CHAPTER 5

Identifying Inhibitory Compounds in Lignocellulosic Biomass Hydrolysates using An Exometabolomics Approach

Ying Zha, Johan A. Westerhuis, Bas Muilwijk, Karin M. Overkamp, Bernadien M. Nijmeijer, Leon Coulier, Age K. Smilde, and Peter J. Punt

This chapter has been submitted to '*BMC Biotechnology*'

Abstract

Background

During the pretreatment process of lignocellulosic biomass, inhibitors are formed that reduce the fermentation performance of the fermenting yeast. To systematically identify inhibitors in lignocellulosic biomass hydrolysates, an exometabolomics approach was applied.

Results

We studied the composition and fermentability of 24 different biomass hydrolysates. To create diversity, the 24 hydrolysates were prepared from six different biomass types, and with four different pretreatment methods. Their composition and that of fermentation samples generated with these hydrolysates were analyzed with two GC-MS methods. To preclude sugars, which obscure the detection of less abundant compounds, either ethyl acetate extraction or ethyl chloroformate derivatization was used to treat samples before conducting GC-MS. Through establishing relationship between fermentability and composition of the hydrolysates, using multivariate PLS-2CV and nPLS-2CV data analysis models, potential inhibitors were identified. These identified compounds were tested for their effects on the growth of the model yeast, confirming that the majority of the identified compounds were indeed inhibitors.

Conclusion

Using a non-targeted systematic approach, metabolomics, inhibitory compounds in lignocellulosic biomass hydrolysates were successfully identified. The identified inhibitors include both known ones, such as furfural, HMF and vanillin, and novel inhibitors, namely sorbic acid and phenylacetaldehyde.

Background

Lignocellulosic biomass, like bagasse, wheat straw, and corn stover, is the 2nd generation feedstock for biofuel production. Compared to fossil fuel, it is abundant, renewable and environmental friendly, while compared to 1st generation feedstock, like corn, it does not compete with world food supply [1,2]. Lignocellulosic biomass is composed of cellulose, hemicellulose and lignin, of which cellulose is the homopolymer of glucose, while hemicellulose is a heteropolymer mainly composed of glucose and xylose [3,4]. To produce biofuel from lignocellulosic biomass, a pretreatment step is required to break down its structure and expose cellulose for enzymatic hydrolysis [5,6]. The hydrolysis product, so-called biomass hydrolysate, is used as substrate for biofuel production through fermentation processes [7]. During most biomass pretreatment processes, harsh conditions, like high temperature and high pressure, were adopted. This causes sugars and lignin in biomass hydrolysates to degrade, forming products that possess inhibitory effects towards fermenting hosts, thus resulting in reduced growth and productivity [8-11].

Research has been conducted to identify the compounds in the biomass hydrolysates that cause inhibitory effects [12-14]. In these studies, it was found that the inhibitors are mainly of the following three categories, weak acids, furans and phenolic compounds, and the most frequently studied representatives are acetic acid, furfural, HMF and vanillin, respectively [15-17]. A variety of experimental and analytical methods were used in these studies for identifying inhibitory compounds, and a common feature of these studies was that the approach was targeted [18]. In another words, a group of compounds were selected prior to hydrolysate compositional analysis, based on knowledge of lignocellulosic biomass structure and previous research. The selected compounds were analyzed for their presence in the biomass hydrolysates and their toxicity towards the fermenting microorganisms [11,19,20].

Besides the identified inhibitors, evidence was obtained showing that other compounds present in biomass hydrolysates also display inhibitory effects [21,22]. They were observed as unknown peaks in hydrolysate compositional analysis results, which reduced in size after detoxification [23]. To identify novel inhibitory compounds in biomass hydrolysates, in this study a non-targeted exometabolomics approach was applied. Generally, metabolomics is one of the ‘omics’ tools that studies the performance of research objects by analyzing their overall compositions [24,25]. In this study, research objects are lignocellulosic biomass hydrolysates, which are used as fermentation media for bioethanol production. The performance of biomass hydrolysates as fermentation media vary due to the difference in their compositions, i.e. inhibitory compounds and their concentrations. Through establishing

the relation between the composition and the performance of different biomass hydrolysates statistically, compounds that possess inhibitory effect could be indicated in an unbiased way (Figure 1).

In metabolomics, the search for important metabolites responsible for a certain response, e.g. fermentability, is often performed with multivariate data analysis methods [26,27]. These multivariate methods are able to search for the interactions between metabolites that are responsible for the response that is modeled. Partial least squares (PLS) is a multivariate data analysis method that is commonly used in metabolomics to search for the important metabolites [28]. As an extension of the PLS method, also n-way PLS may be used when the data-set consists of a time series such as in the case of our metabolomics experiments. As multivariate data analysis methods may lead to false positive correlations, rigorous validation of these models is necessary [29,30]. Therefore, we decided to use double cross validation (2CV) to obtain unbiased prediction errors for the (n)PLS models [31,32].

