Modélisation mathématique de la dégradation des hydrates de carbone dans le côlon humain

Rafael Munoz Tamayo

N° D'ORDRE

THÈSE DE DOCTORAT

SPECIALITE : PHYSIQUE

Ecole Doctorale « Sciences et Technologies de l'Information des Télécommunications et des Systèmes »

Présentée par : Rafael Muñoz-Tamayo

Sujet **: Modélisation Mathématique de la Dégradation des Hydrates de Carbone dans le Côlon Humain**

Soutenue le 22 mars 2010 devant les membres du jury :

M Denis Dochain	Rapporteur
M Karel Keesman	Rapporteur
M Robbert Kleerebezem	Examinateur
Mme Béatrice Laroche	Encadrant de thèse
Mme Marion Leclerc	Encadrant de thèse
M Silviu-Iulian Niculescu	Examinateur
M Éric Walter	Directeur de thèse

Mathematical Modelling of Carbohydrate Degradation in the Human Colon

Rafael Muñoz Tamayo

Director: Éric Walter
Supervisors:
Béatrice Laroche
Marion Leclerc
Kiên Kiêu

France, 2010

A Olguita y sus mariposas

¿Por qué la tierra es mi casa?
¿Por qué la noche es oscura?
¿Por qué la luna es blancura
que engorda como adelgaza?
¿Por qué una estrella se enlaza
con otra, como un dibujo?
Y ¿por qué el escaramujo
es de la rosa y del mar?
Silvio Rodríguez

Contents

Acknowledgments * Remerciements * Agradecimientos

Je tiens tout d'abord à remercier les équipes de recherche qui ont fait que mon travail soit possible. Ils sont : l'Unité d'Ecologie et Physiologie du système Digestif (UEPSD), département MIcrobiologie et Chaîne Alimentaire (MICA), l'unité Mathématiques et Informatique Appliquées (MIA), département MIA de l'Institut national de la recherche agronomique (INRA), ainsi que le Laboratoire des Signaux et Systèmes (L2S) et le projet AlimIntest. Merci pour m'avoir pourvu des excellentes conditions de travail.

Naturellement, ce chapitre est consacré à dire merci, mais pas un merci comme ce qu'on utilise tous les jours. Non, moi je veux dire un merci avec un chapeau à tous les gens qui m'on soutenu dans ce projet. Donc, je veux dire $\widehat{Merci}$.

Je me sens très honoré d'avoir été encadré par Béatrice Laroche, Marion Leclerc et Éric Walter. Nous avons formé une équipe de travail intéressante, où j'ai pu étendre mon horizon de connaissance et personnel. Je leur dois quelques changements dans ma façon de penser, et ça je l'apprécie vraiment. Merci à eux pour avoir eu aussi l'intérêt et le temps de me connaître, pour l'indépendance et la confiance.

Béatrice, je te remercie pour l'enthousiasme que tu as mis dans ce projet, pour ta patience à m'expliquer tant des choses sur l'univers de la théorie de systèmes. Merci pour le temps partagé dans lequel nous avons discuté d'autres choses que des équations. Il a été très important et réconfortant de pouvoir compter toujours sur ton soutien personnel

et ton amitié.

Marion, je te remercie pour ton attitude critique et *open minded*, pour voir plus loin de l'écran, pour tes idées éloignées de la ligne standard, pour chercher des justifications et des arguments biologiques aux équations. Merci pour toutes les discussions au sujet de métabolisme bactérien. Merci pour ton amitié, pour les échanges musicaux et pour ton intérêt à me faire connaître la culture française.

Éric, travailler et partager avec toi a été une expérience très enrichissante. J'ai été toujours impressionné par ta curiosité et ta connaissance de multiples sujets et ton sens de l'humour. Je te remercie pour ton exigence, tes appels au monde factuel, tes critiques et ton rigueur. Merci de m'avoir appris à analyser ma thèse au-delà de son contexte spécifique.

Je voudrais remercier spécialement Jean-Philippe Steyer (LBE-INRA-Narbonne). Merci Jean-Phi pour avoir ouvert cette porte pour moi. Je te remercie sincèrement pour ton aide et accompagnement, ton soutien et ta confiance. Merci d'avoir toujours eu l'intérêt de discuter de mon travail et aussi pour le diffuser dans la communauté scientifique de la digestion anaérobie. Ces lignes sont vraiment insuffisantes.

J'ai eu de la chance de faire partie de plusieurs équipes de recherche. Chacun d'entre elles m'as apporté une diversité des choses.Je voudrais remercier tout l'ensemble de l'UEPSD, dirigé par Joël Doré pour son accueil. Merci pour enrichir ma connaissance en biologie. C'était un plaisir de partager avec vous ce temps, où j'ai appris de vous plus que ce que vous pouvez imaginer. Joël, je te remercie pour toutes les discussions et les innombrables réponses à mes questions. Je veux remercier Hervé Blottière pour les discussions autour des AGCC. Merci aux gens avec qui j'ai partagé au jour le jour, Patou, Laure, Antonella, Karine, Omar, Stan, Julien (merci pour nos discussions sur ton noyau). Je veux remercier Florence, Fabienne et Charlène, qui m'ont expliqué les détails sur les expériences *in vivo*.

Je tiens à remercier l'unité MIA, dirigée par Alain Trubuil. Merci pour les bons moments autour du café du matin. Je remercie Kiên Kiêu, qui a été toujours disponible à discuter et m'aider. Merci pour l'échange d'idées. Alain, je te remercie pour ton intérêt

sur mon sujet et ta disposition à discuter. Éric et Fabrice, merci pour la bonne ambiance et le support informatique. Merci Caro pour les tips de Matlab et LaTeX. A Suzanne y Jean-Baptiste, todo mi aprecio por la amistad, el intercambio de palabras e imágenes.

Toute ma gratitude à Violaine Rochet et Gaël Kozlowski qui on rendu mon arrivé plus facile. Merci Gaël pour les moments en Jazz. Merci à Xavier, Mathieu et toute l'équipe de volley de l'INRA pour la bonne ambiance.

During my thesis I had the opportunity to exchange ideas, collaborate and discuss with people from various institutes. In particular, I would like to thanks Dr. Sylvia Duncan and Dr. Harry J. Flint (Microbial Ecology Group, Rowett Institute of Nutrition and Health, UK) for the discussion we had about butyrate metabolism. It was very motivating to meet Dr. Roderick Mackie (Microbiology Department of Animal Sciences, University of Illinois, USA) with who I could discuss about bacterial metabolism. I would like to express my gratitude to Dr. Ivan Ramirez (LBE-INRA) and Dr. Jorge Rodríguez (Sustainable Environment Research Centre, University of Glamorgan, UK) for kindly providing their source codes of anaerobic digestion modelling. Juan Rodrigo Bastidas (LBE-INRA), gracias por tu hospitalidad en Narbonne y los incontables artículos que me enviaste.

Este viaje no hubiese sido el mismo sin los amigos encontrados. A Lina y Eduardo, muchas gracias por la amistad de siempre. Liliana, te agradezco tu apoyo incondicional, las conversaciones y la sensación de estar con alguien de la familia. Andrés, gracias por nuestras divagaciones y las horas compartidas entre lo que creemos cercano al silencio. Tomas, gracias por los cafés y los buenos miércoles. Fabrizio, Luis, les agradezco por la alegría del tiempo compartido. Cynthia y Noemi, gracias por el aprecio sincero. A mis amigos del ayer y a los invisibles.

Oriana, por ser en mi y conmigo.

Mi familia es el pilar de mi vida. Para agradecerles su apoyo y amor, tal vez otro lenguaje sería necesario.

Context of the thesis

This work was developed as a multidisciplinary collaboration between the following research teams:

- Unité d'Ecologie et Physiologie du Système Digestif (UEPSD), Institut National de la Recherche Agronomique (INRA) Jouy-en-Josas
- Unité de Mathématiques et Informatique Appliquées (MIA), INRA Jouy en Josas
- Laboratoire des Signaux et Systèmes (L2S), Univ Paris Sud-CNRS-SUPÉLEC

The thesis work belonged to the project AlimIntest ANR-PNRA, which focused on the develop and use of tools to describe the effects of dietary fibers on the structure and functions of the human intestinal microbiota.

Abstract

The human colon is an anaerobic bioreactor with an enormous number of microorganisms, which collectively are named the human colonic microbiota. This community performs the breakdown of polysaccharides that are not digested in the upper intestine, mediating some of the effects of diet on health. As a result of its limited accessibility and its complexity, the human colon remains largely unexplored.

Mathematical models can play a central role for a better insight of the human colon complexity. In this context, this work presents the development of a mathematical model of carbohydrate degradation. Our aim was to provide an *in silico* approach to contribute to a better understanding of the fermentation pattern in such an ecosystem. Our mathematical model is knowledge-based, derived by writing down mass-balance equations. It incorporates physiology of the intestine, metabolic reactions and transport phenomena. As a result, it has a complex structure with a large number of state variables and parameters. In order to handle such a complexity, we made use of mathematical and engineering tools. They included theoretical concepts on identification and model order reduction, and practical aspects of parameter estimation and parameter reduction by knowledge integration. Model simulations provided an adequate qualitative representation of the human colon. The methodology that we have contributed to develop should enable one to assess nutritional scenarios and to address questions that are difficult or impossible to elucidate by experimentation.

Our model is a promising tool in predicting the fermentation pattern by human colonic microbiota. It should help in development of strategies aimed at promoting healthy gut function.

Keywords: anaerobic digestion; complex systems; human colon; mathematical modelling; parameter estimation

Résumé

Le côlon humain (ou gros intestin) fait partie du tractus gastro-intestinal. Il constitue un écosystème anaérobie, dans lequel on trouve de nombreuses espèces de microorganismes, donc l'ensemble forme le microbiote intestinal. Cette communauté est composée de plus de 1000 espèces (Rajilic-Stojanovic et al., 2007). Ce microbiote est spécifique à l'hôte et il a pour fonction la dégradation des polysaccharides qui n'ont pas été digérés dans la partie supérieure de l'intestin. Il joue de ce fait le rôle de médiateur de l'effet du régime alimentaire sur la santé (Flint et al., 2007).

Malgré l'importance de son rôle dans la santé humaine, le côlon reste un organe mal connu en raison de sa complexité et de son accessibilité difficile. Les études *in vivo* sur l'humain sont en effet restreintes pour des raisons éthiques et techniques. Ces contraintes ont motivé le développement de modèles *in vitro* pour simuler le comportement du côlon humain (voir, par exemple, Molly et al. (1994) ; Macfarlane et al. (1998a)). Les modèles *in vitro* ont été particulièrement utiles dans l'étude de la fermentation des hydrates de carbone provenant de différents substrats. Cependant, la plupart d'entre eux ne prend pas en compte des facteurs biotiques tel que l'absorption des acides gras à chaîne courte (AGCC), l'interaction avec l'hôte, et les échanges entre les micro-habitats du système que sont le lumen et le mucus intestinal. Parallèlement à ces études *in vitro*, des modèles animaux (principalement des rongeurs et des cochons) ont aussi été utilisés pour étudier différents phénomènes comme l'absorption des AGCC (Fleming et al., 1991), la distribution spatiale bactérienne (Sarma-Rupavtarm et al., 2004), et les interactions symbiotiques hôte-bactérie (Samuel and Gordon (2006) ; Mahowald et al. (2009)).

La contribution des modèles *in vitro* et *in vivo* sur modèles animaux a été cruciale dans la compréhension de l'organisation et du fonctionnement du tractus gastro-intestinal humain. Il faut néanmoins être prudent dans l'extrapolation de ces études au côlon humain. Les différences physiologiques entre le côlon des animaux et celui de l'être humain sont importantes. De plus, les approches *in vitro* et *in vivo* utilisent des cultures bactériennes provenant de matériels fécaux. Or seulement 20 à 40% du microbiote du côlon humain peut être cultivé (Suau et al., 1999) et des différences majeures entre le microbiote fécal et le microbiote du côlon ont été mises en évidence.

Les modèles mathématiques peuvent grandement contribuer à une meilleure compréhension de la dynamique du côlon humain. Les travaux ici présentés ont été consacrés au développement d'un modèle de la dégradation des hydrates de carbone dans le côlon humain. Nous avons construit un modèle à base de connaissances. Notre objectif a donc été de contribuer, par une approche *in silico*, à une meilleure compréhension du processus de fermentation dans cet écosystème microbien.

Malgré ses spécificités, notre cas d'étude est représentatif d'une grande classe de systèmes biologiques d'intérêt où les données expérimentales sont très limitées et où plusieurs sources d'information hétérogènes doivent être utilisées pour obtenir un modèle utile.

Description du côlon humain

Du point de vue biochimique, le côlon humain est un environnement caractérisé par un potentiel redox bas et une température contrôlée constante de 37°C (Savage, 1977). Il est consistué de trois régions anatomiques bien définies qui sont: le côlon proximal (cecum et côlon ascendant), le côlon transversal et le côlon distal (sigmoïde et rectum). Le côlon est alienté par les effluents en provenance de l'iléon. Le contenu colonique est retenu dans le côlon proximal et le chyme va dans le côlon distal où il est, finalement, dégradé ou excrété.

Les hydrates de carbone constituent environ 85% des substrats disponibles pour le processus de fermentation dans le côlon. Les substrats sont principalement d'origine alimentaire. Les hydrates de carbone principaux sont l'amidon résistant (RS) et les polysaccharides d'amidon non-résistant (NSP). La première étape de dégradation des hydrates de

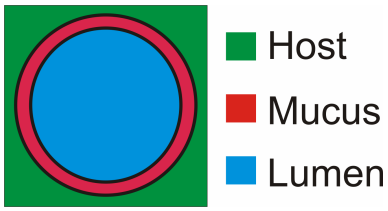

Figure 1: Représentation schématique de la section transversale du côlon humain. Le côlon comprend le lumen et le mucus. L'épithélium est l'interface des interaction entre l'hôte et l'intestin.

carbone est l'hydrolyse des polysaccharides. L'hydrolyse est effectuée par des hydrolases associées aux cellules bactériennes. Les polymères sont décomposés en monomères, qui sont ensuite assimilés par les micro-organismes dans la suite du processus de fermentation. Les principaux produits de la fermentation sont les AGCC, le CO_2, l'hydrogène, la biomasse, et le méthane chez certains individus. Ce processus de fermentation confère au système une distribution spatiale de métabolites et un profil de pH avec des valeurs qui sont approximativement de 5.5 dans le côlon proximal, 6.2 dans le côlon transversal et 6.9 dans le côlon distal (Macfarlane and Cummings, 1991).

L'intestin est délimité par une paroi épithéliale, dont les cellules sécrètent une couche de mucus. Le mucus a les caractéristiques d'une matrice polymérique qui permet l'attachement de micro-organismes et leur résistance à l'arrachement (Sonnenburg et al., 2004). C'est également une source de carbone qui peut contribuer à la survie des bactéries intestinales *in vivo* en l'absence de fibres alimentaires (Macfarlane and Cummings, 1991). Le microbiote du gros intestin est donc principalement localisé dans deux micro-habitats, le lumen et le mucus (voir la figure reffig:corteA). Des études montrent de plus que les espèces bactériennes présentes dans le mucus sont distinctes de celles présentes dans le lumen (Zoetendal et al. (2002) ; Eckburg et al. (2005)).

Base des connaisances

Grâce à la similarité de fonctionnement entre le côlon humain et les réacteurs anaérobie pour le traitement des eaux usées, nous avons pu tirer profit des efforts de modélisation déjà accomplis dans le domaine des réacteurs anaérobie pour construire notre modèle. Le

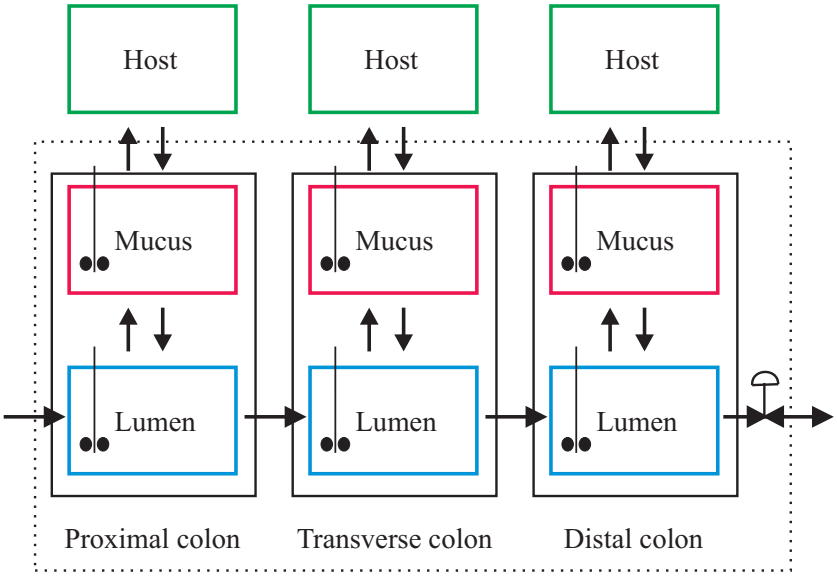

Figure 2: Représentation hydraulique du côlon humain. Le rectangle pointillé représente la limite du système modélisé.

modèle de digestion anaérobie ADM1 (Batstone et al., 2002) a été utilisé comme point de départ. Nous avons modifié sa structure pour prendre en compte des caractéristiques spécifiques du côlon humain, telles que la présence dans le processus de fermentation d'une étape d'acétogénèse réductrice ou des spécificités du métabolisme du lactate.

Dans notre modèle, le côlon humain est divisé en trois régions qui correspondent aux trois régions physiologiques. Dans chacune d'elles, on distingue le lumen et le mucus (figure 2). Le lumen et le mucus sont modélisés comme des réacteurs à alimentation continue parfaitement mélangés. Comme dans les réacteurs à haut rendement (voir, par exemple, Lettinga et al. (1980)), les micro-organismes du côlon humain forment des agrégats denses pour résister aux forces hydrodynamiques et prévenir le phénomène de "lessivage". Ces agrégats sont constitués de microbes, de particules alimentaires, de cellules épithéliales et de mucus (Sonnenburg et al., 2004). Dans notre modèle, afin de prendre en compte l'agrégation microbienne, nous considérons que les micro-organismes ont un temps de résidence plus élevé que le temps de résidence hydraulique.

Le microbiote du côlon humain catalyse un réseau de transformations métaboliques com-

plexes pour dégrader les sources de carbone disponibles. Dans notre modèle, ces transformations sont représentées de façon agrégée. Le microbiote est associé au micro-habitat où il réside, lumen ou mucus. Il est structuré en groupes fonctionnels jouant chacun un rôle dans la chaîne trophique. Ainsi, à chaque consommation de substrat est associé un groupe microbien qui catalyse la transformation. Outre les réactions biochimiques, on observe dans le côlon humain des phénomènes de transport tels que l'absorption de métabolites. L'absorption de l'eau et celle des métabolites sont indispensables pour maintenir une physiologie normale du système (Minekus et al., 1999). Dans notre modèle, nous considérons qu'il se produit une diffusion des sucres entre le lumen et le mucus, une absorption d' AGCC, de lactate et d'eau dans le lumen et le mucus, un transfert entre les phases liquide et gazeuses, et un phénomène de détachement pour le microbiote de la muqueuse.

La figure 3 montre un schéma qui intègre la dégradation des hydrates de carbone et les phénomènes du transport, en accord avec Macfarlane et al. (2006) et Louis et al. (2007b). Les figures 2 et 3 résument donc les bases sur lesquelles repose notre modèle.

Formulation mathématique

Nous avons divisé le côlon humain en ses trois régions physiologiques. Chacune de ces régions comprend deux compartiments, un pour le lumen et un pour le mucus. Notre modèle comprends donc six compartiments. Les équations du modèle s'obtiennent en écrivant les bilans de masse des métabolites et des groupes microbiens dans chacun des compartiments, en supposant que les concentrations dans chaque compartiment sont homogènes. Le système d'équations différentielles ordinaires obtenu a comme vecteur d'état

$$\begin{aligned}\xi = [&s_{su},\ s_{la},\ s_{H_2},\ s_{ac},\ s_{pro},\ s_{bu}, s_{CH_4},\ s_{CO_2},\ s_{H_2O},\ z,\\ &x_{su},\ x_{la},\ x_{H_2a},\ x_{H_2m},\ s_{g,H_2},\ s_{g,CH_4},\ s_{g,CO_2}]^{T},\end{aligned}$$

où s_i est la concentration du composant soluble i, z est la concentration en polysaccarides complexes, x_i est la concentration du groupe microbien i, et $s_{g,i}$ la concentration du composant i dans la phase gazeuse. Notre modèle a donc 102 ($17\times2\times3$) variables d'état.

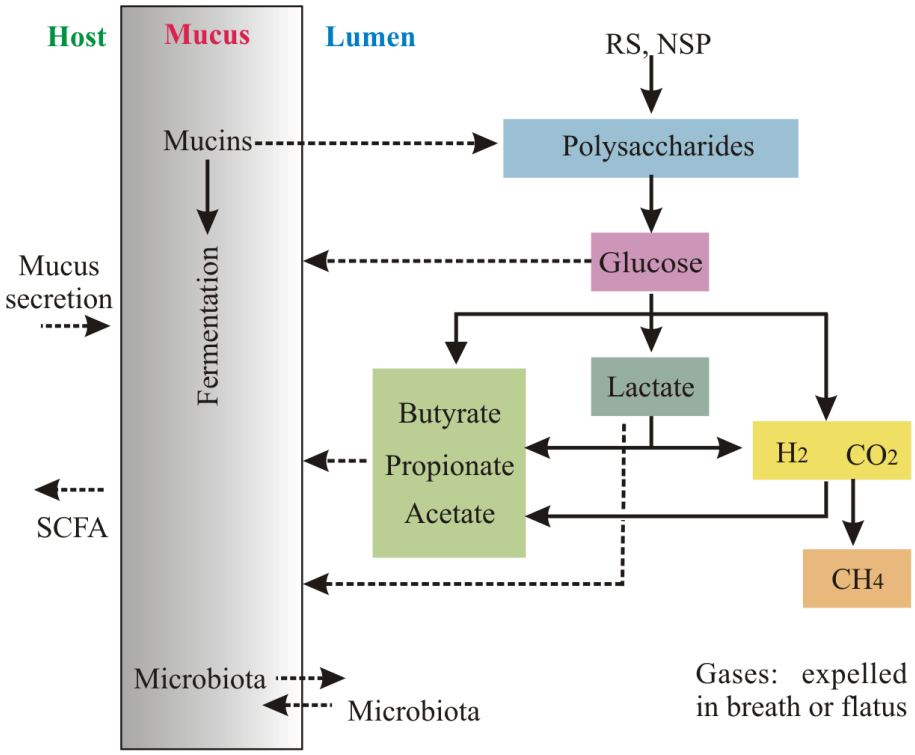

Figure 3: Schéma réactionnel de la dégradation des hydrates de carbone dans le côlon humain, et transferts de masse entre le lumen et le mucus et entre le mucus et l'hôte. Les lignes pointillées représentent des phénomènes du transport, et les lignes continues des transformations métaboliques.

Détaillons, par exemple, les équations associées au lumen dans le côlon proximal. Elles sont définies par

$$\dot{s}_i^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} s_{i,\mathrm{in}}^{\mathrm{l}} - \frac{q_{\mathrm{out}}}{V^{\mathrm{l}}} s_i^{\mathrm{l}} - \gamma_i^{\mathrm{l}} s_i^{\mathrm{l}} + \sum_{j=1}^{5} Y_{i,j}^{\mathrm{l}} \rho_j(\xi^{\mathrm{l}}) - Q_i^{\mathrm{l}}, \tag{1}$$

$$\dot{z}^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} z_{\mathrm{in}}^{\mathrm{l}} + \frac{q_{\mathrm{out}}^{\mathrm{m}}}{V^{\mathrm{l}}} z^{\mathrm{m}} - \frac{q_{\mathrm{out}}}{V^{\mathrm{l}}} z^{\mathrm{l}} - \rho_1(\xi^{\mathrm{l}}), \tag{2}$$

$$\dot{x}_i^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} x_{i,\mathrm{in}}^{\mathrm{l}} - \frac{1}{\tau_i + V^{\mathrm{l}}/q_{\mathrm{out}}} x_i^{\mathrm{l}} + b_i \frac{V^{\mathrm{m}}}{V^{\mathrm{l}}} x_i^{\mathrm{m}} - a_i x_i^{\mathrm{l}} + \sum_{j=2}^{9} Y_{i,j}^{\mathrm{l}} \rho_j(\xi^{\mathrm{l}}), \tag{3}$$

où V est le volume du compartiement et q est le flux à l'entrée (in) et à la sortie (out) du système. Les exposants l et m désignent les micro-habitats "lumen" et "mucus". Tous les flux sont des fonctions positives de l'état du système, sauf celui provenant du tractus digestif supérieur, qui est une entrée exogène. Q est le taux de transfert liquide-gaz et

γ_i est un coefficient du transport. Le coefficient a_i se rapporte à l'adhésion de groupes microbiens provenant du lumen au mucus. Le coefficient b_i se rapporte au détachement des micro-organismes du mucus. τ_i est le temps de résidence supplémentaire du groupe microbien x_i mentionné plus haut. La cinétique du processus de transformation j est dénoté par $\rho_j(\xi)$. $Y_{i,j}$ est le rendement du composant i dans le processus j. La cinétique de l'hydrolyse des polysaccharides $\rho_1(\xi)$ est décrite par l'équation de Contois pour prendre en compte la dépendance de cette transformation en la concentration des bactéries qui hydrolysent les sucres longs. Ensuite, la cinétique d'utilisation des autres substrats est décrite par l'équation de Monod. La mortalité des biomasses est supposée suivre une cinétique du premier ordre.

Le bilan de masse dans la phase gazeuse donne

$$\dot{s}_{\mathrm{g},i} = \frac{q_{\mathrm{g,in}}}{V_{\mathrm{g}}} s_{\mathrm{g},i,\mathrm{in}} - \frac{q_{\mathrm{g,out}}}{V_{\mathrm{g}}} s_{\mathrm{g},i} + Q_i \frac{V}{V_{\mathrm{g}}}, \tag{4}$$

où V_{g} et q_{g} sont le volume et le flux dans la phase gazeuse. Le taux de transfert liquide-gaz est représenté par la loi de Henry

$$Q_i = k_L a(s_i - K_{H,i} RT s_{\mathrm{g},i}). \tag{5}$$

Les équations du modèle pour le mucus sont obtenues de la même façon. Tous les compartiments du mucus ont la même structure. Les compartiments du lumen dans les sous-modèles du côlon proximal et du côlon transversal ont la même structure. Dans le côlon distal, le lumen est modélisé comme un réacteur semi-batch, donc l'équation de la dynamique du volume est incluse. Le modèle complet comporte 333 paramètres.

Stratégie face à un modèle complexe

Notre modèle comprend un grand nombre de variables d'état et de paramètres. Sa mise en œuvre, son analyse et sa simulation ne sont par conséquent pas des tâches simples. En plus de la complexité, l'insuffisance des données expérimentales est un réel obstacle à l'estimation des paramètres. Pour contourner ces difficultés, nous avons tout d'abord collecté des informations sur certains des paramètres en procédant à l'identification de sous-modèles correspondant à des données expérimentales *in vitro*. Nous avons en outre utilisé deux approches pour simplifier le modèle. L'exploitation des relations stoïchiométriques

nous a permis de réduire le nombre des paramètres inconnus, tandis que la méthode des perturbations singulières nous a permis de réduire le nombre des variables d'état. Ces approches sont décrites ci-dessous.

Exploitation des données expérimentales disponibles

Il y a peu de données expérimentales disponibles concernant la dégradation des hydrates de carbone dans le côlon humain. La plupart des connaissances sur le sujet proviennent d'expériences *in vitro* de croissance bactérienne. Ces études fournissent des informations sur des réactions spécifiques de la voie métabolique complète, et permettent l'estimation de leurs paramètres cinétiques. Nous avons réalisé la modélisation et l'identification de deux systèmes *in vitro* qui avaient pour objectif l'analyse de la cinétique de l'homoacétogénèse (Bernalier et al., 1996b) pour l'un, et l'analyse de la production de butyrate par des bactéries du côlon humain (Duncan et al., 2004b) pour l'autre.

Homoacétogénèse par la bactérie du côlon humaine *Ruminococcus hydrogenotrophicus*

L'homoacétogénèse est une des trois voies pour l'utilisation de hydrogène par le microbiote du côlon humain. Cette réaction joue un rôle central dans le transfert d'hydrogène entre micro-organismes, qui est une condition indispensable au bon déroulement de la fermentation (Thauer et al., 1977). On a démontré l'existence de liens entre l'excès de production de H_2 et des maladies comme la maladie de Crohn. Cet excès provoque des douleurs abdominales, des distensions, des flatulences et des borborygmes (Boever et al., 2000). De plus, la consommation de H_2 par les homoacétogènes, les réducteurs de sulfate et les méthanogènes (Archaea) représente la dernière étape d'extraction d'énergie dans la dégradation des hydrates de carbone. Elle affecte aussi la thermodynamique du processus biochimique dans son ensemble (Bäckhed et al. (2005) ; Ley et al. (2006a)).

L'étude cinétique effectuée par Bernalier et al. (1996b) a été réalisé sur l'espèce bactérienne *Ruminococcus hydrogenotrophicus*, cultivée dans des conditions de température contrôlées. Le substrat était un mélange gazeux H_2/CO_2, qui a diffusé dans la phase liquide pour être utilisé comme source d'énergie par les bactéries. L'acétate produit l'a été par la voie réductrice. Dans ce métabolisme autotrophique, les bactéries réduisent deux

moles de CO_2 à une mole d'acétate. La cinétique a été suivie en mesurant la concentration de H_2 dans la phase gazeuse, la concentration d'acétate, et la densité optique de cultures à 600 nm (OD_{600}).

Le modèle cinétique est

$$\dot{x}_{\mathrm{H_2a}} = \mu_{\mathrm{H_2a}} - k_d x_{\mathrm{H_2a}}, \tag{6}$$

$$\dot{\chi} = k_d x_{\mathrm{H_2a}} - k_i \chi, \tag{7}$$

$$\dot{s}_{\mathrm{g,H_2}} = k_L a (s_{\mathrm{H_2}} - K_{H,\mathrm{H_2}} RT s_{\mathrm{g,H_2}}) \frac{V}{V_{\mathrm{g}}}, \tag{8}$$

$$\dot{s}_{\mathrm{ac}} = \frac{1 - Y_{\mathrm{H_2a}}}{Y_{\mathrm{H_2a}}} \mu_{\mathrm{H_2a}}, \tag{9}$$

$$\dot{s}_{\mathrm{H_2}} = -\frac{1}{Y_{\mathrm{H_2a}}} \mu_{\mathrm{H_2a}} - k_L a (s_{\mathrm{H_2}} - K_{H,\mathrm{H_2}} RT s_{\mathrm{g,H_2}}), \tag{10}$$

où V est le volume dans la phase liquide, V_{g} le volume dans la phase gazeuse, et $\mu_{\mathrm{H_2a}}$ la cinétique de croissance bactérienne, exprimée par la loi de Monod,

$$\mu_{\mathrm{H_2a}} = \mu_{\mathrm{max,H_2a}} \frac{s_{\mathrm{H_2}} x_{\mathrm{H_2a}}}{K_{s,\mathrm{H_2a}} + s_{\mathrm{H_2}}}. \tag{11}$$

Les cinq variables d'état sont la concentration $x_{\mathrm{H_2a}}$ de la biomasse active, la concentration χ des bactéries inactives qui ont encore une structure cellulaire, la concentration $s_{\mathrm{g,H_2}}$ de l'hydrogène dans la phase gazeuse, la concentration s_{ac} de l'acétate, et la concentration $s_{\mathrm{H_2}}$ de l'hydrogène dans la phase liquide. Toutes les concentrations sont exprimées en kilogrammes de demande chimique d'oxygène (COD) par mètre cube (kg COD m^{-3}). Les valeurs numériques des paramètres $k_L a$ et $K_{H,\mathrm{H_2}}$ ont été tirés de Batstone et al. (2002).

Le vecteur de sortie du modèle à l'instant t_i est

$$\mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta}) = (\alpha(x_{\mathrm{H_2a}}(t_i) + \chi(t_i)),\ s_{\mathrm{g,H_2}}(t_i),\ s_{\mathrm{ac}}(t_i))^{\mathrm{T}}, \tag{12}$$

où α est un facteur de conversion entre la densité optique (OD) et les kg COD m^{-3}. Les cinq paramètres inconnus à identifier sont $\mu_{\mathrm{max,H_2a}}$, $K_{s,\mathrm{H_2a}}$, k_d, k_i, et $Y_{\mathrm{H_2a}}$. Le vecteur des données collectés à l'instant t_i est modélisé par

$$\mathbf{y}(t_i) = \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta}^*) + \boldsymbol{\epsilon}_i,\ i = 1, \ldots, n_{\mathrm{t}}, \tag{13}$$

où n_{t} est le nombre d'instants d'observation et $\boldsymbol{\theta}^*$ la valeur vraie du vecteur des paramètres. On suppose que les vecteurs d'erreurs de la mesure $\boldsymbol{\epsilon}_i$ $(i = 1, \ldots, n_{\mathrm{t}})$ sont indépendants

et identiquement distribués. Ils suivent une loi Gaussienne de moyenne nulle, donc $\boldsymbol{\epsilon}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

Avant d'estimer les paramètres du modèle, nous avons étudié leur identifiabilité. Comme le modèle n'est pas commandé et est non linéaire, nous avons utilisé une variante de l'approche par isomorphisme d'état locale qui peut être trouvée dans Denis-Vidal and Joly-Blanchard (2004). Nous avons montré que le modèle est structuralement globalement identifiable. Pour estimer les paramètres de sous-modèles, nous avons développé la boîte à outils Matlab IDEAS (Muñoz-Tamayo et al., 2009), qui réalise l'estimation de paramètres au sens du maximum de vraisemblance pour des modèles multi sorties, sous diverses hypothèses concernant le bruit de mesure. IDEAS calcule aussi les fonctions de sensibilité et la matrice d'information de Fisher à l'optimum, permettant de ce fait la quantification approchée de l'incertitude sur les valeurs estimées.

Nous avons identifié les paramètres du modèle en minimisant plusieurs critères qui correspondent à différentes hypothèses sur le bruit de mesure. Les meilleurs résultats ont été obtenus sous l'hypothèse que $\boldsymbol{\Sigma}$ est inconnue et diagonale. Dans ce cas, l'estimateur au sens du maximum de vraisemblance de $\boldsymbol{\theta}$ minimise

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{n_{\mathrm{y}}} \frac{n_{\mathrm{t}_k}}{2} \ln \left[\sum_{i=1}^{n_{\mathrm{t}_k}} [y_k(t_{i_k}) - y_{\mathrm{m}_k}(t_{i_k}, \boldsymbol{\theta})]^2 \right], \tag{14}$$

où t_{i_k} est le i-ième instant de mesure pour la sortie y_k. La valeur estimée de la matrice de covariance du bruit au sens de maximum de vraisemblance est

$$\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}(\widehat{\sigma}_1^2, \cdots, \widehat{\sigma}_{n_{\mathrm{y}}}^2), \tag{15}$$

avec

$$\widehat{\sigma}_k^2 = \frac{1}{n_{\mathrm{t}_k}} \sum_{i=1}^{n_{\mathrm{t}_k}} [y_k(t_{i_k}) - y_{\mathrm{m}_k}(t_{i_k}, \widehat{\boldsymbol{\theta}})]^2. \tag{16}$$

Le modèle cinétique de l'homoacétogénèse, identifié avec IDEAS, représente de façon satisfaisante les données expérimentales (figure 4). Cependant, les intervalles de confiance pour les paramètres de l'équation Monod, à savoir le taux de croissance spécifique maximal ($\mu_{\mathrm{max,H_2a}}$) et la constante de Monod ($K_{s,\mathrm{H_2a}}$), sont très mauvais (voir la deuxième colonne de la table 1). Ce résultat n'est pas vraiment surprenant, à cause d'une part du

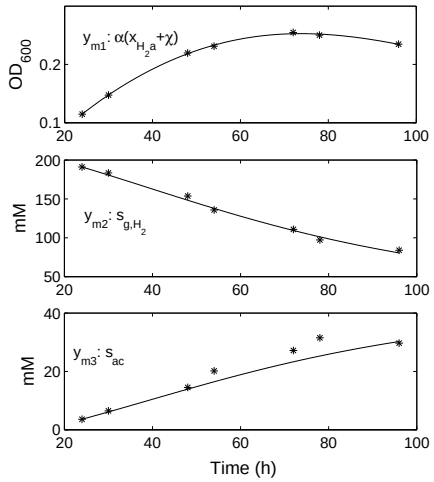

Figure 4: Modèle identifié pour l'homoacétogénèse *in vitro*. * : données expérimentales, trait plein : sorties du modèle.

nombre limité de données et d'autre part de problèmes d'identifiabilité pratique connus pour ce type de modèle (voir, par exemple, Dochain and Vanrolleghem (2001)). La simulation du modèle avec les valeurs estimées des paramètres indique que les valeurs prises par la variable d'état s_{H_2} sont toujours bien inférieures à la valeur du paramètre K_{s,H_2a}. À la fin de l'expérience, la simulation fournit une valeur de $s_{H_2} = 0.0038$ qui est 7 fois inférieure à la valeur estimée de K_{s,H_2a}. Cela nous a conduits à approcher l'équation Monod par une cinétique quadratique

$$\frac{\mu_{\max,H_2a} s_{H_2} x_{H_2a}}{K_{s,H_2a} + s_{H_2}} \approx k_{r,H_2a} s_{H_2} x_{H_2a}, \tag{17}$$

avec k_{r,H_2a} le rapport $\mu_{\max,H_2a}/K_{s,H_2a}$.

La représentation des données avec le modèle cinétique quadratique est qualitativement aussi satisfaisante que celle du modèle original (les courbes pour les deux modèles sont superposées). L'estimation du modèle modifié avec IDEAS a amélioré grandement les intervalles de confiance, comme on peut le constater dans la troisième colonne de la table 1. Les valeurs estimées des paramètres communs aux deux modèles sont du même ordre de grandeur, mais les intervalles de confiance sont plus étroits pour le modèle modifié.

Table 1: Paramètres estimés avec leur intervalles de confiance approximatifs à 95% pour les modèles de Monod et à cinétique quadratique.

Paramètres	**Cinétique de Monod**	**Cinétique quadratique**
$K_{s,\mathrm{H_2a}}$	0.027 ± 0.165	
$\mu_{\max,\mathrm{H_2a}}$	0.195 ± 0.900	
$k_{r,\mathrm{H_2a}}$		5.45 ± 0.465
$Y_{\mathrm{H_2a}}$	0.039 ± 0.013	0.040 ± 0.011
k_d	0.031 ± 0.026	0.027 ± 0.003
k_i	0.061 ± 0.128	0.071 ± 0.088

La même procédure a été appliquée pour analyser des données expérimentales *in vitro* concernant la production de butyrate à partir de lactate. Le butyrate est la source d'énergie préférée des cellules épithéliales du côlon et il joue un rôle important dans la santé humaine (voir, *e.g.*, McIntyre et al. (1993) ; Hamer et al. (2008)). A l'opposé, l'accumulation de lactate dans le côlon est nuisible (voir, par exemple, Vernia et al. (1988) ; Belenguer et al. (2007)). Des études *in vitro* ont permis d'identifier la voie butyryl CoA:acetate CoA transferase comme étant le processus dominant dans la formation du butyrate dans l'écosystème du côlon humain (Louis et al., 2004).

Nous avons analysé des données expérimentales obtenues par Duncan et al. (2004b). Cette étude cinétique a été realisée avec *Eubacterium hallii* et la souche SS2/1. Ces deux espèces sont capables d'utiliser le lactate et l'acétate en suivant la voie du butyryl-CoA:acetate-CoA-transferase. Nous avons utilisé le critère d'information d'Akaike pour comparer les trois cinétiques données par l'équation de Monod, la cinétique quadratique et une cinétique du premier ordre et choisir le meilleur modèle. Ce critère établit un compromis entre l'ajustement des données et la complexité du modèle. Nous avons ainsi retenu le modèle à cinétique quadratique. Il a représenté de façon satisfaisante les données expérimentales pour les deux bactéries (Fig. 5). Les données pour la souche SS2/1 sont plus dispersées, ce qui explique pourquoi l'ajustement est moins satisfaisant que pour *E. hallii* . Les écarts types d'estimation sont satisfaisants.

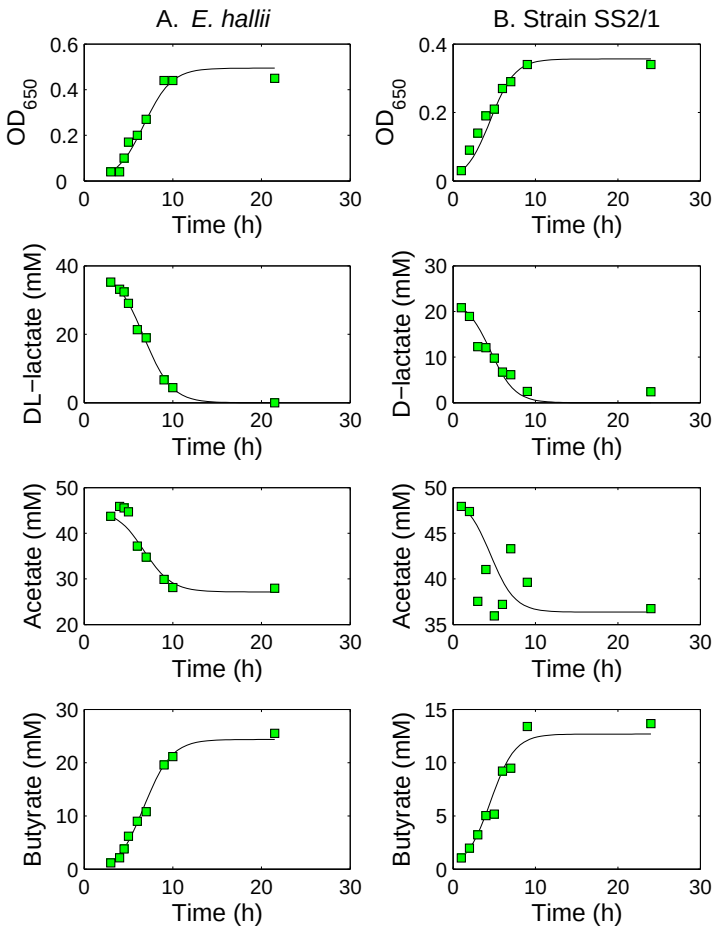

Figure 5: Modèle identifié pour la production de butyrate *in vitro*. Carrés : données expérimentales, trait plein : sorties du modèle.

Les valeurs des paramètres obtenues grâce à l'identification de ces deux sous-modèles ont été utilisées dans le modèle complet. Les valeurs des paramètres cinétiques de l'hydrolyse, l'utilisation du glucose et la méthanogénèse ont été tirées principalement de Batstone et al. (2002). Les coefficients de transport ont été sélectionnés pour garantir la positivité des flux tout en restant en accord avec des valeurs données dans la littérature sur le pourcentage d'absorption d'eau et des AGCC (voir, par exemple, Macfarlane and Cummings (1991)).

La méthodologie que nous avons utilisée pour réduire les dimensions des vecteurs de paramètres et de l'état est décrite ci-dessous.

Réduction du nombre de paramètres

Nous avons tout d'abord fait les hypothèses simplificatrices suivantes : (i) la constante d' Henry $K_{H,i}$ est la même dans tous les compartiments, (ii) $k_L a$ a la même valeur pour les trois gaz dans tous les compartiments, (iii) les constantes de mortalité, d'adhésion et de détachement sont les mêmes pour tous les groupes microbiens dans tous les compartiments, (iii) les paramètres de la cinétique pour un processus donné sont les mêmes dans tous les compartiments, (iv) le facteur de rendement $Y_{i,j}$ est le même dans tous les compartiments. De plus, les facteurs de rendement sont fixés pour garantir le respect des bilans de carbone et hydrogène. Cette méthodologie est basée sur la procédure proposée par Batstone et al. (2002), expliquée ci-dessous.

On peut obtenir des rapports entre les facteurs du rendement à partir de la stoïchiométrie. Considérons, par exemple, les réactions impliquées dans l'utilisation du glucose

$$C_6H_{12}O_6 \Rightarrow 2CH_3CHOHCOOH \quad (R_1)$$
$$C_6H_{12}O_6 + 2H_2O \Rightarrow 2CH_3COOH + 2CO_2 + 4H_2 \quad (R_2)$$
$$3C_6H_{12}O_6 \Rightarrow 4CH_3CH_2COOH + 2CH_3COOH + 2CO_2 + 2H_2O \quad (R_3)$$
$$C_6H_{12}O_6 \Rightarrow CH_3CH_2CH_2COOH + 2CO_2 + 2H_2 \quad (R_4)$$
$$5C_6H_{12}O_6 + 6NH_3 \Rightarrow 6C_5H_7O_2N + 18H_2O \quad (R_5)$$

Le glucose est utilisée dans les réactions cataboliques R_1-R_4, et dans la réaction anabolique R_5 qui représente la croissance bactérienne. Soit f_{su} la fraction de glucose utilisée dans le catabolisme. D'après R_5, le rendement de la production de biomasse Y_{su} est

alors

$$Y_{\text{su}} = \frac{6}{5}(1 - f_{\text{su}}).$$

Le glucose utilisé dans le catabolisme est dégradée via R_1-R_4. Notons η_k ($k = 1,2,3,4$) la fraction transformée dans la réaction R_k. Il est clair que $\eta_4 = 1 - (\eta_1 + \eta_2 + \eta_3)$.

Tous les facteurs de rendement impliqués peuvent donc être exprimés comme des fonctions de f_{su}, η_1, η_2 et η_3, de la façon suivante

$$\begin{aligned}
Y_{\text{la,su}} &= 2 f_{\text{su}} \eta_1, \\
Y_{\text{ac,su}} &= 2 f_{\text{su}} \eta_2 + \frac{2}{3} f_{\text{su}} \eta_3, \\
Y_{\text{pro,su}} &= \frac{4}{3} f_{\text{su}} \eta_3, \\
Y_{\text{bu,su}} &= f_{\text{su}} \eta_4, \\
Y_{\text{H}_2\text{,su}} &= 4 f_{\text{su}} \eta_2 + 2 f_{\text{su}} \eta_4, \\
Y_{\text{CO}_2\text{,su}} &= 2 f_{\text{su}} \eta_2 + \frac{2}{3} f_{\text{su}} \eta_3 + 2 f_{\text{su}} \eta_4, \\
Y_{\text{H}_2\text{O,su}} &= -2 f_{\text{su}} \eta_2 + \frac{2}{3} f_{\text{su}} \eta_3 + \frac{18}{5}(1 - f_{\text{su}}).
\end{aligned}$$

Outre le fait d'assurer la conservation des espèces chimiques, cette approche permet de réduire le nombre de paramètres. Par exemple, le nombre de paramètres à estimer pour l'utilisation du glucose serait huit sans ces considérations. En utilisant la stoïchiométrie, le nombre de paramètres est réduit à quatre. Cette procédure a été utilisée pour tous les substrats.

En prenant en compte toutes les simplifications ci-dessus, le nombre de paramètres du modèle global est réduit de 333 à 60.

Réduction du nombre de variables d'etat

La méthode des perturbations singulières est une technique classique de réduction d'ordre des modèles qui présentent des échelles de temps multiples. Elle repose sur le théorème Tikhonov (voir, par exemple, Khalil (2000)). Notre modèle, comme beaucoup de modèles mathématiques pour les réacteurs de digestion anaérobie (voir, par exemple, Rosen

et al. (2006)), présente des échelles de temps multiples. C'est donc un bon candidat à l'application de cette technique. Nous avons tout d'abord vérifié l'applicabilité de la réduction au petit modèle de l'homoacétogénèse *in vitro* ci-dessus. Le résultat a ensuite été étendu au modèle complet.

Réduction du modèle cinétique de l'homoacétogénèse *in vitro*

Considérons le modèle décrit par (6)-(10). Puisque la concentration d'acétate est obtenue par intégration d'une expression qui ne dépend que des autres concentrations, et qu'elle n'apparait pas dans les autres équations du modèle, nous pouvons ignorer l'acétate et nous focaliser sur le reste du modèle.

Un modèle à deux échelles de temps (lente/rapide) peut être mis sous la forme standard des perturbations singulières

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{z}, t, \varepsilon), \qquad \mathbf{x}(t_0) = \boldsymbol{\xi}(\varepsilon), \tag{18}$$

$$\delta\dot{\mathbf{z}} = \mathbf{g}(\mathbf{x}, \mathbf{z}, t, \varepsilon), \qquad \mathbf{z}(t_0) = \boldsymbol{\zeta}(\varepsilon), \tag{19}$$

où $\mathbf{x}$ comprend les n variables d'état lentes et $\mathbf{z}$ les m variables d'état rapides. Nous supposons les fonctions vectorielles $\mathbf{f}$ et $\mathbf{g}$ continûment différentiables par rapport à leurs arguments $(\mathbf{x}, \mathbf{z}, t, \delta)$.

Quand $\delta \approx 0$, nous obtenons une approximation quasi-statique dans laquelle (19) devient

$$\mathbf{g}(\mathbf{x}, \mathbf{z}, t, 0) = \mathbf{0}. \tag{20}$$

En résolvant (20), nous obtenons le vecteur quasi-statique $\mathbf{z}^*$

$$\mathbf{z}^* = \mathbf{h}(\mathbf{x}, t). \tag{21}$$

Le modèle réduit est alors donné par

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{h}(\mathbf{x}, t), t, 0), \quad \mathbf{x}(t_0) = \boldsymbol{\xi}(0). \tag{22}$$

Pour mettre en évidence la séparation d'échelle de temps dans le modèle de l'homoacétogénèse, nous avons normalisé les variables d'état du modèle. Les équations d'état pour les variables normalisées sont toutes de la forme

$$\dot{x}_i = d_i(\boldsymbol{x})x_i + p_i(\tilde{\boldsymbol{x}}_i), \tag{23}$$

Table 2: Comparaison des échelles de temps locales pour le modèle de l'homoacétogénèse.

x_i	$\tau_i(\boldsymbol{x})$
Biomasse active	$\geq$20
Biomasse inactive	≈ 20
H_2 dans la phase gazeuse	≈ 27
H_2 dans la phase liquide	$\leq$0.12

où x_i est la variable normalisée, $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^{\mathrm{T}}$ et $\tilde{\boldsymbol{x}}_i = (x_j)_{\substack{j=1,\ldots,4 \\ j\neq i}}$. On pose alors

$$\tau_i(\boldsymbol{x}) = \frac{1}{|d_i(\boldsymbol{x})|}, \tag{24}$$

où $\tau_i(\boldsymbol{x})$ est l'échelle de temps locale associée à x_i.

La table 2 compare les échelles de temps locales des variables normalisées. Il y a au moins un facteur 100 entre l'échelle de temps associée à l'hydrogène dans la phase liquide et les autres échelles de temps ; donc les variables d'état du modèle peuvent être séparées en variables lentes et variables rapides, $\mathbf{x} = (x_{\mathrm{H_2a}}, \quad \chi, \quad s_{\mathrm{g,H_2}})$, $\mathbf{z} = s_{\mathrm{H_2}}$. Ensuite, nous avons vérifié que les trois hypothèses de l'énoncé du théorème Tikhonov proposé dans Khalil (2000) étaient satisfaites.

Les simulations ont confirmé la validité de la réduction du modèle. La proximité entre le modèle réduit et le modèle complet dépend de la condition initiale pour la concentration d'hydrogène dans la phase liquide, ici inconnue. Cependant, nous savons que

$$0 < s_{\mathrm{H_2}}(0) < K_{H,\mathrm{H_2}} RT s_{\mathrm{g,H_2}}(0) \approx 0.013. \tag{25}$$

Nous avons simulé les modèles complet et réduit en donnant à $s_{\mathrm{H_2}}(0)$ les valeurs extrêmes de l'intervalle (25). Dans les deux cas, les réponses des deux modèles ont été superposées. Dans un but d'illustration, nous avons fixé $s_{\mathrm{H_2}}(0) = 0.1$ (ce qui est physiquement impossible pour les conditions expérimentales) pour observer le comportement de l'approximation par perturbation singulière sur un horizon de temps long. Les réponses des deux modèles sont montrées en figure 6. Nous observons que la concentration de

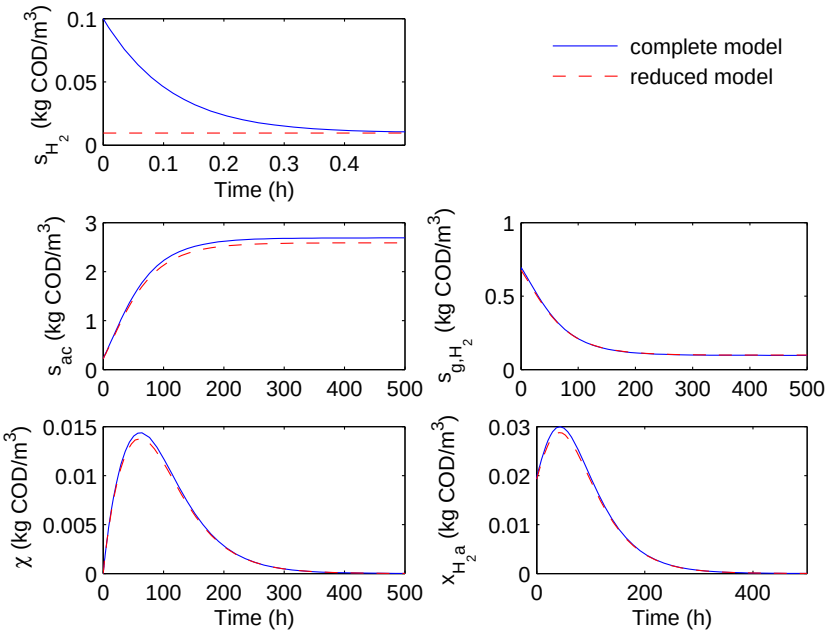

Figure 6: Comparaison du modèle complet et modèle réduit pour l'homoacétogénèse. La différence entre les conditions initiales de s_{H_2} a été exagérée à des fins d'illustration.

l'hydrogène donnée par le modèle complet converge rapidement vers la concentration quasi-statique de l'hydrogène calculée par perturbation singulière. Les réponses des variables lentes dans les deux modèles sont presque identiques.

Réduction du modèle complet de la dégradation des hydrates de carbone

Le résultat précédent nous poussé à étendre l'approche de réduction par perturbation singulière au modèle complet de la dégradation des hydrates de carbone. Considérons, par exemple, le modèle associé au lumen de la section du côlon proximal, défini par (1)-(4). Ces équations peuvent également s'écrire sous la forme générique (23), ce qui nous permet de borner les constantes de temps locales. En appliquant la même procédure que dans le modèle de l'homoacétogénèse, on obtient un modèle à deux échelles de temps dans la forme standard de perturbation singulière. Le vecteur des variables rapides $\mathbf{z}$ est

$$\mathbf{z} = \left[s_{H_2},\, s_{CH_4},\, s_{CO_2}\right]^{\mathrm{T}}.$$

Les hypothèses du théorème de Tikhonov sont également satisfaites. Nous avons illustré par simulation la proximité du comportement du modèle réduit avec celui du modèle complet pour le micro-habitat du lumen dans la section proximale. Comme observé pour le modèle de l'homoacétogénèse, lorsque les conditions initiales dans le modèle complet sont physiquement acceptables, la réponse du modèle réduit est superposée à celle du modèle complet. Nous avons simulé le système avec des conditions initiales des concentrations en hydrogène, méthane et dioxyde de carbone dans la phase liquide anormalement élevées pour observer des différences entre les deux modèles. Le résultat est illustré en figure 7, et confirme la validité de la réduction.

