SELECTION OF A MATHEMATICAL MODEL FOR THE KINETICS OF *Haemophilus influenzae* TYPE B USING AKAIKE'S INFORMATION CRITERION

F. de O. Cintra[1] and M. Takagi[1]*

[1] Laboratório de Bioprocessos, Centro de Biotecnologia, Instituto Butantan, Av. Vital Brasil, 1500, CEP 05503-900, Butantã, São Paulo, SP, Brazil.

Abstract - *Haemophilus influenzae* type b (Hib) vaccine is made up from its capsular polysaccharide (PRP). Low productivity of the polysaccharide during cell growth increases the final cost of this vaccine. Hib achieves low levels of cellular concentration *in vitro* due to the inhibition caused by acetate. The Akaike Information Criterion (*AIC*) was used in this work for selecting models of microbial growth. The application to the case of the multivariate models is outlined and the procedure is carried out using data from Hib cultures. From 4 models of biomass and 15 of acetate and PRP, one could be selected with great evidence for support. The use of *AIC* has shown to be robust and free of subjectivity, and it was able to define a kinetic model that is adequate for the cell growth and the production of its PRP over a wide range of culture conditions. The exponential inhibition factor was found to be the best for modelling inhibition of cell growth by acetate, while the hyperbolic factor was the best for inhibition of PRP formation. The acetate formation was found to have both growth associated and non-associated types. PRP formation was found to be only growth associated.
Keywords: *Haemophilus influenzae* type b; Akaike's Information Criterion; Mathematical model; Biomass; Polysaccharide.

INTRODUCTION

Haemophilus influenzae is a bacterium present in the nasopharyngeal tract of humans, and is potentially pathogenic. Of all its six serotypes, type b is the most harmful one, having once been reported as the main cause of bacterial meningitis in infants until the introduction of the conjugated polysaccharide vaccine in the 1980's (Wilfert, 1990; Yogev et al., 1990). This polysaccharide forms the extracellular capsule of the bacterium, acting as a virulence factor and preventing phagocytosis. The capsular polysaccharide of *H. influenzae* type b (Hib) is a linear polysaccharide composed of poly-ribosylribitol-phosphate units, defined as PRP (Crisel et al., 1975). Upon submerged growth *in vitro* the majority of the PRP content present in the capsule is released into the supernatant, whence it can be purified and chemically conjugated to a protein, originating the conjugate vaccine (Anderson et al., 1976).

Low productivity of PRP is observed in the vaccine production process due to many aspects. One of them comes from the fact that Hib is a fastidious microorganism, requiring specific growth factors (NAD and Hemin) and enriched nutrients (amino acids and vitamins) (Biberstein et al., 1963; Loeb, 1995; Reidl et al., 2000). Also, it is difficult to achieve high cell densities due to metabolic inhibition, which is caused by high levels of secondary metabolites that are produced during growth. Acetate is the main byproduct

* Corresponding author: Dra. Mickie Takagi, e-mail: mickie.takagi@butantan.gov.br, Phone: +55 11 26279720 / +55 11 26279887

found in Hib cultivations, being a dead-end product in energy metabolism. The buildup of acetate in the culture broth cannot be avoided by supplying oxygen as in the case of other facultative bacteria, and thus optimization of the polysaccharide production process must account for acidic inhibition (Fleischmann et al., 1995; Schilling and Palsson, 2000).

Several models of microbial growth, metabolic inhibition and biomolecule production are described in the literature. Typically, these models are included in a system of ordinary differential equations (SODE) for the solution of the mass balances in the control volume of the reactor. The SODE describes the rates of growth and production as functions of cell concentration, cell growth rate and metabolite or substrate concentration (which are inhibiting or limiting factors). In a batch culture, the variables predicted by the models are the rates of variation over time, and not the concentration itself, therefore the experimental data must somehow be transformed into a velocity measurement. This can be done in two ways. The most accurate and reliable alternative is to conduct continuous cultures. In this type of cultivation, the process can be driven into a stationary state, where the cell multiplication rate can be kept constant. As there is a continuous flow of culture broth out of the reactor, the rates of variation can be obtained from the association of concentrations with the flow rate applied. Multiple flow rates will provide multiple formation rates, which can be plotted against the growth rate or the metabolite/substrate concentration, and thus yield sufficient data for the fitting of the equations. Despite being a satisfactory method for obtaining the rates of interest, continuous cultures are troublesome when taken into practical consideration. For instance, adequate equipment must be available, there is higher risk of contamination and a high quantity of medium is needed, which is unattractive for the case of Hib, among others. Furthermore, the stationary state is not achieved before several retention times, and for this reason only few experimental data can be drawn without an overly longstanding experiment.

The second alternative is to estimate the derivative at each experimental point. For this, there are several methods. A broadly used one is polynomial fitting or interpolation of experimental data from discontinuous batches. The problem with this method is that it is highly affected by experimental error: as the polynomial curve tends to pass exactly over the data points, any experimental error causes a significant change in the curve obtained, and differentiation of this curve only amplifies the error. Furthermore, the choice of the right order or the cutting out of outliers in extremely biased by the subjectivity of the operator, making the process of model selection non-robust.

A more accurate method of using discontinuous batch data is directly using non-linear fitting with several candidate models. The goodness of fit for the models can be assessed and compared, and the best fitting model can be determined as the true model.

Hypothesis testing is one way of assessing the goodness of fit for the models. It can give a p-value and a notion of statistical significance for the model chosen. However, this technique is troublesome as it can only be used to compare nested models (Motulsky and Christopoulos, 2003). With it, the comparison between the varieties of models proposed for microbial kinetics becomes unfeasible.

