

SIMPLIFIED APPROACHES TO MID-INFRARED SPECTROMETERS CALIBRATION FOR BIOPROCESS MONITORING APPLICATIONS

THÈSE N° 3917 (2007)

PRÉSENTÉE LE 23 NOVEMBRE 2007

À LA FACULTÉ DES SCIENCES DE BASE

LABORATOIRE DE GÉNIE CHIMIQUE ET BIOLOGIQUE

PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Jonas SCHENK

ingénieur chimiste diplômé EPF
de nationalité suisse et originaire de Coffrane (NE)

acceptée sur proposition du jury:

Prof. M. Mutter, président du jury
Prof. U. von Stockar, Dr I. W. Marison, directeurs de thèse
Prof. D. Bonvin, rapporteur
Prof. J. Nielsen, rapporteur
Prof. B. Sonnleitner, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2007

Remerciements

J'ai passé quatre années extraordinaires à effectuer cette thèse au LGCB, et si c'était à refaire, je n'hésiterais pas un instant. Cette période a été non seulement incroyablement formatrice et intéressante, mais elle m'a donné à vivre de très grands moments. Je tiens donc à remercier ici très sincèrement toutes les personnes qui ont contribué à l'aboutissement de cette thèse, par leurs conseils, leur support et leur enthousiasme.

En premier lieu, je tiens à chaleureusement remercier le Professeur von Stockar, mon directeur de thèse. J'ai joui d'une grande liberté pour effectuer ma recherche, tout en ayant l'assurance de bénéficier de son expertise scientifique, et je le remercie pour m'avoir témoigné pareille confiance. J'aimerais aussi dire qu'il a été très agréable d'avoir un chef toujours de bonne humeur, intéressé et intéressant bien au-delà du cadre de notre recherche, et qui a su créer une ambiance conviviale dans son groupe, notamment par les journées de ski et les séminaires hors-ville !

J'aimerais témoigner également ma reconnaissance au Professeur Ian Marison, pour son inébranlable enthousiasme, ainsi que sa grande disponibilité. Se préparer une tasse de thé au fond du couloir, c'était aussi l'occasion de faire le plein d'idées et de motivation !

Je voudrais aussi remercier ceux dont la contribution scientifique est moins visible. Ce travail doit en effet beaucoup au Dr. Jean-Marie Fuerbringer, qui m'a rappelé qu'il ne fallait que deux points pour tracer une droite ! J'aimerais aussi remercier Fred Meuwly, dont les bons principes de stratégie m'ont été grandement utiles.

Merci infiniment à ma collègue Carmen, avec qui j'ai passé plus de trois ans dans le même bureau. C'est facile de faire une thèse quand on partage son bureau avec quelqu'un qui est toujours de bonne humeur, positif, enthousiaste et motivé ! En plus d'avoir bien collaboré, on a passablement rigolé et partagé beaucoup de bons moments !

Un grand merci aussi à tous les collègues, les anciens, les moins anciens et les derniers survivants. Merci en particulier à Anne et Flo, les voisines qui nous ont tracé la voie, pour leurs sages conseils et les franches rigolades ! Merci à Michaël, pour avoir lutté, à mes côtés et avec courage, contre le FTIR, mais aussi pour les moments mémorables vécus au labo, à Sat ou en congrès. Merci à Patrick, pour son

humour et son aide précieuse. Merci à tous ceux qui m'ont motivé pendant cette thèse, par l'ambiance qu'ils mettaient au LGCB (Manu, Adona, Marcel, Agnes, Thierry, etc.). Merci aux diplômantes (Krisztina, Carole), aux stagiaires (Antoine, Julien, Carla), apprentis (Fabrice, Mireille) et étudiants qui ont travaillé sur le sujet avec moi : non seulement ils m'ont fourni une quantité appréciable de résultats, mais leur présence fut une grande motivation !

J'aimerais remercier tous ceux qui ont fourni un support appréciable à cette thèse, que ce soit logistique (Evelyne), informatique (PAP), mécanique ou électronique (les ateliers). Merci notamment à ceux qui m'ont montré les circuits de course à pied proches de l'EPFL !

Ce travail de thèse doit également beaucoup à tous ceux qui m'ont soutenu et encouragé pendant cette période ! Je tiens donc à remercier mes amis, mes proches et ma famille : ceux qui écoutent quand ça ne va pas, et avec qui on oublie vite les expériences qui n'ont pas marché. Plus particulièrement, j'aimerais remercier mon frère, avec qui j'ai fait pas mal de kilomètres ces dernières années, et qui n'a jamais hésité une seconde pour me donner un coup de main quand il fallait. Un grand merci à mes parents, pour leur soutien indéfectible et tout ce qu'ils m'ont transmis. Merci à toi Val, pour ta patience, ta confiance et tes encouragements !



Jonas Schenk

Lausanne, le 09 août 2007

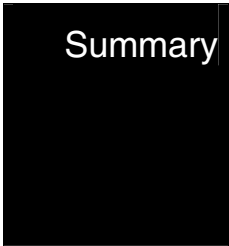
Table of Contents

Thesis

Summary	p. III
Preface	p. IX
Chapter 1 – Introduction	p. 1
Chapter 2 – The use of virtual calibrations to facilitate understanding of factor analysis	p. 11
Chapter 3 – Simplified Fourier-transform mid-infrared spectroscopy calibration based on a spectra library for the on-line monitoring of bioprocesses	p. 35
Chapter 4 – On-line monitoring of nine different batch cultures of <i>E. coli</i> by mid-infrared spectroscopy, using a single spectra library for calibration	p. 61
Chapter 5 – Mid-infrared spectroscopy as a tool for pH prediction and control in bioprocesses	p. 89
Chapter 6 – Effect of temperature on mid-infrared spectroscopy calibrations for the on-line monitoring of bioprocesses	p. 117
Chapter 7 – A simple method to monitor and control methanol feeding of <i>Pichia pastoris</i> fermentations using mid-infrared spectroscopy	p. 139
Chapter 8 – Conclusions and perspectives	p. 161

Supplementary material

Note on supplementary material	p. 175
Part A. The ongoing quest for truly on-line bioprocess monitoring using spectroscopy	p. 177
Part B. Influence of growth rate on specific productivity and glycosylation of a recombinant avidin produced by a <i>Pichia pastoris</i> Mut ⁺ strain	p. 205



Summary

Abstract & Résumé

ABSTRACT

This thesis aimed at developing new, simple methods of calibration for on-line monitoring of bioprocesses by Fourier-transform mid-infrared spectroscopy (FTIR). The conventional calibration approach implies the preparation and measurement of a large number of standards, usually around 50, and involves advanced mathematical tools, such as Principal Component Regression or Partial Least Squares for data treatment. This procedure is time-consuming, and requires, in addition, a fair level of expertise in linear algebra. For these reasons, there is a big need for calibration approaches that can be performed rapidly, by non-experts, in order to allow a routine operation of mid-infrared spectroscopy for the on-line monitoring of cultures.

Mid-infrared spectroscopy deals with the vibrational energy of molecules, which means that a vast majority of compounds can be detected in this wavelength range. Data treatment is therefore complex, due to strong peak overlapping, but also because of the low concentrations involved in culture media (< 3%) and unavoidable drift of the signal instrument. Multi-level calibration designs have been developed to tackle these problems, and they proved to achieve a robust process modeling. However, in the case where a method to correct for signal drift is available and compounds do not interact with each other –which is likely to happen in dilute solutions as such as culture media– multi-level calibration designs are clearly oversized, since a 2-level design provides the same quantity of information for a much lower work load.

It was shown in this research project that, for most common substances found in culture media, the absorbance is linear with respect to concentration (i.e. that the Lambert-Beer law is followed) and that species do not significantly interact with each other in solution. Based on this evidence, a relatively simple calibration method has been developed, which consists in correcting for signal drift through an anchorage method, and calculating the concentration using a library of pure component spectra as calibration set. Pure component refers here to a single compound dissolved in water at a concentration around 0.1 mol L⁻¹. Anchorage points were set in the spectra in regions where none of the compounds of interest absorb, but close enough from the calculation range to insure an appropriate correction. Batch cultures of the yeast *S. cerevisiae* were carried out to validate this approach, and it was shown that the results were as accurate and as robust as if they were found using the conventional approach (standard error of calibration of 0.86, 0.98, 0.15 g L⁻¹ for glucose, ethanol and ammonium

respectively). Pulse additions of compounds during the culture could be successfully monitored by the calibration, which therefore proved to be truly predictive.

A second calibration approach was developed and tested with batch cultures of the bacteria *E. coli*. While a spectra library was also used in this method to calculate concentrations from process data, signal instabilities were corrected for by including in the library a few “drift spectra”. These drift spectra were found by a Principal Component Analysis of water absorbance spectra, measured in aseptic conditions, every 5 minutes over a representative period of 24 hours. The least squares algorithm could therefore use the molar absorbance of the main metabolites as well as two drift spectra in order to calculate concentrations from process measurements. Rather than water, culture medium was used as reference intensity to calculate absorbances during the processes, which allows for including all compounds that present small concentration changes into a background that is eliminated by subtraction. Although this method led to concentration differences instead of absolute values, the same spectra library could be used to monitor different cultures, regardless of the medium composition, therefore saving additional time during calibration. A single library of 5 spectra, including three molar absorbances and two drift spectra, was able to monitor several batch cultures of *E. coli* performed in different media, with a precision similar to what could be expected from the conventional approach.

The effect of temperature and pH on these new calibration approaches was also studied. It was shown that the pH does not directly affect infrared spectra, and that it only influences deprotonation equilibria of weak acids, which in turn induce changes in the absorbance spectra. Temperature also proved not to interfere with the proposed calibration approaches.

The results presented in this thesis pave the way to the implementation of mid-infrared spectroscopy in a high-throughput platform for medium and strain screening. The calibration approaches that have been developed can certainly be easily automated, in order to provide a wealth of relevant information on the main metabolite concentration at a very low labour cost.

Keywords: Mid-infrared spectroscopy; On-line monitoring;
Bioprocess control; Spectra library;

RÉSUMÉ

Ce travail de thèse avait pour objectif le développement de méthodes simples pour la calibration des spectromètres infrarouges en vue de leur utilisation pour le suivi en ligne des bioprocédés. L'approche traditionnelle, qui consiste à préparer et mesurer un grand nombre de standards - typiquement plus de 50 - et à construire un modèle de calibration à l'aide d'outils mathématiques sophistiqués comme la Régression en Composantes Principales, implique un fastidieux travail de laboratoire et nécessite de la part de l'utilisateur une bonne maîtrise de l'algèbre linéaire. Il existe donc une grande demande pour des approches alternatives, passant par des calibrations rapides et simples sur le plan mathématique, afin de pouvoir utiliser la spectroscopie infrarouge de manière routinière.

La grande majorité des molécules peut être détectée par spectroscopie infrarouge, puisque ce domaine de longueurs d'onde correspond aux changements d'énergie vibrationnelle des liaisons chimiques. Si cela présente l'avantage de permettre l'analyse de presque tous les composés, il en résulte des spectres difficiles à interpréter, à cause du chevauchement des différentes contributions. Le traitement des données est rendu encore plus complexe par le fait que les concentrations en jeu lors des cultures sont extrêmement faibles ($< 3\%$) et que le signal de l'instrument est sujet à une importante dérive. Afin de surmonter ces problèmes, des calibrations se basant sur les plans d'expériences fractionnaires à plusieurs niveaux ont été élaborées. Ces plans garantissent une certaine solidité au modèle, mais ils s'avèrent cependant surdimensionnés dans le cas où l'on dispose d'une méthode de correction de la dérive du signal et que les composés n'interagissent pas entre eux - une hypothèse généralement vérifiée pour les solutions diluées que sont les milieux de culture. Dans ce cas de figure, un plan d'expérience à deux niveaux fournit autant d'information qu'un plan à plusieurs niveaux, mais demande une bien moindre quantité de travail.

Il a été montré avec des composés usuels de milieu de culture que l'absorbance est parfaitement linéaire en fonction de la concentration (en d'autres termes, que la loi de Lambert-Beer est respectée), et qu'elle est de plus additive, ce qui sous-entend que les différents composés n'interagissent pas entre eux. Sur cette base, une méthode de calibration relativement simple a été développée, qui consiste à, d'une part, éliminer la dérive du signal en utilisant des points d'ancrage, et d'autre part, à modéliser les variations d'absorbance durant le procédé à l'aide d'une bibliothèque de spectres de composés purs. A noter que l'état entendu comme pur ici fait référence à un composé dissout dans l'eau, à une concentration avoisinant 0.1 mol L^{-1} . Les points d'ancrage ont été fixés dans des zones du spectre où aucun

des composés d'intérêt n'absorbe, mais suffisamment proches de la zone utilisée pour les calculs afin d'obtenir une correction efficace. Des cultures aérobiques en bioréacteur fermé (batch) de *S. cerevisiae* ont été conduites afin de valider cette approche, et il a été montré que les résultats étaient aussi précis que s'ils avaient été fournis par l'approche conventionnelle (erreurs standard de calibration de 0.86, 0.98, 0.15 g L⁻¹, respectivement pour le glucose, l'éthanol et l'ammonium).

Une deuxième méthode de calibration a été développée et testée avec des cultures de la bactérie *E. coli*. Cette deuxième méthode, similairement à la première, utilise une bibliothèque de spectres pour la modélisation des données du procédé. Elle se distingue cependant par une approche différente de la correction de la dérive du signal. La dérive a été en premier lieu caractérisée par une Décomposition en Composantes Principales d'une collection de spectres pris dans de l'eau stérile à température contrôlée pendant 24 heures, et donc représentatifs de la dérive du signal. Les deux principales caractéristiques spectrales de cette dérive ont été intégrées à la bibliothèque des spectres. En utilisant à nouveau la méthode des moindres carrés, les spectres de culture ont pu être modélisés comme une somme de concentration des composés d'intérêt, ainsi que de deux « quantités de dérive ». Cette méthode a été testée avec neuf différentes cultures de souches sauvages de *E. coli*, faites dans des milieux de culture de compositions variées, et contenant notamment différentes sources de carbone. A nouveau, les erreurs standards de prédiction étaient dans la gamme de ce qui aurait pu être obtenu grâce à l'approche traditionnelle. A noter que l'intégration dans une ligne de base de tous les composés dont la concentration ne varie guère durant le procédé a permis l'utilisation d'une seule et même bibliothèque de 5 spectres, en lieu et place de plus d'une centaine de standards s'il avait fallu procéder par la méthode traditionnelle.

L'effet de la température et du pH sur ces méthodes de calibration a été également étudié, en vue d'une généralisation de l'approche proposée. Il a été montré que le pH n'a pas d'influence directe sur les spectres infrarouges, et que son effet peut être parfaitement expliqué par les changements d'équilibre acide-base. Il a également été montré que la température n'interfère guère avec les méthodes de calibration développées.

Les résultats présentés dans cette thèse ouvrent de nouveaux horizons à la spectroscopie infrarouge. Les méthodes de calibration proposées permettent d'envisager l'intégration de la spectrométrie infrarouge dans des plateformes dédiées aux tests massivement parallèles des souches et de milieu de culture, comme moyen non-invasif de mesure des concentrations.

Mots-clés : Spectroscopie infrarouge; Suivi en ligne des bioprocédés ;
Bibliothèque de spectres ; Méthodes de calibration

Preface

What Varnish Production and
Infrared Spectroscopy Can
Have in Common

Primo Levi is best known for being the author of *If This Is a Man*, this stunning account of the prisoner's life in Auschwitz. But he was also a passionate chemist, and despite the racial laws, he completed a doctorate with merit in 1941. As he pointed it out later on, this had an enormous impact on his life, since chemistry shaped his mind and helped him survive the extermination camp¹.

Primo Levi wrote several books that are related to chemistry. *The Periodic Table* is a collection of short stories that are all connected in some ways to chemical elements. Most of them are true and depict, with a subtle sense of humor and a great clarity, Primo Levi's experiences in the chemical industry, or souvenirs of graduation courses and practical classes.

I particularly like in this book the 12th story, entitled *Chromium*. Primo Levi shows how a blind respect of procedures can, as time goes by, result in awkward ways of doing things. I think that every scientist has been confronted once in his career with such a situation, and this story pleasantly encourages challenging all established theories and practices.

In my opinion, calibration procedures for bioprocess monitoring by mid-infrared spectroscopy have also not been called enough into question. These methods were developed for the analysis of undefined mixtures by near-infrared spectroscopy, but they did not really evolve when they were brought into the field bioprocess monitoring, and further on applied to mid-infrared spectrometers.

As an appetizer, I therefore propose an excerpt from Primo Levi's short story *Chromium*². This will, by far, not replace a complete reading, but should nevertheless show what varnish production and infrared calibration can have in common!

"The entrée was fish, but the wine was red. Versino, head of maintenance, said that it was all a lot of nonsense, provided the wine and fish were good; he was certain that the majority or those who upheld the orthodox view could not, blindfolded, have distinguished a glass of white wine from a glass of red. Bruni, from the Nitro Department, asked whether somebody knew why fish goes with white wine: various joking remarks were made but nobody was able to answer properly. Old man Cometto added that life is full of customs whose roots can no longer be traced: the color of sugar paper, the buttoning from different sides for men and women, the shape of a gondola's prow, and the innumerable alimentary compatibilities and incompatibilities, of which in fact the one in question was a particular case: but in any event, why were pig's feet obligatory with lentils, and cheese on macaroni.

¹ *Primo Levi ou la Tragédie d'un optimiste*, Myriam Anissimov, Ed : LGF – Livre de Poche, 2003

² In *The Periodic Table*, by Primo Levi. Reproduced with kind permission of Schocken Books Inc., owner of the english translation copyright.

I made a rapid mental review to be sure that none of those present had as yet heard it, then I started to tell the story of the onion in the boiled linseed oil. [...] I told my companions at table that in a prescription book published about 1942 I had found the advice to introduce into the oil, toward the end of the boiling, two slices of onion, without any comment on the purpose of this curious additive. I had spoken about it in 1949 with Signor Giacomasso Olindo, my predecessor and teacher, who was then more than seventy and had been making varnishes for fifty years, and he, smiling benevolently behind his thick white mustache, had explained to me that in actual fact, when he was young and boiled the oil personally, thermometers had not yet come into use: one judged the temperature of the batch by observing the smoke, or spitting into it, or, more efficiently, immersing a slice of onion in the oil on the point of a skewer; when the onion began to fry, the boiling was finished. Evidently, with the passing of the years, what had been a crude measuring operation had lost its significance and was transformed into a mysterious and magical practice.

[...]

Bruni told us about an episode in which he himself had been involved, and as he told the story, I felt myself invaded by sweet and tenuous sensations which later I will try to explain. I must say first of all that Bruni worked from 1955 to 1965 in a large factory on the shores of a lake, the same one in which I had learned the rudiments of the varnish-making trade during the years 1946-47. So he told us that, when he was down there in charge of the Synthetic Varnishes Department, there fell into his hands a formula of a chromate-based anti-rust paint that contained an absurd component: nothing less than ammonium chloride, the old, alchemical sal ammoniac of the temple of Ammon, much more apt to corrode iron than preserve it from rust. He had asked his superiors and the veterans in the department about it: surprised and a bit shocked, they had replied that in that formulation, which corresponded to at least twenty or thirty tons of the product a month and had been in force for at least ten years, that salt "had always been in it," and that he had his nerve, so young in years and new on the job, criticizing the factory's experience, and looking for trouble by asking silly hows and whys. If ammonium chloride was in the formula, it was evident that it had some sort of use. What use it had nobody any longer knew, but one should be very careful about taking it out because "one never knows." Bruni is a rationalist, and he took all this very badly; but he is a prudent man, and so he accepted the advice, according to which in that formulation and in that lakeshore factory, unless there have been further developments, ammonium chloride is still being put in; and yet today it is completely useless, as I can state from firsthand experience because it was I who introduced it into the formula.

The episode cited by Bruni, the rustproof formula with chromates and ammonium chloride, flung me back in time, all the way to the freezing cold January of 1946, when meat and coal were still rationed, nobody had a car, and never in Italy had people breathed so much hope and so much freedom.

But I had returned from captivity three months before and was living badly. The things I had seen and suffered were burning inside of me; I felt closer to the dead than the living, and felt guilty at being a man, because men had built Auschwitz, and Auschwitz had gulped down millions of human beings, and many of my friends, and a woman who was dear to my heart. It seemed to me that I would be purified if I told its story, and I felt like Coleridge's Ancient Mariner, who waylays on the street the wedding guests going to the feast, inflicting on them the story of his misfortune. I was writing concise and bloody poems, telling the story at breakneck speed, either by talking to people or by writing it down, so much so that gradually a book was later born: by writing I found peace for a while and felt myself become a man again, a person like everyone else, neither a martyr nor debased nor a saint: one of those people who form a family and look to the future rather than the past.

Since one can't live on poetry and stories, I looked feverishly for work and found it in the big lakeshore factory, still damaged from the war, and during those months besieged by mud and ice. Nobody was much concerned with me: colleagues, the director, and workers had other things to think about—the son who wasn't returning from Russia, the stove without wood, the shoes without soles, the warehouses without supplies, the windows without panes, the freezing cold which split the pipes, inflation, famine, and the virulent local feuds. I had been benignly granted a lame-legged desk in the lab, in a corner full of crashing noise, drafts, and people coming and going carrying rags and large cans, and I had not been assigned a specific task. I, unoccupied as a chemist and in a state of utter alienation (but then it wasn't called that), was writing in a haphazard fashion page after page of the memories which were poisoning me, and my colleagues watched me stealthily as a harmless nut. The book grew under my hands, almost spontaneously, without plan or system, as intricate and crowded as an anthill. Every so often, impelled by a feeling of professional conscience, I would ask to see the director and request some work, but he was much too busy to worry about my scruples. I should read and study; when it came to paints and varnishes I was still, if I didn't mind his saying so, an illiterate. I didn't have anything to do? Well, I should praise God and sit in the library; if I really had the itch to do something useful, well, look, there were articles to translate from German.

One day he sent for me and with an oblique glint in his eyes announced that he had a little job for me. He took me to a corner of the factory's yard, near a retaining wall: piled up at random, the lowest crushed by the highest, were thousands of square blocks of a bright orange color. He told me to touch them: they were gelatinous and softish; they had the disagreeable consistency of slaughtered tripes. I told the director that, apart from the color, they seemed to me to be livers, and he praised me: that's just how it was described in the paint manuals! He explained that the phenomenon which had produced them was called just that in English, "livering"; under certain conditions certain paints turned from liquids into solids, with the consistency precisely of the liver or lungs, and must be thrown out. These parallelepiped shares had been cans of paint: the paint had livered, the cans had been cut away, and the contents had been thrown on the garbage dump.

The paint, he told me, had been produced during the war and immediately after; it contained a basic chromate and alkyd resin. Perhaps the chromate was too basic or the resin too acidic: these were exactly the conditions under which a "livering" can take place. All right, he made me the gift of that pile of old sins; I should think about it, make tests and examinations, and try to say with precision why the trouble had occurred, what should be done so that it was not repeated, and if it were possible to reclaim the damaged goods.

[...]

The first skirmish took place in the archives. The two partners, the two fornicators from whose embrace had sprung our orange-colored monsters, were the chromate and the resin. The resin was fabricated on the spot: I found the birth certificate of all the batches, and they did not offer anything suspicious; the acidity was variable, but always inferior to 6, as prescribed. One batch that was found to have a pH of 6.2 had been dutifully discarded by an inspector with a flowery signature. In the first instance the resin could not be faulted.

The chromate had been purchased from different suppliers, and it too had been duly inspected batch by batch. According to Purchase Specification 480/0 it should have contained not less than 28 percent of chromium oxide in all; and now here, right before my eyes I had the interminable list of tests from January 1942 until today (one of the least exciting forms of reading imaginable), and all the values satisfied the specification, indeed were equal among themselves: 29.5 percent, not one percent more, not one less. I felt my inner being as a chemist writh confronted by that abomination; in fact, one should know that the natural oscillations in the method of preparation of such chromate, added to the inevitable analytical errors, make extremely improbable that the many values found in different batches and on different days could coincide so exactly. How come nobody had gotten suspicious? But in fact at that time I

did not yet know the frightening anesthetic power of company papers, their capacity to hobble, douse, and dull every leap of intuition and every spark of talent. It is well known to the scholarly that all secretions can be harmful or toxic: now under pathological conditions it is not rare that the paper, a company secretion, is reabsorbed to an excessive degree, and puts to sleep, paralyzes, or actually kills the organism from which it has been exuded.

The story of what had happened began to take shape. For some reason, some analyst had been betrayed by a defective method, or an impure reagent, or an incorrect habit; he had diligently totted up those so obviously suspicious but formally blameless results; he had punctiliously signed each analysis, and his signature, swelling like an avalanche, had been consolidated by the signatures of the lab chief, the technical director, and the general director. I could see him, the poor wretch, against the background of those difficult years: no longer young, since all the young men were in the military services; perhaps chivied by the Fascists, perhaps himself a Fascist being looked for by the partisans; certainly frustrated, because being an analyst is a young man's job; on guard in his lab within the fortress of his minuscule specialty, since the analyst is by definition infallible and derided and regarded with a hostile eye outside the lab just because of his virtues as an incorruptible guardian, a severe, pedantic, unimaginative little judge, a stick poked in the wheels of production. To judge from the anonymous, neat handwriting, his trade must have exhausted him and at the same time brought him to a crude perfection, like a pebble in a mountain stream that has been twirled over and over all the way to the stream's mouth. It was not surprising that, with time, he had developed a certain insensitivity to the real significance of the operations he was performing and the notes he was writing. I planned to look into his particular case but nobody knew anything more about him; my questions were met with discourteous or absentminded replies. Moreover, I was beginning to feel around me and my work a mocking and malevolent curiosity: who was this Johnny-come-lately, this pipsqueak earning 7,000 lire a month, this maniac scribbler who was disturbing the nights of the guest quarters typing away at God knows what, and sticking his nose into past mistakes and washing a generation's dirty linen? [...]

It was not hard for me to procure, besides the Purchase Specification (the PS), also the equally inviolable CS, the Checking Specifications: in a drawer in the lab there was a packet of greasy file cards, typewritten and corrected several times by hand, each of which contained the way to carry out a check of a specific raw material. The file card on prussian blue was stained with blue, the file card on glycerine was sticky, and the file card on fish oil smelled like sardines. I took out the file card on chromate, which due to long use had become the

color of a sunrise, and read it carefully. It was all rather sensible and in keeping with my not-so-far-off scholastic notions; only one point seemed strange to me. Having achieved the disintegration of the pigment, it prescribed adding twenty-three drops of a certain reagent. Now, a drop is not so definite a unit as to entail so definite a numerical coefficient; and besides, when all is said and done, the prescribed dose was absurdly high: it would have flooded the analysis, leading in any case to a result in keeping with the specification. I looked at the back of the file card: it bore the date of the last review, January 4, 1944; the birth certificate of the first livered batch was on the succeeding February 22.

At this point I began to see the light. In a dusty archive I found the CS collection no longer in use, and there, lo and behold, the preceding edition of the chromate file card bore the direction to add "2 or 3" drops, and not "23": the fundamental "or" was half erased and in the next transcription had gotten lost. The events meshed perfectly: the revision of the file card had caused a mistake in transcription, and the mistake had falsified all succeeding analyses, concealing the results on the basis of a fictitious value due to the reagent's enormous excess and thus bringing about the acceptance of shipments of pigment which should have been discarded; these, being too basic, had brought about the livering.

But there is trouble in store for anyone who surrenders to the temptation of mistaking an elegant hypothesis for a certainty: the readers of detective stories know this quite well. I got hold of the sleepy man in charge of the storeroom, requested from him all the samples of all the shipments of chromate from January 1944 on, and barricaded myself behind a workbench for three days in order to analyze them according to the incorrect and correct methods. Gradually, as the results lined up in a column on the register, the boredom of repetitious work was being transformed into nervous gaiety, as when as children you play hide and seek and discover your opponent clumsily squatting behind a hedge. With the mistaken method you constantly found the fateful 29.5 percent; with the correct method, the results were widely dispersed, and a good quarter, being inferior to the prescribed minimum, corresponded to the shipments which should have been rejected. The diagnosis was confirmed, the pathogenesis discovered: it was now a matter of defining the therapy.

This was found pretty soon, drawing on good inorganic chemistry, that distant Cartesian island, a lost paradise, for us organic chemists, bunglers, "students of gunks": it was necessary to neutralize in some way, within the sick body of that varnish, the excess of basicity due to free lead oxide. The acids were shown to be noxious from other aspects: I thought of ammonium chloride, capable of combining stably with lead oxide, producing an insoluble and inert chloride and freeing the ammonia. Tests on a small scale gave promising results:

now quick, find the chloride, come to an agreement with the head of the Milling Department, slip into a small ball mill two of the liven disgusting to see and touch, add a weighed quantity of the presumed medicine, start the mill under the skeptical eyes of the onlookers. The mill, usually so noisy, started almost grudgingly, in a silence of bad omen, impeded by the gelatinous mass which stuck to the balls. All that was left was to go back to Turin to wait for Monday. [...]

The following Monday the mill had regained its voice: it was in fact crunching away gaily with a full, continuous tone, without that rhythmic roaring that in a ball mill indicates bad maintenance or bad health. I stopped it and cautiously loosened the bolts on the manhole; there spurted out with a hiss an ammoniacal puff, as it should. Then I took off the cover. Angels and ministers of grace!-the paint was fluid and smooth, completely normal, born again from its ashes like the Phoenix. I wrote out a report in good company jargon and the manager increased my salary. Besides, as a form of recognition, I received the assignment of two tyres for my bike

Since the storeroom contained several shipments of perilously basic chromate, which must be utilized because they had been accepted by the inspection and could not be returned to the supplier, the chloride was officially introduced as an antilivering preventive in the formula of that varnish. Then I quit my job: ten years went by, the postwar years were over, the deleterious, too basic chromates disappeared from the market, and my report went the way of all flesh: but formulas are as holy as prayers, decree-laws, and dead languages, and not an iota in them can be changed. And so my ammonium chloride, the twin of a happy love and a liberating book, by now completely useless and probably a bit harmful, is religiously ground into the chromate anti-rust paint on the shore of that lake, and nobody knows why anymore."

Chapter 1

Introduction

1/ INTRODUCTION

Why struggling with on-line monitoring techniques?

When one has to monitor a bioprocess and can go for either on-line or off-line monitoring, it is very likely that an off-line solution will be chosen. While this implies a lot of work for withdrawing and filtering samples, and carrying out HPLC or enzymatic analyses, results are generally reliable, and if there is any doubt, samples can be tested again or differently. Opting for on-line monitoring is more risky. Time must be invested in developing a method, for which robustness and reliability must be thoroughly addressed. The entire analytical procedure must be automated through a multi-disciplinary project involving hardware and software implementation. However, the time investment may be worthwhile, if the process is to be repeated a certain number of times and little work is necessary for supervision. Other motivations to choose an on-line monitoring approach can be the need of a quick response for process control, restrictions on sampling for security reasons or because of broth viscosity, or also the will to get results as quickly as possible to accelerate process development.

Due to the evolution of the life science industry, the pressure is currently high on two such applications that require on-line monitoring. More and more strains and media are tested in early bioprocess development, and for that purpose, micro-scale set-up have been developed for high-throughput screening. The quantity of relevant information that can be obtained by these massive screenings remains limited, which means that one of the next challenges will consist in implementing on-line monitoring tools that can give access to concentration profiles and product titer. For such high-throughput applications, on-line techniques must be non-invasive, and must easily be multiplexed in terms of both hardware and work load. Accuracy and robustness are in this case less critical criteria. The second main opening for on-line monitoring tools, in the foreseeable future, is certainly within the deployment of Process Analytical Technologies, which is encouraged by regulation authorities to improve drug safety and quality. As opposed to high-throughput applications, process control applications require the highest level of robustness and accuracy, simplicity of handling, and to a lesser extend, low work load for calibration.

These prospects for on-line monitoring make the framework of the current thesis. This work is part of a long-term research effort that has for objective to make of spectroscopic techniques, and in particular mid-infrared spectroscopy, a well-established monitoring tool in the afore-mentioned applications.

On the need of simplified calibrations for mid-infrared spectroscopy

Among the very many techniques available for the on-line monitoring of bioprocesses, Fourier-transform mid-infrared spectroscopy (FTIR) presents several interesting advantages. Mid-IR deals with the vibrational energy of molecular bonds, which means that any compound can be detected in this wavelength domain. This confers an enormous potential to mid-IR regarding the diversity of product manufactured in the present day. In addition, it is fast (a spectrum can be acquired every 20 seconds), non-invasive and robust from an industrial point of view, since it can be easily cleaned and sterilized in place.

More than ten years ago, infrared spectroscopy was presented as a technique “with appealing characteristics, but not yet mature enough to be routinely applied for on-line fermentation monitoring” (Konstantinov et al., 1996). Unfortunately, this statement still largely holds after a decade of research in the field, and it must be admitted that mid-IR monitoring has not yet become a popular and widely used technique. A recent review even presented the “potential” of mid-infrared spectroscopy, rather than enumerating successful industrial applications (Roychoudhury et al., 2006). It is now clear that apart from the cost of these instruments and their relatively low long-term stability, the main roadblock to this development is the time-consuming and complex calibration procedures (Clementschitsch and Bayer, 2006; Vojinovic et al., 2006).

Multivariate calibrations (also called chemometrics) are considered as appropriate, and therefore recommended, for the on-line monitoring of complex multi-component mixtures, in which absorbance peaks overlap and can be significantly shifted through interactions between species. The state-of-the-art procedure for the development of a multivariate calibration model involves the preparation and measurement of a large set of standards, which must span the entire range of concentration for all compounds of interest. According to ASTM recommendations (ASTM, 1997), the number of standards should be higher than six times the number of relevant components, in order to insure a reproducible calibration set. Attention must be paid at this stage to avoid correlated combinations, by the use of appropriate multi-level experimental design (commonly four to five levels) (Brereton, 2000; Rhiel et al., 2002a). Building the mathematical model itself involves the use of one of the advanced mathematical tools gathered under the denomination Principal Components Regression (PCR), such as Partial Least Squares (PLS), Principal Component Analysis (PCA), Evolving Factor Analysis (EFA), etc. (Martens and Naes, 1992). The whole procedure commonly requires two days of laboratory work and has to be repeated if the medium composition, and/or optics alignment, is changed.

Although this approach has been successfully applied to various kinds of cultures from laboratory to pilot scale (Crowley et al., 2000; Doak and Phillips, 1999; Kornmann et al., 2003; Pollard et al., 2001; Rhiel et al., 2002b; Riley et al., 1998), it is particularly unsuitable for process development, for which a large number of new media and conditions (pH, temperature, and so on) are generally tested. Ironically enough, that is exactly where FTIR is expected to play an important role. While on-line monitoring of pH, dissolved oxygen and cell concentration can now be implemented relatively easily in screening platforms (Betts and Baganz, 2006; Harms et al., 2005), the on-line and in-situ monitoring of metabolite concentrations remains rather impracticable. Mid-infrared spectroscopy would be therefore highly complementary with the standard bioprocess measurements.

In order to make mid-infrared spectroscopy a powerful and flexible tool for bioprocess development, simplified calibrations are therefore absolutely necessary. Such calibrations would hold for various media and conditions, should not require a sound knowledge of multivariate calibration and could be performed within a few hours; the long-term objective being to integrate FTIR hardware and calibration routines into a completely automated environment. The current thesis aims at contributing to this research effort, by developing such simple calibration approaches.

Two assumptions for a spectra library

The starting point for simplifying the calibration procedure was a question. Why should we actually use a multi-level design of the calibration? Given that the time spent in the laboratory for the preparation and measurement of standards increases exponentially with the number of levels, what is the good reason for going to 5- or even 7-level designs? Is this simply to comply with the ASTM recommendation, which states that the number of standards must be at least six times the number of compounds of interest (ASTM, 1997)?

It seems important, at this stage, to distinguish between the need of a multi-level design and the need of a large number of standards for statistics purposes. Multi-level designs lead ineluctably to very large calibration sets, since a full design for n levels and k variables involves n^k experiments. Taking for instance a case with 4 compounds of interest (e.g. glucose, ethanol, ammonium and acetate) and 5 levels, $5^4 = 625$ standards would have to be prepared. Methods were developed to reduce the size of such calibration sets, while avoiding the creation of large correlations among variables, but 30 to 50 seems the limit for such a case (Brereton, 2000). A large calibration set, therefore, has to be considered as a negative but unavoidable consequence of a multi-level design. It is true that it allows an amplification of

signal to noise ratio, as well as a statistical analysis on reproducibility and variance, but it is not its principal aim. Besides, a 2-level design can be repeated a certain number of times to address these statistics issues.

Multi-level designs are used when complex interactions between variables are likely to occur, and/or when non-linear responses are expected. If all variables show linear responses and do not interact with each other, a 2-level design should be obviously preferred because it provides the same quantity of information, but at a much lower labour cost. Based on this observation, these two assumptions were investigated and validated, and the multi-level design was substituted by a spectra library, which is nothing than a 2-level calibration set.

Presented that way, the proposed approach seems quite trivial, and it would be fair to question why such a simplification has never been done before, or why conventional calibrations are that complex if the real situation is so simple. This may be explained from an historical perspective. The first applications in this field were performed using near-infrared spectroscopy, which exhibits far more peak overlapping than mid-IR, and therefore makes qualitative analysis and compound identification more difficult (Cooper et al., 1997). In addition, as opposed to mid-IR spectroscopy, process parameters as for instance aeration, stirring rate or broth rheological properties, have an influence on near-IR spectra, and therefore, interactions are very common. Obviously, these variables can not be studied independently of the process, since there is no standard available for such parameters. It means that chemometrics tools such as PCR or PLS have to be used for modeling, and numerous process standards must be used as calibration set. Regarding these aspects, Mid-IR spectroscopy is much simpler, especially when it comes to bioprocesses for which synthetic standards can be produced, as it is the case for cultures carried out in chemically defined medium.

The signal drift problem

Single-beam infrared instruments show signal instabilities that considerably complicate the calibration task. This issue is particularly critical for bioprocess applications that can last for days and deal with very low concentration, typically below 3 %.

The measurement of 50 standards takes about two days, and for this reason, the calibration set contains some information about the signal drift that occurred during this period of time. Even though wavenumber-dependent, this latter has some characteristic patterns that may be predicted (Doak and Phillips, 1999). Therefore, when subsequently building a PLS model, additional factors are included in the model to account for the drift,

which gives a certain robustness to the prediction with respect to signal instabilities that occur during the process.

Using traditional least squares and a small calibration set, such as a spectra library, does not allow such a correction, and therefore, signal drift has to be addressed separately. The importance of this topic was highlighted in many chapters of this work, and two methods were proposed for signal drift correction. One is based on signal anchorage, and the second involves the use of so-called “drift spectra” in the spectra library.

About the novelty of this work

The algorithms and calculation methods presented in this work are not new; this thesis does not pretend to present novel mathematical tools for more robust or more comprehensive process modeling. On the contrary, this thesis aims at simplifying as much as possible the calibration of mid-infrared spectrometers for their use in bioprocess monitoring. For this reason, well-known methods such as traditional least squares have been used in order to facilitate data treatment.

The novelty of this work rather lays in the proposition of a new perception of mid-infrared calibration. While it is commonly considered that numerous standards are always necessary, this thesis aims at showing that in some cases, small calibration sets can achieve the same results. It also has for objective to show that addressing signal drift and concentration modeling separately can lead to significant time savings.

2/ OUTLINE OF THE THESIS

The thesis is structured in eight main chapters, and two additional chapters that were provided as “supplementary material”, which contain relevant information that deserved to be included in this manuscript.

The second chapter is complementary to the present introduction. It tries to discuss with as little mathematical formalism as possible how Factor Analysis works, and in which cases traditional least-squares is as efficient as more sophisticated tools such as Partial Least Squares.

The third chapter plays a central role in this thesis. It relates the preliminary experiments that gave the evidence for the use of a spectra library as calibration set, and it addresses the accuracy and robustness of this approach.

Chapter 4 extends the calibration method developed in chapter 3, and shows the potential of this approach for high-throughput applications.

The fifth chapter discusses the influence of the pH on infrared spectra, and shows how the proposed calibration approach can be adapted to this new parameter. In a similar manner to chapter 5, chapter 6 investigates the effect of temperature on infrared calibrations.

Chapter 7 is slightly apart from the previous chapter, since it reports an on-line process control application, based on a very straightforward calibration approach.

The eighth, and last chapter gives conclusions to this work.

Supplementary material consists of a chapter that shows how critical infrared signal drift is for bioprocess application, as well as a chapter with process optimization results that were obtained while developing the method presented in chapter 7.

3/ REFERENCES

- ASTM. 1997. Standard practices for infrared, multivariate, quantitative analysis (E 1655-97). Annual book of ASTM standards. Philadelphia: American Society for Testing and Materials. p 844-869.
- Betts JJ, Baganz F. 2006. Miniature bioreactors: current practices and future opportunities. *Microb. Cell. Fact.* 5:21.
- Brereton RG. 2000. Introduction to multivariate calibration in analytical chemistry. *The Analyst* 125:2125-2154.
- Clementschitsch F, Bayer K. 2006. Improvement of bioprocess monitoring: development of novel concepts. *Microb. Cell. Fact.* 5:19.
- Cooper BC, Wise KL, Welch JT, Sumner MB, Wilt KW, Bledsoe RR. 1997. Comparison of near-IR, raman, and mid-IR spectroscopies for the determination of BTEX in petroleum fuels. *Appl. Spectrosc.* 51(11):1613-1620.
- Crowley J, McCarthy B, Nunn NS, Harvey LM, McNeil B. 2000. Monitoring a recombinant *Pichia pastoris* fed batch process using Fourier transform mid-infrared spectroscopy (FT-MIRS). *Biotechnol. Lett.* 22:1907-1912.
- Doak DL, Phillips JA. 1999. In situ monitoring of an *Escherichia coli* fermentation using a diamond composition ATR probe and mid-infrared spectroscopy. *Biotechnol. Prog.* 15:529-539.
- Harms P, Kostov Y, French JA, Soliman M, Anjanappa M, Ram A, Rao G. 2005. Design and performance of a 24-station high-throughput microbioreactor. *Biotechnol. Bioeng.* 93(1):6-13.
- Konstantinov KB, Tsai Y, Moles D, Matanguihan R. 1996. Control of Long-Term Perfusion Chinese Hamster Ovary Cell Culture by Glucose Auxostat. *Biotechnol. Prog.* 12(1):100-109.
- Kornmann H, Rhie M, Cannizzaro C, Marison I, von stockar U. 2003. Methodology for real-time, multi-analyte monitoring of fermentations using an in-situ mid-infrared sensor. *Biotechnol. Bioeng.* 82(6):702-709.
- Martens H, Naes T. 1992. Multivariate calibration. New York: John Wiley & Sons.
- Pollard DJ, Buccino R, Connors NC, Kirschner TF, Olewinski RC, Saini K, Salmon PM. 2001. Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy. *Bioprocess. Biosyst. Eng.* 24:13-24.

- Rhiel MH, Amrhein MI, Marison IW, von Stockar U. 2002a. The influence of correlated calibration samples on the prediction performance of multivariate models based on mid-Infrared spectra of animal cell cultures. *Anal. Chem.* 74(20):5227-5236.
- Rhiel MH, Ducommun P, Bolzonella I, Marison IW, von Stockar U. 2002b. Real-time in situ monitoring of freely suspended and immobilized cell cultures based on mid-infrared spectroscopic measurements. *Biotechnol. Bioeng.* 77(2):174-185.
- Riley MR, Arnold MA, Murhammer DW, Walls EL, DeLaCruz N. 1998. Adaptive Calibration Scheme for Quantification of Nutrients and Byproducts in Insect Cell Bioreactors by Near-Infrared Spectroscopy. *Biotechnol. Prog.* 14(3):527-533.
- Roychoudhury P, Harvey LM, McNeil B. 2006. The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. *Analytica Chimica Acta* 571:159-166.
- Vojinovic V, Cabral JMS, Fonseca LP. 2006. Real-time bioprocess monitoring. Part I: In situ sensors. *Sensors and Actuators B* 114:1083-1091.

Chapter 2

The Use of Virtual Calibrations to Facilitate Understanding of Factor Analysis

1/ INTRODUCTION

Factor Analysis (FA), which includes Principal Component Analysis (PCA) and Partial Least Squares (PLS), is more and more employed in academia and industry for various purposes such as spectrometer calibration, process modeling, data mining, quality control, etc. While software offering friendly interfaces have contributed to make this approach extremely popular, FA remains far from being straightforward, and examples of inappropriate use are not rare.

Why do we have more factors than compounds? Can Partial Least Squares deal with non-linear responses? It is not that easy to find easy to understand answers to such questions in chemometrics textbooks, which frequently give explanations through large doses of mathematics. This chapter therefore aims at providing non-experts in this field a practical understanding of Factor Analysis using simple examples and with as few equations as possible.

Nine different “virtual calibrations” were used to study how Factor Analysis can deal with signal drift, random noise, interactions between compounds and non-linear responses. The calibration models were developed for mixtures of three hypothetical compounds characterized by artificially assumed IR spectra. The absorption spectrum of each hypothetical calibration standard was calculated for a completely ideal case and for further eight cases assuming the spectra are affected with one or several of the problems mentioned above. All data were, therefore, simulated. The idea was to study the quality of the calibration model in each case by calculating standard errors of calibration, percentage of explained variance, and similar criteria. In order to mimic the traditional calibration approach, which was described and discussed in the previous chapter, a typical 50-standard calibration set was used for all cases.

In order to facilitate the reading of this study, all the matrices used, with their size and a brief explanation, were listed in a table of symbols that can be found at the end of the chapter.

The main reference used to write this chapter is the excellent, recent textbook of Brereton (Brereton RG. 2007. Applied chemometrics for scientists. Chichester (UK): John Wiley & Sons. 379 p.).

This chapter was published as a tutorial in August 2007 on the website *www.chemometrics.se*

2/ METHODOLOGY

Case study description

The absorbance spectra of three imaginary compounds were arbitrarily produced using Gaussian-curve functions (Fig. 1). Peak heights, widths and locations were tuned to produce large overlaps, which are representative of real spectroscopy applications. One hundred imaginary wavelengths were used as the calculation range, which confers on the system a redundancy level of 97 (because $100 - 3 = 97$).

Each absorbance shown in Figure 1 was considered as a molar absorbance, i.e. the absorbance of a 1 mol L⁻¹ solution of a single compound in water. The three spectra were gathered in a matrix **M** (**M** stands for molar) of size 3 x 100.

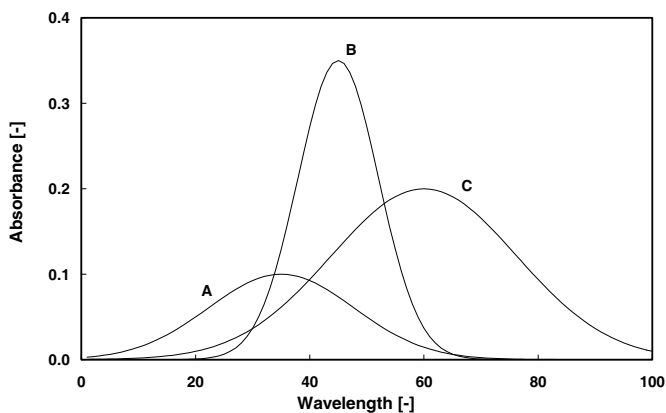


Figure 1. Molar absorbance of the three compounds A B and C. All spectra are Gaussian curves.

Design of calibration set

A full experimental design for three compounds and seven levels would involve $7^3 = 343$ standards. Although such a comprehensive calibration set implies an enormous amount of work, it guarantees the absence of correlations between the three variables. Decreasing the number of standards, while keeping the number of levels, generates ineluctably correlation among the variables, which decreases the model robustness. Mathematically, this translates into the matrix of correlation coefficients **K**. For a full experimental design, the matrix **K** is an identity matrix, whereas for sub-sets the correlations among variables lead to non-zero elements outside the diagonal.

A 7-level calibration set adapted from the template proposed by Munoz and Brereton¹ was used for all examples and is reported in appendix. Such a design is often used to calibrate spectroscopic instruments; it aims at producing a large number of standards, typically 50, that are as uncorrelated as possible and that span the entire experimental domain with several concentration levels for each compound. It can theoretically be used for as many as 20 species, since it features 20 columns that can be considered as orthogonal (the published 7-level calibration design is therefore a matrix of size 50 × 20). The three first columns were used to create the matrix of concentrations **C**. The concentration range was chosen arbitrarily to be from 0 to 6 mol L⁻¹ for the purpose of readability. The size of **C** was therefore 50 × 3. The correlations between the three compounds were all equal to 0.043, which can be considered as not significant (the matrix **K** is reported in appendix). Such a calibration design, even though realistic in terms of number of standards, is sufficiently close to ideality for the current study. It could therefore be assumed that correlations within the calibration set did not influence modeling results.

Singular Value Decomposition

Singular value decomposition (SVD) was used for pattern recognition. SVD is a generalized approach of eigenvalue decomposition that can deal with non-square matrices, and corresponding algorithms have been implemented in most mathematics software (Matlab, Mathcad, Maple, etc.). Singular value decomposition can be, for each calibration matrix of absorbance **E**, formulated as:

$$\mathbf{E} = \mathbf{U} \times \mathbf{S} \times \mathbf{V}^T \quad [1]$$

where **U** is 50 × 50 (i.e. the number of standards) and **V^T** is the transpose of **V** and is 100 × 100 (i.e. the number of wavelengths). **S** is diagonal and of the same size as **E**, 50 by 100, namely the number of standards by the number of wavelengths at which absorbance is measured. The diagonal of **S** contains the singular values λ , in decreasing order, whereas all other elements are equal to zero. The relative importance of the singular values is used to distinguish relevant from non-relevant information. A high singular value means that the corresponding factor explains a large part of the variance observed in the calibration set, while small singular values usually refers to noise, or negligible parameters. In order to facilitate the comparison between analyses, each singular value can be divided by the sum of

¹ Munoz J.A., Brereton R.G. 1998. Partial factorial design for multivariate calibration: extension to seven levels and comparison of strategy. *Chemometrics and Intelligent Laboratory Systems* 43:89-105.

all of them, which gives relative singular values λ_{rel} . These relative values are sometimes referred to as the fraction of variance explained by the factor.

The product of $\mathbf{U} \times \mathbf{S}$ is frequently designated as the scores matrix, \mathbf{T} . In the current work, the columns of \mathbf{T} display the contribution of each factor in the absorbance spectrum of the standards. Following this terminology, \mathbf{V} is traditionally called the loadings matrix, and contains the spectral features of each factor.

Principal Components Regression

Principal Components Regression (PCR) was used to assay quantitatively the quality of modeling. The method consists in extracting the principal components, or factors, using SVD, and in building a calibration model by regression on the relevant factors. In the current work, the term “factor” was preferred to “components”, in order to avoid confusion with chemical compounds. Whereas PCR extracts the factors from the matrix of experimental data only (i.e. in the spectra), Partial Least Squares (PLS) searches them in both the matrix of experimental data and the experimental design matrix (i.e. matrix of concentrations). In other words, PCR, as opposed to PLS, implicitly assumes that there is no error in the experimental design matrix, which would mean for real applications that the standards would have been prepared with an infinite precision. To account for unavoidable errors in standard concentration, PLS searches the principal factors in both the experimental design matrix and the experimental data. PCR was chosen for the current work, in order to keep the case study as simple as possible.

Once an “appropriate” number of factors f has been selected using SVD, the first f columns of the score matrix \mathbf{T} are extracted to produce a smaller matrix \mathbf{T}_b , which should contain only significant information. Similarly, the first f columns of \mathbf{V} are taken to form \mathbf{V}_b . While the size of \mathbf{T} was in the current work 50×100 , \mathbf{T}_b was $50 \times f$ and \mathbf{V}_b was $100 \times f$, with $f \leq 50$.

The next step consists in the calculation of the regression matrix \mathbf{R} , which can be considered as the core of the model. \mathbf{R} relates the concentration of the standards in \mathbf{C} to the main spectral features obtained using the loadings matrix \mathbf{V}_b :

$$\mathbf{C} = \mathbf{E} \times \mathbf{V}_b \times \mathbf{R} \quad [2]$$

where \mathbf{C} is 50×3 , \mathbf{E} is 50×100 , \mathbf{V}_b is $100 \times f$ and \mathbf{R} is $f \times 3$. It can clearly be seen here that the size of the regression matrix depends on the number of factors used. Using equation 1, equation 2 can be written in a simpler manner using the scores matrix:

$$\mathbf{C} = \mathbf{T}_b \times \mathbf{R} \quad [3]$$

\mathbf{R} is then calculated by least-squares regression:

$$\mathbf{R} = \mathbf{T}_b^+ \times \mathbf{C} = (\mathbf{T}_b^T \times \mathbf{T}_b)^{-1} \times \mathbf{T}_b^T \times \mathbf{C} \quad [4]$$

The superscript “+” refers to the pseudo-inverse of the matrix, which is explicitly developed in the second part of the equation, and is a generalized form of matrix inversion for non-square matrices. The prediction of unknown concentrations \mathbf{C}_{unk} from a set of, for example, 10 measured spectra \mathbf{E}_{unk} is obtained, similarly to Equation 2, by:

$$\mathbf{C}_{\text{unk}} = \mathbf{E}_{\text{unk}} \times \mathbf{V}_b \times \mathbf{R} \quad [5]$$

where \mathbf{C} is 10×3 , \mathbf{E} is 10×100 , \mathbf{V}_b is $100 \times f$ and \mathbf{R} is $f \times 3$.

Traditional least squares regression (LS), which is sometimes referred to as Multiple Linear Regression (MLR), was also used to allow for comparison with PCR modeling. In that case, the matrix \mathbf{M} (size 3×100) of the molar absorbance of A, B and C (Fig. 1) was used to calculate the unknown concentration by finding a solution for the equation:

$$\mathbf{E}_{\text{unk}} = \mathbf{C}_{\text{unk}} \times \mathbf{M} \quad [6]$$

which is explicitly given after rearrangement by:

$$\mathbf{C}_{\text{unk}} = \mathbf{E}_{\text{unk}} \times \mathbf{M}^+ \quad [7]$$

The accuracy of PCR and LS modeling was determined by a leave-one out cross-validation. An overall standard error of calibration (SEC), calculated for the three compounds, was used to evaluate the quality of cross-validation (Equation 8).

$$SEC = \sqrt{\frac{\sum_{k=1}^m \sum_{x=A,B,C} (y_{k,x} - \tilde{y}_{k,x})^2}{3 \cdot m}} \quad [8]$$

where $y_{k,x}$ is any concentration of a compound x and $\tilde{y}_{k,x}$ is the corresponding value predicted by the model. m refers to the total number of standards, which was 50 in the current work.

3/ VIRTUAL CALIBRATIONS

Linear, ideal case

For the first case studied, it was assumed that absorbance was linear with respect to concentration (i.e. the Lambert-Beer law was followed) and that there was no interaction between the three compounds. In addition, it was considered that no noise or drift interfered with the absorbance signal. The calibration matrix of this ideal case E_1 , of size 50×100 , was produced using the following equation:

$$E_1 = C \times M \quad [9]$$

where C and M are the matrices of concentrations (50×3) and molar absorbance (3×100) respectively. As expected, a decomposition of the E_1 into singular values gave three non-null singular values, which were 55.0, 10.3 and 4.7. The meaning of these values, as well as the significance of the corresponding loadings, is not obvious. A plot of the three loadings multiplied by their respective singular values shows that they significantly differ from the spectra of the three compounds A, B and C (Fig. 2). The absolute amplitude of the loadings spectra is not relevant, since it depends on the concentration range used for calibration.

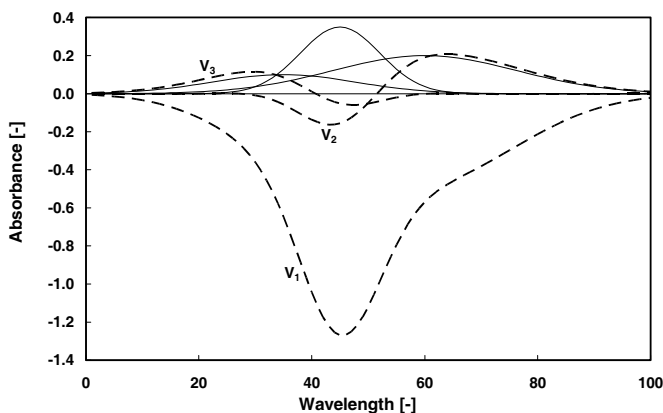


Figure 2. Simulated molar absorbance of the three compounds of interest (A, B and C) and loadings of the three non-null factors multiplied by their corresponding singular value (V_1 , V_2 and V_3). Spectra shapes are very different, even though no interaction, no noise and no drift were added to the linear model.

