The metabolic network of *Clostridium acetobutylicum*: Comparison of the approximate Bayesian computation via sequential Monte Carlo (ABC-SMC) and profile likelihood estimation (PLE) methods for determinability analysis

Graeme J. Thorn [a,*], John R. King [b]

[a] *School of Pharmacy, East Drive, University of Nottingham, University Park, Nottingham, NG7 2RD, UK*
[b] *School of Mathematical Sciences, Mathematical Sciences Building, University of Nottingham, University Park, Nottingham, NG7 2RD, UK*

ARTICLE INFO

ABSTRACT

The Gram-positive bacterium *Clostridium acetobutylicum* is an anaerobic endospore-forming species which produces acetone, butanol and ethanol via the acetone–butanol (AB) fermentation process, leading to biofuels including butanol. In previous work we looked to estimate the parameters in an ordinary differential equation model of the glucose metabolism network using data from pH-controlled continuous culture experiments. Here we combine two approaches, namely the approximate Bayesian computation via an existing sequential Monte Carlo (ABC-SMC) method (to compute credible intervals for the parameters), and the profile likelihood estimation (PLE) (to improve the calculation of confidence intervals for the same parameters), the parameters in both cases being derived from experimental data from forward shift experiments. We also apply the ABC-SMC method to investigate which of the models introduced previously (one non-sporulation and four sporulation models) have the greatest strength of evidence. We find that the joint approximate posterior distribution of the parameters determines the same parameters as previously, including all of the basal and increased enzyme production rates and enzyme reaction activity parameters, as well as the Michaelis–Menten kinetic parameters for glucose ingestion, while other parameters are not as well-determined, particularly those connected with the internal metabolites acetyl-CoA, acetoacetyl-CoA and butyryl-CoA. We also find that the approximate posterior is strongly non-Gaussian, indicating that our previous assumption of elliptical contours of the distribution is not valid, which has the effect of reducing the numbers of pairs of parameters that are (linearly) correlated with each other. Calculations of confidence intervals using the PLE method back this up. Finally, we find that all five of our models are equally likely, given the data available at present.

1. Introduction

The acetone–butanol–ethanol (ABE) fermentation process [1] uses the Gram-positive anaerobic bacterium *Clostridium acetobutylicum* to produce the three named solvents from sugars. ABE fermentation can be used to produce butanol for use in biofuels, and recent work has aimed to increase fermentation efficiency. In a simple batch culture, the metabolism of *Clostridium acetobutylicum* proceeds through an initial acidogenic phase followed by a later solventogenic phase. The primary driver of the shift is the external pH—a break-point exists at around pH 5.2, with acidogenesis strongly prevalent (primarily producing acetate and butyrate) above this value, and solventogenesis strongly prevalent (primarily producing acetone and butanol) below this value. Ethanol is produced in both phases in small amounts. In a batch culture, the acids produced in acidogenesis build up and the pH fall triggers a shift to solventogenesis, with the acids reabsorbed. In a continuous (phosphate-limited) culture, the transition between the two states can be directly induced through external pH control [2], without necessarily requiring acid or solvent building up. In this case, internal cellular pH is held at approximately 1 pH unit below the external conditions [3].

Following from our previous work [4] (based on a model originally introduced in [5]), a reduced version of the full metabolic network is shown in Fig. 1. Hexose sugars are internalized via a phosphoenolpyruvate-dependent phosphotransferase system, and

* Corresponding author. fax.: +44 115 951 3423; fax: +44 115 951 5102.
E-mail addresses: graeme.thorn@nottingham.ac.uk (G.J. Thorn), john.king@nottingham.ac.uk (J.R. King).
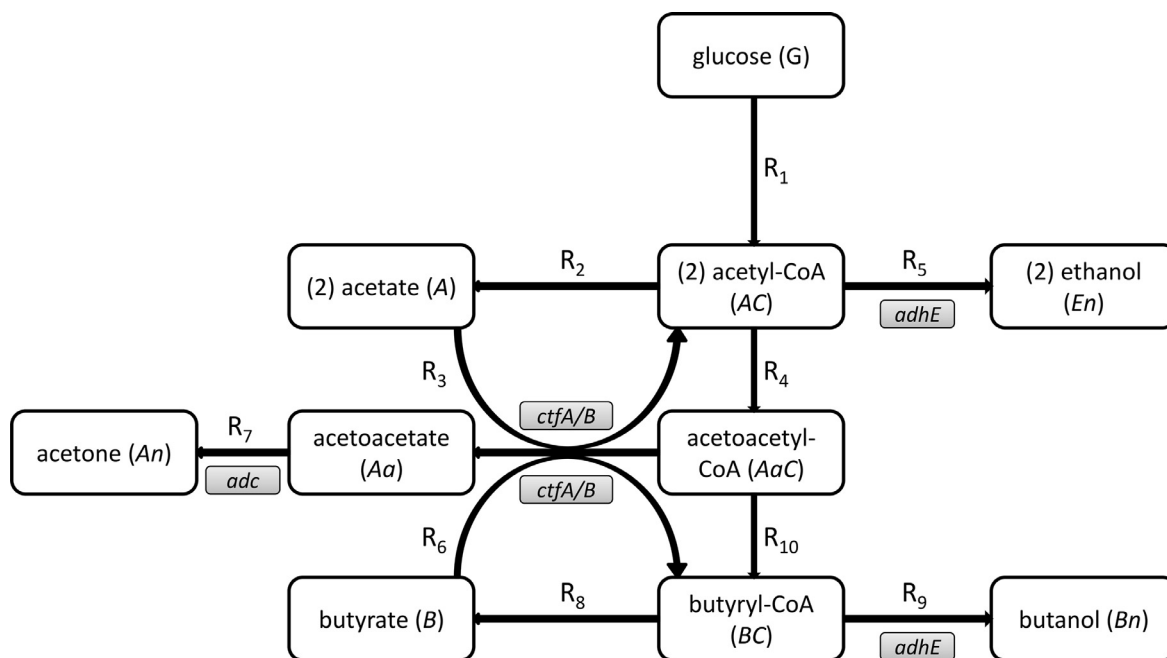
Fig. 1. Diagram of the simplified *Clostridium acetobutylicum* metabolism network (after Haus et al. [5]), showing the ten reactions considered here, and the relevant metabolic enzymes (G: glucose, AC: acetyl-CoA, A: acetate, AaC: acetoacetyl-CoA, BC: butyryl-CoA, B: butyrate, En: ethanol, Bn: butanol, Aa: acetoacetate, An: acetone, *ctfA/B*: CoA-transferase (with two subunits), *adc*: acetoacetate decarboxylase, *adhE*: aldehyde dehydrogenase). Note that the gene marked *adhE* in this network is often called *adhE1*.