La figure 8 compare les réponses, pour les modèles complet et réduit, du compartiment du lumen dans la partie distale. La réduction d'ordre du modèle semble améliorer la stabilité des simulations. Toutefois, plus de recherche sur cette question est nécessaire.

La réduction du modèle a permis une économie d'au moins 30% en temps de calcul, parce que la raideur du modèle est réduite. L'application de cette approche dans les compartiments du mucus et du lumen dans le modèle complet permet de réduire le nombre de variables d'état de 102 à 84.

Exemples d'applications du modèle

Le modèle présenté ici peut être considérée comme une plate-forme expérimentale virtuelle, utilisable pour répondre à des questions biologiques. Le but des simulations suivantes est d'illustrer le caractère plausible du modèle et son utilisation potentielle pour étudier des interactions et des phénomènes difficiles à évaluer par un travail expérimental.

Une image dynamique et spatiale

Les figures 9 et 10 montrent une image dynamique et spatiale de la fermentation prédite par le modèle. L'entrée des fibres est passée de 10 g par jour à 25 g par jour à $t = 0$. Le côlon proximal semble être la partie la plus active. La plupart des fibres y sont dégradées (figure 9). On y observe le niveau de monosaccharides le plus haut; les monosaccharides sont rapidement épuisés dans les sections suivantes. Les concentrations simulées

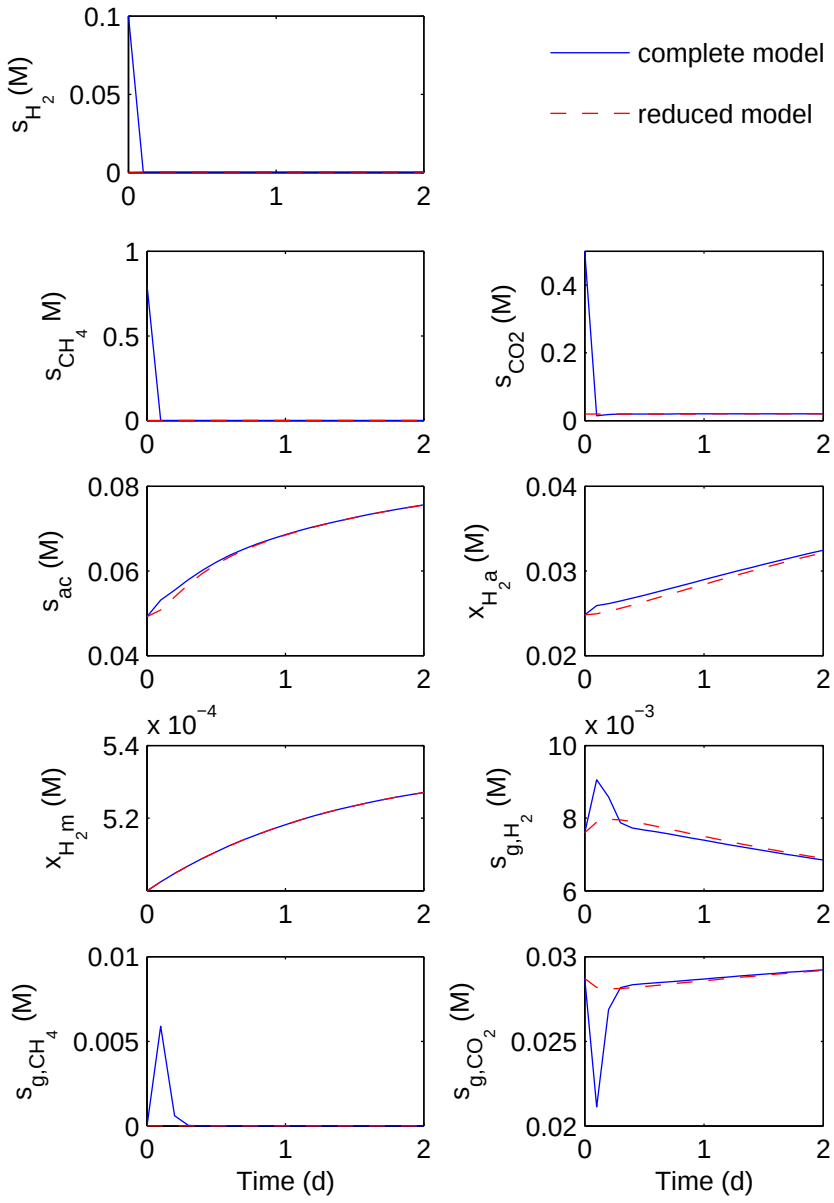

Figure 7: Comparaison des modèles complet et réduit pour le processus de fermentation complet dans le lumen du côlon proximal. La différence des conditions initiales a été exagérée à des fins d'illustration.

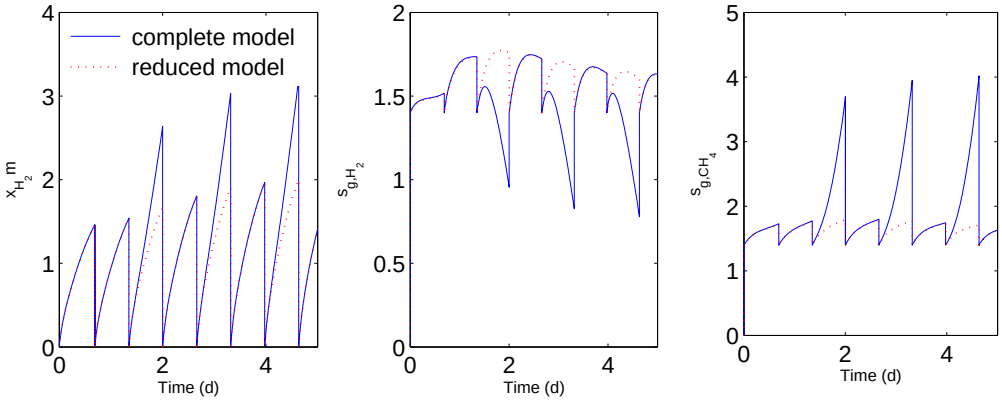

Figure 8: Comparaison de la stabilité numérique pour les simulations des modèles complet et réduit dans le lumen du côlon distal. Le comportement oscillatoire est causé par la réinitialisation périodique du volume due à l'excrétion. La réduction d'ordre du modèle semble améliorer la stabilité des simulations.

des AGCC sont très élevées dans le côlon proximal, et décroissent progressivement vers le côlon distal (figure 10). Les profils de concentrations milimolaires prédits par le modèle en acetate: propionate: butyrate dans les sections luminales sont 82:37:39 dans le côlon proximal, 74:35:37 dans le côlon transversal, et 57:31:33 dans le côlon distal. Ces valeurs sont en accord avec des études *in vivo* (voir Topping and Clifton (2001) et les références citées).

La concentration microbienne prédite dans le côlon distal est d'environ 70 g par jour. Si nous supposons une densité de la matière fécale de 1 g/ml et que 15% des matières fécales sont solides, ceci implique que les micro-organismes représentent 47% des solides, ce qui est en accord avec les valeurs expérimentales (Stephen and Cummings, 1980).

La figure 10 résume la variation temporelle et spatiale des concentrations microbiennes et en métabolites. Si le profil de la concentration microbienne dans le lumen est croissant de la section proximale à la section distale (figure 10C), la concentration microbienne dans le mucus présente un comportement inverse (figure 10D). Ce profil peut être expliqué par la diminution du sucre dans les compartiments du lumen, qui réduit la disponibilité du substrat pour la diffusion vers le mucus. Dans les compartiments du lumen, le chyme perd de l'eau en voyageant le long du côlon et donc la densité microbienne augmente.

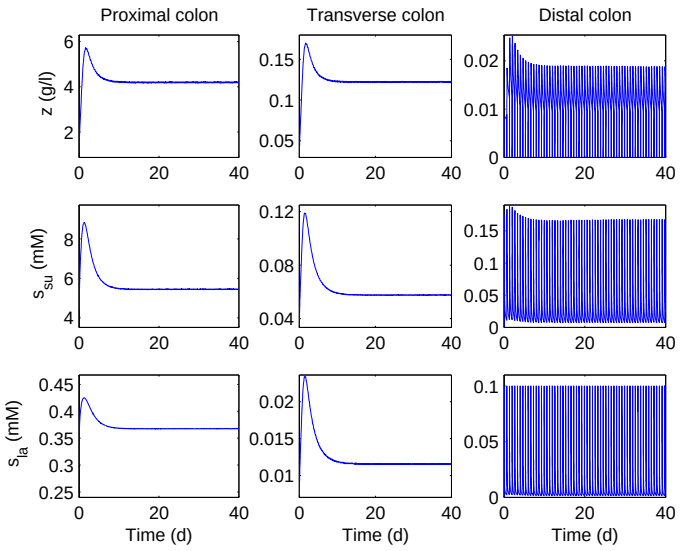

Figure 9: Modèle dynamique de la fermentation dans les compartiments du lumen. Le comportement oscillatoire dans les résultats du côlon distal est causé par la réinitialisation périodique du volume due à l'excrétion.

Les figures 10A et 10B montrent l'évolution des métabolites et des concentrations de micro-organismes au long du côlon lorsque l'entrée des fibres change de 10 g par jour à 25 g par jour. Elles suggèrent que, quand une perturbation sur le régime alimentaire est appliquée sur le système, au moins 10 jours sont nécessaires pour voir l'effet global d'un tel changement. On peut également remarquer que les données recueillies lors des études de la section distale, qui est la seule accessible expérimentalement, ne sont pas suffisantes pour avoir une image complète du comportement du système. Après d'avoir incorporé les phénomènes de transport tels que l'agrégation microbienne, et l'absorption des AGCC et d'eau, il est possible d'obtenir une représentation globale du système du côlon humain qui soit cohérente avec des données physiologiques.

Quel est le rôle du mucus ?

Le mucus affecte la dynamique du lumen sous trois formes. (i) il rend possible l'agrégation microbienne, (ii) les mucines représentent une source supplémentaire de carbone et, enfin, (iii) les micro-organismes dans le mucus peuvent enrichir la communauté microbienne

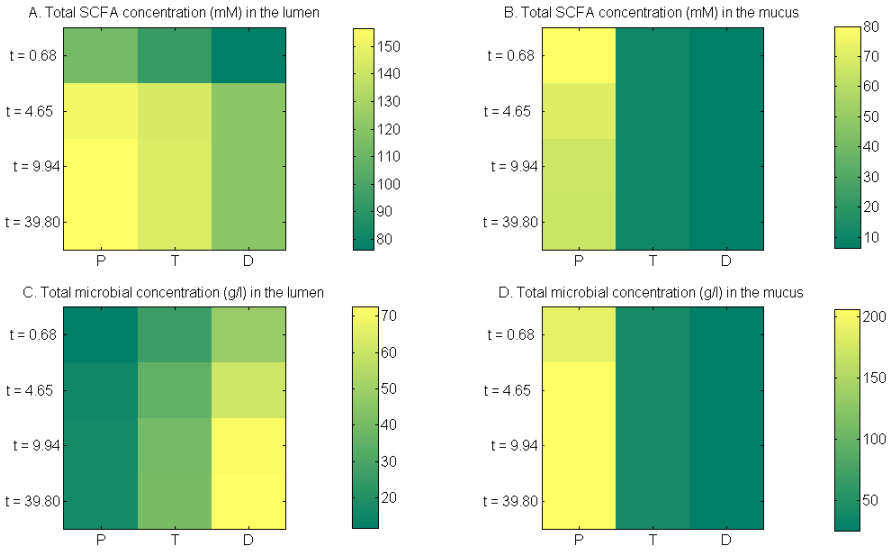

Figure 10: Schéma global de fermentation dans le côlon humain. P : côlon proximal, T : côlon transversal, D : côlon distal. La concentration en AGCC décroit au long du côlon humain, tant dans le lumen que dans le mucus. La densité microbienne totale augmente de la section proximale à la section distale dans le lumen, tandis que la flore microbienne dans le mucus suit un comportement inverse.

dans le lumen grâce au phénomène de détachement. Nous avons utilisé notre modèle mathématique afin d'évaluer l'importance de chacun de ces facteurs sur la dynamique du côlon humain. Ces effets sont illustrés dans la figure 11. La courbe continue représente la réponse du modèle qui prend en compte le mucus et les trois mécanismes mentionnés ci-dessus. Les autres courbes ont été obtenues en mettant tout à tour à zéro les paramètres associés à chacun des trois mécanismes.

Quand aucune mucine n'est disponible comme substrat pour la microflore du lumen (pas de transfert), la dynamique du système reste identique. Cela signifie que la quantité de glucides dans les fibres exogènes est beaucoup plus élevée que celle contenue dans les mucines qui ne sont pas utilisées dans le mucus. La figure 11 suggère également que le microbiote du mucus enrichit celui du lumen. Ces simulations suggèrent que l'agrégation microbienne, qui est représenté par le temps de rétention supplémentaire τ pour les micro-organismes, est nécessaire pour atteindre des conditions physiologiques, à savoir la dégra-

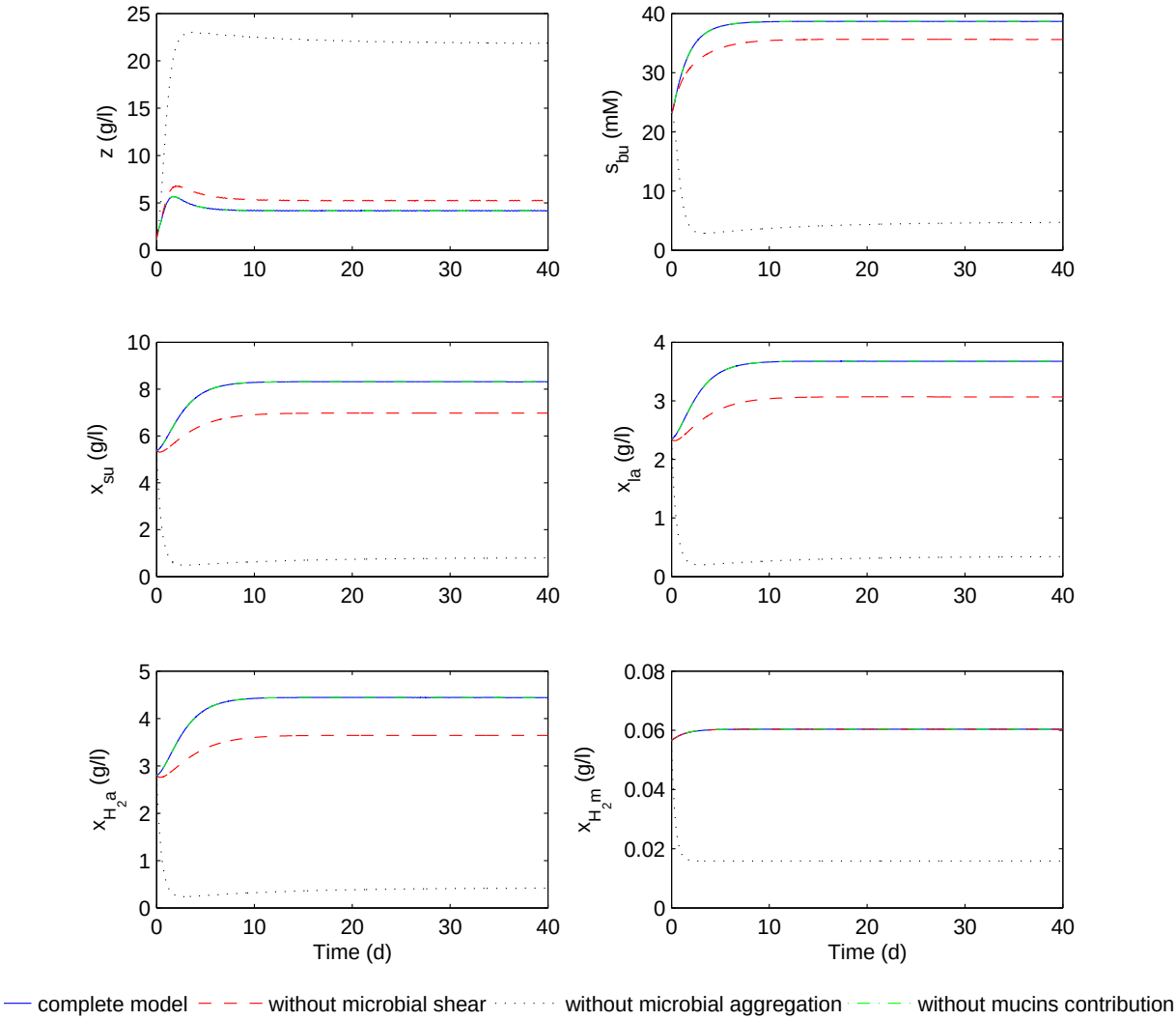

Figure 11: Le rôle du mucus dans le côlon humain. Les concentrations correspondent au compartiment du lumen proximal. L'entrée de fibre été fixée à 25 g par jour. L'agrégation apparaît comme le facteur le plus pertinent pour maintenir des conditions physiologiques.

dation élevée de fibres et la densité microbienne élevée. Ainsi, ce mécanisme apparaît comme le facteur le plus pertinent pour expliquer le rôle du mucus sur la dynamique du côlon humain.

Conclusions

Nous avons présenté ici le développement d'un modèle à base de connaissance de la dégradation des hydrates de carbone dans le côlon humain. Pour autant que nous sachions, ce modèle est le premier à intégrer la physiologie de l'intestin, des réactions biochimiques et les phénomènes de transport entre les micro-habitats du lumen et du mucus. Pour faire face à la complexité du modèle, nous avons fait usage d'outils et de concepts de la théorie des systèmes, tels que la méthode des perturbations singulières, l'identifiabilité et l'estimation de paramètres.

Les simulations du modèle suggèrent que nous avons obtenu une représentation qualitativement adéquate du côlon humain. Notre modèle permet de représenter la dynamique du processus de fermentation tout le long des sections physiologiques, incluant les compartiments du lumen et du mucus. Il permet de faire des analyses qui sont difficiles à réaliser expérimentalement. Ce modèle permet d'aborder facilement une longue liste de questions de type : *et si ?* Nous en avons donné ci-dessus quelques exemples. Aborder ces questions à travers la modélisation mathématique devrait contribuer à notre compréhension de l'écosystème du côlon humain. Un facteur clé pour parvenir à cette compréhension est l'intégration des informations provenant de nouvelles expériences. L'estimation des paramètres du modèle complet est la principale tâche à poursuivre dans un avenir proche. Compte tenu de la complexité du modèle, de la rareté et l'hétérogénéité des données, du caractère asynchrone des observations et de la méconnaissance des conditions initiales, l'estimation Bayésienne nous semble être l'approche la plus prometteuse.

Notre modèle dois être considéré comme complémentaire des travaux expérimentaux (à la fois *in vivo* et *in vitro*). Nous espérons qu'il contribuera à donner un meilleur aperçu de la dynamique du côlon humain. Il pourrait également être un outil pour l'élaboration de stratégies nutritionnelles axées sur la promotion de métabolites bénéfiques pour la santé.

Notation and nomenclature

Abbreviations / Acronyms

AD	Anaerobic digestion
AIC	Akaike's information criterion
COD	Chemical oxygen demand
CSTR	Continuous-flow stirred-tank reactor
DAE	Differential algebraic equation
FIM	Fisher information matrix
GI	Gastrointestinal
IBD	Inflammatory bowel disease
IDEAS	IDEntification and Analysis of Sensitivity
LI	Linear in its inputs
LP	Linear in its parameters
ML	Maximum-likelihood
NSP	Nonstarch polysaccharides
OD	Optical density
ODE	Ordinary differential equation
OTU	Operational taxonomic unit
PCA	Principal component analysis
PFR	Plug-flow reactor
RS	Resistant starch
SCFA	Short chain fatty acids
s.g.i.	structurally globally identifiable
w.r.t.	With respect to

Typography

Matrix	Upper case letter in bold, *e.g.*, $\mathbf{A}$
Vector	Lower-case letter in bold, *e.g.*, $\mathbf{z}$
Scalar	Italic, *e.g.*, k, Y
Set	Upper case letter in italic, *e.g.*, S

Symbols

Cor	Parameter correlation matrix
$\mathscr{M}(\cdot)$	Model structure
$\mathbf{N}(0, \sigma^2)$	Distribution of a Gaussian random variable with zero mean and variance σ^2
$\mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$	Distribution of a Gaussian random vector with zero mean and covariance $\boldsymbol{\Sigma}$
n_p	Number of parameters
n_x	Number of state variables
n_y	Number of outputs
$\mathbf{P}$	Covariance matrix of the ML estimator
$\mathbb{R}$	Set of real numbers
$\mathbf{x}$	State vector
$\mathbf{y}$	Output vector
$\boldsymbol{\varepsilon}$	Vector of measurement errors
$\boldsymbol{\theta}$	Parameter vector
$\boldsymbol{\theta}^*$	True value of $\boldsymbol{\theta}$
$\widehat{\boldsymbol{\theta}}$	Estimated value of $\boldsymbol{\theta}$

Table 3: Nomenclature of the model. For hydrolysis, $k_{m,i}$ is given in g s_j d^{-1} /mol x_j, and $Y_{i,j}$ in mol s_i/ g s_j. γ_i for sugars is given in l/d.

	Definition	**Units**
State variables		
s_i	Concentration of soluble component i in liquid phase	M (mol/l)
$s_{g,i}$	Concentration of soluble component i in gas phase	M
x_i	Concentration of bacterial group i	M or OD
z	Concentration of polysaccharides	g/l
Parameters		
a_i	Adherence coefficient of bacteria in lumen	d^{-1}
b_i	Shear loss coefficient of bacteria in mucus	d^{-1}
f_i	Fraction of the component i used for catabolic reactions	
k_d	Decay constant of bacteria	d^{-1}
$K_{H,i}$	Henry's law coefficient of component i	M/bar
$k_{hyd,i}$	Maximum specific hydrolysis rate	d^{-1}
k_i	Inactivation constant of bacteria	d^{-1}
$k_L a$	Liquid-gas transfer coefficient multiplied by the specific transfer area	d^{-1}
$k_{m,i}$	Maximum specific rate of substrate consumption	mol s_j d^{-1} /mol x_j
$K_{s,i}$	Half-saturation constant for microbial growth (constant of Monod)	M
$K_{x,i}$	Half-saturation coefficient for ratio s/x	M s_j / M x_j
r_i	Density of component i	g/l
w_i	Molecular weight of component i	g/mol
$Y_{i,j}$	Yield of component i in process j	mol s_i/ mol s_j
Y_j	Biomass yield factor for bacteria utilizing substrate j	mol x_j / mol s_j
γ_i	Transport coefficient of component i	d^{-1}
η_i	Reaction yield	
λ_i	Reaction coordinate	
$\mu_{max,i}$	Maximum specific growth rate	d^{-1}
τ_i	Additional residence time of bacteria in lumen	d

Nomenclature of the model (continued).

Functions	
ρ_j	Kinetic rate of process j
Q_i	Liquid-gas transfer rate of component i
Subscripts and superscripts	
g	Gas phase
in	Influent
l	Lumen
lgt	Liquid-gas transfer
m	Mucus
out	Effluent
n	net
For soluble components	
ac	Acetate
bu	Butyrate
la	Lactate
pro	Propionate
su	Sugar (glucose)
For particulate components	
H_{2a}	Hydrogen-utilizing bacteria (acetogenic)
H_{2m}	Hydrogen-utilizing bacteria (methanogenic)
la	Lactate-utilizing bacteria
su	Glucose-utilizing bacteria

Nomenclature of the model (continued).

Physiological parameters	
L	Length of colon section (cm)
P_{atm}	Atmospheric pressure (bar)
p_{H_20}	Water pressure (bar)
R	Ideal gas constant (bar/(M K))
q	Flow rate (l/d)
q_{n}	Net flow rate in the distal lumen (l/d)
T	Temperature (K)
V	Volume in liquid phase (l)
V_{g}	Volume in gas phase(l)
Γ	Endogenous production of mucins (g/d)
ϕ	Diameter of the colon (cm)

Chapter 1

Introduction

I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics.
Charles Darwin

Mathematical models (sometimes called *in silico* models) are central for a better understanding of the behavior of systems in all fields of pure and applied sciences. They can be used to estimate quantities that cannot be measured, to test hypotheses, to predict behaviors, to control processes, to teach, among others. Basically, mathematical models can be derived by induction (data-driven modelling) or by deduction (see, *e.g.*, Kell and Oliver (2004); Teusink and Smid (2006), and the references therein for discussion of both approaches in biological systems).

Inductive models describe the input-output behavior of a system by means of equations that are physically meaningless. By contrary, in the deductive modelling approach, the model construction is based on fundamentals laws (e.g., the law of conservation of mass) and relationships that are supposed to govern the system behavior. Therefore, the modeller is interested to identify the phenomena that take place on the system, and to translate such a prior information into equations that contain parameters with physical meaning. Estimation of such parameters provides a link between the model and the real world.

Mathematical models that are derived deductively are often called knowledge-based models. Given the comprehensive structure of such models, they allow for lateral transfer of insight between scientific domains (Picioreanu and van Loosdrecht, 2003).

In recent decades, mathematical modelling has come to play an essential role in biology (May, 2004). Biological systems are often complex. Their complexity resides in the large numbers of elements with diverse functionalities that interact nonlinearly to produce coherent behaviors (Kitano, 2002a). As rational and formal representations, mathematical models are central to achieve a system-level understanding of such a complexity (Baldwin et al. (1984); Legay (1997); Bailey (1998); Stelling (2004)).

In this work, we tackle the construction of a knowledge-based model of a biological system: the human colon. More precisely, the modelling of carbohydrate degradation in the human colon. Our work is built on three main pillars, namely biology, system theory and chemical engineering.

Why a mathematical model of human colonic fermentation? A biological point of view

The human colonic microbiota is recognized as a key component in gastrointestinal (GI) tract homeostasis. This microbial population is involved in the development of immune function (Hooper and Gordon, 2001) and in pathologies such as inflammatory bowel disease (IBD) (Manichanh et al. (2006); Seksik et al. (2006); Frank et al. (2007); Sokol et al. (2008)) and obesity (Ley et al. (2005); Ley et al. (2006b); DiBaise et al. (2008)). Moreover, fermentation produces essential vitamins and co-factors (Zoetendal et al., 2008), and contributes up to 10% of the energy of the body's metabolic requirements (Macfarlane and Cummings, 1991). The principal products of fermentation are short chain fatty acids (SCFA), microbial biomass, H_2, CO_2, and CH_4 in some individuals. SCFA (mainly acetate, propionate and butyrate) are recognized for their health-promoting effects (Topping and Clifton, 2001).

Despite its important role in human health, the human colonic ecosystem remains largely unexplored due to the complexity of the microbiota. This dynamic microbial consortium

is host-specific, spatially distributed along the GI tract (Zoetendal et al. (2004); Dethlefsen et al. (2006)). Human *in vivo* studies are restricted by ethical considerations and the limited accessibility of the GI tract. These constraints have motivated the development of *in vitro* models to simulate the dynamic conditions of the human colon (see Section 2.4). In parallel to *in vitro* studies, animal models have also been used to investigate different aspects of the colonic fermentation.

Both *in vitro* and *in vivo* models have been crucial for the understanding of the human GI tract. Nevertheless, extrapolation of such studies to the human colon has to be done with caution. *In vitro* models do not account for biotic factors such as SCFA absorption, interaction with the host, and exchange between the lumen and mucus microhabitats. Animal intestines present physiological differences with the human colon. In addition, both *in vitro* and *in vivo* approaches use mixed culture of fecal matter, but only $20-40\%$ of human colonic microbiota can be cultivated (Amann et al. (1995); Suau et al. (1999)) and major differences have been found between fecal microbiota and colonic microbiota (Marteau et al. (2001); Zoetendal et al. (2002); Sarma-Rupavtarm et al. (2004)).

Mathematical modelling can complement *in vitro* and *in vivo* studies. It is a promising approach to circumvent some obstacles associated with experimental work. Although mathematical models have shown to be useful in animal nutrition (see, *e.g.*, Baldwin et al. (1981); Jumars (2000); Hanigan et al. (2006); Dumas et al. (2008)), up to now, mathematical models have rarely been used to study the complexity of the carbohydrate degradation in the human colon and few results have been reported in the literature. Mathematical approaches do not seem to have been applied to the study of the human colonic ecosystem as a whole. Only parts of the system have been considered. Studies in fermentation addressed the modelling of bacterial *in vitro* experiments and attempted to analyze specific reactions of the fermentation process by human colonic strains (Amaretti et al., 2007). Modelling studies based on ^{13}C-labeled carbohydrates have been used to evaluate fermentation in the small and large intestine of infants (Christian et al., 2002). Belenguer et al. (2006) also used labeled components to identify carbon flows in the metabolism of lactate-utilizing bacteria to produce butyrate. This metabolism is affected by exogenous acetate. In this direction Duncan et al. (2004a) and Morrison et al. (2006) estimated, by modelling, the contribution of acetate in butyrate formation. Other phenomena taking

place in the human colon have been studied by mathematical modelling. The colonization of microorganisms in the gut and dynamics of pathogens have received most attention (Kirschner and Blaser (1995); Ballyk et al. (2001); Davis and Gordon (2002); de Jong et al. (2007)). Absorption of organic pollutants in the intestine (Moser and Mclachlan, 2002) and drug metabolism (Haddish-Berhane et al., 2009) have also been modelled. A mathematical approach was presented to describe the spatial distribution of intestinal bacteria with respect to their tolerance to oxygen, taking into account the variable geometry of the intestine (Wilkinson, 2002).

To the best of our knowledge, today there is still no mathematical model integrating the physiology of the intestine, the carbohydrate fermentation process and mass transfer between the lumen and mucus. The purpose of the work reported in this thesis was to develop such a mathematical model in order to provide a virtual platform describing the dynamics of the fermentation pattern in the human colon. We hope that the mathematical model presented here will contribute usefully to coping with the complexity of the human colon, enabling further insight into the microbial metabolism in the human GI tract.

Why a mathematical model of human colonic fermentation? Mathematical challenges

In the human intestinal ecosystem many processes take place at different time scales. No mechanistic explanations are available yet for some phenomena. We are often left with hypotheses, some of which may be contradicted by experimental studies.

Phenomena occurring in the human colon comprise metabolic conversions and mass transfer. Such processes are made possible by interactions (most of them nonlinear) of a large number of components, namely host cells, microorganisms and metabolites. The human colon is a distributed system. A full representation of the system in state-space form would then lead to a mathematical model with an infinite number of state variables. We were then interested in reducing such a complexity and providing a model that could represent the main phenomena with a reasonable number of state variables. In order to handle model complexity, we used mathematical tools, such as parameter identification, singular perturbation, and optimization algorithms to provide a model with an acceptable qualita-

tive representation of system behavior.

One of the main obstacles towards model validation (or rather model falsification) is the scarcity of data. Experimental studies usually provide information on the last section of the colon and no data is available along its length. Few studies provide dynamic data. In addition, the measurements are often asynchronous, some variables can only be measured once, and others are not measured.

Most of the knowledge of human colonic fermentation has come from studies on *in vitro* and *in vivo* models. *In vitro* bacterial growth experiments provide information on specific reactions of the full metabolic pathway. In this work, we performed the modelling and identification of some of such experiments to define parameters of the complete model. Estimating the parameters of the complete model is not trivial. For the time being, we do not prove a quantitative validation of the model. However, we showed that the model predictions were qualitatively in agreement with existing knowledge reported in the literature.

The strategy presented here is a generic procedure. It may be useful to tackle other modelling tasks, where the model is too complex and data are scarce.

Outline of the thesis and reader's guide

The work presented here integrates concepts of biology, mathematics and biochemical engineering. To facilitate navigation, we have tried to indicate the relative weights of each knowledge domain for each chapter, which is illustrated in Fig. 1.1. The route we take to construct our mathematical model begins with a description of the human colonic ecosystem with a brief overwiew of the existing approaches to its study (**Chapter 2**). **Chapter 3** is the thesis core. It comprises the integration of biological knowledge of the system towards the formulation of the mathematical model. In **Chapter 4**, we describe the methodological tools that we used, and developed, to handle our mathematical model. In **Chapter 5**, we tackle the mathematical modelling and parameter estimation of two bacterial growth experiments. The reactions that take place in such experiments are part of the complete fermentation process. The results in this chapter provide useful information in the definition of parameter values of the full model. **Chapter 6** describes our results

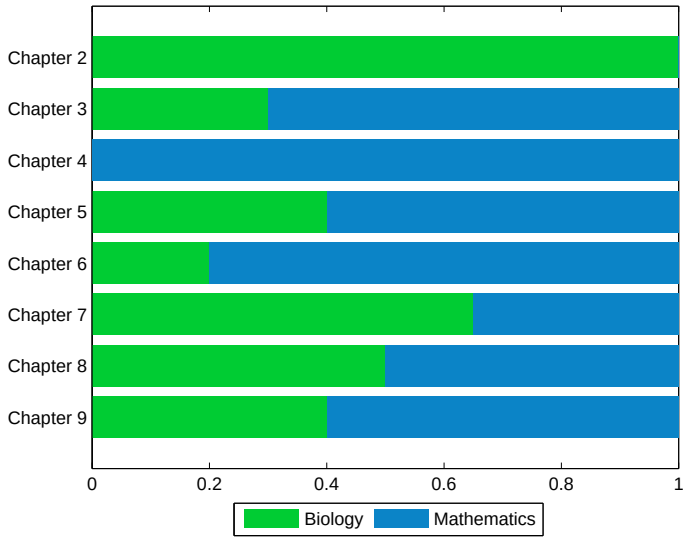

Figure 1.1: Map of the thesis

in reducing the number of parameters of the model and the number of its state variables. **Chapter 7** presents the simulations of the mathematical model under various nutritional scenarios. This chapter aims to illustrate a potential use of the model to get insight into the dynamics of the human colonic ecosystem. **Chapter 8** describes a preliminary study on selection of bacterial strains, in the framework of an experimental work with gnotobiotic rats to analyze, *in vivo*, the dynamics of the fermentation pattern. **Chapter 9** summarizes the results of the thesis and propose future directions of research.

Chapter 2

Description of the human colon

. . . These (the microbes) were located in the choicest places the interiors of the Family could furnish: in the lungs, in the heart, in the brain, in the kidneys, in the blood, in the guts. In the guts particularly. The great intestine was the favorite resort. There they gathered, by countless billions, and worked, and fed, and squirmed, and sang hymns of praise and thanksgiving; and at night when it was quiet you could hear the soft murmur of it. The large intestine was in effect their heaven.

Letters From the Earth. Mark Twain

The human colon (also called large intestine) is an anaerobic ecosystem within the gastrointestinal tract (Fig. 2.1). Its function is the absorption of water and sodium, and breakdown of alimentary polysaccharides that escape degradation in the upper digestive tract. A brief description of this ecosystem is presented below.

2.1 Physiology and anatomy

The human colon is a unique biochemical environment, characterized by low redox potential (Mackie et al., 1999) and a constant controlled temperature of 37°C (Savage, 1977). Its surface has been estimated somewhere between 250 and 400 m^2 (Bengmark, 1998).

It consists of three well-defined anatomical regions: the proximal colon (cecum and ascending colon), the transverse colon and the distal colon (sigmoid and rectum) (Fig. 2.1). The colon receives the food material from the ileum. The contents are retained in the proximal colon and the digesta (chyme) goes to the distal colon to be, finally, degraded or excreted. The intestinal region where the flow of digesta takes place is called the lumen. The passage of the digesta through the intestinal lumen is carried out by peristaltic movements. During this travel, the chyme changes its physical and chemical properties, because of fermentation, and water and metabolite absorption (Minekus, 1998). The pattern of fermentation products confers a spatial distribution of metabolites and a pH profile with values of about 5.5 in the proximal colon, 6.2 in the transverse colon and 6.9 in the distal colon (Macfarlane and Cummings, 1991).

The intestine is lined with the epithelium, where the secretion of a mucus gel layer is carried out by goblet cells. The mucus layer can be partitioned into an inner layer and a sloppy outer layer (Matsuo et al. (1997); Atuma et al. (2001)). The outer layer has a lower resistance to flow and acts primarily as a lubricant. The inner layer is a shear-resistant gel that provides a physical barrier against invading pathogens. Mucus thickness varies along the sections of the large intestine, and takes its largest value in the rectum. Measurements of mucus thickness are very sensitive to manipulation. It appears that an important loss of mucus in the floppy layer takes place during the procedure. This may explain divergent results. Studies report mucus thickness varying from 26 μm to 300 μm (Lichtenberger (1995); Matsuo et al. (1997); Pearson and Brownlee (2005); Swidsinski et al. (2007)). By extrapolating results from *in vivo* studies with rats, where the two layers could be quantified (Atuma et al. (2001); Strugala et al. (2003)), one can estimate that the complete bilayer may have a thickness up to 2000 μm.

Figure 2.2 displays a cross section view of the large intestine. Table 2.1 shows an approximation of the volumes of the anatomical regions of the colon. Calculations are based on the dimensions reported in literature (de Jong et al. (2007); Rajilic-Stojanovic (2007) and the references therein).

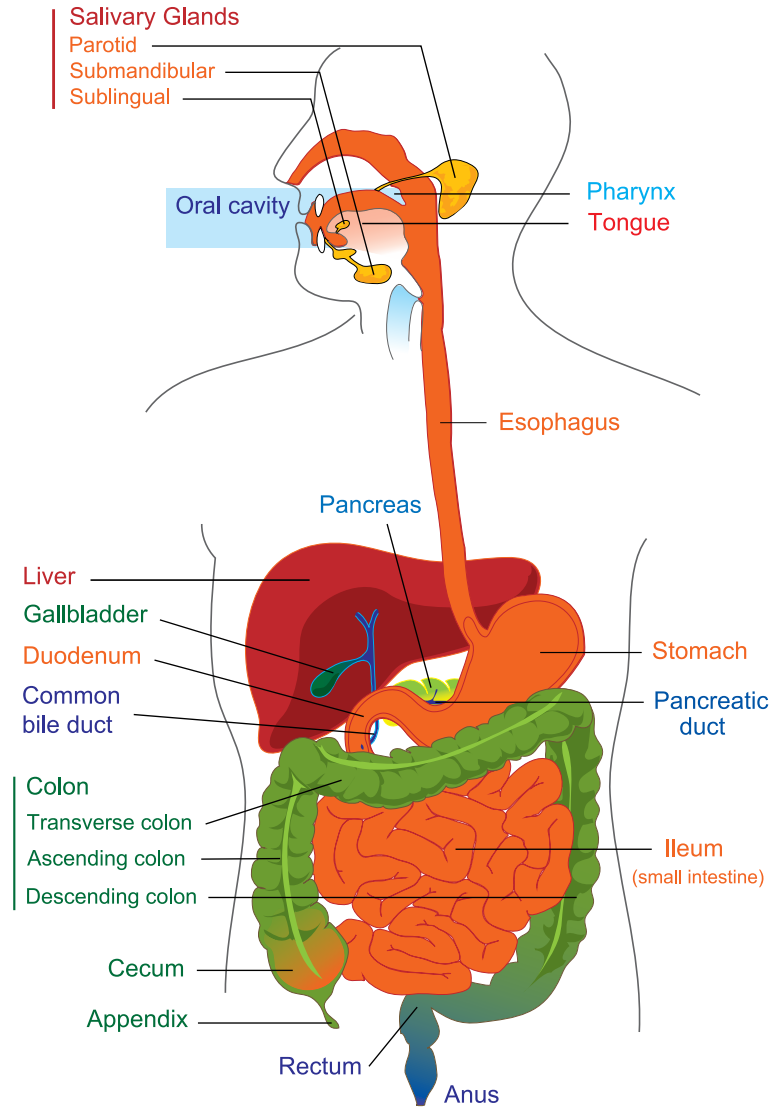

Figure 2.1: Representation of the human GI tract. (source: wikipedia)

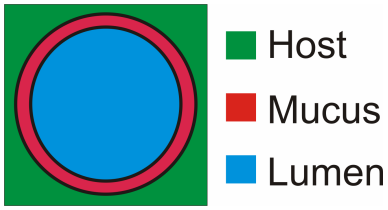

Figure 2.2: Cross section view of the human colon. The large intestine comprises the lumen and the mucus. The epithelium is the interface of interaction between the host and the intestine.

Table 2.1: Approximate dimensions of the human colon.

Anatomical region	Length (cm)	Volume of lumen microhabitat (l)	Volume of mucus [a] microhabitat (l)
Cecum	6		
Ascending colon	15		
Proximal colon	21	0.41	0.017
Transverse colon	50	0.98	0.042
Descending colon	25		
Sigmoid colon	40		
Rectum	18		
Distal colon	83	1.63	0.070
Total	154	3.02	0.129

[a] The volume of the mucus V^{m} is calculated assuming a perfect cylinder shape as $V^{\mathrm{m}} = 0.25\pi(\phi^2 - (\phi - 2 * e_{\mathrm{m}})^2)L$. The thickness of the mucus e_{m} is taken to be equal approximately to 0.0830 cm; ϕ is the diameter of the colon (5 cm), and L is the length of each section. The volume of the lumen is calculated as $0.25\pi\phi^2 L - V^{\mathrm{m}}$.

The large intestine is a very active organ. Many of its functions are assigned to the microbes that it harbors. These microbes conform the human colonic microbiota. Some features about this microbial community are introduced below.

2.2 Human colonic microbiota

The human gut is one of the densest microbial systems on Earth (Whitman et al., 1998), with population densities ranging from 10^8 cell per ml in the cecum to 10^{11} cells per ml in the distal colon. It has been estimated that microbial communities associated with the human body surfaces outnumber the human somatic and germ cells by a factor of ten (Savage, 1977).

The microbial consortium in the human large intestine is host-specific, appearing as a personalized fingerprint (Zoetendal et al. (1998); Eckburg et al. (2005)), which is stable over time in healthy adult individuals (Tannock et al. (2000); Seksik et al. (2003)). However, the term of stability has to be used with caution, as the human colon is not well characterized in terms of its dynamics and the timescale of the observations is limited. The microbial community is affected by dynamic factors such as diet, antibiotic treatment, physiological stress, lifestyle, environmental and stochastic factors (Dethlefsen et al. (2006); Flint et al. (2007); Zoetendal et al. (2008); Phillips (2009)). For example, it has been demonstrated that bacterial species have distinctive dynamic sensitivities when humans are exposed to stress (Holdeman et al., 1976) and that some members of this community can show temporal populations shifts while others remain stable (Vanhoutte et al., 2004).

The human colonic microbiota is mainly located in two microhabitats, namely lumen and mucus (Fig. 2.2). The mucus gel is a polymer-based matrix secreted by goblet cells. It mainly consists of water (95%) and mucins, which are high molecular weight glycoproteines with an important role in microbial life (Laux et al. (2005); Pearson and Brownlee (2005)). This matrix favors the attachment of microorganisms and their resistance to shear forces (Sonnenburg et al., 2004). In the words of Krause et al. (2006), microbes in the large intestine cannot hope a more suitable microhabitat than mucus. Some of the advantages of this bacterial niche are the increase of nutrients availability, protection from antimicrobial agents and proximity to other bacteria. In addition, the mucus is a carbon source that can support intestinal bacteria *in vivo* in the absence of any dietary input (Macfarlane and Cummings, 1991). Bacteria specialized in mucin degradation have been isolated (Salyers et al. (1977); Derrien et al. (2004)). Bacterial species such as *B. thetaiotaomicron* can turn to host glycans when dietary polysaccharides become scarce (Bäckhed et al. (2005); Sonnenburg et al. (2005)). In terms of the bacterial species present, the predominant mucosa-associated population is distinct from the luminal one (Zoetendal et al. (2002); Eckburg et al. (2005); Lepage et al. (2005)). Their metabolic activity can also differ (Probert and Gibson, 2002). For example, *in vitro* studies with biofilm chemostats in series showed that biofilm-associated bacteria were highly active for acetate production while planktonic bacteria made a higher contribution to propionate and butyrate formation (Macfarlane et al., 2005). In the lumen microhabitat, the microbiota is attached to the

surfaces of particulate food material or it can be free-living (not surface associated) (Akin (1976); Walker et al. (2008)). Functional differences between particulate-associated microbiota and free-living microbiota have also been suggested (Macfarlane et al. (1997); Macfarlane and Macfarlane (2006)).

The complexity of the human colonic microbiota has been tackled by ecology to elucidate two central questions: who are present and what are they doing (Zoetendal et al., 2004). Culture-based methods provided an initial picture of the microbial diversity in the human GI tract (Moore and Holdeman (1974); Savage (1977); Miller and Wolin (1982); Finegold et al. (1983)). This approach is still useful to study microbial species of interest in the human large intestine (Bernalier et al. (1996b); Barcenilla et al. (2000); Duncan et al. (2002b); Pryde et al. (2002)). The picture of the human colonic microbiota has remarkably improved thanks to the development of molecular microbial ecology, where diversity is assessed by culture-independent approaches, such as first targeting 16S rRNA genes (Wilson et al. (1997); Suau et al. (1999)). Today, the microbial diversity in the human colon is estimated about 1000 species (Rajilic-Stojanovic et al., 2007). Taking into account interindividual variability, it is estimated that more than 45000 bacterial species comprise the microbial consortium in lower human GI tract (Frank et al., 2007). This microbial community is dominated by relatively few divisions that are highly diverse at the level of species and strains. Most of the members of this consortium belong either to the bacterial phyla *Firmicutes* or *Bacteriodetes*. The domain *Archaea* also inhabits the human colonic ecosystem, with *Methanobrevibacter smithii* the dominant phylotype (Miller and Wolin (1982); Eckburg et al. (2005); Mihajlovski et al. (2008)). Minor populations of *Eukarya* have been recently found (Scanlan and Marchesi, 2008). Viruses and bacteriophages are also present in the human colon (Breitbart et al. (2003); Lepage et al. (2008)).

The second ecological question relates to the metabolic function of community members in the ecosystem. The human intestinal microbiota can be considered as an additional organ. Microbes are provided with a set of enzymes catalyzing metabolic functions that confer physiological properties that the human genome lacks (Gill et al. (2006); O'Hara and Shanahan (2006); Goodacre (2007)). However, due to the complexity of the human colonic microbiota, elucidation of the function of microorganisms is a very difficult task. Advances on the genomics-related techniques at the level of gene expression (DNA-based

approach: genomics; RNA-based approach: transcriptomics, protein: proteomics, and metabolites: metabolomics) have greatly improved our understanding of bacterial functionality in the human gut ecosystem (Gill et al. (2006); Zoetendal et al. (2006); Gloux et al. (2007) Kurokawa et al. (2007); Louis et al. (2007a); Li et al. (2008); Zoetendal et al. (2008); Brugère et al. (2009); Mahowald et al. (2009)). Metagenomics is therefore central to shed light on the diversity and function of the microbial ecosystem residing in the human large intestine.

Today, the microbiota is recognized as a key component in GI tract homeostasis by mediating many of the effects of diet upon gut health (Flint et al., 2007). Fermentation produces essential vitamins and co-factors. Fermentation products such as short chain fatty acids (SCFA) are recognized for their health-promoting effects (Topping and Clifton, 2001). The next section is devoted in describing the main mechanisms of carbohydrate degradation by the human colonic microbiota.

2.3 Anaerobic trophic chain

The human colonic microbiota performs a series of metabolic reactions to breakdown and ferment complex polysaccharides. Carbohydrates constitute 85% of available substrates for colonic fermentation. The substrates are mostly from dietary origin. Main dietary carbohydrates are resistant starch (RS) and nonstarch polysaccharides (NSP). NSP are a diverse group of homo and heteropolymers, including xyloglucans, glucuronomanans, arabinoxylans and xylans (Chassard et al., 2007). Whereas starch is a glucose homopolymer found in the amylose and amylopectin forms. NSP is the major component of dietary fiber. However, on a high starch diet, the amount of RS may exceed the amount of NSP (Macfarlane and Cummings, 1991). The quantity of NSP entering the human colon may be up to 28 g/day for a western diet (Baghurst et al., 1996). The estimation of RS entering to the colon is more difficult to assess, because of the lack of data on food composition, partial digestion in the small intestine, and individual physiological influences such as chewing. In addition to dietary carbon sources, endogenous production of mucins contribute to carbon supply (Bernalier et al., 1999). Mucins are glycoproteins consisting of 80% carbohydrates (Pearson and Brownlee, 2005). Table 2.2 shows the available substrates for the human colonic microbiota.

	Substrate	Amount (g/day)
Carbohydrates	Resistant starch	5-40
	NSP	8-30
	Sugars and sugar alcohols	2-10
	Oligosaccharides	2-8
	Mucins	3-5
	Chitin and amino sugars	1-2
	Synthetic carbohydrate	Variable
	Food additives	Variable
	Therapeutic agents	Variable
Protein	Dietary origin	1-12
	Pancreatic enzymes	4-8
	Sloughed epithelial cells	30-50
Others	Urea, nitrate	0.5
	Bacterial recycling	Unknown

Table 2.2: Substrates available for the human colonic microbiota (from Macfarlane and Cummings (1991); Egert et al. (2006)).

Microbial metabolism consists of catabolic reactions that yield energy and anabolic reactions that lead to bacterial growth. The first stage of carbohydrate degradation is the hydrolysis of polysaccharides that cannot be used by the human (Topping and Clifton, 2001). Hydrolysis is effected by bacterial cell-associated and secreted hydrolases. The polymers are therefore broken down to their constituent monomers, which can then be assimilated by the microorganisms in the fermentation. This process occurs in the absence of external electron acceptors, such as oxygen or nitrate. The principal products of the fermentation are SCFA, CO_2, H_2, biomass, and CH_4 in some individuals (it has been estimated that approximately 1/3 of the human population produces methane (Cummings and Macfarlane, 1991)). SCFA are further absorbed by the host.

The carbohydrate fermentation process is affected by a number of factors, including host genotype, ingestion of microorganisms and microbial interactions, stochastic events and physicochemical factors such as pH, retention time, physical structure of the food matrix and composition of substrates (Macfarlane and Gibson (1997); Dethlefsen et al. (2006); Louis et al. (2007a)). In addition, the methane status of the individual appears to be linked to the activity and structure of cellulose-degrading species (Robert and Bernalier-Donadille (2003); Chassard et al. (2007)).

Anaerobic digestion has a low energy yield in comparison with aerobic processes. This forces the microbial community to cooperate efficiently and thus obtain the energy required for its survival. A special case of such cooperations is syntrophism, in which two metabolic types of microorganisms depend on each other to perform the degradation of a given substrate (Schink, 1997). The term syntrophy has also been used to denote harmful and neutral relationships (Relman, 2008). Syntrophic associations (in the sense described by Schink (1997)) favor the thermodynamics of the conversion process. The fermentation products released from the metabolism of microbial species are utilized by other microbes, driving in this manner the energetics of the reactions and improving the overall utilization of substrates (Voolapalli and Stuckey (1999); Kleerebezem and Stams (2000); Jackson and McInerney (2002); Rodríguez et al. (2008); Madigan et al. (2009)). The human colon can thus be viewed as an extremely efficient reactor, where the thermodynamic feasibility of the reactions is enhanced by microbial syntrophism and metabolite absorption.

2.3.1 Main fermentation pathways

The metabolic pathways of anaerobic fermentation are well documented (Gottschalk (1986); Miller and Wolin (1996); Bernalier et al. (1999); Macfarlane and Gibson (1997); Moat et al. (2002); Madigan et al. (2009)). Anaerobic digestion by the human colonic bacteria comprises several reaction pathways. The main fermentation pathways are described below.

2.3.1.1 Embden-Meyerhoff-Parnas pathway

The Embden-Meyerhoff-Parnas (EMP) pathway has been identified as the major route of glucose metabolism (Miller and Wolin, 1996). However, bacteria can use other pathways, such as the Entner-Doudoroff pathway (see, *e.g.*, Madigan et al. (2009)). The EMP is divided into three stages. In Stage I, glucose is activated and the intermediate glyceraldehide 3-phospate is produced. This stage is characterized by the absence of redox reactions and thus no energy is released. In Stage II, glyceraldehide 3-phospate is oxidized to pyruvate. Two molecules of pyruvate are formed by one molecule of glucose. During the oxidation reaction, NAD^+ is reduced to NADH and four molecules of ATP are formed. Stage I requires two molecules of ATP. Thus, the energy yield at the second stage is two moles of ATP per molecule of glucose fermented. Stage III comprises the formation of final products. All products are made from pyruvate, which appears as the central pivot of fermentation. At this stage the NADH produced from NAD^+ is oxidized back to NAD^+ in order for glycolysis to continue.

2.3.1.2 Pyruvate metabolism

Pyruvate is oxidized to acetyl coenzyme A (acetyl-CoA) that can be routed in various pathways. Acetyl-CoA is routed to acetate via acetate kinase, in which ATP is generated. This reaction requires low partial pressure of H_2. Hence, the action of H_2-utilizing microorganims is central for thermodynamic feasibility. Acetyl-CoA can also be routed to butyrate. There, two molecules of acetyl-CoA are condensed to one molecule of acetoacetyl-CoA, which is subsequently reduced to butyryl-CoA. Finally, butyryl-CoA is converted to butyrate by the enzyme butyrate kinase. ATP yield of this reaction is one half of the one generated in acetate formation. Another route for butyrate synthesis is the butyryl CoA:acetate CoA transferase pathway. There, butyryl-CoA reacts with

acetate to produce butyrate and acetyl-CoA. This mechanism has been identified as the main metabolic pathway for butyrate synthesis in the human large intestine. Propionate is formed either from pyruvate or the intermediates succinate and lactate. The reaction pathways of propionate formation are the succinate pathway or the lactate-acrylate pathway. In the succinate pathway, pyruvate is carboxylated to form methylmalonyl-CoA that produces propionyl-CoA. This latter reacts with succinate, in a conversion catalyzed by the enzyme CoA transferase, producing succinyl-CoA and propionate. In the lactate-acrylate pathway, pyruvate is reduced from lactate, which is converted into the ester lactyl-CoA. This ester is dehydrated into acrylyl-CoA and subsequently reduced into propionyl-CoA to form propionate. In general, propionate production is associated with acetate production. In the overall reaction three moles of pyruvate produce two moles of propionate, one mole of acetate, one mole of CO_2, one mole of H_2O and three to five moles of ATP.

2.3.1.3 Lactate formation and utilization

Pyruvate forms lactate in its two isomers (D, L). Separate lactate dehydrogenases are responsible of the isomers to be produced. Lactate is generally produced by the homofermentative or heterofermentative pathway. In the homofermentative pathway, one mole of glucose yields two moles of lactate and two moles of ATP. The heterofermentative pathway yields one mole each of lactate, ethanol, CO_2 and ATP per mol of glucose. *Bifidobacteria* use a distinctive pathway that yields two moles of lactate, three moles acetate and five moles of ATP from two moles of glucose. Lactate participates in a reversible reaction with pyruvate. It is thus a precursor for the formation of acetate, propionate and butyrate. In these conversions, lactate is oxidized to pyruvate and the corresponding pathways are those described above for pyruvate.

2.3.1.4 Hydrogen formation and utilization

Hydrogen can be formed from pyruvate by the enzyme pyruvate:ferredoxin oxidoreductase linked to hydrogenase or through the action of NADH:ferredoxin oxidoreductase and hydrogenase (Macfarlane and Macfarlane, 2003). In the human colon, hydrogen is used by three main routes, described below.

