRA et du département Automatique de CentraleSupélec pour leur accueil et leur soutien pendant 3 années.

Enfin, merci à toutes les personnes qui m’ont fait la joie de m’entourer tout au long ma thèse, de mon sujet sur les algues jusqu’aux derniers petits fours de la soutenance : ma famille au grand complet, (la) Dégaine, la MPSIV, les amis des éclais, FMR, le Guynemer Brass Band Climb Orchestra, le GT Math-Bio, Hystérésis, la rue des Terres de Bordes, la rue du Tage, le petit bureau noir sous la lucarne, et bien entendu Laureline et sa famille.

Table des matières

Introduction	9
1 Démarches usuelles pour la production d'un composé par bioprocédé	15
1.1 Approche « conduite du procédé »	16
1.1.1 Modèles macroscopiques phénoménologiques	16
1.1.2 Utilisation des modèles pour la commande des bioprocédés en vue de la production d'un composé d'intérêt	27
1.1.3 Conclusion sur l'approche macroscopique phénoménologique	35
1.2 Approche de type « conception de souche »	36
1.2.1 Modèles intracellulaires	36
1.2.2 Utilisation des modèles intracellulaires pour l'optimisation de souche et de bioprocédés	48
1.3 Conclusion du chapitre	53
2 Modélisation et optimisation du comportement de micro-organismes	57
2.1 Modèle agrégé	58
2.1.1 Objectifs de la modélisation et de l'agrégation	58
2.1.2 Formalisation du problème	58
2.1.3 Dynamiques & contraintes	60
2.1.4 Focus sur l'expression du taux spécifique de croissance	63
2.1.5 Bilan du modèle	65
2.1.6 Paramètres du modèle	67
2.2 Comportement nominal du modèle	72
2.2.1 Caractérisation à l'équilibre : lien avec un modèle macroscopique	74
2.2.2 Caractérisation à l'équilibre : lien avec le modèle RBA	81
2.2.3 Conclusion des différentes comparaisons	87
2.3 Optimisation de la production du bioprocédé	87
2.3.1 Position du problème	88
2.3.2 Commande optimale	91
2.3.3 Formulation discrète	93
2.3.4 Résolution numérique	96
2.3.5 Analyse des résultats	97
2.4 Conclusion du chapitre	103

3	Principe de conception rationnelle de souche	107
3.1	Vers une conception rationnelle de souche	108
3.1.1	Éléments de mise en œuvre réaliste d'une stratégie de commande de la cellule	108
3.1.2	Relation entre ν_{E_i} et μ dans le cas d'un gène constitutif	108
3.1.3	Choix de la paramétrisation pour le modèle agrégé	110
3.1.4	Choix des gènes constitutifs dans le cadre du modèle agrégé	110
3.2	Problèmes d'optimisation de la conception de bioprocédé batch	111
3.2.1	Contraintes complémentaires pour l'autonomie de la cellule	111
3.2.2	Reconfiguration complète d'une souche (Cas n° 1)	113
3.2.3	Introduction d'un gène de synthèse constitutif (Cas n° 2)	115
3.2.4	Introduction d'un gène de synthèse constitutif, activable à un ins- tant donné (Cas n° 3)	118
3.2.5	Introduction d'un gène de synthèse constitutif avec action sur la croissance (Cas n° 4)	121
3.2.6	Conclusion à propos de l'optimisation de procédé en mode de fonc- tionnement batch	121
3.3	Optimisation de la conception de bioprocédés fed-batch	123
3.3.1	Adaptation générale du problème	124
3.3.2	Problème avec cellules connues à l'avance	127
3.3.3	Optimisation conjointe du débit d'alimentation et de l'expression d'un gène de synthèse	130
3.4	Conclusion du chapitre	133
4	Extension de l'approche à des descriptions cellulaires plus détaillées	135
4.1	Éléments de généralisation	135
4.1.1	Protocole de construction d'une représentation dRBA	135
4.1.2	Optimisation de la production d'un composé d'intérêt	141
4.1.3	Bilan	141
4.2	Exemple d'un modèle à deux voies de synthèse de l'énergie	142
4.2.1	Éléments biologiques & Construction du modèle	142
4.2.2	Caractérisation de l'état nominal du modèle	146
4.2.3	Utilisation du modèle pour la production de composé d'intérêt	151
4.2.4	Conclusions relatives au modèle à deux voies	160
4.3	Conclusion du chapitre	160
	Conclusions & Perspectives	161
	Bibliographie	165
A	Introduction au fonctionnement d'une cellule bactérienne	171
A.1	Mode de vie des cellules bactériennes	172
A.1.1	Échelles de temps et d'espace	172
A.1.2	Un moyen d'action commun : les protéines	174
A.2	Les processus cellulaires nécessaires à la croissance exponentielle	174
A.2.1	Échanges avec le milieu extérieur	175
A.2.2	Le métabolisme : synthèse des métabolites et macrocomposants	176

A.2.3	Synthèse des ARN : la transcription	178
A.2.4	Synthèse des protéines : la traduction	179
A.2.5	Résumé	180
A.3	Régulation de l'expression des gènes	180
A.3.1	Régulation d'un groupe de gènes	180
A.3.2	Régulation globale : la réponse stringente	183
B	Un exemple d'utilisation du principe du maximum de Pontryaguine	185
B.1	Position du problème	185
B.2	Détermination des conditions nécessaires d'optimalité de Pontryaguine .	186
B.3	Caractérisation de l'arc singulier	188
B.4	Caractérisation de la commande de suivi de contrainte sur l'état	191
B.5	Caractérisation de la trajectoire optimale	191
B.6	Représentation de la trajectoire optimale	192

Introduction

Chaque année, des associations fêtent amèrement le jour du dépassement de la Terre de plus en plus tôt dans l'année. Cette date correspond au jour de l'année auquel l'humanité est supposée avoir consommé l'ensemble des ressources que la planète est capable de régénérer en un an. Ce déséquilibre entre consommation et capacité de production est dû à la fois à l'augmentation de la population, à l'évolution des modes de vie et à un appauvrissement des réserves de la Terre. L'épuisement des réserves naturelles est bien une conséquence de la manière dont les ressources naturelles sont consommées (dilapidées ?) par notre société. Pour ne prendre qu'un exemple, on constate que l'on épuise en à peine deux siècles des réserves de pétrole qui ont mis des centaines de millions d'années à se former. Ceci revient à dire que nous avons consommé le pétrole un million de fois plus vite qu'il n'a été produit.

Cette situation nous conduit droit vers l'inconnu. Il n'est pas raisonnable de continuer dans cette direction de consommation irréfléchie et il est nécessaire de revoir en profondeur la manière dont sont utilisées les ressources de la Nature. Notre devoir en tant que scientifiques est de préparer la transition vers un mode de consommation de ressources adapté aux modes de production, et inversement.

Au-delà de la problématique de la réduction de la consommation, une question pertinente est de savoir s'il existe des modes de production qui s'inscrivent dans les cycles de la nature et sont en accord avec la capacité de la Terre. Dans cette optique, les organismes vivants font partie de la solution. En prenant du recul, tous les organismes vivants transforment les molécules les unes en les autres. Les ressources dont nous avons besoin pour vivre sont elles-mêmes des molécules. Ainsi, n'y a-t-il pas en les organismes vivants un système potentiellement capable de produire nos ressources ? Depuis des milliards d'années, les espèces vivantes prospèrent sur Terre en utilisant comme ressource ce qui est produit par les autres espèces. L'objectif est d'inclure les transformations de ressources de l'espèce humaine dans ce réseau de production et consommation.

Un moyen existant pour faire produire des composés à des organismes est de recourir à la conception de souche. Cette méthode consiste à modifier des micro-organismes en leur ajoutant ou leur retirant des fonctions. Typiquement, il s'agit de modifier le fonctionnement du réseau métabolique (c'est-à-dire les réactions élémentaires que peuvent réaliser les cellules), en agissant sur les gènes de la cellule. Les potentiels ouverts par la modification des organismes vivants sont immenses. Dans ce contexte, les micro-organismes deviennent des usines moléculaires qu'il est possible d'enrichir en ajoutant les mécanismes nécessaires aux transformations désirées. La conception de souche revient à choisir une

cellule dite châssis à laquelle on ajoute des gènes. Le châssis doit être suffisamment bien connu et étudié pour pouvoir prédire le comportement une fois modifié par l'ajout de gènes de synthèse. En particulier, tous les mécanismes intracellulaires doivent être suffisamment connus pour qu'il soit possible d'en extrapoler le comportement dans une souche modifiée. Ainsi, lors de la conception de souches, les chercheurs optimisent en laboratoire le fonctionnement des cellules en trouvant la meilleure coordination des processus intracellulaires. Les modifications prennent racine dans la biologie, c'est-à-dire le fonctionnement connu et observé des souches. Les paramètres introduits sont relatifs à ce comportement biologique. Les modèles de comportement de souche qui sont utilisés peuvent contenir des milliers de variables et de paramètres.

Le passage à l'échelle préindustrielle se fait une fois qu'une souche suffisamment efficace a été assemblée en laboratoire. Ce passage d'échelle consiste à trouver l'environnement adapté à une production optimale d'un composé d'intérêt par la souche. Typiquement, il peut s'agir de maintenir une concentration constante en un certain composé dans l'environnement des souches. Pour une conduite efficace du procédé, des modèles macroscopiques du fonctionnement des souches sont développés. Pour pouvoir être utilisé en temps réel, ces modèles reposent sur un nombre de paramètres bien plus faible que lors de l'étape de conception. Aussi, le rôle de la conduite du procédé étant d'optimiser les conditions de culture, les modèles utilisés en conduite du procédé sont relatifs à l'évolution de l'environnement des souches et ne décrivent le comportement des organismes qu'à travers quelques variables. En ce sens, ils sont donc différents en termes de complexité de ceux évoqués précédemment.

Ainsi, même si l'on voit s'accroître de façon importante le nombre de petites molécules qui sont bio-produites, le développement des bioprocédés reste long et coûteux. La transition de la paillasse à l'échelle industrielle est un maillon faible de la chaîne de développement. Une étude de 2016 [26] indique que des projets d'ingénierie métabolique à l'échelle pilote (production en volume de quelques litres à une centaine) prennent un peu plus de trois ans et appellent sur cette période au moins deux personnes par an (soit six années-hommes). Le passage de la bio-production à l'échelle préindustrielle et industrielle est souvent encore plus coûteux comme le démontre le projet d'ingénierie métabolique de production de l'artémisinine par la bactérie *Escherichia coli* qui a nécessité plus de cent trente années-hommes pour être réalisé.

Pour pouvoir tester efficacement de nouvelles souches et de nouvelles modifications, il est nécessaire de gérer efficacement la jonction entre l'optimisation de la conception de souche et l'optimisation de la conduite du procédé. La biologie des systèmes propose une logique cohérente aux deux étapes d'optimisation. En effet, la biologie des systèmes offre une vision systémique de la structure et du comportement des organismes vivants. Cette vision s'intègre parfaitement à l'approche systémique habituellement considérée pour la conduite du procédé. Dans ce paradigme, la vie est décomposable en sous-systèmes (ou processus) qui interagissent entre eux. L'émergence de cette vision permet de développer une réelle ingénierie du vivant, et donc une modification en profondeur des cycles de développement des bioprocédés. L'optimisation de la conception de souche et de la conduite du procédé revient alors à l'optimisation d'un système dans son ensemble. La perspec-

tive attachée à la biologie des systèmes est une mutation vers une possible conception assistée par ordinateur (CAO) des organismes vivants et des procédés. Cette promesse résulte en grande partie des progrès actuels du champ biologique qui passe par l'accroissement significatif de la compréhension du fonctionnement de la cellule microbienne aux échelles cellulaire et subcellulaire et de la modélisation systémique de la cellule aux mêmes échelles. Ce mouvement est accompagné par des progrès significatifs dans l'ingénierie des génomes et dans les technologies d'observation et de quantification des processus vivants. L'ensemble de ces éléments ouvre sans conteste la voie au développement d'une réelle ingénierie des systèmes pour les microbes, incluant explicitement les étapes de design (*in silico*), de mise en œuvre (l'implémentation des lois de commande du procédé) et de tests.

Les travaux développés dans cette thèse s'inscrivent dans la lignée de cette volonté d'intégration de la biologie des systèmes. Nous proposons une approche alliant la conception de la souche et la conduite du procédé à proprement parlé. L'enjeu est de mettre en œuvre une optimisation qui prenne simultanément en compte les contraintes attachées aux deux étapes. Le but est de réduire les efforts d'intégration et de transition entre optimisation de souche et conduite du procédé. En effet, l'approche classique en deux temps laisse supposer qu'une partie de l'information se perd entre les travaux de conception et ceux de conduite du procédé. L'objectif de cette thèse est de proposer une manière de réduire cette perte d'information via une modélisation à l'interface entre les deux disciplines.

L'objectif est d'optimiser le procédé via le contrôle de l'environnement (conduite du procédé classique) et le comportement intracellulaire (optimisation de souche). Concrètement, nous nous plaçons du point de vue de la conduite du procédé et nous souhaitons proposer un modèle de procédé représentatif des processus intracellulaires. Pour cela, il est nécessaire de déterminer les fondements d'une représentation d'un bioprocédé par les processus intracellulaires des organismes impliqués dans le procédé macroscopique. L'émergence de modèles cellulaires autonomes semble permettre la modélisation et l'optimisation de part et d'autre. Le but de ces travaux est d'explorer cette voie.

Au cours des chapitres de cette thèse, nous présentons les éléments de construction d'un modèle à l'interface entre l'échelle intracellulaire et l'échelle macroscopique. Une telle modélisation permet en effet de donner une origine intracellulaire aux phénomènes observés à l'échelle macroscopique. Ainsi, il devient possible de pointer les processus intracellulaires limitant le procédé à l'échelle macroscopique. En d'autres termes, l'optimisation d'un procédé macroscopique ne passe plus uniquement par l'optimisation de paramètres macroscopiques mais aussi par l'optimisation de processus intracellulaires.

Ce mémoire de thèse se structure en quatre chapitres et deux annexes.

Chapitre 1, Démarches usuelles pour la production d'un composé par bioprocédé

Ce chapitre se focalise sur les techniques d'optimisation de bioprocédés. Il ne cherche pas l'exhaustivité mais parcourt de manière orientée différentes méthodes qui seront utiles dans la construction du modèle des chapitres suivants. Ainsi, on y présente des techniques usuelles de construction de modèles macroscopiques, d'utilisation de ces modèles pour l'optimisation de la conduite de procédé. On s'intéresse à des modèles intracellulaires de différentes natures ainsi qu'aux utilisations de ces modèles pour l'optimisation de la conception de souche. Dans ce chapitre, des tentatives de couplages entre modèles cellulaires et modèles de bioréacteurs sont aussi présentées. Leur but se rapproche du nôtre : optimiser la conduite de procédé en prenant en compte le comportement intracellulaire.

Chapitre 2, Modélisation et optimisation du comportement de micro organismes

Ce chapitre présente un exemple de modélisation *dynamical Resource Balance Analysis* (dRBA) qui est au cœur de cette thèse. L'exemple présenté dans ce chapitre est une représentation agrégée d'une cellule bactérienne. L'idée est de déterminer les contraintes structurelles de l'optimisation du comportement de micro-organismes comme des bactéries. Dans ce chapitre, un problème d'optimisation de conduite du procédé est proposé via le contrôle des processus intracellulaires, c'est-à-dire via une reconfiguration du comportement intracellulaire. A l'issue de ce chapitre, on propose une optimisation de la conduite d'un procédé en agissant à tout instant sur le comportement des processus intracellulaires.

Chapitre 3, Principe de conception rationnelle de souche

Ce chapitre poursuit les travaux du chapitre 2 en introduisant des contraintes de reconfiguration du comportement intracellulaire. On y introduit les notions de régulation génétique. L'objectif est de ne plus optimiser le procédé macroscopique en agissant sur les procédés intracellulaires mais sur les gènes de protéines catalysant les processus intracellulaires. Ce chapitre présente aussi une optimisation du procédé via une optimisation des conditions extracellulaires sous forme d'optimisation du débit d'un réacteur fed-batch.

Chapitre 4, Extension de l'approche à des descriptions cellulaires plus détaillées

Ce dernier chapitre est une ouverture vers une extension de la méthode à des réseaux de processus plus élaborés que le modèle agrégé des chapitres 2 et 3. Le principe de généralisation de la méthode y est présenté avant de l'appliquer à un réseau plus détaillé que le modèle agrégé, introduisant deux alternatives pour la dégradation du substrat. L'idée de cet exemple est de montrer que la méthode appliquée à un réseau plus réaliste donne des possibilités plus riches d'optimisation. En particulier, on réalise dans ce chapitre une optimisation conjointe de la conception de gènes et du débit d'alimentation d'un réacteur fed-batch.

Annexe A, Introduction au fonctionnement d'une cellule bactérienne

Cette annexe rassemble les éléments de biologie cellulaire nécessaires à la compréhension des développements de cette thèse. Elle s'adresse au lecteur curieux ou ayant besoin de quelques rappels.

Annexe B, Un exemple d'utilisation du principe du maximum de Pontryaguine

Dans cette annexe, on développe une méthode de caractérisation analytique d'un problème de conduite du procédé. L'application de cette méthode est présentée sur un modèle macroscopique standard de la littérature.

Les travaux présentés dans cette thèse ont donné lieu aux publications suivantes lors de conférences internationales avec actes :

- **ICSTCC 2016** : G. Jeanne, S. Tebbani, A. Goelzer, V. Fromion, D. Dumur. Modelling and optimizing of metabolic pathways in bacteria, *20th IEEE International Conference on System Theory, Control and Computing (ICSTCC)*, Sinaia, Roumanie, octobre 2016.
- **MED 2017** : G. Jeanne, S. Tebbani, A. Goelzer, V. Fromion, D. Dumur. Optimization of a micro-organisms culture in a fedbatch bioreactor using an intracellular model, *25th IEEE Mediterranean Conference on Control and Automation (MED)*, La Valette, Malte, juillet 2017.
- **FOSBE 2018** : G. Jeanne, A. Goelzer, S. Tebbani, D. Dumur, V. Fromion. Dynamical resource balance models for bioreactor optimization, *7th IFAC Conference on Foundation of Systems Biology in Engineering (FOSBE)*, Chicago, États-Unis, aout 2018.

Ils ont également été présentés lors de la conférence internationale sans acte :

- **Benelux Meeting 2018** : G. Jeanne, S. Tebbani, D. Dumur, A. Goelzer, V. Fromion. Integration of protein dynamics on batch bioreactor control, *37th Benelux Meeting on Systems and Control*, Soesterberg, Pays-Bas, mars 2018.

Un article a été soumis à la conférence internationale avec actes suivante :

- **CDC 2018** : G. Jeanne, A. Goelzer, S. Tebbani, D. Dumur, V. Fromion. Towards a realistic and integrated strain design in batch bioreactor, *57th IEEE Conference on Decision and Control*, Miami Beach, États-Unis, décembre 2018.

Chapitre 1

Démarches usuelles pour la production d'un composé par bioprocédé

Notre objectif est la production de composés d'intérêt par voie biologique, c'est-à-dire utilisant les mécanismes des organismes vivants pour transformer un substrat en un produit de plus grande valeur (y compris en biomasse). Ces techniques ne sont pas nouvelles. La fermentation alcoolique, par exemple, est un bioprocédé utilisé depuis des millénaires, pour produire entre autre, de la bière. Dès le IV^{ème} millénaire avant notre ère, les sumériens buvaient de la bière dont ils maîtrisaient la production. Cette transformation revient à utiliser des levures pour produire de l'alcool à partir de sucre.

Dans ce chapitre, nous allons voir un ensemble de démarches usuelles pour optimiser la bioproduction d'un composé. Nous commencerons par une approche macroscopique liée à la conduite du procédé, en section 1.1. Elle consiste à établir un modèle descriptif et prédictif du procédé à un niveau macroscopique. Puis nous utiliserons ces modèles pour déduire la stratégie optimale de fonctionnement et la structure de commande permettant de la suivre. Nous verrons que ces modèles macroscopiques peuvent devenir très complets mais qu'ils restent limités pour décrire le facteur le plus déterminant du procédé : les cellules.

Pour combler ce manque, nous verrons en section 1.2 une approche qui permet d'optimiser un bioprocédé en se concentrant avant tout sur les cellules. La compréhension et la modélisation des micro-organismes sont fondamentales pour la compréhension et l'optimisation du procédé. Nous parcourrons différentes techniques de modélisation, en partant de modèles systémiques quasi-mécaniques, en passant par des modèles sous contraintes pour finir sur des modèles autonomes qui allient contraintes mécaniques (d'implémentation) et contraintes d'optimisation. Nous verrons également comment ces modèles peuvent être utilisés pour optimiser des souches et que nous pouvons envisager de les coupler à des réacteurs. Nous terminerons ce chapitre en établissant un cahier des charges pour un modèle qui optimise le procédé en agissant sur son environnement et sur son comportement intracellulaire.

1.1 Approche « conduite du procédé »

Il s'agit de comprendre de quelles manières on peut modéliser et utiliser un modèle macroscopique de bioréacteur pour optimiser un bioprocédé.

Optimiser un bioprocédé, c'est maximiser un critère portant sur la production de biomasse et/ou d'un composé d'intérêt. On s'intéressera plus particulièrement à la bio-production de composés d'intérêt.

Tous les bioprocédés que nous présentons s'opèrent en bioréacteur, c'est-à-dire dans une phase aqueuse à laquelle il est possible d'ajouter ou de soutirer du milieu de culture. On considère les modes de fonctionnement :

- batch : pas d'ajout, ni de soutirage de milieu réactionnel,
- fed-batch : ajout de milieu réactionnel au cours du temps,
- continu : ajout et soutirage de milieu réactionnel égaux à tout instant, de sorte que le volume réactionnel ne varie pas.

Nous ne reprenons pas l'intégralité des techniques et des tentatives d'optimisation de bioproduction, d'autant plus que la littérature est très riche dans cette thématique. Nous nous focalisons sur l'aspect modélisation, optimisation et commande, au sens de l'automatique. Les approches les plus usuelles dans la littérature sont présentées ici en accord avec ce qui est développé dans les travaux de cette thèse.

1.1.1 Modèles macroscopiques phénoménologiques

Pour un même bioprocédé (ou d'une manière générale pour un même objet), il existe autant de manière de le modéliser qu'il y a de modélisateurs. Tout modèle est approximatif, on doit simplement connaître le niveau d'approximation atteint. Et en effet, un modèle s'accompagne d'un domaine de validité et d'une finalité. Si l'on perd de vue la finalité, on perd de vue le domaine de validité et on risque d'utiliser le modèle à tort.

Les finalités de modélisation peuvent être diverses. On peut, par exemple, établir un modèle pour représenter ou résumer un phénomène connu, pour proposer une origine à un phénomène, ou pour prédire un phénomène. La finalité pour laquelle on passe par l'élaboration d'un modèle peut aussi être variée. On peut souhaiter agir sur le phénomène que l'on modélise, détecter un écart à un comportement modélisé, retrouver une ou plusieurs variables en transposant la modélisation, ou tout autre dessein propre au modélisateur.

En tant que modélisateurs de bioprocédés, nous abordons la question des modèles en cherchant ceux qui visent à agir sur le bioprocédé par des techniques d'automatique. Nous allons également rencontrer des modèles qui n'ont pas été conçus pour cette finalité.

Il y a deux composantes principales aux modèles de bioprocédés, le bilan de matière (il rend compte des échanges de matière entre les composés), et les expressions mathématiques des échanges (elles caractérisent les différents échanges en fonction des composés.)

1.1.1.1 Schémas réactionnels & bilans de matière

Exemple introductif. Nous allons définir les notions de schéma réactionnel et de bilan de matière au travers d'exemples.

Le premier exemple canonique est le cas d'une population de micro-organismes, $\mathcal{X}$, qui croît en consommant un substrat unique, $\mathcal{G}$, on peut écrire ceci comme :



c'est une première manière de modéliser le phénomène de croissance.

Néanmoins, cette représentation ne rend pas compte de la conservation de la masse. Pour cela, il est nécessaire d'introduire un coefficient de rendement, α , le coefficient de conversion entre les deux entités, pour signifier que α moles (ou molécules) de $\mathcal{G}$ sont nécessaires pour créer une mole (ou une cellule) de $\mathcal{X}$:



On note au-dessus de la flèche la vitesse de réaction, r . On obtient le schéma réactionnel suivant :

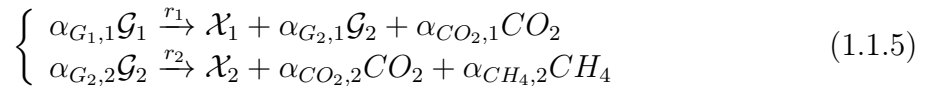


On n'a pas besoin de préciser ce que vaut la vitesse de réaction, r . On introduit X et G les concentrations de $\mathcal{X}$ et $\mathcal{G}$ dans leur milieu respectif, un bilan de matière permet d'aboutir aux équations différentielles suivantes [4] :

$$\begin{cases} \dot{G}(t) = -\alpha r(t) \\ \dot{X}(t) = +1 r(t) \end{cases} \quad (1.1.4)$$

Cela suppose le système isolé (*i.e.* sans aucun échange avec l'extérieur), que le milieu est homogène, et que seule cette réaction (1.1.3) a une influence sur G et X .

Exemple plus avancé. Donnons un exemple, adapté de [21], qui présente un système de traitement des eaux par digestion anaérobie. Dans ce contexte, un substrat $\mathcal{G}_1$ est consommé par une espèce bactérienne $\mathcal{X}_1$ pour croître à la vitesse r_1 . Cette croissance est accompagnée d'une production d'un composé $\mathcal{G}_2$ et de CO_2 . Dans le même temps, une seconde espèce bactérienne $\mathcal{X}_2$ consomme le composé $\mathcal{G}_2$ pour croître à la vitesse r_2 . Sa croissance est accompagnée d'une production de CO_2 et de CH_4 . Les schémas réactionnels suivants résument les deux réactions de croissance :



avec $\alpha_{\xi,1}$ et $\alpha_{\xi,2}$ les coefficients de rendement assurant la conservation des différentes espèces.

En supposant le système isolé et que les gaz restent dissous dans la phase aqueuse, en notant C la concentration en CO_2 dissous et M la concentration en méthane (CH_4) dissous, X_1 , X_2 , G_1 et G_2 respectivement les concentrations en $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{G}_1$ et $\mathcal{G}_2$, alors le

système de réactions (1.1.5) est équivalent au bilan de matière présenté sous forme du système d'équations différentielles suivant :

$$\begin{cases} \dot{X}_1(t) = r_1(t) \\ \dot{X}_2(t) = r_2(t) \\ \dot{G}_1(t) = -\alpha_{G_1,1}r_1(t) \\ \dot{G}_2(t) = \alpha_{G_2,1}r_1(t) - \alpha_{G_2,2}r_2(t) \\ \dot{C}(t) = \alpha_{CO_2,1}r_1(t) + \alpha_{CO_2,2}r_2(t) \\ \dot{M}(t) = \alpha_{CH_4,2}r_2(t) \end{cases} \quad (1.1.6)$$

On peut alors introduire la formulation matricielle générale :

$$\dot{\xi}(t) = Ar(t) \quad (1.1.7)$$

où ξ est le vecteur des concentrations des composés en présence, A la matrice des coefficients de rendement, r le vecteur des vitesses de réaction.

$$\text{Dans l'exemple (1.1.6), } \xi(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ G_1(t) \\ G_2(t) \\ C(t) \\ M(t) \end{pmatrix}, A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\alpha_{G_1,1} & 0 \\ \alpha_{G_2,1} & -\alpha_{G_2,2} \\ \alpha_{CO_2,1} & \alpha_{CO_2,2} \\ 0 & \alpha_{CH_4,2} \end{pmatrix}, r(t) = \begin{pmatrix} r_1(t) \\ r_2(t) \end{pmatrix}.$$

Avec ce formalisme, plus on souhaite rendre compte d'un nombre important de phénomènes et de l'évolution d'un nombre important d'espèces, plus on ajoute de réactions ou on fait figurer d'éléments dans les réactions.

Nous ne nous préoccupons pas des échanges gazeux dans ces travaux, mais on peut aussi introduire le dioxyde de carbone présent à l'état de gaz par opposition au dioxyde de carbone dissous en phase aqueuse, en ajoutant le schéma réactionnel :



avec q_C la vitesse de transfert du CO_2 de la phase liquide vers la phase gazeuse. Habituellement, on introduit un vecteur $q(t)$ d'échanges gazeux que l'on ajoute à la formulation (1.1.7).

Nous ne développerons pas davantage la notion de gaz dans ces travaux.

Ajout et retrait de matière. Si le système n'est pas considéré isolé (échange de matière avec l'extérieur), il faut introduire des facteurs qui prennent en compte ces échanges dans le bilan de matière.

Supposons que le réacteur soit alimenté en milieu de culture par un débit F_{in} (en L/h). Dans ce milieu, les concentrations en composés Ξ sont notées ξ_F . Supposons que l'on soutire du milieu de culture avec un débit F_{out} (en L/h). Sous couvert que le milieu soit homogène, on peut compléter l'équation (1.1.7) comme suit pour décrire le système :

$$\frac{d(\xi V)}{dt}(t) = \underbrace{Ar(t)V(t)}_{\text{échanges}} + \underbrace{\xi_F F_{in}(t)}_{\text{ajout}} - \underbrace{\xi(t)F_{out}(t)}_{\text{retrait}} \quad (1.1.9)$$

avec $V(t)$ le volume effectif du bioréacteur qui varie lui-même selon :

$$\frac{dV}{dt}(t) = F_{in}(t) - F_{out}(t) \quad (1.1.10)$$

Comme $\xi(t)$ est le vecteur des concentrations en éléments Ξ par rapport au volume en fonction du temps, l'équation (1.1.9) représente l'évolution temporelle des quantités de matières (échanges de matière dus aux réactions, apport par le débit d'alimentation et retrait par le débit de soutirage).

En ramenant aux variations de concentrations, on obtient :

$$\frac{d(\xi V)}{dt}(t) = \frac{d\xi}{dt}(t)V(t) + \xi(t)\frac{dV}{dt}(t)$$

i.e.

$$\frac{d(\xi V)}{dt}(t) = \frac{d\xi}{dt}(t)V(t) + \xi(t)(F_{in}(t) - F_{out}(t))$$

L'équation (1.1.9) devient alors :

$$\frac{d\xi}{dt}(t)V(t) + \xi(t)(F_{in}(t) - F_{out}(t)) = Ar(t)V(t) + \xi_F F_{in}(t) - \xi(t)F_{out}(t)$$

ou encore,

$$\frac{d\xi}{dt}(t) = Ar(t) + (\xi_F - \xi(t))\frac{F_{in}(t)}{V(t)} \quad (1.1.11)$$

Il peut paraître surprenant que le débit soutiré F_{out} n'intervienne pas directement dans l'évolution de la concentration de ξ . En fait : le soutirage est indépendant de la nature du composé. Le fait de soutirer de la matière ne change pas les concentrations des espèces en présence.

Pour conclure, on voit que l'on dispose d'un moyen de modéliser plus ou moins finement le procédé, en distinguant les espèces en présence et en détaillant les schémas réactionnels, comme on l'a évoqué avec l'exemple du dioxyde de carbone aqueux et gazeux.

1.1.1.2 Expressions mathématiques des vitesses de croissance

Notion de taux spécifique de croissance. Le second pan de la modélisation de bioprocédés est de donner une expression aux vitesses des réactions, *i.e.* au vecteur r de l'équation (1.1.7). Nous allons chercher à donner à $r(t)$ une expression en fonction des concentrations, $\xi(t)$, en partant du principe que ce sont les concentrations qui influencent les vitesses de réaction (comme c'est le cas dans toute la théorie de la chimie) : $r(t) = r(\xi(t))$.

Intéressons-nous dans un premier temps au phénomène de croissance.

Dès lors qu'une espèce est un des réactifs d'une réaction, (*i.e.* il est nécessaire pour que la réaction se produise), il est possible de mettre en facteur sa concentration dans la vitesse de la réaction. On tient ainsi compte du fait qu'en l'absence de réactif, il n'y a pas de réaction.

L'exemple canonique concerne la croissance. En notant $r_{\text{croissance}}$ la vitesse de croissance d'un organisme $\mathcal{X}$ de concentration $X(t)$ à l'instant t , i.e. $\frac{dX}{dt}(t) = r_{\text{croissance}}$, il vient la factorisation suivante :

$$r_{\text{croissance}}(\xi(t)) = \mu(\xi(t))X(t) \quad (1.1.12)$$

où μ est, par définition, ce que l'on appelle le taux spécifique de croissance.

On a ainsi,

$$\frac{dX}{dt}(t) = \mu(\xi(t))X(t) \quad (1.1.13)$$

et on retrouve la définition qui est parfois donnée pour le taux spécifique de croissance, comme étant la dérivée logarithmique de la concentration en $\mathcal{X}$,

$$\mu(\xi(t)) = \frac{1}{X(t)} \frac{d(X(t))}{dt} = \frac{d \ln(X(t))}{dt}$$

Formule de Monod. Les travaux de Monod portent, en particulier, sur la caractérisation mathématique de taux spécifique de croissance sur un unique substrat.

Dans ses travaux, Monod [60] mesure pour différentes concentrations de divers substrats (glucose, fructose, mannose, ...) l'augmentation massique de colonies de différentes espèces bactériennes (*B. subtilis*, *E. coli*, ...). Ses travaux ont montré que le taux spécifique de croissance augmente avec la concentration en substrat et que ce taux est majoré par une valeur maximale.

Mathématiquement, en notant G la concentration en substrat, les deux hypothèses s'écrivent :

$$\begin{cases} \text{Croissance nulle si } G = 0 & : \mu(G) \propto G \\ \text{Croissance suprémale quand } G \text{ tend vers l'infini} & : \mu(G) \underset{G \rightarrow +\infty}{\sim} \mu_{\max} \end{cases} \quad (1.1.14)$$

Ainsi, en introduisant $\bar{\mu}(G) \triangleq \mu(G)/G$,

$$\begin{cases} \mu(G) = G\bar{\mu}(G) \\ G\bar{\mu}(G) \underset{G \rightarrow +\infty}{\sim} \mu_{\max} \end{cases} \quad (1.1.15)$$

Comme $\mu_{\max}$ est strictement positif (sinon, μ est nulle pour tout G), on peut écrire :

$$\frac{1}{G\bar{\mu}(G)} \underset{G \rightarrow +\infty}{\sim} \frac{1}{\mu_{\max}}$$

ou encore,

$$\begin{aligned} \frac{\mu_{\max}}{G\bar{\mu}(G)} &= 1 + \underset{G \rightarrow +\infty}{o(1)} \\ \frac{\mu_{\max}}{\bar{\mu}(G)} &= G + \underset{G \rightarrow +\infty}{o(G)} \end{aligned}$$

et alors,

$$\frac{\bar{\mu}(G)}{\mu_{\max}} = \frac{1}{G + \underset{G \rightarrow +\infty}{o(G)}}$$

$$\bar{\mu}(G) = \frac{\mu_{max}}{G + \lim_{G \rightarrow +\infty} o(G)} \quad (1.1.16)$$

Pour satisfaire cette équation, il suffit de prendre : $\lim_{G \rightarrow +\infty} o(G) = K_G = cste$. C'est typiquement ce qui est supposé par la formule de Monod dans la formulation hyperbolique suivante :

$$\mu_{Monod}(G) = \frac{\mu_{max}G}{G + K_G} \quad (1.1.17)$$

avec μ_{max} le taux spécifique de croissance maximal, G la concentration en $\mathcal{G}$ et K_G la concentration de $\mathcal{G}$ pour laquelle le taux spécifique de croissance vaut la moitié de sa valeur maximale, ($\mu_{Monod}(K_G) = \mu_{max}/2$) est appelée constante de demi-saturation. Un exemple d'allure de taux spécifique de croissance de Monod est donné figure 1.1.

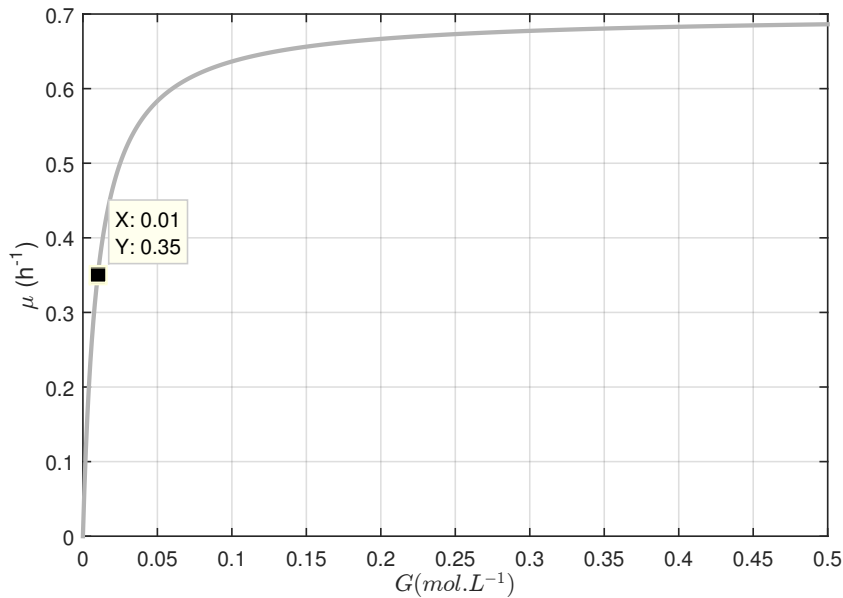


FIGURE 1.1 – Allure d'un taux spécifique de croissance selon une relation de Monod par rapport à la concentration G . $\mu_{max} = 0,7h^{-1}$ et $K_G = 0,01mol.L^{-1}$.

Dès l'introduction de cette formule qui prit son nom, Monod indique que la constante μ_{max} est définie avec précision, tandis que pour K_G , "l'expérience ne fournit qu'un ordre de grandeur".

Cette expression est l'expression la plus répandue pour caractériser la croissance d'une espèce sur un unique substrat, ici $\mathcal{G}$.

Par la suite, elle a été étendue à d'autres organismes sur différents substrats.

L'intérêt de cette formulation est sa simplicité. De plus, elle intègre le fait que la croissance ne peut se faire à concentration nulle en substrat et que la croissance est bornée et tend asymptotiquement vers sa valeur maximale. La limitation de la croissance par le substrat est ainsi modélisée.

Il devient possible de raffiner l'expression du taux spécifique de croissance pour prendre en compte d'autres phénomènes plus fins que la simple limitation. Pour ce faire, on modifie l'expression prise pour $\mu(G)$.

Plus on veut prendre en compte de phénomènes, comme par exemple (1.1.14), plus on devra complexifier l'expression du taux spécifique de croissance.

Nous allons présenter quelques phénomènes et la manière dont ils sont modélisés.

Inhibition par un produit secondaire. Il a été observé des phénomènes d'inhibition du taux spécifique de croissance par les produits secondaires. Ceci peut être pris en compte dans la formulation du $\mu(G)$ dans la relation (1.1.16).

Par exemple, en notant C la concentration d'un composé secondaire qui est connu pour limiter la croissance et en prenant $\mu(G) = \frac{\mu_{max}G}{G + K_G}$ (avec K_G et K_C des paramètres à déterminer) dans la relation (1.1.16), il vient :

$$\mu_{limitation}(G, C) = \frac{\mu_{max}}{G + C/K_C + K_G} \quad (1.1.18)$$

Ceci ne change pas l'adéquation aux observations qui sont faites dans (1.1.14) si l'on se place à une concentration fixe en composé C (celle-ci passe alors dans la catégorie des paramètres).

Par ailleurs, à concentration G fixée, le taux de croissance est décroissant en C et tend même vers 0 lorsque C tend vers l'infini.

Biologiquement, le composé inhibant peut être de l'acétate, du lactate ou tout autre composé connu pour bloquer la croissance.

Inhibition par la biomasse. Sur la même idée d'une croissance inhibée par une grandeur du modèle, on peut citer le modèle de Contois [20] qui permet de représenter une inhibition de la croissance par la concentration en biomasse X :

$$\mu_{Contois}(G, X) = \frac{\mu_{max}G}{G + K_G X} \quad (1.1.19)$$

On note qu'ici, le $\mu(G)$ est choisi égal à $\frac{\mu_{max}G}{G + K_G X}$ (avec K_G paramètre à déterminer) qui est négligeable devant G lorsque G tend vers $+\infty$, toute chose restant égale par ailleurs. Il s'agit bien d'une extension des observations de Monod.

On exprime ainsi le fait que le taux spécifique de croissance chute avec la concentration de la population. De cette manière on représente la difficulté d'accès aux ressources, qui est supposée augmenter avec la concentration en biomasse.

Inhibition par le substrat. Des études faites à forte concentration de substrat complètent les observations de Monod (1.1.14). On ajoute l'hypothèse que le taux spécifique de croissance est croissant en G pour des valeurs faibles de concentration puis décroissant pour des valeurs élevées de concentration [23].

Cette inhibition par le substrat lui-même peut se modéliser par une expression dite de Andrews donné par la formule suivante (elle est couramment appelée à tort relation de Haldane alors que celle-ci est réservée aux expressions de taux de production, voir infra) :

$$\mu_{Andrews}(G) = \frac{\mu_{max}G}{G + K_G + \frac{G^2}{K_i}} \quad (1.1.20)$$

avec μ_{max} , K_G et K_i des constantes positives.

L'allure de cette fonction est donnée figure 1.2. On peut noter que μ_{max} n'est pas le taux spécifique de croissance maximal que peut atteindre le taux spécifique de croissance. On peut noter qu'à faibles concentrations G , le taux spécifique de croissance est très similaire à un modèle de Monod, mais lorsque G augmente, le taux spécifique de croissance diminue jusqu'à s'annuler lorsque G tend vers l'infini.

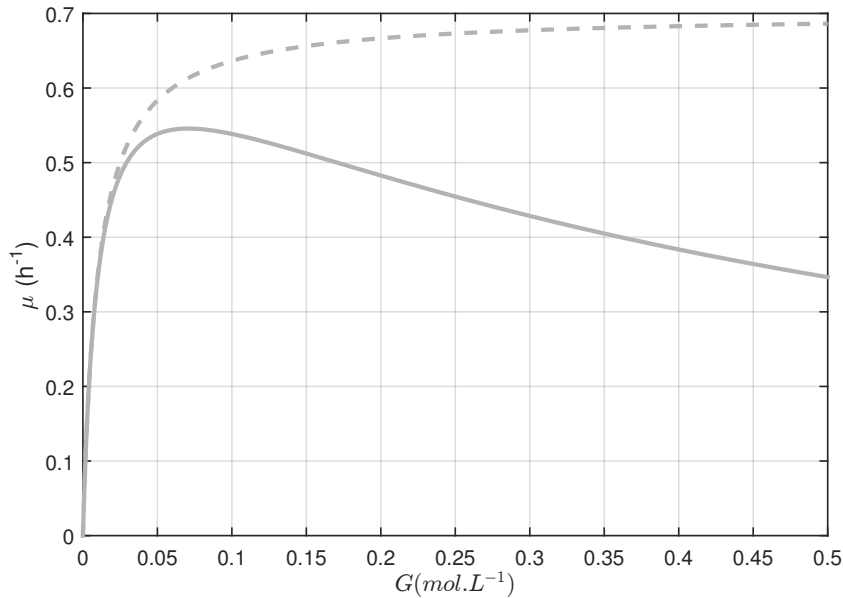


FIGURE 1.2 – Allure d'un taux spécifique de croissance qui suit une relation de Andrews par rapport à la concentration G . $\mu_{max} = 0,7 h^{-1}$, $K_G = 0,01 mol.L^{-1}$ et $K_i = 0,5 mol.L^{-1}$. Comparaison avec une allure de Monod de paramètres identiques ($\mu_{max} = 0,7 h^{-1}$ et $K_G = 0,01 mol.L^{-1}$), en tirets.

Cette observation montre qu'en présence de trop de substrat, les cellules changent de fonctionnement et finissent par s'écarter d'un comportement qui les mène à utiliser les ressources disponibles de manière efficace.

Maintenance cellulaire. Pour finir on peut introduire dans les expressions littérales, le phénomène de maintenance cellulaire, c'est-à-dire le fait qu'il y ait en permanence besoin de ressources pour assurer d'autres tâches que la croissance cellulaire.

C'est ce qui est pris en compte dans le modèle de Herbert qui pose pour formule générale du taux spécifique de croissance [77] :

$$\mu_{Herbert}(G) = \max(\mu(G) - m, 0) \quad (1.1.21)$$

avec $\mu(G)$ formulation initiale pour le taux spécifique de croissance spécifique et m paramètre de maintenance positif. m représente un détournement des ressources de la croissance vers les mécanismes de maintenance, de sorte que pour une concentration G inférieure à une certaine constante G_m , la croissance soit nulle.

m est homogène à un taux spécifique de croissance, c'est-à-dire à un flux de matière par unité de biomasse. Ce paramètre du modèle peut être vu comme le flux de matière qui est nécessaire à la maintenance cellulaire, par unité de biomasse.

La figure 1.3 donne un exemple de taux spécifique de croissance de Monod et de taux spécifique de croissance de Herbert fondé sur un Monod. On note que pour $G \leq 0,01 \text{ mol.L}^{-1}$, il n'y a pas de croissance.

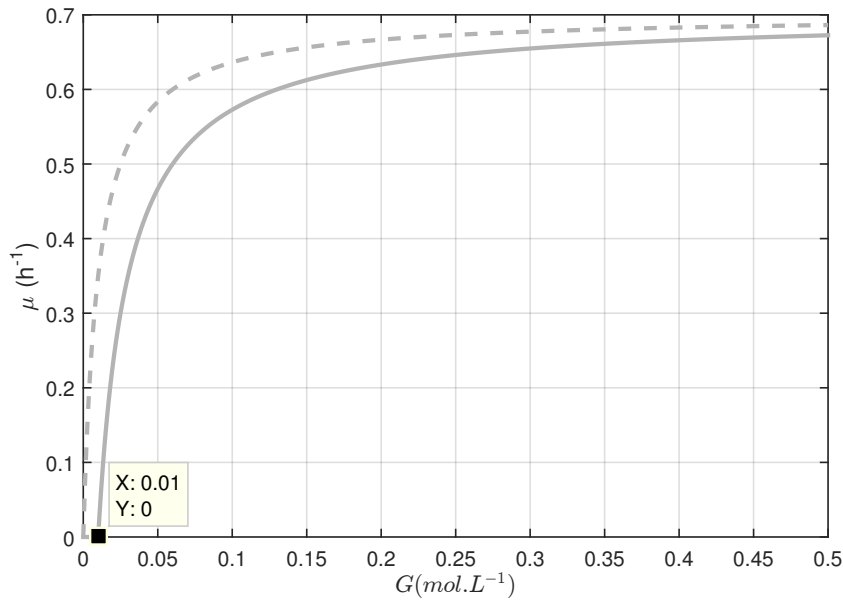


FIGURE 1.3 – Allure d'un taux spécifique de croissance qui suit une relation de Herbert par rapport à la concentration G . La formulation initiale de croissance est une formulation de Monod avec $\mu_{max} = 1,4 \text{ h}^{-1}$ et $K_G = 0,01 \text{ mol.L}^{-1}$. Le paramètre de maintenance est $m = 0,7 \text{ h}^{-1}$. Comparaison avec une allure de Monod de paramètres $\mu_{max} = 0,7 \text{ h}^{-1}$ et $K_G = 0,01 \text{ mol.L}^{-1}$, en tirets.

Résumé. En résumé, le taux spécifique de croissance d'un organisme par une réaction de production de biomasse est donné par le produit de plusieurs facteurs représentant l'activation, la limitation, la saturation et l'inhibition. Les principaux termes que l'on trouve dans la littérature sont résumés dans le tableau 1.1.

Si deux sources sont nécessaires à la croissance d'un organisme, une source de carbone et une source d'azote par exemple, et que la croissance est saturée lorsque leur concen-

TABLE 1.1 – Quelques phénomènes que l'on peut modéliser à travers l'expression du taux spécifique de croissance. Le composé dont il est question dans la première colonne du tableau a pour concentration ξ . On se place dans le cas où le taux spécifique de croissance est une fraction rationnelle des concentrations en présence. Il est envisageable de raffiner les expressions en faisant appel à d'autres structures de formulations (typiquement des formulations avec des exponentielles).

Phénomène	Expression	[Source]
Nécessité d'un composé	$\mu = \cdots \times \xi$	[21]
Saturation	$\mu = \cdots \times \frac{\xi}{\cdots + \xi}$	
Inhibition par un composé	$\mu = \cdots \times \frac{1}{\xi + K_i}$	[21]
	$\mu = \frac{\cdots}{\cdots + \xi/K_i}$	[1]
Maintenance	$\mu(\xi) = \max(\mu(\xi) - m, 0)$	[77]

tration est trop importante (*i.e.* qu'il y a des phénomènes de saturation) on pose comme expression pour le taux spécifique de croissance :

$$\mu_{\text{bisource}}(C, N) = \mu_{\text{max}} \frac{C}{C + K_C} \frac{N}{N + K_N} \quad (1.1.22)$$

où C et N sont respectivement les concentrations en sources carbonée et azotée, et μ_{max} , K_C et K_N sont des constantes à déterminer.

1.1.1.3 Expressions mathématiques des autres éléments des schémas réactionnels

Cinétiques de production. De la même manière, on joue sur l'expression des différents taux de réaction en partant de considérations biologiques de nécessité, de saturation, d'inhibition, *etc.* pour élaborer les expressions mathématique des cinétiques de production.

Ainsi, pour une réaction de bioproduction d'un composé $\mathcal{P}$, $\mathcal{G} \xrightarrow{p} \mathcal{P}$, catalysé par la biomasse $\mathcal{X}$ de concentration X , on écrit :

$$\frac{dP}{dt} = +p = +\pi(\cdot)X \quad (1.1.23)$$

avec P concentration du composé $\mathcal{P}$, p vitesse de production, et π taux spécifique de production (ramené par unité de biomasse).

On exprime le fait que π dépend de plusieurs phénomènes en complexifiant son expression.

Par exemple, pour de nombreux composés, on remarque que le taux spécifique de production suit une relation proche de celle introduite figure 1.2, c'est-à-dire avec une concentration extérieure en substrat optimale pour la production.

On introduit la fonction de Haldane pour la production. C'est la même expression que la relation de Andrews (1.1.20) qui est réservée à l'expression de la croissance :

$$\pi(G) = \frac{\pi_{max}G}{G + K_G + G^2/K_i} \quad (1.1.24)$$

On prend ainsi en compte qu'il existe une concentration de $\mathcal{G}$ pour laquelle le taux de production par unité de biomasse (ou par cellule) est maximal. Cette concentration particulière, G_{max} , est la racine de la dérivée de la fonction $\pi(\cdot)$:

$$0 = \frac{d\pi}{dG} = \pi_{max} \frac{(G + K_G + G^2/K_i) - G(1 + 2G/K_i)}{(G + K_G + G^2/K_i)^2}$$

d'où,

$$K_G - G^2/K_i = 0$$

ou encore,

$$G = \sqrt{K_i K_G} \quad (1.1.25)$$

Rendement des procédés. Un raffinement supplémentaire est apporté aux modèles en faisant varier l'efficacité des procédés en fonction de paramètres extérieurs. Ceci revient à modifier substantiellement le bilan de matière $\alpha \cdot \mathcal{G} \rightarrow 1 \cdot \mathcal{X}$ en

$$\alpha(G, X, \dots) \cdot \mathcal{G} \rightarrow 1 \cdot \mathcal{X}$$

On pose alors que le nombre de moles de substrat $\mathcal{G}$ nécessaire à la formation d'une mole de biomasse $\mathcal{X}$ n'est pas constant. Il évolue.

C'est une subtilité de modélisation discutée dans [1].

Biologiquement, on essaie de rendre compte qu'en fonction des conditions extérieures, la croissance est plus ou moins efficace. Par exemple il est courant de distinguer la croissance aérobie et anaérobie pour une même espèce et un même substrat.

1.1.1.4 Conclusion

Les modèles que l'on a parcourus dans cette section décrivent les variables macroscopiques du bioprocédé et en particulier modélisent le comportement des micro-organismes en fonction de leur environnement. Dès lors, on envisagera le contrôle de l'environnement pour agir sur le comportement des micro-organismes. C'est ce que nous allons voir dans la section suivante.

1.1.2 Utilisation des modèles pour la commande des bioprocédés en vue de la production d'un composé d'intérêt

Nous allons parcourir un certain nombre de techniques qui proviennent du monde de l'automatique et qui ont fait leurs preuves dans cette discipline.

La philosophie de l'automatique, consiste à diviser les grandeurs physiques en 4 catégories :

- les paramètres, fixes par nature et sur lesquels on n'a pas d'influence,
- les variables de commande, qui évoluent au cours du temps et sur lesquelles on peut agir,
- les perturbations, qui varient au cours du temps et sur lesquelles on n'a pas d'influence,
- les variables d'état, qui évoluent au cours du temps en fonction des paramètres, des variables de commande et des perturbations. Les variables d'état caractérisent le système à un instant donné. On peut essayer d'agir sur elles par le biais des variables de commandes. Parmi les variables d'état, on peut distinguer les mesures dont on dispose que l'on appelle les variables de sortie. On les distingue des variables d'état dont les valeurs ne sont pas toujours accessibles.

L'esprit de la commande d'un système revient à trouver les variables de commande adéquates pour que les variables d'état ou de sortie atteignent un objectif désiré.

Avec le formalisme de l'automatique, les problèmes d'optimisation de bioprocédé s'écrivent de la manière suivante :

$$\begin{aligned} \max_{u(t), t_f} J = \phi(x(t_f), t_f) + \int_{t_0}^{t_f} \psi(x(t), u(t), t) dt \\ \text{sous contraintes, } \begin{cases} \frac{dx}{dt}(t) = f(x(t), u(t), t) \\ x(t_0) = x_0 \\ h(x(t), u(t), t) \geq 0, t \in [t_0, t_f] \\ g(x(t_f), t_f) \geq 0 \end{cases} \end{aligned} \quad (1.1.26)$$

- J est le critère à maximiser, décomposé en un critère terminal, ϕ , et un critère intégral, ψ ,
- t_0 est l'instant initial, supposé fixe,
- t_f est l'instant final du procédé, qui est a priori un paramètre que l'on cherche à déterminer,
- x est le vecteur des variables d'état du système,
- u est le vecteur des variables de commande du système,
- f est la dynamique du système, c'est en particulier le modèle du système comme on l'a vu à la section précédente,
- x_0 est le vecteur des variables d'état à l'état initial, supposé connu,
- h regroupe les contraintes sur l'état et la commande à respecter à chaque instant,
- g regroupe les contraintes à l'état final.

A noter qu'en choisissant judicieusement g et h , on peut définir dans ce cadre des contraintes d'égalité. Typiquement, la conjonction des contraintes $x \geq 3$ et $-x \geq -3$ est

équivalente à $x = 3$.

Une première étape dans la recherche de la trajectoire optimale est de définir judicieusement les fonctions h , g , ϕ et ψ .

Les éléments de (1.1.26) diffèrent selon le mode de fonctionnement. On donnera ici quelques exemples en fonction du mode de fonctionnement batch, fed-batch et continu.

La dynamique du système est en grande partie la même d'un mode de fonctionnement à l'autre. Les variables d'état du système sont les concentrations des espèces en présence, et éventuellement le volume du milieu réactionnel pour le fed-batch et le mode continu.

Pour la définition de f , on reprend les éléments des relations (1.1.7), complétés dans le cas du fed-batch et du continu par (1.1.10) (pour donner (1.1.11)). Il est possible de compléter davantage la description du système par des variables d'état auxiliaires, comme la consommation de substrat ou l'évolution d'une quantité secondaire.

Les commandes sont le débit d'alimentation F_{in} en mode de fonctionnement fed-batch, complété éventuellement par le débit de soutirage F_{out} en continu. Il n'y a pas de variables de commande en fonctionnement batch, car par définition il s'agit d'un mode de fonctionnement autonome. L'optimisation porte alors seulement sur le temps final t_f et éventuellement les concentrations initiales des composés.

Les contraintes h sont des contraintes à respecter à tout instant. En fed-batch, on fait figurer le fait que la commande F_{in} est bornée et que le volume du milieu est limité par un volume maximal :

$$\begin{cases} F_{min} \leq F_{in}(t) \leq F_{max} \\ V(t) \leq V_{max} \end{cases} \quad (1.1.27)$$

La première contrainte porte sur la commande et la seconde porte sur l'état.

En mode de fonctionnement continu, la contrainte sur la commande suit la même formule. Dans certains cas, on limitera également la concentration d'une espèce particulière comme la biomasse.

Les contraintes g portent sur l'état final. Il s'agit de contraintes sur l'état final ou de contraintes sur la valeur t_f du temps final. On peut ainsi contraindre t_f à valoir une certaine valeur, ou à ne pas dépasser une durée maximale. En mode de fonctionnement fed-batch, on peut formuler le fait que l'on souhaite consommer une certaine quantité de substrat ou produire une certaine quantité de produit. En mode de fonctionnement batch, on introduit ici le fait qu'à la fin de la culture tout le substrat est consommé.

Le vecteur d'état à l'instant initial, x_0 , est donné ici comme un paramètre du système mais on peut l'ajouter aux variables d'optimisation. En mode de fonctionnement batch et fed-batch, on peut, par exemple, optimiser la concentration initiale en substrat.

Enfin, le critère J est la grandeur à optimiser. On constate que dans la formulation (1.1.26) J comprend une composante à l'instant final, ϕ , et une composante intégrale, $\int \psi$. $\int \psi$ permet de contraindre une concentration ou une commande à être toujours la plus petite ou la plus grande possible. La composante à l'instant final reflète un critère de productivité. Ceci peut être la concentration ou la quantité de produit d'intérêt, éventuellement ramené par unité de temps, par unité de substrat consommé, ou par unité de produit secondaire.

Après avoir traité de la philosophie de l'automatique on s'intéressera à la bioproduction d'un composé d'intérêt en mode de fonctionnement fed-batch. On traitera de la stratégie à adopter pour maximiser la production d'un composé d'intérêt.

Öztürk et al. présentent différents types de stratégies mises en œuvre pour la culture de *B. subtilis* en fed-batch [66]. Ils présentent la stratégie :

- par impulsion,
- par flux d'alimentation constant,
- par flux exponentiel,
- par flux linéaire,
- par optimisation hors-ligne
- par commande en boucle fermée.

Ce sont à ces deux dernières que nous allons nous intéresser pour la maximisation d'un certain critère de productivité d'un composé d'intérêt car elles englobent toutes les autres, d'un point de vue théorique. De plus, ces deux dernières approches sont les plus utilisées dans la littérature.

Dans la suite, comme il est commun en automatique, on caractérisera tout d'abord la trajectoire optimale au sens d'un certain critère. Cette caractérisation aura lieu hors-ligne à partir d'un modèle du système. Puis, dans un deuxième temps, on cherchera un moyen pour que le système réel (et non le modèle) suive cette trajectoire optimale. Pour cela, nous bouclerons le système, c'est-à-dire que l'on définira la commande à lui appliquer en fonction de son comportement réel.

1.1.2.1 Détermination hors-ligne de la trajectoire optimale

La première étape de l'optimisation d'un procédé consiste à chercher la trajectoire du système qui maximise un certain critère.

Formulation sous forme du principe du maximum de Pontryaguine. La formulation de Pontryaguine permet de déterminer l'ensemble des conditions nécessaires que la solutions optimale doit satisfaire (1.1.26). Cette méthode est présentée en détail dans [38], [81] ou [15]. L'annexe B présente les notations standards et une utilisation canonique des conditions nécessaires d'optimalité de Pontryaguine pour la résolution analytique d'un problème type (1.1.26).

Le principe du maximum de Pontryaguine affirme que la solution au problème (1.1.26) est nécessairement solution du problème suivant :

$$\begin{aligned}
\max_{u(t), t_f} \mathcal{L}(t) &= \lambda_0 \psi(x(t), u(t), t) + \lambda(t)^\top f(x(t), u(t), t) + \eta(t)^\top h(x(t), u(t), t) \\
\text{sous contraintes, } &\left\{ \begin{array}{l} \frac{dx}{dt}(t) = f(x(t), u(t), t) \\ x(t_0) = x_0 \\ h(x(t), u(t), t) \geq 0, t \in [t_0, t_f] \\ g(x(t_f), t_f) \geq 0 \\ \frac{d\lambda}{dt}(t) = -\frac{\partial \mathcal{L}}{\partial x}(x^*(t), \nu^*(t), \lambda(t), \eta(t)) \\ \lambda(t_f) = \frac{\partial \phi}{\partial x}(t_f) + \alpha^\top \frac{\partial g}{\partial x}(t_f) \\ \lambda_0 \geq 0 \\ \eta(t) \geq 0, t \in [t_0, t_f] \\ \alpha \geq 0 \\ \eta(t)h(x(t), u(t), t) = 0, t \in [t_0, t_f] \\ \alpha g(x(t_f), t_f) = 0 \\ h(x(t), u(t), t) \geq 0, t \in [t_0, t_f] \\ g(x(t_f), t_f) \geq 0 \end{array} \right. \quad (1.1.28)
\end{aligned}$$

$\mathcal{L}$ est appelé Lagrangien, il appartient à $\mathbb{R}$. Le vecteur $\lambda(t)$ est appelé vecteur des états adjoints. Il est de la même dimension que $x(t)$. Le vecteur $\eta(t)$ est le vecteur des multiplicateurs associés aux contraintes. Il a autant de composantes qu'il y a de contraintes à respecter à tout instant (même dimension que h). Le vecteur α est le vecteur des multiplicateurs associés aux contraintes terminales sur l'état. Il est de même dimension que g .

On constate que cette formulation est un problème aux deux bouts. En posant $(x(t), \lambda(t))$ le vecteur d'état décrivant le système, on a des contraintes sur $x(t_0)$ et sur $\lambda(t_f)$.

Nous reviendrons sur cette formulation dans la suite de la thèse.

Méthodes de caractérisation de la trajectoire optimale. Dans [72], Srinivasan et al. présentent une revue approfondie des techniques pour résoudre le problème (1.1.26). Ils distinguent les méthodes analytiques et les méthodes numériques. Ces travaux sont un outil précieux pour se familiariser avec les différentes méthodes.

Après avoir dérivé le principe du maximum de Pontryaguine et analysé ses conséquences et implications, les auteurs proposent différentes manières d'approcher la solution du problème (1.1.26) par des méthodes numériques. La Table 1.2 regroupe ces méthodes introduites en fonction de la manière dont on considère les états et les commandes.

L'expression commande paramétrée (respectivement état paramétré) signifie que la commande (l'état) est discrétisée et que ses valeurs aux instants de discrétisation sont des paramètres d'optimisation.

Les méthodes de la colonne "Formulation directe" raisonnent sur une résolution du problème (1.1.26). Les méthodes de la colonne "Formulation de Pontryaguine" se fondent sur la résolution d'un problème (1.1.28).

Une approche par "état continu, commande paramétré, formulation directe" consiste à chercher les valeurs optimales des commandes u_k à chaque instant de discrétisation kT avec un état $x(kT)$ donné par une intégration de la dynamique $\dot{x} = f(x(t), u(t), t)$ entre $(k-1)T$ et kT avec une commande $u(t)$ interpolée par rapport à la séquence de commandes (u_k) .

TABLE 1.2 – Méthodes de caractérisation de la trajectoire optimale [72].

	Formulation directe	Formulation de Pontryaguine
État paramétré Commande paramétrée	Approche simultanée (CP) ou Collocation	Paramétrisation de l'état et des adjoints
État continu Commande paramétrée	Approche séquentielle (CVP)	Méthode du gradient
État continu Commande continue	Paramétrisation analytique	Méthode de tir

C'est au cours des années 1990, grâce à la diffusion de l'informatique, que l'utilisation de méthodes numériques s'est démocratisée, Biegler a alors commencé à proposer la conversion du problème d'optimisation de la trajectoire en dimension infinie, en un problème non linéaire de dimension finie. Il a discrétisé la dynamique et les contraintes puis a paramétré le problème par les commandes et les états à chaque pas de temps [20] (formulation CP de la Table 1.2). Le problème non linéaire est résolu par approximations quadratiques successives en introduisant les états adjoints de Lagrange (du problème en dimension finie). C'est ce que l'on appelle la collocation, ou paramétrisation complète (par opposition à la paramétrisation par variable de commande, ou CVP). Le problème résolu n'est pas celui initial (1.1.26) mais on peut montrer que lorsque le pas d'échantillonnage tend vers zéro, la solution du problème discrétisé tend vers la solution du problème optimal [6].

Trajectoires optimales pour la production d'un composé d'intérêt. Des exemples d'optimisations sont disponibles dans [72]. Les exemples concernant la production d'un composé d'intérêt montrent qu'il est préférable de faire croître la population de micro-organismes avant de produire le composé d'intérêt.

Ce résultat s'applique ainsi au cas de maximisation de la productivité du procédé, dans [51]. Le problème est ici formulé à l'aide du principe du maximum de Pontryaguine, sans paramétrer la commande, ni l'état (méthode de tir) et avec deux débits d'alimentations, l'un qui commande la croissance, l'autre la production. Ils montrent que l'optimum est de croître puis de produire.

Les méthodes analytiques ont de la valeur si elles donnent une justification théorique aux profils optimaux proposés. Ce qui ressort de la littérature est la difficulté de la caractérisation de la trajectoire à cause de la linéarité du modèle vis-à-vis de la commande F_{in} , comme présenté dans [67].

Par ailleurs, lorsque le nombre de commande est plus grand que 1, il faut considérer un plus grand nombre de cas, ce qui rend la caractérisation difficile. On trouve une discussion de ce point par exemple dans [51] ou [59].

Les méthodes qui paramètrent les commandes et l'état numériquement vont également dans le sens d'une production qui succède à une croissance.

Cuthrell et Biegler ont appliqué la méthode de paramétrisation de l'état et de la commande à un modèle de bioprocédé prenant en compte la croissance de micro-organisme et la production d'un composé d'intérêt [20]. Dans ces conditions, ils montrent que l'optimum est d'abord de croître puis de produire à croissance nulle.

En résumé, la stratégie en deux temps (croissance puis production) semble être optimale pour maximiser la production d'un composé d'intérêt en utilisant les modèles macroscopiques.

1.1.2.2 Méthodes de commande en boucle fermée

Il est important de comprendre que la résolution de problème d'optimisation de trajectoire ne sert pas tant à trouver la fonction $t \mapsto u^*(t)$ optimale qu'à intégrer cette commande et étudier la trajectoire de l'état $x^*(t)$ définie comme :

$$\begin{cases} x^*(t_0) = x_0 \\ \frac{dx^*}{dt}(t) = f(x^*(t), u^*(t), t) \end{cases} \quad (1.1.29)$$

En raison des écarts entre le modèle et l'objet réel, il n'est pas question d'espérer retrouver les mêmes allures si l'on soumet le système réel et le modèle mathématique à la commande optimale. En particulier, les perturbations que subit le système réel par opposition au modèle mathématique vont dévier le système de la trajectoire attendue par l'analyse du modèle.

Il est nécessaire de boucler la commande du système par rapport à son évolution réelle, *i.e.* de le piloter en boucle fermée autour de la trajectoire optimale déterminée hors-ligne. La trajectoire ainsi obtenue sera d'autant plus optimale que le modèle est représentatif du système réel.

Nous allons nous intéresser à des méthodes qui permettent de commander l'état d'un bioprocédé réel. On parlera de commande en boucle ouverte, ou commande optimale, lorsque l'on appliquera telle quelle la commande calculée hors-ligne pour maximiser le critère sur le modèle.

Pour placer le système réel à son optimum, on peut utiliser la trajectoire optimale sur l'état définie en (1.1.29) et trouver à l'instant t la commande $F_{in}(t)$ qui assure un écart minimum entre l'état du procédé réel, $x(t)$, et l'état de référence, $x^*(t)$.

Plus précisément, en voyant la trajectoire $x^*(t)$ comme une succession de références à suivre sur des composantes de x^* , on peut chercher à suivre cette même succession de références pour x . En clair, si la trajectoire x^* est d'abord une phase où $G = G_1$ puis une phase où $X = X_2$, on peut faire suivre au système réel la référence $G = G_1$ jusqu'à ce que $X = X_2$ où l'on permute la référence à suivre sur $X = X_2$.

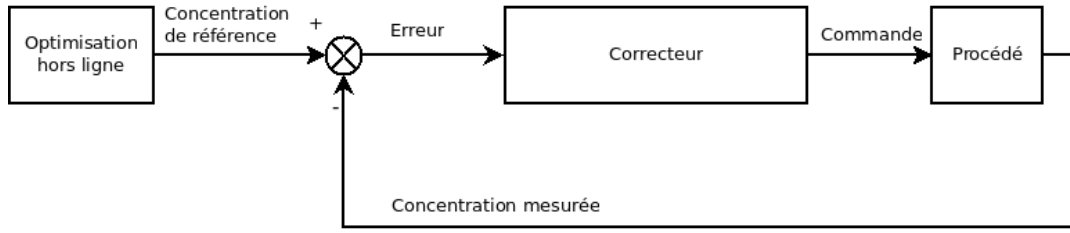


FIGURE 1.4 – Schéma de fonctionnement de la commande par suivi de références. L’optimisation hors-ligne permet de déterminer la référence à faire suivre aux composantes de l’état du système.

La figure 1.4 présente sommairement le principe de fonctionnement de l’utilisation de l’optimisation hors-ligne pour en déduire une trajectoire à suivre. Le système peut être asservi autour de cette succession de références par un correcteur. Le défi est alors de concevoir un correcteur efficace. Nous allons voir avec la commande linéarisante une conception possible pour un correcteur simple et efficace.

Commande linéarisante. Une manière de concevoir le correcteur est d’utiliser la commande de systèmes non linéaires et en particulier la commande linéarisante, comme présentée dans [40] et mise en application dans [73]. Avec cette conception on montre le passage d’une trajectoire optimale à une commande en boucle fermée pour maintenir le procédé autour de cette trajectoire.

L’idée est de faire apparaître un comportement linéaire entre l’entrée et la référence à suivre en introduisant une commande qui rende le système [commande $\rightarrow$ sortie à suivre] équivalent à un certain nombre d’intégrateurs en série.

Cette approche est utilisée dans le cas de systèmes non linéaires affines en la commande. Les bioprocédés décrits jusqu’ici où la commande est le débit d’alimentation, vérifient bien cette hypothèse.

En prenant l’exemple d’une concentration ξ_i à réguler autour d’une valeur de référence, si la variation de ξ_i suit la relation (1.1.11) :

$$\frac{d\xi_i}{dt}(t) = Ar(\xi(t)) + (\xi_{i,F} - \xi_i(t)) \frac{F_{in}(t)}{V(t)}$$

alors on pose $\hat{U}(t) = Ar(\xi(t)) + (\xi_{i,F} - \xi_i(t)) \frac{F_{in}(t)}{V(t)}$ de sorte que :

$$\frac{d\xi_i}{dt}(t) = \hat{U}(t)$$

Ensuite, on régule à l’aide d’un correcteur linéaire classique la commande $\hat{U}(t)$ de sorte à suivre la référence sur ξ_i . On peut introduire un régulateur proportionnel-intégral, $K(1 + \frac{1}{T_{ip}})$ en choisissant le temps caractéristique, T_i , de l’ordre de grandeur de celui du temps de réponse souhaité, et le gain, K , pour avoir une dynamique de l’erreur de suivi assez rapide.