The field of information theory supplies an alternative for model comparison that is much more versatile. Akaike's Information Criterion (*AIC*) is a concept introduced in 1973 which balances the proximity of the model to reality with the bias inherent to the equations used in the model, assessed by the number of estimable parameters used (Burnham and Anderson, 2002). For an experimental data set, each model will have a value of *AIC*; the differences between these values can be used in formulating a confidence set of models, in a Bayesian inference sense (Burnham and Anderson, 2004; Wagenmakers and Farrell, 2004; Li et al. 2017).

In this work, the *AIC* was applied to models of microbial growth, considering the necessary adjustments, in order to select a model for kinetics of Hib growth and polysaccharide production.

MATERIAL AND METHODS

Experimental Design

A complete 2² Rotational Central Composite Design (RCCD) with two replicates of the central point was used to set the conditions of temperature and pH. The temperature range was set between 29.0 and 37.0 °C, while the pH range was set between 6.30 and 7.50. Table 1 specifies the conditions used at each fermentation run.

Table 1. Values of the independent variables used in the Hib cultivation as determined by the experimental design.

Experiment	T (°C)	pH
1	33.0	6.90
2	33.0	6.90
3	35.8	7.32
4	35.8	6.48
5	30.2	7.32
6	30.2	6.48
7	33.0	7.50
8	33.0	6.30
9	37.0	6.90
10	29.0	6.90

Experimental and Analytical Methods

Strain GB3291 of *Haemophilus influenzae* type b was purchased from the Culture Collection Section of Instituto Adolpho Lutz (São Paulo, Brazil). A seed composed of a suspension of cells in Greaves medium was stored in liquid nitrogen (Greaves, 1960). An aliquot of 200 µL of this seed was transferred to 100 mL of the MMP medium as described by Takagi et al. (2003) and incubated in static anaerobic conditions at 37 °C for about 8 h, when the optical density at 540 nm (OD_{540}) exceeded a reading of 0.400 AU. At this moment, a volume of this suspension was transferred into 200 mL of fresh MMP medium in a 1 L Erlenmeyer flask, resulting in an initial OD_{540} of 0.050 AU. This flask was then incubated in an orbital shaker (INNOVA 44 – New Brunswick Scientific) at 150 rpm and 37 °C, for about 12 h, when the OD_{540} was expected to exceed a value of 5.0 AU. A volume of this suspension was calculated in order to achieve 0.150 AU for the OD_{540} in a total volume of 6 L of culture, and inoculated in the reactor. The reactor (BioFlo 2000 bioreactor – New Brunswick Scientific) consisted of a 13 L vessel with 6 L of a modified MMP medium, where the concentrations of glucose and yeast extract were raised to 20 g.L⁻¹. Air was supplied at 0.5 vvm, and the pO2 was controlled at 30% of the saturation. pH was controlled by addition of NaOH 5 M. Samples were taken periodically and centrifuged at 4 °C and 16,000 g for 10 min; the pellets were washed with NaCl 150 mM and centrifuged again in the same conditions, dried at 70 °C and used for measuring the dry mass concentration. The cell free supernatant was frozen at -20 °C for further analyses. Acetate concentration was measured by ion exclusion chromatography as described by Takagi *et al* 2006.

Two fractions of PRP were measured. The free fraction was determined from the culture broth supernatant, and the cell associated fraction was obtained by extraction from the cell pellets as described by Kroll and Moxon (1988). The amount of PRP in both fractions was measured by ion exchange chromatography with pulsed amperometric detection, using a CarboPac PA10 column and a ICS-5000 chromatography system (Thermo Fisher Scientific) and with the protocols described by de Hann et al. (2013). The two fractions were summed to give the total amount of PRP produced by the bacterium.

All the measured data obtained in these experiments are available for consultation in Cintra (2014).

MODELLING AND THEORETICAL ASPECTS

Candidate Models

Three variables were considered to be modeled: biomass, acetate and PRP formation. Acetate inhibition was proposed to act upon all three variables, each with a dedicated inhibition factor. Four inhibition factors were tested. The first inhibition factor evaluated in this work was the one proposed by Aiba et al. (1968) and codified as I_A in eq. 1, and the second is equivalent but with an additional exponent, codified as I_B in eq. 2. The third was proposed by Stepanova et al. (1965), denoted as I_C in eq. 3, and the fourth is equivalent to the third with an additional exponent, as was also proposed by Yano and Koga (1973) and described as I_D in eq. 4. In these four equations I is the inhibition factor, k_i is the inhibition constant, A is the concentration of acetate and n_i is the inhibition exponent.

$$I_A = \exp(-k_i\, A) \tag{1}$$

$$I_B = \exp\left(-(k_i\, A)^{n_i}\right) \tag{2}$$

$$I_C = \frac{1}{1 + \dfrac{A}{k_i}} \tag{3}$$

$$I_D = \frac{1}{1 + \left(\dfrac{A}{k_i}\right)^{n_i}} \tag{4}$$

For the modeling of biomass formation, the maximum specific rate of growth was simply multiplied by one of the four inhibition factors. Biomass formation was identified as X_f, and the models were coded from A to D, as specified in Table 2. In the equations shown, X denotes biomass concentration and μ_{max} denotes the maximum specific rate of growth.

For the formation of products, identical constructions were used for both the formation of acetate (models A_f) and of PRP (models P_f). First, the equation of Luedeking and Piret (1959) was used in the three forms: growth-associated (code A), non-growth-

Table 2. Balance equations with the candidate models of biomass formation.

Code	X_f Model
A	$\dfrac{dX}{dt} = \mu_{max}\, \exp(-(k_X A))\, X$
B	$\dfrac{dX}{dt} = \mu_{max}\, \exp(-(k_X A)^{n_X})\, X$
C	$\dfrac{dX}{dt} = \mu_{max}\, \dfrac{1}{1 + \left(A/k_X\right)}\, X$
D	$\dfrac{dX}{dt} = \mu_{max}\, \dfrac{1}{1 + \left(A/k_X\right)^{n_X}}\, X$

associated (code B) and mixed formation (code C). Each one of the formation models is multiplied by an inhibition factor, as shown in Table 3. In the equations shown, P denotes the concentration of PRP, α denotes the constant of growth-associated formation and β denotes the constant of non-associated formation.