Methanogenesis. It is attributed to *Archaea* members. Interestingly, the two methanogens that have been identified in the large intestine, namely *Methanobrevibacter smithii*

and *Methanosphaera stadtmaniae* (Miller and Wolin, 1985) carry out the conversion through two different pathways. While *M. smithii* uses four moles of H_2 to reduce one mole of CO_2 to one mole of CH_4 following

$$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O,$$

M. stadtmaniae consumes H_2 to reduce methanol to CH_4 (Miller and Wolin (1985); Fricke et al. (2006)) by the reaction

$$H_2 + CH_3OH \rightarrow CH_4 + H_2O.$$

Homoacetogenesis. It is also known as reductive acetogenesis. Colonic bacteria reduce CO_2 to produce acetate following

$$4H_2 + 2CO_2 \rightarrow CH_3COOH + 2H_2O.$$

Sulfate reduction. In this route, H_2 reacts with one mole of sulfate to produce H_2S.

For more details on H_2 utilization pathways, see Macfarlane and Gibson (1997).

The exclusion and coexistence between H_2-utilizers microorganisms in the human colonic ecosystem are not yet entirely elucidated (see, *e.g.*, Cummings and Macfarlane (1991); Doré et al. (1995); Bernalier et al. (1996a)). In our model, we assumed that the main pathways for H_2 utilization are homoacetogenis and methanogesis, and so, sulfate reduction is not accounted for.

2.4 *In vitro* and *in vivo* models to study the human colonic ecosystem

The term model is used to denote an approximate representation of a given system. Such a representation can be physical or abstract. In this thesis we developed a mathematical model (abstract representation) of the human colonic fermentation. This section is devoted to a brief review of physical models that have been developed to study the human gut ecosystem. They include *in vitro* and *in vivo* models.

Figure 2.3: Simulator of the human intestinal microbial ecosystem (SHIME model).

2.4.1 *In vitro* models

In vitro models have been designed as simplified representations of the human GI tract. Complexity of these systems varies according to the goal of the study to be pursued. In terms of their operation, they can be categorized with respect to the number of bacterial strains cultivated (pure or mixed culture), the carbon sources for bacterial growth, the operation mode (batch, semibatch or continuous), the metabolic reactions taking place and their complexity. Miller and Wolin (1981) developed a semicontinuous culture system, consisting of a single glass vessel. The fermenter was inoculated with fecal sample to study the fermentation by human intestinal microbiota. More sophisticated configurations were developed to incorporate continuous multistage culture systems (chemostats) connected in series to mimic the physiological parts of the large intestine. In these systems, temperature, pH and residence time of the compartments are controlled to reproduce physiochemical characteristics of the human colon (Gibson et al. (1988); Macfarlane et al. (1998a)). Molly et al. (1993) developed the SHIME model (Simulator of the human intestinal microbial ecosystem). This *in vitro* model couples compartments representing the stomach, small intestine and large intestine (see Fig. 2.3). In addition, the TNO gastro-Intestinal Model (TIM), developed by the Dutch company TNO Nutrition and Food Research, accounts for biotic factors such as water and SCFA absorption (Minekus (1998); Minekus et al. (1999)). It is equipped with a control system to regulate food transit. Probet

and Gibson (2004) developed a fermentation model of the proximal colon that includes mucin beads to mimic mucus gel layer microhabitat. To represent the colonic fermentation in infants, Cinquin et al. (2004) developed a single-cultured chemostat where bacteria were immobilized in polysaccaharide gel beads. Thus, microbial population was either in suspension or biofilm-associated. This model was further improved by incorporating two additional chemostats to represent the three physiological regions of the human colon (Cinquin et al., 2006b). It was used to study the effects of exopolysaccharides excreted by lactic acid bacteria on the infant colonic microbiota (Cinquin et al., 2006a). Spratt et al. (2005) developed a three-stage tubular model equipped with a membrane, allowing for water and SCFA removal. The membrane also enabled biofilm growth and led to a higher concentration of microorganisms than in conventional chemostat models. Interestingly, in this work, a simplified mathematical model was presented to define, in steady state, the hydraulic conditions that prevent bacterial washout.

In vitro models with various degrees of complexity are still being used. The *in vitro* system TIM has been used to study biopharmaceutical behavior under various physiological gastrointestinal conditions (Blanquet et al., 2003). Macfarlane and colleages have been using their model to study the impact of retention time on bacterial population and fermentation patterns (Macfarlane et al. (1998b); Child et al. (2006)), microbial colonization of mucus gel layer and metabolic activities between biofilm-associated bacteria and planktonic bacteria (Macfarlane et al., 2005). The SHIME model has been used for the screening of potential probiotics and to evaluate the effect of prebiotics such as inulin and soy on bacterial metabolism (Alander et al. (1999); Boever et al. (2000); van de Wiele et al. (2004); van de Wiele et al. (2007); Vanhaecke et al. (2009)). Other applications of *in vitro* models include studies on interspecies interactions (cross-feeding) (Belenguer et al. (2006); Chassard and Bernalier-Donadille (2006); Falony et al. (2006)), effects of diet on fermentation pattern and microbial growth (Duncan et al. (2003); Langlands et al. (2004); Amaretti et al. (2006), van der Meulen. et al. (2006), Stewart et al. (2008); Falony et al. (2009), among others).

The main limitations of *in vitro* models reside in that they do not account for host interactions, absorptive process, and differentiation between lumen and mucus microhabitats. These physiological factors are difficult to reproduce artificially, hence the use of animal

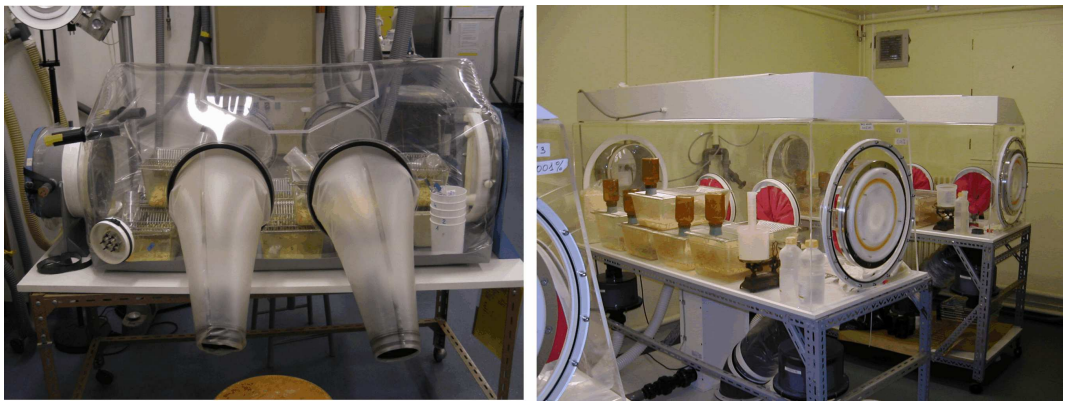

Figure 2.4: Sterile isolators for experiments with animal rodents. Facilities at UEPSD-INRA.

models is of great value.

2.4.2 *In vivo* models

The use of animals to study the human colon offers various advantages. Their diet can be controlled. In addition, intestinal contents, tissues and organs are accessible at autopsy. However, in comparison with *in vitro* models, animal experiment work is expensive and require specialist facilities (Smith et al., 2007) (see Fig. 2.4). Rodents, pigs and dogs have been used as animal models for studying human colonic fermentation. Rodents are coprophagic and their large intestine differs substantially from humans. Pigs and dogs appear to be better models (Topping and Clifton, 2001), however rodents remain the most broadly used models. Le-Blay et al. (1999) showed the butyrogenic effect of long-term ingestion of resistant starch in rats. Lan et al. (2007) studied the survival and metabolic activity of propionibacteria on the gastrointestinal tract of rats. Sato et al. (2008) studied the effect of administration of lactate-utilizing bacteria, coupled with ingestion of galacto-oligosaccarides (GOS), on the fermentation pattern in rats. Fleming et al. (1991) used rat models to study the SCFA absorption in the cecum.

A novel approach to the study of symbiotic host-bacterial relationships in the human GI tract is based on the use of germ-free (axenic) rodents (Hooper et al. (2002); Xu and Gordon (2003); Phillips (2009)). These animals are born and raised in sterile environments (see Fig. 2.4). They can be colonized with specified members of the human colonic mi-

crobiota. The colonized animal is referred to as gnotobiotic (known-life). Studies with gnotobiotic rodents includes colonization, interspecies interaction, host responses, among others. For example, Sarma-Rupavtarm et al. (2004) used mice inoculated with the altered Schaedler flora (Orcutt et al., 1987) to study the spatial distribution of this microbial consortium along the GI tract. Quantitative PCR analysis showed that bacterial distributions in feces and the colon largely differ.

The bacterial species *Bacteriodes thetaiotaomicron* has special characteristics that favored its selection as candidate for gnotobiotic experiments (see *e.g.* Xu and Gordon (2003); Bäckhed et al. (2005)). Samuel and Gordon (2006) used a gnotobiotic model to demonstrate cross-feeding interactions between *B. thetaiotaomicron* and *M. smithii*. *M. smithii* drove the metabolic activity of *B. thetaiotaomicron* and this latter produced formate that was used by *M. smithii*. Interestingly, cocolonization with these strains produced an increase in host adiposity. Mahowald et al. (2009) studied bacterial and host interactions in gnotobiotic mice inoculated with *B. thetaiotaomicron* and *E. rectale*. Metagenomic analysis in this study suggested that both bacterial species worked in symbiosis. Moreover, such a cross-feeding produced a bacterial signal that induced the host to synthesize mucosal glycans that can be metabolized by *B. thetaiotaomicron*. These approaches have shed substantial light on *in vivo* metabolism of the human colonic microbiota and its interaction with the host.

2.5 Conclusions

The human colon is a complex ecosystem with three well defined physiological regions: the proximal colon, the transverse colon and the distal colon. It harbors a myriad of microorganisms that constitutes the human colonic microbiota, which is located in two microhabitats: lumen and mucus. This microbial collection can be regarded as a specialized organ that confers to mankind specific metabolic functions that human genome lacks. The colonic microbiota mediates many of the effects of diet upon gut health. It performs a complex series of anaerobic reactions to degrade polysaccarides from diet and host origin. Main products of this conversion are SCFA, CO_2, H_2, biomass (microorganisms), and CH_4. Due to its physiological characteristics and microbial syntrophism, the human colon can be regarded as a high-performance reactor.

The large intestine is largely inaccessible for routine investigation. Therefore, a variety of *in vitro* and *in vivo* models have been developed to get insight on microbial metabolism and interactions that shape the dynamics of such a complex system. Those models have been crucial for the understanding of the human GI tract. Nevertheless, extrapolation of such studies to the human colon ecosystem has to be done with caution, because *in vitro* models do not account for biotic factors and animals models are physiologically different from human beings. Hence, a mathematical model appears as an interesting approach to complement the existing physical models.

Chapter 3

Mathematical model construction

The construction of a model, therefore, was for him a miracle of equilibrium between principles (left in shadow) and experience (elusive), but the result should be more substantial than either. In a well-made model, in fact, every detail must be conditioned by the others, so that everything holds together in absolute coherence, as in a mechanism where if one gear jams, everything jams.

Palomar. Italo Calvino

3.1 Knowledge basis

Part of this section was published In Proc. 11th World Congress Anaerobic Digestion Bio-energy for our Future. Brisbane, Australia (Muñoz-Tamayo et al., 2007).

The human colon can be viewed as a complex bioreactor. As chemical reactor models do (see, *e.g.*, Fogler (1999)), our mathematical model structure integrates a hydraulic representation of the system, and a description of the transport phenomena and reaction

mechanisms. These three aspects are described below.

3.1.1 Hydraulic representation

A hydraulic representation of the large intestine is displayed in Fig. 3.1. The three physiological sections described in the previous chapter (see Fig. 2.1, page 61) are accounted for. In each of these sections, the lumen and mucus microhabitats are distinguished. Lumen and mucus compartments are described as continuous-flow stirred tank reactors (CSTR). The lumen reactors are connected in a series configuration. The representation of the human colon as series of complete-mixed reactors is in agreement with similar studies on animal guts (Hume (1997); Jumars (2000)).

We assume that the compartments are completely mixed. This is partially justified by the presence of mixing forces such as peristaltic movement and gas production. However, its is clearly an approximation. In its travel, the chyme looses water, increases its viscosity, and thus the mixing is only partial, so the hydraulic regime in the lumen may be closer to that of a plug-flow reactor (PFR). Because of the lack of data to justify an increase in complexity, we do not, however, consider changes in the hydraulic regime along the human colon.

Roughly speaking, the human colon has a cylinder shape. For hand-made reactors, such geometry does not seem to favor microbial growth. If the growth rate of microorganisms is slower than the dilution rate (the inverse of the retention time), then microbial washout occurs. The human colon lacks carrier material to enable the microbes to anchor. Similarly to what takes place in high-rate reactors (see, *e.g.*, Lettinga et al. (1980); Liu et al. (2003)), human colonic microorganisms form dense aggregates (a process known as self-immobilization) to resist hydrodynamic forces and prevent washout. These aggregates consist of microbes, food particles, shed epithelial cells, and mucus gel (Sonnenburg et al., 2004).

To take into account microbial aggregation in our model, we consider that microorganisms have a higher residence time than the hydraulic residence time.

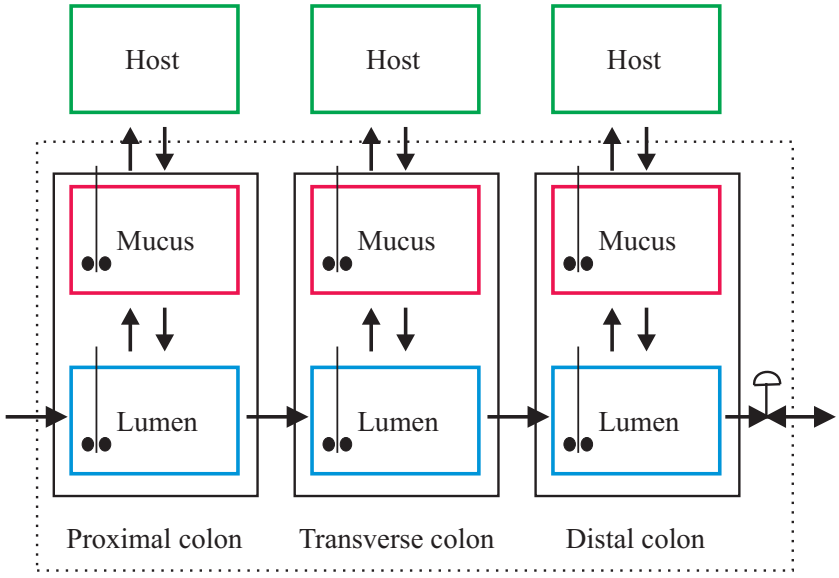

Figure 3.1: Hydraulic representation of the human colon. The dotted rectangle represents the boundary of the system to be modelled.

3.1.2 Transport phenomena

The human colon involves transport phenomena such as absorption of metabolites by the host. Removal of water and metabolites are prerequisites to maintaining the normal physiology of the system (Minekus et al., 1999). In our model, we consider diffusion of sugars between the lumen and mucus microhabitats, SCFA, lactate and water absorption in the lumen and mucus, liquid-gas transfer, and shear-loss phenomenon for the mucus-associated microbiota.

3.1.3 Carbohydrate degradation

The human colonic microbiota catalyzes a complex metabolic network to degrade available carbon sources. In Section 2.3.1, we described the main conversions taking place in the large intestine. In our mathematical model, these conversions are represented in an aggregated pathway. Figure 3.2 displays a scheme integrating carbohydrate fermentation and transport phenomena, in agreement with Macfarlane et al. (2006) and Louis et al. (2007b). Figures 3.1 and 3.2 thus summarize the basis of the mathematical model.

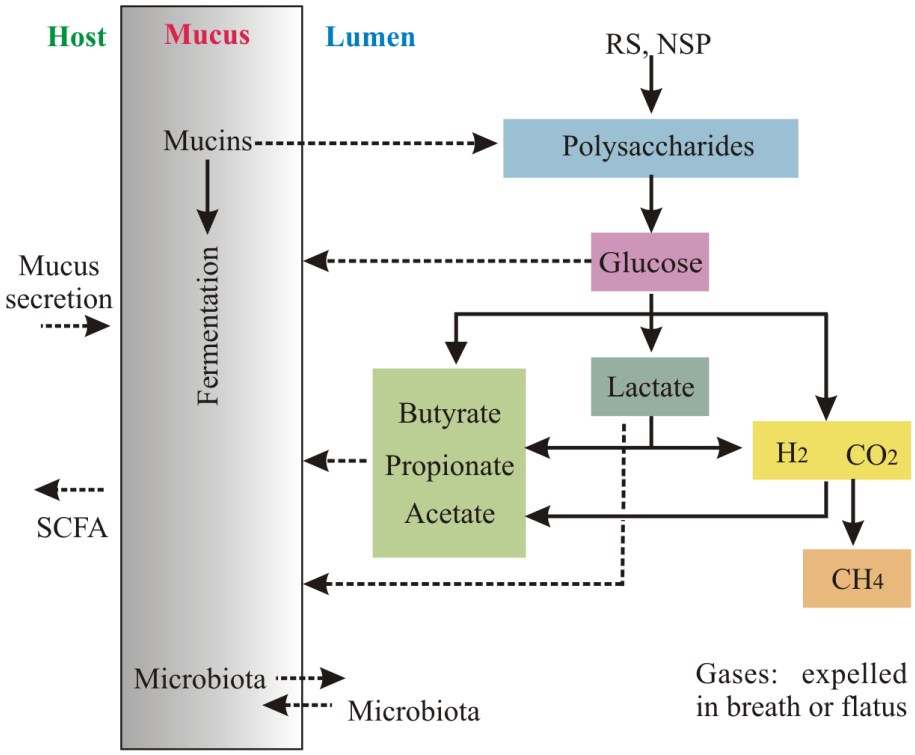

Figure 3.2: Reaction pathway of carbohydrate degradation in the human colon and mass transfer between the lumen and mucus, and mucus and host. Dotted lines represent transport phenomena, while continuous lines represent metabolic transformations.

The microbiota is associated to the microhabitat where it resides, lumen or mucus. It is represented in functional groups according to its role in the trophic chain. Thus, for each substrate utilization, there is a microbial group that catalyzes the conversion. This approach is often used for modelling ecological systems (Ballyk et al., 2001). Functional representation of the microbiota enables one to conciliate the paradox of host specificity of the GI tract microbiota and functional homogeneity among individuals. Studies in anaerobic reactors indicated that complex communities of different compositions may have similar functional characteristics (Fernández et al. (1999); Fernández et al. (2000)). For the human colon, it has been suggested that the limited number of biochemical routes in the metabolic pathway of fermentation leads to functional redundancy in the microbial consortium (Egert et al., 2006). Functional redundancy in the microbial community ensures that key processes are robust to diversity (Goebel and Stackebrandt, 1994). A very recent study suggests that humans share a common phylogenetic core of microorganisms

in their lower GI tract (Tap et al., 2009). The existence of a set of bacterial species present in all healthy individuals may be linked to a common functional core. Indeed, a large range of metabolic functions of carbohydrate metabolism were inferred from the species identified in the phylogenetic core. Moreover, it has been suggested that host-driven selection favors functionally stable communities (Ley et al., 2006a) and that in spite of inter-individual bacterial diversity, there is a core gut microbiome responsible for a uniform pattern of functional categories of genes and metabolic pathways (Turnbaugh et al., 2009). In the same direction, metagenomic studies have showed high functional uniformity between healthy individuals (Gill et al. (2006); Kurokawa et al. (2007)). Thus, a functional representation of the microbiota in the mathematical model appears feasible.

3.1.4 Gaining knowledge from anaerobic digestion modelling

We took advantage of the extensive research that has been carried out on anaerobic reactors. Anaerobic modelling dates from 1970 (see, *e.g.*, Lyberatos and Skiadas (1999); Gavala et al. (2003) for reviews). The model structure proposed by Mosey (1983) and further developed by Costello et al. (1991) served as basis of various model extensions. Many of further developments kept the same structure with slight differences (see, *e.g.*, Batstone et al. (2000); Skiadas et al. (2000); Bernard et al. (2001)). In 2002, the IWA Task Group for Mathematical Modelling of Anaerobic Processes developed the generic model ADM1 (Batstone et al., 2002), with the aim of establishing a common platform for anaerobic process analysis. Nowadays, ADM1 represents the state of the art in the modelling of anaerobic digestion processes. It has been extensively used and extended to represent a large number of wastewater processes, and limitations and need for further improvements have been identified (Batstone et al. (2006a); Kleerebezem and van Loosdrecht (2006)). ADM1 has also been used to explore novel approaches such as individual-based modelling (Batstone et al., 2006b) and variable stoichiometry (Rodríguez et al. (2006b); Penumathsa et al. (2008)). Recently, ADM1 was extended to account for microbial diversity (Ramirez et al., 2009b).

We used ADM1 as a basis and modified it to account for the specific characteristics of the human colon. For example, in our model, reductive acidogenesis is included, as well as the contribution of lactate to butyrate production. Experimental observations have in-

dicated the role of lactate as a precursor for butyrate production (Duncan et al. (2004b); Bourriaud et al. (2005)). Oxidative acetogenesis is not considered in our model, because SCFA are rapidly absorbed by the host. As in ADM1, glucose is considered as the sugar gathering the constituent monomers from polysaccharides hydrolysis. The current version of the model does not include acid-base reactions, pH calculation and inhibition terms for the kinetic rates. The mathematical model is derived from mass balances of the metabolites and microbial groups that participate in the conversion process.

3.2 Formulation of the mathematical model

3.2.1 Proximal colon and transverse colon

The sections corresponding to the proximal colon and transverse colon have the same structure. It is assumed that the volumes of all the phases in each compartment are constant. For the lumen, the mass balance equations of the components in the planktonic phase yield the following differential equations, where the dot symbol denotes derivation with respect to time. The model nomenclature is given in Table 3, page 49 .
For soluble component s_i

$$\dot{s}_i^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} s_{i,\mathrm{in}}^{\mathrm{l}} - \frac{q_{\mathrm{out}}}{V^{\mathrm{l}}} s_i^{\mathrm{l}} - \gamma_i^{\mathrm{l}} s_i^{\mathrm{l}} + \sum_{j=1}^{5} Y_{i,j}^{\mathrm{l}} \rho_j^{\mathrm{l}} - Q_i^{\mathrm{l}}, \tag{3.1}$$

where V is the volume and q denotes the flow rate at the input (in) and output (out) of the system. The superscript l indicates the lumen microhabitat. Q is the liquid-gas transfer rate and γ is a transport coefficient. In the case of SCFA, γ is the absorption factor. Here, absorption is modelled as a convective transport phenomenon. For sugars, the transport term is associated to a diffusion phenomenon. Therefore, $\gamma_i^{\mathrm{l}} s_i^{\mathrm{l}}$ in (3.1), is replaced by $\gamma_i (s_i^{\mathrm{l}} - s_i^{\mathrm{m}})/V^{\mathrm{l}}$, with the superscript m indicating the mucus microhabitat. Microbial shear and adherence are expressed by first-order equations, as proposed by Hsien and Lin (2005) and de Jong et al. (2007). The kinetic rates of the transformation processes (hydrolysis, substrate utilization and decay of microorganisms) are represented by ρ_j, with j denoting the process. $Y_{i,j}$ is the yield of component i in process j. The hydrolysis rate of polysaccharides (ρ_1) is described by the Contois equation to account for dependence on sugar-utilizing bacteria. Contois equation is well adapted to a wide range of organic wastes in anaerobic digestion and is considered as a general model of the complex process

of hydrolysis (Vavilin et al., 2008). The kinetic rates for substrate utilization are described by the Monod equation. Decay rate for the biomasses are assumed to follow a first-order biomass-based kinetics. The biochemical process rates and stoichiometry matrix are depicted in Table 3.1. This representation is often called as the Petersen matrix (Petersen, 1965).

The carbohydrates brought in by the alimentary diet are gathered into the single polysaccharide z, which obeys the equation

$$\dot{z}^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} z^{\mathrm{l}}_{\mathrm{in}} + \frac{q^{\mathrm{m}}_{\mathrm{out}}}{V^{\mathrm{l}}} z^{\mathrm{m}} - \frac{q_{\mathrm{out}}}{V^{\mathrm{l}}} z^{\mathrm{l}} - \rho^{\mathrm{l}}_{1}. \tag{3.2}$$

Microbial group x_i satisfies

$$\dot{x}^{\mathrm{l}}_{i} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} x^{\mathrm{l}}_{i,\mathrm{in}} - \frac{1}{\tau_i + V^{\mathrm{l}}/q_{\mathrm{out}}} x^{\mathrm{l}}_{i} + b_i \frac{V^{\mathrm{m}}}{V^{\mathrm{l}}} x^{\mathrm{m}}_{i} - a_i x^{\mathrm{l}}_{i} + \sum_{j=2}^{9} Y^{\mathrm{l}}_{i,j} \rho^{\mathrm{l}}_{j}. \tag{3.3}$$

The coefficients a_i and b_i refer to the adherence and shear phenomena. Aggregation of the microbial group x_i is represented by incorporating an additional residence time τ_i.

For the hydrogenotrophic methanogenesis, inhibition due to pH is taken into account as proposed in Batstone et al. (2002).

$$I_{\mathrm{pH}} = \begin{cases} \exp\left(-3\left(\frac{pH - pH_{\mathrm{U}}}{pH_{\mathrm{U}} - pH_{\mathrm{L}}}\right)^2\right) & \text{if} \quad pH < pH_{\mathrm{U}}, \\ 1 & \text{if} \quad pH \geq pH_{\mathrm{U}}, \end{cases} \tag{3.4}$$

where I_{pH} is the pH inhibition function, pH_{U} and pH_{L} are upper and lower values of pH in which microorganisms are not inhibited.
The kinetic rate ρ_5 (see Table 3.1) for hydrogenotrophic methanogenesis is then given by

$$\rho_5 = k_{m,\mathrm{H_2m}} \frac{s_{\mathrm{H_2m}} x_{\mathrm{H_2m}}}{K_{s,\mathrm{H_2m}} + s_{\mathrm{H_2m}}} I_{\mathrm{pH}} \tag{3.5}$$

To maintain a constant volume in lumen, the output flow rate is calculated as

$$q_{\mathrm{out}} = q_{\mathrm{in}} + \sum_{i=11}^{14} b_i x^{\mathrm{m}}_{i} V^{\mathrm{m}} w_i / r_i + q^{\mathrm{m}}_{\mathrm{out}} - \sum_{i=2, i\neq 3,7,8}^{9} \gamma^{\mathrm{l}}_{i} s^{\mathrm{l}}_{i} V^{\mathrm{l}} w_i / r_i - \sum_{i=11}^{14} a_i x^{\mathrm{l}}_{i} V^{\mathrm{l}} w_i / r_i, \tag{3.6}$$

with r_i the density and w_i the molecular weight of component i.

Table 3.1: Petersen matrix of the biochemical reactions. In the lumen, z is the aggregated concentration of polysaccharides from dietary origin plus the mucins coming from the mucus. In the mucus, z is the concentration of mucins originating from the host.

For soluble components											
	Component$\rightarrow i$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Kinetic**
j	**Process** $\downarrow$	s_{su}	s_{la}	s_{H_2}	s_{ac}	s_{pro}	s_{bu}	s_{CH_4}	s_{CO_2}	s_{H_2O}	**rate**
1	Hydrolysis	$Y_{su,z}$									$\rho_1 = k_{hyd,z}\frac{zx_{su}}{K_{x,z}x_{su}+z}$
2	Glucose utilization	-1	$Y_{la,su}$	$Y_{H_2,su}$	$Y_{ac,su}$	$Y_{pro,su}$	$Y_{bu,su}$		$Y_{CO_2,su}$	$Y_{H_2O,su}$	$\rho_2 = k_{m,su}\frac{s_{su}x_{su}}{K_{s,su}+s_{su}}$
3	Lactate utilization		-1	$Y_{H_2,la}$	$Y_{ac,la}$	$Y_{pro,la}$	$Y_{bu,la}$		$Y_{CO_2,la}$	$Y_{H_2O,la}$	$\rho_3 = k_{m,la}\frac{s_{la}x_{la}}{K_{s,la}+s_{la}}$
4	Hydrogen utilization: Homoacetogenesis			-1	Y_{ac,H_2a}				Y_{CO_2,H_2a}	Y_{H_2O,H_2a}	$\rho_4 = k_{m,H_2a}\frac{s_{H_2a}x_{H_2a}}{K_{s,H_2a}+s_{H_2a}}$
5	Hydrogen utilization: Methanogenesis			-1				Y_{CH_4,H_2m}	Y_{CO_2,H_2m}	Y_{H_2O,H_2m}	$\rho_5 = k_{m,H_2m}\frac{s_{H_2m}x_{H_2m}}{K_{s,H_2m}+s_{H_2m}}I_{pH}$

For particulate components							
	Component$\rightarrow i$	**10**	**11**	**12**	**13**	**14**	**Kinetic**
j	**Process** $\downarrow$	z	x_{su}	x_{la}	x_{H_2a}	x_{H_2m}	**rate**
1	Hydrolysis	-1					ρ_1
2	Glucose utilization		Y_{su}				ρ_2
3	Lactate utilization			Y_{la}			ρ_3
4	Hydrogen utilization: Homoacetogenesis				Y_{H_2a}		ρ_4
5	Hydrogen utilization: Methanogenesis					Y_{H_2m}	ρ_5
6	Decay of x_{su}		-1				$\rho_6 = k_d x_{su}$
7	Decay of x_{la}			-1			$\rho_7 = k_d x_{la}$
8	Decay of x_{H_2a}				-1		$\rho_8 = k_d x_{H_2a}$
9	Decay of x_{H_2m}					-1	$\rho_9 = k_d x_{H_2m}$

Equation (3.6) imposes conditions for the selection of the transport coefficients to ensure the positivity of q_{out}. In the mucus layer, the mass balance equations in the gel phase lead to the following differential equations.

For soluble component i

$$\dot{s}_i^{\text{m}} = \gamma_i^{\text{l}} s_i^{\text{l}} \frac{V^{\text{l}}}{V^{\text{m}}} - \gamma_i^{\text{m}} s_i^{\text{m}} + \sum_{j=1}^{5} Y_{i,j}^{\text{m}} \rho_j^{\text{m}} - Q_i^{\text{m}}. \tag{3.7}$$

The polysaccharides are the mucins (glycans) produced by the host. Their concentration is assumed to satisfy

$$\dot{z}^{\text{m}} = \frac{\Gamma}{V^{\text{m}}} - \frac{q_{\text{out}}^{\text{m}}}{V^{\text{m}}} z^{\text{m}} - \rho_1^{\text{m}}, \tag{3.8}$$

with Γ the endogenous production of mucins.

The concentration of microbial group i is assumed to satisfy

$$\dot{x}_i^{\text{m}} = a_i \frac{V^{\text{l}}}{V^{\text{m}}} x_i^{\text{l}} - b_i x_i^{\text{m}} + \sum_{j=2}^{9} Y_{i,j}^{\text{m}} \rho_j^{\text{m}}. \tag{3.9}$$

To maintain a constant volume in mucus, the output flow rate is calculated as

$$q_{\text{out}}^{\text{m}} = \frac{\Gamma}{1000 * 0.05} + \sum_{i=11}^{14} a_i x_i^{\text{l}} V^{\text{l}} w_i / r_i - \sum_{i=2, i\neq 3,7,8}^{9} \gamma_i^{\text{m}} s_i^{\text{m}} V^{\text{m}} w_i / r_i - \sum_{i=11}^{14} b_i x_i^{\text{m}} V^{\text{m}} w_i / r_i. \tag{3.10}$$

The denominator for Γ in (3.10) comes from the mucus density (1000 g/l), and the assumption that mucins represent 5% of the mucus. The remaining 95% is considered as water.

For each microhabitat, the mass balance equation for the components in the gas phase (subscript g) yields

$$\dot{s}_{\text{g},i} = \frac{q_{\text{g,in}}}{V_{\text{g}}} s_{\text{g},i,\text{in}} - \frac{q_{\text{g,out}}}{V_{\text{g}}} s_{\text{g},i} + Q_i \frac{V}{V_{\text{g}}}. \tag{3.11}$$

The liquid-gas transfer rate is expressed as

$$Q_i = k_L a (s_i - K_{H,i} R T s_{\text{g},i}), \tag{3.12}$$

where $k_L a$ is the liquid-gas transfer coefficient multiplied by the specific transfer area, K_H the Henry law coefficient, R the ideal gas constant and T the absolute temperature.

The output gas flow rate is calculated as

$$q_{\text{g,out}} = q_{\text{g,in}} + q_{\text{g,lgt}}, \tag{3.13}$$

where $q_{\text{g,lgt}}$ is the gas flow rate due to the liquid-gas transfer, calculated as proposed in Batstone et al. (2002)

$$q_{\text{g,lgt}} = \frac{RT}{P_{\text{atm}} - p_{\text{H}_2\text{0}}} V \left(Q_{\text{H}_2} + Q_{\text{CH}_4} + Q_{\text{CO}_2}\right), \tag{3.14}$$

where P_{atm} is the atmospheric pressure and $p_{\text{H}_2\text{0}}$ is the water vapor pressure. At 37°C, $p_{\text{H}_2\text{0}} = 0.08274$ bar (Smith et al., 2005).
In reactors, when the headspace pressure is variable, the gas flow is calculated as

$$q_{\text{g,out}} = k_p (P_{\text{gas}} - P_{\text{atm}}), \tag{3.15}$$

with k_p a resistance coefficient and P_{gas} the sum of the partial pressures of the gaseous components. We consider that the headspace of the human colon reactor is the atmospheric pressure, hence (3.13) is used in our model.

All concentrations are in molar units (M), except for polysaccharides concentration, expressed in g/l.

3.2.2 Distal colon

The lumen compartment in the distal colon section is modelled as a semi-batch reactor. Therefore, its volume cannot be considered as a constant. It satisfies

$$\dot{V}^{\text{l}} = q_{\text{n}}. \tag{3.16}$$

The net flow rate q_{n} is calculated as the right-hand side of (3.6). Consequently, the behaviors of concentrations in the lumen, derived from the mass balance equations, differ from what was obtained in the proximal and transverse sections.

Soluble component i satisfies

$$\dot{s}_i^{\text{l}} = \frac{q_{\text{in}}}{V^{\text{l}}} s_{i,\text{in}}^{\text{l}} - \frac{q_{\text{n}}}{V^{\text{l}}} s_i^{\text{l}} - \gamma_i^{\text{l}} s_i^{\text{l}} + \sum_{j=1}^{5} Y_{i,j}^{\text{l}} \rho_j^{\text{l}} - Q_i^{\text{l}}. \tag{3.17}$$

For polysaccharide z

$$\dot{z}^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} z^{\mathrm{l}}_{\mathrm{in}} + \frac{q^{\mathrm{m}}_{\mathrm{out}}}{V^{\mathrm{l}}} z^{\mathrm{m}} - \frac{q_{\mathrm{n}}}{V^{\mathrm{l}}} z^{\mathrm{l}} - \rho^{\mathrm{l}}_{1}. \tag{3.18}$$

The evolution of the concentration of microbial group i is described by

$$\dot{x}^{\mathrm{l}}_{i} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} x^{\mathrm{l}}_{i,\mathrm{in}} - \frac{q_{\mathrm{n}}}{V^{\mathrm{l}}} x^{\mathrm{l}}_{i} + b_i \frac{V^{\mathrm{m}}}{V^{\mathrm{l}}} x^{\mathrm{m}}_{i} - a_i x^{\mathrm{l}}_{i} + \sum_{j=2}^{9} Y^{\mathrm{l}}_{i,j} \rho^{\mathrm{l}}_{j}. \tag{3.19}$$

Excretion is modelled as a periodic reset of the volume V^{l}.

In view of the dynamics of the volume of the liquid phase, the gas phase volume also evolves and the concentrations of gas components follow

$$\dot{s}^{\mathrm{l}}_{\mathrm{g},i} = \frac{q^{\mathrm{l}}_{\mathrm{g,in}}}{V^{\mathrm{l}}_{\mathrm{g}}} \left(s^{\mathrm{l}}_{\mathrm{g},i,\mathrm{in}} - s^{\mathrm{l}}_{\mathrm{g},i} \right) - \frac{q^{\mathrm{l}}_{\mathrm{g,lgt}} - q_{\mathrm{n}}}{V^{\mathrm{l}}_{\mathrm{g}}} s^{\mathrm{l}}_{\mathrm{g},i} + Q^{\mathrm{l}}_{i} \frac{V^{\mathrm{l}}}{V^{\mathrm{l}}_{\mathrm{g}}}, \tag{3.20}$$

where $V^{\mathrm{l}}_{\mathrm{g}} = V_{\mathrm{d}} - V^{\mathrm{l}}$ with V_{d} the total volume of the luminal distal compartment.

For the sake of simplicity, the equations do not include an index referencing the physiological region to which they correspond (*i.e.* proximal colon (P), transverse colon (T), distal colon (D)). The reader should consider this implicit assignment.

Although the scope of this work does not cover the modelling of the small intestine, we included a simple model of three CSTRs in series, to dampen the oscillatory nature of feeding. The concentration of polysaccharide z in the jth CSTR for the small intestine satisfies

$$\dot{z}_j = D(z_{j-1} - z_j), \tag{3.21}$$

with D the dilution rate, set to 9 d^{-1} according to Boever et al. (2000).

The resulting model consists of 102 first-order ordinary differential equations. Table 3.2 shows the number of model parameters. The parameter vector has 333 components. In Section 6.1, we show how the dimension of the parameter vector is reduced.

Table 3.2: Number of model parameters

Parameters	**Lumen subsystem**	**Mucus subsystem**	**In each physiological region**	**Whole system**
Reactions				
Hydrolysis				
Contois parameters	2	2	4	12
Yield coefficients	1	1	2	6
Glucose utilization				
Monod parameters	2	2	4	12
Yield coefficients	8	8	16	48
Lactate utilization				
Monod parameters	2	2	4	12
Yield coefficients	7	7	14	42
Hydrogen utilization				
Homoacetogenesis				
Monod parameters	2	2	4	12
Yield coefficients	4	4	8	24
Methanogenesis				
Monod parameters	2	2	4	12
Yield coefficients	4	4	8	24
Decay rate				
Decay constants	4	4	8	24
Subtotal	38	38	76	228

Number of model parameters(continued).

Parameters	**Lumen subsystem**	**Mucus subsystem**	**In each physiological region**	**Whole system**
Transport				
Glucose	1		1	3
Lactate	1	1	2	6
Acetate	1	1	2	6
Propionate	1	1	2	6
Butyrate	1	1	2	6
Water	1	1	2	6
Microbial phenomena				
Shear constants		4	4	12
Adherence constants	4		4	12
Additional residence time	4		4	12
Liquid-gas transfer				
H_2	2	2	4	12
CH_4	2	2	4	12
CO_2	2	2	4	12
Subtotal	20	15	35	105
Total	58	53	111	**333**

3.3 Conclusions

A major part of the time devoted to this work was dedicated to the understanding of the phenomena involved in carbohydrate degradation in the human colon ecosystem, and to the incorporation of such a knowledge into a mathematical model. Hence, this chapter is the core of the thesis. The model developed here is to the best of our knowledge the first one that integrates the physiology of the intestine, biological reactions and transport phenomena between the lumen and mucus microhabitats. It should provide a close picture of the complexity of the human colon ecosystem.

Our model construction process was based on mathematical models for anaerobic reactors, that we adapted to the specificities of the human colon. Our model is unstructured and unsegregated. It has a high dimension structure, both in its state vector and its pa-

rameter vector. In the following chapter, the strategy that we advocate for handling such a model complexity is described. In Chapter 7, the strengths and limitations of the model are discussed.

Chapter 4

Handling a complex model

Diviser chacune des difficultés que j'examinerais, en autant de parcelles qu'il se pourrait, et qu'il serait requis pour les mieux résoudre. Conduire par ordre mes pensées, en commençant par les objets les plus simples et les plus aisés à connaître, pour monter peu à peu comme par degrés jusques à la connaissance des plus composés, et supposant même de l'ordre entre ceux qui ne se précèdent point naturellement les uns les autres.
Discours de la méthode pour bien conduire sa raison, et chercher la vérité dans les sciences. René Descartes

Although our mathematical model is already a simplification of the real system, its structure remains highly complex. It comprises a large number of state variables and parameters. Hence, model implementation, analysis and simulation are not simple tasks. In addition to model complexity, scarcity of data is a serious obstacle for parameter estimation (identification). This section is dedicated to describing how we dealt with this situation. While some of the methods we use are natural and their analysis is straightforward, other methods deserve a more detailed description. The theoretical background of our strategy is recalled.

4.1 Model simplification

Model simplification relates to the principle of parsimony (also know as Occam's razor). This principle states, as a general rule, that simpler theories should be preferred to more complex ones. Translating to model construction, one would like to build the less complex model that can still adequately represent the system under study. Model simplification offers various advantages in identification, computer simulation, state observation and design of control strategies. We may be interested in reducing the number of parameters of our model, the dimension of its state vector, or both. The approaches that we used in this work to simplify our model are described below.

4.1.1 Model structure simplification

When dealing with complex systems, model structure can be defined at different levels of aggregation. For example, the reaction scheme of our model (Fig. 3.2, page 80) is an aggregated representation of the multiple intermediate steps implied in the microbial metabolism, where the main fermentation pathways are included. In our first attempt towards conceptual structure definition (Muñoz-Tamayo et al., 2007), we thought to incorporate the intermediate component succinate into the model. Further analysis led us to discard this idea. Succinate is rapidly metabolized by the microorganisms and, in normal conditions, does not accumulate. This is why we decided to embed succinate metabolism into both glucose and lactate metabolisms. Including succinate in the model structure would have implied twelve additional ODEs and a subsequent increase in the number of parameters.

Microbial population models can be categorized with respect to what the modeller assumes on the homogeneity of individuals of the population, and on the intracellular behavior. If the model accounts for the presence of heterogeneous individuals, the model is called *segregated*. If the model accounts for various chemical components and their interactions within the cell, the model is called *structured* (see, *e.g.*, Bailey (1998)). In this sense, our mathematical model is unstructured and unsegregated. We already mentioned that the colonic microbiota in a given human being consists of more than 1000 microbial species. Considering the dynamics of individual species would lead to an intractable model. Instead of having an individually-based description, we represent the microbial

consortium in functional groups according to their role in the trophic chain.

Structure simplification may also be directed to reducing the number of model parameters. Indeed, when various model structures compete for the representation of a given system, the dimension of the parameter vector provides criteria for model selection (see, *e.g.*, Akaike's information criterion (AIC) (Akaike, 1974)). Mathematical tools such as sensitivity analysis are available to identify the most relevant parameters of the model and thus neglect those with little effect (see Section 4.4).

4.1.2 Knowledge-based parameter reduction

The use of knowledge on the system under study can considerably reduce the number of parameters to be estimated. We refer here to relationships between parameters or laws that have to be fulfilled (*e.g.* the law of mass conservation). In Section 6.1.1 we define a set of reactions describing carbohydrate fermentation. By taking stoichiometry into account, relationships between yield factors can be obtained. Thanks to this procedure, we maintain the balance of chemical elements and reduce considerably the number of model parameters.

4.1.3 Model order reduction

Model order reduction aims at reducing the dimension of the state vector. Order reduction can be carried out by methods based on singular-value decomposition (SVD), or by singular perturbation. For example, Bernard et al. (2006) showed, by means of principal component analysis (PCA), that the complex reaction pathway of anaerobic digestion can be efficiently represented by simpler models. They showed that with two bacterial groups and two reactions it was possible to represent the global behavior of the ADM1 model which comprises nineteen reactions and seven bacterial groups. The singular perturbation approach relies on the dynamical properties of the model. Biochemical models may have multiple time scales. In numerical computation, this property is also referred to as stiffness. In linear models, model stiffness can be checked by comparing the eigenvalues of the state matrix. The higher the ratio between the real parts of the largest and the smallest eigenvalues is, the stiffer the model. Multiple time scales hamper integration, and an appropriate method must be used to solve the ODEs. Otherwise, the simulation can

be time-consuming or even numerically unstable. The Matlab software provides several ODE solvers adapted to stiff models.

When a model has several time scales, the singular perturbation method can be used to reduce the number of ODEs, by considering the fast variables as instantaneous. After such a transformation, the model is referred to as a differential algebraic equation (DAE) model or a quasi-steady-state model. The singular perturbation method has proved to be useful in analyzing complex chemical reaction networks (Okino and Mavrovouniotis (1998); Pécou (2005); Ropers et al. (2009)). Application of this approach on wastewater treatment plant models are presented by Breusegem and Bastin (1991); Steffens et al. (1997). The principle of the singular perturbation method is recalled in the next section (see Khalil (2000) for more details).

4.1.3.1 Singular perturbation method

When some model variables move faster than others, the state can be represented as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{z}, t, \delta), \qquad \mathbf{x}(t_0) = \boldsymbol{\xi}(\delta), \tag{4.1}$$

$$\delta\dot{\mathbf{z}} = \mathbf{g}(\mathbf{x}, \mathbf{z}, t, \delta), \qquad \mathbf{z}(t_0) = \boldsymbol{\zeta}(\delta). \tag{4.2}$$

Here, $\mathbf{x}$ comprises the n slow state variables and $\mathbf{z}$ the m fast state variables. The vector-valued functions $\mathbf{f}$ and $\mathbf{g}$ are assumed to be continuously differentiable in their arguments $(\mathbf{x}, \mathbf{z}, t, \delta) \in D_x \times D_z \times [0, t_1] \times [0, \delta_0]$, where $D_x \subset \mathbb{R}^n$, $D_z \subset \mathbb{R}^m$ are open connected sets. The state depends on the positive scalar parameter δ. When $\delta = 0$, (4.2) is transformed into the algebraic equation

$$\mathbf{g}(\mathbf{x}, \mathbf{z}, t, 0) = \mathbf{0}, \tag{4.3}$$

thus reducing the number of ODEs. The model (4.1)-(4.2) is called a *standard singular perturbation* model if (4.3) has $k \geq 1$ isolated real roots. Let $\mathbf{z}^*$ be one of these roots, defined by

$$\mathbf{z}^* = \mathbf{h}(\mathbf{x}, t). \tag{4.4}$$

The reduced model is then given by

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{h}(\mathbf{x}, t), t, 0), \quad \mathbf{x}(t_0) = \boldsymbol{\xi}(0). \tag{4.5}$$

This model is referred to as a quasi-steady state model if $\mathbf{z}$ rapidly converges to $\mathbf{z}^*$.

With $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^m$, the initial value problem for the ODE model shifts to solving n ODEs instead of $n+m$. The quasi-steady state $\mathbf{z}^*$ is computed by solving (4.5) and using (4.4). Hence, $\mathbf{z}^*$ is not free to start from a prescribed value at t_0, so the computed initial condition may largely differ from that of the original model (4.2). Thus, $\mathbf{z}^*$ cannot be a uniform approximation of $\mathbf{z}$. To guarantee that $\mathbf{z}$ will converge to $\mathbf{z}^*$, certain stability conditions must be satisfied. Such conditions are presented below.

Consider the change of variables $\mathbf{y} = \mathbf{z} - \mathbf{h}(\mathbf{x},t)$. The vector $\mathbf{y}$ satisfies

$$\delta\dot{\mathbf{y}} = \mathbf{g}(\mathbf{x},\mathbf{y}+\mathbf{h}(\mathbf{x},t),t,\delta) - \delta\frac{\partial \mathbf{h}(\mathbf{x},t)}{\partial t} - \delta\frac{\partial \mathbf{h}(\mathbf{x},t)}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x},\mathbf{y}+\mathbf{h}(\mathbf{x},t),t,\delta),$$
$$\mathbf{y}(t_0) = \boldsymbol{\zeta}(\delta) - \mathbf{h}(\boldsymbol{\xi}(\delta),t_0). \tag{4.6}$$

Denote by υ the new time variable $\upsilon = (t-t_0)/\delta$. In the υ time scale, (4.6) is represented by

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\upsilon} = \mathbf{g}(\mathbf{x},\mathbf{y}+\mathbf{h}(\mathbf{x},t),t,\delta) - \delta\frac{\partial \mathbf{h}(\mathbf{x},t)}{\partial t} - \delta\frac{\partial \mathbf{h}(\mathbf{x},t)}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x},\mathbf{y}+\mathbf{h}(\mathbf{x},t),t,\delta),$$
$$\mathbf{y}(0) = \boldsymbol{\zeta}(\delta) - \mathbf{h}(\boldsymbol{\xi}(\delta),t_0). \tag{4.7}$$

The variables t and $\mathbf{x}$ will be slowly varying since, in the υ time scale, they are given by

$$t = t_0 + \delta\tau, \quad \mathbf{x} = \mathbf{x}(t_0+\delta\tau,\delta).$$

Setting $\delta = 0$, we get

$$\mathbf{x} = \boldsymbol{\xi}(0) = \boldsymbol{\xi}_0, \quad t = t_0$$

and (4.7) is reduced to

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\upsilon} = \mathbf{g}(\boldsymbol{\xi}_0,\mathbf{y}+\mathbf{h}(\boldsymbol{\xi}_0,t_0),t_0,0), \quad \mathbf{y}(0) = \boldsymbol{\zeta}_0 - \mathbf{h}(\boldsymbol{\xi}_0,t_0), \tag{4.8}$$

which has an equilibrium at $\mathbf{y} = \mathbf{0}$. If this equilibrium point is asymptotically stable and $\mathbf{y}(0)$ belongs to its region of attraction, it is expected that the solution of (4.8) will reach an $\mathbf{O}(\delta)$ neighborhood of the origin during an $\mathbf{O}(\delta)$ time interval. Beyond this interval, we need to guarantee that $\mathbf{y}(\upsilon)$ will remain close to zero, while the slowly varying variables $(t,\mathbf{x})$ move away from the initial values $(t_0,\boldsymbol{\xi}_0)$. In order to guarantee this property, the following model is introduced

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\upsilon} = \mathbf{g}(\mathbf{x},\mathbf{y}+\mathbf{h}(\mathbf{x},t),t,0), \tag{4.9}$$

which is referred to as the boundary-layer model. It is derived from (4.8), where the slow variables are not frozen to the initial values anymore. The crucial stability property to guarantee is the exponential stability of the origin of (4.9). This stability property is defined below.

Definition 1 *Let D_x be a domain of $\mathbb{R}^n$, and $t_1 > 0$. The equilibrium point $\mathbf{y} = \mathbf{0}$ of the boundary-layer model* (4.9) *is exponentially stable, uniformly in*

$$(t, \mathbf{x}) \in [0, t_1] \times D_x,$$

if there exist positive constants κ, ω, and ν such that the solutions of (4.9) *satisfy*

$$\|\mathbf{y}(\upsilon)\| \leq \kappa \|\mathbf{y}(0)\| e^{-\omega\upsilon}, \quad \forall \|\mathbf{y}(0)\| < \nu, \forall (t, \mathbf{x}) \in [0, t_1] \times D_x, \forall \upsilon \geq 0.$$

The singular-perturbation theory relies on the Tikhonov theorem (see Khalil (2000)), recalled below.

Theorem 1 *Assume that the following conditions are satisfied for all*

$$(t, \mathbf{x}, \mathbf{z} - \mathbf{h}(\mathbf{x}, t), \delta) \in [0, t_1] \times D_x \times D_y \times [0, \delta_0],$$

for some domains $D_x \subset \mathbb{R}^n$, $D_y \subset \mathbb{R}^m$, in which D_x is convex and D_y contains the origin.

- *The functions $\mathbf{f}$, $\mathbf{g}$ and their first partial derivatives with respect to $(\mathbf{x}, \mathbf{z}, \delta)$ are continuous.*
  The first partial derivatives of $\mathbf{g}$ with respect to t are continuous.
  The function $\mathbf{h}(\mathbf{x}, t)$ and the Jacobian $[\partial \mathbf{g}(\mathbf{x}, \mathbf{z}, t, 0)/\partial \mathbf{z}]$ have continuous first partial derivatives with respect to their arguments.
  The initial data $\boldsymbol{\xi}(\delta)$ and $\boldsymbol{\zeta}(\delta)$ are smooth functions of δ.

- *The reduced model* (4.5) *has a unique solution $\overline{\mathbf{x}} \in S$, for $t \in [t_0, t_1]$, where S is a compact subset of D_x.*

- *The origin is an exponentially stable equilibrium point of the boundary-layer model* (4.9)*, uniformly in $(t, \mathbf{x})$.*

Let $R_y \subset D_y$ be the region of attraction of (4.8) *and Ω_y a compact subset of R_y. Then, there exists a positive constant δ^* such that for all $\boldsymbol{\zeta}_0 - \mathbf{h}(\boldsymbol{\xi}_0, t_0) \in \Omega_y$ and $0 < \delta < \delta^*$,*

the singular perturbation model (4.1)-(4.2) *has a unique solution* $\mathbf{x}(t,\delta)$, $\mathbf{z}(t,\delta)$ *on* $[t_0,t_1]$, *and*

$$\mathbf{x}(t,\delta)-\overline{\mathbf{x}}(t)=\mathbf{O}(\delta)$$
$$\mathbf{z}(t,\delta)-\mathbf{h}(\overline{\mathbf{x}}(t),t)-\hat{\mathbf{y}}(\upsilon)=\mathbf{O}(\delta)$$

hold uniformly for $t\in[t_0,t_1]$, *where* $\hat{\mathbf{y}}(\upsilon)$ *is the solution of boundary-layer model* (4.8). *Moreover, given any* $t_b>t_0$, *there is* $\delta^{**}\leqslant\delta^{*}$ *such that*

$$\mathbf{z}(t,\delta)-\mathbf{h}(\overline{\mathbf{x}}(t),t)=\mathbf{O}(\delta)$$

holds uniformly for $t\in[t_b,t_1]$ *whenever* $\delta<\delta^{**}$.

The representation of the model in singularly perturbed form may be difficult. Such a representation is often omitted and the identification of the fast state variables is based on prior knowledge of the dynamical properties of the model (see, *e.g.*, Volcke et al. (2005); Rosen et al. (2006)). Singular perturbation is then straightforward to implement, and the analysis described above is seldom presented. Our intention in this section was to recall such theoretical issues that help understand the conditions under which singular perturbation can be applied validly. In Section 5.1.3, the singular perturbation method is applied on a small part of our model. This result is extended to the complete model in Section 6.2.

4.2 Analyzing the whole by its parts

Dividing a complex problem into smaller subproblems is a natural way of reasoning (Descartes, 1637). For our model, an aggregated representation of the human colonic fermentation was depicted in Fig. 3.2 (page 80). Literature reports *in vitro* bacterial growth experiments dedicated to specific reactions, corresponding to susbsystems of our system of interest. These simpler subsystems can then be modelled, and their parameters estimated. The resulting estimates can then be used to tackle the estimation problem for the complete model. Integration of this knowledge can be carried out in various ways. We may simply use the estimated parameters of the submodels as known parameters for the complete model. Or the information of the estimates may be used to define prior distributions of the parameters in the context of Bayesian estimation. Indeed, given the scarcity of

data on the human colon, the Bayesian approach appears as promising tool for estimating the parameters of the complete model. Although we did not have time yet to fully implement it, we have included a brief description of such an approach, attempting to provide some directions for future research.