On voit figure 1.5, que le correcteur global du système est composé de deux blocs. Le premier est le correcteur linéaire défini ci-dessus ; le second est le bloc de linéarisation

qui prend en entrées $\hat{U}$ et l'état du système et qui renvoie la vraie commande F_{in} . Dans notre cas, ce deuxième bloc est défini comme :

$$F_{in}(t) = \frac{\hat{U}(t) - Ar(\xi(t))}{(\xi_{i,F} - \xi_i(t))} V(t)$$

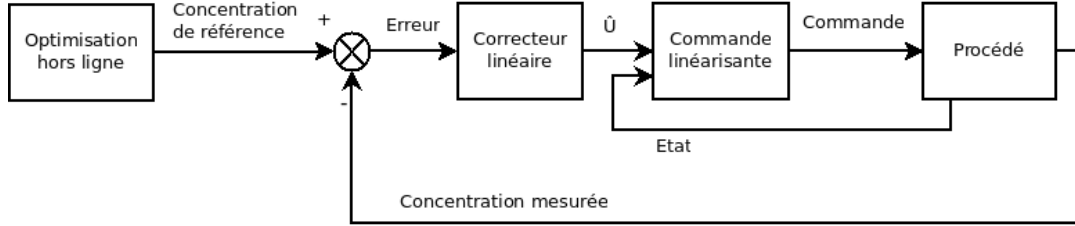


FIGURE 1.5 – Schéma de fonctionnement de la commande linéarisante.

Si l'on n'a pas accès directement à l'état que l'on souhaite, il est possible d'avoir recours à un observateur dans la boucle de retour. C'est ce qui est développé dans [19] où une preuve de stabilité est fournie.

Commande prédictive. Une autre stratégie intéressante est la commande prédictive.

Introduite entre autre dans [16] et [3], la commande prédictive est une optimisation en temps réel qui vise à trouver la commande à appliquer à un système discret pour maximiser un critère étant donné son état actuel et la trajectoire de référence calculée hors-ligne.

On résout alors à chaque pas d'échantillonnage k un problème du type :

$$\begin{aligned} \min_{u_k, \dots, u_{k+n_C}} & J(x_{k+1}, \dots, x_{k+n_H}, u_k, \dots, u_{k+n_C}) \\ \text{sous contraintes, } & \begin{cases} \dots, x_{k-1}, x_k, \text{ connus} \\ \dots, u_{k-2}, u_{k-1}, \text{ connus} \\ x_{k+1}, \dots, x_{k+n_H}, \text{ prédits par } x_{j+1} = f(x_j, u_j) \\ h(x_j, \dots, x_{j+n_H}, u_j, \dots, u_{j+n_C}) \leq 0, \forall j \end{cases} \end{aligned} \quad (1.1.30)$$

avec n_H l'horizon de prédiction, n_C l'horizon de commande, J le critère, f le modèle de l'évolution du système, h des contraintes sur l'état et/ou la commande.

On infère ici la trajectoire à venir pour l'état du système.

Dans la fonction h on trouvera des conditions portant sur les bornes de la commande ou des incréments de commande.

A chaque pas de temps, on résout un problème de ce type, ceci donne un vecteur de commande à appliquer aux n_C prochains pas d'échantillonnage, $(u_k \ u_{k+1} \ \dots \ u_{k+n_C})$, dont on n'applique que la première composante, *i.e.* la commande u_k . Puis on attend l'instant d'échantillonnage suivant et on résout de nouveau le problème (1.1.30) pour $k = k + 1$.

C'est le principe de l'horizon glissant ou horizon fuyant.

L'avantage de cette méthode est que l'on peut avoir pris un modèle éloigné de la réalité et voir que la commande permet d'atteindre un objectif satisfaisant. En effet, on peut envisager une commande prédictive qui permet la prise en compte de la différence entre la sortie mesurée et la sortie présumée du système dans la résolution du problème (1.1.30) (voir partie bleue 1.6). Ce raffinement se révèle efficace dans le cadre des bioprocédés, présentés dans [74]. Dans cet article, on présente les performances d'une commande prédictive en tenant compte de la différence objet modèle comportant des incertitudes sur le modèle et un bruit de mesure. Les résultats montrent une grande robustesse de la méthode.

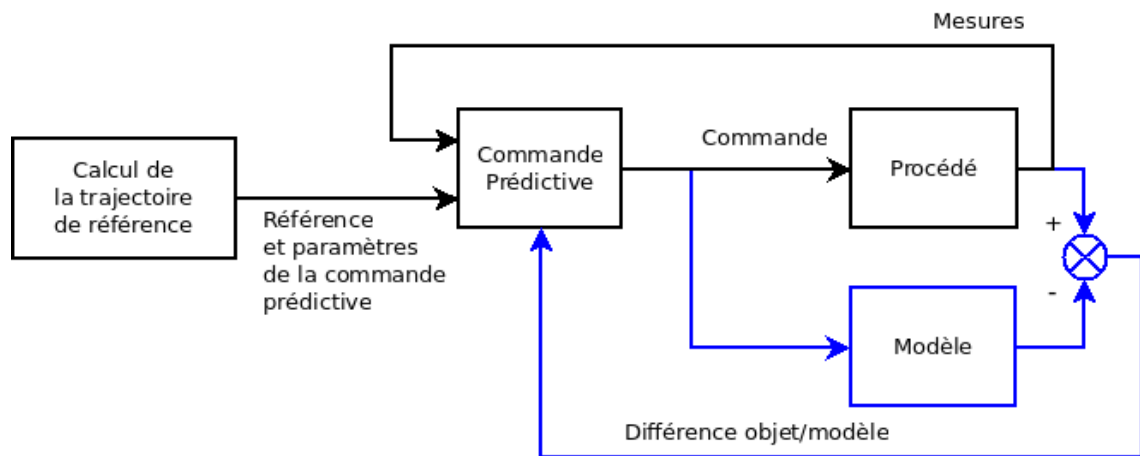


FIGURE 1.6 – Schéma de fonctionnement de la commande prédictive. En bleu, prise en compte en temps réel de la différence entre le procédé réel et son modèle, comme présenté dans [74].

La littérature est riche en application de commande prédictive aux bioprocédés. [74] est un exemple dans lequel le modèle non linéaire est pris tel quel. On peut également citer [58] qui linéarise le modèle, supposant que les variations sont assez lentes pour pouvoir être linéarisées entre deux pas de temps.

Ce type de résolution à chaque pas d'échantillonnage est possible avec des ordinateurs puissants et parce que l'on dispose de plusieurs minutes (voire dizaines de minutes) pour les bioprocédés entre deux instants d'échantillonnage.

1.1.3 Conclusion sur l'approche macroscopique phénoménologique

Dans cette section, on a présenté la manière dont sont construits les modèles macroscopiques de la formulation des échanges (plus ou moins indépendants et plus ou moins raffinés), à la formulation mathématique des cinétiques des réactions.

On s'est ensuite intéressé à la résolution de problème d'optimisation de bioproduction de composés d'intérêt en agissant sur le milieu de culture au travers de la commande de réacteurs fed-batch. Des profils optimaux calculés hors-ligne montrent que pour produire un composé d'intérêt de manière optimale, il faut commencer par une phase de croissance avant de passer à la production.

On a ensuite vu que la séquence de commande issue directement de l'optimisation était mise en défaut par les différences entre le procédé réel et son modèle. Des méthodes de la commande prédictive ou de la commande linéarisante introduisent un bouclage qui tient compte du procédé réel. Ce faisant, elles réduisent la dépendance au modèle. Ces méthodes s'appliquent alors à un système réel.

Au vu de la modélisation, l'optimisation des bioprocédés ne peut se faire que par l'optimisation du milieu de culture, et donc par l'intermédiaire du débit d'ajout de matière dans le cadre des réacteurs fed-batch.

Qu'en est-il du comportement des micro-organismes ? Est-il possible d'optimiser directement le comportement de ces derniers, par exemple par biologie de synthèse, pour optimiser le bioprocédé dans son ensemble ? C'est ce que nous allons chercher à faire dans la prochaine section.

1.2 Approche de type « conception de souche »

Les modèles macroscopiques présentés jusqu'ici intègrent des phénomènes complexes et représentent les échanges entre différents composés en quelques relations mathématiques. La construction itérative que nous avons vu affine le comportement du modèle au gré des besoins de son application.

Toutefois, les mécanismes intracellulaires à l'origine des observations macroscopiques ne sont pas décrits dans ces modèles. Il est alors hors d'atteinte d'optimiser le comportement des cellules en agissant sur ces mécanismes. On pense en particulier aux niveaux d'expression des gènes ou à l'adaptation du fonctionnement cellulaire selon son environnement. Pourtant, des modèles existent pour décrire le fonctionnement des micro-organismes en se fondant sur les mécanismes intracellulaires. Des méthodes existent également pour optimiser le comportement de micro-organismes et pour donner un sens aux paramètres introduits. Dans cette section, on présente des modèles qui décrivent le comportement des micro-organismes et comment on peut améliorer leur fonctionnement en jouant sur les paramètres introduits dans la modélisation.

Après avoir présenté différents types de modèles, nous montrerons comment on les utilise pour optimiser les souches puis les bioprocédés.

1.2.1 Modèles intracellulaires

On s'intéresse aux familles de modèles susceptibles de rendre compte du lien entre structure, mécanismes et fonctionnement des micro-organismes. Cette partie fait appel à des connaissances en mécanismes intracellulaires qui sont rappelées en annexe A de cette thèse.

1.2.1.1 Modèles phénoménologiques

Premier jalon. La première famille de modèles de comportement des cellules est la famille des modèles systémiques, quasi-mécaniques. Allen Marr reprend de manière théorique les travaux expérimentaux de l'école de Copenhague dans l'article [57].

Constatant que des mesures existent et que des comportements cellulaires généraux se dégagent pour le taux spécifique de croissance (travaux de Monod), Marr entreprend de construire un modèle de cellule qui permette de déterminer l'origine de la croissance de *Escherichia coli*.

D'où vient la limite de la croissance bactérienne ? En 1990, cette question était très polémique et ne faisait pas consensus au sein de la communauté des biologistes. La théorie la plus répandue au moment où Marr publie ses travaux, revient à dire que la croissance est limitée par le flux d'ATP que la cellule peut produire. Marr explique que le moteur de la croissance est lié à la production des protéines. L'efficacité de l'appareil de traduction étant limitée, la croissance est limitée *de facto*. Par ailleurs, Marr parvient, en prenant des règles élémentaires de régulation, d'activation et de saturation en fonction du taux spécifique de croissance, à prédire la composition interne des cellules bactériennes en termes de ribosomes, de protéines non-ribosomales, d'acides aminés libres, d'ARN de transferts chargés et non chargés, et de ppGpp (molécule jouant un rôle dans la régulation de la traduction, voir annexe A). Ces prédictions sont confirmées par des mesures expérimentales et quantitatives à différents taux spécifiques de croissance faites par Bremer et Dennis [14]. Dans ces travaux, la ressource n'est pas directement prise en compte. Marr montre que le taux spécifique de croissance ne peut être infini, et il cherche les paramètres intracellulaires qui pourraient être modifiés pour augmenter le taux spécifique de croissance maximal (par exemple le nombre de copies des gènes codants pour les protéines ribosomales). Marr s'intéresse également à la réponse du système cellulaire lors d'une augmentation ou diminution du taux spécifique de croissance. Il réussit à prédire l'inertie des composants.

Distribution des ressources en fonction du taux spécifique de croissance et limitations du taux spécifique de croissance. Les travaux de l'équipe de Terence Hwa s'inscrivent dans la lignée de ceux de Marr. Leur objectif est de caractériser la répartition des ressources entre les différents processus pour différentes conditions de croissance, ou pour différentes valeurs de taux spécifique de croissance.

Dans [47], les auteurs étudient en fonction du taux spécifique de croissance, l'expression des gènes constitutifs et de ceux régulés positivement et négativement par un gène lui-même constitutif. Ils trouvent alors un ensemble de profils fonctions du taux de croissance pour les concentrations de protéine. Ces profils sont présentés pour des cas généraux de régulations dans l'article de [45]. On y voit que la concentration d'une protéine constitutive peut être :

- soit affine croissante ou décroissante en fonction du taux de croissance pour les gènes constitutifs,
- soit hyperboliquement décroissante,
- soit globalement constante lors d'une régulation par un gène lui-même constitutif.

Dans une publication de 2010 [71] ils ajoutent à l'influence du taux spécifique de croissance, l'influence des nutriments. Ils prédisent la concentration d'une protéine en fonction de son rôle et des conditions intra- et extra-cellulaire (respectivement, le taux spécifique de croissance et l'abondance en nutriments). Pour leur étude, ils partagent l'ensemble des protéines en trois groupes :

- le groupe R de l'appareil de traduction,
- le groupe Q des protéines toujours présentes quelles que soient les conditions,

— le groupe P des autres protéines.

Ils s'intéressent à la répartition de la matière entre ces trois groupes en fonction de la qualité des nutriments et de l'efficacité des processus internes dont la traduction.

Leurs conclusions sont en accord avec l'intuition biologique : si la qualité des nutriments augmentent, la part de P diminue, la part de R augmente et la croissance augmente. D'une manière équivalente, si l'efficacité de la traduction diminue, la part de R augmente, la part de P diminue et le taux spécifique de croissance diminue également. Ces résultats sont présentés figure 1.7.

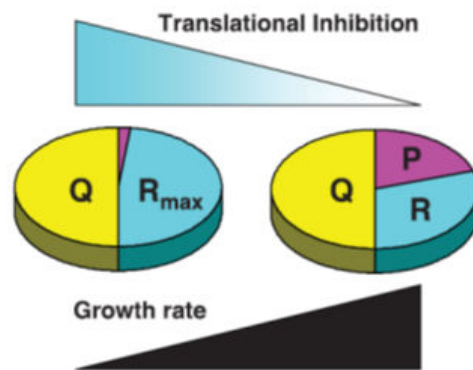


FIGURE 1.7 – Répartitions massiques des trois groupes de protéines P, Q, R en fonction du taux de croissance et de l'efficacité de traduction. Figure issue de [71].

Dans ce dernier travail, ils traitent aussi de l'influence d'une protéine de synthèse inutile à la croissance ajoutée à la cellule. D'après eux, la croissance diminue d'autant plus que l'on augmente la part d'une protéine inutile à la croissance.

En 2013, l'équipe de Terence Hwa divise le groupe des protéines P en un groupe catabolique C et un groupe anabolique A [88]. Ils connaissent les mesures de chaque groupe pour différents taux spécifiques de croissance et dans des conditions de substrats différents. Ils se servent de ces mesures pour calibrer leur modèle puis prédisent le reste des mesures.

Par exemple, ils utilisent la relation phénoménologique qui lie le taux spécifique de croissance et la portion de ribosomes. Ils en déduisent que lorsque le milieu s'appauvrit, ou qu'il est plus coûteux d'importer de la matière dans le compartiment intracellulaire, la part de protéines dédiées au catabolisme augmente mais le taux spécifique de croissance diminue.

Dans [46], ils retrouvent, comme dans Marr [57], une manière de limiter le taux spécifique de croissance maximal : ils développent le groupe de traduction R en ribosomes Rb et ARN de transfert T ; ils expriment le fait que les ribosomes sont utiles à la croissance, que les ARN de transfert agissent sur l'efficacité des ribosomes et que ce sont les proportions ϕ qui comptent (*i.e.* les concentrations massiques). Le taux spécifique de

croissance est alors donné par :

$$\begin{cases} \mu = k_R(\phi_{Rb} - K_0) & (a) \\ k_R = k_R^{max} \frac{\phi_T}{\phi_T + K_R} & (b) \end{cases} \quad (1.2.1)$$

La relation (1.2.1.a), *i.e.* la relation affine entre quantité de ribosomes et taux spécifique de croissance, est phénoménologique : le coefficient K_0 provient d'observations mais n'a pas de sens à proprement parlé. La relation (1.2.1.b) est issue de considérations physico-biologiques.

En combinant ces équations avec un mélange d'autres considérations en partie phénoménologiques et en partie physico-biologiques, ils trouvent une borne sur le taux spécifique de croissance.

En fait, les deux relations (1.2.1) sont suffisantes pour trouver une borne maximale sur le taux spécifique de croissance. Si l'on considère $\phi \leq 1$ (par définition), on peut écrire :

$$\mu \leq k_R^{max} \underbrace{\frac{\phi_T}{\phi_T + K_R}}_{\leq 1} \underbrace{(\phi_{Rb} - K_0)}_{\leq 1} \leq k_R^{max}$$

On note que, comme dans Marr [57], c'est la production des protéines qui limite la croissance.

Pour conclure sur les travaux de modélisation de ce groupe, l'article [45] synthétise les différentes expressions de gènes en fonction de leur caractère constitutif ou de la manière dont ils sont régulés. Les concentrations modélisées montrent que la concentration en protéine peut être affine (croissante ou décroissante et même constante), ou hyperboliquement décroissante. Des cas de bistabilité par rapport au taux spécifique de croissance sont également à prévoir dans les cas de double inhibition.

Cette étude nous intéresse dans la conception de nouvelles souches pour orienter la régulation quand un profil d'expression est établi en fonction du taux spécifique de croissance.

Les travaux de cette équipe donnent de bonnes idées sur le fonctionnement des cellules.

Tant que l'on reste à une échelle descriptive, la manière dont les observations sont utilisées pour dériver des règles générales est tout à fait licite. Toutefois, il est de fait exclus de se servir des règles établies par ces méthodes pour une souche qui n'a pas été étudiée au préalable. Typiquement, la modélisation développée ici utilise à la fois des constantes physico-biologiques comme les efficacités de certains processus (par exemple, k_R^{max}) et des constantes artificielles (comme K_0) qui proviennent de régressions sur des mesures.

Les observations et les principes d'implémentation (qui viennent de la physique ou de la biologie) se mélangent tout comme les paramètres artificiels et les paramètres physico-biologiques.

Comme le cœur du raisonnement repose sur des observations il est difficilement envisageable de généraliser l'approche à un organisme pour lequel des mesures n'ont pas été calibrées, comme une souche modifiée par exemple.

Un modèle de comportement dynamique de la cellule. Le modèle présenté dans [85] se fonde sur les travaux et observations de l'équipe de Hwa. Il propose un ensemble d'équations différentielles pour décrire l'évolution de la cellule en fonction de paramètres issus en partie d'observation et en partie de causes physico-biologiques. Ces équations décrivent l'évolution temporelle d'entités représentatives à savoir :

- un substrat interne,
- un pool d'énergie,
- un pool de ribosomes,
- une protéine de transport,
- une protéine métabolique,
- une protéine de maintenance,
- les 4 ARN messagers associés aux ribosomes et aux trois protéines susmentionnées,
- 4 entités correspondant aux complexes entre ribosomes et chaque messenger.

L'intérêt de ce modèle est qu'il rend compte des interactions dans la cellule. Son caractère descriptif est indéniable. L'inconvénient est qu'il ne permet pas de prédire l'impact des modifications potentielles de souche au travers des paramètres qui peuvent ne pas avoir de sens, ou dont on ne connaît pas *a priori* la plage de modification possible. En particulier, on ne peut pas anticiper l'impact de l'ajout d'un gène sur le taux spécifique de croissance.

Conclusion à propos des modèles systémiques. Les modèles que nous avons parcourus jusqu'à présent ont pour but de décrire la cellule en en comprenant les enjeux, les compromis et les interactions qui surviennent entre les différentes entités intracellulaires.

Néanmoins, on peut penser que cette approche systémique est trop rigide pour prédire le comportement d'un micro-organisme mal connu. En effet, les modèles que nous avons vus partent d'observations de tendances pour en déduire rationnellement des modes de fonctionnement. Ces derniers ont l'air de décrire la cellule mais on doit revenir aux observations pour justifier les résultats fournis par les modèles.

1.2.1.2 Modèles sous contraintes

Nous nous intéressons à présent à une vision orthogonale du fonctionnement cellulaire : l'approche sous contraintes.

Historiquement, l'étude de la cellule est centrée sur les échanges plutôt que sur la répartition de la matière entre les entités. Elle se focalise sur les réactions elles-mêmes plutôt que sur les produits. Grâce au travail des biologistes, on a une idée précise des réactions métaboliques qui ont lieu dans une cellule. On s'intéresse aux vitesses de réaction de ces échanges, que l'on appelle flux et que l'on note en général ν . L'objectif est de caractériser le plus précisément possible les flux cellulaires dans des conditions données. Le problème est alors de trouver les flux de matière qui satisfont des contraintes de réalisation, et dans cet espace des possibles, de choisir la répartition la plus vraisemblable selon un critère.

Origine des modèles sous contraintes : la méthode FBA. La méthode FBA (Flux Balance Analysis) cherche à caractériser l'état du réseau métabolique dans un milieu donné. Cette méthode est apparue au milieu des années 1990, dans l'article fondateur [79].

A la base, Varma et Palsson cherchent la répartition du réseau métabolique qui maximise la croissance en respectant des observations de divers flux extracellulaires. Ils ont travaillé à partir des flux d'absorption d'oxygène et de glucose et sur les flux excrétés quand ils pouvaient les mesurer. En notant Ω_t la matrice stœchiométrique du réseau de transport, ils affirment que le vecteur des flux ν doit satisfaire :

$$\Omega_t \cdot \nu = u \quad (1.2.2)$$

avec u le vecteur qui contient les flux nets d'absorption, observés.

En supposant que le réseau métabolique est à l'équilibre (ce qui est le cas en phase de croissance exponentielle) et que le taux spécifique de croissance est constant, ils posent :

$$\frac{dm}{dt}(t) = \Omega \cdot \nu = 0 \quad (1.2.3)$$

avec m le vecteur des concentrations des différents métabolites, Ω la matrice stœchiométrique de l'ensemble du réseau.

$\Omega_{i,j}$ est alors le coefficient stœchiométrique du $i^{\text{ème}}$ métabolite pour la $j^{\text{ème}}$ réaction.

Pour le taux spécifique de croissance et la production de biomasse, on définit artificiellement une entité "biomasse", X , et une réaction :

$$c^\top m = \sum_i c_i m_i \xrightarrow{\mu, \nu_{biomass}} X \quad (1.2.4)$$

Le flux de cette réaction est noté soit μ , soit $\nu_{biomass}$. Les coefficients c_i sont des coefficients stœchiométriques de la réaction de production de biomasse.

μ ou $\nu_{biomass}$ est ajouté au vecteur ν des flux, X est ajouté au vecteur m . La matrice Ω est augmentée d'une ligne et d'une colonne pour prendre en compte X et la réaction (1.2.4).

Dans la formulation classique des FBA, c'est le flux de la réaction (1.2.4) que l'on cherche à maximiser.

Toutefois, dans [9], les auteurs discutent du critère et estiment que plusieurs critères peuvent finalement entrer dans une configuration vraisemblable : par exemple, la maximisation du flux d'ATP ou la minimisation de la production de dioxyde de carbone. Néanmoins, en maximisant deux critères différents sur le même réseau, les configurations trouvées sont distinctes.

On choisi le taux spécifique de croissance comme critère à maximiser parce qu'il correspond à la théorie de l'évolution, c'est-à-dire la théorie des avantages compétitifs entre individus de la même espèce. On part de l'idée que les cellules qui optimisent leur croissance instantanée ont un avantage si les conditions extérieures sont stables et propices à la croissance. Dans le cas d'un milieu perturbé et dangereux, on pense que pour survivre, les cellule maximiseraient plutôt leur défense face au danger.

Le problème de maximisation du taux spécifique de croissance sous contrainte du respect de croissance exponentielle, et sous contrainte des flux observés d'absorption de composés extracellulaires s'écrit dès lors :

$$\begin{aligned} & \max_{\nu \in \mathbb{R}_+^m} \nu_{biomass} \\ & \text{sous contraintes } \begin{cases} \Omega \cdot \nu = 0 \\ \Omega_t \cdot \nu = u \end{cases} \end{aligned} \quad (1.2.5)$$

La formulation plus générale, plus répandue dans la littérature et à celle-ci que nous nous référerons en parlant des FBA est la suivante :

$$\begin{aligned} & \max_{\nu \in \mathbb{R}^m} \nu_{biomass} \\ & \text{sous contraintes} \quad \begin{cases} \Omega \cdot \nu = 0 & (1.2.6.a) \\ \alpha_i \leq \nu_i \leq \beta_i & (1.2.6.b) \end{cases} \end{aligned} \quad (1.2.6)$$

les α_i et β_i correspondent aux valeurs minimales et maximales que l'on autorise pour les flux. L'intérêt est que l'on peut ici forcer une réaction à être irréversible (en posant $\alpha_i = 0$ et $\beta_i = +\infty$). Dans le cas contraire, la réaction peut avoir lieu dans les deux sens, comme c'est le cas pour les réactions de transport. On retrouve la formulation (1.2.5) en augmentant le vecteur des flux ν par les combinaisons linéaires $\Omega_t \cdot \nu$ et en prenant $\alpha_i = \beta_i = u_i$ pour ces flux.

On remarque donc que la formulation (1.2.6) est un problème d'optimisation linéaire. Il est soluble par des méthodes classiques et la solution est soit unique, soit un sous-ensemble convexe de $\mathbb{R}^m$ avec $m = \dim(\nu)$. De plus, le nombre de paramètres reste limité.

Dans un article de 1994 [80], les mêmes auteurs présentent une analyse de sensibilité de leur modèle. Il apparaît que les bornes α_i et β_i sont les paramètres les plus déterminants, en particulier pour les flux d'absorption de substrat et d'oxygène. L'influence des valeurs de Ω (et en particulier du vecteur c de définition de $\nu_{biomass}$) semble être secondaire, comme présenté dans [9]. Dans cet article, on fait varier les paramètres de plus ou moins 20%. Les auteurs observent que l'influence des bornes est bien plus importante que l'influence des coefficients de c .

Cette observation est un peu dérangeante car les paramètres qui ont le plus d'importance n'ont pas de sens biologique. En effet, que signifie le fait qu'un flux est limité par une valeur donnée ? On retrouve la même cohabitation que dans les travaux de Hwa entre coefficients physico-biologiques (Ω) et coefficients phénoménologiques arbitraires issus d'observations (α_i, β_i).

Jusqu'à présent, la détermination des flux ν est faite de manière statique, c'est-à-dire pour des conditions stationnaires. Voyons à présent une formulation dynamique du problème.

Formulation dynamique des FBA. Dans l'article fondateur [79], les auteurs proposent une version dynamique discrète des FBA décrivant l'évolution de concentrations extracellulaires lorsque l'on intègre les flux d'absorption, de sécrétion, et de croissance sur chaque pas de temps.

A chaque pas de temps, on résout un problème (1.2.6) avec les bornes α_i et β_i données. Le métabolisme (*i.e.* les concentrations en métabolites) est supposé à l'équilibre à chaque instant d'échantillonnage et aucune contrainte n'est formulée sur les incréments des flux.

Ainsi, il n'y a pas de mémoire sur l'état du réseau entre deux pas de temps consécutifs.

En 2002, Mahadevan et al. [56] proposent deux approches dynamiques de la méthode FBA en ajoutant une contrainte sur la dérivée des flux entre deux pas de temps. Ils

ajoutent également une description des métabolites dans la caractérisation de l'état. Deux formulations sont proposées, une dite dynamique et une autre dite séquentielle.

Dans la formulation séquentielle, la période de simulation est divisée en intervalles de durée ΔT . Au début de chaque intervalle, on résout le problème suivant :

$$\begin{aligned} & \max_{\nu(t)} \nu_{biomass}(t) \\ & \text{sous contraintes} \quad \begin{cases} \alpha_i \leq \nu_i(t) \leq \beta_i \\ \dot{\alpha}_i \leq \frac{\nu_i(t) - \nu_i(t - \Delta T)}{\Delta T} \leq \dot{\beta}_i \\ m(t + \Delta T) = m(t) + \Omega \cdot \nu(t) \Delta T \\ m(t + \Delta T) \geq 0 \end{cases} \end{aligned} \quad (1.2.7)$$

avec les mêmes notations que celles introduites jusqu'à présent, $\dot{\alpha}_i$ et $\dot{\beta}_i$ les valeurs minimales et maximales de variation des flux.

La formulation dynamique quant à elle consiste à résoudre l'unique problème suivant :

$$\begin{aligned} & \max_{\nu(t_0), \nu(t_0 + \Delta T), \dots, \nu(t_0 + N\Delta T)} J(\nu(t_0), \nu(t_0 + \Delta T), \dots, \nu(t_0 + N\Delta T)) \\ & \text{s.c.} \quad \begin{cases} \alpha_i \leq \nu_i(t + k\Delta T) \leq \beta_i, & k \in \llbracket 0, N \rrbracket \\ \dot{\alpha}_i \leq \frac{\nu_i(t + k\Delta T) - \nu_i(t + (k-1)\Delta T)}{\Delta T} \leq \dot{\beta}_i, & k \in \llbracket 1, N \rrbracket \\ m(t + (k+1)\Delta T) = m(t + k\Delta T) + \Omega \cdot \nu(t + k\Delta T) \Delta T, & k \in \llbracket 0, N-1 \rrbracket \\ m(t + k\Delta T) \geq 0, & k \in \llbracket 0, N \rrbracket \end{cases} \end{aligned} \quad (1.2.8)$$

où N est le nombre d'instants d'échantillonnage.

On constate que la formulation dynamique cherche à déterminer directement les trajectoires des flux et des concentrations de métabolites sur toute la période de travail, tandis que la méthode séquentielle cherche la meilleure stratégie sur les flux à chaque instant d'échantillonnage afin d'optimiser un critère local. Pour ce faire, cette méthode s'appuie sur l'état des flux et des métabolites à l'instant précédent.

La difficulté de l'approche dynamique est de trouver le critère à maximiser quand on travaille sur toute la durée de la culture. Sachant que les micro-organismes semblent appliquer une stratégie à courte échéance, il est difficile d'envisager qu'ils poursuivent un but à long terme comme la maximisation de la biomasse à l'instant final. En revanche, il est possible de recourir à un critère intégral qui prenne en compte une sorte d'effet sans mémoire sur la croissance, pondérant les quantités de biomasse produite à chaque instant par une exponentielle décroissante, soit :

$$J = \sum_k \mu(t + k\Delta T) \exp\left(-\frac{t + k\Delta T}{\tau}\right) \quad (1.2.9)$$

avec τ un paramètre relatif à la mémoire du procédé.

Le taux spécifique de croissance à l'instant t a ainsi plus d'importance que le taux spécifique de croissance à l'instant $t + \Delta T$.

C'est en jouant sur l'expression de ce critère que la méthode dynamique arrive à surpasser l'approche séquentielle.

Mais le problème décrit-il si bien les phénomènes sous-jacents ? La solution qui consiste à introduire de nouveaux paramètres artificiels $\dot{\alpha}_i$ et $\dot{\beta}_i$ est-elle vraiment la meilleure solution pour décrire le fait que la configuration du réseau métabolique ne peut avoir une évolution instantanée ? Il n'y a pas de réponse évidente.

Raffinement des FBA. Suite à l'article de Varma et Palsson [79], beaucoup d'articles proposent des améliorations à cette méthode. Ils partent tous du fait que la formulation initiale (1.2.6) est intéressante car simple et descriptive du métabolisme mais qu'elle ne tient pas compte d'aspects cellulaires qui peuvent se révéler déterminants.

Sans chercher l'exhaustivité, les articles introduits ci-après présentent des exemples de raffinements proposés à la méthode FBA afin de la rapprocher de la réalité physico-biologique.

Lee et al. présentent dans [52] des avancées dans le raffinement des FBA. Outre le changement de critère, l'ajout de contraintes sur la régulation des gènes (pour refléter le fait que des réactions sont mutuellement exclusives, par exemple) et sur la thermodynamique (car certaines réactions sont autant réalisables que d'autres mais bien plus lentes), les auteurs présentent les défis qui subsistent pour faire coïncider les observations métaboliques et les prévisions des FBA. L'article concerne également les utilisations faites des méthodes assimilées aux FBA pour le design de souche. Nous y reviendrons en section 1.2.2.

En 2012, Palsson et son équipe présentent une revue des méthodes d'analyses sous contrainte de micro-organismes [53]. Une vraie phylogénie d'une centaine de méthodes est présentée, permettant de distinguer des familles phylogénétiques de méthodes :

- les méthodes FBA avec contraintes biologiques
- les méthodes qui prennent en compte la thermodynamique
- les méthodes qui concilient les données omiques
- les méthodes qui prennent en compte les mécanismes de régulation
- les méthodes qui étudient en particulier le graphe associé au réseau métabolique
- les méthodes d'inférence de données manquantes
- les méthodes qui cherchent des ensembles d'optimums alternatifs possibles

Il demeure la question de donner un sens biologique aux paramètres introduits par les modèles. Ceci permettrait d'avoir une formalisation autonome.

Un travail qui mérite une attention particulière est présenté dans [5]. La méthode FBAwMC (pour *Flux Balance Analysis with Molecular Crowding*) y est introduite. Elle va dans le sens d'une autonomie pour la modélisation de la cellule. Dans cette approche, on parvient à faire correspondre un flux et l'entité qui le catalyse. Une contrainte générale est donnée sur l'ensemble des entités, en affectant à chaque entité un volume et en affirmant que le volume est limité. Typiquement,

$$\begin{cases} \nu_i = k_i E_i \\ \sum_i \ell_i E_i \leq D \end{cases} \quad (1.2.10)$$

avec ν_i le flux catalysé par la machinerie de concentration E_i , k_i un paramètre d'efficacité du processus Σ_i , ℓ_i l'encombrement de E_i et D l'encombrement maximal du cytosol.

On obtient la contrainte :

$$\sum_i \frac{\ell_i}{k_i} \nu_i \leq D \quad (1.2.11)$$

Cette contrainte globale, d'origine physico-biologique, s'ajoute aux contraintes sur les bornes α_i et β_i , (1.2.6.b).

Le travail publié par Waldherr et al. dans [83] quant à lui, s'approche d'une formalisation autonome. Les auteurs présentent une méthode dite deFBA (pour *dynamic enzyme-cost Flux Balance Analysis*). Se fondant sur les travaux des FBA dynamiques de [56], les auteurs proposent un ensemble de contraintes sur l'efficacité des processus et sur la place que prennent les machineries qui catalysent ces processus.

Cependant, les contraintes sur les bornes des flux (α_i et β_i) sont conservées donnant l'impression d'une redondance sur les contraintes. Une discussion est faite sur l'aspect dynamique comme nous l'avons faite en présentant [56] mais rien n'est dit sur la redondance des flux. Difficile alors de comprendre ce qui provient des contraintes biologiques et ce qui vient des contraintes artificielles sur les α_i et β_i .

Conclusion à propos des modèles sous contraintes. Les modèles sous contraintes ont un potentiel de description mais les contraintes qu'ils introduisent paraissent artificielles. On ne distingue pas suffisamment ce qui est issu de la biologie et ce qui est lié aux artifices introduits pour être conforme aux observations ou à des comportements présupposés.

1.2.1.3 Modèles autonomes

Un modèle est dit autonome s'il peut être établi uniquement à partir de données physico-biologiques. Les modèles autonomes remplacent les artifices mathématiques par des considérations d'ordre biologique, et permettent de comprendre l'origine des phénomènes qui ressortent des simulations.

Modélisation RBA. C'est avec cette idée qu'a été formulée en 2008 la méthode RBA (pour *Resource Balance Analysis*, analyse de l'équilibre des ressources), [34]. Il s'agit de rapprocher l'approche systémique et phénoménologique, d'une approche sous contraintes. On peut voir cette méthode comme une optimisation sous contraintes systémiques, c'est-à-dire que les contraintes introduites sont le reflet des sous-unités du système cellulaire. Les défauts des deux approches précédentes, systémiques et sous contraintes, sont contournés : avec les modèles autonomes,

- (i) il n'y a pas besoin d'introduire de paramètres artificiels comme c'est le cas avec les méthodes sous contraintes précédentes,
- (ii) la partie sous contrainte permet d'être confiant dans l'idée que la partie systémique qui est introduite n'a pas besoin d'être complète pour être prédictive.

La modélisation RBA s'approche du fonctionnement réel des micro-organismes en posant le problème sous la forme qui est mise en œuvre lors de la sélection naturelle. On résout une succession de problèmes de faisabilité (qui sont tous convexes) à μ fixé, et on

cherche le μ maximal qui donne lieu à une configuration intracellulaire faisable. On peut justifier, de nouveau, qu'entre deux populations de croissance légèrement différente, celle qui a le taux spécifique de croissance instantané le plus élevé est prépondérante devant l'autre en temps fini, les croissances étant exponentielles. Le problème RBA cherche finalement la configuration qui est probablement la plus répandue parmi les populations de micro-organismes.

Pour les contraintes de faisabilité, les RBA en formulent quatre :

- 1) On suppose que tous les composés intracellulaires sont produits en quantités suffisantes par rapport à la demande de la cellule pour soutenir sa croissance. Ainsi, la stœchiométrie générale est prise en compte, comme elle l'est dans la formulation FBA.
- 2) On suppose que les constituants de l'appareil de traduction (ribosomes, facteur de traduction, ...) sont présents en quantité suffisante dans la cellule pour produire toutes les protéines, y compris celles de l'appareil de traduction. L'efficacité de traduction est supposée limitée, éventuellement donnée par une fonction du taux spécifique de croissance.
- 3) Les valeurs des flux de tous les processus intracellulaires sont limitées par la concentration de la machinerie moléculaire catalysant la réaction, multipliée par l'efficacité de cette réaction. Cette contrainte se rapproche de la contrainte précédente.
- 4) La densité totale de protéine est limitée, ce qui signifie que l'accumulation de masse conduit nécessairement à la croissance.

On peut formuler ces contraintes sous la forme suivante :

$$\left\{ \begin{array}{ll} \mu \xi_j = \sum_i \Omega_{j,i} \nu_i & (a) \\ \mu \sum_i \ell_{\mathcal{E}_i} E_i \leq k_R E_R & (b) \\ \nu_i \leq k_i E_i & (c) \\ \sum_i \ell_{\mathcal{E}_i}^D E_i \leq D_0 & (d) \end{array} \right. \quad (1.2.12)$$

avec :

- ξ_j la concentration totale en entité Ξ_j ,
- $\Omega_{j,i}$ le nombre algébrique de Ξ_j produits par la réaction associée au flux ν_i ,
- $\ell_{\mathcal{E}_i}$ le nombre d'acides aminés compris dans la machinerie moléculaire $\mathcal{E}_i$,
- E_i la concentration en entité $\mathcal{E}_i$,
- k_R l'efficacité de production des protéines,
- ν_i le flux de la réaction catalysé par la machinerie $\mathcal{E}_i$,
- k_i l'efficacité du processus Σ_i ,
- $\ell_{\mathcal{E}_i}^D$ le nombre d'acides aminés équivalent de la machinerie $\mathcal{E}_i$,
- D_0 la densité de l'organisme considéré.

Les grandeurs ℓ et ℓ^D sont les mêmes pour les protéines mais peuvent différer pour les complexes comme, par exemple, les ribosomes qui sont composés en partie d'ARN ribosomaux. Ces ARN comptent dans ℓ^D mais pas dans ℓ , car ils participent à l'encombrement du cytosol mais ne doivent pas être assemblés par les ribosomes.

Les concentrations ξ_j sont soit fixées (éventuellement à zéro), soit déterminées en fonction des entités $\mathcal{E}_i$ si Ξ_j est un élément constitutif de $\mathcal{E}_i$.

La contrainte (1.2.12.a) revient à limiter la demande en métabolite Ξ_i pour assurer la croissance, par ce qui est produit. Les termes à gauche du signe '=' représentent la croissance et ceux de droite les productions. Le terme de gauche de (1.2.12.b) est le nombre d'acides aminés à assembler par unité de temps pour assurer une croissance μ . Il est limité par la capacité des ribosomes à assembler les acides aminés (terme de droite). Les contraintes (1.2.12.c) sont du même ordre que la contrainte (1.2.12.b) mais pour chaque réaction catalysée par une entité $\mathcal{E}_i$: le flux de la réaction est majoré par le produit entre l'efficacité de la réaction et la concentration de cette entité. Enfin, la contrainte (1.2.12.d) impose que la quantité en entité catalysant des réactions soit inférieure à la densité maximale observée.

Une configuration cellulaire est alors un ensemble de flux, ν , un ensemble de concentrations en machinerie $\mathcal{E}$, E , et un taux spécifique de croissance, μ : (ν, E, μ) . La faisabilité d'une configuration cellulaire est satisfaite si les inégalités (1.2.12) sont satisfaites. On cherche par dichotomie la valeur maximale, μ_{max} , du taux spécifique de croissance pour laquelle il existe une configuration (ν, E, μ_{max}) faisable.

Ce qui est remarquable dans cette formulation est qu'elle se situe à la confluence entre les modèles systémiques et les modèles sous contraintes. Elle cherche à la fois la répartition des flux de réactions et la répartition des entités qui catalysent ces flux. C'est un point avantageux par rapport aux modèles de type FBA qui ne permettent pas de validation expérimentale quantitative. Là, on peut prédire une répartition en protéines dans une certaine configuration extracellulaire, puis valider cette prédiction expérimentalement, en mesurant les quantités de chaque espèce et en comparant avec la prédiction. C'est ce qui a été fait avec succès dans [35].

Un autre point fort de cette formulation est sa complexité algorithmique théorique. Si l'on pose directement le problème $\max \mu$ sous contraintes (1.2.12), on voit que les termes de demande en machineries moléculaires dans (1.2.12.b) sont quadratiques en les variables d'optimisation $(\mu \times E_i)$. Comme présenté dans [34], le fait de poser le problème en deux temps rend le problème semi-convexe : chaque problème de faisabilité est convexe, et la recherche du μ maximal par dichotomie conserve la caractéristique de semi-convexité.

Par ailleurs, la formulation des contraintes reste générale et systémique dans le sens où ce que l'on dit a un sens à n'importe quelle échelle. Il est possible d'établir une formulation RBA pour des procédés dans une cellule aussi bien que pour des organes dans un organisme multicellulaire. Ce qui compte est la structure du problème, comme présenté dans [33].

Travaux approchants. Certains auteurs ont essayé d'approcher une formulation aussi autonome que celle des RBA. On a déjà évoqué la méthode deFBA introduite dans [83], mais les contraintes RBA y sont mélangées aux contraintes FBA avec des bornes α_i et β_i sur les flux. Ceci ne permet pas de classer cette méthode du côté des méthodes autonomes car il n'est pas possible de construire un modèle de cellule uniquement à partir de considérations physico-biologiques : on a nécessairement besoin de bornes pour les flux

pour construire ce modèle.

De la même manière, les modèles CAFBA (*Constrained Allocation Flux Balance Analysis*) publié dans [61] et ME-Model (*Metabolism & gene Expression models*) publié dans [64] remplacent tous les deux les contraintes sur les flux de la formulation FBA classique par des contraintes sur les efficacités de processus et les concentrations des machineries qui les catalysent. Ils introduisent aussi tous les deux une contrainte sur l'occupation limitée des cellules par ces machineries. En revanche, ces contraintes sur l'occupation des cellules sont dans les deux cas déduites d'observations empiriques (par exemple sur la proportionnalité observée entre ARN et protéines (ME-Models), ou entre concentration de ribosomes et taux spécifique de croissance (CAFBA)). Ceci ne permet donc pas non plus de les classer dans la catégorie des modèles autonomes car il y a besoin d'observations propres à la cellule pour établir son modèle.

Aussi, ces articles ne permettent pas toujours de trancher entre ce qui est déduit des observations et ce qui est déduit de la structure de la modélisation.

La méthode cFBA (*conditional FBA*) proposée dans [70] prédit la configuration d'un organisme phototrophe sur une période d'une journée. Les auteurs ont repris les idées directrices des RBA en libérant les contraintes sur les métabolites intermédiaires (précédemment présentées dans la première contrainte (1.2.12.a)) et en forçant la répartition intracellulaire entre composés à être la même à $t = 24h$ et à $t = 0h$. La maximisation est faite sur l'accumulation de masse sur un jour. Ce travail semble adapté à l'étude d'organisme n'atteignant jamais leur équilibre, comme c'est le cas pour les plantes ou des micro-algues qui ne cessent d'osciller entre fonctionnement diurne et fonctionnement nocturne.

Conclusion à propos des modèles autonomes. En conclusion, les modèles autonomes sont, à notre connaissance, très peu nombreux et peu développés. Ils offrent néanmoins un avantage significatifs pour l'étude des micro-organismes peu connus et pour lesquels il n'est pas possible de définir des paramètres issus d'observations.

1.2.2 Utilisation des modèles intracellulaires pour l'optimisation de souche et de bioprocédés

On a vu dans la section précédente différentes manières de modéliser une cellule. La finalité de ces travaux est variée. Pour optimiser des bioprocédés, nous proposons d'explorer l'utilisation de ces modèles pour la production d'un composé d'intérêt. Ceci revient à modifier le comportement des micro-organismes, en jouant sur leurs gènes (de manière statique ou dynamique) ou en incluant leur description dans une modélisation de l'évolution du bioprocédé. Nous allons voir que la plupart du temps, la modélisation FBA est celle qui est retenue pour représenter le comportement des micro-organismes.

1.2.2.1 Les modèles sous contraintes pour l'optimisation statique de souche

Dans [24], l'auteur propose plusieurs solutions pour augmenter le flux de production d'un composé particulier. Il envisage :

- L'activation sélective de l'expression de certains gènes qui vont dans le sens du caractère désiré. Même si cette méthode semble la plus naturelle, l'auteur nous met en garde en précisant que la sélection peut être difficile vu l'intrication des fonctions des gènes.
- L'augmentation globale de l'expression d'une partie du réseau métabolique. Ceci aurait tendance également à favoriser l'accumulation de biomasse en parallèle de la production du composé d'intérêt. Le rendement ne serait pas significativement meilleur.
- La réduction des phénomènes de rétroaction sur l'efficacité des réactions. Mais il est difficile d'agir sur un grand nombre de réaction, à cause de l'intrication des régulations métaboliques et génétiques (voir [32]).

La première approche est la plus naturelle à aborder avec la formulation FBA, qui n'intègre pas les mécanismes de régulation. Il s'agit alors de chercher une modification de souche en modifiant les bornes des flux α_i et β_i .

Les travaux de l'équipe de Costas Maranas dans les années 2000 ont popularisé cette approche. Ils proposent une formulation à deux niveaux :

- au niveau supérieur, on cherche à maximiser le flux de production d'un composé,
- au niveau inférieur, on subit le fait que la cellule maximise son taux spécifique de croissance instantané compte tenu de la stœchiométrie et les bornes sur les flux.

Nous présentons ici trois méthodes développées par cette équipe.

La méthode OptKnock, développée en 2003 et résumée dans l'article [18], est chronologiquement la première de la série. En reprenant les mêmes notations que précédemment, le problème peut s'écrire sous la forme suivante :

$$\begin{aligned} & \max_y \nu_{prod} \\ & \text{sous contraintes, } \begin{cases} \max_{\nu} \mu \text{ avec : } \begin{cases} \Omega \cdot \nu = 0 \\ \alpha_i(1 - y_i) \leq \nu_i \leq \beta_i(1 - y_i) \end{cases} \\ \sum_j y_j = \kappa \\ y_j \in \{0,1\} \end{cases} \end{aligned} \quad (1.2.13)$$

On s'autorise ici un nombre fixe, κ , d'extinctions de gènes qui correspondent à évaluer les flux à 0.

Cette formulation à deux niveaux est finalement un problème MILP (*Mixed Integer Linear Programming*) étant donné que les contraintes sont toutes linéaires et que l'on cherche des modifications y binaires et des flux éléments de $\mathbb{R}$. Ce problème peut se convertir en un problème d'optimisation à un seul niveau en augmentant les paramètres d'optimisation (y, ν) , puis en reformulant le problème d'optimisation sur le taux spécifique de croissance par les conditions de dualité.

Si l'on exclut d'éteindre les gènes connus pour être nécessaires à la vie de la cellule, on diminue drastiquement l'espace de recherche, et le problème semble se résoudre plus aisément.

Les autres méthodes ne sont que des extensions de ce cas d'étude.

OptStrain [68], par exemple, envisage des délétions de gènes et des ajouts de voies qui seraient favorables. Cette méthode se place dans le même cadre que OptKnock et

propose de fixer les bornes de certains flux natifs à zéro et de fixer à des valeurs à déterminer des flux qui ne sont pas natifs, *i.e.* les flux de voies de synthèse.

OptForce [69] propose d'inactiver l'expression de gènes ou de les réguler positivement ou négativement. En raisonnant sur les systèmes de régulation, on voit apparaître des stratégies équivalentes, comme par exemple augmenter l'expression d'un gène ou diminuer l'effet de son inhibiteur. Ces classes d'équivalence sont fournies par la méthode.

Des travaux s'inspirent de ces méthodes, comme notamment [55] qui cherche la configuration génétique optimale par algorithme génétique, ou où la prise en compte de la viabilité complète les contraintes classiques [84].

[39] donne une synthèse des méthodes développées entre 2003 et 2010. En conclusion de cet article, on trouve une discussion sur le fait que l'optimisation dynamique est incomparablement plus efficace que les méthodes d'optimisation statique car elle permet d'envisager une succession de fonctionnement pour la cellule : passer d'une phase de croissance à une phase de production, par exemple.

1.2.2.2 Optimisation dynamique

L'optimisation de souche statique, *i.e.* le fait d'agir sur les expressions des gènes, est un moyen trop rigide d'agir sur les cellules. En effet, on a vu, pour le cas de la production macroscopique de composés d'intérêt en bioprocédé, que la succession d'une phase de croissance puis de production était préférable à une unique phase de croissance et de production simultanée+6.

Il est donc préférable d'agir sur les cellules dynamiquement pour avoir une succession de comportements différents.

C'est le problème auquel se confrontent les auteurs de l'article [27], qui sont en partie ceux qui ont introduit la méthode dFBA dans [56]. Le problème est posé de la manière suivante :

$$\begin{aligned} & \max_{\nu_1^m(t_1), \dots, \nu_1^m(t_f), \dots, \nu_{n_m}^m(t_1), \dots, \nu_{n_m}^m(t_f)} P(t_f) \\ \text{sous contraintes, } & \left\{ \begin{array}{l} \max_{\nu(t_k)} \mu(t_k) \\ \text{avec : } \left\{ \begin{array}{l} \Omega \cdot \nu(t_k) = 0 \\ \alpha_i \leq \nu_i(t_k) \leq \beta_i \\ \nu_i^m(t_k) \text{ donné} \end{array} \right. \\ \text{Dynamique sur les substrats, le produit et la biomasse} \\ \alpha_i^m \leq \nu_i^m(t_k) \leq \beta_i^m \end{array} \right. \end{aligned} \quad (1.2.14)$$

où les éléments de ν^m sont les flux dont on autorise la manipulation. Les valeurs de ces flux sont des paramètres pour les problèmes FBA de niveau inférieur, résolus à chaque instant d'échantillonnage.

On reconnaît une formulation dFBA séquentielle.

La solution à ce problème réintroduit la stratégie en deux temps (croissance puis production). Les auteurs ont remarqué que la transition entre les deux phases était instantanée si l'on n'ajoute pas de contrainte explicite sur la différence $\nu_i^m(t_k) - \nu_i^m(t_{k-1})$ comme on peut le voir dans certaines formulations dFBA.

Aussi, ils ont cherché l'instant de commutation t_s entre les deux stratégies, en pré-supposant que les configurations de croissance et de production étaient connues. Ceci revient à chercher t_s et t_f solutions de :

$$\begin{aligned} & \max_{t_s, t_f} P(t_f) \\ & \text{sous contraintes, } \left\{ \begin{array}{l} \max_{\nu(t_k)} \mu(t_k) \\ \text{avec : } \left\{ \begin{array}{l} \Omega \cdot \nu(t_k) = 0 \\ \alpha_i \leq \nu_i(t_k) \leq \beta_i \\ \nu_i^m(t_k) = \nu_i^{m,-} \text{ pour } t_k \leq t_s \\ \nu_i^m(t_k) = \nu_i^{m,+} \text{ pour } t_k > t_s \end{array} \right. \\ \text{Dynamique sur les substrats, le produit et la biomasse} \end{array} \right. \end{aligned} \quad (1.2.15)$$

avec les couples $(\nu_i^{m,-}, \nu_i^{m,+})$ connus pour tous les flux sur lesquels on suppose pouvoir agir. La configuration ν^- correspond à une configuration de croissance et celle ν^+ à un état de production. On force un flux nécessaire à la croissance à valoir 0 dans la phase de croissance et on impose un flux de production à atteindre une certaine valeur dans la configuration ν^+ .

Le temps de calcul pour résoudre le problème passe de l'ordre de l'heure à quelques secondes. On ne cherche donc plus des trajectoires sur les flux mais des configurations de croissance et de production.

Les configurations avant et après t_s sont calculées par les méthodes d'optimisation statiques de gènes vues précédemment. C'est ce qui est sous-entendu dans les articles.

La suite de ce travail se trouve dans [28] où cette fois, il est question d'optimiser la consommation de substrat, avec l'idée que l'on commute entre deux fonctionnements à un instant t_s mais que l'on peut changer les micro-organismes de milieu de culture à tout instant pour limiter les effets inhibants de composés présents dans le milieu.

Le travail présenté dans [41] explore l'utilisation d'un critère alternatif : la productivité, *i.e.* le rapport entre la quantité produite de composé d'intérêt et le temps nécessaire pour le produire. Ce critère est repris dans [42], où l'on connaît à l'avance les flux que l'on va substituer à l'instant t_s , comme dans (1.2.15). Les contraintes qui ressortent du problème interne sont une sorte de FBA augmentée de contraintes sur les protéines (cf. [83]).

Les auteurs de [42] explorent différentes stratégies de modification de souche :

- la délétion de gènes statiques,
- la répression de l'expression de gène,
- la répression de flux métabolique,
- une combinaison de délétion statique avec une stratégie binaire sur la présence d'oxygène dans le milieu.

En conclusion, la répression statique apparaît plus efficace pour la production du com-

posé d'intérêt (alors que l'on ne peut pas avoir deux phases), et les résultats numériques ne paraissent pas en accord avec le fonctionnement observé de micro-organisme (pas de phase exponentielle pour la phase supposée de croissance). Il n'y a pas dans cet article de discussion sur la faisabilité de l'implémentation.

Donner la possibilité à des flux de commuter instantanément d'une valeur à une autre (voire forcer les flux à évoluer de cette manière), revient à une sorte de transmutation des protéines. Ceci équivaut à négliger l'inertie que subit une cellule (inertie due aux protéines qu'elle produit dans son cytosol) ou à oublier le fait qu'une cellule ne peut se défaire d'une protéine qu'en la diluant, c'est-à-dire en croissant.

Avec la formulation dFBA (1.2.15), on force la répartition des flux à commuter d'une configuration à une autre. Comme un flux ne peut se faire sans sa protéine, il faut donc que les protéines qui catalysent les flux qui basculent se métamorphosent instantanément pour catalyser une réaction, puis une autre.

C'est biologiquement impossible.

Une autre approche pour traiter du problème de l'optimisation dynamique des cellules est présentée dans [22].

Le modèle simplifié de la cellule qui y est décrit est repris de [30] et [87]. Cette modélisation contient de nombreuses simplifications qui l'écartent de la réalité biologique. La cellule y est décrite par deux processus : le métabolisme et l'appareil de traduction. Le métabolisme produit un composé qui peut par la suite être transformé en protéines du métabolisme ou en protéines de l'appareil de traduction en fonction d'un paramètre. Les travaux présentés dans [30] et [87] cherchent une relation entre la composition en protéines métaboliques et protéines de l'appareil de traduction qui soit fonction de la concentration en substrat disponible pour croître. Ils supposent que le flux de consommation du composé intermédiaire est divisé entre α et $(1 - \alpha)$ pour la partie métabolique et la partie traduction. Cet α est ensuite traduit en fonction de paramètres intracellulaires. Ainsi, le modèle semble autonome.

Dans [22], l'auteur ajoute une protéine de synthèse et cherche comment détourner des ressources pour la produire. Étant donné que le modèle est simplifié, les auteurs ne disposent pas de leviers réalistes pour ajouter un gène et pour que la cellule le transcrive puis le traduise naturellement. Ils sont obligés de recourir à des mécanismes pour reconcevoir la souche entière de zéro, ou pour ajouter un nouvel appareil de traduction qui fonctionne de manière orthogonale à l'appareil natif, en plus du gène codant pour la fonction désirée.

Cette approche semble difficilement implémentable.

1.2.2.3 Couplage avec un modèle de bioréacteur

Pour conclure nous présentons maintenant une partie des couplages qui ont été entrepris entre modèles FBA et descriptions macroscopiques.

Une façon naturelle de coupler ces modèles est de partir d'un modèle macroscopique où l'on suppose connus les états, c'est-à-dire les concentrations en substrats, en produits et en biomasse. On donne en entrée ces concentrations à un modèle FBA, qui suppose

que les cellules sont en régime quasi stationnaire et peuvent inférer un taux spécifique de croissance et d'éventuels flux de production de produits. Cette logique est présentée figure 1.8. C'est exactement ce qui est proposé dans [75] pour comprendre la production de NADPH lors de la consommation de furfural.

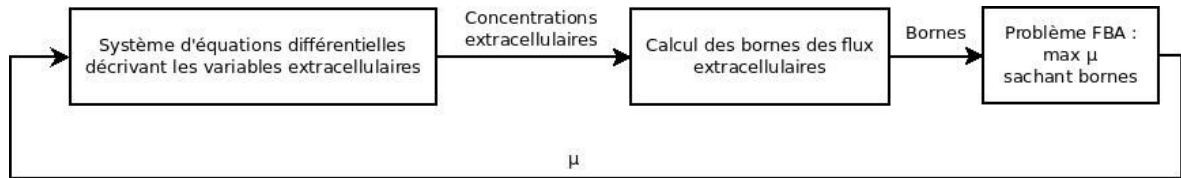


FIGURE 1.8 – Schéma de fonctionnement du couplage entre un problème FBA et une description. Les signaux sont discrétisés. A chaque pas de temps, le bloc de droite résout un problème FBA discret.

L'auteur de [8] fait de même avec un modèle à deux sources dans le but de caractériser proprement le mécanisme d'overflow.

[17] puis [76] proposent tous les deux de combiner la prédiction de l'état du métabolisme faite par une méthode FBA (ou plutôt dFBA) avec une commande prédictive en ligne. Dans [17], les performances de commande sont comparées avec une commande prédictive sur modèle macroscopique et avec une commande en boucle ouverte. Dès lors que les paramètres intracellulaires d'absorption de substrat ou d'oxygène s'écartent de 20% de leur valeur nominale, les modèles macroscopiques sont moins performants que la commande prédictive couplée au modèle FBA.

C'est une perspective encourageante pour le couplage entre modèles intracellulaires et modèles macroscopiques.

1.3 Conclusion du chapitre

Nous avons vu l'approche macroscopique classique de modélisation de bioprocédés par les méthodes de génie des procédés, qui développent un bilan de matière plus ou moins détaillé pour rendre compte des échanges qui ont lieu dans le bioréacteur. Ces méthodes donnent ensuite une formulation mathématique à la vitesse de chaque réaction introduite. La formulation mathématique est plus ou moins riche en fonction des phénomènes qu'ils prennent en compte.

Nous avons vu comment sont utilisés ces modèles pour définir des stratégies de commande de réacteurs fed-batch pour la production de composés d'intérêt. Les problèmes d'optimisation sont bien adaptés à la définition de stratégie de commande, mais qu'il était nécessaire de boucler la réponse du bioprocédé réel, étant donné l'écart entre le modèle et la réalité. Par exemple, la commande prédictive ou la commande linéarisante permettent ce bouclage et le suivi de la trajectoire optimale.

Ces descriptions macroscopiques de bioprocédés sont d'une bonne fiabilité, mais ne permettent pas d'agir sur le bioprocédé par le contrôle des cellules elles-mêmes. Pour cela, il est nécessaire de définir plus clairement le rôle et le fonctionnement des cellules

dans le bioprocédé. Les méthodes présentées dans la deuxième partie de ce chapitre, répondent à ce problème.

En effet, nous avons présenté par la suite des modèles systémiques de micro-organismes qui se focalisent sur le comportement des composants de la cellule dans des conditions de fonctionnement général. D'autres modélisations s'appuient sur un critère d'optimalité du comportement cellulaire, ce sont les méthodes dérivées de la formulation FBA. Le critère relaxe la définition que l'on a des éléments de la cellule. En revanche, la formulation des contraintes est un réel problème car il est aisé d'introduire des contraintes artificielles dans le seul but de se rapprocher des observations biologiques.

Enfin, nous avons présentés les modèles autonomes qui allient une part de représentation systémique et une part de représentation sous contrainte pour s'abstenir de faire appel à des artifices de modélisation tout en gardant une certaine souplesse de modélisation.

Pour finir, nous avons vu comment il était possible de faire appel à ces modélisations pour envisager l'optimisation de la bioproduction d'un composé d'intérêt. Nous avons passé en revue des travaux qui s'intéressent à l'intervention au niveau des gènes en supprimant ou en ajoutant des fonctions à la cellule, que ce soit de manière statique ou dynamique. Puis nous avons présenté des travaux qui mettent en relation des modélisations de cellules sous contraintes avec une représentation macroscopique (et automatique) de bioprocédé.

La mise en parallèle des travaux de modélisation macroscopique de bioprocédé avec les avancées en terme de modélisation intracellulaires nous mène à croire qu'il est pertinent de coupler la description macroscopique d'un bioprocédé avec une représentation du comportement des micro-organismes qui y prennent part. L'objectif est de pouvoir agir sur le bioprocédé pour optimiser la conduite du procédé et la conception de la souche.

En conséquence, nous développons dans le chapitre suivant une modélisation qui combine les deux échelles. Le cahier des charges à satisfaire est :

- La modélisation doit saisir les aspects dynamiques concernant l'évolution du milieu, du comportement des micro-organismes.
- La partie de la modélisation concernant les micro-organismes doit être autonome et viser à limiter les artifices de modélisation. Elle doit également faire part des contraintes biologiques qui agissent sur le comportement des micro-organismes.
- L'origine des phénomènes doit respecter la réalité. Le taux spécifique de croissance, les rendements et les échanges de matière entre les micro-organismes et leur environnement doivent être exclusivement le résultat de phénomènes intracellulaires.
- La modélisation doit prendre en compte le fait que les commandes du bioprocédés sont à la fois des commandes de son environnement (typiquement, le débit d'alimentation en milieu de culture) et des commandes du comportement cellulaire (typiquement, l'expression des gènes).
- L'objectif de la modélisation n'est pas de prendre en compte la globalité du comportement cellulaire mais d'en faire ressortir l'essence du fonctionnement. En d'autres termes, la modélisation doit être simple. On doit comprendre dans quelles conditions le couplage est envisageable.

Aucune des modélisations passées en revue ici ne permet de satisfaire la totalité de ces spécifications.

Le prochain chapitre présente les fondements d'une proposition de couplage qui repose sur la méthode RBA.

Chapitre 2

Modélisation et optimisation du comportement de micro-organismes

Dans le chapitre précédent, nous avons vu différentes approches classiques pour optimiser les bioprocédés : d'un côté les modèles macroscopiques adaptés à la commande et à la régulation du milieu de culture, de l'autre les modèles intracellulaires sous contraintes qui donnent une description fiable de la croissance et du comportement des micro-organismes à l'échelle intracellulaire. Les modèles macroscopiques sont d'un grand intérêt par leur simplicité mais le rôle des cellules est limité à leur concentration dans le milieu. Les modèles intracellulaires sont beaucoup plus précis sur la description des cellules en intégrant les mécanismes essentiels de la Vie, mais leur complexité ne permet pas de les utiliser pour la conduite du procédé en temps réel. Pour proposer une nouvelle méthode d'optimisation de la bioproduction de composé d'intérêt à travers la conduite du procédé et l'optimisation de souche, nous avons montré le besoin d'un modèle à la confluence entre la description macroscopique du procédé et la description des mécanismes intracellulaires qui prendrait le meilleur des deux approches. L'objectif principal de cette modélisation est alors de centrer la description du procédé macroscopique sur les mécanismes intervenant à l'intérieur des cellules.

Cet objectif fait l'objet de ce chapitre, au sein duquel nous proposons une représentation simplifiée et agrégée de la cellule. Dans un premier temps, en section 2.1, nous présentons un modèle agrégé de la cellule qui capte l'essence du problème d'allocation de ressource dans la cellule. Ce modèle sera utile pour comprendre les enjeux de notre problème : l'optimisation de la bioproduction d'un composé d'intérêt en bioréacteur. Le formalisme de ce modèle rend compte des principes généraux régissant le fonctionnement des cellules et du couplage avec leur environnement. Une fois le modèle exposé, une étude est proposée en section 2.2 pour montrer qu'il s'agit bien à la fois d'un prolongement de modèles macroscopiques et dans le même temps de l'extension de modèles intracellulaires validés expérimentalement. Enfin, en section 2.3, nous proposons une solution au problème d'optimisation de la production d'un composé d'intérêt en utilisant le modèle développé auparavant. Ceci permet d'envisager la trajectoire optimale des processus intracellulaires pris en compte dans l'aggrégation de la section 2.1. Nous donnons ainsi le premier pas vers la résolution du problème global : l'optimisation de la production de composé d'intérêt par le contrôle en temps réel des processus intracellulaires.

2.1 Modèle agrégé

2.1.1 Objectifs de la modélisation et de l'agrégation

L'objectif est de réaliser le couplage entre un modèle macroscopique de bioréacteur et un modèle intracellulaire de bactérie. Nous avons vu que les modèles macroscopiques résument le fonctionnement très complexe des cellules à leur simple concentration dans le milieu, alors même que des progrès phénoménaux sont réalisés en modélisation pour comprendre et prédire le fonctionnement général des micro-organismes aux échelles intracellulaires.

Pour coupler les deux approches, même si nous pourrions techniquement aborder un modèle à l'échelle du génome et le coupler avec le milieu extérieur, nous ne souhaitons pas nous intéresser en détail à la complexité de la cellule mais plutôt au principe de fonctionnement de celle-ci. La cellule est en effet un système bien particulier, capable de fabriquer tous ses composants (sa membrane, son ADN, ses protéines, *etc.*, voir annexe A) à partir des espèces chimiques à sa disposition dans son milieu (du glucose, du dioxygène, des ions, *etc.*). C'est le principe même de la Vie et une condition nécessaire pour la reproduction des organismes. C'est un élément que nous souhaitons faire ressortir de notre modélisation.

En outre, les actions élémentaires que la cellule est en mesure de réaliser sont, pour la quasi totalité, sujettes à l'action de protéines qui catalysent les processus intracellulaires. Il est donc important que les protéines aient un rôle de première importance dans la modélisation adoptée.

Au vu des considérations sur le manque d'autonomie des modèles intracellulaires de la littérature présentés au chapitre précédent, il nous apparaît nécessaire de développer une nouvelle modélisation autonome et dynamique de la cellule.

Les contraintes biologiques auxquelles sont sujettes les cellules sont très complexes et nous retiendrons l'ensemble de contraintes présentées dans le modèle RBA, qui se sont avérées capter de manière autonome l'essence du problème d'allocation après validation expérimentale [35].

2.1.2 Formalisation du problème

On considère une unique espèce bactérienne immergée dans un bioréacteur. Ce dernier est supposé fonctionner en mode batch, c'est-à-dire sans ajout ou retrait de matière. Le milieu, de volume constant noté V , est supposé parfaitement agité et homogène. On considère qu'il n'y a qu'un substrat $\mathcal{G}$ disponible dans le milieu et qu'il est suffisant pour assurer la croissance des bactéries.

Les bactéries sont supposées toutes identiques et on résume l'étude de la population à un individu représentatif. On parlera indistinctement de la cellule, des cellules ou de la population.

Du point de vue de la biologie des systèmes, la cellule est décrite à travers trois processus cellulaires principaux :

- (i) le premier, noté Σ_T , regroupe tous les processus cellulaires liés à l'import de matière et au catabolisme, c'est-à-dire les processus (principalement enzymatiques) impliqués dans la production d'énergie et des briques élémentaires nécessaires à la survie et à la croissance des bactéries ;
- (ii) le second, noté Σ_R , regroupe les processus nécessaires à la synthèse des protéines, c'est-à-dire l'appareil de traduction, y compris les ribosomes et un grand nombre de protéines accessoires, comme par exemple les facteurs d'élongation ;
- (iii) le troisième, noté Σ_B , comprend tous les processus cellulaires impliqués dans la production de macrocomposants cellulaires qui ne sont pas des protéines comme l'ADN, la membrane, la paroi cellulaire et les ARN.

Dans la suite, dans un but de lisibilité, nous supposons que les processus sont catalysés par des machines moléculaires composées exclusivement de protéines. On notera $\mathcal{E}_i$ la protéine qui catalyse la processus Σ_i .

Dans ce contexte, le substrat $\mathcal{G}$ est importé et transformé par Σ_T , en un métabolite interne $\mathcal{S}$. Le métabolite $\mathcal{S}$ peut être utilisé indifféremment par Σ_R et Σ_B afin de construire respectivement des protéines ou d'autres macrocomposants cellulaires $\mathcal{B}$ nécessaires à la croissance cellulaire. Nous associons alors pour chaque unité de $\mathcal{E}_T$, $\mathcal{E}_R$ et $\mathcal{E}_B$:

- un coût de ressource (notation introduite par la suite), correspondant au nombre de $\mathcal{S}$ requis pour la construction de chaque unité,
- une efficacité, k_i , du processus Σ_i .

Enfin, puisque le but du modèle est d'optimiser la production d'un produit d'intérêt $\mathcal{P}$, nous ajoutons un processus cellulaire supplémentaire, appelé Σ_P et catalysé par la protéine $\mathcal{E}_P$, pour produire $\mathcal{P}$ à partir de $\mathcal{S}$ et le sécréter dans le milieu. Nous associons enfin à $\mathcal{E}_P$, son coût en ressources et son efficacité, k_P .

On note ν_i le flux de matière qui passe à travers le procédé Σ_i .

En introduisant α_i les coefficients de rendement des différents processus, les schémas réactionnels s'écrivent comme suit :

- $\Sigma_T : 1 \cdot \mathcal{G} \xrightarrow{\nu_T} \alpha_T \cdot \mathcal{S}$: import de $\mathcal{G}$ depuis le milieu extracellulaire et production de $\mathcal{S}$
- $\Sigma_P : 1 \cdot \mathcal{S} \xrightarrow{\nu_P} \alpha_P \cdot \mathcal{P}$: production et sécrétion de $\mathcal{P}$ à partir de $\mathcal{S}$
- $\Sigma_B : 1 \cdot \mathcal{S} \xrightarrow{\nu_B} \alpha_B \cdot \mathcal{B}$: production des macrocomposants $\mathcal{B}$ à partir de $\mathcal{S}$
- $\Sigma_R : 1 \cdot \mathcal{S} \xrightarrow{\nu_R} \alpha_{E_i} \cdot \mathcal{E}_i$: production des protéines catalysant le processus Σ_i , pour $i \in \{T, P, B, R\}$. Ainsi, $\frac{1}{\alpha_{E_i}}$ est le coût en ressource de l'unité $\mathcal{E}_i$ introduit plus haut.