Five possibilities of acidic inhibition were considered for the formation of the products, and were denoted as models A_i for inhibition by acetate or models P_i for inhibition of PRP: without inhibition (code O) and by the four forms of inhibition as used for biomass (identified similarly by the subscripts A through D). The equations used are specified in Table 4.

The global kinetic model was constructed in two steps. First, the four models for biomass formation, the three for acetate formation and the five for acetate inhibition were combined, resulting in a testing set of sixty partial models (they were named partial as they do not include the models of product formation). These models were denoted by a code PM, as listed in Table 5.

From these sixty partial models, only the most adequate ones were chosen, as determined by the *AIC*. The selected partial models were then combined to the three models of PRP formation and to the five of PRP inhibition. Therefore, there were fifteen complete models to be tested for each selected partial model. The codification of the complete models was done by extending the PM numbers of Table 5 with a CM code, as listed in Table 6.

Table 3. Candidate models of acetate and PRP formation.

Código	Modelo A_f	Modelo P_f
A	$\dfrac{dA}{dt} = \alpha_A \dfrac{dX}{dt} I$	$\dfrac{dP}{dt} = \alpha_P \dfrac{dX}{dt} I$
B	$\dfrac{dA}{dt} = \beta_A X I$	$\dfrac{dP}{dt} = \beta_P X I$
C	$\dfrac{dA}{dt} = \left(\alpha_A \dfrac{dX}{dt} + \beta_A X \right) I$	$\dfrac{dP}{dt} = \left(\alpha_P \dfrac{dX}{dt} + \beta_P X \right) I$

Table 4. Candidate models of acetate and PRP inhibition.

Code	A_i Model	P_i Model
O	$I = 1$	$I = 1$
A	$I = \exp\left(-(k_A A)\right)$	$I = \exp\left(-(k_P A)\right)$
B	$I = \exp\left(-(k_A A)^{n_A}\right)$	$I = \exp\left(-(k_P A)^{n_P}\right)$
C	$I = \dfrac{1}{1 + \left(A / k_A \right)}$	$I = \dfrac{1}{1 + \left(A / k_P \right)}$
D	$I = \dfrac{1}{1 + \left(A / k_A \right)^{n_A}}$	$I = \dfrac{1}{1 + \left(A / k_P \right)^{n_P}}$

Table 6. Models for PRP kinetics used in the testing set of complete models.

	Codes			Codes			Codes	
CM	P_f	P_i	CM	P_f	P_i	CM	P_f	P_i
01	A	O	06	B	O	11	C	O
02	A	A	07	B	A	12	C	A
03	A	B	08	B	B	13	C	B
04	A	C	09	B	C	14	C	C
05	A	D	10	B	D	15	C	D

Model Fitting

The system of ordinary differential equations (SODE) obtained by the combinations of the models were solved by the Runge-Kutta algorithm of order (4.5). The models were then fitted to the experimental data by the Levenberq-Marquart algorithm. The implementations of both numerical methods were taken from the functions ode45 and leasqr from GNU Octave 3.8.0.

AIC Application to Multivariate Functions

AIC is described as the balance between the log likelihood of the model and the number of estimable

Table 5. Models for biomass and acetate kinetics used in the testing set of partial models.

	Codes				Codes				Codes				Codes		
PM	X_f	A_f	A_i	PM	X_f	A_f	A_i	PM	X_f	A_f	A_i	PM	X_f	A_f	A_i
01	A	A	O	16	B	A	O	31	C	A	O	46	D	A	O
02	A	A	A	17	B	A	A	32	C	A	A	47	D	A	A
03	A	A	B	18	B	A	B	33	C	A	B	48	D	A	B
04	A	A	C	19	B	A	C	34	C	A	C	49	D	A	C
05	A	A	D	20	B	A	D	35	C	A	D	50	D	A	D
06	A	B	O	21	B	B	O	36	C	B	O	51	D	B	O
07	A	B	A	22	B	B	A	37	C	B	A	52	D	B	A
08	A	B	B	23	B	B	B	38	C	B	B	53	D	B	B
09	A	B	C	24	B	B	C	39	C	B	C	54	D	B	C
10	A	B	D	25	B	B	D	40	C	B	D	55	D	B	D
11	A	C	O	26	B	C	O	41	C	C	O	56	D	C	O
12	A	C	A	27	B	C	A	42	C	C	A	57	D	C	A
13	A	C	B	28	B	C	B	43	C	C	B	58	D	C	B
14	A	C	C	29	B	C	C	44	C	C	C	59	D	C	C
15	A	C	D	30	B	C	D	45	C	C	D	60	D	C	D

parameters in the same model (Burnham and Anderson, 2002). It is detailed in eq. 5, where $\mathcal{L}(\beta|\boldsymbol{D})$ denotes the likelihood of the model with the maximum likelihood estimator (MLE) β of the parameters, calculated on the set D of experimental data, and K is the number of estimable parameters in the model.

$$AIC = -2\ln\left(\mathcal{L}\left(\hat{\beta}\,|\,\mathbf{D}\right)\right) + 2K \qquad (5)$$

The function $\mathcal{L}(\hat{\beta}|\boldsymbol{D})$ is calculated under the premise that the deviation of the model for the data approximates a multivariate normal distribution, as stated in eq.6, where D is a $n \times p$ matrix, n being the number of samples and p the dimensionality of the model, which is equivalent to the number of dependent variables measured and modeled; y_i is the vector of the experimental values for the i-th sample, or the i-th line in D; $f(t_i, \hat{\beta})$ is the value predicted by the resolution of the SODE in the model at time t_i; and Σ_i is the $p \times p$ covariance matrix of the errors.