For the time being, we have performed the estimation of two kinetic submodels of the complete fermentation pathway (see Chapter 5). The estimates of these models, and values selected from literature review are used to define the parameters of the complete model. Parameters whose values are not reported in literature were selected in regard of available biological knowledge, in order to provide plausible results. Integration of such a knowledge into the parameter estimation framework is essential. Note, for example, that transport coefficients in (3.6) (page 83) must be selected to guarantee positivity of flux.

4.3 Parameter identification

Once a mathematical model has been defined to represent the behavior of a system, we must address the following questions (i) Can the parameters of the model be identified, (ii) What are their numerical values? and (iii) What is the accuracy of the estimation? These questions are analyzed below.

4.3.1 Identifiability

Before attempting numerical estimation of the model parameters, it is important to investigate their identifiability. The question of identifiability can be addressed on the sole basis of the model structure (theoretical or structural identifiability) but also taking available data into account (practical identifiability).

4.3.1.1 Structural identifiability

Let $\mathscr{M}(\cdot)$ be a mathematical model structure. Denote by $\mathscr{M}(\boldsymbol{\theta}^*) = \mathscr{M}(\widehat{\boldsymbol{\theta}})$ the fact that the model $\mathscr{M}(\cdot)$ with the parameter vector $\boldsymbol{\theta}^*$ has the same input-output behavior as the model $\mathscr{M}(\cdot)$ with the parameter vector $\widehat{\boldsymbol{\theta}}$. Theoretical identifiability addresses the question of whether this identical input-output behavior implies that the parameter vector $\boldsymbol{\theta}^*$ is equal to the parameter vector $\widehat{\boldsymbol{\theta}}$. The parameter θ_i will be structurally globally identifiable

(s.g.i.) if, for almost any $\boldsymbol{\theta}^*$,

$$\mathscr{M}(\boldsymbol{\theta}^*) = \mathscr{M}(\widehat{\boldsymbol{\theta}}) \Rightarrow \theta_i^* = \widehat{\theta}_i.$$

Moreover, the model structure $\mathscr{M}(\cdot)$ will be s.g.i. if all its parameters are s.g.i. (see, *e.g.*, Walter and Pronzato (1997)). Various methods are available for testing structural identifiability. For models that are linear in their inputs (LI), they include the Laplace transform approach and the similarity transformation approach. For nonlinear models, they include methods based on Taylor series expansion and local state isomorphism. For further details on these methods, the interested reader is referred to (Walter (1982); Walter (1987) and the references therein; Vajda et al. (1989); Vajda and Rabitz (1989); Chappell et al. (1990); Walter and Pronzato (1997); Dochain and Vanrolleghem (2001)). Denis-Vidal and Joly-Blanchard (2004) proposed a variant of the local state isomorphism approach to test the identifiability of nonlinear uncontrolled systems. This method was used in one of our studies on parameter estimation and it is discussed in Section 5.1.4.

Testing the structural identifiability of complex models turns out to be difficult. Computer algebra provides useful tools (Raksanyi et al., 1985), and softwares have become available (see, *e.g.*, Bellu et al. (2007); Sedoglavic (2001).

4.3.1.2 Practical identifiability

Contrary to structural identifiability, which is defined independently of properties of the actual values taken by the data, practical identifiability depends on the quality of the available measurements. The question to be addressed becomes: Are the available data informative enough to allow an accurate identification the model parameters? Data are always corrupted by noise and usually in short supply. Hence, even if the model is s.g.i., the parameter estimates may be correlated and the optimization problem ill-conditioned. Estimates may even take values that are physically meaningless.

To illustrate the principle of practical identifiability, consider the example described in Dochain et al. (1995). The model $y = \theta_1 x_1 + \theta_2 x_2$ is s.g.i. if y, x_1, x_2 are measured. However, it is clear that if x_1 and x_2 are almost proportional ($x_1 \simeq \alpha x_2$), the parameters θ_1, θ_2 cannot be estimated accurately. Only the combination $\theta_1\alpha + \theta_2$ can be identified. Optimal selection of experimental conditions may help to improve the quality of the esti-

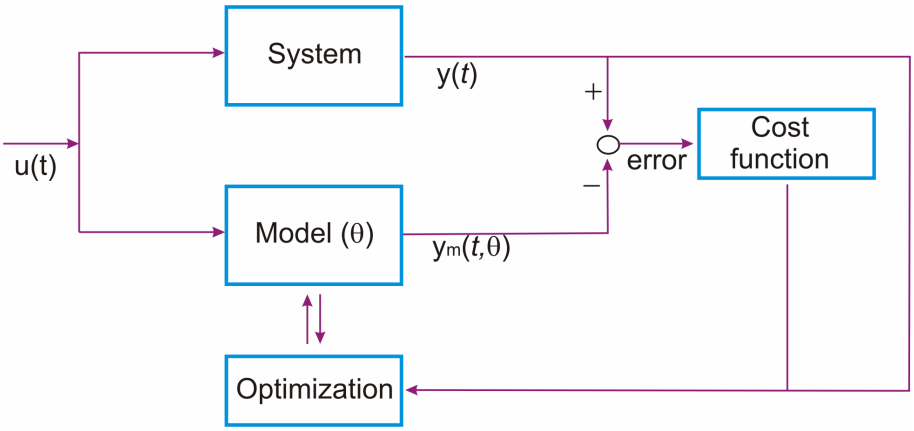

Figure 4.1: Scheme of the estimation process.

mates, when there are some degrees of freedom in the procedure for data collection. In Section 4.3.3, we return to practical identifiability and related concepts are introduced.

4.3.2 Parameter estimation

Once experimental data are available, estimation is classically translated into an optimization problem, namely that of finding a feasible parameter vector that best fits the model to the data. Figure 4.1 displays a possible scheme for this purpose.
Consider the state-space model

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}, t),\ \mathbf{x}(0) = \mathbf{x}_0(\boldsymbol{\theta}), \tag{4.10}$$

where $\mathbf{x}(t, \boldsymbol{\theta})$ is the state vector ($\mathbf{x} : \mathbb{R}^+ \times \mathbb{R}^{n_\mathrm{p}} \longrightarrow \mathbb{R}^{n_\mathrm{x}}$), $\boldsymbol{\theta}$ is the parameter vector ($\boldsymbol{\theta} \in \mathbb{R}^{n_\mathrm{p}}$), and $\mathbf{f}$ is a C^1 (continuous with continuous first-order partial derivatives) vector-valued function of the state and parameters ($\mathbf{f} : (\mathbb{R}^{n_\mathrm{x}} \times \mathbb{R}^{n_\mathrm{p}} \times \mathbb{R}^+) \longrightarrow \mathbb{R}^{n_\mathrm{x}}$).

In the special case of synchronous observations, the model output is the $\mathbb{R}^{n_\mathrm{y}}$ vector

$$\mathbf{y}_\mathrm{m}(t, \boldsymbol{\theta}) = \mathbf{h}(\mathbf{x}(t, \boldsymbol{\theta}), \boldsymbol{\theta}, t), \tag{4.11}$$

where $\mathbf{h}$ is a C^1 vector-valued function ($\mathbf{h} : (\mathbb{R}^{n_\mathrm{x}} \times \mathbb{R}^{n_\mathrm{p}} \times \mathbb{R}^+) \longrightarrow \mathbb{R}^{n_\mathrm{y}}$).

We would like to find $\boldsymbol{\theta}$ to bring the error as close to $\mathbf{0}$ as possible in the sense of some appropriately chosen cost function. The definition of such a cost function depends on the statistical hypotheses about the data collected. In the sequel, the maximum likelihood (ML) approach and the maximum *a posteriori* (MAP) estimators are introduced.

4.3.2.1 Maximum likelihood

The ML estimator is broadly used in parameter estimation. For the sake of simplicity, we shall present it here assuming that the vector of data collected at time t_i is modelled as

$$\mathbf{y}(t_i) = \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta}^*) + \boldsymbol{\varepsilon}_i,\ i = 1, ..., n_{\mathrm{t}}, \tag{4.12}$$

with n_{t} the number of observation times, $\mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta}^*)$ the output of a deterministic model and $\boldsymbol{\theta}^*$ the true value of the parameter vector. We will assume that the measurement errors $\boldsymbol{\varepsilon}_i$ $(i = 1, ..., n_{\mathrm{t}})$ are independent, homoscedastic, zero mean and Gaussian, so $\boldsymbol{\varepsilon}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

The likelihood of $\mathbf{y}$ is then defined as the probability density $\pi_y(\mathbf{y}|\theta, \boldsymbol{\Sigma})$ of the data $\mathbf{y}$ being generated by (4.12). The ML estimator maximizes $\pi_y(\mathbf{y}|\theta, \boldsymbol{\Sigma})$, or equivalently its logarithm. Under our hypotheses, the ML estimator is

$$(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Sigma}}) = \arg \min_{\boldsymbol{\theta}, \boldsymbol{\Sigma}} L(\boldsymbol{\theta}, \boldsymbol{\Sigma}), \tag{4.13}$$

where $L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = -\ln \pi_y(\mathbf{y}|\theta, \boldsymbol{\Sigma})$, given by

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{n_{\mathrm{y}} n_{\mathrm{t}}}{2} \ln 2\pi + \frac{n_{\mathrm{t}}}{2} \ln \det \boldsymbol{\Sigma} + \frac{1}{2} \sum_{i=1}^{n_{\mathrm{t}}} [\mathbf{y}(t_i) - \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{y}(t_i) - \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta})]. \tag{4.14}$$

The cost function derived from (4.14) depends on the hypotheses made on the covariance matrix (see, *e.g.*, Goodwin and Payne (1977) or Walter and Pronzato (1997)). In the sequel, some of such hypotheses are presented.

If $\boldsymbol{\Sigma}$ is known, then the ML estimator corresponds to the Gauss-Markov estimator, which minimizes

$$J_1(\boldsymbol{\theta}) = \sum_{i=1}^{n_{\mathrm{t}}} [\mathbf{y}(t_i) - \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{y}(t_i) - \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta})]. \tag{4.15}$$

If $\boldsymbol{\Sigma}$ is proportional to the identity matrix, the ML estimator for $\boldsymbol{\theta}$ is the unweighted least-squares estimator, which minimizes

$$J_2(\boldsymbol{\theta}) = \sum_{i=1}^{n_{\mathrm{t}}} [\mathbf{y}(t_i) - \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta})]^{\mathrm{T}} [\mathbf{y}(t_i) - \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta})], \tag{4.16}$$

and the ML estimate of the covariance matrix is

$$\widehat{\boldsymbol{\Sigma}} = \frac{J_2(\widehat{\boldsymbol{\theta}})}{n_{\mathrm{t}}} \mathbf{I}_{n_{\mathrm{p}}}, \tag{4.17}$$

with $\mathbf{I}_{n_\mathrm{p}}$ the $(n_\mathrm{p} \times n_\mathrm{p})$ identity matrix.

If the covariance matrix is completely unknown, the ML estimator for $\boldsymbol{\theta}$ minimizes

$$J_3(\boldsymbol{\theta}) = \ln \left[\det \sum_{i=1}^{n_\mathrm{t}} [\mathbf{y}(t_i) - \mathbf{y}_\mathrm{m}(t_i, \boldsymbol{\theta})][\mathbf{y}(t_i) - \mathbf{y}_\mathrm{m}(t_i, \boldsymbol{\theta})]^\mathrm{T} \right], \tag{4.18}$$

and the ML estimate of the covariance matrix is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n_\mathrm{t}} \sum_{i=1}^{n_\mathrm{t}} [\mathbf{y}(t_i) - \mathbf{y}_\mathrm{m}(t_i, \widehat{\boldsymbol{\theta}})][\mathbf{y}(t_i) - \mathbf{y}_\mathrm{m}(t_i, \widehat{\boldsymbol{\theta}})]^\mathrm{T}. \tag{4.19}$$

The above approaches require all components of $\mathbf{y}$ to be measured synchronously. In the case of asynchronous measurements, it can be shown that if $\boldsymbol{\Sigma}$ is unknown and diagonal, the ML estimator for $\boldsymbol{\theta}$ minimizes

$$J_4(\boldsymbol{\theta}) = \sum_{k=1}^{n_\mathrm{y}} \frac{n_{\mathrm{t}_k}}{2} \ln \left[\sum_{i=1}^{n_{\mathrm{t}_k}} [y_k(t_{i_k}) - y_{\mathrm{m}_k}(t_{i_k}, \boldsymbol{\theta})]^2 \right], \tag{4.20}$$

where t_{i_k} is the ith measurement time for y_k. The ML estimate of the covariance matrix is

$$\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}(\widehat{\sigma}_1^2, \cdots, \widehat{\sigma}_{n_\mathrm{y}}^2), \tag{4.21}$$

with

$$\widehat{\sigma}_k^2 = \frac{1}{n_{\mathrm{t}_k}} \sum_{i=1}^{n_{\mathrm{t}_k}} [y_k(t_{i_k}) - y_{\mathrm{m}_k}(t_{i_k}, \widehat{\boldsymbol{\theta}})]^2. \tag{4.22}$$

Note that (4.20)-(4.22) can also be applied for the synchronous case $(t_{i_k} = t_i,\ n_{\mathrm{t}_k} = n_\mathrm{t})$.

The estimators in (4.17), (4.19) and (4.22) are biased. However, under some technical conditions that include global identifiability, the bias tends to zero as n_t tends to infinity. The minimization of the cost function corresponding to the hypothesis on $\boldsymbol{\Sigma}$ provides $\widehat{\boldsymbol{\theta}}$.

4.3.2.2 Bayesian estimation: maximum *a posteriori*

Parameter estimation is an inverse problem that is often ill posed . Non-Bayesian estimators require at least that the number of measurements equals the number of parameters. This condition is not mandatory for Bayesian estimation, because it allows the incorporation of prior knowledge of previous studies into the problem formulation. Literature in this approach is extensive. The description presented here is mainly taken from Walter

and Pronzato (1997). More details can be found for example in Gelman et al. (2004).

In the framework of ML estimation, $\boldsymbol{\theta}$ is considered as unknown but with a single actual value. Bayesian approaches instead consider a distribution of possible values for $\boldsymbol{\theta}$. Hence, $\boldsymbol{\theta}$ is assumed to have a known prior probability density $\pi_p(\boldsymbol{\theta})$. The joint probability density of $\mathbf{y}$ and $\boldsymbol{\theta}$ satisfies

$$\pi(\mathbf{y},\boldsymbol{\theta}) = \pi_y(\mathbf{y}|\boldsymbol{\theta})\pi_p(\boldsymbol{\theta}) = \pi_p(\boldsymbol{\theta}|\mathbf{y})\pi_y(\mathbf{y}), \tag{4.23}$$

where $\pi_y(\mathbf{y})$ is the marginal distribution of the observed data, defined by

$$\pi_y(\mathbf{y}) = \int_\Theta \pi_y(\mathbf{y}|\boldsymbol{\theta})\pi_p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}. \tag{4.24}$$

The posterior probability density for $\pi_p(\boldsymbol{\theta}|\mathbf{y})$ is given by the Bayes rule

$$\pi_p(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi_y(\mathbf{y}|\boldsymbol{\theta})\pi_p(\boldsymbol{\theta})}{\pi_y(\mathbf{y})}. \tag{4.25}$$

The maximum *a posteriori* (MAP) estimator maximizes $\pi_p(\boldsymbol{\theta}|\mathbf{y})$. Since $\pi_y(\mathbf{y})$ does not depend on $\boldsymbol{\theta}$, taking logarithm on (4.25), we see that the MAP estimator maximizes

$$J_{\mathrm{MAP}}(\boldsymbol{\theta}) = \ln \pi_y(\mathbf{y}|\boldsymbol{\theta}) + \ln \pi_p(\boldsymbol{\theta}). \tag{4.26}$$

An important feature of Bayesian estimation is the need for the prior distribution $\pi_p(\boldsymbol{\theta})$, which may be considered as the most critical issue. The prior distribution can be estimated from previous measurements on the same process or on similar processes, which is the case when observations are performed on individuals of a population (see Walter and Pronzato (1997) for further details). Methods are available to determine prior distributions. Some of them are for example the *maximum entropy* approach, and the use of noninformative priors. The interested reader is referred to (Box and Tiao (1973); Robert (2001)). Bayesian theory also provides criteria for model selection (see, *e.g.*, Robert (2001); Vyshemirsky and Girolami (2008)).

We recall that parameter estimation of the complete model is hampered by data scarcity. The Bayesian approach will make it possible by incorporating information resulting from our studies on estimation of submodels (Chapter 5), and also information from existing knowledge on the human colonic ecosystem into the parameter estimation framework. Bayesian estimation of the complete model is thus a central subject that must be pursued in our future work.

4.3.2.3 Optimization

Many methods are available to the solution of optimization problems (see, *e.g.*, Fletcher (1987). Optimization itself is a major subject in mathematical research. Applications of optimization methods on modelling of biological systems are many (see, *e.g.*, Mendes and Kell (1998); Banga (2008)). They include traditional local methods such as the Nelder-Mead simplex approach, Levenberg-Marquardt algorithm, and Quasi-Newton methods (see, *e.g.*, Haag et al. (2003); Hoops et al. (2006); Schmidt and Jirstrand (2006); Checchi et al. (2007)). Metaheuristic approaches aiming to global optimization have also been developed to circumvent difficulties of local methods (*i.e.*, sensitivity to initial guess of the parameter vector). In our work, we did not tackle global optimization methods. We used the Quasi-Newton algorithm implemented in Matlab (Mathworks, 2008), which is based on the method developed by Broyden, Fletcher, Goldfarb and Shanno (BFGS). The BFGS method is considered as the most effective of the Quasi-Newton methods. Vanrolleghem and Keesman (1996) evaluated the performance of a miscellaneous group of optimization methods on parameter estimation for Monod-based models. The BFGS method was the most efficient algorithm, but also the most sensitive to local minima. The BFGS method is summarized below.

Let J a cost function of $\boldsymbol{\theta}$ to be optimized and $\mathbf{H}$ its Hessian. Starting from a initial point $\widehat{\boldsymbol{\theta}}^k$, the general iterative equation of the Quasi-Newton method is

$$\widehat{\boldsymbol{\theta}}^{k+1} = \widehat{\boldsymbol{\theta}}^k - \beta^k \mathbf{B}^k \mathbf{g}^k, \tag{4.27}$$

with β^k is the step length, to be optimized by line search, $\mathbf{g}^k$ the gradient of J at $\widehat{\boldsymbol{\theta}}^k$, and $\mathbf{B}^k$ a symmetric positive definite matrix that approximates the inverse of the Hessian $(\mathbf{H}^k)^{-1}$. The BFGS method updates the matrix $\mathbf{B}$ by

$$\mathbf{B}^{k+1} = \mathbf{B}^k + \left(1 + \frac{(\Delta\mathbf{g}^k)^{\mathrm{T}}\mathbf{B}^k\Delta\mathbf{g}^k}{(\Delta\boldsymbol{\theta}^k)^{\mathrm{T}}\Delta\mathbf{g}^k}\right)\frac{\Delta\boldsymbol{\theta}^k(\Delta\boldsymbol{\theta}^k)^{\mathrm{T}}}{(\Delta\boldsymbol{\theta}^k)^{\mathrm{T}}\Delta\mathbf{g}^k} - \frac{\Delta\boldsymbol{\theta}^k(\Delta\mathbf{g}^k)^{\mathrm{T}}\mathbf{B}^k + \mathbf{B}^k\Delta\mathbf{g}^k(\Delta\boldsymbol{\theta}^k)^{\mathrm{T}}}{(\Delta\boldsymbol{\theta}^k)^{\mathrm{T}}\Delta\mathbf{g}^k}, \tag{4.28}$$

where $\Delta\boldsymbol{\theta}^k = \widehat{\boldsymbol{\theta}}^{k+1} - \widehat{\boldsymbol{\theta}}^k$ and $\Delta\mathbf{g}^k = \mathbf{g}^{k+1} - \mathbf{g}^k$.

4.3.3 Parameter uncertainty

The parameters of knowledge-based models are usually physically meaningful. They need to be estimated from experimental data, with confidence intervals associated to the

estimates. Once $\widehat{\boldsymbol{\theta}}$ has been found, one must ask about the accuracy of the estimation. The uncertainty attached to the estimation can be assessed by different approaches, including Monte-Carlo methods and methods based on the density of the estimator. This latter group comprises uncertainty evaluation via the computation of the Fisher information matrix (FIM). In this work, the FIM was selected as our approach of uncertainty assessment because of its simplicity of implementation. The FIM satisfies

$$\mathbf{F}(\boldsymbol{\theta}^*) = -\mathrm{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\ln \pi_y(\mathbf{y}|\theta,\boldsymbol{\Sigma})\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}. \tag{4.29}$$

In the general case, when $\boldsymbol{\Sigma}$ is unknown, the (j,k)th element of the FIM satisfies

$$\mathbf{F}(\boldsymbol{\theta}^*)_{j,k} = \sum_{i=1}^{n_\mathrm{t}} \left[\frac{\partial \mathbf{y}_\mathrm{m}}{\partial\theta_j}\right]^{\mathrm{T}}_{(t_i,\boldsymbol{\theta}^*)} \boldsymbol{\Sigma}^{-1} \left[\frac{\partial \mathbf{y}_\mathrm{m}}{\partial\theta_k}\right]_{(t_i,\boldsymbol{\theta}^*)} + \frac{1}{2}\sum_{i=1}^{n_\mathrm{t}} \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\left[\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right]_{(t_i,\boldsymbol{\theta}^*)} \boldsymbol{\Sigma}^{-1}\left[\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\right]_{(t_i,\boldsymbol{\theta}^*)}\right), \tag{4.30}$$

where $\boldsymbol{\theta}$ is the extended parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_\mathrm{m} \\ \boldsymbol{\theta}_\sigma \end{bmatrix},$$

with $\boldsymbol{\theta}_\mathrm{m}$ the vector of the parameters of the deterministic model and $\boldsymbol{\theta}_\sigma$ the vector of the parameters of the covariance matrix $\boldsymbol{\Sigma}$.

If the covariance matrix is known, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ and the FIM can be written as

$$\mathbf{F}(\boldsymbol{\theta}^*) = \sum_{i=1}^{n_\mathrm{t}} \left[\frac{\partial \mathbf{y}_\mathrm{m}}{\partial\boldsymbol{\theta}}\right]^{\mathrm{T}}_{(t_i,\boldsymbol{\theta}^*)} \boldsymbol{\Sigma}_0^{-1} \left[\frac{\partial \mathbf{y}_\mathrm{m}}{\partial\boldsymbol{\theta}}\right]_{(t_i,\boldsymbol{\theta}^*)}. \tag{4.31}$$

In practice, when $\boldsymbol{\Sigma}$ is unknown, instead of using (4.30) to calculate the FIM, a widely used approach is to approximate $\boldsymbol{\Sigma}$ by its ML estimate $\widehat{\boldsymbol{\Sigma}}$ and to use (4.31) with $\boldsymbol{\Sigma}_0 = \widehat{\boldsymbol{\Sigma}}$.

Under a number of technical assumptions that include theoretical identifiability, the covariance matrix $\mathbf{P}$ of the ML estimator satisfies

$$\mathbf{P} \geq \mathbf{F}^{-1}(\boldsymbol{\theta}^*). \tag{4.32}$$

The inequality (4.32) is known as the Cramér-Rao inequality. Often, the covariance matrix for $\boldsymbol{\theta}$ is approximated by the inverse of the FIM computed at $\widehat{\boldsymbol{\theta}}$ as

$$\widehat{\mathbf{P}} = \mathbf{F}^{-1}(\widehat{\boldsymbol{\theta}}). \tag{4.33}$$

However it must be kept in mind that, except in very special cases, approximating the covariance of the parameter estimates by the inverse of the FIM is only valid asymptotically, when the number of data points tends to infinity, the statistical hypotheses on the noise are satisfied, and the estimate $\widehat{\boldsymbol{\theta}}$ is close to the true parameter vector $\boldsymbol{\theta}^*$. When these idealized conditions are far from being satisfied, this evaluation of the uncertainty on the estimates has to be considered with care.

The covariance matrix $\widehat{\mathbf{P}}$ can be used to compute approximate confidence intervals of the estimated parameters since the square root η_j of the jth diagonal element of $\widehat{\mathbf{P}}$ is an estimate of the standard deviation of $\widehat{\theta}_j$. The confidence interval calculation depends on whether $\boldsymbol{\Sigma}$ is known or not (see, *e.g.*, Walter and Pronzato (1997)). In the case when $\boldsymbol{\Sigma}$ is known, an approximate 95% confidence interval for θ_j is

$$[\widehat{\theta}_j - 2\eta_j, \widehat{\theta}_j + 2\eta_j]. \tag{4.34}$$

The expression above is derived from the properties of the ML estimator. Instead of using the FIM, the Hessian of the cost function can also be used to compute confidence intervals. This approach consider the curvature of the expected surface (also called solution locus). In non-LP models, the curvature may be significant and confidence intervals calculated from the FIM and the Hessian may differ (see, *e.g.*, Marsili-Libelli et al. (2003)). Discussion about the effect of the curvature on the computation of confidence intervals can be found in Seber and Wild (1989).

To evaluate (4.31), the first-order sensitivity of the output with respect to the parameters must be computed. Let $\mathbf{s}_j$ denote the first-order sensitivity $\frac{\partial \mathbf{x}}{\partial \theta_j}$ of the state w.r.t. the parameter θ_j. It is the solution of

$$\dot{\mathbf{s}}_j = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_{(\mathbf{x},\boldsymbol{\theta},t)} \mathbf{s}_j + \left[\frac{\partial \mathbf{f}}{\partial \theta_j}\right]_{(\mathbf{x},\boldsymbol{\theta},t)}, \tag{4.35}$$

where $\mathbf{x}$ is computed via (4.10) and the initial conditions are

$$\mathbf{s}_j(0) = \frac{\partial \mathbf{x}(0)}{\partial \theta_j},\ j = 1, \ldots, n_\mathrm{p}. \tag{4.36}$$

When the initial conditions on the state are supposed to be known $\mathbf{s}_j(0) = \mathbf{0}$. According to (4.11), the sensitivity of the output w.r.t. the parameter θ_j at t_i is then evaluated as

$$\left[\frac{\partial \mathbf{y}_\mathrm{m}}{\partial \theta_j}\right]_{t_i} = \left[\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right]_{(\mathbf{x},\boldsymbol{\theta},t_i)} \mathbf{s}_j + \left[\frac{\partial \mathbf{h}}{\partial \theta_j}\right]_{(\mathbf{x},\boldsymbol{\theta},t_i)}. \tag{4.37}$$

Equation (4.35) is also called *forward sensitivity equation*. The evaluation of the sensitivities can be handled in various ways. They include finite-difference approximation, analytic differentiation, automatic differentiation (Schittkowski (2002)) and the complex-step approximation method (Squire and Trapp (1998)). The most popular method is certainly finite-difference approximation. It is easy to implement, but its accuracy is subordinated to an adequate selection of the step sizes on the parameters. For complex models, handmade analytic differentiation requires complicated, error-prone and time-consuming manipulations. An interesting option to overcome these difficulties is symbolic manipulation.

The condition number of the FIM for the spectral norm, *i.e.*, the ratio of the largest singular value of the FIM to the smallest, provides information on the numerical conditioning of the optimization problem and is a useful indicator of the practical identifiability of the model given the available data. The higher the condition number, the more difficult the optimization is and the lower practical identifiability. Since the FIM is symmetric, positive definite, this amounts to computing the ratio of its largest eigenvalue to its smallest. Moreover, the matrix $\widehat{\mathbf{P}}$ is used to compute an estimate **Cor** of the parameter correlation matrix, with the element $\mathbf{Cor}(i,j)$ an approximate correlation coefficient between the ith and jth estimated parameters given by

$$\mathbf{Cor}(i,j) = \frac{\widehat{\mathbf{P}}(i,j)}{[\widehat{\mathbf{P}}(i,i) * \widehat{\mathbf{P}}(j,j)]^{1/2}}. \tag{4.38}$$

In addition to the applications described above, the FIM can be exploited to look for suitable experimental conditions providing high informative content. This problem is tackled by optimal experiment design. Various methods are devoted to such a problem (see, *e.g.* Seber and Wild (1989); Walter and Pronzato (1990); Walter and Pronzato (1997); Dochain and Vanrolleghem (2001); Franceschini and Macchietto (2008)). Here, we do not enter into the details associated to this problem. However, we may intuitively think that a good experiment will be one that makes the FIM well-conditioned. We can also seek to minimize the determinant of $\widehat{\mathbf{P}}$ (or maximizes the determinant of the FIM). As we mentioned above, optimal experiment design is only possible when there are some degrees of freedom in the procedure for data collection.

4.4 Local sensitivity analysis

Sensitivity equations are not only instrumental in uncertainty assessment. Other applications include for example sensitivity analysis, which aims to study how the outputs of a model are affected by different sources of variation (Saltelli et al., 2000). Sensitivity analysis is an important tool to identify influential parameters and variables in the model. This information is, indeed, very useful. We could realize that some parameters have little influence on model behavior and thus decide to discard them to simplify the model structure. In knowledge-based models, sensitivity analysis is central to identifying the phenomena that play a major role in system behavior. Theoretical background and methods for sensitivity analysis can be found in Saltelli et al. (2000). Here, we are particularly interested in local sensitivity analysis, which is based in the computation of (4.37) at $\widehat{\boldsymbol{\theta}}$. Let $\overline{s}_j^k$ be the normalized first-order sensitivity of the model output y_{m_k} w.r.t. the parameter θ_j, defined as (Turányi and Rabitz, 2000)

$$\overline{s}_j^k(t_i,\widehat{\boldsymbol{\theta}}) = \frac{\widehat{\theta}_j}{y_{\mathrm{m}_k}(t_i,\widehat{\boldsymbol{\theta}})}\left[\frac{\partial y_{\mathrm{m}_k}}{\partial \theta_j}\right]_{(t_i,\widehat{\boldsymbol{\theta}})}. \tag{4.39}$$

The normalized sensitivity is dimensionless, allowing for comparison of the effect of various parameters on the model output. We can construct the matrix $\mathbf{S}_y$ where the element (k,j) is the L_1 norm of $\overline{s}_j^k$ along t

$$\mathbf{S}_y(k,j) = \|\overline{s}_j^k\|_1 = \sum_{i=1}^{n_\mathrm{t}} \left|\overline{s}_j^k(t_i,\widehat{\boldsymbol{\theta}})\right|. \tag{4.40}$$

When the model has high dimension both in its state and parameter vectors, some methods have to be used for summarizing the information in $\mathbf{S}_y$ (Saltelli et al., 2000). The matrix $\mathbf{S}_y$ has very interesting applications. We can represent it graphically. In our study, we used the function *imagesc* implemented in Matlab to represent grafically the matrix $\mathbf{S}_y$. Such a function scales matrix data to the full range of a given colormap and displays the associated image. This gives us the opportunity to visualize the effect of the parameters on the model outputs. Moreover if we sum the components in each column, we will obtain a quantification of the overall sensitivity for each parameter. Then, a parameter ranking can be performed that enables one to identify the more sensitive parameters of the model. This information can be used to reduce the number of parameters in the model or in the parameter estimation framework. For example, we could conduct the optimization by targeting the most sensitive parameters (Abusam et al. (2001); Lei and Jørgensen (2001);

Degenring et al. (2004); Juillet et al. (2006)).

In the following, the development of a toolbox for parameter estimation with sensitivity analysis capabilities is described. This software integrates many theoretical aspects that have been underlined above.

4.5 IDEAS: a Matlab toolbox for parameter estimation

The main part of this work was published in Preprints of 15th Symposium on System Identification, Saint-Malo, France. (Muñoz-Tamayo et al., 2009).

In the expanding field of systems biology a number of software packages devoted to modelling, simulation, parameter estimation and sensitivity analysis have been developed. Software development is a critical component of research in systems biology, which is characterized by a multidisciplinary work that assembles technology, computational modelling, biological foundations and the analysis of dynamic systems (Kitano, 2002b). Given the disparity of modelling environment between research groups, an important international effort was undertaken to establish a common software environment. The result of such a effort is the creation of the Systems Biology Markup Language (SBML; http://www.sbml.org/), which is particularly applied to represent biochemical reaction networks (Hucka et al., 2003). SBML has become accepted as a standard format for representing formal, quantitative and qualitative models at the level of biochemical reactions and regulatory networks (Hucka et al., 2004). However, to our knowledge, the use of SBML has not been pervasive in the field of anaerobic modelling where many model implementations have been performed in the software package Matlab (see, *e.g.*, the work of Rosen et al. (2006) and the recent work of Rodríguez et al. (2009)). However, SBML applications to anaerobic modelling may be object of further discussion and some tools are available to facilitate interfacing between SBML and Matlab environments (see, *e.g.*, Keating et al. (2006)).

Among the numerous softwares devoted to systems biology, COPASI (Hoops et al., 2006) and SBML-PET (Zi and Klipp, 2006) allow for parameter estimation (among other fea-

tures). SBML-SAT (Zi et al., 2008) is devoted to sensitivity analysis. Other softwares have been built as toolboxes to be run in the Matlab environment, which allows the use of functions already implemented in Matlab. They include the recently launched commercial software SimBiology® and the freely available toolboxes PEAS (Checchi et al., 2007), System Biology (SBtoolbox) (Schmidt and Jirstrand, 2006) and PottersWheel (Maiwald and Timmer, 2008). All of the parameter-estimation softwares mentioned above use a (possibly weighted) quadratic cost function as their measure of optimality. PEAS and PottersWheel quantify the accuracy of the estimated parameters through a *finite-difference* evaluation of the FIM or an approximation of this matrix at the estimates. They also provide a Monte-Carlo analysis particularly adapted for models with non-differentiable dynamics. Even though SBtoolbox and SimBiology estimate parameters, they do not assess their uncertainty. Both toolboxes provide routines for sensitivity analysis, which can be used to assess the accuracy of the estimates, but this not so trivial calculation is left to the user. For sensitivity analysis, SimBiology uses the complex-step approximation method (Squire and Trapp, 1998). SBtoolbox only provides sensitivity analysis at steady state, but an extension to non-steady state trajectories was developed (Schmidt, 2007) to accomplish this task through an interface with the free software CVODES, using the forward or adjoint method (Hindmarsh and R.Serban, 2006). The current version SBToolbox2 allows forward sensitivity analysis.

To tackle the parameter estimation of submodels describing bacterial growth experiments dedicated to specific reactions of the trophic chain, we decided to develop the Matlab toolbox IDEAS (*IDEntification and Analysis of Sensitivity*). It is devoted to knowledge-based ODE models. Therefore, its application is not restricted to biological models.

4.5.1 Why develop IDEAS?

Our interest in developing the toolbox IDEAS (*IDEntification and Analysis of Sensitivity*) was to provide a convenient and easy-to-use tool to tackle parameter estimation for knowledge-based models from available data with carefully designed sensitivity analysis capabilities. The estimation is carried out via the ML approach, and can easily be extended to MAP estimation. As some of the above mentioned toolboxes, IDEAS takes advantage of the functions implemented in Matlab for the optimization and simulation of ODE models. The main originality of IDEAS compared to the previous toolboxes resides

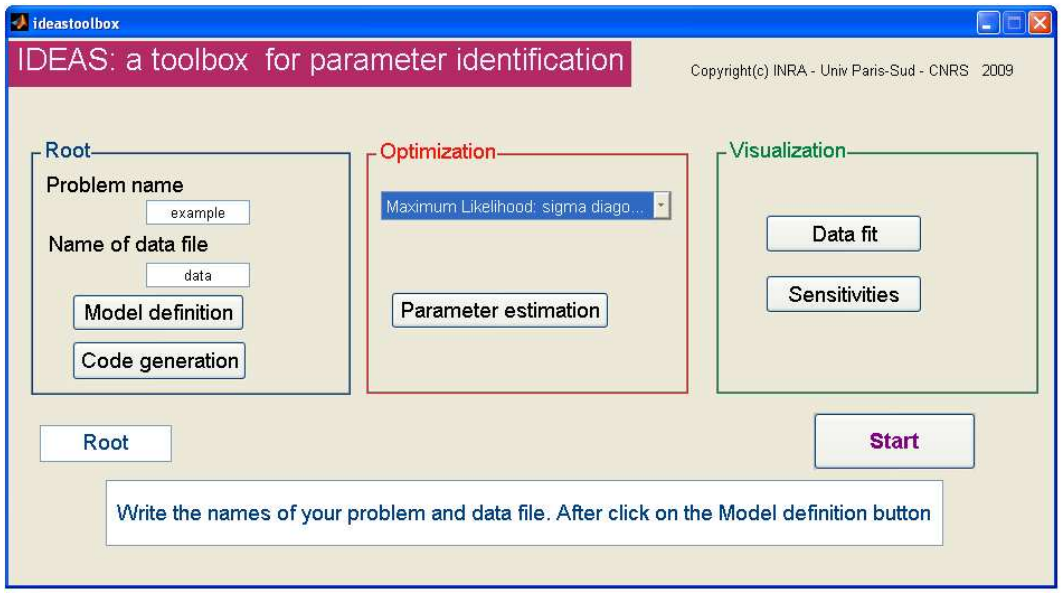

Figure 4.2: IDEAS Interface.

in the analytic computation of the sensitivity functions that are used for the evaluation of the FIM. IDEAS performs an exact and automatically generated formal evaluation of the right-hand sides of (4.35) and (4.37), using the Symbolic Toolbox of Matlab to differentiate $\mathbf{f}$ and $\mathbf{h}$ with respect to the state and parameters. There is thus no need for the user to provide supplementary routines to evaluate the right-hand sides of (4.35) and (4.37). Then, the toolbox integrates the augmented model (4.10), (4.35) at $\widehat{\boldsymbol{\theta}}$ to solve (4.37). The solutions for the sensitivity functions are used to compute the FIM according to (4.31) with $\boldsymbol{\Sigma}_0 = \widehat{\boldsymbol{\Sigma}}$, and thus assess parameter uncertainty. IDEAS also provides the parameter correlation matrix and condition number of the FIM for the spectral norm, which are useful indicators of the practical identifiability of the model given the available data. A graphical representation of the local sensitivity analysis is also provided, similar in spirit to that in Zi et al. (2008) (see Section 4.4). Another original feature of IDEAS is the opportunity offered to use four different optimization criteria suggested by the ML approach (see Section 4.3). The minimization of any of these cost functions is performed using the BFGS method, as implemented in Matlab for unconstrained optimization (Mathworks, 2008).

4.5.2 How to use IDEAS?

IDEAS is freely available at http://www.inra.fr/miaj/public/logiciels/ideas/index.html for academic usage. A user's guide is also available at the same address. The toolbox runs in Matlab v7.0 or later versions. It requires the optimization and symbolic toolbox to be executed.

The software interface not only facilitates the use of the software itself, but also plays an important role as a learning tool in the topic of identification (Ljung, 2003). Our aim was to provide a friendly interface with open source code available to the user. IDEAS is operated through the graphical interface displays in Fig. 4.2 and dialog boxes in the command window.

To illustrate its main features, consider the simplistic model

$$\dot{x} = \mu \frac{x}{K+x} - k_d x, \; x(0) = 5, \tag{4.41}$$

$$y_{\mathrm{m}} = x, \tag{4.42}$$

with μ, K and k_d to be estimated.

The user provides the data in a .txt file. Afterwards, the names of the parameters and the state variables are requested in the command window (Fig. 4.3). The user defines the mathematical model following the instructions of a template automatically generated (Fig. 4.4a). From the input files, the software automatically generates the routines required for estimation and visualization, using symbolic differentiation to generate the sensitivity computation routine from the equations provided by the user. Fig. 4.4b shows the code generated to evaluate the augmented model. The terms d$Fdpv$ and d$Fdxv$ are the automatically generated derivatives of the vector $\mathbf{f}$ in (4.10) with respect to the parameters and the state, respectively.

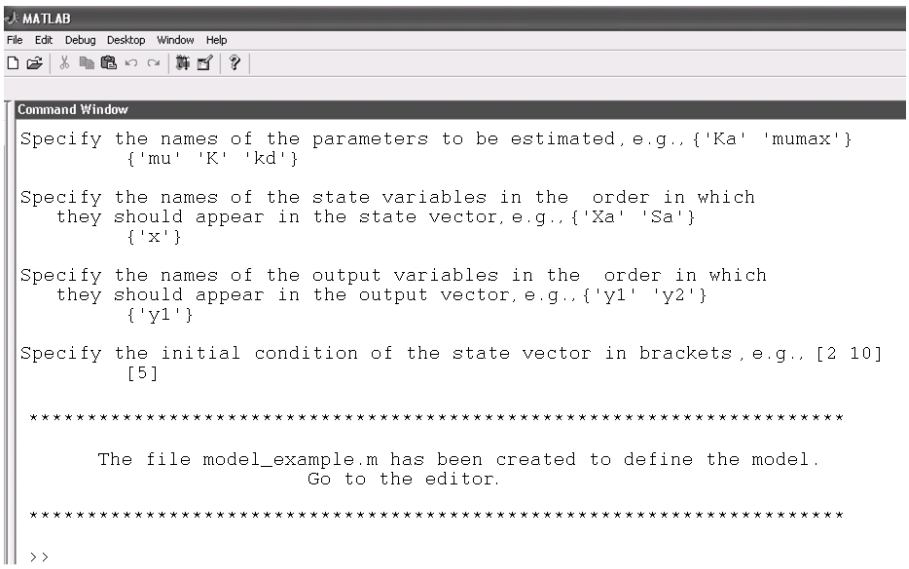

Figure 4.3: Naming the parameters, state and output variables of the model.

(a)

```
...
% Write the elements of the vector F, e.g., F(1)= Kh*x
F(1) = mu *x /(K + x) -kd*x;
% Write the elements of the vector H, e.g., H(1) = x*s
H(1) = x;
% save the file and return to the interface
```

(b)

```
function [Fv] = examplese(t,X)
global param
...
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% State variables
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
x = X(1);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% State equations
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        Fv(1,:)=mu*x/(K+x)-kd*x;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Derivative of F with respect to the parameters
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
       dFdpv(1,1)=x/(K+x);
       dFdpv(1,2)=-mu*x/(K+x)^2;
       dFdpv(1,3)=-x;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Derivative of F with respect to the state variables
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
       dFdxv(1,1)=mu/(K+x)-mu*x/(K+x)^2-kd;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Sensitivity equations
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
j =Nx+1;
  for k=1:Np;
Fv(j:j+Nx-1,:)=dFdxv*X(j:j+Nx-1)+dFdpv(:,k);
         j=j+Nx;
   end
```

Figure 4.4: Operation of IDEAS: (a) Defining the state equation; (b) Automatically generated code to evaluate the augmented model (4.10), (4.35).

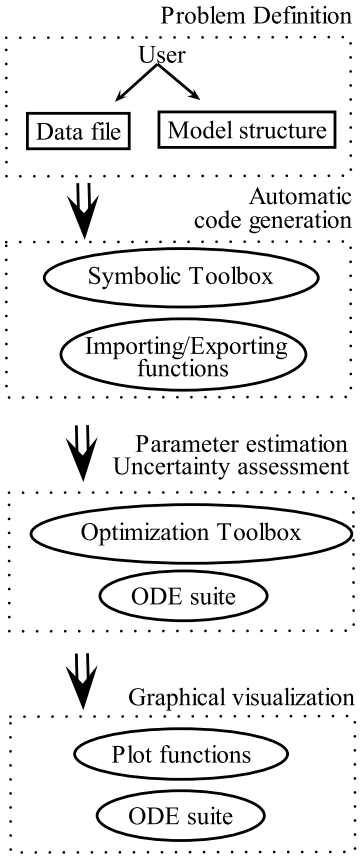

Figure 4.5: Steps followed in a typical execution of IDEAS.

The Optimization Toolbox and ODE solver are used to find the estimates and compute the FIM. Finally, at the visualization step, plots of the data fit and the sensitivity trajectories are displayed. Figure 4.5 shows the integration of these steps in a typical execution. All the functions generated are accessible and can be used in other user-defined routines, and modified if needed.

4.5.3 Future extensions

IDEAS is still at an early stage of its development. Some extensions are planned in the near future in order to enhance its capabilities. The additional features in the next version will include the possibility to consider the initial conditions as parameters to be estimated. This issue is a serious matter, especially for biomass concentrations, which are technically difficult to measure for mixed culture systems. The next version will also take

into account missing data in the output vector components at some measurement times, or asynchronous measurements. This situation is frequently encountered in biological experiments.

Concerning the statistical framework used in IDEAS, the main extensions will be to allow for heteroscedastic errors and for Bayesian estimation. Special interest will be devoted to the maximum a posteriori estimator, which enables one to deal with situations where few experimental data are available. Another extension will be the computation of the exact FIM given by (4.29) for each estimator.

At the optimization stage, a multi-start routine will be included in order to decrease the sensitivity to the user-provided initial guess on the parameters, and stand-alone optimization and automatic differentiation routines will be implemented to make IDEAS usable without the Optimization Toolbox. Another additional feature will be the symbolic computation of the gradient of the cost function, which should speed-up optimization. This computation will be based on the sensitivity functions that are already being evaluated by IDEAS.

The finite-difference approximation method as well as direct integration of the augmented model can become numerically intensive for large-scale models. Under this condition, a very attractive alternative is the use of adjoint methods (Valko and Vajda (1984); Cao et al. (2003)).

4.6 Conclusions

In this chapter, we presented the strategy that we pursued in our modelling task. The nature of our model implied the interpretation of biological knowledge to provide a mechanistic representation of the human colon. Given the complexity of the the system, we decided to aggregate certain phenomena to reduce the level of such a complexity.

We described technical background on the mathematical tools employed in this work. They included the singular perturbation approach and the parameter estimation frame-

work. For the singular perturbation method, we emphasized the conditions under which it can be applied validly. For parameter estimation, we recalled the central concept of identifiability, and the maximum likelihood and Bayesian estimation approaches. Bayesian estimation appears to be the most promising approach for the estimation of the complete model. Finally, we presented the development of the parameter estimation toolbox IDEAS. The strategy presented here is a generic approach that could be applied to many other types of modelling problems where the experimental data are scarce.

Examples of applications of the above methods are presented in the following two chapters. The next chapter presents our results on the modelling and identification of two small parts of the full fermentation pathway performed by human colonic bacteria.

Chapter 5

Kinetic modelling of *in vitro* bacterial growth experiments

Kinetics is nature's way of preventing everything from happening all at once.
Steven E. LeBlanc

In this chapter, we present our studies on parameter estimation of mathematical models describing two kinetic processes, namely homoacetogenesis and butyrate production by human colonic bacteria. According to our strategy (see Section 4.2), the parameters obtained for the kinetics studied here are used as known parameters for the complete model.

5.1 Homoacetogenesis by the human colonic bacteria *Ruminococcus hydrogenotrophicus*

The main part of this work was published in Proc. 16th IEEE Mediterranean Conference on Control and Automation. Ajaccio, France (Muñoz-Tamayo et al., 2008).

Homoacetogenesis is one of the three routes for hydrogen utilization by the human colonic microbiota (see Section 2.3.1.4). Figure 5.1 shows the location of homoacetogenesis in the

whole process of carbohydrate degradation. Homoacetogenesis is an autotrophic process, in which colonic microorganisms reduce two moles of CO_2 to one mole of acetate. It has been estimated that acetate from the reductive pathway represents about 40% of the total acetate produced in the human (Bernalier et al., 1999). This conversion process is central for transfer of hydrogen between microorganisms, which is required for fermentation (Thauer et al., 1977). Excess of H_2 production has been related to irritable bowel syndrome. It causes abdominal pain, distention, flatulence and borborygmus (Boever et al., 2000). Moreover, H_2 utilization by homoacetogens, sulfate reducers and methanogens (Archaea) provides the final step in energy extraction from carbohydrate degradation and affects the thermodynamics of the whole biochemical process (Bäckhed et al. (2005); Ley et al. (2006a)). H_2 levels affect the NAD/NADH redox couple (Costello et al., 1991) and thus the thermodynamic feasibility of some of the fermentation reactions (Madigan et al., 2009).

We present now a mathematical model that we developed to represent the *in vitro* experiments described in Bernalier et al. (1996b) to study the kinetics of homoacetogenesis by the human colonic bacteria *Ruminococcus hydrogenotrophicus*.

5.1.1 Experimental setup

The kinetic study was carried out in culture tubes connected to pressure sensors (Fig. 5.2). The bacteria were cultivated in the medium AC-21 modified using H_2/CO_2 (60:40, v/v) at 202 kPa. Temperature was controlled at 37°C. The substrate was a gas mixture H_2/CO_2 that diffused to the liquid phase to be used as energy source for the bacteria. Acetate was produced via the reductive pathway. The variables measured in the experiment were the concentration of H_2 in the gas phase, estimated from manometric measurements and gas chromatography, the concentration of acetate, measured enzymatically, and the optical density of cultures at 600 nm (OD_{600}).

5.1.2 Mathematical model

Two main phenomena occur, namely biological reaction and mass transfer between the gas and liquid phases. The decay of bacteria is described by a first-order equation. Bacterial growth was quantified through optical density measurements, which relate light

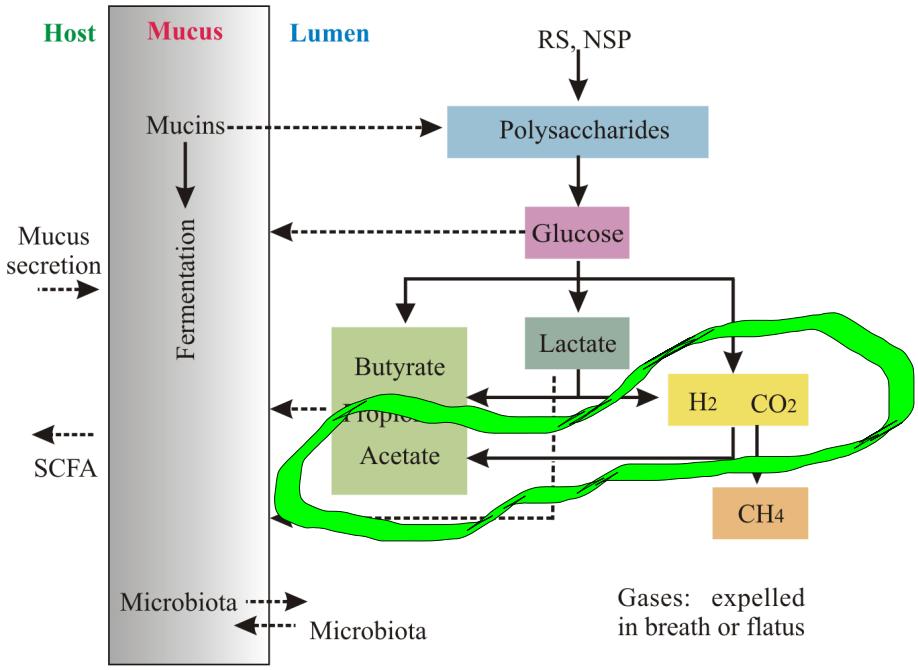

Figure 5.1: Location of the homoacetogenesis in the full process of carbohydrate degradation.

dispersion with cell concentration. Unfortunately, these measurements are imprecise because the bacteria that become biologically inactive keep their structure, so they will still be counted although they no longer participate to the reaction. For these reasons, we explicitly model the inactive bacteria concentration and assume a first-order dynamic for this variable. This makes it possible to relate optical density as given by the model with the experimental data. For liquid-gas transfer, a modification of the Henry law is used (Stumm and Morgan, 1996).

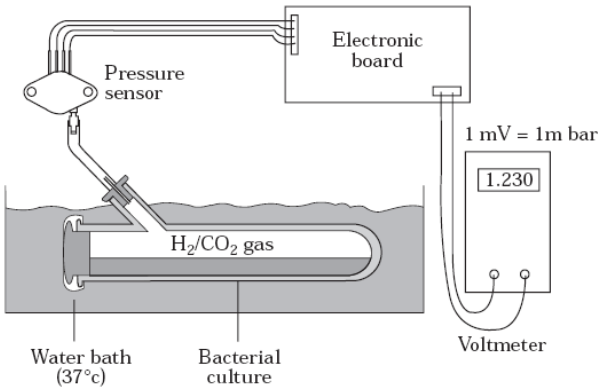

Figure 5.2: *In vitro* system to study homoacetogenesis reaction (taken from Leclerc et al. (1997)).

The equations of the model are

$$\dot{x}_{\mathrm{H_2a}} = \mu_{\mathrm{H_2a}} - k_d x_{\mathrm{H_2a}}, \qquad x_{\mathrm{H_2a}}(0) = 0.0193, \tag{5.1}$$

$$\dot{\chi} = k_d x_{\mathrm{H_2a}} - k_i \chi, \qquad \chi(0) = 0, \tag{5.2}$$

$$\dot{s}_{\mathrm{g,H_2}} = k_L a(s_{\mathrm{H_2}} - K_{H,\mathrm{H_2}} RT s_{\mathrm{g,H_2}}) \frac{V}{V_{\mathrm{g}}}, \qquad s_{\mathrm{g,H_2}}(0) = 0.677, \tag{5.3}$$

$$\dot{s}_{\mathrm{ac}} = \frac{1 - Y_{\mathrm{H_2a}}}{Y_{\mathrm{H_2a}}} \mu_{\mathrm{H_2a}}, \qquad s_{\mathrm{ac}}(0) = 0.218, \tag{5.4}$$

$$\dot{s}_{\mathrm{H_2}} = -\frac{1}{Y_{\mathrm{H_2a}}} \mu_{\mathrm{H_2a}} - k_L a(s_{\mathrm{H_2}} - K_{H,\mathrm{H_2}} RT s_{\mathrm{g,H_2}}), \qquad s_{\mathrm{H_2}}(0), \text{ unknown}, \tag{5.5}$$

with V the volume in liquid phase, V_{g} the volume in gas phase, and $\mu_{\mathrm{H_2a}}$ the bacterial growth kinetics, expressed by the Monod law (Monod, 1949)

$$\mu_{\mathrm{H_2a}} = \mu_{\max,\mathrm{H_2a}} \frac{s_{\mathrm{H_2}} x_{\mathrm{H_2a}}}{K_{s,\mathrm{H_2a}} + s_{\mathrm{H_2}}}. \tag{5.6}$$

The five state variables are the concentration $x_{\mathrm{H_2a}}$ of active biomass, the concentration χ of inactive bacteria that keep the cell structure, the concentration $s_{\mathrm{g,H_2}}$ of hydrogen in the gas phase, the concentration $s_{\mathrm{H_2}}$ of hydrogen in the liquid phase, and the concentration s_{ac} of acetate. All of them expressed in kilograms of chemical oxygen demand (COD) per cubic meter (kg COD m^{-3}). Initial condition for $s_{\mathrm{H_2}}$ is unknown.
The output vector is

$$\mathbf{y}_{\mathrm{m}} = (\alpha(x_{\mathrm{H_2a}} + \chi),\ s_{\mathrm{g,H_2}},\ s_{\mathrm{ac}})^{\mathrm{T}}, \tag{5.7}$$

where α is a conversion factor between optical density and kg COD m^{-3}, which depends on the physical properties of the bacteria. We looked at the literature and 1 unit of optical density may be equal from $1*10^7$ cells/ml to $1*10^{10}$ cells/ml. Assuming an average weight of a cell of $4.2*10^5$ kg and considering a conversion factor of 1.4 kg COD/kg for the biomass, the factor α ranges from 0.1 OD/(kg COD m^{-3}) to 40 OD/(kg COD m^{-3}). Including this factor in the estimation lead to practical identifiability problems, thus we decided to set it to a value that provided acceptable results for the other kinetic parameters. α was set to 5.95 OD/(kg COD m^{-3}).