A noter que le processus Σ_R est la conjonction des différents processus de production de chaque protéine particulière. On divisera le flux ν_R qui passe à travers Σ_R en quatre flux, ν_{E_i} , qui produisent chacun une protéine, $\mathcal{E}_i$:

$$\nu_R = \nu_{E_T} + \nu_{E_B} + \nu_{E_P} + \nu_{E_R} \quad (2.1.1)$$

avec les schémas réactionnels unitaires, $1 \cdot \mathcal{S} \xrightarrow{\nu_{E_i}} \alpha_{E_i} \cdot \mathcal{E}_i$ pour $i \in \{T, P, B, R\}$.

Le schéma figure 2.1 résume les transferts de matière entre les différentes entités.

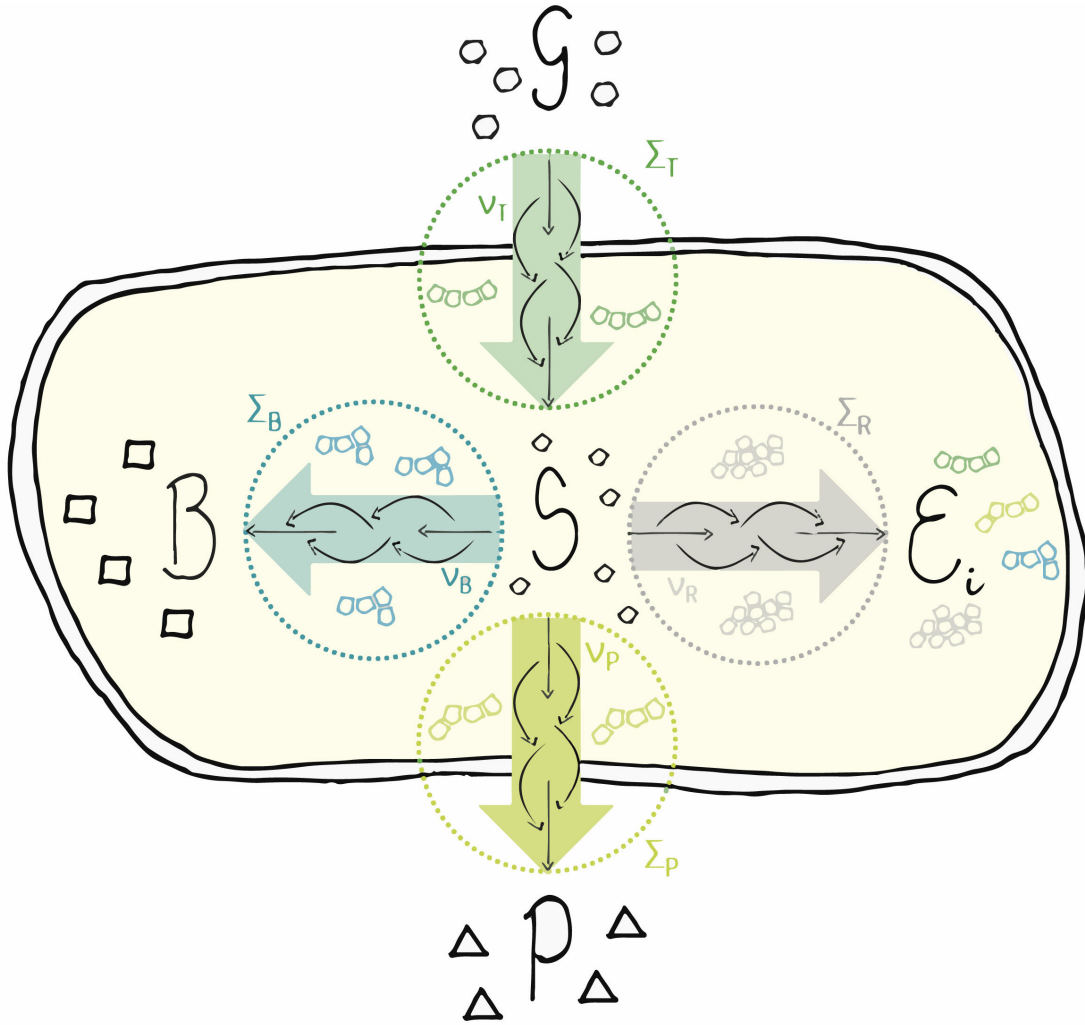


FIGURE 2.1 – Présentation des différents flux et processus intracellulaires du modèle agrégé.

2.1.3 Dynamiques & contraintes

On cherche à caractériser l'évolution des concentrations des différents composés en présence.

Concernant les composés $\mathcal{G}$ et $\mathcal{P}$, on exprimera leur concentration en quantité de matière par unité de volume du milieu de culture. L'unité de concentration est la millimole par litre de réacteur ($mmol.L^{-1}$). Les composés intracellulaires, $\mathcal{S}$, $\mathcal{B}$ et les $\mathcal{E}_i$, seront eux caractérisés par leur quantité de matière ramenée par unité de masse sèche de cellules. Ceci est justifié par le fait que les mesures sur les espèces intracellulaires sont habituellement faites de cette manière. L'unité de concentration des espèces intracellulaire est alors : $mol.g_{CDW}^{-1}$ (mole par gramme de matière sèche, *gram of Cell Dry Weight*, en anglais). Dans le but de ne pas introduire de coefficient de conversion, la biomasse, notée X (comme classiquement dans la littérature relative à la conduite du procédé), sera caractérisée par sa concentration en masse sèche par unité de volume du milieu de culture, c'est à dire en gramme de masse sèche par litre de réacteur.

Dans la suite, nous noterons G (respectivement P , S , B , E_B , E_P , E_R et E_T) la concentration de l'espèce $\mathcal{G}$ (respectivement $\mathcal{P}$, $\mathcal{S}$, $\mathcal{B}$, $\mathcal{E}_B$, $\mathcal{E}_P$, $\mathcal{E}_R$ et $\mathcal{E}_T$).

2.1.3.1 Dynamiques de S et de G

En considérant le bilan de matière des différents processus pris en considération dans la description, l'équation différentielle régissant l'évolution de la concentration S de $\mathcal{S}$ est donnée par :

$$\dot{S}(t) = \alpha_T \nu_T(t) - \nu_B(t) - \nu_P(t) - \nu_{E_T}(t) - \nu_{E_B}(t) - \nu_{E_P}(t) - \nu_{E_R}(t) - \mu(t)S(t) \quad (2.1.2)$$

Le terme $-\mu(t)S(t)$ prend en compte l'augmentation du volume de biomasse avec le temps, avec $\mu(t)$ le taux spécifique de croissance, défini plus spécifiquement en section 2.1.4. C'est ce que l'on appelle le terme de dilution.

$\nu_T(t)$ est le flux positif de matière qui passe à travers le processus Σ_T , c'est-à-dire la quantité de $\mathcal{G}$ importée et transformée en $\mathcal{S}$ à l'instant t , par unité de biomasse et de temps. On le retrouve également dans la dynamique de la concentration en $\mathcal{G}$ dans le milieu extracellulaire :

$$\dot{G}(t) = -\nu_T(t)X(t) \quad (2.1.3)$$

Ce flux d'import est contraint par la capacité maximale du processus Σ_T , qui vaut la concentration de la protéine $\mathcal{E}_T$ multipliée par son efficacité, k_T , du processus Σ_T . Ceci mène à une première contrainte :

$$\nu_T(t) \leq k_T E_T(t) = \frac{v_{m,T} G(t)}{G(t) + K_T + K_S S(t)} E_T(t) \quad (2.1.4)$$

avec E_T la concentration en protéines catalysant le processus Σ_T . L'efficacité k_T du processus Σ_T suit une cinétique du premier ordre par rapport à la concentration en $\mathcal{G}$, c'est-à-dire qu'elle est hyperbolique par rapport à la concentration G (asymptote linéaire en $G = 0$ et asymptote horizontale en $G \rightarrow +\infty$). Les paramètres $v_{m,T}$, K_T et K_S sont tous positifs. $v_{m,T}$ est l'efficacité maximale du processus Σ_T . K_S introduit une inhibition par le produit du processus, *i.e.* $\mathcal{S}$. Ainsi, l'efficacité diminue lorsque la concentration en $\mathcal{S}$ augmente.

2.1.3.2 Synthèse des macrocomposants $\mathcal{B}$ par Σ_B

L'équation différentielle associée à la production de macrocomposants $\mathcal{B}$ est donnée par :

$$\dot{B}(t) = \alpha_B \nu_B(t) - \mu(t)B(t) \quad (2.1.5)$$

avec ν_B le flux positif de matière qui passe à travers le processus Σ_B , et $-\mu(t)B(t)$ le terme de dilution.

La concentration des différents composés inclus dans $\mathcal{B}$, c'est-à-dire l'ADN, les ARN, les composés de la membrane et de la paroi, est périodique en fonction du cycle cellulaire. Ainsi, si l'on raisonne sur un individu et sur un cycle cellulaire, il faudrait imposer une contrainte de périodicité à la concentration B . Cependant, comme nous nous plaçons à l'échelle de la population et sur une période de temps bien plus grande qu'un cycle cellulaire, nous pouvons considérer que la concentration de $\mathcal{B}$ peut être assimilée à sa

valeur moyenne sur le cycle cellulaire et supposée constante. Cette valeur moyenne est supposée la même dans toutes les conditions de croissance, comme indiquée dans [65].

Ceci donne directement à partir de (2.1.5) la contrainte d'égalité entre production et dilution de $\mathcal{B}$:

$$\nu_B(t) = \frac{B_0}{\alpha_B} \mu(t) \quad (2.1.6)$$

où B_0 est la concentration nominale en $\mathcal{B}$.

De plus, la capacité du processus Σ_B est limitée, ce qui mène à une deuxième contrainte d'inégalité :

$$\nu_B(t) \leq k_B E_B(t) = \frac{v_{m,B} S(t)}{S(t) + K_B} E_B(t) \quad (2.1.7)$$

avec E_B la concentration en protéines catalysant le processus Σ_B et l'efficacité du processus Σ_B , k_B , qui suit une relation de Michaelis-Menten (justifiée par la suite) de coefficients $v_{m,B}$ et K_B par rapport au substrat de la réaction, $\mathcal{S}$.

Dans le même esprit que (2.1.4), on pourrait introduire une inhibition par le produit du processus, en l'occurrence $\mathcal{B}$. Étant donné que la concentration en $\mathcal{B}$ est supposée constante, ceci modifie la constante K_B mais pas le modèle utilisé.

2.1.3.3 Synthèse de $\mathcal{P}$ par Σ_P

D'une manière équivalente, l'équation différentielle associée à la production de $\mathcal{P}$ est donnée par :

$$\dot{P}(t) = \alpha_P \nu_P(t) X(t) \quad (2.1.8)$$

où le flux de production de $\mathcal{P}$ par unité de milieu de culture est le produit entre la production de $\mathcal{P}$ par unité de volume cellulaire ($\alpha_P \nu_P$) et la concentration de cellule par unité de milieu de culture (X).

Comme pour les autres processus, le flux ν_P est limité par l'efficacité du processus Σ_P comme suit :

$$\nu_P(t) \leq k_P E_P(t) = \frac{v_{m,P} S(t)}{S(t) + K_P} E_P(t) \quad (2.1.9)$$

avec E_P la concentration en protéines catalysant Σ_P et k_P l'efficacité de Σ_P , qui suit une relation de Michaelis-Menten (justifiée par la suite) par rapport au substrat $\mathcal{S}$.

2.1.3.4 Synthèse des protéines par Σ_R

Enfin, toutes les protéines $\mathcal{E}_i$, $i \in \{T, B, P, R\}$ impliquées dans les processus cellulaires sont produites par Σ_R , ce qui conduit aux équations différentielles suivantes :

$$\dot{E}_i(t) = \alpha_{E_i} \nu_{E_i}(t) - \mu(t) E_i(t), i \in \{T, B, P, R\} \quad (2.1.10)$$

où $\frac{1}{\alpha_{E_i}}$ est le coût nécessaire pour construire la protéine $\mathcal{E}_i$ à partir de $\mathcal{S}$.

Une des conséquences de cette formulation est que l'on considère que la dilution est prépondérante et même la seule manière de faire diminuer la concentration d'une protéine. Les protéines étant des structures très stables dont la constante de temps de dégradation est de l'ordre de plusieurs dizaines d'heures [2], on négligera donc dans cette modélisation la dégradation active des protéines par les protéases.

La contrainte de capacité du processus Σ_R s'exprime cette fois sur la somme des flux qui passent à travers, c'est-à-dire la somme des flux ν_{E_i} :

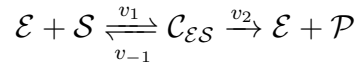
$$\sum_{P,B,R,T} \nu_{E_i}(t) \leq k_R E_R(t) = \frac{v_{m,R} S(t)}{S(t) + K_R} E_R(t) \quad (2.1.11)$$

avec E_R la concentration de la protéine qui catalyse le procédé Σ_R et l'efficacité k_R de Σ_R qui suit une relation de Michaelis-Menten par rapport à la concentration de $\mathcal{S}$.

2.1.3.5 Justification biologique des efficacités

Le fait que l'efficacité de Σ_R soit croissante avec la concentration de $\mathcal{S}$ et bornée permet de rendre compte du fait que les ribosomes sont plus efficaces en milieux riches qu'en milieux pauvres, comme indiqué dans [57] et [13]; et par ailleurs que les concentrations des métabolites sont globalement plus élevées en milieu riche qu'en milieu pauvre [7].

L'argument est légèrement différent pour les efficacités k_B et k_P qui sont aussi croissantes en S et bornées. Ceci fait écho au fait que les processus Σ_B et Σ_P sont des processus catalytiques et que leurs efficacités suivent des cinétiques enzymatiques usuelles (en l'occurrence de type Michaelis-Menten par rapport au substrat de la réaction enzymatique). La formulation de Michaelis-Menten provient quant à elle de la formulation des réactions enzymatiques comme étant une succession de deux réactions dont une est irréversible. En notant $\mathcal{E}$ la protéine qui catalyse une réaction de transformation d'un composé $\mathcal{S}$ en un composé $\mathcal{P}$, on peut introduire un composé $\mathcal{C}_{\mathcal{E}\mathcal{S}}$ qui est le complexe que forme la protéine avec $\mathcal{S}$. On peut alors détailler la réaction $\mathcal{S} \rightarrow \mathcal{P}$ comme suit :



où la première réaction est réversible (le complexe peut se désagréger) mais où la seconde ne l'est pas (la protéine $\mathcal{E}$ ne peut transformer $\mathcal{P}$ en $\mathcal{C}_{\mathcal{E}\mathcal{S}}$). Dès lors, en prenant pour les vitesses de réaction $v_1 = k_1[\mathcal{E}][\mathcal{S}]$, $v_{-1} = k_{-1}[\mathcal{C}_{\mathcal{E}\mathcal{S}}]$, $v_2 = k_2[\mathcal{E}][\mathcal{P}]$, où la notation entre crochets signifie la concentration, et en supposant que l'espèce $\mathcal{C}_{\mathcal{E}\mathcal{S}}$ est à l'équilibre, on peut montrer que la vitesse de la réaction $\mathcal{S} \rightarrow \mathcal{P}$ lorsque la concentration en $\mathcal{P}$ est nulle, est donnée par $v = \frac{k_2[\mathcal{S}]}{[\mathcal{S}] + \frac{k_2+k_{-1}}{k_1}}[\mathcal{E}]$. C'est la formulation de Michaelis-Menten. Dès lors que la concentration en $\mathcal{P}$ n'est pas nulle, la vitesse diminue.

Pour le processus Σ_T on prend une efficacité proche de celle de Michaelis-Menten en son substrat $\mathcal{G}$ avec un terme d'inhibition sur le produit du processus $\mathcal{S}$ (voir équation (2.1.4)). Ce raffinement permet notamment de prendre explicitement en compte le fait que l'augmentation de la concentration $\mathcal{S}$ rend plus difficile l'import de matière dans l'enceinte de la cellule.

2.1.4 Focus sur l'expression du taux spécifique de croissance

Il a été montré que la densité intracellulaire de bactéries comme *E. coli* ou *B. subtilis* est constante pour n'importe quelle condition de croissance et tout au long du cycle cellulaire, voir [50] et [49]. Ainsi, en suivant l'approche de [57], qui permet d'exprimer

la densité intracellulaire D_0 comme étant le résultat des contributions des différents composés peptiques (les protéines), il vient par définition de la densité :

$$\sum_{P,B,R,T} \frac{1}{\alpha_{E_i}} E_i(t) = D_0 \quad (2.1.12)$$

Les coûts $\frac{1}{\alpha_{E_i}}$ étant sans unité, D_0 s'exprime de la même manière que les concentrations E_i , *i.e.* en $mol.g_{CDW}^{-1}$.

Pour maintenir sa densité constante, la cellule doit donc augmenter son volume si de nouvelles protéines sont produites. Ceci revient à contraindre le taux spécifique de croissance des bactéries selon :

$$\mu(t) = \frac{1}{D_0} \sum_{P,B,R,T} \nu_{E_i}(t) \quad (2.1.13)$$

En effet, en dérivant (2.1.12) par rapport au temps,

$$\sum_{P,B,R,T} \frac{1}{\alpha_{E_i}} \dot{E}_i(t) = 0$$

Puis, en substituant $\dot{E}_i$ par leur expressions (2.1.10),

$$\sum_{P,B,R,T} \frac{1}{\alpha_{E_i}} (\alpha_{E_i} \nu_{E_i}(t) - \mu(t) E_i(t)) = 0$$

ou encore,

$$\sum_{P,B,R,T} \nu_{E_i}(t) = \mu(t) \underbrace{\sum_{P,B,R,T} \frac{1}{\alpha_{E_i}} E_i(t)}_{=D_0, \text{ d'après (2.1.12)}}$$

On trouve bien la relation (2.1.13).

La relation (2.1.13) garantit le caractère autonome du modèle : le taux spécifique de croissance est défini par des paramètres associés aux processus intracellulaires.

2.1.4.1 Évolution de la biomasse

D'après la définition du taux spécifique de croissance (2.1.13) et dû au fait qu'il n'y a qu'une seule espèce bactérienne en présence dont la densité est supposée constante au cours du temps, on peut affirmer que la quantité de cellules dans le réacteur, ainsi que la concentration de cellules, X , par unité de réacteur, suit la relation suivante :

$$\dot{X}(t) = \mu(t) X(t) \quad (2.1.14)$$

2.1.4.2 Lien avec les travaux de Hwa [71]

On voit que l'on obtient une formule proche de celle des travaux de Hwa, présentés au chapitre précédent. Pour rappel de la relation (1.2.1), le taux spécifique de croissance est égal pour eux à :

$$\mu(t) = k(\phi_R(t) - \phi_{R,0}(t))$$

où ϕ_R est la proportion massique de ribosomes dans le cytosol, $\phi_{R,0}$ la proportion massique de ribosomes inactifs dans le cytosol, k est un coefficient proportionnel à l'efficacité de traduction des ribosomes.

Par ailleurs, en combinant l'équation (2.1.13) avec la relation (2.1.11), on peut écrire :

$$\mu(t) = \frac{1}{D_0} \sum_{P,B,R,T} \nu_{E_i}(t) \leq \frac{1}{D_0} k_R(t) E_R(t)$$

d'où,

$$\mu(t) \leq k_R(t) \underbrace{\frac{E_R(t)}{D_0}}_{\propto \phi_R(t)}$$

Ce qui est proche de la formule de Hwa donnée en (1.2.1) ($\mu = k_R(\phi_R - K_0)$). On peut voir que l'on néglige dans notre modèle la part des ribosomes inactifs par rapport aux travaux de l'équipe de Hwa. Cependant, nous n'imposons pas d'égalité stricte mais voyons cette relation comme une limitation du système, exprimée sous la forme d'une contrainte d'inégalité. C'est une contrainte de fonctionnement qui doit être satisfaite à tout instant, quelles que soient les conditions. Elle provient de considérations physico-biologiques, notamment sur l'efficacité des processus internes. Pour l'équipe de Hwa, l'origine de leur formule provient d'observations et de mesures qui ont montré qu'en croissance exponentielle, le nombre de ribosomes était plus ou moins proportionnel à la valeur du taux spécifique de croissance. L'argument est bien différent.

D'une certaine manière, notre approche automatique nous fait voir le taux spécifique de croissance comme une *commande* du système. C'est une grandeur qui va avoir un impact sur la cellule en tant que système. Dans les travaux du groupe de Hwa, il semblerait que leur approche soit plus physicienne et qu'ils voient le taux spécifique de croissance comme un paramètre explicatif du système qu'ils subissent. Typiquement, ceci les empêchent de prédire le comportement du système si l'on ajoute un processus.

2.1.5 Bilan du modèle

On souhaite écrire le système d'équations de (2.1.2) à (2.1.14) de manière compacte en utilisant le formalisme générique de l'automatique : $\dot{x}(t) = f(x(t), u(t), t)$ où $x(t)$ est le vecteur des variables d'état et $u(t)$ est le vecteur des variables de commande. Ici, l'état correspond aux concentrations et la commande est l'ensemble des flux des processus intracellulaires. Le vecteur des variables de commande sera alors plutôt noté $\nu(t)$.

On pose

$$x(t) = (S(t) \quad E_B(t) \quad E_P(t) \quad E_R(t) \quad E_T(t) \quad G(t) \quad P(t) \quad X(t))^T$$

et

$$\nu(t) = (\nu_T(t) \ \nu_B(t) \ \nu_P(t) \ \nu_{E_B}(t) \ \nu_{E_P}(t) \ \nu_{E_R}(t) \ \nu_{E_T}(t))^T$$

On note $n = \dim(x) = 8$ et $m = \dim(\nu) = 7$. On suppose que les concentrations sont positives ou nulles ainsi que les flux. Ceci signifie notamment que les processus sont irréversibles.

D'après les équations (2.1.2), (2.1.3), (2.1.8), (2.1.10) et (2.1.14) et en substituant le taux spécifique de croissance μ par son expression (2.1.13) en fonction des commandes ν_{E_i} , on peut écrire la dérivée temporelle de x comme suit :

$$\frac{d}{dt} \begin{pmatrix} S \\ E_B \\ E_P \\ E_R \\ E_T \\ G \\ P \\ X \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_T & -1 & -1 & -1 - \frac{S}{D_0} & -1 - \frac{S}{D_0} & -1 - \frac{S}{D_0} & -1 - \frac{S}{D_0} \\ 0 & 0 & 0 & \alpha_{E_B} - \frac{E_B}{D_0} & -\frac{E_B}{D_0} & -\frac{E_B}{D_0} & -\frac{E_B}{D_0} \\ 0 & 0 & 0 & -\frac{E_P}{D_0} & \alpha_{E_P} - \frac{E_P}{D_0} & -\frac{E_P}{D_0} & -\frac{E_P}{D_0} \\ 0 & 0 & 0 & -\frac{E_R}{D_0} & -\frac{E_R}{D_0} & \alpha_{E_R} - \frac{E_R}{D_0} & -\frac{E_R}{D_0} \\ 0 & 0 & 0 & -\frac{E_T}{D_0} & -\frac{E_T}{D_0} & -\frac{E_T}{D_0} & \alpha_{E_T} - \frac{E_T}{D_0} \\ -X & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_P X & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{X}{D_0} & \frac{X}{D_0} & \frac{X}{D_0} & \frac{X}{D_0} \end{pmatrix}}_{\triangleq F(x)} \begin{pmatrix} \nu_T \\ \nu_B \\ \nu_P \\ \nu_{E_B} \\ \nu_{E_P} \\ \nu_{E_R} \\ \nu_{E_T} \end{pmatrix} \quad (2.1.15)$$

où l'on a omis le paramètre temporel t par soucis de lisibilité et où $F(x)$ est la matrice d'évolution.

La dynamique du système s'écrit alors :

$$\dot{x}(t) = F(x(t))\nu(t) \quad (2.1.16)$$

Les contraintes d'efficacité des processus, (2.1.4), (2.1.7), (2.1.9), (2.1.11), s'écrivent matriciellement sous la forme :

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}}_{\triangleq A} \begin{pmatrix} \nu_T \\ \nu_B \\ \nu_P \\ \nu_{E_B} \\ \nu_{E_P} \\ \nu_{E_R} \\ \nu_{E_T} \end{pmatrix} \leq \underbrace{\begin{pmatrix} k_T(G,S)E_T \\ k_B(S)E_B \\ k_P(S)E_P \\ k_R(S)E_R \end{pmatrix}}_{\triangleq b(x)} \quad (2.1.17)$$

et la contrainte (2.1.6) sur l'équilibre de B devient en substituant de nouveau μ par son

expression en fonction des commandes ν_{E_i} :

$$\underbrace{\begin{pmatrix} 0 & \frac{\alpha_B D_0}{B_0} & 0 & -1 & -1 & -1 & -1 \end{pmatrix}}_{\triangleq L} \begin{pmatrix} \nu_T \\ \nu_B \\ \nu_P \\ \nu_{E_B} \\ \nu_{E_P} \\ \nu_{E_R} \\ \nu_{E_T} \end{pmatrix} = 0 \quad (2.1.18)$$

Au final, le problème s'écrit de manière compacte :

$$\begin{cases} \dot{x}(t) = F(x(t))\nu(t) \\ A\nu(t) \leq b(x(t)) \\ L\nu(t) = 0 \\ x(t) \geq 0 \\ \nu(t) \geq 0 \end{cases} \quad (2.1.19)$$

On appellera cet ensemble d'équations et de contraintes la modélisation *dynamical Resource Balance Analysis* (dRBA) de la cellule.

Les paramètres du modèle sont :

- les coefficients de rendement α_i , et les coûts des protéines $\frac{1}{\alpha_{E_i}}$;
- la concentration cible en macrocomposants, B_0 ;
- la densité intracellulaire, D_0 ;
- les paramètres des efficacités des processus, K_i et $v_{m,i}$.

2.1.6 Paramètres du modèle

Les paramètres du modèle sont donnés ci-dessous à titre d'exemple pour comprendre la philosophie de l'agrégation. On cherche à avoir des valeurs qui ont un sens pour une cellule la plus proche possible de *B. subtilis*. Si besoin, c'est-à-dire si les données ne sont pas disponibles pour *B. subtilis*, on prendra les valeurs pour *E. coli* à défaut de mieux. On utilisera pour cela des documents complémentaires fournis en annexe de l'article [65] qui donnent une bonne partie de la composition intracellulaire pour *B. subtilis* à un taux spécifique de croissance de $0,1 \text{ h}^{-1}$.

2.1.6.1 Coefficients de rendement et coûts des processus

Les coefficients α_ξ sont les coefficients de rendement des réactions $1 \cdot \mathcal{G} \longrightarrow \alpha_T \cdot \mathcal{S}$ et $1 \cdot \mathcal{S} \longrightarrow \alpha_\xi \cdot \xi$, $\xi \in \{\mathcal{P}, \mathcal{B}, \mathcal{E}_B, \mathcal{E}_R, \mathcal{E}_T, \mathcal{E}_P\}$. Ainsi, par conservation de la masse, la masse d'une mole de $\mathcal{G}$ vaut exactement la masse de α_T moles de $\mathcal{S}$. De même, la masse d'une mole de $\mathcal{S}$ vaut la masse de α_ξ moles de ξ , pour $\xi \in \{\mathcal{P}, \mathcal{B}, \mathcal{E}_B, \mathcal{E}_R, \mathcal{E}_T, \mathcal{E}_P\}$

On a alors,

$$\left\{ \begin{array}{l} \alpha_T = \frac{M_G}{M_S} \\ \alpha_P = \frac{M_S}{M_P} \\ \alpha_B = \frac{M_S}{M_B} \\ \alpha_{E_i} = \frac{M_S}{M_{E_i}} \quad i \in \{P, B, R, T\} \end{array} \right. \quad (2.1.20)$$

avec M_ξ la masse molaire de l'espèce ξ .

Concernant le contenu de $\mathcal{S}$, on suppose que l'entité $\mathcal{S}$ agrège les éléments nécessaires à la construction des macrocomposants et des protéines (les métabolites). Plus spécifiquement, on suppose que les acides aminés libres, les nucléotides et les différentes formes de dinucléotides (NAD, NADH, NADP, NADPH) font partie de cette entité.

En supposant que l'on a accès aux concentrations nominales des composants de $\mathcal{S}$, marquées d'un 0 en indice, et à leur masse molaire, la concentration nominale de $\mathcal{S}$, S_0 , est la somme des concentrations de ses constituants, et sa masse molaire est la moyenne des masses molaires de ses constituants pondérée par la concentration relative de ce constituant :

$$\left\{ \begin{array}{l} S_0 = \sum_{\xi \in \mathcal{S}} \xi_0 \\ M_S = \sum_{\xi \in \mathcal{S}} \frac{\xi_0}{S_0} M_\xi \end{array} \right. \quad (2.1.21)$$

Concernant les acides aminés, [65] ne distingue pas la composition des protéines de la composition des acides aminés libres. Cependant, ce qui nous intéresse pour caractériser $\mathcal{S}$ sont uniquement les acides aminés libres. On se reporte alors au modèle de Marr [57] qui prédit la composition de *E. coli* à différents taux spécifique de croissance et trouve une concentration en acides aminés libres de 1 à $57 \mu\text{mol}.g_{CDW}^{-1}$ en fonction des conditions de croissance. On prendra comme valeur nominale, la valeur moyenne :

$$AA_0^{libre} \approx 25 \mu\text{mol}.g_{CDW}^{-1} \quad (2.1.22)$$

La masse molaire d'un acide aminé moyen est donnée dans [65] et est de l'ordre de :

$$M_{AA} \approx 129,1 \text{ g.mol}^{-1} \quad (2.1.23)$$

Les masses molaires et concentrations mesurées des nucléotides et dinucléotides sont directement données par [65] et sont rappelées Table 2.1.

En appliquant la relation (2.1.21), pour S , on trouve :

$$\left\{ \begin{array}{l} S_0 = 57 \mu\text{mol}.g_{CDW}^{-1} \\ M_S = 358 \text{ g.mol}^{-1} \end{array} \right. \quad (2.1.24)$$

D'autre part, dans la littérature, on trouve $M_{glucose} = M_G = 180 \text{ g.mol}^{-1}$. Par (2.1.20), on en déduit :

$$\alpha_T = 0,50 \quad (2.1.25)$$

TABLE 2.1 – Masses molaires et concentrations mesurées [65] des constituants de $\mathcal{S}$.

Composés	Masses molaires (g/mol)	Concentrations relevées dans [65] ($\mu\text{mol.g}_{CDW}^{-1}$)
NAD	662,4	16,2
AMP	345,2	4,7
ATP	503,2	3
ADP	424,2	2,6
CMP	321,2	1
NADP	740,4	0,9
CTP	479,1	0,5
GMP	361,2	0,5
GTP	519,1	0,4
CDP	400,2	0,3
NADPH	741,4	0,2
GDP	440,2	0,2
Acides aminés libres	(équ. (2.1.23)) 129,1	25 (équ. (2.1.22))

Pour $\mathcal{P}$, il s'agit d'une molécule proche d'un métabolite, on peut supposer qu'il a la même masse molaire que $\mathcal{S}$:

$$\alpha_P = \frac{M_{\mathcal{S}}}{M_{\mathcal{P}}} = 1 \quad (2.1.26)$$

En ce qui concerne $\mathcal{B}$, on peut regrouper sous cette entité ce qui constitue la membrane, la paroi, les ARN et l'ADN de la cellule.

Une relation similaire à (2.1.21) s'applique pour la concentration et la masse molaire équivalente de $\mathcal{B}$.

Les constituants de $\mathcal{B}$ sont détaillés dans la Table 2.2.

Ainsi,

$$\begin{cases} B_0 &= 387 \mu\text{mol.g}_{CDW}^{-1} \\ M_B &= 1013 \text{ g.mol}^{-1} \end{cases} \quad (2.1.27)$$

et finalement,

$$\alpha_B = \frac{M_{\mathcal{S}}}{M_B} = \frac{358}{1013} = 0,35 \quad (2.1.28)$$

En ce qui concerne les protéines, α_{E_i} peut s'écrire en première approximation sous la forme :

$$\alpha_{E_i} = \frac{M_{\mathcal{S}}}{M_{\mathcal{E}_i}} = \frac{M_{\mathcal{S}}}{M_{AA}\ell_{\mathcal{E}_i}} \quad (2.1.29)$$

avec M_{AA} la masse molaire d'un acide aminé moyen, $\ell_{\mathcal{E}_i}$ la longueur de la protéine $\mathcal{E}_i$. On reconnaît ici le fait que $\frac{1}{\alpha_{E_i}}$ est proportionnel à la longueur de la protéine, qui est une manière de quantifier le coût en ressources d'une protéine.

On prend pour les longueurs des protéines $\mathcal{E}_B$, $\mathcal{E}_T$ et $\mathcal{E}_P$, la longueur moyenne des protéines de *Bacillus subtilis*. Pour $\mathcal{E}_R$, on prend le nombre total d'acides aminés dans

TABLE 2.2 – Masses molaires et concentrations mesurées [65] des constituants de $\mathcal{B}$.

Composés	Masses molaires ($g.mol^{-1}$)	Concentrations relevées dans [65] ($\mu mol.g_{CDW}^{-1}$)
<i>ARN :</i>		
AMP	345,2	50,5
GMP	361,2	62,2
CMP	321,2	38,4
UMP	322,2	41,5
<i>ADN :</i>		
dAMP	329,2	23
dGMP	345,2	17,4
dCMP	305,2	17,4
dTMP	320,2	22,9
<i>Lipides (membrane) :</i>		
Monoglucosyldiacylglycerol	72 208,8	0,08575
Diglucosyldiacylglycerol	88 422,9	0,110292
Triglucosyldiacylglycerol	104 637	0,065833
Cardiolipin	133 389,9	0,004642
Phosphatidylglycerol	71 299,7	0,175859
Lysylphosphatidylglycerol	84 218	0,022057
Phosphatidylethanolmaine	68 198,7	0,559509
<i>Composants de la paroi :</i>		
Peptidoglycane	991	101,82
Acide teichoïque	11 313	10,9

un ribosome de *B. subtilis* :

$$\begin{cases} \ell_{\mathcal{E}_P} = 360 \\ \ell_{\mathcal{E}_B} = 360 \\ \ell_{\mathcal{E}_R} = 10100 \\ \ell_{\mathcal{E}_T} = 360 \end{cases} \quad (2.1.30)$$

ce qui conduit à

$$\begin{cases} \alpha_{E_P} = \alpha_{E_B} = \alpha_{E_T} = 7,7.10^{-3} \\ \alpha_{E_R} = 2,7.10^{-4} \end{cases} \quad (2.1.31)$$

2.1.6.2 Efficacité des processus

Pour K_T , $v_{m,T}$, $v_{m,P}$, $v_{m,B}$ on choisit des valeurs cohérentes avec celles considérées dans [35] (justifié plus loin).

$$\begin{cases} K_T = 5 \mu mol.g_{CDW}^{-1} \\ v_{m,R} = 20\,000 \, h^{-1} \\ v_{m,T} = 420 \, h^{-1} \\ v_{m,B} = 750 \, h^{-1} \\ v_{m,P} = 750 \, h^{-1} \end{cases} \quad (2.1.32)$$

K_S , K_P , K_B rendent compte de l'impact qu'ont les constituants de $\mathcal{S}$ sur les efficacités métaboliques. Ceci n'est pas pris en compte par les RBA, qui ne décrivent pas les

concentrations en métabolites. On prend des valeurs en accord avec l'ordre de grandeur de la valeur nominale S_0 de la concentration en $\mathcal{S}$ calculée en (2.1.24). On prendra

$$\begin{cases} K_S &= 0,5 \text{ mmol.g}_{CDW}^{-1} \\ K_P &= 3 \text{ } \mu\text{mol.g}_{CDW}^{-1} \\ K_B &= 3 \text{ } \mu\text{mol.g}_{CDW}^{-1} \\ K_R &= 10 \text{ } \mu\text{mol.g}_{CDW}^{-1} \end{cases} \quad (2.1.33)$$

Les efficacités maximales $v_{m,B}$ et $v_{m,T}$ des protéines qui catalysent les processus Σ_B et Σ_T sont plutôt dans la tranche basse des valeurs d'efficacité que l'on retrouve chez *Bacillus subtilis* (ici, de l'ordre de $0,2\text{-}0,5 \text{ s}^{-1}$, pour une efficacité médiane mesurée chez *B. subtilis* de l'ordre de 7 s^{-1}). On explique ce choix par le fait que l'on ait agrégé différents processus enzymatiques et le fait que ce soit l'efficacité la plus faible qui détermine l'efficacité d'une succession de réactions.

Par ailleurs, l'efficacité maximale de Σ_R , $v_{m,R}$, peut être assimilée à l'efficacité maximale des ribosomes. Dans les RBA, cette efficacité est modulée par le taux spécifique de croissance et prend des valeurs autour de 15 acides aminés par secondes, en accord avec les observations faites dans [14]. En réalité, c'est la même valeur qui est retenue ici à un facteur de conversion près entre les acides aminés et les éléments qui constituent $\mathcal{S}$. En effet, une équation fondamentale des RBA (énoncée au chapitre précédent, voir section 1.2.1.3) est la limitation du taux spécifique de croissance par la capacité des ribosomes qui s'écrit :

$$\mu(t) \sum \ell_{\mathcal{E}_i} E_i(t) \leq k_R^{RBA}(t) E_R(t) \quad (2.1.34)$$

avec les mêmes notations que celles introduites jusque là, et k_R^{RBA} l'efficacité des ribosomes, valant 15 AA/s (acides aminés par seconde). $\ell_{\mathcal{E}_i} = \frac{M_S}{M_{AA}} \frac{1}{\alpha_{E_i}}$ est la longueur en acides aminés de la protéine $\mathcal{E}_i$, proportionnelle au coût de $\mathcal{E}_i$ en $\mathcal{S}$.

De plus, pour les RBA, le régime quasi statique est atteint, ce qui signifie que $\dot{E}_i(t) = 0$. D'après la relation (2.1.10), $\nu_{E_i}(t) = \frac{1}{\alpha_{E_i}} \mu(t) E_i(t)$. Ainsi, la contrainte des RBA (2.1.34) s'écrit avec notre formalisme :

$$\frac{M_S}{M_{AA}} \sum \frac{1}{\alpha_{E_i}} \mu(t) E_i(t) \leq k_R^{RBA}(t) E_R(t)$$

ou encore,

$$\sum \nu_{E_i}(t) \leq \frac{M_{AA}}{M_S} k_R^{RBA}(t) E_R(t)$$

On reconnaît là la contrainte d'efficacité du processus Σ_R , (2.1.11), et on identifie :

$$\frac{M_{AA}}{M_S} k_R^{RBA}(t) = k_R(t) = \frac{v_{m,R} S(t)}{S(t) + K_R}$$

Numériquement, $\frac{M_{AA}}{M_S} k_R^{RBA} = \frac{129,1}{358} 15 = 5,4 \text{ s}^{-1}$ et ramené en s^{-1} , l'efficacité maximale choisie pour le processus Σ_R vaut $v_{m,R} = 20000/3600 = 5,5 \text{ s}^{-1}$. La valeur choisie est bien en accord.

2.1.6.3 Densité intracellulaire

Enfin, pour la densité, on se réfère à la méthode RBA et aux mesures expérimentales qui ont été faites sur *B. subtilis*. La densité intracellulaire en acides aminés par gramme de poids sec est de l'ordre de $D_{RBA} = 6,67 \text{ mmolAA}/g_{CDW}$, voir [35].

Il apparait cependant que la méthode RBA considère en plus de ce qui a été détaillé précédemment, un pool de protéines, $\mathcal{P}_G$, qui s'expriment mais sans contribuer à la croissance de l'organisme. Ces protéines sont manifestement des mécanismes dont sont pourvues les cellules qui ont été sélectionnées au cours des millions d'années d'évolution et de sélection naturelle, bien qu'ils ne lui servent pas à court terme pour croître. Il s'agit notamment des mécanismes relatifs à la prévention des divers stress. Ces protéines représentent à peu près la moitié des protéines et nous ne les avons pas considérées dans notre modèle simplifié.

Il nous faut alors retrancher de la densité D_{RBA} la part des protéines qui ne contribuent pas à la croissance, ainsi que la part des ribosomes qui participent à la production de ces dernières.

Finalement, d'après les résultats de [35], ceci représente en moyenne 45% de la part totale des protéines qu'il faut retirer, et on prend ainsi une densité pour l'ensemble des protéines privé du sous-ensemble $\mathcal{P}_G$:

$$D_{RBA}^{\setminus \mathcal{P}_G} = 3,68 \text{ mmolAA}/g_{CDW}$$

Encore une fois, il faut convertir cette densité en revenant aux formules :

$$\begin{cases} D_{RBA}^{\setminus \mathcal{P}_G} = \sum \ell_{\mathcal{E}_i} E_i \\ D_0 = \sum \frac{1}{\alpha_{E_i}} E_i = \frac{M_{AA}}{M_S} \sum \ell_{\mathcal{E}_i} E_i \end{cases}$$

Ainsi,

$$D_0 = 1,33 \text{ mmol} \cdot g_{CDW}^{-1} \quad (2.1.35)$$

2.1.6.4 Synthèse des paramètres du modèle

La table 2.3 synthétise les paramètres du modèle agrégé, c'est-à-dire les valeurs données équations (2.1.25), (2.1.26), (2.1.27), (2.1.28), (2.1.31), (2.1.32), (2.1.33), (2.1.35).

2.2 Comportement nominal du modèle

Le but de cette section est de caractériser le comportement nominal du modèle agrégé. Comme pour les RBA et les FBA, on caractérise le comportement du modèle agrégé en supposant qu'il cherche à maximiser la croissance instantanée des micro-organismes, étant données des conditions extérieures invariantes.

Ce critère consistant à maximiser la croissance instantanée est examiné dans [9]. Le but est simplement de proposer un critère raisonnable du point de vue évolutif pour rendre compte du fait que les organismes qui ont subsisté jusqu'à nos jours sont ceux qui ont le mieux survécu à l'évolution. Le "vrai" critère serait sans doute de maximiser la "viabilité" sachant que les conditions extérieures peuvent varier d'une manière proche de

TABLE 2.3 – Valeurs des paramètres pour le modèle agrégé

Paramètre	Valeur
α_T	0,50
α_B	0,35
α_P	1
α_{E_T}	$7,7.10^{-3}$
α_{E_B}	$7,7.10^{-3}$
α_{E_P}	$7,7.10^{-3}$
α_{E_R}	$2,7.10^{-4}$
B_0	$387 \mu mol.L^{-1}$
D_0	$1,33 mmol.L^{-1}$
$v_{m,R}$	$20\,000 h^{-1}$
$v_{m,T}$	$420 h^{-1}$
$v_{m,B}$	$750 h^{-1}$
$v_{m,P}$	$750 h^{-1}$
K_S	$0,5 mmol.g_{CDW}^{-1}$
K_P	$3 \mu mol.g_{CDW}^{-1}$
K_B	$3 \mu mol.g_{CDW}^{-1}$
K_R	$10 \mu mol.g_{CDW}^{-1}$
K_T	$5 \mu mol.g_{CDW}^{-1}$

la façon dont elles ont pu varier dans la niche écologique des organismes que l'on considère. Ceci permettrait notamment de prédire l'ensemble des protéines qui s'expriment sans servir à la croissance (voir le pool $\mathcal{P}_G$ présenté ci-dessus). Cet aspect n'est pas développé dans ce document.

Dans notre cas, le comportement nominal du système est défini comme l'état stationnaire qui est atteint par une population de micro-organismes qui maximise sa croissance dans un milieu de culture invariant, typiquement, dans lequel la concentration en $\mathcal{G}$ est constante égale à G_0 . En reprenant les notations compactes introduites dans (2.1.19), on cherche donc à résoudre le problème d'optimisation (Π_{G_0}) suivant :

$$\begin{aligned}
 \max_{(\bar{x}, \bar{\nu}) \in \mathbb{R}^{n \times m}} \mu &= \frac{1}{D_0} (0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1) \bar{\nu} \\
 (\Pi_{G_0}) : \quad &\text{sous contraintes} \begin{cases} F(\bar{x})\bar{\nu} = 0 \\ A\bar{\nu} \leq b(\bar{x}) \\ L\bar{\nu} = 0 \\ \bar{x} \geq 0 \\ \bar{\nu} \geq 0 \\ \bar{G} = G_0 \end{cases} \quad (2.2.1)
 \end{aligned}$$

où l'état stationnaire $\bar{x}$ est garanti par le fait que $\dot{\bar{x}} = F(\bar{x})\bar{\nu} = 0$, et $\bar{\nu}$ est la commande à l'équilibre.

Par la suite, pour alléger les notations, on notera l'équilibre simplement x et ν au lieu de $\bar{x}$ et $\bar{\nu}$, respectivement.

2.2.0.1 Considération générale sur l'optimum

Il est clair qu'à l'optimum, les contraintes d'inégalités entre les concentrations de protéines et les flux doivent être saturées. En effet, en raisonnant par l'absurde, si dans la configuration qui maximise la croissance, il y a un écart entre un flux et la capacité maximale du procédé correspondant, cela signifie qu'il y a un excès de protéines pour ce procédé. On peut alors diminuer la quantité de ressource allouée à ces protéines tout en respectant la contrainte. On répartit alors entre tous les procédés participant à la croissance la quantité de protéine qui a été retirée : on augmente ainsi la croissance, ce qui contredit l'hypothèse selon laquelle la configuration initiale maximise la croissance.

La contrainte $A\nu \leq b(x)$ devient :

$$A\nu = b(x)$$

Nous allons à présent caractériser mathématiquement l'optimum et montrer que l'on retrouve alors des allures de modèles macroscopiques standards ainsi que ce qui peut être obtenu par la méthode RBA.

2.2.1 Caractérisation à l'équilibre : lien avec un modèle macroscopique

Réolvons ici une série de problèmes (Π_{G_0}) donnés par (2.2.1) pour différentes valeurs de G_0 . Cette démarche est proche de l'élaboration des modèles macroscopiques, où des cellules sont placées en réacteurs continus pour lesquels on fait varier le débit d'alimentation normalisé par rapport au volume du réacteur. A l'équilibre, on note les diverses concentrations dans le réacteur et on établit ainsi un modèle entre les grandeurs macroscopiques.

Nous allons manipuler les contraintes du problème (Π_{G_0}) pour trouver des relations simples entre les grandeurs. En particulier, on cherche à paramétrer le taux de croissance en fonction des autres grandeurs lorsque les contraintes du problème (Π_{G_0}) sont satisfaites.

2.2.1.1 Caractérisation analytique

Soit G_0 la concentration de $\mathcal{G}$ dans le réacteur. On résout analytiquement le problème (Π_{G_0}) donné par (2.2.1) avec les contraintes d'inégalités saturées $A\nu = b(x)$.

La contrainte (2.1.11) et la relation (2.1.13) donnent :

$$\sum_{P,B,R,T} \nu_{E_i} = D_0\mu = k_R(S)E_R \quad (2.2.2)$$

De même, (2.1.7) et (2.1.6) donnent :

$$\nu_B = \frac{B_0}{\alpha_B}\mu = k_B(S)E_B \quad (2.2.3)$$

Puis, l'équilibre en $\mathcal{S}$ s'écrit grâce à l'équation (2.1.2) :

$$\alpha_T \nu_T = \nu_B + \nu_P + \sum_{P,B,R,T} \nu_{E_i} + \mu S$$

i.e. par les équations (2.2.2) et (2.2.3)

$$\alpha_T \nu_T = \frac{B_0}{\alpha_B} \mu + \nu_P + D_0 \mu + \mu S$$

ou encore,

$$\alpha_T \nu_T = \left(\frac{B_0}{\alpha_B} + D_0 + S \right) \mu + \nu_P$$

et avec (2.1.4),

$$\nu_T = \left(\frac{B_0}{\alpha_B} + D_0 + S \right) \mu / \alpha_T + \nu_P / \alpha_T = k_T(G_0, S) E_T \quad (2.2.4)$$

Par ailleurs, la densité, $\sum_i E_i / \alpha_{E_i}$, étant constante, c.f. (2.1.12), on peut écrire :

$$\underbrace{\frac{D_0 \mu}{\alpha_{E_R} k_R(S)}}_{=E_R/\alpha_{E_R}, \text{ eq. (2.2.2)}} + \underbrace{\frac{\frac{B_0}{\alpha_B} \mu}{\alpha_{E_B} k_B(S)}}_{=E_B/\alpha_{E_B}, \text{ eq. (2.2.3)}} + \underbrace{\frac{(\frac{B_0}{\alpha_B} + D_0 + S) \mu / \alpha_T + \nu_P / \alpha_T}{\alpha_{E_T} k_T(G_0, S)}}_{=E_T/\alpha_{E_T}, \text{ eq. (2.2.4)}} + E_P / \alpha_{E_P} = D_0$$

En isolant les termes en μ , il vient :

$$\mu = \frac{D_0 - E_P / \alpha_{E_P} - \frac{\nu_P}{\alpha_T \alpha_{E_T} k_T(G_0, S)}}{\frac{D_0}{\alpha_{E_R} k_R(S)} + \frac{\frac{B_0}{\alpha_B}}{\alpha_{E_B} k_B(S)} + \frac{(\frac{B_0}{\alpha_B} + D_0 + S) / \alpha_T}{\alpha_{E_T} k_T(G_0, S)}}$$

On note que E_P et ν_P ont une influence négative sur la croissance. Ceci est cohérent avec l'idée de compétition entre croissance et production.

Ainsi, pour maximiser la croissance, il est nécessaire d'avoir E_P et $\nu_P = 0$, quelle que soit la valeur G_0 :

$$\begin{cases} E_P(G_0) = 0, \forall G_0 \\ \nu_P(G_0) = 0, \forall G_0 \end{cases} \quad (2.2.5)$$

μ est alors donné comme une fonction de G_0 et S uniquement :

$$\mu = \frac{D_0}{\frac{D_0}{\alpha_{E_R} k_R(S)} + \frac{\frac{B_0}{\alpha_B}}{\alpha_{E_B} k_B(S)} + \frac{(\frac{B_0}{\alpha_B} + D_0 + S) / \alpha_T}{\alpha_{E_T} k_T(G_0, S)}} \quad (2.2.6)$$

Concentration en $\mathcal{S}$ qui permet de maximiser le taux spécifique de croissance.

A G_0 fixée, on cherche la concentration en $\mathcal{S}$ qui maximise μ sous contraintes de fonctionnement possible des cellules.

D'après les relations ci-dessus, si $S \geq 0$ et $\mu \geq 0$, alors les contraintes $x \geq 0$ et $\nu \geq 0$ du problème (2.2.1) sont satisfaites. Le domaine de définition $(x, \nu) \in \mathbb{R}_+^{n+m}$ se limite donc à $(\mu, S) \in \mathbb{R}_+^2$.

Par ailleurs, d'après la relation (2.2.6), pour $S = 0$, on a $\mu = 0$.

Montrons que pour $S > 0$ (i.e. à l'intérieur du domaine de définition), on peut trouver un taux de croissance $\mu > 0$ (i.e. meilleur que le la frontière du domaine de définition).

On cherche donc une concentration $S > 0$ telle que la dérivée de μ par rapport à S s'annule :

$$\frac{\partial \mu}{\partial S}|_{G_0, \text{fixée}}(S) = 0$$

Il en est de même pour la dérivée de l'inverse du taux spécifique de croissance, $\frac{\partial(1/\mu)}{\partial S} = \frac{\partial \mu}{\partial S} \times \frac{-1}{\mu^2} = 0$ et on peut donc chercher S qui annule la dérivée de l'inverse de μ , pour un taux spécifique de croissance non nul :

$$\mu^{-1} = \frac{\frac{D_0}{\alpha_{E_R} k_R(S)} + \frac{\frac{B_0}{\alpha_B}}{\alpha_{E_B} k_B(S)} + \frac{(\frac{B_0}{\alpha_B} + D_0 + S)/\alpha_T}{\alpha_{E_T} k_T(G_0, S)}}{D_0}$$

Par substitution des k_i par leurs expressions, on en déduit :

$$\mu^{-1} = \frac{1}{\alpha_{E_R} v_{m,R}} (1 + K_R/S) + \frac{B_0}{\alpha_B \alpha_{E_B} D_0 v_{m,B}} (1 + K_B/S) + \frac{(\frac{B_0}{\alpha_B} + D_0 + S)}{\alpha_T \alpha_{E_T} D_0} \frac{(G_0 + K_T + K_S S)}{v_{m,T} G_0}$$

qui peut s'écrire sous la forme générale :

$$\mu^{-1} = \alpha + \frac{\beta}{S} + \gamma(G_0) S^2 + \delta(G_0) S + \eta(G_0)$$

où α et β sont des constantes, γ, δ, η des fonctions de G_0 . donc nécessairement,

$$\frac{\partial \mu^{-1}}{\partial S} = 0 = \frac{-\beta}{S^2} + 2\gamma(G_0) S + \delta(G_0)$$

i.e.

$$0 = 2\gamma(G_0) S^3 + \delta(G_0) S^2 - \beta$$

avec

$$\begin{cases} \gamma(G_0) = \frac{1}{G_0} \frac{K_S}{\alpha_T \alpha_{E_T} D_0 v_{m,T}} \\ \delta(G_0) = \frac{1}{G_0} \frac{(B_0/\alpha_B + D_0) K_S + K_T}{\alpha_T \alpha_{E_T} D_0 v_{m,T}} + \frac{G_0}{\alpha_T \alpha_{E_T} D_0 v_{m,T}} \\ \beta = \frac{K_R}{\alpha_{E_R} v_{m,R}} + \frac{K_B B_0}{\alpha_B \alpha_{E_B} D_0 v_{m,B}} \end{cases}$$

Après substitution et en rassemblant les termes en G_0 , il vient :

$$G_0 = \frac{2K_S S^3 + ((B_0/\alpha_B + D_0) K_S + K_T) S^2}{\alpha_T \alpha_{E_T} D_0 v_{m,T} \left(\frac{K_R}{\alpha_{E_R} v_{m,R}} + \frac{K_B B_0}{\alpha_B \alpha_{E_B} D_0 v_{m,B}} \right) - S^2} \quad (2.2.7)$$

Sous réserve d'existence, $S(G_0)$ est donnée par la réciproque de cette fonction. On va montrer le caractère bijectif de cette application pour en déduire que la réciproque existe bien.

La fonction $G_0(S)$ définie en (2.2.7) est le quotient d'une fonction de S croissante par une fonction de S décroissante. $G_0(S)$ est donc croissante sur $\mathbb{R}_+$.

Dès lors que G_0 et S sont des concentrations positives, le domaine de validité de cette relation est restreint à $S \in [0, S_{max}[$ avec S_{max} la racine du dénominateur :

$$S_{max} = \sqrt{\alpha_T \alpha_{E_T} D_0 v_{m,T} \left(\frac{K_R}{\alpha_{E_R} v_{m,R}} + \frac{K_B B_0}{\alpha_B \alpha_{E_B} D_0 v_{m,B}} \right)} = 0,0701 \text{ mmol.g}_{CDW}^{-1} \quad (2.2.8)$$

L'application $\begin{pmatrix} [0, S_{max}[& \rightarrow & \mathbb{R}_+ \\ S & \mapsto & G_0(S) \end{pmatrix}$ est bijective. La fonction inverse $G_0 \mapsto S(G_0)$ définie sur $\mathbb{R}_+$ a un sens.

La relation (2.2.8) permet une étude préliminaire de l'impact des différents paramètres sur la valeur de S_{max} . Ceci permet de s'assurer que le fonctionnement est cohérent en première approximation. Si l'efficacité $v_{m,T}$ augmente, S_{max} augmente. A l'inverse, si les efficacités $v_{m,R}$ ou $v_{m,B}$ augmentent, S_{max} diminue. Si D_0 ou B_0 augmentent, *i.e.* si la demande en $\mathcal{S}$ est plus importante, alors S_{max} augmente. Ceci semble cohérent.

Biologiquement, étudier l'influence des paramètres sur S_{max} revient à se placer à une concentration G_0 qui tend vers l'infini, *i.e.* dans un milieu où la ressource n'est pas limitée. Quand on cherche à maximiser la croissance des micro-organismes dans ce milieu, on se confronte au fait qu'une même ressource est partagée par plusieurs processus. Il faut alors trouver l'équilibre dans l'allocation de ressource qui est faite à ces processus. En faisant varier les paramètres, on observe l'influence des processus sur cet équilibre.

Ainsi, modifier les efficacités $v_{m,i}$ revient à modifier les processus Σ_i . On voit alors que le composé $\mathcal{S}$ est un tampon entre le substrat $\mathcal{G}$ et les éléments à produire pour assurer la croissance (les macrocomposants $\mathcal{B}$ et les protéines $\mathcal{E}$). Si l'apport en substrat diminue, ou si la demande en éléments à produire augmente, le résultat est le même : les réserves de $\mathcal{S}$ diminuent. Ce comportement est bien en accord avec le fait que $\mathcal{S}$ soit une sorte de métabolite qui a un rôle tampon entre le substrat $\mathcal{G}$ et les composés à produire pour assurer la croissance.

D'autre part, pour S tendant vers 0, G est équivalent à un terme en S^2 . Autrement dit, S devient proportionnel à la racine carrée de G pour des faibles valeurs de G .

$$S(G) = \mathcal{O}_{G \rightarrow 0}(\sqrt{G}) \quad (2.2.9)$$

Ceci sera analysé de nouveau plus loin.

Taux spécifique de croissance maximal en condition stationnaire. Le taux spécifique de croissance maximal est ensuite donné en fonction de G en reportant l'expression $S(G)$ (réciproque de la relation (2.2.7)) dans l'équation (2.2.6). Il y a bien sûr

un effet dû à $\mathcal{S}$ mais à S fixée, l'équation (2.2.6) s'écrit en fonction de G comme :

$$\mu = \frac{D_0}{\frac{D_0}{\alpha_{E_R} k_R(S)} + \frac{\frac{B_0}{\alpha_B}}{\alpha_{E_B} k_B(S)} + \frac{(\frac{B_0}{\alpha_B} + D_0 + S)/\alpha_T}{\alpha_{E_T}}} \cdot \frac{G + K_T + K_S S}{v_{m,T} G}$$

ou encore, en multipliant haut et bas par G :

$$\mu = \frac{D_0 G}{\left(\frac{D_0}{\alpha_{E_R} k_R(S)} + \frac{\frac{B_0}{\alpha_B}}{\alpha_{E_B} k_B(S)}\right) G + \frac{(\frac{B_0}{\alpha_B} + D_0 + S)/\alpha_T}{\alpha_{E_T} v_{m,T}} (G + K_T + K_S S)}$$

c'est-à-dire,

$$\mu = \frac{\mu_{max}(S) G}{G + K_\mu(S)} \quad (2.2.10)$$

avec

$$\left\{ \begin{array}{l} \mu_{max}(S) = \frac{D_0}{\frac{D_0}{\alpha_{E_R} k_R(S)} + \frac{\frac{B_0}{\alpha_B}}{\alpha_{E_B} k_B(S)} + \frac{(\frac{B_0}{\alpha_B} + D_0 + S)/\alpha_T}{\alpha_{E_T} v_{m,T}}} \\ K_\mu(S) = \frac{\frac{(\frac{B_0}{\alpha_B} + D_0 + S)/\alpha_T}{\alpha_{E_T} v_{m,T}} (K_T + K_S S)}{\frac{D_0}{\alpha_{E_R} k_R(S)} + \frac{\frac{B_0}{\alpha_B}}{\alpha_{E_B} k_B(S)} + \frac{(\frac{B_0}{\alpha_B} + D_0 + S)/\alpha_T}{\alpha_{E_T} v_{m,T}}} \end{array} \right. \quad (2.2.11)$$

Ces deux paramètres dépendent de S . Cependant, pour le taux spécifique de croissance maximal, en injectant les expressions de $k_B(S)$ on peut trouver une relation du type $\mu_{max}(S) = \frac{\dots S}{\dots + S}$, qui est croissante en S . Comme S prend valeur dans $[0, S_{max}[$ avec S_{max} définie en (2.2.8), ceci donne comme valeur maximale pour le taux spécifique de croissance $\mu_{max}^{max} = \mu_{max}(S_{max})$. Avec les valeurs de nos paramètres, on trouve :

$$\mu_{max}^{max} = 0,67 h^{-1}$$

Cette valeur est asymptotiquement atteinte pour des concentrations de $\mathcal{G}$ qui tendent vers l'infini. En effet, quand G tend vers l'infini, S tend vers S_{max} , alors $K_\mu(S)$ tend vers $K_\mu(S_{max})$ et, par la relation (2.2.10), on trouve que le taux spécifique de croissance est équivalent à $\mu_{max}(S)$ qui tend également vers $\mu_{max}(S_{max})$.

En revenant à la relation (2.2.6), et en considérant la concentration S comme un paramètre, on peut étudier l'influence des différents paramètres et donc des différents processus sur le taux spécifique de croissance.

Typiquement, si l'efficacité k_i d'un processus Σ_i augmente (que ce soit par l'intermédiaire d'une concentration S ou d'un paramètre $v_{m,i}$ ou K_i), le taux spécifique de croissance général, μ , augmente. Si la demande en macrocomposants B_0 augmente, le taux spécifique de croissance diminue. En revanche, si la densité D_0 augmente, le taux spécifique de croissance augmente. Cette observation tient au fait que si la densité est

plus importante, la cellule dispose de plus de protéines pour accomplir ses fonctions, y compris la croissance.

A concentration S fixée, on obtient une relation de Monod entre la croissance maximale et la concentration extracellulaire en substrat $\mathcal{G}$ (relation (2.2.10)). Par ailleurs, on note que les paramètres μ_{max} et K_μ donnés en (2.2.11) prennent en compte la totalité des paramètres introduits dans la modélisation. On a ainsi introduit les ingrédients essentiels pour retrouver une formulation classique de taux spécifique de croissance en ne faisant appel qu'à des considérations physico-biologiques et des paramètres qui ont un sens biologique. A titre d'exemple, les coefficients (2.2.11) ne sont pas des paramètres arbitraires mais ils ont bien une origine biologique.

2.2.1.2 Résolution numérique

Dans le but de valider l'étude analytique que nous venons de proposer, on résout numériquement un ensemble de problèmes (Π_{G_0}) donnés par l'équation (2.2.1) pour des valeurs discrètes de G_0 , de manière complètement indépendante de l'étude analytique.

Le problème (2.2.1) est codé en Matlab et résolu à l'aide d'un solveur classique, comme `fmincon`. On cherche alors un vecteur d'état et de commande (dimension $n + m = 15$) qui satisfait un ensemble de contraintes non linéaires d'égalités et d'inégalités.

Il n'est pas utile de chercher la valeur de la concentration extracellulaire en $\mathcal{P}$, car cette concentration n'influence pas le comportement des micro-organismes. On peut indistinctement se placer à la valeur $P = 0 \text{ mmol.L}^{-1}$ ou $P = 300\,000 \text{ mmol.L}^{-1}$. On retire donc P des variables d'optimisation, diminuant ainsi la dimension de l'espace de recherche. Il en va de même pour la valeur de X , que l'on peut fixer arbitrairement à $X = 1 \text{ g}_{CDW}.\text{L}^{-1}$. Concernant la concentration G , il est raisonnable de la considérer davantage comme un paramètre de l'optimisation que comme une variable à optimiser (sachant que l'on impose $G = G_0$). On ne cherche donc plus qu'un vecteur de dimension $(n - 3) + m = 12$.

Tout d'abord, en accord avec l'analyse précédente (équation (2.2.5)) et avec la vision que la production de $\mathcal{P}$ est antagoniste à la croissance, l'optimisation trouve quelle que soit la concentration G_0 :

$$\begin{cases} E_P = 0 \\ \nu_{E_P} = 0 \end{cases} \quad (2.2.12)$$

Ensuite, l'ensemble des concentrations optimales S pour les différents problèmes (Π_{G_0}) est donné en fonction de la valeur G_0 par la figure 2.2. Une comparaison est donnée avec les résultats de l'étude analytique.

On reconnaît là l'allure en racine carrée prédite analytiquement précédemment (2.2.9). On n'atteint pas la borne S_{max} pour les valeurs de G choisies, les valeurs de S restent bien inférieures à cette borne. D'autre part, les valeurs numériques sont du bon ordre de grandeur (dizaine de micro mole par gramme de masse sèche) par rapport aux données

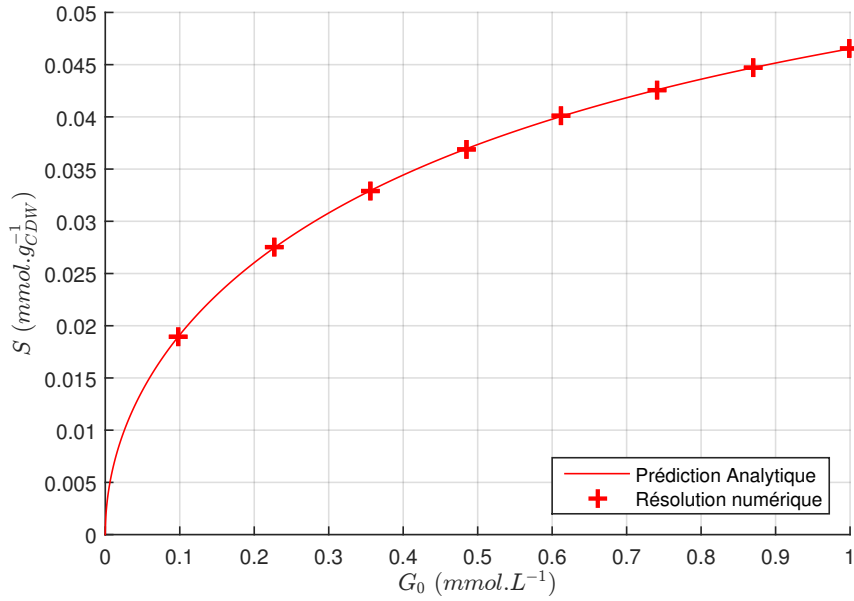


FIGURE 2.2 – Concentration de S en fonction de la concentration de G maximisant le taux spécifique de croissance. Solution numérique du problème (2.2.1) pour différentes valeurs de G_0 .

expérimentales de [65] et aux calculs développés plus haut (2.1.24).

2.2.1.3 Analyse de l'allure $\mu(G_0)$

En ce qui concerne le taux spécifique de croissance, l'allure est donnée figure 2.3.

On retrouve une allure croissante bornée avec une asymptote linéaire à l'origine. C'est bien en accord avec les observations de Monod [60]. D'autre part, l'ordre de grandeur de la concentration de glucose $G_{1/2}$ qui permet d'atteindre la moitié de la valeur maximale du taux spécifique de croissance, $\mu(G_{1/2}) = \frac{\mu(G_0 \rightarrow +\infty)}{2}$, est bien en accord avec les travaux de Monod [60], c'est-à-dire une dizaine de micromoles par litre. En effet, on calcule d'après la courbe présentée figure 2.3 :

$$G_{1/2} = 8,0 \mu\text{mol.L}^{-1} \quad (2.2.13)$$

A noter que le taux spécifique de croissance tend bien asymptotiquement vers la valeur calculée analytiquement $\mu_{max}^{max} = 0,67 \text{ h}^{-1}$, lorsque G_0 tend vers l'infini.

L'allure de la figure 2.3 est rassurante et nous conforte dans l'idée que la modélisation proposée est réaliste. En effet, il n'était en rien garanti d'obtenir une allure croissante, hyperbolique et saturée, pas tout à fait comme une allure de Monod (au sens de la formule de Monod) mais dans l'idée des observations qui ont conduit Monod à établir son modèle, comme expliqué au chapitre précédent [60].

En effet, contrairement à des articles comme [75] où un modèle intracellulaire est couplé à un modèle de bioréacteur par l'intermédiaire du taux spécifique de croissance

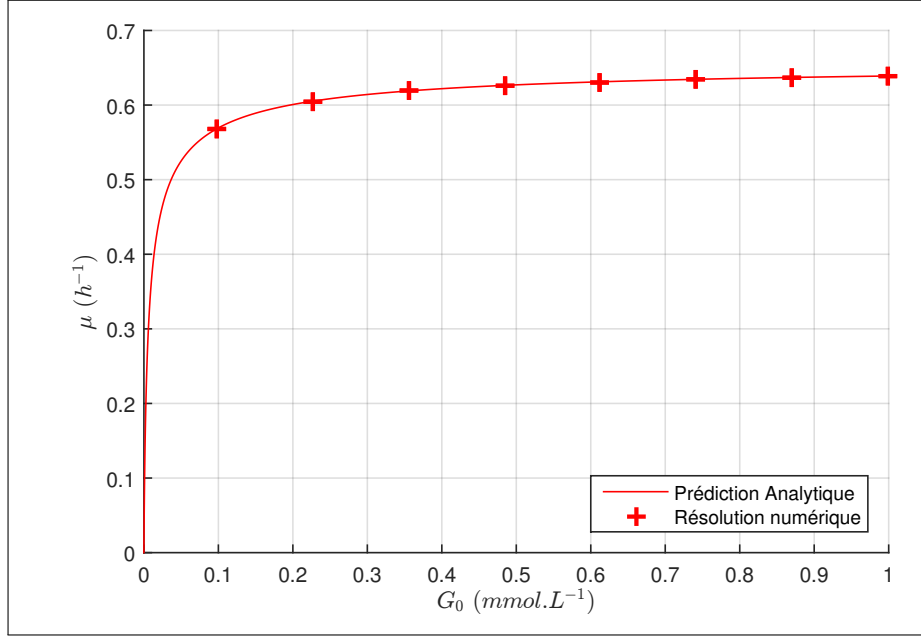


FIGURE 2.3 – taux spécifique de croissance maximal μ en fonction de la concentration extracellulaire en $\mathcal{G}$. Solution numérique du problème (2.2.1) pour différentes valeurs de G_0 .

qui est une fonction de Monod, ici, on a bien retrouvé l'allure d'une cinétique de type Monod en ne considérant que les phénomènes intracellulaires et le couplage n'est fait que par le processus Σ_T . On utilise pleinement le caractère autonome du taux spécifique de croissance, comme expliqué précédemment.

2.2.1.4 Conclusion

Dans cette section, nous avons montré que notre modèle se comportait bien macroscopiquement de la même manière qu'un modèle macroscopique comme on a pu en voir dans le chapitre précédent : le taux spécifique de croissance est bien croissant, asymptotiquement majoré par une valeur maximale.

De plus, les ordres de grandeur des paramètres sont en accord avec ceux des modèles macroscopiques standards donnés dans la littérature.

Ceci nous pousse à avoir confiance en la pertinence des phénomènes qui ont été modélisés à l'échelle intracellulaire. L'objectif de donner un sens aux paramètres de la modélisation macroscopique est atteint.

2.2.2 Caractérisation à l'équilibre : lien avec le modèle RBA

Dans cette section, on cherche à retrouver les profils de concentrations des protéines $\mathcal{E}_i$ en fonction du taux spécifique de croissance. En effet, il est connu que les concentrations des protéines constitutives de la cellule dépendent du taux spécifique de croissance, d'après des travaux comme ceux vus dans le chapitre précédent [44, 29, 45, 12]. On

souhaite vérifier que la manière dont l'agrégation a été faite entre groupe de processus conserve cette caractéristique.