The relative values of the three singular values were 0.79, 0.15 and 0.07 (relative values of singular values are summarized in Table 1), which corresponds to the relative size of the

areas under the curve V_1 , V_2 and V_3 in Figure 2. To compare with, the relative size of the areas under the curve of the molar absorbance of the three compounds A, B, and C are much closer to each other, precisely 0.46, 0.35 and 0.18. This highlights how the SVD algorithm works: factor after factor, it tries to explain as much variance as possible, even if this implies the use of negative spectra. As expected, the standard error of calibration found by leave-one out cross-validation was equal to zero for the model involving three factors, meaning that the prediction within the calibration set was infinitely accurate. The SEC value of the least squares model was also equal to zero (all SEC values are summarized in Table 2).

Using PCR for such an ideal case is obviously not required. It would even make the situation more complex, by involving scores and loadings that have no physical significance. As a matter of fact, for this ideal situation, an accurate modeling can be performed with only the measurements at three different wavelengths, by writing a system of three equations that allows solving for the unknown concentrations of A, B and C

Table 1. Summary of the nine cases studied, with the corresponding explained variance (relative singular values) of the first five factors. Y and N stand for Yes and No respectively. Linear means that the Lambert-Beer law was followed. n refers to the level of noise added to the signal.

Case No	Linear	Interactions	Noise	Drift	Explained variance by factor [%]				
					1	2	3	4	5
1	Y	N	N	N	78.6	14.7	6.7	0.0	0.0
2	Y	N	N	Y	74.8	14.3	7.1	3.8	0.0
3	Y	N	$n = 1$	N	11.0	3.2	3.0	3.0	3.0
4	Y	N	$n = 1/3$	N	25.4	5.0	2.9	2.5	2.4
5	Y	N	$n = 1/10$	N	48.3	9.1	4.2	1.4	1.4
6	Y	Y*	N	N	77.2	14.4	7.3	1.0	0.0
7	Y	Y**	N	N	77.2	14.4	7.3	1.0	0.0
8	Y	Y***	N	N	72.6	14.0	8.5	2.5	0.0
9	N	N	N	N	80.1	13.2	6.7	0.0	0.0

*) A and B form a complex that absorb; the molar absorbance of A and B are unaffected by this interaction

**) B binds to A, which results in peak shift of the molar absorbance peak of A proportional to the concentration of B

***) B binds to A, which results in peak shift of the molar absorbance peak of A proportional to the square root of the concentration of B

Signal drift

The second example is aimed at discussing the ability of Factor Analysis to deal with signal drift, which is typically wavelength-dependent. Signal drift was assumed to have a characteristic, constant shape (Fig. 3), which was given by a polynomial function with additional sinus and exponential terms.

A random number β between -1 and 1 was used to vary the amplitude of the drift Z from one standard to another. The calibration matrix E_2 was therefore obtained using:

$$E_2(i,:) = E_1(i,:) + \beta \cdot Z \quad [10]$$

The notation $(i,:)$ is used to refer to the i -th row of a matrix, which corresponds to the absorbance spectrum of the i -th standard.

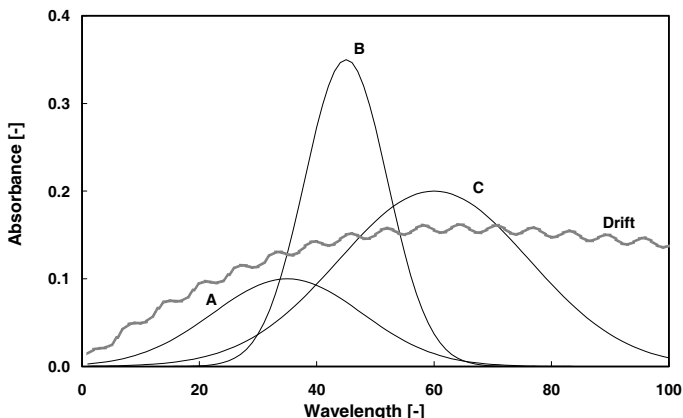


Figure 3. Simulated molar absorbance of the three compounds of interest (A, B and C) and simulated drift (bold line). The drift spectrum is a polynomial function with additional sinus and exponential terms.

SVD decomposition of the calibration matrix gave four non-null factors, which were on a relative scale equal to 0.75, 0.14, 0.07 and 0.04. As compared to the previous case, one extra factor was found by the algorithm in order to explain the variations induced by the drift. Standard error of calibration was equal to zero for the model involving these 4 factors (SEC was equal to 1.45 with 3 factors), whereas the least squares model led to a SEC value of 1.35. This shows that factor analysis is much more reliable than traditional least squares when the signal is subjected to a drift, which, even though wavelength-dependent, presents a predictable pattern.

Influence of random noise

This third example is aimed at studying how factor analysis can handle random noise, as opposed to a drift, which in the previous case was assumed to have a predictable behaviour. For that purpose, white gaussian noise was added to the calibration matrix of the ideal case E_1 to produce the matrices E_3 , E_4 and E_5 :

$$E_{3-5}(i,:) = E_1(i,:) + n \cdot N \quad [11]$$

where N refers to a vector (size 1×100) of white gaussian noise, chosen with an arbitrary amplitude and variance. A new N was generated for each row of the calibration matrix. The

term n is a noise level and was equal to 1, $1/3$ and $1/10$ for the matrices E_3 , E_4 and E_5 respectively. Among the three, the calibration matrix E_5 was thus the closest to the ideal, noiseless case.

SVD decomposition gave for the three matrices 50 non-null singular values (Fig. 4, A).

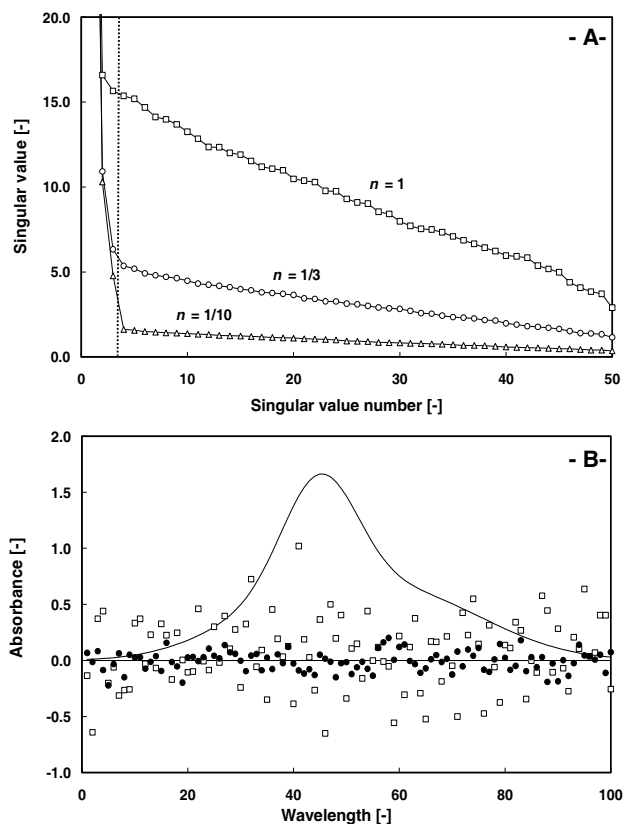


Figure 4. A) All 50 absolute singular values found in the calibration matrix for a linear case, without drift and interactions, but with three different noise levels n . A high n value corresponds to a noisy signal. The three first factors are on the left of the vertical dashed line. B) Absorbance of the first calibration standard and associated noise for $n = 1/10$ (black circles) and $n = 1/3$ (open squares).

For a low noise level (i.e. $n = 1/10$), it is easily possible to distinguish the three relevant factors that are associated to the compounds A, B and C from the noise-related factors: The sharp decrease of singular values stopped with the fourth value and completely flattened out

afterwards. While this elbow was still observable for $n = 1/3$, noise-related factors could not be identified anymore for the highest noise level ($n = 1$). Figure 4, frame B, provides an order of magnitude of the noise levels used for the current case, by plotting together a matrix \mathbf{N} and the absorbance spectrum of the first standard. It can be seen that $n = 1/3$ already corresponds to a very noisy situation, which can be generally avoided by a limitation of environmental interferences.

For reasonably low noise levels, the PCR approach performs in a similar way to traditional least squares (Fig. 5). Including a large number of factors in the PCR model reduces only partially the SEC value. For $n = 0.1$ to $n = 0.7$, it can be observed that a model with 10 factors is not significantly more accurate than a model using 3 parameters, these being either factors (for PCR) or molar absorbance spectra (for LS).

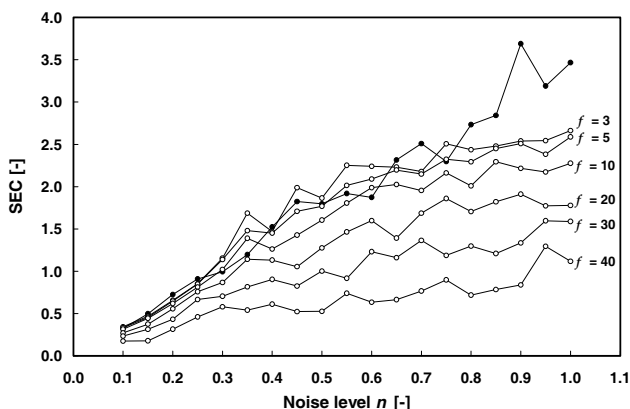


Figure 5. Standard error of calibration (SEC) as a function of the noise level n added to the signals. SEC values were calculated for the least squares method (black circles), and for several PCR models built with a number of factors f ranging from 1 to 40 (open circles). Including a large number of factors to the model does not significantly improve the modeling.

While Factor Analysis is extremely efficient to tackle drifts of predictable pattern, it seems not appropriate to deal with random noise in the signal. Even though certainly less robust than PCR, the traditional least-squares method is, for such a case, much simpler and leads to similar results. In addition, it has the advantage of avoiding virtual factors, which allows a better qualitative interpretation of data. Filtering methods, as for instance a simple moving average or a low-pass filter, could be implemented prior to LS modeling, in order to enhance accuracy of modeling.

Interactions between compounds

Different sorts of interaction between compounds A, B and C could be imagined. For example, in a simple case, A and B can, without losing their IR activity, form a complex that produces a new absorbance peak (Fig. 6). The amplitude of this peak would be proportional to the concentration of A and B. For such a case, the calibration matrix can be expressed by:

$$\mathbf{E}_6 = \mathbf{C}_6 \times \mathbf{M}_6 \quad [12]$$

where \mathbf{M}_6 is given by the molar spectra matrix \mathbf{M} , with an additional, fourth row that contains the interaction peak due to the complex A-B. The matrix \mathbf{C}_6 is created from the concentration matrix \mathbf{C} (i.e. the experimental design), by adding an additional, fourth column that contains the concentration of A multiplied by the concentration of B and an interaction factor, set to 0.05:

$$C_6(i,4) = 0.05 \cdot C_6(i,1) \cdot C_6(i,2) \quad [13]$$

Applying the SVD algorithm to the calibration matrix \mathbf{E}_6 gave four non-null factors, which is analogue to what was observed previously with the drift case (Table 1). Similarly, the standard error of calibration was equal to zero for the model that includes four factors. The least squares approach led, however, to a non-null SEC value, precisely 0.88, which is completely understandable, since the spectra library (i.e. \mathbf{M}) did not contain the interaction peak, and was unable to predict it.

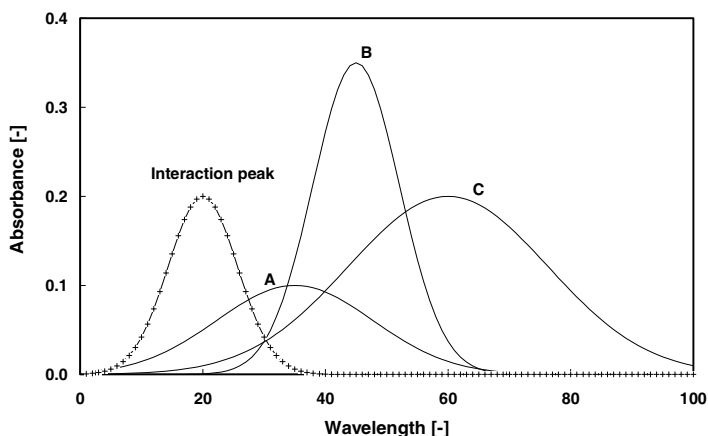


Figure 6. Molar absorbance of the three compounds A, B and C (plain lines) and interaction peak (line with crosses). A and B form a complex A-B that shows an absorbance peak denominated “interaction peak”. The complexation does not influence the molar absorbance of A and B themselves.

This shows that Factor Analysis is more suitable than traditional LS for the modeling of such simple, linear interactions. It must be emphasized, however, that such a result can be achieved with a much smaller set of standards. Seven standards, i.e. solutions of A, B, C, A with B, B with C, A with C and the three together leads to exactly the same SEC values, with a completely uncorrelated calibration design, since the interaction between A and B is linearly dependent on the concentrations of A and B.

It could be argued that it is not fair to compare a PCR model, which includes the interaction peak, and the least square method, which does not. However, this corresponds to a situation that may arise in reality. In the absence of preliminary experiments dedicated to the identification of the interaction peak, modeling would be performed using the molar absorbance of the three known compounds.

Spectral distortion is a more complicated type of interaction. One could imagine that the absorbance peak of A, under the influence of B, would be shifted to the left and significantly narrowed. To represent such a case, it was assumed that the molar absorbance of A was equal to the interaction peak shown in figure 6 when the concentration of the compound B was equal to 10 mol L⁻¹ or above. For any concentration of B between 0 and 10 mol L⁻¹, the molar absorbance of A is given by a linear combination of the interaction peak and the original molar absorbance of A. The calibration matrix E_7 was obtained using:

$$E_7(i,:) = C(i,:) \times M_{7,i} \quad [14]$$

where $M_{7,i}$ is a matrix that is similar to M , but contains in the first row an apparent molar concentration of A, which is a function of the concentration of B (i.e. $C(i,2)$) and is therefore recalculated for each standard.

The SVD of E_7 gave exactly the same results as for E_6 , which is fairly logical since the same four spectra were used to generate the calibration matrix. The leave-one out cross-validation, using the four non-null factors, led to a SEC value of zero, meaning that the interaction can be explained by the model. The least squares model, again, gave a large SEC value of 0.62, for the same reason as discussed previously for the case E_6 .

The ability of PCR to deal with interactions was further tested with a non-linear relation, based on the previous example. The molar absorbance of compound A was no longer related to the concentration of B (as for E_7), but was set proportional to the square root of this concentration. A coefficient was used so that the function fulfills the same boundary criteria, namely that at a concentration of B of 10 mol L⁻¹ the molar absorbance of A was equal to the interaction peak. As an example, at a concentration of B of 5 mol L⁻¹, the molar absorbance of A was given by a mix of 70.7% of the interaction peak and 29.3% of the original molar

absorbance of A, whereas the fractions would have been 50/50 and 50/50 for the previous case E6. Surprisingly enough, the PCR model was able to predict correctly the concentration of three compounds, since the SEC value was equal to zero when using the four non-null factors found by the SVD decomposition. This shows that Factor Analysis can effectively predict interactions between compounds, even non-linear ones, whereas traditional least-squares is completely inefficient.

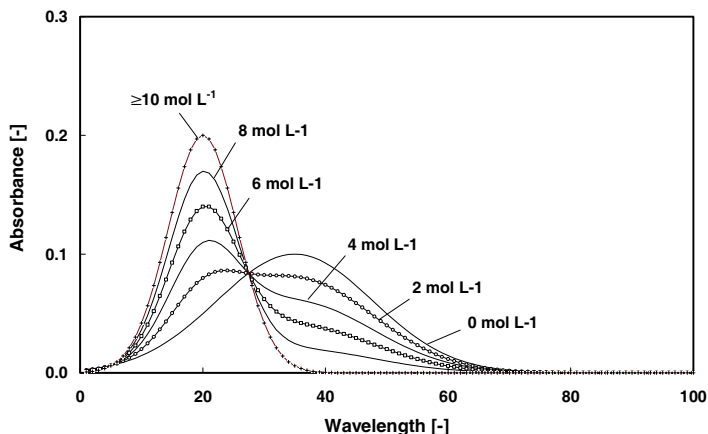


Figure 7. Apparent molar absorbance of A for different concentrations of the compound B. B has a tendency to bind to A, which results in distortions in the molar absorbance spectrum of A. This latter is narrowed and shifted to the left under the influence of B. For any concentration of B between 0 and 10 mol L⁻¹, the molar absorbance of A is given by a linear combination of the molar absorbance at concentrations of B of 0 and 10 mol L⁻¹.

Non-linear absorbance

The ninth example aimed at investigating whether Factor Analysis is a suitable approach for calibration when the absorbance is not linear with respect to concentration, namely when the Lambert-Beer law is not followed. For the three compounds A, B and C, a simple polynomial function was used to relate the absorbance a at a concentration y to the molar absorbance at the same wavenumber ν :

$$\frac{a_{\nu}(y)}{a_{\nu}(y=1)} = \frac{1}{\alpha_1 + \alpha_2 + \alpha_3} \cdot (\alpha_1 \cdot y + \alpha_2 \cdot y^{1.5} + \alpha_3 \cdot y^{0.5}) \quad [15]$$

where the function parameters α_1 , α_2 and α_3 were chosen arbitrarily to produce different absorbance profiles for the three compounds (Fig. 8). It must be emphasized here that, again, such a case is a very particular, and other non-linear responses could be imagined.

This function did not produce distortions in the spectra; it only multiplies the molar absorbance spectrum by a factor that was not linearly related to the concentration. Taking as an example all α_i parameters to be one, it can be found that the spectra of a 2 and a 10 M solutions are given by the molar absorbance spectrum multiplied by 2.1 and 14.9 respectively. This means that even though the increase is not linear, the shape of the spectrum is conserved. Such a response would be expected for compounds that strongly self-interact. The absorbance profile used for compound C could be, for instance, explained by a tendency of the molecules of C to form aggregates of lower IR absorbance.

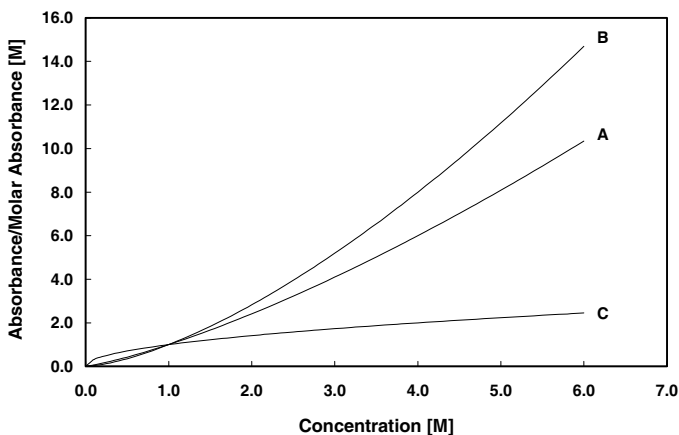


Figure 8. Ratio of absorbance as a function of concentration to the molar absorbance for the compounds A, B and C. The same polynomial function was used for all wavenumbers, in order to avoid the creation of distortion in the spectra.

These nonlinear responses for A, B and C translate into a biased experimental design C_0 , which was calculated from the calibration design used previously, C , by applying to each of the concentrations of C the nonlinear function (Equation 15). It has been chosen to calculate here apparent concentrations to express the non-linearity, but the same results could have been obtained by using C and calculating apparent molar absorbance spectra. The function parameters α_i are compound-dependent, therefore one set of parameters was used per column of the matrix C . The calibration matrix E_0 was then given by the product of C_0 and M , in a similar manner to Equation 9:

$$E_0 = C_0 \times M \quad [16]$$

A decomposition of E_9 gave only three non-null singular values, the relative values being equal to 0.80, 0.13 and 0.07. This result is very similar to what has been found for the ideal linear case (E_1), and although surprising at first glance, it turns out to be logical, since the spectra shapes were conserved by the nonlinear function used. The leave-one cross-validation, using the three factors in the PCR model, led to a SEC value of 0.93, which by itself indicates a poor modeling. A comparison of the predicted concentrations and the real concentrations highlights the low accuracy of calculated values (Fig. 9). The PCR model was unable to match the nonlinear response; it simply found a linear regression that minimized the sum of squared residues, which resulted in an underestimation of the low concentrations and an overestimation of the high concentrations. Even though Factor Analysis has some capabilities to deal with non-linear responses, it should be limited to linear cases. If the Lambert-Beer law is not respected, the use of nonlinear modeling, as for instance Neural Networks, should be considered.

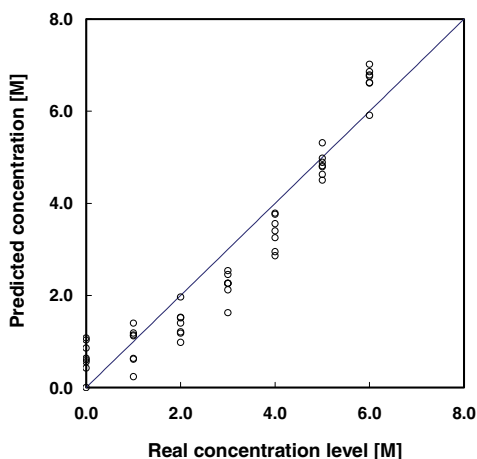


Figure 9. Predicted concentrations found by leave-one out cross-validation versus the real concentration for compound B. The PCR approach is clearly not able to deal with a non-linear response, as studied in the current case.

The least squares model was also not able to predict the deviation from Lambert-Beer's law. The SEC was even higher, with a value of 3.28, and the discrepancy at high concentration much larger. The reason for that is that the least squares algorithm finds solutions by extrapolating from the molar absorbances, which are on the edge of the experimental domain, whereas PCR finds a linear solution including all data.

The difference between the case E_8 and E_9 lies in the fact that for E_9 (Lambert-Beer law not respected), the nonlinearity did not induce any change in the spectral shapes, whereas for E_8 (interactions), the nonlinearity involved a new spectrum (i.e. the interaction peak). This new spectrum brought supplementary information that allowed modeling the interaction between the two compounds, which explains why in one case the SEC was equal to zero and in the other not.

Overall diagnostics

The standard errors of calibration calculated for the least squares and PCR models, with a number of factor ranging from 1 to 10, were summarized in Table 2. Results for the PCR models including one or two factors were also included in the table, even though they were not extensively discussed previously. Unsurprisingly, models with fewer factors than the number of known compounds led to large SEC values.

Table 2. Standard error of calibration obtained by leave-one out cross-validation for the nine examples studied. Calculations were performed for a number of factor f from 1 to 10, and also using traditional least-squares (LS). Summary of the cases description are given below this table but can also be found in Table 1.

f [-]	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	
1	2.90	2.96	2.97	2.91	2.89	2.89	2.92	2.92	2.98	
2	2.11	2.30	2.53	2.17	2.11	2.11	2.14	2.15	2.16	
3	0.00	1.45	2.38	1.20	0.37	0.35	0.40	0.59	0.93	
4	0.00	0.00	2.36	1.17	0.37	0.00	0.00	0.00	0.93	
5	0.00	0.00	2.28	1.16	0.36	0.00	0.00	0.00	0.93	
6	0.00	0.00	2.22	1.14	0.36	0.00	0.00	0.00	0.93	
7	0.00	0.00	2.20	1.13	0.36	0.00	0.00	0.00	0.93	
8	0.00	0.00	2.08	1.12	0.35	0.00	0.00	0.00	0.93	
9	0.00	0.00	2.08	1.10	0.35	0.00	0.00	0.00	0.93	
10	0.00	0.00	2.05	1.08	0.34	0.00	0.00	0.00	0.93	
LS	0.00	1.35	3.32	1.05	0.38	0.88	0.62	0.94	3.28	
E1:	Ideal case									
E2:	Drift case									
E3:	High level of noise			E4:	Medium level of noise			E5:	Low level of noise	
E6:	Complex formation			E7:	Linear peak shift			E8:	Nonlinear peak shift	
E9:	Non-linear relationship between absorbance and concentration									

For the simple, ideal E_1 case, both approaches led to a perfect modeling with three variables, these being molar absorbance spectra or factors. The drift case (E_2) showed that PCR, unlike LS, is able to detect and model a signal drift of predictable pattern, since the SEC value was zero for the model with four factors. Both approaches were found to be very similar regarding random noise ($E_3 - E_5$; in decreasing noise level order). In order to reduce the SEC value, a large number of factors have to be used in the PCR model, which has for consequence to undermine its predictive ability and robustness. Similarly to the drift case, cases E_6 to E_8 showed that factor analysis can deal with interactions. While traditional least

squares is limited, for modeling, to the use of the three molar absorbance spectra, PCR can integrate an additional factor that allows a perfect fitting of the interactions, even in case of peak distortion. The last example studied (E₉) showed that both algorithms have a poor ability to deal with a nonlinear relationship between absorbance and concentration, even though PCR, by finding the best linear fit within the calibration data, gave slightly lower SEC values.

The nine examples developed in this work have shown that, whereas Factor Analysis has better modeling abilities than traditional least-squares when compounds interact with each other or when the signal is subjected to a drift of predictable pattern. Both methods gave the same results when facing random noise, and they both turned out to be unable to model a case in which the absorbance was not linear with respect to concentration (Table 1 and 2).

4/ CONCLUSIONS

The nine examples studied using virtual calibrations have shown that Factor Analysis is superior to traditional least-squares when the spectrometer signal is subjected to a rather predictable drift, and when compounds interact with each other therefore inducing peak distortion. Both methods showed similar performances when facing random noise, and none of them was able to deal with the case in which absorbance was not linear with respect to concentration.

However, it must be said that PCR and PLS are in reality more robust than traditional least squares, and able to deal to some extent with non-linear response. This was not illustrated by the very simple, simulated cases presented here, but should be taken into account when having to choose a calibration approach.

In terms of real applications, these virtual calibrations show that preliminary experiments on the validity of the Lambert-Beer law and the independence of the species present should be conducted before developing a calibration model. If the absorbance is linear with respect to concentration, and if the compounds do not interact in solution (a condition which is likely to be satisfied, keeping in mind the degree of dilution of most culture media), traditional least squares should be preferred to Factor Analysis if keeping modeling simple is a priority. The presence of signal drift of predictable pattern would promote the use of Factor Analysis, but the problem can also be tackled separately by using specific signal drift correction methods such as anchoring of Savitzky-Golay filtering. Models based on Factor Analysis are also certainly more robust regarding unexpected perturbations, since they focus on points of major variance.

Checking for linearity and interactions not only has an impact on data treatment, i.e. on the approach chosen for regression, but it also largely influences the design of the calibration set itself. Apart from noise and drift considerations, a multi-level design does not bring more information than a 2-level design if the Lambert-Beer law is strictly followed and if all compounds are independent. Since the number of levels influences in an exponential manner the size of the calibration set, preliminary experiments are therefore absolutely worthwhile, because they can lead to a significant reduction of the calibration set and a simplification of the modeling approach.

5/ LIST OF SYMBOLS

For clarity reasons, the matrix sizes are given for the current case, which involved 3 compounds (A, B and C), 50 standards for calibration and measurement of spectra at 100 wavelengths. The number of factors used in modeling depended on the example, and is referred to here as f .

Matrices

C	50×3	Matrix of concentrations (calibration design: the concentration level of each compound in each standard)
E	50×100	Calibration matrix (the spectra simulated for each standard, or in other words the experimental data one would obtain after having measured the 50 standards)
K	3×3	Correlation coefficients matrix
M	3×100	Molar spectra matrix (molar absorbance of A, B and C)
N	1×100	Additive white Gaussian noise
R	$f \times 3$	Regression matrix (matrix determined by regression during the calibration and used for concentration prediction)
S	50×100	Singular values matrix (diagonal elements are singular values in decreasing order; others are equal to zero)
T	50×100	Scores matrix (is equal to $\mathbf{U} \times \mathbf{S}$, contains the weighed contribution of each factor to each standard)
T_b	$50 \times f$	Scores matrix used for regression (size depends on the number of factors retained for modeling)
U	50×50	No special name
V	100×100	Loadings matrix (contain spectral features of the factors)
V_b	$100 \times f$	Loadings matrix used for regression (size depends on the number of factors retained for modeling)
V_j	100×1	j-th loading (i.e. j-th column of matrix V)
Z	1×100	Drift vector (spectrum of the drift)

Other symbols

a	Absorbance
f	Number of factors chosen
i	Column of matrix row
j	Position of matrix column
(i,j)	(i,j) -th element of a matrix (row i , column j)
$(i,:)$	i -th row of a matrix
$(:,j)$	j -th column of a matrix
m	Total number standards (e.g. 50 for the current work)
n	Coefficient for noise level
y	Standard concentration

Greek symbols

α	Model Parameter
β	Random number between -1 and 1
λ	Singular value
ν	Wavenumber [cm^{-1}]

Superscripts and subscripts

T	superscript	Transpose of a matrix
-1	superscript	Inverse of a matrix
$+$	superscript	Pseudo-inverse of a matrix
k	subscript	Standard number ($k \leq m$)
x	subscript	Compound A, B or C
unk	subscript	Unknown concentration
rel	subscript	Relative to the sum of all singular values

6/ APPENDIX

A-1

Design of the 50-standard calibration set (matrix **D**), for the three compounds A, B, and C. Adapted from Brereton². The numbers in the matrix, form 0 to 6, correspond to the 7 concentration level.

Standard No	A	B	C
1	3	3	3
2	3	0	1
3	0	1	2
4	1	2	2
5	2	2	5
6	2	5	4
7	5	4	5
8	4	5	3
9	5	3	4
10	3	4	0
11	4	0	5
12	0	5	5
13	5	5	6
14	5	6	2
15	6	2	6
16	2	6	3
17	6	3	2
18	3	2	4
19	2	4	6
20	4	6	6
21	6	6	1
22	6	1	5
23	1	5	1
24	5	1	3
25	1	3	5
26	3	5	2
27	5	2	1
28	2	1	1
29	1	1	0
30	1	0	6
31	0	6	0
32	6	0	3
33	0	3	6
34	3	6	5
35	6	5	0
36	5	0	0
37	0	0	4
38	0	4	1
39	4	1	4
40	1	4	3
41	4	3	1
42	3	1	6
43	1	6	4
44	6	4	4
45	4	4	2
46	4	2	0
47	2	0	2
48	0	2	3
49	2	3	0
50	0	0	0

² Brereton RG. 1997. Multilevel multifactor designs for multivariate calibration. The Analyst 122:1521-1529.

A-2

K, the correlation coefficients matrix of **D**:

	A	B	C
A	1.0000	0.0431	0.0431
B	0.0431	1.0000	0.0431
C	0.0431	0.0431	1.0000

Chapter 3

Simplified FTIR Calibration Based on a Spectra Library for the On-line Monitoring of Bioprocesses

1/ ABSTRACT

In order to significantly reduce the time involved in mid-infrared spectroscopy calibrations, a novel approach based on a library of pure component spectra was developed and tested with an aerobic *S.cerevisiae* fermentation. Instead of the 30-50 standards that would have been required for to build a chemometric model, only 5 solutions were used to assemble the library, namely one for each compound (glucose, ethanol, glycerol, ammonium and acetate). Concentration profiles of glucose, ethanol and ammonium were monitored with a fair accuracy, leading to standard error of prediction (SEP) values of 0.86, 0.98 and 0.15 g L⁻¹. Prediction of the two minor metabolites, acetate and glycerol, was less accurate and presented a detection limit of around 0.5 g L⁻¹. The overall performance of the library-based method proved to be very similar to a 49-standard chemometrics model. The model was shown to be very robust and uncorrelated, since it was able to predict accurately concentration changes during a spiking experiment. Even though simple, this method allows more advanced calculations, such as determination of the explained variance and detection of unexpected compounds using residuals analysis.

This chapter was published in *Analytica Chimica Acta*¹:

Schenk J, Marison IW, von Stockar U. 2007. Simplified Fourier-transform mid-infrared calibration based on a spectra library for the on-line monitoring of bioprocesses. *Anal. Chim. Acta* 591(1):132-140

¹ This version slightly differs from the published article on one detail: Wavenumber and time dimensions of the process data matrix have been inverted in order to comply with traditional chemometrics rules. Equations 5 to 9 were consequently adapted.

2/ INTRODUCTION

Fourier-transform mid-infrared spectroscopy (FTIR) has been used in bioprocess monitoring for more than a decade and despite very appealing characteristics – sterilization in place, non-invasiveness, fast detection of many compounds simultaneously – it has not yet become a widely-used instrument in the bioprocess industries. This observation can be largely explained by the related calibration procedures that are labor-intensive and require a high level of expertise.

Best practice calibration procedures for infrared spectroscopy involve the preparation and measurement of more than six times the number of standards than the total number of compounds to be measured (ASTM, 1997). Care must be taken to avoid correlations within the calibration set that would considerably decrease the model robustness (Rhiel et al., 2002a), therefore specific designs are recommended to produce standards that are independent and uniformly distributed over experimental space (Brereton, 1997). A wide range of mathematical tools gathered under the “multivariate analysis” denomination are available to build a model, the most popular being Principal Component Analysis (PCA) and Partial Least Squares (PLS) (Martens and Naes, 1992). The efficiency of the chemometrics approach has been demonstrated in various applications, such as *E. coli*, fungi and yeast fermentations at various scales (Crowley et al., 2000; Kornmann et al., 2003; Pollard et al., 2001), monitoring of mammalian and insect cell cultures (Rhiel et al., 2002b; Riley et al., 1998), and even simultaneous measurement in culture medium of up to 19 components, including amino acids (Riley et al., 2001). However, the overall calibration procedure takes a minimum of 2 days of intensive work and requires a certain level of expertise to deal with the advanced mathematical tools. Moreover, the calibration has to be repeated for every culture medium and operating condition employed, which may lead to an important increase of work load in the development of a bioprocess.

Due to the strong signal peak overlapping and culture medium complexity, it is commonly considered that a multivariate analysis approach is necessary for bioprocess applications, and indeed FTIR is invariably presented under this perspective (Harms et al., 2002; Kornmann et al., 2003; Vojinovic et al., 2006). Historically, the first applications in this field were performed using near-infrared spectroscopy, which exhibits far more peak overlapping and therefore makes qualitative analysis and compound identification more difficult (Cooper et al., 1997). As a consequence, only a few attempts to simplify the calibration task have been published, and the focus was mainly on decreasing the number of

standards while keeping advanced data analysis such as PLS (Fairbrother et al., 1991; Navarro-Villoslada et al., 1995). A more significant simplification has been reported, with the accurate monitoring of sugars and ethanol in plant cell culture, using a linear combination of the four components at four wavenumbers (Hashimoto et al., 2005). However, all measurements were performed off-line with withdrawn samples, which allows for corrections of long-term instrument instability. This suggests that extended calibration models may actually account more for instrument instability than for mixture complexity itself.

In this paper, we present a simplified FTIR calibration technique, based on a library of pure component spectra, which assumes additivity of absorbances from the various species, as well as a linear relationship between concentration and absorbance. Such hypotheses are not trivial concerning infrared spectroscopy, and they were therefore validated with preliminary experiments. The reliability of this approach was assessed and validated with an aerobic fermentation of the yeast *S. cerevisiae* and was shown to be at least as accurate as commonly employed chemometrics-based calibrations. *S. cerevisiae*, the baker's yeast, was used as a case study for its well-known, though not simple, metabolism. More details on this widely used organism can be found elsewhere (Sonnleitner and Käppeli, 1986; von Schalien et al., 1995). The library consisted of molar absorbance spectra for each of the important compounds to be monitored during the process, namely glucose, ethanol, glycerol, acetate and ammonium. As a result, the time dedicated to calibration could be considerably reduced, since only five spectra had to be measured. This novel approach allows testing a broad choice of strains and defined media without the need for further work dedicated to calibration, and therefore paves the way for implementation of FTIR into high-throughput screening platforms.

3/ MATERIALS AND METHODS

Bioreactor Fermentation

The *Saccharomyces cerevisiae* wild-type strain CBS 8066 (CBS, Utrecht, The Netherlands) was used for this experiment. The inoculum was prepared by growing the cells in a shake flask for 24 hours in 100 mL of YPG medium (10 g L⁻¹ yeast extract, 10 g L⁻¹ peptone, 20 g L⁻¹ glucose). Cells were then centrifuged (10 minutes at 1500 g) and re-suspended in 10 mL of bioreactor medium for fermentor inoculation. Bioreactor medium was defined and contained per litre: 20 g glucose, 5 g (NH₄)₂SO₄, 9 g KH₂PO₄, 0.5 g MgSO₄·7H₂O; pH was adjusted to a value of 5 by addition of NaOH 5N. After filter-sterilization (0.2 µm, using Steritop®, Millipore, MA), 0.4 mL of antifoam (Struktol SB2121, Schill and Seilacher, Hamburg, Germany), and 10 mL of sterile vitamins and trace elements solution were added (composition per litre: 1 g CaCl₂·2H₂O, 267 mg H₃BO₃, 80 mg CuSO₄·5H₂O, 27 mg KI, 267 mg MnCl₂, 107 mg Na₂MoO₄·2H₂O, 1.2 g ZnSO₄·7H₂O, 4 g EDTA, 80 mg CoCl₂, 800 mg FeSO₄·7H₂O, 267 mg Ca-penthotenate, 13 mg biotin, 6.67 g m-inositol, 2.67 mg nicotinic acid, 53 mg para-amino benzoic acid, 267 mg pyridoxine.HCl, 267 mg thiamine.HCl). Bioprocess was carried out at 30°C in a 3.6-L bioreactor (Bioengineering, Wald, Switzerland) with a working volume of 2 litres, with pH controlled at a value of 5 by the automated addition of KOH 2 M. Agitation was set at 1300 rpm, using three 6-blade Rushton impellers, and total aeration was maintained constant at 3 NL min⁻¹, which was sufficient to maintain the dissolved oxygen above 30% of air saturation. Oxygen and carbon dioxide in the exhaust gas was measured using a paramagnetic and infra-red gas analyzers respectively (Dr. Marino Müller Systems, Esslingen, Switzerland).

FTIR Set-up and spectra acquisition

A Mettler Toledo FTIR 1000 (Mettler-Toledo, Greifensee, Switzerland) with an Attenuated Total Reflectance probe (ATR, DiComp) was mounted within the bioreactor thereby, allowing cleaning and sterilization in place. More details on this set-up can be found in previous publications (Doak and Phillips, 1999; Rhiel et al., 2002b). Sixty-four single-beam spectra with a resolution of 4 cm⁻¹ were collected and averaged for each measurement. Single-component spectra for the library were acquired within the reactor, in order to keep the FTIR alignment constant and to control the temperature at 30°C. For the five compounds of the library, the same procedure was followed: 2 litres of thermostated water supplemented with 1 mL of antifoam Struktol SB2121 served as intensity background, and the absorbance was

determined after addition of 250 mL of concentrated solution of known molarity (within the range 0.5 – 1 M). Molar absorbance spectra were obtained by dividing the measured absorbance spectra by the final concentration. During the bioprocesses, spectra were acquired every 2 minutes. Data treatment was performed using Matlab software (Mathworks, Natick, MA). Due to a technical breakdown of the FTIR 1000, a newer version of the instrument (FTIR 4000, also from Mettler-Toledo) was used for the spiking experiment, which implies that a second library was built to account for alignment changes.

Off-line analysis

Samples (5 mL) were collected at intervals (10 to 30 min depending on the batch phase) and immediately cooled to 2°C. Sample treatment was carried out within 2 hours. Dry cell weight was determined by gravimetric analysis: 3 ml of sample were filtered on 0.2 µm pre-weighed filters, washed and dried to constant weight. The remaining volume of the sample was filtered through 0.2 µm filters for off-line analysis of metabolites. Glucose, ethanol, glycerol, ammonium ion and acetic acid concentrations were determined by enzymatic assay, using commercially available methods (R-Biopharm AG, Darmstadt, Germany) and an automated analyzer (Cobas Mira, Roche, Basel, Switzerland).

4/ RESULTS AND DISCUSSION

Equations

Assuming that 1) the absorbance of every species is linear with respect to concentration and that 2) all absorbances are additive, the measured absorbance at a wavenumber ν can be expressed as a linear combination of the contribution of the m different species (Equation 1).

$$A_{tot,\nu} = \sum_{i=1}^{i=m} A_{i,\nu} + \varepsilon_{\nu} \quad [1]$$

where the subscript i indicates any compound. The term ε is added to account for experimental error. Using the Lambert-Beer law for each species absorbance, the following expression is obtained:

$$A_{tot,\nu} = \sum_{i=1}^{i=m} \kappa_{i,\nu} \cdot C_i + \varepsilon_{\nu} \quad [2]$$

where κ is the molar absorbance (or extinction coefficient) and C the molarity. The majority of the components of culture media such as vitamins, trace elements, buffer and salts show a relatively constant concentration during the time-course of a cultivation. They can therefore be gathered into a baseline absorbance in order to decrease significantly the number of species that must be integrated into the library to a value j , with $j < m$:

$$A_{tot,\nu} = A_{baseline,\nu} + \sum_{i=1}^{i=j} \kappa_{i,\nu} \cdot C_i + \varepsilon_{\nu} \quad [3]$$

To add the time dimension into the equations, the subscripts t and o are used to indicate an absorbance measured during the process and at the start respectively. The baseline term can then be eliminated:

$$A_{tot,\nu,t} - A_{tot,\nu,o} = \Delta A_{\nu,t} = \sum_{i=1}^{i=j} \kappa_{i,\nu} \cdot \Delta C_i + \Delta \varepsilon_{\nu} \quad [4]$$

where the delta symbol indicates the difference between the initial state (i.e. before inoculation) and any time t during the process. It must be emphasized that working with concentration differences instead of absolute concentrations considerably simplifies the calibration task. The absorbance contribution of all compounds that show small concentration changes during the culture (e.g. phosphates, trace elements, sodium, vitamins, etc.) is included into the baseline and, therefore, eliminated by the subtraction. This presents the advantage that a wide range of medium compositions can be studied with the same calibration, as long as no new carbon source or other main metabolite is used or appears. Even though a defined medium was used in this study, this method should therefore also

withstand limited amounts of chemically undefined components such as yeast extract or peptone, as far as they are not consumed massively during the cultivation. Equation 4 can be rewritten in matrix notation:

$$\Delta \mathbf{A}_{s,wn} = \Delta \mathbf{C}_{s,j} \times \mathbf{K}_{j,wn} + \mathbf{E}_{s,wn} \quad [5]$$

The subscripts refer to the matrix dimensions (row,column), j , wn and s being the number of compounds monitored, of wavenumber and the number of spectra taken during the process respectively. \mathbf{K} is the spectra library, or in other words, the molar absorbance measured in water for the compounds that are monitored. This system of linear equations is clearly overdetermined, since the absorbance is measured at more than hundred wavenumbers and the number of compounds in most cases does not exceed 10 (precisely 5 for the current work). It can therefore be solved in the least square sense, by minimizing the modeling error for each spectrum.

An analysis of the residuals provides a wealth of information on the model quality and can be used for the identification of unexpected compounds and process deviations (Equation 6). However, interpretation of a 3-dimensional view of the residues is not easy, and projections in time and wavenumber dimensions (SSQt and SSQwn respectively) give an overview on the model reliability (Equations 7-8).

$$\mathbf{E}_{s,wn} = \Delta \mathbf{A}_{s,wn} - \Delta \mathbf{C}_{s,j} \times \mathbf{K}_{j,wn} \quad [6]$$

$$SSQt = row_sum(\mathbf{E} \times \mathbf{E}) \quad [7]$$

$$SSQwn = column_sum(\mathbf{E} \times \mathbf{E}) \quad [8]$$

where *row_sum* and *column_sum* are used to indicate a sum of the matrix elements by column and row respectively, and the operator “ \times ” an element-by-element multiplication. Note that unlike \mathbf{E} , SSQt and SSQwn are not matrices but vectors. The sum of squared residues obtained for every time can be compared to the total variance found in the system, which gives the explained variance EV, expressed here in percentage:

$$EV = 100 \cdot [1 - SSQt ./ row_sum(\Delta \mathbf{A} \times \Delta \mathbf{A})] \quad [9]$$

where the “ $./$ ” operator refers to an element-by-element division. Since the least-squares algorithm minimizes the sum of the squared residues over a wide range of wavenumbers, SSQwn may become larger at some point of the spectrum than even the sum of the squares of the absorbencies (*column_sum*($\Delta \mathbf{A} \times \Delta \mathbf{A}$)) themselves. This can typically happen in regions where no metabolite absorbs. It therefore makes less sense to define an explained variance in the wavenumber dimension, since it could become negative and thus renders any interpretation highly questionable.

Validation of assumptions: linearity and additivity of mid-IR absorbances

The entire library concept is based on the hypothesis that the measured absorbance can be expressed as a linear combination of pure component spectra. This is valid only if the Lambert-Beer law is strictly followed and if the contribution from each component is additive. To address how linear the absorbance is with respect to concentration in the fingerprint region, infrared (IR) spectra of glucose aqueous solutions were measured at 30°C, with concentrations ranging from 1 to 500 g L⁻¹ (Fig. 1). For every wavenumber, the absorbance was plotted versus the concentration and the regression coefficient R^2 was determined as an evaluation of the linearity of the relationship. With regression coefficients very close to 1, the absorbance can be considered as very linear with respect to concentration within the 900-1600 cm⁻¹ region. The drop of R^2 at 1500 cm⁻¹ is not relevant, since the very low glucose absorbance in this region results in large relative experimental errors.

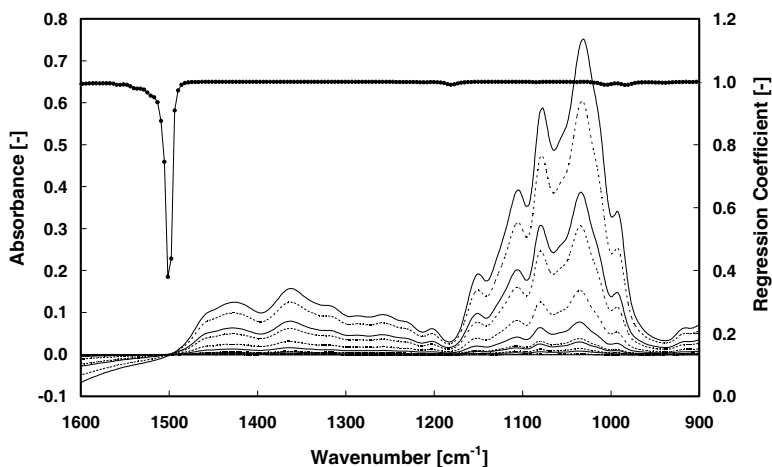


Figure 1. Absorbance of glucose solutions, for concentrations from 1 to 500 g L⁻¹ (lines, left axis), and regression coefficient from the linear fitting of absorbance vs. concentration (line with black symbols, right axis). The highest peaks correspond to 500 g L⁻¹ of glucose.

To address the spectra additivity, the molar absorbance of glucose, sodium acetate, ammonium chloride and ethanol was determined in water, at 30°C. In a second step, the absorbance of a solution consisting of 0.10 M sodium acetate, 0.13 M ammonium chloride and 0.11 M glucose was predicted by assuming additivity of the pure component spectra. This was then confronted to the real measured spectrum (Fig. 2). Profiles match almost perfectly within the 900-1600 cm⁻¹ region, which confirms that absorbances can indeed be considered as

additive. This suggests that all species in solution do not interact with each other from a spectroscopy point of view. On the basis of these results, the assumptions made for the development of the equations could be safely validated, as long as temperature and pH, which both influence infrared absorption, remain constant.

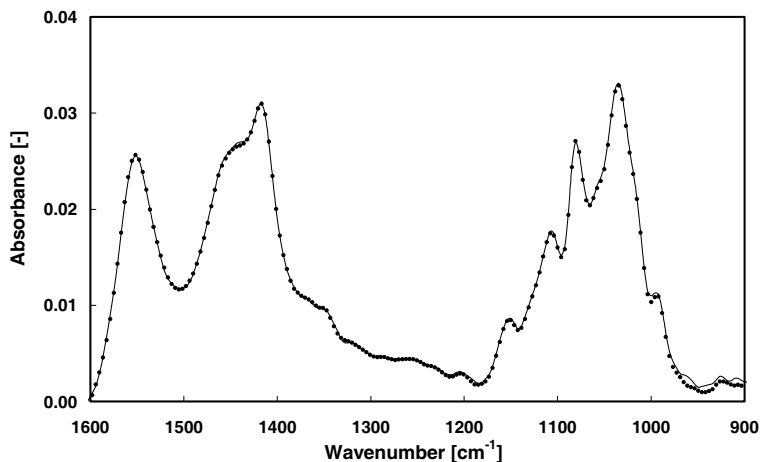


Figure 2. Measured (symbols) and predicted (plain line) absorbances of an aqueous solution with a known composition of glucose, sodium acetate and ammonium chloride.

Setting up the library

For each of the five compounds found in *S.cerevisiae* cultures, the absorbance of a solution of known concentration was measured in order to determine the molar absorbance spectra. Care was taken to control the temperature of the flow-through cell and the solutions at the process temperature, 30°C, since infrared spectra are known to be temperature-dependent (Wülfert, 1998). The pH of the solutions was adjusted to 5. Instrument specificity, especially regarding to optics alignment, restrains the use of literature data or commercially available libraries. This laboratory work, although unavoidable, can be completed within a couple of hours, due to the small number of standards that have to be prepared and measured.

Pure component spectra of ions such as ammonium and acetate ions cannot be isolated from the total salt absorbance, which includes peaks due to the counter-ion. This causes a problem for bioprocess monitoring, since microorganisms can for instance consume ammonium cations and produce acetate anions. The counter-ion interference had to be limited as far as possible and a series of salts were tested in order to find electrolytes that are as transparent as possible to IR. Since infrared spectroscopy relates to vibrational bond

energy, monoatomic ions such as Na^+ , K^+ or Cl^- are better candidates than larger molecules, such as phosphate or sulphate. Indeed sodium chloride showed a rather transparent behavior, with only a small peak at 1630 cm^{-1} (Fig. 3). Compared to the molar spectra of ammonium chloride or sodium acetate, the influence of Na^+ and Cl^- was considered as not significant.

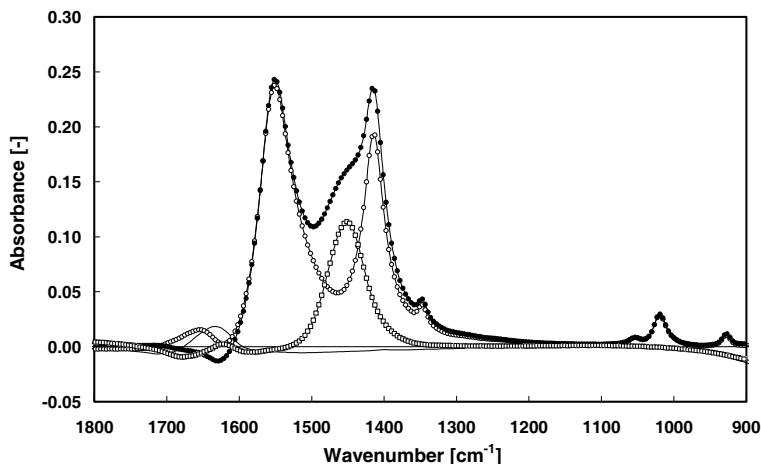


Figure 3. Absorption of 1 M sodium chloride (plain line), ammonium chloride (open squares), sodium acetate (open circles) and ammonium acetate (black circles).

Molar absorbances shown in Figure 3 also confirm that the contributions from the different compounds can be added linearly. Thus the spectrum of ammonium acetate can be predicted by subtracting sodium chloride from the sum of ammonium chloride and sodium acetate.

Correction of baseline drift and spectral range used for calculation

Spectral range for calculation was set to $1000\text{--}1600\text{ cm}^{-1}$, in the “fingerprint region”, where the signal to noise ratio is the highest for aqueous solutions. This range was chosen in order to include all significant peaks (ethanol peak: $1028\text{--}1065\text{ cm}^{-1}$), but also to exclude noisy regions observed below 1000 cm^{-1} and above 1600 cm^{-1} (Fig. 4). Signal instability of the FTIR single-beam instrument has already been described in detail by Doak and Phillips (Doak and Phillips, 1999). To highlight this point, intensity drifts of up to 10% could be observed for the instrument used in this study, which corresponds to the absorbance change induced by a glucose concentration of 10 g L^{-1} . Even though baseline instability is a critical issue that must

be addressed in order to achieve a fair accuracy, it has received surprisingly little attention. This can be explained by the fact that the signal drift can be indirectly corrected by process-related information, such as the use of in-process standards (Kornmann et al., 2004; Pollard et al., 2001). Although this approach ensures a relative robustness, it necessitates sample withdrawal and off-line analysis, with the result that monitoring is less predictive. A convenient solution consists in performing the FTIR measurements off-line with samples, which allow compensating for the baseline drift by measuring a reference solution such as water (Crowley et al., 2000; Hashimoto et al., 2005).

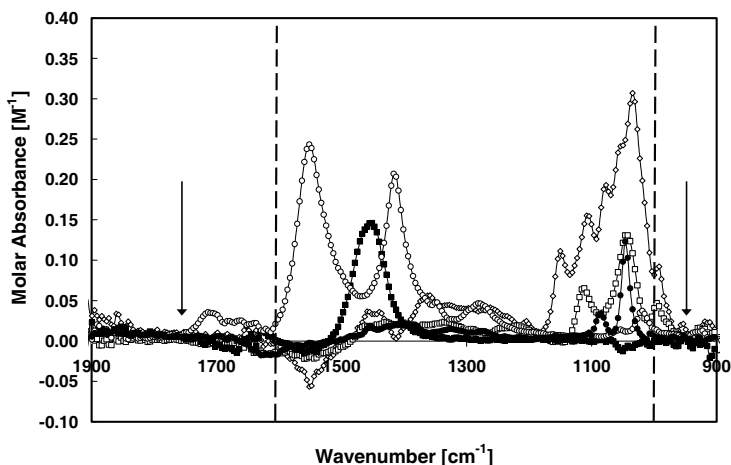


Figure 4. Molar absorbance spectra of glucose (○), ethanol (●), acetate (○), glycerol (□) and ammonium (■). Dashed lines indicate the range used for calculations (1000 – 1600 cm⁻¹); arrows indicate the two anchorage points (950 and 1750 cm⁻¹).

In the current study, a previously described anchorage method has been used for simplicity (Schenk et al., 2007). It is based on the observation that even though the signal drift is wavenumber-dependant, it can be considered as relatively linear within the fingerprint region. Two anchors were set at 950 and 1750 cm⁻¹ (Fig. 4), and all spectra intensities were corrected using Equation 10, in order to calculate absorbances and resulting concentrations.

$$I_{i,v}^{corrected} = \frac{I_{i,v}}{0.5 \cdot \left(\frac{I_{i,950}}{I_{o,950}} + \frac{I_{i,1750}}{I_{o,1750}} \right)} \quad [10]$$

On-line monitoring of a *S. cerevisiae* culture

The library-based calibration method was tested with an aerobic fermentation of the wild-type *S. cerevisiae* strain CBS 8066. As indicated by off-line analysis, glucose consumption was coupled to a massive ethanol production (Fig. 5). During this phase, glycerol and acetate were also produced, but in much lower amounts. The optical density at 600 nm reached the value of 4, which corresponds to a dry cell weight of 3.4 g L^{-1} . After a short lag phase ethanol was consumed simultaneously with acetate and later glycerol, which led to a final OD600 of 10.3 (equivalent to 7.0 g L^{-1}). The culture lasted for about 15 hours and is representative of a typical baker's yeast fermentation (Sonnleitner and Käppeli, 1986). A carbon balance over the process closed to 98.5%, which suggests that no major carbon-containing compounds have been unaccounted for.

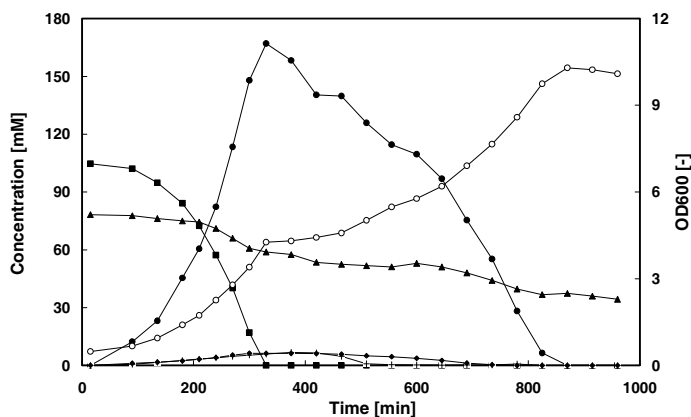


Figure 5. Concentration profiles measured off-line during an aerobic batch fermentation of a wild-type *S. cerevisiae*. Black symbols refer to the left axis: glucose (squares), ammonium (triangles), ethanol (circles), glycerol (diamonds) and acetate (crosses). Open circles show the optical density at 600 nm on the right axis.