Parameter k_d (h^{-1}) is the decay factor for the biomass, $\mu_{\max,H_2a}$ (h^{-1}) is the maximum growth rate, K_{s,H_2a} (kg COD m^{-3}) is the saturation constant of the Monod law, k_i (h^{-1}) is a lysis constant. Y_{H_2a} is the yield factor that gives the biomass produced by hydrogen consumed, k_La (h^{-1}) is the liquid-gas transfer coefficient multiplied by the specific transfer area, K_{H,H_2} is the Henry law coefficient for hydrogen, in molar units per bar (M bar^{-1}), R is the ideal gas law constant and T is the absolute temperature. The numerical values of the physical parameters involved are:

$k_La = 8.33\ h^{-1}$, $K_{H,H_2} = 7.28\ 10^{-4}$ M bar^{-1},

$R = 8.31\ 10^{-2}$ bar m^3 $kmol^{-1}$ K^{-1},

$T = 310.15$ K, $V_g = 22.57$ ml, and $V = 5.3$ ml.

The five unknown parameters to be identified are $\mu_{\max,H_2a}$, K_{s,H_2a}, k_d, k_i, and Y_{H_2a}.

5.1.3 Reducing the number of state variables

The dynamics of hydrogen concentration in the liquid phase appears to be fast, leading to a multi time scale model, as proposed by Rosen et al. (2006). Thus, we can apply the singular perturbation method to perform model order reduction (See Section 4.1.3.1). We first observe that the dynamics of the active and inactive biomass and the liquid and gaseous hydrogen concentrations are decoupled from the dynamic of the acetate concentration. This last concentration is obtained by integration of an expression depending on the other concentrations, and plays no role in the first four equations. Hence we can discard this last equation from our study and concentrate on the others.

For the sake of readability, we adopt the following notation: $\mu_{\max,\mathrm{H_2a}}$ will be denoted as $\mu_{\max}$, $K_{s,\mathrm{H_2a}}$ as K, $Y_{\mathrm{H_2a}}$ as Y and $K_{H,\mathrm{H_2}}$ as K_H.

We first put the model in singularly perturbed form (Eqs. (4.1)-(4.2), page 94). This is is usually carried out by scaling and possibly changing variables. Here, scaling is sufficient, so we introduce constants m, h_{g} and h_{l} such that the normalized variables x_1, x_2, x_3, x_4 defined by

$$x_1 = \frac{x_{\mathrm{H_2a}}}{m}, \quad x_2 = \frac{\chi}{m}, x_3 = \frac{s_{\mathrm{g,H_2}}}{h_{\mathrm{g}}}, \quad x_4 = \frac{s_{\mathrm{H_2}}}{h_{\mathrm{l}}},$$

all belong to the interval $[0,1]$. These scaling coefficients can be obtained from the experimental data. m is chosen to be the maximum value of the measured OD, divided by α, h_{g} is taken to be the maximum value of hydrogen concentration in the gas phase, and we set $h_{\mathrm{l}} = K_H RT h_{\mathrm{g}}$.

The state equations for these normalized variables are

$$\dot{x}_1 = (\mu_{\max} \frac{x_4}{\frac{K}{h_{\mathrm{l}}} + x_4} - k_d) x_1, \tag{5.8}$$

$$\dot{x}_2 = k_d x_1 - k_i x_2, \tag{5.9}$$

$$\dot{x}_3 = (x_4 - x_3) k_L a K_H RT \frac{V}{V_{\mathrm{g}}}, \tag{5.10}$$

$$\dot{x}_4 = -\left(\frac{\mu_{\max} m}{h_{\mathrm{l}} Y} \frac{x_1}{\frac{K}{h_{\mathrm{l}}} + x_4} + k_L a \right) x_4 + k_L a x_3. \tag{5.11}$$

The scaling factors and the order of magnitude of the parameters are computed from the experimental data or drawn from the literature

$$m = \frac{0.26}{5.974} = 0.044, \quad h_{\mathrm{g}} = 0.677, \quad h_{\mathrm{l}} = 0.013,$$
$$K, k_d, k_i, Y \approx 0.05, \quad \mu_{\max} \approx 0.1\,.$$

Equations (5.8)-(5.11) share the generic form

$$\dot{x}_i = d_i(\boldsymbol{x}) x_i + p_i(\tilde{\boldsymbol{x}}_i), \tag{5.12}$$

where $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^{\mathrm{T}}$ and $\tilde{\boldsymbol{x}}_i = (x_j)_{\substack{j=1,\ldots,4 \\ j \neq i}}$.

We set the local time scale $\tau_i(\boldsymbol{x})$ associated to x_i as

$$\tau_i(\boldsymbol{x}) = \frac{1}{|d_i(\boldsymbol{x})|}. \tag{5.13}$$

As the dynamical system is positive, we easily obtain that starting from a non-negative initial state

$$k_L a \leq |d_4(\boldsymbol{x})| = \frac{\mu_{\max} m}{h_{\mathrm{l}} Y} \frac{x_1}{\frac{K}{h_{\mathrm{l}}} + x_4} + k_L a \leq \frac{\mu_{\max} m}{KY} + k_L a,$$
$$-k_d \leq d_1(\boldsymbol{x}) = \mu_{\max} \frac{x_4}{\frac{K}{h_{\mathrm{l}}} + x_4} - k_d \leq \mu_{\max} - k_d$$

so

$$\frac{1}{\frac{\mu_{\max} m}{KY} + k_L a} \leq \tau_4(\boldsymbol{x}) \leq \frac{1}{k_L a} \approx 0.12,$$

and

$$\tau_1(\boldsymbol{x}) \geq \frac{1}{\max(k_d, |\mu_{\max} - k_d|)} \approx 20.$$

Moreover, $\tau_2(\boldsymbol{x}) = 1/k_d \approx 20$, and finally $\tau_3(\boldsymbol{x}) = \frac{V_{\mathrm{g}}}{k_L a V K_H R T} \approx 27$. Hence, there is at least a 100 fold change between the time scale associated to x_4 and the other time scales, and the state variables of the model can be separated into slow and fast ones.

We perform the time scale change $t_s = \frac{t}{\tau_s}$, where $\tau_s = 20$, in order to express the model in the slow time scale and obtain the standard form for the singular perturbation problem. The dot symbol will now refer to time differentiation w.r.t. t_s. We set $\delta = \frac{1}{k_L a \tau_s} = 0.006$ and obtain

$$\dot{x}_1 = (\tau_s \mu_{\max} \frac{x_4}{\frac{K}{h_{\mathrm{l}}} + x_4} - \tau_s k_d) x_1,$$
$$\dot{x}_2 = \tau_s k_d x_1 - \tau_s k_i x_2,$$
$$\dot{x}_3 = (x_4 - x_3) \tau_s k_L a K_H R T \frac{V}{V_{\mathrm{g}}},$$
$$\delta \dot{x}_4 = -\left(\frac{\mu_{\max} m}{k_L a h_{\mathrm{l}} Y} \frac{x_1}{\frac{K}{h_{\mathrm{l}}} + x_4} + 1 \right) x_4 + x_3,$$

where all the coefficients have orders of magnitude ranging from 0.5 to 4. Coming back to the original variables, the model is in the standard form (4.1-4.2), with

$$\mathbf{x} = \begin{pmatrix} x_{\mathrm{H_2a}} \\ \chi \\ s_{\mathrm{g,H_2}} \end{pmatrix}, \ \mathbf{z} = (s_{\mathrm{H_2}}),$$

$$\mathbf{f}(\mathbf{x},\mathbf{z}) = \begin{pmatrix} \tau_s \mu_{\mathrm{H_2a}} - \tau_s k_d x_{\mathrm{H_2a}} \\ \tau_s k_d x_{\mathrm{H_2a}} - \tau_s k_i \chi \\ \tau_s k_L a (s_{\mathrm{H_2}} - K_H RT s_{\mathrm{g,H_2}}) \frac{V}{V_{\mathrm{g}}} \end{pmatrix},$$

$$\mathbf{g}(\mathbf{x},\mathbf{z}) = -\frac{\mu_{\max}}{k_L a Y} \frac{s_{\mathrm{H_2}} x_{\mathrm{H_2a}}}{K + s_{\mathrm{H_2}}} - (s_{\mathrm{H_2}} - K_H RT s_{\mathrm{g,H_2}}),$$

and initial conditions $\mathbf{x}(0) = \mathbf{x_0}$ and $\mathbf{z}(0) = \mathbf{z_0}$. Note that in our case, $\mathbf{f}$ and $\mathbf{g}$ are time invariant and do not depend on δ. They are continuously differentiable on $D_x \times D_z$, with $D_x = \mathbb{R}^3_+$ and $D_z = (-K, +\infty)$.

The quasi-steady-state hydrogen concentration in the liquid phase $l(\mathbf{x})$ satisfies the algebraic equation

$$-\frac{\mu_{\max}}{k_L a Y} \frac{l(\mathbf{x}) x_{\mathrm{H_2a}}}{K + l(\mathbf{x})} - (l(\mathbf{x}) - K_H RT s_{\mathrm{g,H_2}}) = 0, \tag{5.14}$$

which can be reformulated as a second-degree polynomial equation. It is easy to check that it has a positive discriminant and a unique positive root $l(\mathbf{x})$, which is the quasi-steady hydrogen concentration in the liquid phase, given by

$$l(\mathbf{x}) = \frac{-c_1 + \sqrt{c_1^2 + 4c_2}}{2}, \tag{5.15}$$

with

$$c_1 = \frac{\mu_{\max} x_{\mathrm{H_2a}}}{Y k_L a} + K - K_H RT s_{\mathrm{g,H_2}},$$

$$c_2 = K_H K RT s_{\mathrm{g,H_2}}.$$

The reduced model is then given by $\dot{\bar{\mathbf{x}}} = \mathbf{f}(\bar{\mathbf{x}}, l(\bar{\mathbf{x}}))$, that is

$$\dot{\bar{x}}_{\mathrm{H_2a}} = \tau_s \mu_{\max} \frac{l(\bar{\mathbf{x}})}{K + l(\bar{\mathbf{x}})} \bar{x}_{\mathrm{H_2a}} - \tau_s k_d \bar{x}_{\mathrm{H_2a}}, \tag{5.16}$$

$$\dot{\bar{\chi}} = \tau_s k_d \bar{x}_{\mathrm{H_2a}} - \tau_s k_i \bar{\chi}, \tag{5.17}$$

$$\dot{\bar{s}}_{\mathrm{g,H_2}} = \tau_s k_L a (l(\bar{\mathbf{x}}) - K_H RT \bar{s}_{\mathrm{g,H_2}}) \frac{V}{V_{\mathrm{g}}}, \tag{5.18}$$

with initial condition $\bar{\mathbf{x}}(0) = \mathbf{x_0}$.

We can now check that the three assumptions of the Tikhonov theorem are satisfied.

Assumption 1 (regularity). The vector functions **f**, **g** and their first partial derivatives w.r.t. **x** and **z** are continuously differentiable on $D_x \times D_z$. The function $l(\mathbf{x})$ and Jacobian matrix $\frac{\partial \mathbf{g}}{\partial \mathbf{z}}$ are also continuously differentiable on $D_x \times D_z$.

Assumption 2 (existence of bounded solutions of the reduced-order model). It must be shown that for all $T > 0$, there exists a compact subset S of D_x such that any trajectory of the reduced model starting in S is well defined, unique on $[0,T]$, and remains in S. We proved a stronger result. Notice that the complete and reduced models are both positive: the vector fields of both models have no outward component when reaching the boundary of the positive cone ($\mathbb{R}^4_+$ and $\mathbb{R}^3_+$ respectively) from the inside. Moreover, setting $A_r(t) = \bar{x}_{\mathrm{H_2a}}(t) + \bar{\chi}(t) + \frac{YV_\mathrm{g}}{V}\bar{s}_{\mathrm{g,H_2}}(t)$, we have from (5.16), (5.17) and (5.18)

$$\dot{A}_r(t) = -\tau_s k_i \bar{\chi} \leq 0.$$

Therefore, for all $C > 0$, if we define the following compact subset of D_x:

$$S_C = \{\mathbf{x} \in \mathbb{R}^3_+,\, x_1 + x_2 + \frac{YV_\mathrm{g}}{V} x_3 \leq C\},$$

we have

$$\bar{\mathbf{x}}_\mathbf{0} \in S_C \;\Rightarrow\; \forall t \geq 0,\, \bar{\mathbf{x}}(t) \in S_C.$$

Hence, trajectories starting in S_C are uniquely defined and remain in S_C for ever.

Assumption 3 (stability of the origin of the boundary-layer model). Consider now the boundary layer differential equation

$$\frac{dy}{d\upsilon} = -\frac{\mu_{\max}}{k_L a Y} \frac{\bar{x}_{\mathrm{H_2a}}(y + l(\bar{\mathbf{x}}))}{K + y + l(\bar{\mathbf{x}})} - (y + l(\bar{\mathbf{x}}) - K_H R T \bar{s}_{\mathrm{g,H_2}}), \tag{5.19}$$

where $\bar{\mathbf{x}}$ is considered to be constant in the $\upsilon = \frac{t_\mathrm{s}}{\delta}$ time scale, and the unknown function $\upsilon \mapsto y(\upsilon)$ takes its values in D_z. By construction, $l(\bar{\mathbf{x}})$ is positive when $\bar{\mathbf{x}} \in D_x$, so $y \in D_z$ implies $K + y + l(\bar{\mathbf{x}}) > 0$. Term to term subtraction of (5.19) and (5.14) yields

$$\frac{dy}{d\upsilon} = -\frac{\mu_{\max} \bar{x}_{\mathrm{H_2a}} K}{k_L a Y (K + l(\bar{\mathbf{x}}))} \frac{y}{K + y + l(\bar{\mathbf{x}})} - y. \tag{5.20}$$

Setting $K_1 = K + l(\overline{\mathbf{x}})$ and $K_2 = \frac{\mu_{\max} \bar{x}_{\mathrm{H_2a}} K}{k_L a Y (K + l(\overline{\mathbf{x}}))}$, we can rewrite (5.20) as

$$\frac{dy}{d\upsilon} = -K_2 \frac{y}{y + K_1} - y. \tag{5.21}$$

It is easy to see that when starting from an initial condition $y(0) > -K_1$, we have $y(\upsilon) > -K_1$ for all $\upsilon > 0$, since $\lim_{y \to -K_1^+} -\left(K_2 \frac{y}{y+K_1} + y \right) = +\infty$. Therefore, if $y(0) > -K > -K_1$, we have, for all $\upsilon > 0$,

$$\frac{1}{2}\frac{dy^2}{d\upsilon} = -y^2 \left(\frac{K_2}{y + K_1} + 1 \right) \leq -y^2.$$

It implies that $\forall \upsilon \geq 0$, $|y(\upsilon)| \leq |y(0)| e^{-\upsilon}$. Hence, the boundary layer model is exponentially stable, uniformly in $(t, \overline{\mathbf{x}}) \in \mathbb{R}_+ \times D_x$, with region of attraction $R_y = D_z$.

The hypotheses of the Tikhonov theorem are thus verified, so there exists $\delta^\star > 0$ and $\rho > 0$ such that for all time intervals $[0, T]$ where T is of order τ_s, we can find $C > 0$ such that

$$\forall t \in [0, T],\ \|\mathbf{x}(t) - \overline{\mathbf{x}}(t)\| < C\delta,$$

for all $\delta < \delta^\star$ and all initial conditions $\mathbf{z}(0)$ that satisfy $\|\mathbf{z}(0) - l(\mathbf{x}(0)\| \leq \rho$.

Simulations confirm the validity of the model reduction. The closeness between the two models depends on the initial condition of hydrogen in the liquid phase, which is unknown in our study. However, we know that

$$0 < s_{\mathrm{H_2}}(0) < K_{H,\mathrm{H_2}} RT s_{\mathrm{g,H_2}}(0) \approx 0.013. \tag{5.22}$$

We simulated the complete and reduced models with $s_{\mathrm{H_2}}(0)$ taking the values of the bounds in (5.22). For the two conditions, the responses of both models were superimposed. For illustration purpose, we set $s_{\mathrm{H_2}}(0) = 0.1$ (which is physically impossible for the experimental conditions) to observe the singular perturbation approximation in a large time horizon. The responses of both models are displayed in Fig. 5.3. We observe that the hydrogen concentration given by the complete model converges rapidly to the hydrogen concentration calculated from the singular perturbation approach. The responses of the slow variables in both models are almost identical.

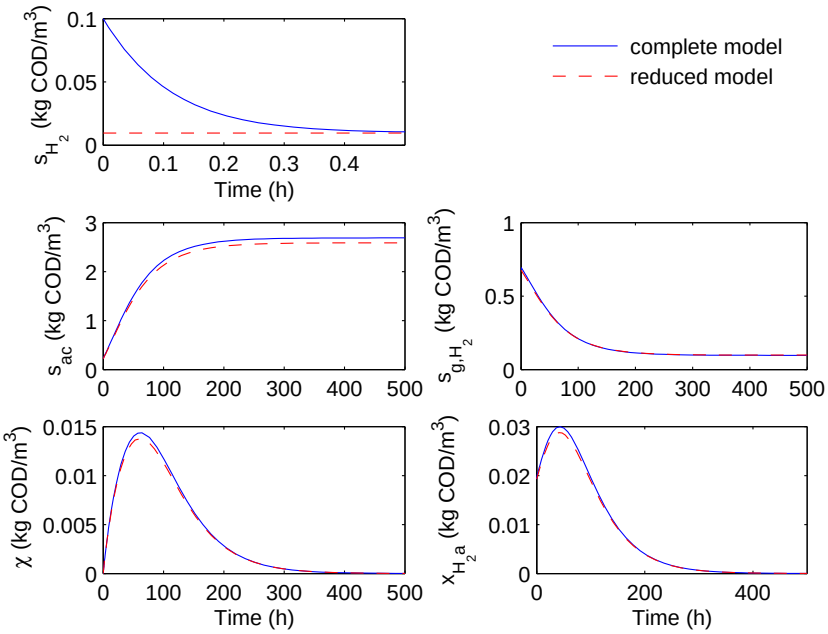

Figure 5.3: Comparison of the complete model and reduced model for homoacetogenesis. Discrepancy in initial conditions for s_{H_2} has been exaggerated for the purpose of illustration.

5.1.4 Identifiability study

The study of identifiability is important before tackling the estimation problem (see Section 4.3.1.1). This issue is particularly central when dealing with biological systems for which critical significance is often attached to the value of the parameters.

Let $\boldsymbol{\theta} = (\mu_{\max},\, K,\, k_d,\, k_i,\, Y)^{\mathrm{T}}$ be the parameter vector in the parameter space $U \subset \mathbb{R}^5_+$ and $\mathbf{x} = (x_{\mathrm{H_2a}},\, \chi,\, s_{\mathrm{g,H_2}},\, s_{\mathrm{ac}})^{\mathrm{T}}$ be the state vector of the model. $\mathbf{x}(t,\boldsymbol{\theta})$ denotes the trajectory of the model with parameter $\boldsymbol{\theta}$ at time t, and $\mathbf{x}_0$ is the vector of initial conditions. Denote by $l(\mathbf{x},\boldsymbol{\theta})$ the unique positive solution of (5.14), by $\mathbf{f}(\mathbf{x},\boldsymbol{\theta})$ the vector function whose entries are the right-handsides of (5.1-5.4) and by $\mathbf{h}(\mathbf{x},\boldsymbol{\theta}) = (\alpha(x_{\mathrm{H_2a}}+\chi),\, s_{\mathrm{g,H_2a}},\, s_{\mathrm{ac}})^{\mathrm{T}}$ the vector of measurements.

As the model under consideration is an uncontrolled nonlinear ODE model with initial conditions, we use a sufficient condition for the model to be globally identifiable that can

be found in Denis-Vidal and Joly-Blanchard (2004):

Proposition 1 *Let $(\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}) \in U^2$ be such that $\mathbf{f}(\mathbf{x}_0(\boldsymbol{\theta}),\boldsymbol{\theta}) \neq \mathbf{0}$ and $\mathbf{f}(\mathbf{x}_0(\tilde{\boldsymbol{\theta}}),\tilde{\boldsymbol{\theta}}) \neq \mathbf{0}$. If the existence of an open neighborhood $\tilde{V}$ of $\mathbf{x}_0(\tilde{\boldsymbol{\theta}})$ and an analytic diffeomorphism $\sigma : \tilde{V} \rightarrow \sigma(\tilde{V})$ such that*

$$
\begin{aligned}
&(i)\ \sigma(\mathbf{x}_0(\tilde{\boldsymbol{\theta}})) = \mathbf{x}_0(\boldsymbol{\theta}),\\
&(ii)\ \forall \tilde{\mathbf{z}} \in \tilde{V},\ \mathbf{f}(\sigma(\tilde{\mathbf{z}}),\boldsymbol{\theta}) = \frac{\partial \sigma(\tilde{\mathbf{z}})}{\partial \tilde{\mathbf{z}}} \mathbf{f}(\tilde{\mathbf{z}},\tilde{\boldsymbol{\theta}}),\\
&(iii)\ \mathbf{h}(\sigma(\mathbf{x}(t,\tilde{\boldsymbol{\theta}})),\boldsymbol{\theta}) = \mathbf{h}(\mathbf{x}(t,\tilde{\boldsymbol{\theta}}),\tilde{\boldsymbol{\theta}}),\\
&\quad \textit{for all } t \textit{ such that } \tilde{\mathbf{x}}(t,\tilde{\boldsymbol{\theta}}) \in \tilde{V},
\end{aligned}
$$

implies $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, then the model is globally identifiable at $\boldsymbol{\theta}$.

Definition 2 *The model is said to be globally structurally identifiable if it is globally identifiable for almost any $\boldsymbol{\theta}$ in U.*

Initial conditions are denoted by $\mathbf{x}_0$ and $\tilde{\mathbf{x}}_0$. We assume that the initial conditions are identical for both models and known. Moreover we assume that $x_{01} > 0$, $x_{02} = 0$, and that the experiment starts at the bacterial growth phase, which implies that $\dot{\sigma}_1(\tilde{\mathbf{x}}(0)) > 0$ and $\dot{\tilde{x}}_1(0) > 0$. The parameter set U is such that

$$
U = \left\{ \begin{array}{l} \mathbf{v} \in \mathbb{R}^5_+ / \ \frac{v_1}{v_2 + 0.5\left(-c_1 + \sqrt{c_1^2 + 4c_2}\right)} - v_3 > 0, \\ \text{where } c_1 = \frac{v_1 x_{01}}{v_5 k_L a} + v_2 - K_H R T x_{03}, \text{ and } c_2 = K_H R T x_{03} v_2 \end{array} \right\}
$$

We prove that the sufficient condition in Proposition 1 is met, so that the model is structurally globally identifiable.

proof:
For conciseness we use in the sequel the notation $\tilde{\mathbf{x}}(t) = \mathbf{x}(t,\tilde{\boldsymbol{\theta}})$ and $\mathbf{x}(t) = \mathbf{x}(t,\boldsymbol{\theta})$. For all t such that $\tilde{\mathbf{x}}(t) \in \tilde{V}$ we have from (iii)

$$\sigma_3(\tilde{\mathbf{x}}(t)) = \tilde{x}_3(t), \tag{5.23}$$

$$\sigma_4(\tilde{\mathbf{x}}(t)) = \tilde{x}_4(t), \tag{5.24}$$

$$\sigma_1(\tilde{\mathbf{x}}(t)) + \sigma_2(\tilde{\mathbf{x}}(t)) = \tilde{x}_1(t) + \tilde{x}_2(t). \tag{5.25}$$

Let us now differentiate (5.23)-(5.25) w.r.t. time in combination with (ii) in Proposition 1.

Differentiating (5.23) in time and using (ii) we get

$$(ii_3) \quad k_L a(l(\sigma(\tilde{\mathbf{x}}),\theta) - K_{H,\mathrm{H}_2} RT\,\sigma_3(\tilde{\mathbf{x}}(t))\frac{V}{V_\mathrm{g}} = k_L a(l(\tilde{\mathbf{x}},\tilde{\theta}) - K_{H,\mathrm{H}_2} RT\,\tilde{x}_3(t))\frac{V}{V_\mathrm{g}}.$$

It follows from (5.23) that

$$l(\tilde{\mathbf{x}}(t),\tilde{\theta}) = l(\sigma(\tilde{\mathbf{x}}(t)),\theta) = l(t). \tag{5.26}$$

Eq. (5.26) can be used in (5.14) and we get

$$\frac{Y}{\tilde{Y}}\frac{\tilde{\mu}_{\max}\tilde{x}_1(t)}{\tilde{K}+l(t)} = \frac{\mu_{\max}\sigma_1(\tilde{\mathbf{x}}(t)}{K+l(t)}. \tag{5.27}$$

Differentiating (5.24) in time and using (ii) we get

$$(ii_4) \quad \frac{1-Y}{Y}\mu_{\max}\frac{\sigma_1(\tilde{\mathbf{x}}(t))}{K+l(t)} = \frac{1-\tilde{Y}}{\tilde{Y}}\tilde{\mu}_{\max}\frac{\tilde{x}_1(t)}{\tilde{K}+l(t)}.$$

The combination of (ii_4) with (5.27) yields $Y=\tilde{Y}$, and

$$\frac{\tilde{\mu}_{\max}\tilde{x}_1(t)}{\tilde{K}+l(t)} = \frac{\mu_{\max}\sigma_1(\tilde{\mathbf{x}}(t))}{K+l(t)}. \tag{5.28}$$

Differentiating (5.25) in time and using (ii), we obtained

$$(ii_1+ii_2) \quad \mu_{\max}\frac{\sigma_1(\tilde{\mathbf{x}}(t))l(t)}{K+l(t)} - k_i\sigma_2(\tilde{\mathbf{x}}(t)) = \tilde{\mu}_{\max}\frac{\tilde{x}_1(t)l(t)}{\tilde{K}+l(t)} - \tilde{k}_i\tilde{x}_2(t). \tag{5.29}$$

By using (5.28) on (5.29), we obtain

$$k_i\sigma_2(\tilde{\mathbf{x}}(t)) = \tilde{k}_i\tilde{x}_2(t). \tag{5.30}$$

Assuming $k_i \neq 0$, the substitution of (5.30) in (5.25) proves that

$$\sigma_1(\tilde{\mathbf{x}}(t)) = \tilde{x}_1 + (1-\frac{\tilde{k}_i}{k_i})\tilde{x}_2(t). \tag{5.31}$$

Now, if we assume $k_d \neq 0$, differentiating (5.30) in time and using (ii), we get

$$k_d\sigma_1(\tilde{\mathbf{x}}(t)) - k_i\sigma_2(\tilde{\mathbf{x}}(t)) = \frac{\tilde{k}_i}{k_i}(\tilde{k}_d\tilde{x}_1(t) - \tilde{k}_i\tilde{x}_2(t)).$$

Substituting (5.30) in the above equation yields

$$\sigma_1(\tilde{\mathbf{x}}(t)) = \frac{\tilde{k}_i\tilde{k}_d}{k_i k_d}\tilde{x}_1(t) + \frac{\tilde{k}_i}{k_d}(1-\frac{\tilde{k}_i}{k_i})\tilde{x}_2(t). \tag{5.32}$$

At $t=0$, using (i) in (5.32), we get that $\frac{\tilde{k}_i\tilde{k}_d}{k_i k_d} = 1$.
Now, combining (5.31) with (5.32) we get

$$(1-\frac{\tilde{k}_i}{k_d})(1-\frac{\tilde{k}_i}{k_i})\tilde{x}_2(t) = 0. \tag{5.33}$$

There exists at least one time t such that $\tilde{x}_2(t) \neq 0$, (5.33) yields that $(1-\frac{\tilde{k}_i}{k_d})(1-\frac{\tilde{k}_i}{k_i}) = 0$. Then, we have

$$k_i = \tilde{k}_i \text{ and } k_d = \tilde{k}_d, \tag{5.34}$$

or

$$k_d = \tilde{k}_i \text{ and } k_i = \tilde{k}_d. \tag{5.35}$$

If (5.34) *holds.* Combining (5.31) and (5.34), we obtain that $\sigma_1(\tilde{\mathbf{x}}(t)) = \tilde{x}_1$, which implies in (5.28) that

$$\frac{\tilde{K}+l(t)}{\tilde{\mu}_{\max}} = \frac{K+l(t)}{\mu_{\max}}. \tag{5.36}$$

It follows that

$$(\mu_{\max}-\tilde{\mu}_{\max})l(t) = \tilde{\mu}_{\max}K - \mu_{\max}\tilde{K}.$$

Two cases arises: either $\mu_{\max} = \tilde{\mu}_{\max}$ which implies $K = \tilde{K}$, or $\mu_{\max} \neq \tilde{\mu}_{\max}$ and in this case, $l(t)$ is constant.
In this last situation, let us define $\zeta = \frac{l(\tilde{\mathbf{x}})}{\tilde{K}+l(\tilde{\mathbf{x}})}$, then differentiating (5.14) in time leads to the following differential equation for ζ

$$\dot{\zeta} = \frac{-\tilde{\mu}_{\max}\tilde{x}_1}{\tilde{\mu}_{\max}\tilde{x}_1 + \frac{Yk_L a\tilde{K}}{(1-\zeta)^2}}\zeta(\tilde{\mu}_{\max}\zeta - \kappa), \tag{5.37}$$

where $\kappa = \tilde{k}_d - \frac{k_L a K_H RTV}{V_g}$. Due to the assumptions on the initial conditions, we have $\zeta(0) > \frac{\tilde{k}_d}{\tilde{\mu}_{\max}} > \max(0, \frac{\kappa}{\tilde{\mu}_{\max}})$, and therefore $\dot{\zeta}(0) < 0$. However, a constant value of $l(t)$ implies a constant value of $\zeta(t)$, and therefore $\dot{\zeta}(0) = 0$. This contradiction shows that this case is not possible.

If (5.35) *holds.* At $t = 0$, (5.28) yields

$$\frac{\tilde{K} + l(0)}{\tilde{\mu}_{\max}} = \frac{K + l(0)}{\mu_{\max}}. \tag{5.38}$$

It follows that

$$(\mu_{\max} - \tilde{\mu}_{\max}) l(0) = \tilde{\mu}_{\max} K - \mu_{\max} \tilde{K}.$$

In this equation, either $\mu_{\max} = \tilde{\mu}_{\max}$, which implies $K = \tilde{K}$, or $\mu_{\max} \neq \tilde{\mu}_{\max}$ and

$$l(0) = \frac{\tilde{\mu}_{\max} K - \mu_{\max} \tilde{K}}{\mu_{\max} - \tilde{\mu}_{\max}}.$$

Let us consider the case where $\mu_{\max} = \tilde{\mu}_{\max}$. From (5.28), we deduce that

$$\tilde{x}_1(t) = \sigma_1(\tilde{\mathbf{x}}(t)). \tag{5.39}$$

The substitution of (5.39) in (5.31) leads to $\tilde{k}_i = k_i$, and $\tilde{k}_d = k_d$ from (5.32). From (5.35), we finally obtain that $k_d = \tilde{k}_i = k_i = \tilde{k}_d$.

In the case where $\mu_{\max} \neq \tilde{\mu}_{\max}$, we can eliminate l between the two second degree equations that we obtain writing (5.14) for both systems, and eliminate $\sigma_1(\tilde{\mathbf{x}})$ using (5.31) in the resulting equation.
The computation were performed with the symbolic computation software Maple, and led to a nonlinear polynomial relationship $N_1(t) = 0$ between the parameters of both systems and the first three components of the state vector $\tilde{x}$. Evaluating this expression at $t = 0$ leads to

$$\begin{aligned} &k_L a Y K_H R T (\mu_{\max} - \tilde{\mu}_{\max})(K - \tilde{K}) x_{03} + \\ &(\tilde{\mu}_{\max} K - \mu_{\max} \tilde{K})((\mu_{\max} - \tilde{\mu}_{\max}) x_{01} + k_L a Y (K - \tilde{K})) = 0 \end{aligned} \tag{5.40}$$

Setting $\Delta_\mu = \tilde{\mu}_{\max} - \mu_{\max}$ and $\Delta_K = \tilde{K} - K$, the above expression becomes

$$N_{01}(\Delta_K, \Delta_\mu) = k_L a Y K_H R T x_{03} \Delta_K \Delta_\mu - (K \Delta_\mu - \mu_{\max} \Delta_K)(x_{01} \Delta_\mu + k_L a Y \Delta_K) = 0. \tag{5.41}$$

Note that N_{01} is a second degree nonlinear multivariate polynomial equation in Δ_K and Δ_μ, whose coefficients depend on the initial conditions and on parameters $\mu_{\max}$ and K.
Differentiating the nonlinear polynomial relationship $N_1(t) = 0$ obtained above along

the system's trajectories, we obtain an additional nonlinear polynomial relationship $N_2(t) = 0$, which we can evaluate at $t = 0$. We obtain a new expression $N_{02}(\Delta_K, \Delta_\mu) = 0$, where N_{02} is a fourth degree nonlinear multivariate polynomial equation in Δ_K and Δ_μ, whose coefficients depend on the initial conditions and on the full vector of parameters $(\mu_{max}, K, k_d, k_i)^{\mathrm{T}}$, thanks to the assumption (5.35).
Elimination of Δ_K between the two equations $N_{01}(\Delta_K, \Delta_\mu) = 0$ and $N_{02}(\Delta_K, \Delta_\mu) = 0$ for a given initial condition and parameter vector $(\mu_{max}, K, k_d, k_i)^{\mathrm{T}}$ shows that, except possibly on a zero measure set of parameter $(\mu_{\max}, K, k_i, k_d)^{\mathrm{T}}$, we have either $\Delta_\mu = \Delta_K = 0$ (and it follows that the model is globally identifiable), or $\Delta_\mu = f_1(\mu_{max}, K, k_d, k_i, x_{30}, x_{10})$ and $\Delta_K = f_2(\mu_{max}, K, k_d, k_i, x_{30}, x_{10})$.
In this last situation we can differentiate $N_2(t) = 0$, and obtain an additional nonlinear polynomial relationship $N_3(t) = 0$, which can be evaluated at $t = 0$. This provides a nonlinear multivariate polynomial equation in Δ_K and Δ_μ, $N_{03}(\Delta_K, \Delta_\mu) = 0$, whose coefficients depend on the initial conditions and on the full vector of parameters $(\mu_{max}, K, k_d, k_i)^{\mathrm{T}}$. Substituting the expressions of Δ_μ and Δ_K in $N_{03}(\Delta_K, \Delta_\mu) = 0$ shows that the existence of the non zero solution $\Delta_\mu = f_1(\mu_{max}, K, k_d, k_i, x_{30}, x_{10})$ and $\Delta_K = f_2(\mu_{max}, K, k_d, k_i, x_{30}, x_{10})$ may be possible on the zero measure set for $(\mu_{max}, K, k_d, k_i)^{\mathrm{T}}$ defined by

$$N_{03}(f_2(\mu_{max}, K, k_d, k_i, x_{30}, x_{10}), f_1(\mu_{max}, K, k_d, k_i, x_{30}, x_{10})) = 0,$$

and is not possible elsewhere.

In all cases, we have shown that except on a zero measure set, the model is globally identifiable in U. It follows that the model is structurally globally identifiable. ■

5.1.5 Parameter estimation

Estimation was performed by the maximum likelihood (ML) approach, using the toolbox IDEAS (see Section 4.5). Three estimators were evaluated. The estimator **C1** minimizes the cost function (4.15) (page 101)

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{n_t} [\mathbf{y}(t_i) - \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{y}(t_i) - \mathbf{y}_{\mathrm{m}}(t_i, \boldsymbol{\theta})],$$

with

$$\boldsymbol{\Sigma} = \sigma^2 \mathrm{diag}[v_1, v_2, v_3]$$

and $v_k = (y_{k\max} - y_{k\min})^2$, where $y_{k\max}$ and $y_{k\min}$ are respectively the maximum and minimum value among all the collected data for output k. The estimator **C2** minimizes again the cost function defined by (4.15) but with a diagonal matrix Σ computed from the data reported in Bernalier et al. (1996b), which are mean values of three determinations. At each time the standard deviations were calculated and used to compute the matrix. The kth diagonal element σ_k^2 of Σ was calculated as

$$\sigma_k^2 = \frac{1}{n_\mathrm{t}} \sum_{i=1}^{n_\mathrm{t}} \sigma_{k,i}^2. \tag{5.42}$$

The estimator **C3** minimizes (4.18) (page 102)

$$J(\boldsymbol{\theta}) = \ln \left[\det \sum_{i=1}^{n_\mathrm{t}} [\mathbf{y}(t_i) - \mathbf{y}_\mathrm{m}(t_i, \boldsymbol{\theta})][\mathbf{y}(t_i) - \mathbf{y}_\mathrm{m}(t_i, \boldsymbol{\theta})]^\mathrm{T} \right],$$

and finally, the estimator **C4** minimizes the cost function defined by (4.20) (page 102)

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{n_\mathrm{y}} \frac{n_\mathrm{t}}{2} \ln \left[\sum_{\mathrm{i}=1}^{n_\mathrm{t}} [y_k(t_i) - y_{\mathrm{m}_k}(t_i, \boldsymbol{\theta})]^2 \right].$$

5.1.6 Results and discussion

Figure 5.4 compares the experimental data and the outputs of the mathematical model with the parameters obtained for the four criteria. From a qualitative point of view, the optimal set of parameters for each cost function gives a very close fit for the optical density. However for the acetate and hydrogen concentrations, it is seen that more weight is given either to acetate or to hydrogen, depending of the cost function considered. The estimators **C1** and **C2** tend to favor the fit of acetate concentrations. To the contrary, the more sophisticated estimators **C3** and **C4** tend to favor the fit of hydrogen concentrations. This is consistent with the experimental conditions, from which it seems reasonable to give more importance to the fit of the measured hydrogen concentrations, because hydrogen is only consumed through homoacetogenesis. As for acetate, it could be produced by other reactions, for instance within the yeast fermentation pathway, not modeled here.

The resulting parameter estimates are shown in Table 5.1 with their approximate confidence intervals calculated from (4.34) (page 106). The values obtained for the Monod constant $K_{s,\mathrm{H_2a}}$ and the yield factor $Y_{\mathrm{H_2a}}$ are in accordance with the information reported

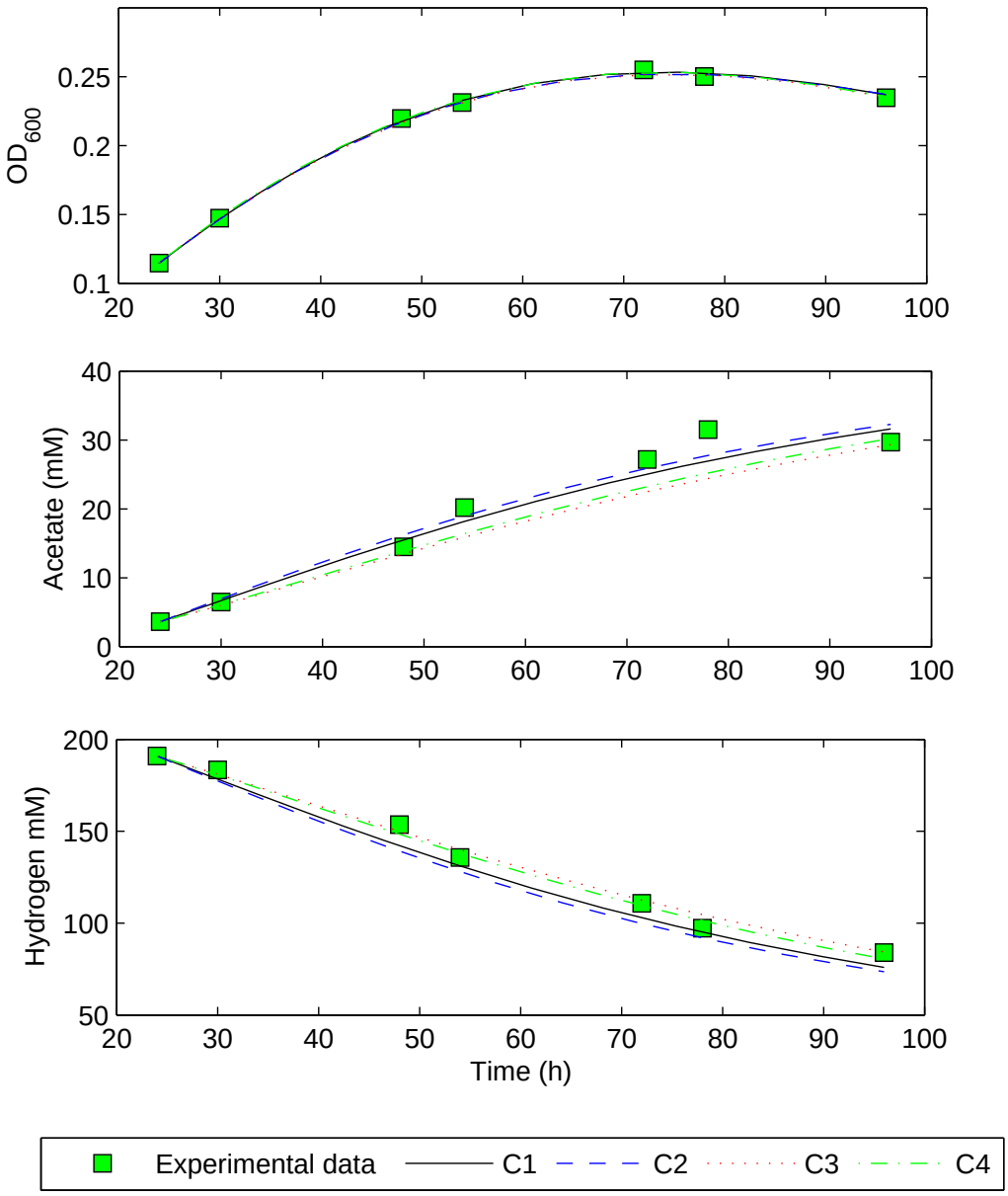

Figure 5.4: Comparison of the ML estimators.

in Lokshina and Vavilin (1999) for different anaerobic systems. The maximum growth rate $\mu_{\max,H_2a}$ is higher than the maximum value reported. This could be due to the fact that, in the experiment considered here, the medium resistance to growth is low compared to the systems referenced in Lokshina and Vavilin (1999) (all of them solids). We did not find theoretical values to compare the estimates of k_d for homoacetogens. However, the estimated values of k_d are in the range reported in Batstone et al. (2002) for hydrogen-utilizing methanogenic bacteria.

The confidence intervals of the estimates turn out to be very disappointing, owing to the limited set of data and well-known practical identifiability problems (Dochain and Vanrolleghem, 2001). The parameters $\mu_{\max,H_2a}$ and K_{s,H_2a} show a very strong correlation, as shown in Table 5.2. Other parameters are also strongly correlated. In Fig. 5.5, we observe the sensitivities of the model output s_{ac} with respect to the parameters K_{s,H_2a} and $\mu_{\max,H_2a}$. The curves are nearly proportional. Simulations of the model with the estimated parameter values indicate that the values taken by the state variable s_{H_2} are always lower than the value of the parameter K_{s,H_2a}. At the end of the experience the simulation provides a value of $s_{H_2} = 0.0038$, which is 7 times lower than the estimate of K_{s,H_2a}. We decided to approximate the Monod equation by a quadratic kinetics

$$\frac{\mu_{\max,H_2a} s_{H_2} x_{H_2a}}{K_{s,H_2a} + s_{H_2}} \approx k_{r,H_2a} s_{H_2} x_{H_2a}, \tag{5.43}$$

with k_{r,H_2a} the ratio $\mu_{\max,H_2a}/K_{s,H_2a}$.

Table 5.1: Comparison of estimated parameters with their approximate 95% confidence intervals.

	C1	**C2**	**C3**	**C4**
K_{s,H_2a}	0.025 ± 6.940	0.020 ± 2.326	0.020 ± 0.338	0.027 ± 0.165
$\mu_{\max,H_2a}$	0.186 ± 3.760	0.157 ± 12.354	0.154 ± 1.760	0.195 ± 0.900
Y_{H_2a}	0.030 ± 0.030	0.029 ± 0.022	0.040 ± 0.020	0.039 ± 0.013
k_d	0.034 ± 1.392	0.036 ± 0.645	0.033 ± 0.097	0.031 ± 0.026
k_i	0.029 ± 1.040	0.025 ± 0.338	0.052 ± 0.270	0.061 ± 0.128

Table 5.2: Correlation matrix for the estimates in the Monod model.

	K_{s,H_2a}	μ_{max,H_2a}	Y_{H_2a}	k_d	k_i
K_{s,H_2a}	1	0.9999	0.4423	-0.9830	0.8357
μ_{max,H_2a}		1	0.4542	-0.9848	0.8429
Y_{H_2a}			1	-0.5705	0.8501
k_d				1	-0.9147
k_i					1

Table 5.3: Estimates with their approximate 95% confidence intervals for the Monod and quadratic kinetic models.

Parameters	**Monod kinetics**	**Quadratic kinetics**
K_{s,H_2a}	0.027 ± 0.165	
μ_{max,H_2a}	0.195 ± 0.900	
k_{r,H_2a}		5.45 ± 0.465
Y_{H_2a}	0.039 ± 0.013	0.040 ± 0.011
k_d	0.031 ± 0.026	0.027 ± 0.003
k_i	0.061 ± 0.128	0.071 ± 0.088

The estimator **C4** was the best ML estimator for the Monod-based model. It was also used to estimate the parameters of the modified model. Confidence intervals are greatly improved, as shown in Table 5.3, column 3. The estimates of the parameters that are shared by the two models have the same order of magnitude, but the confidence intervals are smaller for the modified model.

The data fit with the parameters estimated from the modified-kinetics model are qualitatively as satisfactory as those of the original model. The responses for the two models are superimposed and cannot be distinguished. Without the assessment of the uncertainty of the estimates, we would have had no reason to prefer the modified model to the Monod-based model.

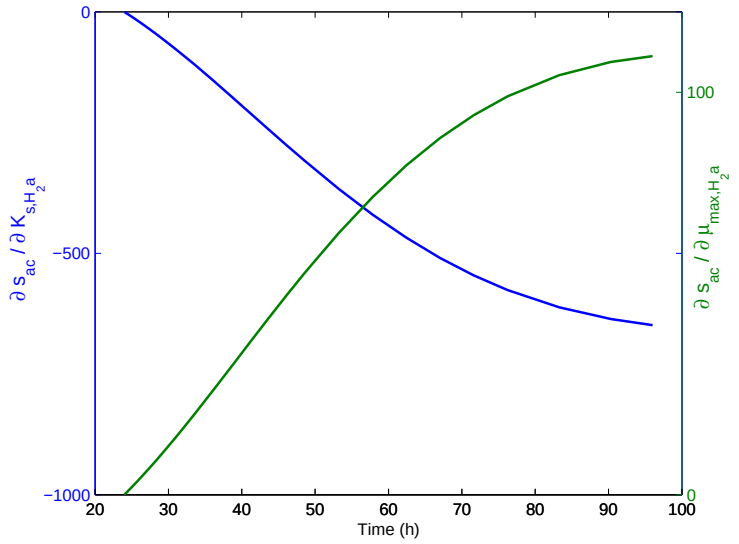

Figure 5.5: Sentivities of acetate concentration s_{ac} with respect to the parameters of Monod equation.

The procedure employed here is generic, and can be applied to estimate model parameters related to others *in vitro* experiments. In the following section, we perform the modelling and identification of bacterial growth experiments devoted to study the metabolism of butyrate production by two human colonic bacteria.

5.2 Lactate utilization and butyrate production by key human colonic bacterial species

Butyrate is the preferred energy source for the colonocytes (colonic epithelial cells) and plays an important role in human health (see, *e.g.*, McIntyre et al. (1993); Pryde et al. (2002); Scheppach and Weiler (2004); Ewaschuk et al. (2005); Hamer et al. (2008)), whilst accumulation of high concentrations of lactate is detrimental (see, *e.g.*, Vernia et al. (1988); Chan et al. (1994); Hove et al. (1994); Kaneko et al. (1997); Belenguer et al. (2007)). Butyrate can be synthesized from two metabolic pathways: butyrate kinase and butyryl CoA:acetate CoA transferase (Miller and Wolin (1996); Diez-Gonzalez et al.

(1999); Duncan et al. (2002a); Macfarlane2003). *In vitro* studies have identified the latter mechanism as the dominant route for butyrate formation in the human colonic ecosystem (Louis et al., 2004). Lactate and and acetate are considered as important precursors for butyrate production (Bourriaud et al. (2005); Morrison et al. (2006); Belenguer et al. (2008)). Lactate is the energy source for the bacteria, while acetate participates as electron sink. Different groups of butyrate producing bacteria, mostly belonging to the Firmicutes, have been isolated from the human GI tract (Barcenilla et al., 2000). *Eubacterium hallii* and strain SS2/1 are dominant member of clostridial cluster XIVa, and are able to utilize lactate and acetate via the butyryl-CoA:acetate-CoA-transferase route.

Here, we develop a mathematical model to describe the conversion of lactate and acetate into butyrate by human colonic bacteria. Figure 5.6 shows the location of the metabolism analyzed here in the whole process of carbohydrate degradation. The model is assessed by comparing, for the species *Eubacterium halli* and strain SS2/1, the model output with *in vitro* data (Duncan et al., 2004b).

5.2.1 Mathematical model and reactions stoichiometry

The mathematical model is derived by writing down mass-balance differential equations for a batch system. The kinetic model is an aggregated representation of the butyryl-CoA:acetate-CoA-transferase pathway represented in Fig 5.7. *In vitro* experiments have shown that acetate cannot be used as a sole source energy for the bacteria (see, *e.g.*, Hino et al. (1991); Duncan et al. (2002a)). Therefore, bacterial growth is assumed to be dependent on lactate concentration only.
According to this, the model equations are

$$\dot{x}_{\mathrm{la}} = Y_{\mathrm{la}}\rho_{\mathrm{la}}, \tag{5.44}$$

$$\dot{s}_{\mathrm{la}} = -\rho_{\mathrm{la}}, \tag{5.45}$$

$$\dot{s}_{\mathrm{ac}} = Y_{\mathrm{ac,la}}\rho_{\mathrm{la}}, \tag{5.46}$$

$$\dot{s}_{\mathrm{bu}} = Y_{\mathrm{bu,la}}\rho_{\mathrm{la}}, \tag{5.47}$$

with ρ_{la} is the specific consumption rate of lactate. A miscellany of expressions have been developed to represent biological kinetic rates (see, *e.g.*, Dochain and Bastin (1990)). In this work we evaluate the following equations

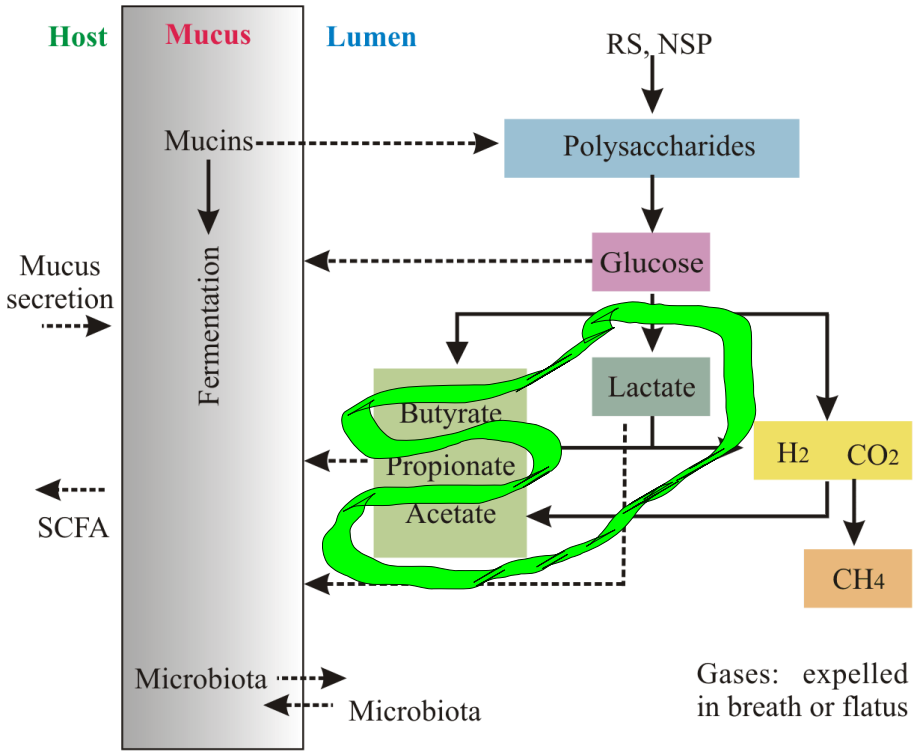

Figure 5.6: Location of butyrate metabolism in the full process of carbohydrate degradation.

Monod equation:

$$\rho_{\mathrm{la}} = k_{m,\mathrm{la}} \frac{s_{\mathrm{la}} x_{\mathrm{la}}}{K_{s,\mathrm{la}} + s_{\mathrm{la}}}, \tag{5.48}$$

Quadratic kinetics:

$$\rho_{\mathrm{la}} = k_{m,\mathrm{la}} s_{\mathrm{la}} x_{\mathrm{la}}, \tag{5.49}$$

First-order kinetics

$$\rho_{\mathrm{la}} = k_{m,\mathrm{la}} s_{\mathrm{la}}, \tag{5.50}$$

The state variables of the model are the bacterial concentration x_{la} in units of optical density (OD), lactate concentration s_{la}, acetate concentration s_{ac} and butyrate concentration s_{bu}, in mM units. $k_{m,\mathrm{la}}$ is the consumption rate constant. For the Monod kinetics, the units of $k_{m,\mathrm{la}}$ are mM Lactate (OD h)$^{-1}$, for the quadratic kinetis they are (OD h)$^{-1}$ and for first-order kinetics h^{-1}. The Monod constant $K_{s,\mathrm{la}}$ is in mM units. Y_{la} (OD/mM Lactate) is the yield factor for biomass with respect to the consumed lactate. The yield factors $Y_{\mathrm{ac,la}}$ (mM Acetate/mM Lactate) and $Y_{\mathrm{bu,la}}$ (mM Butyrate/mM Lactate) represent the molar change of

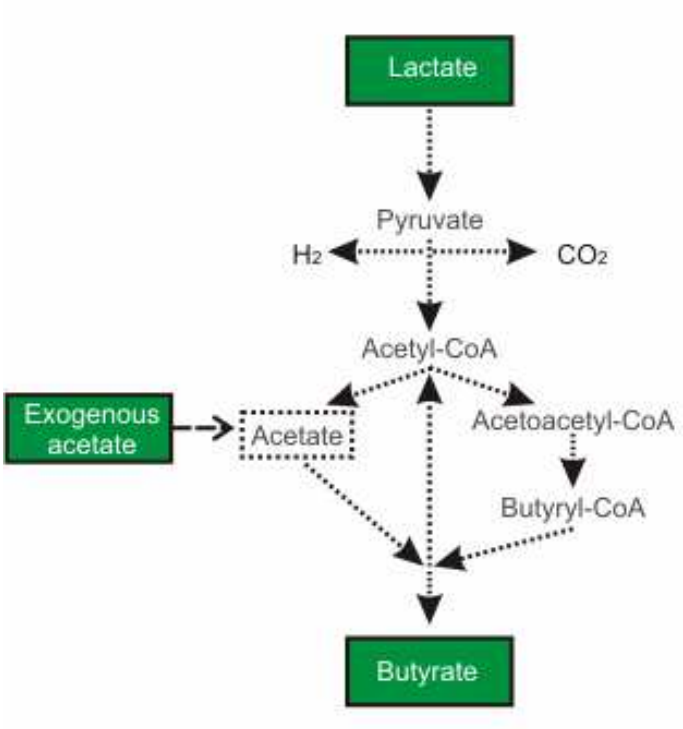

Figure 5.7: Simplified scheme of the reaction pathway for butyrate production; pyruvate is the central intermediate of the conversion. It is oxidized to acetyl-CoA, that can be routed either into acetate or butyrate formation, by mediation of the enzymatic activities of acetate kinase and butyryl-CoA:acetate-CoA-transferase respectively. Adapted from Papoutsakis (1984) and Duncan et al. (2004b).

acetate and butyrate with respect to the moles of lactate that are consumed. These factors are related to the stoichiometry of the metabolic conversions. It is thus interesting to look at the reactions implied in the pathway to maintain balance of the elements participating in the conversion.