2.2.2.1 Étude qualitative

De l'étude analytique précédente, et des équations (2.2.2), (2.2.3) et (2.2.4), on exprime les concentrations en protéines en fonction du taux spécifique de croissance à l'équilibre (et des concentrations en $\mathcal{G}$ et $\mathcal{S}$) :

$$\begin{cases} E_R = \frac{D_0}{k_R(S)}\mu & (a) \\ E_B = \frac{B_0}{\alpha_B k_B(S)}\mu & (b) \\ E_T = \frac{B_0/\alpha_B + D_0 + S}{\alpha_T k_T(G,S)}\mu & (c) \end{cases} \quad (2.2.14)$$

Cherchons, à G et S fixées, comment évoluent ces expressions. Elles sont toutes linéaires et croissantes en μ . Bien entendu, comme la densité est limitée, il n'y a qu'un point de fonctionnement envisageable : celui pour lequel la somme des termes E_i/α_{E_i} vaut la densité.

Pour la compréhension du modèle, on trace, figure 2.4, en fonction de μ , $E_B/\alpha_{E_B} + E_T/\alpha_{E_T}$ d'une part et $D_0 - E_R/\alpha_{E_R}$ d'autre part, à G et S fixées, *i.e.* pour des efficacités fixées.

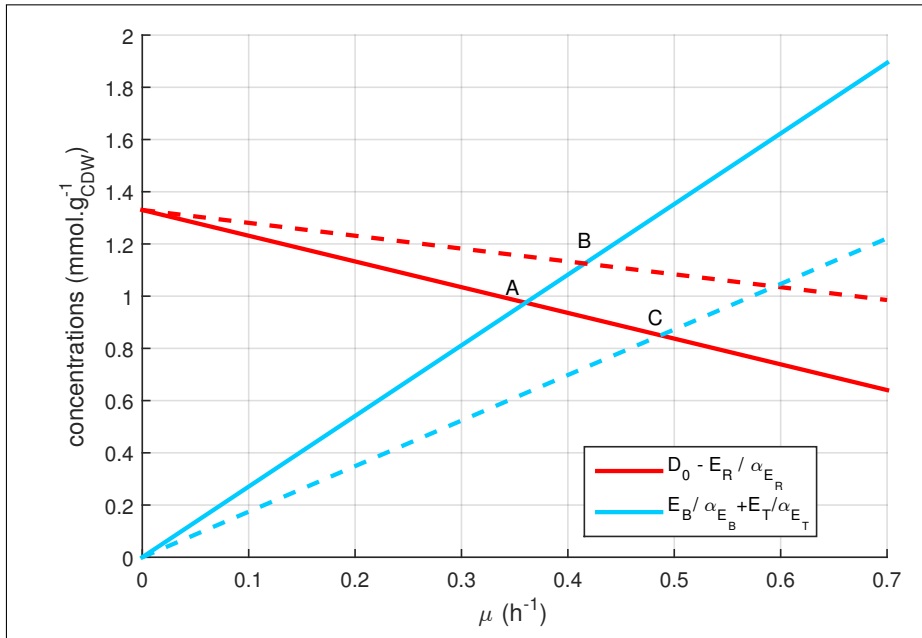


FIGURE 2.4 – Allure de $E_B/\alpha_{E_B} + E_T/\alpha_{E_T}$ et $D_0 - E_R/\alpha_{E_R}$ données par les équations (2.2.14) pour différentes valeurs de μ et des efficacités k_i fixées. En rouge, $D_0 - E_R/\alpha_{E_R}$, l'efficacité k_R est deux fois plus petite dans le cas du trait plein que dans le cas de la ligne discontinue. De même, en bleu, $E_B/\alpha_{E_B} + E_T/\alpha_{E_T}$, l'efficacité k_T est la même dans les deux cas, k_B est deux fois plus petite pour la ligne continue que pour la ligne discontinue.

On note qu'il n'y a qu'un point d'intersection, noté A sur la figure 2.4. Le taux spécifique de croissance à cette intersection est le taux spécifique de croissance maximal que la cellule ne peut dépasser pour des concentrations de $\mathcal{G}$ et de $\mathcal{S}$ fixées.

Étudions à présent l'influence des différents paramètres du modèle :

Si l'efficacité du processus Σ_R , $k_R(S)$, augmente, la pente de E_R diminue, et alors la droite $D_0 - E_R/\alpha_{E_R}$ se redresse, donnant lieu à un nouveau point d'intersection entre les deux droites, voir intersection B entre la ligne pointillée rouge et la ligne bleue en trait plein sur la figure 2.4. En comparant par rapport au point d'intersection A, le point B correspond à un taux spécifique de croissance plus important. De plus, la part de protéines $\mathcal{E}_B$ et $\mathcal{E}_T$ est plus importante dans le cas B que dans le cas A (l'ordonnée du point d'intersection est plus grande), la part de ribosomes est plus faible.

De même, si l'on augmente une des efficacités k_B ou k_T ou si l'on diminue la valeur de B_0 , la pente de $E_B/\alpha_{E_B} + E_T/\alpha_{E_T}$ diminue. Sur la figure 2.4, on a représenté le cas où l'efficacité k_B doublait par rapport à sa valeur de la ligne bleue en trait plein. Toutes choses restant égales par ailleurs, le point de fonctionnement est à ce moment là le point C.

Par rapport au point A, le taux spécifique de croissance est plus important bien que la part de protéines des processus Σ_B et Σ_T soit plus faible (car l'ordonnée du point C est plus faible que l'ordonnée du point A). La part de ribosomes est plus grande.

Ces observations (à S fixée) sont en accord avec ce qui a été prédit par la méthode RBA. Si l'efficacité d'un processus augmente, le taux spécifique de croissance augmente. Biologiquement, si l'efficacité d'un processus augmente, alors il y a besoin de moins de ressource pour catalyser le processus correspondant et alors de la ressource peut être distribuée entre tous les autres processus pour augmenter globalement la croissance.

Malgré tout, les non linéarités et l'imbrication de la concentration S dans les trois efficacités compliquent quelque peu l'élaboration de conclusions. Par exemple, on voit dans (2.1.7) et (2.1.11) que les efficacités k_B et k_R sont croissantes avec S . Comme S est croissante avec G d'après l'étude analytique précédente, on s'attend à ce que les efficacités k_B et k_R soient croissantes avec G .

Ainsi, si G augmente, k_R augmente, par l'analyse précédente, la pente de $D_0 - E_R/\alpha_{E_R}$ augmente et mène alors à une part de ribosomes plus faible. Dans le même temps, k_B augmente, et la pente de $E_B/\alpha_{E_B} + E_T/\alpha_{E_T}$ diminue et mène à une part de ribosomes plus importante... On ne peut pas conclure si la concentration en $\mathcal{E}_R$ est croissante ou décroissante avec la concentration de $\mathcal{G}$. On peut toutefois affirmer que la contrainte de densité mène à une imbrication du comportement des protéines à laquelle il est difficile d'échapper.

2.2.2.2 Approche quantitative

On cherche l'évolution des concentrations en $\mathcal{E}_B$, $\mathcal{E}_R$ et $\mathcal{E}_T$ qui maximisent le taux spécifique de croissance pour différentes concentrations extérieures en substrat G . D'après l'étude faite précédemment, et en particulier les équations (2.2.7) et (2.2.10), et avec les équations rappelées en (2.2.14), on est capable de trouver la répartition intracellulaire

en protéines qui maximise la croissance pour une concentration extérieure donnée de $\mathcal{G}$. On obtient les allures figure 2.5.

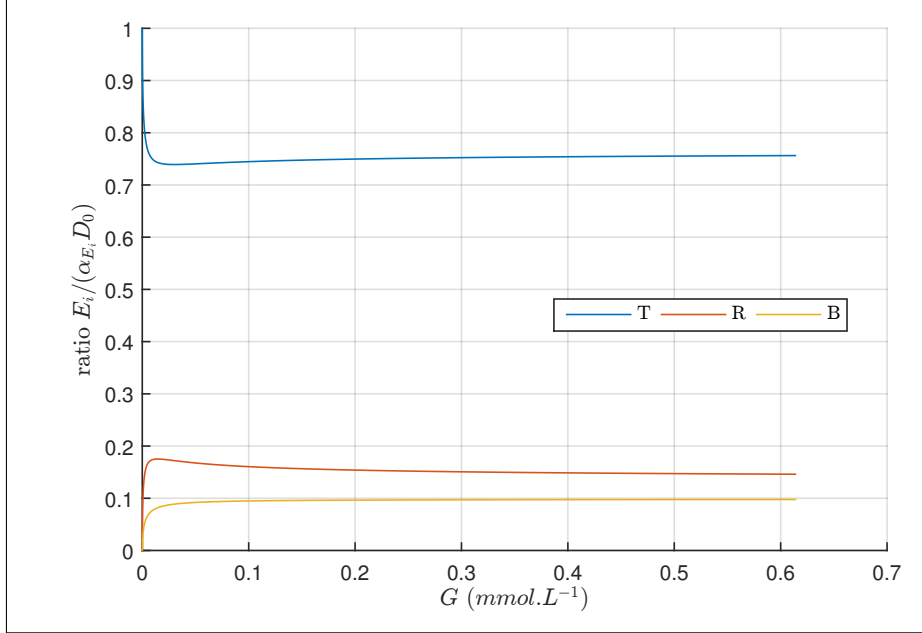


FIGURE 2.5 – Les ratios $\frac{E_i}{\alpha_{E_i} D_0}$ dans le cadre du modèle agrégé pour différentes concentrations extérieures de substrat $\mathcal{G}$.

On voit que les allures ne sont pas monotones, comme l'étude qualitative précédente le laissait penser.

Comme expliqué en début de section, pour caractériser notre modèle, il semble plus raisonnable de tracer les allures de ces ratios en fonction du taux spécifique de croissance. Il sera alors possible de comparer les tendances avec d'autres modélisations intracellulaires. On choisit comme référence le modèle RBA de [35] qui est validé expérimentalement sur plusieurs points de fonctionnement. Comme cette modélisation est faite à l'échelle du génome, il est nécessaire de procéder à une agrégation des protéines en pools de protéines pour pouvoir effectuer une comparaison entre les deux modèles.

On forme ainsi trois groupes, R, T et B, qui représentent les trois protéines $\mathcal{E}_R$, $\mathcal{E}_T$ et $\mathcal{E}_B$ du modèle agrégé.

- Dans le groupe R, on place les protéines ribosomales, les chaperones et les facteurs de traduction, c'est-à-dire les protéines qui participent au processus de traduction des protéines (voir définition du processus Σ_R).
- Dans le groupe T, on agrège toutes les protéines de transport et d'assimilation de carbohydrates, les protéines du métabolisme carboné central, de la respiration aérobie, de synthèse d'ATP, et des voies de synthèse des acides aminés. Il s'agit globalement des voies qui permettent de produire les composés présentés dans le tableau 2.1, donc des voies qui transforment le substrat $\mathcal{G}$ en $\mathcal{S}$.
- Le groupe B comprend toutes les protéines des voies de biosynthèse, dégradation et conversion des nucléotides et déoxy-nucléotides (ARN et ADN), des voies de dégradation des acides aminés, des voies de biosynthèse des vitamines et cofacteurs,

et des voies de biosynthèse des lipides et des composants de la paroi cellulaire. Il s'agit des réactions pour construire les ARN et les composés du tableau 2.2 à partir de ceux du tableau 2.1.

De fait, on omet une partie des protéines, y compris le groupe de protéines $\mathcal{P}_G$ dont il a été question au moment de donner une valeur à la densité.

On résout numériquement une série de problème RBA [35] pour des milieux saturés en ions et oxygène, et dans lequel le glucose est la seule source de carbone. Ce dernier est supposé être le limitant de la croissance. On fait varier la concentration de glucose de 0 à $0,5 \text{ mmol.L}^{-1}$ avec un pas de $0,1 \text{ } \mu\text{mol.L}^{-1}$. On trace par la suite la masse relative de chaque groupe par rapport à la masse totale des trois groupes en fonction du taux spécifique de croissance, ce qui donne les courbes en traits discontinus figure 2.6.

On trace également les expressions (2.2.14) en fonction du taux spécifique de croissance μ . Pour ce faire, on part d'un ensemble de concentrations de $\mathcal{S}$ variant de 0 à S_{max} (expression donnée en (2.2.8)). On en déduit les concentrations de $\mathcal{G}$ correspondantes par la formule (2.2.7), puis on trouve les taux spécifique de croissance correspondants par la formule (2.2.6). Il reste à substituer les valeurs de S , G et μ dans (2.2.14) pour avoir les concentrations des $\mathcal{E}_i$.

Les proportions massiques des groupes R, T et B sont dans notre contexte les ratios $\frac{E_i/\alpha_{E_i}}{\sum E_i/\alpha_{E_i}}$, c'est à dire la contribution de $\mathcal{E}_i$ dans la densité (se référer à la formule (2.1.12)). Ils sont tracés en trait plein sur la figure 2.6.

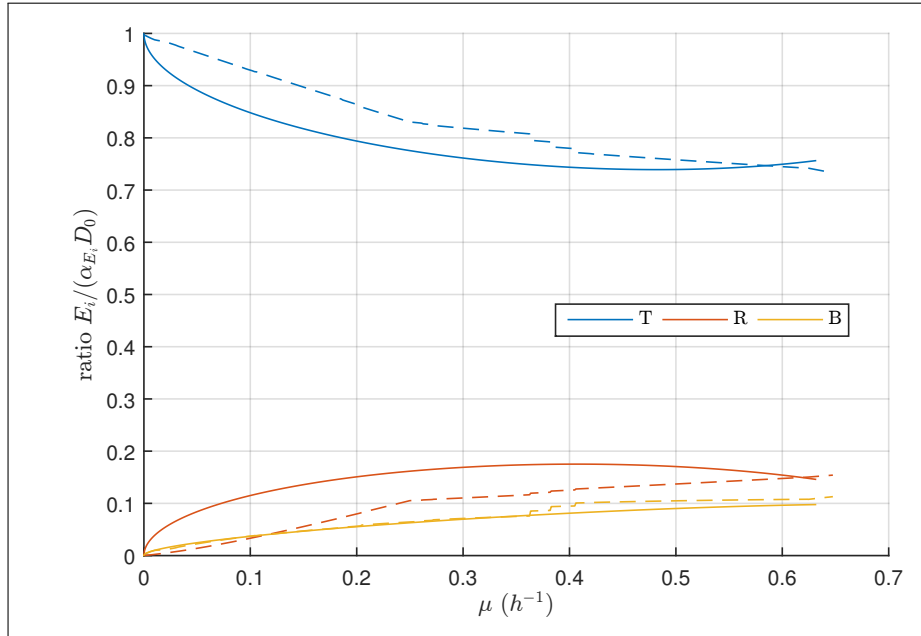


FIGURE 2.6 – En traits discontinus, la proportion massique des trois groupes B , R et T pour la méthode RBA. En traits pleins, les ratios $\frac{E_i}{\alpha_{E_i} D_0}$ dans le cadre du modèle agrégé.

Dans une première approximation, on peut voir que les comportements sont similaires : la part de $\mathcal{E}_T$ (et de protéines du groupe T) a tendance à diminuer avec le taux spécifique

de croissance, tandis que les parts de $\mathcal{E}_R$ et $\mathcal{E}_B$ (et des protéines des groupes R et B) augmentent avec le taux spécifique de croissance.

Les écarts sont de l'ordre de 5% entre la part de $\mathcal{E}_T$ et la part du groupe T. Idem pour la différence entre la part de $\mathcal{E}_R$ et la part du groupe R. Pour le groupe B, les écarts sont bien moindre, à peine 1%.

Pour une agrégation de plusieurs centaines de protéines en seulement 3 groupes, les écarts sont acceptables. Par ailleurs, les écarts sont significatifs surtout pour des valeurs de taux spécifique de croissance intermédiaires ($\mu = 0,1 - 0,4 \text{ h}^{-1}$). Pour des valeurs du taux spécifique de croissance proches du taux spécifique maximal, l'écart entre le modèle agrégé présenté ici et le modèle RBA n'est que de 2%.

2.2.2.3 Expression du taux spécifique de croissance en fonction de la concentration extérieure en glucose calculé par les RBA

Pour aller encore plus loin dans la comparaison avec une méthode validée expérimentalement, on compare ici les allures des taux spécifique de croissance maximaux pour la méthode RBA et pour notre modèle en fonction de la concentration en glucose (pour les RBA) et en $\mathcal{G}$ (pour notre modèle) dans le milieu extérieur. Les résultats sont présentés figure 2.7.

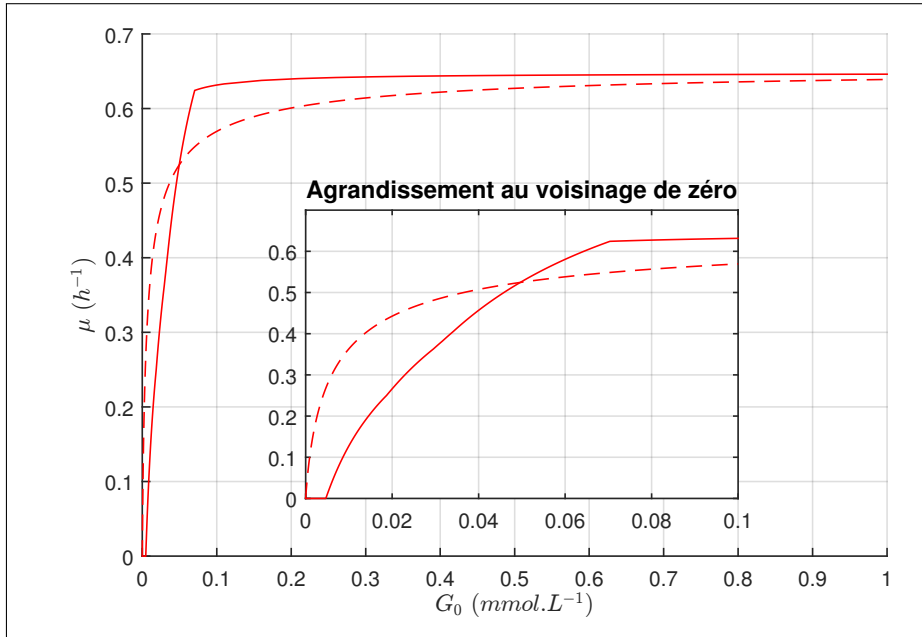


FIGURE 2.7 – En trait plein, le taux spécifique de croissance maximal en fonction de la concentration de glucose pour la méthode RBA ; en trait discontinu, un rappel de la figure 2.3 *i.e.* le taux spécifique de croissance de la méthode dRBA. Au centre, un agrandissement de la figure pour des valeurs de G_0 de 0 à $0,1 \text{ mmol.L}^{-1}$.

Il est intéressant de noter que la méthode RBA donne une expression du taux spécifique de croissance en fonction de la concentration de glucose proche d'une formule de Herbert (présenté en section 1.1.1.2 [77]), c'est-à-dire qui prend en compte une cer-

taine forme de maintenance : quand la concentration de glucose est inférieure à une concentration critique strictement positive, la croissance est rigoureusement nulle.

Les taux spécifique de croissance maximaux et les concentrations de substrat de demi-saturation du taux spécifique de croissance sont du même ordre de grandeur pour la méthode RBA à l'échelle du génome et pour notre modélisation donnée figure 2.3. Il y a certes un écart pour les concentrations de substrat de demi-saturation du taux spécifique de croissance mais on conserve un ordre de grandeur similaire (similaire également avec les travaux de Monod [60]), ce qui est satisfaisant. Par ailleurs, on a bon espoir qu'en détaillant les processus et en allant vers ce qui est fait avec la méthode RBA, on retrouve directement un profil de taux de croissance qui prenne en compte la maintenance.

2.2.3 Conclusion des différentes comparaisons

Dans la section 2.2.1, on a étudié le comportement macroscopique de notre modèle en régime quasi-stationnaire lorsqu'il maximise sa croissance instantanée. On a réussi à retrouver un taux spécifique de croissance maximal croissant en fonction de la concentration de substrat et asymptotiquement borné par une valeur maximale. Ceci est en accord avec des modèles empiriques issus d'observations. On a pu étudier sommairement l'impact des paramètres du modèle sur son comportement et constater que les tendances qui se dégageaient étaient en accord avec les phénomènes biologiques.

Par ailleurs, dans la section 2.2.2, on a réussi à rapprocher notre modèle agrégé de modèles intracellulaires plus fins et en particulier de la méthode RBA. Le fait que les allures observées soient en accord entre notre modèle agrégé et la modélisation RBA validée expérimentalement [35] nous conforte dans l'idée que le modèle agrégé est représentatif du comportement du modèle plus finement détaillé.

On a alors confiance en notre modèle pour considérer qu'il est une extension des modèles macroscopiques usuels, en apportant une description des processus intracellulaires. De plus, les correspondances entre les processus intracellulaires introduits dans notre modélisation et l'agrégation a posteriori de résultats de simulation de RBA nous conforte dans l'idée que notre agrégation a un sens et qu'il est possible de restreindre l'étude du comportement cellulaire à ces quelques processus agrégés.

Nous disposons à présent d'un modèle suffisamment représentatif et en accord avec les modélisations existantes pour que son utilisation dans la résolution de problèmes d'optimisation de bioprocédé donne des solutions réalistes. On a bon espoir que l'utilisation de notre modèle (extension des représentations classiques) donne des résultats qui soient des extensions de ce qui aurait été trouvé par l'utilisation de représentations classiques.

2.3 Optimisation de la production du bioprocédé

Le caractère descriptif de notre modèle étant validé, on s'intéresse, dans cette partie, à sa capacité prédictive en utilisant le modèle pour établir un scénario de commande du bioprocédé pour maximiser la production d'un composé d'intérêt, $\mathcal{P}$. Pour cela, on

cherche les profils $\nu(t)$ en fonction du temps qui maximisent un critère de productivité en accord avec la modélisation vue précédemment. On s'intéresse désormais à des situations dynamiques, sans se restreindre au régime permanent comme on l'a fait jusqu'à présent.

2.3.1 Position du problème

En reprenant le formalisme compact introduit en (2.1.19), le problème que l'on cherche ici à résoudre est de la forme :

$$\begin{aligned} & \max_{\nu(t), t_f} J(P(t_f), t_f) \\ \text{sous contraintes} & \begin{cases} \dot{x}(t) = F(x(t))\nu(t) \\ A\nu(t) \leq b(x(t)) \\ L\nu(t) = 0 \\ x(t) \geq 0 \\ \nu(t) \geq 0 \\ \Gamma(x(t_0), x(t_f), t_f) = 0 \end{cases} \end{aligned} \quad (2.3.1)$$

où Γ est un ensemble de contraintes aux bornes que doivent respecter les concentrations à l'instant initial et à l'instant final, t_f , ainsi que des contraintes potentielles sur le temps final. On reconnaît là un problème de commande optimale.

2.3.1.1 Critère

Le critère J est le reflet de l'efficacité du procédé à produire le composé d'intérêt, $\mathcal{P}$. A noter que l'efficacité du procédé dépend à la fois de la quantité (ou indifféremment de la concentration) de produit à la fin de la culture, mais également de la durée de la culture. En effet, il est aisé d'imaginer un procédé qui dure des jours et des jours et qui convertisse l'intégralité du substrat disponible en produit d'intérêt, le tout sous l'action d'une seule et unique bactérie ne se reproduisant pas. Effectivement, toute quantité de substrat qui sert à produire de la biomasse est de la matière qui ne devient pas du composé d'intérêt. C'est donc une perte nette du point de vue de la production. Aussi est-il souhaitable de ne pas produire du tout de biomasse si l'on cherche à maximiser le transfert vers le composé d'intérêt. Néanmoins, ce n'est pas la solution convenable pour le problème de la production de composé d'intérêt, ne serait-ce qu'en admettant que notre modèle ne prenne pas en compte la mortalité probable des cellules. Nous y reviendrons, en particulier au travers de la figure 2.10.

Dans la littérature, un compromis est classiquement trouvé en cherchant la solution qui maximise le rendement du procédé par rapport à la durée de culture [83], c'est-à-dire :

$$J = \frac{P(t_f)}{t_f} \quad (2.3.2)$$

où t_f est le temps final.

De cette manière, la durée de culture représente une pression pour la solution qui doit non seulement maximiser la production mais aussi en un temps le plus limité possible.

Il est à noter qu'un tel critère permet, de manière détournée, de représenter le stress qui agit sur les cellules pour faire le plus vite possible ce qu'elles ont à faire. Entre une

solution rapide et une plus lente, nous privilégierons ainsi la plus rapide pour se rapprocher de la réalité et des phénomènes de l'ordre de la sélection naturelle, qui favorise les organismes aux taux de croissance les plus efficaces. Ceci se rapproche des remarques faites sur la maximisation du taux spécifique de croissance, section 2.2.

D'un point de vue industriel, ce critère est aussi un bon indicateur de performance de l'efficacité d'un procédé à convertir la matière disponible en composé d'intérêt, comme rappelé dans [41].

2.3.1.2 Conditions aux bornes

On note t_0 l'instant initial. En ce qui concerne les conditions aux bornes, pour décrire le procédé, il suffit de fixer les concentrations extracellulaires, $G(t_0)$, $X(t_0)$, $P(t_0)$, ainsi que la concentration intracellulaire, $S(t_0)$. On peut tout à fait laisser les concentrations de protéines libres sous réserve qu'elles respectent la contrainte de densité : $\sum_i E_i(t_0)/\alpha_{E_i} = D_0$. Dans ces conditions, la densité intracellulaire reste égale à la densité D_0 à tout instant puisque μ est défini pour maintenir constante la densité au sens $\sum_i E_i(t)/\alpha_{E_i}$.

Les contraintes à l'instant initial s'écrivent :

$$\begin{cases} G(t_0), X(t_0), P(t_0), \text{ fixées} \\ S(t_0), \text{ fixée} \\ \sum_{P,B,R,T} \frac{E_i(t_0)}{\alpha_{E_i}} = D_0 \end{cases} \quad (2.3.3)$$

Concernant le temps final, t_f , on restera général en prenant : $0 < t_{min} \leq t_f \leq t_{max}$ en choisissant t_{min} (resp. t_{max}) assez petit (resp. assez grand). Dans le cas où une des deux bornes serait atteinte, on n'hésitera pas à la faire varier. L'objectif de t_{min} est de s'assurer que le critère reste défini (*i.e.* d'éviter une division par zéro). La raison visant à imposer un t_{max} sera exposée au moment d'établir la preuve d'existence. On veillera à ce que t_{max} ne soit pas une contrainte forte sur l'optimisation.

En posant C_0 et c_0 telles que :

$$\begin{cases} C_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{1}{\alpha_{E_B}} & \frac{1}{\alpha_{E_P}} & \frac{1}{\alpha_{E_R}} & \frac{1}{\alpha_{E_T}} & 0 & 0 & 0 \end{pmatrix} \\ c_0 = (S_0 \ G_0 \ P_0 \ X_0 \ D_0)^T \end{cases} \quad (2.3.4)$$

avec S_0 (respectivement G_0 , P_0 , X_0) la concentration initiale fixée de S (respectivement G , P , X), les contraintes (2.3.3) à l'instant initial s'écrivent :

$$C_0 x(t_0) = c_0$$

2.3.1.3 Simplification du problème

Lors de la phase de modélisation, nous avons précisé que les concentrations devaient être positives ou nulles à tout instant. La question qui se pose est, au vu de la formulation

du problème d'optimisation, de savoir s'il est nécessaire de conserver cette contrainte sur l'état du système. En effet, comme discuté dans [38], les contraintes sur les trajectoires des états sont difficiles à gérer dans le cadre d'un problème de commande optimale.

Nous allons voir qu'il est possible de remplacer cette contrainte $x(t) \geq 0, \forall t$ simplement par $G(t_f) \geq 0$. Cette contrainte sur l'état final sera bien plus facile à manipuler lors de l'optimisation.

A partir du moment où les flux sont positifs et que $X(t_0) \geq 0$, par (2.1.14), X est croissante et alors $X(t) \geq 0, \forall t$. Ainsi, par (2.1.8) et avec ν_P positif, on trouve que P est croissante et donc $P(t_0) \geq 0$ suffit pour affirmer que $P(t) \geq 0, \forall t$.

La dérivée de G , donnée par (2.1.3), étant négative ou nulle (car ν_T et $X \geq 0$), G est décroissante, il suffit alors que $G(t_0)$ et $G(t_f)$ soient positives pour que $G(t)$ le soit à tout instant.

Concernant les concentrations E_i , au vu de la dynamique (2.1.10), on peut affirmer pour $i \in \{P, B, R, T\}$ que : si E_i est positif à un instant particulier τ_0 , alors, cela est vrai pour tout instant $t \geq \tau_0$.

En effet, en raisonnant par l'absurde, s'il existe un instant τ_2 pour lequel E_i est strictement négatif, alors, par continuité, c'est qu'il existe un instant τ_1 entre τ_0 et τ_2 , pour lequel $E_i(\tau_1)$ est nul et de dérivée temporelle strictement négative. Or la dérivée de E_i en τ_1 s'écrit, d'après (2.1.10) : $\dot{E}_i(\tau_1) = \alpha_{E_i} \nu_{E_i}(\tau_1) - 0$, qui est positive ou nulle. D'où la contradiction.

Il suffit donc de contraindre $E_i(t_0)$ à être positif pour que $E_i(t)$ soit positif pour tout $t \geq t_0$.

Étant donné que l'on fixe $S(t_0)$ à une valeur S_0 (positive), les efficacités $k_B(t_0)$, $k_P(t_0)$ et $k_R(t_0)$ sont toutes trois positives. De même, comme $G(t_0)$ est positive, $k_T(t_0)$ est positive.

Les contraintes (2.1.4), (2.1.7), (2.1.9) et (2.1.11) prises en t_0 , avec le fait que les $\nu_i(t_0)$ sont positifs, donnent alors la positivité des $E_i(t_0)$.

Il ne reste plus qu'à montrer que $S(t)$ est positive à chaque instant. Raisonnons par l'absurde. $S(t)$ est absolument continue au vu de la relation (2.1.2). Ainsi, par continuité, si $S(t)$ n'est pas toujours positive, on peut trouver un $\varepsilon \geq 0$ aussi petit que l'on veut et un intervalle $[\tau_1, \tau_2]$ tels que $-\varepsilon < S(t) < 0, \dot{S}(t) < 0, \forall t \in [\tau_1, \tau_2]$. On suppose que ε est plus petit que le minimum entre K_R , K_P et K_B . Dès lors, $k_B(S(t))$, $k_P(S(t))$ et $k_R(S(t))$ sont négatifs pour $t \in [\tau_1, \tau_2]$. Donc, d'après les contraintes (2.1.4), (2.1.7), (2.1.9) et (2.1.11) et le fait que les $\nu_i(t)$ soient positifs, ν_P , ν_B , les ν_{E_i} et E_P , E_B et E_R sont nuls sur l'intervalle $[\tau_1, \tau_2]$. Comme tous les ν_{E_i} sont nuls, μ l'est aussi par la relation (2.1.13). Ainsi, la dérivée temporelle de S s'écrit d'après (2.1.2) : $\dot{S} = \alpha_T \nu_T(t)$ qui est positif, ce qui contredit l'hypothèse selon laquelle $\dot{S}(t) < 0$.

La concentration $S(t)$ est bien positive pour tout t .

En résumé, $x(t) \geq 0$ peut être substitué, dans la formulation du problème, par :

$$G(t_f) \geq 0 \quad (2.3.5)$$

Matriciellement, on pose C_f, c_f telles que :

$$\begin{cases} C_f = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \\ c_f = (0) \end{cases} \quad (2.3.6)$$

et on écrira :

$$C_f x(t_f) \geq c_f$$

2.3.1.4 Formulation compacte

Une forme compacte du problème (2.3.1) est alors :

$$\begin{aligned} & \max_{\nu(t), t_f} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{t_f} & 0 \end{pmatrix} x(t_f) \\ & \text{s.c.} \begin{cases} \dot{x}(t) = F(x(t))\nu(t) & (2.3.7.a) \\ A\nu(t) \leq b(x(t)) & (2.3.7.b) \\ L\nu(t) = 0 & (2.3.7.c) \\ \nu(t) \geq 0 & (2.3.7.d) \\ C_0 x(t_0) = c_0 & (2.3.7.e) \\ C_f x(t_f) \geq c_f & (2.3.7.f) \\ t_{min} \leq t_f \leq t_{max} & (2.3.7.g) \end{cases} \end{aligned} \quad (2.3.7)$$

On cherche la solution à ce problème de commande optimale où :

- le critère porte sur l'état à l'instant final (formulation de Mayer) ;
- la dynamique est linéaire en la commande, (2.3.7.a) ;
- les contraintes sur la commande sont linéaires, (2.3.7.c) (2.3.7.d), ou dépendent de l'état, (2.3.7.b) ;
- il y a des conditions sur l'état initial et l'état final, (2.3.7.e) et (2.3.7.f) ;
- le temps final est libre mais borné, (2.3.7.g).

Nous allons voir ce que nous pouvons dire de ce problème.

2.3.2 Commande optimale

Les contraintes sur la commande dépendant de l'état du système ne rentrent pas dans le cadre habituellement considéré du principe du maximum de Pontryaguine. Ainsi, la formulation du problème, bien que compacte, n'est pas si classique.

En effet, soient :

$$\Omega(x) = \{\nu \in \mathbb{R}_+^m \mid L\nu = 0, A\nu \leq b(x)\} \quad (2.3.8)$$

et

$$N(x) = \{F(x)\nu \mid \nu \in \Omega(x)\}$$

pour un x donné, les variables de commande prennent valeur dans $\Omega(x)$. Dans ce contexte, le principe du maximum est remplacé par des conditions plus avancées telles que celles associées à un problème de Mayer où la dynamique du système est définie par une inclusion différentielle non linéaire (voir par exemple [25] ou [81]).

2.3.2.1 Existence

Il est cependant possible de prouver l'existence d'une solution au problème. Pour cela, on applique le théorème de Fillipov-Cesari [38], qui garantit l'existence d'une solution optimale mesurable sous réserve que l'état, la commande et le temps final soient bornés, et que $N(x)$ soit un ensemble convexe quel que soit x .

2.3.2.2 Preuve

Par conservation de la masse, la masse de chaque composé est majorée par $m_{max} = m_G + m_S + m_P + \sum m_{E_i}$, où m_ξ est la quantité de ξ à l'instant initial. La population XV étant par ailleurs croissante, et le volume du réacteur V étant constant, les concentrations de chaque composé rapportées au volume de population sont bien majorées (*i.e.* S , E_i) ainsi que les concentrations de chaque composé dont la concentration est rapportée au volume du réacteur (*i.e.* G , P , X).

Dès lors, les contraintes d'efficacité des processus permettent de trouver des majorants généraux pour les éléments de ν .

Le temps final t_f est lui majoré par définition du problème par t_{max} .

Il reste à vérifier que la condition de Roxin est satisfaite, *i.e.* que $N(x)$ est bien convexe pour tout $x \in \mathbb{R}^n$.

Soient $x_0 \in \mathbb{R}^n$, n_1 et n_2 deux éléments de $N(x_0)$. Il existe ν_1 et ν_2 dans $\Omega(x_0)$ tels que $n_1 = F(x_0)\nu_1$ et $n_2 = F(x_0)\nu_2$. Comme ν_1 et ν_2 sont dans $\Omega(x_0)$, ils vérifient : $L\nu_1 = 0$, $A\nu_1 \leq b(x_0)$ et $\nu_1 \in \mathbb{R}_+^m$ et de même pour ν_2 , $L\nu_2 = 0$, $A\nu_2 \leq b(x_0)$ et $\nu_2 \in \mathbb{R}_+^m$.

Soient $t \in [0,1]$, $n_m = tn_1 + (1-t)n_2$ et $\nu_m = t\nu_1 + (1-t)\nu_2$.

— On a $t > 0$, $1-t > 0$, $\nu_1, \nu_2 \in \mathbb{R}_+^m$ donc on a bien $\nu_m \in \mathbb{R}_+^m$,

— $L\nu_m = tL\nu_1 + (1-t)L\nu_2 = 0$,

— $A\nu_m = tA\nu_1 + (1-t)A\nu_2 \leq tb(x_0) + (1-t)b(x_0) = b(x_0)$

ainsi, $\nu_m \in \Omega(x_0)$. Par ailleurs, $F(x_0)\nu_m = tF(x_0)\nu_1 + (1-t)F(x_0)\nu_2 = tn_1 + (1-t)n_2 = n_m$, donc n_m est bien dans $N(x_0)$. D'où la convexité de $N(x_0)$, puis de tout $N(x)$ pour $x \in \mathbb{R}^n$.

Ainsi, $N(x)$ étant convexe pour tout x ; x , ν et t_f étant bornés, on applique le théorème de Filippov-Cesari et on conclut qu'il existe un triplet optimal $(x^*(\cdot), \nu^*(\cdot), t_f^*)$ avec $\nu^*(\cdot)$ mesurable. $\square$

Ceci nous donne le fait que la commande optimale $\nu^*(t)$ existe et est mesurable. En tant que telle, elle peut donc être approchée par une suite de fonctions en escaliers.

2.3.2.3 Dérivation de conditions nécessaires d'optimalité

Si l'on suppose en outre que la commande optimale a une certaine propriété de continuité, on peut essayer de déterminer un certain nombre de conditions nécessaires que trajectoire optimale doit satisfaire, comme présenté dans [38]. Pour ce faire, il suffit de supposer que la commande optimale ν^* est continue à droite avec des limites à gauche.

Introduisons l'Hamiltonien $\mathcal{H}$ défini par

$$\mathcal{H}(x(t), \nu(t), \lambda(t)) = \lambda(t)^\top F(x(t)) \nu(t) \quad (2.3.9)$$

où λ appartient à $\mathbb{R}^n$ et est l'adjoint de l'état x .

Soit également $\mathcal{L}$, le Lagrangien associé défini par

$$\begin{aligned} \mathcal{L}(x(t), \nu(t), \lambda(t), \eta_1(t), \eta_2(t), \eta_3(t)) = & \mathcal{H}(x(t), \nu(t), \lambda(t)) \\ & + \eta_1(t) L \nu(t) + \eta_2(t)^\top (b(x(t)) - A \nu(t)) + \eta_3(t)^\top \nu(t) \end{aligned} \quad (2.3.10)$$

où η_1, η_2, η_3 sont les multiplicateurs associés aux contraintes.

Ensuite, si l'on pose $(x^*(t), \nu^*(t))$ un couple optimal, alors, il existe une fonction adjoint absolument continue par morceaux λ , des fonctions multiplicatrices continues par morceaux η telles que :

- (a) $\nu^*(t) = \arg \max_{\nu \in \Omega(x^*(t))} \mathcal{H}(x^*(t), \nu, \lambda(t))$;
- (b) $\dot{\lambda}(t) = -\frac{\partial \mathcal{L}}{\partial x}(x^*(t), \nu^*(t), \lambda(t), \eta_1(t), \eta_2(t), \eta_3(t))$ (cette équation est valable uniquement car il n'y a pas de contrainte pure sur l'état) ;
- (c) $\frac{\partial \mathcal{L}}{\partial \nu}(x^*(t), \nu^*(t), \lambda(t), \eta_1(t), \eta_2(t), \eta_3(t)) = 0$;
- (d) $\eta_1(t) \in \mathbb{R}, \eta_1(t) L \nu^*(t) = 0$;
- (e) $\eta_2(t) \in \mathbb{R}_+^{n_2}, \eta_2(t) \circ [b(x^*(t)) - A \nu^*(t)] = 0_{n_2 \times 1}$, avec le produit matriciel terme à terme de Hadamard noté $\circ$;
- (f) $\eta_3(t) \in \mathbb{R}_+^{n_3}, \eta_3(t) \circ \nu^*(t) = 0_{n_3 \times 1}$;
- (g) $\lambda(t_0) = \frac{\partial(\alpha_0^\top (C_0 x - c_0))}{\partial x}$, $\lambda(t_f) = \frac{\partial(J + \alpha_f(C_f x - c_f))}{\partial x}$ avec $\alpha_0 \in \mathbb{R}^{n_0}$ et $\alpha_f \in \mathbb{R}_+$, avec $n_0 = \dim(c_0)$.

Habituellement, on peut arriver à trouver une expression de $\nu^*(t)$ en fonction de $x(t)$ et de $\lambda(t)$ en partant de la condition (a). En fonction du signe de $\mathcal{H}_{\nu_i}(t) = \frac{\partial \mathcal{H}}{\partial \nu_i}(t)$, $\nu_i^*(t)$ vaut soit sa valeur minimale (si $\mathcal{H}_{\nu_i}(t) < 0$), soit sa valeur maximale (si $\mathcal{H}_{\nu_i}(t) > 0$), soit une valeur intermédiaire donnée par le reste des contraintes (si $\mathcal{H}_{\nu_i}(t) = 0$). Les valeurs minimales et maximales sont données par rapport à l'ensemble $\Omega(x^*(t))$, défini en (2.3.8).

Ce qui complique la résolution ici, est le fait que les ensembles $\Omega(x^*(t))$ ont des valeurs extrémales qui sont définies par rapport à l'état mais surtout par rapport aux commandes, à cause de la contrainte $A \nu \leq b(x)$ et au fait que A ne soit pas inversible. La caractérisation analytique semble compromise. Il faut se tourner vers d'autres types de caractérisation de la solution optimale.

2.3.3 Formulation discrète

Les méthodes analytiques traditionnelles semblent compromises par la formulation du problème et en particulier les contraintes entre l'état et la commande. En reprenant l'idée que la commande optimale est mesurable, et approchable par des fonctions en escaliers, on se dirige alors vers des méthodes numériques pour résoudre le problème (2.3.7).

L'idée est de convertir le problème de commande optimale de dimension infinie en un problème non linéaire de dimension finie. Pour ce faire, on procède comme dans [6], c'est-à-dire la discrétisation de la dynamique sur $N + 1$ pas de temps de t_0 à t_f puis on cherche les valeurs des commandes et des états à ces instants de discrétisation. Le pas de discrétisation vaut $\frac{t_f - t_0}{N}$. Il convient de déterminer également des équivalents des différentes contraintes pour ce cas discret.

2.3.3.1 Variable d'optimisation

La variable d'optimisation $\mathbf{X}$ comprend l'ensemble des états et commandes aux instants de discrétisation ainsi que le temps final :

$$\mathbf{X} = \{\{x_k, \nu_k\}_{k \in \{0, \dots, N\}}, t_f\} \in \mathbb{R}^{(n+m) \times (N+1) + 1} \quad (2.3.11)$$

x_k est l'état du système discrétisé à l'instant $t_k = t_0 + (t_f - t_0)\frac{k}{N}$ et ν_k la valeur de la commande à t_k .

x_N est l'état à l'instant final et on cherche donc à maximiser $\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{t_f} & 0 \end{pmatrix} x_N$. Rapporté à $\mathbf{X}$, on cherche à maximiser $\frac{1}{t_f} c^T \mathbf{X}$ avec c le $((n+m) \times N + 7)$ -ième élément de la base canonique de $\mathbb{R}^{(n+m) \times (N+1) + 1}$.

2.3.3.2 Contrainte liée à la dynamique

La discrétisation du système peut être faite de différentes manières. La plus simple, et qui sera utilisée ici, consiste à appliquer un bloqueur d'ordre zéro à la commande entre deux pas de temps et à utiliser une formulation d'Euler explicite pour approcher la dérivée de l'état entre deux pas de temps. On préférera augmenter le nombre de pas d'échantillonnage plutôt que de complexifier la formulation, par exemple via une approximation polynomiale de la commande entre deux pas d'échantillonnage et une estimée x_{k+1} calculée par une formule plus élaborée, voir [73]. On veillera à choisir un pas de discrétisation assez faible pour que le schéma d'Euler soit correct.

On considère que :

$$x_k = x_{k-1} + \frac{t_f - t_0}{N} F(x_{k-1}) \nu_{k-1} \quad (2.3.12)$$

pour $k \in \llbracket 1, N \rrbracket$.

On pose $\mathbf{R}_k(\mathbf{X}) = x_{k+1} - x_k - \frac{t_f - t_0}{N} F(x_k) \nu_k$, la k -ième contrainte de raccord.

La contrainte sur la dynamique du système peut alors être remplacée par

$$\begin{aligned} \dot{x}(t) = F(x(t))\nu(t) & \rightarrow x_k = x_{k-1} + \frac{t_f - t_0}{N} F(x_{k-1}) \nu_{k-1} & \rightarrow \mathbf{R}(\mathbf{X}) = \begin{pmatrix} \mathbf{R}_1(\mathbf{X}) \\ \vdots \\ \mathbf{R}_N(\mathbf{X}) \end{pmatrix} = 0 \\ \forall t \in [t_0, t_f] & \quad \forall k \in \llbracket 1, N \rrbracket \end{aligned} \quad (2.3.13)$$

A noter que dans le cas d'un bloqueur d'ordre zéro pour estimer la dynamique, la dernière commande ν_N n'intervient pas. Dans le formalisme développé dans la suite, on présente le cas général, sans s'en soucier.

2.3.3.3 Autres contraintes

Les contraintes sur la commande et mixtes entre l'état et la commande deviennent immédiatement :

$$\begin{cases} \nu(t) \geq 0 \\ A\nu(t) \leq b(x(t)) \\ L\nu(t) = 0 \\ \forall t \in [t_0, t_f] \end{cases} \rightarrow \begin{cases} \nu_k \geq 0 \\ A\nu_k \leq b(x_k) \\ L\nu_k = 0 \\ \forall k \in \llbracket 1, N \rrbracket \end{cases} \rightarrow \begin{cases} \mathbf{P} \cdot \mathbf{X} \geq 0 \\ \mathbf{A} \cdot \mathbf{X} \leq \mathbf{b}(\mathbf{X}) \\ \mathbf{L} \cdot \mathbf{X} = 0 \end{cases} \quad (2.3.14)$$

avec $\mathbf{P}$, $\mathbf{A}$ et $\mathbf{L}$ des matrices construites par produits de Kronecker bien choisis, et $\mathbf{b}$ une fonction matricielle de $\mathbf{X}$ bien choisie également.

A noter que le respect des contraintes aux instants de discrétisation ne garantit pas le respect des contraintes entre ces instants. Il conviendra de vérifier a posteriori qu'il n'y a pas de dépassement hors des instants de discrétisation.

2.3.3.4 Formulation discrète

Finalement les contraintes aux bornes deviennent :

$$\begin{cases} C_0 x(t_0) = c_0 \\ C_f x(t_f) \geq c_f \\ t_{min} \leq t_f \leq t_{max} \end{cases} \rightarrow \begin{cases} C_0 x_0 = c_0 \\ C_f x_N \geq c_f \\ t_{min} \leq t_f \leq t_{max} \end{cases} \rightarrow \begin{cases} \mathbf{C}_0 \cdot \mathbf{X} = c_0 \\ \mathbf{C}_f \cdot \mathbf{X} \geq c_f \\ t_{min} \leq \mathbf{e}_{(n+m) \times (N+1)+1} \cdot \mathbf{X} \leq t_{max} \end{cases} \quad (2.3.15)$$

avec $\mathbf{C}_0$ et $\mathbf{C}_f$ des matrices bien choisies et $\mathbf{e}_{(n+m) \times (N+1)+1} = (0 \ \cdots \ 0 \ 1)$.

Le problème discrétisé est alors le suivant :

$$\begin{aligned} & \max_{\mathbf{X} \in \mathbb{R}^{(n+m) \times (N+1)+1}} \frac{1}{t_f} c^T \mathbf{X}(t_f) \\ & \text{s.c.} \begin{cases} \mathbf{R}(\mathbf{X}) = 0 \\ \mathbf{P} \cdot \mathbf{X} \geq 0 \\ \mathbf{A} \cdot \mathbf{X} \leq \mathbf{b}(\mathbf{X}) \\ \mathbf{L} \cdot \mathbf{X} = 0 \\ \mathbf{C}_0 \cdot \mathbf{X} = c_0 \\ \mathbf{C}_f \cdot \mathbf{X} \geq c_f \\ t_{min} \leq \mathbf{e}_{(n+m) \times (N+1)+1} \cdot \mathbf{X} \leq t_{max} \end{cases} \end{aligned} \quad (2.3.16)$$

La question est tout de même de déterminer le lien qui existe entre les solutions au problème de commande optimale (2.3.7) et au problème discrétisé (2.3.16). Une piste de réponse se trouve section 10.4 de [6]. Il apparaît qu'en affinant le pas de discrétisation, la solution du problème (2.3.16) tend vers la solution du problème (2.3.7).

Ainsi, dans la mesure où les trajectoires solutions de (2.3.16) seront régulières, le risque de converger vers une solution complètement différente restera maîtrisé. On pourra réduire le pas de discrétisation pour s'assurer que la solution obtenue n'est pas radicalement différente.

2.3.4 Résolution numérique

Le passage de (2.3.7) à (2.3.16) a été réalisé grâce au logiciel **Bocop** développé par une équipe de l'INRIA [10]. **Bocop** est un logiciel de résolution de problème de commande optimale par la méthode présentée en section 2.3.3 dite méthode directe. Une fois le problème écrit sous la forme d'un problème non linéaire de dimension finie, la solution est trouvée par un algorithme du point intérieur. Il s'agit essentiellement d'une descente de gradient du critère qui est augmenté d'un ensemble de pénalités relatives aux contraintes d'égalités et d'inégalités. On se référera à l'article [82] pour plus de détails.

Avec les conditions initiales $G(t_0) = 0,1 \text{ mmol/L}$, $P(t_0) = 0 \text{ mmol/L}$, $X(t_0) = 1 \text{ mg}_{CDW}/L$, et $S(t_0) = 0,019 \text{ mmol/g}_{CDW}$ (pris en accord avec la figure 2.2), $t_{min} = 0,1 \text{ h}$, $t_{max} = 50 \text{ h}$, un nombre d'échantillonnage $N = 100$, et les paramètres de la table 2.3, on obtient les allures de la figure 2.8. En testant plusieurs initialisations au problème d'optimisation, on trouve toujours la même solution. Le temps de résolution dépend lui de l'initialisation (de quelques dizaines de secondes à quelques minutes).

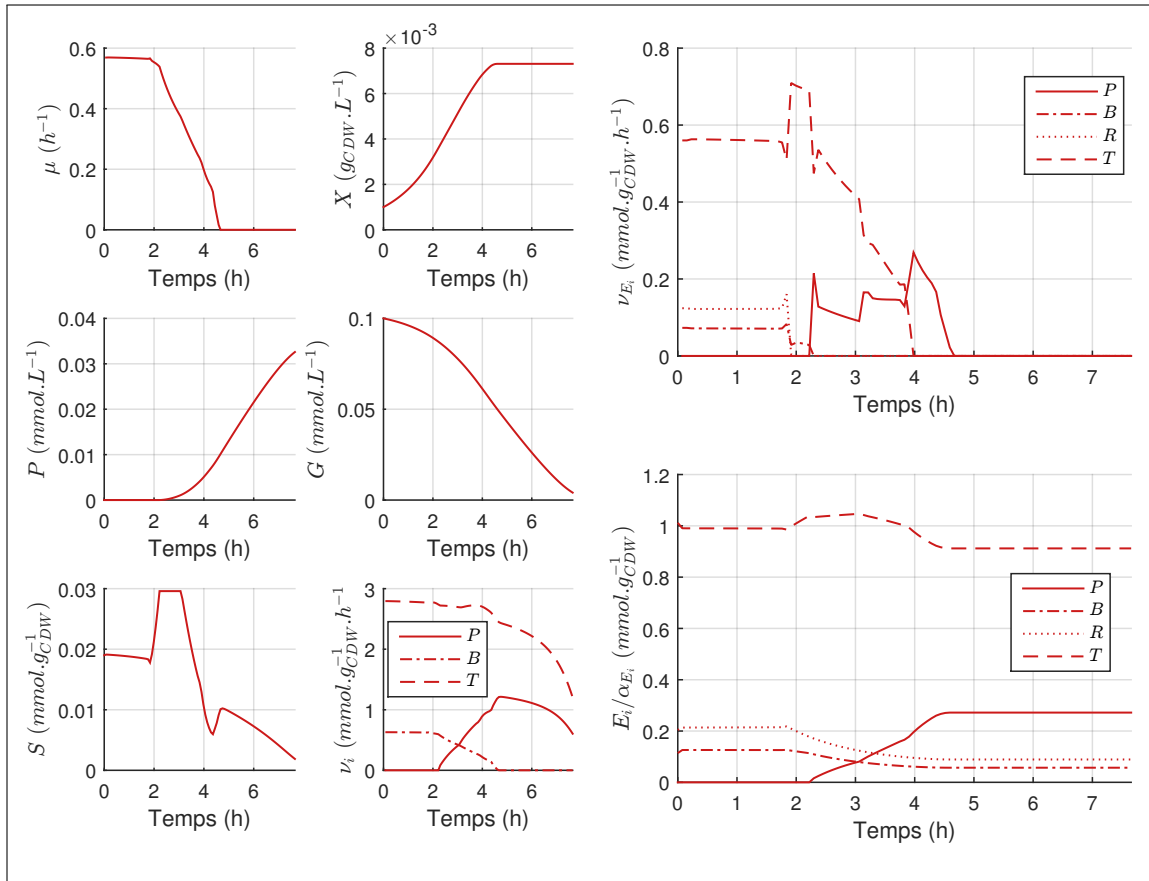


FIGURE 2.8 – Solution du problème (2.3.16).

Si l'on augmente le nombre de pas de discrétisation (tests effectués jusqu'à $N = 1000$), les résultats sont inchangés. On passe cependant d'une résolution en quelques secondes à plus d'une heure avec un ordinateur de bureau muni d'un processeur Intel Core i7. Par

ailleurs, nous nous sommes assurés que les résultats présentés dans ce chapitre étaient indépendants des valeurs numériques choisies pour les conditions initiales. Pour peu que l'on satisfasse la relation qui lie G à S à l'équilibre (équation (2.2.7)), on obtient des allures similaires à celles présentées par la suite.

2.3.5 Analyse des résultats

On trouve les valeurs optimales suivantes :

$$\begin{cases} P(t_f) = 32,6 \cdot 10^{-3} \text{ mmol.L}^{-1} \\ t_f = 7,66 \text{ h} \\ \frac{P(t_f)}{t_f} = 4,26 \cdot 10^{-3} \text{ mmol.L}^{-1} \cdot \text{h}^{-1} \end{cases}$$

On obtient trois phases successives :

- (i) La première, de $t = 0h$ à $t = 1,8h$, est une phase de croissance exponentielle durant laquelle les concentrations intracellulaires sont constantes. C'est bien un régime quasi-statique dans la mesure où le taux spécifique de croissance est constant. On retrouve la valeur du taux spécifique de croissance maximal que l'on puisse obtenir à la concentration $G = G(t_0)$, *c.f.* figure 2.3. Les valeurs pour des concentrations intracellulaires en protéines et métabolite S correspondent à la solution du problème statique ($\Pi_{G(t_0)}$) (formulation (2.2.1)).
- (ii) La deuxième phase, de $t = 1,8h$ à $t = 4,7h$, est une période de transition durant laquelle la configuration de la cellule change complètement. Le taux spécifique de croissance diminue progressivement, la production du composé d'intérêt $\mathcal{P}$ commence, mais ce qui est le plus important est la composition en protéines : tandis que les concentrations en $\mathcal{E}_B$ et $\mathcal{E}_R$ diminuent, la concentration en $\mathcal{E}_P$ augmente. Celle en $\mathcal{E}_T$ reste relativement constante avant de diminuer à son tour en fin de phase. Les ressources sont dirigées vers le passage à une configuration favorable pour produire le produit d'intérêt. Les flux ν_{E_i} vont clairement dans ce sens.
- (iii) Enfin, la dernière phase, de $t = 4,7h$ à $t = 7,7h$, est une période de production pure du composé d'intérêt. La croissance est nulle, les concentrations intracellulaires en protéines sont fixes, permettant un transfert efficace de $\mathcal{G}$ à S puis de S à $\mathcal{P}$ sans utiliser de ressource pour d'autres tâches. Les cellules sont ici de pures machines catalytiques convertissant parfaitement $\mathcal{G}$ en $\mathcal{P}$.

Comme on l'a vu au chapitre précédent, la première et la troisième phases sont classiquement utilisées en bioproduction de composés d'intérêt en utilisant des modèles macroscopiques phénoménologiques. Il est intéressant de constater qu'on retrouve la même politique en partant de la description des processus intracellulaires.

En utilisant une modélisation FBA, c'est-à-dire sans prendre en compte le fait que les bornes sur les flux ($k_i E_i$) résultent d'un compromis entre entités du même type (en l'occurrence, d'un compromis entre composés peptidiques), on arrive à retrouver les deux phases de croissance (i) puis de production (iii). La phase de transition (ii) reste cependant absente des prédictions. On retrouve bien la même idée séquentielle dans la solution au problème d'optimisation posé avec notre modèle. Un réel avantage de notre approche par rapport aux méthodes présentes dans la littérature est la caractérisation de cette phase de transition par des considérations physico-biologiques. En effet, les

phases de transition obtenues par des méthodes FBA apparaissent grâce à l'introduction de limitations artificielles sur la dérivée temporelle des flux : des contraintes de type $\dot{\alpha}_i \leq \frac{\nu_i(t_k) - \nu_i(t_{k-1})}{t_k - t_{k-1}} \leq \dot{\beta}_i$ sont ajoutées au problème avec $\dot{\alpha}_i$ et $\dot{\beta}_i$ des coefficients supplémentaires déterminés de sorte à faire coïncider les résultats expérimentaux et les simulations numériques. Dans notre travail, on obtient le phénomène de façon autonome à travers les contraintes d'allocation de ressources.

Par ailleurs, sur le plan de la stratégie de la cellule lors du changement de phase (i) à (ii), c'est-à-dire avant de modifier son programme interne, on note que la concentration en $\mathcal{S}$ augmente. Ceci semble dû au fait que le flux de production de $\mathcal{S}$ est supérieur à la somme des flux de consommation. En particulier, le flux de Σ_R chute à cet instant. Pour le reste, $S(t)$ décroît au cours du temps, tout comme la concentration en substrat extérieure, $G(t)$.

On peut étudier les allures des concentrations en protéines en fonction du taux spécifique de croissance, comme présenté figure 2.9.

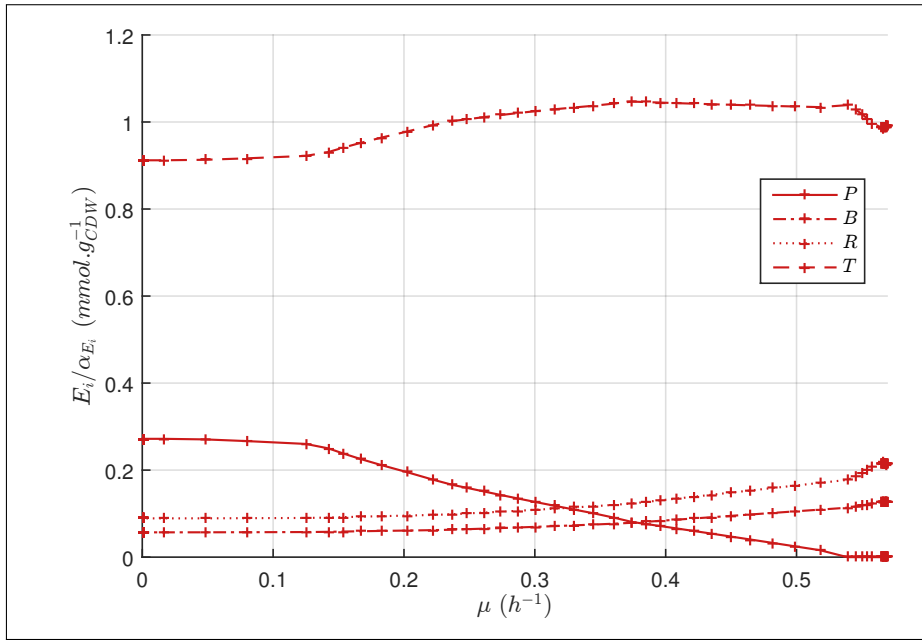


FIGURE 2.9 – Correspondance entre la concentration des protéines et le taux spécifique de croissance pour la solution au problème (2.3.16).

On constate que la stratégie optimale revient à une sorte de correspondance plus ou moins linéaire par morceaux entre les concentrations en protéines et le taux spécifique de croissance. Nous reviendrons sur cette correspondance dans le chapitre suivant.

2.3.5.1 Retour sur le critère d'optimisation

Revenons sur le choix du critère, discuté au début de la section 2.3.1. On résout le problème (2.3.16) en prenant comme critère à maximiser :

$$J_0 = P(t_f) \quad (2.3.17)$$

le temps final étant par ailleurs libre, borné par t_{min} et t_{max} . Le résultat obtenu par résolution numérique est donné figure 2.10. On notera que la valeur du temps final est égale à sa borne maximale (ici, $50h$). Conformément à nos attentes, la population reste constante à sa valeur initiale. La croissance est rigoureusement nulle et toutes les concentrations restent constantes, égales à leur valeur à l'instant initial. Le choix des concentrations est libre à l'instant initial, pourvu que (2.3.3) soit satisfaite.

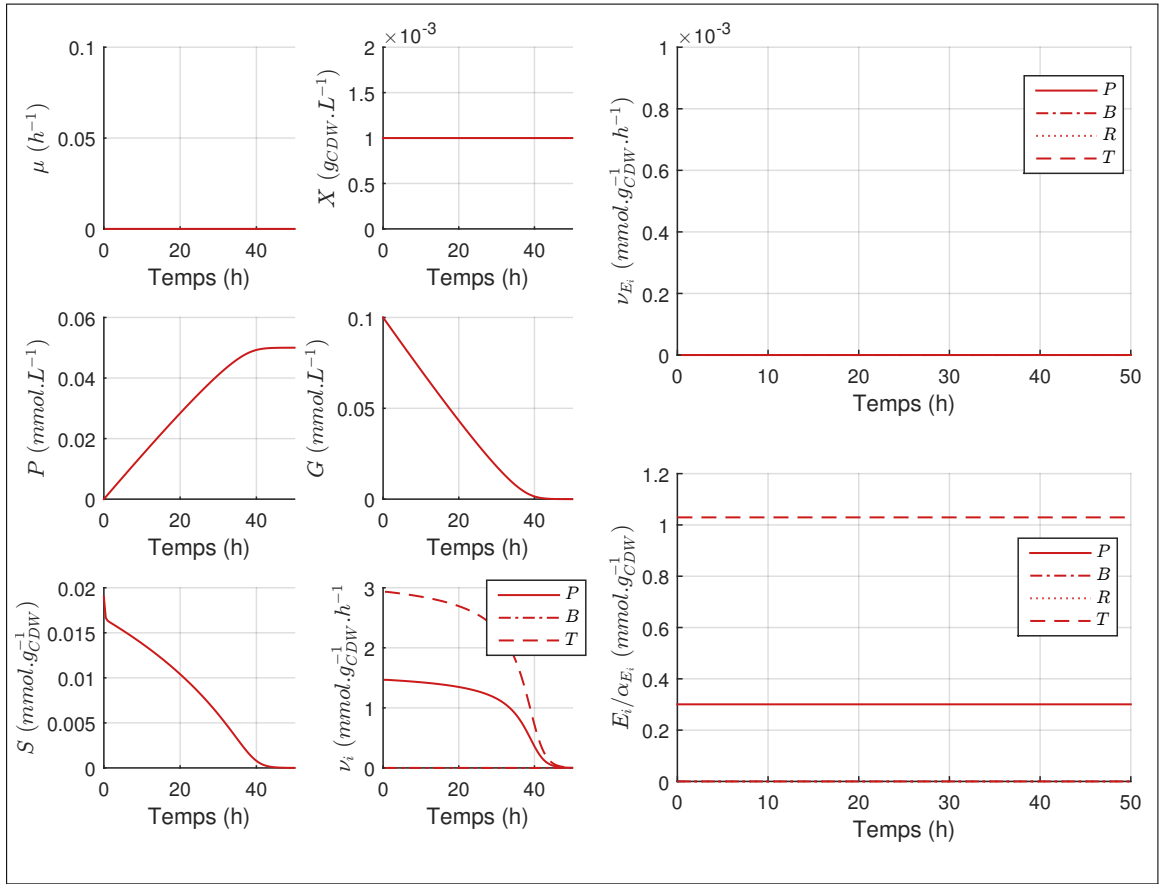


FIGURE 2.10 – Solution du problème (2.3.16) avec comme critère $J_0 = P(t_f)$.

La répartition optimale des protéines est alors :

$$\begin{cases} E_T = 1,03 \text{ mmol.g}_{CDW}^{-1} \\ E_P = 0,30 \text{ mmol.g}_{CDW}^{-1} \\ E_B = 0 \text{ mmol.g}_{CDW}^{-1} \\ E_R = 0 \text{ mmol.g}_{CDW}^{-1} \end{cases}$$

En effet, comme il n'y a pas de croissance, il n'y a pas besoin de $\mathcal{E}_R$ ni de $\mathcal{E}_B$ et les deux protéines $\mathcal{E}_P$ et $\mathcal{E}_T$ sont nécessaires pour transformer $\mathcal{G}$ en $\mathcal{S}$, puis $\mathcal{S}$ en $\mathcal{P}$.

Cette simulation nous révèle la quantité maximale de $\mathcal{P}$ produite dans le cas d'une conversion parfaite de $\mathcal{G}$ en $\mathcal{P}$. On trouve en l'occurrence le rapport des masses molaires $M_{\mathcal{G}}/M_{\mathcal{P}}$: on est bien dans le cas d'une conversion parfaite de masse entre $\mathcal{G}$ et $\mathcal{P}$. Cette borne maximale sur la production de $\mathcal{P}$ pourra être retenue comme référence :

$$P_r = \frac{M_{\mathcal{G}}}{M_{\mathcal{P}}} G(t_0) \quad (2.3.18)$$

A noter que l'on néglige ici la masse de $\mathcal{S}$ et des autres éléments qui pourraient être convertis en $\mathcal{P}$.

2.3.5.2 Phase de transition

La période de transition entre les deux phases exclusives de croissance et de production mérite une attention particulière. Elle est le seul moyen dont le système dispose pour faire évoluer sa configuration interne. Il est en effet connu que les protéines sont des structures stables et qu'en phase de croissance, leur dégradation active est négligeable devant la dilution (constante de temps de l'ordre de plusieurs heures [2]). Aussi, en phase de croissance, il est possible de négliger leur dégradation, comme nous l'avons fait dans (2.1.10). Le seul moyen de diminuer la place que prend une protéine dans la cellule est de la diluer. Ce phénomène de dilution n'est pas instantané mais a un temps caractéristique qui vaut $1/\mu$. De plus, la densité intracellulaire étant constante, pour augmenter la concentration d'une protéine, il est nécessaire de faire de la place en réduisant la concentration d'une autre protéine *i.e.* de diluer cette autre protéine. Néanmoins, la production de protéine ne peut se faire instantanément car le taux de production d'une protéine est majoré par la concentration en ribosome et l'efficacité de ces derniers.

La phase de transition, (ii), est le reflet de l'inertie du système biologique et en particulier de la rémanence des protéines dans la cellule.

Par ailleurs, on peut noter que cette phase se décompose en une succession de plusieurs sous-phases :

- Tout d'abord, la concentration intracellulaire en protéine $\mathcal{E}_T$ augmente. La production de ribosomes et de protéines $\mathcal{E}_P$ est nulle ;
- Ensuite, la production de $\mathcal{E}_P$ commence, en parallèle de la production de $\mathcal{E}_T$. En effet, les deux processus Σ_P et Σ_T doivent être actifs pour transformer $\mathcal{G}$ en $\mathcal{P}$. Peu à peu, $\mathcal{E}_P$ prend la place de $\mathcal{E}_B$ et $\mathcal{E}_R$ dans la répartition des protéines ;
- Enfin, la production de $\mathcal{E}_T$ tombe à zéro, permettant à $\mathcal{E}_P$ de gagner encore en concentration.

Le rôle du processus Σ_T est double. Il sert à produire $\mathcal{S}$ comme intermédiaire entre $\mathcal{G}$ et $\mathcal{P}$, mais aussi pour augmenter l'efficacité de tous les processus.

2.3.5.3 Inertie des protéines

Pour se convaincre du rôle essentiel des protéines, on retire la contrainte sur l'inertie de leur concentration. Toute chose restant inchangée par ailleurs, on autorise les flux de production de protéines, ν_{E_i} , à être négatifs. Ainsi, on autorise les protéines à disparaître.

Montrons que dans ces conditions, il devient également possible de substituer une protéine à une autre instantanément.

On suppose qu'à un instant t_s on a une répartition (E_1, E_2) entre deux protéines. La question est de savoir si l'on peut passer à la configuration $(E_1 + \delta, E_2 - \delta)$ à $t_s + \varepsilon$ avec $\varepsilon > 0$.

$$\text{On a } \dot{E}_i = \nu_{E_i} - \mu E_i, \text{ d'où, } \begin{cases} E_1(t_s + \varepsilon) = E_1(t_s) + \int_{t_s}^{t_s + \varepsilon} (\nu_{E_1}(t) - \mu(t)E_1(t))dt \\ E_2(t_s + \varepsilon) = E_2(t_s) + \int_{t_s}^{t_s + \varepsilon} (\nu_{E_2}(t) - \mu(t)E_2(t))dt \end{cases} .$$

$$\text{Peut-on avoir : } \begin{cases} \delta = \int_{t_s}^{t_s + \varepsilon} (\nu_{E_1}(t) - \mu(t)E_1(t))dt \\ -\delta = \int_{t_s}^{t_s + \varepsilon} (\nu_{E_2}(t) - \mu(t)E_2(t))dt \end{cases} ?$$

Par continuité, pour ε assez proche de 0, l'intégrande reste constante :

$$\begin{cases} +\delta = \varepsilon(\nu_{E_1}(t_s) - \mu(t_s)E_1(t_s)) \\ -\delta = \varepsilon(\nu_{E_2}(t_s) - \mu(t_s)E_2(t_s)) \end{cases}$$

D'où,

$$\begin{cases} \nu_{E_1}(t_s) = +\delta/\varepsilon + \mu(t_s)E_1(t_s) \\ \nu_{E_2}(t_s) = -\delta/\varepsilon + \mu(t_s)E_2(t_s) \end{cases} \quad (2.3.19)$$

On retrouve qu'en imposant $\nu_2 \geq 0$, il est nécessaire d'avoir $\mu > 0$ pour faire diminuer E_2 . Mais, si l'on retire la contrainte $\nu_2 \geq 0$ et que l'on autorise ν_2 à être négatif, on peut prendre les formules (2.3.19) quels que soient μ , δ et ε , y compris pour un μ nul.

Par ailleurs, comme on a ajouté et retiré la même quantité de protéine, la substitution n'a pas d'influence sur la densité intracellulaire et ce qui en provient.

On vient de montrer que l'on pouvait substituer une entité à une autre en relaxant la contrainte sur la positivité des flux ν_{E_i} .

On trace figure 2.11 les mêmes allures qu'à la figure 2.8 mais pour la solution au problème d'optimisation (2.3.16) dans lequel on a retiré la contrainte de positivité des flux ν_{E_i} . La modélisation est alors dénommée dRBAu (pour dRBA *unconstrained*).

Sur ces courbes, on trouve la stratégie classique en deux temps : croissance puis production. On note que la configuration de la cellule pendant la phase de croissance est rigoureusement identique dans les deux cas, en accord avec les figures 2.3 et 2.6. L'instant de bascule, t_s , entre les deux stratégies est $t_s = 3,40$ h et le temps de culture total est de $t_f = 7,66$ h.