FTIR concentration profiles were obtained using signal drift correction (Equation 10) and subsequent least-squares modeling (Equation 5), using the five spectra from the library as a calibration set (Fig. 6a). No filtering or moving average was applied to the data. Since concentration differences from the initial medium are calculated with this method, substrates appear with negative values whereas products are positive.

The main features of the baker's yeast fermentation observed with the off-line analysis are also described by the FTIR on-line data. Profile shapes of glucose, ethanol and ammonium

clearly highlight the three-stage process, with the first glucose phase, followed by concomitant consumption of ethanol, acetate and glycerol, and the final ethanol-only phase. The glycerol and acetate concentrations were not accurately predicted, but it could be qualitatively concluded that they never exceeded 15 mM, which was confirmed by off-line analysis (maximum acetate and glycerol concentrations: 7 and 6 mM respectively). Glycerol and ethanol monitoring is especially delicate due to the weak infrared activity of these species. Their maximum molar absorbances are both around 0.13 M^{-1} , whereas at the same wavenumbers glucose shows a three-fold higher molar absorbance (0.31 M^{-1}). This largely explains why the ethanol profile has more associated noise compared with the others.

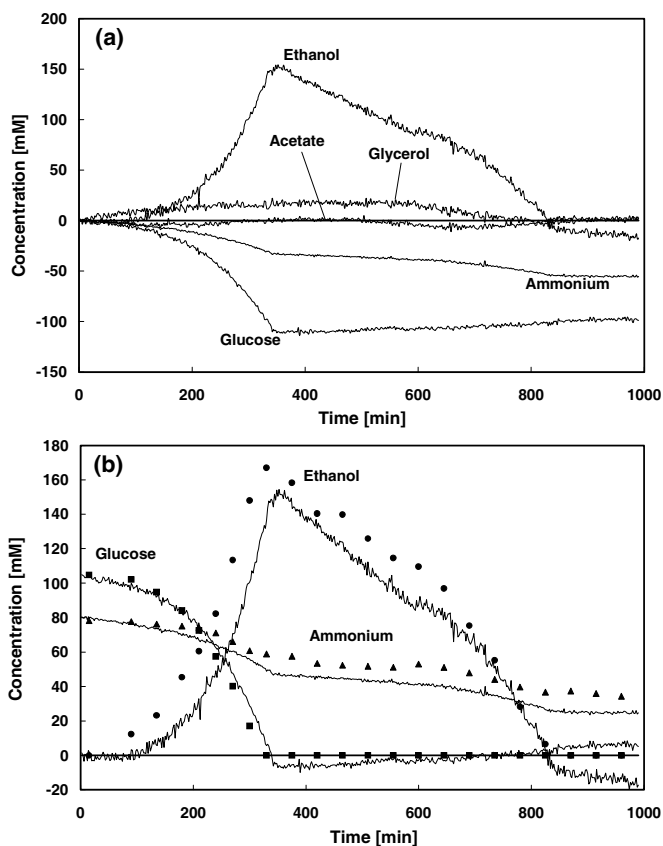


Figure 6. a) Concentration difference from the initial state obtained by on-line FTIR for the aerobic *S. cerevisiae* fermentation. b) Absolute concentration profiles re-calculated from the medium composition. Only glucose, ethanol and ammonium profiles are reported for clarity reasons.

Absolute concentration profiles could be readily recalculated using the initial concentrations, given by the medium composition (Fig. 6b). While glucose was predicted with a fair precision, both ethanol and ammonium showed a higher offset, which can be explained by the smaller molar absorbance of these two latter compounds.

The accuracy of the library-based calibration was assessed with the standard error of prediction (SEP), which can be calculated from the following formula:

$$SEP = \sqrt{\frac{\left(\sum_k y_{k,Off-line} - y_{k,FTIR} \right)^2}{n}} \quad [11]$$

where y_k indicates any off-line data or its corresponding FTIR value, and n refers to the total number of samples. Calculated SEP values for the five compounds are shown in table 1.

Table 1. Standard error of prediction (SEP) for the five main compounds found in the *S.cerevisiae* fermentation. Total number of data for SEP calculation: 28.

Compound	SEP [mM]	SEP [g L ⁻¹]
Glucose	4.8	0.86
Ethanol	21.4	0.98
Ammonium (NH ₄ ⁺)	8.4	0.15
Glycerol	5.4	0.32
Acetate (CH ₃ COO ⁻)	8.8	0.81

Even though all SEP values are of the same order of magnitude, they have to be confronted with the total concentration variations. For glucose, ethanol, and ammonium the SEP values are between 10 to 20 % of the total concentration change, but this value rises to 50 and 200 % for acetate and glycerol respectively. Consequently, the method can be considered as relatively accurate, but not reliable below 0.5 g L⁻¹. For a second culture performed in exactly the same conditions, a 49-standard calibration set, with Partial Least-Squares modeling, yielded to SEP values of 1.42, 0.92 and 0.38 g L⁻¹ for glucose, ethanol and ammonium respectively. This demonstrates that the precision of more complete chemometrics calibrations is very similar to what was obtained with a library-based approach.

Residuals analysis and explained variance

In a similar way to chemometric calibrations, the library-based method allows addressing the quality of the modeling. It also tracks the appearance of unexpected compounds and even points to their identity. In order to highlight how relevant information can be obtained from a

residuals analysis, the cultivation data were reprocessed with an incomplete model containing only the spectra of glucose and ammonium (e.g. ethanol, glycerol and acetate were not included into the library) (Fig. 7). The two-spectrum library could predict the ammonium profile as accurately as the 5-component model (SEP: 6.7 mM), but the error of prediction on glucose was 4-time higher (SEP: 19.2 mM). Not including ethanol into the model had a larger on the glucose prediction, because these two compound absorb in the same wavenumber range (glucose: 1000 - 1180 cm^{-1} , ethanol: 1046 cm^{-1}), whereas ammonium absorbs at higher frequencies (1420 - 1500 cm^{-1}). It is interesting to note that such a simple model already allowed predicting the end of the glucose batch phase that occurred after 350 minutes.

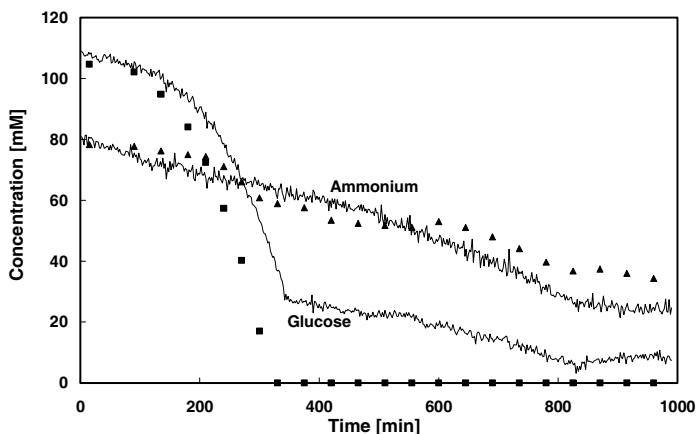


Figure 7. Glucose and ammonium concentration profiles calculated using only the glucose and ammonium spectra in the library. Even though the three other compounds were missing, it is possible to identify the end of the batch phase after 350 min.

For every spectrum taken during the cultivation, the least-squares algorithm minimized the error of prediction, which left an array of residues that can be plotted against the wavenumbers. This led to 3-dimensional overview of the unmodeled absorbance observed during the process (Fig. 8). The “shark fin” that can be observed in the residual plot clearly indicates that an unexpected compound was produced and subsequently consumed during the process, which is obviously in this case ethanol.

Such a conclusion can also be drawn from 2-dimensional projections of the residual plot. More precisely, the projection in the time dimension is used to detect the appearance of unexpected metabolites, whereas the projection in the wavenumbers dimension points to their identity.

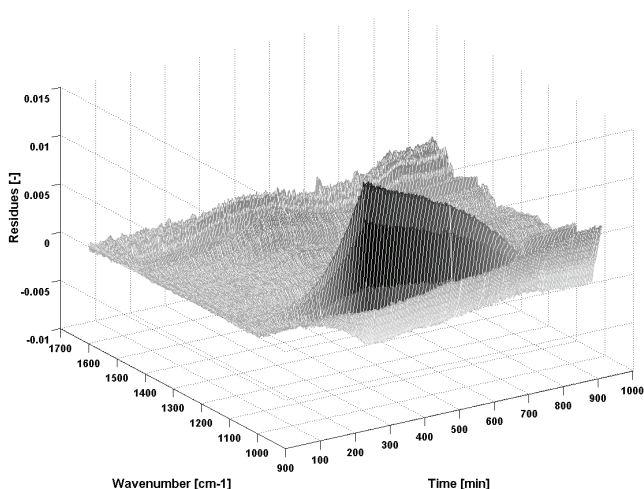


Figure 8. Three-dimensional view of the residuals left by the modeling made with only the glucose and ammonium spectra in the library. The production and subsequent consumption of ethanol that could not be predicted by the incomplete model can be here clearly identified with the residuals.

The evolution of the sum of the squared residues (SSQt, equation 7) and the explained variance (EV, equation 9) during the time-course of the process were plotted in figure 9a. The SSQt increased exponentially during the batch phase on glucose and then returned to a 3-fold lower value by the end of the cultivation, which suggests that a growth-associated metabolite was produced and subsequently consumed. The explained variance reached 90% within 200 minutes, when the cells entered exponential phase and thus provoked significant concentration changes in the medium. This value increased slowly while ethanol was consumed to finally stabilize around 98%. The small ethanol molar absorbance explains why the explained variance averaged over the entire process was so high (89.7%).

The identification of unexpected compound is considerably facilitated by the projection of the residues in the wavenumbers dimension (Fig. 9b). The sum of the squared residues at every wavenumber showed a sharp peak at 1046 cm^{-1} , typical of ethanol IR absorption. As opposed to more sophisticated chemometric methods that work with factors and loadings that do not always have a physical meaning, the library-based calibration approach allows identification of unexpected compounds since it provides their absorbance fingerprint.

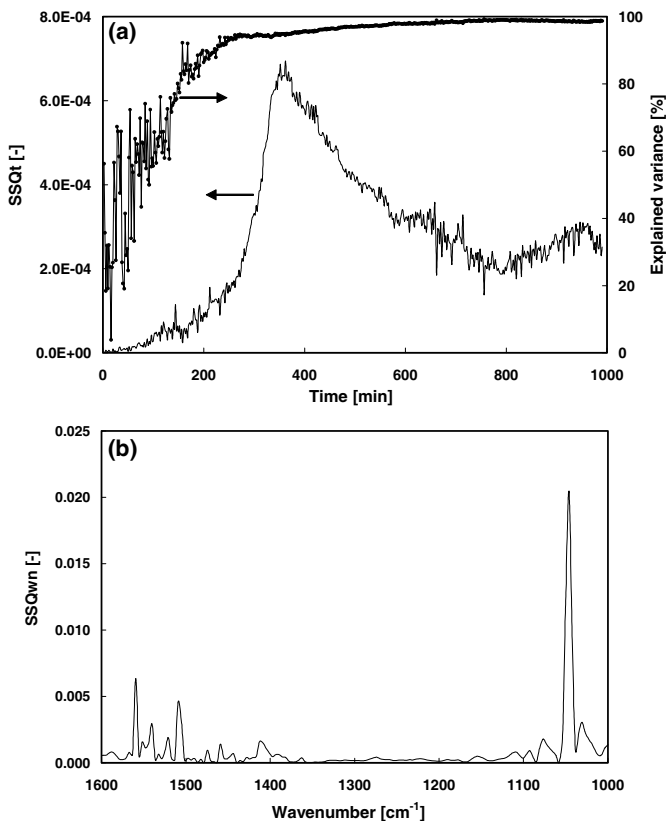


Figure 9. a) Sum of the squared residues in the time dimension (SSQt, line without symbols, left axis) and explained variance during the time-course of the process (line with symbols, right axis). Even though the explained variance was above 90% during most of the cultivation, the sum of squared residues clearly indicates that an unexpected compound was produced and subsequently produced during the process b) Sum of the squared residues in function of the wavenumbers (SSQwn) for the modeling with only glucose and ammonium. This projection of the squared residues helps identifying the missing compound: the sharp peak at 1046 cm⁻¹ can be attributed to ethanol.

The cultivation data were also reprocessed with other models in order to discuss whether the averaged explained variance could be used as a global indicator of the model quality, since this value can be computed quickly and automatically and does not require any off-line information (Table 2). The glucose alone explained 86.1% of the total absorbance variation during the process. This is due to the large glucose absorbance and the concentration change that amounted to 20 g L⁻¹ during the cultivation. For the same reasons, the model including

only ethanol and ammonium led to high error of prediction and low averaged explained variance.

Table 2. Averaged explained variance (Avge EV) and Standard error of prediction (SEP) for glucose, ethanol and ammonium, depending on the compounds used in the model. Glucose alone explained more than 85% of the total variance, due to its high initial concentration and strong absorbance. Addition of two extra minor compounds (acetate and glycerol) did not significantly improve the modeling accuracy. A star means that this compound was not included in the model and therefore it was not possible to extract its concentration profile.

Number	Compounds in models	SEP [mM] (Glu / EtOH / NH ₄ ⁺)	Avge EV [%]
1	Glucose	18 / * / *	86.1
2	Ethanol	* / 378 / *	54.6
3	Ammonium (NH ₄ ⁺)	* / * / 31	9.1
4	Glucose - Ethanol	12 / 93 / *	87.3
5	Glucose - NH ₄ ⁺	19 / * / 6.7	89.7
6	Ethanol - NH ₄ ⁺	* / 91 / 15	56.1
7	Glucose – Ethanol - NH ₄ ⁺	8.2 / 19 / 9.9	91.3
8	All five compounds	4.8 / 21 / 8.4	94.0
9	Glucose - NH ₄ ⁺ - Lactate	19 / * / 6.5	91.4
10	Glucose – Ethanol – NH ₄ ⁺ - Lactate	8.9 / 16 / 9.5	93.1

Using in the model only the three main metabolites (glucose, ethanol and ammonium, model number 7) the concentration profiles could be already accurately predicted, since the addition of the two minor metabolites only increased the averaged explained variance of less than 3% but did not significantly decrease the SEP values (model 8). Unsurprisingly, the averaged explained variance increased with the number of component included into the library, since this made the model more flexible and able to fit the noise by playing with the minor metabolites. The statistics of the eight first models tested shows that the library-approach is quite robust with respect to missing compounds: an estimate of the growth rate and the ethanol yield could be obtained with only three standards in the library instead of five. Lactate, which was measured off-line and never found during this cultivation, was included into two other models. In both cases, the lactate concentration profile was correctly computed as being almost constant and close to zero over the entire time-course of the process. SEP values of the other compounds were not significantly affected by this unnecessary standard (models 5 vs. 9 and 7 vs. 10), which also confirms the high robustness of the prediction. However, lactate amounted to 2% of the averaged explained variance, which is as much as ethanol. This indicator cannot therefore be used to automatically select the right compounds from a large spectra database. In conclusion, having a high explained variance is a mandatory but not sufficient condition to consider a model as reliable. It

therefore seems that a manual inspection of the residual remains unavoidable to address modeling quality.

Model robustness regarding to spiking of compounds

As pointed out by Rhiel and co-workers (Rhiel et al., 2002a), model robustness can be significantly reduced by correlations within the calibration set. This frequently happens when samples from previous cultivations are used as standards to calibrate the instrument. In such a case, the different concentrations are bound to each other by the cell metabolism, and the model is therefore unable to predict variations that are not due to the cell activity. In the present study, the library-based model robustness was addressed by a spiking experiment. A batch culture was carried out in the same conditions, but after 95, 245 and 350 minutes, 30 mmoles of NH_4Cl , 50 mmoles of glucose and 150 mmoles of ethanol respectively were added with a syringe to the 2-litre bioreactor in order to produce concentration changes not related to the yeast metabolism. The culture was then stopped at end of the batch phase. The 5-component library (glucose, ammonium, ethanol, glycerol, acetate) was perfectly able to predict the additions of ammonium chloride, glucose and ethanol (Fig. 10).

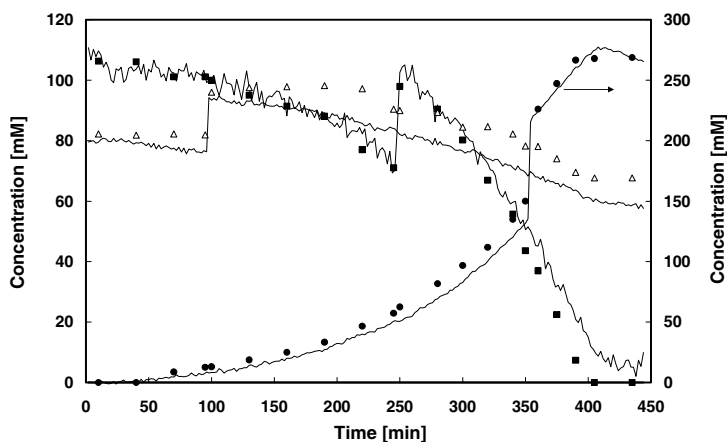


Figure 10. FTIR (lines) and off-line concentration profiles of glucose (black squares, left axis), ammonium (open triangles, left axis) and ethanol (black circles, right axis) during a batch phase on glucose of *S.cerevisiae*. In order to address the robustness of the model, 15 mM of ammonium, 25 mM of glucose and 75 mM of ethanol were added after 95 min, 245 min and 350 min respectively. Concentration profiles of acetate and glycerol were omitted for clarity reasons.

This demonstrates that the model was not correlated, since a sudden change in the concentration of a compound due to an external perturbation did not perturb the prediction of the other concentrations. Acetate and glycerol profiles were not reported for clarity reasons because they remained below 15 mM. SEP values were 1.26, 0.43 and 0.14 g L⁻¹ for glucose, ethanol and ammonium respectively, which is very similar to what was obtained in the first experiment.

This spiking experiment demonstrates that library-based calibrations are certainly not correlated. The reason for that is that synthetic and single-component standards were used to make the library, without any data from previous cultivations, which could induce correlations due to the metabolism. More generally, since the number of spectra in the library equals the number of concentrations predicted, the standards are expected to be linearly independent and therefore correlations can arise only if the molar absorbances of two compounds are perfectly proportional over the entire range used for calculations (1000 -1600 cm⁻¹).

5/ CONCLUSIONS

A library made of five single-component spectra was shown to be able to predict the concentration profiles of the five major metabolites found in *S.cerevisiae* aerobic fermentations. This new calibration method for the on-line monitoring of bioprocesses by FTIR showed a similar accuracy and robustness to what is generally obtained through the chemometrics approach. However, it is less time-consuming, much simpler and still allows detecting unexpected compounds, as indicated by the residuals analysis variance. Once a complete library is collected for a given instrument, many strains, organisms and defined media can be tested without any further work, which is the most evident benefit gained by this method.

While the actual trend in medium formulation encourages the use of defined media, the ability of the library-based method to predict concentration profiles in rich and semi-defined media remains unclear. To our opinion, this calibration should withstand the addition of chemically undefined compounds such as yeast extract, as long as they are not massively consumed during the process. This issue is particularly relevant and should be investigated in a future work.

As pointed out by Harms *et al.*, recent advances in microfluidics and sensors are turning bioprocess optimization into massively parallel experiments in fully automated and monitored environment (Harms et al., 2002). For such systems, microsenors for pH and dissolved oxygen are now commercially available, but there is to date no ready-to-use solution for the non-invasive measurement of key metabolites. FTIR, together with the library-based calibration approach, is foreseen to fill this gap and therefore plays an important role in future high-throughput platforms for strain and medium screening.

6/ REFERENCES

- ASTM. 1997. Standard practices for infrared, multivariate, quantitative analysis (E 1655-97). Annual book of ASTM standards. Philadelphia: American Society for Testing and Materials. p 844-869.
- Brereton RG. 1997. Multilevel multifactor designs for multivariate calibration. *The Analyst* 122:1521-1529.
- Cooper BC, Wise KL, Welch JT, Sumner MB, Wilt KW, Bledsoe RR. 1997. Comparison of near-IR, raman, and mid-IR spectroscopies for the determination of BTEX in petroleum fuels. *Appl. Spectrosc.* 51(11):1613-1620.
- Crowley J, McCarthy B, Nunn NS, Harvey LM, McNeil B. 2000. Monitoring a recombinant *Pichia pastoris* fed batch process using Fourier transform mid-infrared spectroscopy (FT-MIRS). *Biotechnol. Lett.* 22:1907-1912.
- Doak DL, Phillips JA. 1999. In situ monitoring of an *Escherichia coli* fermentation using a diamond composition ATR probe and mid-infrared spectroscopy. *Biotechnol. Prog.* 15:529-539.
- Fairbrother P, George WO, Williams JM. 1991. Whey fermentation: On-line analysis of lactic acid by FTIR spectroscopy. *Appl. Microbiol. Biotechnol.* 35:301-305.
- Harms P, Kostov Y, Rao G. 2002. Bioprocess monitoring. *Current Opinion in Biotechnology* 13(2):124-127.
- Hashimoto A, Yamanaka A, Kanou M, Nakanishi K, Kameoka T. 2005. Simple and rapid determination of metabolite content in plant cell culture medium using an FTIR/ATR method. *Bioprocess. Biosyst. Eng.* 27:115-123.
- Kornmann H, Rhiel M, Cannizzaro C, Marison I, von stockar U. 2003. Methodology for real-time, multi-analyte monitoring of fermentations using an in-situ mid-infrared sensor. *Biotechnol. Bioeng.* 82(6):702-709.
- Kornmann H, Valentinotti S, Marison IW, von Stockar U. 2004. Real-time update of calibration model for better monitoring of batch processes using spectroscopy. *Biotechnol. Bioeng.* 87(5):593-601.
- Martens H, Naes T. 1992. Multivariate calibration. New York: John Wiley & Sons. 432 p.
- Navarro-Villoslada F, Perez-Arribas LV, Leon-Gonzalez ME, Polo-Diez LM. 1995. Selection of calibration mixtures and wavelengths for different multivariate calibration methods. *Anal. Chim. Acta* 313(1-2):93-101.

- Pollard DJ, Buccino R, Connors NC, Kirschner TF, Olewinski RC, Saini K, Salmon PM. 2001. Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy. *Bioprocess. Biosyst. Eng.* 24:13-24.
- Rhiel MH, Amrhein MI, Marison IW, von Stockar U. 2002a. The influence of correlated calibration samples on the prediction performance of multivariate models based on mid-Infrared spectra of animal cell cultures. *Anal. Chem.* 74(20):5227-5236.
- Rhiel MH, Ducommun P, Bolzonella I, Marison IW, von Stockar U. 2002b. Real-time in situ monitoring of freely suspended and immobilized cell cultures based on mid-infrared spectroscopic measurements. *Biotechnol. Bioeng.* 77(2):174-185.
- Riley MR, Arnold MA, Murhammer DW, Walls EL, DeLaCruz N. 1998. Adaptive Calibration Scheme for Quantification of Nutrients and Byproducts in Insect Cell Bioreactors by Near-Infrared Spectroscopy. *Biotechnol. Prog.* 14(3):527-533.
- Riley MR, Crider HM, Nite ME, Garcia A, Woo J, Wegge RM. 2001. Simultaneous Measurement of 19 Components in Serum-Containing Animal Cell Culture Media by Fourier Transform Near-Infrared Spectroscopy. *Biotechnol. Prog.* 17(2):376-378.
- Schenk J, Marison IW, von Stockar U. 2007. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2):344-353.
- Sonnleitner B, Käppeli O. 1986. Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. *Biotechnol. Bioeng.* 28:927-937.
- Vojinovic V, Cabral JMS, Fonseca LP. 2006. Real-time bioprocess monitoring. Part I: In situ sensors. *Sensors and Actuators B* 114:1083-1091.
- von Schalien R, Fagervik K, Saxen B, Ringbom K, Rydstrom M. 1995. Adaptive online model for aerobic *Saccharomyces cerevisiae* fermentation. *Biotechnol. Bioeng.* 48(6):631-638.
- Wülfert F. 1998. Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Anal. Chem.* 70(9):1761-1767.

7/ LIST OF SYMBOLS

Matrices and vectors

A	$s \times wn$	Absorbance matrix: s absorbances measured during the process, at wn different wavenumbers
C	$s \times j$	Matrix of concentrations: the concentrations of the j species calculated from the s spectra measured during the process
E	$s \times wn$	Error matrix: the residues found at wn wavenumbers, for the s spectra measured during the process
K	$j \times wn$	Spectra library: the molar absorbance of j species at wn different wavenumbers
SSQ_{wn}	$1 \times wn$	Sum of square residues in wavenumber dimension (vector)
SSQ_t	$s \times 1$	Sum of square residues in time dimension (vector)

Greek symbols

ε	Experimental error on absorbance [-]
κ	Molar absorptivity (or molar extinction coefficient) [M ⁻¹]
ν	Wavenumber [cm ⁻¹]

Other symbols

A	Absorbance [-]
C	Concentration [g L ⁻¹], or [mol L ⁻¹]
EV	Explained variance [%]
I	Spectrum intensity [-]
j	Number of species included in the library ($j \leq m$)
m	Number of different species involved in the process
SEP	Standard error of prediction [-]
y	Predicted or measured concentration [g L ⁻¹], or [mol L ⁻¹]

Superscripts and subscripts

<i>baseline</i>	subscript	Baseline (refers to the initial culture medium)
k	subscript	k-th sample
i	subscript	i-th species
n	subscript	Number of off-line samples
s	subscript	Number of spectra measured during the process
tot	subscript	Total
wn	subscript	Number of wavenumber measured

Chapter 4

On-line Monitoring of Nine Different Batch Cultures of *E. coli* by Mid-infrared Spectroscopy, Using a Single Spectra Library for Calibration

1/ ABSTRACT

A single spectra library was used to monitor on-line, by mid-infrared spectroscopy, nine different batch cultures of *Escherichia coli* performed with various medium compositions, including chemically complex formulations. Whereas the classic chemometrics approach would have required the preparation and measurement of hundreds of standards, only 6 spectra were included in the library. These included the molar absorbance of the four main metabolites (i.e. glucose, glycerol, ammonium and acetate), and the remaining two were drift spectra found by factor analysis. The accuracy of the prediction was not altered by a change of the carbon source, the ammonium concentration or even the addition of chemically undefined compounds, such as yeast extract and peptone. The standard errors of prediction averaged over the nine experiments were 8.0, 12.3, 5.9 and 5.6 mM for glucose, glycerol, ammonium and acetate respectively. Inclusion of two drift spectra in the library provided an estimation of how noisy an experiment was. This also allowed detection of batch cultures that require further investigation, namely runs which were subject to large signal drift or during which an unexpected compound was produced, without having to carry out time-consuming off-line analyses.

This chapter was submitted for publication in *Journal of Biotechnology*:

Schenk J, Viscasillas C, Marison IW, von Stockar U. On-line monitoring of nine different batch cultures of *E. coli* by mid-infrared spectroscopy, using a single spectra library for calibration.

2/ INTRODUCTION

Fourier-transform mid-infrared spectroscopy (FTIR) is commonly presented as a technique with very appealing characteristics, since it can monitor simultaneously a high number of compounds, is fast, non-invasive, and can be coupled to “attenuated total reflectance” (ATR) probes that can be cleaned and sterilized in-situ (Clementschietsch and Bayer, 2006; Kornmann et al., 2004; Roychoudhury et al., 2006b; Vojinovic et al., 2006). Published work in this area clearly stress that calibrating the instrument is heavily time-consuming, and requires a high level of expertise in *process chemometrics* for data treatment. This important drawback, together with the cost of instrumentation, has largely hindered the widespread, routine utilization of FTIR for bioprocess monitoring.

The recommended procedure for the development of multivariate calibration models involves the preparation and measurement of numerous standards that span the entire range of concentration for all compounds of interest. According to ASTM recommendations (ASTM, 1997) on standard practices, the number of solutions to be produced to insure a reproducible calibration set should be higher than six times the number of relevant components. Attention must be paid at this stage to avoid correlated combinations, by the use of appropriate multi-level experimental design (commonly four to five levels) (Brereton, 2000; Rhiel et al., 2002). Building the mathematical model involves the use of advanced mathematical tools, in order to extract the relevant information, including Partial Least Squares (PLS), Principal Component Analysis (PCA) and Evolving Factor Analysis (EFA) (Martens and Naes, 1992). This entire procedure commonly requires two days of labour and has to be repeated if the medium composition is changed.

While this approach has been successfully applied to various kinds of cultures from laboratory to pilot scale (Crowley et al., 2000; Doak and Phillips, 1999; Kornmann et al., 2003; Pollard et al., 2001; Roychoudhury et al., 2006a), it is particularly unsuitable for process development, during which a large number of new media and conditions (pH, temperature, and so on) must generally be tested. Ironically enough, it is exactly for such applications that FTIR is expected to play an important role (Harms et al., 2002). While on-line monitoring of pH, dissolved oxygen and cell concentration can now be implemented relatively easily in screening platforms (Betts and Baganz, 2006; Harms et al., 2005), the on-line and in-situ monitoring of metabolite concentrations remains impracticable. Mid-infrared spectroscopy is among the very few techniques that can achieve this, and would therefore be highly complementary with standard bioprocess measurements, were it not for the calibration need

requiring the preparation and testing of a number of standards that is clearly prohibiting for a high-throughput platform. If FTIR is to deploy its very considerable potential in bioprocess development, dramatically simplified calibration procedures need to be developed.

Starting from the observation that a multi-level design of the calibration set could be oversized for subsequent linear modeling, we showed in a previous publication that the calibration procedure could be dramatically simplified by using a library of pure component spectra, without influencing significantly the quality of prediction (Schenk et al., 2007b). Briefly, this approach consists in modeling any process spectra as a linear combination of the molar absorbance of the principal metabolites ("pure" compound spectra), while intensity drifts and noise are accounted for by a signal anchoring method. This approach assumes that absorbances add linearly and that the Lambert-Beer law is respected, two conditions that are generally met at constant pH and temperature. Beside the 6-fold decrease in the number of standards to be prepared, this approach also presents the advantage of allowing a more knowledge-based interpretation of results, such as the analysis of residuals to detect the appearance of unexpected metabolites and point to their identity. At a more conceptual level, this approach, as opposed to process chemometrics methodology, addresses separately the metabolite concentration modeling and signal drift correction, rather than lumping them together by the use of virtual factors.

The aim of this paper is to show that a single spectra library can be used to predict the concentration of the principal metabolites during cultures conducted on different media. For that purpose, *Escherichia coli* was chosen as a model organism, due to its wide use in research and industrial production. Nine batch experiments were carried out, which allowed addressing the reliability of the library-based monitoring with respect to different medium compositions, strain type, and the use of chemically complex additives, such as peptone and yeast extract. Signal intensity drifts and noise problems were solved by including into the spectra library, in addition to the four pure component spectra (i.e. glucose, glycerol, ammonium and acetate), two "drift spectra", obtained by factor analysis.

3/ MATERIALS AND METHODS

Bioreactor Fermentation

The wild-type strains of *Escherichia coli* ML308, W3100 and W3110 were used in these experiments (DSMZ No 1329, 6302 and 5911 respectively, Braunschweig, Germany). The inoculum was prepared by growing the cells in two shake flasks containing 100 ml of the bioreactor medium for 14 hours. Medium composition for each culture is summarized in Table 1.

Table 1. Media composition for the nine different batches. Indicated quantities are for 1 litre of medium. Media A, B and C have different carbon sources. Medium D is nitrogen-limited, whereas the others are carbon-limited. Media E, F and G are chemically complex.

	A	B	C	D	E	F	G
Glucose [g]	15.0	-	7.5	15.0	15.0	15.0	15.0
Glycerol [g]	-	15.0	7.5	-	-	-	-
KH ₂ PO ₄ [g]	13.3	13.3	13.3	13.3	13.3	13.3	13.3
(NH ₄) ₂ PO ₄ [g]	5.0	5.0	5.0	2.5	-	5.0	5.0
NH ₄ Cl [g]	2.0	2.0	2.0	1	-	2.0	2.0
Citric acid anhydrous [g]	1.7	1.7	1.7	1.7	-	1.7	1.7
MgSO ₄ ·7H ₂ O [g]	1.2	1.2	1.2	1.2	-	1.2	1.2
Na ₂ EDTA·2H ₂ O [mg]	10.7	10.7	10.7	10.7	-	10.7	10.7
Tryptone [g]	-	-	-	-	10	-	-
Yeast extract [g]	-	-	-	-	5	5	1
Trace elements solution [ml]	10.0	10.0	10.0	10.0	10.0	10.0	10.0
Vitamin solution [ml]	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Antifoam [ml]	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Medium used for Batches	B01* B08** B09***	B02*	B03*	B04*	B05*	B06*	B07*

*) Strain used: ML308

**) Strain used: W3100

***) Strain used: W3110

The pH of the medium was adjusted to 6.5 by addition of NaOH 5 M. Sterile solutions of antifoam (Struktol SB2121, Schill and Seilacher, Hamburg, Germany), vitamins (thiamine.HCl 4.5 g L⁻¹) and trace elements (composition per litre: 0.3 g H₃BO₃, 0.15 g CuCl₂·4H₂O, 1.35 g MnSO₄·H₂O, 0.25 g Na₂MoO₄·2H₂O, 1.3 g Zn(CH₃CHOO)₂·2H₂O, 0.25 g CoCl₂·6H₂O, 11 g FeCl₃·6H₂O) were added to the filter-sterilized medium (0.2 µm Steritop®, Millipore, MA). Fermentations were carried out at 37°C in a 3.6-L bioreactor (Bioengineering, Wald,

Switzerland) with a working volume of 2 litres. Agitation was set at 1000 rpm, with three 6-blade Rushton impellers, and total aeration was constant at 3 NL/min, a value sufficient to maintain the dissolved oxygen above 20% of air saturation. Oxygen and carbon dioxide composition of the exhaust gas was measured using a gas analyzer (Dr. Marino Müller Systems, Esslingen, Switzerland).

Off-line sample analysis

Cultures samples (5 mL) were collected at intervals of 10 to 30 minutes (depending on the growth kinetics) and immediately cooled to 2°C. Dry cell weight (g-DCW L⁻¹) was determined by gravimetric analysis: 3 ml of sample were filtered on 0.2 µm pre-weighed filters, and then dried to constant weight. Optical density was measured at 600 nm (OD₆₀₀). The remaining volume of the sample was filtered through 0.2 µm filters for off-line analysis of metabolites. Glucose, glycerol, ammonium ion, acetic acid, and other organic acid concentrations were determined by enzymatic assay, using commercially available kits (R-Biopharm AG, Darmstadt, Germany) and an automated analyzer (Cobas Mira, Roche, Basel, Switzerland).

FTIR Set-up and spectra acquisition

A Mettler Toledo FTIR 4000 (Mettler-Toledo, Greifensee, Switzerland) with an Attenuated Total Reflectance probe (ATR, DiComp) was mounted in a flow-through cell, which allows convenient measurement of standards in an off-line configuration, but also connection to a bioreactor for on-line operation. This assembly is very similar to an in-situ ATR-probe, since it allows cleaning and sterilization in place. More details on the instrument and the set-up can be found in previous publications (Doak and Phillips, 1999; Schenk et al., 2007a). In order to measure the molar absorbance of pure components for the library, the intensity spectra (I) of the standards were determined in parallel with the intensity spectrum of water, which was used as reference intensity (I_o). Measurements were performed at 37°C and a pH of 6.5 (pH was adjusted using NaOH 1 M or HCl 1 M). Each standard contained a single compound at a precisely known concentration within the range 50 – 100 mM. Three series of 64 single-beam intensity spectra with a resolution of 4 cm⁻¹ were collected and averaged for each measurement, and the absorbance A at a wavenumber ν was calculated using:

$$A_{\nu} = \log_{10} \left(\frac{I_{\nu,o}}{I_{\nu}} \right) \quad [1]$$

where I_0 is the reference intensity (i.e. intensity spectrum of water when working off-line) and I is the intensity of the sample. The 1000 - 1500 cm^{-1} range was used for calculations, which is sometimes referred to as the “fingerprint region”, since it features high sensitivity and distinctive absorption bands for almost all compounds. This range was chosen to include all significant peaks, while excluding noisy frequencies (below 1000 cm^{-1} and above 1500 cm^{-1}). Other interesting regions of the mid-IR spectrum (e.g. 2800 – 2950 cm^{-1}) were not used for calculation, in order to keep modeling as simple as possible.

During cultures, a recirculation loop driven by a membrane pump (Prominent Gamma/L, Heidelberg, Germany) was used to connect the reactor to the FTIR flow-through cell. The residence time within the loop was less than 20 seconds, while the response time was 15 seconds. Spectra were acquired every 2 minutes.

Calculation of concentrations from infrared spectra

The method used to calculate the concentration profiles from the process spectra has been described in detail previously and is only summarized here (Schenk et al., 2007b). In brief, it consists of modeling any absorbance measured during the process as a linear combination of pure component spectra, the best combination being found through the least-squares approach. The first intensity spectrum, measured with the culture medium under process conditions, was used as reference intensity I_0 . Using equation 1, the absorbance difference from the initial state ΔA was calculated for every spectrum acquired during the process (at time t):

$$\Delta A_{\nu,t} = \sum_{i=1}^{i=j} \kappa_{i,\nu} \cdot \Delta C_i + \varepsilon_{\nu} \quad [2]$$

where $\kappa_{i,\nu}$ is the molar absorbance of a compound i at a wavenumber ν , and ΔC its concentration change with respect to the medium composition. ε is a residual due to experimental errors that is minimized by the least-squares algorithm. This equation is clearly over-determined since the total number of compounds j , is in all cases below 10, while the number of equations is close to 200, since a measurement is made every 4 cm^{-1} , from 1000 to 1500 cm^{-1} . Equation 2 can be rewritten in matrix notation, which gives a single equation (Equation 3) for the whole process:

$$\Delta \mathbf{A}_{s,wn} = \Delta \mathbf{C}_{s,j} \times \mathbf{K}_{j,wn} + \mathbf{E}_{s,wn} \quad [3]$$

The subscripts refer to the matrix dimensions (row,column), j , wn and s being the number of compounds, of wavenumber monitored and the number of spectra taken during the process respectively. \mathbf{K} is the spectra library, which is set-up prior to the experiments.

Concentration profiles are therefore found by resolving equation 3 to find ΔC , followed by the consecutive addition of the initial concentration, given by the medium composition. The size of ΔA was chosen to be $s \times wn$, instead of $wn \times s$ as in our previous work (Schenk et al., 2007b), in order to comply with the traditional chemometrics notation.

Analysis of the residuals can be used to detect the appearance of unexpected compounds, and even point to their identity. Residuals can be displayed in three dimensions, but projections in the time ($SSQt$) and wavelength ($SSQwn$) dimension can facilitate interpretation (Equation 4 – 5).

$$SSQt = row_sum(\mathbf{E} \cdot \mathbf{E}) \quad [4]$$

$$SSQwn = column_sum(\mathbf{E} \cdot \mathbf{E}) \quad [5]$$

where *row_sum* and *column_sum* are used to indicate a sum of the matrix elements by row and column respectively, and the operator “.” an element-by-element multiplication. Note that unlike \mathbf{E} , $SSQt$ and $SSQwn$ are not matrices but vectors and can be plotted in two dimensions. Since this method deals with concentration differences rather than absolute concentrations, the sign of residuals can also be used to distinguish between consumed and produced missing compounds, whereas sum of squared residuals can not. However, there is a risk that negative and positive residuals cancel each other during the process, and therefore, averaged residuals are preferentially used as a complementary analysis.

The sum of squared residues obtained for every time can be compared to the total variance found in the system, which gives the explained variance EV, expressed here as a percentage:

$$EV = 100 \cdot [1 - SSQt ./ row_sum(\Delta \mathbf{A} \cdot \Delta \mathbf{A})] \quad [6]$$

where the “./” operator refers to an element-by-element division. The averaged explained variance (AEV) is defined as the mean of the explained variance over the entire process. It does not make much sense to calculate an explained variance in the wavenumber dimension, since in some region of the spectra, typically where variations are very low, the solution found by the least-square algorithm can be larger than the observed change, since the minimization of the residual error is made over a wide wavenumber range.

The accuracy of the FTIR prediction calibration was assessed from the standard error of prediction (SEP), which was calculated using the following formula:

$$SEP = \sqrt{\frac{\sum_k (y_{k, Off-line} - y_{k, FTIR})^2}{n}} \quad [7]$$

where y_k indicates any off-line data or its corresponding FTIR value, and n refers to the total number of samples (typically 18-25).

4/ RESULTS AND DISCUSSION

Setting up the library

The library-based approach, as presented it in a previous article (Schenk et al., 2007b), addressed signal instability and the modeling issues separately. Signal drift was corrected using fixed points in the spectra (called anchor points), whereas the library (i.e. the calibration model) contained the molar absorbance of the main metabolites in aqueous solution, at the same temperature and pH as used during the process. Setting up this library could be done within a couple of hours, since it implied the preparation and the measurement of only five synthetic standards, instead of 50 for an equivalent result using the traditional chemometrics procedure.

The approach used for the current nine experiments was slightly different, since the library contained not only pure compound spectra, but also “drift spectra”, obtained by factor analysis. This incorporation allows for simultaneous measurement of metabolite concentrations and correction of signal drift. Compared to the other method, where drift was addressed separately by signal anchoring, the latter approach is more accurate. However, it requires an update if the drift features change, as would occur after a change in optics alignment.

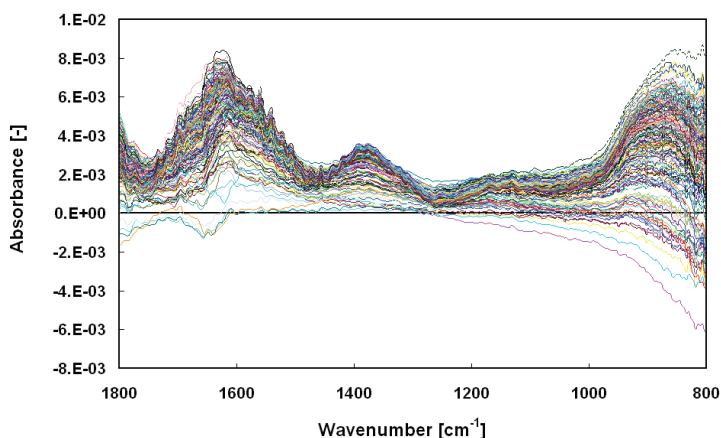


Figure 1. Absorbance difference (from initial state) of water, measured every 2 minutes for 8 hours under sterile conditions. The first spectrum was used as reference intensity, and therefore, a perfectly stable instrument should display no absorbance difference. The signal drift is wavenumber dependant and of large amplitude, since the change observed at 1000 cm^{-1} corresponds to the absorbance of a 4 g L^{-1} glucose solution.

Including both pure component and drift spectra into the library can be seen as an intermediate approach between Multiple Linear Regression (MLR) and Principal Component Regression (PCR). Whereas MLR is commonly used when the parameters influencing the response are known and pure standards are available, PCR is appropriate when dealing with unidentified variables, and/or when standards are unavailable (Brereton, 2003). A PCR model does not relate directly the measured spectra to concentration, but rather to virtual variables, called factors. Poorly defined mixtures that necessitate this approach occur frequently in the food industry (milk, wine, oils, etc.), but can also arise when working with chemically complex media. However, for a large number of bioprocesses, the main metabolites expected to show significant concentration changes in the infrared range are known, and synthetic standards can be prepared. Taking *E. coli* cultures as an example, the relevant compounds are the carbon source from the medium (glucose, glycerol, etc.), ammonium that serves as nitrogen source, and organic acids, the most abundant being acetic acid. For these compounds, their molar absorbance in water was simply used in the library.

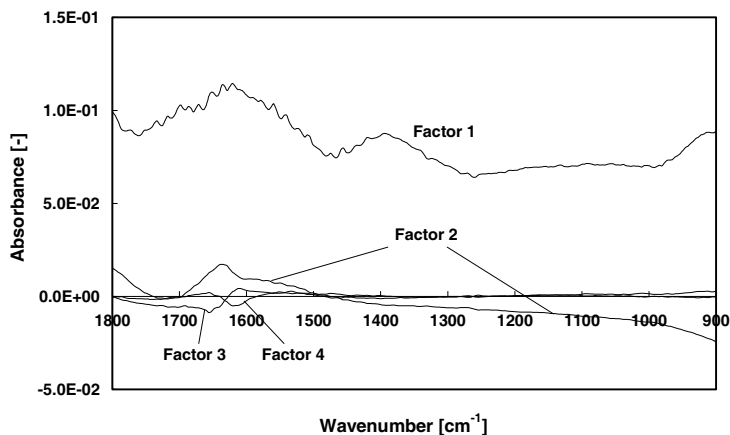


Figure 2. The four principal drift spectra found by singular value decomposition. They were multiplied by their respective singular value, in order to show the relative importance of one to another. Only the first two were included in the spectra library for on-line monitoring.

Since noise and drift are by nature undefined, they had to be included in the library as factors, determined by Principal Component Analysis (PCA). Signal drift was observed by measuring every 2 minutes, for 8 hours, the spectra of water placed under sterile conditions, in the flow-through cell thermostated at 37°C. The first intensity spectrum was used as reference (I_0) to calculate the absorbance spectra, using equation 1. While a perfectly stable

instrument should display no change in absorbance, a large drift occurred, which was, in addition, wavenumber dependent (Fig.1). To give an order of magnitude, the peak observed at 1000 cm^{-1} could have been produced by a 4 g L^{-1} glucose solution.

Singular value decomposition (SVD) was used to extract the key features from the signal drift. SVD is a generalized approach of eigenvalue decomposition that can deal with non-square matrices. Algorithms have been implemented in most mathematics software (Matlab, Mathcad, Maple, etc.), and more details on this method can be found elsewhere (Martens and Naes, 1992). Singular value decomposition can be, for the absorbance change shown in Figure 1, formulated as:

$$\Delta\mathbf{A} = \mathbf{U} \times \mathbf{S} \times \mathbf{V}^T \quad [8]$$

where \mathbf{U} is of size s by s , \mathbf{V}^T is the transpose of \mathbf{V} and is wn by wn . \mathbf{S} is diagonal and of the same size as $\Delta\mathbf{A}$, s by wn , namely the number of spectra acquired during the experiment by the number of wavenumbers at which absorbance is determined. \mathbf{V} contains the spectral features of the signal drift (the loadings, in chemometrics terminology), whereas \mathbf{S} and \mathbf{U} contain their respective weight and their evolution during the process. Plotting the first four rows of $\mathbf{S} \times \mathbf{V}^T$ allows visualizing the shape and importance of the first four factors found to represent the best the changes in absorbance (Fig. 2).

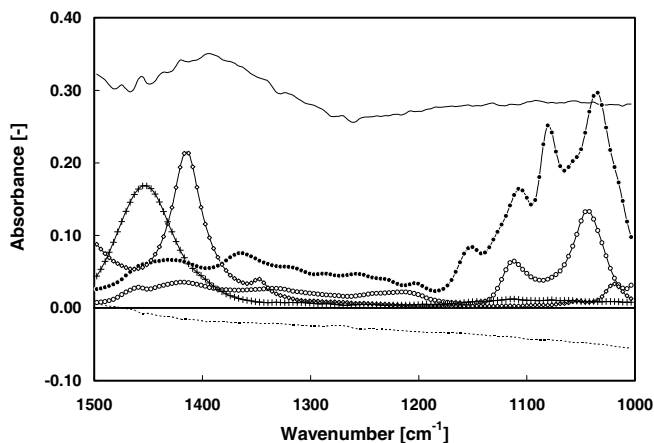


Figure 3. The six spectra used in the library, in the calculation range $1000 - 1500\text{ cm}^{-1}$. Molar absorbance of glucose (black circles), glycerol (open circles), acetate (open diamonds) and ammonium (crosses). The two drift spectra were amplified by a factor 4 for clarity reasons, and they are represented by a plain line (first drift spectrum), and a dashed line (second drift spectrum).

The first four singular values were 1.35, 0.19, 0.04 and 0.01. Only the drift spectra of the two first factors were included in the library. Incorporating more factors could potentially improve the prediction, but it would almost certainly undermine the model robustness by inducing overfitting. Practically, the size of the library \mathbf{K} (equation 3) was increased from $3 \times wn$ to $5 \times wn$, by adding the two first rows of $\mathbf{S} \times \mathbf{V}^T$ to the molar absorbance spectra of glucose, ammonium and acetate. Any multiple of the drift of these two rows could have been included in the library, since no drift amount or concentration can be reasonably defined. Solving equation 3 led, therefore, to molar concentrations for glucose, ammonium and acetate and arbitrary amount of the two first drift feature.

The six spectra that were used in this study were reported in figure 3. For clarity reasons, the two drift spectra were multiplied by a factor four. The molar absorbance was represented for the four real compounds, namely glucose, glycerol, ammonium and acetate. It can be seen that even the range used for calculation was not very broad (1000 – 1500 cm^{-1}), it contains all significant peaks.

Cultures on different carbon sources

Batch cultures of *E. coli* (strain ML 308) were carried out on glucose, glycerol and a mixture of the two, in order to address the reliability of the library-based method with respect to changes in carbon source. The library consisted, for the three experiments, of the molar absorbance spectra of ammonium and acetate, as well as the first two drift spectra. Depending on the medium composition, the pure component spectra of glucose, glycerol or both were also included. Concentration profiles were calculated using equation 5, by applying traditional least-squares regression. All standard errors of prediction are reported in Table 2.

The glucose, ammonium and acetate profiles found by FTIR agreed well with off-line data for both batch growth on glucose (Batch B01, Fig. 4) and glycerol (Batch B02, Fig. 5). The FTIR profiles clearly showed that these cultures were limited by the carbon source, and that only 60% of the nitrogen source was consumed. The biomass concentration, measured by optical density at 600 nm, reached the value of 14 (results not shown). Standard errors of prediction were 3.2 and 8.7 mM for glucose and glycerol respectively, whereas for NH_4^+ and acetate, they were around 8 mM. For all metabolites, with the exception of acetate, such SEP values correspond to less than 10% of the observed concentration change.

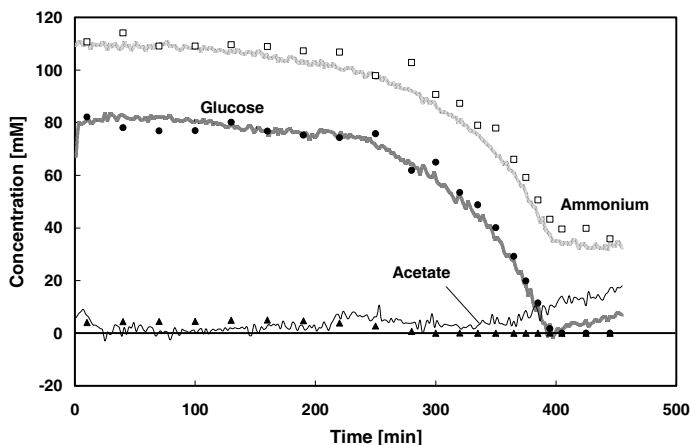


Figure 4. Glucose (black circles), ammonium (open squares) and acetate (black triangles) concentration profiles measured off-line (symbols) and on-line by FTIR (corresponding lines), during batch growth on glucose of the *E. coli* strain ML308 (batch B01).

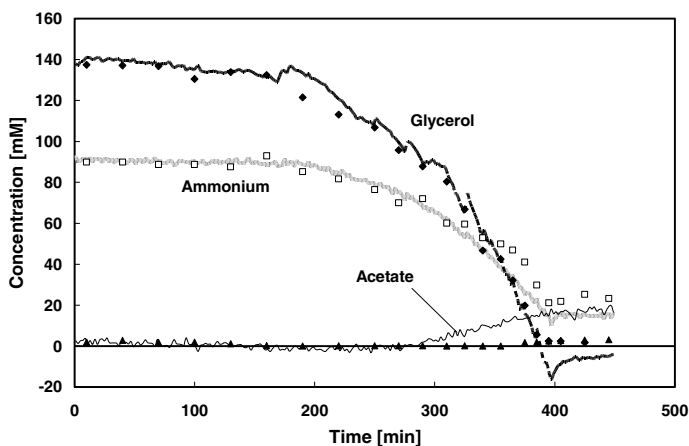


Figure 5. Glycerol (black diamonds), ammonium (open squares) and acetate (black triangles) concentration profiles measured off-line (symbols) and on-line by FTIR (corresponding lines), during batch growth on glycerol of the *E. coli* strain ML308 (batch B02).

Surprisingly, organic acids were not produced during either of the cultures. As indicated by off-line measurements, the small amount of acetate brought by the inoculum was consumed simultaneously with the main carbon source, and no acetate could be detected during the second half of the process. Enzymatic off-line analyses showed that neither lactate,

succinate nor formate was produced during the process. In contrast, significant amounts of acetate could be measured in the inoculum and during cultures on rich media, which suggests that a chemically defined medium limits the production of growth-associated products for the *E. coli* strain used. The FTIR model predicted for both cultures an increase of the acetate concentration at the end of the culture. This is certainly an artifact caused by the least-squares algorithm.

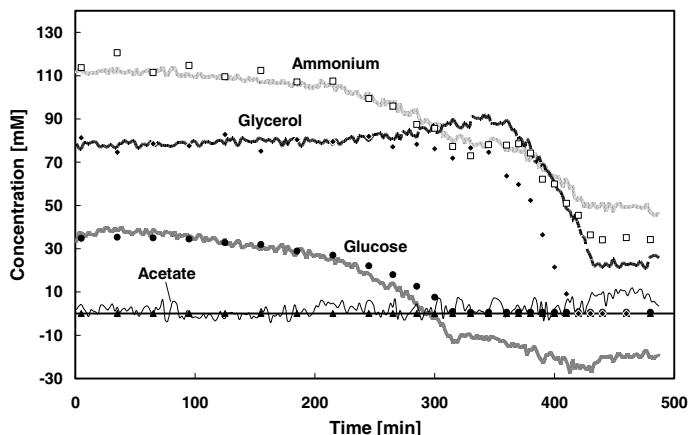


Figure 6. Glucose (black squares), glycerol (black diamonds), ammonium (open squares) and acetate (black triangles) concentration profiles measured off-line (symbols) and on-line by FTIR (corresponding lines), during batch growth on glucose and glycerol of the *E. coli* strain ML308 (batch B03).

The third batch experiment was performed with a medium that contained two carbon sources, namely glucose and glycerol, at the same concentration (in C-mol). From a qualitative point of view, the FTIR model and the off-line analyses displayed the same profile shapes (B03, Fig. 6). The two carbon sources were not consumed simultaneously. Glucose was used first, and after a short adaptation phase of about 30 minutes, glycerol was consumed rapidly. The concentration profiles found by infrared spectroscopy matched well the data found by the reference analysis during the glucose phase, but diverged significantly during the second phase. The SEP values for the culture were 13.3, 15.8, 5.9 and 3.4 mM for glucose, glycerol, ammonium and acetate respectively. The SEP values, concerning the carbon sources, were 4 and 2-times higher than for the runs B01 and B02. This much lower precision can largely be explained by the fact that the two carbon sources have very similar functional groups and therefore absorption peaks, in the range 1000 – 1150 cm^{-1} . While library-based

modeling can be considered to be an accurate tool if a single carbon source is present in the medium, it allows only qualitative analysis when two substrates, or products, present very similar molar absorbances. Such a situation can be expected with molecules that have very similar functional groups and structures, as for instance compounds within the carbohydrate family.

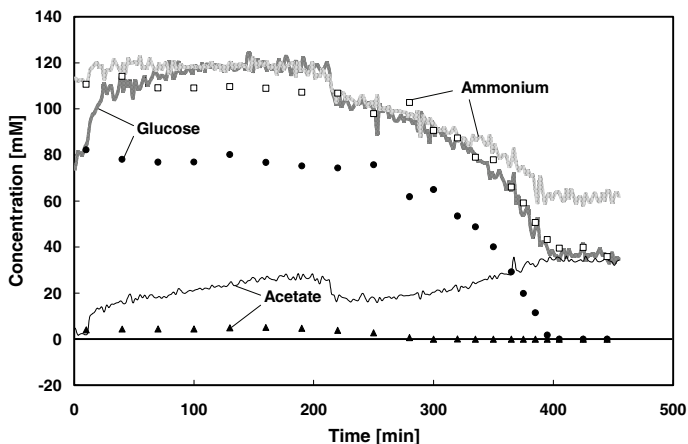


Figure 7. Glucose (black circles), ammonium (open squares) and acetate (black triangles) concentration profiles measured off-line (symbols) and on-line by FTIR (corresponding lines), during batch growth on glucose of the *E. coli* strain ML308 (batch B01), without correction of signal drift. The spectra library did not include the two drift spectra, and therefore only comprised the molar absorbance of glucose, ammonium and acetate.