Based on the reactions for butyrate production (Papoutsakis, 1984) and lactate consumption (Costello et al., 1991), the next set of reactions are derived to group the intermediate steps of butyrate production

$$CH_3CHOHCOOH + H_2O \Rightarrow CH_3COOH + 2H_2 + CO_2 \quad \text{(a)}$$

$$CH_3CHOHCOOH + CH_3COOH \Rightarrow CH_3CH_2CH_2COOH + H_2O + CO_2 \quad \text{(b)}$$

In addition to the above reactions, we consider the following anabolic reaction

$$5CH_3CHOHCOOH + 3NH_3 \Rightarrow 3C_5H_7O_2N + 9H_2O \quad \text{(c)},$$

which allows for the calculation of biomass to be produced. Reaction (c) is based on the work of Costello et al. (1991). It expresses lactate contribution to bacterial growth. The empirical formula $C_5H_7O_2N$ is used to represent biomass as proposed in Henze et al. (1987); Batstone et al. (2002).

We can use the above metabolic reactions to define the yield factors in the kinetic model. Consider $(1 - f_{\text{la}})$ the fraction of lactate that is used for reaction (c), in which biomass is produced. Denote by η_a the part of the fraction f_{la} of lactate that is converted in reaction (a). η_b applies for the reaction (b), such that $\eta_a + \eta_b = 1$. By means of the stoichiometry, we get

$$Y_{\text{la}} = \alpha \frac{3}{5}(1 - f_{\text{la}}), \tag{5.51}$$

$$Y_{\text{ac,la}} = f_{\text{la}}(\eta_a - \eta_b), \tag{5.52}$$

$$Y_{\text{bu,la}} = f_{\text{la}}\eta_b, \tag{5.53}$$

where α is a conversion factor between optical density and mM. This procedure is further recalled in Section 6.1.1,

Note that $Y_{\text{ac,la}}$ is a net factor representing the difference between production and consumption of acetate . Thus, if $Y_{\text{ac,la}}$ is positive, this means that lactate is routed to reaction (a) in higher proportion than to reaction (b).

Another application of setting the reactions implied in the conversion is to establish stoichiometric relationships between the components.
Consider the following reaction

$$|\nu_1|\text{A} + |\nu_2|\text{B} \Rightarrow |\nu_3|\text{C} + |\nu_4|\text{D},$$

with $|\nu_j|$ the stoichiometric coefficient of the component j. The stoichiometric number ν_j is negative when the component j is a reactant, and is positive when the component j is a product.

The changes of moles for the components participating of the reaction are related to their stoichiometric numbers as follows

$$\frac{\mathrm{d}n_{\mathrm{A}}}{\nu_1} = \frac{\mathrm{d}n_{\mathrm{B}}}{\nu_2} = \frac{\mathrm{d}n_{\mathrm{C}}}{\nu_3} = \frac{\mathrm{d}n_{\mathrm{D}}}{\nu_4} = \mathrm{d}\lambda,$$

where $\mathrm{d}n_j$ is the change in the number of moles of the component j. The term $\mathrm{d}\lambda$ characterizes the differential change of moles for all the components. λ is called the reaction coordinate (see, *e.g.*, Smith et al. (2005)).

For multiple reactions, it is satisfied that

$$\mathrm{d}n_j = \sum_{i=1}^{N} \nu_{j,i}\mathrm{d}\lambda_i,\ i = 1,\ldots,N, \tag{5.54}$$

where $\nu_{j,i}$ is the stoichiometric number of the component j in the reaction i, λ_i the reaction coordinate for the reaction i, and N the number of reactions. It is assumed that at the initial state, there is no reaction and thus $\lambda_i(0) = 0$. Equation (5.54) is then transformed to

$$\mathrm{d}n_j = \sum_{i=1}^{N} \nu_{j,i}\lambda_i,\ i = 1,\ldots,N. \tag{5.55}$$

Assuming that the reaction coordinate of the reaction for biomass production is much lower than the reaction coordinates for reactions (a) and (b) (which is usually the case for anaerobic systems), and applying (5.55) for lactate, acetate and butyrate, we get

$$\mathrm{d}n_{\mathrm{la}} = -\lambda_{\mathrm{a}} - \lambda_{\mathrm{b}}, \tag{5.56}$$

$$\mathrm{d}n_{\mathrm{ac}} = \lambda_{\mathrm{a}} - \lambda_{\mathrm{b}}, \tag{5.57}$$

$$\mathrm{d}n_{\mathrm{bu}} = \lambda_{\mathrm{b}}, \tag{5.58}$$

so,

$$\mathrm{d}n_{\mathrm{bu}} \approx -\frac{\mathrm{d}n_{\mathrm{la}} + \mathrm{d}n_{\mathrm{ac}}}{2}, \tag{5.59}$$

which gives an approximative relationship between the butyrate produced and the lactate and acetate utilized.

The mathematical model does not take into account biomass decay, and no distinction between D-lactate and L-lactate is attempted. For *E. hallii*, the sum of the two isomers is considered as a sole substrate. In the case of strain SS2/1, D-lactate is the single substrate, because this strain was not able to utilize L-lactate.

5.2.2 Parameter estimation and model selection

Experimental data formerly obtained (Duncan et al., 2004b) were analyzed to estimate the parameters of our mathematical model. The published data are means of triplicate cultures experiments with pure cultures of *E. hallii* and strain SS2/1. Bacterial strains were grown in YCFA medium with 45 mM DL-lactate, 10 mM glucose or DL-lactate plus glucose as added carbon sources in Hungate tubes under anaerobic conditions. Fermentation was monitored by synchronous measurements of the optical density (OD_{650}) and concentrations of glucose, lactate (DL and L), acetate and butyrate. Concentrations of L-lactate and glucose were determined by enzymatic methods. Concentrations of SCFA and DL-lactate were determined by capillary gas chromatography. D-lactate was obtained by difference between total lactate and L-lactate concentrations.

Parameter estimation focused on *E. hallii* and strain SS21/1 growth experiments in DL-lactate containing medium. Data in the lag phase were not considered.
Parameter estimation of the Monod, quadratic and first-order kinetic models was performed by using the toolbox IDEAS. The cost function to be optimized was derived from the assumption that measurement errors followed a normal distribution with unknown and diagonal covariance matrix (see Eq. (4.20), page 102), here recalled

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{n_\mathrm{y}} \frac{n_\mathrm{t}}{2} \ln \left[\sum_{\mathrm{i}=1}^{n_\mathrm{t}} [y_k(t_{i_k}) - y_{\mathrm{m}_k}(t_{i_k}, \boldsymbol{\theta})]^2 \right].$$

In section 4.1.1, we mentioned Akaike's information criterion (AIC) (Akaike, 1974) as a broadly used criterion for model selection. Denote by $\mathscr{M}_i(\boldsymbol{\theta}^i)$ a fixed model structure. The estimated information criterion AIC is defined by

$$AIC(\mathscr{M}_i, \widehat{\boldsymbol{\theta}}^i) = \frac{1}{n_\mathrm{t}} \left[-\ln \pi_y(\mathbf{y}|\widehat{\boldsymbol{\theta}}^i, \widehat{\boldsymbol{\Sigma}}^i) + \dim(\widehat{\boldsymbol{\theta}}^i) \right], \tag{5.60}$$

where $\widehat{\boldsymbol{\theta}}^i, \widehat{\boldsymbol{\Sigma}}^i$ are the optimal values for $\boldsymbol{\theta}^i$ and the covariance matrix $\boldsymbol{\Sigma}^i$.
The log-likelihood is computed by (4.14) (page 101) at $\widehat{\boldsymbol{\theta}}^i, \widehat{\boldsymbol{\Sigma}}^i$

$$-\ln \pi_y(\mathbf{y}|\widehat{\boldsymbol{\theta}}^i, \widehat{\boldsymbol{\Sigma}}^i) = \frac{n_\mathrm{y} n_\mathrm{t}}{2} \ln 2\pi + \frac{n_\mathrm{t}}{2} \ln \det \widehat{\boldsymbol{\Sigma}}^i + \frac{1}{2} \sum_{j=1}^{n_\mathrm{t}} [\mathbf{y}(t_j) - \mathbf{y}_\mathrm{m}(t_j, \widehat{\boldsymbol{\theta}}^i)]^\mathrm{T} (\widehat{\boldsymbol{\Sigma}}^i)^{-1} [\mathbf{y}(t_j) - \mathbf{y}_\mathrm{m}(t_j, \widehat{\boldsymbol{\theta}}^i)].$$

When there are several competing models, the best model in Akaike's information criterion sense is the one with the smallest AIC. After parameter estimation, Eq. (5.60) was

Table 5.4: Comparison of model candidates.

Model	$-\ln \pi_y(\mathbf{y}\|\widehat{\boldsymbol{\theta}}^i, \widehat{\Sigma}^i)$	$AIC(\mathscr{M}_i, \widehat{\boldsymbol{\theta}}^i)$
Monod kinetics	39.762	4.974
Quadratic kinetics	25.182	3.242
First-order kinetics	55.451	6.606

evaluated for each model candidate.

5.2.3 Results and discussion

The construction of the mathematical model describing the kinetics of butyrate production by *E. hallii* and strain SS2/1 is inspired by standard models of anaerobic reactors. We adapted such models to the specificities of human colonic bacteria. For instance, lactate contribution to butyrate is not considered in mathematical models for anaerobic reactors (Costello et al. (1991), Skiadas et al. (2000)), whereas lactate is known to be an important precursor of butyrate production in the human colon.

5.2.3.1 Selecting the best model

Table 5.4 compares the candidate models to represent the experimental data for *E. hallii*. The model with quadratic kinetics is the best one with respect to Akaike's criterion, and also provides the smallest value of the optimized cost function. The model with quadratic

Table 5.5: Estimates of the parameters for the quadratic kinetic model with their approximate 95% confidence intervals.

	$k_{m,\mathrm{la}}$	f_{la}	η_{la}	α
E. hallii	1.307 ± 0.089	0.844 ± 0.097	0.222 ± 0.072	0.140 ± 0.087
SS2/1	1.894 ± 0.249	0.562 ± 0.231	0.005 ± 0.375	0.060 ± 0.032

kinetics adequately represents the *in vitro* experimental data of both strains (Fig. 5.8). Data for strain SS2/1 is more scattered, thus the fit is less satisfactory than for *E. hallii*. A

discrepancy for acetate concentration is observed in the experiments for both strains. In the case of *E. hallii*, this mismatch occurs for the first three sampling times after the initial condition. According to the model, acetate concentration will either always increase or always decrease. It is therefore impossible to account for the first four data points if the decreasing trend observed on the remaining data is consistent. This transient mismatch may be due to extra carbon sources in the medium, such as yeast extract, not accounted for by the model. This argument is also valid for strain SS2/1. For *E. hallii*, the mathematical model slightly overestimates the biomass concentration at the final point of the experiment. Apart from measurement uncertainty, this could be explained by the fact that the kinetic model considers that biological activity is growth-associated. It is known that bacteria can remain metabolically active without cell division. Thus, the optical density will not always be a good indicator for bacterial activity.

The estimates of the parameters are shown in Table 5.5 with their approximate 95% confidence intervals, calculated from Eq. (4.34), page 106. The confidence intervals are remarkably narrow for *E. hallii* and less satisfactory for strain SS2/1, specially for η_{la}. For strain SS2/1, it appears that reaction (a) can be neglected to represent biochemical conversions, explaining the large standard deviation for η_{la}. Keeping in mind that the data for strain SS2/1 is very scattered, large variance of η_{la} may be due to measurement uncertainty. The yield factors of the quadratic kinetic model are computed by using (5.51)-(5.53). Their values are given in Table 5.6.

Table 5.6: Yield factors for the quadratic kinetic model.

	Y_{la}	$Y_{\text{ac,la}}$	$Y_{\text{bu,la}}$
E. hallii	0.013	-0.470	0.657
SS2/1	0.016	-0.556	0.559

We also assessed the accuracy of parameters for the Monod-kinetic based model. Results are shown in Table 5.7. While the standard deviations of the parameters f_{la}, η_{la}, α have the same order of magnitude that those for the quadratic kinetic model, $k_{m,\text{la}}$ and the

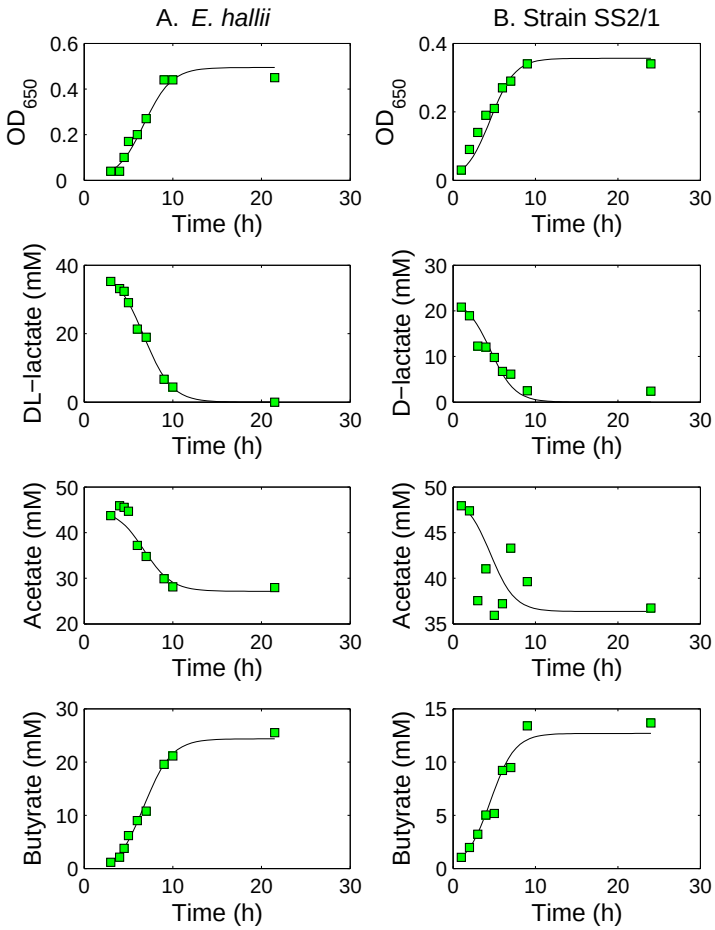

Figure 5.8: Comparison of the model outputs (solid lines) with the experimental data (squares).

affinity constant $K_{s,\mathrm{la}}$ present high standard deviations in the Monod-based model. Difficulties in attaining good accuracy for estimate the Monod affinity constant may relate to sampling times, limited amount of data available, and their poor informative content.

5.2.3.2 Reactions stoichiometry

We integrated reactions stoichiometry with our kinetic model. The equations resulting from the simplified set of reactions for butyrate production has enabled us to present stoichiometric relationships between metabolites. For instance, the number of moles of

Table 5.7: Estimates of the parameters for Monod kinetic model with their approximate 95% confidence intervals.

	$K_{s,la}$	$k_{m,la}$	f_{la}	η_{la}	α
E. hallii	6.626 ± 10.168	43.998 ± 14.331	0.791 ± 0.123	0.219 ± 0.075	0.098 ± 0.059
SS2/1	7.722 ± 16.099	42.632 ± 29.683	0.574 ± 0.249	0.035 ± 0.390	0.060 ± 0.036

butyrate formed is half the number of reacting moles of lactate plus acetate. This relationship was indeed verified in the experimental data. In addition, the reaction equations explain why acetate could not be utilized when lactate was depleted in the medium. This phenomenon was also observed for other butyrate-producing bacteria grown in glucose or fructose, where the conversion of acetate and bacterial growth ceased with the disappearance of glucose or fructose (Duncan et al. (2004a); Falony et al. (2006)). Acetate is not an energy source for the bacteria in this process and its utilization is regulated by the availability of other carbon sources, such as fructose or here lactate.

5.2.3.3 Model limitations and applications

The model adequately describes the kinetics for butyrate production by *E. hallii* and strain SS2/1. The parameters were estimated from *in vitro* data obtained under a specific set of conditions and these could change for experiments carried out under other conditions. For instance, lactate consumption rate and SCFA production are known to depend on hydrogen partial pressure (Costello et al. (1991); Macfarlane and Macfarlane (2003)). Belenguer et al. (2007) showed that lactate utilization and fermentation products formed by human fecal microbiota depended on pH. Amaretti et al. (2007) reported that the estimates of kinetic parameters can also change with pH conditions. The model presented here does not take into account the influence of pH or of hydrogen partial pressure on the kinetics.

Environmental conditions impose thermodynamic limitations on the spectrum of fermentation products. The Gibbs free energy change ($\Delta G'$) of the reactions is the driving force that determines the concentration profile of the metabolites. Thus, initial concentration

and accumulation of metabolites will strongly affect the progress of the reactions (Rodríguez et al. (2006a); Madigan et al. (2009)). *In vitro experiments* with the ruminal butyrate-producing bacterium *M. elsdenii* showed the impact of the initial concentration of acetate on butyrate metabolism. Under low initial concentrations of extracellular acetate (< 4mM), the production rate of acetate was higher than its consumption rate. When the initial concentration increased, however, production decreased and consumption increased, concomitant with an increase in butyrate production (Hino et al., 1991). For the experimental data analyzed here, the acetate consumption rate was higher than its production rate. In the kinetic model, the estimated value of η_{la} was lower than 0.5, which implies a negative value of the yield factor $Y_{\mathrm{ac,la}}$. Under other experimental conditions, where acetate production is higher than its consumption, the model structure advocated here will remain identical, but η_{la} will be higher than 0.5 and $Y_{\mathrm{ac,la}}$ will become positive. This shift in the progress of the reaction may be handle mathematically by using thermodynamic-based models as proposed in Rodríguez et al. (2008).

The mathematical model presented here focuses on *E. hallii* and strain SS2/1. These two strains differ in their affinity for lactate enantiomers. While *E. hallii* was able to use both D and L-lactate, strain SS2/1 only utilized D-lactate. The model does account for these differences. In addition, the mathematical model should not be restricted to *E. hallii* or strain SS2/1 as it could be applied to other phylogenetically distinct bacterial groups that share enzymatic reactions. Kinetic rates may differ between species as demonstrated for the two strains analyzed here, but the model structure will be identical. Changing the key carbon source from, for example, lactate to glucose would allow the model to be employed in studying other key butyrate producing bacteria that are unable to utilize lactate, such as *R. intestinalis* and *F. prausnitzii*.

The aggregated model structure could also be applied for butyrate producing strains regardless of their different enzymatic activities. For example, *in vitro studies* with *R. intestinalis* DSM 14610 and *A. caccae* DSM 14662 with fructose as the main carbon source showed that exogenous acetate stimulates bacterial growth in both strains. Acetate however was essential for the growth of *R. intestinalis*, but not for *A. caccae* (Falony et al., 2006). In this case, endogenous acetate production compensates for the lack of exogenous acetate. This could be explained by a higher acetate kinase activity, such as shown for the

strain *A. caccae* L1-92 compared to *R. intestinalis* related strains (Louis et al., 2004).

Integration of the results presented here into the complete model may enables one to assess strategies aimed at reducing colonic L-lactate concentrations and enhance butyrate production.

5.3 Conclusions

In this chapter, we presented the parameter estimation of two submodels of the full reaction pathway of carbohydrate fermentation performed by the human colonic microbiota. We tackled such a task with a systematic methodology, that can be extended to other reaction steps of the trophic chain. The toolbox IDEAS was instrumental to detect practical identifiability problems. For the kinetic model of butyrate production, it also enabled us to easily test various model structures. Quadratic kinetics were selected to describe microbial growth rates. The resulting models described satisfactorily *in vitro* experimental data and provided a better accuracy in the estimates than Monod-based models.

In vitro experiments have shown transfer of hydrogen and acetate between butyrate producing bacteria and hydrogen utilizers (Chassard and Bernalier-Donadille, 2006). It would be interesting to reproduce such experiments in order to integrate the resulting data into a model that couples the two kinetic models studied here. Such a model would allow valuable cross-feeding interactions to be studied. In addition, we could assess the role of hydrogen in the metabolic process.

The application of the singular perturbation method on the model of homoacetogenesis is a preliminary step towards model order reduction of the complete model. Such an approach will enable us to reduce simulation time, which may be time consuming for the parameter estimation of the full model.

The Monod equation, which is derived from the Michaelis-Menten equation (1913), is the most widely used expression to describe microbial growth rate. However, parameter estimation of Monod based-models is known to be hampered by practical identifiability problems (see, *e.g.*, Holmberg (1982); Vanrolleghem et al. (1995); Kesavan and Law

(2005)), as we also showed in the present study. Other kinetics expressions with less parameters can be used to avoid those difficulties. For instance, several authors (Jeppsson and Olsson (1993); Castillo et al. (1999); Soto-Cruz et al. (2002); Kong et al. (2006)) have showed that simple equations such as first-order kinetics can be effective to represent biological activity. Holmberg (1982) suggests that the practical identifiability obstacles of Monod based models should not limit its applicability, in virtue of its excellent explanation of bacterial growth. In this direction, Monod kinetics has been recommended for comprehensive models of biological processes (Batstone, 2006). In the complete model structure, Monod kinetics is kept to represent microbial activity.

In our modelling studies, more informative data would have helped improving the estimation of Monod parameters. Selection of experimental conditions is not a simple task, and, as mentioned in Section 4.3.3, it is an area of research. It would be interesting to identify the degrees of freedom in the procedure for data collection of the two submodels here studied. This information could be used in future studies on optimal experiment design. We think that the toolbox IDEAS could be exploited in such a task.

The parameter estimation performed here provides useful information for building prior knowledge to tackle the estimation problem for the complete model. Currently, the parameter estimates of the two kinetic models, together with values extracted from the literature, are used to set the parameters of the complete model. In the future, this information could be integrated in a Bayesian estimation framework, when data from *in vivo* experiments become available.

We analyzed here the whole by its parts. In the following chapter, we show the methodology used to reduce the dimensions of the parameter and state vectors of the complete model.

Chapter 6

Reducing model complexity

A theory should be made as simple as possible,
but no simpler.
Albert Einstein[1]

The first part of this chapter describes the integration of knowledge to reduce the dimension of the parameter vector in our model. In the second part, we show preliminary results on model order reduction of the complete model by the singular perturbation method.

6.1 Reducing the dimension of the parameter vector

In order to reduce the number of unknown parameters, the following is assumed: (i) all compartments share the same Henry law constant, (ii) $k_L a$ takes the same value for the three gases in all compartments; (iii) decay constants, adherence and detachments coefficients are the same for all microbial groups in all compartments, (iii) kinetics parameters for process j are the same in all compartments, (iv) the yield factor $Y_{i,j}$ is the same in all compartments. In addition, the yield factors are set to guarantee carbon and hydrogen balances as proposed by Batstone et al. (2002). This procedure is explained below.

[1]This quotation is often attributed to Einstein. However, no one has found yet its original source. The quotation may be a paraphrase of Einstein's statements about simplicity (Caprice, 2000).

6.1.1 Determination of yield factors based on stoichiometry

The reactions considered in the model are illustrated in Table 6.1. Relationships between the yield factors can be obtained from stoichiometry (see, *e.g.*, Batstone et al. (2002); Kleerebezem and van Loosdrecht (2006)).

Consider, for example, the reactions implied in glucose utilization. Glucose is used in catabolic processes R_1-R_4, and the anabolic process R_5 for bacterial growth. Let f_{su} denote the fraction of glucose used for catabolism. From R_5, the yield of biomass Y_{su} is then

$$Y_{su} = \frac{6}{5}(1 - f_{su}).$$

Glucose used in the catabolism is degraded via R_1-R_4. The fraction that is converted in reaction R_k is denoted by η_k ($k = 1,2,3,4$). For R_4, it is clear that $\eta_4 = 1-(\eta_1+\eta_2+\eta_3)$. In a reacting system, η_k can change as a function of the thermodynamic conditions. We take them as constants.

The yield factors for the components can be expressed as functions of f_{su}, η_1, η_2 and η_3, as

$$\begin{aligned}
Y_{la,su} &= 2 f_{su}\eta_1,\\
Y_{ac,su} &= 2 f_{su}\eta_2 + \frac{2}{3} f_{su}\eta_3,\\
Y_{pro,su} &= \frac{4}{3} f_{su}\eta_3,\\
Y_{bu,su} &= f_{su}\eta_4,\\
Y_{H_2,su} &= 4 f_{su}\eta_2 + 2 f_{su}\eta_4,\\
Y_{CO_2,su} &= 2 f_{su}\eta_2 + \frac{2}{3} f_{su}\eta_3 + 2 f_{su}\eta_4,\\
Y_{H_2O,su} &= -2 f_{su}\eta_2 + \frac{2}{3} f_{su}\eta_3 + \frac{18}{5}(1 - f_{su}).
\end{aligned}$$

In addition to maintaining elementary balances, this approach enables us to reduce the number of parameters. For example, without these considerations, the number of parameters to be estimated for glucose utilization would be 8. Using the stoichiometry, it is reduced to 4. This procedure is used for all substrates.

Table 6.1: Stoichiometric reactions used in the model. The biomass is assumed to have the empirical formula $C_5H_7O_2N$ (Batstone et al., 2002).

Glucose utilization	
$C_6H_{12}O_6 \Rightarrow 2CH_3CHOHCOOH$	(R_1)
$C_6H_{12}O_6 + 2H_2O \Rightarrow 2CH_3COOH + 2CO_2 + 4H_2$	(R_2)
$3C_6H_{12}O_6 \Rightarrow 4CH_3CH_2COOH + 2CH_3COOH + 2CO_2 + 2H_2O$	(R_3)
$C_6H_{12}O_6 \Rightarrow CH_3CH_2CH_2COOH + 2CO_2 + 2H_2$	(R_4)
$5C_6H_{12}O_6 + 6NH_3 \Rightarrow 6C_5H_7O_2N + 18H_2O$	(R_5)
Lactate utilization	
$3CH_3CHOHCOOH \Rightarrow 2CH_3CH_2COOH + CH_3COOH + CO_2 + H_2O$	(R_6)
$2CH_3CHOHCOOH \Rightarrow CH_3CH_2CH_2COOH + 2H_2 + 2CO_2$	(R_7)
$5CH_3CHOHCOOH + 3NH_3 \Rightarrow 3C_5H_7O_2N + 9H_2O$	(R_8)
Hydrogen utilization: homoacetogenesis reaction	
$4H_2 + 2CO_2 \Rightarrow CH_3COOH + 2H_2O$	(R_9)
$10H_2 + 5CO_2 + NH_3 \Rightarrow C_5H_7O_2N + 8H_2O$	(R_{10})
Hydrogen utilization: methanogenesis reaction	
$4H_2 + CO_2 \Rightarrow CH_4 + 2H_2O$	(R_{11})
$10H_2 + 5CO_2 + NH_3 \Rightarrow C_5H_7O_2N + 8H_2O$	(R_{12})

Hydrogen has to be considered carefully, as it participates as substrate in two reactions. In our model we defined a total fraction f_{H_2}, which is split in the reactions of homoacetogenesis and methanogenesis in the proporpotions f_{H_2a}, f_{H_2m}, such that $f_{H_2} = f_{H_2a} + f_{H_2m}$.

Taking into account the above assumptions, the number of the model parameters is reduced from 333 to 60 (Table 6.2).

Table 6.2: Reduction of model parameters.

Parameters	**Lumen subsystem**	**Mucus subsystem**	**In each physiological region**	**Whole system**
Reactions				
Hydrolysis				
Contois parameters	2	2	2	2
Yield coefficients	1	1	1	1
Glucose utilization				
Monod parameters	2	2	2	2
Yield coefficients	4	4	4	4
Lactate utilization				
Monod parameters	2	2	2	2
Yield coefficients	2	2	2	2
Hydrogen utilization				
Homoacetogenesis				
Monod parameters	2	2	2	2
Yield coefficients	1	1	1	1
Methanogenesis				
Monod parameters	2	2	2	2
Yield coefficients	1	1	1	1
Decay rate				
Decay constants	1	1	1	1
Subtotal	20	20	20	20

Reduction of model parameters (continued).

Parameters	**Lumen subsystem**	**Mucus subsystem**	**In each physiological region**	**Whole system**
Transport				
Glucose	1		1	3
Lactate	1	1	2	6
Acetate	1	1	2	6
Propionate	1	1	2	6
Butyrate	1	1	2	6
Water	1	1	2	6
Microbial phenomena				
Shear constants		1	1	1
Adherence constants	1		1	1
Additional residence time	1		1	1
Liquid-gas transfer				
H_2	1	1	1	1
CH_4	1	1	1	1
CO_2	1	1	1	1
$k_L a$	1	1	1	1
Subtotal	12	10	18	40
Total	32	30	38	**60**

6.2 Reducing the dimension of the state vector

The previous result of model order reduction for the submodel of homoacetogenesis (see Section 5.1.3) motivated us to extend the singular perturbation approach to the complete model of carbohydrate degradation.

Consider, for example, the model associated with the lumen in the proximal colon section, defined by

$$\dot{s}_i^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} s_{i,\mathrm{in}}^{\mathrm{l}} - \frac{q_{\mathrm{out}}}{V^{\mathrm{l}}} s_i^{\mathrm{l}} - \gamma_i^{\mathrm{l}} s_i^{\mathrm{l}} + \sum_{j=1}^{5} Y_{i,j}^{\mathrm{l}} \rho_j(\boldsymbol{\xi}^{\mathrm{l}}) - Q_i^{\mathrm{l}}, \tag{6.1}$$

$$\dot{z}^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} z_{\mathrm{in}}^{\mathrm{l}} + \frac{q_{\mathrm{out}}^{\mathrm{m}}}{V^{\mathrm{l}}} z^{\mathrm{m}} - \frac{q_{\mathrm{out}}}{V^{\mathrm{l}}} z^{\mathrm{l}} - \rho_1(\boldsymbol{\xi}^{\mathrm{l}}), \tag{6.2}$$

$$\dot{x}_i^{\mathrm{l}} = \frac{q_{\mathrm{in}}}{V^{\mathrm{l}}} x_{i,\mathrm{in}}^{\mathrm{l}} - \frac{1}{\tau_i + V^{\mathrm{l}}/q_{\mathrm{out}}} x_i^{\mathrm{l}} + b_i \frac{V^{\mathrm{m}}}{V^{\mathrm{l}}} x_i^{\mathrm{m}} - a_i x_i^{\mathrm{l}} + \sum_{j=2}^{9} Y_{i,j}^{\mathrm{l}} \rho_j(\boldsymbol{\xi}^{\mathrm{l}}), \tag{6.3}$$

$$\dot{s}_{\mathrm{g},i} = \frac{q_{\mathrm{g,in}}}{V_{\mathrm{g}}} s_{\mathrm{g},i,\mathrm{in}} - \frac{q_{\mathrm{g,out}}}{V_{\mathrm{g}}} s_{\mathrm{g},i} + Q_i \frac{V}{V_{\mathrm{g}}}. \tag{6.4}$$

The scaling factors and the order of magnitude of the parameters are again deduced from those obtained from *in vitro* experimental data or drawn from the literature. Equations (6.1)-(6.4) also share the generic form (5.12)

$$\dot{x}_i = d_i(\boldsymbol{x}) x_i + p_i(\tilde{x}_i),$$

which allows us to bound the local time constants. By applying the same procedure as in the homoacetogenesis model, we obtain a two-scale model in the standard singular perturbed form. The vector of fast variables $\mathbf{z}$ is

$$\mathbf{z} = \left[s_{\mathrm{H_2}},\, s_{\mathrm{CH_4}},\, s_{\mathrm{CO_2}} \right]^{\mathrm{T}}.$$

Denote by $\mathbf{x}$ the vector of slow variables, and consider for instance the dynamics of $s_{\mathrm{H_2}}$, expressed in the slow time scale. It can be written as

$$\begin{aligned} \frac{1}{\tau_s k_L a} \dot{s}_{\mathrm{H_2}} = \mathbf{k}(\mathbf{x}) - \left(\frac{q_{\mathrm{out}}(\mathbf{x}) + \gamma_{\mathrm{H_2}}^{l}}{k_L a} + 1 \right) s_{\mathrm{H_2}} - \mu_{\max,\mathrm{H_2a}} \frac{x_{\mathrm{H_2}a}}{K_{s,\mathrm{H_2a}} + s_{\mathrm{H_2}}} s_{\mathrm{H_2}} \\ - \mu_{\max,\mathrm{H_2m}} \frac{x_{\mathrm{H_2}m}}{K_{s,\mathrm{H_2m}} + s_{\mathrm{H_2}}} s_{\mathrm{H_2}}, \end{aligned} \tag{6.5}$$

where $\mathbf{k}(\mathbf{x})$ is a function of the slow variables. Consequently, the quasi steady state value for the concentration in the liquid phase, $l(\mathbf{x})$, is the unique positive solution of the third-degree polynomial equation that can be deduced from

$$\mathbf{k}(\mathbf{x}) - \left(\frac{q_{\text{out}}(\mathbf{x}) + \gamma^{l}_{\text{H}_2}}{k_L a} + 1\right) l(\mathbf{x}) - \mu_{\max,\text{H}_2\text{a}} \frac{x_{\text{H}_2 a}}{K_{s,\text{H}_2\text{a}} + l(\mathbf{x})} l(\mathbf{x}) - \mu_{\max,\text{H}_2\text{m}} \frac{x_{\text{H}_2\text{m}}}{K_{s,\text{H}_2\text{m}} + l(\mathbf{x})} l(\mathbf{x}) = 0. \tag{6.6}$$

As we did for the homoacetogenesis model, we obtain a boundary-layer equation of the form

$$\frac{dy}{d\upsilon} = -\left(\frac{K_1}{K_2 + y} + \frac{K_3}{K_4 + y} + K_5\right) y, \tag{6.7}$$

where the coefficient K_i is a positive constant in the fast time scale. The origin of the boundary layer model is easily shown to be exponentially stable. The other hypotheses of the Tikhonov theorem are also satisfied. The computation for the other fast variables are similar, so the model can be simplified.

This result is very convenient on many aspects. First, concentrations of hydrogen, methane, and carbon dioxide are difficult to measure, and their initial conditions are usually unknown. By applying the singular perturbation approach, we are no longer interested in estimating the initial conditions of such state variables. In addition, model order reduction leads to savings in simulation time and contributes to simulation stability.

We illustrated, by means of simulations, the closeness of the behavior of the reduced model to that of the complete model for the lumen microhabitat in the proximal section. As observed for the homoacetogenesis model, when initial conditions in the complete model are in the acceptable physical range, the response of the reduced model is superimposed to the response of the complete model. We set the initial conditions of hydrogen, methane and carbon dioxide in the liquid phase to unrealistically high values to observe differences between the two models. The result is illustrated in Fig. 6.1, which confirms the validity of the reduction.

Figure 6.2 compares the responses for the complete (continuous curves) and reduced (dashed curves) models of the lumen compartment in the distal section. Model-order reduction seems to improve the stability of the simulations. However, further investiga-

tion in this issue is needed.

Model reduction allows a saving of at least 30% in computational time, because the stiffness of the model is reduced. Applying this approach to the mucus and lumen compartments in the complete model reduces the number of state variables from 102 to 84.

6.3 Conclusions

Under the assumptions that some parameters of the model take the same value in all compartments, and by taking stoichiometry into account, we substantially reduced the dimension of the parameter vector of our model. However, we should keep in mind that the reduced parameter vector may not be sufficient to represent experimental data and thus its reformulation might be needed. To provide a more conclusive argument, it would be necessary to perform sensitivity analysis to evaluate to what extent the parameters affect the system behavior, and to confirm the validity of these assumptions. Such an analysis is an important step to be performed in future work. Given the complexity of the model, the local sensitivity approach described in Section 4.4 may not be suitable and more sophisticated methods should be needed (see, *e.g.*, the methods described in Saltelli et al. (2000)).

Results on model-order reduction for the homoacetogenesis model were extended to the complete model. Theoretical analysis and simulations results confirmed the validity of the singular perturbation approach. Model-order reduction seemed to improve the stability of the simulations, and allowed savings of computational time.

Up to now, we have presented the methodology to construct our model and to define its parameters. In the next chapter, we show our first results on simulating the model under various nutritional scenarios.

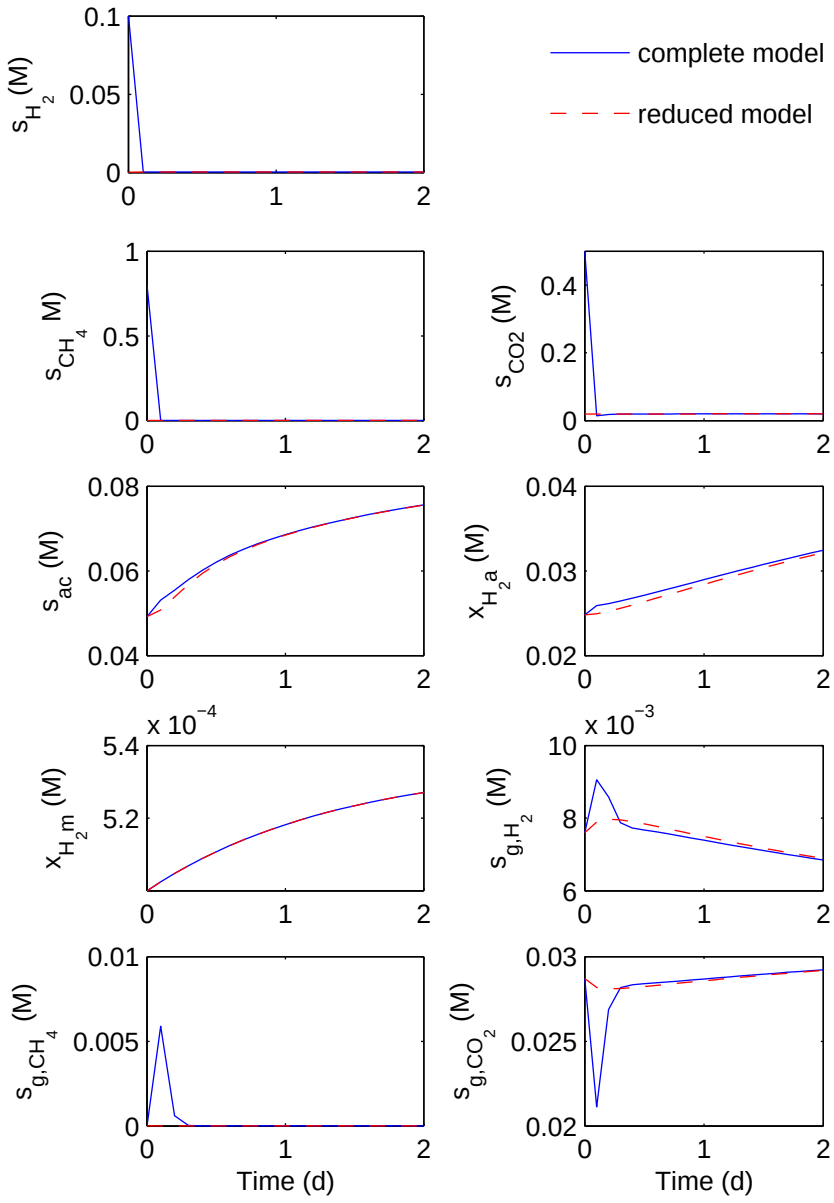

Figure 6.1: Comparison of the complete and reduced model for the full fermentation pathway in the lumen of the proximal colon. Discrepancy in initial conditions has been exaggerated for the sake of illustration.

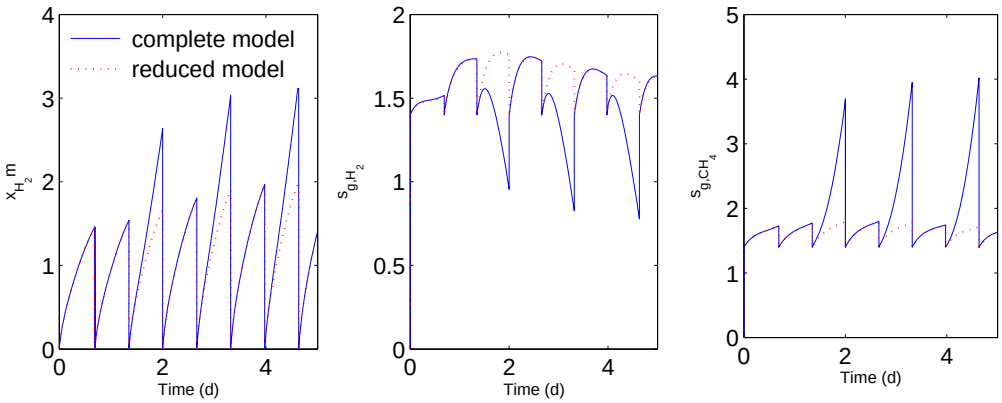

Figure 6.2: Comparison of the numerical stability for the simulations of the full (continuous curves) and reduced model (dashed curves) in the lumen of the distal colon. The oscillatory behavior results from the periodic reset of the volume due to excretion. Model-order reduction seems to improve the stability of the simulations.

Chapter 7

The mathematical model as a virtual experimental platform

Il n'existe pas une catégorie de sciences auxquelles on puisse donner le nom de sciences appliquées. Il y a la science et les applications de la science, liées entre elles comme le fruit à l'arbre qui l'a porté.

Louis Pasteur

This chapter is dedicated to illustrate the potential use of our mathematical model in the analysis of the human colonic ecosystem. The simulations presented here address specific questions. Some of them aim to assess the plausibility of the model on the basis of existing knowledge, while others attempt to study interactions and phenomena that are difficult to evaluate through experimental work. The mathematical model is implemented in Simulink/Matlab (Fig. 7.1). Simulink provides a graphical interface that facilitates the communication with biologists. The ordinary differential equation solver ODE15s was selected, because it is well adapted to stiff models. It is based on an implicit method proposed by Klopfenstein (1971) (see Shampine and Reichelt (1997) for more details).

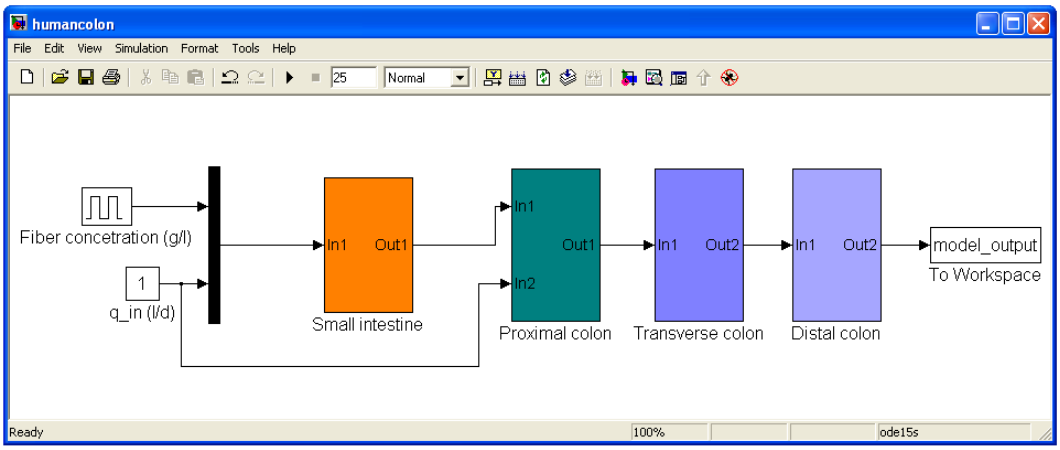

Figure 7.1: Model implementation in Simulink.

Simulation conditions

The input flow rate was set to 1.5 l/d. Fiber intake was set as a periodical pulse, with three meals per day of 30 min duration each. The amplitude of the pulse was calculated according to the desired fiber intake. Initial conditions were determined by simulating the model with a total periodic feed of 10 g of fiber per day until stationary state was reached in the proximal and transverse colon. Afterward, the stationary state variables were saved to be used as initial conditions for the evaluation of various scenarios. Mucins production was set as 5 g/d. Volume in the distal colon was initialized to 100 ml. The volume of one excretion was adjusted to 200 ml according to values reported in literature (Birkett et al. (1997); Topping and Clifton (2001)). Simulations were carried out on a Dell PowerEdge 1950 server (32 GB RAM). In average, each run took about 30 min of computer time. The distal colon is the most demanding section in terms of computational time. Numerical values of the model parameters used for the simulations are presented in Appendix A.

7.1 A dynamic and spatial picture

The purpose of this section is to illustrate the dynamics of the fermentation pattern along the physiological regions of the human colon and to assess the plausibility of our model. Fiber intake was taken as a periodical signal providing 25 g/d of fiber. Figures 7.2-7.8 dis-

play a dynamic and spatial picture of the fermentation pattern as predicted by the model.

The proximal colon appears to be the most active section. Most of the fiber is degraded there (Fig. 7.2). It presents the highest level of monosaccharides that are depleted in the following sections . Simulated SCFA concentrations turn out to be high in the proximal colon and fall progressively towards the distal colon (Fig. 7.3). The predicted molar ratio Acetate:Propionate:Butyrate in each luminal section were:

in the proximal colon 82:37:39,

in the transverse colon 74:35:37, and

in the distal colon 57:31:33.

These values are in agreement with *in vivo* studies (Cummings et al. (1987); Topping and Clifton (2001) and the references therein).

It is interesting to observe in Fig. 7.4 that homoacetogens and methanogens in the luminal microhabitats coexist. However, in the mucus the competition is more critical. In Fig. 7.8, it is observed that in the proximal colon, hydrogen utilizers are mainly homoacetogens, while for the transverse and distal colon methanogens predominate. Little is known about the dynamics of the mucus. The simulations presented here provide a view of such a microhabitat.

The predicted microbial concentration in the distal colon is about 70 g/l. If we assume a density for fecal matter of 1 g/ml and that 15% of fecal matter are solids, this implies that microorganisms represent 47% of the solids, which agrees with the experimental values reported in the literature (Moore and Holdeman (1974); Stephen and Cummings (1980)).

The responses of the concentration of methanogens (Fig. 7.4), and gas concentrations of H_2 and CO_2 in the distal colon (Fig. 7.5) illustrate some problems of graphical representation (note the peaks). This can be linked to the fast dynamics of the liquid-gas transfer phenomenon, which lead to multiple time scales.

Figure 7.9 summarizes the spatial and time variation of microbial and metabolite concentrations. While the profile of microbial concentration in the lumen is increasing from

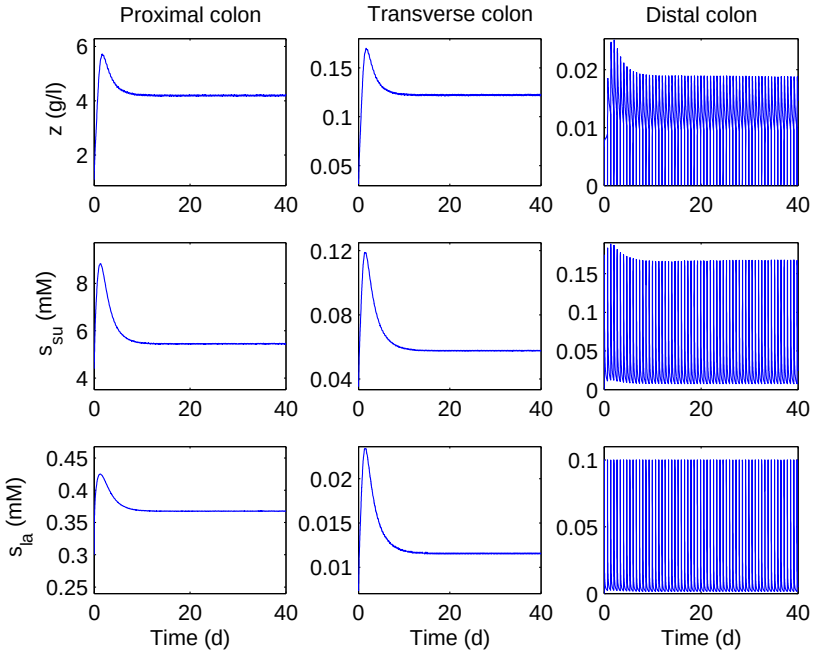

Figure 7.2: Dynamic fermentation pattern along the lumen compartments.

the proximal to the distal section (Fig. 7.9 C), the microbial concentration in the mucus presents an opposite behavior (Fig. 7.9 D). This may be explained by the sugar depletion in luminal compartments, which reduce substrate availability for diffusion to the mucus. In luminal compartments, the chyme looses water traveling along the colon and thus the microbial density increases. Figure 7.10 shows the frequency of excretion. The model predicts a number of 1.5 defecations per day, which is in the range reported in Birkett et al. (1997).

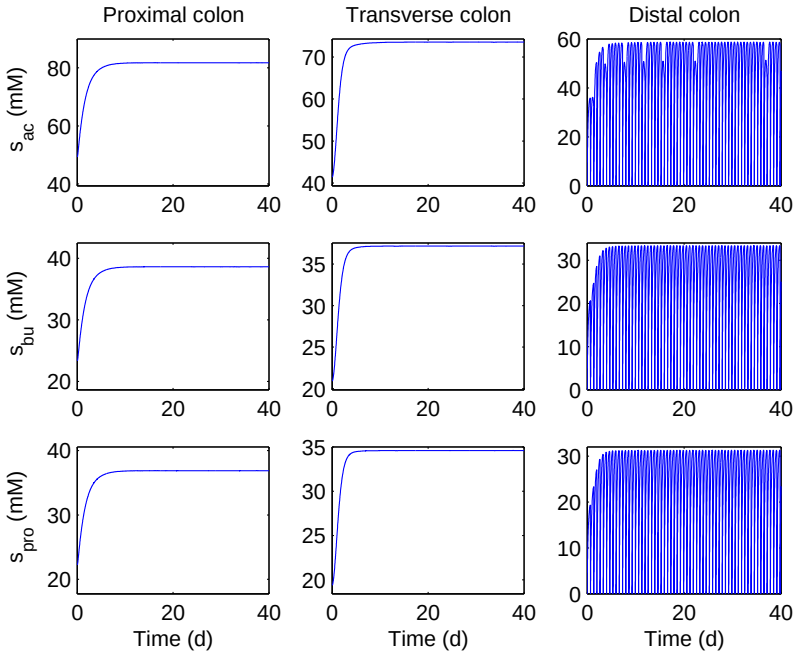

Figure 7.3: Dynamic fermentation pattern along the lumen compartments.

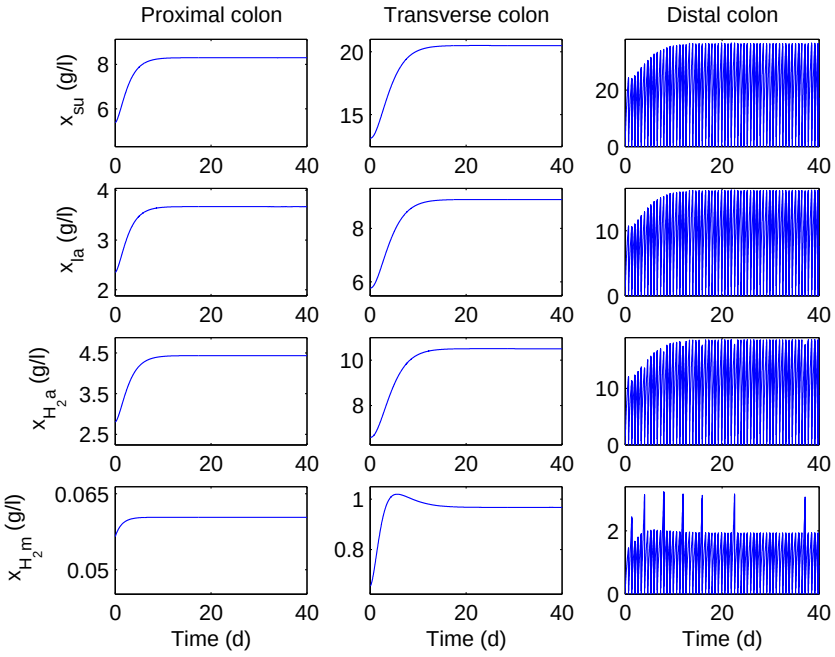

Figure 7.4: Dynamic fermentation pattern along the lumen compartments.

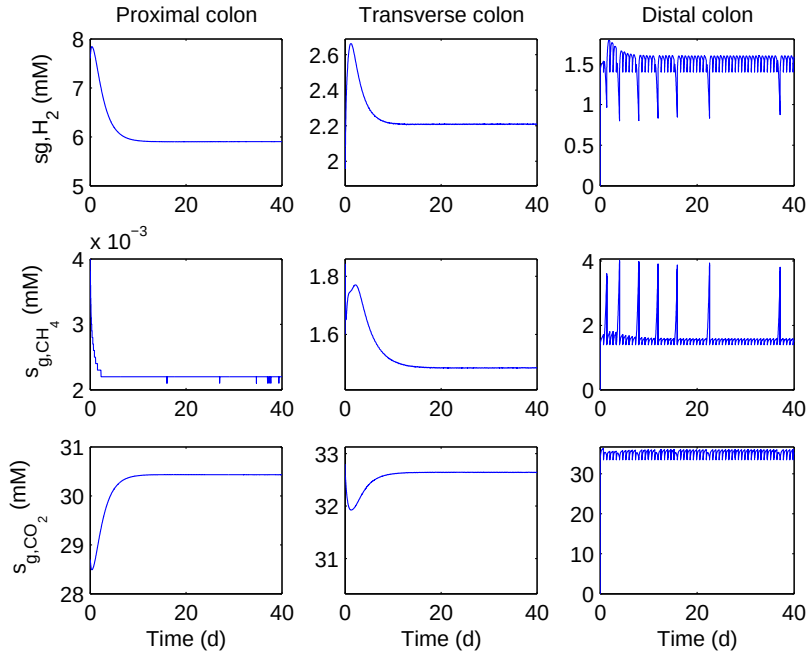

Figure 7.5: Dynamic fermentation pattern in the gas phase along the lumen compartments.

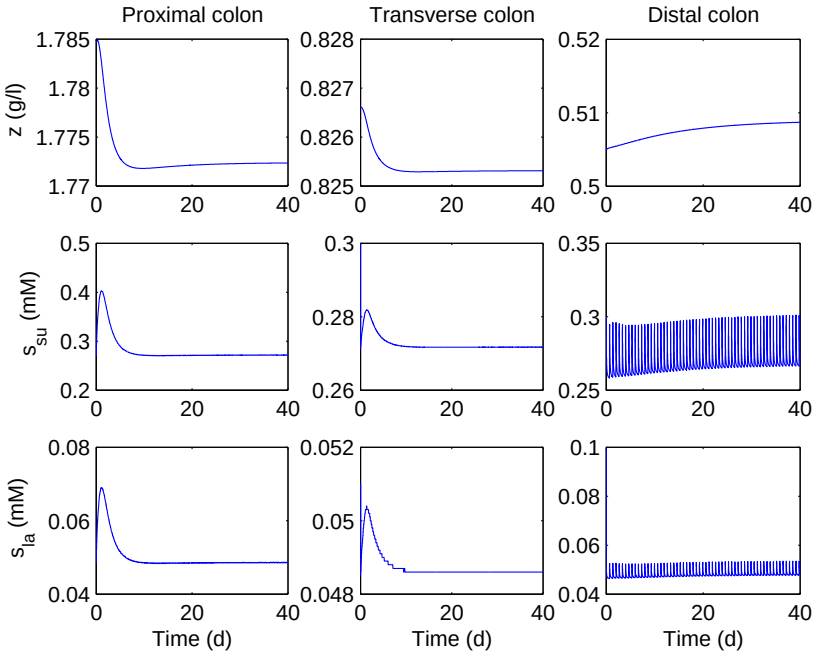

Figure 7.6: Dynamic fermentation pattern along the mucus compartments.