$$\mathcal{L}\left(\hat{\beta}\,|\,\mathbf{D}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^p\,|\Sigma_i|}} \exp\left(-\frac{1}{2}\left(y_i - f\left(t_i,\hat{\beta}\right)\right)' \Sigma_i^{-1}\left(y_i - f\left(t_i,\hat{\beta}\right)\right)\right) \quad (6)$$

In some applications, the variance of the errors is unknown and is estimated as 1, increasing the number of estimable parameters in the model (Motulsky and Christopoulos, 2003). In the multivariate case, this increase would equal $p(p + 1)/2$, or the number of independent values in Σ_i (Bedrick and Tsai, 1994). If the experimental error is accessible, Σ_i can be formulated as a diagonal matrix having these values of error for each sample i, thus considering that the measured variables are inter-independent and minimizing the number of parameters to penalize for. As for K, its value equals the number of parameters estimated in $\hat{\beta}$ plus one, because of the estimation of the initial value y_0 necessary for the resolution of the SODE.

It has been stated that, in small samples, a correction must be made to account for the bias incurred from the low number of real data. This correction is necessary in cases where $n/K < 40$ (Hurvich and Tsai, 1989). The corrected AIC, or AIC_c, is calculated by eq. 7 below. It can be seen that the minimum number of samples must be such that $n \geq K + 2$.

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1} \qquad (7)$$

Comparison and Selection of the Models

In order to compare models and determine which are the most adequate to describe the experimental data, the value of AIC_c is calculated for each model in the set being tested. The model with the smallest AIC_c,

which can be denoted as the S-model, then has its value subtracted from the remaining ones. The difference is used to calculate the evidence ratio ε_i for each model compared to the apparent best, as stated by eq. 8.

$$\varepsilon_i = \exp\left(-\frac{1}{2}\left(AIC_{cS} - AIC_{ci}\right)\right) \qquad (8)$$

This ratio represents the relative likelihood supported by the available data that the i-th model must be considered to represent the reality over the S-model (Motulsky and Christopoulos, 2003). The whole set of evidence ratios can then be normalized to sum 100, giving the Akaike weights $w_\%$ shown in eq. 9.

$$w_{\%i} = \frac{\varepsilon_i}{\sum_{j=1}^{n}\varepsilon_j}100 \qquad (9)$$

It has been demonstrated that the weight of a model can be used as the probability of that model being the true model, given the data available and the set of models chosen. It has also been demonstrated that, if the true model is present in the set, its weight value will approach 1 as the number of samples tends into infinity, and that with a limited number of samples the true model may not show the highest weight, implying that simply choosing the model with the highest value may not be a robust procedure (Burnham and Anderson, 2002). Instead, a method for defining a confidence set of models must be carried out. This can be done by ranking the models in order of their Akaike weights, from highest to lowest; then, the confidence set is generated by adding the models one by one until the total sum of their weights surpasses a defined threshold, which for this work was set at 99.9 %. This set represents the set of models whose probability of being true is supported by the data available, for the total set of proposed models. Because of this, special care must be taken when formulating the candidate models, as an increased number of models in the set may negatively influence the results for the confidence set (Burnham and Anderson, 2002).

RESULTS AND DISCUSSION

In this work biomass, acetate and polysaccharide were considered for the construction of models of Hib growth kinetics. Acetic acid is the main byproduct of Hib's metabolism, originated from energetic homeostasis, and it was investigated because it causes inhibition of growth and biosynthesis. Substrate, despite the fact that it is usually considered in kinetic modelling, was left out from the study for several

reasons. For instance, when the regression algorithms are applied to Monod's equation and its variations, they yield unreal values for the parameters, which are highly correlated; this happens in our specific case because bacterial growth ceases much earlier than the substrate achieves limiting concentrations due to the high susceptibility of Hib to acetate concentration. In this way, the experiments were carried out on complex medium with excess amounts of substrates, in order to eliminate the limitation factors from the model construction.

The differential equations in the mass balance that are used to solve the kinetic model are interdependent, i.e., the resolution of the equation for one variable depends on the resolution of the others. In the case of biomass, the rate of formation is determined by the concentration of acetate. The acetate concentration is in turn obtained from the resolution of the Luedeking and Piret's equation, which depends upon cell formation rate and/or concentration. In this way, they must be solved together. In the same way as acetate, PRP formation is dependent on the resolution of the equations of biomass and acetate. However, as neither biomass nor acetate formation are dependent on PRP concentration, this variable can be left out of the resolution in the first moment, leaving a model

that describes only biomass and acetate formation. This also simplifies the selection procedure by reducing the number of model combinations in the testing set. The four models for biomass formation in Table 2 were combined to the three models of acetate formation listed in Table 3 and to the five models of inhibition by acetate formation listed in Table 4, resulting in sixty partial models which were fitted to the experimental data.

After the fitting procedure, it was observed that not all models yielded satisfactory convergence, i.e., in some experiments either the values of the fitted parameters were unrealistic (negative or divergent towards infinity) or the fitting algorithm was unable to find a MLE for the parameters after several attempts. Thus, the models that did not converge satisfactorily in any experiments were excluded from the study. From the sixty candidate models for biomass and acetate formation, only twenty-nine could achieve convergence in all experiments.

For each one of the twenty-nine successfully adjusted partial models, the values of AIC_c were calculated followed by the determination of the respective Akaike weights. These values were then ordered in descent, and the 99.9% confidence set was selected. This procedure was replicated for each of

Table 7. Akaike weights calculated for the partial models of biomass, acetate formation and acetate inhibition.