Ces résultats sont plus proches de ce qui ressort des modèles classiques, que ce soit des modèles macroscopiques ou des modèles fondés sur les FBA. Les non linéarités sur la concentration en $\mathcal{S}$ introduisent cependant de légères modifications. Pour se rapprocher encore davantage des FBA, on pourrait retirer l'influence de la concentration de $\mathcal{S}$ sur les processus en prenant les efficacités k_i indépendantes de la concentration S .

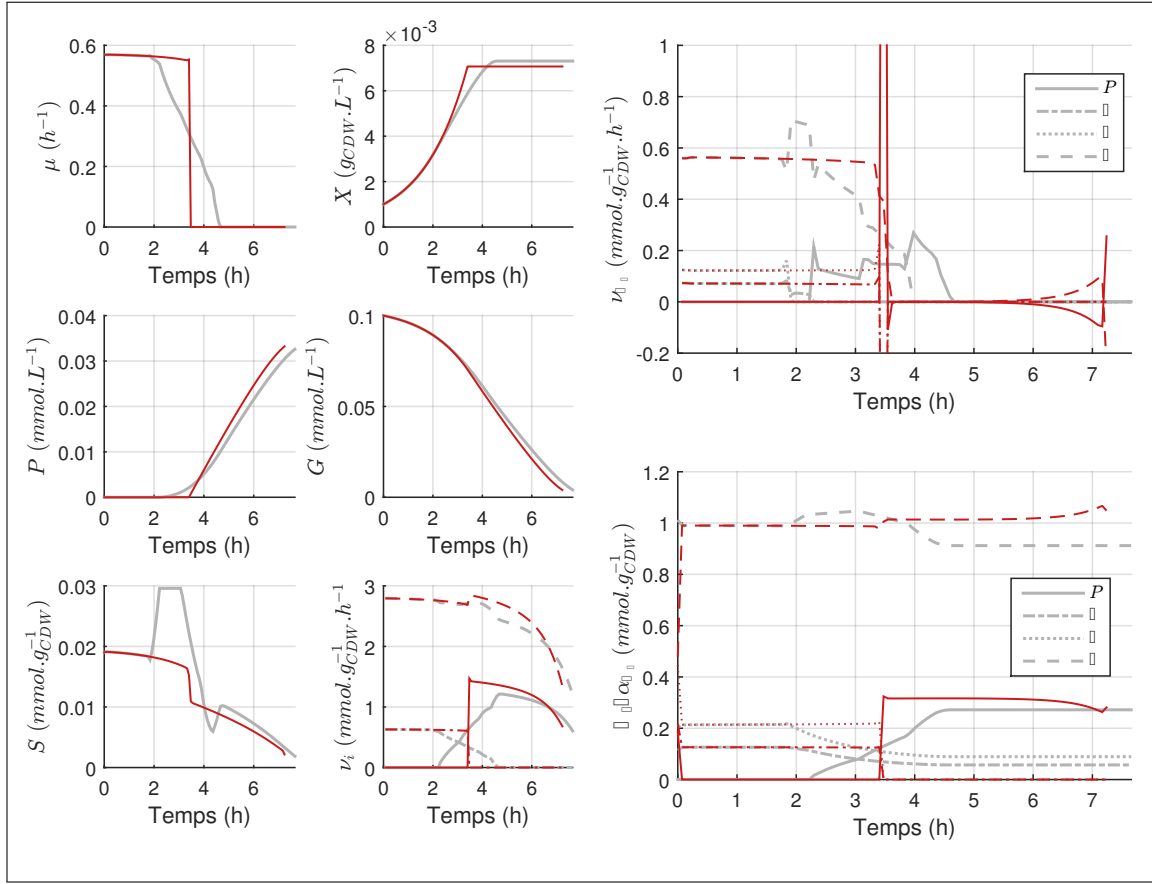


FIGURE 2.11 – En rouge, solution du problème (2.3.16) où les flux ν_{E_i} peuvent être négatifs (dRBAu). En gris clair, rappel de la figure 2.8, solution du problème dRBA.

2.3.5.4 Discussion sur les profils de production du composé d'intérêt

D'après la figure 2.11, les trajectoires $P(t)$ sont très proches dans les cas où on prend en compte les contraintes d'évolutions des machineries moléculaires (dRBA), et dans le cas où on ne les prend pas en compte (dRBAu). Malgré cette observation macroscopique, on peut voir que l'origine de ces deux profils est bien différents et donneraient lieu à des conduites du procédé différentes.

En effet, ces expériences visent à trouver le fonctionnement optimal d'un procédé. En particulier, on cherche à identifier le compromis entre croissance et production. Doivent-elles être successives ou concomitantes ? La réponse en utilisant des modèles macroscopiques et des modèles reposant sur les FBA montrent qu'il est préférable de croître puis de produire. Il est rassurant qu'en décrivant les processus intracellulaires, on conserve cette même idée d'une croissance suivie d'une production.

Par ailleurs, la conduite du procédé est déterminée par les actions à effectuer sur le procédé. C'est-à-dire qu'au delà du profil $P(t)$, il est primordial de connaître l'instant optimal auquel la production de $\mathcal{P}$ doit commencer. Montrons qu'il est important de prendre en compte les considérations de la section 2.1 pour la conduite du procédé en tant que telle.

Ce qui est déterminant est l'instant auquel la transition doit s'opérer. Certes, les profils $P(t)$ sont très proches mais les instants auxquels la production de $\mathcal{P}$ commence sont différents. Dans le cas dRBA, la production commence en $t_s = 2,2$ h, dans le cas dRBAu, elle commence en $t_s = 3,4$ h. En ayant sous-évalué la période de transition, on se retrouve avec un résultat différent.

La figure 2.12 présente une expérience durant la production commence à $t_s = 3,4$ h, *i.e.* l'instant déterminé sans prendre en compte la phase de transition, mais où l'on prend en compte l'évolution des concentrations des machineries qui catalysent les processus intracellulaires. Cette expérience est appelée expérience D. Elle permet d'identifier la perte d'optimalité due à une mauvaise identification de l'instant t_s .

TABLE 2.4 – Valeurs optimales obtenues par les diverses méthodes.

	t_f (en h)	$P(t_f)$ (en $\mu\text{mol.L}^{-1}$)	$X(t_f)$ (en $\text{mg}_{CDW}.\text{L}^{-1}$)	t_s (en h)	$\frac{P(t_f)}{t_f}$ (en $\mu\text{mol.L}^{-1}.\text{h}^{-1}$)
Référence absolue	-	50	1,00	-	0
Méthode dRBA	7,66	32,6	7,31	2,22	4,26
Méthode dRBAu	7,24	33,2	7,07	3,40	4,59
L'expérience D	7,30	29,5	8,47	3,40	4,04

Les valeurs numériques sont données table 2.4. La référence absolue est la solution du problème où le critère est $\max P(t_f)$ (résultats présentés figure 2.10). Le paramètre t_s est, le cas échéant, l'instant auquel commence la production de $\mathcal{P}$. On note un écart de 10% entre la concentration de $\mathcal{P}$ prédite par dRBAu et l'expérience D. Par ailleurs, par conservation de la matière, on note que l'expérience D conduit à produire plus de biomasse que l'optimum donné par la méthode dRBA. Bien que le temps de culture soit plus court pour l'expérience D que pour l'optimum de la méthode dRBA, le critère de performance $P(t_f)/t_f$ est meilleur avec la méthode dRBA.

La différence entre les deux tracés de la figure 2.12 pourrait expliquer certaines différences entre les expériences pratiques et les prédictions des différents modèles qui ne tiennent pas compte des processus intracellulaires. Ceci fera l'étude du chapitre 4 où la question de la production d'acétate sera spécifiquement abordée.

Le modèle développé dans ce chapitre prend tout son sens dans l'optimisation du bioprocédé dans son ensemble. La prise en compte des contraintes sur les processus intracellulaires introduit des contraintes sur le fonctionnement du bioprocédé dans son ensemble.

2.4 Conclusion du chapitre

Dans ce chapitre, on a développé un modèle agrégé du comportement de cellules bactériennes. Ce modèle a été couplé à un modèle d'évolution d'un bioréacteur opérant en fonctionnement batch.

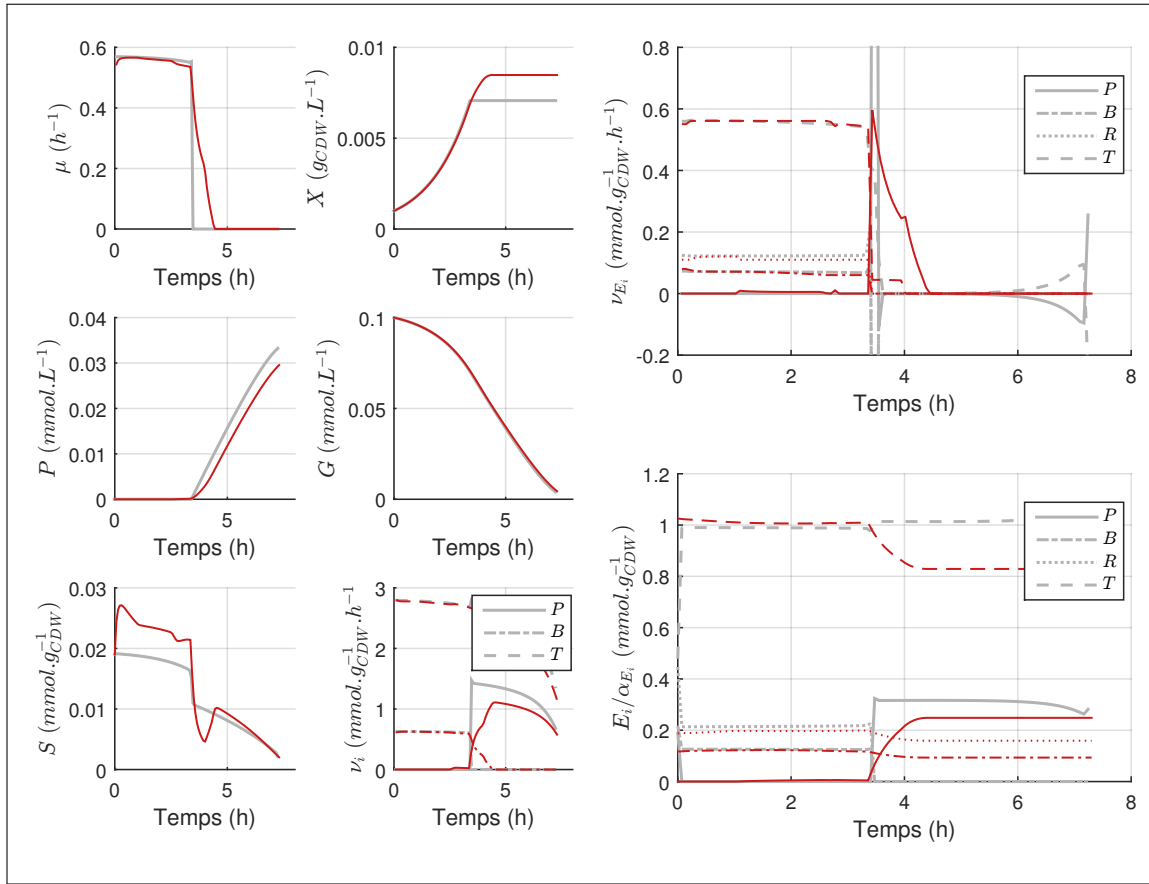


FIGURE 2.12 – Évolutions des variables du système dans le cas où la dynamique des protéines est négligée et où l'on peut substituer une protéine à une autre instantanément (dRBAu) en gris, voir figure 2.11, et en bordeaux, dans le cas réaliste où les protéines sont contraintes, voir figure 2.8, mais où l'on force la production de $\mathcal{P}$ à ne pas démarrer avant $t_s = 3,4h$ (Expérience D, voir texte).

Des tests ont ensuite été effectués pour le rapprocher à la fois de modèles macroscopiques existants et du comportement de modèles intracellulaires détaillés validés expérimentalement. Nous avons pu constater que le taux spécifique de croissance de notre modèle suivait une relation de Monod. Ainsi, notre approche complète bien les modèles macroscopiques usuels et n'est pas en désaccord avec ce qui a été observé depuis le milieu du XX^{ème} siècle sur le plan macroscopique. Par ailleurs, on a montré que, malgré l'agrégation, la composition intracellulaire suivait les mêmes tendances et ordres de grandeur que ce qui a pu être observé biologiquement à l'échelle intracellulaire. On peut donc dire que les tests effectués ont été suffisamment concluants pour pouvoir considérer qu'il est proche de la réalité tant du point de vue du comportement macroscopique que de celui intracellulaire. Par construction, ce modèle sera appelé dRBA.

Dès lors, il devient possible de l'utiliser pour optimiser le procédé dans son ensemble. Les variables de contrôle du bioprocédé sont les flux des processus intracellulaires y compris ceux de l'expression des gènes. Le critère d'optimisation du bioprocédé est choisi pour être la production de composé d'intérêt par unité de temps. Ce critère est à la fois judicieux du point de vue de la rentabilité industrielle mais permet également d'exercer

une pression sur les cellules en les forçant à aller le plus vite possible. La caractérisation de l'optimum n'est pas triviale, nécessitant le recours à une discrétisation du problème qui nous conduit à une solution numérique.

Cette solution présente un grand intérêt à plusieurs égards. Tout d'abord, on reconnaît la politique traditionnelle de croissance puis de production qui est courante dans la conduite de procédé, mais on constate également l'apparition d'une phase de transition entre la croissance et la production. Cette phase de transition est due au fait que la production tout comme la croissance sont des processus intracellulaires catalysés par des protéines et qu'il y a besoin de changer la répartition de ces protéines pour passer d'un processus à l'autre. Ce changement de répartition a besoin de temps et ne peut être fait instantanément, comme supposé indirectement par les modèles de la littérature qui ne font pas intervenir les protéines.

Se pose désormais la question de l'implémentation biologiquement réalisable de cette trajectoire optimale. Comment peut-on contrôler en temps réel le flux qui passe à travers les processus intracellulaires ? Cet aspect fait l'objet du prochain chapitre.

Chapitre 3

Principe de conception rationnelle de souche

Nous avons étudié dans le chapitre précédent un modèle agrégé du fonctionnement de la cellule qui a permis de poser un problème d'optimisation de bioprocédé en agissant sur les processus intracellulaires. Nous avons pu trouver les profils optimaux des espèces intracellulaires et des flux des différents processus intracellulaires. Les résultats se sont révélés en accord avec les considérations usuelles de la littérature : une phase de croissance exponentielle, une phase de transition et une phase de production pure.

Toutefois, cette étude a été réalisée en laissant libre les processus intracellulaires, ce qui laisse supposer qu'il est possible de contrôler indépendamment chaque processus intracellulaire en temps réel. D'un point de vue de la conduite du procédé, ce n'est pas réaliste. Nous allons présenter dans ce chapitre une manière de formaliser ces contraintes et optimiser la conception de souche pour une conduite du procédé réaliste, c'est-à-dire qui demande un nombre limité d'actions lors de la conduite du procédé.

Comme nous allons le voir, certains processus intracellulaires sont régulés de manière autonome par la cellule. En particulier, la régulation de la synthèse des protéines (*i.e.* l'expression des gènes) peut être programmée de manière nominale par le taux spécifique de croissance. Dans le but d'intervenir le moins possible sur le procédé, nous allons forcer le comportement cellulaire optimal à être réaliste au niveau de l'implémentation dans la cellule. L'objectif de ce chapitre est de comprendre la perte d'optimalité qu'entraîne des contraintes d'implémentation réalisables biologiquement.

En section 3.1, nous présentons un ensemble de contraintes lié à l'implémentation réaliste de gènes dans la cellule. Cet ensemble de contrainte est ajouté à la formulation du problème d'optimisation de la conception de bioprocédé en mode de fonctionnement batch à la section 3.2. Ensuite, en section 3.3, nous proposons une formulation du même problème en mode de fonctionnement fed-batch. Cette formulation nous permet d'envisager l'optimisation de la conduite du procédé avec des cellules rationnellement modifiées et également l'optimisation conjointe de la conduite du procédé et de la conception rationnelle de souche.

3.1 Vers une conception rationnelle de souche

3.1.1 Éléments de mise en œuvre réaliste d'une stratégie de commande de la cellule

Nous cherchons ici un moyen de mettre en œuvre la stratégie optimale que nous avons présentée au chapitre précédent. Pour cela, on a notamment supposé pouvoir contrôler en temps réel l'expression de tous les gènes.

D'un point de vue pratique, contrôler en temps réel l'expression d'un gène peut se faire par l'intermédiaire de facteurs de transcription, via leur effecteur. En introduisant un effecteur dans le milieu, celui-ci va se lier à un facteur de transcription particulier présent dans la cellule. Le complexe effecteur-facteur de transcription va réagir avec le promoteur du gène à transcrire et moduler la transcription du gène, soit en l'inhibant, soit en la stimulant activement. Le nombre des facteurs de transcription connus à ce jour reste limité [63]. Dès lors, il ne va être possible d'agir que sur un nombre très limité de gènes.

Comme présenté dans [31], le réseau de régulation d'un organisme comme *Bacillus subtilis* est très structuré. Outre les facteurs sigma qui permettent de changer complètement de programme (voir annexe A), les quelques milliers de gènes se partagent finalement un nombre restreint de régulations. En agissant directement sur ces régulations, il est possible de choisir le programme que l'on souhaite que la bactérie effectue.

Comme indiqué en annexe A, un des mécanismes principaux qui permet à la cellule de moduler l'expression de ses gènes est lié au taux spécifique de croissance. Nous pouvons alors envisager d'adopter une commande de l'expression des gènes qui dépende du taux spécifique de croissance. Ceci rend la commande du procédé réalisable. Nous appellerons ces contraintes, des contraintes d'implémentation.

Concernant les processus métaboliques (Σ_T , Σ_B et Σ_P dans le cadre du modèle agrégé), dans la plupart des cas, dès lors qu'une protéine est au contact de son substrat, le processus qu'elle catalyse se produit. Des exceptions peuvent avoir lieu, en particulier via la régulation des processus par le produit du processus. Typiquement, un processus de production d'un composé intracellulaire sera inactivé dès lors que la concentration de ce composé atteint un seuil.

3.1.2 Relation entre ν_{E_i} et μ dans le cas d'un gène constitutif

L'objectif de cette partie est de donner une justification du lien entre les flux de production de protéines et le taux spécifique de croissance. Nous donnons des éléments biologiques qui permettent de justifier ces liens pour les gènes constitutifs, c'est-à-dire qui ne sont pas régulés.

L'annexe A présente les étapes principales de la synthèse des protéines. On peut rappeler le fait qu'un gène est transcrit par les ARN-Polymérases en un ARN messager qui est ensuite traduit par les ribosomes en protéine.

Considérons un gène constitutif. On note g_i son nombre de copies présentes dans la cellule, $m_i(t)$ la concentration des ARN messagers issus de la transcription de ce gène à

l'instant t , $E_i(t)$ la concentration des protéines issus de la traduction des messagers de ce même gène. La dynamique de $m_i(t)$ et $E_i(t)$ s'écrit :

$$\begin{cases} \dot{m}_i(t) = \nu_{m_i}(t) - \delta m_i(t) \\ \dot{E}_i(t) = \nu_{E_i}(t) - \mu(t)E_i(t) \end{cases} \quad (3.1.1)$$

où $\nu_{m_i}(t)$ est le flux de transcription du gène, ν_{E_i} est le flux de traduction des protéines, δ est le taux de dégradation de l'ARN messager, $\mu(t)$ le taux spécifique de croissance.

Différents travaux comme ceux récents de Gerosa et al. [29] pour la transcription et ceux de Borkowski et al. [11] [12] pour la traduction, permettent de déduire les expressions suivantes des flux ν_{m_i} et ν_{E_i} :

$$\begin{cases} \nu_{m_i}(t) = \alpha_{m,i} \frac{[Pol_f]}{[Pol_f] + K_{m,i}} g_i \\ \nu_{E_i}(t) = \alpha_{E,i} \frac{[R_f]}{[R_f] + K_{E,i}} m_i(t) \end{cases} \quad (3.1.2)$$

où $[Pol_f]$ est la concentration en ARN-Polymérase libres, $[R_f]$ est la concentration en ribosomes libres, $\alpha_{m,i}$, $K_{m,i}$, $\alpha_{E,i}$, $K_{E,i}$ sont des paramètres positifs.

En supposant que m_i est à l'équilibre, on peut écrire :

$$m_i = \frac{1}{\delta} \nu_{m_i} \quad (3.1.3)$$

Puis, en injectant cette relation dans (3.1.2), il vient :

$$\nu_{E_i} = \alpha_{E,i} \frac{[R_f]}{[R_f] + K_{E,i}} \frac{1}{\delta} \nu_{m_i} \quad (3.1.4)$$

ou encore, d'après (3.1.2),

$$\nu_{E_i} = \alpha_{E,i} \alpha_{m,i} g_i \frac{1}{\delta} \frac{[R_f]}{[R_f] + K_{E,i}} \frac{[Pol_f]}{[Pol_f] + K_{m,i}} \quad (3.1.5)$$

Par ailleurs, d'après les travaux de Bremer et Dennis [14], la concentration $[Pol_f]$ est croissante avec le taux spécifique de croissance. De façon parallèle, les travaux de Borkowski [11] montrent que la concentration $[R_f]$ est décroissante en fonction du taux spécifique de croissance. De sorte que le flux ν_{E_i} est le produit entre :

- une fonction croissante par rapport au taux spécifique de croissance : $\frac{[Pol_f]}{[Pol_f] + K_{m,i}}$
- une fonction décroissante par rapport au taux spécifique de croissance : $\frac{[R_f]}{[R_f] + K_{E,i}}$
- un terme de modulation : $\alpha_{E,i} \alpha_{m,i} g_i \frac{1}{\delta}$

En fonction des paramètres $K_{E,i}$ et $K_{m,i}$, on peut avoir des allures de ν_{E_i} en fonction du taux spécifique de croissance qui sont soit croissantes, soit décroissantes, soit stationnaires, soit paraboliques (croissants puis décroissants, ou inversement).

De cette caractérisation, on conclut que le flux de production de protéine constitutives peut être localement approximé par un polynôme en fonction du taux spécifique de croissance.

3.1.3 Choix de la paramétrisation pour le modèle agrégé

Nous revenons au modèle agrégé développé au chapitre précédent et souhaitons caractériser à présent la perte d'optimalité à forcer des gènes à être constitutifs. Pour cela, il est possible d'ajouter au modèle les notions de messagers, de ribosomes libres, d'ARN-polymérases, *etc.* comme on vient de le faire.

Néanmoins, pour rester dans un contexte général et simplifié, on va simplement ajouter des contraintes entre les flux de production de protéines et le taux spécifique de croissance. La question est de choisir une paramétrisation adaptée à la simplicité du modèle agrégé.

Étant donnés les profils des concentrations en protéines en fonction du taux spécifique de croissance donnés par la méthode RBA, *c.f.* figure 2.6 du chapitre précédent, on va approcher les concentrations en protéines dans la cellule à l'état stationnaire par des polynômes du deuxième ordre par rapport au taux spécifique de croissance :

$$\bar{E}_i(\mu)/\alpha_{E_i} = a_i\mu^2 + b_i\mu + c_i \quad (3.1.6)$$

En passant la relation (2.1.10) (*i.e.* $\dot{E}_i(t) = \alpha_{E_i}\nu_{E_i}(t) - \mu(t)E_i(t)$) à l'équilibre, on obtient :

$$\alpha_{E_i}\bar{\nu}_{E_i} = \mu\bar{E}_i \quad (3.1.7)$$

Ainsi, en reportant l'approximation polynomiale (3.1.6), on aboutit finalement à paramétrer les $\nu_{E_i}(t)$ à tout instant de la manière suivante :

$$\nu_{E_i}(t) = \nu_{E_i}(\mu(t)) = a_i\mu(t)^3 + b_i\mu(t)^2 + c_i\mu(t) \quad (3.1.8)$$

3.1.4 Choix des gènes constitutifs dans le cadre du modèle agrégé

Il importe de déterminer quels gènes peuvent être considérés comme constitutifs dans la modélisation agrégée.

La relation (3.1.8) établit une contrainte réaliste sur le flux de production des gènes constitutifs. Dans le cas de notre modèle agrégé, on supposera que la protéine $\mathcal{E}_B$ est une protéine constitutive. En tant que protéine d'un processus d'import et de dégradation, on peut considérer que la protéine $\mathcal{E}_T$ est contrôlée par la présence ou l'absence de substrat $\mathcal{G}$ dans le milieu. Comme notre étude est restreinte au cas où il n'y a que ce substrat $\mathcal{G}$ dans le milieu, on considérera que $\mathcal{E}_T$ est constitutive. Concernant les ribosomes, $\mathcal{E}_R$, il semble difficile de négliger l'effet de la protéine relA (présenté en annexe A) sur la régulation de la production en machineries de traduction. On ne peut donc pas considérer que $\mathcal{E}_R$ soit constitutive.

Pour ce qui est de la protéine $\mathcal{E}_P$, nous allons supposer que son expression est donnée par le taux de croissance (*i.e.* constitutive). Ainsi, il y aura simplement à concevoir un gène et à l'introduire dans la cellule pour que la cellule produise $\mathcal{E}_P$. Pour étudier l'impact d'actions supplémentaires, on pourra éventuellement envisager des techniques de régulations supplémentaires pour $\mathcal{E}_P$, comme nous le verrons dans la suite de ce chapitre.

3.2 Problèmes d'optimisation de la conception de bioprocédé batch

On se place en mode de fonctionnement batch. On suppose donc avoir peu de moyen d'action sur le bioprocédé. L'objectif est alors de concevoir un procédé qui nécessite le moins d'intervention possible. On va passer en revue plusieurs cas d'opération et on va évaluer la perte de performance liée à l'autonomie de chacun. L'idée est de trouver un compromis acceptable entre autonomie du procédé (le moins d'interventions extérieures) et sa performance. Pour ce faire, on va repartir du problème (2.3.7) en ajoutant des contraintes du type (3.1.8). Les cas vont différer les uns des autres en fonction de la manière dont on va contraindre les polynômes (3.1.8). Leurs coefficients des polynômes des gènes constitutifs vont être fixés a priori ou bien optimisés.

On étudie les quatre cas décrits dans la table 3.1. On lit dans la table les interventions nécessaires pour chaque cas.

TABLE 3.1 – Cas d'études présentés dans ce chapitre.

Cas	Interventions nécessaires
n° 1 : Reconfiguration complète d'une souche et introduction d'un gène de synthèse constitutif	Reconfigurer tous les gènes constitutifs + Configurer le gène de synthèse
n° 2 : Introduction d'un gène de synthèse constitutif	Configurer le gène de synthèse
n° 3 : Introduction d'un gène de synthèse constitutif "activable" à un instant donné	Configurer le gène de synthèse + Activer ce gène à un instant t_s
n° 4 : Introduction d'un gène de synthèse constitutif avec action sur la croissance	Configurer le gène de synthèse + Interrompre la croissance à un instant t_s

3.2.1 Contraintes complémentaires pour l'autonomie de la cellule

En plus des contraintes de type (3.1.8) à ajouter pour chaque protéine constitutive, il est nécessaire d'adapter le problème d'optimisation. On se place dans le contexte d'optimisation de la bioproduction d'un composé d'intérêt, comme au chapitre précédent.

3.2.1.1 Conditions initiales sur les concentrations en $\mathcal{E}_i$

Concernant les concentrations initiales en protéines, on a dit qu'il fallait uniquement contraindre la densité initiale à être égale à D_0 . A présent que les concentrations en protéines constitutives sont liées à la valeur du taux spécifique de croissance, il est nécessaire de contraindre les concentrations initiales en protéines constitutives comme suit :

$$E_i(t_0)/\alpha_{E_i} = a_i\mu(t_0)^2 + b_i\mu(t_0) + c_i \quad (3.2.1)$$

où $\mu(t_0)$ est la valeur du taux spécifique de croissance à l'instant initial. On prend ainsi en compte le fait que les protéines constitutives sont à l'équilibre.

3.2.1.2 Écriture compacte du taux spécifique de croissance

Dans l'étude du chapitre précédent, on a volontairement retiré le taux spécifique de croissance de la formulation du problème en le substituant par son expression en fonction des flux de production de protéines, $\mu(t) = \frac{1}{D_0} \sum \nu_{E_i}(t)$. Dans les expressions (3.1.8) et (3.2.1), on substitue μ par son expression en fonction des flux. Ces équations s'écrivent en fonction des éléments de x et ν :

$$\begin{cases} \nu_{E_i}(\nu(t)) = a_i(\frac{1}{D_0} \sum \nu_{E_i}(t))^3 + b_i(\frac{1}{D_0} \sum \nu_{E_i}(t))^2 + c_i(\frac{1}{D_0} \sum \nu_{E_i}(t)) \\ E_i(t_0)/\alpha_{E_i} = a_i(\frac{1}{D_0} \sum \nu_{E_i}(t_0))^2 + b_i(\frac{1}{D_0} \sum \nu_{E_i}(t_0)) + c_i \end{cases} \quad (3.2.2)$$

Pour simplifier et compacter les expressions, on utilisera la notation $\mu(t)$ en lieu et place de $\frac{1}{D_0} \sum_i \nu_{E_i}(t)$.

3.2.1.3 Contrainte sur les flux ν_T , ν_P et ν_R

Dans la résolution précédente, les flux ν_T , ν_P et ν_R étaient libres, sous réserve de respecter la contrainte d'efficacité des processus Σ_T , Σ_P et Σ_R . Il est important de noter que la résolution du problème d'optimisation a conduit à ne pas toujours saturer ces contraintes. On a vu en particulier que le flux de production des protéines n'était pas saturé dans la phase de production du composé d'intérêt. Techniquement, cela signifie qu'il faudrait avoir un moyen d'action pour limiter l'efficacité des processus.

Comme on souhaite ne pas agir sur les micro-organismes pendant la durée du batch, on va contraindre les flux à être égaux à leur borne. En particulier :

$$\begin{cases} \nu_T(t) = k_T(G(t), S(t))E_T(t) \\ \nu_P(t) = k_P(S(t))E_P(t) \\ \sum_{P,B,R,T} \nu_{E_i}(t) = k_R(S(t))E_R(t) \end{cases} \quad (3.2.3)$$

Pour ce qui est de Σ_B , on peut raisonnablement supposer que le flux est contrôlé par la concentration en macrocomposants $\mathcal{B}$. On conservera la contrainte

$$\nu_B(t) \leq k_B(S(t))E_B(t) \quad (3.2.4)$$

On synthétise ces contraintes par la notation matricielle suivante :

$$\begin{cases} A_0\nu(t) = b_0(x(t)), \text{ pour la relation (3.2.3)} \\ A_1\nu(t) \leq b_1(x(t)), \text{ pour la relation (3.2.4)} \end{cases} \quad (3.2.5)$$

Ainsi, ν_P et ν_T ne sont plus des commandes du système mais des variables d'état auxiliaires. On pourrait les retirer de la description du modèle en les remplaçant par leur relation (3.2.3). Pour que le lien avec le chapitre précédent soit clair et éviter de spécifier un nouveau système, nous continuerons à les considérer comme des variables de commande.

3.2.2 Reconfiguration complète d'une souche (Cas n° 1)

Dans cette section, on s'intéresse au design réaliste complet d'une souche, c'est à dire que l'on suppose que tous les flux de production de protéine suivent la relation (3.1.8) et on cherche les coefficients a_i , b_i , c_i pour $\mathcal{E}_B$, $\mathcal{E}_T$ et $\mathcal{E}_P$. Avec les notations introduites précédemment (voir chapitre 2), le problème à résoudre prend la forme :

$$\begin{aligned}
 & \max_{\substack{\nu_B(t), \nu_{E_R}(t), t_f \\ a_B, b_B, c_B \\ a_T, b_T, c_T \\ a_P, b_P, c_P}} \left(0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \frac{1}{t_f} \ 0 \right) x(t_f) \\
 & \text{s.c.} \left\{ \begin{array}{ll} \dot{x}(t) = F(x(t))\nu(t) & (3.2.6.a) \\ A_0\nu(t) = b_0(x(t)) & (3.2.6.b_0) \\ A_1\nu(t) \leq b_1(x(t)) & (3.2.6.b_1) \\ L\nu(t) = 0 & (3.2.6.c) \\ \nu(t) \geq 0 & (3.2.6.d) \\ C_0x(t_0) = c_0 & (3.2.6.e) \\ C_fx(t_f) \geq c_f & (3.2.6.f) \\ t_{min} \leq t_f \leq t_{max} & (3.2.6.g) \\ \nu_{E_B}(t) - (a_B\mu(t)^3 + b_B\mu(t)^2 + c_B\mu(t)) = 0 & (3.2.6.h_B) \\ \nu_{E_T}(t) - (a_T\mu(t)^3 + b_T\mu(t)^2 + c_T\mu(t)) = 0 & (3.2.6.h_T) \\ \nu_{E_P}(t) - (a_P\mu(t)^3 + b_P\mu(t)^2 + c_P\mu(t)) = 0 & (3.2.6.h_P) \\ E_B(t_0)/\alpha_{E_B} - (a_B\mu(t_0)^2 + b_B\mu(t_0) + c_B) = 0 & (3.2.6.i_B) \\ E_T(t_0)/\alpha_{E_T} - (a_T\mu(t_0)^2 + b_T\mu(t_0) + c_T) = 0 & (3.2.6.i_T) \\ E_P(t_0)/\alpha_{E_P} - (a_P\mu(t_0)^2 + b_P\mu(t_0) + c_P) = 0 & (3.2.6.i_P) \end{array} \right. \quad (3.2.6)
 \end{aligned}$$

Dans cette configuration, le flux de production de $\mathcal{E}_R$ est lui aussi contraint d'être polynomial en fonction du taux spécifique de croissance. En effet, dans la contrainte (3.2.6.c) figure la contrainte $\sum_i \nu_{E_i} = D_0\mu$. Au vu des contraintes (3.2.6.h), il vient :

$$\nu_{E_R}(t) = -(a_B + a_P + a_T)\mu(t)^3 - (b_B + b_P + b_T)\mu(t)^2 + (D_0 - c_B - c_P - c_T)\mu(t) \quad (3.2.7)$$

Comme précédemment, ce problème est résolu à l'aide du logiciel Bocop [10], initialisé avec la solution du chapitre précédent et les mêmes données numériques. Les résultats sont présentés figure 3.1 où la solution optimale au problème dRBA du chapitre précédent est rappelé. On se sert de cet optimum comme étalon.

Bien que les variables macroscopiques G et P aient des évolutions similaires dans les deux cas, la stratégie intracellulaire est grandement différente. Le taux spécifique de croissance n'est plus constant en début de simulation et n'est plus nul à la fin. En ce qui concerne les flux, on constate que les allures sont globalement similaires entre les deux cas, à la différence que la situation de reconfiguration de souche donne lieu à des allures plus régulières. C'est également le cas pour les concentrations de protéines qui ont quasiment les mêmes valeurs initiales (taux spécifique de croissance élevé) et qui augmentent et diminuent de la même manière au premier ordre.

En résumé, l'écart entre les deux solutions est caractérisé par l'absence de phases strictes (de croissance ou de production). Avoir rendu la cellule autonome ne lui permet

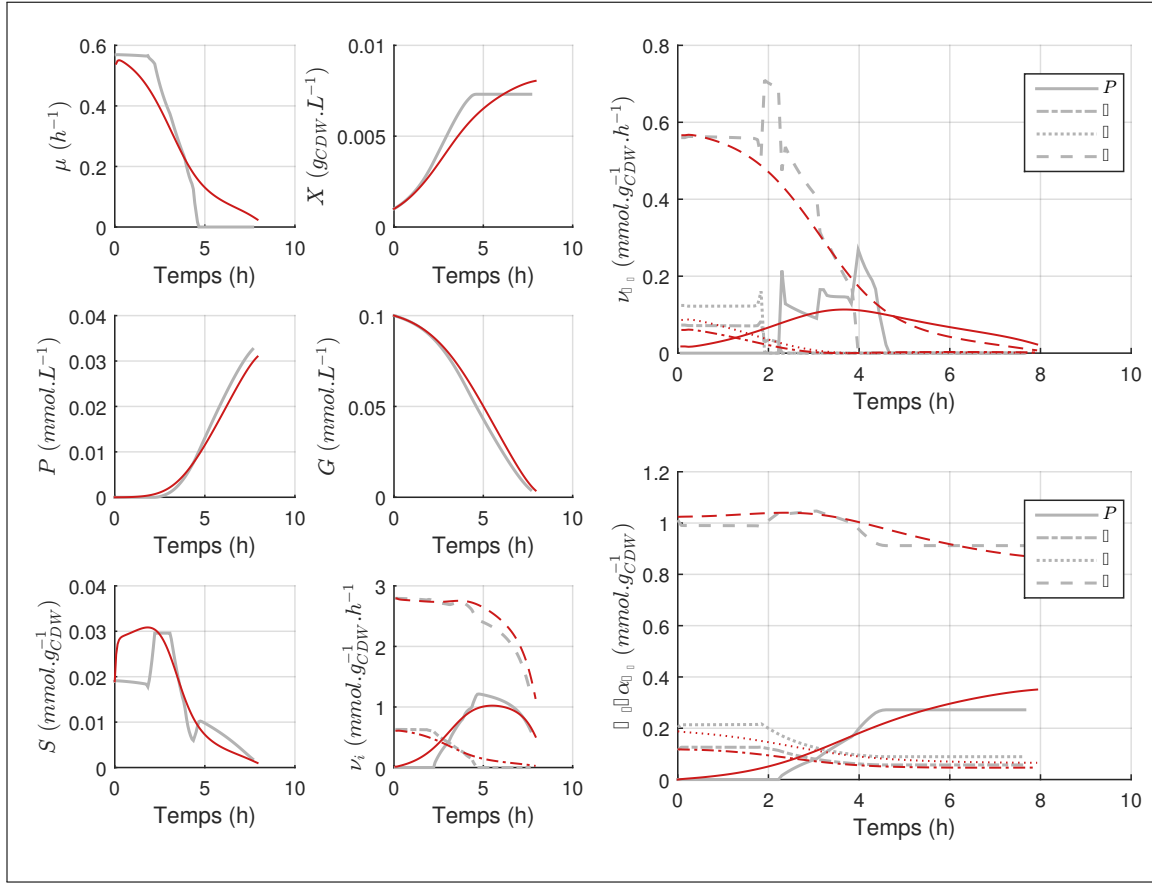


FIGURE 3.1 – Comparaison entre la solution du problème d’optimisation en agissant sur tous les processus (*c.f.* chapitre précédent) (en gris), et la solution au problème de reconfiguration complète de souche (Cas n° 1) (en rouge).

pas d’annuler brusquement ses flux : la cellule est contrainte de produire du produit à tout instant et de croître à tout instant. Ainsi, on ne retrouve plus de phases distinctes de croissance puis de production comme précédemment.

Plus que les valeurs des coefficients a_i , b_i , c_i , il est intéressant de représenter les allures des flux de production de protéines en fonction du taux spécifique de croissance. Ceci est représenté figure 3.2.

On note que le flux de production de la protéine $\mathcal{E}_P$ est décroissant en fonction du taux spécifique de croissance, de sorte qu’à croissance faible on ne produise quasiment que des protéines $\mathcal{E}_P$ et $\mathcal{E}_T$, c’est-à-dire les protéines nécessaires à la transformation de $\mathcal{G}$ en $\mathcal{S}$ puis en $\mathcal{P}$. Le flux de production de $\mathcal{E}_B$ et $\mathcal{E}_R$ sont des paraboles. Le flux de production de $\mathcal{E}_T$ est globalement croissant en μ . Pour des valeurs élevées de taux spécifique de croissance, on retrouve des valeurs de flux sensiblement identiques à celles obtenues pour les ratios $E_i/(\alpha_{E_i} D_0)$ lorsque l’on maximise le taux spécifique de croissance, voir figure 2.6 du chapitre précédent. Cependant, les allures générales des figures 3.2 et 2.6 du chapitre précédent n’ont rien de comparable. En se limitant à n’utiliser que des gènes constitutifs, il faudrait donc reconfigurer l’intégralité des gènes de la cellule. Ce n’est pas

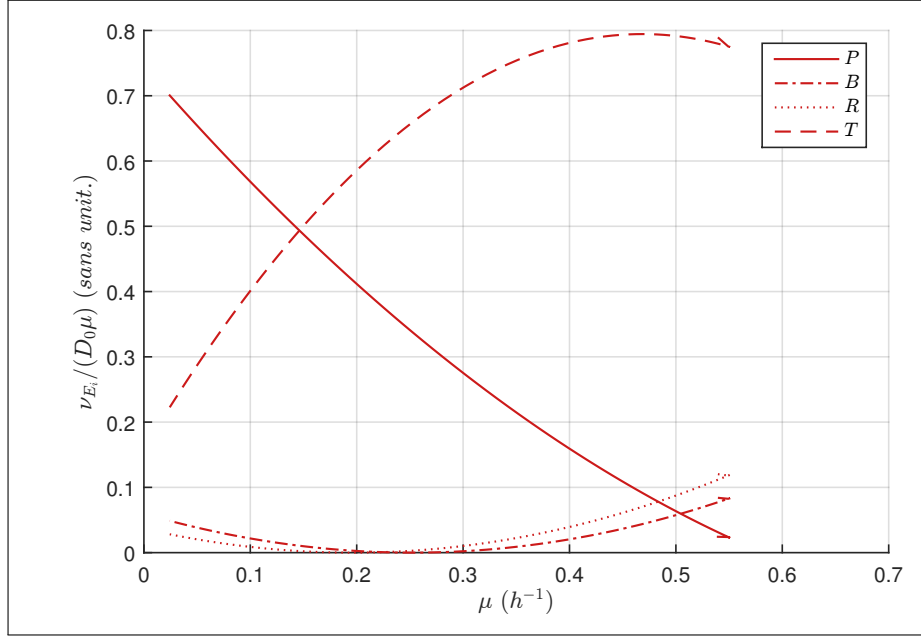


FIGURE 3.2 – Allure des flux $\nu_{E_i}(t)$ tracés en fonction du taux spécifique de croissance $\mu(t)$, pour la solution au problème (3.2.6).

techniquement envisageable à ce jour.

3.2.3 Introduction d'un gène de synthèse constitutif (Cas n° 2)

On va à présent supposer que le comportement d'une partie de la cellule est donné. On suppose que l'on ajoute à cette cellule un gène de synthèse associé à la protéine $\mathcal{E}_P$. Pour se dispenser du contrôle du bioprocédé en temps réel, on suppose que ce gène est constitutif.

On résout donc le problème suivant :

$$\begin{aligned} & \max_{\substack{\nu_B(t), \nu_{E_R}(t), t_f \\ a_P, b_P, c_P}} \left(0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \frac{1}{t_f} \ 0 \right) x(t_f) \\ & \text{sous contraintes du problème (3.2.6) et } \begin{cases} (a_B, b_B, c_B) = (a_B^N, b_B^N, c_B^N) \\ (a_T, b_T, c_T) = (a_T^N, b_T^N, c_T^N) \end{cases} \end{aligned} \quad (3.2.8)$$

avec a_B^N, b_B^N, c_B^N (respectivement, a_T^N, b_T^N, c_T^N) les coefficients de l'expression de $\mathcal{E}_B$ (respectivement, $\mathcal{E}_T$) dans le cas nominal, c'est-à-dire lorsque la cellule maximise sa croissance. Ils sont calculés en section 3.2.3.1.

Par le même argument que précédemment, ceci revient à dire que le flux de production des ribosomes $\nu_{E_R}(t)$ est polynomial par rapport à μ . On ne cherche donc que le profil $\nu_B(t)$ et un nombre fini de paramètres. Outre la simplicité du contrôle du bioprocédé, l'optimum de ce cas de figure est aussi plus simple à formuler.

3.2.3.1 Détermination des coefficients polynomiaux dans le cas nominal

On cherche l'expression des flux en conditions nominales, *i.e.* lorsque les cellules maximisent leur taux spécifique de croissance.

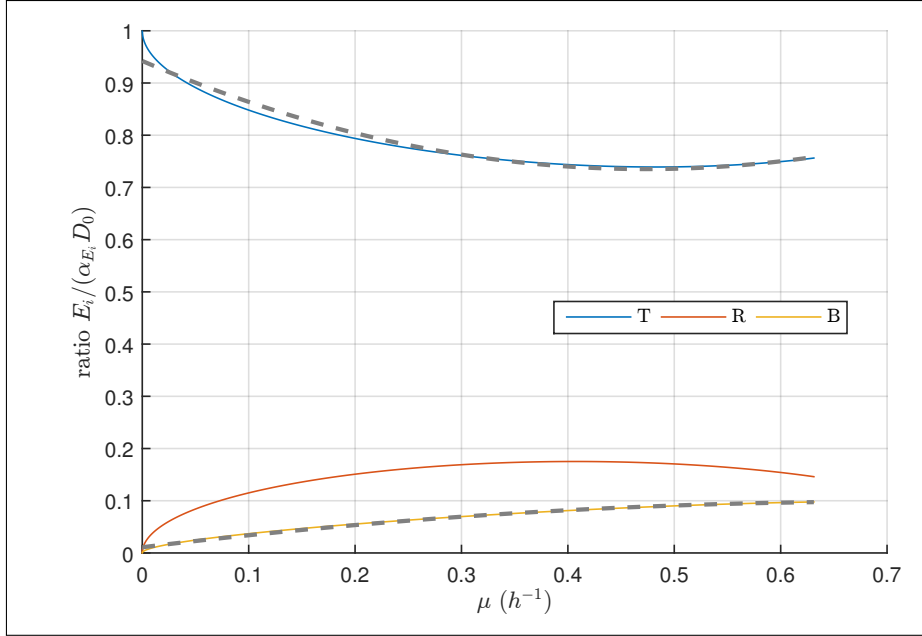


FIGURE 3.3 – Rappel de la figure 2.6 du chapitre 2 et superposition avec les approximations polynomiales (degré 2) des concentrations de $\mathcal{E}_T$ et $\mathcal{E}_B$ en fonction du taux spécifique de croissance. Les coefficients des polynômes sont donnés équation (3.2.9).

On revient à la figure 2.6 du chapitre précédent. Sur cette figure, on identifie, à l'aide d'une méthode des moindres carrés les paramètres a_B^N , b_B^N , c_B^N et a_T^N , b_T^N , c_T^N qui approximent le caractère quadratique des concentrations E_B et E_T . Les approximations sont données figure 3.3. Par une régression des moindres carrés, on trouve les valeurs suivantes pour les paramètres a_B^N , b_B^N , c_B^N et a_T^N , b_T^N , c_T^N :

$$\begin{cases} a_B^N = -0,2327 \\ a_T^N = 1,2304 \\ b_B^N = 0,3288 \\ b_T^N = -1,1641 \\ c_B^N = 0,0145 \\ c_T^N = 1,2531 \end{cases} \quad (3.2.9)$$

D'après les relations (3.1.8) et (3.1.6), ces coefficients sont les mêmes pour les concentrations $1/(\alpha_{E_i})E_i(\mu)$ à l'équilibre et pour les flux $\nu_{E_i}(\mu)$.

3.2.3.2 Résolution

A présent, on fixe les paramètres a_B^N , b_B^N , c_B^N et a_T^N , b_T^N , c_T^N aux valeurs (3.2.9) et on résout le problème (3.2.8) avec Bocop en initialisant l'algorithme d'optimisation par

la solution du cas n° 1 sauf les coefficients a_P^N , b_P^N , c_P^N qui sont initialisés à des valeurs proches de zéro pour que l'initialisation soit faisable. Les résultats sont présentés figure 3.4 où la solution du problème de reconfiguration complète est également rappelée.

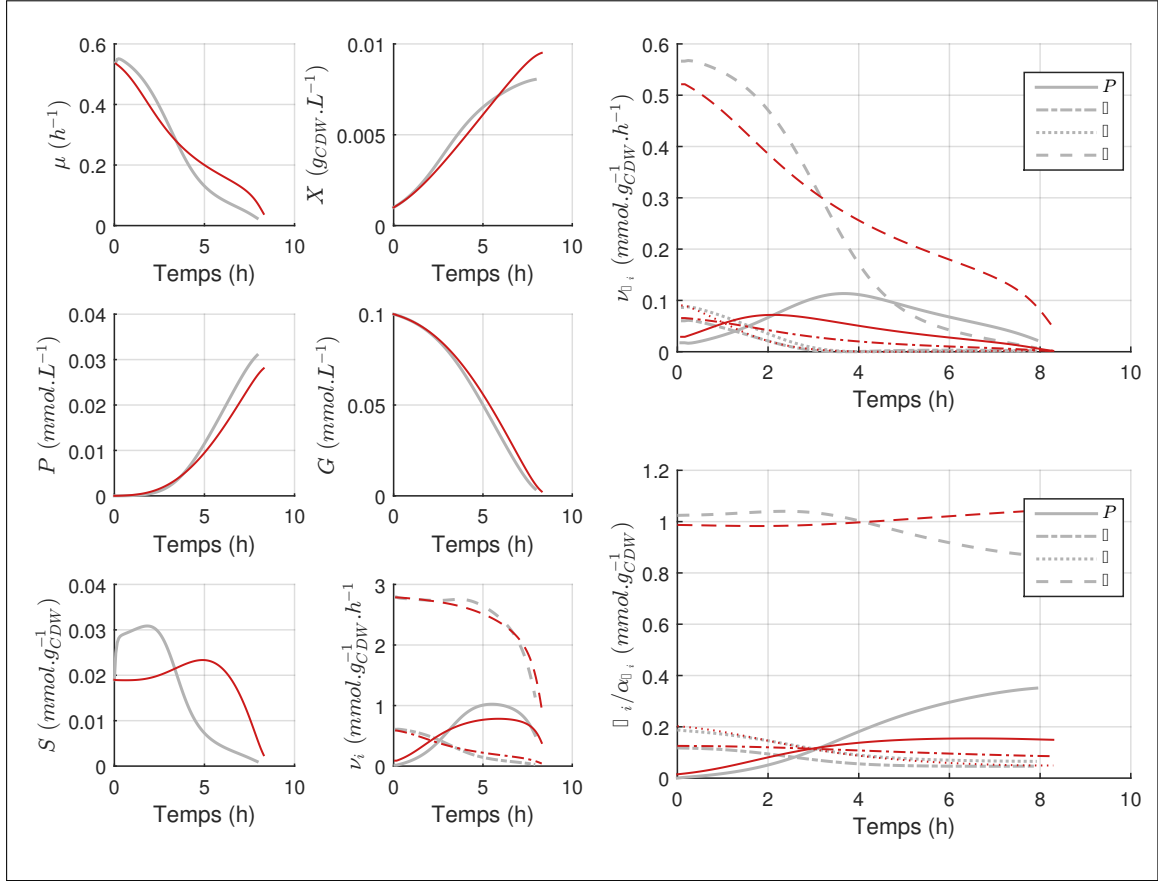


FIGURE 3.4 – Comparaison entre la solution au problème de reconfiguration complète de souche (Cas n° 1, en gris), et la solution au problème d'introduction d'un gène constitutif (Cas n° 2, en rouge).

Les profils sont aussi réguliers pour la solution au problème (3.2.8) que pour le problème (3.2.6). Simplement, on constate que la concentration de la protéine $\mathcal{E}_T$ tend à augmenter avec le temps, c'est-à-dire lorsque le taux spécifique de croissance diminue. Ceci est bien en accord avec l'allure de la figure 3.3. En revanche, comme la concentration de $\mathcal{E}_T$ augmente, il y a moins de place disponible dans la cellule pour les protéines $\mathcal{E}_P$. On voit ainsi que la concentration finale de $\mathcal{E}_P$ est moitié moins importante dans le cas n° 2 que dans le cas n° 1. La production du composé $\mathcal{P}$ est ainsi nécessairement moins élevée dans le cas où moins de machineries $\mathcal{E}_P$ ont été produites.

De la même manière que pour la figure 3.2, on peut s'intéresser aux allures des flux de production de protéine en fonction du taux spécifique de croissance. Celles-ci sont données figure 3.5. On reconnaît les interpolations de la figure 3.3 pour ν_{E_B} et ν_{E_T} . On note également que la somme des flux de production de $\mathcal{E}_R$ et $\mathcal{E}_P$ équivaut au flux de production de $\mathcal{E}_R$ dans le cas d'une maximisation de la croissance instantanée. Ceci semble

être raisonnable au vu du fait que la densité est limitée : si l'on ne produit pas $\mathcal{E}_R$, c'est pour former $\mathcal{E}_P$. On retrouve ici le fait que la production de $\mathcal{E}_P$ n'est pas aussi importante à faible taux spécifique de croissance qu'elle l'est dans le cas d'une reconfiguration totale de la cellule. En effet, comme le flux ν_{E_T} est élevé lorsque le taux de croissance est faible, le flux ν_{E_P} ne peut être aussi important à faible μ qu'il l'était dans le cas n° 1.

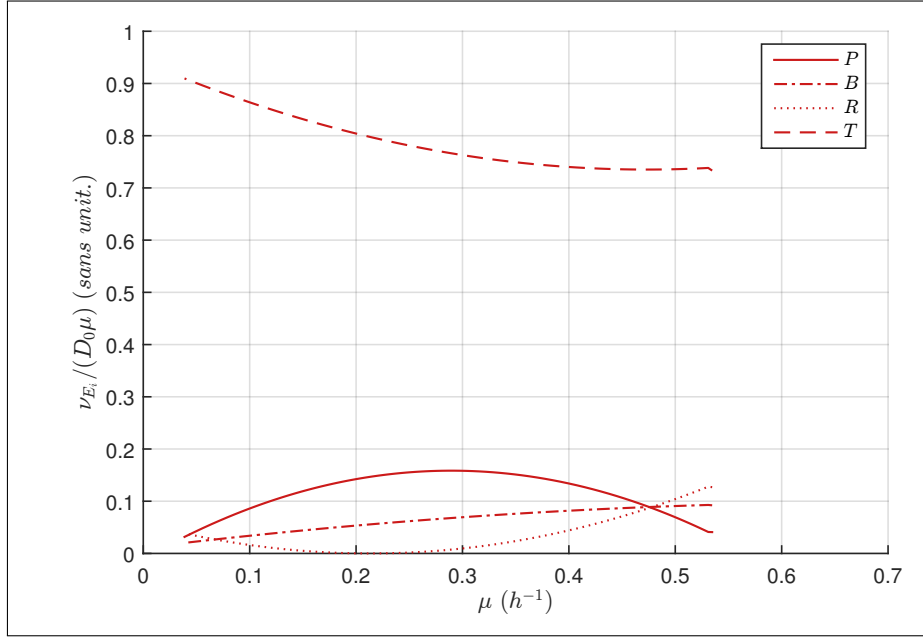


FIGURE 3.5 – Allure des flux $\nu_{E_i}(t)$ tracés en fonction du taux spécifique de croissance $\mu(t)$, pour la solution au problème (3.2.8).

3.2.4 Introduction d'un gène de synthèse constitutif, activable à un instant donné (Cas n° 3)

Nous venons de définir une conception de souche qui mène à la conduite du procédé en mode batch sans intervention extérieure. Il n'y a pas d'autre intervention à faire sur le procédé que d'ajouter un gène convenablement conçu.

Tout en restant dans l'idée que l'on souhaite agir le moins possible sur le procédé au cours du temps, on peut envisager de concevoir un gène de synthèse dont l'activité soit déclenchable à un instant particulier. Autrement dit, on cherche, ici, un profil de $\nu_{E_P}(t)$ tel que :

$$\begin{cases} \nu_{E_P}(t) = 0, & \text{pour } t \leq t_s \\ \nu_{E_P}(t) = a_P\mu(t)^3 + b_P\mu(t)^2 + c_P\mu(t), & \text{pour } t > t_s \end{cases} \quad (3.2.10)$$

avec a_P , b_P , c_P et t_s quatre paramètres à déterminer.

Il est possible d'ajouter cette contrainte au problème (3.2.8), en ajoutant le paramètre t_s aux variables d'optimisation et la contrainte (3.2.10) en lieu et place la contrainte

(3.2.8.h_P). La contrainte (3.2.8.i_P) est également remplacée par :

$$E_P(t_0) = 0 \quad (3.2.11)$$

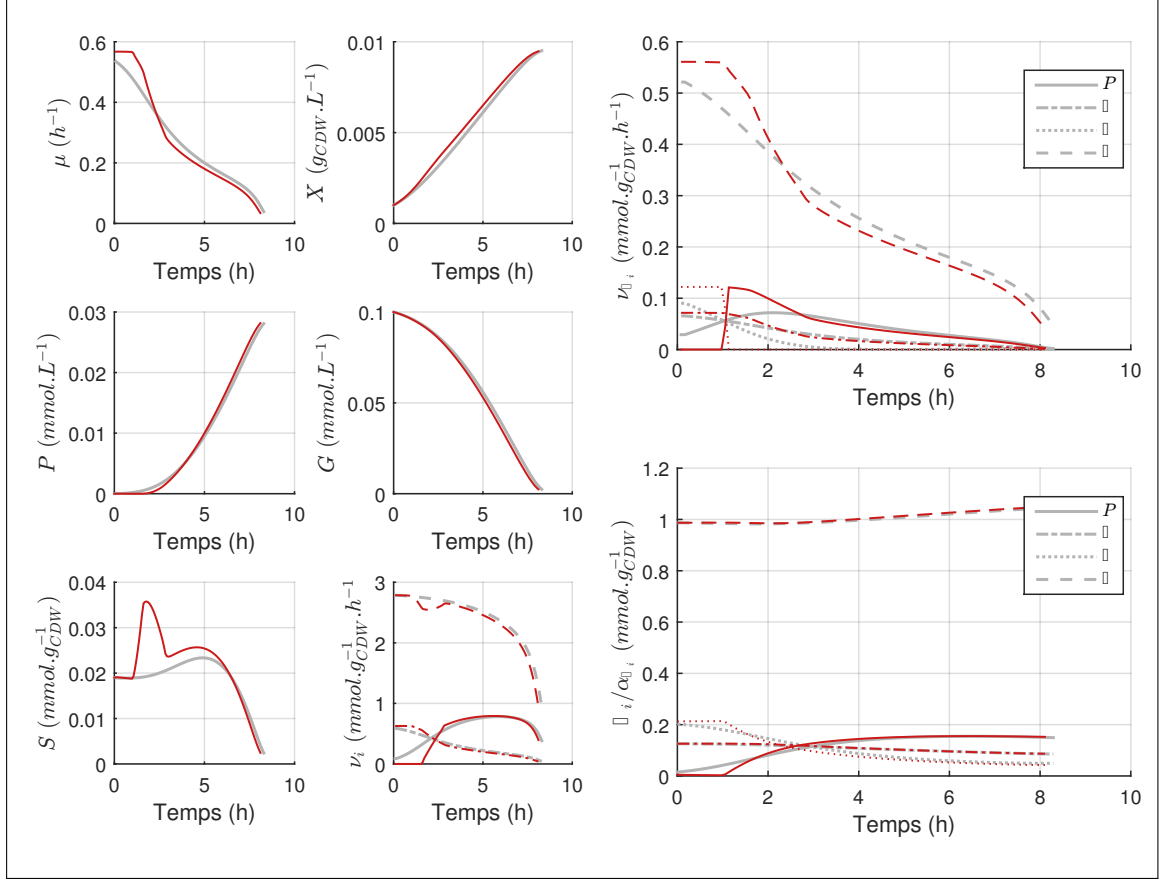


FIGURE 3.6 – Comparaison entre la solution au problème de reconfiguration complète de souche (Cas n° 1, en gris), et la solution au problème d'introduction d'un gène constitutif avec activation à un instant donné (Cas n° 3, en rouge).

Le problème est résolu en utilisant le logiciel Bocop, avec t_s initialisé à une valeur proche de zéro. Les résultats sont présentés figure 3.6. La valeur optimale de t_s est : $t_s = 0,97 \text{ h}^{-1}$. En comparant par rapport au cas n° 2, on constate qu'il y a une phase de croissance exponentielle à μ constant et pendant laquelle la configuration intracellulaire est constante : la cellule est en régime quasi-stationnaire. Par ailleurs, on retrouve exactement la valeur du taux spécifique de croissance maximal atteignable pour une concentration extérieure en substrat $\mathcal{G}$ égale à $0,1 \text{ mmol.L}^{-1}$ (figure 2.3). La répartition entre les $\mathcal{E}_i$ correspond également à celle obtenue pour cette valeur du taux spécifique de croissance (figure 3.3). La phase de croissance exponentielle correspond donc à une croissance maximale.

A $t_s = 0,97 \text{ h}$, la production de $\mathcal{E}_P$ commence et on constate sur la figure 3.6 en haut à droite (ν_{E_i} en fonction du temps) qu'elle se substitue à la production de $\mathcal{E}_R$.

Sur la figure 3.7 qui représente les flux de production de protéine en fonction du taux spécifique de croissance, on retrouve l'effet de commutation entre $\mathcal{E}_P$ et $\mathcal{E}_R$ pour μ de

l'ordre de $0,56 \text{ h}^{-1}$. L'instant de permutation t_s correspond d'une certaine manière à une commutation entre deux stratégies. Comme pour le cas de la figure 3.5, la somme des flux de production de $\mathcal{E}_P$ et $\mathcal{E}_R$ équivaut à la production de $\mathcal{E}_R$ dans le cas de la maximisation du taux spécifique de croissance. On constate que le profil de $\nu_{E_P}(\mu)$ est rigoureusement celui de E_R/α_{E_R} pour la maximisation du taux spécifique de croissance. Autrement dit, la solution optimale est de produire la protéine de synthèse à la place des ribosomes.

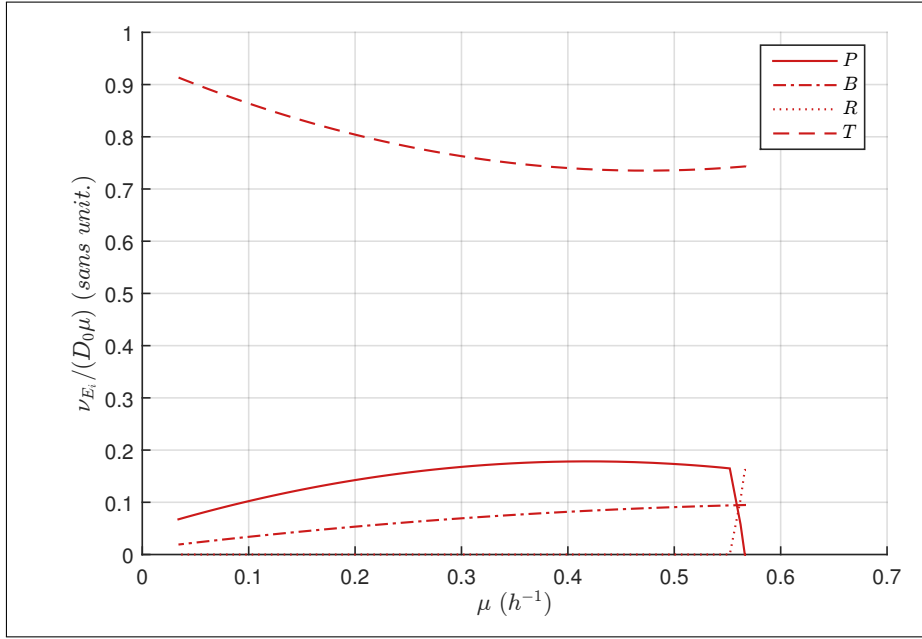


FIGURE 3.7 – Allure des flux $\nu_{E_i}(t)$ tracés en fonction du taux spécifique de croissance $\mu(t)$, pour la solution au problème (3.2.8). On reconnaît les interpolations de la figure 3.3 pour ν_{E_B} et ν_{E_T} .

D'un point de vue de la conduite de procédé, la stratégie optimale est de faire croître la population de micro-organismes de manière nominale jusqu'à un certain instant t_s , puis de substituer la production de ribosomes par des protéines qui catalysent la production de composé d'intérêt. Ceci semble réalisable en bloquant à $t = t_s$ la production de ribosomes. D'autre part, le promoteur du gène de synthèse pourrait dans ce cas être fortement semblable au promoteur d'un gène d'une protéine ribosomale.

Un intérêt notoire est d'avoir une phase où le gène de synthèse n'est pas du tout produit. Il faut toutefois reconnaître que la différence macroscopique avec le cas où il n'est pas possible d'activer à un instant donné la synthèse du gène de synthèse, ainsi que le gain de productivité entre les deux cas, sont minimes. Cet aspect sera détaillé en section 3.2.6. Il est raisonnable de penser que l'activation n'est pas ce qui est déterminant pour la production du composé d'intérêt.

3.2.5 Introduction d'un gène de synthèse constitutif avec action sur la croissance (Cas n° 4)

L'étude précédente a permis d'identifier une manière de retrouver la phase de croissance exponentielle à un taux spécifique de croissance maximal. On cherche à présent un moyen de faire apparaître la phase de production de $\mathcal{P}$ sans croissance, c'est-à-dire la dernière phase dans le cadre de l'optimisation du chapitre précédent. Pour cela, on s'autorise à moduler l'efficacité du processus de production des protéines.

On autorise donc le flux du processus Σ_R à ne pas valoir sa borne maximale. Ainsi, la somme des flux de production de protéine peut être modulé, et avec elle le taux spécifique de croissance. D'un point de vue de l'implémentation, il semble possible de moduler le taux spécifique de croissance en carençant le milieu de culture en un élément secondaire essentiel à la croissance des micro-organismes, par exemple en phosphore ou en azote.

Dans le problème (3.2.8), on modifie les contraintes (3.2.8. b_0) et (3.2.8. b_1) comme suit :

$$\begin{aligned} (3.2.8.b_0) : & \begin{cases} \nu_T(t) = k_T(G(t), S(t))E_T(t) \\ \nu_P(t) = k_P(S(t))E_P(t) \end{cases} \\ (3.2.8.b_1) : & \begin{cases} \sum_{P,B,R,T} \nu_{E_i}(t) \leq k_R(S(t))E_R(t) \\ \nu_B(t) \leq k_B(S(t))E_B(t) \end{cases} \end{aligned} \quad (3.2.12)$$

Le résultat du problème d'optimisation avec ces contraintes et les mêmes conditions de simulation et d'initialisation que le cas n° 3, est donné figure 3.8. Cette fois-ci la différence provient bien de la phase de production de $\mathcal{P}$ qui a lieu sans croissance, comme c'était le cas pour l'optimisation sans contrainte d'implémentation au chapitre précédent. Il est intéressant de retrouver également une augmentation de la concentration du pool interne $\mathcal{S}$ au moment où la croissance est interrompue. Cette augmentation s'accompagne d'une inhibition sur le flux ν_T mais est sans grande incidence sur le flux de production ν_P , qui augmente même légèrement.

Finalement, le fait de pouvoir interrompre la croissance permet de produire moins de biomasse et ainsi, par conservation de la masse, il y a plus de composé d'intérêt.

Les flux de production des $\mathcal{E}_i$ en fonction du taux spécifique de croissance ne sont pas représentés. Il s'agit sensiblement des mêmes polynômes que ceux obtenus sans la possibilité d'interrompre la croissance (figure 3.5).

3.2.6 Conclusion à propos de l'optimisation de procédé en mode de fonctionnement batch

Pour fixer les idées sur les tentatives d'optimisation de procédé en mode de fonctionnement batch, on regroupe dans le tableau 3.2 les caractéristiques quantitatives des différentes solutions.

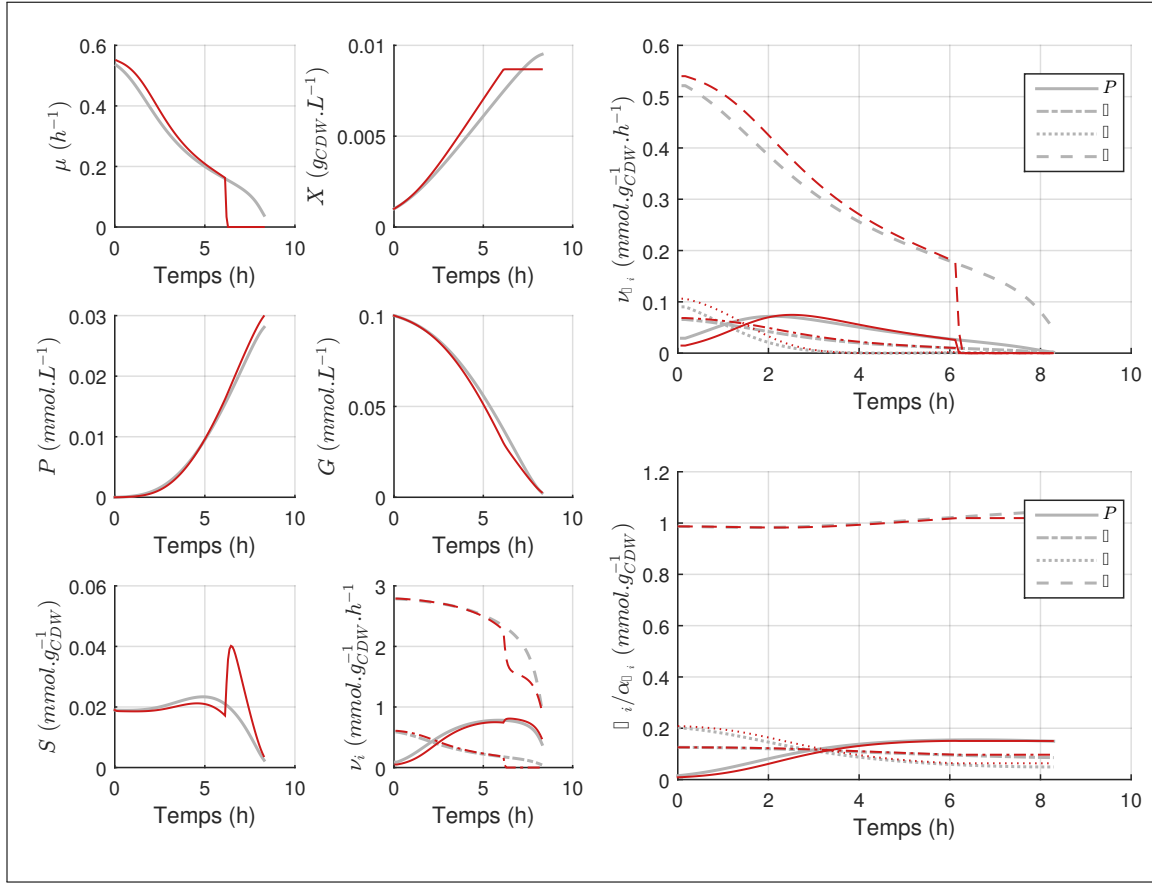


FIGURE 3.8 – Comparaison entre la solution au problème de reconfiguration complète de souche (Cas n° 1, en gris), et la solution au problème d'introduction d'un gène constitutif avec action sur la croissance (Cas n° 4, en rouge).

On rappelle à titre d'étalon l'optimum trouvé au chapitre précédent (Contrôle en temps réel de tous les processus intracellulaires). Cet étalon est un cas idéal qui est difficilement réalisable en pratique, comme expliqué en introduction de ce chapitre. On constate que le critère $P(t_f)/t_f$ est maximisé lorsque l'optimisation a le plus de degrés de liberté. La reconfiguration complète de la cellule donne des performances proches du cas où tous les processus sont contrôlables. Les trois cas pour lesquels les flux de production de $\mathcal{E}_T$ et $\mathcal{E}_B$ sont fixés donnent des résultats plus limités.