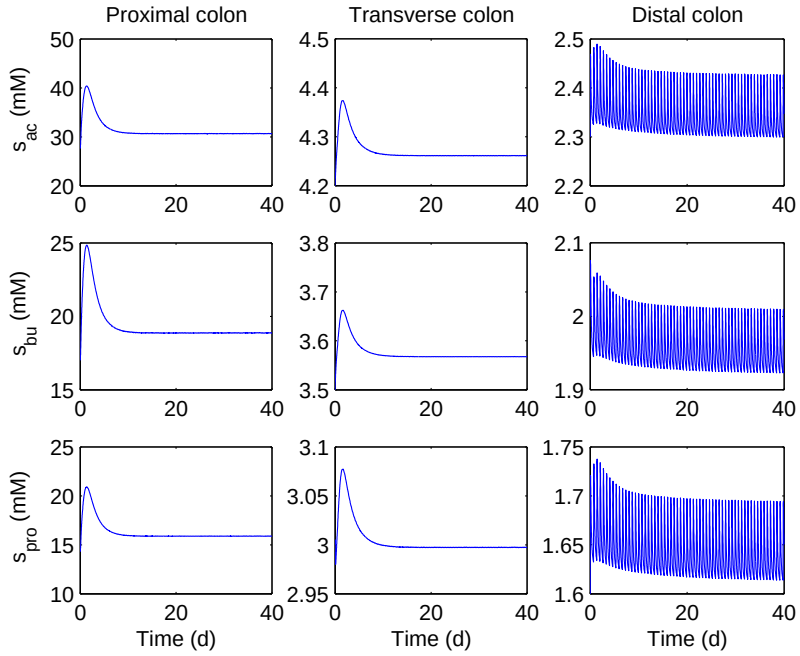

Figure 7.7: Dynamic fermentation pattern along the mucus compartments.

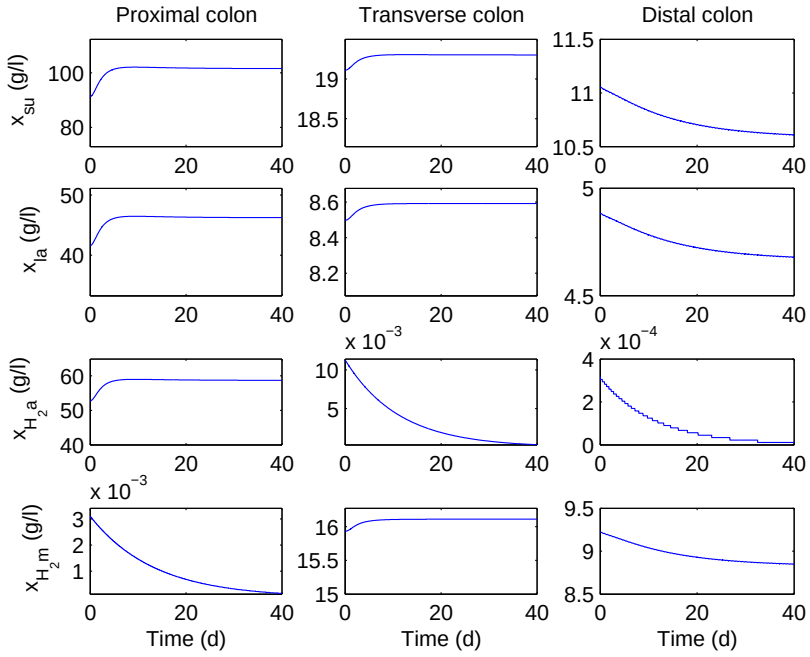

Figure 7.8: Dynamic fermentation pattern along the mucus compartments.

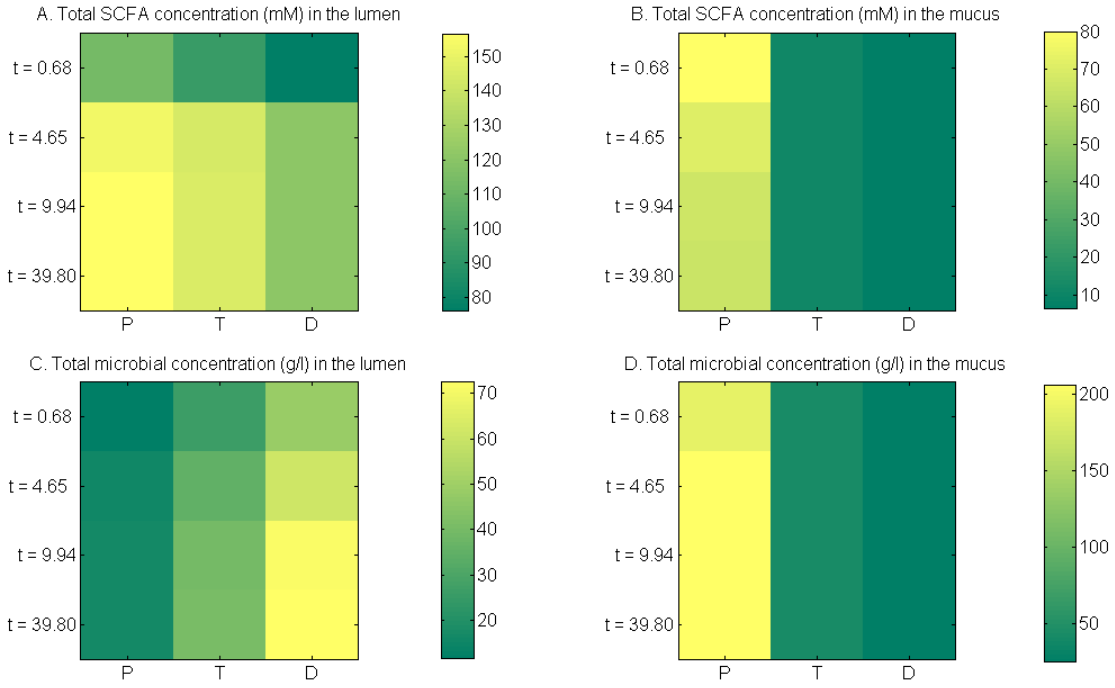

Figure 7.9: Overall fermentation pattern in the human colon. P: proximal colon, T: transverse colon, D: distal colon. Concentration of SCFA falls along the human colon, both in lumen and mucus. Total microbial density increases from proximal to distal section in the lumen, while the microbiota in the mucus follows an opposite behavior.

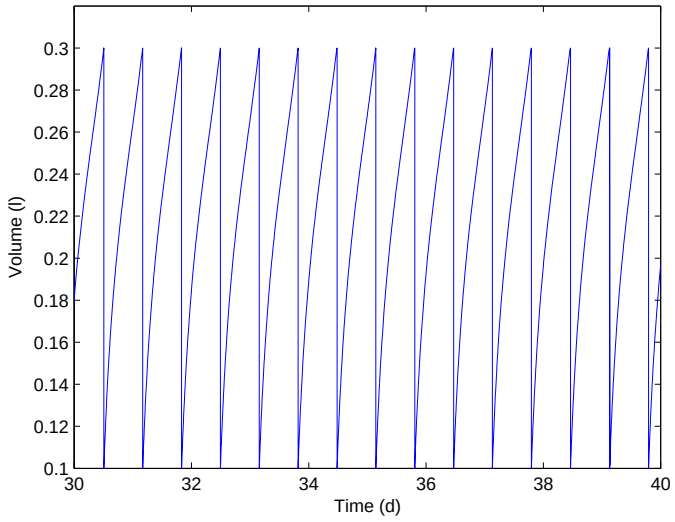

Figure 7.10: Excretion frequency. The number of defecations per day is 1.5.

These figures summarize the prediction of the evolution of metabolites and microorganisms concentrations along the colon when fiber intake changes from 10 g/d to 25 g/d. From there, we can infer that when a perturbation on dietary regime is applied on the system, at least 10 days are needed to see the overall effect of such a change. It is also observed that studying data collected from the distal section, which is the only one experimentally accessible, is not enough to have a full picture of the behavior of the system. By including transport phenomena such as microbial aggregation, SCFA and water absorption, it was possible to obtain a global representation of the human colonic ecosystem that was consistent with physiological data.

7.2 What is the impact of dietary fiber?

Literature relates the benefits of fiber intake on the large intestine function. The assessment of the impact of fibers on the fermentation pattern can be based on human nutritional trial, on model simulations, or both. In this section, we address that question by means of simulation of our model. Three nutritional strategies were evaluated, varying the daily amount of dietary fiber as follows: 10 g/d (low-fiber diet), 25 g/d (average-fiber diet) and 40 g/d (high-fiber diet). The selection of the values of fiber intake was based on a real

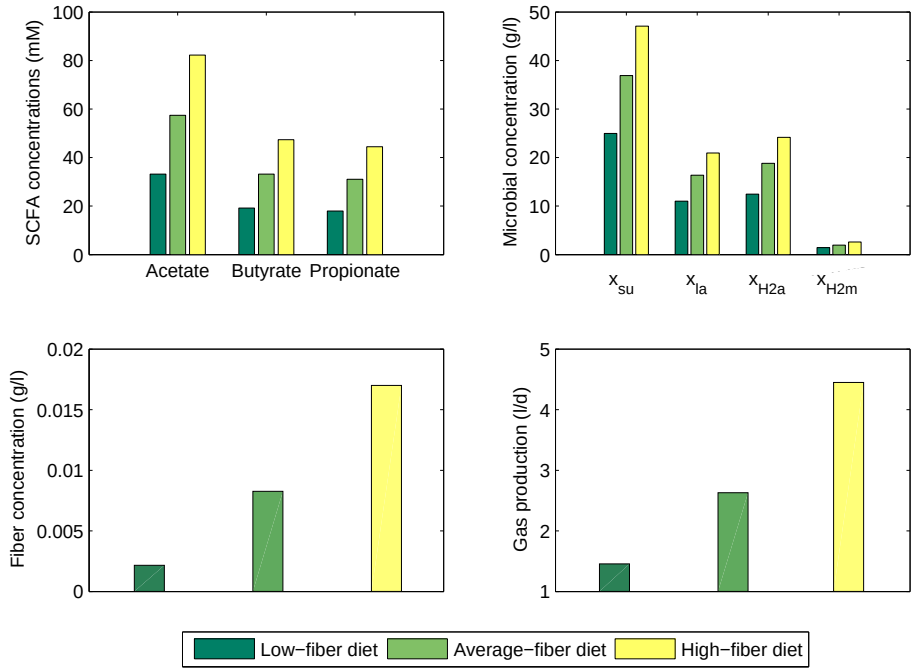

Figure 7.11: Predicted impact of fiber intake on fermentation pattern. Concentrations are taken at last excretion. They correspond to the lumen compartment in the distal colon.

human clinical trial performed within the AlimIntest project. Fig. 7.11 compares the predicted effect of these strategies after a simulation period of 40 days. Simulations show the fermentation pattern in the lumen of the distal section. They predict that increasing fiber intake leads to an increase in SCFA production, as has been observed (reviewed in Topping and Clifton (2001)). In addition, gas production increases, which could be detrimental. While SCFA production has a health-promoting effect, increase of gas causes abdominal pain, distention, and may be related to inflammatory bowel disease. Total amount of gas produced from fermentation is estimated from 0.5 l/d to 4 l/d (Macfarlane and Cummings (1991); Bouhnik et al. (1997)), and the values given by the mathematical model are in accordance with the literature.

Our model could be useful to evaluate the relationship between fiber intake and gas production, which is difficult to carry out both *in vivo* and *in vitro*, with contradictory experimental results (see, *e.g.*, Topping and Clifton (2001)). We can also use the model to characterize the dynamic response of the system and thus determine when the main

changes in the fermentation pattern occur. In the future, our model may be helpful to the establishment of nutrition guidelines.

7.3 What is the role of the mucus?

The mucus affects the dynamics of the lumen in three forms. (i) it provides conditions for microbial aggregation. (ii) the mucins represent an additional carbon source and, finally, (iii) microorganisms in the mucus can enrich the luminal microbial community by means of the shear phenomenon. We have used our mathematical model to assess the importance of each of the above factors on the human colon dynamics. Such effects are illustrated in Fig. 7.12. The continuous blue curve represents the response of the model taking into account the mucus and the action of the three mechanisms mentioned above. The other curves were obtained by setting to zero the parameters associated to each of the three mechanisms in turn. When no mucins were available as substrates for luminal microbiota, the dynamics of the system remained identical. This means that the amount of carbohydrates in the exogenous fiber is much higher than that contained in the mucins that are not utilized in the mucus. Figure 7.12 also suggests that microbiota from the mucus enrich that of the lumen. The simulations predict that microbial aggregation, which is represented by the additional retention time τ for the microorganisms, is necessary to achieve physiological conditions, namely high degradation of fiber and high microbial density. Thus, this mechanism appears as the most relevant factor to explain the role of the mucus on human colon dynamics.

7.4 Model limitations

The current version of our model does not account for some physiological mechanisms that are important for the system behavior. For example, dietary fiber is known to modify colonic transit time. The water-holding capacity of fiber contributes to fecal bulking, which has been shown to be a health-promoting factor by diminishing risk of large bowel cancer (Cummings and Stephen (1980); Topping and Clifton (2001)). Fecal bulking relates to structural properties of the fiber, mechanisms of digesta flow through the intestine, such as peristaltic movement and change in the rheology. Viscosity of the digesta passing along the colon is progressively increasing, imposing a physical barrier to substrate up-

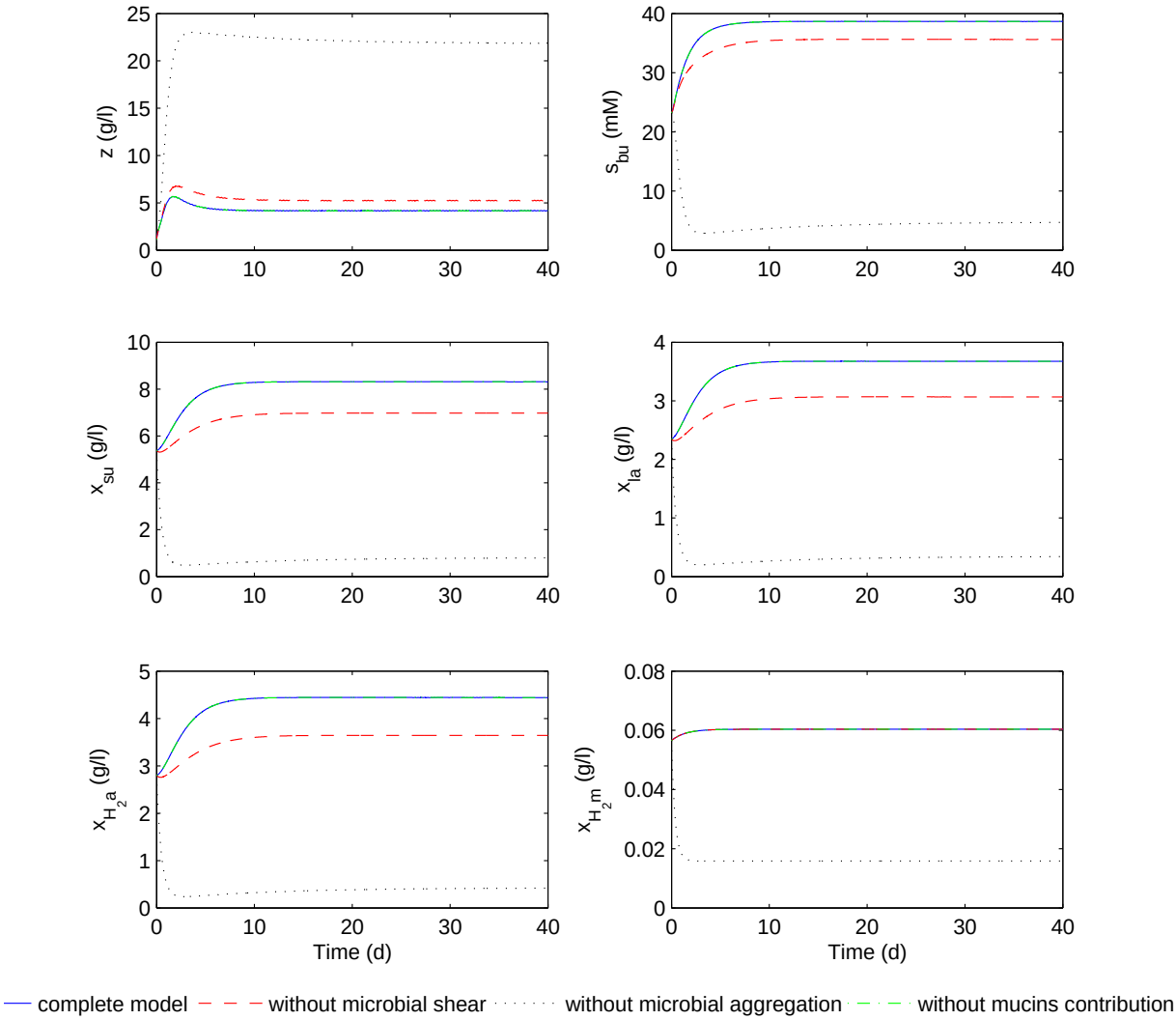

Figure 7.12: Role of mucus in the human colon. Concentrations correspond to the proximal lumen compartment. Fiber intake was set to 25 g/d. Aggregation appears as the most relevant factor to maintain physiological conditions.

take by the microbiota. In biofilm reactors, as in heterogeneous systems (see, *e.g.*, Bird et al. (1960)), the process can be diffusion controlled. The conversion rate may then differ to the one predicted by a model where conversions are assumed to be controlled by biological kinetics (Bernet et al. (2005); Gonzalez-Gil et al. (2001)). In addition, the increased chyme viscosity changes the flow characteristics of the system which are more likely to be that of a plug flow reactor. These factors are not yet included in our model, and should be considered in future work to take into account both biochemical and physical effects of fiber intake on the human colonic ecosystem.

In our model, we did not include protein as carbon source both in lumen and mucus. Protein contribution from dietary origin ranges from 1-12 g/d (see Table 2.2, page 66). Thus, in certain individuals, protein may be an important substrate for the microbiota. Incorporation of protein into the model, implies the integration of its degradation pathway. Extension of our model to account for protein degradation should then consider amino acids (which are the constitutes of proteins) and the functional bacterial group for amino acids utilization.

As previously mentioned in Section 2.3.1.4, the mathematical model does not account for sulfate reduction. This reaction has deleterious consequences for gut health via the formation of toxic sulfide (Flint et al., 2007). Incorporation of such a metabolic conversion into the model may be useful to evaluate strategies for reducing colonic sulfide concentrations.

7.5 Conclusions

In this chapter, we illustrated the potential use of our mathematical model. Simulation results strengthen the plausibility of the model with respect to information reported in the literature. Our model enables one to represent the dynamics of the fermentation process along the physiological sections, including lumen and mucus compartments. It was tested under various nutritional strategies to assess the impact of fiber intake on fermentation pattern. We could evaluate such strategies in a reasonable computational time. Simulations showed that increases in fiber intake lead to increases in SCFA and microbial concentration, but also in gas production, as expected. Furthermore, the simulations indicated

interesting features about the dynamics in the mucus. An advantage of the model is that it allows for analysis that are difficult to carry out by experimental routines. By means of simulations we could study the mechanisms in which the mucus affects the behavior of the luminal subsystem. It was suggested that the most relevant of such mechanisms relates to providing conditions for microbial aggregation.

We also identified the main limitations of the model. They include for example the lack of modelling of the rheology of the chyme, which is an important physiological factor. Such a mechanism may be an interesting element to be considered in future research.

This model makes it easy to address a wide list of *What if* questions, of which those presented here are only examples. Tackling these questions through mathematical modelling should contribute to our insight into the human colonic ecosystem. A key factor towards such an understanding is the integration of information of new bench experiments. In this direction, gnotobiotic rats has been used to evaluate the dynamics of carbohydrate degradation. Preliminaries of such a study are discussed in the following chapter.

Our model should be a complementary approach to the experimental work (both *in vivo* and *in vitro*). It could be a useful tool in developing nutritional strategies focusing on health-promoting metabolites and diminishing gas production.

Chapter 8

Animal rodents inoculated with a minimal human colonic microbiota: an ongoing experimental work

What matters it how far we go? his scaly friend replied.
There is another shore, you know, upon the other side.
Alice's adventures in wonderland. Lewis Carroll

Animal rodents have been broadly used as *in vivo* models of the human colonic ecosystem, as already discussed in Section 2.4.2. Information from metabolic studies of such *in vivo* models is a keystone to tackle the parameter estimation of our mathematical model. Currently, in our lab, a study with gnotobiotic rats has been evaluating the dynamics of carbohydrate degradation. This study comprises three main steps. (i) selection of microorganisms from the human colonic microbiota to be inoculated into the axenic rodents, (ii) preliminary study of microbial colonization, and (iii) study on the impact of fiber intake on fermentation pattern. In this chapter, we describe the methodology used to select human intestinal microorganisms.

8.1 Selection of a minimal functional microbiota

The main part of this work was published in Environmental Microbiology (Tap et al., 2009).

In our mathematical model, the microbiota is represented in metabolic functional groups. Our criteria to select the group of microorganisms for inoculation were (i) coverage of the metabolic pathway illustrated in Fig. 3.2 (page 80), and (ii) prevalence in the human colonic microbiota. The recent study presented by Tap et al. (2009) suggested that humans share a common phylogenetic core of microorganisms in their large intestine. The identification of such a microbial core is briefly presented below.

Sequencing of 16S rRNA genes was carried out from 17 human fecal DNA samples. Clustering of the sequences led to the identification of 3180 Operational Taxonomic Units (OTU). Generally speaking, each OTU represents a bacterial species.

Under the hypothesis that there was not dependence between the individuals, it was assumed that the presence/absence of the OTUs followed a binomial distribution on the prevalence. Let Ψ_j be the probability that the OTU j is detected in an individual. The probability that OTU j is detected in k individuals is

$$\binom{n}{k} \Psi_j^n (1-\Psi_j)^{n-k}, \tag{8.1}$$

with n the number of individuals ($n = 17$). The estimated probability $\widehat{\Psi}_j$ is given by $\widehat{\Psi}_j = \frac{k_j}{n}$, where k_j is the number of individuals where OTU j was detected.
In order to assess the uncertainty related to the sample size, confidence intervals were computed by the method proposed by Wilson (1927) and discussed by Agresti and Coull (1998).
The confidence interval has the form

$$\frac{\widehat{\Psi}_j + z^2_{\alpha/2}/2n \pm z_{\alpha/2}\left[\left(\widehat{\Psi}_j(1-\widehat{\Psi}_j) + z^2_{\alpha/2}/4n\right)/n\right]^{1/2}}{1+z^2_{\alpha/2}/n} \tag{8.2}$$

where $z^2_{\alpha/2}$ denotes the quantile of the standard normal distribution. It was set to a value of 1.96 to correspond to a 95% interval confidence.

Table 8.1: Phylogenetic core.

OTU	pj (%)		Confidence interval	OTU	pj (%)		Confidence interval
Faecalibacterium prausnitzii	4.76	0.94	[0.73 ; 0.99]	Roseburia intestinalis	0.8	0.59	[0.36 ; 0.79]
Anaerostipes caccae	3.11	0.88	[0.65 ; 0.97]	Clostridium clostridioforme	1.84	0.59	[0.36 ; 0.79]
Clostridium spiroforme	0.95	0.88	[0.65 ; 0.97]	Clostridium nexile	1.39	0.59	[0.36 ; 0.79]
Bacteroides uniformis	4.97	0.82	[0.59 ; 0.94]	Alistipes putredinis	1.53	0.59	[0.36 ; 0.79]
Dorea longicatena	1.23	0.76	[0.52 ; 0.90]	Eubacterium ramulus	0.77	0.59	[0.36 ; 0.79]
Bifidobacterium longum biovar Longum	0.86	0.76	[0.52 ; 0.90]	Faecalibacterium prausnitzii	1.15	0.59	[0.36 ; 0.79]
Clostridium sp	2.17	0.76	[0.52 ; 0.90]	Bacteroides capillosus	1.22	0.59	[0.36 ; 0.79]
Faecalibacterium prausnitzii	1.95	0.76	[0.52 ; 0.90]	Ruminococcus obeum	0.8	0.59	[0.36 ; 0.79]
Clostridium bolteae	2.41	0.76	[0.52 ; 0.90]	Ruminococcus luti	1.11	0.59	[0.36 ; 0.79]
Clostridium sp	1.9	0.76	[0.52 ; 0.90]	Clostridium amygdalinum	0.56	0.59	[0.36 ; 0.79]
Lachnospira pectinoschiza	1.22	0.71	[0.47 ; 0.87]	Clostridium straminisolvens	0.9	0.59	[0.36 ; 0.79]
Faecalibacterium prausnitzii	2.96	0.71	[0.47 ; 0.87]	Ruminococcus obeum	0.73	0.53	[0.31 ; 0.74]
Roseburia intestinalis	2.12	0.71	[0.47 ; 0.87]	Ruminococcus luti	1.7	0.53	[0.31 ; 0.74]
Ruminococcus obeum	0.55	0.71	[0.47 ; 0.87]	Bacteroides capillosus	1.04	0.53	[0.31 ; 0.74]
Eubacterium hallii	1.39	0.71	[0.47 ; 0.87]	Clostridium orbiscindens	0.58	0.53	[0.31 ; 0.74]
Hespellia porcina	2.61	0.71	[0.47 ; 0.87]	Coprococcus eutactus	4.52	0.53	[0.31 ; 0.74]
Bacteroides vulgatus	8.18	0.71	[0.47 ; 0.87]	Eubacterium hallii	0.93	0.53	[0.31 ; 0.74]
Ruminococcus obeum	1.04	0.71	[0.47 ; 0.87]	Clostridium sp	0.97	0.53	[0.31 ; 0.74]
Eubacterium rectale	8.16	0.65	[0.42 ; 0.83]	Clostridium xylanolyticum	0.77	0.53	[0.31 ; 0.74]
Faecalibacterium prausnitzii	1.58	0.65	[0.42 ; 0.83]	Eubacterium rectale	0.66	0.53	[0.31 ; 0.74]
Faecalibacterium prausnitzii	1.42	0.65	[0.42 ; 0.83]	Collinsella aerofaciens	0.73	0.53	[0.31 ; 0.74]
Dorea formicigenerans	0.85	0.65	[0.42 ; 0.83]	Ruminococcus bromii	9.01	0.53	[0.31 ; 0.74]
Eubacterium rectale	1.23	0.65	[0.42 ; 0.83]	Eubacterium eligens	1.47	0.53	[0.31 ; 0.74]
Eubacterium rectale	4.74	0.65	[0.42 ; 0.83]	Clostridium bolteae	0.7	0.53	[0.31 ; 0.74]
Ruminococcus schinkii	1.74	0.59	[0.36 ; 0.79]				

The phylogetic core was defined as the subset of OTUs found in at least 50% of the individuals. The resulting core was a set of 49 candidates (see Table 8.1). Taxonomic characterization of the OTUs was performed using the RDP II Classifier program (RDP II Release 9.58). The set of candidates were compared to four human GI tract 16S rRNA external databases obtained from different countries (Li (2003); Eckburg et al. (2005); Manichanh et al. (2006); Gill et al. (2006)). The 81.6% of the candidates were found in more than three databases, which confirmed the selection. The uncertainty of the probabilities estimates in the binomial distribution is high due to the small sample size. According to the confidence intervals, the 10 most frequent OTUs were very likely to be part of the core. After the assignation to the closet isolated type strain from RDP II, the ten most prevalent species were: *Faecalibacterium prausnitzii, Anaerostipes caccae, Clostridium spiroforme, Bacteroides uniformis, Dorea longicatena, Bifidobacterium longum biovar Longum, Clostridium sp, Faecalibacterium prausnitzii species 2, Clostridium bolteae, Clostridium sp*.

The parameters Ψ_j do not provide information about the abundance of the OTUs in the global dataset. In order to have a representation of the core in terms of abundance, the numbers of sequences of each OTU were averaged on the subset of individuals where the OTU was detected. So

$$p_l = \frac{1}{k_l} \sum_{i=1}^{17} c_{i,l}, \tag{8.3}$$

with p_l the average abundance of the the OTU l belonging to the core and $c_{i,l}$ the number of sequences of the OTU l in the individual i.
Average abundances were normalized as

$$\bar{p_l} = \frac{p_l}{\sum_{l=1}^{49} p_l}, \tag{8.4}$$

with $\bar{p_l}$ the normalized abundance of the OTU l.

The representation of the core is depicted in Fig 8.1. The normalized abundance of the core OTUs varied from 0.5% to 9%. Each of the ten OTUs with highest normalized abundance was assigned to the closet isolated type strain from RDPII. *Ruminoccocus bromii, Bacteroides vulgatus, Eubacterium rectale and Faecalibacterium prausnitzii* had an important contribution in the phylogenetic core with respect to their normalized abundance.

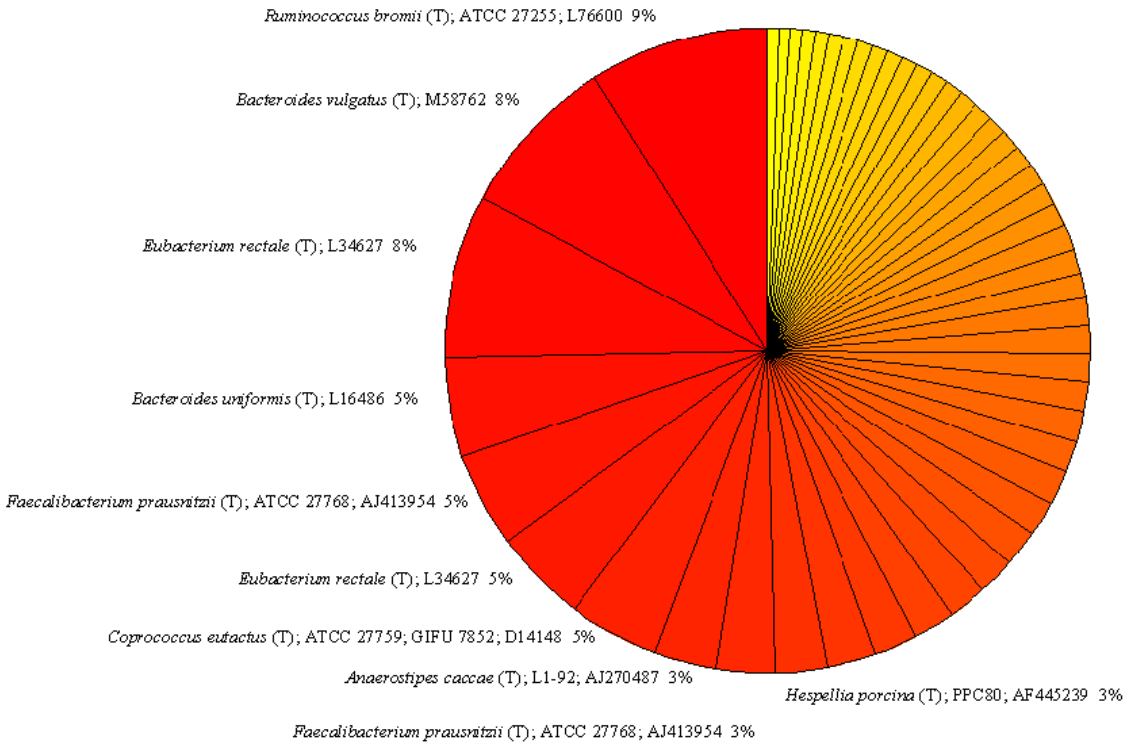

Figure 8.1: Representation of the phylogenetic core. Each fraction corresponds to the normalized abundances of the bacterial strains.

The normalized abundances in Fig. 8.1 may be linked to the trophic chain and the stress resistance. From the metabolic point of view, a high value on the core composition could represent a high flexibility of the species to utilize diverse substrates. It also can be related to a high yield of biomass with respect to the substrate(s). In terms of conditions linked to the stress induced by environmental variables such as the pH or mechanical forces (peristaltic movement, shed of epithelial cells), the value of the normalized abundance can be associated to the capability of the species to resist the wash-out phenomenon.

Finally, taking into consideration the identification of a possible phylogenetic core for the human intestinal microbiota, and metabolic functions of bacterial strains, a group of seven bacterial strains was selected for the *in vivo* experiments (Fig. 8.2). This minimal human colonic microbiota was inoculated in seven rats. Evaluation of bacterial colonization (step (ii) of the experimental work) is being carried out by temporal temperature gradient gel electrophoresis (TTGE) and polymerase chain reaction (PCR) techniques.

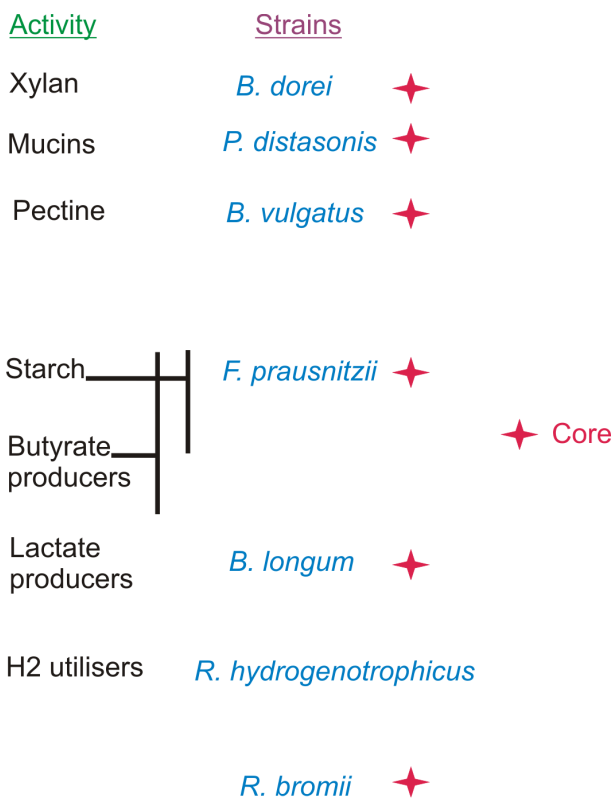

Figure 8.2: Minimal human colonic microbiota selected for *in vivo* experiments.

8.2 Conclusions

In this chapter, we showed a preliminary step of an ongoing experimental work with gnotobiotic rats. The aim of such experiments is to provide information on the dynamics of carbohydrate degradation, and on the location of microorganisms of the trophic chain. Data resulting from there are expected to be integrated into the further validation of our mathematical model. We presented here the strategy that we used to select a minimal group of the human colonic microbiota to be inoculated in the axenic rats. The selection was based on information of bacterial metabolic activities, and on the prevalence of bacterial strains in a study with 17 individuals. Currently, members of domain *Archaea* are not yet incorporated, thus the methane reaction will not be part of the metabolic pathway. The *in vivo* experiments referred here comprise one of the perspectives of this work, which are discussed in the following chapter.

Chapter 9

Concluding remarks and perspectives

The rest is not silence[1]

We present here the main conclusions of our work. Some of them were already mentioned in the previous chapters. Because of time constraints, there are very interesting subjects that we could not tackle. They are mentioned here. We believe that they constitute important axes for future research towards the strengthening of a comprehensive and plausible mathematical model of the human colonic ecosystem.

Conclusions

We have constructed a mathematical model of a complex process, namely the carbohydrate degradation in the human colon. We decided to build a knowledge-based model in order to provide a structure with physical meaning that enable system theoreticians or process engineers to discuss with microbiologists. The model building process was enriched by points of views of microbiologists, physiologists, mathematicians and process engineers.

We considered the human colon as a biological reactor. This is way our mathematical

[1]The original phrase, pronounced by Hamlet is: *the rest is silence*. William Shakespeare

model was based on modelling approaches usually employed for chemical reactors, particularly for anaerobic bioreactors. Our model takes into account the physiological regions of the human colon. It integrates a hydraulic representation of the system, biological reactions and transport phenomena. To the best of our knowledge, this is the first mathematical model that achieves such an integration, providing a close picture of the complexity of the human colon ecosystem.

The resulting model comprises a large number of state variables and parameters. To deal with such a complexity, we applied a generic strategy that could be extended to other modelling tasks where the model is too complex to be built from experimental data alone and where large body of prior knowledge must be taken advantage of. This strategy includes aggregation of mechanisms, integration of knowledge to reduce the number of model parameters, modelling and identification of parts of the complete process. Such parts corresponded here to *in vitro* bacterial growth experiments dedicated to studying specific reactions of the complete trophic chain. The development of the toolbox IDEAS enabled us to perform parameter estimation of such submodels and to detect practical identifiability problems. We also applied the singular perturbation method to a submodel related to homoacetogenesis. This study showed that it was possible to reduce the order of the submodel. The extension of such an approach to the complete model was satisfactory, leading to improvements on the stability of simulations and savings in computational time.

We tested our mathematical under various nutritional scenarios. Simulation results showed the plausibility of the model with respect to information reported in the literature. In addition, the mathematical model that we have developed appears as promising tool to analyze phenomena that are difficult to study experimentally. It should be viewed as complementary to *in vivo* and *in vitro* models. We hope that the mathematical model presented here will help to provide a better insight into the dynamics of the human colon ecosystem. In fact, our model was already used to simulate a real clinical trial aimed at evaluating the impact of two dietary regimes (10 g and 40 g of fibers per day) on human volunteers. Model simulations provided useful information to analyze the collected data (Tap, 2009) (the results are not shown in this work).

Outlook

A major part of the time available for the preparation of this thesis was dedicated to the understanding of the phenomena involved in carbohydrate degradation by human colonic microbiota. The construction of the model and its implementation were also quite time demanding. As a result, we could not tackle some interesting and important features as much as we would have wished to. In the following, we mention some of them.

Parameter estimation

Estimation of the parameters of the complete model is, from our point of view, the main task to be pursued in the near future. Such a step is obviously dependent on availability of data. However, our mathematical model can be used as virtual system to mimic an experimental scenario, taking into consideration the nature and limitations of the real measurements. *In silico* data could be further used to explore the estimation in the Bayesian framework, which is for us a promising way to deal with the full estimation problem.

We mentioned that a current experimental work with gnotobiotics rats is being carried out in our lab. Information from these experiments will provide useful data for validation of our model. Microbiologists use to provide the concentration of bacteria by number of cells per ml or by optical density. Such measurements impose an additional obstacle to the estimation, because to relate the data with the model we need information about physical properties of each bacteria. Such an information is not always available. Therefore, hypotheses on these properties have to be formulated to allow the use of the data in the estimation framework, which add other level of uncertainty. For the monitoring of biological reactors, a broadly used measurement for microbial concentration is the volatile suspended solids (VSS) (See the Standard Methods of the American Public Health Association, APHA (2005)). We suggest to include this measurement in the experimental protocol. The measurement of VSS is given in units of mass per volume, which could facilitate the incorporation of the data in the model.

Some ideas on parameter estimation of *in vitro* experiments

Modelling and identification of *in vitro* bacterial growth experiments are of great value in the definition of prior distributions of the parameters in the Bayesian framework. In our

estimation studies on homoacetogenesis and butyrate production, identification of Monod parameters was hampered by practical identifiability problems. We think that it would be interesting to study the degrees of freedom of the procedure for data collection in both experiments, in order to assess the possibility of optimal experiment design. The toolbox IDEAS may also be useful in this context, because it provides an approximate of the Fisher information matrix on which most design optimality criteria are based.

The Monod equation has interesting properties that we did not exploit. In simple models, Monod kinetics can be arranged in LP form (see, *e.g.*, Lineweaver and Burk (1934); Ratkowsky (1986); Dochain and Vanrolleghem (2001); Doeswijk and Keesman (2009)). For the estimation of the transformed LP model, it is then possible to apply the ordinary least-squares estimator, which has a much lower computational complexity than nonlinear estimation. It is known that nonlinear estimation is highly sensitive to the initial guess of the parameters, so estimates given by the ordinary least-squares estimator could be used as initial guesses of the nonlinear estimation algorithm.

Sensitivity analysis

Another interesting subject to be treated is the identification of the most influential parameters in the model. We already mentioned that this information could be used in the estimation framework. In addition, it also provides insight into the phenomena that play a major role in system behavior, and could lead to modifications on the model structure. Methods for sensitivity analysis can be found, *e.g.*, in Saltelli et al. (2000).

Model structure

Our mathematical model has already a complex structure. Any new incorporation of additional mechanisms in the current structure must be done with care. We identified some limitations of the model in representing the physiology of the human colon. We think that the main one relates to the simplification of the rheology in the system and the characterization of flow pattern. A further study in modelling such features may help to provide a closer approach. Residence time theory used in chemical reactor engineering provides a useful tool that may be applied to the human colon. For details in methods for residence time distributions, see, *e.g.*, Nauman (2002).

We did not include acid-base reactions in our model. These equations are useful to calculate the pH, which is an important factor that may be related to physiological disorders. We think that including such reactions should be considered in future work (see Batstone et al. (2002) for further details).

We gathered all the available complex polysaccharides into a single variable. Some *in vitro* studies have suggested that fermentation rates differ among the available substrates (see, *e.g.*, Stewart et al. (2008)). In order to evaluate the impact of specific functional aliments, it may be useful to split the polysaccharide concentration in its components, such as xylan, starch, fructo-oligosaccharides and inulin.

Our model does not account for protein degradation. To incorporate such a process, it will be needed to include the proteolysis step and further conversion of amino acids. This implies an additional functional bacterial group for amino acids utilization. Information on protein degradation pathway can be found, *e.g.*, in (Macfarlane and Cummings (1991); Bernalier-Donadille (2004)). Sulfate reduction was not included in the model. This reaction has a deleterious effect on gut health via the formation of toxic sulfide. Its incorporation into the model structure may be useful to evaluate strategies aimed at reducing colonic sulfide concentrations.

We oversimplified the dynamics of the mucus. The mucus is a biofilm. Literature reports complex models to represent biofilms (see, *e.g.*, Picioreanu et al. (2004); Xavier et al. (2005b), Xavier et al. (2005a)). Such approaches, often called individual-based modelling (IbM), are also hampered by lack of experimental data, which is our main obstacle to model validation. As a theoretical framework, IbM approaches can provide interesting qualitative results on the distribution of bacterial species along the intestine.

Advances in metagenomics should offer a better insight into the diversity and functions of the human colonic microbiota. How to incorporate information given by molecular ecology techniques into kinetic modelling is an interesting topic of research. For biological reactors some approaches have been recently proposed to include microbial diversity and functional assigning (see, *e.g.*, Harmand et al. (2008); Dumont et al. (2009); Ramirez

et al. (2009b)). These approaches are still in open discussion and further work is needed for their validation and acceptance, providing interesting opportunities for future research. It will be interesting to model mathematically the dynamics of *in vivo* metabolic cross-feeding studies, where functional genomic data are available. The ongoing experimental work with gnotobiotic rats could be exploited in such a context.

Finally, it has been suggested that studies on bioreactors can contribute to the understanding of the factors that shape the microbiota (Sonnenburg et al. (2004); Bäckhed et al. (2005); Ley et al. (2008)). We also think that our work could have, in the long term, implications on biological reactor design. Living systems are high performance reactors and their study could be used in an attempt at reverse engineering. Given the strong link between the human colon and anaerobic reactors, a continuous and fruitful feedback both in modelling and system behavior analysis could be expected. For example, the particularity of colonic SCFA absorption may inspire the design of reactor configurations in order to drive the thermodynamics of the metabolic conversions. As a chemical engineer who has borrowed concepts of process engineering and wastewater treatment for the construction of a model in the human colonic microbiology, I would be most happy if the microbiology of the large intestine could in turn provide useful concepts to chemical engineering.

For sure, the rest is not silence.

Publications

This work led to the following publications:

Papers in proceedings

R. Muñoz-Tamayo, J. P. Steyer, B. Laroche and M. Leclerc. 2007. Human colon: a complex bioreactor. Conceptual modelling for the anaerobic digestion of the functional trophic chain. In Proc. 11th World Congress Anaerobic Digestion Bio-energy for our Future. Brisbane, Australia. 6 pages on CD-Rom proceedings.

R. Muñoz-Tamayo, B. Laroche, M. Leclerc and E. Walter. 2008. Modelling and identification of *in vitro* homoacetogenesis by human-colon bacteria. In Proc. 16th IEEE Mediterranean Conference on Control and Automation. Ajaccio, France. 1717-1722.

R. Muñoz-Tamayo, B. Laroche, M. Leclerc and E. Walter. 2009. IDEAS: a parameter identification toolbox with symbolic analysis of uncertainty and its application to biological modelling. In Preprints of the 15th IFAC Symposium on System Identification. Saint-Malo, France. 1271-1276.

Contribution in

J. Tap, S. Mondot, F. Levenez, E. Pelletier, C. Caron, J.P. Furet, E. Ugarte, **R. Muñoz-Tamayo**, D. Le Paslier, R. Nalin, J. Doré and M. Leclerc. 2009. Towards the human intestinal microbiota phylogenetic core. Environmental Microbiology 11(10), 2574-2584.

Journal papers: submitted

R. Muñoz-Tamayo, B. Laroche, E. Walter, J. Doré, S. H. Duncan, H. J. Flint and M. Leclerc. Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species.

R. Muñoz-Tamayo, B. Laroche, E. Walter, J. Doré and M. Leclerc. Mathematical modelling of carbohydrate degradation by human colonic microbiota.

R. Muñoz-Tamayo, B. Laroche, E. Walter, J. Doré and M. Leclerc. How can control theory and system identification contribute to the modelling of complex biological processes from scarce and heterogeneous data?

Appendix

Appendix A

Numerical values of the model parameters

Parameters of the model were defined according to our studies on estimation of kinetic submodels (Chapter 5). Other parameters were selected from literature review. Parameters whose values we did not find in the literature were defined based on available biological knowledge. Physiological constants and parameters values used in the simulations presented in Chapter 7 are given below.

Physiological parameters

Physiological parameters	**Value**	**Sources**	**Comments**
P_{atm}	1.013 bar		
p_{H_20}	0.08274bar	a	
R	0.08314 bar/(M K)	a	
T	310.15 K	b	
q_{in}	1.5 l/d	c	Input flow in the proximal lumen
V_g			It is assumed that gas phase occupies 10% of the volume in each microhabitat
Γ	5 g/d	d,e	Each of the three sections is assumed to produce $\Gamma/3$

a. Smith et al. (2005); b. Savage (1977); c. Cummings and Macfarlane (1991); d. Macfarlane and Cummings (1991); e. Egert et al. (2006).

Kinetic parameters

Kinetic parameters	**Value**	**Sources**	**Kinetic parameters**	**Value**	**Sources**
$k_{hyd,z}$	$1.20 * 10^3$	f	$Y_{pro,la}$	0.267	
$K_{x,z}$	29.99	f	$Y_{bu,la}$	0.200	
$Y_{su,z}$	0.005	f	$Y_{CO_2,la}$	0.533	
$k_{m,su}$	7.92	g	$Y_{H_2O,la}$	0.493	
$K_{s,su}$	0.0026	h	k_{m,H_2a}	108.837	
Y_{su}	0.120	h	K_{s,H_2a}	0.0017	
$Y_{la,su}$	0.499		Y_{H_2a}	0.043	
$Y_{H_2,su}$	1.440		Y_{ac,H_2a}	0.143	
$Y_{ac,su}$	0.567		Y_{CO_2,H_2a}	-0.5	
$Y_{pro,su}$	0.240		Y_{H_2O,H_2a}	0.629	
$Y_{bu,su}$	0.270		k_{m,H_2m}	22.581	h
$Y_{CO_2,su}$	1.100		K_{s,H_2m}	$1.563*10^{-6}$	h
$Y_{H_2O,su}$	1.440		Y_{H_2m}	0.062	
$k_{m,la}$	103		Y_{CH_4,H_2m}	0.095	
$K_{s,la}$	6.626e-3		Y_{CO_2,H_2m}	-0.450	
Y_{la}	0.120		Y_{H_2O,H_2m}	0.686	
$Y_{H_2,la}$	0.400		k_d	0.01	h
$Y_{ac,la}$	0.133				

f. Ramirez et al. (2009a); g. Rodríguez et al. (2006b); h. Batstone et al. (2002). When the reference is not specified, parameters were derived from our kinetic studies in Chapter 5. Yield factors were calculated following the procedure explained in Section 6.1.1. Kinetics parameters are the same in all compartments.

Parameters associated to mass-transfer phenomena

Parameter	Value
$k_L a$	200 d^{-1}
K_{H,H_2}	7.29*10−4 M/bar
K_{H,CO_2}	0.0255 M/bar
K_{H,CH_4}	0.0011 M/bar

All values were taken from Batstone et al. (2002).

Parameter	Value		
	Proximal colon	**Transverse colon**	**Distal colon**
γ^{l}_{su}	1.60 l/d	3.80 l/d	6.30 l/d
γ^{l}_{la}	0.88 d^{-1}	0.43 d^{-1}	2.03 d^{-1}
γ^{l}_{ac}	1.32 d^{-1}	0.64 d^{-1}	3.05 d^{-1}
γ^{l}_{pro}	1.07 d^{-1}	0.62 d^{-1}	2.47 d^{-1}
γ^{l}_{bu}	0.90 d^{-1}	0.57 d^{-1}	2.49 d^{-1}
$\gamma^{l}_{H_2O}$	1.60 d^{-1}	0.77 d^{-1}	3.66 d^{-1}
γ^{m}_{la}	12.60 d^{-1}	12.60 d^{-1}	12.60 d^{-1}
γ^{m}_{ac}	18.90 d^{-1}	18.90 d^{-1}	18.90 d^{-1}
γ^{m}_{pro}	15.32 d^{-1}	15.32 d^{-1}	15.32 d^{-1}
γ^{m}_{bu}	12.88 d^{-1}	12.88 d^{-1}	12.88 d^{-1}
$\gamma^{m}_{H_2O}$	0.01 d^{-1}	0.01 d^{-1}	0.01 d^{-1}

Definition of these parameters was based on information reported in the literature on percentage of water absorption and SCFA concentrations along the human colon.

Parameters associated to microbial phenomena

We did not consider microbial adherence, thus a_i was set to zero. The shear coefficient for all microrganisms was defined as $b_i = 0.08$ d^{-1}. The additional residence time τ_i was set to 1.0 d^{-1}, in order to achieve high microbial density, as reported in the literature.

Bibliography

Abusam, A., Keesman, K. J., van Straten, G., Spanjers, H., Meinema, K., 2001. Parameter estimation procedure for complex non-linear systems: calibration of asm no. 1 for n-removal in a full-scale oxidation ditch. Water Sci Technol 43 (7), 357–365.

Agresti, A., Coull, B. A., 1998. Approximate is better than exact for interval estimation of binomial proportions. The American Statistician 52, 119–125.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Akin, D. E., 1976. Ultrastructure of rumen bacterial attachment to forage cell walls. Appl Environ Microbiol 31, 562–568.

Alander, M., Smet, I. D., Nollet, L., Verstraete, W., von Wright, A., Mattila-Sandholm, T., Jan 1999. The effect of probiotic strains on the microbiota of the simulator of the human intestinal microbial ecosystem (SHIME). Int J Food Microbiol 46 (1), 71–79.

Amann, R. I., Ludwig, W., Schleifer, K. H., Mar 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59 (1), 143–169.

Amaretti, A., Bernardi, T., Tamburini, E., Zanoni, S., Lomma, M., Matteuzzi, D., Rossi, M., 2007. Kinetics and metabolism of Bifidobacterium adolescentis mb 239 growing on glucose, galactose, lactose, and galactooligosaccharides. Appl Environ Microbiol 73 (11), 3637–3644.

Amaretti, A., Tamburini, E., Bernardi, T., Pompei, A., Zanoni, S., Vaccari, G., Matteuzzi, D., Rossi, M., Dec 2006. Substrate preference of Bifidobacterium adolescentis mb 239: compared growth on single and mixed carbohydrates. Appl Microbiol Biotechnol 73 (3), 654–662.

APHA, 2005. Standard Methods for the Examination of Water and Wastewater. American Public Health Association (APHA), American Water Works Association (AWWA) & Water Environment Federation (WEF).

Atuma, C., Strugala, V., Allen, A., Holm, L., May 2001. The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. Am J Physiol Gastrointest Liver Physiol 280 (5), G922–G929.

Baghurst, P. A., Baghurst, K. I., Record, S. J., 1996. Dietary fibre, non-starch polysaccharides and resistant starch : A review. Food Aust 48, S3–S35.

Bailey, J. E., 1998. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol Prog 14 (1), 8–20.

Baldwin, R. L., Bywater, A. C., Taylor, J., Murphy, M., 1984. Hervibore Nutrition in the Subtropics and Tropics. The Science Press (Pty) ltd, Ch. Integration of metabolism in the rumen and host animal, pp. 478–502.

Baldwin, R. L., Crist, K., Waghorn, G., Smith, N. E., Jan 1981. The synthesis of models to describe metabolism and its integration. Proc Nutr Soc 40 (1), 139–145.

Ballyk, M. M., Jones, D. A., Smith, H. L., 2001. Microbial competition in reactors with wall attachment: a mathematical comparison of chemostat and plug flow models. Microbial Ecology 41 (3), 210–221.

Banga, J. R., 2008. Optimization in computational systems biology. BMC Syst Biol 2, 47.

Barcenilla, A., Pryde, S. E., Martin, J. C., Duncan, S. H., Stewart, C. S., Henderson, C., Flint, H. J., Apr 2000. Phylogenetic relationships of butyrate-producing bacteria from the human gut. Appl Environ Microbiol 66 (4), 1654–1661.

Batstone, D. J., 2006. Mathematical modelling of anaerobic reactors treating domestic wastewater: rational criteria for model use. Reviews in Environmental Science and Bio/Technology 5, 57–71.

Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostahis, S. G., Rozzi, A., Sanders, W. T. M., Siegrist, H., Vavilin, V. A., 2002. Anaerobic Digestion Model No. 1 (ADM1), IWA Task Group for Mathematical Modelling of Anaerobic Digestion Processes. IWA Publishing, London.

Batstone, D. J., Keller, J., Newell, R. B., Newland, M., 2000. Modelling anaerobic degradation of complex wastewater. I: model development. Bioresource Technology 75, 67–74.

Batstone, D. J., Keller, J., Steyer, J. P., 2006a. A review of ADM1 extensions, applications, and analysis: 2002-2005. Water Sci Technol 54 (4), 1–10.

Batstone, D. J., Picioreanu, C., van Loosdrecht, M. C. M., 2006b. Multidimensional modelling to investigate interspecies hydrogen transfer in anaerobic biofilms. Water Res 40, 3099–3108.

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., Gordon, J. I., Mar 2005. Host-bacterial mutualism in the human intestine. Science 307 (5717), 1915–1920.

Belenguer, A., Duncan, S. H., Calder, A. G., G.Holtrop, Louis, P., Lobley, G. E., Flint, H. J., May 2006. Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. Appl Environ Microbiol 72 (5), 3593–3599.