The data of the batch culture on glucose (B01) were reprocessed without any drift correction, with a library consisting of only the molar absorbance of glucose, ammonium and acetate, in order to illustrate how critical drift correction is (Fig. 7). Compared to figure 3, for which calculations were performed with a library that included the two drift spectra, on-line FTIR concentration profiles exhibited sudden changes, after 15 and 200 minutes, and more generally they did not match the off-line, reference data. Standard errors of prediction were 34.5, 12.4, and 24.2 mM for glucose, ammonium and acetate respectively. The glucose prediction was 10 times less accurate, which demonstrates the importance of the method for signal drift correction.

Carbon versus nitrogen-limited cultures

A nitrogen-limited culture was monitored with exactly the same library as for the first batch run, in order to show that the library-based calibration is not correlated to the medium composition. The spectra library contained the same five spectra as for the first experiment (B01), namely glucose, ammonium, acetate, and the two drift contributions.

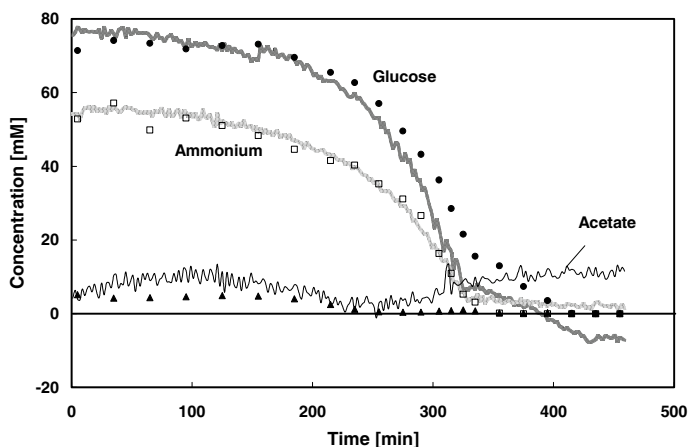


Figure 8. Glucose (black squares), glycerol (black diamonds), ammonium (open squares) and acetate (black triangles) concentration profiles measured off-line (symbols) and on-line by FTIR (corresponding lines), during batch growth on glucose, with nitrogen limitation, of the *E. coli* strain ML308 (batch B04).

The FTIR modeling correctly predicted the nitrogen limitation, which occurred at 320 minutes (Fig. 8). After this point, the predicted ammonium concentration remained constant, and very close to zero, while the glucose concentration began to decrease linearly for another 80 minutes, until depletion. The acetate concentration profile significantly diverged by the end of the run, probably for similar reasons to those discussed earlier. The standard errors of prediction were 6.9, 2.1 and 7.2 mM for glucose, ammonium and acetate respectively. These SEP values are very similar to what was obtained for the first two runs. The accuracy of the prediction proves that there is no correlation between the medium composition and the modeling, and therefore, a large number of different media can be tested with the same library.

Robustness of the calibration with respect to chemically undefined media

While the actual trend in medium formulation encourages the use of defined media, additives such as yeast extract or peptone are still commonly employed for their stimulating effect on growth. The ability of the library-based calibration to monitor concentration profiles in chemically undefined solutions was addressed with three different batches. One was performed on a typical rich medium (Batch B05), while two others on the same defined medium as used for the first batch (B01), supplemented with 5 g L⁻¹ (Batch B06) or 1 g L⁻¹ (Batch B07) of yeast extract. For all media, the carbon source was glucose, at an initial concentration of 15 g L⁻¹.

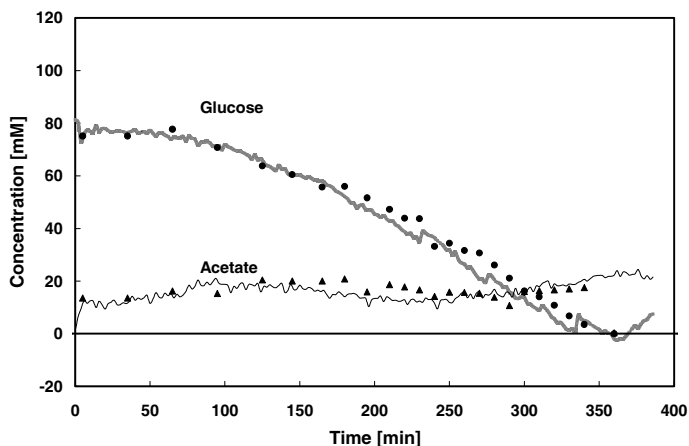


Figure 9. Glucose (black circles) and acetate (black triangles) concentration profiles measured off-line (symbols) and on-line by FTIR (corresponding lines), during batch growth on a rich medium of the *E. coli* strain ML308 (batch B05).

The rich medium contained, in addition to phosphate buffer, yeast extract, peptone and glucose. No defined nitrogen source was supplied, and therefore, the library comprised only the molar absorbance spectra of glucose and acetate, in addition to the two usual drift spectra. Concentration profiles found by FTIR were in excellent agreement with the off-line reference data (Fig. 9). In contrast to cultures conducted with chemically defined media, the amount of acetate brought by the addition of the inoculum was not subsequently consumed, and the acetate concentration remained almost constant.

The SEP values were 6.1 mM for glucose and 3.4 mM for acetate. The consumption of nitrogen from various sources found in peptone and yeast extract did not significantly induce perturbations in the infrared spectra, since the standard error of prediction of glucose and

acetate were similar to those obtained in the defined medium. This observation holds for the two cultures performed with a semi-defined medium (B06 and B07; SEP values are reported in Table 2), and it can therefore be stated that the library-based approach is robust enough to withstand the addition of chemically undefined compounds. This conclusion is not as surprising as it appears at first glance, since peptone and yeast extract are supplemented to chemically defined media to provide certain nutrients (e.g. growth factors, vitamins, etc.) that are consumed in very limited amounts, therefore inducing small changes in infrared spectra. In addition, since the calibration approach uses the culture medium as reference intensity, the absorbance of the species found in yeast extract and tryptone is included in the background intensity and therefore does not interfere with the calibration itself. Concerning the rich medium, which did not contain any defined ammonium source, exhaust gas analysis (data not shown) showed a high number of small adaptation phases. This indicates that the bacteria had to switch from one nitrogen source to another, which means that various compounds were consumed in small amounts, rather than only one in large quantities. It certainly resulted in negligible changes distributed over a broad range of wavenumber in the infrared spectra, which did not undermine the prediction.

Three different strains

The same defined medium was used to grow three different *E. coli* wild type strains, namely ML308, W3100 and W3110. The idea was to test whether it is possible to identify strains that grow quickly and produce small amounts of organic acids, as it could be done in a screening step during process development. The culture of the strain W3100 (Batch B08) showed the same characteristics as the ML308 run (Batch B01), in terms of SEP values (Table 2), but also in terms of acetate production. For the culture W3110, however, the SEP value for glucose was two-fold higher than for the other runs (except for the batch B03, on glucose and glycerol). Such a large discrepancy between the off-line data and the FTIR results could also be observed for ammonium, which presented a SEP value of 13 mM, which is three times higher than for the other runs.

The modeling certainly failed because a growth-associated compound was produced during the process. Off-line HPLC analyses showed a distinct peak, which increased in size during the culture, but no acetate, formate, succinate, lactate or ethanol could be detected by enzymatic analysis. The fingerprint of this unidentified compound could also be observed through residuals analysis. Plotting the sum of the squared residues (SSQ_{wn}, equation 5) for all nine experiments clearly shows that modeling of run B09, with the strain W3110, showed

much more variation in the process spectra (Fig. 10). A net peak at 1240 cm^{-1} could be observed for run B09, with an amplitude three times higher than all other residuals peaks found during the other cultures. This provides the fingerprint of the unexpected compound, although it was not possible to identify it, and it also indicates that the modeling of this culture was not accurate.

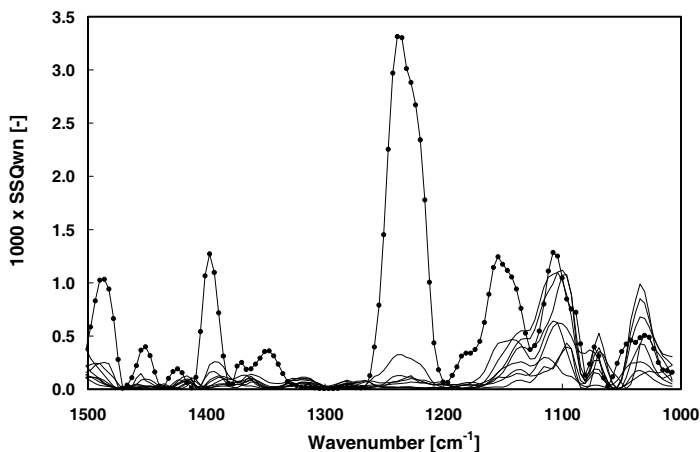


Figure 10. Sum of the squared residues in the wavenumber dimension (SSQwn) for culture B09 (strain W3110; line with symbols) and for the eight other batch cultures (line without symbols). The sharp peak at 1240 cm^{-1} points to the identity of an unidentified compound.

The drift profiles found during the runs can also be used to detect outlier runs that necessitate further investigation. A plot of the amount of drift attributed to the first and second drift factor for all nine experiments shows that batch B09 was subjected to an apparently larger signal drift (Fig. 11). The run that presented the second highest change in the factor 1 was batch B03, on glucose and glycerol. The SEP values for this experiment were also higher than the average. In presence of an unknown peak caused by a missing compound, the least squares algorithm tries to minimize the sum of squared residuals using all spectra from the library. While this gives erroneous predicted concentrations, it may also lead to underestimated, or more likely, to overestimated drift amounts. Drift concentrations can therefore be used, along with the inspection of SSQwn, as a tool to detect experiments that contain too much drift, or during which unexpected compounds appear. Repeating the culture can help to distinguish between these two possible causes. Batch B09 was repeated

three times, and for all replicates, the drift concentration was comparable, which confirms the hypothesis of the production of a growth-associated metabolite.

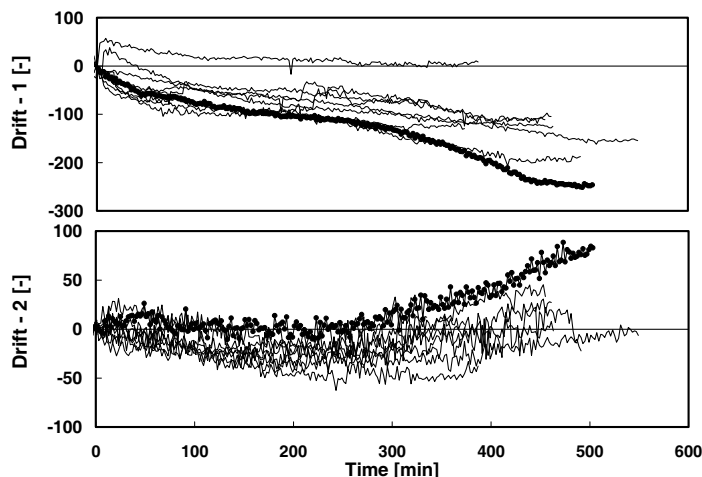


Figure 11. Amount of the first (top) and second (below) drift factors for the culture B09 (strain W3110; line with symbols) and for the eight other batch cultures (line without symbols). The amount of drift during the culture B09 was higher than usual, which indicates either a large signal drift or the appearance of unexpected compounds.

Overall diagnostics

Standard errors of prediction for all nine batch experiments are summarized in Table 2. The standard errors of prediction averaged over the nine experiments were 8.0, 12.3, 5.9 and 5.6 mM for glucose, glycerol, ammonium and acetate respectively. As a rule of thumb, an error of prediction lower than 10 mM, or 0.5 – 1.5 g L⁻¹, can be expected with this method. This value is very similar to the precision of traditional chemometrics approaches used in a truly predictive mode, i.e. without including in-process standards (Rhiel et al., 2002). For instance, SEP values within the range 0.22 – 1.54 g L⁻¹ were obtained for the on-line monitoring of *Gluconacetobacter xylinus* cultures, with a 56-standard calibration set and PLS modeling (Kornmann et al., 2004). Much lower standard errors of prediction have been reported, with values ranging from 0.01 to 0.5 g L⁻¹, but measurements were made at-line, which allows referencing the process spectra to a reference intensity and, therefore, removing most of the drift and noise (Kansiz et al., 2001; Roychoudhury et al., 2006a).

No clear relationship between the SEP values and the averaged explained variance (AEV) could be observed for the nine experiments. The averaged explained variance is, therefore,

not a relevant indicator for the detection of cultures during which unexpected compounds are produced, and more generally, for evaluating modeling quality. This point has been made previously, but more in terms of the number of spectra included in the library (Schenk et al., 2007b). The concentration of drift, therefore, seems to be the best criterion for an automated detection of culture that requires further investigation.

Table 2. Modeling statistics for the nine batch cultures of *E. coli*. The standard errors of prediction were calculated with more than 18 points. Growth rates were estimated by FTIR by regression over more than 100 points.

Batch No	Medium	C-Source	Strain	Standard error of prediction [mM]				Growth rate [h ⁻¹]		
				Glu	Gly	NH4	AcO	OD600	FTIR	AEV [%]
B01	CDM	Glu	ML308	3.2	N/A	5.3	7.0	0.75	0.77	96.1
B02	CDM	Gly	ML308	N/A	8.7	7.0	8.0	0.60	0.62	96.9
B03	CDM	Glu + Gly	ML308	13.3	15.8	5.9	3.4	0.80/0.66	0.73/1.91	99.1
B04	CDM	Glu (N-limited)	ML308	6.9	N/A	2.1	7.2	0.85	0.79	98.1
B05	BLB + 15 g L ⁻¹ Glu	Glu	ML308	6.1	N/A	N/A	3.4	0.99	1.07	97.1
B06	CDM + 5 g L ⁻¹ YE	Glu	ML308	8.8	N/A	4.3	5.8	0.79	0.77	95.8
B07	CDM + 1 g L ⁻¹ YE	Glu	ML308	6.2	N/A	4.6	6.3	0.78	0.76	97
B08	CDM	Glu	W3100	7.9	N/A	4.6	5.3	0.47	0.41	96.6
B09	CDM	Glu	W3110	12.0	N/A	13.4	4.0	0.61	0.41	96.1

Medias: CDM = Chemically defined medium; BLB = Buffered Luria-Bertani; YE stands for Yeast extract.

C-Sources: Glu = Glucose; Gly = Glycerol.

AEV: Averaged Explained Variance.

*) For the culture B03, the two growth rates displayed are for the glucose and glycerol phases respectively (Glu/Gly).

In order to test whether FTIR monitoring can be used to select for fast-growing strains, the specific growth rate, μ , was calculated for each experiment. For that purpose, the natural logarithm of the carbon source concentration was plotted against time. The specific growth rate was calculated by linear regression over the linear range, which corresponds to the exponential growth phase. The computed values for the nine experiments are reported in Table 2. Except for runs B03 and B09, which showed much higher SEP values, the growth rate predicted by FTIR was within $\pm 8\%$ of the value determined by off-line measurements. On average, the error amounted to 3.9%. With only the growth rate calculated from the FTIR profiles, it could be already possible to observe that growth was faster on glucose than on glycerol (B01 vs. B02), and that μ was 30% higher on a rich medium, whereas the addition of up to 5 g L⁻¹ of yeast extract did not significantly speed up growth (B01 vs B05 – B07). In addition, the low specific growth rate of strain W3100 could be clearly identified (B01 vs. B09). From the standpoint of medium and strain screening, this result shows that off-line sampling becomes useless in most cases, and remains necessary only for some cultures during which unexpected compounds are produced. By balancing the simplicity of the

method and the amount of valuable information that can be obtained for a large variety of medium compositions and strains, it turns out that this use of mid-infrared spectroscopy features a very competitive ratio of work load to useful data.

5/ CONCLUSIONS

Traditional least squares modeling, using for calibration set a single library consisting of six spectra, proved to be as accurate and predictive as the conventional process chemometrics approach, for the on-line monitoring of several batch cultures of *Escherichia coli*. Such a result represents a significant improvement in terms of time invested in calibrating the instrument, since only six infrared spectra had to be measured, instead of more than 40 for each experiment. This library-based approach does not involve the use of Principal Component Regression (PCR), and is therefore readily accessible for non-experts in process modeling. In addition, dealing with pure component spectra allows much easier data interpretation, as for instance the detection and identification of unexpected compounds.

While an optical conduit was used in this study to connect the FTIR instrument to the *in-situ* sensor, it would be highly interesting to investigate how the proposed calibration approach could be used with several bioreactors connected in parallel by fiber optics probes, which allow multiplexing. The performance of the calibration approach, in comparison with the single-vessel application, would have to be addressed (Roychoudhury et al., 2007), as well as the practical use of fiber optics probes, which are known to cover a narrower range of the mid-IR domain and have poorer transmission characteristics (Roychoudhury et al., 2006b). The advances in that field, coupled with the library-based calibration approach that could easily be automated, will certainly make the implementation of FTIR into high-throughput platforms for medium and strain screening feasible.

6/ REFERENCES

- ASTM. 1997. Standard practices for infrared, multivariate, quantitative analysis (E 1655-97). Annual book of ASTM standards. Philadelphia: American Society for Testing and Materials. p 844-869.
- Betts JJ, Baganz F. 2006. Miniature bioreactors: current practices and future opportunities. *Microb. Cell. Fact.* 5:21.
- Brereton RG. 1997. Multilevel multifactor designs for multivariate calibration. *The Analyst* 122:1521-1529.
- Brereton RG. 2000. Introduction to multivariate calibration in analytical chemistry. *The Analyst* 125:2125-2154.
- Brereton RG. 2003. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. New York: Wiley & Sons.
- Clementschtitsch F, Bayer K. 2006. Improvement of bioprocess monitoring: development of novel concepts. *Microb. Cell. Fact.* 5:19.
- Crowley J, McCarthy B, Nunn NS, Harvey LM, McNeil B. 2000. Monitoring a recombinant *Pichia pastoris* fed batch process using Fourier transform mid-infrared spectroscopy (FT-MIRS). *Biotechnol. Lett.* 22:1907-1912.
- Doak DL, Phillips JA. 1999. In situ monitoring of an *Escherichia coli* fermentation using a diamond composition ATR probe and mid-infrared spectroscopy. *Biotechnol. Prog.* 15:529-539.
- Harms P, Kostov Y, French JA, Soliman M, Anjanappa M, Ram A, Rao G. 2005. Design and performance of a 24-station high-throughput microbioreactor. *Biotechnol. Bioeng.* 93(1):6-13.
- Harms P, Kostov Y, Rao G. 2002. Bioprocess monitoring. *Curr. Opini. Biotech.* 13(2):124-127.
- Kansiz M, Gapes JR, McNaughton D, Lendl B, Schuster KC. 2001. Mid-infrared spectroscopy coupled to sequential injection analysis for the on-line monitoring of the acetone-butanol fermentation process. *Anal. Chim. Acta* 438(1-2):175-186.
- Kornmann H, Rhiel M, Cannizzaro C, Marison I, von stockar U. 2003. Methodology for real-time, multi-analyte monitoring of fermentations using an in-situ mid-infrared sensor. *Biotechnol. Bioeng.* 82(6):702-709.
- Kornmann H, Valentinotti S, Duboc P, Marison IW, von Stockar U. 2004. Monitoring and control of *Gluconacetobacter xylinus* fed-batch cultures using in situ mid-IR spectroscopy. *J. Biotechnol.* 113(1-3):231-245.

- Martens H, Naes T. 1992. Multivariate calibration. New York: John Wiley & Sons.
- Pollard DJ, Buccino R, Connors NC, Kirschner TF, Olewinski RC, Saini K, Salmon PM. 2001. Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy. *Bioprocess. Biosyst. Eng.* 24:13-24.
- Rhiel MH, Amrhein MI, Marison IW, von Stockar U. 2002. The influence of correlated calibration samples on the prediction performance of multivariate models based on mid-Infrared spectra of animal cell cultures. *Anal. Chem.* 74(20):5227-5236.
- Roychoudhury P, Harvey LM, McNeil B. 2006a. At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total reflectance mid-infrared (ATR-MIR) spectroscopy. *Anal. Chim. Acta* 561:218-224.
- Roychoudhury P, Harvey LM, McNeil B. 2006b. The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. *Anal. Chim. Acta* 571:159-166.
- Roychoudhury P, O'Kennedy R, McNeil B, Harvey LM. 2007. Multiplexing fibre optic near infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses. *Anal. Chim. Acta* 590(1):110-117.
- Schenk J, Marison IW, von Stockar U. 2007a. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2):344-353.
- Schenk J, Marison IW, von Stockar U. 2007b. Simplified Fourier-transform mid-infrared calibration based on a spectra library for the on-line monitoring of bioprocesses. *Anal. Chim. Acta* 591(1):132-140.
- Vojinovic V, Cabral JMS, Fonseca LP. 2006. Real-time bioprocess monitoring. Part I: In situ sensors. *Sensor. Actuat. B* 114:1083-1091.

7/ LIST OF SYMBOLS

Matrices and vectors

A	$s \times wn$	Absorbance matrix: s absorbances measured during the process, at wn different wavenumbers
C	$s \times j$	Matrix of concentrations: the concentrations of the j species calculated from the s spectra measured during the process
E	$s \times wn$	Error matrix: the residues found at wn wavenumbers, for the s spectra measured during the process
K	$j \times wn$	Spectra library: the molar absorbance of j species at wn different wavenumbers
U	$s \times s$	Matrix of Singular Value Decomposition that contains the weight of each singular value in each sample
S	$s \times wn$	Singular Value Matrix
V	$wn \times wn$	Loadings matrix (spectral features)
SSQ_{wn}	$1 \times wn$	Sum of square residues in wavenumber dimension (vector)
SSQ_t	$s \times 1$	Sum of square residues in time dimension (vector)

Greek symbols

ϵ	Experimental error on absorbance [-]
κ	Molar absorptivity (or molar extinction coefficient) [M ⁻¹]
ν	Wavenumber [cm ⁻¹]

Other symbols

A	Absorbance [-]
C	Concentration [g L ⁻¹], or [mol L ⁻¹]
EV	Explained variance [%]
I	Spectrum intensity [-]
j	Number of species included in the library ($j \leq m$)
m	Number of different species involved in the process
SEP	Standard error of prediction [-]
y	Predicted or measured concentration [g L ⁻¹], or [mol L ⁻¹]

Superscripts and subscripts

<i>baseline</i>	subscript	Baseline (refers to the initial culture medium)
k	subscript	k-th sample
i	subscript	i-th species
n	subscript	Number of off-line samples
s	subscript	Number of spectra measured during the process
<i>tot</i>	subscript	Total
wn	subscript	Number of wavenumber measured

Chapter 5

Mid-Infrared Spectroscopy as a Tool for pH Prediction and Control in Bioprocesses

1/ ABSTRACT

An online pH monitoring method based on mid-infrared spectroscopy relevant to bioprocesses is presented. This approach is non-invasive and does not require the addition of indicators or dyes, since it relies on the analysis of species of common buffers used in culture media, such as phosphate buffer. Starting with titrations of phosphoric and acetic solutions over almost the entire pH range (2 – 12), it was shown that the infrared spectra of all samples can be expressed as a linear combination of the molar absorbance of the acids and their deprotonated forms. In other words, pH had no direct influence on the molar infrared spectra themselves, but only on deprotonation equilibria. Accurate prediction (standard error of prediction for pH < 0.15 pH units) was achieved by taking into account the non-ideal behavior of the solutions, using the Debye-Hückel theory to estimate the activity coefficients. Batch cultures of *E. coli* were chosen as a case study to show how this approach can be applied to bioprocess monitoring. The discrepancy between the spectroscopic prediction and the conventional electrochemical probe never exceeded 0.12 pH units, and the technique was fast enough to implement a feedback controller to maintain the pH constant during cultivation.

This chapter was submitted for publication in *Biotechnology and Bioengineering* as:

Schenk J, Marison IW, von Stockar U. pH prediction and control in bioprocesses using mid-infrared spectroscopy.

2/ INTRODUCTION

Mid-infrared spectroscopy is an interesting technique for the monitoring of bioprocesses, because it relates to the vibrational energy of molecules, and can therefore be used to measure the concentration of almost all species in solution (e.g. glucose, ethanol, ammonium, acetate, glycerol, etc.). This paper aims at presenting a relatively unknown aspect of this technique, which is the additional ability to monitor pH through the determination of the relative concentration of any acid to its conjugate base.

The determination of pH based on spectroscopic measurements, which is sometimes referred to as a “pH optode”, is more complex than the use of a conventional electrochemical probe and is therefore attractive only in specific applications, such as when space is limiting. This approach is generally based on UV-VIS absorbance or fluorescence spectroscopy. In such cases, it relies on the measurement of a colored indicator or a fluorescent dye. Among the acid-base colored indicators, phenol red has been the most frequently used, due to its low cytotoxicity. Reported applications include the on-line monitoring of tissue cultures (Xu et al., 2006) and the pH control of a perfused bioreactor (Jeevarajan et al., 2002). Concerning fluorescence spectroscopy, several different dyes, such as carboxy-fluorescein (John et al., 2003), 8-hydroxy-1,3,6-pyrene trisulfonic acid trisodium salt (HPTS) (Kermis et al., 2002) and cyroxynaphtofluorescein (Song et al., 1997) have been used for pH monitoring in culture media.

Since Fourier-transform mid-infrared spectroscopy (FTIR) can potentially be used to measure all compounds in solution, pH prediction based on this technique does not necessitate the addition of a potentially toxic dye and can rely on any acid and its conjugate basis, such as the phosphate buffer. In addition, it features several interesting characteristics for bioprocess applications: it is fast, non-invasive and attenuated total reflectance (ATR) probes can be sterilized *in-situ*. Examples of on-line monitoring of metabolite concentrations have been reported for *E.coli*, fungi and yeast cultures at various scales (Crowley et al., 2000; Doak and Phillips, 1999; Kornmann et al., 2003; Pollard et al., 2001) as well as for mammalian cell cultures (Rhiel et al., 2002; Riley et al., 1998). However, to our knowledge, no utilization of FTIR spectroscopy for the direct measurement of pH has been reported to date. Instead studies have been focused on the detection of partial and total alkalinity (Steyer et al., 2002), minor phosphate complexes (Baril et al., 2000), or on food control through multivariate analysis (Karoui et al., 2006; Lobo et al., 2006) rather than pH monitoring based on modeling of the deprotonation equilibria.

This paper aims at discussing the influence of pH on mid-infrared spectra, using weak acids titrations. It also shows how this technique can be applied to the measurements of pH measurement in common buffer solutions and during microorganism cultivation. Experiments were carried out in a 2-litre bioreactor, in order to establish a proof of principle for future applications in miniaturized systems. While most of the studies mentioned above circumvent the problems caused by the non-ideality of the solutions by calibrating the dye response in the culture medium, this study aims at addressing these thermodynamic issues from a theoretical angle, using Debye-Hückel theory to estimate activity coefficients.

3/ THEORY

Infrared spectroscopy modeling

The recommended calibration procedure for FTIR instruments involves the measurement of a number of standards greater than 6 times the number of compounds to be monitored (ASTM, 1997), followed by subsequent multivariate modeling, using techniques such as Partial-Least Squares (PLS) or Principal Component Analysis (PCA) (Martens and Naes, 1992). However, it has been shown that this calibration procedure can be considerably simplified without significantly decreasing the accuracy of the prediction by the use of a spectra library (Schenk et al., 2007b), and this latter approach was followed in the current study. Briefly, the method consists of modeling a process spectrum as a linear combination of pure component absorbances, where the pure component is defined to be in aqueous solution. The best combination is found using the least-squares algorithm, and residual analysis can be used to evaluate the quality of modeling. As opposed to more sophisticated chemometric methods such as PLS, this approach uses only one variable (or factor) per component for calculations, which is its pure component spectrum. The main advantage of this approach therefore lies in the fact that the number of standards required to calibrate the instrument can be reduced to the number of compounds to be measured, which corresponds to a 6-fold decrease compared to what is recommended for chemometric modeling.

The key equations of this library-based approach are summarized here. The absorbance of a sample A at a wavenumber ν is given by:

$$A_{\nu} = \log_{10} \left(\frac{I_{\nu,o}}{I_{\nu}} \right) \quad [1]$$

where I_{ν} and I_{ν^o} stands for the sample and reference intensity respectively. The water signal is taken as reference intensity when working off-line (i.e. in the present case to set up the library and for off-line titrations), whereas during on-line operation, the first spectrum of the process is set as reference. In the latter case, it includes the absorbance of all the compounds at their initial concentration and can be regarded as a baseline. The absorbance measured later during the process is therefore expressed as an absorbance difference ΔA , since it is produced by a concentration change ΔC . Assuming that the infrared activity of all species in solution add linearly, the absorbance difference of an experimental measurement can be expressed as:

$$\Delta A_{\nu,t} = \sum_{i=1}^{i=j} \kappa_{i,\nu} \cdot \Delta C_i + \varepsilon_{\nu} \quad [2]$$

where κ_i refers to the molar absorbance of the compound i , and is given by the spectra library. ϵ stands for the experimental error and is minimized by the least-squares algorithm. This equation can also be used for off-line experiments presented below, when expressed in terms of absolute absorbance and concentration.

Thermodynamic Modeling

The first deprotonation of a diprotic acid can be expressed as:



The thermodynamic equilibrium constant K_{a1} for this reaction states:

$$K_{a1} = \frac{a_{\text{H}^+} \cdot a_{\text{HA}^-}}{a_{\text{H}_2\text{A}}} = \frac{\gamma_{\text{H}^+} \cdot \gamma_{\text{HA}^-}}{\gamma_{\text{H}_2\text{A}}} \cdot \frac{C_{\text{H}^+} \cdot C_{\text{HA}^-}}{C_{\text{H}_2\text{A}}} \quad [4]$$

where γ_i and C_i are the activity coefficient and the concentration of a species i respectively.

Using equation 4, the pH can be expressed as:

$$\text{pH} = -\log_{10}(a_{\text{H}^+}) = \text{p}K_a + \log_{10}\left(\frac{\gamma_{\text{HA}^-}}{\gamma_{\text{H}_2\text{A}}}\right) + \log_{10}\left(\frac{C_{\text{HA}^-}}{C_{\text{H}_2\text{A}}}\right) \quad [5]$$

The measurement of the concentrations of the acid H_2A and its conjugate anion HA^- by mid-infrared spectroscopy can therefore lead to the pH of the solution, provided that the other terms in equation 5 can be estimated. Values for the thermodynamic equilibrium constant $\text{p}K_{a1}$ at infinite dilution are tabulated in the literature for most acids. The determination of the activity coefficients γ_{AH} and γ_{A} is more difficult, since their direct measurement is far from being straightforward and therefore not readily compatible with bioprocess applications. Fortunately, concerning common compounds, there is a great deal of data and models available to estimate activity coefficients in non-ideal solution.

An empirical extension of the Debye-Hückel theory that accounts for short range interactions is often used to predict activity coefficients of ionic and non-ionic compounds in solution for ionic strength up to 1 mol kg^{-1} (Zemaitis et al., 1986)

$$\log_{10}(\gamma) = \frac{-\alpha \cdot z^2 \cdot \sqrt{I_c}}{1 + \beta \cdot \sqrt{I_c}} + b \cdot I_c \quad [6]$$

where z is the ion charge and I_c the ionic strength. α is a function of the temperature only, and has the value of 0.5224 $\text{kg}^{1/2} \text{mol}^{-1/2}$ at 37°C. β depends on the ion radius but as a first approximation can be set to 1.5 $\text{kg}^{1/2} \text{mol}^{-1/2}$ for all temperatures and all compositions, which is generally referred to as the “Bates-Guggenheim convention” (Siggaard-Andersen et al., 1984). b , which is sometimes called the “salting-out” parameter, is an empirical constant that is ion-

dependant. Values for b are tabulated for numerous compounds (Meier, 1982), but a default value of $0.055 \text{ kg mol}^{-1}$ can be used when no data can be found (Zemaitis et al., 1986).

For neutral species, equation 6 becomes:

$$\log_{10}(\gamma) = b \cdot I_c \quad [7]$$

This latter equation has no real physical meaning but has useful practical application at not excessively high concentrations. More elaborated relationships have been proposed to account for interactions between neutral species and ions (Zemaitis et al., 1986), but they lead to very small corrections, as long as the concentrations remain sufficiently low (Pitzer and Silvester, 1976). Indeed, some authors simply assume a constant activity coefficient of 1 for neutral species, since according to equation 7, $\gamma = 1.065$ at an ionic strength of 0.5 mol kg^{-1} .

The equations 3 to 5 can be easily extended to the second dissociation of a diprotic acid H_2A with an equilibrium given by:



The equation for pH prediction then becomes

$$\text{pH} = -\log_{10}(a_{\text{H}^+}) = \text{p}K_{a2} + \log_{10}\left(\frac{\gamma_{\text{A}^{2-}}}{\gamma_{\text{HA}^-}}\right) + \log_{10}\left(\frac{C_{\text{A}^{2-}}}{C_{\text{HA}^-}}\right) \quad [9]$$

where both activity coefficients have to be computed for an ionic species (Equation 6).

In order to compare computed values to tabulated data, it is convenient to define apparent equilibrium constants $\text{p}K_{a1}'$ and $\text{p}K_{a2}'$ which depend on the composition of the solution:

$$\text{p}K_{a1}' = \text{p}K_{a1} + \log_{10}\left(\frac{\gamma_{\text{HA}^-}}{\gamma_{\text{H}_2\text{A}}}\right) \quad [10]$$

$$\text{p}K_{a2}' = \text{p}K_{a2} + \log_{10}\left(\frac{\gamma_{\text{A}^{2-}}}{\gamma_{\text{HA}^-}}\right) \quad [11]$$

It must be emphasized that the apparent equilibrium constant was defined in some publications to also include the activity coefficient of H^+ . Comparison is, therefore, often not possible between these two definitions, since γ_{H^+} cannot be easily computed with the necessary accuracy.

By examining equation 5 and 9, it can be observed that the pH prediction becomes highly sensitive to errors in the measurement of concentrations at pH far from pKa, or in other word, when the fraction of the acid that is deprotonated is close to zero or one. This can be illustrated by rearranging equation 5 for a monoprotic acid HA, and by using the definition of $\text{p}K_{a1}'$:

$$pH - pK'_a = \log_{10} \left(\frac{C_{A^-}}{C_{HA}} \right) \quad [12]$$

A graphical representation of equation 12 shows that if the fraction of the total acid concentration that is deprotonated is above 90% or below 10%, minor errors in the estimation of this fraction leads to large changes in the pH prediction (Fig. 1). In order to insure a certain robustness to the pH determination, measured fraction that are outside this range should be discarded. A reliability criterion can therefore be formulated as:

$$0.1 < \frac{C_{A^-}}{C_{A^-} + C_{HA}} < 0.9 \quad [13]$$

While this criterion prevents gross errors of prediction, it also limits the range of prediction for pH values that are within ± 1 pH units around the pKa.

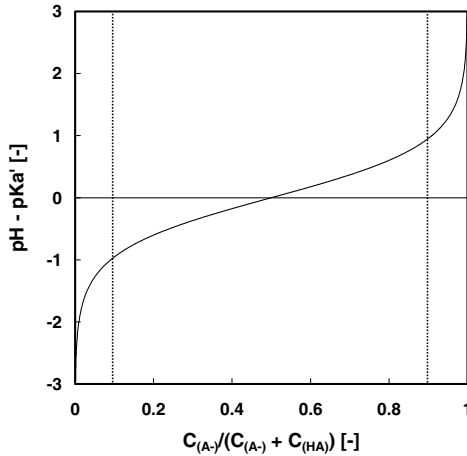


Figure 1. Variation of the pH from pKa as a function of the deprotonated fraction of the total acid concentration. Above 90% or below 10% (dashed lines), small error in the estimation of this fraction leads to large pH variations.

4/ MATERIALS AND METHODS

FTIR Set-up

A Mettler Toledo FTIR 4000 (Mettler-Toledo, Greifensee, Switzerland) with an attenuated total reflectance probe (ATR, DiComp) was mounted in a flow-through cell thermostated at 37°C, which allowed convenient measurement of standards in an off-line configuration, but also connection to a bioreactor for on-line operation. This assembly is very similar to an *in-situ* ATR-probe, since it allows cleaning and sterilization in place. More details on the instrument and the set-up can be found elsewhere (Doak and Phillips, 1999; Schenk et al., 2007a). For all off-line measurements, the spectra of the solutions of interest were measured in parallel to water, which was used to determine the reference intensity for absorbance calculation. To produce titration curves, volumes ranging from 0 to 1.7 ml of NaOH 10 N were added to 50 ml aliquots of the acid solutions (approximately 0.1 M) until complete titration ($\text{pH} > 12$). The titration was completed in 13, 22 and 24 steps for acetic acid, phosphoric acid and the mixture of the two respectively. Each standard used to build the library contained only one compound, at a concentration of about 100 mM. For on-line operation, culture medium was continuously recirculated to the flow-through cell using a membrane pump (Prominent Gamma/L, Heidelberg, Germany). The average residence time in the circuit was less than 20 seconds and absorbance spectra were acquired every 2 minutes (except for the control experiment, for which the frequency was increased to 1 min⁻¹).

Calculations were made over the wavenumber range 950 – 1700 cm⁻¹, with a resolution of 4 cm⁻¹. This range was broader than for previous studies, in order to include significant H₂PO₄⁻ and HPO₄²⁻ infrared characteristic peaks. The absorbance was, therefore, determined at more than 250 different frequencies, which was largely redundant for the 2 to 6 compounds to be measured simultaneously. During on-line experiments, two anchors were set at 950 and 1750 cm⁻¹ in order to correct for drifts and noise. Such a correction method, based on fixed points in the infrared spectrum where no expected compound absorb, is absolutely necessary to get a high precision and has been discussed in details in a previous publication (Schenk et al., 2007a). The position of these anchor points had to be slightly adapted for the current work to avoid overlapping with the absorption of the buffer phosphate species.

The accuracy of the library-based calibration was assessed with the standard error of prediction (SEP), which can be calculated by the following formula:

$$SEP = \sqrt{\frac{\left(\sum_k y_{k,Off-line} - y_{k,FTIR} \right)^2}{n}} \quad [14]$$

The confidence interval on the pH prediction was calculated by multiplying the SEP by the value of Student's *t*-distribution at a degree of freedom *n*-1 and for a confidence level of 95%

Bioreactor Fermentation

The wild-type strain of *Escherichia coli* ML308 (DSMZ No 1329, Braunschweig, Germany) was used in the experiments. The inoculum was prepared by growing the cells in a shake flask in 100 ml of the same medium used for bioreactor for 14 hours. The culture medium contained: glucose 15 g L⁻¹, KH₂PO₄ 13.3 g L⁻¹, (NH₄)₂HPO₄ 5 g L⁻¹, NH₄Cl 2 g L⁻¹, MgSO₄·7H₂O 1.2 g L⁻¹, EDTA·Na₂ 10.7 mg L⁻¹ and 10 mL L⁻¹ of trace elements solution (composition per litre: H₃BO₃ 0.3 g, CuCl₂·4H₂O 0.15 g, MnSO₄·H₂O 1.35 g, Na₂MoO₄·2H₂O 0.25 g, Zn(CH₃CHOO)₂·2H₂O 1.3 g, CoCl₂·6H₂O 0.25 g, FeCl₃·6H₂O 11 g). For the on-line titration experiment, the medium was prepared without magnesium sulphate to avoid precipitation at alkaline pH. The pH was adjusted to 6.5 using NaOH 5 M. After filter-sterilization (0.2 µm Steritop®, Millipore, MA), 1 mL L⁻¹ of vitamine solution (Thiamine.HCl 4.5 g/l) and 0.6 mL L⁻¹ of antifoam (Struktol SB2121, Schill and Seilacher, Hamburg, Germany) were added.

Fermentations were carried out at 37°C in a 3.6-L bioreactor (Bioengineering, Wald, Switzerland) with a working volume of 2 litres. Agitation was set at 1000 rpm, with three 6-blade Rushton impellers, and total aeration was constant at 3 NI/min, a value sufficient to maintain the dissolved oxygen above 20% of air saturation. Oxygen and carbon dioxide composition of the exhaust gas was measured using a gas analyzer (Dr. Marino Müller Systems, Esslingen, Switzerland). For pH control, calculations were performed on-line in Matlab (The Mathworks, Natick, MA), and the predicted pH values were introduced in a feedback controller that drove the addition of NaOH 2M.

5 ml culture samples were collected at intervals of 10 to 30 minutes (depending on the batch phase) and immediately cooled to 2°C. Sample treatment was carried out within 2 hours. Biomass concentration was determined by measuring optical density at 600 nm (OD₆₀₀). The remaining volume of the sample was filtered through 0.2 µm filters for off-line analysis of metabolites. Glucose, ammonium ion and acetic acid concentrations were determined by enzymatic assay, using commercially available kits (R-Biopharm AG, Darmstadt, Germany) and an automated analyzer (Cobas Mira, Roche, Basel, Switzerland).

5/ RESULTS AND DISCUSSION

Titration of acetic acid, phosphoric acid and a mixture of the two

A 50 ml solution of acetic acid ca. 0.1 M was titrated by the addition of NaOH 10 M in 13 steps (final volume: 50.6 ml). Spectra were taken off-line at 37°C, taking water as reference intensity for absorbance calculations (Fig. 2). The absorbance of the initial solution was attributed to non-dissociated acetic acid, since at pH = 2.8 more than 99% of the acid is not deprotonated, whereas the spectrum of the titrated solution (pH = 11.9) was attributed to the acetate anion.

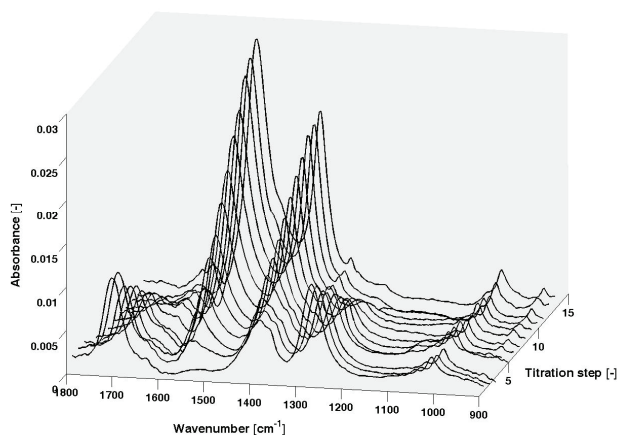


Figure 2. Infrared absorbance spectra during the titration in 13 steps of an acetic acid solution by concentrated NaOH. The first spectrum was attributed to the acetic acid, since at a pH of 2.8, 99% of the acid is not dissociated. The last spectrum, at pH 11.9, was attributed to the acetate anion. Intermediate spectra can be expressed as a linear combination of the first and last spectrum.

The non-dissociated form of acetic acid shows two peaks around 1710 and 1280 cm^{-1} , whereas the acetate anion absorbs around 1560 and 1420 cm^{-1} . It can be clearly observed qualitatively that the other solutions display absorbance spectra that are a combination of these two extremes and that no other complex is likely to be formed. The titration by NaOH brought a significant amount of sodium cations into the solution, but this did not perturb the measurements, since the infrared absorbance of atomic ions such as Na^+ or Cl^- is very low (Schenk et al., 2007b).

The concentration of acetic acid and acetate in the other 11 samples were calculated by the least-squares method, using only the first and the last spectrum as pure component spectra in the library. The pKa of acetic acid at infinite dilution was calculated to be 4.75 at 37°C using published thermodynamic data (Goldberg et al., 2002). The activity coefficients of acetate was calculated from equation 6, with the standard values for α and β (namely 0.5224 kg^{1/2} mol^{-1/2} and 1.5 kg^{1/2} mol^{-1/2} respectively), and using 0.9 kg mol⁻¹ for b , which is an averaged value of published data on sodium and potassium acetate (Meier, 1982). For the non-dissociated acid, the default value of 0.055 kg mol⁻¹ was used for b to calculate the activity coefficient. With respect to the charge balance, the ionic strength was assumed to be equal to the concentration of acetate found by IR spectroscopy, since the counter-ion (Na⁺ from NaOH) carried a single charge. The assumption was made for all calculations that the molality and molarity scales were equal, since the concentrations studied were relatively low (ionic strength < 0.4 mol L⁻¹). In order to insure a certain confidency in the pH prediction, the reliability criterion (equation 13) was applied to filter the results. Predicted pH values that passed this test were plotted against the values measured using a conventional electrochemical pH probe (Fig.3).

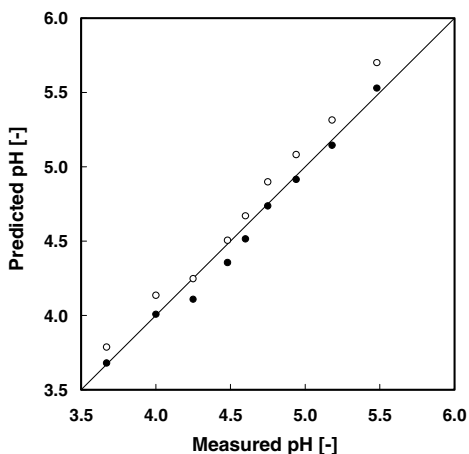


Figure 3. Predicted against measured pH for the titration of acetic acid. Calculations were performed using activity coefficients equal to 1 (open circles), and using activity coefficients given by the empirical extension of the Debye-Hückel law (black circles). Concentrations of acetate and acetic acid were calculated from IR spectra using a two-spectrum library.

The standard error of prediction from the 9 measurements that were within the reliability range was 0.07 pH units, which gives a confidence interval of ± 0.17 pH units. In order to quantify the importance of the correction for non-ideality, the calculations were also performed with activity coefficients equal to 1. The pH values predicted for this ideal case were all 0.15 pH units higher than the values calculated using the activity coefficients given by the empirical extension of the Debye-Hückel law. The SEP value was equal to 0.13 pH units, which is two times higher than with the correction for non-ideality.

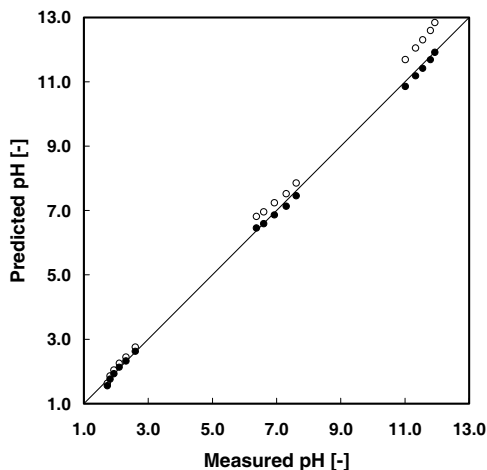


Figure 4. Predicted against measured pH for the titration of phosphoric acid. Calculations were performed using activity coefficients equal to 1 (open circles), and using activity coefficients given by the empirical extension of the Debye-Hückel law (black circles). Concentrations of H_3PO_4 , H_2PO_4^- , HPO_4^{2-} , PO_4^{3-} were calculated from IR spectra using a four-spectrum library.

Phosphoric acid ca. 0.1 M was titrated in a similar way to acetic acid. 1.68 ml of NaOH 10 M were added in 22 consecutive additions. The spectra were taken off-line, at 37°C, with water as background intensity. Due to the triprotic nature of phosphoric acid, the pure-component library had to be built in a slightly different manner to that with acetic acid. Solutions were prepared at the exact equivalence points (given by $0.5 \cdot [\text{pK}_{a1}' + \text{pK}_{a2}']$ and $0.5 \cdot [\text{pK}_{a2}' + \text{pK}_{a3}']$) to obtain the absorbance fingerprint of H_2PO_4^- and HPO_4^{2-} . Since pK_{a1} and pK_{a3} are very low and very high respectively (2.196 and 12.29 at infinite dilution and 25°C), solutions at pH values close to 0 and 14 would have been necessary to get fractions of H_3PO_4 and PO_4^{3-} above 99%. Unfortunately, at these points, the concentration of the hydronium cation or the hydroxide anion are close to 1 M, and their infrared absorbance cannot longer be

neglected. As a consequence, the molar absorbance of H_3PO_4 was calculated from the absorbance spectrum of a phosphoric acid solution of known concentration at $\text{pH} = 1.73$, by subtracting the absorbance caused by the fraction of H_2PO_4^- at this pH . The same operation was performed with the absorbance spectrum of a solution at $\text{pH} = 12.33$, to obtain the absorbance of PO_4^{3-} after subtraction of the HPO_4^{2-} contribution. For all species (H_3PO_4 , H_2PO_4^- , HPO_4^{2-} , PO_4^{3-}), the default values for α , β and b were used in thermodynamic calculations, except for HPO_4^{2-} , for which the published value of $-0.11 \text{ mol kg}^{-1}$ was used for b (Meier, 1982). When the fractions of the two major species were above 10%, the pH was predicted and plotted against the value measured by the conventional pH probe (Fig. 4).

The pH values that fulfilled the reliability criterion were symmetrically dispersed around the three pK_a of phosphoric acid (2.196, 7.163 and 12.23 at infinite dilution and 25°C), which is similar to what was also observed for acetic acid. Interestingly, the simple library-based approach was able to predict the pH with a fair accuracy over almost the entire standard pH range, using only the absorbance of the four main phosphoric acid species. The pH could even be predicted down to 1.6, where the H^+ activity is equal to 0.025 mol L^{-1} , which suggests that the infrared absorbance of the proton is likely to be rather weak. The SEP value for the titration of phosphoric acid was 0.10 pH units, which corresponds, with 95%-confidence interval, to an error range of $\pm 0.22 \text{ pH}$ units. Such a fairly small error of prediction shows that only the four main species of phosphoric acid are necessary to model the titration spectra. This also indicates that for the pH range, and for the relatively low concentration of acid used in these experiments, it is not necessary to account for minor complexes (Baril et al., 2000; Pitzer and Silvester, 1976) or “acidic” and “alkaline” infrared spectra (Max and Chapados, 1998). The second and third deprotonation produce divalent and trivalent species that strongly contribute to the non-ideality of the solution, since activity coefficients are proportional to the square of the charge according to the Debye-Hückel theory. The discrepancy between the real pH value and pH values calculated using activity coefficients equal to 1 was indeed much larger around the second and third dissociation than around the first, reaching even 0.6 pH units (Fig. 4). The SEP value of the ideal case was 0.48 pH units, which shows that correcting for non-ideality enhanced the accuracy of prediction by a factor of almost 5.

A mixture of acetic and phosphoric acid at ca. 0.1 M each was titrated in 24 steps with NaOH 10 M. The libraries used for the two previous titrations were merged to create a new calibration set that therefore contained six species: acetic acid, acetate, H_3PO_4 , H_2PO_4^- , HPO_4^{2-} and PO_4^{3-} . The activity coefficients were calculated with the same parameter values, and the

ionic strength was deduced from the charge balance in a similar manner as before. The same reliability criterion was applied to select data for comparison with the conventional pH-meter measurements (Fig. 5).

Mixing two acids slightly decreased the accuracy of the prediction (SEP : 0.15; 95%-confidence interval: ± 0.33 pH units) but allowed coverage of a larger part of the pH scale. The pH could indeed be predicted almost continuously from 3.9 to 7.1. In this range, the SEP value was 0.13 and the 95%-confidence interval ± 0.33 pH units. The pH could be calculated independently from both acid equilibria, but since the pK_a' (ca. 4.6) of acetic acid is not close to either the pK_{a1}' or pK_{a2}' of phosphoric acid (ca. 2.1 and 6.9 respectively), for none of the samples two predictions were available. It is interesting to notice that a solution of two acids with pK_a values within one pH unit, such as for instance succinic acid ($pK_{a1} = 4.19$) and formic acid ($pK_a = 3.75$), would lead to a redundant measurement of the pH. Such a mixture could also be used to estimate the ionic strength of a solution without having to measure the pH with a conventional electrochemical probe, by equalizing the two pH predictions (equation 5).

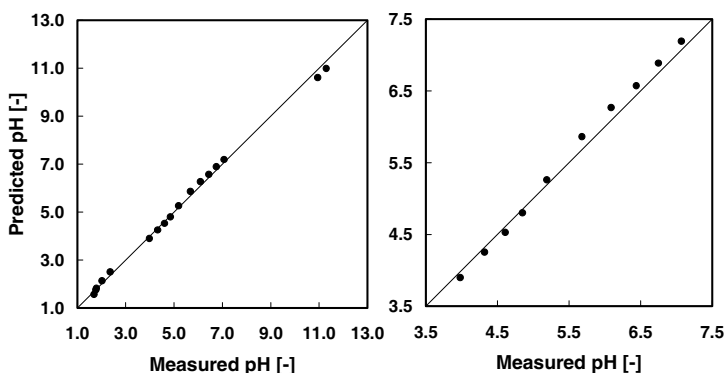


Figure 5. Predicted against measured pH for the titration of a mixture of acetic and phosphoric acids (left: entire scale; right: enlargement). Concentrations of acetic acid, acetate, H_3PO_4 , $H_2PO_4^-$, HPO_4^{2-} , PO_4^{3-} were calculated using a six-spectrum library. Activity coefficients were predicted using the empirical extension of the Debye-Hückel law with the same parameters as previously for the single-acid titrations. (Equation 6).

On-line Titration of a Multi-component System

As a step prior to the pH monitoring during microorganism cultivation, the prediction accuracy of the current FTIR-based method was assessed with a typical culture medium

during on-line operation, under sterile conditions. Previous measurements (data not shown) have provided evidence that the infrared spectra of non-charged species, such as glucose, or strong electrolytes, such as NaCl, are not significantly affected by the pH of the solution, at least within the pH range 3-11. This implies that for culture media, variations in the IR spectrum induced by a pH shift can be explained solely by the changes in weak acids dissociation equilibrium. More specifically, only the weak acids that have a pKa value within the range $\text{pH} \pm 1$ will produce significant changes in the infrared spectrum of the solution, as demonstrated earlier.

The composition of the medium used for this assessment was the same as for the cultures, except that no magnesium sulfate was added in order to avoid precipitation under alkaline conditions. The pH was set to 6.5 and the medium placed in the bioreactor under culture conditions (stirring rate, aeration, temperature) and recirculated to the flow-through cell. Spectra were acquired once per minute, and after 2 hours, HCl 5 M was pumped in at a constant rate until the pH reached the value of 5. Subsequent addition, at the same rate, of NaOH 5 M and HCl 5 M resulted in the pH curve displayed in Figure 6.

Some new equations had to be developed to enable prediction of pH by FTIR spectroscopy for the case where it is desired to calculate a pH difference from the initial state (ΔpH_i), rather than an absolute pH value. The reason for this is that the library-based concept of calibration deals with concentration differences rather than absolute concentrations, which allows for including the absorbance of all components displaying negligible concentration changes into the background intensity (Equation 2) (Schenk et al., 2007b). Using equation 9, it can be easily shown that the difference of pH at time t compared to the initial state (subscript o) can be formulated as:

$$\Delta\text{pH}_i = \text{pH}_t - \text{pH}_o = \log_{10} \left(\frac{\gamma_{A_t^{2-}}}{\gamma_{HA_t^-}} \right) + \log_{10} \left(\frac{C_{A_t^{2-}}}{C_{HA_t^-}} \right) - \log_{10} \left(\frac{\gamma_{A_o^{2-}}}{\gamma_{HA_o^-}} \right) - \log_{10} \left(\frac{C_{A_o^{2-}}}{C_{HA_o^-}} \right) \quad [15]$$

Assuming that the ionic strength remains constant during the process, the expression for ΔpH_i can be simplified to:

$$\Delta\text{pH}_i = \text{pH}_t - \text{pH}_o = \log_{10} \left(\frac{C_{A_t^{2-}}}{C_{HA_t^-}} \right) - \log_{10} \left(\frac{C_{A_o^{2-}}}{C_{HA_o^-}} \right) \quad [16]$$

because in the Debye-Hückel model, the activity coefficient of a given compound depends only on the ionic strength (Equation 6). Since the library-based calibration method calculates

concentration differences between time t and the initial medium composition, the initial concentrations of HA^- and A^{2-} have to be estimated, in order to solve equation 6. This can be easily done by solving the system of two equations resulting from the mass balance (Equation 17-a) and law of mass action (Equation 17-b, which is actually another form of Equation 5):

$$C_{\text{A}_o^{2-}} + C_{\text{HA}_o^-} = C_{\text{H}_2\text{A}_{\text{tot},o}} \quad [17-a]$$

$$\frac{C_{\text{A}_o^{2-}}}{C_{\text{HA}_o^-}} = 10^{\left(\text{pH}_o - \text{pK}_a - \log_{10} \left(\frac{\gamma_{\text{A}_o^{2-}}}{\gamma_{\text{HA}_o^-}} \right) \right)} \quad [17-b]$$

where $\text{C}_{\text{H}_2\text{A}_{\text{tot},o}}$ refers to the initial total concentration of the acid, regardless of the degree of dissociation, and is given by the medium composition. It is important to emphasize that even though the ionic strength does not appear directly in the prediction of ΔpH_t , its value is required for the calculation of the initial ratio of the dissociated and undissociated forms of the acid.