Entre la référence du chapitre précédent et le cas de reconfiguration complète, on perd 8% de performances générales ($= \frac{4,26-3,91}{4,26}$). L'introduction d'un gène de synthèse constitutif dégrade le critère de plus de 13% ($= \frac{3,91-3,39}{3,91}$). Cette perte d'optimalité se répercute sur la quantité de biomasse produite ($9,5 \text{ mg}_{CDW} \cdot L^{-1}$ au lieu de $8,1 \text{ mg}_{CDW} \cdot L^{-1}$). Le fait de pouvoir activer le gène à un instant donné ne permet pas de gagner réellement du produit, mais plutôt de gagner du temps (8,10 h au lieu de 8,29 h pour produire la même quantité de produit). Cependant, on constate que lorsque l'on permet à l'optimisation de limiter la croissance par d'autres moyens que la quantité de ribosome, on est en mesure de produire 6% de plus de produit ($= \frac{29,9-28,1}{28,1}$) en un temps comparable (8,27 h au lieu de 8,29 h).

TABLE 3.2 – Caractéristiques des solutions optimales pour les différents cas d'études.

Cas	$P(t_f)$ ($\mu\text{mol.L}^{-1}$)	t_f (h)	$P(t_f)/t_f$ ($\mu\text{mol.L}^{-1}.\text{h}^{-1}$)	$X(t_f)$ ($\text{mg}_{CDW}.\text{L}^{-1}$)
Étalon : Contrôle en temps réel de tous les processus intracellulaires	32,6	7,66	4,26	7,3
Cas n° 1 : Reconfiguration complète	31,0	7,93	3,91	8,1
Cas n° 2 : Introduction d'un gène de synthèse constitutif	28,1	8,29	3,39	9,5
Cas n° 3 : Introduction d'un gène "activable"	28,2	8,10	3,48	9,5
Cas n° 4 : Introduction d'un gène avec action sur la croissance	29,9	8,27	3,62	8,7

Par ailleurs, nous avons identifié des moyens de mettre en œuvre certaines composantes de la solution optimale où les processus intracellulaires sont tous contrôlables :

- la phase de croissance exponentielle sans production de composé d'intérêt est retrouvée en introduisant une activation sur le gène de production du composé d'intérêt.
- la phase de production pure de composé d'intérêt est retrouvée en inhibant la croissance à un instant donné.

Ainsi, agir sur la croissance des micro-organismes améliore grandement la productivité du procédé. On propose donc d'étudier à présent l'impact du mode de fonctionnement fed-batch, c'est-à-dire le contrôle du procédé par l'ajout de substrat $\mathcal{G}$.

3.3 Optimisation de la conception de bioprocédés fed-batch

On a vu qu'agir sur le bioprocédé, par exemple pour interrompre la croissance à un instant donné, permettait de gagner sur la productivité du bioprocédé. C'est pourquoi on va à présent explorer en quoi le mode de fonctionnement fed-batch, c'est-à-dire l'ajout potentiel de substrat au cours du procédé, peut permettre d'améliorer la bioproduction d'un composé d'intérêt.

On se place en mode de fonctionnement fed-batch. Soit $F_{in}(t)$ le débit d'alimentation en milieu de culture, exprimé en L.h^{-1} . On optimise le profil de $F_{in}(t)$ en parallèle de la conception d'un gène de synthèse. Nous allons présenter l'impact du mode de fonctionnement sur le modèle. Puis nous allons définir une formalisation du problème qui soit adapté à une comparaison avec le mode de fonctionnement batch. En effet, l'objectif de cette section est de déterminer l'avantage de contrôler le fonctionnement par la concentration en $\mathcal{G}$ disponible. Typiquement, on suppose disposer d'une certaine quantité de substrat et on cherche à optimiser la manière dont elle est ajoutée au procédé pour maximiser la production de produit. On aura ainsi étudié l'effet du mode de fonctionnement

fed-batch sur la conduite optimale du procédé.

3.3.1 Adaptation générale du problème

Se placer en mode de fonctionnement fed-batch modifie sensiblement la dynamique du système.

3.3.1.1 Dynamique du volume

Volume du réacteur. En fed-batch, on prend en compte l'augmentation de volume du milieu de culture, $V(t)$, comme suit :

$$\dot{V}(t) = F_{in}(t) \quad (3.3.1)$$

On introduit classiquement une contrainte sur le volume maximal du milieu de culture :

$$V(t) \leq V_{max}, \forall t \quad (3.3.2)$$

où V_{max} est le volume maximal limitant le milieu de culture.

Comme V est croissant, on peut remplacer la contrainte (3.3.2) simplement par :

$$V(t_f) \leq V_{max} \quad (3.3.3)$$

Par ailleurs, pour des raisons techniques, $F_{in}(t)$ doit être borné. Soient F_{min} et F_{max} les bornes minimale et maximale entre lesquelles doit être compris F_{in} à tout instant :

$$F_{min} \leq F_{in}(t) \leq F_{max}, \forall t \quad (3.3.4)$$

Évolutions des concentrations extracellulaires. Le débit d'alimentation $F_{in}(t)$ et la variation temporelle du volume $V(t)$ ont une influence sur les concentrations des espèces $\mathcal{P}$, $\mathcal{G}$ et sur l'expression de la concentration de la biomasse $X(t)$.

Soit $\xi(t)$ la concentration en composé Ξ exprimée en quantité de matière de Ξ ou en masse de Ξ par rapport au volume $V(t)$, on peut écrire :

$$\frac{d\xi(t)V(t)}{dt} = \frac{d\xi(t)}{dt}V(t) + \xi(t)\frac{V(t)}{dt} = \varphi_{\xi}^{prod.}(t) - \varphi_{\xi}^{cons.}(t) + \varphi_{\xi}^{alim.}(t) \quad (3.3.5)$$

où $\varphi_{\xi}^{cons.}$ (respectivement $\varphi_{\xi}^{prod.}$) est le flux absolu de Ξ consommé (respectivement produit) par les cellules à l'instant t , et $\varphi_{\xi}^{alim.}$ est le flux absolu de Ξ apporté par l'alimentation.

Posons $\varphi_{\xi}(t) = \varphi_{\xi}^{prod.}(t) - \varphi_{\xi}^{cons.}(t) + \varphi_{\xi}^{alim.}(t)$. En reportant l'équation (3.3.1) dans l'équation (3.3.5), il vient :

$$\frac{d\xi(t)}{dt}V(t) + \xi(t)F_{in}(t) = \varphi_{\xi}(t)$$

ou encore,

$$\frac{d\xi(t)}{dt} = \varphi_{\xi}(t)\frac{1}{V(t)} - \xi(t)\frac{F_{in}(t)}{V(t)} \quad (3.3.6)$$

D'après les équations (2.1.3), (2.1.8) et (2.1.14) du chapitre précédent, concernant $\mathcal{G}$, $\mathcal{P}$ et X , les flux absolus de production et de consommation s'expriment de la manière suivante, avec les mêmes notations que le chapitre précédent :

$$\begin{cases} \varphi_G^{cons.}(t) = \nu_T(t)X(t)V(t) \\ \varphi_P^{prod.}(t) = \nu_P(t)X(t)V(t) \\ \varphi_X^{prod.}(t) = \mu(t)X(t)V(t) \\ \varphi_G^{prod.}(t) = \varphi_P^{cons.}(t) = \varphi_X^{cons.}(t) = 0 \end{cases} \quad (3.3.7)$$

Les flux absolus d'alimentation dépendent de la composition du milieu d'alimentation, comme suit :

$$\varphi_\xi^{alim.} = \xi_F F_{in}(t) \quad (3.3.8)$$

où ξ_F est la concentration de Ξ dans le milieu d'alimentation.

On peut raisonnablement supposer qu'il n'est composé que de $\mathcal{G}$ (*i.e.* qu'il n'y a ni biomasse, ni produit d'intérêt) et alors,

$$\begin{cases} \varphi_G^{alim.}(t) = G_F F_{in}(t) \\ \varphi_P^{alim.}(t) = 0 \\ \varphi_X^{alim.}(t) = 0 \end{cases} \quad (3.3.9)$$

Ainsi, en reportant les relations (3.3.7) et (3.3.9) dans la relation générale (3.3.6), les évolutions temporelles des quantités $G(t)$, $P(t)$ et $X(t)$ sont données par :

$$\begin{cases} \dot{G}(t) = -\nu_T(t)X(t) + (G_F - G(t))\frac{F_{in}(t)}{V(t)} \\ \dot{P}(t) = \nu_P(t)X(t) - P(t)\frac{F_{in}(t)}{V(t)} \\ \dot{X}(t) = \mu(t)X(t) - X(t)\frac{F_{in}(t)}{V(t)} \end{cases} \quad (3.3.10)$$

Tant que l'on supposera qu'il n'y a pas de cellules dans le débit d'alimentation, les concentrations des espèces intracellulaires ne seront pas impactées par le changement de mode de fonctionnement.

3.3.1.2 Consommation de substrat

La consommation de substrat est définie comme la quantité de substrat qui a été introduite dans le réacteur, *i.e.* qui a été en contact avec les cellules.

Dans notre cas, elle s'exprime sous la forme :

$$\mathcal{C}_G = \underbrace{G(t_0)V(t_0)}_{\text{quantité initiale}} + \underbrace{\int_{t_0}^{t_f} G_F F_{in}(t) dt}_{\text{quantité introduite}}$$

où G_F est la concentration de substrat du milieu d'alimentation.

En intégrant l'équation (3.3.1) entre t_0 et t_f ,

$$\mathcal{C}_G = G(t_0)V(t_0) + G_F(V(t_f) - V(t_0)) \quad (3.3.11)$$

$G(t_0)$, G_F , et $V(t_0)$ sont des paramètres du procédé mais peuvent éventuellement être optimisés, comme nous le verrons au paragraphe 3.3.3.

Par ailleurs, on s'intéresse à des réacteurs pour lesquels la consommation de substrat est limitée, c'est-à-dire que l'on introduit des contraintes du type :

$$\mathcal{C}_G \leq \mathcal{C}_G^{max} \quad (3.3.12)$$

avec $\mathcal{C}_G^{max}$ la consommation maximale autorisée en substrat $\mathcal{G}$. Ceci est équivalent à :

$$G(t_0)V(t_0) + G_F(V(t_f) - V(t_0)) \leq \mathcal{C}_G^{max}$$

ou encore,

$$V(t_f) \leq \frac{\mathcal{C}_G^{max} + V(t_0)(G(t_f) - G(t_0))}{G_F} \quad (3.3.13)$$

Cette inégalité est un cas particulier de l'inégalité (3.3.2). Il y a même équivalence entre limiter $\mathcal{C}_G$, limiter $V(t)$ à chaque instant et limiter $V(t_f)$. Cette équivalence se démontre de manière immédiate par le calcul développé ci-dessus et car $V(t)$ est croissant.

3.3.1.3 Modification du critère d'optimalité du procédé

On définit un critère d'optimalité associé au mode de fonctionnement fed-batch. On cherche avant tout un critère qui permette d'évaluer l'intérêt de procéder en mode de fonctionnement fed-batch par rapport aux procédés batch présentés jusqu'ici.

Le critère à optimiser peut alors être :

$$\max J = \max P(t_f)V(t_f) \quad (3.3.14)$$

qui est la quantité de produit à l'instant final.

Ou, comme on l'a présenté en batch, le rendement par rapport au temps :

$$\max J = \max \frac{P(t_f)V(t_f)}{t_f} \quad (3.3.15)$$

Ou encore, on peut s'intéresser au rendement du procédé par rapport à la quantité de $\mathcal{G}$ consommée :

$$\max J = \max \frac{P(t_f)V(t_f)}{\mathcal{C}_G} \quad (3.3.16)$$

3.3.1.4 Synthèse du problème d'optimisation de bioproduction en mode de fonctionnement fed-batch

Pour tester l'avantage du mode de fonctionnement fed-batch, on fixe la quantité de substrat disponible :

$$\mathcal{C}_G^{max} = 2 \text{ mmol} \quad (3.3.17)$$

Pour le reste, on prend les mêmes paramètres et valeurs initiales que précédemment. Ainsi, la quantité initiale de $\mathcal{G}$ présente dans le réacteur est :

$$G(t_0)V(t_0) = 1 \text{ mmol} = \mathcal{C}_G^{max}/2 \quad (3.3.18)$$

De sorte que cette expérience revient à donner au procédé décrit section 3.2.3 le double de substrat et à se demander de quelle manière il faut l'incorporer au milieu de culture.

Au vu de ces conditions, on choisit le critère (3.3.16) qui semble être un bon compromis pour tester l'intérêt d'ajouter du substrat. Ceci nous permet de tester s'il y a un intérêt à utiliser le substrat de l'alimentation ou s'il est préférable de faire des successions de batch. Le problème d'optimisation du rendement du procédé de production d'un composé d'intérêt en mode de fonctionnement fed-batch s'écrit dans notre cas, sous la forme suivante :

$$\begin{aligned} & \max_{\substack{F_{in}(t), \nu_B(t), \nu_{E_R}(t) \\ a_i, b_i, c_i, t_f}} \frac{P(t_f)V(t_f)}{\mathcal{C}_G} \\ & \text{sous contraintes identiques à (3.2.6) + } \begin{cases} \dot{x}(t) = F(x(t))\nu(t) + F_1(x(t))F_{in}(t) & (3.3.19.a) \\ F_{min} \leq F_{in}(t) \leq F_{max} & (3.3.19.d_F) \\ V(t_f) \leq V_{max} & (3.3.19.f_V) \\ \mathcal{C}_G \leq \mathcal{C}_G^{max} & (3.3.19.f_{\mathcal{C}}) \end{cases} \end{aligned} \quad (3.3.19)$$

Les variables d'optimisation de ce problème sont les flux non contraints $\nu_{E_R}(t)$, $\nu_B(t)$, le débit d'alimentation $F_{in}(t)$, le temps final t_f et éventuellement les coefficients polynomiaux a_i , b_i et c_i de certains processus constitutifs. On considère que ces coefficients sont soit fixés a priori, soit à optimiser.

3.3.2 Problème avec cellules connues à l'avance

On fixe dans un premier temps le comportement des cellules en reprenant la cellule optimale de la section 3.2.3, c'est-à-dire la cellule dont les flux de production de $\mathcal{E}_T$ et $\mathcal{E}_B$ sont des polynômes du taux spécifique de croissance de coefficients (3.2.9) et où le flux de production de $\mathcal{E}_P$ est également un polynôme (donné figure 3.4) de coefficients :

$$\begin{cases} a_P = -2,68 \\ b_P = 1,55 \\ c_P = -0,014 \end{cases} \quad (3.3.20)$$

On considère que le volume initial du réacteur est :

$$V(t_0) = 10 \text{ L} \quad (3.3.21)$$

que la concentration en $\mathcal{G}$ du milieu d'alimentation est :

$$G_F = 3 \text{ mmol.L}^{-1} \quad (3.3.22)$$

On suppose que la contrainte V_{MAX} n'est pas limitante par rapport à la contrainte sur $\mathcal{C}_G$. On pourra prendre $V_{MAX} = 20 \text{ L}$, sachant que d'après (3.3.13), (3.3.17) et (3.3.22), le volume ne peut excéder 13,34 L.

On pose également :

$$\begin{cases} F_{min} = 0 \text{ L.h}^{-1} \\ F_{max} = 1 \text{ L.h}^{-1} \end{cases} \quad (3.3.23)$$

En initialisant le problème d'optimisation avec différentes valeurs de F_{in} constant à tout instant, il vient la solution présentée figures 3.9 et 3.10. A noter que la borne maximale, t_{max} , sur le temps final n'est pas atteinte. Dans le cas du résultat de la figure 3.9, $t_f = 28,44$ h et $t_{max} = 50$ h.

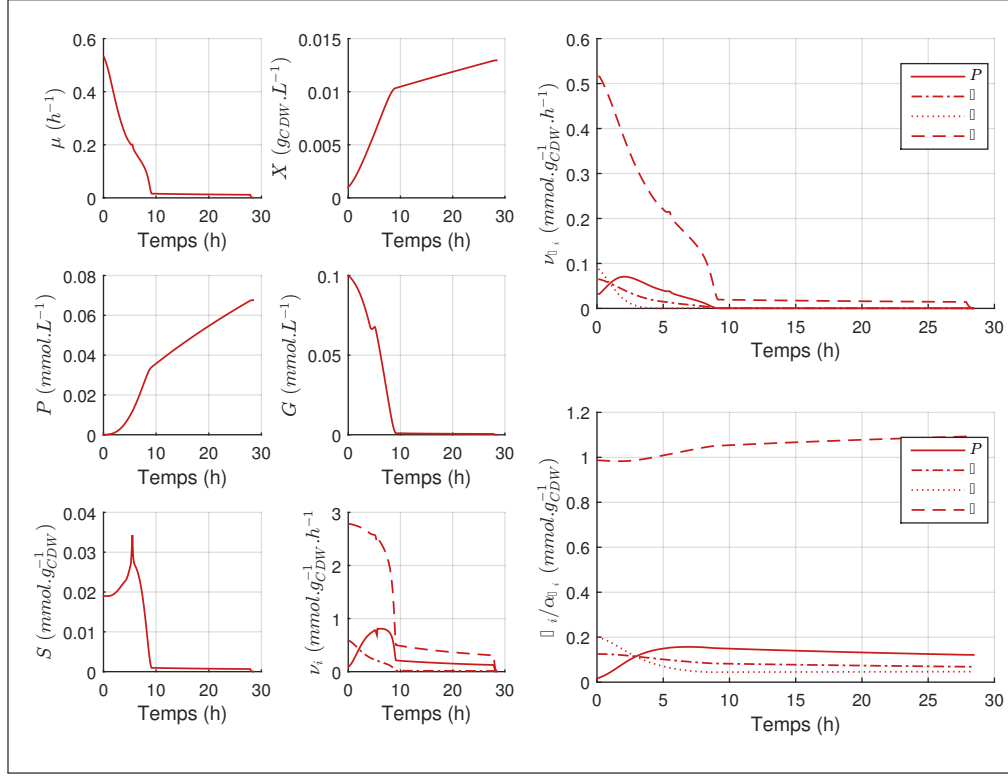


FIGURE 3.9 – Allures de la solution au problème (3.3.19).

Le profil d'alimentation de la figure 3.10 a trois composantes :

- une impulsion à $t = 4,9$ h,
- un débit uniforme de $t = 9,2$ h à $t = 27,9$ h,
- un mode batch $F_{in} = 0$ sinon.

On constate que tout le substrat disponible a été consommé, ce qui revient à dire que l'optimisation a choisi de transformer le gramme d'alimentation de $\mathcal{G}$ en plus du gramme disponible à l'instant initial.

Analysons ce résultat. En termes de taux de croissance, l'impulsion a lieu lorsque $\mu = 0,2$ h⁻¹. En se reportant à la figure 3.5, on constate qu'il s'agit de la valeur du taux de croissance pour laquelle la production de $\mathcal{E}_R$ est nulle. Aussi, le débit uniforme de la deuxième phase correspond à maintenir la cellule dans une situation particulière où la croissance est très faible et où la concentration intracellulaire en $\mathcal{S}$ est très proche de zéro.

Ces observations laissent paraître que pour maximiser la quantité de $\mathcal{P}$ produite, l'optimisation tente de maximiser le flux $\nu_P(t)$ et de minimiser le taux spécifique de croissance $\mu(t)$. D'une certaine manière, ceci revient à maximiser $k_P(S(t))$ et $E_P(t)$ (dont le produit est $\nu_P(t)$), et à minimiser $k_R(t)$ et $E_R(t)$ (dont le produit est proportionnel à

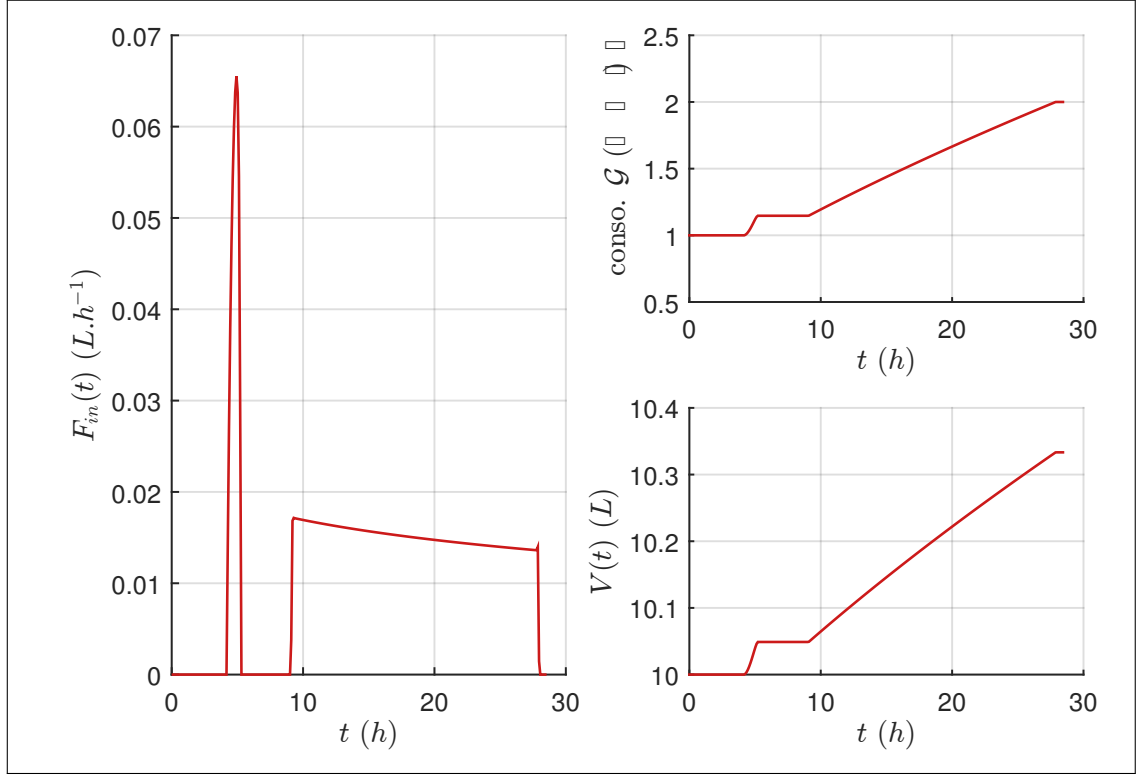


FIGURE 3.10 – Allures du débit d'alimentation pour la solution au problème (3.3.19).

$\mu(t)$). En ajoutant du substrat à $\mu = 0,2 \text{ h}^{-1}$, l'optimisation favorise la production de cellules avec peu de $\mathcal{E}_R$, donc plus de $\mathcal{E}_P$. L'impulsion maximise en quelques sortes le ratio E_P/E_R .

L'alimentation uniforme de la deuxième phase intervient à un instant où la concentration de $\mathcal{S}$ est particulièrement faible. Nous allons montrer que ceci a tendance à maximiser k_P/k_R .

En effet, à partir de 9,2 h, la concentration de $\mathcal{S}$ vaut moins de $0,8 \text{ } \mu\text{mol.g}_{CDW}^{-1}$. En injectant cette valeur dans les expressions de $k_P(S)$ et $k_R(S)$, on trouve :

$$\begin{cases} k_R(S = 0,8 \mu\text{mol.g}_{CDW}^{-1}) = 7,4\% \text{ } k_R^{max} \\ k_P(S = 0,8 \mu\text{mol.g}_{CDW}^{-1}) = 21\% \text{ } k_P^{max} \end{cases}$$

ainsi,

$$\frac{k_P(S = 0,8 \mu\text{mol.g}_{CDW}^{-1})/k_P^{max}}{k_R(S = 0,8 \mu\text{mol.g}_{CDW}^{-1})/k_R^{max}} = 2,84 \quad (3.3.24)$$

Sachant qu'à la limite, pour S qui tend vers 0,

$$\frac{k_P(S)/k_P^{max}}{k_R(S)/k_R^{max}} = \frac{S + K_R}{S + K_P} \xrightarrow{S \rightarrow 0} \frac{K_R}{K_P} = 3,33 \quad (3.3.25)$$

mais qu'alors, l'efficacité du processus Σ_P tend également vers 0.

Il semble donc que la méthode d'optimisation considère qu'un compromis acceptable soit de se placer à $S = 0,8 \mu\text{mol.g}_{CDW}^{-1}$ pour maximiser le ratio k_P/k_R et avoir $k_P \neq 0$.

On constate que le mode fed-batch est pleinement utilisé pour placer la cellule à un point de fonctionnement plus efficace pour un certain critère.

3.3.3 Optimisation conjointe du débit d'alimentation et de l'expression d'un gène de synthèse

Naturellement, se pose désormais la question de pouvoir optimiser conjointement le débit d'alimentation et l'expression d'un gène de synthèse avec notre modèle en mode de fonctionnement fed-batch.

La question sous-jacente est de trouver la manière optimale d'approvisionner le réacteur en substrat. En définitive, le mode de fonctionnement fed-batch apporte-t-il un avantage par rapport au mode de fonctionnement batch, dès lors que l'on autorise les modifications de souche ?

Avec notre approche, le problème se formalise sous la forme suivante :

$$\begin{aligned} & \max_{\substack{F_{in}(t), \nu_B(t), \nu_{ER}(t), t_f \\ a_P, b_P, c_P \\ G(t_0), G_F}} J \\ & \text{avec les contraintes (3.3.19) et avec } \begin{cases} (a_B, b_B, c_B) = (a_B^N, b_B^N, c_B^N) \\ (a_T, b_T, c_T) = (a_T^N, b_T^N, c_T^N) \end{cases} \end{aligned} \quad (3.3.26)$$

où les variables de commande sont le débit d'alimentation, les flux qui ne sont pas contrôlés par la cellule, le temps final, les coefficients a_P , b_P , c_P , ainsi que $G(t_0)$ et G_F c'est-à-dire les concentrations initiales et du milieu d'alimentation.

La quantité $\mathcal{C}_G$ de la contrainte (3.3.19. $f_{\mathcal{C}}$) est définie en (3.3.11). Elle dépend directement des paramètres d'optimisation $G(t_0)$ et G_F .

J est le critère pour un procédé fed-batch. On pourra se référer aux critères présentés en (3.3.14), (3.3.15) et (3.3.16). Nous allons voir que pour ces trois critères, le contrôle du débit d'alimentation n'apporte pas de réel avantage par rapport à l'optimisation en batch.

3.3.3.1 Maximisation de la quantité produite : $J = P(t_f)V(t_f)$

Au chapitre précédent, on a vu la solution optimale en batch pour ce genre de critère (section 2.3.5.1, figure 2.10) : il suffisait d'avoir une croissance nulle et de transformer tout le substrat en produit.

Il est tout à fait possible de reproduire la même situation. En effet, pour garantir une croissance nulle à l'instant initial, il suffit de prendre :

$$\begin{cases} a_P = -a_B - a_T \\ b_P = -b_B - b_T \\ c_P = D_0 - c_B - c_T \end{cases} \quad (3.3.27)$$

de sorte que $\nu_{ER}(t) = 0, \forall t$ et en particulier à l'instant initial. Ainsi, la croissance est nulle à l'instant initial t_0 et le reste pendant toute la durée de l'expérience. Par ailleurs, la concentration de $\mathcal{E}_P$ à l'instant initial est non nulle.

Avec t_f assez grand, on retrouve la situation dans laquelle tout le substrat est transformé en produit. Le débit d'alimentation F_{in} n'intervient pas dans la caractérisation de l'optimum.

D'après la conservation de la masse, cette solution conduit à une quantité de $\mathcal{P}$ produite :

$$P(t_f)V(t_f) = \frac{M_{\mathcal{G}}}{M_{\mathcal{P}}} \mathcal{C}_G \quad (3.3.28)$$

où $M_{\mathcal{G}}$ (respectivement $M_{\mathcal{P}}$) est la masse molaire de $\mathcal{G}$ (respectivement $\mathcal{P}$) et où $\mathcal{C}_G$ est la quantité de $\mathcal{G}$ consommée.

Ainsi, on conclut que $J_{opt} = \frac{M_{\mathcal{G}}}{M_{\mathcal{P}}} \mathcal{C}_G$.

La solution est optimale tant que $\mathcal{C}_G = \mathcal{C}_G^{max}$, quelle que soit la manière dont est introduit le substrat $\mathcal{G}$ dans le réacteur.

3.3.3.2 Maximisation du rendement du procédé : $J = \frac{P(t_f)V(t_f)}{\mathcal{C}_G}$

Pour caractériser la solution au problème (3.3.26) avec ce critère, on se place dans un premier temps dans la situation où la consommation de substrat $\mathcal{C}_G$ est fixée. Dans ce cas, on est ramenée à la situation précédente : $\max P(t_f)V(t_f)$ sachant que $\mathcal{C}_G = cste$.

On vient de voir que la solution optimale à ce problème était indépendante du mode de fonctionnement fed-batch et qu'une condition suffisante était de prendre les coefficients a_P , b_P , c_P selon (3.3.27), et $F_{in}(t)$, G_F et $G(t_0)$ de sorte à respecter la contrainte sur $\mathcal{C}_G$.

Par ailleurs, d'après (3.3.28), pour chacun des problèmes particuliers, la valeur maximale du critère est donnée par :

$$J_{opt} = \frac{P_{opt}(t_f)V_{opt}(t_f)}{\mathcal{C}_G} = \frac{M_{\mathcal{G}}}{M_{\mathcal{P}}}$$

Cette quantité est indépendante de la valeur de $\mathcal{C}_G$. Aucun choix particulier de $\mathcal{C}_G$ n'est donc privilégié.

Si l'on revient au problème général où la valeur de $\mathcal{C}_G$ n'est pas fixée, on conclut que la solution est indépendante de $\mathcal{C}_G$ et du mode d'alimentation. Il suffit de transformer tout le substrat en produit et ceci est possible en choisissant les coefficients a_P , b_P et c_P de sorte à ne rendre possible aucune croissance.

3.3.3.3 Maximisation du rendement temporel : $J = \frac{P(t_f)V(t_f)}{t_f}$

On va montrer que le mode de fonctionnement batch est optimal au sens de ce critère. En effet, nous allons montrer par l'absurde qu'il n'y a pas d'avantage à ajouter le substrat au fur et à mesure par rapport au fait de l'avoir à disposition à l'instant initial.

La configuration de ce problème est un cas particulier d'un problème batch dans laquelle on relaxe la contrainte $\nu_T(t) = k_T(G(t), S(t))E_T(t)$ en $\nu_T(t) \leq k_T(G(t), S(t))E_T(t)$. Ce faisant, on autorise les cellules à importer au fur et à mesure le substrat.

On peut alors définir l'efficacité effective du processus Σ_T comme étant l'efficacité du $k_{T,eff}$ telle que :

$$\nu_T(t) = k_{T,eff}(t)E_T(t) \leq k_T(G(t), S(t))E_T(t)$$

S'ensuit la définition de la concentration effective en $\mathcal{G}$, $G_{eff}(t)$ telle que :

$$k_{T,eff}(t) = k_T(G_{eff}(t), S(t))$$

qui est la concentration en $\mathcal{G}$ minimale qui est effectivement utilisée par la souche à l'instant t .

Comme k_T est croissant en G , $G_{eff}(t) \leq G(t)$.

Une partie du problème est de trouver l'allure optimale $G_{eff,opt}(t)$. Il faudra ensuite trouver un profil $F_{in}(t)$ qui puisse suivre cette allure quand on repassera au problème (3.3.26).

Nous allons montrer que le profil $G_{eff} = G$ est optimal. Le profil $F_{in}(t)$ à suivre sera alors simplement $F_{in} = 0$.

On considère une trajectoire $G_{eff,i}$ telle qu'à l'instant t_0 , $G_{eff,i}(t_0) < G(t_0)$. On définit alors la trajectoire $G_{eff,e}$ qui est égale à $G_{eff,i}$ en tout point sauf en t_0 où $G_{eff,e}(t_0) = G(t_0)$.

Montrons que substituer $G_{eff,e}$ à $G_{eff,i}$ conduit à une solution meilleure au sens du critère $J = \frac{P(t_f)V(t_f)}{t_f}$.

En prenant $G_{eff,e} > G_{eff,i}$, on augmente la concentration S en t_0 . Ainsi, on augmente l'efficacité des processus Σ_R et Σ_P . Ceci conduit donc à une diminution du temps nécessaire pour produire la même quantité de biomasse et de produit d'intérêt : pour la même quantité $P(t_f)$, t_f est plus faible pour le cas $G_{eff,e}$. On a bien amélioré le critère.

Ce qu'il fallait démontrer : $G_{eff}(t) = G(t), \forall t$.

On a ainsi montré que le fonctionnement en batch avec la quantité maximale de substrat introduite dès le début est meilleur du point de vue du critère $J = \frac{P(t_f)V(t_f)}{t_f}$ que tous les fed-batch possibles. Il n'est donc pas pertinent d'étudier le mode de fonctionnement fed-batch pour ce critère. On peut se contenter du batch.

3.3.3.4 Conclusion : un problème surparamétré

Au travers de ces trois exemples, nous avons montré que le problème d'optimisation conjointe de la conception de souche et de la conduite du procédé en mode de fonctionnement fed-batch revenait à un problème d'optimisation de conception de souche en mode de fonctionnement batch. La simplicité de cette solution et la sorte d'équivalence entre batch et fed-batch est à mettre sur le compte de la simplicité du modèle à ce stade.

D'après la littérature, le mode de fonctionnement fed-batch est privilégié dans la conduite de procédé pour limiter les effets d'inhibition et de production de cofacteurs,

contrôler les concentrations de composés secondaires, de pH, d'oxygène comme on peut le lire dans [86]. Le modèle fondé sur l'agrégation développée au chapitre précédent ne permet donc pas de révéler l'intérêt du mode de fonctionnement fed-batch par rapport au mode de fonctionnement batch dans ce contexte. Il semble pertinent de raffiner le modèle pour faire apparaître un des effets cité dans [86].

3.4 Conclusion du chapitre

Le travail développé dans ce chapitre a permis de poser un cadre rigoureux pour la conception rationnelle de souche en mode de fonctionnement batch et fed-batch. Nous avons donné des réponses concrètes d'implémentations de stratégies optimales aux problèmes de conception rationnelle de souche et de conduite du procédé. A notre connaissance, cette intégration de contraintes d'implémentation biologique dans un problème de contrôle du fonctionnement de bioprocédé, constitue une première dans la littérature. Le rapprochement entre ces mêmes contraintes d'implémentation et une description dynamique autonome des processus intracellulaires est également une première, bien que sur un modèle agrégé.

Notre étude sur les solutions au problème de conception de souche en relaxant les contraintes d'autonomie des processus intracellulaires nous a permis de comprendre sur quels processus intracellulaires il était potentiellement intéressant d'agir pour améliorer les performances d'un procédé. Nous avons vu qu'activer la production d'un gène à un instant donné permettait de faire apparaître une phase de croissance exponentielle pure (*i.e.* sans production de composé d'intérêt). Nous avons aussi observé qu'agir sur le processus de production des protéines en l'interrompant à partir d'un instant permettait d'avoir une phase de production pure de composé d'intérêt sans dépenser de ressource inutile en biomasse, ce qui procure un net avantage du point de vue du critère considéré.

Nous avons exploité le mode de fonctionnement fed-batch pour maximiser le rendement de conversion en produit d'intérêt avec la souche de synthèse à laquelle un gène de synthèse constitutif a été ajouté. La modélisation intracellulaire a permis lors de la conduite du procédé de donner une interprétation qualitative de la trajectoire optimale du débit d'alimentation. Ceci permettrait dans la suite d'envisager une loi de commande en boucle fermée en accord avec le procédé réel.

Nous avons enfin envisagé d'optimiser conjointement la conduite du procédé et la conception de souche. Une analyse théorique nous a montré qu'il n'y avait pas d'intérêt à se placer en mode de fonctionnement fed-batch à ce stade d'agrégation. L'aspect contrôle en temps réel qu'autorise le mode de fonctionnement fed-batch est plus intéressant pour inhiber des processus intracellulaires que pour maintenir des conditions extérieures à des valeurs adéquates.

Les travaux de ce chapitre ont permis de faire le tour de la question de l'optimisation de la conduite du procédé et de la conception rationnelle de souche. Nous pouvons donc entreprendre sereinement le passage à une échelle de description plus fine. Le prochain chapitre y est consacré.

Chapitre 4

Extension de l'approche à des descriptions cellulaires plus détaillées

Les deux chapitres précédents donnent une preuve de concept de la conception optimale rationnelle de souche adaptée à la conduite de procédé. L'objectif de ce chapitre est de montrer qu'il est possible d'étendre la philosophie de la modélisation à une représentation plus élaborée du fonctionnement des micro-organismes.

Nous présentons en section 4.1 un cadre de généralisation de la démarche de modélisation et de conception de souche (correspondant aux chapitres 2 et 3). L'objectif est de pouvoir étendre l'approche à n'importe quel réseau de réactions intracellulaires.

En section 4.2, nous donnons un exemple d'application de ce protocole à une représentation simplifiée du métabolisme où deux processus réalisent la même fonction. Nous verrons qu'en prenant une représentation plus riche du métabolisme, l'optimisation de la production d'un composé d'intérêt est plus élaborée.

4.1 Éléments de généralisation

Dans cette section, nous nous intéressons à la formulation d'un problème d'optimisation de la production d'un composé d'intérêt en nous appuyant sur une représentation plus détaillée du réseau de processus intracellulaires.

4.1.1 Protocole de construction d'une représentation dRBA

L'objectif de cette partie est de construire à partir d'un réseau de processus intracellulaires, une représentation dRBA telle qu'elle a été présentée au chapitre 2 pour le modèle agrégé. Le but de cette partie est d'établir un protocole général pour aboutir à une formulation cohérente. L'objectif de la démarche est de comprendre quels éléments sont nécessaires à la modélisation.

On suppose que le procédé est opéré dans le mode batch.

Données. On appelle $\mathcal{R}$ un réseau de réactions décrivant le fonctionnement d'une cellule, c'est-à-dire un ensemble de processus. On suppose disposer d'une représentation

de $\mathcal{R}$ sous forme d'une matrice de stœchiométrie étendue Ω , décrivant les échanges entre composés intra et extracellulaires. Cette matrice n'agrège pas uniquement les échanges intracellulaires mais également ceux avec le milieu extérieur. Pour fixer les idées, cette matrice a autant de colonnes qu'il y a de processus, et autant de lignes qu'il y a de composés chimiques intra et extracellulaires. On note n_Σ le nombre de processus, $n_{\mathcal{M}}$ le nombre de composés. Par ailleurs, les processus ne sont pas uniquement des processus métaboliques.

On note $\mathcal{M}$ l'ensemble des composés pris en compte dans la représentation. Cet ensemble est divisé en deux sous-ensemble $\mathcal{M}_{int}$ et $\mathcal{M}_{ext}$ qui regroupent respectivement les métabolites intracellulaires et extracellulaires, respectivement de cardinal $n_{\mathcal{M}_{int}}$ et $n_{\mathcal{M}_{ext}}$.

Ainsi, on introduit le vecteur $M_{int}(t)$ des concentrations des métabolites intracellulaires et le vecteur $M_{ext}(t)$ des concentrations des métabolites extracellulaires. Soient Ω_{int} et Ω_{ext} les sous-matrices de Ω correspondant respectivement aux lignes des composés intracellulaires et extracellulaires.

En parallèle, pour chaque processus (noté Σ_i) du réseau $\mathcal{R}$, on introduit :

- un flux de valeur $\nu_{M,i}(t)$,
- une machinerie moléculaire $\mathcal{E}_i$ catalysant le processus et pour laquelle on définit :
 - une concentration intracellulaire $E_i(t)$,
 - une efficacité $k_i(t)$ du processus Σ_i ,
 - un coefficient d'encombrement volumique ℓ_{E_i} ,
 - un flux de production de la machinerie $\mathcal{E}_i$, noté $\nu_{E_i}(t)$.

On regroupe les valeurs des flux $\nu_{M,i}(t)$ dans un vecteur $\nu_M(t)$, et dans un vecteur $\nu_E(t)$ pour les flux de production $\nu_{E_i}(t)$. On pose aussi $E(t)$ le vecteur des concentrations $E_i(t)$.

Les concentrations intracellulaires sont en $mmol.g_{CDW}^{-1}$ et les concentrations extracellulaires sont en $mmol.L^{-1}$. Les flux sont exprimés en $mmol.g_{CDW}^{-1}.h^{-1}$. Les efficacités sont en h^{-1} . Les coefficients d'encombrement n'ont pas d'unité mais représentent le volume d'une machinerie $\mathcal{E}_i$ en nombre d'acides aminés équivalents.

Processus de traduction. S'il n'est pas compris dans la représentation $\mathcal{R}$, il est nécessaire d'ajouter un processus de traduction, Σ_R , qui regroupe les sous-processus de production de machineries moléculaires. Ce processus produit tous les flux du vecteur $\nu_E(t)$. On associe à ce processus une machinerie moléculaire $\mathcal{E}_R$ représentant toutes les machineries impliqués dans la production des protéines (ribosomes, facteurs de traduction, ...). Sa concentration est notée $E_R(t)$, son efficacité est k_R , son encombrement est noté ℓ_{E_R} . $\mathcal{E}_R$ est produite par elle-même avec un flux $\nu_{E_R}(t)$. $E_R(t)$ est ajoutée au vecteur $E(t)$ et son flux de production ν_{E_R} est ajouté au vecteur $\nu_E(t)$.

On introduit une nouvelle matrice de stœchiométrie, Ω_E , qui a autant de colonnes qu'il y a de $\mathcal{E}_i$ (y compris $\mathcal{E}_R$, donc $n_\Sigma + 1$) et autant de lignes qu'il y a de métabolites internes *i.e.* $n_{\mathcal{M}_{int}}$. Le coefficient $\Omega_{E,i,j}$ est le nombre de $\mathcal{M}_{int,i}$ nécessaire à la production d'un $\mathcal{E}_j$. Ces coefficients sont négatifs pour les cofacteurs produits lors de la synthèse de $\mathcal{E}_j$.

Bilan de matière. De là, on écrit les bilans de matières pour les métabolites et les machines moléculaires qui catalysent les processus de la manière suivante :

$$\begin{cases} \dot{M}_{int}(t) &= \Omega_{int}\nu_M(t) + \Omega_E\nu_E(t) - \mu(t)M_{int}(t) \\ \dot{E}(t) &= \nu_E(t) - \mu(t)E(t) \\ \dot{M}_{ext}(t) &= \Omega_{ext}\nu_M(t)X(t) \end{cases} \quad (4.1.1)$$

avec $\mu(t)$ le taux spécifique de croissance et $X(t)$ la concentration de biomasse dans le milieu extérieur.

La relation suivante lie ces deux quantités :

$$\dot{X}(t) = \mu(t)X(t) \quad (4.1.2)$$

$\mu(t)$ est exprimé en h^{-1} , l'unité de $X(t)$ est le $g_{CDW}.L^{-1}$.

Contraintes d'efficacité des processus. Les efficacités et les concentrations des machineries limitent les flux qui passent à travers les processus de la manière suivante :

$$\nu_M(t) \leq k(t) \circ E_{\mathcal{E}_R}(t) \quad (4.1.3)$$

où $\circ$ est le produit terme à terme de Hadamard, le vecteur $k(t)$ regroupe les efficacités $k_i(t)$ des processus Σ_i , excepté $\mathcal{E}_R$ et où $E_{\mathcal{E}_R}$ est le vecteur des concentrations des machineries $\mathcal{E}_i$ privé de E_R .

Pour le processus Σ_R et la machinerie $\mathcal{E}_R$, on pose :

$$\sum_{\mathcal{E}} \nu_{E_i} \leq k_R(t)E_R(t) \quad (4.1.4)$$

où la somme porte bien sur tous les éléments de $\mathcal{E}$, y compris $\mathcal{E}_R$, puisqu'ils sont tous produits par Σ_R , par définition.

Les efficacités des processus peuvent éventuellement être formulées sous la forme de fonctions des concentrations des métabolites :

$$\begin{cases} k(t) = k(M(t)) \\ k_R(t) = k_R(M(t)) \end{cases} \quad (4.1.5)$$

C'est ce qui a été fait dans le chapitre 2 lorsque l'on a formulé les efficacités des processus comme des fonctions des concentrations $S(t)$ et $G(t)$.

Contraintes de régulation de composés internes. On complète les contraintes de conception en contraignant certaines concentrations intracellulaires à être constantes au cours du temps. Pour ce faire, comme on l'a fait pour le composé $\mathcal{B}$ au chapitre 2, on contraint la dérivée de certaines concentrations à être nulle.

Soit $\mathcal{M}_c$ l'ensemble des composés dont la concentration intracellulaire est connue pour être constante. Soient $M_c(t)$ le vecteur de ces concentrations, $M_{c,0}$ les concentrations nominales des composés internes constants. La contrainte de régulation de la concentration des composés $\mathcal{M}_c$ s'écrit :

$$\dot{M}_c(t) = \Omega_c \nu_M(t) - \Omega_{E,c} \nu_E(t) - \mu(t) M_{c,0} = 0 \quad (4.1.6)$$

avec Ω_c (respectivement $\Omega_{E,c}$) la sous-matrice de Ω (respectivement, de Ω_E) correspondant aux éléments de $\mathcal{M}_c$ parmi $\mathcal{M}$.

Taux de croissance. Pour le taux de croissance, on reprend l'assertion selon laquelle la densité intracellulaire est constante au cours du temps [49, 50]. Avec les coefficients d'encombrement ℓ_{E_i} , et D_0 la densité intracellulaire supposée connue, on a :

$$D_0 = \ell_E^\top E(t) \quad (4.1.7)$$

avec ℓ_E le vecteur qui regroupe les coefficients d'encombrement ℓ_{E_i} ordonnés comme le vecteur $E(t)$.

En dérivant cette relation par rapport au temps, on obtient :

$$0 = \ell_E^\top \dot{E}(t)$$

et avec (4.1.1),

$$0 = \ell_E^\top (\nu_E(t) - \mu(t) E(t))$$

ou encore,

$$\ell_E^\top \nu_E(t) = \mu(t) \underbrace{\ell_E^\top E(t)}_{=D_0, (4.1.7)}$$

c'est-à-dire,

$$\mu(t) = \frac{1}{D_0} \ell_E^\top \nu_E(t) \quad (4.1.8)$$

On a là un équivalent direct de la formule (2.1.13) qui avait été trouvée au chapitre 2. On constate que les coefficients ℓ_E jouent le même rôle que les coefficients $\frac{1}{\alpha_{E_i}}$ du chapitre 2. En effet, nous avons rapproché les coefficients $\frac{1}{\alpha_{E_i}}$ des coût des $\mathcal{E}_i$ en ressources. Il en est de même pour ℓ_{E_i} .

Contraintes d'expression rationnelle des gènes. Comme avancé au chapitre précédent, une manière de prendre en compte la régulation de l'expression des gènes par le taux de croissance, c'est-à-dire le fait que certains gènes sont constitutifs, est de contraindre les flux de production des machineries moléculaires à suivre une relation de la forme :

$$\nu_{E_i}(t) = \nu_{E_i}(\mu(t)) \quad (4.1.9)$$

Jusqu'alors, nous avons pris ν_{E_i} sous forme polynomiale :

$$\nu_{E_i}(t) = a_{E_i}^\mu \mu(t)^3 + b_{E_i}^\mu \mu(t)^2 + c_{E_i}^\mu \mu(t) \quad (4.1.10)$$

On note $\mathcal{E}_\mu$ l'ensemble des machineries moléculaires qui sont constitutives.

Matriciellement, avec T une matrice rectangulaire convenablement choisie ne comprenant que des 0 et un seul et unique 1 sur chaque ligne, et a_E^μ (respectivement, b_E^μ , c_E^μ) le vecteur des $a_{E_i}^\mu$ (respectivement, $b_{E_i}^\mu$, $c_{E_i}^\mu$), la contrainte (4.1.10) s'écrit :

$$T\nu_E(t) = \mu(t)^3 a_E^\mu + \mu(t)^2 b_E^\mu + \mu(t) c_E^\mu \quad (4.1.11)$$

Remarque sur les processus réversibles. Certains processus sont réversibles. Ceci signifie que la même machinerie peut transformer le substrat en produit et le produit en substrat. Pour inclure ce cas dans la modélisation actuelle sans compliquer davantage le modèle, on peut procéder comme suit.

Soit Σ un processus réversible. On note ν le flux de ce processus, éventuellement négatif. On a le schéma réactionnel suivant : $\mathcal{S} \xrightarrow{\nu} \mathcal{P}$.

Avec k l'efficacité du processus et E la concentration de la machinerie $\mathcal{E}$ qui catalyse le flux, on doit satisfaire à tout instant :

$$\begin{cases} \nu(t) \leq k(t)E(t) \\ -\nu(t) \leq k(t)E(t) \end{cases}$$

On introduit ν^+ et ν^- deux flux positifs tels que : $\mathcal{S} \xrightleftharpoons[\nu^-]{\nu^+} \mathcal{P}$. Ces deux flux sont catalysés par la même machinerie $\mathcal{E}$. Aussi, la somme de ces flux doit être inférieure à $k(t)E(t)$.

On doit donc respecter la relation suivante :

$$\nu^+(t) + \nu^-(t) \leq k(t)E(t)$$

qui est une extension naturelle de la relation (4.1.3).

Par simplicité d'écriture, on considère dans la suite que tous les flux sont positifs, à cette variation près.

Schéma de construction d'un modèle dRBA. La figure 4.1 présente une hiérarchie entre les variables à définir pour poser le problème de conception rationnelle de souche.

Modèle dRBA. La modélisation du réseau $\mathcal{R}$ selon la méthode dRBA est donnée par les équations (4.1.1), (4.1.2), (4.1.3), (4.1.4), (4.1.5), (4.1.6), (4.1.8), (4.1.11). En résumé,

$$\begin{cases} \dot{M}_{int}(t) &= \Omega_{int}\nu_M(t) + \Omega_E\nu_E(t) - \mu(t)M_{int}(t) \\ \dot{E}(t) &= \nu_E(t) - \mu(t)E(t) \\ \dot{M}_{ext}(t) &= \Omega_{ext}\nu_M(t)X(t) \\ \dot{X}(t) &= \mu(t)X(t) \end{cases} \quad (4.1.12)$$

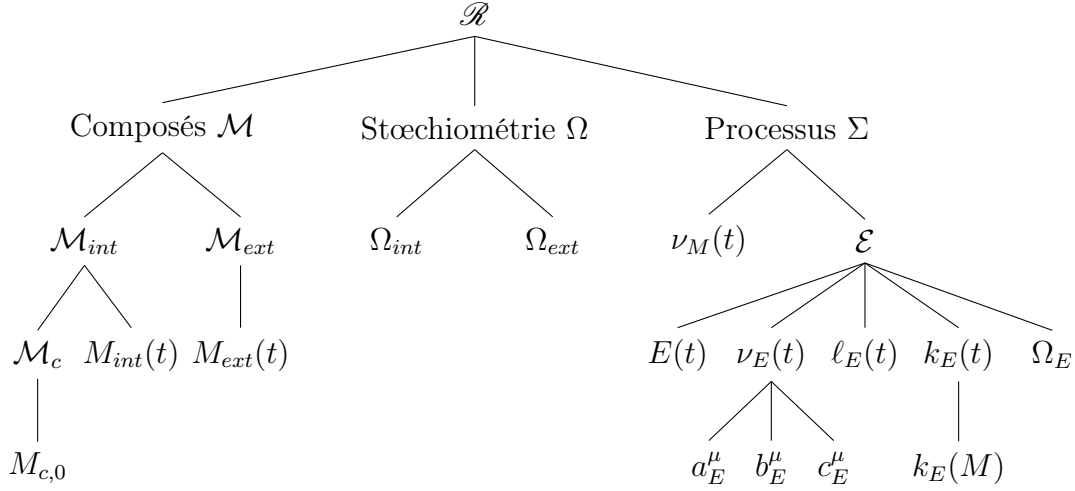


FIGURE 4.1 – Relations entre les variables utilisées dans le modèle.

$$\text{sous contraintes, } \left\{ \begin{array}{l} \nu_M(t) \leq k(t) \circ E_{\setminus \mathcal{E}_R}(t) \\ \sum_{\mathcal{E}} \nu_{E_i} \leq k_R(t) E_R(t) \\ k_i(t) = k_i(M(t)) \\ 0 = \Omega_c \nu_M(t) - \Omega_{E,c} \nu_E(t) - \mu(t) M_{c,0} \\ \mu(t) = \frac{1}{D_0} \ell_E^\top \nu_E(t) \\ T \nu_E(t) = \mu(t)^3 a_E^\mu + \mu(t)^2 b_E^\mu + \mu(t) c_E^\mu \\ \nu_E(t), \nu_M(t) \geq 0 \\ M_{int}(t), E(t), M_{ext}(t), X(t) \geq 0 \end{array} \right. \quad (4.1.13)$$

Comme présenté au chapitre 2, on peut réécrire formellement ce système de la manière suivante :

$$\dot{x}(t) = F(x(t)) \nu(t)$$

$$\text{sous contraintes, } \left\{ \begin{array}{l} A \nu(t) \leq b(x(t)) \\ L \nu(t) = 0 \\ C_{a_E^\mu, b_E^\mu, c_E^\mu}(\nu(t)) = 0 \\ \nu(t) \geq 0 \\ x(t) \geq 0 \end{array} \right. \quad (4.1.14)$$

avec

$$\left\{ \begin{array}{l} \nu(t) = (\nu_E(t)^\top \quad \nu_M(t)^\top)^\top \\ x(t) = (M_{int}^\top(t) \quad E^\top(t) \quad M_{ext}^\top(t) \quad X(t))^\top \end{array} \right.$$

et avec $C_{a_E^\mu, b_E^\mu, c_E^\mu}(\nu(t)) = T \nu_E(t) - (\frac{1}{D_0} \ell_E^\top \nu_E(t))^3 a_E^\mu - (\frac{1}{D_0} \ell_E^\top \nu_E(t))^2 b_E^\mu - (\frac{1}{D_0} \ell_E^\top \nu_E(t)) c_E^\mu$, la contrainte d'expression rationnelle des gènes.

Les contraintes sur la limitation des flux sont regroupées dans $A \nu \leq b(x(t))$, les contraintes sur l'équilibre des concentrations intracellulaires M_c se retrouvent dans $L \nu(t) = 0$. La définition du taux de croissance, $\mu(t) = \frac{1}{D_0} \ell_E^\top \nu_E(t)$, peut se substituer aux occurrences de $\mu(t)$ et n'a donc pas lieu d'apparaître dans la définition du problème.

4.1.2 Optimisation de la production d'un composé d'intérêt

On peut utiliser le modèle (4.1.14) pour optimiser la production d'un composé d'intérêt en respectant le réseau $\mathcal{R}$.

On résout alors un problème du type :

$$\begin{aligned} & \max_{\nu(t), t_f, a_E^\mu, b_E^\mu, c_E^\mu} J \\ & \text{sous contraintes} \quad \left\{ \begin{array}{l} \dot{x}(t) = F(x(t))\nu(t) \\ C_0 x(t_0) = c_0 \\ C_f x(t_f) \geq c_f \\ A\nu(t) \leq b(x(t)) \\ L\nu(t) = 0 \\ C_{a_E^\mu, b_E^\mu, c_E^\mu}(\nu(t)) = 0 \\ \nu(t) \geq 0 \\ x(t) \geq 0 \end{array} \right. \end{aligned} \quad (4.1.15)$$

où C_0 , c_0 , C_f et c_f permettent de définir des contraintes sur l'état initial et final. En particulier, on impose $\ell_E^\top E(t_0) = D_0$ et $M_c(t_0) = M_{c,0}$.

J est le critère à maximiser. Typiquement, on pourra se référer à la discussion 2.3.1.1 et prendre un critère de rendement par rapport au temps :

$$J = \frac{P(t_f)}{t_f}$$

ou tout simplement un critère du type :

$$J = P(t_f)$$

où P est dans ce cas un élément de M_{ext} .

On peut ajouter éventuellement des contraintes sur les paramètres a_E^μ , b_E^μ et c_E^μ . Ainsi, comme au chapitre 2, certains coefficients a_E^μ , b_E^μ et c_E^μ peuvent être fixés à leurs valeurs dans le cas nominal, c'est-à-dire lorsque la configuration du réseau maximise le taux spécifique de croissance.

Dans tous les cas, il s'agit d'un problème de commande optimale avec des contraintes linéaires et non-linéaires sur la commande, et des contraintes sur la commande dépendant de l'état. La résolution de ce type de problème peut se faire comme précédemment par discrétisation et utilisation d'une méthode de type collocation, par exemple à l'aide du logiciel Bocop [10] utilisé précédemment.

4.1.3 Bilan

On a décrit une manière de poser le problème dRBA en mode de fonctionnement batch pour un réseau métabolique $\mathcal{R}$ à condition de pouvoir définir les éléments regroupés figure 4.1.

Nous allons à présent suivre ce protocole pour une représentation légèrement plus élaborée du réseau métabolique que celle présentée au chapitre 2. Nous allons ainsi voir l'influence d'un raffinement de modélisation sur la trajectoire optimale.

4.2 Exemple d'un modèle à deux voies de synthèse de l'énergie

Nous allons nous intéresser à un réseau qui dispose d'une alternative pour produire un composé. Typiquement, cette alternative est présente au niveau de la synthèse de l'énergie chez la plupart des organismes. On pense à la voie de la respiration (qui possède un très bon rendement de conversion, mais qui est coûteuse en protéines et peu efficace) et à la voie du débordement ou de l'*overflow* (dont la conversion de la matière est partielle, mais qui a une bonne efficacité et qui est peu coûteuse en protéines) qui servent toutes les deux à produire de l'ATP à partir d'un composé intermédiaire de la glycolyse : le pyruvate. La voie de l'*overflow* correspond en général à la synthèse de produits comme de l'acétate ou du lactate.

Dans cette section, nous définissons un réseau fictif qui reproduit cette alternative entre deux voies parallèles (section 4.2.1). Nous caractérisons en section 4.2.2 le fonctionnement de ce réseau qui maximise le taux spécifique de croissance. Enfin, en section 4.2.3, nous utilisons ce modèle pour résoudre des problèmes de conception de bioprocédé.

4.2.1 Éléments biologiques & Construction du modèle

Dans cette section, nous définissons le réseau $\mathcal{R}_{2voies}$ associé à la configuration présentée figure 4.1. Nous suivons la méthodologie de la section 4.1 pour représenter ce réseau.

Le réseau $\mathcal{R}_{2voies}$ est une représentation de la cellule qui s'appuie sur le modèle agrégé présenté au chapitre 2. Comme présenté sur la figure 4.2, on affine la description de la voie de synthèse de $\mathcal{S}$ à partir de $\mathcal{G}$. Pour ce faire, on introduit un composé intermédiaire, $\mathcal{Y}$, ainsi que deux processus pour transformer $\mathcal{Y}$ en $\mathcal{S}$: Σ_{CvC} et Σ_{CvP} . Σ_{CvC} est un processus de conversion supposée complète de la matière (**C**onversion **C**omplète). Σ_{CvP} est un processus de conversion partielle (**C**onversion **P**artielle) de la matière donnant lieu également à la production et la sécrétion dans le milieu extracellulaire d'un produit secondaire, $\mathcal{A}$.

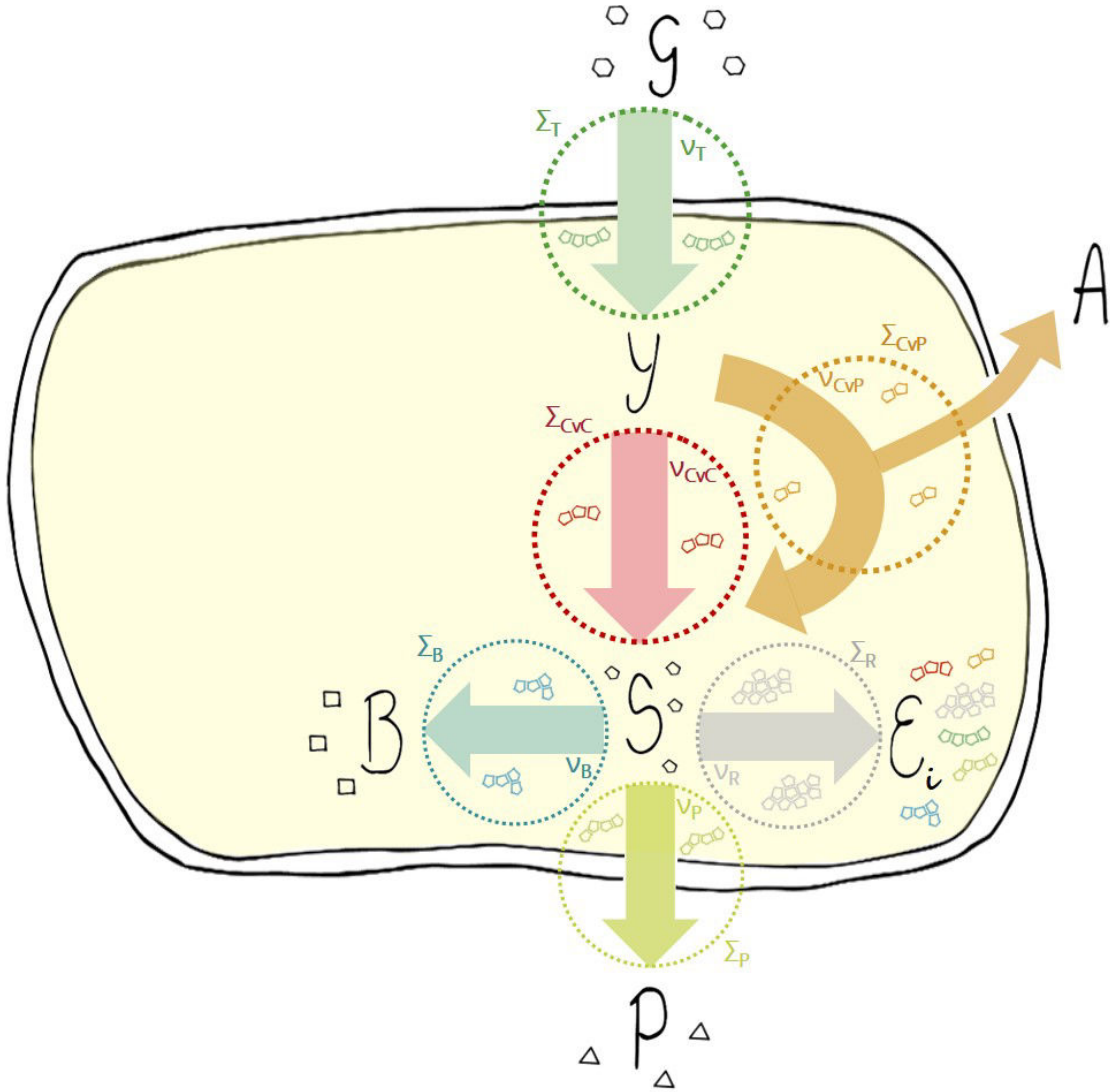
En considérant que $\mathcal{S}$ est le produit du catabolisme, le composé $\mathcal{Y}$ mime le rôle d'un composé intermédiaire du catabolisme. Les deux processus Σ_{CvC} et Σ_{CvP} miment les voies de synthèse de l'énergie dont disposent les organismes pour synthétiser l'énergie, respectivement la voie de la respiration et celle de l'*overflow*.

Avec le formalisme de la section 4.1.1, on définit :

$$\begin{cases} \mathcal{M}_{int} = \{\mathcal{Y}, \mathcal{S}, \mathcal{B}\} \\ \mathcal{M}_{ext} = \{\mathcal{G}, \mathcal{P}, \mathcal{A}\} \end{cases} \quad (4.2.1)$$

avec les vecteurs de concentrations associés :

$$\begin{cases} M_{int}(t) = (Y(t) \quad S(t) \quad B(t))^{\top} \\ M_{ext}(t) = (G(t) \quad P(t) \quad A(t))^{\top} \end{cases} \quad (4.2.2)$$

FIGURE 4.2 – Schéma du réseau $\mathcal{R}_{2voies}$

Pour simplifier la description intracellulaire, on suppose que les concentrations de tous les métabolites intracellulaires sont constantes :

$$\mathcal{M}_c = \mathcal{M}_{int} \quad (4.2.3)$$

Les flux des processus intracellulaires sont regroupés dans le vecteur $\nu_M(t)$ défini comme suit :

$$\nu_M(t) = (\nu_T(t) \quad \nu_B(t) \quad \nu_P(t) \quad \nu_{CvC}(t) \quad \nu_{CvP}(t))^T \quad (4.2.4)$$

La stœchiométrie des processus est donnée par la matrice de stœchiométrie suivante :

$$\Omega = \begin{matrix} & \Sigma_T & \Sigma_B & \Sigma_P & \Sigma_{CvC} & \Sigma_{CvP} \\ \begin{matrix} \mathcal{Y} \\ \mathcal{B} \\ \mathcal{S} \\ \mathcal{G} \\ \mathcal{P} \\ \mathcal{A} \end{matrix} & \begin{pmatrix} \alpha_T & 0 & 0 & -1 & -1 \\ 0 & \alpha_B & 0 & 0 & 0 \\ 0 & -1 & -1 & \alpha_S & \rho\alpha_S \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_P & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_A \end{pmatrix} \end{matrix} \quad (4.2.5)$$

ce qui revient à l'ensemble de schémas réactionnels suivant :

- $\Sigma_T : 1.\mathcal{G} \rightarrow \alpha_T.\mathcal{Y}$
- $\Sigma_B : 1.\mathcal{S} \rightarrow \alpha_B.\mathcal{B}$
- $\Sigma_P : 1.\mathcal{S} \rightarrow \alpha_P.\mathcal{P}$
- $\Sigma_{CvC} : 1.\mathcal{Y} \rightarrow \alpha_S.\mathcal{S}$
- $\Sigma_{CvP} : 1.\mathcal{Y} \rightarrow \rho\alpha_S.\mathcal{S} + \alpha_A.\mathcal{A}$

où les α_i sont des coefficients stœchiométriques. Le paramètre ρ est le rendement du processus Σ_{CvP} pour transformer $\mathcal{Y}$ en $\mathcal{S}$.

La matrice Ω se décompose en

$$\Omega_{int} = \begin{pmatrix} \alpha_T & 0 & 0 & -1 & -1 \\ 0 & \alpha_B & 0 & 0 & 0 \\ 0 & -1 & -1 & \alpha_S & \rho\alpha_S \end{pmatrix} \quad (4.2.6)$$

et

$$\Omega_{ext} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_P & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_A \end{pmatrix} \quad (4.2.7)$$

Comme $\mathcal{M}_c = \mathcal{M}_{int}$, on a également :

$$\Omega_c = \Omega_{int} \quad (4.2.8)$$

Enfin, les machineries moléculaires qui catalysent les processus Σ_i sont notées :

$$\mathcal{E} = \{\mathcal{E}_T, \mathcal{E}_B, \mathcal{E}_P, \mathcal{E}_{CvC}, \mathcal{E}_{CvP}, \mathcal{E}_R\} \quad (4.2.9)$$

dont les concentrations sont regroupées dans le vecteur $E(t)$ défini comme suit :

$$E(t) = (E_T(t) \ E_B(t) \ E_P(t) \ E_{CvC}(t) \ E_{CvP}(t) \ E_R(t))^T \quad (4.2.10)$$

Ces machineries catalysent les flux suivants :

$$\nu_E(t) = (\nu_{E_T}(t) \ \nu_{E_B}(t) \ \nu_{E_P}(t) \ \nu_{E_{CvC}}(t) \ \nu_{E_{CvP}}(t) \ \nu_{E_R}(t))^T \quad (4.2.11)$$

avec comme efficacités

$$k(M_{int}(t)) = (k_T(G(t)) \ k_B \ k_P \ k_{CvC} \ k_{CvP})^T \quad (4.2.12)$$

avec

$$k_T(G(t)) = \frac{v_T G(t)}{G(t) + K_T} \quad (4.2.13)$$

Les autres efficacités sont supposées indépendantes de $\mathcal{M}_{int}$.

L'efficacité k_R est également supposée constante, indépendante de $\mathcal{M}_{int}$.

On suppose que le seul composé qui entre dans la composition des machineries moléculaires est $\mathcal{S}$. Avec α'_{E_i} le coefficient stœchiométrique de la réaction : $\alpha'_{E_i} \cdot \mathcal{S} \xrightarrow{\nu_{E_i}} 1 \cdot \mathcal{E}_i$. La matrice Ω_E introduite au paragraphe 4.1.1 s'écrit :

$$\Omega_E = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -\alpha'_{E_T} & -\alpha'_{E_B} & -\alpha'_{E_P} & -\alpha'_{E_{CvC}} & -\alpha'_{E_{CvP}} & -\alpha'_{E_R} \end{pmatrix} \quad (4.2.14)$$

et on a

$$\Omega_{E,c} = \Omega_E \quad (4.2.15)$$

On va supposer que les encombrements ℓ_{E_i} des machineries $\mathcal{E}_i$ sont donnés par le vecteur ℓ_E :

$$\ell_E = (\ell_{E_T} \quad \ell_{E_B} \quad \ell_{E_P} \quad \ell_{E_{CvC}} \quad \ell_{E_{CvP}} \quad \ell_{E_R})^\top \quad (4.2.16)$$

Ces coefficients ℓ_{E_i} sont supposés être la longueur en acides aminés de $\mathcal{E}_i$. Pour rester dans un cadre de conservation de la matière, on suppose qu'ils sont proportionnels aux coûts en $\mathcal{S}$, c'est-à-dire :

$$\ell_{E_i} \propto \alpha'_{E_i}$$

En effet, en faisant le bilan de masse de l'équation $\alpha'_{E_i} \cdot \mathcal{S} \xrightarrow{\nu_{E_i}} 1 \cdot \mathcal{E}_i$, il vient :

$$\alpha'_{E_i} M_{\mathcal{S}} = 1 M_{\mathcal{E}_i}$$

et avec ℓ_{E_i} la longueur en acides aminés de $\mathcal{E}_i$,

$$\alpha'_{E_i} M_{\mathcal{S}} = \ell_{E_i} M_{AA}$$

avec $M_{\mathcal{S}}$ (respectivement, $M_{\mathcal{E}_i}$ et M_{AA}) la masse molaire du composé $\mathcal{S}$ (respectivement, de $\mathcal{E}_i$ et d'un acide aminé).

On constate que la définition de α_{E_i} donnée équation (2.1.10) du chapitre 2 est différente de la définition que l'on donne ici des α'_{E_i} . Plus précisément, en considérant que le composé $\mathcal{S}$ est le même dans les deux modèles, $\alpha'_{E_i} = \frac{1}{\alpha_{E_i}}$

On prend les valeurs numériques du tableau 4.1. Ces valeurs sont reprises du modèle agrégé. Les valeurs Y_0 et S_0 sont fixées à 0, dans le but de négliger leur influence sur le modèle.

Comme dans la bactérie, on introduit une différence entre les deux processus Σ_{CvC} et Σ_{CvP} en postulant que le processus correspondant à la respiration (Σ_{CvC}) a besoin de plus de machinerie pour produire le même flux que le processus mimant l'*overflow*

TABLE 4.1 – Valeurs numériques.