Model Codes				Experiments									
PM	X	A_f	A_i	1	2	3	4	5	6	7	8	9	10
01	A	A	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**0.03**
06	A	B	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
08	A	B	B	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**55.11**
11	A	C	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**3.93**
12	A	C	A	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**0.62**
13	A	C	B	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**2.70**
15	A	C	D	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**2.63**
16	B	A	O	< 0.01	< 0.01	0.01	< 0.01	< 0.01	< 0.01	**1.13**	< 0.01	< 0.01	0.03
21	B	B	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
22	B	B	A	**83.78**	**92.50**	**0.63**	0.01	**27.39**	**7.23**	**0.94**	**6.17**	**54.09**	**0.30**
> 23	B	B	B	**1.97**	**4.25**	**3.53**	**55.39**	**5.10**	**56.21**	**4.76**	**60.94**	**2.69**	**10.02**
24	B	B	C	**7.76**	**0.20**	< 0.01	< 0.01	**60.21**	< 0.01	**0.33**	**0.06**	**30.67**	**2.10**
> 25	B	B	D	**1.69**	**1.74**	**4.10**	**18.56**	**5.11**	**28.93**	**5.98**	**29.27**	**1.95**	**19.24**
> 26	B	C	O	**4.77**	**1.26**	**88.21**	**1.96**	**1.44**	**0.13**	**13.00**	**0.20**	**10.18**	**2.69**
28	B	C	B	0.01	0.03	0.02	**12.16**	**0.23**	**3.72**	**0.12**	**2.28**	**0.05**	**0.07**
30	B	C	D	0.01	0.02	0.03	**11.92**	**0.23**	**1.85**	**0.13**	**0.97**	0.04	0.03
31	C	A	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
36	C	B	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
39	C	B	C	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**0.04**
41	C	C	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**0.26**
46	D	A	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**3.96**	< 0.01	< 0.01	< 0.01
51	D	B	O	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
52	D	B	A	< 0.01	< 0.01	**0.04**	< 0.01	**0.08**	**0.24**	**1.80**	0.02	**0.18**	< 0.01
54	D	B	C	< 0.01	< 0.01	< 0.01	< 0.01	**0.18**	< 0.01	**0.67**	< 0.01	**0.12**	0.02
55	D	B	D	< 0.01	< 0.01	**0.17**	< 0.01	0.02	**1.45**	**9.51**	**0.07**	0.01	**0.14**
56	D	C	O	< 0.01	< 0.01	**3.13**	< 0.01	< 0.01	< 0.01	**49.29**	< 0.01	0.02	**0.05**
58	D	C	B	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**0.13**	**0.36**	0.01	< 0.01	< 0.01
59	D	C	C	< 0.01	< 0.01	**0.12**	< 0.01	0.01	0.01	**7.88**	< 0.01	0.01	< 0.01
60	D	C	D	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**0.10**	**0.16**	< 0.01	< 0.01	< 0.01

the 10 experiments, and the values of the weights for each model in each batch are shown in Table 7; the 99.9% confidence sets of models for each batch are highlighted in bold.

The analysis of the confidence set in each batch gives insight into the interpretation of the Akaike weights as probabilities. The weight of a model states the probability, sustained by the data and by the models inserted into the testing set, of that model being equivalent to the true model. One cannot simply select the model with highest probability, as this value may be affected by factors such as the number of available data points, the experimental error inherent to the measured data, the number of models in the testing set, among others (Burnham and Anderson, 2002; Wagenmakers and Farrell, 2004, Li, et al. 2017). In Table 7, the comparison of the values of the Akaike weights obtained in experiments 1 and 2 demonstrates how the construction of a confidence set can overcome the variability inherent to the method. As these two experiments are replicates (see Table 1), it is expected that any one model should show the same goodness of fit; due to experimental variability, the value of the $w_\%$ is not exactly the same, neither is the order of quality of the different models in the set. Nevertheless, the confidence set in both experiments is exactly the same, allowing the conclusion and confirmation that these models are equally good in describing the kinetics in these cultivation conditions.

In the other eight cultivation conditions (experiments 3 to 10) different confidence sets were obtained for each case. The modification of pH and temperature influenced the metabolism in such a way that in one condition the growth kinetics approached a profile that may be described satisfactorily by one model, but this same model cannot describe the kinetics in other conditions. This situation is best exemplified in experiment 10, were the seemingly best partial model is 08, showing an Akaike weight of 55.11, while in any other condition the model fits so badly to the experimental data that it is not even included in the confidence set.

As the purpose of the study is to select a model that can describe the cell kinetics in any condition, the models that are included in all the confidence sets of the 10 experiments were indicated by an arrow at the left side of the MP code in Table 7. Three partial models were included in all 10 confidence sets: 23, 25 and 26. This means that all the data obtained in the 10 experiments are enough to support that any of these three partial models should be chosen over the remaining models in the testing set, but do not give clear evidence on which of the three is the most adequate.

By analyzing the equations that are used in the three selected partial models, it can be seen that only one equation of biomass formation is adequate for the description of Hib's cell growth. The biomass formation (X_f) model B (see Table 2) is used in all the three partial models, thus defining the best choice for this variable independent of the kinetics of acetate formation. For acetate formation itself, it can be seen that partial models 23 and 25 describe almost the same configuration, depicting non-associated formation (A_f model B in Table 3) with inhibition; the only difference is the equation used for the inhibition factor. It is interesting to note that the two possible equations for the inhibition factor are the ones that were extended by the use of an exponential factor (see A_i models B and D in Table 4). The third partial model, 26, discards the inhibition factor and incorporates both associated and non-associated formation (see A_f model C in Table 3).

One last remark can be made before the consideration of the polysaccharide formation model. The acetate kinetics in partial models 23 and 25 are described by one parameter in the formation factor and two parameters in the inhibition factor, totalizing three parameters; in partial model 26, only the two parameters of the formation factor are used. Therefore, the first two models are more complex than the third, and it could be argued that, as the simpler model shows good fitness to the data in all the 10 experiments, the two more complex ones should be ignored. However, the Akaike weights are evidence that the increase in fitting quality for partial models 23 and 25 was enough to compensate for the extra parameter. In order to ensure experimental support in the choice of the complete model, all three partial models were considered in the next steps.