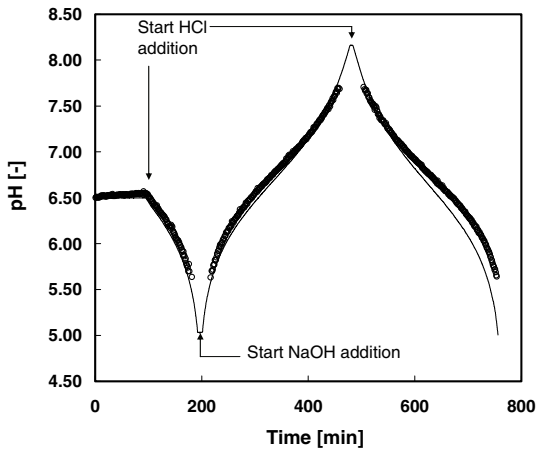


Figure 6. pH measured using a conventional electrochemical pH probe (plain line) and predicted by FTIR spectroscopy (open circles), in a typical culture medium in sterile bioreactor operation. pH change was induced by the addition at constant rate of HCl 5 M or NaOH 5 M.

The pH was predicted during the on-line titration of the culture medium using equations 16 and 17, using the same Debye-Hückel constant as before and an ionic strength of 0.32 mol L^{-1} , a value that was calculated from the medium composition. Concentrations of H_2PO_4^- and HPO_4^{2-} were calculated using a two-spectrum library extracted from the 4-spectrum library used previously. The SEP of the prediction was 0.12 pH units, which corresponds to an error

range of 0.24 pH units within a 95%-confidence interval (Fig. 6). The matching of the pH prediction to the conventional probe value was less good at the end of the experiment (after 10 hours), which can be explained either by a poor compensation of signal drift and/or an increase of the ionic strength due to the addition of concentrated acid and base. This result confirms that the infrared absorbance is not significantly influenced by the pH of the solution for most of the species found in a culture medium, such as strong electrolytes (e.g. NaCl), non-charged compounds (e.g. glucose) and weak acids/bases that have a pK_a/pK_b far enough from the usual cultivation pH (e.g. ammonia).

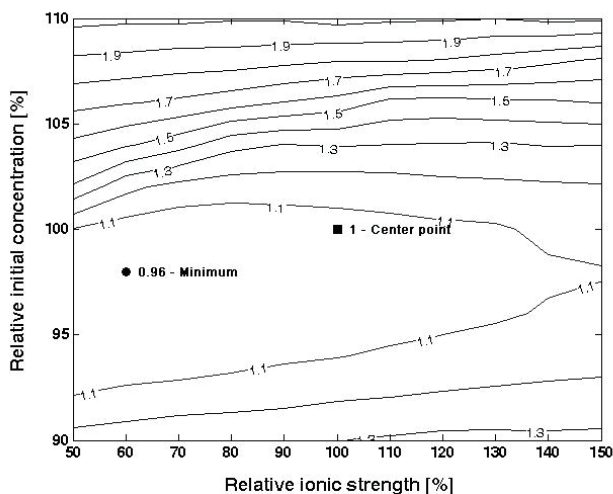


Figure 7. Contour plot (topology) of the relative SEP (normalized to the center point) of pH as a function of the ionic strength and the initial total concentration of the acid. The center point corresponds to an ionic strength of 0.32 mol L^{-1} and a total phosphate concentration of 0.136 mol L^{-1} , which gave a SEP of 0.12 pH units. The ionic strength had little influence on the error of prediction, since an error of 50% in estimating its value increased the SEP by about 20%. The initial concentration had a much higher leverage on the SEP of prediction (axes scales are different).

This calculation was based on the assumption that the ionic strength remained constant, but it also required specifying the total initial acid concentration. The sensitivity of the standard error of prediction to these parameters was tested by reprocessing the data for an ionic strength ranging from - 50% to + 50% around the estimated value (therefore within the range 0.16 to 0.48 mol L^{-1}) and for $\pm 10\%$ around the total initial acid concentration used previously. The two spans were chosen differently to account for the respective expected precision. All SEP values were normalized by the standard error of prediction of the center

point (0.10 pH units) and were represented as a contour plot, which is a projection on the x-y dimension of the response surface topology (Fig. 7). The model was found to be extremely sensitive to the initial total concentration, which means that the medium has to be prepared carefully. On the other hand, the ionic strength had little impact on the SEP value, since even an error of 50% in estimating its value did not decrease the accuracy of the method by more than 20%. The best SEP value found was only 4% lower than the center point, which indicates that the parameter estimation was rather accurate.

pH monitoring during cultures

The ability of the current method to reliably predict the pH during micro-organism cultivations was assessed with a batch culture of *Escherichia coli* on glucose conducted in a bioreactor. The pH was voluntarily not controlled in order to check if the FTIR-based prediction could follow the drop caused by ammonia consumption. pH monitoring was performed exactly the same way as for the on-line titration, except for the fact that the library this time included 6 species: glucose, NH_4^+ , H_2PO_4^- , HPO_4^{2-} , acetic acid and the acetate anion. The pH before inoculation was 6.5, and dropped by 0.1 units when 200 mL of inoculum were added to the bioreactor medium, at time zero (Fig. 8).

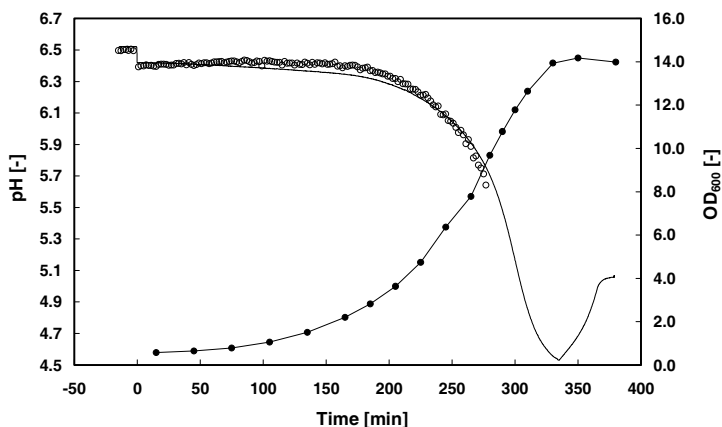


Figure 8. pH measured by a conventional pH probe (full line, left axis) and as predicted by FTIR spectroscopy (open circles, left axis) during a batch growth of *E. coli* on glucose without pH control. Optical density of the biomass is represented on the right axis (line with filled circles). The pH drop could be followed by FTIR accurately down to a pH of 5.7.

The biomass concentration reached an OD_{600} of 14 within less than 6 hours with complete glucose consumption, as indicated by off-line analyses. Over this time, the pH fell to 4.5, which resulted in growth inhibition at the end of the batch phase. The pH increased again to ca. 5 during the stationary phase, as a result of the acetic acid produced during the exponential growth being consumed. Even though this change (0.5 pH units) might appear to be very large, it must be kept in mind that this occurred in a range where phosphate has no buffering capability.

The pH predicted using the $H_2PO_4^-$ to HPO_4^{2-} ratio matched well the conventional pH probe measurements, since the discrepancy between the two measurements never exceeded 0.10 pH units. The SEP was 0.04 pH units, which corresponds to an error range of 0.08 pH units. The pH could not be predicted using the acetic acid to acetate ratio, because the measured concentrations were extremely low (< 15 mM). As a consequence, the pH could not be monitored outside the phosphate buffer range, i.e. below 5.6.

The large pH change had a negative effect on the standard error of prediction of other compounds. SEP values for glucose and ammonium were 12.6 and 20.9 mM respectively (on a total concentration change of 80 and 75 mM respectively), values that are almost three times higher than found with the same method at constant pH (Schenk et al., 2007b). The intense $H_2PO_4^-$ and HPO_4^{2-} infrared absorbance, coupled with the large buffer ratio change, accounts for a major part of the total variation observed in process spectra, with for consequence that the least-squares minimization increased the relative error on the other predictions.

FTIR-based pH control during cultures

In order to show that the developed technique is truly predictive, a feedback controller using FTIR-based pH monitoring was implemented in the bioreactor set-up. An *E. coli* batch culture was performed under the same conditions as previously, but the pH was controlled at the set point of 6.5 by adding NaOH 4 M on demand of the controller. The spectra library used for this experiment, as well as the thermodynamic parameters, were the same as for the previous batch. At time zero, the addition of inoculum produced a drop of 0.12 pH units that was immediately compensated by the controller (Fig. 9). The two pH measurements then slightly diverged and a maximum discrepancy of 0.12 pH units could be observed. This offset did not further increase over the next 3 hours. Due to the pH control, growth remained exponential until glucose was exhausted and the process was, therefore, slightly shorter than the previous batch (20 min), during which cell growth was inhibited by medium acidity.

The pH of the culture medium could always be maintained within ± 0.15 pH units from the desired set-point. Such a range is relatively large compared to that obtained using a controller based on a conventional electrochemical probe; however pH deviations of this order of magnitude have little influence on the metabolism of most bacteria and yeasts (Bailey and Ollis, 1986).

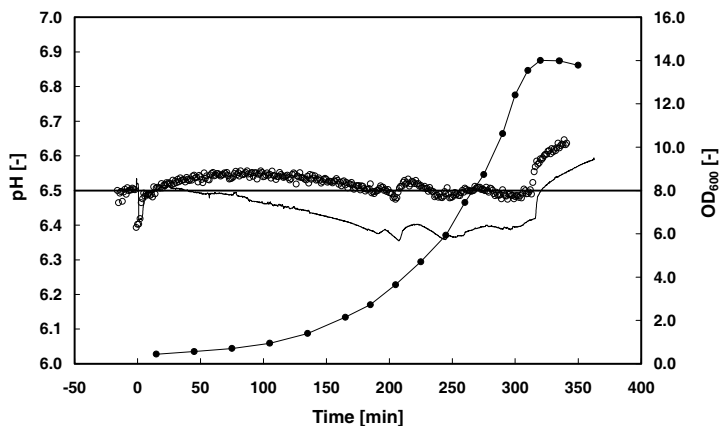


Figure 9. pH measured by a conventional pH probe (line without symbols, left axis) and as predicted by FTIR spectroscopy (open circles, left axis) during batch growth of *E. coli* on glucose. pH control was based on FTIR predictions; the set point was pH = 6.5. Optical density of the biomass is represented on the right axis (line with filled circles). Maximum discrepancy between the two measurements was 0.12 pH units.

While the glucose prediction matched almost perfectly the off-line data, the ammonium concentration slightly diverged from the reference analyses, especially at the end of the culture (Fig.10). the maximum discrepancy between the two sets of results never exceeded 16 mM. The SEP values for glucose and ammonium, 3.9 and 10.9 mM respectively, are nevertheless very similar to what can be expected when a usual pH controller based on an electrochemical probe is used (Schenk et al., 2007b). Total acetic acid concentration again never exceeded 15 mM.

The accuracy of this infrared-based method is slightly less than expected from a pH sensor (Kermis et al., 2002), however the ease of handling and the simultaneous measurement of multiple metabolites renders the spectroscopic approach interesting for bioprocess applications. In addition, it also does not require the modification of the culture medium, such as the addition of potentially toxic dyes, and therefore presents no risk of cell

metabolism perturbation. Moreover, it also presents the advantage of being simple in terms of spectroscopic modeling, since a library of pure component spectrum can achieve the desired accuracy. While it is clear that on-line FTIR cannot compete in terms of accuracy with conventional pH measurement and off-line analyses, this approach gives access to a fairly high amount of information at very low labour cost. This makes FTIR particularly suitable for screening experiments, for which the ratio of information to work load is a critical parameter.

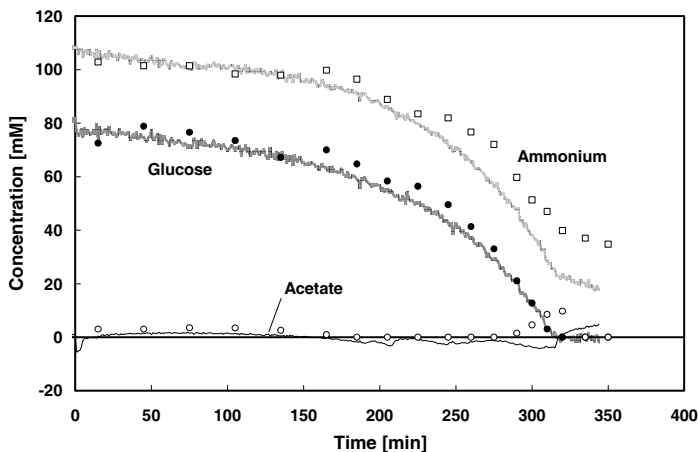


Figure 10. Concentration profiles of glucose (black circles), ammonium (open squares) and acetate (open circles) measured off-line (symbols) and on-line by FTIR (corresponding lines), during a batch culture of *E. coli*. The pH controller was based on the pH prediction calculated from the infrared measurements.

6/ CONCLUSIONS

This work demonstrated that the pH of bioprocess media can be predicted from mid-infrared measurements, using a spectra library to calculate the concentration of the species of buffer in solution, and the extended Debye-Hückel theory to estimate activity coefficients. This shows that the pH does not significantly influence the molar infrared absorbance of most compounds, but only the ratio of the protonated to deprotonated forms of weak acids.

A simple, chemically defined medium was used in the current study, and it remains unclear how medium complexity affects accuracy of prediction. This point is being investigated using rich and semi-defined media, and preliminary results suggest that the library-based calibration approach withstands the addition of compounds such as yeast extract or peptone as long as they are not massively consumed during the culture.

This method presents several advantages that are particularly relevant to bioprocess applications when space is limited (e.g. as in microbioreactors), since it is non-invasive, does not require the addition of a fluorescent dye and it is fast enough to be integrated into a feedback controller. With the recent development of fiber-optics and miniaturization, it should be possible to implement FTIR into a high-throughput platform for medium and strain screening, as an all-in-one probe capable of measuring simultaneously the pH and all major metabolite concentration profiles.

7/ REFERENCES

- ASTM. 1997. Standard practices for infrared, multivariate, quantitative analysis (E 1655-97). Annual book of ASTM standards. Philadelphia: American Society for Testing and Materials. p 844-869.
- Bailey JE, Ollis DF. 1986. Biochemical engineering fundamentals. New York: McGraw-Hill.
- Baril J, Max J-J, Chapados C. 2000. Infrared titration of phosphoric acid. Can. J. Chem. 78:490-507.
- Crowley J, McCarthy B, Nunn NS, Harvey LM, McNeil B. 2000. Monitoring a recombinant *Pichia pastoris* fed batch process using Fourier transform mid-infrared spectroscopy (FT-MIRS). Biotechnol. Lett. 22:1907-1912.
- Doak DL, Phillips JA. 1999. In situ monitoring of an *Escherichia coli* fermentation using a diamond composition ATR probe and mid-infrared spectroscopy. Biotechnol. Prog. 15:529-539.
- Goldberg RN, Kishore N, Lennen RM. 2002. Thermodynamic quantities for the ionization reactions of buffers. J Phys. Chem. Ref. Data 31(2):231-370.
- Jeevarajan AS, Vani S, Taylor TD, Anderson MM. 2002. Continuous pH monitoring in a perfused bioreactor system using an optical pH sensor. Biotechnol. Bioeng. 78(4):467-472.
- John GT, Goelling D, Klimant I, Schneider H, E. H. 2003. pH-Sensing 96-well microtitre plates for the characterization of acid production by dairy starter cultures. J. Dairy Res. 70(3):327-333.
- Karoui R, Mouazen AM, Dufour E, Pillonel L, Schaller E, Picque D, De Baerdemaeker J, Bosset J-O. 2006. A comparison and joint use of NIR and MIR spectroscopic methods for the determination of some parameters in European Emmental cheese. Eur. Food Res. Technol. 223:44-50.
- Kermis HR, Kostov Y, Harms P, Rao G. 2002. Dual excitation ratiometric fluorescent pH sensor for noninvasive bioprocess monitoring: development and application. Biotechnol. Prog. 18(5):1047-1053.
- Kornmann H, Rhiel M, Cannizzaro C, Marison I, von stockar U. 2003. Methodology for real-time, multi-analyte monitoring of fermentations using an in-situ mid-infrared sensor. Biotechnol. Bioeng. 82(6):702-709.
- Lobo AP, Valles BS, Tascon NF, R.R. M, Garcia OF. 2006. Calibration models for routine analysis of cider by mid-infrared spectroscopy. LWT-Food Sci. Technol. 39:1026-1032.

- Martens H, Naes T. 1992. Multivariate calibration. New York: John Wiley & Sons.
- Max J-J, Chapados C. 1998. Subtraction of the water spectra from infrared spectra of acidic and alkaline solutions. *Appl. Spectrosc.* 52(7):963-969.
- Meier PC. 1982. Two-parameter debye-hückel approximation for the evaluation of mean activity coefficients of 109 electrolytes. *Anal. Chim. Acta* 136:363-368.
- Pitzer KS, Silvester LF. 1976. Thermodynamics of electrolytes. VI. Weak electrolytes including H_3PO_4 . *J. Solution Chem.* 5(4):269-278.
- Pollard DJ, Buccino R, Connors NC, Kirschner TF, Olewinski RC, Saini K, Salmon PM. 2001. Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy. *Bioprocess. Biosyst. Eng.* 24:13-24.
- Rhiel MH, Ducommun P, Bolzonella I, Marison IW, von Stockar U. 2002. Real-time in situ monitoring of freely suspended and immobilized cell cultures based on mid-infrared spectroscopic measurements. *Biotechnol. Bioeng.* 77(2):174-185.
- Riley MR, Arnold MA, Murhammer DW, Walls EL, DeLaCruz N. 1998. Adaptive calibration scheme for quantification of nutrients and byproducts in insect cell bioreactors by near-Infrared spectroscopy. *Biotechnol. Prog.* 14(3):527-533.
- Schenk J, Marison IW, von Stockar U. 2007a. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2):344-353.
- Schenk J, Marison IW, von Stockar U. 2007b. Simplified Fourier-transform mid-infrared calibration based on a spectra library for the on-line monitoring of bioprocesses. *Anal. Chim. Acta* 591(1):132-140.
- Siggaard-Andersen O, Durst RA, Maas AHJ. 1984. Physicochemical quantities and units in clinical chemistry with special emphasis on activities and activity coefficients. *Pure & Appl. Chem.* 56(5):567-594.
- Song A, Parus S, Kopelman R. 1997. High-performance fiber-optic pH microsensors for practical physiological measurements using a dual-emission sensitive dye. *Anal. Chem.* 69(5):863-867.
- Steyer JP, Bouvier JC, Conte T, Gras P, Harmand J, Delgenes JP. 2002. On-line measurement of COD, TOC, VFA, total and partial alkalinity in anaerobic processes using infra-red spectrometry. *Water Sci. Technol.* 45(10):133-138.
- Xu X, Smith S, Urban J, Cui Z. 2006. An in line non-invasive optical system to monitor pH in cell and tissue culture. *Med. Eng. Phys.* 28(5):468-474.

Zemaitis JF, Clark DM, Rafal M, Scrivner NC. 1986. Handbook of aqueous electrolyte thermodynamics. New York: American Institute of Chemical Engineers (AIChE).

8/ LIST OF SYMBOLS

Greek symbols

α	First Debye-Hückel parameter [$\text{kg}^{1/2} \text{mol}^{1/2}$]
β	Second Debye-Hückel parameter [$\text{kg}^{1/2} \text{mol}^{1/2}$]
ε	Experimental error on absorbance [-]
γ	Activity coefficient [-]
κ	Molar absorptivity (or molar extinction coefficient) [M^{-1}]
ν	Wavenumber [cm^{-1}]

Other symbols

a	Activity [mol L^{-1}]
A	Absorbance [-]
b	Parameter for empirical extension of Debye-Hückel [kg mol^{-1}]
C	Concentration [g L^{-1}], or [mol L^{-1}]
I_c	Ionic strength [mol kg^{-1}]
I_ν	Spectrum intensity at wavenumber ν [-]
j	Number of compounds in library
K_{a1}	Equilibrium constant of first deprotonation [mol L^{-1}]
K_{a2}	Equilibrium constant of second deprotonation [mol L^{-1}]
K'	Apparent equilibrium constant [mol L^{-1}]
n	Number of off-line samples
SEP	Standard error of prediction [pH units] or [mol L^{-1}] or [g L^{-1}]
y	Predicted or measured concentration [g L^{-1}], or [mol L^{-1}]

Superscripts and subscripts

k	subscript	k-th sample
i	subscript	i-th species
o	subscript	Initial conditions
t	subscript	Time
tot	subscript	Total

Chapter 6

Effect of Temperature on Mid-Infrared Spectroscopy Calibrations for the On-line Monitoring of Bioprocesses

1/ ABSTRACT

Calibration models for mid-infrared spectrometers are generally considered to be highly temperature-sensitive, due to non-linear perturbations induced by this process variable. The influence of temperature on mid-infrared spectra was investigated from this point of view, and it was shown that intensity spectra are largely influenced by the temperature. However, within the fingerprint region ($1000 - 1500 \text{ cm}^{-1}$), absorbance was linearly dependent on the temperature, although the parameters of the linear regression were wavenumber-dependent. As a result, it could be shown that the molar absorbance of glucose and ammonium chloride can be considered as independent of temperature. Five batch cultures of *E.coli*, carried out at temperatures between 20°C and 38°C , were monitored on-line by mid-infrared spectroscopy as a case study. The calibration set consists of a single library of pure component spectra that were measured at 30°C , and modelling was done by traditional least squares. The average standard error of prediction for glucose, ammonium and acetate were 7.0, 6.4 and 5.5 mmol L^{-1} respectively. The simple calibration approach proposed can, therefore, be considered as temperature-independent, and in addition, as accurate as more complex chemometrics approaches.

This chapter was submitted for publication in *Applied Spectroscopy*:

Schenk J, Viscasillas C, Marison IW, von Stockar U. Effect of Temperature on Mid-Infrared Spectroscopy Calibrations for the On-line Monitoring of Bioprocesses.

2/ INTRODUCTION

Many bioprocess developments include the testing of several culture temperatures, in order to gain an understanding of strain yields and kinetics, but also as a means to potentially increase productivity. The use of mid-infrared spectroscopy for the on-line monitoring of such culture runs, although highly desirable in order to reduce the work involved in off-line analyses, can be very problematic, since temperature is known to have a large influence on infrared spectra (Hageman et al., 2005). The development of a calibration model for each temperature is not an interesting option, given the time involved for that purpose in standards preparation and measurement, and alternative solutions have to be investigated. Therefore this paper aims at proposing a calibration approach for mid-infrared spectroscopy that is simple and temperature-independent, and allows the monitoring of processes at different temperatures with a single library of pure component spectra.

Temperature determines the rotational and vibrational states of the molecules, and a large amount of research has been done on its effect on near-infrared spectra (Cozzolino et al., 2007; Czarnik-Matusewicz et al., 2005; Segtnan et al., 2005; Wülfert et al., 1998). Peak shifting, increases and decreases of absorbance have been reported, and perturbations are generally considered as nonlinear. In comparison, less information is available on the influence of temperature on mid-infrared spectroscopy, which is a more recent technique for process monitoring. Concerning the fingerprint region (ca. 900-1800 cm^{-1}), where most compounds absorb, main effects that occur are baseline shifts and a change of the peak height attributed to –OH band (Doak and Phillips, 1999; Pollard et al., 2001)

Various attempts have been made to develop calibration models that can handle the perturbations induced by temperature changes. A Partial Least Squares (PLS) model built using standards measured at different temperatures was reported to be able to accurately predict concentration in the corresponding temperature range (Bellon-Maurel et al., 1995; Wülfert et al., 1998). Another approach consists in including explicitly the temperature in calibration data, by appending this additional information to measured spectra. This method has been successfully applied, with PLS (Wülfert et al., 2000) and artificial neural network (ANNs) (Fayolle et al., 1996) modelling.

However, all of these examples imply the use of advanced mathematical treatment and large sets of standards, which are well-identified roadblocks to the use of mid-infrared spectroscopy for bioprocess development (Roychoudhury et al., 2006b; Vojinovic et al., 2006). An alternative solution for Fourier-transform mid-infrared spectroscopy (FTIR) comes from a

simpler calibration approach based on a library of pure component spectra that was presented in a previous article (Schenk et al., 2007b). This method assumes that absorbance is linear with respect to concentration and that species do not interact in solutions – assumptions that are likely to be verified in the dilute aqueous solution that are culture media. As a consequence, any process spectrum is expressed as a linear combination of pure component spectra, using traditional least-squares. While such a calibration involves the preparation of only a few standards (i.e. one per compound of interest), it proved to be as accurate as a more sophisticated approaches, such as a PLS model based on 49-standard calibration set.

This paper aims at showing how this library-based calibration approach can be extended to handle different process temperatures. For that purpose, the influence of the temperature on mid-infrared intensity spectra and absorbance of common compounds was studied within the range 20 – 55°C. *E.coli* was chosen as a model organism to perform batch cultures at temperatures from 20°C to 38°C.

3/ MATERIALS AND METHODS

Bioreactor Fermentation

The wild-type *Escherichia coli* strain ML30 (DSMZ No 1329, Braunschweig, Germany) was used in all experiments. The inoculum was prepared by growing the cells in two shake flasks containing 100 ml of the bioreactor medium for 14 hours, at process temperature. Medium composition was: glucose 15 g L⁻¹, KH₂PO₄ 13.3 g L⁻¹, (NH₄)₂PO₄ 5 g L⁻¹, NH₄Cl 2 g L⁻¹, citric acid anhydrous 1.7 g L⁻¹, MgSO₄·7H₂O 1.2 g L⁻¹, Na₂EDTA 10.7 mg L⁻¹. The pH of the medium was adjusted to 6.5 by addition of NaOH 5 M. Sterile solutions of antifoam (Struktol SB2121, Schill and Seilacher, Hamburg, Germany), vitamins (thiamine.HCl 4.5 g L⁻¹) and trace elements (composition per litre: 0.3 g H₃BO₃, 0.15 g CuCl₂·4H₂O, 1.35 g MnSO₄·H₂O, 0.25 g Na₂MoO₄·2H₂O, 1.3 g Zn(CH₃CHOO)₂·2H₂O, 0.25 g CoCl₂·6H₂O, 11 g FeCl₃·6H₂O) were added to the filter-sterilized medium (0.2 µm Steritop®, Millipore, MA). Fermentations were carried out in a 3.6-L bioreactor (Bioengineering, Wald, Switzerland) with a working volume of 2 litres. Agitation was set at 1000 rpm, with three 6-blade Rushton impellers, and total aeration was constant at 3 Nl/min. When necessary, pure oxygen was blended with air to maintain dissolved oxygen above 20% of air saturation. Oxygen and carbon dioxide composition of the exhaust gas was measured using a gas analyzer (Dr. Marino Müller Systems, Esslingen, Switzerland).

Off-line sample analysis

Culture samples (5 mL) were collected at intervals of 10 to 45 minutes (depending on the growth kinetics) and immediately cooled to 2°C. Dry cell weight (g L⁻¹) was determined by gravimetric analysis: 3 ml of sample were filtered on 0.2 µm pre-weighed filters, and then dried to constant weight. Optical density was measured at 600 nm (OD₆₀₀). The remaining volume of the sample was filtered through 0.2 µm filters for off-line analysis of metabolites. Glucose, ammonium ion and acetic acid concentrations were determined by enzymatic assay, using commercially available kits (R-Biopharm AG, Darmstadt, Germany) and an automated analyzer (Cobas Mira, Roche, Basel, Switzerland).

FTIR set-up and spectra acquisition

A Mettler Toledo Fourier-transform mid-infrared spectrometer (FTIR) 4000 (Mettler-Toledo, Greifensee, Switzerland) with an Attenuated Total Reflectance probe (ATR, DiComp) was mounted in a flow-through cell, which allows convenient measurement of standards in

an off-line configuration, but also connection to a bioreactor for on-line operation. The temperature of the flow-through cell was tightly controlled by a thermostated bath. This assembly is very similar to an in-situ ATR-probe, since it allows cleaning and sterilization in place. More details on the instrument and the set-up can be found in previous publications (Doak and Phillips, 1999; Pollard et al., 2001; Schenk et al., 2007a). During cultures, a recirculation loop driven by a membrane pump (Prominent Gamma/L, Heidelberg, Germany) was used to connect the reactor to the FTIR flow-through cell. The residence time within the loop was less than 20 seconds, while the response time was 15 seconds. Spectra were acquired every 2 minutes.

Series of 64 single-beam intensity spectra with a resolution of 4 cm^{-1} were collected and averaged for each measurement. The absorbance A at a wavenumber ν was calculated using:

$$A_{\nu} = \log_{10} \left(\frac{I_{\nu,o}}{I_{\nu}} \right) \quad [1]$$

where I_o is a reference intensity and I is the intensity measured with the sample. The choice of the reference intensity varied with the applications and the mode of operation, as it is explained below.

Setting up the library

The spectra library consisted of the molar absorbance of three pure components, namely glucose, ammonium and acetate, as well as three “drift spectra”, which were used to correct for signal instabilities (see below “signal drift correction”). The denomination pure component refers here to a single compound dissolved in water at a concentration around 0.1 mol L^{-1} . Molar absorbances were determined by measuring the intensity spectra (I) of the standards in parallel with the intensity spectrum of water, which was used as reference intensity (I_o). Measurements were performed at 30°C , and the pH of standards were adjusted to 6.5 using NaOH or HCl 1 M. Ammonium chloride and sodium acetate salts were used for the pure component spectrum of ammonium and acetate respectively, since sodium and chloride ions show very little infrared activity (Schenk et al., 2007b).

Calculation of concentrations from infrared data

The experimental evidence supporting the calibration approach based on a spectra library, as well as a complete discussion of the methodology, can be found in previous articles (Schenk et al., 2007b; Schenk et al., 2007c). Only a brief description is given here.

The use of a library of pure component spectra as calibration set implicitly assumes that absorbance is linear with respect to concentration, and that species in solution do not interact with each other. Both hypotheses were verified experimentally for common compounds found in culture medium and should hold for most bioprocesses, which are by nature carried out in dilute aqueous solutions. Any absorbance measured during the process is then expressed as a linear combination of the spectra from the library, the best combination being found by the least-squares algorithm. The first intensity spectrum, measured with the culture medium under process conditions, is used as reference intensity I_0 . This allows including the absorbance of all the compounds that do not show significant concentration during the process into a baseline, which is eliminated by subtraction. Measured absorbances have therefore to be considered as absorbance differences from initial state. This presents the advantage that the library can be used for different cultures, regardless of the medium composition, as long as the main metabolites remain the same (Schenk et al., 2007c).

The absorbance difference from the initial state ΔA (equation 1, with I_0 measured in culture medium before inoculation) was calculated for every spectrum acquired during the process (at time t) and expressed as a linear combination of molar absorbances:

$$\Delta A_{v,t} = \sum_{i=1}^{i=j} \kappa_{i,v} \cdot \Delta C_i + \varepsilon_v \quad [2]$$

where $\kappa_{i,v}$ is the molar absorbance of a compound i at a wavenumber v , and ΔC its concentration change with respect to initial medium composition. ε is a residual due to experimental errors that is minimized by the least-squares algorithm. This equation is clearly over-determined since the total number of compounds j is in the present study equal to 3, while the number of equations is close to 200, since a measurement is made every 4 cm^{-1} , from 1000 to 1500 cm^{-1} . Equation 2 can be rewritten in matrix notation, which gives a single equation (Equation 3) for the whole process:

$$\Delta \mathbf{A}_{s,wn} = \Delta \mathbf{C}_{s,j} \times \mathbf{K}_{j,wn} + \mathbf{E}_{s,wn} \quad [3]$$

The subscripts refer to the matrix dimensions (row,column), j , wn and s being the number of compounds, of wavenumber monitored and the number of spectra taken during the process respectively. \mathbf{K} is the spectra library, which is set-up prior to the experiments. Concentration profiles are therefore found by resolving equation 3 to find $\Delta \mathbf{C}$, followed by the consecutive addition of the initial concentration, given by the medium composition. The size of $\Delta \mathbf{A}$ was chosen to be $s \times wn$, instead of $wn \times s$ as in our previous work (Schenk et al., 2007b), in order to comply with the traditional chemometrics notation. The following equations were adapted in consequence.

Analysis of the residuals can be used to detect the appearance of unexpected compounds, and even point to their identity. Residuals can be displayed in three dimensions, but projections in the time ($SSQt$) and wavelength ($SSQwn$) dimension can facilitate interpretation (Equation 4 – 5).

$$SSQt = row_sum(\mathbf{E} \times \mathbf{E}) \quad [4]$$

$$SSQwn = column_sum(\mathbf{E} \times \mathbf{E}) \quad [5]$$

where row_sum and $column_sum$ are used to indicate a sum of the matrix elements by row and column respectively, and the operator “ \times ” an element-by-element multiplication. Note that unlike E , $SSQt$ and $SSQwn$ are not matrices but vectors and can be plotted in two dimensions. The sum of squared residues obtained for every time can be compared to the total variance found in the system, which gives the explained variance EV, expressed here as a percentage:

$$EV = 100 \cdot [1 - SSQt ./ row_sum(\Delta\mathbf{A} \times \Delta\mathbf{A})] \quad [6]$$

where the “ $./$ ” operator refers to an element-by-element division. The averaged explained variance (AEV) is defined as the mean of the explained variance over the entire process. It does not make much sense to calculate an explained variance in the wavenumber dimension, since in some regions of the spectra, typically where variations are very low, the solution found by the least-square algorithm can be larger than the observed change, since the minimization of the residual error is made over a wide wavenumber range.

The accuracy of the FTIR prediction calibration was assessed from the standard error of prediction (SEP), which was calculated using the following formula:

$$SEP = \sqrt{\frac{\sum_k (y_{k,Off-line} - y_{k,FTIR})^2}{n}} \quad [7]$$

where y_k indicates any off-line data or its corresponding FTIR value, and n refers to the total number of samples (typically 18-25).

Signal drift correction

Two different methods for signal drift correction were compared in this work. Both methods have been described in previous articles. The first approach consists in including in the spectra library, in addition to molar absorbances, “drift spectra” that will account for signal instability when equation 3 is solved (Schenk et al., 2007c). The drift spectra were obtained from a preliminary stability experiment. Water was placed aseptically at 30°C in the bioreactor, under process conditions, with 1.5 mL of antifoam to prevent adherence of air

bubbles to the detector surface. The first intensity spectrum was used as reference. During 24 h, intensity spectra were measured every 2 minutes, and absorbances were calculated using equation 1. Absorbance variations that were observed were attributed to signal drift. Principal Component Analysis (PCA) was used to identify the principal spectral features of the signal drift (or loadings, in chemometrics terminology), and the first three were included as drift spectra in the library. Practically, the size of library **K** (equation 3) was increased from $3 \times wn$ to $6 \times wn$, by adding the three main spectral features to the molar absorbance spectra of glucose, ammonium and acetate. Any multiple of the drift spectra could have been included in the library, since no drift amount or concentration can be reasonably defined. Solving equation 3 led, therefore, to molar concentrations for glucose, ammonium and acetate and arbitrary quantities for the three drift spectra.

The second approach used to correct for signal instabilities is simpler, but also less accurate. The method consists in a simple signal anchorage, and is based on the observation that, even though signal drift is wavenumber-dependent, it is sufficiently homogeneous within the fingerprint region to allow a linear correction (Schenk et al., 2007a). Two points are chosen in the spectra where none of the compounds of interest absorb, but close enough to the calculation range ($1000 - 1500 \text{ cm}^{-1}$) to be representative enough. In the current study, these points were set at 950 and 1750 cm^{-1} , as in previous works. The two anchors were used to calculate, for each spectrum measured during the process (at time t), a corrected signal intensity from the raw signal intensity, I , using:

$$I_{t,v}^{corrected} = \frac{I_{t,v}}{0.5 \cdot \left(\frac{I_{t,950}}{I_{o,950}} + \frac{I_{t,1750}}{I_{o,1750}} \right)} \quad [8]$$

4/ RESULTS AND DISCUSSION

Influence of temperature on signal intensity

In order to observe the effect of temperature on mid-infrared spectra, 2 litres of aseptic water with 1 mL of antifoam were placed in the bioreactor, with moderate stirring, at 20°C. The solution was recirculated to the flow-through cell, which was also controlled at 20°C. After stabilization of the system (after ca. 2 h), a first intensity spectrum I_0 was measured.

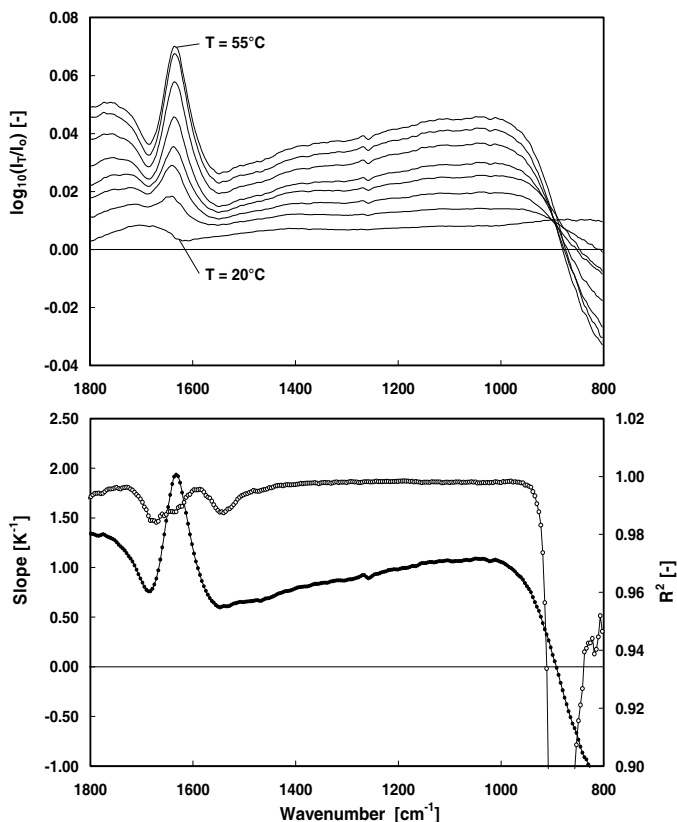


Figure 1. Upper frame: Absorbance spectra of water at 20 °C, 25 °C, 30 °C, 35 °C, 40 °C, 45 °C, 50 °C and 55 °C (lines without symbol in ascending order; left y-axis) calculated using an intensity spectrum of water at 20°C as reference intensity. The absorbance at 20°C was not equal to zero because of signal drift. Lower frame: slope (line with black circles; left y-axis) and regression coefficient (line with open circles; right y-axis) for the linear correlation between absorbance and temperature.

One hour later, the intensity spectrum of the solution at 20°C, $I_{T=20^{\circ}\text{C}}$ was measured again in order to calculate the absorbance (equation 1). This absorbance spectrum should have been equal to zero, since the properties of the solution were exactly the same. However, due to signal drift, a small offset could be observed (Fig. 1, upper frame). Temperature set point was then increased to 25°C and one hour later, the intensity spectrum at 25°C was acquired to calculate the absorbance. This operation was repeated with temperature steps of 5°C up to 55°C. While on most of the fingerprint region, the absorbance increased with the temperature, it decreased within the range 800 – 900 cm^{-1} . The absorbance measured at 55°C corresponds, at 1034 cm^{-1} , to the absorbance of a 30 g L⁻¹ glucose solution, which highlights the very large influence of temperature on infrared intensity spectra.

For each wavenumber, the absorbance was found to be linearly related to the temperature, by plotting these two variables against each other. A regression on the eight data points allowed characterizing the linearity of this relationship, using the correlation coefficient R^2 , as well the dependence to the wavenumber, given by the slope (Fig. 1, lower frame). Within the range 1000 – 1500 cm^{-1} , R^2 was above 0.99, which denotes a clear linear correlation between absorbance and temperature. Outside this range the correlation was slightly lower, with coefficients from 0.94 to 0.98, apart from the low-absorbance region around 980 cm^{-1} . The slope of the regression was strongly wavenumber-dependent, with positive values on most of the fingerprint domain and negative values below 980 cm^{-1} . Previous studies have reported the same trends, however they did not make a quantitative description of the phenomenon (Doak and Phillips, 1999; Pollard et al., 2001). They noted that the overall change induced by the temperature can be roughly represented by a baseline shift between 950 and 1500 cm^{-1} , coupled with an enlargement of the peak height attributed to water –OH bend, located around 1630 cm^{-1} . Interestingly, linearity is poorer in this latter range.

Due to the inertia of the system regarding temperature changes, it was not possible to always reference the measurements to the spectrum of water at 20°C in order to eliminate the influence of signal drift. However, the R^2 coefficients, which are close to 1, demonstrate a clear correlation between absorbance and temperature. The experiment was nevertheless repeated in the reverse order, by starting at 55°C and reducing the temperature to 20°C by steps of 5°C, and the results were very similar (data not shown).

From a purely practical point of view, and in terms of calibration purposes, absorbance could be considered as linear with respect to temperature for each wavenumber in the fingerprint region. The parameters of the linear function, i.e. the slope and the intercept, were,

however, found to be strongly wavenumber-dependent. Deviations from linearity may be observed on a broader temperature range, but with respect to microorganism cultures, it did not make much sense to calibrate above 50°C or below 20°C.

Influence of temperature on the absorbance of glucose and ammonium

In order to show the large influence of temperature on infrared spectra, in the previous section, absorbances were calculated using a reference intensity spectrum measured at a certain temperature and a sample intensity spectrum acquired at another temperature. In practice, however, both intensity spectra are generally measured at the same temperature. Following this approach, the molar absorbance of glucose and ammonium chloride was determined from solutions of known concentration, at every 5°C from 20°C to 45°C, using water at the same temperature to determine the reference intensity (Fig. 2). The temperature of the measurement cell was controlled by a water bath, in which water and standards were submerged. No significant influence of the temperature could be observed in the molar absorbance spectrum of glucose and ammonium within this temperature range. The small variations measured in the middle of the fingerprint region (1180 – 1380 cm^{-1}), where the absorbance of both compounds was very low, were not correlated to temperature and seemed to be due to signal drift.

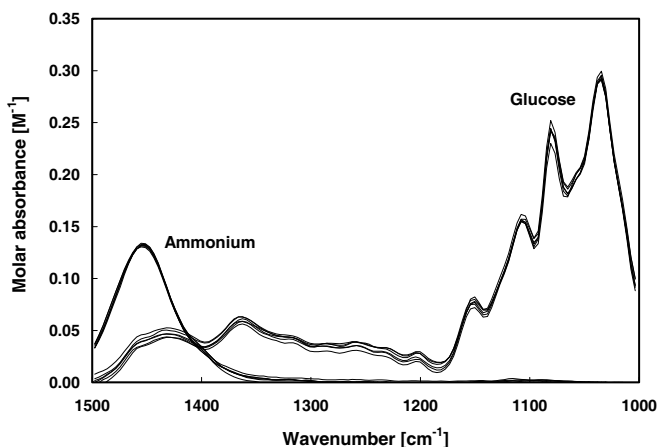


Figure 2. Absorbance spectra of glucose and ammonium at 20 °C, 25 °C, 30 °C, 35 °C, 40 °C and 45 °C. Reference intensity was the spectrum of water at the same temperature as the sample. The temperature showed no significant effect on the molar absorbance of glucose and ammonium.

The fact that these molar absorbances did not vary with temperature implies that the absorbance of glucose and ammonium are affected by temperature exactly the same way as water. This means that if the experiment discussed in the previous section was repeated with a glucose or ammonium solution, the same slopes as shown in figure 1 (lower frame) would be found. In other words, heating the standards and background water multiplies all intensity spectra proportionally, thereby keeping the intensity ratio constant and leading to temperature-invariant molar absorbances.

Batch cultures at different temperatures using the same calibration

The library-based calibration approach, as it has been developed, references all process spectra to the first spectrum of the culture, measured with culture medium prior to inoculation (Schenk et al., 2007b). The original goal of this referencing was to include the absorbance of all the compounds that do not exhibit significant concentration changes into the background. However, this also implicitly eliminates the effect of temperature, as shown with the molar absorbance of glucose and ammonium. Of course, temperature has to be maintained constant throughout the culture, and this approach is therefore not suitable for processes that include cold- or heat-shock, or any kind of temperature shift.

In order to assess the ability of the library-based calibration approach to monitor concentration profiles at different temperatures, batch cultures on glucose of the *E.coli* wild-type strain ML30 were performed at 20, 25, 30, 36, 38 and 42°C. Cells did not grow at 42°C, and it was considered that only five runs were relevant. Three metabolites were monitored, namely glucose, ammonium and acetate. As discussed in the next section, the method involving “drift spectra” to correct for signal instabilities led to better results than the anchoring approach. For this reason, the results shown here were calculated using this method. Calculations were performed as described in materials and methods section, without any additional data pre-treatment and post processing. The results can therefore be considered as truly predictive, as it was demonstrated previously by a spiking experiment (Schenk et al., 2007b). The same spectra library was used for all experiments as calibration set and contained the molar absorbance, measured at 30°C, of glucose, ammonium, and acetate as well as three drift spectra. These were found by a Principal Component Analysis (PCA) of water absorbance spectra, which were measured in aseptic conditions, at 30°C, over a period of time of 24 hours.

The concentration profiles of the batch culture performed at 20°C are represented in figure 3. Growth was very slow (specific growth rate $\mu = 0.22 \text{ h}^{-1}$), and about 17 hours were

necessary to reach glucose depletion. The maximum acetate concentration was 6.5 mM, which is very low but normal for this strain (Schenk et al., 2007c). FTIR and off-line profiles matched well, even though discrepancies up to 7 and 13 mM could be measured for glucose and ammonium respectively, at the end of the culture. The acetate profile also showed a maximum offset of about 12 mM. The standard errors of prediction for glucose, ammonium and acetate were 5.5, 5.5 and 5.4 mM respectively.

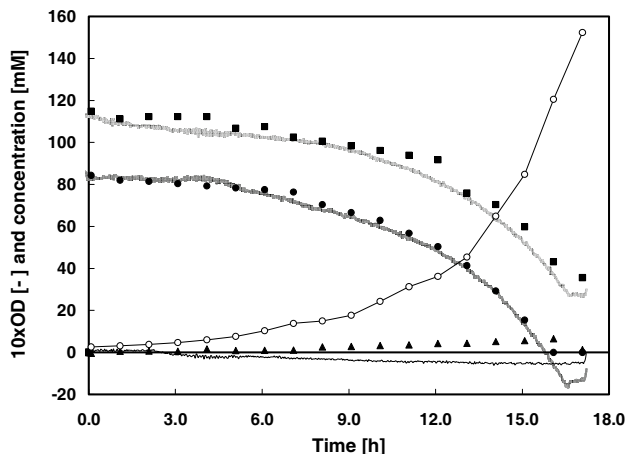


Figure 3. Glucose (black circles), ammonium (black squares) and acetate (black triangles) concentration profiles measured off-line (symbols) and on-line by FTIR (corresponding lines), during a batch culture on glucose at 20°C. The biomass concentration (multiplied by a factor 10; open circles with line) was measured off-line by optical density at 600 nm. The spectra library used to calculate the FTIR profiles was set up at 30°C.

The four other runs led to comparable results, with average SEP values of 7.0, 6.4 and 6.2 mM for glucose, ammonium and acetate respectively. These values are very similar to those obtained through a more complex chemometrics approach in truly real-time applications (Dabros et al., 2006; Kornmann et al., 2004a). Much lower SEP values have been reported, but monitoring was performed at-line, which allows referencing all samples to a background spectrum and thereby correcting for signal drift (Hashimoto et al., 2005; Roychoudhury et al., 2006a).

The specific growth rate, μ , was calculated for each experiment from the FTIR profiles, in order to investigate whether this simple approach is suitable for kinetics study. For that purpose, the natural logarithm of the glucose and ammonium concentrations was plotted against time. The specific growth rate was calculated by linear regression over the linear

range, which included at least 30 points and corresponded to the exponential growth phase (Table 1). The specific growth rates were also calculated from optical density (at 600 nm) and dry cell weight measurements. The values of μ obtained by OD₆₀₀ were systematically higher than the values obtained by dry cell weight, and therefore, these two results were averaged to give a reference value for comparison with FTIR data.

Table 1. Specific growth rate (μ) as a function of the temperature, calculated from on-line and off-line data. For on-line data, the concentration profiles of ammonium and glucose during exponential phase were used to calculate μ . For off-line data, the optical density (OD) and dry cell weight (DCW) measurements were used. The last column compares the averages of the specific growth rates measured on-line and off-line.

Temperature [°C]	FTIR data			Off-line analyses			Difference [%]
	μ_{glu} [h ⁻¹]	μ_{nh4} [h ⁻¹]	$\mu_{\text{Ave-FTIR}}$ [h ⁻¹]	μ_{OD} [h ⁻¹]	μ_{DCW} [h ⁻¹]	$\mu_{\text{Ave-Offline}}$ [h ⁻¹]	
20	0.20	0.21	0.21	0.27	0.17	0.22	6.6
25	0.38	0.39	0.38	0.46	0.39	0.42	9.1
30	0.67	0.54	0.61	0.63	0.59	0.61	0.4
36	0.73	0.67	0.70	0.74	0.68	0.71	1.2
38	0.71	0.75	0.73	0.80	0.66	0.73	0.5

Specific growth rates calculated from glucose and ammonium profiles were always within 12% of the reference, averaged off-line value. The average error over the 10 specific growth rates determined from FTIR results was 0.03 h⁻¹, or 6.6 %. In general, all values agreed well with previously published data on the same strain (Kovarova et al., 1996). These calculations show that valuable kinetic data can be obtained by mid-infrared spectroscopy, at a labour cost much lower than conventional off-line analyses, since the calibration could be performed within a couple of hours, whereas optical density and dry cell weight measurements are time-consuming.

Signal anchorage versus drift spectra

Single beam infrared instruments show large signal instabilities that considerably affect modeling (Doak and Phillips, 1999), and therefore, correcting signal is of critical importance in order to get a reliable and accurate prediction. This can easily be done off-line by referencing all samples to a reference solution that is regularly measured, or by using in-process standards to update the model (Kornmann et al., 2004b). Correcting for signal drift in real-time, without impairing the predictive ability, requires the use of specific mathematical treatments (Dabros et al., 2007). Two methods were chosen in the current study for their simplicity, namely signal anchorage (Schenk et al., 2007b) and the use of drift spectra (Schenk

depend on the process temperature, but rather on external factors such as room temperature, quality of dry purge gas in optic conduit, etc. The simultaneous use of the two methods did not lead to much better results than the correction performed using drift spectra only.

The efficiency difference between the two methods for signal drift correction is consistent with the observations made previously concerning the influence of temperature on intensity and absorbance spectra. Signal anchorage works at the intensity spectrum level, and is therefore strongly affected by temperature changes, while drift spectra are absorbance spectra, and similarly to molar absorbances, are hardly affected by temperature changes. With respect to these results, the use of signal drift to handle instabilities should be preferred to an anchorage method, even though it involves the Principal Component Analysis and is therefore slightly more complicated in terms of data treatment.

5/ CONCLUSIONS

This work has shown that a single library of pure component spectra, coupled with a method for signal drift correction, can be used to monitor by FTIR spectroscopy batch cultures of *E. coli* at different temperatures. This temperature-independent calibration approach does not involve a large calibration set, since three standards were necessary to set up the library, and remains simple in terms of data treatment, because traditional least squares were used for modeling. Nevertheless, it is as predictive and accurate as more sophisticated chemometrics methods.

This approach is based on the observation that the absorbance of glucose and ammonium is affected exactly the same as water by temperature. The large influence of temperature on infrared spectra can, therefore, be eliminated by referencing the sample spectra to a background solution kept at the same temperature. By doing so, molar absorbances of compounds such as glucose and ammonium are temperature-independent. While this observation certainly does not match the real structural rearrangements induced by temperature changes, it is, from a practical point of view, accurate enough for modeling purposes. An extension of the temperature range above 50°C or below 20°C may induce nonlinear changes and the consequent collapse of this calibration approach. However, with respect to most bioprocesses, the range 20 – 50°C seems already rather large.

It remains unclear whether the absorbance of all compounds is affected in a similar manner to water, glucose and ammonium, or if there exist species for which molar absorbance is temperature-dependent. A large investigation dedicated to that particular point would be highly relevant, in order to generalize the calibration approach developed here to other applications.

6/ REFERENCES

- Bellon-Maurel V, Vallat C, Goffinet D. 1995. Quantitative analysis of individual sugars during starch hydrolysis by FT-IR/ATR spectrometry. Part II: influence of external factors and wavelength parameters. *Appl. Spectrosc.* 49(5):563-568.
- Cozzolino D, Liu L, Cynkar WU, Damberg RG, Janik L, Colby CB, Gishen M. 2007. Effect of temperature variation on the visible and near infrared spectra of wine and the consequences on the partial least square calibrations developed to measure chemical composition. *Anal. Chim. Acta* 588(2):224-230.
- Czarnik-Matusiewicz B, Pilorz S, Hawranek JP. 2005. Temperature-dependent water structural transitions examined by near-IR and mid-IR spectra analyzed by multivariate curve resolution and two-dimensional correlation spectroscopy. *Anal. Chim. Acta* 544(1-2):15-25.
- Dabros M, Amrhein M, Gujral P, von Stockar U. 2006. On-line recalibration of spectral measurements using metabolite injections and dynamic orthogonal projection. *Appl. Spectrosc.* 61(5):507-513.
- Dabros M, Schenk J, Marison IW, von Stockar U. 2007. The ongoing quest for truly on-line bioprocess monitoring using spectroscopy. *Biotechnology and Bioengineering Research Trends*. Hauppauge (NY): Nova Science Publishers (*Submitted*).
- Doak DL, Phillips JA. 1999. In situ monitoring of an *Escherichia coli* fermentation using a diamond composition ATR probe and mid-infrared spectroscopy. *Biotechnol. Prog.* 15:529-539.
- Fayolle P, Picque D, Perret B, Latrille E, Corrieu G. 1996. Determination of major compounds of alcoholic fermentation by middle-infrared spectroscopy : study of temperature effects and calibration methods. *Appl. Spectrosc.* 50(10):1325-1330.
- Hageman JA, Westerhuis JA, Smilde AK. 2005. Temperature robust multivariate calibration: an overview of methods for dealing with temperature influences on near infrared spectra. *J. near Infrared Spec.* 13(2):53-62.
- Hashimoto A, Yamanaka A, Kanou M, Nakanishi K, Kameoka T. 2005. Simple and rapid determination of metabolite content in plant cell culture medium using an FT-IR/ATR method. *Bioprocess. Biosyst. Eng.* 27(2):115-123.
- Kornmann H, Valentinotti S, Duboc P, Marison IW, von Stockar U. 2004a. Monitoring and control of *Gluconacetobacter xylinus* fed-batch cultures using in situ mid-IR spectroscopy. *J. Biotechnol.* 113(1-3):231-245.

- Kornmann H, Valentinotti S, Marison IW, von Stockar U. 2004b. Real-time update of calibration model for better monitoring of batch processes using spectroscopy. *Biotechnol. Bioeng.* 87(5):593-601.
- Kovarova K, Zehnder AJB, Egli T. 1996. Temperature-dependent growth kinetics of *Escherichia coli* ML 30 in glucose-limited continuous cultures. *J. Biotechnol.* 178(15):4530-4539.
- Pollard DJ, Buccino R, Connors NC, Kirschner TF, Olewinski RC, Saini K, Salmon PM. 2001. Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy. *Bioprocess. Biosyst. Eng.* 24:13-24.
- Roychoudhury P, Harvey LM, McNeil B. 2006a. At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total reflectance mid-infrared (ATR-MIR) spectroscopy. *Anal. Chim. Acta* 561:218-224.
- Roychoudhury P, Harvey LM, McNeil B. 2006b. The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. *Analytica Chimica Acta* 571:159-166.
- Schenk J, Marison IW, von Stockar U. 2007a. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2):344-353.
- Schenk J, Marison IW, von Stockar U. 2007b. Simplified Fourier-transform mid-infrared calibration based on a spectra library for the on-line monitoring of bioprocesses. *Anal. Chim. Acta* 591(1):132-140.
- Schenk J, Viscasillas C, Marison IW, von Stockar U. 2007c. On-line monitoring of nine different batch cultures of *E. coli* by mid-infrared spectroscopy, using a single spectra library for calibration. *J. Biotechnol.* *J. Biotechnol.: Accepted.*
- Segtnan VH, Mevik BH, Isaksson T, Naes T. 2005. Low-cost approaches to robust temperature compensation in near-infrared calibration and prediction situations. *Appl. Spectrosc.* 59(6):816-825.
- Vojinovic V, Cabral JMS, Fonseca LP. 2006. Real-time bioprocess monitoring. Part I: In situ sensors. *Sensors and Actuators B* 114:1083-1091.
- Wülfert F, Kok WT, de Noord OE, Smilde AK. 2000. Linear techniques to correct for temperature-induced spectral variation in multivariate calibration. *Chemometr. Intell. Lab.* 51:189-200.

et al., 2007c). They have proved to handle efficiently signal instabilities, but their ability to correct for signal drift at different temperatures has never been investigated.