Paramètre	Valeur	(unité)	Paramètre	Valeur	(unité)
v_T	420	(h^{-1})	α_T	0,5	
k_B	750	(h^{-1})	α_B	0,35	
k_P	750	(h^{-1})	α_P	1	
k_{CvC}	600	(h^{-1})	α_S	1	
k_{CvP}	5000	(h^{-1})	ρ	0,8	
k_R	20000	(h^{-1})	α_A	0,2	
K_T	$5,0 \cdot 10^{-3}$	$(mmol.L^{-1})$			
ℓ_T	180		α'_{ET}	64,9	
ℓ_B	360		α'_{EB}	129,7	
ℓ_P	360		α'_{EP}	129,7	
ℓ_{CvC}	180		α'_{ECvC}	64,9	
ℓ_{CvP}	60		α'_{ECvP}	21,6	
ℓ_R	10100		α'_{ER}	3639,4	
D_0	1,33	$(mmol.g_{CDW}^{-1})$	B_0	387	$(\mu mol.g_{CDW}^{-1})$

(Σ_{CvP}) : l'efficacité de la machinerie du processus de respiration est plus faible que celle du processus d'*overflow*. Le composé $\mathcal{A}$ est supposé être rejeté dans le milieu extracellulaire. Par conservation de la masse, le rendement du processus Σ_{CvP} est moins bon que celui Σ_{CvC} .

Pour mimer l'effet de l'alternative entre les voies de respiration et de l'*overflow*, on impose une efficacité, k_i , plus importante et un coût en machineries moléculaires, α'_{E_i} et ℓ_i , plus faible pour le processus Σ_{CvP} que pour le processus Σ_{CvC} .

4.2.2 Caractérisation de l'état nominal du modèle

On caractérise ce modèle dans la situation où le taux spécifique de croissance est maximisé. L'idée est de comprendre les interactions entre les processus et de voir de quelle manière la ressource se répartit entre les machineries moléculaires.

Le but est aussi de caractériser les coefficients a_E^μ , b_E^μ et c_E^μ , de la contrainte (4.1.11).

Comme dans la section 2.2.2 du chapitre 2, on cherche la configuration du réseau qui maximise le taux spécifique de croissance en conditions extracellulaires stationnaires, c'est-à-dire pour des concentrations de substrat constantes valant G_0 . On résout un ensemble de problèmes (Π_{G_0}) qui s'écrivent de la manière suivante :

$$(\Pi_{G_0}) : \left\{ \begin{array}{l} \max_{(\bar{x}, \bar{\nu}) \in \mathbb{R}^{n \times m}} \mu = \frac{1}{D_0} (1 \ 1 \ 1 \ 1 \ 1 \ 1) \bar{\nu}_E \\ \text{sous contraintes} \left\{ \begin{array}{l} F(\bar{x}) \bar{\nu} = 0 \\ A \bar{\nu} \leq b(\bar{x}) \\ L \bar{\nu} = 0 \\ \bar{x} \geq 0 \\ \bar{\nu} \geq 0 \\ \bar{G} = G_0 \end{array} \right. \end{array} \right. \quad (4.2.17)$$

où l'on cherche les concentrations à l'équilibre, $\bar{x}$, et les flux à l'équilibre, $\bar{\nu}$, qui maximisent le taux spécifique de croissance.

De la même manière qu'au chapitre 2, on résout numériquement un ensemble de problèmes (Π_{G_0}) pour des valeurs G_0 variant de 0 à $0,2 \text{ mmol.L}^{-1}$, avec un pas de $0,01 \text{ mmol.L}^{-1}$. La figure 4.3 représente le taux de croissance maximal atteignable pour une concentration extérieure $G = G_0$.

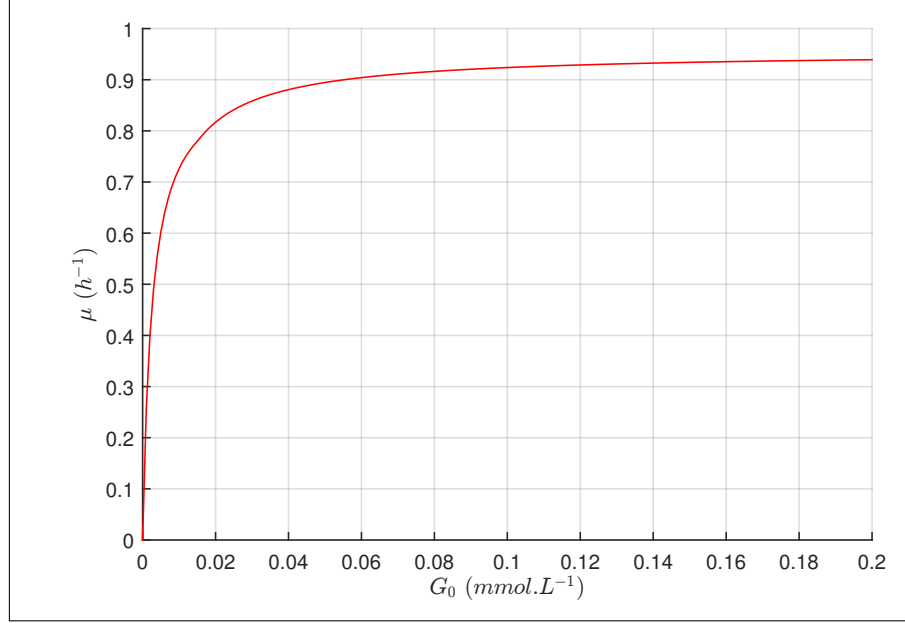


FIGURE 4.3 – Taux spécifique de croissance maximal μ en fonction de la concentration extracellulaire en $\mathcal{G}$. Solution numérique du problème (4.2.17) pour différentes valeurs de G_0 .

L'allure est proche de celle présentée en figure 2.3 pour le modèle agrégé. On retrouve une allure hyperbolique, croissante et majorée, comme c'était le cas avec le modèle agrégé.

On souhaite examiner si cette allure suit une relation de Monod, $\mu = \frac{vG_0}{G_0 + K}$ avec des paramètres v et K identifiables. Pour $G_0 > 0$, montrons qu'il est équivalent de dire :

- (i) $\mu(G_0)$ suit une relation de Monod
- (ii) la dérivée de $1/\mu$ par rapport à $1/G_0$ existe et est constante

Si (i), i.e. $\mu(G_0) = \frac{vG_0}{G_0 + K}$, alors

$$\frac{1}{\mu(G_0)} = \frac{G_0 + K}{vG_0} = \frac{1}{v} + \frac{K}{v} \frac{1}{G_0}$$

Donc, $\frac{d\frac{1}{\mu}}{d\frac{1}{G_0}}$ existe et vaut :

$$\frac{d\frac{1}{\mu}}{d\frac{1}{G_0}}(G_0) = \frac{K}{v}$$

indépendant de G_0 . D'où (ii).

Par ailleurs, si (ii), *i.e.*

$$\frac{d\frac{1}{\mu}}{d\frac{1}{G_0}}(G_0) = C, \text{ constante}$$

alors, on peut intégrer entre G_a et G_0 :

$$\frac{1}{\mu(G_0)} - \frac{1}{\mu(G_a)} = \int_{G_a}^{G_0} C d\frac{1}{G} = C \frac{1}{G_0} - C \frac{1}{G_a}$$

i.e.

$$\frac{1}{\mu(G_0)} = C \frac{1}{G_0} - C'$$

avec C' une constante d'intégration. En prenant l'inverse de cette relation,

$$\mu(G_0) = \frac{1}{C \frac{1}{G_0} + C'} = \frac{G_0}{C + C'G_0} = \frac{G_0/C'}{C/C' + G_0} \quad (4.2.18)$$

c'est-à-dire que μ suit une relation de Monod. $\square$

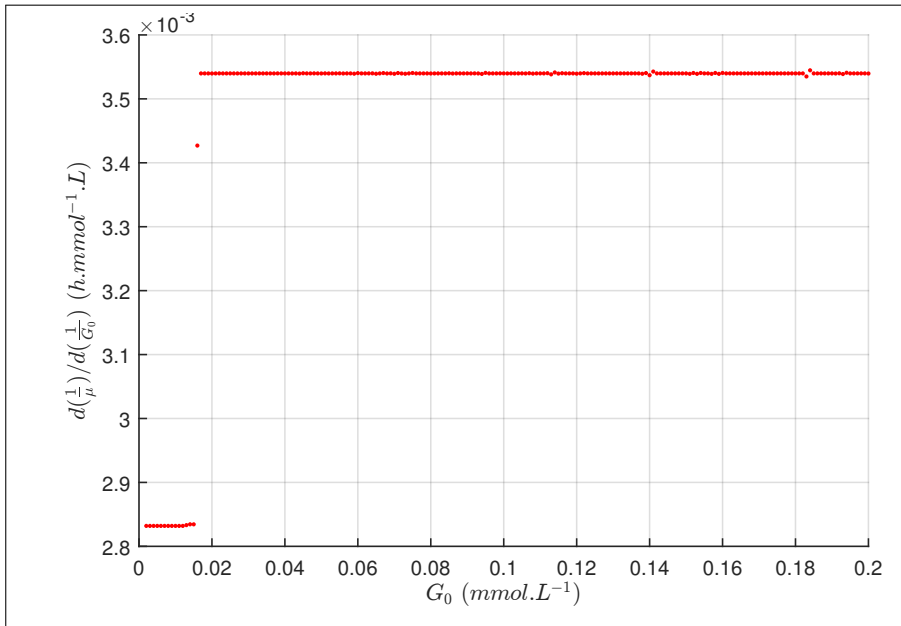


FIGURE 4.4 – Dérivée $\frac{d\frac{1}{\mu}}{d\frac{1}{G_0}}$ tracée en fonction de G_0 pour la solution aux problèmes (Π_{G_0}) (4.2.17).

Dans le cas du taux spécifique de croissance de la figure 4.3, on trace la dérivée $\frac{d\frac{1}{\mu}}{d\frac{1}{G_0}}$ en fonction de G_0 sur la figure 4.4. On obtient deux paliers correspondant à deux régimes. Ainsi, l'allure de la figure 4.3 est en réalité la succession de deux allures de Monod que

l'on peut écrire rigoureusement :

$$\mu(G_0) = \begin{cases} \frac{v_1 G_0}{K_1 + G_0}, & \text{si } G \leq G_s \\ \frac{v_2 G_0}{K_2 + G_0}, & \text{sinon} \end{cases} \quad (4.2.19)$$

avec v_1 , v_2 , K_1 , K_2 et G_s des paramètres intrinsèques au réseau $\mathcal{R}_{2\text{voies}}$, qu'il serait possible d'identifier.

L'évolution des concentrations des machineries moléculaires $\mathcal{E}$ en fonction du taux de croissance présente également deux régimes (*c.f.* 4.5). A faible taux spécifique de croissance (et faible concentration G_0), le processus coûteux mais à fort rendement, Σ_{CvC} , est actif. Par contre, à taux de croissance (ou concentration G_0) élevé, c'est le processus Σ_{CvP} qui est utilisé pour synthétiser. Dans les deux cas, l'autre processus est totalement inactif.

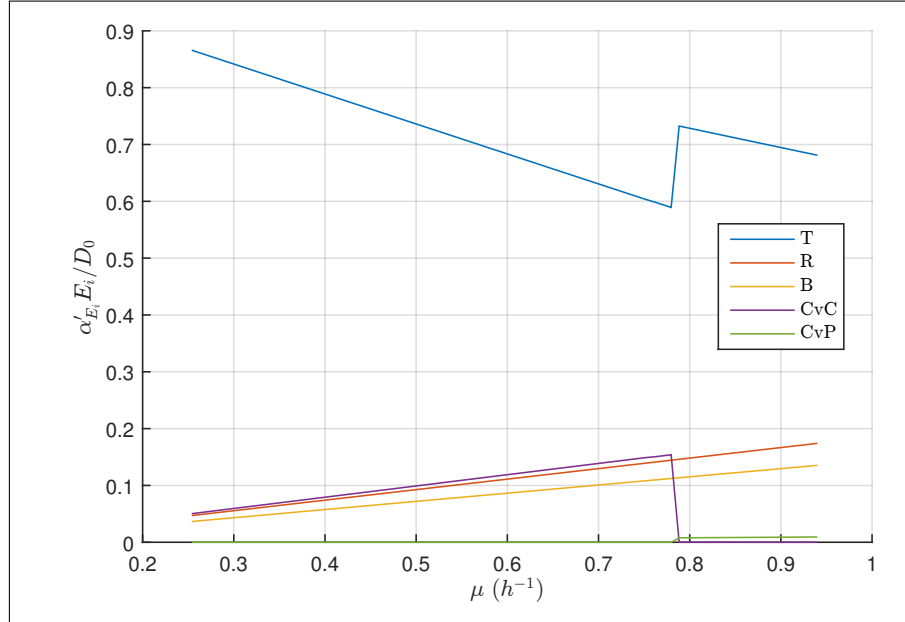


FIGURE 4.5 – Répartition des $\mathcal{E}_i$ en fonction du taux de croissance pour la solution au problème (4.2.17).

Ce changement de régime est en accord avec le fonctionnement observé des cellules [33]. En effet, les cellules utilisent les processus d'*overflow* lors de phases d'abondance de ressources. Ces processus d'*overflow* leur permettent un gain de taux de croissance instantané. Dans la plupart des cas, les processus d'*overflow* ne sont pas utilisés quand le substrat est limitant. Le compromis trouvé par les cellules pour maximiser leur taux de croissance dépend donc de l'abondance en ressources. Le coût des ressources varie en fonction de leur concentration extracellulaire : lorsqu'il y a peu de ressources, leur coût est élevé et lorsqu'il y en a beaucoup, leur coût est plus faible. En abondance de ressources, il est préférable d'utiliser une voie qui permet de gagner de la puissance de croissance, même si le taux de conversion en biomasse est moins bon. En revanche, lorsqu'il y a peu de ressources, le coût d'import est si élevé qu'il est préférable de ne pas

faire de compromis sur le taux de conversion pour maximiser la croissance.

Pour caractériser le changement de régime, on trace sur la figure 4.6 en fonction de la concentration G_0 les taux spécifiques de croissance maximaux solutions du problème (4.2.17) avec la contrainte additionnelle $E_{CvC} = 0$ (courbe bleue), et avec la contrainte additionnelle $E_{CvP} = 0$ (courbe rouge). Le cas bleu (respectivement rouge) correspond à forcer le réseau à utiliser uniquement le processus Σ_{CvP} (respectivement Σ_{CvC}) pour synthétiser $\mathcal{S}$. Les deux courbes s'intersectent en $(G_0; \mu) = (0,015 \text{ mmol.L}^{-1}; 0,78 \text{ h}^{-1})$. Ces valeurs sont celles des changements de régime des figures 4.4 et 4.5.

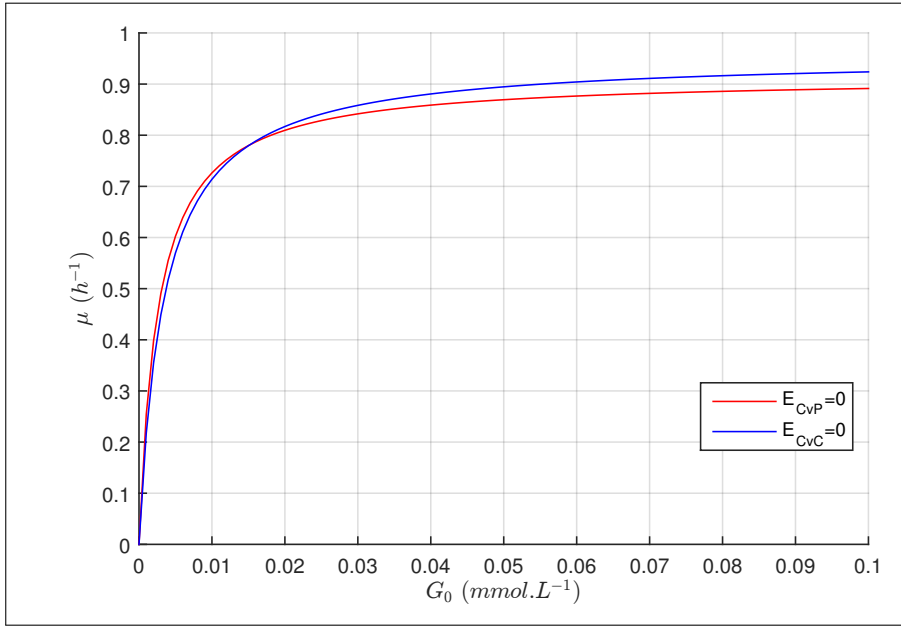


FIGURE 4.6 – Taux de croissance maximal μ en fonction de la concentration extracellulaire en $\mathcal{G}$ pour le cas où l'on impose $E_{CvC} = 0$ et le cas où $E_{CvP} = 0$.

A faible concentration de substrat, le processus Σ_{CvC} est plus efficace pour croître que le processus Σ_{CvP} . A forte concentration de substrat, c'est l'inverse.

Observations analogues. Le comportement retrouvé ici à l'aide de considérations physico-biologiques uniquement peut être rapproché d'observations semblables de la littérature.

En effet, on retrouve cette prise en compte du fait que les cellules utilisent des voies métaboliques moins rentables mais plus efficaces, par exemple dans les travaux [36]. Dans ce modèle macroscopique phénoménologique, les cellules sont supposées utiliser différents régimes (régime oxydatif et régime oxydo-fermentatif). En fonction des conditions extérieures, la croissance est donnée par une combinaison de ces régimes. D'une manière équivalente, lorsqu'il y a peu de ressources, la croissance se fait à fort taux de conversion tandis qu'il y a production d'acétate en abondance de ressources.

Pour aller plus loin, cet aspect peut être rapproché de phénomènes qui mènent les cellules à reconfigurer un très grand nombre de gènes en fonction de la valeur du flux de glycolyse. On pense aux phénomènes de répression catabolique [60] ou d'effet Warburg [78]. Pour ne citer que deux effets, chez *Bacillus subtilis*, cela conduit à réprimer le cycle de Krebs et à activer les enzymes produisant de l'acétate.

4.2.3 Utilisation du modèle pour la production de composé d'intérêt

Nous allons à présent utiliser le modèle du réseau $\mathcal{R}_{2voies}$ développé en section 4.2.1 pour optimiser la production du composé d'intérêt $\mathcal{P}$.

Comme cela a été fait aux chapitres 2 et 3 pour le modèle agrégé, nous allons envisager plusieurs cas correspondant à des degrés différents de conduite du procédé. Pour caractériser un optimum, nous envisageons de contrôler en temps réel tous les processus intracellulaires (cas n° 1). Cette caractérisation est faite en mode de fonctionnement batch. Puis, à l'aide de l'étude menée en section 4.2.2, nous réduisons le degré d'interventions sur le procédé en fixant le comportement de certains processus à leur état nominal (cas n° 2). Ce cas d'étude sera aussi fait en mode de fonctionnement batch mais nous supposons pouvoir contrôler en temps réel le flux qui passe à travers le processus Σ_T . Pour finir, nous reprendrons le cas n° 2 en mode de fonctionnement fed-batch en supposant ne pouvoir agir sur le flux ν_T qu'à travers la concentration en $\mathcal{G}$ (contrôlable par le débit d'alimentation du procédé). Ceci constituera le cas n° 3.

4.2.3.1 Optimisation de la bioproduction en supposant que tous les processus sont contrôlables (Cas n° 1)

On résout un problème du type (4.1.15) où $\nu(t)$, $x(t)$, F , A , b et L sont définis d'après la procédure de la section 4.1 avec le modèle du réseau $\mathcal{R}_{2voies}$ introduit à la section 4.2.1.

Pour les conditions initiales, on suppose :

$$\begin{cases} P(t_0) = A(t_0) = 0 \text{ mmol.L}^{-1} \\ G(t_0) = 0,1 \text{ mmol.L}^{-1} \\ \ell_E^\top E(t_0) = D_0 \\ M_c(t_0) = M_{c,0} \end{cases} \quad (4.2.20)$$

Ceci permet de définir les paramètres C_0 et c_0 de la formulation (4.1.15).

On ne spécifie pas de condition finale particulière.

A ce stade, on ne considère pas non plus de contrainte d'implémentation $C_{a_E^\mu, b_E^\mu, c_E^\mu}$.

En résolvant ce problème avec Bocop [10] avec 100 pas d'échantillonnages et une discrétisation implicite d'Euler, et en initialisant depuis plusieurs points où toutes les concentrations sont constantes, on trouve la solution présentée figure 4.7.

Comme pour la solution optimale du modèle agrégé du chapitre 2, on retrouve trois phases :

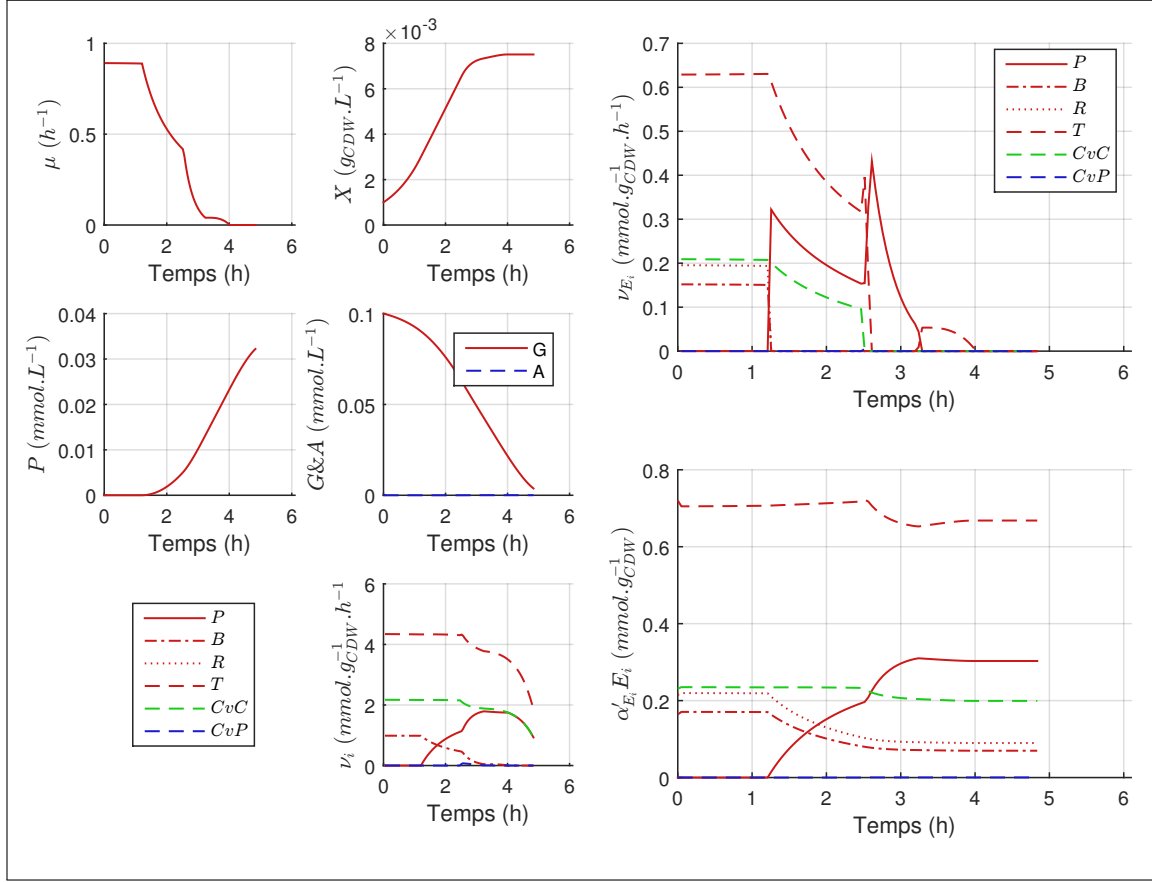


FIGURE 4.7 – Trajectoire optimale, solution du problème de conception de souche de la section 4.2.3.1.

- (i) De $t = 0$ h à $t = 1,21$ h, les cellules sont en régime de croissance exponentielle. La valeur du taux spécifique de croissance correspond à la valeur de la courbe $\mu(G_0 = G(t_0))$ dans le cas $E_{CvP} = 0$, figure 4.6. Les concentrations E sont constantes.
- (ii) De $t = 1,21$ h à $t = 4,01$ h, le taux de croissance diminue et la concentration de la protéine $\mathcal{E}_P$ augmente au détriment de celles des autres $\mathcal{E}$. La répartition intracellulaire des $\mathcal{E}_i$ change séquentiellement. D'abord, les concentrations de $\mathcal{E}_R$ et $\mathcal{E}_B$ diminuent, puis celles de $\mathcal{E}_{CvC}$ et $\mathcal{E}_T$ diminuent également. La production de $\mathcal{P}$ commence. Dans cette phase, la cellule passe d'une configuration de croissance sans production à une configuration de production sans croissance.
- (iii) De $t = 4,01$ h à $t = 4,84$ h, la croissance est nulle et la cellule est dans une configuration stationnaire qui permet de transformer tout le substrat $\mathcal{G}$ restant en produit d'intérêt $\mathcal{P}$.

Le taux spécifique de croissance dans la phase (i) n'est pas le taux maximal que la cellule peut atteindre à la concentration $G_0 = 0,1$ mmol.L⁻¹ (figure 4.3). On a $\mu = 0,89$ h⁻¹ au lieu de $\mu = 0,92$ h⁻¹. La répartition des machineries $\mathcal{E}_i$ n'est pas non plus celle de la cellule à ce taux de croissance à l'état nominal (figure 4.5). En effet, les cellules utilisent le processus Σ_{CvC} au lieu du processus Σ_{CvP} comme elles sont censés le faire à cette concentration G pour maximiser leur croissance. On constate que le processus

Σ_{CvP} n'est pas du tout utilisé dans ce contexte. Le composé $\mathcal{S}$ est exclusivement produit à partir du processus Σ_{CvC} .

Ceci est à la fois contraire au comportement du cas nominal et également à l'idée que l'on s'était faite de la phase de croissance avec le modèle agrégé : la phase de croissance ne correspond pas tout à fait à une phase de croissance maximale. Il semblerait toutefois que le choix du processus de conversion complète Σ_{CvC} soit privilégié car il permet un meilleur compromis entre conservation de la masse et vitesse de croissance.

Pour suivre la trajectoire présentée figure 4.7, il est nécessaire de contrôler en temps réel les processus intracellulaires. Ce contrôle semble difficile à mettre en place. C'est pourquoi on se dirige vers la prise en compte de contraintes d'implémentation, comme au chapitre 3, pour établir une trajectoire optimale réaliste en termes de conduite du procédé.

4.2.3.2 Optimisation réaliste de la bioproduction d'un composé d'intérêt (Cas n° 2)

On désire à présent étudier un mode de fonctionnement réaliste pour la maximisation de la bioproduction d'un composé d'intérêt. Par "réaliste", on entend trouver une trajectoire qu'il soit possible de suivre sans avoir à reconfigurer entièrement la cellule, ni avoir à contrôler en temps réel chaque processus intracellulaire.

On propose de ne pas reconfigurer la partie catabolique de la cellule, *i.e.* les processus Σ_T , Σ_{CvP} et Σ_{CvC} . Ainsi, on suppose que les flux de production des composés $\mathcal{E}_T$, $\mathcal{E}_{CvP}$ et $\mathcal{E}_{CvC}$ sont donnés par la configuration nominale établie en section 4.2.2, figure 4.5.

Pour prendre en compte le contrôle de l'expression des gènes $\mathcal{E}_T$, $\mathcal{E}_{CvP}$ et $\mathcal{E}_{CvC}$ par le taux spécifique de croissance μ , on introduit trois relations (4.1.9) pour les trois entités $\mathcal{E}_T$, $\mathcal{E}_{CvP}$ et $\mathcal{E}_{CvC}$. Par rapport aux allures de la figure 4.5, il semble raisonnable de paramétrer les flux de production de ces machineries, de la manière suivante :

$$\nu_{E_i} = \begin{cases} \mu(t)(a_{E_i,1}^\mu \mu(t)^2 + b_{E_i,1}^\mu \mu(t) + c_{E_i,1}^\mu), & \text{si } \mu \leq \mu_s \\ \mu(t)(a_{E_i,2}^\mu \mu(t)^2 + b_{E_i,2}^\mu \mu(t) + c_{E_i,2}^\mu), & \text{sinon} \end{cases} \quad (4.2.21)$$

où μ_s est la valeur du taux de croissance à laquelle on observe un changement de mode de fonctionnement. On prendra $\mu_s = 0,78 \text{ h}^{-1}$.

On identifie par une méthode des moindres carrés, les flux de production des machineries $\mathcal{E}_T$, $\mathcal{E}_{CvP}$ et $\mathcal{E}_{CvC}$ en fonction du taux spécifique de croissance sur la figure 4.5 :

$$\text{pour } \mu \leq 0,78 \text{ h}^{-1}, \begin{cases} \nu_{E_T}(\mu) = \mu(-0,70\mu + 1,33) \\ \nu_{E_{CvC}}(\mu) = \mu(0,26\mu + 0) \\ \nu_{E_{CvP}}(\mu) = 0 \end{cases} \quad (4.2.22)$$

$$\text{pour } \mu > 0,78 \text{ h}^{-1}, \begin{cases} \nu_{E_T}(\mu) = \mu(-0,45\mu + 1,33) \\ \nu_{E_{CvC}}(\mu) = 0 \\ \nu_{E_{CvP}}(\mu) = \mu(0,01\mu + 0) \end{cases} \quad (4.2.23)$$

Par ailleurs, on suppose ne pas agir sur les processus Σ_{CvC} et Σ_{CvP} . On force donc leurs flux à valoir leurs bornes maximales, c'est-à-dire :

$$\begin{cases} \nu_{CvP}(t) = k_{CvP}E_{CvP}(t) \\ \nu_{CvC}(t) = k_{CvC}E_{CvC}(t) \end{cases} \quad (4.2.24)$$

On laisse libre $\nu_T(t) \leq k_T(G(t))E_T(t)$ pour une raison sur laquelle nous reviendrons en section 4.2.3.3.

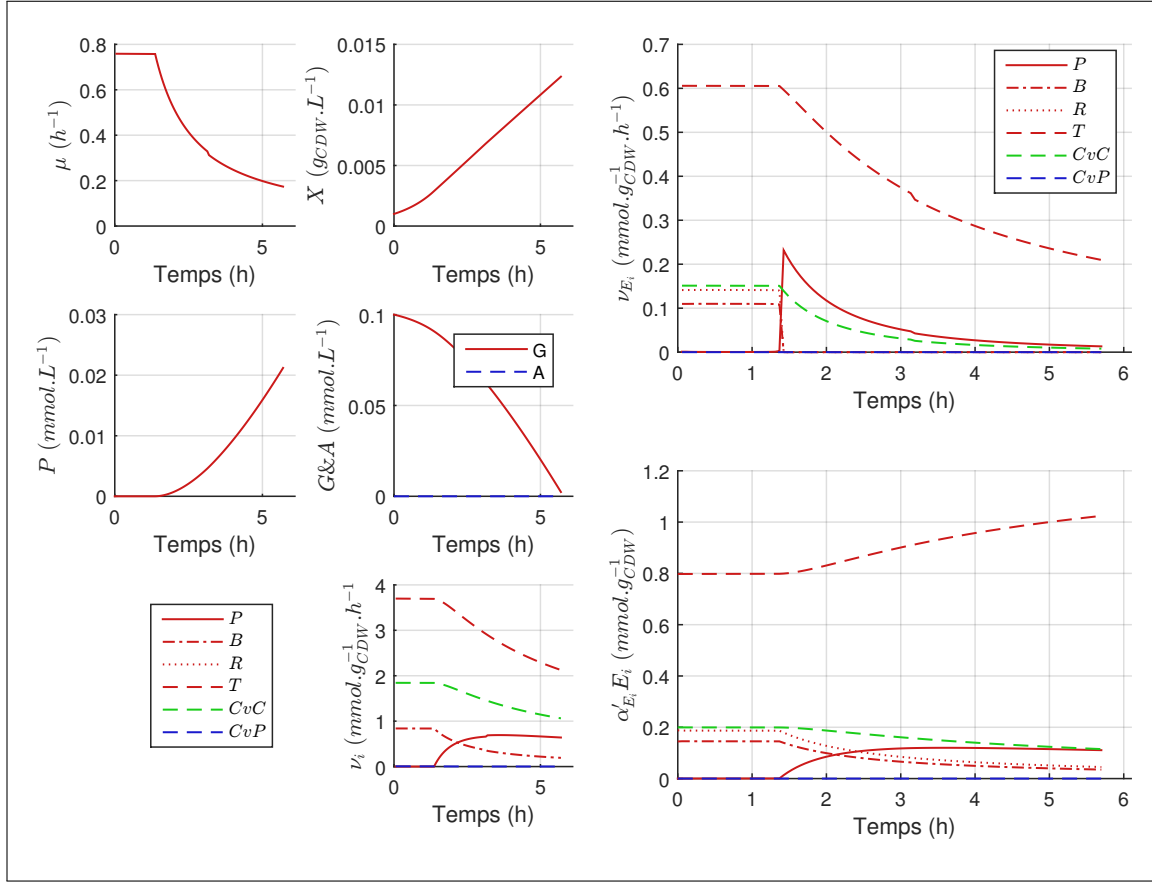


FIGURE 4.8 – Trajectoire optimale, solution du problème de conception de souche de la section 4.2.3.2.

Le cas n° 2 correspond au cas n° 1 complété des contraintes (4.2.22), (4.2.23), (4.2.24). La trajectoire qui optimise ce problème est calculée dans les mêmes conditions que pour le cas n° 1 et est présentée figure 4.8. Les flux de production de protéines ν_E sont présentés en fonction du taux spécifique de croissance, figure 4.9.

Cette trajectoire ne présente plus que deux phases :

- de $t = 0$ h à $t = 1,37$ h, une phase de croissance exponentielle à une valeur de taux spécifique de croissance $\mu = 0,77$ h⁻¹. La répartition des protéines correspond à l'état nominal donné par la figure 4.5.

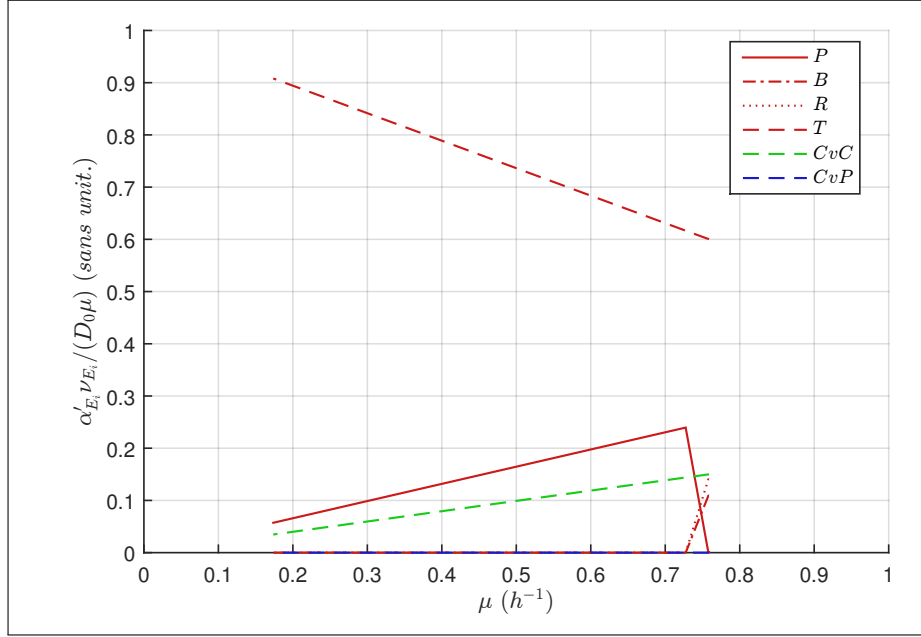


FIGURE 4.9 – Flux de production des protéines ν_{E_i} en fonction du taux spécifique de croissance μ pour la solution du problème de conception de souche de la section 4.2.3.2.

- de $t = 1,37 \text{ h}$ à $t = 5,70 \text{ h}$, une phase qui ressemble à la phase de transition de la section 4.2.3.1. La croissance diminue exponentiellement. La répartition intracellulaire des machineries $\mathcal{E}_i$ change. La production de $\mathcal{E}_B$ et $\mathcal{E}_R$ est remplacée par la production de $\mathcal{E}_P$.

Comme détaillé en section 4.2.3.4, la production de $\mathcal{P}$ est plus faible dans ce cas que dans le cas n° 1.

On se demande comment mettre en œuvre cette stratégie sur un bioprocédé réel. La répartition des protéines étant en accord avec le comportement nominal dans la phase de croissance, il n'y a pas de modification à faire dans ce sens.

L'allure de ν_{E_P} en fonction du taux spécifique de croissance est affine croissante, modulo une activation à l'instant $t_s = 1,37 \text{ h}$. Cette allure est réalisable. Il suffit de trouver un moyen d'activer le gène à $t = 1,37 \text{ h}$.

On peut vérifier que le processus Σ_R est toujours saturé, *i.e.* $\nu_R(t) = k_R E_R(t)$. Ainsi, il n'y a pas à agir sur l'efficacité des ribosomes. La seule mesure à prendre est de couper leur production à $t_s = 1,37 \text{ h}$.

Il en est de même pour le processus Σ_B .

En revanche, le flux du processus Σ_T n'est jamais saturé : il va falloir agir sur le procédé à tout instant. La section 4.2.3.3 présente une manière raisonnable de s'y prendre en mode de fonctionnement fed-batch.

4.2.3.3 Conduite du procédé en fed-batch (Cas n° 3)

Dans notre modélisation, on a supposé que l'efficacité du processus Σ_T dépendait de $G(t)$, selon la relation (4.2.13). Le flux de ce processus est donc contrôlable par la concentration $G(t)$. On va ainsi contrôler le flux ν_T en se plaçant en mode de fonctionnement

fed-batch.

On contraint $\nu_T(t)$ à valoir sa borne maximale :

$$\nu_T(t) = k_T(G(t))E_T(t)$$

avec $G(t)$ contrôlée par le débit d'alimentation, *i.e.* par conduite du procédé. On cherche la manière optimale d'approvisionner le réacteur en substrat.

On adapte le problème de la section 4.2.3.2 pour le mode de fonctionnement fed-batch. En reprenant le raisonnement de la section 3.3, on définit le débit d'alimentation du bioréacteur, $F_{in}(t)$, et le volume du réacteur $V(t)$. Pour les composés extracellulaires, on ajoute un flux de dilution relatif à l'augmentation du volume du réacteur.

On pose $\mathcal{C}_G$ la quantité de $\mathcal{G}$ consommée pendant la culture, définie selon (3.3.11) :

$$\mathcal{C}_G = G(t_0)V(t_0) + \int_{t_0}^{t_f} G_F F_{in}(t) dt \quad (4.2.25)$$

où G_F est la concentration en $\mathcal{G}$ du milieu d'alimentation.

On veut comparer la stratégie fed-batch avec la stratégie de la section 4.2.3.2, c'est-à-dire en batch où tout le substrat $\mathcal{G}$ est introduit à t_0 dans le réacteur.

On fixe $V(t_0) = 10L$. Ainsi, la quantité de $\mathcal{G}$ consommée par le procédé batch vaut :

$$\mathcal{C}_{G,batch} = G(t_0)V(t_0) + \int 0 dt = 1 \text{ mmol} \quad (4.2.26)$$

On ajoute alors la contrainte $\mathcal{C}_G = 1 \text{ mmol}$ pour l'optimisation fed-batch. On laisse ainsi la possibilité à l'optimisation de retrouver le profil batch.

On donne aussi la concentration G_F :

$$G_F = 1 \text{ mmol.L}^{-1}$$

La concentration initiale $G(t_0)$ est laissée libre.

Avec ces conditions, la solution au problème de maximisation du critère $J = \frac{P(t_f)}{t_f}$ est donnée figure 4.10 avec un rappel du cas n° 2 en gris. La démarche de modélisation et de résolution du problème est similaire à celle des cas n° 1 et 2. L'algorithme d'optimisation est initialisé avec un débit nul.

Pour les états intracellulaires, la trajectoire optimale du cas n° 3 se confond avec la trajectoire optimale de l'optimisation batch avec action sur le processus Σ_T , cas n° 2. La différence porte sur les évolutions des concentrations extracellulaires et en particulier sur le profil $G(t)$. Étant donné que les comportements intracellulaires sont exactement les mêmes, les différences entre les profils $P(t)$ et $X(t)$ sont dues à l'augmentation du volume qui fait baisser les concentrations extracellulaires.

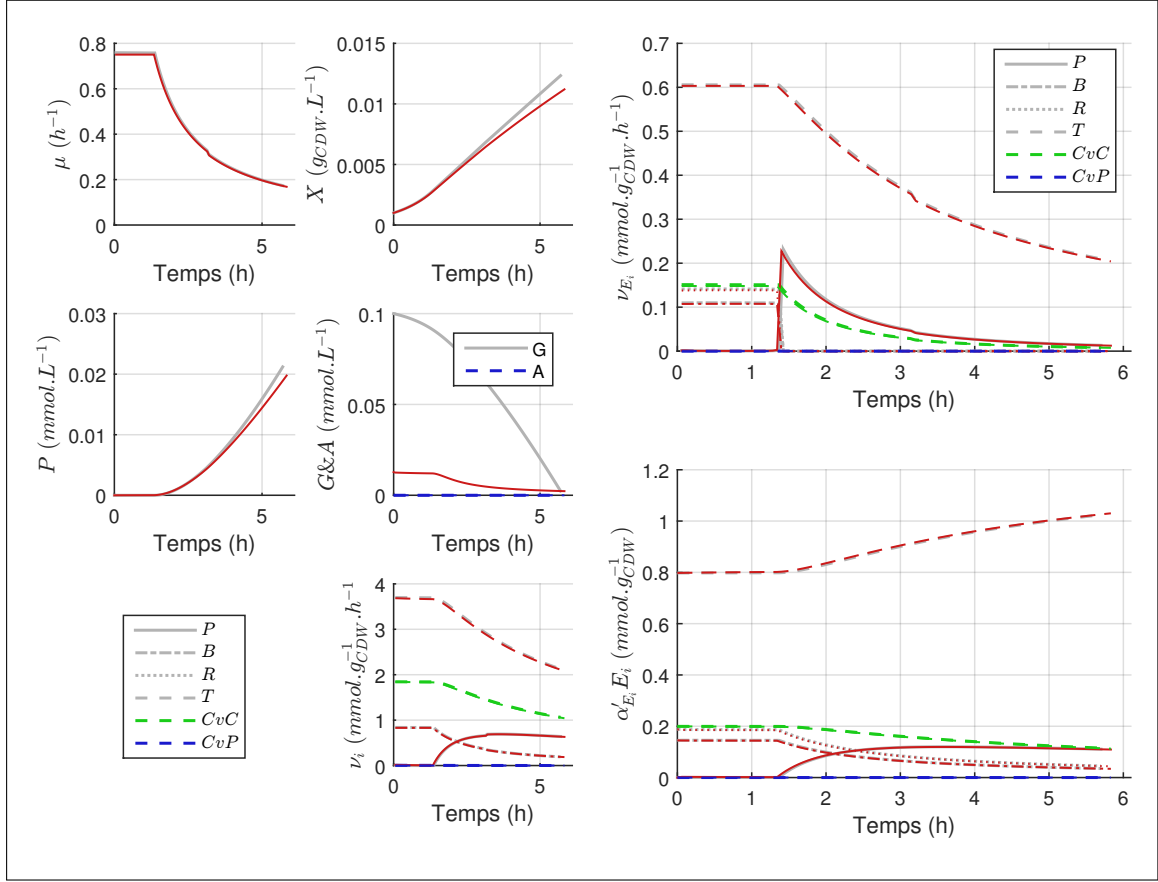


FIGURE 4.10 – En rouge, trajectoire optimale, solution cas n° 3. En gris, rappel du cas n° 2.

Le profil $G(t)$ se décompose en deux phases :

- Dans un premier temps, G est quasiment constante à une valeur qui correspond à la valeur du taux de croissance $\mu = 0,78 \text{ h}^{-1}$ à l'état nominal (figure 4.3).
- Dans un second temps, G décroît.

Le profil de F_{in} est donné figure 4.11. Il présente aussi deux phases dont la première est une allure exponentielle. Elle permet de garder la concentration G relativement constante dans un premier temps.

On peut donner une justification de l'optimalité du profil F_{in} en revenant au cas n° 2, dans lequel le flux ν_T est libre. On pose ν_T^{*batch} le flux du processus Σ_T dans le cas n° 2. A partir de ce flux, définissons l'efficacité effective du processus Σ_T , $k_{T,eff}$, comme suit :

$$k_{T,eff}(t)E_T(t) = \nu_T^{*batch}(t) \leq k_T(t)E_T(t) \quad (4.2.27)$$

D'après la relation (4.2.13), il vient la définition de $G_{eff}(t)$, la concentration effective en $\mathcal{G}$ que la cellule utilise à l'instant t dans le cas n° 2 :

$$k_{T,eff}(t) = \frac{\nu_T G_{eff}(t)}{G_{eff}(t) + K_T} \quad (4.2.28)$$

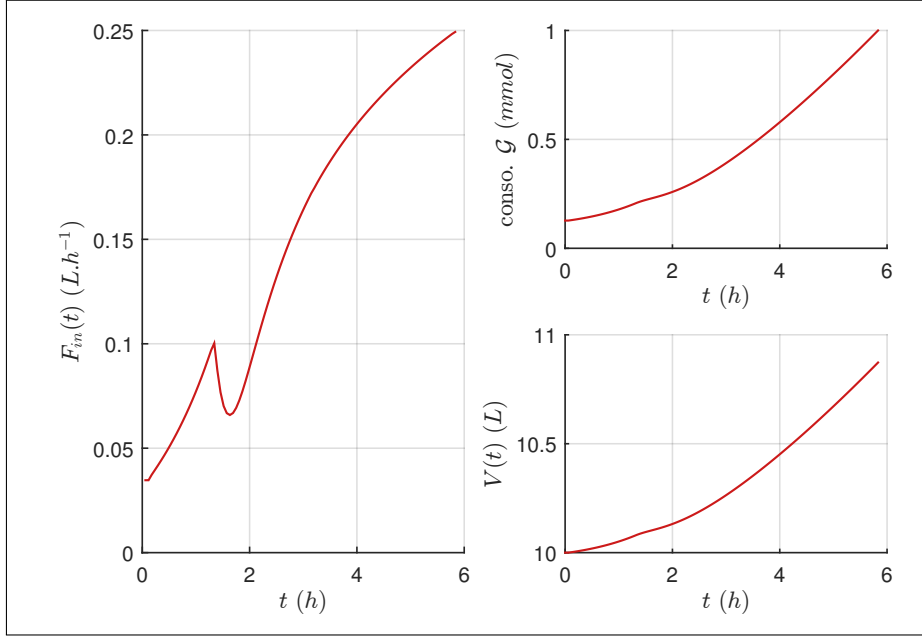


FIGURE 4.11 – Allures du débit d'alimentation F_{in} , de la quantité de $\mathcal{G}$ consommée et du volume effectif du bioréacteur pour le cas n° 3.

Pour faire suivre l'allure ν_T^{*batch} au flux ν_T en conditions fed-batch, il suffit de faire suivre l'allure $G_{eff}(t)$ à la concentration en $\mathcal{G}$. En traçant sur la figure 4.12 l'évolution de la concentration $G_{eff}(t)$ pour la solution au cas n° 2 et la concentration $G(t)$ pour la trajectoire optimale en fed-batch, on constate qu'elles sont très proches. Ainsi, on peut conclure que le profil optimal $F_{in}(t)$ et le paramètres $G(t_0)$ du cas n° 3 sont bien définis de sorte à suivre l'allure $G_{eff}(t)$ du cas n° 2.

Expérimentalement, en utilisant Bocop et en partant de différents points initiaux, le problème fed-batch de la section (4.2.3.3) est bien plus long à résoudre que le problème batch de la section (4.2.3.2) (empiriquement, de l'ordre de l'heure au lieu de quelques secondes).

La procédure ci-dessus permet de résoudre le problème en deux temps. D'abord, on résout le problème en batch avec tout le substrat disponible dans le milieu à l'état initial. Puis, en fonction des flux de consommation, on trouve le profil optimal de substrat en fonction du temps. Enfin la trajectoire optimale du débit d'alimentation est celle qui suit G_{eff} . Cette procédure d'optimisation en fed-batch pourra être transposée et utilisée dans le cas de la modélisation de l'inhibition par le substrat.

On a présenté ici un contrôle réaliste de l'expression des gènes et des conditions extracellulaires pour la maximisation de la concentration de $P(t_f)$ par unité de temps.

Il est intéressant de constater qu'on retrouve la même stratégie que pour l'optimisation de la conduite du procédé présentée dans [36], *i.e.* avec un modèle macroscopique phénoménologique décrivant le même comportement nominal. En effet, dans ces travaux, on cherche à maximiser la production de biomasse, et la stratégie revient à se placer à la limite entre le régime oxydo-fermentatif et celui de respiration.

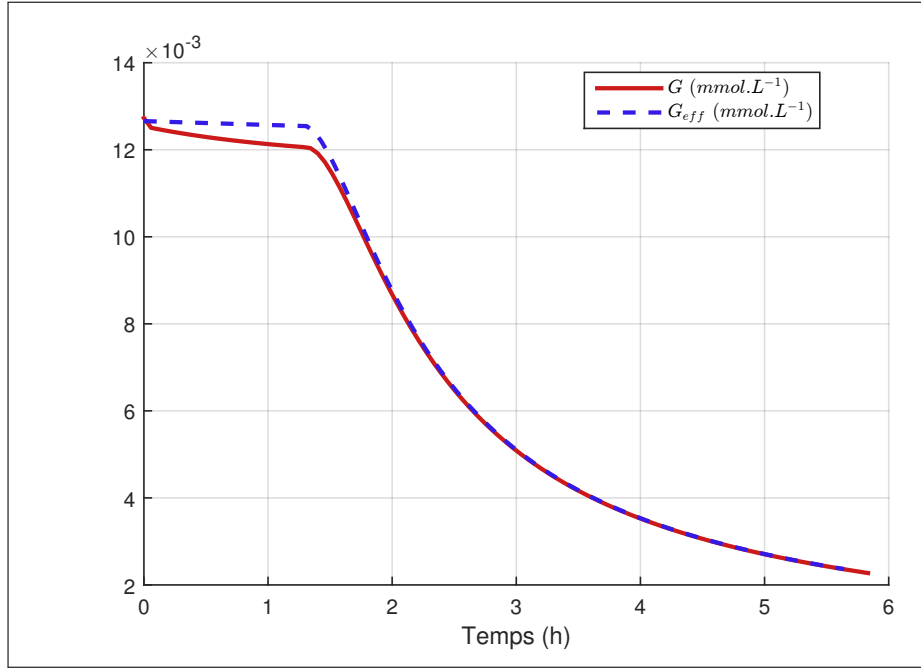


FIGURE 4.12 – Comparaison entre la concentration effective G_{eff} définie au cas n° 2 en fonctionnement batch et la concentration G du cas n° 3 en mode de fonctionnement fed-batch.

4.2.3.4 Conclusion des problèmes d'optimisation

Au travers de ces trois cas, on a défini pour le réseau $\mathcal{R}_{2voies}$ une procédure rationnelle d'optimisation de la bioproduction d'un composé d'intérêt par conception de souche et conduite du procédé. En effet, la modification de souche proposée à l'issue du cas n° 3 revient à ajouter un gène dont on connaît l'expression en fonction du taux de croissance, et à activer ce gène à un instant donné. La conduite du procédé fed-batch revient à suivre l'allure G_{eff} donnée figure 4.12. On peut alors utiliser une technique de commande linéarisante ou de commande prédictive pour suivre ce profil, comme présenté en 1.1.2.2.

Le tableau 4.2 présente la concentration finale en produit $\mathcal{P}$, le temps final et la valeur du critère pour les solutions optimales présentées en section 4.2.3.1, 4.2.3.2 et 4.2.3.3.

TABLE 4.2 – Valeurs numériques associées aux solutions optimales des cas n° 1, n° 2 et n° 3.

	$P(t_f) (\mu mol.L^{-1})$	$t_f (h)$	$\frac{P(t_f)}{t_f} (\mu mol.L^{-1}.h^{-1})$
Cas n° 1	32,2	4,84	6,65
Cas n° 2	21,2	5,70	3,72
Cas n° 3	19,8	5,83	3,39

Le passage du cas n° 1 au cas n° 2 revient à une dégradation de performance de l'ordre de 44%. Cette dégradation vient de l'implémentation de la régulation de l'expression des gènes par le taux de croissance. En revanche, le passage du cas n° 2 au cas n° 3 est quasiment neutre, comme cela a été discuté au point 4.2.3.3.

4.2.4 Conclusions relatives au modèle à deux voies

Le modèle $\mathcal{R}_{2voies}$ introduit dans cette section a permis de tester la procédure définie en 4.1. Nous avons défini avec succès le modèle avec le formalisme de la section 4.1. Ensuite, nous avons abordé des problèmes d'optimisation de conception de souche. Comme au chapitre 2, nous avons d'abord trouvé la solution optimale étant données les contraintes de fonctionnement de la cellule. Nous avons ensuite cherché la solution qui maximise la productivité du composé d'intérêt et qui puisse être réalisée au travers de la conception de gènes constitutifs.

Nous avons vu que le passage à un modèle plus détaillé pouvait permettre d'agir différemment sur le procédé. En particulier, nous avons proposé une trajectoire optimale pour le débit d'alimentation dans le cas d'un bioréacteur en mode de fonctionnement fed-batch.

4.3 Conclusion du chapitre

Les travaux présentés dans ce chapitre sont encourageants pour donner une suite à la modélisation développée aux chapitres 2 et 3. Nous avons détaillé un protocole permettant de représenter selon la méthode dRBA n'importe quel réseau décrivant le fonctionnement cellulaire. Ce faisant, nous avons établi la liste des paramètres à définir. Cette étude permet de saisir quels éléments sont nécessaires pour définir la modélisation sur un réseau. Finalement, il apparait que ce sont les mêmes éléments que ceux utilisés par la méthode RBA statique : tous ces paramètres peuvent être établis sur une base commune car ils proviennent tous de considérations physico-biologiques. On voit de nouveau que le modèle développé est bien un modèle autonome dans le sens où il n'y a pas besoin d'avoir d'observations du fonctionnement de la cellule pour calibrer une partie des paramètres.

La section 4.2 a considéré l'application de la méthode à un raffinement du modèle agrégé du chapitre 2. Le processus de synthèse de $\mathcal{S}$ à partir de $\mathcal{G}$ a été divisé en trois processus dont deux sont redondants. Cette section a permis de tester le protocole de construction de modèle et de retrouver un mode de fonctionnement connu des micro-organismes. Ce mode de fonctionnement est le rejet de cofacteurs en présence de concentrations importantes de substrat. On appelle cela également processus de débordement (ou *overflow*). Nous avons été en mesure d'appliquer le protocole de construction du modèle et avons effectivement retrouvé les modes de fonctionnement typiques des micro-organismes.

Le raffinement du modèle a impacté la politique de la conduite du procédé. En effet, la solution optimale au problème de maximisation de la production de composé d'intérêt est plus élaborée ici qu'avec le modèle agrégé. En particulier, nous avons obtenu une stratégie de contrôle en deux temps avec la régulation de la concentration extracellulaire en substrat à une valeur constante dans le premier temps. Cette perspective est intéressante et nous encourage à détailler davantage le modèle pour améliorer la conduite du procédé.

Conclusions & Perspectives

Au cours de cette thèse, nous avons cherché à intégrer les étapes d'optimisation de la conduite de procédés avec les étapes de conception de souche. Pour ce faire, nous avons défini un modèle à la jonction entre les deux domaines. Ainsi, nous avons posé les bases d'une intégration rationnelle des processus intracellulaires dans l'optimisation de la conduite de procédé. Ce faisant, nous avons rapproché les modèles de bioprocédé des éléments biologiques sous-jacents. Cette intégration repose sur une modélisation de bioprocédé qui est le cœur de ces travaux.

Dans le premier chapitre de ce manuscrit, nous avons vu des méthodes classiques pour optimiser un bioprocédé. Elles consistent à développer un modèle macroscopique des échanges entre les entités. Cette idée a été reprise au chapitre 2 pour élaborer un modèle de bioprocédé qui englobe les aspects de conception de souche. Comme présenté au chapitre 1, avec une vision du fonctionnement fondée sur les processus, la conception de souche revient à modifier le fonctionnement des micro-organismes en trouvant les flux optimaux des processus intracellulaires.

La modélisation que nous avons développée au chapitre 2 permet de formuler des problèmes de conception de souche, à partir de considérations physico-biologiques. Ce modèle agrégé du fonctionnement d'une cellule bactérienne a été rapproché de modèles phénoménologiques macroscopiques. En particulier, on a montré que l'allure du taux spécifique de croissance en fonction de la concentration de substrat dans le milieu était hyperbolique, croissante majorée, comme une allure classique de Monod. Par ailleurs, on a vu que l'évolution du fonctionnement intracellulaire et la répartition des ressources entre les processus était en accord avec des méthodes de modélisation fine du fonctionnement des bactéries, en particulier la méthode RBA validée expérimentalement à l'échelle du génome.

En résumé, le modèle développé au chapitre 2 correspond bien à une modélisation à la jonction entre celles utilisées en conduite du procédé et en conception de souche.

A l'aide de ce modèle, nous avons proposé la formulation d'un problème d'optimisation de la bioproduction d'un composé d'intérêt. En utilisant le modèle pour décrire le procédé et en particulier le fonctionnement des organismes, nous avons été en mesure de proposer une trajectoire optimale qui satisfait les contraintes liées à la structure des principaux mécanismes des micro-organismes. Nous avons trouvé une stratégie optimale étonnamment proche des stratégies optimales obtenues pour l'optimisation de la production de composé d'intérêt formulées à l'aide de modèles macroscopiques. Nous avons obtenu une stratégie en trois phases : croissance, transition, production. Par rapport à des modèles phénoménologiques à l'échelle intracellulaire, nous avons montré que notre méthode permettait de caractériser la phase de transition à l'aide de considérations

physico-biologiques. A notre connaissance, il s'agit d'une première dans la caractérisation de la transition sans artifices phénoménologiques de modélisation.

A ce stade, la trajectoire concerne les flux et les états de tous les processus intracellulaires. Elle est optimale sous respect des contraintes structurelles des processus. Toutefois, à ce stade, aucun élément n'est donné concernant la mise en œuvre de cette solution, en particulier en terme de conduite du procédé. Le chapitre 3 lève le voile de l'implémentation. On y propose une manière d'implémenter le contrôle des processus intracellulaires par la conception de gènes constitutifs des processus intracellulaires. Concrètement, on observe dans ce chapitre la perte d'optimalité liée à l'ajout de contraintes réalistes sur le contrôle de l'expression des gènes par le taux de croissance. Les trajectoires obtenues dans ce chapitre sont ainsi respectueuses du fonctionnement des processus et des contraintes d'implémentation réalistes. Cet ajout de contraintes de commande permet d'envisager la conduite de procédé à partir de souche modifiée. On est alors en mesure de proposer une loi de commande macroscopique qui s'appuie sur les processus intracellulaires.

Le dernier chapitre donne un cadre de généralisation de la méthode à des représentations plus détaillées du fonctionnement de micro-organismes. On applique en particulier le protocole de généralisation à un réseau de processus qui mime le fonctionnement des voies de synthèse de l'énergie dans les micro-organismes : un processus de conversion parfaite du substrat en énergie qui est coûteux à mettre en place, et un processus de conversion partielle qui est peu coûteux. L'extension de la modélisation à ce réseau a permis de justifier la généralisation de la méthode. Par ailleurs, nous avons vu qu'en raffinant la modélisation, on raffinaient également la trajectoire à suivre pour l'optimisation du procédé.

Ces résultats nous encouragent à poursuivre sur cette voie pour traiter des sujets d'optimisation de bioprocédés, en particulier dans les cas où il y a la possibilité de modifier les souches.

Une perspective naturelle à ces travaux au sens de la conduite du procédé serait d'implémenter, en simulation, une loi de commande en mode de fonctionnement fed-batch, en reprenant la trajectoire optimale obtenue en section 3.3.2 avec le modèle agrégé. Cette trajectoire avait été obtenue avec l'utilisation du modèle à deux voies et via l'optimisation conjointe de la conduite du procédé et de la conception de souche (section 4.2.3.3). L'idée serait de mettre en place une commande linéarisante ou une commande prédictive (présentées en section 1.1.2.2) pour asservir le fonctionnement réel à la trajectoire calculée hors-ligne. On comparerait alors les performances entre une loi de commande fondée sur un modèle macroscopique classique et un modèle de processus intracellulaire, comme développé dans cette thèse. Pour une comparaison entre modèles équivalents, on calibrerait un modèle macroscopique sur les équilibres atteints par le modèle des processus à l'équilibre (comme on l'a fait en section 2.2.1). Le modèle pourrait ensuite être perturbé et on observerait l'intérêt de se baser sur les processus intracellulaires plutôt que sur les observations macroscopiques pour commander le bioprocédé.

Si l'on souhaite utiliser la méthode développée dans ces travaux dans un cadre réel, il est envisageable voire nécessaire de développer des techniques d'estimation d'état des processus intracellulaires. En effet, jusqu'alors on a supposé que les états intracellulaires

étaient parfaitement connus. Ceci n'a posé aucun problème pour la définition de trajectoires optimales, mais pour suivre ces trajectoires et déterminer la commande en boucle fermée à appliquer au système, il faut connaître l'évolution des états.

En lien avec cette définition d'observateurs et d'estimateurs, il serait bon de poursuivre les travaux présentés dans [54] sur l'identifiabilité des paramètres. Ces travaux ont été développés au cours d'un projet interne au département Automatique de CentraleSupélec. Ils présentent un protocole d'identification des paramètres d'une version intermédiaire du modèle intracellulaire présenté [43]. Les étapes sont reprises de [37]. L'identification du modèle est un point central pour l'estimation d'état. Mener une étude d'identifiabilité est également intéressant pour étudier l'influence des paramètres du modèle sur les états.

Une perspective naturelle aux travaux développés dans cette thèse consiste à détailler la modélisation pour prendre en compte le plus de processus possible, jusqu'à aller à l'échelle du génome pour un organisme type, comme *Bacillus subtilis* ou *Escherichia coli*. Au vu du protocole défini au chapitre 4, il ne semble pas y avoir de limitation à cela. Comme ce modèle repose quasiment sur les mêmes entités que la modélisation RBA, et qu'un modèle RBA existe pour *B. subtilis* à l'échelle du génome, il serait possible d'étendre notre modèle à cet organisme. Le point bloquant pourrait être de définir les efficacités des processus comme des fonctions des métabolites, mais comme nous l'avons vu en section 4.2, prendre dans un premier temps des efficacités constantes permet tout de même de caractériser le système.

Parmi les suites à donner à ces travaux, en restant à une échelle agrégée des processus intracellulaires, il serait bon de définir un processus de maintenance. Selon nous, la maintenance peut être prise en compte de différentes manières. Comme dans les RBA, on peut définir un ensemble de protéines à produire en permanence. On définit soit que le flux de production de ces protéines doit être supérieur à un seuil, soit que la concentration de ce groupe est fixe (ou représente une part fixe dans la densité). Comme on l'a vu en comparant les allures du taux de croissance de notre modèle et du modèle RBA (figure 2.7), cette prise en compte de la maintenance permettrait sans doute de retrouver la possibilité de ne pas croître sans être à concentration nulle en substrat. Il se pourrait alors que la prise en compte de ce phénomène modifie (voire simplifie) les allures des trajectoires à suivre pour optimiser la bioproduction du composé d'intérêt.

Plus subtilement, la suite de ce travail consiste à remettre en question le critère utilisé pour l'optimisation des bioprocédés. Nous avons discuté des différents critères en particulier du rendement de production du produit par rapport au temps. Ce critère nous a permis de retrouver des modes de fonctionnements classiques en conduite du procédé, en particulier des phases exponentielles de croissance. Il peut paraître décevant de ne pas avoir réussi à retrouver le même genre de comportement avec le critère de maximisation du produit d'intérêt ($\max J = \max P(t_f)$).

Toutefois, des résultats intermédiaires non reproduits dans ce manuscrit ont révélé que la prise en compte de la maintenance dans la modélisation des processus intracellulaires couplée à la maximisation de la quantité de produit à l'instant final permettait de retrouver des allures typiques (phase de croissance exponentielle suivie de production du produit d'intérêt). Cette piste reste à explorer.

Enfin, pour un raffinement supplémentaire de modélisation, on pourrait envisager la prise en considération du fonctionnement de la cellule de manière plus détaillée. En particulier en prenant en compte explicitement les messagers et le processus de transcription, nous aurions une couche de détails supplémentaire qui permettrait de mettre en avant de nouvelles stratégies de contrôle des processus. De la même façon, nous pourrions détailler davantage les mécanismes de régulations intracellulaires afin de permettre une conception de souche plus élaborée. La conception de souche reviendrait alors à définir un certain nombre de gènes constitutifs et à définir les niveaux de régulation des processus.

Bibliographie

- [1] P. Agrawal, G. Koshy, and M. Ramseier. An algorithm for operating a fed-batch fermentor at optimum specific-growth rate. *Biotechnology and bioengineering*, 33(1) :115–125, 1989.
- [2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Biologie moléculaire de la cellule, Cinquième édition*. Médecine Sciences Publications, Lavoisier, 2011.
- [3] F. Allgöwer and A. Zheng. *Nonlinear Model Predictive Control*, volume 26. Birkhäuser, 2012.
- [4] G. Bastin and D. Dochain. *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier, 1990.
- [5] Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A.-L. Barabási, and Z. N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, 104(31) :12663–12668, 2007.
- [6] L. T. Biegler. *Nonlinear Programming : Concepts, Algorithms, and Applications to Chemical Processes*, volume 10. Siam, 2010.
- [7] V.M. Boer, C.A. Crutchfield, P.H. Bradley, D. Botstein, and J.D. Rabinowitz. Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations. *Molecular biology of the cell*, 21(1) :198–211, 2010.
- [8] P. Bogaerts, K. M. Gziri, and A. Richelle. From MFA to FBA : Defining linear constraints accounting for overflow metabolism in a macroscopic FBA-based dynamical model of cell cultures in bioreactor. *Journal of Process Control*, 60 :34–47, 2017.
- [9] H. P.J. Bonarius, G. Schmid, and J. Tramper. Flux analysis of underdetermined metabolic networks : the quest for the missing constraints. *Trends in Biotechnology*, 15(8) :308–314, 1997.
- [10] F. J. Bonnans, D. Giorgi, V. Grelard, B. Heymann, S. Maindrault, P. Martinon, O. Tissot, and J. Liu. Bocop – A collection of examples. Technical report, INRIA, 2017.
- [11] O. Borkowski. *Growth-rate-dependent protein production in bacteria*. PhD thesis, Paris 5, 2013.
- [12] O. Borkowski, A. Goelzer, M. Schaffer, M. Calabre, U. Mäder, S. Aymerich, M. Jules, and V. Fromion. Translation elicits a growth rate-dependent, genome-wide, differential protein production in *Bacillus subtilis*. *Molecular Systems Biology*, 12(5) :870, 2016.

- [13] H. Bremer and P. P. Dennis. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1), 2008.
- [14] H.D.P.P. Bremer, P.P. Dennis, et al. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella : cellular and molecular biology*, 2(2) :1553–69, 1996.
- [15] A. E. Bryson and Y.-C. Ho. *Applied Optimal Control : Optimization, Estimation and Control*. Routledge, 1975.
- [16] E. F. Camacho and C. B. Alba. *Model Predictive Control*. Springer Science & Business Media, 2013.
- [17] L. Chang, X. Liu, and M. A. Henson. Nonlinear model predictive control of fed-batch fermentations using dynamic flux balance models. *Journal of Process Control*, 42 :137–149, 2016.
- [18] A. Chowdhury, A. R. Zomorodi, and C. D. Maranas. Bilevel optimization techniques in computational strain design. *Computers & Chemical Engineering*, 72 :363–372, 2015.
- [19] D. Coutinho and A. Vande Wouwer. A robust non-linear feedback control strategy for a class of bioprocesses. *IET Control Theory & Applications*, 7(6) :829–841, 2013.
- [20] J.E. Cuthrell and L. T. Biegler. Simultaneous optimization and solution methods for batch reactor control profiles. *Computers & Chemical Engineering*, 13(1-2) :49–62, 1989.
- [21] Dochain D. *Automatique des Bioprocédés*. Hermes Science Publications, 2001.
- [22] H. De Jong, J. Geiselman, and D. Ropers. Resource reallocation in bacteria by reengineering the gene expression machinery. *Trends in microbiology*, 25(6) :480–493, 2017.
- [23] V. H. Edwards. The influence of high substrate concentrations on microbial kinetics. *Biotechnology and Bioengineering*, 12(5) :679–712, 1970.
- [24] D. A. Fell. Increasing the flux in metabolic pathways : a metabolic control analysis perspective. *Biotechnology and Bioengineering*, 58(2-3) :121–124, 1998.
- [25] H. Frankowska. The maximum principle for a differential inclusion problem. In *Analysis and Optimization of Systems*, pages 517–531. Springer, 1984.
- [26] V. Fromion et al. Biologie de synthèse pour la chimie, l’énergie et l’environnement. Technical report, Allenvi, 2016.
- [27] K. G. Gadkar, F. J. Doyle III, J. S. Edwards, and R. Mahadevan. Estimating optimal profiles of genetic alterations using constraint-based models. *Biotechnology and bioengineering*, 89(2) :243–251, 2005.
- [28] K. G. Gadkar, R. Mahadevan, and F. J. Doyle III. Optimal genetic manipulations in batch bioreactor control. *Automatica*, 42(10) :1723–1733, 2006.
- [29] L. Gerosa, K. Kochanowski, M. Heinemann, and U. Sauer. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Molecular Systems Biology*, 9(1) :658, 2013.
- [30] N. Giordano, F. Mairet, J.-L. Gouzé, J. Geiselman, and H. De Jong. Dynamical allocation of cellular resources as an optimal control problem : novel insights into microbial growth strategies. *PLoS computational biology*, 12(3) :e1004802, 2016.