then glycolysed into pyruvate [1]. Further steps convert pyruvate into the branch-point intermediate acetyl coenzyme A (acetyl-CoA). The 2-carbon products (acetate and ethanol) are formed from this intermediate, but further metabolism of acetyl-CoA into acetoacetyl-CoA and butyryl-CoA provides the starting point for the 3-carbon products (butyrate, acetone and butanol). The three coenzyme A products form the important (pH-independent) backbone of ABE fermentation and the two end points acetyl-CoA and butyryl-CoA are branch points where acidogenesis and solventogenesis diverge. During acidogenesis, acetyl-CoA and butyryl-CoA are phosphory-lated, converted into acetate and butyrate, and then excreted [1]. Post-metabolic shift, acetate and butyrate are reabsorbed, converted back into acetyl-CoA and butyryl-CoA via Ping-Pong-Bi-Bi reactions (mediated by a CoA-transferase of two subunits CtfA/B), which has the effect of converting acetoacetyl-CoA into acetoacetate and then into acetone (via the acetoacetate decarboxylase Adc). Finally, acetyl-CoA and butyryl-CoA are converted into acetaldehyde and butyraldehyde, and then into ethanol and butanol. Both reactions are mediated by an acetaldehyde dehydrogenase (AdhE, sometimes referred to as AdhE1), which is only active during solventogenesis. The relative priority of the acidogenic and solventogenic pathways in *C. acetobutylicum* strongly depends upon the external conditions (particularly pH [2]), but the organism produces ethanol indepen-dently of acetone and butanol fermentation whilst in continuous culture—ethanol is produced at the same rate in both phases of the metabolism. The solventogenesis-influencing enzymes CtfA/B and Adc are produced in higher quantities during solventogenesis compared to acidogenesis [1], measured in both steady-state and forward shift experiments. Solventogenesis itself is part of *C. ace-tobutylicum*'s survival mechanism in low pH conditions: the other part is sporulation, which is not included in the initial metabolism model. There is little evidence of spore formation during continuous culture experiments; however, spores have been observed during batch culture where population levels can vary widely [2].

Our previous work [4] using this model suggested that some key parameters in the model were ill-determined and that parameters related to the main metabolic backbone through the intermediates acetyl-CoA, acetoacetyl-CoA and butyryl-CoA were highly correlated,

and by introducing a very simple sporulation model we could im-prove the model's fitting, but the evidence for sporulation was not conclusive. Here we improve on this result, by using the Bayesian method of ABC-SMC [6] and its model-selection extension [7]. This allows us to quantify actively which parameters can be determined from the dataset (and which are indeed correlated), and to explore which of the models (including the non-sporulation model) has the most evidence given the data. Our new results show that the com-mon assumption in determinability analysis (such as in [8]) that con-tours of the posterior distribution are ellipsoidal, appropriate in the asymptotic limit of large numbers of data points (indicating a joint Gaussian distribution) does not always hold; the implication in the current context is that the previous post-hoc discovery of apparent correlations between parameters turn out to be spurious. Comparing the results of the approximate posterior distribution with the results of post-hoc confidence intervals derived using the profile likelihood [9], which uses the true shape of the residual-sum-of-squares func-tion rather than the asymptotic function, backs this finding of spu-rious correlations up, further strengthening the case for the use of determinability analysis using non-asymptotic methods. Moreover, the current work demonstrates the importance of determining the concentrations of the internal intermediates acetyl-CoA, acetoacetyl-CoA and butyryl-CoA during shift experiments in order to calculate the undetermined reaction constants of the metabolism model, and that assays to determine the population composition of the cells are required to determine any potential spore fraction. These limitations suggest that in particular the basic non-sporulation model is the best one to use for determining the potential metabolic behaviour of *C. acetobutylicum* mutants deficient in any of the three genes (*ctfA/B, adc* or *adhE*) explicitly included in the model.

The combination of the ABC and PLE methodologies in this pa-per has enabled us to make these conclusions, and as far as the au-thors are aware, this has not been attempted previously on such a high-dimensional problem. Neither has either method been used on a system with this many parameters. By stretching the methods to the limits of their computational feasibility, this study has given us a unique insight into how the two methods behave in such a situa-tion. The current results thus not only have implications specific to