To evaluate the polysaccharide formation kinetics, the three models of biomass and acetate formation were combined to the three models of polysaccharide formation (see Table 3) and to the five models of polysaccharide inhibition (see Table 4), thus resulting in a testing set of forty five complete models. The models were fitted to the experimental data in the same way, and again not all the equations could be fitted satisfactorily to all the ten experiments. From the forty five complete models, only twelve converged properly, and the remaining ones were discarded. The same procedure for the calculation of AIC_c and $w_\%$ was used, and the results are shown in Table 8; the 99.9% confidence sets are highlighted in bold.

In the case of the complete model, the confidence sets for the two replicate experiments (1 and 2) were not exactly the same, but they were very similar, with only one model (26/07) not being included in both sets. This may have arisen due to the extra information added by

Table 8. Akaike weights calculated for the complete models with PRP formation.

Model Codes			Experiments									
PM/CM	P_f	P_i	1	2	3	4	5	6	7	8	9	10
23/01	A	O	< 0.01	< 0.01	< 0.01	<u>0.08</u>	< 0.01	0.02	< 0.01	< 0.01	< 0.01	< 0.01
25/01	A	O	< 0.01	< 0.01	< 0.01	0.03	< 0.01	0.01	< 0.01	< 0.01	< 0.01	< 0.01
25/06	B	O	< 0.01	< 0.01	< 0.01	< 0.01	0.01	< 0.01	< 0.01	< 0.01	< 0.01	<u>0.09</u>
25/15	C	D	< 0.01	< 0.01	< 0.01	0.03	< 0.01	<u>28.12</u>	< 0.01	<u>25.09</u>	< 0.01	0.01
26/01	A	O	< 0.01	< 0.01	< 0.01	0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
26/06	B	O	< 0.01	< 0.01	< 0.01	< 0.01	0.01	< 0.01	0.01	< 0.01	< 0.01	0.02
26/07	B	A	<u>5.75</u>	< 0.01	<u>0.12</u>	< 0.01	<u>29.09</u>	< 0.01	<u>0.08</u>	< 0.01	< 0.01	<u>17.35</u>
26/08	B	B	<u>6.33</u>	<u>34.04</u>	<u>0.12</u>	<u>66.16</u>	<u>1.03</u>	< 0.01	< 0.01	<u>37.53</u>	<u>78.24</u>	<u>30.13</u>
26/09	B	C	< 0.01	< 0.01	< 0.01	< 0.01	<u>23.41</u>	< 0.01	<u>0.07</u>	< 0.01	< 0.01	<u>11.81</u>
> 26/10	B	D	<u>1.93</u>	<u>54.69</u>	<u>0.20</u>	<u>32.50</u>	<u>1.06</u>	<u>71.84</u>	<u>2.46</u>	<u>35.78</u>	<u>21.11</u>	<u>31.62</u>
26/11	C	O	<u>85.99</u>	<u>11.25</u>	<u>99.56</u>	<u>0.86</u>	<u>45.38</u>	< 0.01	<u>97.35</u>	< 0.01	< 0.01	<u>7.95</u>
26/13	C	B	< 0.01	0.01	< 0.01	<u>0.33</u>	0.01	< 0.01	0.03	<u>1.61</u>	<u>0.65</u>	<u>1.02</u>

the consideration of the PRP formation kinetics, the differences in the distribution of the experimental data, the increased number of parameters in the model to be penalized for and the reduced number of models in the testing set.

Similar to what happened with the partial models, the confidence sets for the complete models in the other eight experiments differed significantly. However, while the previous selection procedure using the partial model disregarding PRP formation could only outline clearly the best choice for biomass formation leaving uncertainty in the model of acetate synthesis, the ten confidence sets of complete models showed only one common model, as seen in Table 8. With the addition of one more variable in the model, more statistical information was given to the test and so the data obtained in the experiments could give stronger support to the choice of the best complete model, increasing the penalty for model complexity and broadening the differences in the evidence ratios of the models.

The complete model that is chosen in this procedure is the one of code 26/10 for which the parameters are shown in Table 9. This model states that biomass is subjected to inhibition from acetate, following an inhibition kinetics similar to the proposed by Aiba et al. (1968) but with the addition of an extra exponent in the formula; acetate follows uninhibited mixed formation kinetics, while PRP formation is non-growth associated (P_f model B from Table 3) and is subjected to acidic inhibition, following the inhibition factor of Yano and Koga (1973) (P_i model D from Table 4). Therefore, the complete model proposed in this work as supported by experimental data to describe the kinetics of Hib growth and polysaccharide production is described by eqs. 10 through 12 below.

$$\frac{dX}{dt} = \mu_{max} \exp\left(-\left(k_X\, A\right)^{n_X}\right) X \tag{10}$$

$$\frac{dA}{dt} = \left(\alpha_A \frac{dX}{dt} + \beta_A\, X\right) \tag{11}$$

$$\frac{dP}{dt} = \beta_P \frac{1}{1 + \left(\frac{A}{k_P}\right)^{n_P}} X \tag{12}$$

The overall quality of the fitted model is demonstrated in Figures 1 and 2, where the results of experiments 1 and 2 with the measured data and the experimental error estimates can be compared to the fitted model. All the other experiments and fitted curves can be assessed in Cintra (2014).

Table 9. Fitted values for the parameters of the final chosen model, at each run of the experimental design.