For the five culture runs, the standard errors of prediction for glucose and ammonium were calculated using the two different approaches to signal drift correction (Fig. 4). For comparison purposes, SEP values were also calculated using both methods simultaneously. The efficiency of signal anchorage was highly temperature-dependent, since the SEP values of glucose and ammonium were much higher at 20°C and 38°C than at 30°C. This method was originally developed for yeast cultures performed at 30°C, which explains why the SEP values go to a minimum at this temperature. Since the influence of temperature on infrared intensity is wavenumber-dependent, the position of the anchor points should be set as a function of the temperature, which considerably alters the robustness of the method.

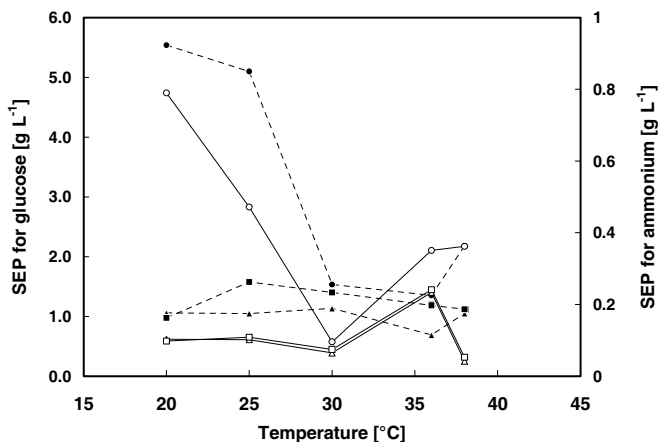


Figure 4. Standard error of prediction as a function of temperature for glucose (black symbols and dashed line; left y-axis) and for ammonium (open symbols with plain line; right y-axis). SEP values were calculated using the three different methods for signal drift correction: with signal anchorage (circles), with drift spectra (squares) and with both signal anchorage and drift spectra (triangles). The anchorage method is much more temperature-dependent than the correction based on the use of drift spectra.

The drift spectra that were included in the spectra library were also measured at 30°C, however, their ability to correct for signal instabilities was almost constant within the range 20-38°C. The signal drift was measured and analyzed by PCA at different temperatures (data not shown), and no significant variation could be observed, which explains the robustness of this approach with respect to temperature. This also suggests that signal instabilities do not

Wulfert F, Kok WT, Smilde AK. 1998. Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Anal. Chem.* 70(9):1761-1767.

7/ LIST OF SYMBOLS

Matrices and vectors

A	$s \times wn$	Absorbance matrix: s absorbances measured during the process, at wn different wavenumbers
C	$s \times j$	Matrix of concentrations: the concentrations of the j species calculated from the s spectra measured during the process
E	$s \times wn$	Error matrix: the residues found at wn wavenumbers, for the s spectra measured during the process
K	$j \times wn$	Spectra library: the molar absorbance of j species at wn different wavenumbers
SSQ_{wn}	$1 \times wn$	Sum of square residues in wavenumber dimension (vector)
SSQ_t	$s \times 1$	Sum of square residues in time dimension (vector)

Greek symbols

ε	Experimental error on absorbance [-]
κ	Molar absorptivity (or molar extinction coefficient) [M ⁻¹]
ν	Wavenumber [cm ⁻¹]

Other symbols

A	Absorbance [-]
C	Concentration [g L ⁻¹], or [mol L ⁻¹]
EV	Explained variance [%]
I	Spectrum intensity [-]
j	Number of species included in the library ($j \leq m$)
m	Number of different species involved in the process
SEP	Standard error of prediction [-]
y	Predicted or measured concentration [g L ⁻¹], or [mol L ⁻¹]

Superscripts and subscripts

<i>baseline</i>	subscript	Baseline (refers to the initial culture medium)
k	subscript	k-th sample
i	subscript	i-th species
n	subscript	Number of off-line samples
s	subscript	Number of spectra measured during the process
tot	subscript	Total
wn	subscript	Number of wavenumber measured

Chapter 7

A Simple Method to Monitor and Control Methanol Feeding of *Pichia pastoris* Fermentations using mid-IR Spectroscopy

1/ ABSTRACT

Mid-infrared FTIR spectroscopy is an efficient tool for the monitoring of bioprocesses, since it is fast and able to detect many compounds simultaneously. However, complex and time-consuming calibration procedures are still required, and have inhibited the spreading of these instruments. A simple and quick method to calibrate a FTIR instrument was developed for the control of fed-batch fermentations of the methylotrophic yeast *Pichia pastoris*. Based on the assumptions that (1) only substrate concentration may change significantly during a fed-batch process and (2) absorbance can be considered as proportional to concentration, a linear two-point calibration was implemented. Long term instability of the instrument had to be addressed in order to get accurate results: two fixed points, on both sides of substrate absorbance peak, were used to perform on-line a linear correction of the signal drift. Fed-batch experiments at constant methanol (substrate) concentration ranging from 0.8 g L⁻¹ to 15 g L⁻¹ were carried out. Off-line HPLC control analysis showed a good agreement with on-line FTIR data, with standard error of prediction values < 0.12 g L⁻¹. Even though methanol acts both as carbon source and inducer of protein expression, no significant effect was observed on the level of protein expression in the recombinant strain used.

This chapter was published in *Journal of Biotechnology*:

Schenk J, Marison IW, von Stockar U. 2007. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2):344-353.

2/ INTRODUCTION

Mid-infrared spectroscopy (MIR; 650 – 4000 cm^{-1}) relates to vibrational modes of functional groups and molecules and can potentially provide information on nearly all compounds. Although this is a clear advantage for qualitative identification of compounds, quantitative analysis remains complex and time-consuming. The latter requires the preparation of a considerable number of standards, as well as the use and understanding of advanced mathematical tools, such as Partial Least-Squares (PLS). This bottleneck has delayed the spread of infrared (IR) instruments for the monitoring of bioprocesses and there is a need for simplified and fast calibration procedures.

In addition to the ability of IR spectroscopy to detect several compounds simultaneously, many other aspects make IR instruments attractive for the monitoring of bioprocesses. It is fast and can be coupled to an Attenuated Total Reflectance (ATR) probe, made of diamond, which is therefore inert, robust, non-invasive and can be sterilized and cleaned in-situ. Fourier Transform signal processing facilitates data treatment and the overall set-up, the so-called ATR/FTIR, has become widely used for the monitoring of bioprocesses. Applications have been reported for different culture types, including *E.coli*, fungi and yeast cultivations at various scales (Crowley et al., 2000; Kornmann et al., 2003; Pollard et al., 2001). Monitoring of mammalian and plant cells cultures has also been reported (Hashimoto and Kameoka, 2000; Rhiel et al., 2002b).

All of the examples cited use of multivariate calibration (also called chemometrics). These methods are considered as appropriate for the predictive on-line analysis of complex multi-component mixtures, where absorbance peaks overlap and can be significantly shifted through interactions between species. The development of such a calibration model involves the preparation and measurement of numerous standards, followed by the use of adequate mathematical tools to extract the information, including PLS, Principal Component Analysis (PCA), Evolving Factor Analysis (EFA), etc. (Martens and Naes, 1998). Of these two parts, the first is the most time-consuming, and thus offers the biggest opportunity for improvement. Guidelines for the preparation of representative standard sets have been reported elsewhere (Brereton, 2000), and only a very short summary is presented here. The first step in the development of the calibration model consists in listing metabolites, substrates and products related to the metabolism of the cells under investigation. Secondly, a large number of solutions spanning the entire concentration range of all compounds must be prepared and analyzed. Attention must be paid at this stage to avoid correlated combinations, by the use of

appropriate multi-level experimental design (commonly four to five levels) (Rhiel et al., 2002a). According to ASTM recommendations (ASTM, 1997) on standard practices, the number of solutions to be produced to insure a reproducible calibration set should be higher than six times the number of relevant components. This whole procedure usually requires two days of labour and must generally be partially repeated to update the model to any new conditions, such as changes in the alignment of instrument optics. As a result, such an approach is particularly unsuitable for process development, during which new media and conditions (pH, temperature, and so on) are tested. Some studies have shown that smaller calibration sets can be used significantly without increasing the error of predictions, (Kornmann et al., 2004b; Navarro-Villoslada et al., 1995), although 15 standards appears to be the lower limit.

Even though FTIR spectrometers are able to monitor several compounds simultaneously, many applications and studies focus on only one. *Pichia pastoris* cultivations fall into this category and simpler as well as faster calibration procedures would be particularly appropriate. *Pichia pastoris* is able to synthesize proteins to high titres as well as to secrete and glycosylate, thereby making this organism a very interesting host for the production of recombinant proteins at both laboratory and large scale. Indeed complete kits for the expression of heterologous proteins in *Pichia pastoris*, together with process cultivation guidelines are commercially available (Invitrogen, 2002). One remarkable advantage of *Pichia* as a host for heterologous protein production is that, as eukaryote, it has an expression system designed for protein processing and folding, as well as post translational modifications. Dozens of different proteins of various origins (bacteria, fungi, plants, animal, human) have been reported to be produced, at titers up to 12 g L⁻¹ (Cereghino and Cregg, 2000). The first step in methanol (MeOH) utilization by all methylotrophic microorganisms is catalyzed by the alcohol oxidase enzyme (AOX). This enzyme has a poor affinity for oxygen - which is necessary for methanol oxidation - and is therefore expressed at high levels by the cell, typically above 30% of total soluble protein (Sreekrishna et al., 1997). This specificity and high levels of expression have been exploited to express foreign proteins in *Pichia pastoris*, through isolation of the gene and promoter for alcohol oxidase and the creation of vectors, strains, and the corresponding protocols for molecular genetic manipulation.

Pichia is usually grown in fed-batch cultivations that reach very high cell density. The general process involves three stages: Stage one, batch growth on glycerol to produce a significant amount of biomass rapidly, followed by stage two, glycerol-limited fed-batch growth in order to de-repress the AOX promoter and stage three, fed-batch growth on

methanol to induce protein expression. Since methanol induces protein expression, complete depletion is not desirable during induction phase, while care must also be taken to avoid accumulation to inhibitory levels. Some authors have found inhibitory levels from 3.7 g L⁻¹ (Cunha et al., 2004) to 20 g l⁻¹ (Stratton et al., 1998), whereas others have reported an optimum concentration with respect to specific productivity around 2 g L⁻¹ (Zhang et al., 2000) or 3.5 g L⁻¹ (Curvers et al., 2002). This influence of the residual methanol concentration on process productivity has stimulated the development of on-line monitoring techniques, which are generally based on the analysis of the fermentor exhaust gas. Specific methanol sensors are commercially available and relatively inexpensive, but require non linear calibrations and have long response times (ca. 5 min) (Cunha et al., 2004; Hellwig et al., 2001; Zhang et al., 2000). The use of at-line gas chromatography has also been reported (Curvers et al., 2002; Minning et al., 2001). Acquisition frequency is slower with this method, although more compounds may be monitored simultaneously. Fourier Transform Mid-Infrared Spectroscopy (FTIR) has also proven to be an efficient tool for off-line measurements of methanol concentration in *Pichia* cultivations (Crowley et al., 2000), although there are no reports of the technique having been used for on-line process monitoring.

In this article we describe a simple calibration procedure for an ATR/FTIR that was developed for control of the residual methanol concentration during *Pichia pastoris* cultivations. Based on the assumptions that only the methanol concentration changes significantly during the time-course of the methanol induction phase and that absorbance is linear with respect to concentration, a linear two-point calibration was implemented. The critical influence of instrument baseline stability on the precision of the results is also discussed and a method to address this problem presented.

3/ MATERIALS AND METHODS

Strain and Inoculum Preparation

The construction of a recombinant avidin (with lowered isoelectric point, pI 5.4) secreting Mut⁺ strain of *Pichia pastoris* has been described in a previous paper (Zocchi et al., 2003). A 30-vial cell bank was used for seeding all cultivations. The bank was prepared by inoculation of a flask containing 100 ml of YPG medium (6 g L⁻¹ yeast extract, 5 g L⁻¹ peptone, 20 g L⁻¹ glycerol) and incubation at 30°C. After 24 h the culture was centrifuged 10 min at 1500 g, then resuspended in 36 ml of saline glycerol solution (9 g L⁻¹ NaCl, 20 g L⁻¹ glycerol), followed by aliquoting in 2 mL Nalgene tubes (Rochester, NY) and stored at -80°C. For each cultivation, an inoculum was prepared from a new vial from the cell bank, grown overnight in 100 mL of YPG medium. Cells were then centrifuged (10 min at 1500 g) and resuspended in 10 mL of UHP water for fermentor inoculation.

Bioreactor Cultivations

Medium composition is based on the standard *Pichia* basal salts medium, to which EDTA was supplemented to avoid precipitation. No carbon source was added initially, since this latter is supplied during the calibration procedure (26.7 mL L⁻¹ H₃PO₄ 85% in weight, 0.93 g L⁻¹ CaSO₄·2H₂O, 18.2 g L⁻¹ K₂SO₄, 14.9 g L⁻¹ MgSO₄·7H₂O, 4.13 g L⁻¹ KOH, 0.8 g L⁻¹ EDTA·Na₂·2H₂O, NH₄OH 28% in weight to adjust pH at a value of 5.0 or 3.0). After filter-sterilization (0.2 µm, Steritop®, Millipore, MA), 0.4 mL of antifoam (Struktol SB2121, Schill and Seilacher, Hamburg, Germany), 0.125 µg of biotin (0.125 µL of 1 g L⁻¹ biotin stock solution in 1 mol L⁻¹ NaOH) and 4.35 mL of sterile trace elements solutions were added (PTM: 5 mL L⁻¹ H₂SO₄ 98% in weight, 0.02 g L⁻¹ H₃BO₃, 6 g L⁻¹ CuSO₄·5H₂O, 0.08 g L⁻¹ NaI, 3 g L⁻¹ MnSO₄·H₂O, 0.2 g L⁻¹ Na₂MoO₄·2H₂O, 0.5 g L⁻¹ Ca₂SO₄·2H₂O, 20 g L⁻¹ ZnCl₂, 65 g L⁻¹ FeSO₄·7H₂O). The medium was weighed before and after introduction into the reactor. Specific density was also measured (1.035 ± 0.02 kg L⁻¹) in order to calculate accurately the initial medium volume.

Cultivations were carried out at 30°C in a fully automated 2-Litre bioreactor (RC1, Mettler-Toledo, Greifensee, Switzerland). pH was controlled at a value of 5.0 (or at 3.0 in one experiment) by the addition of NH₄OH 28% in weight. Agitation was set at 1000 rpm, with two 6-blade Rushton impellers, giving a power input of more than 10 W L⁻¹ (torque measurement). Total aeration was controlled to provide 2 NL min⁻¹, and dissolved oxygen was maintained above 80% air saturation by blending air and pure oxygen using thermal mass flow controllers (MFC 5850E, Brooks, Veenendaal, The Netherlands). The maximum

cooling capacity of the reactor was approximately 140 W L^{-1} . Oxygen and carbon dioxide composition of the exhaust gas was measured using a gas analyzer (Dr. Marino Müller Systems, Esslingen, Switzerland).

Calibration of the FTIR instrument was performed within 30 min after bioreactor inoculation. The amount of methanol added for this step was estimated to be as close as possible to the desired set point of the experiment. Closed-loop control of the methanol concentration based on FTIR monitoring was then started and maintained as long as the cooling power enabled the temperature within the reactor being controlled at 30°C (for 44 to 58 h).

FTIR Set-up and Methanol Controller

A Mettler Toledo FTIR 4000 (Mettler-Toledo, Greifensee, Switzerland) with an Attenuated Total Reflectance probe (ATR, DiComp) was mounted in a flow-through cell to monitor the culture broth (Fig. 1). This configuration is identical to the in-situ ATR-probe device, but is more convenient for small reactors and allows linkage to several reactors. A recirculation loop driven by a membrane pump (Prominent Gamma/L, Heidelberg, Germany) was used to connect the reactor to the FTIR measurement cell. The residence time within the loop was less than 20 seconds, and the response time was 15 seconds. Spectra were acquired every 4 min (2 min at low methanol concentrations), and then routinely treated using Matlab (Mathworks, Natick, MA) to calculate the methanol concentration. The values obtained were then sent by tcp/ip to the automated controller of the reactor (Fieldpoint, National Instrument, Austin, TX) and the feeding rate adjusted using a purpose-built PI controller running under Labview RT (National Instrument, Austin, TX). A reservoir containing feeding solution (pure methanol, supplemented with 12 ml l⁻¹ of PTM solution) was placed on a balance in order to calculate the feeding rate.

Off-line analysis

Samples (5 mL) were collected every 2 hours and immediately cooled to 4°C using a purpose-built automatic device (Cannizzaro, 2002). Sample treatment was carried out within 12 hours after sampling. Dry cell weight concentration (g L^{-1}) was determined by gravimetric analysis: 3 ml of sample were filtered on $0.2 \mu\text{m}$ pre-weighed filters, and then dried at 100°C to constant weight. At biomass concentrations higher than 10 g L^{-1} this analysis was performed as follows: 3 mL of sample were centrifuged for 30 minutes at 1200 g in a pre-

weighed glass tube, the supernatant was discarded and biomass dried to constant weight at 100°C.

The remaining volume of the sample was filtered through 0.2 µm filters and used for the determination of recombinant avidin and HPLC analysis (1100 series, Agilent, Palo Alto, CA). Methanol concentration was determined by ion-exchange chromatography (Supelcogel H 300 mm, Supelco, Bellefonte, PA) by isocratic elution at 0.6 mL min⁻¹ with a solution of 5 mmol L⁻¹ sulfuric acid. The detection limit was 0.05 g L⁻¹. Recombinant avidin concentration was determined by concentration (2 to 4 times) of samples by ultrafiltration (Centriplus 30kDa, Millipore, Billerica, MA) followed by titration with biotin-4-fluorescein (Fluka, Buchs, Switzerland), according to the published procedure (Kada et al., 1999).

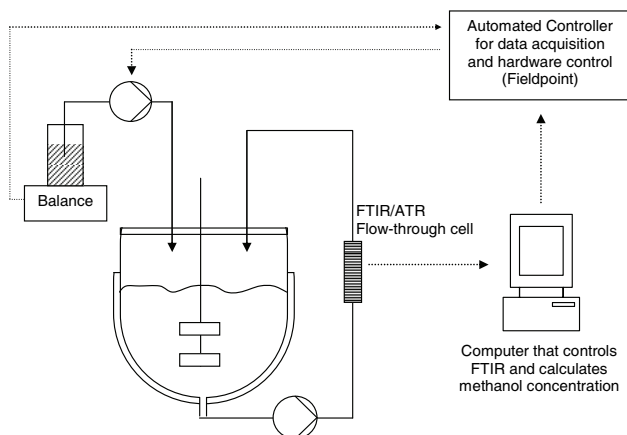


Figure 1. Implementation of closed-loop controller of methanol concentration. Culture broth is recirculated to an external loop through a non invasive FTIR/ATR cell. Spectra are taken every 4 minutes and the methanol concentration calculated and sent by a computer to a remote automated controller that adjusts the feeding rate.

Data consistency

A check for data consistency was carried out for every run, using the standard procedure (Stephanopoulos et al., 1998). Methanol, carbon dioxide and biomass yields were used for the reconciliation, with an error range of 3% on each. All runs passed the statistical test with a confidence interval set at 95% (redundancy = 1, threshold $\chi^2 = 3.84$).

Biomass composition was determined by elemental analysis to be $\text{CH}_{1.69}\text{O}_{0.50}\text{N}_{0.18}$. Methanol losses by stripping were measured for each experiment during the lag phase of the

run. Methanol added during the first 7 hours after inoculation was assumed to compensate for stripping only.

4/ FTIR IN-SITU CALIBRATION

Linear calibration

Even though complex calibration procedures are efficient for the simultaneous measurement of several compounds, more straightforward approaches are desirable for less complex cases. A simple method has been developed for such a “basic” process and is based on two assumptions, namely that only one concentration varies significantly during the time-course of the process and that absorbance is linearly related to concentration.

The first assumption, that only one concentration varies significantly during the process, can be made for certain microorganisms under certain conditions. *Pichia pastoris* cultivations come under this category, since during induction phase, they consume methanol as carbon source, ammonium hydroxide as nitrogen source and produce biomass and recombinant protein. Gases are not taken into account, since both CO₂ and O₂ have undetectable (by FTIR) concentrations in liquid phase (around 12 and 7 mg L⁻¹, respectively). Theoretically this makes four components to be measured, however biomass shows no IR signal with an ATR probe (data not shown), ammonium hydroxide concentration remains rather constant, since ammonia is added to control pH, and recombinant protein titers are usually too low to be detected (< 200 mg L⁻¹). Methanol can therefore be considered as the only measurable compound, which considerably simplifies the initial situation. Even though *Pichia pastoris* is a special case, this approach can be generalized to other organisms, for which no overflow metabolite is produced.

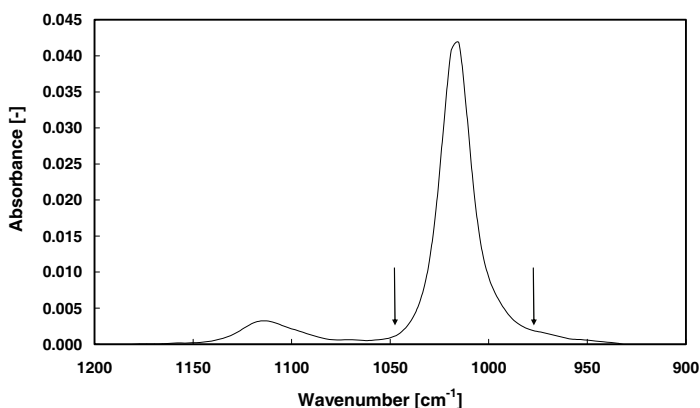


Figure 2. Absorbance of a 12 g L⁻¹ methanol aqueous solution. The arrows represent the anchorage points for drift correction.

Methanol aqueous solutions exhibit a sharp peak on IR spectra at 1018 cm^{-1} , due to C-O stretching of the alcohol functional group (Fig. 2). The relationship between absorbance and concentration is not necessarily expected to follow the Lambert-Beer law strictly, since methanol is known to form several water complexes in aqueous solution (Puxty et al., 2005), which show different extinction coefficients. Consequently the overall methanol absorbance must be calculated as a sum of all contributions, which depends on the equilibrium constants of different complexes. However, results show that considerable accuracy ($R^2 = 0.9996$) can be achieved by assuming linearity (data not shown).

Even for a single component, four or five standards covering the entire concentration range are often used for calibration. This approach, commonly applied to IR spectroscopy but also for instance to GC or HPLC methods, provides information on the linearity of the response. However, if it is known that the response should be linear, or that a linear regression is considered as appropriate, for the same amount of work, a two-level design is more efficient. Widely spaced points have a stronger leverage, and repeated measurements enable estimating the reproducibility of the method. Once reproducibility of the technique has been determined, it is subsequently necessary for only two standards to be measured (one at each level), which further decreases the work load.

In-situ calibration procedure

A linear two-point calibration can be carried out directly in-situ, thereby eliminating the need for the preparation of standards. Apart from saving time, this method can correct for changes in alignment of optics from run to run. The first point of the calibration is the baseline. Medium is prepared, without carbon source, and placed under culture conditions (stirring, aeration, temperature, pH). The medium volume, denominated V_R , must be accurately determined gravimetrically and the specific density determined. The reactor is then inoculated and a first spectrum, namely the baseline, is taken and signal intensity at 1018 cm^{-1} extracted (I_0). The second step consists in the addition of a known amount of the substrate, m_s , to give a concentration $C_s = m_s/V_R$ close to the desired set point. A second spectrum is then taken and the signal intensity at 1018 cm^{-1} (I_s) determined.

The Lambert-Beer law can be written using a constant K that includes both the extinction coefficient and the length of the measurement cell:

$$A = \log_{10} \frac{I_0}{I} = K \cdot C \quad [1]$$

where A is the absorbance, I the signal intensity and C a concentration. By setting the baseline spectrum as the reference spectrum ($I_B = I_0$), the absorbance of both points can be calculated as:

$$A_B = \log_{10} \frac{I_0}{I_B} = \log_{10} \frac{I_B}{I_B} = 0 \quad [2]$$

$$A_S = \log_{10} \frac{I_0}{I_S} = \log_{10} \frac{I_B}{I_S} \quad [3]$$

The calibration coefficient K is thus calculated as:

$$K = \frac{A_S}{C_S} \quad [4]$$

Signal Drift Correction

Long-term instability of FTIR signals has been described in detail (Doak and Phillips, 1999), but has received little attention in papers reporting the use of these instruments for the monitoring of bioprocesses. However, this is a critical issue, particularly due to the long duration of cultivations. The accuracy needed for such applications is typically of the order of 0.2 g L^{-1} , or in other words 0.02% (200 ppm). To address the long term stability of the signal, spectra of UHP water were collected over 4 days at a frequency of 30 min in the sealed and thermostated FTIR flow-through cell. Under these conditions drifts of up to 3% were measured at 1018 cm^{-1} , which clearly illustrates the need for baseline correction.

During online operation signal drift and noise correction is not trivial, since no reference standard, such as water, can be measured during the time-course of the process. To make things even worse, baseline variations are wavenumber-dependant. However, as described previously, no significant change in the medium composition is expected to occur, with the exception of methanol, during the fed-batch *Pichia* cultivation. Signal intensity at wavenumbers around the methanol absorbance peaks can therefore be used to correct the signal drift within this interval. These so-called anchor points have to be set as close as possible from the peak of interest to get a reliable drift correction, but far enough to be unaffected by the methanol absorbance. A compromise has therefore to be found, and they were set at 984 and 1042 cm^{-1} (Fig. 2), on each side of the main methanol peak (1018 cm^{-1}).

The anchor points were assumed to be unaffected by any change in the culture broth, and consequently any deviation from the initial value was attributed to signal drift. For each of these points, the ratio of the new signal (subscript t) to the initial signal (subscript 0) was calculated, and the ratios were averaged to give a global correction coefficient L_{drift} (Eq. 5-7):

$$L_{984} = I_{t,984} / I_{o,984} \quad [5]$$

$$L_{1042} = I_{t,1042} / I_{o,1042} \quad [6]$$

$$L_{drift} = 0.5 \cdot (L_{984} + L_{1042}) \quad [7]$$

The entire spectrum was then linearly reprocessed (superscript star), by dividing the signal at every wavenumber ν by the correction coefficient L_{drift} :

$$I_{t,\nu}^* = I_{t,\nu} / L_{drift}$$

5/ RESULTS AND DISCUSSION

Methanol Control for Fed-Batch Cultivations

Six fed-batch cultivations were conducted with control of residual methanol concentration at levels ranging from 0.8 to 15 g L⁻¹. These so-called nutrstatic experiments lasted for 40 to 60 hours, depending on the growth rate, and had to be stopped when the maximum cooling capacity of the reactor was reached (140 W). Cells were directly grown on methanol, without prior batch phase on glycerol, in order to avoid the delay caused by the adaptation to substrate change. An example of such a run is presented in Fig. 3, for which the methanol concentration was controlled at 6 g L⁻¹. Since substrate was limiting and never depleted, growth was exponential and reached 70 g L⁻¹ (grams of Dry Cell Weight per litre). Off-line HPLC control analysis showed a good agreement with on-line FTIR methanol measurements, considering an error range of ± 0.2 g L⁻¹ for both instruments. Methanol control was turned off after 42.5 hours, and the remaining methanol was completely consumed within 20 minutes. FTIR monitoring showed a final offset of + 0.2 g L⁻¹ at the end of the experiment, which may be due to accumulation of metabolic by-products during the cultivation. If such products were formed the first assumption made for calibration was invalid.

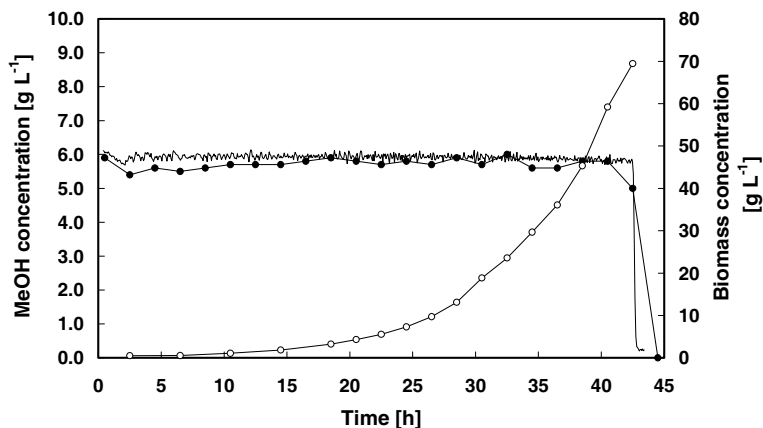


Figure 3. Fed-batch fermentation with control of residual methanol concentration at 6 g L⁻¹. Methanol concentration is shown on the left axis, as measured by on-line FTIR (plain line) and off-line HPLC (black dots). Biomass concentration (white circles) is shown on the right axis.

The residual methanol concentration profiles of the other fed-batch cultivations are represented in Fig. 4. All of them were obtained on-line, and were not reprocessed in any way after the culture. The run, at a set point of 4 g L⁻¹ methanol, was conducted at a pH of 3, instead of 5 as in all the others. An interesting aspect of this simple monitoring technique is that various culture conditions can be explored without leading to a significant increase in the labour needed for calibration. Standard error of prediction (SEP) values were calculated for all cultivations and were below 0.12 g L⁻¹, which is similar to values reported using multivariate calibration (Kornmann et al., 2004a; Pollard et al., 2001; Rhiel et al., 2002b).

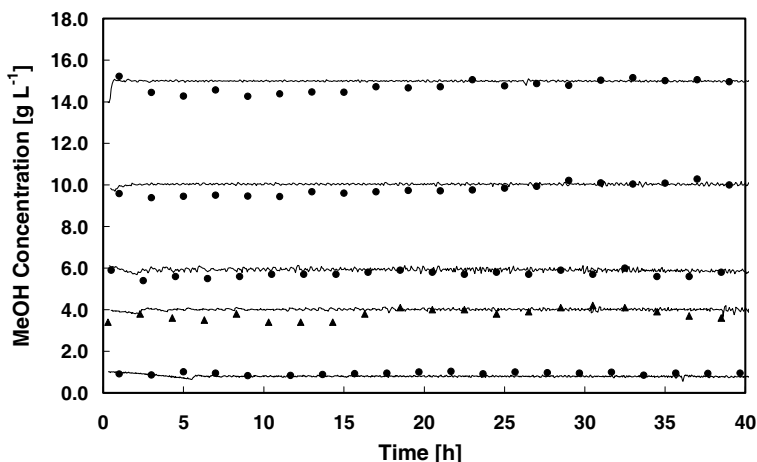


Figure 4. Methanol profiles for five fed-batch experiments with control of residual methanol concentration. FTIR on-line measurements are indicated by lines and HPLC off-line analysis by symbols. All runs were carried out at pH 5 (circles), except one at pH 3, with a residual methanol concentration of 4 g L⁻¹ (triangles).

Baseline stability was of critical importance for accurate monitoring. FTIR data were reprocessed without signal drift correction and compared to on-line data (Fig. 5). Long-term signal drift led to differences of up to 6 g L⁻¹ for the set point of 10 g L⁻¹ (overall SEP > 14 g L⁻¹). The daily additions of liquid nitrogen to the cooling device of the infrared detector were found to be followed by a short period of high instability. The proposed method of signal anchorage and linear correction was therefore absolutely necessary, and even though very simple, sufficient to achieve small SEP values.

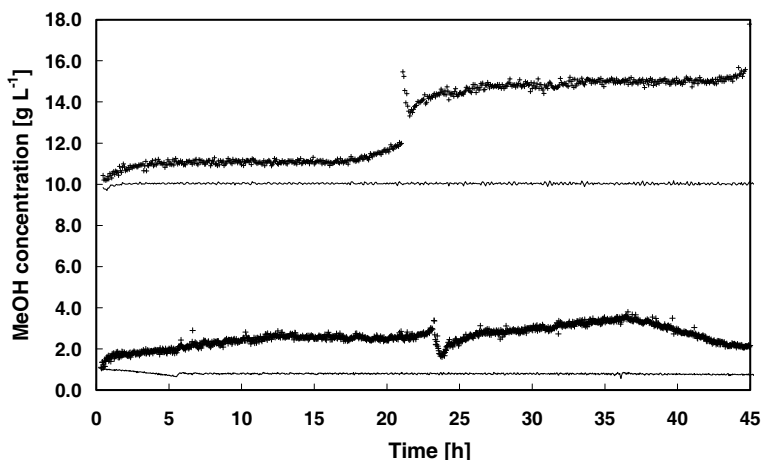


Figure 5. Residual methanol concentration profiles measured by FTIR for two fed-batch cultivations at 0.8 and 10 g L⁻¹. Lines indicate on-line data with drift correction, whereas crosses represent data without correction of signal drift by the anchorage method. Sudden changes after 22 and 24 hours are due to the daily addition of liquid nitrogen to the cooling system of the infrared detector.

Effect on Growth and Productivity

Growth rate (μ) and avidin specific productivity (q_p) were determined for all runs at pH 5, and are reported in Fig. 6. Growth rate is at its maximum (0.139 h⁻¹) at methanol concentrations between 1 to 6 g L⁻¹. Above 6 g L⁻¹ methanol has an inhibitory effect, although a concentration of 15 g L⁻¹ reduces the specific growth rate by only 30%. A residual methanol concentration of 0.8 g L⁻¹ shows a growth rate (0.124 h⁻¹) close to μ_{max} , which is in good agreement with previously published kinetic data ($\mu_{max} = 0.14$ h⁻¹; Monod constant = 0.1 g L⁻¹) (Jahic et al., 2002; Jahic et al., 2003).

Specific productivity of recombinant proteins has been reported to be related to residual methanol concentration. Since methanol acts both as carbon source and as inducer of promoter, it is generally considered that a methanol concentration above 1 g L⁻¹ enhances protein expression significantly (Curvers et al., 2002; Stratton et al., 1998; Zhang et al., 2000). Our results show that the specific productivity was relatively constant for residual methanol concentrations ranging from 0.8 to 15 g L⁻¹ (Fig. 6). A control run was carried out at a growth rate of 0.024 h⁻¹, corresponding to a very low residual methanol concentration (< 0.05 g L⁻¹), using a predefined exponential feed, and an avidin specific productivity of 24.4 μ g g⁻¹ h⁻¹ was

measured. It can therefore be concluded that the residual methanol concentration does not significantly influence the protein expression in the strain used in this work.

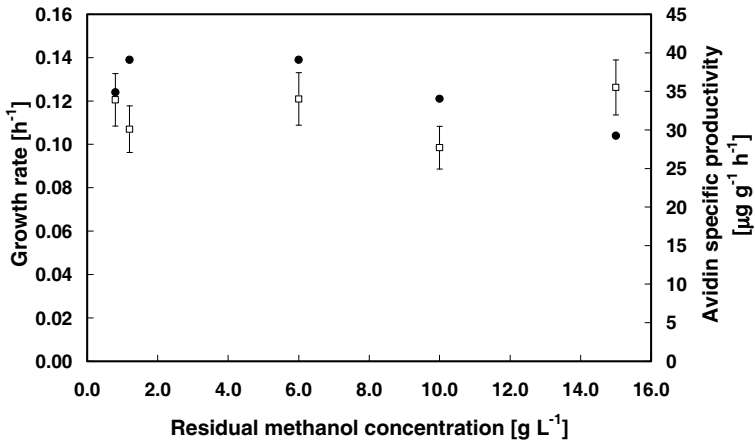


Figure 6. Growth rate (black dots, left axis) and avidin specific productivity of recombinant protein (open squares, right axis) at different residual methanol concentrations. No significant variation of specific productivity was observed.

6/ CONCLUSIONS

It has been shown in this study that in some cases, such as *Pichia pastoris* cultivations, multivariate calibration can be replaced by an easier approach based on the measurement of a single compound of interest. Also, by making the assumption that absorbance was linear with respect to concentration (i.e. Lambert-Beer law was followed), the calibration procedure could be considerably simplified, without significantly decreasing the accuracy of the prediction. Since only two measurements were required to calibrate the FTIR, it was then possible to perform the calibration in-situ, at the beginning of the cultivation, using spectra of culture medium with and without carbon source (here methanol). Even though some information was lost by making these simplifications, off-line HPLC control analysis confirmed that methanol was accurately controlled during six cultivations, with residual concentrations controlled within the range 0.8 – 15 g L⁻¹. SEP values below 0.12 g L⁻¹ were obtained for all runs, and time dedicated to the calibration procedure reduced from about two days to less than an hour.

Long-term baseline instability turned out to be a critical parameter influencing the accuracy of the monitoring. Two fixed points, so-called anchors, were set in the spectra on each side of the methanol peak, to allow a linear correction of the signal. Off-line reprocessing of data showed that without this pre-treatment of spectra, differences of up to 6 g L⁻¹ (overall SEP > 14 g L⁻¹) would have been measured.

The influence of the residual methanol concentration on the growth rate and specific productivity was studied for our current avidin-producing strain. It was found that methanol concentration inhibited growth rate at concentrations higher than 6 g L⁻¹, but no significant effect on protein expression could be observed.

The proposed method for FTIR calibration, even though developed for a very special case, may be applied to other microorganisms and/or bioprocesses. More generally, the use of linear calibrations and related two-level experimental designs should be explored to make the use of FTIR spectroscopy easier and help the spread of these instruments.

7/ REFERENCES

- ASTM. 1997. Standard Practices for Infrared, Multivariate, Quantitative Analysis (E 1655-97). Annual book of ASTM standards. Philadelphia: American Society for Testing and Materials. 844-869.
- Brereton RG. 2000. Introduction to multivariate calibration in analytical chemistry. The Analyst. 125:2125-2154.
- Cannizzaro C. 2002. Spectroscopic monitoring of bioprocesses: a study of carotenoid production by *Phaffia rhodozyma* yeast [PhD]: EPFL. 294 p.
- Cereghino JL, Cregg JM. 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol. Rev. 24(1):45-66.
- Crowley J, McCarthy B, Nunn NS, Harvey LM, McNeil B. 2000. Monitoring a recombinant *Pichia pastoris* fed batch process using Fourier transform mid-infrared spectroscopy (FT-MIRS). Biotechnol. Lett. 22:1907-1912.
- Cunha AE, Clemente JJ, Gomes R, Pinto F, Thomaz M, Miranda S, Pinto F, Moosmayer D, PDonner P, Carrondo MJT. 2004. Methanol Induction Optimization for scFv Antibody Fragment Production in *Pichia pastoris*. Biotechnol. Bioeng. 86(4):458-467.
- Curvers S, Linnemann J, Wandrey C, Takors R. 2002. Recombinant Protein Production with *Pichia pastoris* in Continuous Fermentation - Kinetic Analysis of Growth and Product Formation. Eng. Life Sci. 2(8):229-235.
- Doak DL, Phillips JA. 1999. In Situ Monitoring of an *Escherichia coli* Fermentation using a Diamond Composition ATR Probe and Mid-infrared Spectroscopy. Biotechnol. Prog. 15:529-539.
- Hashimoto A, Kameoka T. 2000. Mid-infrared spectroscopic determination of sugar contents in plant-cell culture media using an ATR method. Appl. Spectrosc. 54(7):1005-1011.
- Hellwig S, Emde F, Raven NPG, Henke M, van der Logt P, Fischer R. 2001. Analysis of single-chain antibody production in *Pichia pastoris* using on-line methanol control in fed-batch and mixed-feed fermentations. Biotechnol. Bioeng. 74(4):344-352.
- Invitrogen. 2002. *Pichia* Fermentation Process Guidelines. Carlsbad (California).
- Jahic M, Rotticci-Mulder JC, Martinelle M, Hult K, Enfors SO. 2002. Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. Bioprocess. Biosyst. Eng. 24:385-393.

- Jahic M, Wallberg F, Bollock M, Garcia P, S-O. E. 2003. Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures. *Microb. Cell. Fact.* 2:6.
- Kada G, Falk H, Gruber HJ. 1999. Accurate measurement of avidin and streptavidin in crude biofluids with a new, optimized biotin-fluorescein conjugate. *Biochim. Biophys. Acta* 1427:33-43.
- Kornmann H, Rhiel M, Cannizzaro C, Marison I, von stockar U. 2003. Methodology for real-time, multi-analyte monitoring of fermentations using an in-situ mid-infrared sensor. *Biotechnol. Bioeng.* 82(6):702-709.
- Kornmann H, Valentinotti S, Duboc P, Marison IW, von Stockar U. 2004a. Monitoring and control of *Gluconacetobacter xylinus* fed-batch cultures using in situ mid-IR spectroscopy. *J. Biotechnol.* 113(1-3):231-245.
- Kornmann H, Valentinotti S, Marison IW, von Stockar U. 2004b. Real-time update of calibration model for better monitoring of batch processes using spectroscopy. *Biotechnol. Bioeng.* 87(5):593-601.
- Martens H, Naes N. 1998. *Multivariate Calibration*. New York: John Wiley & Sons.
- Minning S, Serrano A, Ferrer P, Solá C, Schmid RD, Valero F. 2001. Optimization of the high-level production of *Rhizopus oryzae* lipase in *Pichia pastoris*. *J. Biotechnol.* 86(1):59-70.
- Navarro-Villoslada F, Perez-Arribas LV, Leon-Gonzalez ME, Polo-Diez LM. 1995. Selection of Calibration Mixtures and Wavelengths for Different Multivariate Calibration Methods. *Anal. Chim. Acta* 313(1-2):93-101.
- Pollard D, Buccino R, Connors N, Kirschner T, Olewinski R, Saini K, Salmon P. 2001. Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy. *Bioprocess. Biosyst. Eng.* 24:13-24.
- Puxty G, Maeder M, Radack KP, Gemperline PJ. 2005. Equilibrium Modeling of Mixtures of Methanol and Water. *Appl. Spectrosc.* 59(3):329-334.
- Rhiel MH, Amrhein MI, Marison IW, von Stockar U. 2002a. The Influence of Correlated Calibration Samples on the Prediction Performance of Multivariate Models Based on Mid-Infrared Spectra of Animal Cell Cultures. *Anal. Chem.* 74(20):5227-5236.
- Rhiel MH, Ducommun P, Bolzonella I, Marison IW, von Stockar U. 2002b. Real-Time In Situ Monitoring of Freely Suspended and Immobilized Cell Cultures Based on Mid-Infrared Spectroscopic Measurements. *Biotechnol. Bioeng.* 77(2):174-185.

- Sreekrishna K, Brankamp RC, Kropp KE, Blankenship DT, Tsay JT, Smith PL, Wierschke JD, Subramaniam A, Birkenberger LA. 1997. Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* 190(1):55-62.
- Stephanopoulos GN, Aristidou AA, Nielsden J. 1998. *Metabolic engineering: principles and methodologies*. San Diego: Academic Press.
- Stratton, J., Chiruvolu, V., Meagher, M., 1998. High cell-density fermentation. In: Higgins, D.R., Cregg, J.M. (Eds), *Methods in Molecular Biology*. Totowa, New Jersey: Humana Press. 107-120.
- Zhang W, Bevins MA, Plantz BA, Smith LA, Meagher MM. 2000. Modeling *Pichia pastoris* growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A. *Biotechnol. Bioeng.* 70(1):1-8.
- Zocchi A, Jobe A-M, Neuhaus J-M, Ward TR. 2003. Expression and purification of a recombinant avidin with a lowered isoelectric point in *Pichia pastoris*. *Protein Expr. Purif.* 32(2):167-174.

8/ LIST OF SYMBOLS

Latin symbols

A	Absorbance [-]
C	Concentration [g L ⁻¹], or [mol L ⁻¹]
I	Intensity of infrared spectrum [-]
K	Proportionality constant [M ⁻¹]
L	Correction coefficient for drift [-]
m	Quantity [m]
q_p	Avidin specific productivity [μg g ⁻¹ h ⁻¹]
t	Time [h]
V_R	Broth volume [L]
X	Biomass concentration (in dry cell weight) [g L ⁻¹]

Greek symbols

μ	Specific growth rate [h ⁻¹]
-------	---

Superscripts and subscripts

$*$	superscript	Corrected intensity
B	subscript	Baseline
s	subscript	Standard
o	subscript	Reference spectrum

Chapter 8

Conclusions & Perspectives

1/ THREE WAYS OF CONCLUDING

Three conclusions to this work are presented here. A first “horizontal conclusion” goes through the different chapters, and reviews what has been studied and shown in each of them studied. A second conclusion, called “vertical conclusion”, aims at discussing this thesis as a whole, and in particular at highlighting the novelty of this work and the main achievements that were delivered. Lastly, there is a more personal conclusion, under the form of “the ten golden rules of FTIR monitoring”.

2/ A HORIZONTAL CONCLUSION

The goal of this thesis was to develop simple calibration methods for the on-line monitoring and control of bioprocesses by mid-infrared Fourier-transform spectroscopy (FTIR)

The traditional chemometrics approach is commonly regarded as very time-consuming and complex, thereby making monitoring by FTIR not suitable for either research and development or process control in an industrial environment. The recommended calibration procedure requires the tedious preparation of a high numbers of standards and involves, for modeling, advanced mathematical tools such as Principal Components Regression (PCR) or Partial Least Squares (PLS). Work load and complexity of data treatment are well-identified roadblocks for the widespread of these instruments, and a motivation for this thesis was to develop calibration methods that require less time than off-line analyses, and that are accessible to non-experts in chemometrics.

The introduction (chapter 1 and 2) discussed under which conditions a 5-level design of the calibration set can be reduced to a library of pure component spectra. Multi-level designs are appropriate when the variables are expected to display a non-linear response and/or interact with each other. If a linear response is expected (or assumed to be sufficiently accurate) and all variables are independent, such a multi-level design can be reduced to a 2-level design. This reduction of the number of levels has an enormous impact on the size of the calibration set, since this latter should be proportional to v^n , where v is the number of variables and n is the number of levels. A large calibration set increases the relevant information to noise ratio. A reduction of the number of different standards should therefore be compensated by the measurement of replicates, or with some other methods to correct for noise interference.

Chapter three aimed at verifying if the hypotheses made on linearity and variables independence could be validated, taking *Saccharomyces cerevisiae* cultures as a case study. Preliminary experiments, made off-line with aseptic solutions, showed that in the fingerprint region ($900 - 1800 \text{ cm}^{-1}$), the Lambert-Beer law is strictly followed for common compounds found in culture media, and that absorbance spectra can be added linearly, without interfering with each others. This demonstrated that a library of pure component spectra can be used as a calibration set, as long as it is coupled with a method to address signal instabilities. Concerning ionic species, such as ammonium and acetate, the importance of the counter-ions for the spectra library was also discussed, and it was shown that sodium and

chloride salts should be preferentially selected. Most compounds found in culture media, e.g. phosphate buffer, trace elements, vitamins, magnesium, etc. exhibit small concentration changes. The influence on these compounds can be eliminated by subtraction, by referencing all process spectra to the first spectrum taken during the process, before inoculation. The calculated concentrations are then expressed as concentration differences from the initial state, but absolute concentrations can easily be found by adding the initial, known, concentrations of the culture medium. The simple least squares modeling, using a library of 5 pure component spectra, as shown to be as accurate as a PCR model based on a 49-standard calibration set. Standard errors of prediction were 4.8, 21.0 and 8.4 mM for glucose, ethanol and ammonium respectively. A spiking experiment demonstrated that such an approach is not correlated to the cell metabolism and is consequently truly predictive.

Chapter four took, as an example, nine different batch cultures of *Escherichia coli*, to highlight another aspect of library-base calibrations: including into an intensity background all the compounds that are not relevant, in terms of monitoring, allows a use of the same spectra library for different bioprocesses. Using a single library of six spectra as calibration set, the concentration profiles of seven of the cultures could be monitored with an accuracy similar to what could be expected from a PCR approach, which would have required hundreds of standards. The two other cultures showed higher error of predictions, but this relative failure could be detected by analyzing model residuals. The standard error of predictions averaged over the nine experiments were 8.0, 12.3, 5.9 and 5.6 mM for glucose, glycerol, ammonium and acetate respectively. In this chapter, another approach to signal drift correction was also presented. Instead of using signal anchorage, as in chapter 3, 5 and 7, principal features of the signal distortion were included as “drift spectra” in the calibration library. These contributions were found by applying Principal Component Analysis to the absorbance spectra of aseptic water recorded during 24 hours. The calibration library therefore not only consisted in pure absorbance of the main metabolites, but also in two drifts spectra. The least squares algorithm could play with two drift contributions in order to model process spectra, which led, for each batch culture, to a profile of “drift concentration” that could be used to detect outliers.

Chapter 5 focused on the influence of pH on FTIR calibrations. Experiments carried out off-line, which enables referencing samples to a water reference and by that means eliminating noise, proved that the pH has no direct influence on the infrared absorbance of compounds. Variations observed in the infrared absorbance of solutions submitted to a pH shift only reflect changes in thermodynamic equilibria of weak acids and bases. Additivity

and linearity of molar absorbances remain valid, and the library-based calibration concept developed in the previous chapters could be extended so to include pH monitoring. The pH values could be predicted by measuring the concentration ratio of the protonated to deprotonated forms of any acid. Since culture media are far from being ideal solutions, the dissociation equilibrium constant at infinite dilution (K_a) could not be used per se. Instead of determining experimentally apparent dissociation equilibrium constants, the Debye-Hückel theory was used to predict activity coefficients, as a function of the ionic strength. This approach turned out to be quite accurate, since the pH fall caused by ammonium consumption during and *E. coli* culture could be measured with a standard error of prediction of 0.12 pH units. The technique was shown to be fast enough to develop a feedback controller for the pH, based on the infrared measurements.

Chapter 6 had for subject the effect of temperature on mid-infrared spectra. It was shown that the temperature has a very large influence on intensity spectra. However, this influence can be eliminated by referencing the samples to a background intensity measured at the same temperature. By this means, the molar absorbance of common compounds, such as glucose and ammonium chloride, were demonstrated to be independent of temperature in the fingerprint region. The same spectra library was used to monitor 5 batch cultures of *E. coli* carried out at 20, 25, 30, 36 and 38°C. The average standard errors of prediction over the 5 runs were 7.0 and 6.4 mmol L⁻¹ for glucose and ammonium respectively, which are very comparable to the results displayed in chapter 3. The robustness of the two methods for signal drift correction was discussed, and it was shown that the approach based on drift spectra worked efficiently within the range 20 – 38 °C, whereas the anchorage method was highly sensitive to temperature.

Chapter 7 presented another simple approach to FTIR calibration. This method did not imply the use of a spectra library, but also assumed that the Lambert-Beer is followed and that species do not interact with each other. This chapter focused on a particular application, in which the residual concentration of the substrate must be controlled at a certain level during the fed-batch phase of a process. The studied case is the production of recombinant protein in a Mut⁺ strain of *Pichia pastoris*. In order to further simplify the calibration task, it was assumed that during the fed-batch phase, only the substrate concentration may produce changes in the infrared spectra, since the product, a protein, was produced in very limited amounts (< 0.5 g L⁻¹). Ammonium concentration was assumed constant, due to the regulated addition of ammonia for pH control. The calibration therefore resulted in a simple univariate model, using the height of the sharp methanol infrared peak, at 1018 cm⁻¹. High accuracy

(standard error for the prediction of methanol < 3.7 mM) was obtained by placing the two anchors for signal drift correction very close to the methanol peak, and by calibrating the instrument *in-situ* at the beginning of each process.

3/ A VERTICAL CONCLUSION

The goal of this thesis was to develop simple calibration methods for the on-line monitoring and control of bioprocess by mid-infrared spectroscopy (FTIR). This has been achieved successfully for the cultures of microorganisms on known substrates, preferentially in defined media. The methods developed during this work do not concern cultures grown on chemically undefined substrates and bioprocesses implying metabolites for which synthetic standards cannot be produced. Nevertheless, the actual trend is to encourage the use of chemically defined medium, and most bioprocesses can be monitored using the presented calibration methods.

For such bioprocesses, the traditional process chemometrics approach could be simplified into a library-based calibration, therefore reducing dramatically work load and modeling complexity. Three main observations have enabled this achievement:

1. Mid-infrared absorbances can be added linearly and follow the Lambert-Beer's law. This means that species do not interact in solution, and peak shifting does not occur. As a result, a multi-level calibration design does not provide more information than a 2-level design with 3 replicates. While this does not reduce the number of standards to measure, this divides by a factor of three the number of solutions to be prepared.
2. Noise/drift issues and concentration changes can be addressed separately. The drift problem can be effectively overcome using an anchorage method, or by modeling it with a few principal spectral features. Working with a very clean signal reduces the need for replicates during calibration, and therefore, the number of standards to be measured. In addition, it renders useless Principal Component Regression modeling, which has for goal to distinguish between relevant information and noise from calibration data. Modeling becomes much simpler, and can be performed by non-experts, in any calculation software treatment – even Microsoft Excel.
3. Working with concentration differences from the initial state rather than absolute concentration allows for using a single spectra library for different processes, regardless of the medium composition and microorganism used. It results in a further reduction of the size of the calibration set, since the first level of the 2-level design is included for all species in the medium background, while the second level is expressed by pure component spectra. Such a simplification has an enormous impact on the number of standards to measure if a large number of medium compositions have to be tested.

In addition, it was also shown that the pH has no direct influence on the infrared absorbance of compounds. The pH only acts on the deprotonation equilibria, which in turn, change the concentration of the deprotonated and protonated forms of weak acids in solution. These variations can be explained by inclusion of the buffer species into the library. It was also demonstrated that although the temperature has a large influence on the intensity spectra, it has a negligible influence on the absorbance coefficients. Working with concentration differences eliminates the effect of temperature on the modeling. The spectra library can therefore be easily extended for bioprocess developments that include the testing of various pH and temperatures.

The calibration methods developed in this work are, intrinsically, not really new, and equations and algorithm used are very common. But the novelty of this work does not come from the methods themselves, but rather from how they can be applied to the monitoring of cultures by mid-infrared spectroscopy. Whereas Principal Components Regression is considered as absolutely necessary for bioprocess modeling, this work has shown, somewhat against the mainstream, that simpler tools can be much more efficient. The decisive action during this thesis was, therefore, to challenge and question the established, state-of-the-art, practices.

4/ THE TEN GOLDEN RULES OF FTIR

1. Do not focus only on modeling, the spectroscopic measurement must come first

Developing highly sophisticated models for spectroscopic measurements of low quality is rather inadvisable, and it could be a trap for chemometrics experts. Although methods exist to correct for signal instabilities, it is much more preferable to limit noise sources from the environment, and spending some time on this point is definitely worth it. The dry purge air is a critical parameter to improve the quality of the spectra. Large variations in room temperature and light exposure have a negative impact on signal stability. Do not underestimate the importance of the infrared spectroscopy theory. This technique has been mainly used for qualitative analysis, and knowing the absorption bands of the main functional groups may help tracking unexpected metabolites.

2. Keep modeling as simple as you can

Principal Components Regression is a powerful, but quite complex tool. Use it only when simpler approaches have failed. Multiple Linear Regression (i.e. traditional least squares) works fine with defined media, and even withstands the addition of limited amounts of chemically undefined compounds, including yeast extract or peptone. More advanced process chemometric modeling should be used only when the main metabolites are unidentified and/or synthetic standards cannot be produced (chapter 1 and 2)

3. Address noise issue separately from the calibration itself

Signal instability is by far the most problematic issue in on-line monitoring by mid-infrared spectroscopy. The traditional process chemometrics approach addresses the signal instability and calibration issues, by employing additional factors for noise compensation simultaneously. Tackling these two problems separately (if possible, see point 2) results in enormous time savings. Signal instabilities can be corrected for by using several different methods (supplementary material, part A). Signal anchoring (chapter 3, 5 and 8) and incorporation of drift spectra into the library (chapter 4, 6 and 7) are quite robust methods, which, in addition, allow easy detection of gross drift.

4. Identify the species that are likely to show significant concentration changes

Most species found in culture media display very small concentration changes during the bioprocess. Phosphate buffer, trace elements, vitamins, salts and poorly soluble

gases (e.g. oxygen) fall in this category. Under usual stirring conditions, biomass does not get close enough to the Attenuated Total Reflectance (ATR) probe to be detected. The compounds of interest, in terms of monitoring by spectroscopy, therefore consist in the carbon and nitrogen sources, and the products that reach a certain titer. All the species that show small concentration changes can be included in a “background” term in order to eliminate them (chapter 3). Practically, this is done by referencing all process spectra (I) to the first spectrum taken, before inoculation (I_0).