- [31] A. Goelzer, F. B. Briki, I. Martin-Verstraete, P. Noirot, P. Bessières, S. Aymerich, and V. Fromion. Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC systems biology*, 2(1) :20, 2008.
- [32] A. Goelzer and V. Fromion. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1810(10) :978–988, 2011.
- [33] A. Goelzer and V. Fromion. Resource allocation in living organisms. *Biochemical Society Transactions*, page BST20160436, 2017.
- [34] A. Goelzer, V. Fromion, and G. Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6) :1210–1218, 2011.
- [35] A. Goelzer, J. Muntel, V. Chubukov, M. Jules, E. Prestel, R. Nölker, M. Mariadasou, S. Aymerich, M. Hecker, P. Noirot, et al. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic engineering*, 32 :232–243, 2015.
- [36] G. Hafidi. *Application de la commande prédictive non-linéaire à la commande de culture de bactéries Escherichia coli*. PhD thesis, Université Paris Sud-Paris XI, 2008.
- [37] G. Hafidi, S. Tebbani, D. Dumur, and A. Vande Wouwer. Analyse de sensibilité pour l’identification d’un modèle paramétrique de culture *E. coli*. In *Automatique avancée et Informatique appliquée*, ISBN 978-973-27-1806-3, pages 129–135. Edition de l’Académie Roumaine, 2009.
- [38] R. F. Hartl, S. P. Sethi, and R. G. Vickson. A survey of the maximum principles for optimal control problems with state constraints. *SIAM review*, 37(2) :181–218, 1995.
- [39] W. J. Holtz and J. D. Keasling. Engineering static and dynamic control of synthetic pathways. *Cell*, 140(1) :19–23, 2010.
- [40] A. Isidori. *Nonlinear Control Systems*. Springer Science & Business Media, 2013.
- [41] B. Jabarivelisdeh and S. Waldherr. Improving bioprocess productivity using constraint-based models in a dynamic optimization scheme. *IFAC-PapersOnLine*, 49(26) :245–251, 2016.
- [42] B. Jabarivelisdeh and S. Waldherr. Optimization of bioprocess productivity based on metabolic-genetic network models with bilevel dynamic programming. *Biotechnology and bioengineering*, 2018.
- [43] G. Jeanne, S. Tebbani, A. Goelzer, V. Fromion, and D. Dumur. Modelling and optimization of metabolic pathways in bacteria. In *System Theory, Control and Computing (ICSTCC), 2016 20th International Conference on*, pages 312–317. IEEE, 2016.
- [44] S. Klumpp and T. Hwa. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of Sciences*, 105(51) :20245–20250, 2008.
- [45] S. Klumpp and T. Hwa. Bacterial growth : global effects on gene expression, growth feedback and proteome partition. *Current opinion in biotechnology*, 28 :96–102, 2014.

- [46] S. Klumpp, M. Scott, S. Pedersen, and T. Hwa. Molecular crowding limits translation and cell growth. *Proceedings of the National Academy of Sciences*, 110(42) :16754–16759, 2013.
- [47] S. Klumpp, Z. Zhang, and T. Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7) :1366–1375, 2009.
- [48] J. E. Krebs, E. S. Goldstein, and S. T. Kilpatrick. *Lewin’s Genes XI*. Jones & Bartlett Learning, 2014.
- [49] H. E. Kubitschek. Buoyant density variation during the cell cycle in microorganisms. *CRC Critical reviews in microbiology*, 14(1) :73–97, 1987.
- [50] H. E. Kubitschek, W. W. Baldwin, S. J. Schroeter, and R. Graetzer. Independence of buoyant cell density and growth rate in *Escherichia coli*. *Journal of bacteriology*, 158(1) :296–299, 1984.
- [51] J. Lee and W.F. Ramirez. Optimal fed-batch control of induced foreign protein production by recombinant bacteria. *AIChE Journal*, 40(5) :899–907, 1994.
- [52] J. M. Lee, E. P. Gianchandani, and J. A. Papin. Flux balance analysis in the era of metabolomics. *Briefings in bioinformatics*, 7(2) :140–150, 2006.
- [53] N. E. Lewis, H. Nagarajan, and B. Ø. Palsson. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4) :291, 2012.
- [54] G. Loup and E. Portier. Identification des paramètres d’une cellule bactérienne. Technical report, CentraleSupélec, 2017.
- [55] D. S. Lun, G. Rockwell, N. J. Guido, M. Baym, J. A. Kelner, B. Berger, J. E. Galagan, and G. M. Church. Large-scale identification of genetic design strategies using local search. *Molecular Systems Biology*, 5(1) :296, 2009.
- [56] R. Mahadevan, J. S. Edwards, and F. J. Doyle III. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical journal*, 83(3) :1331–1340, 2002.
- [57] A. G. Marr. Growth rate of *Escherichia coli*. *Microbiological reviews*, 55(2) :316–333, 1991.
- [58] M. Z. Mehrjerdi, B. Moshiri, G. Amoabediny, A. Ashktorab, and B. N. Araabi. Model predictive control with sequential linearization approach for a nonlinear time varying biological system. In *Control and Decision Conference (CCDC), 2013 25th Chinese*, pages 4465–4470. IEEE, 2013.
- [59] J. M. Modak. Choice of control variable for optimization of fed-batch fermentation. *The chemical engineering journal*, 52(3) :B59–B69, 1993.
- [60] J. Monod. Recherches sur la croissance des cultures bacteriennes. 1942.
- [61] M. Mori, T. Hwa, O. C Martin, A. De Martino, and E. Marinari. Constrained allocation flux balance analysis. *PLoS computational biology*, 12(6) :e1004913, 2016.
- [62] P. Nicolas, U. Mäder, E. Dervyn, T. Rochat, A. Leduc, N. Pigeonneau, E. Bidnenko, E. Marchadier, M. Hoebeke, S. Aymerich, et al. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science*, 335(6072) :1103–1106, 2012.

- [63] P. S. Novichkov, A. E. Kazakov, D. A. Ravcheev, S. A. Leyn, G. Y. Kovaleva, et al. Regprecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC genomics*, 14(1) :745, 2013.
- [64] E. J. O’Brien, J. A. Lerman, R. L. Chang, D. R. Hyduke, and B. Ø. Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(1) :693, 2013.
- [65] Y.-K. Oh, B. Ø. Palsson, S. M. Park, C. H. Schilling, and R. Mahadevan. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry*, 282(39) :28791–28799, 2007.
- [66] S. Öztürk, P. Calik, and T. H. Özdamar. Fed-batch biomolecule production by *Bacillus subtilis* : a state of the art review. *Trends in biotechnology*, 34(4) :329, 2016.
- [67] S. Park and W.F. Ramirez. Optimal production of secreted protein in fed-batch reactors. *AIChE Journal*, 34(9) :1550–1558, 1988.
- [68] P. Pharkya, A. P. Burgard, and C. D. Maranas. Optstrain : a computational framework for redesign of microbial production systems. *Genome research*, 14(11) :2367–2376, 2004.
- [69] S. Ranganathan, P. F. Suthers, and C. D. Maranas. Optforce : an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS computational biology*, 6(4) :e1000744, 2010.
- [70] M. Rügen, A. Bockmayr, and R. Steuer. Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. *Scientific reports*, 5, 2015.
- [71] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression : Origins and consequences. *Science*, 330(6007) :1099–1102, nov 2010.
- [72] B. Srinivasan, S. Palanki, and D. Bonvin. Dynamic optimization of batch processes : I. Characterization of the nominal solution. *Computers & Chemical Engineering*, 27(1) :1–26, 2003.
- [73] S. Tebbani, D. Dumur, and G. Hafidi. Open-loop optimization and trajectory tracking of a fed-batch bioreactor. *Chemical Engineering and Processing : Process Intensification*, 47(11) :1933–1941, 2008.
- [74] S. Tebbani, D. Dumur, G. Hafidi, and A. Vande Wouwer. Nonlinear predictive control of fed-batch cultures of *Escherichia coli*. *Chemical Engineering & Technology*, 33(7) :1112–1124, 2010.
- [75] P. Unrean and C. J. Franzen. Dynamic flux balancing elucidates NAD(P)H production as limiting response to furfural inhibition in *saccharomyces cerevisiae*. *Biotechnology journal*, 10(8) :1248–1258, 2015.
- [76] P. Unrean, S. Khajeeram, and V. Champreda. Combining metabolic evolution and systematic fed-batch optimization for efficient single-cell oil production from sugar-cane bagasse. *Renewable Energy*, 111 :295–306, 2017.

- [77] P. Van Bodegom. Microbial maintenance : a critical review on its quantification. *Microbial ecology*, 53(4) :513–523, 2007.
- [78] M.G. Vander Heiden, L.C. Cantley, and C.B. Thompson. Understanding the warburg effect : the metabolic requirements of cell proliferation. *Science*, 324(5930) :1029–1033, 2009.
- [79] A. Varma and B. Ø. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* w3110. *Applied and environmental microbiology*, 60(10) :3724–3731, 1994.
- [80] A. Varma and B. Ø. Palsson. Parametric sensitivity of stoichiometric flux balance models applied to wild-type *Escherichia coli* metabolism. *Biotechnology and bioengineering*, 45(1) :69–79, 1995.
- [81] R. Vinter. *Optimal Control*. Springer Science & Business Media, 2010.
- [82] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1) :25–57, 2006.
- [83] S. Waldherr, Diego A Oyarzún, and A. Bockmayr. Dynamic optimization of metabolic networks coupled with gene expression. *Journal of theoretical biology*, 365 :469–485, 2015.
- [84] R.-S. Wang, Z. Li, X.-S. Zhang, and L. Chen. Optimization and control of metabolic networks. In *2010 29th Chinese Control Conference (CCC)*, pages 6251–6255. IEEE, 2010.
- [85] A. Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9) :E1038–E1047, 2015.
- [86] L. Yee and H.W. Blanch. Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. *Nature Biotechnology*, 10(12) :1550, 1992.
- [87] I. Yegorov, F. Mairet, and J.-L. Gouzé. Optimal resource allocation for bacterial growth with degradation. *IFAC-PapersOnLine*, 50(1) :9858–9863, 2017.
- [88] C. You, H. Okano, S. Hui, Z. Zhang, M. Kim, C. W. Gunderson, Y.-P. Wang, P. Lenz, D. Yan, and T. Hwa. Coordination of bacterial proteome with metabolism by cyclic amp signalling. *Nature*, 500(7462) :301–306, 2013.

Annexe A

Introduction au fonctionnement d'une cellule bactérienne

Cette annexe s'adresse en priorité au lecteur n'ayant pas de connaissance en biologie et qui aurait besoin d'une mise à niveau initiale pour être en mesure de lire le manuscrit présent. Nous donnons une représentation du fonctionnement général d'une cellule vivante en centrant notre vision sur les mécanismes et comportements présentés dans le corps de la thèse.

La vision développée est centrée sur le monde des bactéries, (procaryotes) et précisément sur l'espèce *Bacillus subtilis*. Les bactéries ne sont qu'une partie de tous les organismes vivants qui sont réparties en trois domaines : les bactéries, les eucaryotes et les archées. Néanmoins, la plupart des points présentés dans cette annexe sont d'ordre général et s'appliquent à tout être vivant. Nous ne saurions que renvoyer le lecteur à des références telles que [2] ou [48] pour de plus amples détails sur la manière dont fonctionnent les processus que nous décrivons. Les pages Wikipedia en version anglaise, associées aux différents termes, mécanismes et principes présentés ci-après sont par ailleurs souvent bien fournies et globalement de bonne qualité.

Une cellule est une entité capable de produire chacun de ses composants à l'aide uniquement des composés disponibles dans son milieu. C'est le principe de la vie : se contenter de ce qui est présent dans son environnement et en faire le meilleur usage possible pour se reproduire. Avec cette vision idéaliste de cellules qui ne font que croître en conditions parfaites de culture. La cellule ne fait rien d'autre que de produire une cellule complète. C'est en un sens la définition que donne Schrödinger, pour qui la vie est un système ouvert utilisant l'énergie de son environnement pour entretenir son ordre intérieur, en somme un puits à entropie.

Le principe de la cellule à avoir à l'esprit est la conservation de la matière au niveau atomique : chaque molécule qui entre dans une cellule est avant tout un ensemble d'atomes. La cellule produit une nouvelle cellule à partir de ces atomes. Lorsque l'on se représente le monde à cette échelle, il n'y a pas de distinction entre la physique et la chimie : les réactions chimiques sont des transferts d'atomes, l'énergie se trouve sous forme d'énergie mécanique, l'étude de la polarité des composés fait appels aux lois de l'électrostatique, *etc.*

A.1 Mode de vie des cellules bactériennes

Les cellules bactériennes considérées dans ce travail sont présentes en milieu aqueux dans lequel elles consomment des nutriments qui permettant de croître.

En conditions favorables, on observe que les colonies bactériennes se développent à taux de croissance constant, *i.e.* leur augmentation massique est proportionnelle à leur masse. Mathématiquement, en notant $X(t)$ la masse sèche de bactéries présente à l'instant t dans un certain milieu favorable, il a donc été observé que :

$$\frac{\dot{X}(t)}{X(t)} \triangleq \mu(t) = cste \quad (\text{A.1.1})$$

où $\dot{X}(t)$ est la dérivée temporelle de $X(t)$ et où $\mu(t)$ est le taux spécifique de croissance. En conditions favorables, cette quantité a été observée constante, posons μ_0 sa valeur. On a donc $\dot{X}(t) = \mu_0 X(t)$, ou encore

$$X(t) = X(t_0) \exp(\mu_0(t - t_0))$$

avec t_0 un certain instant.

C'est ce que l'on appelle une phase exponentielle de croissance. La composition intracellulaire est alors en moyenne constante. On parle ici de moyenne sur la population bactérienne. Cette composition dépend du taux de croissance.

Lorsque les conditions ne sont pas réunies pour la croissance, typiquement lorsque la quantité de nutriments est insuffisante, on a alors $\mu(t) = 0$. Les cellules sont dites en phase stationnaire. La population est alors constante, éventuellement décroissante à cause de la mortalité des cellules. Des comportements et des stratégies spécifiques à chaque espèce de micro-organisme ont alors lieu comme la sporulation ou la formation de biofilm. La sporulation consiste pour un organisme à emballer son ADN en une spore extrêmement résistante et à attendre que les conditions extérieures soient de nouveau favorables à la croissance.

Les phases transitoires entre des phases de croissance ou entre des phases de croissance et des phases stationnaires sont caractérisées par le changement de mode de fonctionnement des cellules. Ce sont des phases difficiles à décrire et à prédire.

A.1.1 Échelles de temps et d'espace

Il est difficile de parler d'une bactérie en tant que cellule isolée, puisqu'il est difficile d'identifier un individu particulier parmi la population, et que les bactéries se reproduisent en se dédoublant quasiment parfaitement. Dès lors, il est impossible de distinguer la mère de la fille. Il reste plus raisonnable de parler de population cellulaire.

Pour la compréhension de l'objet, nous tentons néanmoins de définir ici des notions d'espace et de temps pour un individu.

A.1.1.1 Quelques ordres de grandeur spatiaux

L'ordre de grandeur de la longueur caractéristique d'une bactérie est le micromètre, 10^{-6} m. Si l'on prend 10^{-10} m pour la longueur caractéristique d'un atome, 10^{-9} m

pour la longueur caractéristique d'une molécule (un substrat ou un produit), une cellule bactérienne est 1000 fois plus grande qu'une petite molécule. On voit alors qu'il y a besoin d'intermédiaires pour "manipuler" les molécules et créer des cellules (*i.e.* se reproduire). C'est là que les protéines entrent en jeu. Si l'on imagine qu'un acide aminé fait 10^{-9} m, que l'on suppose qu'une protéine contient 1000 acides aminés, et que cette protéine est complètement recroquevillée, elle a alors la taille de $\sqrt[3]{1000} = 10$ acides aminés, soit $\pm 10^{-8}$ m. Étant dix fois plus grande que les petites molécules à manipuler, on peut dire que les protéines sont de bonnes candidates pour faire interagir des molécules entre elles et produire l'ensemble des composés cellulaires.

Une analogie rapide nous mène à considérer un atelier (une cellule) qui ferait une dizaine de mètres de longueur caractéristique. Cet atelier contient des outils (les protéines) qui sont 100 fois plus petits, de l'ordre de la dizaine de centimètres. Ces outils servent finalement à construire des ateliers complets (des cellules) en ayant à leur disposition des composés (les molécules) qui font quelques centimètres. La particularité de ces ateliers est que les outils travaillent tout seuls et qu'ils sont tous faits de la même manière comme nous allons le voir.

A.1.1.2 Temps caractéristiques intracellulaires

L'ordre de grandeur du taux de croissance μ défini équation (A.1.1) est de l'ordre de 0,1 à 1 h^{-1} pour des bactéries à croissance rapide comme *Escherichia coli* ou *Bacillus subtilis*. La limite avec le régime stationnaire se situe vers $0,1 \text{ h}^{-1}$ et en milieu particulièrement riche (notamment en acides aminés), le taux de croissance peut monter jusqu'aux alentours de $1,8 \text{ h}^{-1}$. Pour des milieux composés de glucose et d'éléments essentiels (ions, etc.), il atteint $0,5\text{-}0,6 \text{ h}^{-1}$. A $\mu_0 = 0,5 \text{ h}^{-1}$, ceci signifie que le temps pour doubler la population est de l'ordre de :

$$\Delta t = \frac{\ln(2)}{\mu_0} \approx 1,3h \quad (\text{A.1.2})$$

On peut prendre cette valeur comme la "durée de vie" d'une cellule, le temps entre deux duplications, aussi appelé temps de doublement.

Dans la cellule, certains composés sont très stables (comme les protéines ou l'ADN) et d'autres ont des vitesses de dégradation plus élevées (comme les ARN ou les petites molécules intermédiaires).

La dilution est le phénomène le plus lent intervenant dans les concentrations des composés les plus stables. Le temps de dédoublement est alors la constante de temps la plus grande à laquelle est sensible la variation de leurs concentrations. Dès lors, la constante de temps la plus grande à laquelle est sensible la variation de la concentration est alors le temps de doublement. On considérera ainsi que l'équilibre des concentrations intracellulaires en protéine est atteint au bout de 3 fois le temps de doublement (soit 4 heures pour une croissance en milieu riche en glucose exclusivement).

Les composés instables ont des dynamiques beaucoup plus rapides. Par exemple, la constante de temps associée à la dégradation des ARN est de quelques minutes (2-4 minutes).

Les concentrations des molécules intermédiaires utilisées par la cellule pour construire ses composants à partir du milieu de culture ont des constantes de temps aussi faibles.

En résumé, il y a plusieurs échelles de temps qui se juxtaposent au sein de la cellule. Celles qui nous intéressent sont principalement l'échelle rapide, de l'ordre de la minute, pour décrire les variations des petits composés, et l'échelle de l'ordre de l'heure pour les composés stables comme les protéines et le taux de croissance. On peut ajouter une échelle supplémentaire, de l'ordre de la dizaine d'heures pour décrire la culture en elle-même.

A.1.2 Un moyen d'action commun : les protéines

Chez les organismes vivants, les protéines jouent un rôle essentiel dans la réalisation des tâches élémentaires.

En effet, d'un point de vue thermodynamique, les réactions ayant lieu dans les cellules ont des énergies d'activation plus ou moins élevées. C'est pourquoi on a besoin de catalyseur pour rendre possible ces réactions. C'est le rôle des protéines.

A cette échelle, les réactions "chimiques" sont quasiment des réactions "physiques" : les molécules s'assemblent, s'échangent des électrons (réactions d'oxydoréduction), se transfèrent des groupements fonctionnels de l'une à l'autre, les liaisons covalentes peuvent se rompre, des groupements fonctionnels peuvent se réarranger pour créer des isomères, ... Toutes ces tâches sont accomplies en grande partie par les protéines, véritables ouvrières pouvant s'assembler en complexes pour effectuer des tâches plus élaborées.

Grâce à la coordination de toutes ces tâches élémentaires que l'ensemble des éléments de la cellule est construit.

A.1.2.1 Structure

Les protéines sont des successions de quelques centaines d'acides aminés. Les acides aminés sont des métabolites qui comportent un groupe -COOH , un groupe -NH_2 et une chaîne latérale. Les deux groupes -COOH et -NH_2 peuvent former des liaisons covalentes, permettant aux acides aminés d'être assemblés les uns aux autres comme une chaînette. Les chaînes latérales R interagissent entre elles (s'attirent ou se repoussent plus ou moins en fonction de leur charge, de leur polarité et de leur forme) pour former des structures en trois dimensions. Ces structures 3D peuvent prendre une infinité de formes en fonction de la succession d'acides aminés. En général, le repliement en 3D va laisser apparaître un site (une cavité) dans lequel la réaction est catalysée. C'est ce que l'on appelle le site actif.

A.2 Les processus cellulaires nécessaires à la croissance exponentielle

Nous nous focalisons sur la caractérisation des phases exponentielles car ce sont les régimes dans lesquels les cellules peuvent être utilisées pour la transformation de composés

chimiques. Nous présentons ici les principaux processus intracellulaires qui ont lieu en phase exponentielle et qui sont coordonnés pour assurer la croissance. Chaque processus est une tâche essentielle mettant en jeu les différents composants intracellulaires. Pour simplifier, une cellule bactérienne se compose d'une paroi extérieure, d'une membrane, d'un chromosome d'ADN, d'ARN, de lipides, de métabolites, d'ions et de protéines. Les ARN sont présents majoritairement sous trois formes : les ARN messagers (ARNm), les ARN ribosomiaux (ARNr), et les ARN de transfert (ARNt).

A.2.1 Échanges avec le milieu extérieur

L'origine de tout processus intracellulaire est l'import de matière depuis le milieu extracellulaire. Cet échange avec l'extérieur est primordial car il fournit les molécules nécessaires à la construction de tous les constituants de la cellule. L'interface physique avec le milieu extérieur est la paroi, mais l'interface fonctionnelle est la membrane. En effet, la paroi assure une structure à la cellule, en particulier contre les effets de la pression osmotique. En fonction de la pression osmotique du milieu extérieur, la cellule pourrait exploser ou imploser. La paroi s'oppose à ces effets. Par ailleurs, sa charge électrique permet d'attirer des ions et composés chargés nécessaires à la croissance de la cellule. La membrane n'a pas le même rôle. C'est ce que nous allons voir.

A.2.1.1 Structure de la membrane

La membrane est plus ou moins perméable aux molécules en fonction de leurs propriétés physico-chimiques des molécules. Elle se compose de deux couches de phospholipides, de longues molécules avec une tête hydrophile et une double queue hydrophobe. Elle comprend également des protéines (pour le transfert de matière et de signaux entre le milieu intérieur et extérieur) et des molécules comme des lipides (pour la rigidité). Parmi les protéines, on citera les transporteurs de nutriments, les oxydases et ATPases impliquées dans la respiration aérobie ou encore divers complexes de protéines impliqués dans la sécrétion de protéines. Parmi ces protéines se trouvent également celles assemblant la membrane elle-même.

La membrane désigne donc ce compartiment entre les deux couches de phospholipides. C'est un milieu relativement fluide qui permet la diffusion de composés. On parle du cytosol pour désigner le milieu intracellulaire (membrane exclue).

A.2.1.2 Différents types d'échanges avec le milieu extérieur

L'échange le plus élémentaire passe par la différence de potentiel chimique qui provoque un phénomène de diffusion. Ce phénomène est plus ou moins naturel en fonction des propriétés des molécules. Typiquement, la membrane est totalement perméable aux molécules hydrophobes (comme l' O_2 , le CO_2 , le N_2), plutôt perméable aux petites molécules non chargées (comme l'eau, l'urée, le glycérol), plutôt imperméable aux grosses molécules polaires non chargées (comme le glucose) et totalement imperméable aux ions et molécules chargées en général.

Les molécules auxquelles la membrane n'est pas totalement perméable peuvent être importées ou exportées par l'intermédiaire de protéines. Parmi elles, on peut distinguer les

protéines canaux et les protéines perméases. Les canaux permettent la diffusion de petites espèces chimiques, en particulier des ions. Les perméases (ou translocases) quant à elles assurent le transport de plus grosses molécules. Sans entrer dans les détails, parmi les perméases, on peut distinguer les transporteurs passifs (qui n'ont pas besoin d'énergie pour fonctionner) et les transporteurs actifs (qui ont besoin d'apport d'énergie, typiquement sous forme d'ATP (molécule d'adénosine triphosphate), pour faire passer des molécules d'un côté à l'autre).

A.2.2 Le métabolisme : synthèse des métabolites et macrocomposants

Le métabolisme désigne l'ensemble des réactions de catabolisme (réactions chimiques de dégradation de molécules qui fournissent de l'énergie) et d'anabolisme (réactions chimiques de synthèse/construction de molécules qui consomment de l'énergie) qui ont lieu dans la cellule. Cette énergie dont il est question s'échange entre le catabolisme et l'anabolisme au travers de la molécule d'ATP (adénosine triphosphate) essentiellement. La réaction associée est la suivante : $\text{ATP} \rightarrow \text{ADP} + \text{P}$, où ADP désigne l'adénosine diphosphate, P est un atome de phosphate.

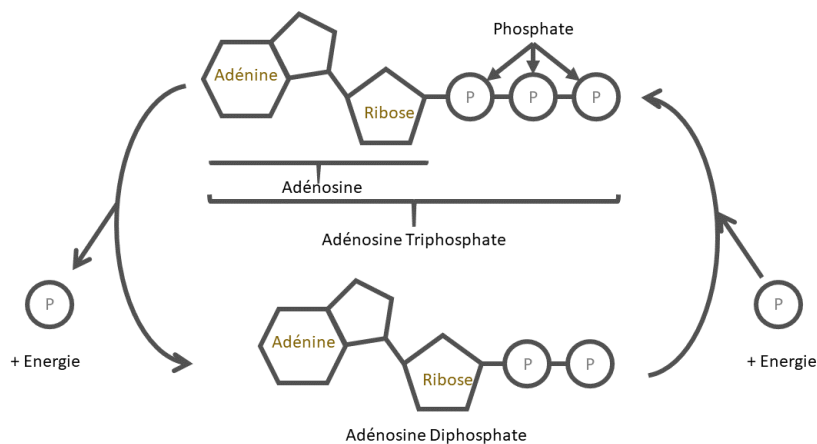


FIGURE A.1 – Réaction fondamentale d'échange d'énergie dans la cellule.

Cette relation physique est à la base des échanges énergétiques entre catabolisme et anabolisme. Une autre source d'énergie repose sur le couple guanosine triphosphate (GTP) et guanosine diphosphate (GDP) où une molécule de guanine remplace la molécule d'adénine dans la réaction de la figure A.1. Comme nous le verrons, cette source d'énergie est davantage utilisée pour la synthèse des protéines que pour le métabolisme.

A.2.2.1 Métabolites

Les petites molécules intervenant dans le métabolisme sont appelées des métabolites. Ce sont en quelque sorte des formes intermédiaires entre les constituants essentiels de la

cellule et les substrats présents dans l'environnement. D'un point de vue automatique, on peut suggérer de les voir comme des capteurs de l'état du métabolisme, *i.e.* des réactions qui ont lieu ou pas. Mais ces composés sont, par essence, voués à être transformés et assemblés pour former de plus grandes unités : les macrocomposants. Typiquement, les sucres (métabolites) s'assemblent pour former des polysaccharides, les acides gras (métabolites) pour former des lipides, les acides aminés (métabolites) pour former des protéines, les nucléotides (métabolites) pour former des acides nucléiques comme l'ADN (acide désoxybo-nucléique) ou des ARN (acides ribo-nucléiques).

A.2.2.2 Macrocomposants

Parmi les macrocomposants, on distingue principalement la paroi, la membrane, l'ADN, les ARN et les protéines. Nous présentons succinctement les mécanismes de synthèse de la paroi, de la membrane et de l'ADN avant de nous intéresser plus en détail à la synthèse des ARN et des protéines.

A.2.2.3 Synthèse de la paroi/membrane

La paroi est composée principalement de peptidoglycane, composé de chaînes carbonées et azotées, d'acides teichoïques et lipoteichoïques. Elle protège la cellule tout en permettant le passage des nutriments vers la membrane. La paroi est assemblée au cours de la phase de croissance par des complexes protéiques ancrés dans la membrane et permettant ainsi le passage des composants à travers celle-ci.

A.2.2.4 Synthèse de l'ADN

L'ADN est l'essence de l'information génétique d'un organisme. Il contient toute l'information génétique permettant le fonctionnement de l'organisme. Cette information génétique aussi appelée génome est divisé en gènes qui contiennent chacun l'information pour une fonction précise.

L'ADN bactérien est présent dans la cellule sous la forme d'un unique chromosome circulaire composé de deux brins formant une hélice et reliés par des liaisons hydrogènes. Chaque brin est composé de nucléotides correspondant aux 4 bases azotées connues : adénine (A), thymine (T), guanine (G) et cytosine (C). Les deux brins sont identiques à la correspondance suivante près :

$$\left\{ \begin{array}{l} A \rightarrow T \\ T \rightarrow A \\ G \rightarrow C \\ C \rightarrow G \end{array} \right. \quad (\text{A.2.1})$$

Lors de la croissance, un complexe de protéines appelé ADN-polymérase a pour rôle d'assembler des déoxy-nucléotides (métabolites) pour former une copie du chromosome. Cette opération est réalisée en ouvrant le chromosome et en complétant les deux brins simples par une succession de déoxy-nucléotides selon la règle (A.2.1).

A.2.3 Synthèse des ARN : la transcription

Les ARN sont des brins uniques de ribo-nucléosides. Ils sont composés des 4 bases A, U, C, G, c'est-à-dire les mêmes que l'ADN où l'Uracile (U) remplace la Thymine (T). Il existe trois grands types d'ARN : les ARN messagers, les ARN de transfert et les ARN ribosomiaux. Chacun a sa spécificité fonctionnelle, mais leur composition est la même. La transcription est le nom donné à la synthèse des ARN.

Au même titre que les autres réactions de synthèse de macrocomposants, la transcription est sujette à l'action d'un complexe de protéines appelé ARN-polymérase. L'ARN-polymérase se fixe sur une séquence particulière appelée promoteur et située en amont du gène (phase d'initiation). L'allongement de l'ARNm (phase d'élongation) s'effectue par insertion de nucléotides (A,U,C,G) par lecture directe du code génétique (respectivement T,A,G,C). La transcription s'arrête lorsque l'ARN-polymérase rencontre un terminateur (séquence particulière de bases). L'ARN-polymérase se décroche alors de l'ADN. L'ARN qui résulte de ce processus contient rigoureusement la même information que le gène qui a été parcouru. De plus, un même gène peut être transcrits simultanément par plusieurs ARN-polymérases.

L'ARN-polymérase est composée de plusieurs sous-unités protéiques et d'une petite protéine particulière appelée facteur σ . Le facteur σ reconnaît la séquence promotrice du gène et permet l'accrochage de l'ARN-polymérase sur le promoteur. Il est ensuite libéré dans le cytosol dès la fin de la phase d'initiation de la transcription. Il existe différents facteurs σ dans les bactéries qui sont en compétition pour former la forme active de l'ARN-polymérase capable de s'accrocher sur l'ADN. Tant que le facteur σ ne s'est pas détaché, la transcription n'est pas irréversible et l'ARN-polymérase peut se décrocher du promoteur. Le promoteur est donc une région déterminante pour la transcription du gène, en particulier pour le taux de production d'un ARN à partir d'un gène. Nous y reviendrons dans la partie A.3.

Le cas des opérons : Un opéron est une succession de gènes sur l'ADN qui ont le même promoteur et sont donc transcrits en même temps. Dès lors que l'ARN-polymérase a terminé la transcription du premier gène, elle initie la transcription du suivant, *et cætera*. Généralement, les gènes qui se trouvent en opérons codent pour des protéines qui ont des fonctions en cascade, par exemple des protéines impliquées dans la même voie métabolique. En plaçant des gènes en opérons, on force leur expression à être simultanée. Il suffit alors de réguler le promoteur de l'opéron pour réguler l'expression de tous les gènes de l'opéron.

Étant donné les interactions entre les bases A, U, C, G, il est probable que des séquences interagissent et que le brin se replie sur lui-même pour adopter une configuration de double brin par endroits, et de boucles à d'autres. C'est ce que l'on appelle la structure secondaire de l'ARN. Par ailleurs, des interactions peuvent avoir lieu entre des régions de la structure secondaire de l'ARN pour mener à des repliements en trois dimensions, c'est ce que l'on appelle la structure tertiaire de l'ARN. Ces structures sont importantes car elles permettent jouer le rôle de régulation.

A.2.4 Synthèse des protéines : la traduction

La traduction est le processus de synthèse de la protéine à partir de l'ARN messager. Le principal acteur impliqué est un énorme complexe appelé ribosome. Il est composé de deux sous-unités, appelées ribosome 30S et 50S, qui regroupent au total environ 50 protéines et trois ARNr. Le ribosome se fixe sur l'ARNm sur une séquence spécifique appelée *Ribosome Binding Sequence* (RBS) (phase d'initiation). Le ribosome se déplace le long de l'ARNm tout en produisant une chaîne peptidique. A chaque pas, le ribosome lit un codon (*i.e.* un triplet de nucléotides) dans la séquence codante et insère l'acide aminé correspondant à la fin de la chaîne peptidique grâce à des ARNt chargés (phase d'élongation). L'ARNt se compose d'une partie où l'acide aminé (AA) se fixe (ARNt chargé), et d'une partie appelée anticodon assurant la correspondance entre le code génétique porté par l'ARNm et l'acide aminé selon le tableau A.2. Enfin, lorsque le ribosome rencontre un codon particulier, le codon STOP, la traduction s'arrête. La chaîne polypeptidique est libérée dans le cytosol et le ribosome se décroche (phase de terminaison). Pour assurer la traduction du messenger, le ribosome est aidé par des facteurs accessoires : les facteurs d'initiation qui vont aider le ribosome à reconnaître le RBS et à engager la traduction, les facteurs d'élongation nécessaires à l'allongement de la chaîne polypeptidique et à la translocation du ribosome le long de l'ARNm, et enfin les facteurs de terminaison nécessaires au décrochage du ribosome à la fin du messenger. On parlera d'appareil de traduction l'ensemble des machineries moléculaires impliquées dans la traduction. Le ribosome a également besoin d'énergie, principalement sous forme de GTP selon la réaction $\text{GTP} \rightarrow \text{GDP} + \text{P}$, et ce pour les trois phases de la traduction.

Remarque : Chaque ARN de transfert ne peut se lier qu'à un seul acide aminé. Autrement dit, un codon correspond strictement à un seul acide aminé. La table donnée figure A.2 donne la correspondance entre les codons de l'ARN messager et les acides aminés de la protéine assemblée.

A l'image de la transcription, on notera que la phase d'initiation de la traduction est réversible. Le ribosome, et plus spécifiquement le ribosome 30S en complexe avec les facteurs d'initiation, se fixe au RBS de l'ARN messager, et cherche le premier codon (codon start) (cf. figure A.3). Lors de cette phase, le complexe 30S-facteurs d'initiation peut se décrocher. Lorsque l'ARNt correspondant au codon start se présente, la partie 50S est recrutée pour compléter le ribosome, les facteurs d'initiation sont relâchés. Cette étape est irréversible et conclut la phase d'initiation. L'élongation de la traduction peut commencer.

Après la traduction, la chaîne polypeptidique se replie sur elle-même pour former la protéine fonctionnelle. Pour certaines protéines, ce repliement ne s'effectue pas tout seul et nécessite d'autres protéines, appelées chaperonnes. Ces chaperonnes sont aussi nécessaires pour ajouter certains cofacteurs, en particulier les ions Fe^{2+} au sein de clusters fer-soufre.

		deuxième lettre					
		U	C	A	G		
première lettre	U	UUU } phénylanine	UCU } sérine	UAU } tyrosine	UGU } cystéine	U	troisième lettre
		UUC } leucine	UCC } sérine	UAC } tyrosine	UGC } cystéine	C	
		UUA } leucine	UCA } sérine	UAA } codons stop	UGA } codon stop	A	
		UUG } leucine	UCG } sérine	UAG } codons stop	UGG } tryptophane	G	
	C	CUU } leucine	CCU } proline	CAU } histidine	CGU } arginine	U	
		CUC } leucine	CCC } proline	CAC } histidine	CGC } arginine	C	
		CUA } leucine	CCA } proline	CAA } glutamine	CGA } arginine	A	
		CUG } leucine	CCG } proline	CAG } glutamine	CGG } arginine	G	
	A	AUU } isoleucine	ACU } thréonine	AAU } asparagine	AGU } sérine	U	
		AUC } isoleucine	ACC } thréonine	AAC } asparagine	AGC } sérine	C	
		AUA } méthionine	ACA } thréonine	AAA } lysine	AGA } arginine	A	
		AUG } méthionine	ACG } thréonine	AAG } lysine	AGG } arginine	G	
	G	GUU } valine	GCU } alanine	GAU } acide	GGU } glycine	U	
		GUC } valine	GCC } alanine	GAC } aspartique	GGC } glycine	C	
		GUA } valine	GCA } alanine	GAA } acide	GGA } glycine	A	
		GUG } valine	GCG } alanine	GAG } glutamique	GGG } glycine	G	

Ce tableau donne les diverses combinaisons possibles des 4 nucléotides pris 3 par 3 et leur « signification ».

FIGURE A.2 – Table de correspondance entre codon et acide aminé.

A.2.5 Résumé

Schématiquement, tous les processus cellulaires sont catalysés par des protéines. L'objectif du réseau métabolique est de produire les précurseurs métaboliques et l'énergie nécessaires à la croissance. Les autres processus cellulaires consomment des métabolites et de l'énergie pour synthétiser les différents composants cellulaires. Parmi ces processus, on notera le statut particulier de l'appareil de traduction (et des ribosomes) dont l'objectif est de produire des protéines. Il est bien évident que la cellule ne se limite pas aux seuls processus cités ci-dessus. On citera notamment les processus impliqués dans la maintenance de la pression osmotique, ou dans la prévention et la réponse au stress (oxydatif, thermique, ...) que nous n'avons pas décrit ici.

A.3 Régulation de l'expression des gènes

Différents mécanismes existent pour réguler la vitesse avec laquelle se déroule une réaction au sein d'une cellule. Nous ne prétendons pas en faire ici une revue complète mais nous allons nous focaliser sur les régulations qui ont lieu d'une manière générale sur les machineries moléculaires et principalement sur la synthèse des protéines. Les ARN étant une étape intermédiaire à la production des protéines, nous allons également nous intéresser aux mécanismes de régulation de la synthèse des ARN. Il y a plusieurs façons de réguler la production des protéines. Certains paramètres sont généraux à la cellule, d'autres sont particulier à un gène ou un messenger.

A.3.1 Régulation d'un groupe de gènes

Il existe différents mécanismes de régulation permettant de moduler l'expression d'un gène. On distinguera les mécanismes modulant l'initiation de l'élongation de la trans-

Si le type de facteur σ permet d'allumer ou d'éteindre des fonctions générales de la cellule, il existe au sein d'un même type de facteur σ des motifs de fixation différents. Chez la bactérie *B. subtilis*, 6 motifs différents pour le facteur σ_A ont été identifiés [62]. Ces familles de motifs permettent de moduler l'affinité de la polymérase pour le promoteur, et donc *in fine* de moduler la transcription du gène en fonction de la concentration de l'ARN-polymérase libre.

Cette caractéristique nous intéresse particulièrement dans la conception de souche car on dispose ici d'un moyen de pouvoir changer globalement les gènes qui sont exprimés, tout simplement en concevant des promoteurs adaptés à un facteur σ particulier et en exprimant ce facteur σ , qui est lui même une protéine.

Les facteurs de transcription (FT) : Un facteur de transcription est une protéine qui augmente (activateur) ou diminue (répresseur) l'affinité de l'ARN-polymérase pour son promoteur en se fixant sur des séquences spécifiques de l'ADN proches du promoteur. De plus, l'activité du facteur de transcription peut être modulée par un effecteur (généralement un métabolite). L'effecteur module l'accrochage du FT sur l'ADN. Dans certains cas, il peut stimuler ou au contraire empêcher l'accrochage du FT sur l'ADN. Le résultat final sur la transcription du gène dépend donc d'une part de l'effet positif ou négatif du FT, et d'autre part de l'effet positif ou négatif de l'effecteur sur le FT. Le régulon associé à un régulateur correspond à l'ensemble des gènes sous le contrôle direct de ce régulateur.

Les autres mécanismes de régulation génétique : Les mécanismes modulant l'élongation de la transcription, autrefois minoritaires et peu étudiés, jouent en fait un rôle central notamment dans la régulation des voies métaboliques. Le principe général de ces régulations est le suivant. Après l'initiation, l'ARNm naissant possède deux motifs particulier, un terminateur précoce, et un motif anti-terminateur (ou anti-anti-terminateur) sur lequel l'élément régulateur va se fixer. L'effet de l'élément régulateur dépend de la stabilité de l'ARNm naissant vis à vis de la boucle de terminaison. Soit le terminateur est stable, et dans ce cas l'accrochage de l'élément régulateur va permettre de le déstabiliser et ainsi d'autoriser la transcription. Soit le terminateur est instable, et la transcription a lieu en l'absence d'élément régulateur. Les principaux éléments régulateurs identifiés chez *Bacillus subtilis* sont :

- les ARNt non chargés : la transcription a lieu lorsque l'ARNt non chargé s'accroche sur l'ARNm ;
- les métabolites : la transcription s'arrête généralement lorsque le métabolite est fixé sur l'ARNm ;
- des protéines : ici il n'y a pas de règles générales concernant l'effet.

Le cas des gènes “constitutifs” : Toutefois la majorité des gènes d'une bactérie comme *Bacillus subtilis* ne sont soumis à aucune régulation génétique. On dit alors qu'ils sont constitutifs. Le seul mécanisme de régulation qui agit sur leur niveau d'expression est l'affinité entre le promoteur du gène et l'ARN-polymérase et l'affinité entre les sous-unités 30S des ribosomes et la zone RBS du messenger. Chaque gène possède une affinité propre du promoteur vis-à-vis de l'ARN-polymérase, et du RBS vis-à-vis du ribosome 30S. Comme les concentrations des ARN-polymérases libres et des ribosomes libres

dépendent du taux de croissance, l'expression du gène est également modulée par le taux de croissance. A l'inverse des autres mécanismes de régulation qui peuvent dépendre de conditions extérieures, la régulation de l'expression des gènes par le taux de croissance est littéralement codée "en dur" dans la séquence ADN. Les nucléotides du promoteur du gène et les nucléotides en amont de la partie codante d'un ARN messager sont donc déterminants sur le niveau d'expression d'une protéine. Leur choix peut entièrement spécifier l'expression d'une protéine particulière en fonction du taux de croissance. Grâce à ce mécanisme, une cellule change de configuration en fonction du taux de croissance, toute chose restant égale par ailleurs.

A.3.2 Régulation globale : la réponse stringente

Parmi les mécanismes globaux de régulations, les concentrations d'ARN-polymérases et de ribosomes jouent donc un rôle important. Plus précisément, en modulant les concentrations d'ARN-polymérases libres et de ribosomes libres, l'expression de tous les gènes de la cellule est impactée.

La concentration de GTP joue aussi un rôle important car c'est cette source d'énergie qui est surtout utilisée pour la synthèse des protéines. Diminuer la concentration de GTP revient d'une manière ou d'une autre à diminuer le nombre de ressources dirigées vers l'appareil de traduction.

C'est ce que l'on observe en cas de pénurie de ressource ou de stress. Une protéine, nommée *relA*, permet de mesurer l'activité des ribosomes. En cas de pénurie de ressource, la quantité d'ARNt non chargée augmente. Statistiquement, cela signifie qu'un nombre important d'ARNt non chargé entre dans le site actif du ribosome. Lorsqu'un ARNt non chargé est détecté, la protéine *RelA* produit du guanosine pentaphosphate, ppGpp à partir de GTP. La concentration de GTP diminue et avec elle la production des ribosomes. Il y a moins de ressources utilisées pour produire des ribosomes, leur concentration diminue jusqu'à ce que la concentration de l'appareil de traduction soit en accord avec le chargement des ARN de transfert. Ce mécanisme de régulation est aussi appelé "réponse stringente" dans la littérature. D'un point de vue de l'automatique, on peut considérer ici que la concentration en ribosome est asservie à une valeur maximale donnée par la capacité du métabolisme.

Annexe B

Un exemple d'utilisation du principe du maximum de Pontryaguine

On présente dans cette annexe les étapes de résolution d'un problème d'optimisation pour la maximisation de la production d'un composé d'intérêt, en s'inspirant d'un des exemples présentés dans [72].

On suppose une croissance de biomasse qui suit une allure de Andrews et une production de composé d'intérêt qui est tout simplement proportionnelle à la concentration de biomasse en présence. En partant d'un réacteur n'ayant pas de substrat, la stratégie optimale est de se placer à la concentration qui maximise le taux spécifique de croissance ($G = \sqrt{K_i K_G}$, cf. formule (1.1.25)), puis lorsque la biomasse a atteint une certaine concentration, de basculer vers $G = 0$ pour arrêter la croissance. La production étant proportionnelle à la biomasse, il y a bien production dans cette dernière phase.

B.1 Position du problème

On considère le modèle de bioprocédé suivant :

$$\left\{ \begin{array}{lcl} \dot{X}(t) = \mu_m \frac{G(t)}{G(t) + K_G + \frac{G(t)^2}{K_I}} X(t) & - & \frac{X(t)F_{in}(t)}{V(t)} \\ \dot{P}(t) = \pi X(t) & - & \frac{P(t)F_{in}(t)}{V(t)} \\ \dot{G}(t) = -\alpha \mu_m \frac{G(t)}{G(t) + K_G + \frac{G(t)^2}{K_I}} X(t) - \beta \pi X(t) & + & \frac{(G_F - G(t))F_{in}(t)}{V(t)} \\ \dot{V}(t) = F_{in}(t) & & \end{array} \right. \quad (\text{B.1.1})$$

avec μ_m , K_G , K_I , π , α , β et G_F des paramètres.

On pose

$$\mu(G(t)) = \mu_m \frac{G(t)}{G(t) + K_G + \frac{G(t)^2}{K_I}} \quad (\text{B.1.2})$$

Et on cherche le profil $F_{in}(t)$ qui maximise la concentration de produit à l'instant

final :

$$\max_{F_{in}(t), t_f} J = \max_{F_{in}(t), t_f} P(t_f) \quad (\text{B.1.3})$$

On considère les contraintes suivantes sur le débit d'alimentation et l'instant final :

$$\begin{cases} 0 \leq F_{in}(t) \leq F_{MAX}, \forall t \\ 0 \leq t_f \leq t_{MAX} \end{cases} \quad (\text{B.1.4})$$

On doit ajouter des contraintes de modélisation sur la positivité des concentrations des entités ($X(t)$, $P(t)$ et $G(t) \geq 0$). Toutefois, il suffit de contraindre $G(t) \geq 0$ pour que le taux spécifique de croissance μ soit positif et donc que $X(t)$ soit a fortiori positif si $X(t_0) > 0$. Si $X(t)$ est positif, $P(t)$ l'est aussi.

On ajoute simplement la contrainte :

$$G(t) \geq 0, \forall t \quad (\text{B.1.5})$$

B.2 Détermination des conditions nécessaires d'optimalité de Pontryaguine

L'objectif des conditions nécessaires de Pontryaguine est de caractériser la solution optimale au problème défini ci-dessus. Pour ce faire, on ajoute des contraintes pour définir un nouveau problème pour lequel une solution au problème initial est nécessairement solution. C'est une sorte d'extension des conditions de Lagrange ou de Karush-Kuhn-Tucker, pour des problèmes d'optimisation en dimension infinie.

On pose donc le $\mathcal{H}$, l'Hamiltonien défini comme suit :

$$\begin{aligned} \mathcal{H}(x, u, \lambda, \eta) = & \lambda_X(\mu(G)X - X \frac{F_{in}}{V}) + \lambda_P(\pi X - P \frac{F_{in}}{V}) \dots \\ & + \lambda_G(-\alpha\mu(G)X - \beta\pi X + (G_F - G) \frac{F_{in}}{V}) + \eta G \end{aligned}$$

où $x(t)$ est le vecteur d'état (X, P, G, V), $u(t)$ la commande du système ($F_{in}(t)$), $\lambda_{x_i}(t)$ le multiplicateur de l'état x_i , $\eta(t)$ le multiplicateur associé à la contrainte sur l'état à respecter à chaque instant $G(t) \geq 0$. Les multiplicateurs $\lambda(t)$ sont absolument continus, tandis que η est simplement continu par morceaux.

Les conditions nécessaires d'optimalité de Pontryaguine s'écrivent de la manière suivante : Si $F_{in}^*(t)$ est une trajectoire optimale au problème défini à la section B.1, alors elle est solution du problème suivant :

$$\begin{aligned}
& \max_{F_{in}(t), t_f} \mathcal{H}(x, u, \lambda, \eta) \\
& \text{sous contraintes, } \left\{ \begin{array}{l}
\dot{X} = \mu(G)X - \frac{XF_{in}}{V} \\
\dot{P} = \pi X - \frac{PF_{in}}{V} \\
\dot{G} = -\alpha\mu(G)X - \beta\pi X + \frac{(G_F - G)F_{in}}{V} \\
\dot{V} = F_{in} \\
X(t_0) = X_0 \\
P(t_0) = P_0 \\
G(t_0) = G_0 \\
V(t_0) = V_0 \\
0 \leq F_{in}(t) \leq F_{MAX} \\
0 \leq t_f \leq t_{MAX} \\
G(t) \geq 0 \\
\eta(t) \geq 0 \\
\eta(t)G(t) = 0 \\
\dot{\lambda}_X = -\frac{\partial \mathcal{H}}{\partial X} \\
\dot{\lambda}_P = -\frac{\partial \mathcal{H}}{\partial P} \\
\dot{\lambda}_G = -\frac{\partial \mathcal{H}}{\partial G} \\
\dot{\lambda}_V = -\frac{\partial \mathcal{H}}{\partial V} \\
\lambda_X(t_f) = \frac{\partial J}{\partial X}(t_f) \\
\lambda_P(t_f) = \frac{\partial J}{\partial P}(t_f) \\
\lambda_G(t_f) = \frac{\partial J}{\partial G}(t_f) \\
\lambda_V(t_f) = \frac{\partial J}{\partial V}(t_f)
\end{array} \right. \quad (B.2.1)
\end{aligned}$$

En omettant la dépendance au temps par souci de lisibilité et avec la notation $\dot{x} \triangleq \frac{dx}{dt}$.

Ainsi, la dynamique des états adjoints λ s'écrit :

$$\left\{ \begin{array}{l}
\dot{\lambda}_X = -\frac{\partial \mathcal{H}}{\partial X} = -\lambda_X\mu(G) + \lambda_X\frac{F_{in}}{V} - \lambda_P\pi + \lambda_G\alpha\mu(G) + \lambda_G\beta\pi \\
\dot{\lambda}_P = -\frac{\partial \mathcal{H}}{\partial P} = +\lambda_P\frac{F_{in}}{V} \\
\dot{\lambda}_G = -\frac{\partial \mathcal{H}}{\partial G} = -\lambda_X X\mu'(G) + \lambda_G\alpha\mu'(G)X + \lambda_G\frac{F_{in}}{V} + \eta \\
\dot{\lambda}_V = -\frac{\partial \mathcal{H}}{\partial V} = -\frac{1}{V^2}(\lambda_X XF_{in} + \lambda_P PF_{in} + \lambda_G(G - G_F)F_{in})
\end{array} \right. \quad (B.2.2)$$

avec $\mu'(G) = \frac{d\mu}{dG}$, et avec comme conditions finales :

$$\begin{cases} \lambda_X(t_f) = \frac{\partial J}{\partial X}(t_f) = 0 \\ \lambda_P(t_f) = \frac{\partial J}{\partial P}(t_f) = 1 \\ \lambda_G(t_f) = \frac{\partial J}{\partial G}(t_f) = 0 \\ \lambda_V(t_f) = \frac{\partial J}{\partial V}(t_f) = 0 \end{cases} \quad (\text{B.2.3})$$

Par ailleurs, traduisons le fait que F_{in} maximise $\mathcal{H}$. On voit que $\mathcal{H}$ est affine en F_{in} . Ainsi, F_{in} maximise $\mathcal{H}$, revient à la disjonction suivante :

- Soit $\frac{\partial \mathcal{H}}{\partial F_{in}} > 0$ et alors sans condition supplémentaires, $F_{in}(t) = F_{MAX}$;
 - Soit $\frac{\partial \mathcal{H}}{\partial F_{in}} < 0$ et alors sans condition supplémentaires, $F_{in}(t) = 0$;
 - Soit $\frac{\partial \mathcal{H}}{\partial F_{in}} = 0$ et alors $0 < F_{in}(t) < F_{MAX}$.
- avec

$$\mathcal{H}_{F_{in}} \triangleq \frac{\partial \mathcal{H}}{\partial F_{in}} = -\frac{\lambda_X X}{V} - \frac{\lambda_P P}{V} + \frac{\lambda_G (G_F - G)}{V} + \lambda_V \quad (\text{B.2.4})$$

B.3 Caractérisation de l'arc singulier

Caractérisons la trajectoire, appelée arc singulier, $t \mapsto F_{in,s}(t)$ qui correspond au cas où $\mathcal{H}_{F_{in}} = 0$. On s'intéresse au cas où la contrainte sur l'état $G(t) \geq 0$ n'est pas saturée. On suppose donc dans cette partie que $\eta(t) = 0 \forall t$.

Si $\mathcal{H}_{F_{in}}(t) = 0$ sur un intervalle de temps de mesure non nulle, alors, sa dérivée temporelle est aussi nulle sur l'intérieur de cet intervalle :

$$\frac{d}{dt}(\mathcal{H}_{F_{in}}) = 0 \quad (\text{B.3.1})$$

Pour faciliter les calculs, on peut dériver plutôt $V(t)\mathcal{H}_{F_{in}}$. En effet, si $\mathcal{H}_{F_{in}} = 0$ et $\frac{d}{dt}(\mathcal{H}_{F_{in}}) = 0$, alors,

$$\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}}) = \frac{d}{dt}(V(t)) \underbrace{\mathcal{H}_{F_{in}}}_{=0} + V(t) \underbrace{\frac{d}{dt}(\mathcal{H}_{F_{in}})}_{=0} = 0$$

par ailleurs,

$$\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}}) = \frac{d}{dt}(-\lambda_X X - \lambda_P P + \lambda_G (G_F - G) + \lambda_V V)$$

c'est-à-dire,

$$\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}}) = -\dot{\lambda}_X X - \lambda_X \dot{X} - \dot{\lambda}_P P - \lambda_P \dot{P} + \dot{\lambda}_G (G_F - G) - \lambda_G \dot{G} + \dot{\lambda}_V V + \lambda_V \dot{V}$$

en substituant chaque dérivée temporelle par son expression,

$$\begin{aligned}
\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}}) &= \lambda_X \mu(G)X - \lambda_X \frac{F_{in}}{V}X + \lambda_P \pi X - \lambda_G \alpha \mu(G)X - \lambda_G \beta \pi X \\
&\quad - \lambda_X \mu(G)X + \lambda_X X \frac{F_{in}}{V} \\
&\quad - \lambda_P \frac{F_{in}}{V}P - \lambda_P \pi X + \lambda_P P \frac{F_{in}}{V} \\
&\quad + (G - G_F) \lambda_X X \mu'(G) - (G - G_F) \alpha \lambda_G \mu'(G)X - (G - G_F) \lambda_G \frac{F_{in}}{V} \\
&\quad + \lambda_G \alpha \mu(G)X + \lambda_G \beta \pi X - \lambda_G (G_F - G) \frac{F_{in}}{V} \\
&\quad + \lambda_V F_{in} - \lambda_X X \frac{F_{in}}{V} - \lambda_P P \frac{F_{in}}{V} - \lambda_G (G - G_F) \frac{F_{in}}{V}
\end{aligned}$$

Avec les simplifications suivantes :

$$\begin{aligned}
\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}}) &= \underbrace{\lambda_X \mu(G)X}_{=\heartsuit} \underbrace{- \lambda_X \frac{F_{in}}{V}X}_{=\diamond} \underbrace{+ \lambda_P \pi X}_{=\square} \underbrace{- \lambda_G \alpha \mu(G)X}_{=\circ} \underbrace{- \lambda_G \beta \pi X}_{=\triangle} \\
&\quad \underbrace{- \lambda_X \mu(G)X}_{=-\heartsuit} \underbrace{+ \lambda_X X \frac{F_{in}}{V}}_{=-\diamond} \\
&\quad \underbrace{- \lambda_P \frac{F_{in}}{V}P}_{=\spadesuit} \underbrace{- \lambda_P \pi X}_{=-\square} \underbrace{+ \lambda_P P \frac{F_{in}}{V}}_{=-\spadesuit} \\
&\quad + (G - G_F) \lambda_X X \mu'(G) - (G - G_F) \alpha \lambda_G \mu'(G)X - (G - G_F) \lambda_G \frac{F_{in}}{V} \\
&\quad \underbrace{+ \lambda_G \alpha \mu(G)X}_{=-\circ} \underbrace{+ \lambda_G \beta \pi X}_{=\triangle} \underbrace{- \lambda_G (G_F - G) \frac{F_{in}}{V}}_{=\clubsuit} \\
&\quad + \lambda_V F_{in} - \lambda_X X \frac{F_{in}}{V} - \lambda_P P \frac{F_{in}}{V} \underbrace{- \lambda_G (G - G_F) \frac{F_{in}}{V}}_{=-\clubsuit}
\end{aligned}$$

i.e.

$$\begin{aligned}
\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}}) &= (G - G_F) \lambda_X X \mu'(G) - (G - G_F) \alpha \lambda_G \mu'(G)X \\
&\quad \underbrace{F \left(-(G - G_F) \lambda_G \frac{1}{V} + \lambda_V - \lambda_X X \frac{1}{V} - \lambda_P P \frac{1}{V} \right)}_{=\mathcal{H}_{F_{in}}=0}
\end{aligned}$$

ainsi,

$$\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}}) = (G - G_F) \lambda_X X \mu'(G) - (G - G_F) \alpha \lambda_G \mu'(G)X$$

ou encore,

$$\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}}) = (G - G_F) X \mu'(G) (\lambda_X - \alpha \lambda_G) \quad (\text{B.3.2})$$

Dès lors, la commande $F_{in,s}(t)$ qui annule $\frac{d}{dt}(V(t)\mathcal{H}_{F_{in}})$ annule l'un des termes du produit.

- Soit $G(t) = G_F$ sur un intervalle de temps, et alors d'après la dynamique de G , ceci revient à $X(t) = 0$.
- Soit $X(t) = 0$, et alors il n'y a pas biomasse, donc pas de bioprocédé.
- Soit $\mu'(G) = 0$, et alors $G = \sqrt{K_i K_G}$, d'après la relation (1.1.25).
- Soit $\lambda_X = \alpha \lambda_G$.

Montrons que ce dernier cas conduit à une commande hors du domaine de faisabilité. Pour ce faire, on applique le résultat présenté dans [72] en calculant l'ordre de singularité ξ relatif au terme $\lambda_X - \alpha \lambda_G$ du produit (B.3.2). Dans notre cas, ξ est le plus petit l tel que $\frac{d^l}{dt^l}(\lambda_X - \alpha \lambda_G) \neq 0$. Si $\xi = -\infty$, alors la commande $F_{s,in}$ associée n'est pas faisable.

Si $\lambda_X = \alpha \lambda_G$, alors en dérivant cette relation par rapport au temps,

$$\dot{\lambda}_X(t) = \alpha \dot{\lambda}_G(t)$$

En substituant $\dot{\lambda}$ par leur expression venant de (B.2.2) :

$$-\lambda_X \mu(G) + \lambda_X \frac{F_{in}}{V} - \lambda_P \pi + \lambda_G \alpha \mu(G) + \lambda_G \beta \pi = \alpha(-\lambda_X X \mu'(G) + \lambda_G \alpha \mu'(G) X + \lambda_G \frac{F_{in}}{V})$$

En simplifiant par $\lambda_X(t) = \alpha \lambda_G(t)$,

$$\beta \lambda_G = \lambda_P$$

Mais alors, les évolutions de λ_X et λ_G par rapport au temps s'écrivent :

$$\begin{cases} \dot{\lambda}_X = \lambda_X \frac{F}{V} \\ \dot{\lambda}_G = \lambda_G \frac{F}{V} \end{cases}$$

Ce qui signifie que

$$\frac{d}{dt}(\lambda_X + \alpha \lambda_G) = (\lambda_X + \alpha \lambda_G) \frac{F}{V} = 0$$

Ainsi, par argument de récurrence, quel que soit l ,

$$\frac{d^l}{dt^l}(\lambda_X + \alpha \lambda_G) = 0$$

On trouve alors que l'ordre de singularité associé au terme $\lambda_X - \alpha \lambda_G$ est $\xi = -\infty$. Ainsi, la commande associée à ce terme n'est pas réalisable. En d'autres termes, on exclut le cas $\lambda_X = \alpha \lambda_G$ de la recherche d'arc singulier.

Pour conclure cette caractérisation d'arc singulier, un seul cas est possible : $F_{in}(t) = F_{in,s}(t)$ telle que $G = \sqrt{K_I K_P}$. Exprimons alors $F_{in,s}(t)$.

Si $G(t) = \sqrt{K_I K_P}$ sur un intervalle de temps non nul, alors,

$$\dot{G}(t) = 0 = -\alpha \underbrace{\mu(\sqrt{K_I K_P})}_{=\mu_{max}} X(t) - \beta \pi X(t) + \frac{(G_F - \sqrt{K_i K_G}) F_{in,s}(t)}{V(t)}$$

ou encore

$$\frac{(G_F - \sqrt{K_i K_G})}{V(t)} F_{in,s}(t) = (\alpha\mu_{max} + \beta\pi)X(t)$$

finalemt,

$$F_{in,s}(t) = \frac{\alpha\mu_{max} + \beta\pi}{G_F - \sqrt{K_i K_G}} X(t)V(t) \quad (\text{B.3.3})$$

La commande singulière $F_{in,s}(t)$ est proportionnelle à la quantité de biomasse $X(t)V(t)$.

B.4 Caractérisation de la commande de suivi de contrainte sur l'état

On caractérise ici la commande $F_{in,c}(t)$ qui sature la contrainte sur l'état $G(t) = 0$ sur un intervalle de mesure non nulle.

Si $G(t) = 0$ sur un intervalle de temps non nul, alors :

$$\dot{G}(t) = 0 = -\alpha\mu(0)X(t) - \beta\pi X(t) + \frac{(G_F - 0)F_{in,c}(t)}{V(t)}$$

i.e.

$$F_{in,c}(t) = \frac{\beta\pi}{G_F} X(t)V(t)$$

De plus,

$$\dot{X}(t) = \mu(0)X(t) - X(t)\frac{F_{in,c}(t)}{V(t)}$$

donc,

$$\dot{X}(t) = -X(t)\frac{\dot{V}(t)}{V(t)}$$

c'est-à-dire,

$$X(t)V(t) = \text{constante}$$

On pose $X(t)V(t) = c_X$. De sorte que,

$$F_{in,c}(t) = \frac{\beta\pi}{G_F} c_X = \text{constante} \quad (\text{B.4.1})$$

La commande $F_{in,c}$ qui sature la contrainte $G(t) = 0$ est une commande constante.

B.5 Caractérisation de la trajectoire optimale

On peut conclure cette étude en disant que $F(t)$ est une succession de phases de différents types :

- $F_{in}(t) = 0$
- $F_{in}(t) = F_{in,s}(t)$ donnée en (B.3.3)
- $F_{in}(t) = F_{in,c}(t)$ donnée en (B.4.1)
- $F_{in}(t) = F_{MAX}$

On trouve numériquement la succession correcte et les instants de bascule entre les différentes stratégies sur F .

Dans ce cas, avec les valeurs numériques du tableau B.1, on trouve que la succession optimale est :

- 1- $F_{in}(t) = F_{MAX}$
- 2- $F_{in}(t) = F_{in,s}(t)$
- 3- $F_{in}(t) = F_{MAX}$
- 4- $F_{in}(t) = 0$
- 5- $F_{in}(t) = F_{in,c}(t)$

TABLE B.1 – Valeurs numériques utilisées, tirées de [72]).

Paramètre	Valeur	(unité)
μ_m	0,53	(h^{-1})
K_G	1,2	($mmol.L^{-1}$)
K_I	1,2	($mmol.L^{-1}$)
π	0,5	($mmol.h^{-1}.g^{-1}$)
α	2,5	($mmol.g^{-1}$)
β	1	($mmol.mmol^{-1}$)
G_F	20	($mmol.L^{-1}$)
X_0	1	($g.L^{-1}$)
P_0	0	($mmol.L^{-1}$)
G_0	0	($mmol.L^{-1}$)
V_0	2	(L)
F_{MAX}	1	($L.h^{-1}$)
t_{MAX}	10	(h)

B.6 Représentation de la trajectoire optimale

La figure B.1 représente l'allure de la trajectoire optimale obtenue par le programme Bocop [10]. Dans cet exemple, on retrouve la succession des phases de croissance et de production dont on a parlé au long de la thèse. On distingue les 5 phases présentées ci-dessus :

- Dans un premier temps, la commande est maximale de sorte à ce que la concentration G augmente et atteigne $\sqrt{K_i K_G} = 5 \text{ mmol.L}^{-1}$. Le taux de croissance μ augmente également pour atteindre sa valeur maximale.
- Ensuite, on applique la commande singulière $F_{in,s}(t)$ pour maintenir la concentration G à sa valeur $\sqrt{K_i K_G}$ et μ à sa valeur maximale. L'allure $F_{in,s}$ est exponentielle, c'est bien le produit entre $X(t)$ et $V(t)$.
- La troisième phase correspond de nouveau à $F_{in}(t) = F_{MAX}$. En réalité, il s'agit de la commande singulière $F_{in,s}$ qui bute sur la contrainte $F_{in}(t) \leq F_{MAX}$.
- La quatrième phase $F_{in}(t) = 0$ permet de faire chuter G et ainsi d'interrompre la croissance μ . On passe d'un régime de croissance à un régime de production.

- Une fois la concentration de G nulle, on applique la commande $F_{in,c}$ qui maintient cette contrainte d'état saturée. Comme attendu, $F_{in,c}$ est bien constante. On constate que dans cette dernière phase, la concentration X est décroissante. Cette décroissance est due à la dilution de la biomasse (à l'augmentation du volume). La quantité $X(t)V(t)$ est constante.

Par ailleurs, on constate que les multiplicateurs λ satisfont bien les conditions finales (B.2.3).

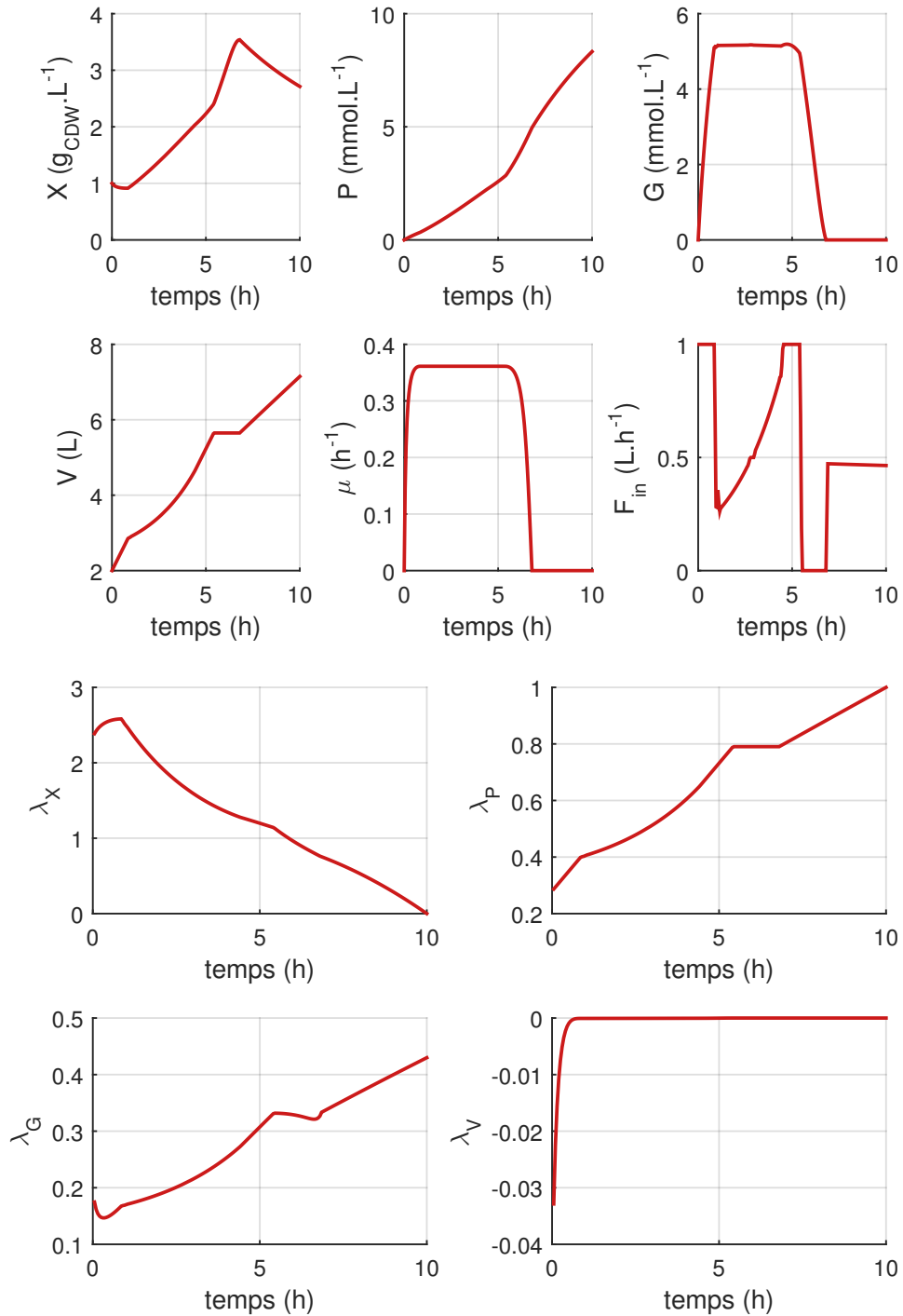


FIGURE B.1 – Trajectoire optimale obtenue par une méthode numérique de paramétrisation complète.

Titre : Optimisation de la conception de bioprocédés : vers une approche intégrée biologie de synthèse et conduite du procédé

Mots clés : Bioprocédé, Optimisation, Conduite du procédé, Conception de souche

Résumé : La conception de souches efficaces pour la production de composés d'intérêt offre des potentiels immenses qui restent trop peu exploités par manque de lien entre les étapes d'optimisation de la conception de souche et celle de la conduite du bioprocédé.

Pour combler ce manque, cette thèse propose une description des bioprocédés intégrant pleinement le fonctionnement interne des micro-organismes impliqués dans la production de composés d'intérêt. Cette description permet d'optimiser simultanément la souche et le procédé pour maximiser la production d'un composé d'intérêt en respectant les contraintes attachées à ces deux étapes.

Dans un premier temps, une nouvelle classe de modélisation de bioprocédés est développée, à l'interface entre les modèles intracellulaires de gestion de ressources et les modèles macroscopiques usuellement utilisés dans la commande de bioprocédés en bioréacteurs. Dans un second temps, des contraintes liées à l'implémentation biologique de la stratégie de contrôle sont intégrées au problème. Ceci permet d'obtenir une conception plus réaliste du point de vue de l'ingénierie des génomes. Enfin, la dernière partie de la thèse montre que la méthodologie présentée jusqu'alors sur un modèle agrégé peut être étendue à des représentations détaillées du comportement des micro-organismes.

Title: Bioprocess design optimization : towards an integrated approach of synthetic biology and process control

Keywords: Bioprocess, Optimization, Process control, Strain design

Abstract: The design of efficient strains for the production of compounds of interest offers tremendous potentials that remain insufficiently exploited due to the lack of link between the optimization stages of strain design and that of bioprocess control.

This thesis proposes a description of bioprocesses that fully integrates the internal functioning of micro-organisms involved in the production of compounds of interest. This description allows the strain and process to be optimized simultaneously to maximize the production of a compound of interest while respecting the

constraints attached to these two stages.

First, a new bioprocess modelling class is developed at the interface between intracellular resource allocation models and macroscopic models commonly used in bioprocess control. In a second stage, constraints linked to the biological implementation of the control strategy are integrated into the problem. This provides a more realistic genome engineering design. Finally, the last part of the thesis shows that the methodology presented so far on an aggregate model can be extended to detailed representations of the behaviour of micro-organisms.