Experimental Design			Biomass Parameters			Acetate Parameters		Polysaccharide Parameters		
Run	T (°C)	pH	μ_{max} (h^{-1})	k_X (g$_A$.L^{-1})	n_X	α_A (g$_A$.g$_X^{-1}$)	β_A (g$_A$.g$_X^{-1}$.h^{-1})	β_P (mg$_P$.g$_X^{-1}$.h^{-1})	k_P (g$_A$.L^{-1})	n_P
1	33.0	6.90	0.50	7.10	2.69	0.62	0.11	39.97	7.71	5.12
2	33.0	6.90	0.49	7.38	3.77	0.65	0.11	37.67	8.08	7.06
3	35.8	7.32	0.50	9.25	4.09	0.88	0.12	44.43	9.64	3.11
4	35.8	6.48	0.60	6.30	3.50	0.60	0.11	32.29	6.91	11.82
5	30.2	7.32	0.29	8.55	3.17	0.66	0.08	23.60	12.63	3.07
6	30.2	6.48	0.37	6.60	7.70	0.54	0.10	24.25	6.98	17.43
7	33.0	7.50	0.34	9.91	3.15	0.79	0.10	31.26	12.08	5.48
8	33.0	6.30	0.40	5.82	7.54	0.44	0.13	20.52	6.54	16.29
9	37.0	6.90	0.61	7.44	3.01	0.56	0.14	41.33	8.39	11.88
10	29.0	6.90	0.26	7.35	4.50	0.61	0.08	20.63	7.52	13.43
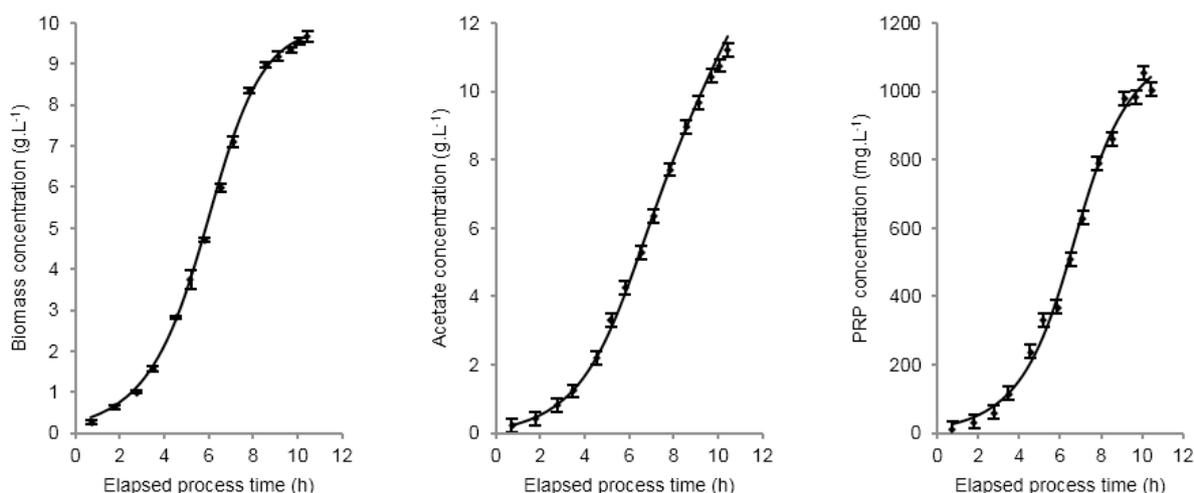
Figure 1. Experimental data and fitted model for experiment 1.
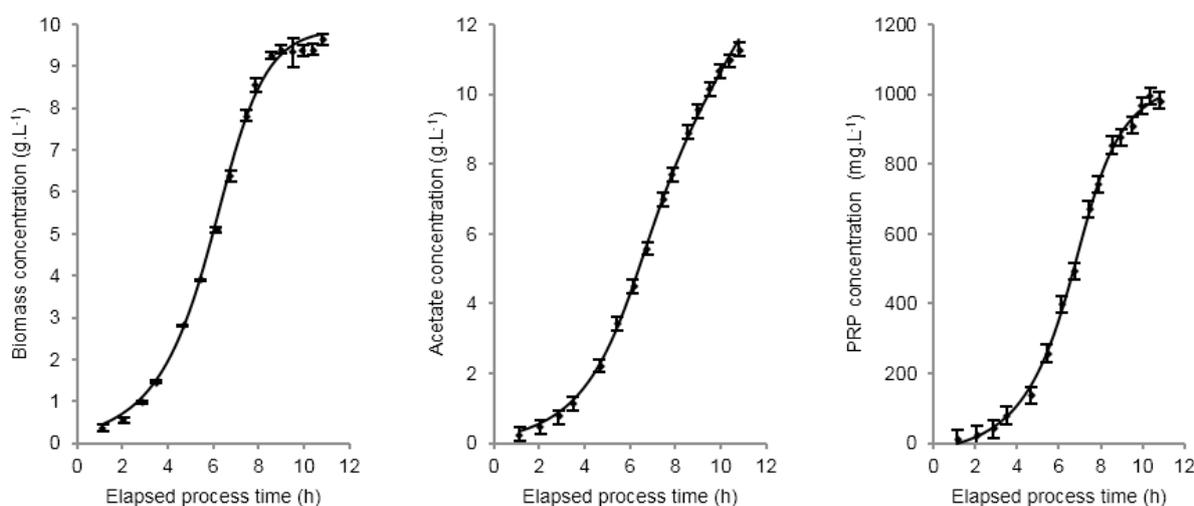
Figure 2. Experimental data and fitted model for experiment 2.