5. Select for the best counter-ions

Cells consume and produce ions, and not only salts. In terms of calibration, this means that, for instance, standards of ammonium and acetate should be prepared. Unfortunately, with respect to charge balance, this is impossible, and salts must be used. Sodium and chloride are the best counter-ion, since they show a very weak mid-infrared absorbance (chapter 3). Use for example ammonium chloride for your ammonium standard, and avoid if possible other salts such as ammonium sulfate or ammonium acetate.

6. Build your library carefully

Only a small number of standards must be prepared and measured for the spectra library, but carry out this work with great care. Measure your sample and the water reference three times and average out the spectra. A qualitative inspection of the molar absorbance spectra can turn out to be useful to detect any problem that could have occurred during this calibration process, as for instance a sudden baseline drift. Temperature and pH have an influence on infrared spectra, either directly, and/or through equilibrium constants (chapter 5 and 6). Include these effects in your model, or take care to adjust the temperature and pH of your standards and baseline to the corresponding process set points.

Consider the use of antifoam when preparing your standards. The surface tension of non-ionic standards can be very low, which may have for a result that air bubbles may adhere on the surface of the ATR probe. This can significantly perturb the measurements and can be avoided by adding a small amount of antifoam, as is it done for culture medium.

7. Select an appropriate wavenumber range for calculations

A broad wavenumber range can potentially provide more information, a narrow window allows avoidance of noisy regions. A compromise must be found, in order to include all relevant peaks while avoiding noisy regions. The “fingerprint region” (900

– 1800 cm^{-1}) presents the highest signal to noise ratio for the aqueous solutions and should be used preferentially. A large number of substrates and products (e.g. glucose, glycerol, ethanol, etc.) show absorption peaks within 950 – 1150 cm^{-1} , whereas charges species, such as organic acids and ammonium, absorb in the range 1400 – 1600 cm^{-1} . Noisy regions occur below 1000 cm^{-1} and above 1600 cm^{-1} . Reasonable compromises for the calculation range are therefore 1000 – 1500 cm^{-1} to 950 – 1650 cm^{-1} (chapter 3, 4, 5 and 6). Working with a single wavenumber for calculation (chapter 7) is limited to special applications, in which a single compound is likely to exhibit concentration changes. Optimizing the calculation range, using for instance segments, may decrease model robustness and hinder the detection of unexpected compounds.

8. Track outliers

Modeling results should be accepted only after careful analysis of residuals. The amplitude of the apparent signal drift should be examined, by plotting against time either the noise concentration profiles (chapter 4) or the signal drift amplitude (chapter 3). The consistency of this correction should be examined: have the two anchors moved in a similar manner? Do the profiles of the two concentration factors present unusual features? The projection of the squared residues in the wavenumber dimension can help identifying the unexpected metabolites.

9. Exploit the library's generality

A single library can be used for a wide variety of bioprocesses, as long as the set up is not modified (chapter 4). It is therefore worth spending some time to build carefully the library, and to include compounds that could appear sooner or later, such as formate, lactate, succinate, etc. These can be used when tracking unidentified process metabolites.

10. Remember that FTIR is not yet a perfect tool

Yes, FTIR can simultaneously monitor many compounds on-line. It is true that it is non-invasive, fast and can be sterilized in-situ. It can indeed replace tedious and boring off-line analyses. But it is not yet a perfect tool. Signal instabilities, a certain percentage of experiments will have to be repeated due sudden, huge, signal drifts or software crashes. Do not get desperate, and just assume that future developments in fiber optics and multiplexing will overcome these problems.

5/ PERSPECTIVES

Simplified calibrations pave the way for the implementation of mid-infrared spectroscopy instruments into high-throughput platforms for strain and medium screening. While micro-devices are now available for the on-line monitoring of pH and dissolved oxygen, no convincing example of simultaneous measurement of principal metabolites has been reported. FTIR spectroscopy could fill this gap, since a single library-based calibration can be used to explore a wide range of culture media and strains. With a few programming developments, data treatment and outliers detection could be performed automatically. pH monitoring could be performed by FTIR, which would save space for complementary probes, e.g. dissolved oxygen or biomass sensors.

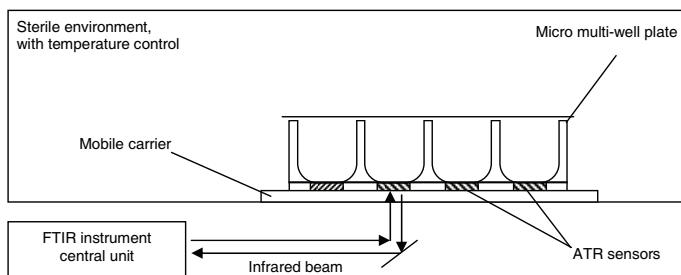


Figure 1. Non-invasive monitoring of a multi-well plate by mid-infrared spectroscopy. The bottom of each well is made of an Attenuated Total Reflectance (ATR) sensor. The plate moves in order to place every well consecutively in the infrared beam. The entire set up is located in a sterile environment.

Practically, such an implementation could be done, at first glance, in two different manners. On the market are available 6 mm probes, connected to the instrument central unit by a flexible fiber-optic conduit. Assuming that a spectrum can be collected every 45 seconds, a 48-well plate could be monitored within 40 minutes by the probe, moving from one well to another. Such a design could be easily developed from existing hardware, but presents a few severe drawbacks. Inserting the same probe in different wells would certainly compromise the sterility of the experiment. Due to unavoidable sample adherence to the probe tip, culture medium would be lost at each measurement. In addition, it would be difficult to guarantee a complete containment, i.e. that no liquid from one well can be end up in another. These drawbacks could be overcome by designing the set up differently. Rather than having one probe that moves from one sample to another, one could imagine having Attenuated Total

Reflectance (ATR) sensors at the bottom of each well. The plate would move, in order to place sequentially every well in the infrared beam (Fig.1). This design would be completely non-invasive and would not compromise sterility and containment. However, purpose-made well plates would have to be manufactured. In order to work with disposable, “one-use” plates, diamond sensors could be replaced by much cheaper silicon or germanium crystals. In this case, preliminary studies should be conducted to study the performance of such sensors (in the present thesis, all measurements have been made with a diamond sensor). It must be emphasized that both of these set ups would allow an easy correction of the signal drift, by placing water in one of the wells and referencing all samples to the water intensity spectra. This correction should lead to a measurement accuracy that could compete with off-line analyses.

On a longer term, the major challenge for FTIR monitoring is the on-line quantification of secondary metabolites (e.g. antibiotics) and recombinant proteins. Bioprocess development seeks high productivities, and product measurement is generally carried out off-line, through expensive, inaccurate and time-consuming analyses, such as electrophoretic gels, immunoassays, etc. The need for fast and reliable on-line measurement for these compounds is therefore extremely high and any advance in that field would make a considerable breakthrough. However, the concentration of these products rarely reaches high titers, which makes monitoring very challenging. In addition, in terms of functional group composition, proteins are all very similar. A solution might come from a microfluidic device combined to the ATR/FTIR probe. Using a dielectrophoretic field, proteins could be concentrated in the close vicinity of the ATR probe, enhancing local concentration and, therefore, resolution. By playing with the chip frequency, it might be even possible to separate proteins according to their size and isoelectric point. Although very appealing, the on-line monitoring of recombinant proteins and other high added-value products remains far from today's instrument capabilities, and only a long-term research effort, carried out by a consortium of groups with the appropriate expertise may perhaps tackle this problem.

Supplement

Note on Supplementary
Material

1/ NOTE ON SUPPLEMENTARY MATERIAL

The two following chapters do not focus on simplified mid-infrared calibrations for bioprocess monitoring. For these reasons, I considered that they would dilute the main message of the present thesis, if they were gathered with the previous chapters. Nevertheless, they provide additional information that deserves to be comprised in this work, and that is why they were included as “supplementary material”.

The first chapter emphasizes the importance of signal drift for bioprocess monitoring by mid-infrared spectroscopy. Several methods to correct for signal instabilities are reviewed and discussed. The first author of this work is Michal Dabros, and this chapter was included in his PhD thesis.

The second chapter is less related to mid-infrared spectroscopy. It reports a study on recombinant protein production that has been done, during this PhD, in parallel with the development of the calibration method presented in chapter 7. Control of residual methanol concentration during *Pichia pastoris* cultures by FTIR was used as a means to investigate the effect of upstream conditions on protein glycosylation and productivity.

Supplement
Part A

The Ongoing Quest for Truly
On-line Bioprocess Monitoring
Using Spectroscopy

1/ ABSTRACT

A general overview of methods that help maintain the on-line reliability and robustness of spectroscopic monitoring instruments used in bioprocess monitoring is presented. Truly predictive spectroscopic monitoring techniques, capable of providing continuous, live information about process analytes, biomass and bioproducts are highly desirable as they set the stage for process automation, control and optimization. In addition, the use of these techniques is in line with the FDA's Process Analytical Technology (PAT) initiative and is increasingly required in industrial biotechnology to ensure consistent product quality and enhance early fault detection. One of the major obstacles currently impeding the technology is the chronic lack of on-line reliability. Instability, drift and offsets in spectral data, often amplified by changing process parameters or equipment-related factors, undermine the validity of the calibration model. Unfortunately, most of the commonly applied signal and model correction methods require off-line analysis to obtain in-process calibration standards or reference samples for model transfer algorithms. Technically, this approach leads to a mere retrofitting of the predicted properties to off-line data. This work aims to review some of the currently available techniques of correcting the signal and adapting the calibration model to the current monitoring conditions without sacrificing the real-time applicability of the technology. A simple case study is presented to illustrate the effect of some of the described methods.

This chapter was published as a chapter of a book:

Dabros M, Schenk J, Marison IW, von Stockar U. 2007. The ongoing quest for truly on-line bioprocess monitoring using spectroscopy. In: *Biotechnology and Bioengineering Research Trends*. Hauppauge (NY): Nova Science Publishers: *In Press*

2/ INTRODUCTION

Recent progress in on-line spectroscopy has created enormous opportunities for the development of processes monitoring, automation, optimization and control in biotechnology. At the same time, the use of Process Analytical Technology (PAT) tools is increasingly encouraged or demanded by the Food and Drug Administration (FDA) and various government regulations in order to increase real-time process supervision, reduce process variability and improve the consistency of product quality (FDA, 2005; Haack et al., 2004; Junker and Wang, 2006; Schügerl, 2001; Wold, 2004). These process enhancement techniques require extensive and, preferably, live information about the system. On-line spectroscopic sensors provide fast, continuous, real-time measurements of multiple process analytes, biomass and bioproducts, helping the user to run the process at the desired conditions and setting the stage for control and optimization applications (Arnold et al., 2002; Bellon, 1993; Cannizzaro, 2002; Cannizzaro et al., 2004; Kornmann et al., 2003; Kornmann et al., 2004a; Olsson et al., 1998; Pollard et al., 2001; Rhiel et al., 2002b; Schenk et al., 2007; Sivakesava et al., 2001; Wilson and Trapp, 1999). In addition, spectroscopic sensors installed *in-situ* offer significant advantages for bioprocesses over traditional at-line and off-line methods: sterility, non-invasiveness, non-destructiveness and generally low maintenance requirements (Arnold et al., 2002; Cannizzaro, 2002; Kornmann et al., 2004a; Pollard et al., 2001; Rhiel et al., 2002b; Schenk et al., 2007). Yet, despite these vast benefits, the spread and acceptance of on-line spectroscopy in industrial applications are still considerably hindered by the general lack of real-time reliability.

Like most other analytical instruments, spectroscopic sensors need to be calibrated before they can be used for prediction. Because the spectral features of the monitored property (properties) of interest frequently overlap, calibration is typically approached through multivariate analysis methods. Algorithms, such as Principal Component Regression (PCR) and Partial Least Squares (PLS) have established their position as the chemometric techniques most commonly applied to model the relationship between the obtained spectral information and the monitored variable(s) (Arnold et al., 2002; Feudale et al., 2002; Haack et al., 2004; Haaland and Thomas, 1988; Hall et al., 1996; Kornmann et al., 2003; Martens and Naes, 1997; Nadler and Coifman, 2005; Sivakesava et al., 2001; Trygg and Wold, 2002; Wold et al., 2001; Wolthuis et al., 2006). However, the difficulty of developing a reliable, robust multivariate calibration model and maintaining the model's validity for future real-time applications present some of the major obstacles in on-line spectroscopy. Decreased model performance in

on-line conditions is often due to variations in process parameters, instrument-related factors, various interferences causing noise and signal distortion, as well as shifts in baseline signal, spectral drift and offsets and a number of other reasons (Arnold et al., 2002; Duponchel et al., 1999; Feudale et al., 2002; Goodacre et al., 1997; Manning and Griffiths, 1997; Mark and Griffiths, 2002; Rhiel et al., 2002b; Riley et al., 1999; Rodig and Siebert, 1999; Schenk et al., 2007; Wolthuis et al., 2006; Yardley et al., 2000; Zhang et al., 2002). As a result, spectrometers are either routinely recalibrated or need to have the calibration model standardized to work under the required conditions (Feudale et al., 2002; Zhang et al., 2002). However, recalibrating the model requires extensive workload and may be impractical in industrial settings, especially with *in-situ* applications, while calibration standardization algorithms typically require off-line analysis to obtain the reference standards. As such, the predictive capacity of the technique is lost. Moreover, even if the model is developed or standardized shortly before a particular process, it may still lose its validity during the course of the run for any number of the reasons mentioned above. Figure 1 shows an example of the impact of short-term signal drift on the predictive performance of a mid-infrared spectrometer calibrated just before the experiment. To overcome such problems, the spectra and / or the calibration model need to be updated in some way in order to adapt them to the current monitoring conditions. Achieving the necessary adjustments without losing the real-time applicability of the technology presents the challenge that will be discussed in this work.

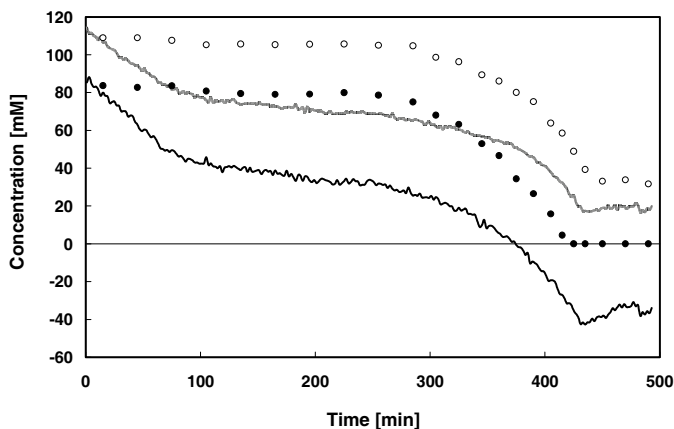


Figure1. Short-term drift experienced with a mid-infrared spectrometer during a batch culture of *E. coli*. Points refer to off-line reference analysis of glucose (black circles) and ammonium (open circles), whereas lines represent concentration calculated from IR spectra (glucose: plain line; ammonium: dashed line).

Basing the classification on the design and maintenance of the calibration model, Kornmann et al. (Kornmann et al., 2003) suggest dividing on-line spectroscopic monitoring applications into two general categories: a) process analysis involving retrospective reprocessing of the obtained data using additionally acquired off-line measurements or standards, and b) truly predictive, real-time monitoring of the process using a previously developed model. Although many applications still fall under the first group, there is a clear trend towards the second category, catering to the increasing importance of the PAT initiative.

The former approach relies on using samples collected during the on-going bioprocess as standards for the calibration model. These in-process standards (IPS) are analyzed off-line in parallel to the run, and progressively added to the calibration set of the monitoring spectrometer as they become available. In-process standards can be used either to develop the model entirely on their own (Arnold et al., 2002; Doak and Phillips, 1999; Fayolle et al., 2000; Fayolle et al., 1996; Pollard et al., 2001; Sivakesava et al., 2001), or as supplementary information used to augment an existing, previously built model (Blanco et al., 2004; Flores-Cerillo and MacGregor, 2003; Kornmann et al., 2003; Riley et al., 1998). With each new sample, the model is recalibrated, and because it is fitted with data from the actual process, its performance typically increases for that particular application. Although generally effective and widely used in various ways, the general disadvantage of this approach is that it requires some kind of off-line sample treatment and/or retrospective refitting of data. As such, the user technically does not have access to continuous live information on the monitored process and the monitoring technique can no longer be considered purely on-line. Furthermore, using standards from the monitored process introduces metabolism-induced correlations into the calibration model, which may further weaken its robustness (Rhiel et al., 2002a; Riley et al., 1998). Methods belonging to this category will not be discussed in this review.

The second category of on-line spectroscopic bioprocess monitoring applications involves the use of strictly predictive models. The general idea is straightforward: the instrument is calibrated beforehand and used to obtain continuous, real-time information on the process, without requiring any type of adjustments based on off-line analysis. Adapting the instrument to make it more suited to the current monitoring conditions is acceptable, so long as it is done in real-time and only using data available on-line (Crowley et al., 2000; Kornmann et al., 2004a; Mazarevica et al., 2004; Olsson et al., 1998; Rhiel et al., 2002b; Schenk et al., 2007; Yardley et al., 2000; Zhang et al., 2002). The aim of this work is to summarize and exemplify some of the main difficulties encountered in the implementation of truly on-line

spectroscopic monitoring practices and to review several recently and currently studied approaches to overcome the obstacles. A simple case study is carried out to illustrate the outcome of some of the described approaches. The study is based on results obtained with a Fourier transform mid-infrared spectrometer (FTIR), a major player in the field of on-line monitoring of bioprocess metabolites and products. The principles and methods presented in this review are, however, broadly applicable to other types of spectroscopy.

3/ REVIEW OF CURRENT METHODS FOR REAL-TIME SPECTROMETER ADAPTATION

Rhiel et al. (Rhiel et al., 2002b) suggest that a calibration model can be considered truly robust for on-line prediction if its validity, based on some predefined criteria, is shown to hold at least a week following its development. During this time, various factors may come into effect leading to decreased predictive performance. These may include baseline shifts, spectral drift and offsets, or unexpected instrument responses due to changes in process parameters, probe alignment, equipment configuration and other external factors. To exemplify the problem of signal drift, Figure 2 shows mid-infrared intensity spectra of water, collected under strictly-controlled conditions, over a continuous period of 72 hours and at an interval of 20 minutes, using an FTIR (ReactIR™ 4000, Mettler Toledo, Greifensee, Switzerland). The instability is particularly pronounced in the *fingerprint* region of the mid-IR range (1500 - 950 cm^{-1}), where most common bioprocess analytes display activity. The average absorbance value at each wavenumber in this region, with the standard deviation in error bars, is shown in Figure 3a. It should be emphasized that unwanted spectral variations are of particular importance in the bioprocess field, since most analyte concentrations are in the range of only 0.01 - 2 %. As a consequence, even the slightest spectral drift or baseline offset can translate into significant prediction errors. To illustrate the relative significance of the problem, Figure 3b shows how the same water spectra compare to a pure component absorbance spectrum of glucose at 10 g L⁻¹.

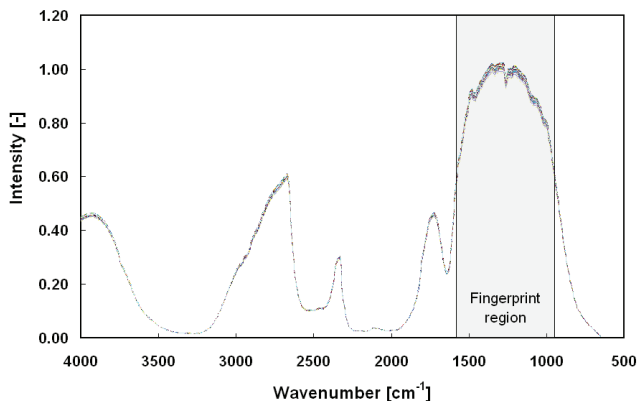


Figure 2. Mid-infrared intensity spectra of water collected over a period of 72 hours.

The frequent lack of signal stability, combined with the increasing need for reliable and precise real-time monitoring, has lead to a shift in the research focus in bioprocess spectroscopy. Much of the current research effort centers on attempting to maintain the robustness of an existing calibration model through various mathematical and chemometric methods of correcting spectral signals and/or adapting the calibration model to current, on-line conditions. This may involve various data pre-treatment schemes, offset removal, baseline corrections and other approaches that will be discussed in this review.

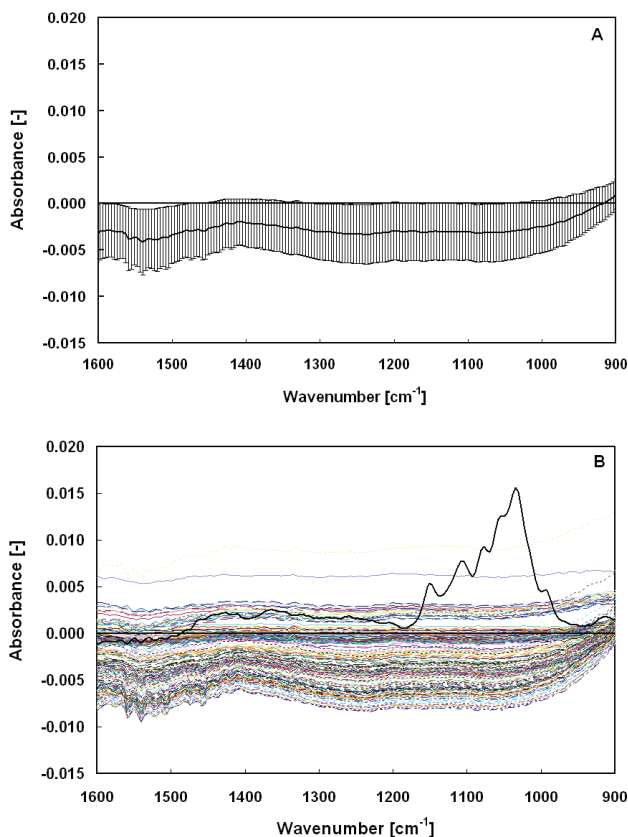


Figure 3. Average absorbance value and standard deviation (in error bars) of mid-infrared spectra of water collected in the fingerprint region over a period of 72 hours (A). The spectra compared to a pure component spectrum of glucose at 10 g L⁻¹ (B).

The optimal strategy will often depend on the anticipated nature of the artifact(s) that compromised the quality of the spectral data or the validity of the instrument's calibration model. The methods reviewed here have been grouped into a few broad categories. First, we describe the most commonly used data pre-processing steps, including variable selection, data scaling, differentiation, wavelet compression and multiplicative signal correction. Offset removal strategies are then described, followed by on-line signal correction through spectral anchoring. Next is a general overview of a variety of techniques based on orthogonalization, including orthogonal signal correction (OSC) and orthogonal partial least squares (O-PLS). Quantification and different ways of modeling of process parameters that influence spectral data are then discussed. Finally, we review the on-line applicability of data reconciliation for on-line spectroscopy and present an overall discussion of the importance of real-time bioprocess monitoring.

Common data pretreatment methods

Multivariate modeling is seldom performed on raw spectral data. Digital filtering, variable selection, data scaling and various data pre-processing techniques, such as wavelet transforms (Alsberg et al., 1997), are usually applied in order to eliminate systematic errors and maximize the quality and usefulness of the data both for the calibration and the prediction steps. Variable selection which, in spectroscopy, usually means selecting the appropriate range(s) of frequencies (or wavenumbers) allows the user to target those spectral regions where the monitored property of interest is the most distinguishable. A comprehensive review of variable selection methods is given by Ferreira et al. (Ferreira et al., 2005). Mean centering and/or variance (auto) scaling of spectral data is also nearly always used as a way to even out the importance of each measured variable and, sometimes, to correct for linear shifts in the spectra (Sjöblom et al., 1998).

Often, efficient correction of noise and baseline artifacts can be performed by filtering, smoothing or differentiation of spectral data (Arnold et al., 2002; Blanco et al., 2004; Blank et al., 1996; Crowley et al., 2000; Giavasis et al., 2003; Hall et al., 1996; Rhiel et al., 2004; Sivakesava et al., 2001; Wold et al., 1998). The technique that appears to have found the most success in spectroscopy is the Savitzky and Golay smoothing and differentiation algorithm (Savitzky and Golay, 1964). Perhaps the most important reason for this success is that through its simple operation, the algorithm is generally very efficient both at removing random noise from the spectral data, as well as handling background offsets. If the measurement interval of the instrument is short, the algorithm can be used to smooth the data with respect to time

using polynomial regression of a given degree. Smoothing and derivation with respect to frequency (or wavenumbers) can be performed on spectra with a sufficiently high resolution. The effectiveness of taking the first order derivative with respect to wavenumbers of the water spectra presented earlier in Figure 3b can be seen in Figure 4a: the average derivative of absorbance is much closer to the expected value of zero, with very minimal standard deviations. Figure 4b shows the derivative water spectra along with the first derivative spectrum of glucose at 10 g L⁻¹ superimposed for comparison. The increase of the signal-to-noise ratio is clearly notable.

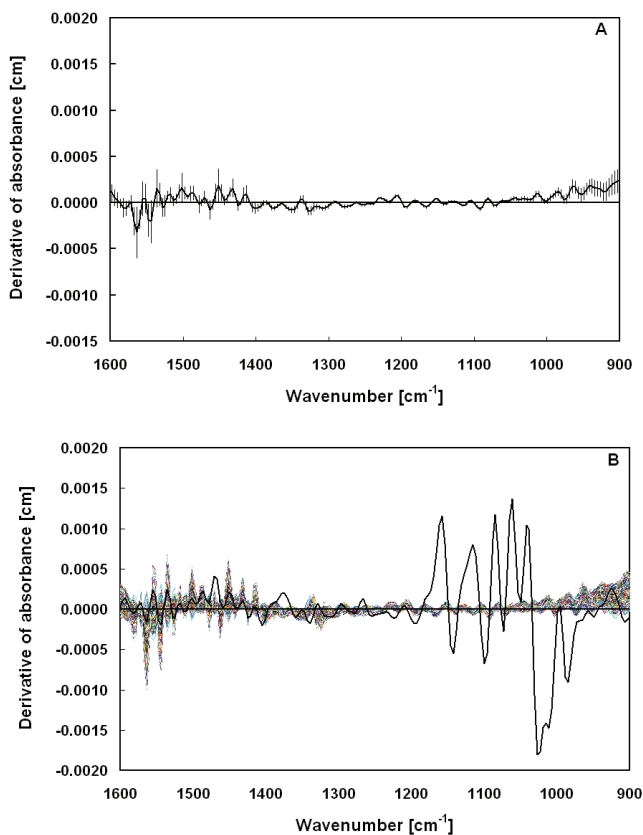


Figure 4. Average value and standard deviation (in error bars) of 1st derivative of the absorbance spectra from Figure 3b (A). The spectra compared to a 1st derivative pure component spectrum of glucose at 10 g L⁻¹ (B).

Another data pretreatment technique worth mentioning is multiplicative signal (scatter) correction (MSC), a method of signal pre-processing useful in eliminating the influence of such instrumental artefacts as light scattering variations and changes in path length. These problems are a common result of slight power fluctuations, frequently experienced with IR and Raman spectrometers. MSC corrects for the multiplicative effects of light scattering on the spectra by expressing each spectrum as a multiple of the mean spectrum of the calibration dataset and an additive residual, thereby simulating equal and “ideal” light scatter levels in all the samples (Geladi et al., 1985; Martens and Naes, 1997; Wolthuis et al., 2006).

Finally, the effects of signal instability in spectrometry can be handled by systematically referencing the measurements against a background standard. This method is routinely applied in various off-line analyzers, such as HPLC or optical density spectrophotometers, as well as in dual-beam instruments. However, background referencing is difficult to implement for single-beam instruments used on-line and installed *in-situ* due to sterility issues (Arnold et al., 2002). Mazarevica et al. (Mazarevica et al., 2004) proposed a solution involving the construction of a sterile automated flow system capable of alternating between recording spectra of the fermentation mixture and a background solution.

Offset removal

Slight variations in the equipment setup, probe alignment or other operational parameters may cause offset to be present in the absolute values predicted by the spectrometer. If the initial value of the monitored property is known *a priori* (such as is often the case, for example with known concentrations of process analytes or biomass prior to inoculation), this information can be used to perform a simple offset adjustment in the predicted results (Cannizzaro, 2002; Rhiel et al., 2004; Rhiel et al., 2002b). The offset in the measured property is calculated at the initial state and then simply subtracted from the estimated values throughout the experiment. The validity of this method can be reconfirmed at the end of the experiment if the final conditions are also known, for example by checking if the predicted concentration of the substrate is zero following the substrate’s depletion.

A variation of the technique involves removing the offset directly in the raw spectra. As such, the correction is performed on the model inputs, rather than on the outputs. This approach requires some *a priori* knowledge about the expected shape and magnitude of the spectrum at a given reference point, most commonly at the beginning of the process. For instance, Cannizzaro et al. (Cannizzaro et al., 2004) corrected on-line frequency scan capacitance spectra by removing the initial spectrum collected before inoculation (which

should have been, but was not – nil) from each subsequent spectrum. A more elaborate approach was investigated by Yardley et al. (Yardley et al., 2000), who developed a special model for on-line baseline flattening that corrected the influence of baseline distortions and electrode polarization on dielectric spectra. The model was built based on a series of frequency-dependent baseline artifacts that were observed in the capacitance spectra under test conditions involving a number of varying process and instrumental factors. The artifacts were modeled as additive or multiplicative terms and subsequently compensated for in real-time, increasing the on-line fidelity of the dielectric data.

Spectral anchoring

A frequently applied technique of correcting signal drifts on-line involves spectral anchoring. One or several anchor points are identified in specific regions of the spectrum that are expected to remain constant during the process. For example in infrared spectra, anchor points would be chosen at wavenumbers where no components present in the fermentation medium are expected to absorb (Fayolle et al., 2000; Schenk et al., 2007; Wilson and Trapp, 1999). During the course of the process, all the subsequent spectra are then corrected based on the movements of the anchor point(s) relative to a reference point in time. The spectral correction can either be constant (independent of the value of the signal at a given frequency, Equation 1a), or proportional to that value (Equation 1b). Using a single anchor point, the corrected spectrum, \mathbf{X}^* , can be calculated at any given time during the experiment as follows:

$$\mathbf{X}^* = \mathbf{X}_i + (X_r - X_{r_0}) \quad \forall i = 1, \dots, p \quad [1a]$$

$$\mathbf{X}^* = \left(\frac{X_r}{X_{r_0}} \right) \mathbf{X} \quad [1b]$$

where \mathbf{X} is the raw spectrum of length p , X_r is the value of the spectrum at the anchor point, and X_{r_0} is the reference value of the spectrum at the anchor point. When more than one anchor point is applied, the difference or the ratio of X_r and X_{r_0} can be calculated as the average value. Figure 5a shows the improvement achieved by anchoring the original spectra of the case study using the second equation at two points: 950 cm^{-1} and 1750 cm^{-1} . A comparison with the anchored spectrum of glucose at 10 g L^{-1} is shown in Figure 5b. Note that in this case, the corrected spectra do not actually pass through zero at the anchor points, as it might intuitively be expected. This is due to the fact that, as mentioned above, an average ratio of X_r and X_{r_0} was used during the correction.

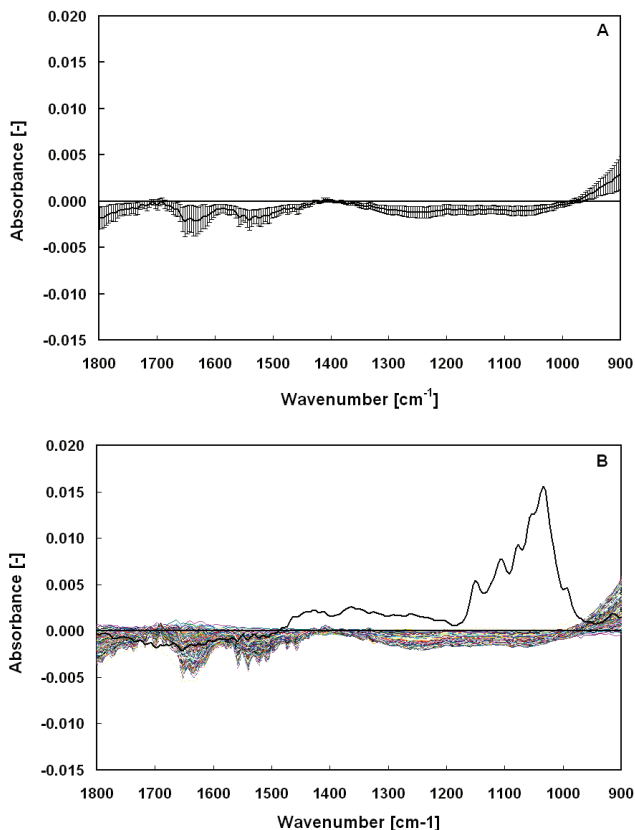


Figure 5. Average absorbance value and standard deviation (in error bars) of the absorbance spectra from Figure 3b, anchored at 950 cm⁻¹ and 1750 cm⁻¹(A). The spectra compared to a pure component spectrum of glucose at 10 g L⁻¹ (B).

The main difficulty with this technique is defining an appropriate anchor point. It needs to be chosen in a region where it will not be influenced by the observed variable(s). At the same time, however, the effect of any external factors on the anchor point needs to provide a representative indication of the effect on the entire spectrum.

Orthogonalization

Most of the signal correction techniques described until now have the common potential drawback that in the process of removing noise, drift or offsets from the spectral data, some

of the relevant information contained in the raw spectra can be lost, too. Methods based on data orthogonalization, such as orthogonal signal correction (OSC), circumvent this problem by removing only those errors and variations in the spectra that are shown to be unrelated to the observed property (Wold et al., 1998). In essence, the general idea of OSC is to filter out from the spectral data matrix (\mathbf{X}) information that is mathematically orthogonal, and thus definitely unrelated, to the matrix of estimated variables (\mathbf{Y}). More specifically for the case of PLS modeling, this can be achieved if the scores (\mathbf{T}) of \mathbf{X} are made orthogonal to \mathbf{Y} , meaning that both $\mathbf{T}^t\mathbf{Y}$ and $\mathbf{Y}^t\mathbf{T}$ are matrices with only zero-valued elements. The orthogonalized scores multiplied by the loadings constitute the undesirable part of the data and are removed from \mathbf{X} . A new model is then developed correlating the filtered \mathbf{X} to \mathbf{Y} . Detailed descriptions of OSC are given in Wold et al. (Wold et al., 1998) and Sjöblom et al. (Sjöblom et al., 1998). The effect of applying OSC to the case study spectra is demonstrated in Figure 6.

Orthogonal signal correction can be combined with and ultimately integrated into the PLS model itself. This vision gave rise to orthogonalized-PLS, also called orthogonal projections to latent structures and abbreviated O-PLS (Trygg and Wold, 2002). The basic approach of the merged method is that variation that is orthogonal to the response matrix is analyzed and removed in each PLS component individually. The orthogonal (“bad”) variation in the predictor matrix is consequently separated from the correlated (“good”) variation. Thus, in addition to correcting the signal prior to modeling, O-PLS provides the opportunity of interpreting the unwanted variation in \mathbf{X} by making it possible to analyse the scores and loadings found to be orthogonal to \mathbf{Y} . At the same time, the corrected PLS model also becomes easier to interpret since the latent orthogonal variation in \mathbf{X} has been filtered out, yielding a simpler, more relevant model (Trygg and Wold, 2002; Wold et al., 2001).

Real-time signal correction by orthogonalization can also be achieved through the use of some reference measurements available on-line. Dabros et al. (Dabros et al., 2006) investigated a dynamic orthogonalization approach for single-beam instruments based on reference points obtained with analyte spiking. Planned pulse additions of known amounts of monitored analytes were injected periodically into the culture medium, generating *delta spectra* (Figure 7) that were used as on-line references for a novel model adaptation method, known as Dynamic Orthogonal Projection (DOP) (Zeaiter et al., 2006). The method was shown to reduce the effect of spectral drift in a single-beam FTIR spectrometer in real-time and without the need for retrospective sample analysis.

Quantification and modeling of influencing process parameters

The influence of various process parameters (such as temperature, pH, pressure, conductivity, viscosity, gas hold-up etc.) on the response of multivariate calibration models can be quantified and reduced either by explicitly correcting for their effect (November and Van Impe, 2000; Yardley et al., 2000) or simply by incorporating the values of these parameters into the model itself (Wülfert et al., 2000). Using this approach, the impact of various influence factors can be accounted for directly at the modeling step.

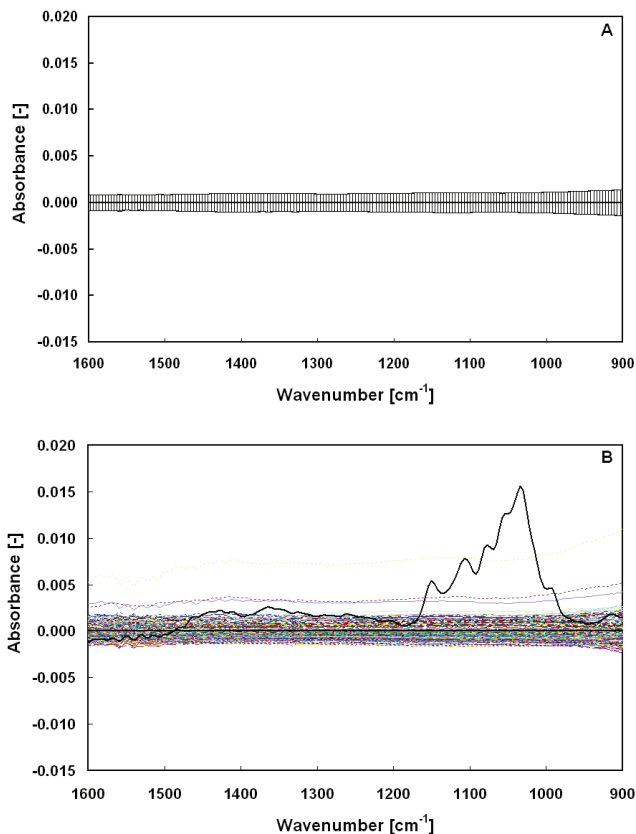


Figure 6. Average absorbance value and standard deviation (in error bars) of the absorbance spectra from Figure 3b, corrected by OSC (A). The spectra compared to a pure component spectrum of glucose at 10 g L⁻¹ (B).

Explicit correction for a given parameter initially requires a quantification of the parameter's influence, usually done under test conditions simulating the actual process. An analytical model describing the relationship between the affecting parameter and the spectral signal is developed empirically, validated on the test run and eventually applied to the actual process. For example, November and Van Impe (November and Van Impe, 2000) developed a linear correction equation describing the additive effect of stirrer speed and air flow rate on the capacitance readings of a biomass monitor.

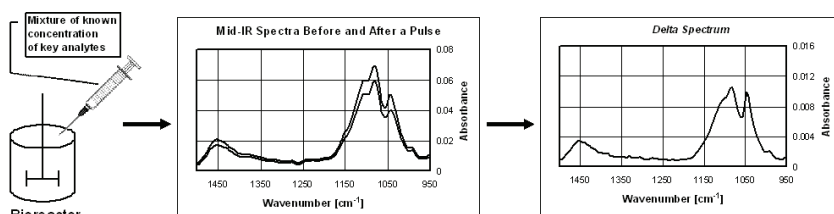


Figure 7. Procedure used to obtain the delta spectrum using a pulse injection of key metabolites into the medium (reprinted with permission).

The inclusion of the influencing parameter into the calibration model is a much more straightforward approach. If the process parameter is measured on-line during the process, its value can be integrated directly into the **X** matrix. For example, the measured temperature can be appended to an IR spectrum, or the value of medium conductance to a capacitance frequency span in dielectric spectroscopy. Since the **X** matrix now contains independent variables of different ranges and units, the data should be pretreated by scaling. Alternatively, the process parameter can be incorporated into the model's dependent variable, **Y**. This approach creates a situation where the model is inversely calibrated with the variable that causes the variation in the model's response. The technique can be applied in the case where the influencing parameter is not measured during the process. In fact, it will actually be predicted by the model from the measured spectra in **X**.

Since the most commonly used multivariate models like PLS are in principle linear, corrections of this nature should technically work only if the influence of the particular process parameter on the spectra is linear. Still, in some circumstances, small non-linear effects (for example, the influence of small changes in temperature or pressure on IR absorption) can be included implicitly into the calibration design by modeling them as unknown interferences. Wülfert et al. (Wülfert et al., 1998) have shown that pooling spectra

collected under variable temperatures into a global PLS model and accounting for the interference with additional latent variables can provide reasonable prediction performance. However, if the non-linear disturbance is strong, this approach may not yield the required model accuracy. In this case, non-linear correction equations should be quantified and non-linear global calibration models should be designed based on, for example, artificial neural networks (Fayolle et al., 1996; Gemperline et al., 1991) or support vector machines (Thissen et al., 2004).

Data reconciliation

Improved monitoring accuracy and precision of spectrometers can be attained by reconciling the predicted properties of interest to their best estimates using additional information available on-line during the process. Data reconciliation, sometimes called data confrontation, is a statistical data adjustment technique that evaluates the consistency of data in an overdetermined system and attempts to reduce errors in measured variables or to estimate unmeasured variables subject to a defined set of balances and physical equality (or inequality) constraints (Crowe, 1996; de Kok and Roels, 1980; Narasimhan and Jordache, 2000; Romagnoli and Sanchez, 2000; van der Heijden et al., 1994; Wang and Stephanopoulos, 1983). In bioprocess applications, this is typically done by performing macroscopic elemental and/or energy balances based on the analytic redundancy achieved with various process analyzers, such as on-line gas sensors, mass-flow controllers or reaction calorimeters (Duboc and von Stockar, 1995; Herwig et al., 2001; Kornmann et al., 2004b; Wang and Stephanopoulos, 1983).

Data reconciliation is traditionally applied for chemical processes involving flow circuits, where mass balance equations are readily definable. In biotechnology, the technique is sometimes used to obtain more accurate conversion rates and yield coefficients (Herwig et al., 2001; Jungo et al., 2007; van der Heijden et al., 1994). However, the method has received little attention in on-line bioprocess spectroscopy, most likely due to the difficulties in properly defining the system and establishing precise dynamic process balances and constraints. Kornmann et al. (Kornmann et al., 2004b) used data reconciliation to correct mid-infrared estimates of medium analyte concentrations and used the reconciled measurements to systematically recalibrate the instrument's model on-line. The approach led to a significant reduction in the number of standards required for regular, pre-process calibration. Unfortunately, due to a lack of appropriate on-line sensors, biomass could not be followed and was omitted in the algorithm. Using simultaneously a mid-infrared spectrometer estimating medium metabolites and a dielectric spectrometer to monitor biomass, Dabros et

al. (Dabros et al., 2007) applied data reconciliation to improve the accuracy of both instruments in real-time.

On-line data reconciliation has enormous potential to play a central role in improving the accuracy of spectrometers and increasing the spread of spectroscopy in bioprocess monitoring and control. One of the main advantages of the technique is that it can easily be coded directly into the prediction routine and thus, form an integral, systematic part of the instrument's model.

4/ DISCUSSION

Rendering spectroscopic sensors more dependable in on-line bioprocess monitoring and control has become an active topic in research. Currently available spectrometers frequently suffer from poor performance in real-time monitoring applications. Variable instrumental and process factors, as well as stochastic drifts and offsets in the spectral data significantly undermine the on-line predictive capacity of spectrometers. Moreover, since analyte or biomass concentration levels in biological processes are often very small, prediction results become sensitive to errors in accuracy. Consequently, calibration models are routinely recalibrated or adapted using standards or transfer samples collected during the monitored process and analyzed off-line. Besides the significant increase in the required workload, this approach compromises the actual on-line applicability of the technology. In order to ensure truly predictive functioning of spectrometers, the techniques used to correct the spectral data or to adapt the calibration model to the current monitoring conditions should be performed in real-time and only using data available on-line. Unfortunately, it was discovered in the process of writing this review that many authors neglect to provide clear explanations as to the exact steps taken during data collection and treatment, modeling and prediction. As a result, it is often not obvious whether the described application was in fact performed in real-time.

Table 1. Main techniques for real-time signal or model adaptation as a function of their key application.

Data preparation for modeling	Noise removal	Offset correction	Drift correction	Correction for influence of external factors
Variable selection	Digital filtering	Differentiation	Anchoring	Orthogonalization
Data scaling	Smoothing	Manual offset removal	Data reconciliation	Modeling of influencing factors
	Data reconciliation			Multiplicative signal correction

Some of the main truly predictive adaptation methods available for use in spectrometry and covered in this review are summarized in Table 1. Given the growing need for on-line biosensors, these techniques will likely receive considerable attention in the coming years. The purpose of the table is to place each method in its array of potential applications and to offer a key into choosing and combining the appropriate techniques according to the user's needs.

The question that naturally arises is, when is truly predictive, on-line monitoring really of essence? Ideally, well calibrated and robust spectroscopic sensors allow the user to follow manufacturing bioprocesses in real-time, but is continuous process information really

necessary or could we get away with some reasonable sampling and analysis delay? Lower prediction accuracy can be acceptable to a certain degree if the instrument is used merely for rough qualitative analysis or easy-to-interpret classification studies. Also, if some delay is tolerable, the performance of spectrometers can be substantially improved with in-process reference or standardization samples. However, on-line instrument stability and model robustness become crucial when collecting samples is impractical or when the availability of reliable continuous process information is required by regulation or necessary for the implementation of control schemes. In his recent editorial (Wold, 2004), Wold provided some insight into the importance of on-line monitoring from the point of view of the PAT initiative. Immediate analytical results have been shown to reduce production costs by minimizing the economical loss caused by unnecessary waiting periods between process steps. Eliminating off-line and post-process sample analysis allows on-line quality control and facilitates early fault detection, which may be of key importance in the food or pharmaceutical industry, for instance. Equally important in various industries is uninterrupted continuous process monitoring for traceability reasons (Watari and Ozaki, 2006). Finally, in the context of research and development, live process information without the need of sampling provides the necessary grounds for rapid strain screening, miniaturized high-throughput process development and other process enhancement studies aimed at improving product yield and quality.

5/ CONCLUSION

Removing variations from the spectral matrix that are unrelated to the monitored properties is essential in successful industrial applications of on-line spectroscopy. Data artifacts may range from plain noise through linear to nonlinear spectral drifts and offsets. The required adaptation technique will invariably depend on the expected nature of the variation. Noise in the spectra can usually be handled through various filtering and smoothing methods that can easily be incorporated into the data pre-treatment and modeling procedures. Cases of “purely” linear signal shifts and drifting can often be handled with mean centering of the data, derivation, offset removal or baseline correction through spectral anchoring. More complex variations can be identified and effectively removed from the spectral data by confirming their orthogonal nature with respect to the observed response variables. Process or instrumental variables that are expected to interfere with the shape of the spectral data can be accounted for by including them in the modeling step. Finally, the predicted values of the parameters of interest can be reconciled to their best estimates by using redundant measurements available on-line to the user.

Naturally, the utility of spectroscopy for real-time bioprocess monitoring does have its limitations. If the process conditions, equipment setup or target application change considerably, the instrument may simply need to be recalibrated. Developing a methodology of assessing when certain adaptation methods can be used to maintain the performance of a spectrometer on-line, and when the instrument can no longer serve as a real-time sensor, should be the subject of a separate investigation.

Achieving the goal of developing and maintaining a robust bioprocess monitoring platform capable of working entirely in real-time has enormous field potential, particularly in terms of process development, control and optimization, on-line product quality checks and tracing, reducing time between process steps, rapid fault detection and improved decision making. As live process information becomes increasingly important and demanded in the biotechnology field, so will the techniques that make it possible. On-going and future research should cater to this need by focusing on improving the stability of spectrometers and defining a standardized methodology of maintaining the accuracy of instruments working in on-line bioprocess monitoring applications.

6/ REFERENCES

- Alsberg BK, A.M. W, Kell DB. 1997. An introduction to wavelet transforms for chemometricians: A time-frequency approach. *Chemometr. Intell. Lab.* 37(2):215-239.
- Arnold SA, Gaensakoo R, Harvey LM, McNeil B. 2002. Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process. *Biotechnol. Bioeng.* 80(4):405-413.
- Bellon V. 1993. Fermentation control using ATR and an FT-IR spectrometer. *Sens. Actuator B-Chem.* 12(1):57-64.
- Blanco M, Peinado AC, Mas J. 2004. Analytical monitoring of alcoholic fermentation using NIR spectroscopy. *Biotechnol. Bioeng.* 88(4):539-542.
- Blank TB, Sum ST, Brown SD, Monfre SL. 1996. Transfer of near-infrared multivariate calibrations without standards. *Anal. Chem.* 68(17):2987-2995.
- Cannizzaro C. 2002. Spectroscopic monitoring of bioprocesses: a study of carotenoid production by *Phaffia rhodozyma* yeast: PhD Thesis No 2620, EPFL. 294 p.
- Cannizzaro C, Valentiniotti S, von Stockar U. 2004. On-line biomass monitoring of CHO perfusion culture with scanning dielectric spectroscopy. *Bioprocess. Biosyst. Eng.* 26(6):377-383.
- Crowe CM. 1996. Data reconciliation - Progress and challenges. *J. Process Contr.* 6(2-3):89-98.
- Crowley J, McCarthy B, Nunn NS, Harvey LM, McNeil B. 2000. Monitoring a recombinant *Pichia pastoris* fed batch process using Fourier transform mid-infrared spectroscopy (FT-MIRS). *Biotechnol. Lett.* 22:1907-1912.
- Dabros M, Amrhein M, Gujral P, von Stockar U. 2006. On-line recalibration of spectral measurements using metabolite injections and dynamic orthogonal projection. *Appl. Spectrosc.* 61(5):507-513.
- Dabros M, Gujral P, Amrhein M, Bonvin D, Marison IW, von Stockar U. 2007. On-line data reconciliation of mid-infrared and dielectric spectral measurements. *To be submitted.*
- de Kok HE, Roels JA. 1980. Method for the statistical treatment of elemental and energy balances with application to steady-state continuous culture growth of *Saccharomyces cerevisiae* in the respiratory region. *Biotechnol. Bioeng.* 22:1097-1104.
- Doak DL, Phillips JA. 1999. In situ monitoring of an *Escherichia coli* fermentation using a diamond composition ATR probe and mid-infrared spectroscopy. *Biotechnol. Prog.* 15:529-539.

- Duboc P, von Stockar U. 1995. Energetic investigation of *Saccharomyces cerevisiae* during transitions. Part 1. Mass balances. *Thermochim. Acta* 251(3):119-130.
- Duponchel L, Ruckebusch C, Huvenne JP, Legrand P. 1999. Standardisation of near infrared spectrometers using artificial neural networks. *J. near Infrared Spec.* 7(3):155-166.
- Fayolle P, Picque D, Corrieu G. 2000. On-line monitoring of fermentation processes by a new remote dispersive middle-infrared spectrometer. *Food Control* 11:291-296.
- Fayolle P, Picque D, Perret B, Latrille E, Corrieu G. 1996. Determination of major compounds of alcoholic fermentation by middle-infrared spectroscopy : study of temperature effects and calibration methods. *Appl. Spectrosc.* 50(10):1325-1330.
- FDA. 2005. Process Analytical Technology (PAT) Initiative. Available from: <http://www.fda.gov/cder/OPS/PAT.htm>.
- Ferreira AP, Alves TP, Menezes JC. 2005. Monitoring complex media fermentations with near-infrared spectroscopy: Comparison of different variable selection methods. *Biotechnol. Bioeng.* 91(4):474-481.
- Feudale RN, Woody NA, Tan HW, Myles AJ, Brown SD, Ferre J. 2002. Transfer of multivariate calibration models: a review. *Chemometr. Intell. Lab.* 64(2):181-192.
- Flores-Cerillo J, MacGregor JF. 2003. Within-batch and batch-to-batch inferential-adaptive control of semibatch reactors: A partial least squares approach. *Ind. Eng. Chem. Res.* 42(14):3334-3345.
- Geladi P, Macdougall D, Martens H. 1985. Linearization and scatter-correction for near-Infrared reflectance spectra of meat. *Appl. Spectrosc.* 39(3):491-500.
- Gemperline PJ, Long JR, Gregoriou VG. 1991. Nonlinear multivariate calibration using Principal Components Regression and Artificial Neural Networks. *Anal. Chem.* 63(20):2313-2323.
- Giaivas I, Robertson I, McNeil B, Harvey LM. 2003. Simultaneous and rapid monitoring of biomass and biopolymer production by *Sphingomonas paucimobilis* using Fourier transform-near infrared spectroscopy. *Biotechnol. Let.* 25(12):975-979.
- Goodacre R, Timmins EM, Jones A, Kell DB, Maddock J, Heginbotham ML, Magee JT. 1997. On mass spectrometer instrument standardization and interlaboratory calibration transfer using neural networks. *Anal. Chim. Acta* 348(1-3):511-532.
- Haack MB, Eliasson A, Olsson L. 2004. On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. *J. Biotechnol.* 114(1-2):199-208.

- Haaland DM, Thomas EV. 1988. Partial Least-Squares methods for spectral analyses .1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* 60(11):1193-1202.
- Hall JW, McNeil B, Rollins MJ, Draper I, Thomson BG, Macaloney G. 1996. Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial *Escherichia coli* fermentation. *Appl. Spectrosc.* 50(1):120-108.
- Herwig C, Marison IW, von Stockar U. 2001. On-line stoichiometry and identification of metabolic state in dynamic process conditions. *Biotechnol. Bioeng.* 75(3):345-354.
- Jungo C, Schenk J, Marison IW, von Stockar U. 2007. A quantitative analysis of the benefits of mixed feeds of sorbitol and methanol for the production of recombinant avidin with *Pichia pastoris*. *J. Biotechnol.* 131(1):57-66.
- Junker BH, Wang HY. 2006. Bioprocess monitoring and computer control: Key roots of the current PAT initiative. *Biotechnol. Bioeng.* 95(2):226-261.
- Kornmann H, Rhiel M, Cannizzaro C, Marison I, von stockar U. 2003. Methodology for real-time, multi-analyte monitoring of fermentations using an in-situ mid-infrared sensor. *Biotechnol. Bioeng.* 82(6):702-709.
- Kornmann H, Valentinotti S, Duboc P, Marison IW, von Stockar U. 2004a. Monitoring and control of *Gluconacetobacter xylinus* fed-batch cultures using in situ mid-IR spectroscopy. *J. Biotechnol.* 113(1-3):231-245.
- Kornmann H, Valentinotti S, Marison IW, von Stockar U. 2004b. Real-time update of calibration model for better monitoring of batch processes using spectroscopy. *Biotechnol. Bioeng.* 87(5):593-601.
- Manning CJ, Griffiths PR. 1997. Noise sources in step-scan FT-IR spectrometry. *Appl. Spectrosc.* 51(8):1092-1101.
- Mark HL, Griffiths PR. 2002. Analysis of noise in Fourier transform infrared spectra. *Appl. Spectrosc.* 56(5):633-639.
- Martens H, Naes T. 1997. *Multivariate Calibration*. New York: John Wiley & Sons.
- Mazarevica G, Diewok J, Baena JR, Rosenberg E, Lendl B. 2004. On-line fermentation monitoring by mid-infrared spectroscopy. *Appl. Spectrosc.* 58(7):804-810.
- Nadler B, Coifman RR. 2005. Partial least squares, Beer's law and the net analyte signal: statistical modeling and analysis. *J. Chemometr.* 19(1):45-54.
- Narasimhan S, Jordache C. 2000. *Data reconciliation & gross error detection*. Houston (TX): Gulf Publishing Company.

- November EJ, Van Impe JF. 2000. Evaluation of on-line viable biomass measurements during fermentations of *Candida utilis*. *Bioproc. Eng.* 23(5):473-477.
- Olsson L, Schulze U, Nielsen J. 1998. On-line bioprocess monitoring - an academic discipline or an industrial tool? *Trac-Trend. Anal. Chem.* 17(2):88-95.
- Pollard DJ, Buccino R, Connors NC, Kirschner TF, Olewinski RC, Saini K, Salmon PM. 2001. Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy. *Bioprocess. Biosyst. Eng.* 24:13-24.
- Rhiel MH, Amrhein MI, Marison IW, von Stockar U. 2002a. The influence of correlated calibration samples on the prediction performance of multivariate models based on mid-Infrared spectra of animal cell cultures. *Anal. Chem.* 74(20):5227-5236.
- Rhiel MH, Cohen MB, Arnold MA, Murhammer DW. 2004. On-line monitoring of human prostate cancer cells in a perfusion rotating wall vessel by near-infrared spectroscopy. *Biotechnol. Bioeng.* 86(7):852-861.
- Rhiel MH, Ducommun P, Bolzonella I, Marison IW, von Stockar U. 2002b. Real-time in situ monitoring of freely suspended and immobilized cell cultures based on mid-infrared spectroscopic measurements. *Biotechnol. Bioeng.* 77(2):174-185.
- Riley M, Arnold M, Murhammer DW. 1998. Matrix-enhanced calibration procedure for multivariate calibration models with near-infrared spectra. *Appl. Spectrosc.* 52(10):1339-1347.
- Riley MR, Okeson CD, Frazier BL. 1999. Rapid calibration of near-infrared spectroscopic measurements of mammalian cell cultivations. *Biotechnol. Prog.* 15(6):1133-1141.
- Rodig C, Siebert F. 1999. Errors and artifacts in time-resolved step-scan FT-IR spectroscopy. *Appl. Spectrosc.* 53(8):893-901.
- Romagnoli JA, Sanchez MC. 2000. Data processing and reconciliation for chemical process operations. San Diego: Academic Press.
- Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36(8):1627-1639.
- Schenk J, Marison IW, von Stockar U. 2007. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2):344-353.
- Schügerl K. 2001. Progress in monitoring, modeling and control of bioprocesses during the last 20 years. *J. Biotechnol.* 85:149-173.
- Sivakesava S, Irudayaraj J, Demirci A. 2001. Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. *J. Ind. Microbiol. Biot.* 26:185-190.