Belenguer, A., Duncan, S. H., Holtrop, G., Anderson, S. E., Lobley, G. E., Flint, H. J., Oct 2007. Impact of ph on lactate formation and utilization by human fecal microbial communities. Appl Environ Microbiol 73 (20), 6526–6533.

Belenguer, A., Duncan, S. H., Holtrop, G., Flint, H. J., Lobley, G. E., 2008. Quantitative analysis of microbial metabolism in the human large intestine. Current Nutr Food Sci 4, 109–126.

Bellu, G., Saccomani, M. P., Audoly, S., D'Angiò, L., 2007. Daisy: a new software tool to test global identifiability of biological and physiological systems. Comput Methods Programs Biomed 88 (1), 52–61.
URL `http://www.dei.unipd.it/~pia/`

Bengmark, S., Jan 1998. Ecological control of the gastrointestinal tract. the role of probiotic flora. Gut 42 (1), 2–7.

Bernalier, A., Doré, J., Durand, M., 1999. Colonic Microbiota, Nutrition and Health. Kluwer Academic Publishers, Ch. Biochemistry of fermentation, pp. 37–54.

Bernalier, A., Lellait, M., Rochet, V., Grivet, J. P., Gibson, G. R., Durand, M., 1996a. Acetogenesis from H_2 and CO_2 by methane- and non-methane- producing human colonic bacterial communities. FEMS Microbiol Ecol 19, 193–202.

Bernalier, A., Willems, A., Leclerc, M., Rochet, V., Collins, M. D., 1996b. Ruminococcus hydrogenotrophicus sp. nov., a new H_2/CO_2-utilizing acetogenic bacterium isolated from human feces. Arch Microbiol 166 (3), 176–183.

Bernalier-Donadille, A., 2004. Flore microbienne intestinale. Physiologie et pathologie digestives. Jhon Libbey Eurotext, Paris, Ch. Principales fonctions métaboliques de la flore intestinale de l'homme, pp. 61–80.

Bernard, O., Chachuat, B., Héllias, A., Rodriguez, J., 2006. Can we assess the model complexity for a bioprocess: theory and example of the anaerobic digestion process. Water Sci Technol 53 (1), 85–92.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A., Steyer, J. P., Nov 2001. Dynamical model development and parameter identification for an anaerobic wastewater treatment process. Biotechnol Bioeng 75 (4), 424–438.

Bernet, N., Sanchez, O., Cesbron, D., Steyer, J. P., Delegènes, J. P., 2005. Modeling and control of nitrite accumulation in a nitrifying biofilm reactor. Biochemical Engineering Journal 24, 173–183.

Bird, R. B., Stewart, W. E., Lightfoot, E. N., 1960. Transpot Phenomena. John Wiley & Songs, New York.

Birkett, A. M., Jones, G. P., de Silva, A. M., Young, G. P., Muir, J. G., 1997. Dietary intake and faecal excretion of carbohydrate by australians: importance of achieving stool weights greater than 150 g to improve faecal markers relevant to colon cancer risk. Eur J Clin Nutr 51 (9), 625–632.

Blanquet, S., Meunier, J. P., Minekus, M., Marol-Bonnin, S., Alric, M., May 2003. Recombinant Saccharomyces cerevisiae expressing p450 in artificial digestive systems: a model for biodetoxication in the human digestive environment. Appl Environ Microbiol 69 (5), 2884–2892.

Boever, P. D., Deplancke, B., Verstraete, W., Oct 2000. Fermentation by gut microbiota cultured in a simulator of the human intestinal microbial ecosystem is improved by supplementing a soygerm powder. J Nutr 130 (10), 2599–2606.

Bouhnik, Y., Flourié, B., D'Agay-Abensour, L., Pochart, P., Gramet, G., Durand, M., Rambaud, J. C., 1997. Administration of transgalacto-oligosaccharides increases fecal bifidobacteria and modifies colonic fermentation metabolism in healthy humans. J Nutr 127 (3), 444–448.

Bourriaud, C., Robins, R. J., Martin, L., Kozlowski, F., Tenailleau, E., Cherbut, C., Michel, C., 2005. Lactate is mainly fermented to butyrate by human intestinal microfloras but inter-individual variation is evident. J Appl Microbiol 99 (1), 201–212.

Box, G. E. P., Tiao, G. C., 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., Rohwer, F., 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol 185 (20), 6220–6223.

Breusegem, V. V., Bastin, G., 1991. Reduced order dynamical modelling of reaction systems: A singularperturbation approach. In: Proc. 30th IEEE Conference on Decision and Control, Brighton, UK. Vol. 2. pp. 1049–1054.

Brugère, J.-F., Mihajlovski, A., Missaoui, M., Peyret, P., 2009. Tools for stools: the challenge of assessing human intestinal microbiota using molecular diagnostics. Expert Rev Mol Diagn 9 (4), 353–365.

Cao, Y., Li, S., Linda, P., Radu, S., 2003. Adjoint sensitivity analysis for differential-algebraic equations: the adjoint DAE system and its numerical solution. SIAM Journal on Scientific Computing 24 (3), 1076–1089.

Caprice, A., 2000. The Expanded Quotable Einstein. Princeton University Press.

Castillo, A., Llabres, P., Mata-Alvarez, J., 1999. A kinetic study of a combined anaerobic-aerobic system for treatment of domestic sewage. Wat. Res. 33, 1742–1747.

Chan, L., Slater, J., Hasbargen, J., Herndon, D. N., Veech, R. I., Wolf, S., 1994. Neurocardiac toxicity of racemic d, l-lactate fluids. Int. Physiol. Behav. Sci 29, 38–394.

Chappell, M. J., Godfrey, K. R., Vajda, S., 1990. Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods. Math Biosci 102 (1), 41–73.

Chassard, C., Bernalier-Donadille, A., 2006. H2 and acetate transfers during xylan fermentation between a butyrate-producing xylanolytic species and hydrogenotrophic microorganisms from the human gut. FEMS Microbiol Lett 254 (1), 116–122.

Chassard, C., Goumy, V., Leclerc, M., Del'homme, C., Bernalier-Donadille, A., Jul 2007. Characterization of the xylan-degrading microbial community from human faeces. FEMS Microbiol Ecol 61 (1), 121–131.

Checchi, N., Giusti, E., Marsili-Libelli, S., 2007. PEAS: A toolbox to assess the accuracy of estimated parameters in environmental models. Environmental Modelling & Software 22 (6), 899–913.

Child, M. W., Kennedy, A., Walker, A. W., Bahrami, B., Macfarlane, S., Macfarlane, G. T., 2006. Studies on the efect of system retention time on bacterial populations colonizing a three-stage continuous culture model of the human large gut using FISH techniques. FEMS Microbial Ecol 55 (2), 299–310.

Christian, M. T., Amarri, S., Franchini, F., Preston, T., Morrison, D. J., Dodson, B., Edwards, C. A., Weaver, L. T., Feb 2002. Modeling 13c breath curves to determine site and extent of starch digestion and fermentation in infants. J Pediatr Gastroenterol Nutr 34 (2), 158–164.

Cinquin, C., Blay, G. L., Fliss, I., Lacroix, C., 2004. Immobilization of infant fecal microbiota and utilization in an in vitro colonic fermentation model. Microb Ecol 48 (1), 128–138.

Cinquin, C., Blay, G. L., Fliss, I., Lacroix, C., 2006a. Comparative effects of exopolysaccharides from lactic acid bacteria and fructo-oligosaccharides on infant gut microbiota

tested in an in vitro colonic model with immobilized cells. FEMS Microbiol Ecol 57 (2), 226–238.

Cinquin, C., Blay, G. L., Fliss, I., Lacroix, C., 2006b. New three-stage in vitro model for infant colonic fermentation with immobilized fecal microbiota. FEMS Microbiol Ecol 57 (2), 324–336.

Costello, D. J., Greenfield, P. F., Lee, P. L., 1991. Dynamic modelling of a single-stage high-rate anaerobic reactor- i. model derivation. Water Research 25 (7), 847–858.

Cummings, J. H., Macfarlane, G. T., Jun 1991. The control and consequences of bacterial fermentation in the human colon. J Appl Bacteriol 70 (6), 443–459.

Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P., Macfarlane, G. T., Oct 1987. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. Gut 28 (10), 1221–1227.

Cummings, J. H., Stephen, A. M., Dec 1980. The role of dietary fibre in the human colon. Can Med Assoc J 123 (11), 1109–1114.

Davis, S. A., Gordon, D. M., May 2002. The influence of host dynamics on the clonal composition of Escherichia coli populations. Environ Microbiol 4 (5), 306–313.

de Jong, P., Vissers, M. M. M., van der Meer, R., Bovee-Oudenhoven, I. M. J., Jan 2007. In silico model as a tool for interpretation of intestinal infection studies. Appl Environ Microbiol 73 (2), 508–515.

Degenring, D., Froemel, C., Dikta, G., Takors, R., 2004. Sensitivity analysis for the reduction of complex metabolism models. Journal of Process Control 14, 729–745.

Denis-Vidal, L., Joly-Blanchard, G., 2004. Equivalence and identifiability analysis of uncontrolled nonlinear dynamical systems. Automatica 40, 287–292.

Derrien, M., Vaughan, E. E., Plugge, C. M., de Vos, W. M., Sep 2004. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol 54 (Pt 5), 1469–1476.

Descartes, R., 1637. Discours de la méthode pour bien conduire sa raison, et chercher la vérité dans les sciences.

URL ExtractedfromTheProjectGutenbergcollection(http://www.gutenberg.org)

Dethlefsen, L., Eckburg, P. B., Bik, E. M., Relman, D. A., 2006. Assembly of the human intestinal microbiota. Trends Ecol Evol 21 (9), 517–523.

DiBaise, J. K., Zhang, H., Crowell, M. D., Krajmalnik-Brown, R., Decker, G. A., Rittmann, B. E., Apr 2008. Gut microbiota and its possible relationship with obesity. Mayo Clin Proc 83 (4), 460–469.

Diez-Gonzalez, F., Bond, D. R., Jennings, E., Russell, J. B., 1999. Alternative schemes of butyrate production in butyrivibrio fibrisolvens and their relationship to acetate utilization, lactate production, and phylogeny. Arch Microbiol 171 (5), 324–330.

Dochain, D., Bastin, G., 1990. On-line estimation and adaptive control of bioreactors. Elsevier, Amsterdam.

Dochain, D., Vanrolleghem, P., 2001. Dynamical Modelling and Estimation in Wastewater Treatment Processes. IWA Publishing, London.

Dochain, D., Vanrolleghem, P. A., Daele, M. V., 1995. Structural identifiability of biokinetic models of activated sludge respiration. Wat. Res. 29, 2571–2578.

Doeswijk, T. G., Keesman, K. J., Jan 2009. Linear parameter estimation of rational biokinetic functions. Water Res 43 (1), 107–116.

Doré, J., Pochart, P., Bernalier, A., Goderel, I., Morvan, B., Rambaud, J. C., 1995. Enumeration of H_2-utilizing methanogenic archaea, acetogenic and sulfate-reducing bacteria from human feces. FEMS Microbiol Ecol 17, 279–284.

Dumas, A., Dijkstra, J., France, J., 2008. Mathematical modelling in animal nutrition: a centenary review. Jounal of Agricultural Science 146, 123–142.

Dumont, M., Harmand, J., Rapaport, A., Godon, J.-J., 2009. Towards functional molecular fingerprints. Environ Microbiol.

Duncan, S. H., Barcenilla, A., Stewart, C. S., Pryde, S. E., Flint, H. J., Oct 2002a. Acetate utilization and butyryl coenzyme A (CoA):acetate-CoA transferase in butyrate-

producing bacteria from the human large intestine. Appl Environ Microbiol 68 (10), 5186–5190.

Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S., Flint, H. J., Sep 2002b. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. Int J Syst Evol Microbiol 52 (Pt 5), 1615–1620.

Duncan, S. H., Holtrop, G., Lobley, G. E., Calder, A. G., Stewart, C. S., Flint, H. J., 2004a. Contribution of acetate to butyrate formation by human faecal bacteria. Br J Nutr 91 (6), 915–923.

Duncan, S. H., Louis, P., Flint, H. J., 2004b. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Appl Environ Microbiol 70 (10), 5810–5817.

Duncan, S. H., Scott, K. P., Ramsay, A. G., Harmsen, H. J. M., Welling, G. W., Stewart, C. S., Flint, H. J., 2003. Effects of alternative dietary substrates on competition between human colonic bacteria in an anaerobic fermentor system. Appl Environ Microbiol 69 (2), 1136–1142.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., Relman, D. A., 2005. Diversity of the human intestinal microflora. Science 308, 1635–1638.

Egert, M., de Graaf, A. A., Smidt, H., de Vos, W. M., Venema, K., 2006. Beyond diversity: functional microbiomics of the human colon. Trends Microbiol 14 (2), 86–91.

Ewaschuk, J. B., Naylor, J. M., Zello, G. A., Jul 2005. D-lactate in human and ruminant metabolism. J Nutr 135 (7), 1619–1625.

Falony, G., Calmeyn, T., Leroy, F., Vuyst, L. D., 2009. Coculture fermentations of Bifidobacterium species and Bacteroides thetaiotaomicron reveal a mechanistic insight into the prebiotic effect of inulin-type fructans. Appl Environ Microbiol 75 (8), 2312–2319.

Falony, G., Vlachou, A., Verbrugghe, K., Vuyst, L. D., 2006. Cross-feeding between Bifidobacterium longum BB536 and acetate-converting, butyrate-producing colon bacteria during growth on oligofructose. Appl Environ Microbiol 72 (12), 7835–7841.

Fernández, A., Huang, S., Seston, S., Xing, J., Hickey, R., Criddle, C., Tiedje, J., Aug 1999. How stable is stable? function versus community composition. Appl Environ Microbiol 65 (8), 3697–3704.

Fernández, A. S., Hashsham, S. A., Dollhopf, S. L., Raskin, L., Glagoleva, O., Dazzo, F. B., Hickey, R. F., Criddle, C. S., Tiedje, J. M., Sep 2000. Flexible community structure correlates with stable community function in methanogenic bioreactor communities perturbed by glucose. Appl Environ Microbiol 66 (9), 4058–4067.

Finegold, S. M., Sutter, V. L., Mathisen, G. E., 1983. Human intestinal microflora in health and disease. Academic press, New York, Ch. Normal indigenous intestinal flora., pp. 3–31.

Fleming, S. E., Choi, S. Y., Fitch, M. D., Nov 1991. Absorption of short-chain fatty acids from the rat cecum in vivo. J Nutr 121 (11), 1787–1797.

Fletcher, R., 1987. Practical Methods of Optimization. John Wiley & Songs, Chichester.

Flint, H. J., Duncan, S. H., Scott, K. P., Louis, P., 2007. Interactions and competition within the microbial community of the human colon: links between diet and health. Environ Microbiol 9 (5), 1101–1111.

Fogler, H. S., 1999. Elements of Chemical Reaction Engineering. Prentice Hall.

Franceschini, G., Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. Chemical Engineering Science 63, 4846–4872.

Frank, D. N., Amand, A. L. S., Feldman, R. A., Boedeker, E. C., Harpaz, N., Pace, N. R., 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A 104 (34), 13780–13785.

Fricke, W. F., Seedorf, H., Henne, A., Krüer, M., Liesegang, H., Hedderich, R., Gottschalk, G., Thauer, R. K., 2006. The genome sequence of Methanosphaera stadtmanae reveals why this human intestinal archaeon is restricted to methanol and h2 for methane formation and atp synthesis. J Bacteriol 188 (2), 642–658.

Gavala, H. N., Angelidaki, I., Ahring, B. K., 2003. Advances in Biochemical Engineering/Biotechnology. Springer, Berlin, Ch. Kinetics and modeling of anaerobic digestion process, pp. 57–93.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis. Chapman and Hall/CRC, Boca Raton.

Gibson, G. R., Cummings, J. H., Macfarlane, G. T., Nov 1988. Use of a three-stage continuous culture system to study the effect of mucin on dissimilatory sulfate reduction and methanogenesis by mixed populations of human gut bacteria. Appl Environ Microbiol 54 (11), 2750–2755.

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., Nelson, K. E., 2006. Metagenomic analysis of the human distal gut microbiome. Science 312 (5778), 1355–1359.

Gloux, K., Leclerc, M., Iliozer, H., L'Haridon, R., Manichanh, C., Corthier, G., Nalin, R., Blottière, H. M., Doré, J., 2007. Development of high-throughput phenotyping of metagenomic clones from the human gut microbiome for modulation of eukaryotic cell growth. Appl Environ Microbiol 73 (11), 3734–3737.

Goebel, B. M., Stackebrandt, E., May 1994. Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. Appl Environ Microbiol 60 (5), 1614–1621.

Gonzalez-Gil, G., Seghezzo, L., Lettinga, G., Kleerebezem, R., 2001. Kinetics and mass-transfer phenomena in anaerobic granular sludge. Biotechnol Bioeng 73 (2), 125–134.

Goodacre, R., Jan 2007. Metabolomics of a superorganism. J Nutr 137 (1 Suppl), 259S–266S.

Goodwin, G., Payne, R., 1977. Dynamic System Identification: Experiment Design and Data Analysis. Academic Press, New York.

Gottschalk, G., 1986. Bacterial Metabolism. Springer, New York.

Haag, J. E., Wouwera, A. V., Queinnec, I., 2003. Macroscopic modelling and identification of an anaerobic waste treatment process. Chemical Engineering Science 58, 4307–4316.

Haddish-Berhane, N., Farhadi, A., Nyquist, C., Haghighi, K., Keshavarzian, A., Mar 2009. Simdot-abme: microphysiologically based simulation tool for quantitative prediction of systemic and local bioavailability of targeted oral delivery formulations. Drug Metab Dispos 37 (3), 608–618.

Hamer, H. M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F. J., Brummer, R.-J., 2008. Review article: the role of butyrate on colonic function. Aliment Pharmacol Ther 27 (2), 104–119.

Hanigan, M. D., Bateman, H. G., Fadel, J. G., McNamara, J. P., Mar 2006. Metabolic models of ruminant metabolism: recent improvements and current status. J Dairy Sci 89 Suppl 1, E52–E64.

Harmand, J., Rapaport, A., Dochain, D., Lobry, C., 2008. Microbial ecology and bioprocess control: opportunities and challenges. Journal of Process Control 18, 865–875.

Henze, M., Grady, C. P. L., Gujer, W., Marais, G., Matsuo, T., 1987. Activated sludge model no. 1. Tech. rep., IWAPRC Scientific and Technical Report.

Hindmarsh, A., R.Serban, 2006. User documentation for CVODES v2.5.0.
URL `http://www.llnl.gov/CASC/sundials/`

Hino, T., Miyazaki, K., Kuroda, S., 1991. Role of extracellular acetate in the fermentation of glucose by a ruminal bacterium, Megasphaera elsdenii. J. Gen. Appl. Microbiol. 37, 121–129.

Holdeman, L. V., Good, I. J., Moore, W. E., Mar 1976. Human fecal flora: variation in bacterial composition within individuals and a possible effect of emotional stress. Appl Environ Microbiol 31 (3), 359–375.

Holmberg, A., 1982. On the practical identifiability of microbial growth models incorporating michaelis-menten type nonlinearities. Math Biosci 62, 23–43.

Hooper, L. V., Gordon, J. I., 2001. Commensal host-bacterial relationships in the gut. Science 292 (5519), 1115–1118.

Hooper, L. V., Midtvedt, T., Gordon, J. I., 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu Rev Nutr 22, 283–307.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. COPASI–a COmplex PAthway simulator. Bioinformatics 22 (24), 3067–3074.

Hove, H., Nordgaard-Andersen, I., Mortensen, P. B., Mar 1994. Faecal dl-lactate concentration in 100 gastrointestinal patients. Scand J Gastroenterol 29 (3), 255–259.

Hsien, T. Y., Lin, Y. H., 2005. Biodegradation of phenolic wastewater in a fixed biofilm reactor. Biochemical Engineering Journal 27, 95–103.

Hucka, M., Finney, A., Bornstein, B., Keating, S., Shapiro, B., Matthews, J., Kovitz, B., Schilstra, M., Funahashi, A., Doyle, J., Kitano, H., 2004. Evolving a lingua franca and associated software infrastructure for computational systems biology: the systems biology markup language (sbml) project. IEEE Proceedings Systems Biology 1 (1), 41–53.

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Novère, N. L., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J., Forum, S. B. M. L., Mar 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19 (4), 524–531.

Hume, I. D., 1997. Gastrointestinal Microbiology vol I: Gastrointestinal ecosystems and fermentations. Chapman and Hall, New York, Ch. Fermentation in the hindgut of mammals, pp. 84–115.

Jackson, B. E., McInerney, M. J., 2002. Anaerobic microbial metabolism can proceed close to thermodynamic limits. Nature 415 (6870), 454–456.

Jeppsson, U., Olsson, G., 1993. Reduced order models for on-line parameter identification of the actvated sludge process. Water Sci Technol 28, 173–183.

Juillet, B., Saccomani, M. P., Bos, C., Gaudichon, C., Tomé, D., Fouillet, H., Dec 2006. Conceptual, methodological and computational issues concerning the compartmental modeling of a complex biological system: Postprandial inter-organ metabolism of dietary nitrogen in humans. Math Biosci 204 (2), 282–309.

Jumars, 2000. Animal guts as nonideal chemical reactors: Partial mixing and axial variation in absorption kinetics. Am Nat 155 (4), 544–555.

Kaneko, T., Bando, Y., Kurihara, H., Satomi, K., Nonoyama, K., Matsuura, N., Dec 1997. Fecal microflora in a patient with short-bowel syndrome and identification of dominant lactobacilli. J Clin Microbiol 35 (12), 3181–3185.

Keating, S., Bornstein, B. J., Finney, A., Hucka, M., 2006. SBML Toolbox: an SBML toolbox for MATLAB users. Bioinformatics 22 (10), 1275–1277.

Kell, D. B., Oliver, S. G., 2004. Here is the evidence, now what is the hypothesis? the complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays 26 (1), 99–105.

Kesavan, P., Law, V. J., 2005. Practical identifiability of parameters in Monod kinetics and statistical analysis of residuals. Biochemical Engineering Journal 24, 95–104.

Khalil, H. K., 2000. Nonlinear Systems, 3rd Edition. Prentice Hall, New Jersey.

Kirschner, D. E., Blaser, M. J., 1995. The dynamics of Helicobacter pylori infection of the human stomach. J Theor Biol 176 (2), 281–290.

Kitano, H., 2002a. Computational systems biology. Nature 420 (6912), 206–210.

Kitano, H., 2002b. Systems biology: a brief overview. Science 295 (5560), 1662–1664.

Kleerebezem, R., Stams, A. J., Mar 2000. Kinetics of syntrophic cultures: a theoretical treatise on butyrate fermentation. Biotechnol Bioeng 67 (5), 529–543.

Kleerebezem, R., van Loosdrecht, M. C. M., 2006. Critical analysis of some concepts proposed in ADM1. Water Sci Technol 54 (4), 51–57.

Klopfenstein, R. W., 1971. Numerical differentiation formulas for stiff systems of ordinary differential equations. RCA Review 32, 447–462.

Kong, Q., He, G. Q., Chen, F., Ruan, H., 2006. Studies on a kinetic model for butyric acid bioproduction by clostridium butyricum. Lett Appl Microbiol 43 (1), 71–77.

Krause, D. O., Gaskins, H. R., Mackie, R. I., 2006. Biofilms in the Food Environment. Blackwell Scientific, Ames, IA., Ch. Prokaryote diversity of gut mucosal biofilms in the human digestive tract, p. 127–151.

Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K., Srivastava, T. P., Taylor, T. D., Noguchi, H., Mori, H., Ogura, Y., Ehrlich, D. S., Itoh, K., Takagi, T., Sakaki, Y., Hayashi, T., Hattori, M., Aug 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res 14 (4), 169–181.

Lan, A., Bruneau, A., Philippe, C., Rochet, V., Rouault, A., Hervé, C., Roland, N., Rabot, S., Jan, G., 2007. Survival and metabolic activity of selected strains of propionibacterium freudenreichii in the gastrointestinal tract of human microbiota-associated rats. Br J Nutr 97 (4), 714–724.

Langlands, S. J., Hopkins, M. J., Coleman, N., Cummings, J. H., 2004. Prebiotic carbohydrates modify the mucosa associated microflora of the human large bowel. Gut 53 (11), 1610–1616.

Laux, D. C., Cohen, P. S., Conway, T., 2005. Colonization of Mucosal Surfaces. ASM Press, Washington DC, Ch. Role of the mucus layer in bacterial colonization of the intestine., pp. 199–212.

Le-Blay, G., Michel, C., Blottière, H. M., Cherbut, C., Nov 1999. Enhancement of butyrate production in the rat caecocolonic tract by long-term ingestion of resistant potato starch. Br J Nutr 82 (5), 419–426.

Leclerc, M., Bernalier, A., Donadille, G., Lelait, M., 1997. H_2/CO_2 metabolism in acetogenic bacteria isolated from the human colon. Anaerobe 3 (5), 307–315.

Legay, J. M., 1997. L'expérience et le modèle. INRA éditions.

Lei, F., Jørgensen, S. B., Jul 2001. Estimation of kinetic parameters in a structured yeast model using regularisation. J Biotechnol 88 (3), 223–237.

Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Doré, J., Leclerc, M., 2008. Dysbiosis in inflammatory bowel disease: a role for bacteriophages? Gut 57 (3), 424–425.

Lepage, P., Seksik, P., Sutren, M., de la Cochetière, M.-F., Jian, R., Marteau, P., Doré, J., May 2005. Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. Inflamm Bowel Dis 11 (5), 473–480.

Lettinga, G., van Velsen, A. F. M., Hobma, S. W., de Zeeuw, W., Klapwijk, A., 1980. Use of the upflow sludge blanket (USB) reactor concept for biological waste water treatment especially for anaerobic treatment. Biotechnol Bioeng 22, 699–734.

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., Gordon, J. I., 2005. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A 102 (31), 11070–11075.

Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., Gordon, J. I., Oct 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 6 (10), 776–788.

Ley, R. E., Peterson, D. A., Gordon, J. I., 2006a. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124 (4), 837–848.

Ley, R. E., Turnbaugh, P. J., Klein, S., Gordon, J. I., Dec 2006b. Microbial ecology: human gut microbes associated with obesity. Nature 444 (7122), 1022–1023.

Li, K.-B., 2003. Clustalw-mpi: Clustalw analysis using distributed and parallel computing. Bioinformatics 19 (12), 1585–1586.

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., Zhang, Y., Shen, J., Pang, X., Zhang, M., Wei, H., Chen, Y., Lu, H., Zuo, J., Su, M., Qiu, Y., Jia, W., Xiao, C., Smith, L. M., Yang, S., Holmes, E., Tang, H., Zhao, G., Nicholson, J. K., Li, L., Zhao, L., Feb 2008. Symbiotic gut microbes modulate human metabolic phenotypes. Proc Natl Acad Sci U S A 105 (6), 2117–2122.

Lichtenberger, L. M., 1995. The hydrophobic barrier properties of gastrointestinal mucus. Annu Rev Physiol 57, 565–583.

Lineweaver, H., Burk, D., 1934. The determination of enzyme dissociation constants. J. Am. Chem. Soc 56, 658–666.

Liu, Y., Xu, H.-L., Yang, S.-F., Tay, J.-H., 2003. Mechanisms and models for anaerobic granulation in upflow anaerobic sludge blanket reactor. Water Res 37 (3), 661–673.

Ljung, L., 2003. Educational aspects of identification software user interfaces. In: Proc. 13th Symposium on System Identification, Rotterdam, The Netherlands. pp. 1590–1594.

Lokshina, L. Y., Vavilin, V. A., 1999. Kinetic analysis of the key stages of low temperature methanogenesis. Ecological Modelling 117, 285–303.

Louis, P., Duncan, S. H., McCrae, S. I., Millar, J., Jackson, M. S., Flint, H. J., Apr 2004. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. J Bacteriol 186 (7), 2099–2106.

Louis, P., McCrae, S. I., Charrier, C., Flint, H. J., 2007a. Organization of butyrate synthetic genes in human colonic bacteria: phylogenetic conservation and horizontal gene transfer. FEMS Microbiol Lett 269 (2), 240–247.

Louis, P., Scott, K. P., Duncan, S. H., Flint, H. J., 2007b. Understanding the effects of diet on bacterial metabolism in the large intestine. J Appl Microbiol 102 (5), 1197–1208.

Lyberatos, G., Skiadas, I. V., 1999. Modelling of anaerobic digestion - a review. Global Nest: the Int. J. 1, 63–76.

Macfarlane, Macfarlane, Gibson, 1998a. Validation of a three-stage compound continuous culture system for investigating the effect of retention time on the ecology and metabolism of bacteria in the human colon. Microb Ecol 35 (2), 180–187.

Macfarlane, G. T., Cummings, J. H., 1991. The Large Intestine: Physiology, Pathophysiology and Disease. Raven Press Ltd, London, Ch. The colonic flora, fermentation, and large bowel digestive function, pp. 51–92.

Macfarlane, G. T., Gibson, G., 1997. Gastrointestinal Microbiology. Vol. 1. Chapman and Hall, London, Ch. Carbohydrate fermentation, energy transduction and gas metabolism in the human large intestine, p. 269–318.

Macfarlane, S., Macfarlane, G. T., 2003. Regulation of short-chain fatty acid production. Proc Nutr Soc 62 (1), 67–72.

Macfarlane, S., Macfarlane, G. T., Sep 2006. Composition and metabolic activities of bacterial biofilms colonizing food residues in the human gut. Appl Environ Microbiol 72 (9), 6204–6211.

Macfarlane, S., Macfarlane, G. T., Cummings, J. H., Sep 2006. Review article: prebiotics in the gastrointestinal tract. Aliment Pharmacol Ther 24 (5), 701–714.

Macfarlane, S., McBain, A. J., Macfarlane, G. T., Apr 1997. Consequences of biofilm and sessile growth in the large intestine. Adv Dent Res 11 (1), 59–68.

Macfarlane, S., Quigley, M. E., Hopkins, M. J., Newton, D. F., Macfarlane, G. T., 1998b. Polysaccharide degradation by human intestinal bacteria during growth under multi-substrate limiting conditions in a three-stage continuous culture system. FEMS Microbiology Ecology 26, 231–243.

Macfarlane, S., Woodmansey, E. J., Macfarlane, G. T., 2005. Colonization of mucin by human intestinal bacteria and establishment of biofilm communities in a two-stage continuous culture system. Appl Environ Microbiol 71 (11), 7483–7492.

Mackie, R. I., Sghir, A., Gaskins, H. R., May 1999. Developmental microbial ecology of the neonatal gastrointestinal tract. Am J Clin Nutr 69 (5), 1035S–1045S.

Madigan, M. T., Martinko, J. M., Dunlap, P. V., Clark, D. P., 2009. Brock Biology of Microorganisms, 12th Edition. Pearson Benjamin Cummings, USA.

Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R. K., Cantarel, B. L., Coutinho, P. M., Henrissat, B., Crock, L. W., Russell, A., Verberkmoes, N. C., Hettich, R. L., Gordon, J. I., 2009. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc Natl Acad Sci U S A 106 (14), 5859–5864.

Maiwald, T., Timmer, J., 2008. Dynamical modeling and multi-experiment fitting with potterswheel. Bioinformatics 24 (18), 2037–2043.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., Dore, J., 2006. Reduced diversity of faecal microbiota in crohn's disease revealed by a metagenomic approach. Gut 55 (2), 205–211.

Marsili-Libelli, S., Guerrizio, S., Checchi, N., 2003. Confidence regions of estimated parameters for ecological systems. Ecological Modelling 165, 127–146.

Marteau, P., Pochart, P., Doré, J., Béra-Maillet, C., Bernalier, A., Corthier, G., Oct 2001. Comparative study of bacterial groups within the human cecal and fecal microbiota. Appl Environ Microbiol 67 (10), 4939–4942.

Mathworks, 2008. Optimization Toolbox User's Guide.

Matsuo, K., Ota, H., Akamatsu, T., Sugiyama, A., Katsuyama, T., Jun 1997. Histochemistry of the surface mucous gel layer of the human colon. Gut 40 (6), 782–789.

May, R. M., Feb 2004. Uses and abuses of mathematics in biology. Science 303 (5659), 790–793.

McIntyre, A., Gibson, P. R., Young, G. P., Mar 1993. Butyrate production from dietary fibre and protection against large bowel cancer in a rat model. Gut 34 (3), 386–391.

Mendes, P., Kell, D., 1998. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics 14 (10), 869–883.

Mihajlovski, A., Alric, M., Brugère, J.-F., 2008. A putative new order of methanogenic archaea inhabiting the human gut, as revealed by molecular analyses of the mcra gene. Res Microbiol 159 (7-8), 516–521.

Miller, T. L., Wolin, M. J., Sep 1981. Fermentation by the human large intestine microbial community in an in vitro semicontinuous culture system. Appl Environ Microbiol 42 (3), 400–407.

Miller, T. L., Wolin, M. J., Feb 1982. Enumeration of methanobrevibacter smithii in human feces. Arch Microbiol 131 (1), 14–18.

Miller, T. L., Wolin, M. J., Mar 1985. Methanosphaera stadtmaniae gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. Arch Microbiol 141 (2), 116–122.

Miller, T. L., Wolin, M. J., May 1996. Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. Appl Environ Microbiol 62 (5), 1589–1592.

Minekus, M., 1998. Development and validation of a dynamic model of the gastrointestinal tract. Ph.D. thesis, University of Utrecht.

Minekus, M., Smeets-Peeters, M., Bernalier, A., Marol-Bonnin, S., Havenaar, R., Marteau, P., Alric, M., Fonty, G., in't Veld, J. H. H., Dec 1999. A computer-controlled system to simulate conditions of the large intestine with peristaltic mixing, water absorption and absorption of fermentation products. Appl Microbiol Biotechnol 53 (1), 108–114.

Moat, A. G., Foster, J. W., Spector, M. P., 2002. Microbial Physiology. John Wiley & Songs, New York.

Molly, K., Woestyne, M. V., Smet, I. D., Verstraete, W., 1994. Validation of the simulator of the human intestinal microbial ecosystem (SHIME) reactor using microorganism-associated activities. Microb Ecol Health Dis 7 (4), 191–200.

Molly, K., Wostyne, V., Verstraete, W., 1993. Development of a 5-step multi-chamber reactor as a simulation of the human intestinal microbial ecosystem. Appl. Microbiol. Biotechnol. 39, 254–258.

Monod, J., 1949. The growth of bacterial cultures. Annual Review of Microbiology 3, 371–394.

Moore, W. E., Holdeman, L. V., May 1974. Human fecal flora: the normal flora of 20 japanese-hawaiians. Appl Microbiol 27 (5), 961–979.

Morrison, D. J., Mackay, W. G., Edwards, C. A., Preston, T., Dodson, B., Weaver, L. T., Sep 2006. Butyrate production from oligofructose fermentation by the human faecal flora: what is the contribution of extracellular acetate and lactate? Br J Nutr 96 (3), 570–577.

Moser, G. A., Mclachlan, M. S., 2002. Modeling digestive tract absorption and desorption of lipophilic organic contaminants in humans. Environmental Science and Technology 36 (15), 3318–3325.

Mosey, F. E., 1983. Mathematical modelling of the anaerobic digestion process: regulatory mechanisms for the formation of short-chain volatile acids from glucose. Water Sci Technol 15, 209–232.

Muñoz-Tamayo, R., Laroche, B., Leclerc, M., Walter, E., 2008. Modelling and identification of *in vitro* homoacetogenesis by human-colon bacteria. In: Proc. 16th IEEE Mediterranean Conference on Control and Automation, Ajaccio, France. pp. 1717–1722.

Muñoz-Tamayo, R., Laroche, B., Leclerc, M., Walter, E., 2009. IDEAS: a parameter identification toolbox with symbolic analysis of uncertainty and its application to biological modelling. In: Prepints of the 15th IFAC Symposium on System Identification, Saint-Malo, France. pp. 1271–1276.
URL `http://www.inra.fr/miaj/public/logiciels/ideas/index.html`

Muñoz-Tamayo, R., Steyer, J. P., Laroche, B., Leclerc, M., 2007. Human colon: a complex bioreactor. conceptual modelling for the anaerobic digestion of the functional trophic chain. In: Proc. 11th World Congress Anaerobic Digestion Bio-energy for our Future, Brisbane, Australia. 6 pages on CD-Rom proceedings.

Nauman, E. B., 2002. Chemical Reactor Design, Optimization and Scaleup. McGraw-Hill, New York.

O'Hara, A. M., Shanahan, F., 2006. The gut flora as a forgotten organ. EMBO Rep 7 (7), 688–693.

Okino, M., Mavrovouniotis, M., 1998. Simplification of mathematical models of chemical reaction systems. Chem Rev 98 (2), 391–408.

Orcutt, R. P., Gianni, F. J., Judge., R. J., 1987. Development of an altered Schaedler flora for NCI gnotobiotic rodents. Microecol. Ther. 17, 59.

Papoutsakis, E. T., 1984. Equations and calculations for fermentations of butyric acid bacteria. Biotechnol Bioeng 26 (2), 174–187.

Pécou, E., 2005. Splitting the dynamics of large biochemical interaction networks. J Theor Biol 232 (3), 375–384.

Pearson, J. P., Brownlee, I. A., 2005. Colonization of Mucosal Surfaces,. ASM Press, Washington DC, Ch. Structure and function of mucosal surfaces., pp. 3–16.

Penumathsa, B. K. V., Premier, G. C., Kyazze, G., Dinsdale, R., Guwy, A. J., Esteves, S., Rodríguez, J., 2008. ADM1 can be applied to continuous bio-hydrogen production using a variable stoichiometry approach. Water Res 42 (16), 4379–4385.

Petersen, E. E., 1965. Chemical Reaction Analysis. Prentice Hall, Englewood Cliffs, New Jersey.

Phillips, M. L., May 2009. Gut reaction: environmental effects on the human microbiota. Environ Health Perspect 117 (5), A198–A205.

Picioreanu, C., Kreft, J. U., van Loosdrecht, M. C. M., May 2004. Particle-based multidimensional multispecies biofilm model. Appl Environ Microbiol 70 (5), 3024–3040.

Picioreanu, C., van Loosdrecht, M. C. M., 2003. Biofilms in Medicine, Industry and Environmental Biotechnology – Characteristics, Analysis and Control. IWA Publishing, Ch. Use of mathematical modelling to study biofilm development and morphology, pp. 413– 437.

Probert, H. M., Gibson, G. R., Sep 2002. Bacterial biofilms in the human gastrointestinal tract. Curr Issues Intest Microbiol 3 (2), 23–27.

Probet, H. M., Gibson, G. R., 2004. Development of a fermentation system to model sessile bacterial populations in the human colon. Biofilms 1, 13–19.

Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S., Flint, H. J., Dec 2002. The microbiology of butyrate formation in the human colon. FEMS Microbiol Lett 217 (2), 133–139.

Rajilic-Stojanovic, M., 2007. Diversity of the human gastrointestinal microbiota: novel perspectives from high throughput analyses. Ph.D. thesis, Wageningen University.

Rajilic-Stojanovic, M., Smidt, H., de Vos, W. M., 2007. Diversity of the human gastrointestinal tract microbiota revisited. Environ Microbiol 9 (9), 2125–2136.

Raksanyi, A., Lecourtier, Y., Walter, E., Venot, A., 1985. Identifiability and distinguishability testing via computer algebra. Math Biosci 77, 245–266.

Ramirez, I., Mottet, A., Carrère, H., Déléris, S., Vedrenne, F., Steyer, J. P., 2009a. Modified ADM1 disintegration/hydrolysis structures for modeling batch thermophilic anaerobic digestion of thermally pretreated waste activated sludge. Water Res 43 (14), 3479–3492.

Ramirez, I., Volcke, E. I. P., Rajinikanth, R., Steyer, J.-P., Jun 2009b. Modeling microbial diversity in anaerobic digestion through an extended ADM1 model. Water Res 43 (11), 2787–2800.

Ratkowsky, D. A., Dec 1986. A suitable parameterization of the michaelis-menten enzyme reaction. Biochem J 240 (2), 357–360.

Relman, D. A., 2008. 'til death do us part': coming to terms with symbiotic relationships. forward. Nat Rev Microbiol 6 (10), 721–724.

Robert, C., Bernalier-Donadille, A., 2003. The cellulolytic microflora of the human colon: evidence of microcrystalline cellulose-degrading bacteria in methane-excreting subjects. FEMS Microbiology Ecology 46, 81–89.

Robert, C. P., 2001. The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation. Springer, New York.

Rodríguez, J., Kleerebezem, R., Lema, J. M., van Loosdrecht, M. C. M., Feb 2006a. Modeling product formation in anaerobic mixed culture fermentations. Biotechnol Bioeng 93 (3), 592–606.

Rodríguez, J., Lema, J. M., Kleerebezem, R., 2008. Energy-based models for environmental biotechnology. Trends Biotechnol 26 (7), 366–374.

Rodríguez, J., Lema, J. M., van Loosdrecht, M. C. M., Kleerebezem, R., 2006b. Variable stoichiometry with thermodynamic control in ADM1. Water Sci Technol 54 (4), 101–110.

Rodríguez, J., Premier, G. C., Dinsdale, R., Guwy, A. J., 2009. An implementation framework for wastewater treatment models requiring a minimum programming expertise. Water Sci Technol 59 (2), 367–380.

Ropers, D., Baldazzi, V., de Jong, H., 2009. Reduction of a kinetic model of the carbon starvation response in Escherichia coli. In: Prepints of the 15th IFAC Symposium on System Identification, Saint-Malo, France. pp. 27–32.

Rosen, C., Vrecko, D., Gernaey, K. V., Pons, M. N., Jeppsson, U., 2006. Implementing ADM1 for plant-wide benchmark simulations in matlab/simulink. Water Sci Technol 54 (4), 11–19.

Saltelli, A., Chan, K., Scott, E. M., 2000. Sensitivity analysis. Wiley Series in Probaility and Statistics. John Wiley & Songs, Chichester.

Salyers, A. A., West, S. E., Vercellotti, J. R., Wilkins, T. D., Nov 1977. Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. Appl Environ Microbiol 34 (5), 529–533.

Samuel, B. S., Gordon, J. I., Jun 2006. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. Proc Natl Acad Sci U S A 103 (26), 10011–10016.

Sarma-Rupavtarm, R. B., Ge, Z., Schauer, D. B., Fox, J. G., Polz, M. F., 2004. Spatial distribution and stability of the eigth microbial species of the altered schaedler flora in the mouse gastrointetinal tract. Appl Environ Microbiol 70 (5), 2791–2800.

Sato, T., Matsumoto, K., Okumura, T., Yokoi, W., Naito, E., Yoshida, Y., Nomoto, K., Ito, M., Sawada, H., 2008. Isolation of lactate-utilizing butyrate-producing bacteria from human feces and in vivo administration of anaerostipes caccae strain l2 and galacto-oligosaccharides in a rat model. FEMS Microbiol Ecol 66 (3), 528–536.

Savage, D. C., 1977. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 31, 107–133.

Scanlan, P. D., Marchesi, J. R., 2008. Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. ISME J 2 (12), 1183–1193.

Scheppach, W., Weiler, F., 2004. The butyrate story: old wine in new bottles? Curr. Opin. Clin. Nutr. Metab. Care 7, 563–567.

Schink, B., Jun 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol Mol Biol Rev 61 (2), 262–280.

Schittkowski, K., 2002. Numerical Data Fitting in Dynamical Systems. Kluwer Academic Publishers.

Schmidt, H., 2007. Sbaddon: high performance simulation for the systems biology toolbox for MATLAB. Bioinformatics 23 (5), 646–647.

Schmidt, H., Jirstrand, M., 2006. Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics 22 (4), 514–515.

Seber, G. A. F., Wild, C. J., 1989. Nonlinear Regression. John Wiley & Songs, New york.

Sedoglavic, A., 2001. A probabilistic algorithm to test local algebraic observability in polynomial time. In: Proc. 2001 International Symposium on Symbolic and Algebraic Computation. pp. 309–316.
URL `http://www2.lifl.fr/~sedoglav/Software/ObservabilityTest/`

Seksik, P., Rigottier-Gois, L., Gramet, G., Sutren, M., Pochart, P., Marteau, P., Jian, R., Doré, J., 2003. Alterations of the dominant faecal bacterial groups in patients with crohn's disease of the colon. Gut 52 (2), 237–242.

Seksik, P., Sokol, H., Lepage, P., Vasquez, N., Manichanh, C., Mangin, I., Pochart, P., Doré, J., Marteau, P., 2006. Review article: the role of bacteria in onset and perpetuation of inflammatory bowel disease. Aliment Pharmacol Ther 24 Suppl 3, 11–18.

Shampine, L. F., Reichelt, M. W., 1997. The Matlab ODE suite. SIAM Journal on Scientific Computing 18, 1–22.

Skiadas, I. V., Gavala, H. N., Lyberatos, G., 2000. Modelling of the periodic anaerobic baffled reactor (PABR) based on the retaining factor concept. Wat. Res. 34 (15), 3725–3736.

Smith, J. M., Ness, H. C. V., Abbot, M. M., 2005. Introduction to Chemical Engineering Thermodynamics. McGraw-Hill, New York.

Smith, K., McCoy, K. D., Macpherson, A. J., 2007. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. Semin Immunol 19 (2), 59–69.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L. G., Gratadoux, J.-J., Blugeon, S., Bridonneau, C., Furet, J.-P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottière, H. M., Doré, J., Marteau, P., Seksik, P., Langella, P., 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of crohn disease patients. Proc Natl Acad Sci U S A 105 (43), 16731–16736.

Sonnenburg, J. L., Angenent, L. T., Gordon, J. I., 2004. Getting a grip on things: how do the communities of bacterial symbionts become established in our intestine? Nature Immunology 5 (6), 569–573.

Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C.-H., Westover, B. P., Weatherford, J., Buhler, J. D., Gordon, J. I., 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science 307 (5717), 1955–1959.

Soto-Cruz, O., Favela-Torres, E., Saucedo-Castañeda, G., 2002. Modeling of growth, lactate consumption, and volatile fatty acid production by Megasphaera elsdenii cultivated in minimal and complex media. Biotechnol Prog 18 (2), 193–200.

Spratt, P., Nicolella, C., Pylle, D. L., 2005. An engineer model of the human colon. Trans IChemE 83(C2), 147–157.

Squire, W., Trapp, G., 1998. Using complex variables to estimate derivatives of real functions. SIAM Review 40 (1), 110–112.

Steffens, M. A., Lant, P. A., Newell, R. B., 1997. A systematic approach for reducing complex biological wastewater treatment models. Wat. Res. 31, 590–606.

Stelling, J., 2004. Mathematical models in microbial systems biology. Curr Opin Microbiol 7 (5), 513–518.

Stephen, A. M., Cummings, J. H., 1980. The microbial contribution to human faecal mass. J Med Microbiol 13 (1), 45–56.

Stewart, M. L., Timm, D. A., Slavin, J. L., 2008. Fructooligosaccharides exhibit more rapid fermentation than long-chain inulin in an in vitro fermentation system. Nutr Res 28 (5), 329–334.

Strugala, V., Allen, A., Dettmar, P. W., Pearson, J. P., Feb 2003. Colonic mucin: methods of measuring mucus thickness. Proc Nutr Soc 62 (1), 237–243.

Stumm, W., Morgan, J. J., 1996. Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters. John Wiley & Songs, New York.

Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D., Doré, J., 1999. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl Environ Microbiol 65 (11), 4799–4807.

Swidsinski, A., Loening-Baucke, V., Theissig, F., Engelhardt, H., Bengmark, S., Koch, S., Lochs, H., Dörffel, Y., 2007. Comparative study of the intestinal mucus barrier in normal and inflamed colon. Gut 56 (3), 343–350.

Tannock, G. W., Munro, K., Harmsen, H. J., Welling, G. W., Smart, J., Gopal, P. K., Jun 2000. Analysis of the fecal microflora of human subjects consuming a probiotic product containing Lactobacillus rhamnosus dr20. Appl Environ Microbiol 66 (6), 2578–2588.

Tap, J., 2009. Impact du régime alimentaire sur la structure et la dynamique fonctionnelle du microbiote intestinal humain. Ph.D. thesis, Université Pierre & Marie Curie-Paris 6.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.-P., Ugarte, E., Muñoz-Tamayo, R., Paslier, D. L. E., Nalin, R., Dore, J., Leclerc, M., 2009. Towards the human intestinal microbiota phylogenetic core. Environ Microbiol 11 (10), 2574–2584.

Teusink, B., Smid, E. J., 2006. Modelling strategies for the industrial exploitation of lactic acid bacteria. Nat Rev Microbiol 4 (1), 46–56.

Thauer, R. K., Jungermann, K., Decker, K., Mar 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev 41 (1), 100–180.

Topping, D. L., Clifton, P. M., July 2001. Short-chain fatty acids and human colonic function: roles of resitant starch and nonstarch polysaccharides. Physiological Reviews 81 (3), 1031–1064.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath,

A. C., Knight, R., Gordon, J. I., 2009. A core gut microbiome in obese and lean twins. Nature 457 (7228), 480–484.

Turányi, T., Rabitz, H., 2000. Sensitivity Analysis. John Wiley & Songs, Chichester, Ch. Local Methods, pp. 81–99.

Vajda, S., Godfrey, K. R., Rabitz, H., 1989. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. Math Biosci 93 (2), 217–248.

Vajda, S., Rabitz, H., 1989. State isomorphism approach to global identifiability of nonlinear systems 34 (2), 220–223.

Valko, P., Vajda, S., 1984. An extented ODE solver for sensitivity calculations. Computers & Chemistry 8 (4), 255–271.

van de Wiele, T., Boon, N., Possemiers, S., Jacobs, H., Verstraete, W., 2004. Prebiotic effects of chicory inulin in the simulator of the human intestinal microbial ecosystem. FEMS Microbiol Ecol 51, 143–153.

van de Wiele, T., Boon, N., Possemiers, S., Jacobs, H., Verstraete, W., 2007. Inulin-type fructans of longer degree of polymerization exert more pronounced in vitro prebiotic effects. J Appl Microbiol 102 (2), 452–460.

van der Meulen., R., Makras, L., Verbrugghe, K., Adriany, T., Vuyst, L. D., 2006. In vitro kinetic analysis of oligofructose consumption by Bacteroides and Bifidobacterium spp. indicates different degradation mechanisms. Appl Environ Microbiol 72 (2), 1006–1012.

Vanhaecke, L., Grootaert, C., Verstraete, W., de Wiele, T. V., 2009. Chemopreventive effects from prebiotic inulin towards microbial 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine bioactivation. J Appl Microbiol 106 (2), 474–485.

Vanhoutte, T., Huys, G., Brandt, E., Swings, J., 2004. Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16s rrna gene primers. FEMS Microbiol Ecol 48 (3), 437–446.

Vanrolleghem, P. A., Daele, M. V., Dochain, D., 1995. Practical identifiability of a biokinetic model of activated sludge respiration. Water Research 29 (11), 2561–2570.

Vanrolleghem, P. A., Keesman, K. J., 1996. Identification of biodegradation models under model and data uncertainty. Water Sci Technol 33, 91–105.

Vavilin, V. A., Fernandez, B., Palatsi, J., Flotats, X., 2008. Hydrolysis kinetics in anaerobic degradation of particulate organic material: an overview. Waste Manag 28 (6), 939–951.

Vernia, P., Caprilli, R., Latella, G., Barbetti, F., Magliocca, F. M., Cittadini, M., Dec 1988. Fecal lactate and ulcerative colitis. Gastroenterology 95 (6), 1564–1568.

Volcke, E. I., Hulle, S. V., Deksissa, T., Zaher, U., Vanrolleghem, P., 2005. Calculation of ph and concentration of equilibrium components during dynamic simulation by means of a charge balance. Tech. rep., BIOMATH, Ghent University.
URL `http://biomath.ugent.be/~peter/ftp/pvr570.pdf`

Voolapalli, R. K., Stuckey, D. C., 1999. Relative importance of trophic group concentrations during anaerobic degradation of volatile fatty acids. Appl Environ Microbiol 65 (11), 5009–5016.

Vyshemirsky, V., Girolami, M. A., 2008. Bayesian ranking of biochemical system models. Bioinformatics 24 (6), 833–839.

Walker, A. W., Duncan, S. H., Harmsen, H. J. M., Holtrop, G., Welling, G. W., Flint, H. J., 2008. The species composition of the human intestinal microbiota differs between particle-associated and liquid phase communities. Environ Microbiol 10 (12), 3275–3283.

Walter, E., 1982. Identifiability of State Space Models. Springer, Berlin.

Walter, E. (Ed.), 1987. Identifiability of Parametric Models. Pergamon Press, Oxford.

Walter, E., Pronzato, L., 1990. Qualitative and quantitative experiment design for phenomenological models. Automatica 26, 195–213.

Walter, E., Pronzato, L., 1997. Identification of Parametric Models from Experimental Data. Springer, London.

Whitman, W. B., Coleman, D. C., Wiebe, W. J., 1998. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A 95 (12), 6578–6583.

Wilkinson, M. H. F., 2002. Model intestinal microflora in computer simulation: a simulatioand modelling package for host-micro flora interactions. IEEE Transactions on Biomedical Engineering 49 (10), 1077–1084.

Wilson, E. B., 1927. inference, the law of succession, and statistical inference. Journal of the American Statistical Association 22, 209–212.

Wilson, K. H., Ikeda, J. S., Blitchington, R. B., Sep 1997. Phylogenetic placement of community members of human colonic biota. Clin Infect Dis 25 Suppl 2, S114–S116.

Xavier, J., Picioreanu, C., van Loosdrecht, M. C. M., 2005a. A framework for multidimensional modelling of activity and structure of multispecies biofilms. Environ Microbiol 7 (8), 1085–1103.

Xavier, J., Picioreanu, C., van Loosdrecht, M. C. M., 2005b. A general description of detachment for multidimensional modelling of biofilms. Biotechnol Bioeng 91 (6), 651–669.

Xu, J., Gordon, J. I., 2003. Inaugural article: Honor thy symbionts. Proc Natl Acad Sci U S A 100 (18), 10452–10459.

Zi, Z., Klipp, E., 2006. SBML-PET: a systems biology markup language-based parameter estimation tool. Bioinformatics 22 (21), 2704–2705.

Zi, Z., Zheng, Y., Rundell, A. E., Klipp, E., 2008. SBML-SAT: a systems biology markup language (SBML) based sensitivity analysis tool. BMC Bioinformatics 9, 342.

Zoetendal, E. G., Akkermans, A. D., Vos, W. M. D., Oct 1998. Temperature gradient gel electrophoresis analysis of 16s rrna from human fecal samples reveals stable and host-specific communities of active bacteria. Appl Environ Microbiol 64 (10), 3854–3859.

Zoetendal, E. G., Cheng, B., Koike, S., Mackie, R. I., Sep 2004. Molecular microbial ecology of the gastrointestinal tract: from phylogeny to function. Curr Issues Intest Microbiol 5 (2), 31–47.

Zoetendal, E. G., Rajilic-Stojanovic, M., de Vos, W. M., Nov 2008. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. Gut 57 (11), 1605–1615.

Zoetendal, E. G., Vaughan, E. E., de Vos, W. M., 2006. A microbial world within us. Mol Microbiol 59 (6), 1639–1650.

Zoetendal, E. G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A. D. L., de Vos, W. M., 2002. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Appl Environ Microbiol 68 (7), 3401–3407.