CONCLUSION

In this work, the techniques of model selection based on Akaike's Information Criterion were applied to models of bacterial kinetics. The procedure for converting the experimental data into AIC_c values for the multivariate kinetic models was outlined and carried out on real data from Hib cultivations. The performance of this technique could be evaluated by using experimental data from ten experiments generated from a RCCD, including experiments in replicate and in different culture conditions of pH and temperature. This procedure enabled visualization and understanding of the meaning of the Akaike weights in the probability sense and for assessing the robustness of the technique. By analyzing the confidence set of models obtained from these ten experiments, it was possible to point out one model with good evidence for support. The model of Hib kinetics selected in this paper describes the inhibition of cellular growth by acetate and the formation of two main metabolites, acetate and PRP. Acetate synthesis was determined as being free of inhibition and following a mixed type formation, while PRP is subjected to acetate inhibition and is not growth-associated. Inhibition of cell growth was found to be of the exponential type, as described by Aiba et al. (1968), and the inhibition of polysaccharide formation was found to be of the hyperbolic type, as described by Yano and Koga (1973).

The selection procedure using AIC was demonstrated to be a noteworthy alternative for defining models for microbial kinetics in discontinuous batches, replacing the methods of approximating derivatives that are traditionally used in the field, but that are error prone and bound to subjectivity of the operator.

REFERENCES

Aiba, S.; Shoda, M. and Nagatani, M., Kinetics of product inhibition in alcohol fermentation. Biotechnol. Bioeng., 10(6), 845–864 (1968).

Anderson, P.; Pitt, J. and Smith, D.H., Synthesis and release of polyribophosphate by *Haemophilus influenzae* type b in vitro. Infect. Immun., 13(2), 581–589 (1976).

Bedrick, E. and Tsai, C., Model selection for multivariate regression in small samples. Biometrics, 50(1), 226–231 (1994).

Biberstein, E.L.; Mini, P.D. and Gills, M.G., Action of *Haemophilus* cultures on delta-aminolevulinic acid. J. Bacteriol. 86, 814–819 (1963).

Burnham, K.P. and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Burnham, K.P. and Anderson, D.R., Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociol. Methods Res., 33(2), 261–304 (2004).

Cintra, F.O., Mathematical model and optimization of exopolysaccharide production by *Haemophilus influenzae* type b. Master thesis. Programa de Pós-Graduação Interunidades em Biotecnologia USP/ Instituto Butantan/IPT (2014).

Crisel, R.M.; Baker, R.S. and Dorman, D.E., Capsular polymer of *Haemophilus influenzae* type b. I. Structural characterization of the capsular polymer of strain Eagan. J. Biol. Chem., 250(13), 4926–4930 (1975).

De Haan, A.; Van der Put, R.M.F. and Beurret, M., HPAEC-PAD method for the analysis of alkaline hydrolyzates of *Haemophilus influenzae* type b capsular polysaccharide. Biomed. Chromatogr., 27(9), 1137–1142 (2013).

Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R; Bult, C.J.; Tomb, J.-F.;, Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.-I.; Glodek, A.; Kelley, J.M., Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; and Craig Ventert, J. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science, 269(5223), 496–512 (1995).

Greaves, R.I., Preservation of living cells by freeze-drying. Ann. N. Y. Acad. Sci., 85(2), 723–728 (1960).

Hurvich, C. and Tsai, C., Regression and time series model selection in small samples. Biometrika, 76(2), 297–307 (1989).

Kroll, J.S. and Moxon, E.R., Capsulation and gene copy number at the cap locus of *Haemophilus influenzae* type b. J. Bacteriol., 170(2), 859–864 (1988).

Li, Y; Shi, X; Yang, M and Liang, L., Variable selection in data envelopment analysis via Akaike's information criteria. Ann. Oper. Res., 253, 453–476 (2017).

Loeb, M.R., Ferrochelatase activity and protoporphyrin IX utilization in *Haemophilus influenzae*. J. Bacteriol., 177(12), 3613–3615 (1995).

Luedeking, R. and Piret, E.L., A kinetic study of the lactic acid fermentation. Batch process at controlled pH. J. Biochem. Microbiol. Technol. Eng., 1(4), 393–412 (1959).

Motulsky, H. and Christopoulos, A., Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting. San Diego: GraphPad Software Inc. (2003)

Reidl, J.; Schlör, S.; Kraiss, A.; Schmidt-Brauns, J.; Kemmer, G. and Soleva, E., . NADP and NAD utilization in *Haemophilus influenzae*. Mol. Microbiol., 35(6), 1573–1581 (2000).

Schilling, C.H. and Palsson, B.O., Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. J. Theor. Biol., 203(3), 249–283 (2000).

Stepanova, N.; Romanovskiĭ, I. and Ierusalimskiĭ, N.,. Mathematical modelling of the growth of microorganisms in the case of continuous (uninterrupted) culture). Dokl. Akad. Nauk. SSSR, 163(5), 1266–1269 (1965).

Takagi, M.; Cabrera-Crespo, J.; Baruque-Ramos, J.; Zangirolami, T.C.; Raw, I. and Tanizaki, M.M., Characterization of polysaccharide production of *Haemophilus influenzae* type b and its relationship to bacterial cell growth. Appl. Biochem. Biotechnol., 110(2), 91–100 (2003).

Takagi, M., Cabrera-Crespo, J., Zangirolami, T. C., Raw, I., and Tanizaki, M., Improved cultivated conditions for polysaccharide production by *H. influenzae* type b. Journal of Chemical Technology and Biotechnology, 81, 82–188 (2006).

Wagenmakers, E.J. and Farrell, S., AIC model selection using Akaike weights. Psychon. Bull. Rev., 11(1), 192–196 (2004).

Wilfert, C.M., . Epidemiology of *Haemophilus influenzae* type b infections. Pediatrics 85(4), 631–635 (1990).

Yano, T. and Koga S., Dynamic behavior of the chemostat subject to product inhibition. J. Gen. Appl. Microbiol., 19(2), 97–114 (1973).

Yogev, R.; Arditi, M.; Chadwick, E.G.; Amer, M.D. and Sroka, P.A., *Haemophilus influenzae* type b conjugate vaccine (meningococcal protein conjugate): immunogenicity and safety at various doses. Pediatrics, 85(4), 690–693 (1990).