- Sjöblom J, Svensson O, Josefson M, Kullberg H, Wold S. 1998. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemometr. Intell. Lab.* 44(1-2):229-244.
- Thissen U, Pepers M, Ustun B, Melssen WJ, Buydens LMC. 2004. Comparing support vector machines to PLS for spectral regression applications. *Chemometr. Intell. Lab.* 73(2):169-179.
- Trygg J, Wold S. 2002. Orthogonal projections to latent structures (O-PLS). *J. Chemometr.* 16(3):119-128.
- van der Heijden RTJM, Romein B, Heijnen JJ, Hellinga C, Luyben KCAM. 1994. Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors. *Biotechnol. Bioeng.* 43(1):11-20.
- Wang NS, Stephanopoulos G. 1983. Application of macroscopic balances to the identification of gross measurement errors. *Biotechnol. Bioeng.* 25:2177-2208.
- Watari M, Ozaki Y. 2006. Practical calibration correction method for the maintenance of an on-line near-infrared monitoring system for molten polymers. *Appl. Spectrosc.* 60(5):539-538.
- Wilson RH, Trapp HS. 1999. Mid-infrared spectroscopy for food analysis: recent new applications and relevant developments in sample presentation methods. *Trac-Trend. Anal. Chem.* 18(2):85-93.
- Wold S. 2004. PAC, PAT and variable selection. Available from: <http://www.acc.umu.se/~tnkjtg/chemometrics/editorial/jan2004.html>.
- Wold S, Antti H, Lindgren F, Ohman J. 1998. Orthogonal signal correction of near-infrared spectroscopy. *Chemometr. Intell. Lab.* 44(1-2):175-185.
- Wold S, Trygg J, Berglund A, Antti H. 2001. Some recent developments in PLS modeling. *Chemometr. Intell. Lab.* 58(2):131-150.
- Wolthuis R, Tjiang GCH, Puppels GJ, Schut TCB. 2006. Estimating the influence of experimental parameters on the prediction error of PLS calibration models based on Raman spectra. *J. Raman. Spectrosc.* 37(1-3):447-466.
- Wülfert F, Kok WT, de Noord OE, Smilde AK. 2000. Linear techniques to correct for temperature-induced spectral variation in multivariate calibration. *Chemometr. Intell. Lab.* 51:189-200.
- Wülfert F, Kok WT, Smilde AK. 1998. Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Anal. Chem.* 70(9):1761-1767.

- Yardley JE, Todd R, Nicholson DJ, Barrett J, Kell DB, Davey CL. 2000. Correction of the influence of baseline artefacts and electrode polarisation on dielectric spectra. *Bioelectrochemistry* 51(1):53-65.
- Zeaiter M, Roger JM, Bellon-Maurel V. 2006. Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations. *Chemometr. Intell. Lab.* 51(2):189-200.
- Zhang L, Small GW, Arnold MA. 2002. Calibration standardization algorithm for partial least-squares regression: Application to the determination of physiological levels of glucose by near-infrared spectroscopy. *Anal. Chem.* 74(16):4097-4108.

7/ LIST OF SYMBOLS

Matrices

X	Spectral data matrix (p spectra measured, at wn different wavenumbers)
X^*	Corrected spectral data matrix (p spectra measured, at wn different wavenumbers)
Y	Matrix of estimated variables (i.e. concentrations of different compounds)
T	Scores matrix of X

Other symbols

X_r	Value of the spectrum at the anchor point
X_{r_0}	Value of the reference spectrum at the anchor point

Supplement
Part B

Influence of Specific Growth
Rate on Specific Productivity
and Glycosylation of a
Recombinant Avidin Produced
by a *Pichia pastoris* Mut⁺ Strain

1/ ABSTRACT

A recombinant avidin-producing Mut⁺ *Pichia pastoris* strain was used as a model organism to study the influence of the methanol feeding strategy on the specific product productivity (q_p) and protein glycosylation. Fed-batch cultivations performed at various specific growth rates (μ) and residual methanol concentrations showed that the specific avidin productivity is growth-dependent. The specific productivity increases strongly with the specific growth rate for μ ranging from 0 to 0.02 h⁻¹, and increases only slightly with the specific growth rate above this limit. N-terminal glycosylation was also found to be influenced by the specific growth rate, since 9-mannose glycans were the most abundant form at low growth rates, whereas 10-mannose carbohydrate chains were favored at higher μ . These results show that culture parameters, such as the specific growth rate, may significantly affect the activity of glycoproteins produced in *Pichia pastoris*. In terms of process optimization, this suggests that a compromise on the specific growth rate may have to be found, in certain cases, to work with an acceptable productivity while avoiding the addition of many mannoses.

This chapter was published in *Biotechnology and Bioengineering*:

Schenk J, Balazs K, Jungo C, Urfer J, Wegmann C, Zocchi A, Marison IW, von Stockar U. 2007. Influence of specific growth rate on specific productivity and glycosylation of a recombinant avidin produced by a *Pichia pastoris* Mut⁺ strain. *Biotechnol. Bioeng.* *In press.*

2/ INTRODUCTION

Pichia pastoris is a popular expression system for the production of heterologous proteins at both laboratory and large scale. Many characteristics make this yeast an interesting host organism for the production of recombinant proteins (Cereghino and Cregg, 2000; Macauley-Patrick et al., 2005; Sreekrishna et al., 1997): i) techniques for genetic modifications are available, ii) this yeast is capable of secreting proteins, iii) post-translational modifications are possible, iv) foreign proteins have been produced at very high levels, both intra- and extracellularly.

In general, the production process consists of three phases, as recommended by Invitrogen in the *Pichia* Fermentation Process Guidelines (Invitrogen, 2002). The first is a batch phase on glycerol that lasts for about 24 hours. During this phase, protein expression is repressed, but biomass concentration increases rapidly. The second phase, which consists in a limited fed-batch on glycerol, is much shorter (4 hours) and aims at further increasing biomass concentration, while preparing the cells for substrate change. At this stage, protein expression is induced by methanol feeding. This last part of the process brings the biomass concentration to $>150 \text{ g L}^{-1}$ of dry cell weight within 2 days.

During this latter phase methanol acts both as carbon source and inducer of protein expression, therefore complete methanol depletion is undesirable, while accumulation to toxic levels must be also avoided. As a consequence, numerous studies have been performed on how the feed profile can be optimised during the induction phase (Cos et al., 2006)). Some authors reported an enhanced productivity when controlling the residual methanol concentration at 0.4 g L^{-1} (Hellwig et al., 2001), 3.7 g L^{-1} (Zhang et al., 2000), or $4\text{--}30 \text{ g L}^{-1}$ (Stratton et al., 1998), whereas some others showed that the specific growth rate can also influence protein expression (Cunha et al., 2004; Kobayashi et al., 2000; Ohya et al., 2005; Zhang et al., 2005).

In contrast, less information is available on the glycosylation performed by recombinant *Pichia pastoris* strains, because in general yeast have been mainly used to produce proteins that do not require proper glycosylation to be active or to have a therapeutic efficiency, such as insulin (Gerngross, 2004). N-glycosylation in yeast has a composition of $\text{Man}_n\text{GlcNAc}_2$ (Man: Mannose; GlcNAc: N-acetylglucosamine), where n is the number of mannose included in the structure. This number has been found to vary in *Pichia pastoris* from 3 to 17, depending on the expressed protein, but is generally within the range 8-10 (Daly and Hearn, 2005; Montesino et al., 1998). A higher number of mannose residues, which is referred to as

hyperglycosylation, is rarely observed in *Pichia pastoris*, as opposed to *S. cerevisiae*. While major glycosylation changes and optimisations are generally effectuated at the genome level, some bioprocess parameters can also influence the structure and homogeneity of the carbohydrate moiety, both during cultivation and purification (Goochee and Monica, 1990; Jenkins et al., 1996). The effect on glycosylation of cultivation parameters such as glucose starvation, pH, perfusion rate, temperature or batch length have been investigated in details for mammalian cells (Andersen et al., 2000; Meuwly et al., 2004; Yuen et al., 2003). Similar studies on product quality were performed with *Pichia pastoris*, but rather than investigating glycosylation, they concentrated on protein activity or level of aggregation. For instance, it has been reported that the product quality can be improved by maintaining a high residual methanol concentration (Kahtri and Hoffman, 2006), growing cells at low rates (Cunha et al., 2004), lowering the cultivation temperature (Hong et al., 2002), or reducing stirring rate and aeration (Woo et al., 2006). The recent advances in humanizing yeast glycosylation and the need to bring to the market therapeutic glycoproteins produced in *Pichia* is likely to stimulate such studies focusing on environmental factors that influence product quality and glycosylation, as is the case with mammalian cells.

The recombinant *Pichia pastoris* strain used in this study produces avidin, a glycoprotein found in egg white, which can form an extraordinary strong complex with the water-soluble vitamin biotin ($K_a > 10^{14} \text{ M}^{-1}$). This unusual particularity has been extensively used in molecular biology and analytical chemistry (Wilchek and Bayer, 1990) and for this reason, the avidin structure and activity has been well characterized. Avidin is a tetrameric protein, composed of four subunits of approximately 16.4 kDa each of which can bind a biotin molecule and have a single glycosylation site, on Asn₁₇ (DeLange, 1970). The carbohydrate unit in natural occurring avidin is known to be heterogeneous, but generally comprises an average of four to five mannoses (Man) and three N-acetylglucosamine (GlcNAc) residues (Bruch and White, 1982). The glycan chain is not involved in the avidin-biotin activity (Hiller et al., 1987), and therefore the glycosylation profile of avidin is not of particular interest regarding applications based on the biotin affinity. However, the carbohydrate moiety has been identified as playing an important role in more complex *in vivo* systems, such as avidin liver uptake (Yao et al., 1999) or blood clearance (Rosebrough and Hartley, 1996).

Recombinant avidin has been produced in *E. coli* in the non-glycosylated form (Airenne et al., 1994; Nardone et al., 1998). Eukaryotic systems such as baculovirus-infected insect cells (Airenne et al., 1997) or maize (Hood et al., 1997), which are able to perform post-translational modifications, have also been used to produce recombinant avidin. Both of these studies

reported that the glycosylation of the recombinant avidin was heterogeneous and differed significantly from that found in egg white avidin, however, no further quantification was possible, due to the relatively low precision of the characterization method (SDS-PAGE).

The construction of the recombinant avidin-producing *Pichia pastoris* used in this study as well as the regulation of the protein expression in continuous cultures has been described in two previous publications (Jungo et al., 2006; Zocchi et al., 2003). This article focuses on how the specific growth rate can influence the avidin specific productivity and its glycan chain, during fed-batch cultivations. Beyond the specific scope of the current avidin case, this article aims at discussing the issues of methanol feeding and specific growth rate control, under the perspective of optimizing both protein productivity and glycosylation.

3/ MATERIALS AND METHODS

Strain and Inoculum Preparation

The construction of a recombinant avidin (with lowered isoelectric point, pI 5.4) secreting Mut⁺ strain of *Pichia pastoris* has been described previously (Zocchi et al., 2003). A 30-vial cell bank was used for seeding all cultivations. The bank was prepared by inoculation of a flask containing 100 mL of YPG medium (6 g L⁻¹ yeast extract, 5 g L⁻¹ peptone, 20 g L⁻¹ glycerol) and incubation at 30°C. After 24 h the culture was centrifuged 10 min at 1500 g, then resuspended in 36 mL of saline glycerol solution (9 g L⁻¹ NaCl, 20 g L⁻¹ glycerol), followed by aliquoting in 2 mL Nalgene tubes (Rochester, NY) and stored at -80°C. For each culture, an inoculum was prepared from a new vial from the cell bank, grown for 24 hours in 100 mL of YPG medium at 30°C and 200 rpm. Cells were then centrifuged (10 min at 1500 g) and resuspended in 10 mL of UHP water for fermentor inoculation.

Bioreactor set-up and cultivations conditions

The Invitrogen “basal salts medium” (Invitrogen, 2002), supplemented with EDTA to avoid precipitation, was used for all experiments. The medium contained, beside carbon source: 26.7 mL L⁻¹ H₃PO₄ 85% in weight, 0.93 g L⁻¹ CaSO₄·2H₂O, 18.2 g L⁻¹ K₂SO₄, 14.9 g L⁻¹ MgSO₄·7H₂O, 4.13 g L⁻¹ KOH, 0.8 g L⁻¹ EDTA·Na₂·2H₂O, NH₄OH 28% in weight to adjust pH at a value of 3.0, 5.0 or 6.0). After filter-sterilization (0.2 µm, Steritop®, Millipore, MA), 0.4 mL of antifoam (Struktol SB2121, Schill and Seilacher, Hamburg, Germany), 125 µg of biotin (125 µL of 1 g L⁻¹ biotin stock solution in 1 mol L⁻¹ NaOH) and 4.35 mL of sterile trace elements solutions PTM were added (PTM: 5 mL L⁻¹ H₂SO₄ 98% in weight, 0.02 g L⁻¹ H₃BO₄, 6 g L⁻¹ CuSO₄·5H₂O, 0.08 g L⁻¹ NaI, 3 g L⁻¹ MnSO₄·H₂O, 0.2 g L⁻¹ Na₂MoO₄·2H₂O, 0.5 g L⁻¹ Ca₂SO₄·2H₂O, 20 g L⁻¹ ZnCl₂, 65 g L⁻¹ FeSO₄·7H₂O). 20 g L⁻¹ of glycerol was supplemented to the medium for the experiments using pre-defined feeding strategy. All chemicals were from Sigma-Aldrich (St. Louis, MO).

Cultivations were carried out at 30°C in a fully automated 2-litre bioreactor (RC1, Mettler-Toledo, Greifensee, Switzerland). pH was controlled by the addition of NH₄OH 28% w/w, which also served as nitrogen source. Off-line analysis of preliminary experiments showed that, by that means, nitrogen was never limiting. Agitation was set at 1000 rpm, with two 6-blade Rushton impellers, giving a power input of more than 10 W L⁻¹ (torque measurement). Total aeration was controlled to provide 2 NL min⁻¹ (NL refers to a litre of gas at normal temperature and pressure), and dissolved oxygen was maintained above 80% air

saturation by blending air and pure oxygen using thermal mass flow controllers (MFC 5850E, Brooks Instrument, Emerson Process Management, Ede, The Netherlands). The maximum cooling capacity of the reactor was approximately 140 W L⁻¹. Oxygen and carbon dioxide composition of the exhaust gas was measured using a gas analyzer (Dr. Marino Müller Systems, Esslingen, Switzerland).

Specific growth rate control strategy

Two different feeding strategies were followed depending on the specific growth rate to be studied. For high specific growth rates ($\mu > 0.08 \text{ h}^{-1}$), the control was effectuated on the residual methanol concentration using on-line mid-infrared spectroscopy, as described in a previous publication (Schenk et al., 2007).

Specific growth rates lower than 0.08 h^{-1} were studied using a pre-defined exponential feeding rate strategy. Pre-defined exponential feeding rates F , in grams of methanol per hour, were determined using the equation:

$$F = F_0 \cdot \exp(\mu \cdot t) \quad [1]$$

where μ is the specific growth rate to be studied. The constant F_0 depends on the biomass to substrate yield $Y_{X/S}$, the specific growth rate set point and the initial amount of biomass:

$$F_0 = \frac{\mu}{Y_{X/S}} \cdot X_0 \cdot V_0 \quad [2]$$

where X_0 and V_0 are the initial biomass concentration (g L⁻¹) and bioreactor volume (L) respectively. $Y_{X/S}$ was assumed to be constant on methanol with the value of 0.45 C-mol C-mol⁻¹, as measured from preliminary experiments. Accurate control of the pre-defined feeding rate was achieved by placing the feeding solution (composition: pure methanol supplemented with 12 mL L⁻¹ PTM solution) on a scale and controlling the addition by a PI feedback controller. Initial feeding rates (F_0) were underestimated by 10% to avoid methanol accumulation at the beginning of the induction phase. In order to account for $Y_{X/S}$ changes, especially at low growth rates, the specific growth rates were determined from off-line measurement of the biomass concentration.

For experiments with a pre-defined feeding rate, cells were first grown on 40 g L⁻¹ of glycerol in order to get a reasonable amount of biomass for the induction phase. For the two experiments at very low specific growth rates (namely 0.001 and 0.002 h⁻¹), biomass concentration was further increased after the initial batch phase to ca. 80 g L⁻¹ dry cell weight by a limited fed-batch phase on glycerol as defined by Equation 1-2, at a specific growth rate set point of 0.2 h⁻¹ (feed composition: pure glycerol with 12 ml L⁻¹ PTM solution). The

induction phase lasted until maximum cooling capacity of the bioreactor was reached, or when biomass concentration was equal to 140 g L⁻¹.

Off-line analysis

5 mL samples were collected regularly and immediately cooled to 4°C, using a custom-made automatic sampling device (Cannizzaro, 2002). Sample treatment was done within 12 hours of sampling. Dry cell weight concentration (g L⁻¹) was determined by gravimetric analysis: 3 mL of sample were filtered on pre-weighed 0.2 µm filters, then dried at 100°C to constant weight. At biomass concentrations higher than 10 g L⁻¹, this analysis was performed as following: 3 mL of sample were centrifuged for 30 minutes at 1200 g in a pre-weighed glass tube, supernatant was discarded and biomass dried to constant weight at 100°C. The remaining volume of the sample was filtered (0.2 µm) for recombinant avidin and HPLC analysis (1100 series, Agilent, Palo Alto, CA). Methanol concentration was determined by ion-exchange chromatography (Supelcogel H 300 mm, Supelco, Bellefonte, PA), using isocratic elution at 0.6 mL min⁻¹ with a solution of 5 mM sulfuric acid. Detection limit was 0.05 g L⁻¹. Recombinant avidin concentration was determined by titration with biotin-4-fluorescein (Fluka, Buchs, Switzerland), following the published procedure (Kada et al., 1999). Preliminary concentration (2 to 4 times) of samples by ultrafiltration was necessary (Centriplus 30kDa, Millipore, Billerica, MA).

During high-cell density cultivations, up to 65% of the total broth volume could be occupied by biomass volume. For this reason, component concentrations measured in filtrated samples exceeded the concentrations of the overall culture broth. This effect was taken into account by calculating the avidin concentration in the culture broth from the concentration found by titration of the supernatant using:

$$C_{\text{broth}} = C_{\text{supernatant}} \cdot \frac{V_{\text{supernatant}}}{V_{\text{broth}}} \quad [3]$$

The volume ratio was determined by measuring the supernatant volume after centrifugation at 14000 g for 2 minutes of 1 mL of cultivation sample. This correction, if not performed, might lead to overestimated protein titers, which explains the significant difference between our current results and previously published data (Zocchi et al., 2003).

Data consistency of cultivations results

Check for data consistency was carried out for every run, following a standard procedure (Stephanopoulos et al., 1998). Methanol, carbon dioxide and biomass yields were used for the

reconciliation, with an error range of 3% on each. All runs shown in this article passed the statistical test with a confidence interval set at 95% (redundancy = 1, threshold $\chi^2 = 3.84$). Biomass composition used in this calculation was determined by elemental analysis to be $\text{CH}_{1.52}\text{O}_{0.51}\text{N}_{0.18}$ (Gurakan et al., 1990).

Recombinant avidin purification and characterization of glycosylation

At the end of the run, the culture broth was immediately cooled down to 10°C within the reactor and then centrifuged 15 minutes at 3000 g. 250 mL of supernatant was collected, filtered at 0.2 μm and stored at -20°C. The thawed harvest was then concentrated to 50 mL (membrane cut-off: 1000 Da; Millipore, MA) and diafiltrated with 250 mL of exchange buffer (NaCl 0.5 M, NaHCO_3 0.05 M, pH = 9.8 adjusted with NaOH 1 M). The retentate was further clarified by 20 min of centrifugation at 47000 g (at 4°C) followed by filtration at 0.2 μm . The clear solution was loaded on a 2-*iminobiotin*-agarose column (Affiland, Ans-Liege, Belgium) and eluted with acetic acid 100 mM (pH adjusted to a value of 3.9). Elution was monitored by UV-VIS spectroscopy in order to collect protein-containing fractions. A final dialfiltration was used to exchange the elution buffer with 200 mL of UHP water and to remove the remaining salts. The resulting solution was frozen at -80 °C and freeze-dried (Secfroid, Switzerland) to yield recombinant avidin in a white powder form, which was stored at 4 °C. The glycosylation of the purified avidin was analyzed by liquid chromatography-mass spectrometry. The samples were introduced to the mass spectrometer by means of a Waters modular CapLC system (Waters, Ruppertswil, Switzerland), comprising a low flow capillary LC pump and an autosampler. A Stream Select module, attached directly to the Z spray source, was configured with one trapping pre-column Symmetry 300TM (300 μm ID x 5 mm) with a C18 stationary phase. Samples were loaded onto this pre-column Symmetry (300 μm ID x 5 mm) and desalted with 0.1% formic acid for 3 minutes at a flow rate of 30 μL per minute. The Stream Select module was then switched such that the samples were eluted directly, at a flow rate of 10 mL min^{-1} for 7 min with the 0.1% formic acid in acetonitrile-water (80/20 in volume), onto the Q-ToF Ultima API mass spectrometer (Micromass, Manchester, UK) equipped with a Z-spray type ESI source. Phosphoric acid was used for the positive ion mass calibration range of 100-2000 m/z . Full control of the pump operation and electrospray optimisation were performed through Masslynx. The mass spectrometer were operated in the positive ion mode with following conditions: capillary voltage, 3.0 kV; source temperature, 100°C; cone voltage, 80 V; and source block temperature, 200°C; and a counter current gas flow rate of 50 L h^{-1} . The ESI nebulization and drying gases

were nitrogen. Data were acquired and processed using Masslynx version 4.0. Due to the nature of samples investigated for non-covalent work, multiply charged envelopes are expected result for ESI acquisitions. In order to transform the multiply charged series into the equivalent uncharged spectrum a deconvolution tool must be used, using Maximum Entropy processing applied to mass spectrometry data. MaxEnt aids the enhancement of complex spectra, and MaxEnt allows the deconvolution of overlapping multiply-charged spectra produced by the analysis of protein mixtures and glycosylated proteins.

4/ RESULTS AND DISCUSSION

Effect of specific growth rate on avidin specific productivity

Under carbon-limited conditions, the specific growth rate and residual substrate concentration cannot be chosen independently, and a wide variety of models can be used to describe this relationship, the most common being Andrews to account for substrate inhibition (Andrews, 1968). Our previous results concerning this recombinant Mut⁺ *Pichia pastoris* strain showed an Andrews-like relation, since growth was found to be inhibited by methanol concentrations higher than 6 g L⁻¹ (Schenk et al., 2007). In this previous work, the maximum specific growth rate (μ_{\max}) was observed for residual methanol concentration from 2 to 6 g L⁻¹ and was determined to be 0.14 h⁻¹. This value is very close to the μ_{\max} of the wild type, even though recombinant protein production often lowers the maximum specific growth rate (Cos et al., 2006).

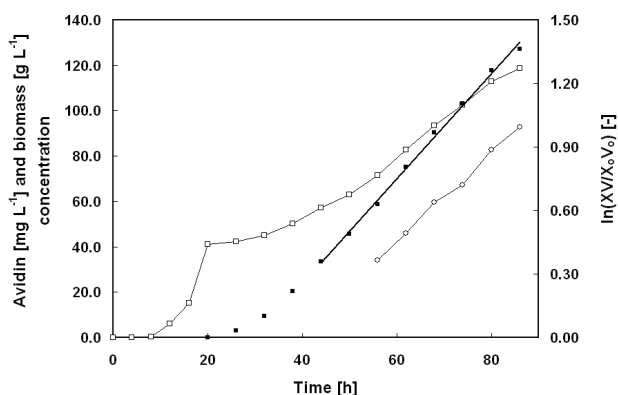


Figure 1: Avidin (open circles, left axis) and biomass concentration in dry cell weight per litre (open squares, left axis), during an experiment with pre-defined feeding of methanol. The natural logarithm of the total amount of biomass (black squares, right axis) was used to determine the specific growth case. In this case, μ , which is given by the slope of the linear regression, was equal to 0.025 h⁻¹ ($R^2 = 0.998$). Avidin specific productivity was found constant, and determined from the 6 points displayed.

In order to study the specific avidin productivity and its glycosylation over the complete range of specific growth rates and methanol concentrations, two different control strategies were necessary. For μ lower than 0.08 h⁻¹ (ca. 60% of μ_{\max}), a direct control of the specific growth rate was operated by applying a pre-defined exponential feeding rate. However, this strategy turns out to be unsuitable at specific growth rates close to μ_{\max} , because a small error

on the estimation of the parameters from Equations 1-2 can lead to significant methanol accumulation and therefore highly unsteady situations. As a consequence, for specific growth rates higher than 0.8 h^{-1} , the control was effectuated by on-line mid-infrared spectroscopy on the residual methanol concentration, which in turn dictated the specific growth rate.

The biomass and avidin concentration profiles during an experiment with pre-defined feed control were reported in Figure 1. After the glycerol batch phase, which lasted for about 20 hours and brought biomass concentration up to 20 g L^{-1} , induction phase with pre-defined methanol feeding was started. The biomass and avidin concentration profiles did not show an exponential shape due to the broth volume increase, which dilutes metabolites. During this phase, the residual methanol concentration was so low that it was practically non-detectable. The natural logarithm of the total amount of biomass (i.e. $X \cdot V$), determined from off-line analyses and accurate calculation of bioreactor volume, was used to calculate the effective specific growth rate, by linear regression (Fig. 1, right axis). About 15 hours after induction, μ remained very constant until the end of the culture. In the present case, the specific growth rate was equal to 0.025 h^{-1} ($R^2 = 0.998$). Avidin concentration was measured from 30 hours after induction on methanol, and until the end of the process. For all runs, the specific avidin productivity was always found constant during that time. More details on cultures performed with methanol control can be found in a previous publication (Schenk et al., 2007).

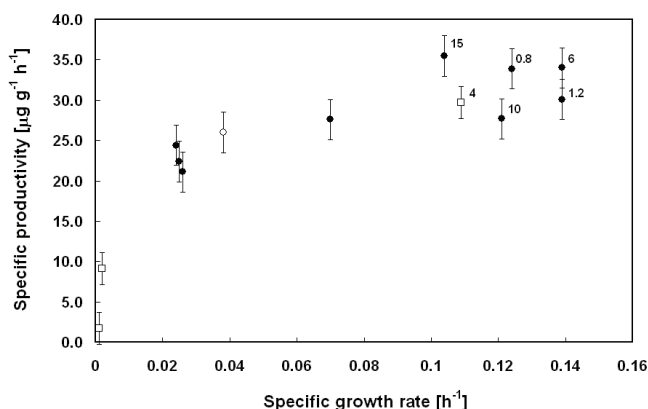


Figure 2: Specific productivity of recombinant avidin as a function of the specific growth rate, in fed-batch cultures performed at a pH of 5 (black circles), 3 (open squares) or 6 (open circle). Figures on the graph indicate experiments with control of residual methanol concentration and the corresponding set point.

The avidin specific productivity was determined for 13 fed-batch experiments performed at specific growth rate between 0.001 h^{-1} and 0.14 h^{-1} (equal to μ_{\max}), with residual methanol concentrations of up to 15 g L^{-1} (Fig. 2). Two of these runs were carried out at pH 3, one at pH 6 and the rest at a pH 5, in order to also investigate the influence of the medium pH on the protein expression (Fig.2).

The specific avidin productivity was found to depend strongly on the specific growth rate, whereas other parameters, namely pH and residual methanol concentration, showed no significant influence. The residual methanol appeared to be unrelated to the specific productivity at concentrations higher than 0.8 g L^{-1} , which corresponds to $\mu > 0.12\text{ h}^{-1}$. For specific growth rates higher than about 0.02 h^{-1} , q_p increased slightly with μ , and was within the range $25 - 35\text{ }\mu\text{g g}^{-1}\text{ h}^{-1}$. However, a large decrease in specific productivity was observed at specific growth rates below 0.02 h^{-1} , and almost no avidin was produced when the specific growth rate reached zero. These results are in agreement with those previously published with the same strain, in chemostat cultures (Jungo et al., 2006). However, since the response time of continuous culture is extremely long at small dilution rates, investigation of specific growth rates lower than 0.035 h^{-1} was not performed, which did not allow observing the drop of specific productivity at low specific growth rates. Such a decrease in specific productivity at low specific growth rate has already been observed with *P. pastoris* Mut⁺ strains (Ohya et al., 2005; Zhang et al., 2005), but other studies have reported very different behaviors, including a decrease of specific productivity with increasing specific growth rate (Cunha et al., 2004; Kobayashi et al., 2000). Experiments performed at a pH of 3 and 6 were aligned on the general trend, which shows that the pH had no significant effect on the specific productivity of the current strain.

Methanol acts both as carbon source and inducer of the protein expression promoter, which means that at low specific growth rates the promoter probably compete with the central metabolic demand for methanol, when the residual substrate concentration becomes very low, as predicted by the Monod relationship. This effect has already been reported with *Pichia pastoris* Mut⁺ strains, but more in terms of residual methanol concentration than in terms of specific growth rates. Indeed, it is generally recommended to control the residual methanol concentration at a non-limiting level, which was reported to be 0.4 g L^{-1} (Hellwig et al., 2001), 3.7 g L^{-1} (Zhang et al., 2000), or $4\text{-}30\text{ g L}^{-1}$ (Stratton et al., 1998). However, it must be pointed out that this implies that cells grow at μ_{\max} if they are not limited by an other substrate (nitrogen, oxygen, etc.), which as a consequence results in cultivations not lasting longer than 24 h, since in this time the biomass concentration should reach 150 g L^{-1} . It might

be very difficult to achieve the required oxygen and heat transfer rates for such a rapid growth.

In terms of process design, the decrease of the avidin specific productivity at low μ can turn out to be very problematic at large scale, or more generally with any bioreactor having a poor cooling power. For a industrial-scale bioreactor with a typical cooling capacity of 30 W L⁻¹, the power released by the cell metabolism can not exceed about 25 W L⁻¹, since at least 5 W L⁻¹ are also dissipated in stirring. A temperature increase in the medium can be avoided if the following relation, which is derived from an enthalpy balance, is satisfied:

$$\mu \leq \frac{Q_{cool} \cdot Y_{X/S}}{X \cdot Y_{Q/S}} \quad [4]$$

where Q_{cool} refers to the cooling power, in the current case 20 W L⁻¹. X is the biomass concentration, $Y_{X/S}$ is the biomass to substrate yield and $Y_{Q/S}$ is the heat to substrate yield. Assuming that $Y_{Q/S}$ on methanol is equal to 510 kJ C-mol⁻¹, $Y_{X/S}$ is equal to 0.45 C-mol C-mol⁻¹ and that the biomass C-molar weight is equal to 25 g C-mol⁻¹ (Jungo et al., 2006), the maximum allowed μ is about 0.02 h⁻¹ at a biomass concentration of 100 g L⁻¹. This implies that with such a bioreactor the specific growth rate would have to be kept in a range where the specific avidin productivity is low, and this already far below the typical biomass concentration target (i.e. 150 g L⁻¹). As a result, overall process performance would be seriously impaired.

Effect of specific growth rate on avidin glycosylation

The N-glycosylation profile of purified avidin was determined by LC-MS. Peak intensities were normalized by the highest intensity, in order to facilitate comparison between runs (Fig. 3). The molecular weight of non-glycosylated avidin was computed to be 14490 Da, whereas the molecular weight of a glycosylated form with n mannoses Avidin-Man $_n$ -GlcNAc $_2$ is given by 14896 + $n \cdot 162$ (Man: Mannose; GlcNAc: N-acetyl-glucosamine). The most abundant forms found were the 9-mannose (16354 Da, indicated by a star in Figure 3) and the 10-mannose (16516 Da). For such glycoproteins, the oligosaccharide chain amounts to more than 10% of the total weight. Using normalized intensities and the predicted molecular weights, the relative fraction of the glycosylated forms were attributed on the basis of spectrograms as shown in Figure 3.

Since a high number of steps were involved in the production, purification and analysis of recombinant avidin, the reproducibility of the whole process in terms of glycosylation profile had to be addressed. In particular, since the carbohydrate moiety is at the protein

surface, it may change the solubility. Modification of the purification procedure (especially concerning membrane and chromatography steps) are known to influence the distribution of glycoforms (Goochee et al., 1991). Two complete process duplicates were therefore performed in order to test the reproducibility of the determination of glycosylation distributions. The fed-batches were operated at 0.024 and 0.025 h^{-1} ($q_p = 24.4$ and $22.4 \mu\text{g g}^{-1} \text{h}^{-1}$ respectively) and purified separately, with a six month delay between each other, during which all other process harvests were purified and analyzed. The glycosylation distributions were found to be very similar, with a maximum difference of 13% between the two (Fig. 4). This certainly indicates that the entire assay, including the culture, purification steps and MS measurement, was very reproducible in terms of glycosylation distribution.

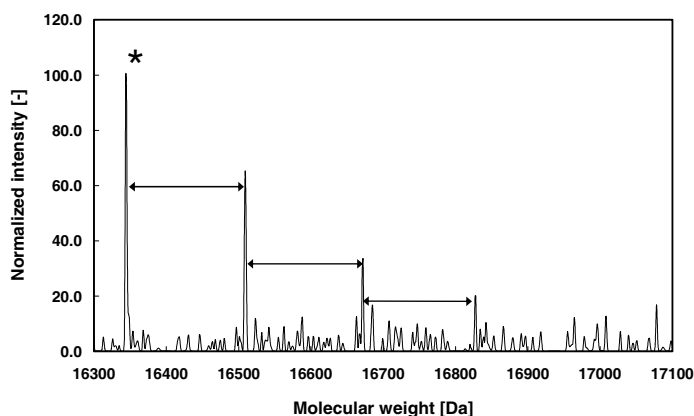


Figure 3: Mass spectroscopy spectrum of the fed-batch grown at a very low specific growth rate ($\mu = 0.001 \text{ h}^{-1}$). Intensity was normalized so that the highest peak is equal to 100. The peak indicated by a star can be attributed to a 9-mannose glycosylation, with a molecular weight of 16354 Da (non-glycosylated avidin: 14490 Da). Arrows represent gaps of 162 Da, the mass of an additional mannose.

In addition to the duplicates ($\mu = 0.025 \text{ h}^{-1}$, $\text{pH} = 5$), four representative cultivation harvests were purified and the resulting recombinant avidin characterized by mass spectroscopy (Fig. 5). One was operated at a very low specific growth rate and small specific productivity ($\mu = 0.001 \text{ h}^{-1}$, $\text{pH} = 3$), one at an intermediary specific growth rate ($\mu = 0.038 \text{ h}^{-1}$, $\text{pH} = 6$), and two at high specific growth rates under control of residual methanol concentration, but at different pH values ($\mu = 0.109 \text{ h}^{-1}$, $\text{pH} = 3$, $C_{\text{methanol}} = 4 \text{ g L}^{-1}$ and $\mu = 0.139 \text{ h}^{-1}$, $\text{pH} = 5$, $C_{\text{methanol}} = 6 \text{ g L}^{-1}$). Knowing that avidin has a single N-glycosylation site (DeLange, 1970), the fractions of protein carrying each glycan structure could be obtained directly from

the relative intensities of the MS peaks. The 9-mannose glycosylation form was the most abundant at very low specific growth rate and amounted to about 50% of the total purified avidin, whereas the 10-mannose fraction was 32%. The duplicates, which were operated at a μ of 0.025 h^{-1} , presented a less pronounced predominance of the 9-mannose form (44%) and more 10-mannose structures (37%). All the runs conducted at higher specific growth rates ($\mu > 0.038\text{ h}^{-1}$) showed a higher fraction of 10-mannose than 9-mannose forms. Zocchi and coworkers, who followed the Invitrogen protocole, for which $\mu < 0.03\text{ h}^{-1}$ during most of the induction phase, also reported for the same recombinant strain a clear predominance of the 9-mannose form (Zocchi et al., 2003).

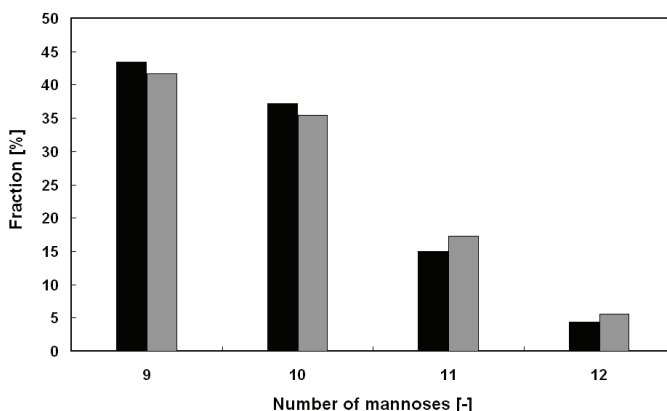


Figure 4: Glycosylation distributions of the recombinant avidin, determined by LC-MS after purification. The number of mannose refers to n in the structure $\text{Man}_n\text{-GlcNAc}_2$. The black and grey bars represent overall process duplicates, with specific growth rate of 0.024 and 0.025 h^{-1} and specific avidin specific productivities of 24.4 and $22.4\text{ }\mu\text{g g}^{-1}\text{ h}^{-1}$ respectively.

The typical protein glycosylation in *Pichia pastoris* is $\text{Man}_{8-9}\text{-GlcNAc}_2$ (Cereghino and Cregg, 2000; Daly and Hearn, 2005), which is in agreement with the distributions observed with the current avidin-producing Mut^+ strain. More highly mannosylated glycans such as $\text{Man}_{10-12}\text{-GlcNAc}_2$ were also found, amounting to about 25% of the total glycoproteins. Such heterogeneity in glycosylation is rather common with *Pichia pastoris* (Choi et al., 2003; Grinna and Tschopp, 1989; Reddy and Dahms, 2002; Trimble et al., 1991), but also in natural egg white avidin (Bruch and White, 1982).

These results show that for the studied recombinant strain, increasing the specific growth rate favors additional mannose on the sugar moiety, whereas pH and residual methanol

concentration did not significantly influence the glycosylation. A similar observation was already made for q_p , which suggests that mannose addition is stimulated when the specific avidin productivity is high. Following up on the short large-scale production example given above, it can be seen that going to industrial scale, and therefore limiting the specific growth rate to low levels, would not only alter specific productivity of this strain, but also affect the sugar moiety. With respect to avidin specifically, the implications of these results are relatively limited, since glycosylation is not involved in the formation and stability of the biotin-avidin complex (Hiller et al., 1987), and avidin is in most of the cases used for *in-vitro* bioanalytical purposes. These findings are, however, more relevant for *in vivo* applications, such as tumor targeting, for which the protein half-life depends on the structure of the carbohydrate moiety (Yao et al., 1999).

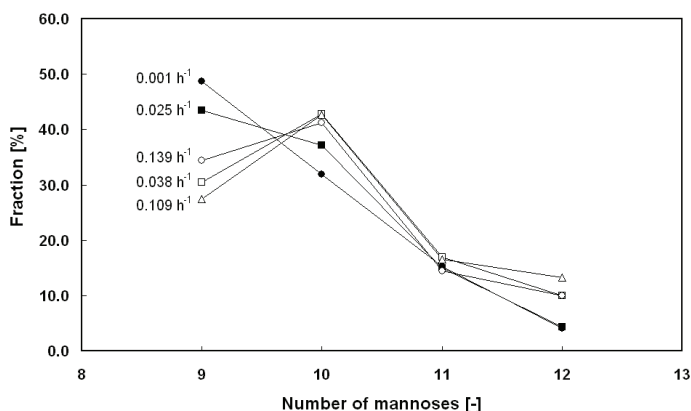


Figure 5: Glycosylation distribution of recombinant avidin, determined by LC-MS after purification, for different specific growth rates during cultivation. Low avidin specific productivities (black circles: $1.7 \mu\text{g g}^{-1} \text{h}^{-1}$; black squares: $22.4 \mu\text{g g}^{-1} \text{h}^{-1}$) tend to favor the 9-mannose form, whereas higher avidin specific productivities (open squares: $26.6 \mu\text{g g}^{-1} \text{h}^{-1}$; open triangles: $29.7 \mu\text{g g}^{-1} \text{h}^{-1}$; open circles: $34.0 \mu\text{g g}^{-1} \text{h}^{-1}$) lead to a predominant 10-mannose form.

Beyond the scope of the specific avidin case, this work should be seen in the more general framework of the production of therapeutic glycoproteins in yeast. The recent spectacular advances in the humanization of *Pichia pastoris* glycosylation is likely to encourage the production in yeast of more and more high added-value proteins (Wildt and Gerngross, 2005). Although glycoengineered strains have been reported to exhibit a high glycan uniformity (Gerngross, 2004; Li et al., 2006), our results suggest that bioprocess development

should include already at an early stage a study of the influence of upstream parameters, and especially the specific growth rate, on the structure of the sugar moieties.

5/ CONCLUSIONS

The specific productivity and the glycosylation of recombinant avidin have been studied as a function of three cultivation parameters, namely the specific growth rate, residual methanol concentration and pH of culture medium. It has been shown that the residual methanol concentration during fed-batch culture and the pH of the medium had little effect on the avidin expression, both in terms of quantity and glycosylation. However, it was found that the specific productivity is higher at high specific growth rates, which tends to favor higher mannosylation, since 10-mannose glycoforms were the most abundant, whereas at low specific growth rates, 9-mannose glycosylation was predominant. While the implications of these findings are limited for the particular avidin case, they are much more important with respect to the production of therapeutic glycoproteins in *Pichia pastoris*. The results suggest that effect of culture parameters, and foremost the specific growth rate, on the glycan structure should be integrated in bioprocess development to avoid quality variations when scaling-up the process.

6/ REFERENCES

- Airenne KJ, Oker-Bolm C, Marjomäki VS, Bayer EA, Wilchek M, Kulomaa MS. 1997. Production of biologically active recombinant avidin in baculovirus-infected insect cells. *Protein Expr. Purif.* 9(1):100-108.
- Airenne KJ, Sarkinen P, Punnonen E-L, Kulomaa MS. 1994. Production of recombinant avidin in *Escherichia coli*. *Gene* 144(1):75-80.
- Andersen DC, Bridges T, Gawlitsek M, Hoy C. 2000. Multiple cell culture factors can affect the glycosylation of Asn-184 in CHO-produced tissue-type plasminogen activator. *Biotechnol. Bioeng.* 70(1):25-31.
- Andrews JF. 1968. A Mathematical model for the continuous culture of microorganisms utilizing inhibitory Substrates. *Biotechnol. Bioeng.* X:707-723.
- Bruch R, White HB. 1982. Composition and structural heterogeneity of avidin glycopeptides. *Biochemistry* 21:5334-5341.
- Cannizzaro C. 2002. Spectroscopic monitoring of bioprocesses: a study of carotenoid production by *Phaffia rhodozyma* yeast: PhD Thesis No 2620, EPFL. 294 p.
- Cereghino JL, Cregg JM. 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* 24(1):45-66.
- Choi B-K, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, Miele RG, Nett JH, Wildt S, Gerngross TU. 2003. Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. *Proc. Natl. Acad. Sci. U. S. A.* 100(9):5022-5027.
- Cos O, Ramón R, Montesinos JL, Valero F. 2006. Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: A review. *Microb. Cell. Fact.* 5:17.
- Cunha AE, Clemente JJ, Gomes R, Pinto F, Thomaz M, Miranda S, Pinto F, Moosmayer D, Donner P, Carrondo MJT. 2004. Methanol Induction Optimization for scFv Antibody Fragment Production in *Pichia pastoris*. *Biotechnol. Bioeng.* 86(4):458-467.
- Daly R, Hearn MTW. 2005. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J. Mol. Recognit.* 18(2):119-138.
- DeLange RJ. 1970. Egg white avidin. I. Amino acid composition; sequence of the amino- and carboxyl-terminal cyanogen bromide peptides. *J. Biol. Chem.* 245(5):907-916.

- Gerngross TU. 2004. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat. Biotechnol.* 22(11):1409-1414.
- Goochee CF, Gramer MJ, Andersen DC, Bahr JB, Rasmussen JR. 1991. The oligosaccharides of glycoproteins - bioprocess factors affecting oligosaccharide structure and their effect on glycoprotein properties. *Biotechnology* 9(12):1347-1355.
- Goochee CF, Monica T. 1990. Environmental-effects on protein glycosylation. *Biotechnology* 8(5):421-427.
- Grinna LS, Tschopp JF. 1989. Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, *Pichia pastoris*. *Yeast* 5(2):107-115.
- Gurakan T, Marison IW, von Stockar U, Gustafsson L, Gnaiger E. 1990. Proposals for a standardized sample handling procedure for the determination of elemental composition and enthalpy combustion of biological material. *Thermochim. Acta* 172:251-266.
- Hellwig S, Emde F, Raven NPG, Henke M, van der Logt P, Fischer R. 2001. Analysis of single-chain antibody production in *Pichia pastoris* using on-line methanol control in fed-batch and mixed-feed fermentations. *Biotechnol. Bioeng.* 74(4):344-352.
- Hiller Y, Gershoni JM, Bayer EA, Wilchek M. 1987. Biotin binding to avidin. Oligosaccharide side chain not required for ligand association. *Biochem. J.* 24(1):167-171.
- Hong F, Meinander NQ, Jönsson LJ. 2002. Fermentation strategies for improved heterologous expression of laccase in *Pichia pastoris*. *Biotechnol. Bioeng.* 79(4):438-449.
- Hood EE, Witcher DR, Maddock S, Meyer T, Baszczynski C, Bailey M, Flynn P, Register J, Marshall L, Bond D and others. 1997. Commercial production of avidin from transgenic maize: characterization of transformant, production, processing, extraction and purification. *Mol. Breeding* 3(4):291-306.
- Invitrogen. 2002. *Pichia* Fermentation Process Guidelines. Carlsbad (California).
- Jenkins N, Parekh RB, James DC. 1996. Getting the glycosylation right: Implications for the biotechnology industry. *Nat. Biotechnol.* 14(8):975-981.
- Jungo C, R  rat C, Marison IW, von Stockar U. 2006. Quantitative characterization of the regulation of the synthesis of alcohol oxidase and of the expression of recombinant avidin in a *Pichia pastoris* Mut⁺ strain. *Enz. Microb. Technol.* 39(4):936-944.
- Kada G, Falk H, Gruber HJ. 1999. Accurate measurement of avidin and streptavidin in crude biofluids with a new, optimized biotin-fluorescein conjugate. *Biochim. Biophys. Acta* 1427:33-43.

- Kahtri NK, Hoffman F. 2006. Impact of methanol concentration on secreted protein production in oxygen-limited cultures of recombinant *Pichia pastoris*. *Biotechnol. Bioeng.* 93(5):871-879.
- Kobayashi K, Kuwae S, Ohya T, Ohda T, Ohyama M, Tomomitsu K. 2000. High level secretion of recombinant human serum albumin by fed-batch fermentation of the methylotrophic yeast, *Pichia pastoris*, based on optimal methanol feeding strategy. *J. Biosci. Bioeng.* 90(3):280-288.
- Li HJ, Sethuraman N, Stadheim TA, Zha DX, Prinz B, Ballew N, Bobrowicz P, Choi BK, Cook WJ, Cukan M and others. 2006. Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. *Nat. Biotechnol.* 24(2):210-215.
- Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM. 2005. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* 22(4):249-270.
- Meuwly F, von Stockar U, Kadouri A. 2004. Optimization of the medium perfusion rate in a packed-bed bioreactor charged with CHO cells. *Cytotechnology* 46(1):37-47.
- Montesino R, García R, Quintero O, Cremata JA. 1998. Variation in N-Linked oligosaccharide structures on heterologous proteins secreted by the methylotrophic yeast *Pichia pastoris*. *Protein Expr. Purif.* 14(2):197-207.
- Nardone E, Rosano C, Santambrogio P, Curnis F, Corti A, Magni F, Siccardi AG, Paganelli G, Losso R, Aprea B and others. 1998. Biochemical characterization and crystal structure of a recombinant hen avidin and its acidic mutant expressed in *Escherichia coli*. *Eur. J. Biochem.* 256(2):453-460.
- Ohya T, Ohyama M, Kobayashi K. 2005. Optimization of human serum albumin production in methylotrophic yeast *Pichia pastoris* by repeated fed-batch fermentation. *Biotechnol. Bioeng.* 90(7):876-887.
- Reddy ST, Dahms NM. 2002. High-level expression and characterization of a secreted recombinant cation-dependent mannose 6-phosphate receptor in *Pichia pastoris*. *Protein Expr. Purif.* 26(2):290-300.
- Rosebrough SF, Hartley DF. 1996. Biochemical modification of streptavidin and avidin: In vitro and in vivo analysis. *J. Nucl. Med.* 37(8):1380-1384.
- Schenk J, Marison IW, von Stockar U. 2007. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2):344-353.
- Sreekrishna K, Brankamp RG, Kropp KE, Blankenship DT, Tsay JT, Smith PL, Wierschke JD, Subramaniam A, Birkenberger LA. 1997. Strategies for optimal synthesis and

- secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* 190(1):55-62.
- Stephanopoulos GN, Aristidou AA, Nielsden J. 1998. *Metabolic engineering: principles and methodologies*. San Diego: Academic Press.
- Stratton J, Chiruvolu V, Meagher M. 1998. High cell-density fermentation. In: DR Higgins and JM Clegg, editors. *Methods in Molecular Biology*: Humana Press. p 107-120.
- Trimble RB, Atkinson PH, Tschopp JF, Townsend RR, Maley F. 1991. Structure of oligosaccharides on *Saccharomyces SUC2* invertase secreted by the methylotrophic yeast *Pichia pastoris*. *J. Biol. Chem.* 266(34):22807-22817.
- Wilchek M, Bayer EA. 1990. Applications of avidin-biotin technology: Literature survey. In: Wilchek M, Bayer EA, editors. *Methods in Enzymology*. San Diego: Academic Press. Volume 148. p 14-45.
- Wildt S, Gerngross TU. 2005. The humanization of *N*-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* 3(2):119-128.
- Woo JH, Liu YY, Neville DM. 2006. Minimization of aggregation of secreted bivalent anti-human T cell immunotoxin in *Pichia pastoris* bioreactor culture by optimizing culture conditions for protein secretion. *J. Biotechnol.* 121:75-85.
- Yao Z, Zhang M, Sakahara H, Nakamoto Y, Higashi T, Zhao S, Sato N, Arano Y, Konishi J. 1999. The relationship of glycosylation and isoelectric point with tumor accumulation of avidin. *J. Nucl. Med.* 40(3):479-483.
- Yuen CT, Storrington PL, Tiplady RJ, Izquierdo M, Wait R, Gee CK, Gerson P, Lloyd P, Cremata JA. 2003. Relationships between the N-glycan structures and biological activities of recombinant human erythropoietins produced using different culture conditions and purification procedures. *Brit. J. Haematol.* 121(3):511-526.
- Zhang W, Bevins MA, Plantz BA, Smith LA, Meagher MM. 2000. Modeling *Pichia pastoris* growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A. *Biotechnol. Bioeng.* 70(1):1-8.
- Zhang W, Sinha J, Smith LA, Inan M, Meagher MM. 2005. Maximization of production of secreted recombinant proteins in *Pichia pastoris* fed-batch fermentation. *Biotechnol. Prog.* 21(2):386-393.
- Zocchi A, Jobe A-M, Neuhaus J-M, Ward TR. 2003. Expression and purification of a recombinant avidin with a lowered isoelectric point in *Pichia pastoris*. *Protein Expr. Purif.* 32(2):167-174.

7/ LIST OF SYMBOLS

Nomenclature

GlcNAC	N-acetylglucosamine
Man	Mannose
pI	Isoelectric point

Latin symbols

C	Concentration [g L^{-1}], or [mol L^{-1}]
F	Fed-batch feeding rate [g h^{-1}]
q_p	Avidin specific productivity [$\mu\text{g g}^{-1} \text{h}^{-1}$]
n	Number of mannose in glycan chain
t	Time [h]
V	Broth volume [L]
$Y_{X/S}$	Biomass to substrate stoichiometric yield [C-mol C-mol^{-1}]
$Y_{Q/S}$	Heat to substrate stoichiometric yield [C-mol C-mol^{-1}]
X	Biomass concentration (in dry cell weight) [g L^{-1}]

Greek symbols

μ	Specific growth rate [h^{-1}]
-------	--

Superscripts and subscripts

o	subscript	Initial state
-----	-----------	---------------

CV

Curriculum vitae

Jonas SCHENK

Swiss citizen

Born the 24th of October 1979

Education

2007	PhD in Bioprocess Engineering Ecole Polytechnique Fédérale de Lausanne Laboratory of Biological and Chemical Engineering (Prof von Stockar)
2002	Diplôme d'Ingénieur-chimiste (Master of Chemical Engineering) Ecole Polytechnique Fédérale de Lausanne, Switzerland
1997	Maturité Fédérale scientifique (type C) Gymnase Cantonal de La Chaux-de-Fonds

Professional experience

2003-2007	Bioprocess junior scientist and assistant Ecole Polytechnique Fédérale de Lausanne Laboratory of Biological and Chemical Engineering (Prof von Stockar)
2002-2003	Associate teacher Engineering School of Unjung Pandang, Indonesia (Swiss Civil Service)

Languages

French	Mother tongue
English	Fluent written/spoken
German	Medium level
Indonesian	Basic level

Computer skills

Labview	Sound knowledge
Matlab	Fair level

Personal interests

Running	Semi-marathon Lausanne 2006: 1h22'07" Marathon Paris 2006: 3h04'50"
---------	--

Publications

- Schenk J, Marison IW, von Stockar U. 2007. Simplified Fourier-transform mid-infrared calibration based on a spectra library for the on-line monitoring of bioprocesses. *Anal. Chim. Acta* 591(1):132-140.
- Schenk J, Marison IW, von Stockar U. 2007. A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy. *J. Biotechnol.* 128(2):344-353.
- Schenk J, Balazs K, Jungo C, Urfer J, Wegmann C, Zocchi A, Marison IW, von Stockar U. 2007. Influence of specific growth rate on specific productivity and glycosylation of a recombinant avidin produced by a *Pichia pastoris* Mut⁺ strain. *Biotechnol. Bioeng. In press* (DOI 10.1002/bit.21565).
- Schenk J, Marison IW, von Stockar U. pH prediction and control in bioprocesses using mid-infrared spectroscopy. *Biotechnol. Bioeng. Submitted*
- Schenk J, Viscasillas C, Marison IW, von Stockar U. On-line monitoring of nine different batch cultures of *E. coli* by mid-infrared spectroscopy, using a single spectra library for calibration. *J. Biotechnol. Submitted*
- Schenk J, Viscasillas C, Marison IW, von Stockar U. Effect of temperature on mid-infrared spectroscopy calibrations for the on-line monitoring of bioprocesses. *Appl. Spectrosc. Submitted*

Oral presentations

- | | |
|------|--|
| 2006 | ESBES 6 (European Symposium on Biochemical Engineering Science)
Salzburg, Austria, 29.08.2006
"Simplified FTIR calibrations based on a spectra library for the on-line monitoring of bioprocesses" |
| 2005 | ECB12 (European Congress on Biotechnology)
Copenhagen, Denmark, 24.08.2005
"Simple and quick in-situ calibration of a FTIR instrument to control fed-batch fermentations of <i>Pichia pastoris</i> "
Abstract published in Journal of Biotechnology, 2005, volume 188S1, p. S37 |
| 2005 | BioTech 2005, 3 rd Swiss-Czech Meeting
Wädenswil, Switzerland, 18.05.2005
"Understanding regulation of recombinant protein expression in <i>Pichia pastoris</i> "
Presented together with Carmen Jungo |

Poster presentation

- | | |
|------|---|
| 2007 | ECB13 (European Congress on Biotechnology)
Barcelona, Spain, 19.09.2007
"Mid-infrared spectroscopy as a tool for pH prediction and control in bioprocesses" |
|------|---|