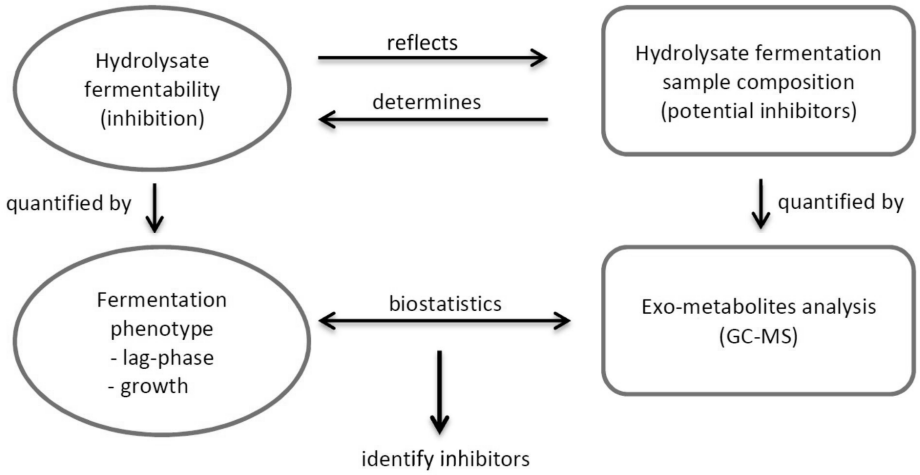


Figure 1 Graphic illustration of the concept of the exometabolomics approach.

We report here the detailed procedure and the results of using the exometabolomics approach for identifying inhibitory compounds in lignocellulosic biomass hydrolysates. This includes the batch fermentability of 24 different biomass hydrolysates using baker's yeast, *S. cerevisiae* CEN.PK113-7D, and the analysis results of the fermentation samples by two GC-MS methods; the statistical model building procedure for identifying potential inhibitory

compounds, and the toxicity testing results of the suggested potential inhibitors. The results of this study show that of the potential inhibitory compounds indicated by the statistical models, a large fraction indeed exhibited inhibitory effects on the growth of fermenting yeast. These compounds consist of both known inhibitors, such as furfural and HMF, and novel inhibitors.

Materials and Methods

Biomass hydrolysate preparation and fermentation

24 different hydrolysates were prepared from six types of biomass, by using four different hydrolysate preparation methods. The six types of biomass were sugar cane bagasse (Zillor, Brazil), corn stover (University of Cape Town, South Africa), wheat straw (Oostwaardshoeve, The Netherlands), barley straw (Oostwaardshoeve, The Netherlands), willow wood chips (Oostwaardshoeve, The Netherlands) and oak sawdust (wood-flooring supplier ESCO, The Netherlands). Prior to pretreatment, biomass (except oak sawdust) was ground to pieces of average length 3 mm and dried at 80°C for at least 16 hours. To prepare 1 l hydrolysate, 300 g dried biomass was used. The four hydrolysate pretreatment methods were dilute acid (2% H_2SO_4), mild alkaline (3% $Ca(OH)_2$), alkaline/peracetic acid and concentrated acid (72% H_2SO_4). The biomass pretreated with the first three methods was hydrolyzed enzymatically, using Accellerase 1500 (Genencor®), while acid hydrolysis was used for biomass pretreated with concentrated acid (40% and 15% H_2SO_4). The detailed pretreatment and hydrolysis procedure was described in Zha *et al.* [33]. After hydrolysis, solid content was separated from the hydrolysate by filtration, and the filtrated hydrolysate was sterilized using filter sterilization and stored at 4°C before use.

Batch fermentations were carried out in 2 l New Brunswick fermentors, using 1 l of sterilized hydrolysate as substrate. The fermenting yeast was *Saccharomyces. cerevisiae* CEN.PK 113-7D (CBS 8340), and the inoculum was prepared in a 500 ml Erlenmeyer flask. The cells were harvested by centrifugation after incubating overnight in mineral medium (MM) [34] with 20 g/l glucose, and inoculated into fermentors with density of 0.1 g cell dry weight per 1 l hydrolysate. All fermentations were carried out at 30°C, under anaerobic conditions by sparging 0.5 l/min N_2 continuously, and pH was set at 5 by adding 1 M H_2SO_4 or 2 M KOH.

For each of the 24 hydrolysates, one batch fermentation was conducted after checking its reproducibility [18]. During the whole fermentation process, CO_2 concentration in the off-gas

was monitored automatically and samples were taken at fixed time intervals. These samples were kept at 4°C and used to measure optical density (OD), glucose and ethanol concentration with either Cobas® Mira Plus (Roche) or Arena® 20 Analyzer (Thermo Scientific).

Hydrolysate fermentation sample analysis

For each of the selected hydrolysate fermentations, cell free time samples were chosen for analyzing their overall compositions. Two GC-MS methods, namely ethyl acetate extraction (EA)-GC-MS and ethyl chloroformate derivatization (EC)-GC-MS, were used to analyze the fermentation samples.

For EA-GC-MS, the extraction was done by adding 550 µl ethyl acetate into 0.5 ml sample and vortex for 2 min. The mixture was centrifuged to separate ethyl acetate fraction, of which 400 µl was transferred to a vial and dried by blowing N_2. The following internal standards in ethyl acetate were added to the same vial: phenylethanol-D5, cinnamic acid-D5 and hydroxybenzaldehyd-D4. The extraction and centrifugation process was repeated, and from the ethyl acetate fraction, another 400 µl was transferred to the same vial, after drying with N_2, the following internal standards in pyridine were added: alanine-D4 and citric acid-D4. The extract was then oxidized by adding 30 µl 56 mg/ml ethoxyamine·HCl, and incubating at 40°C for 90 min. Followed by adding dicyclohexylphtalate (DCHP) and difluorobiphenyl (DFB) in pyridine as injection standard, the oxidized extract was silylated by adding 100 µl *N*-methyl-*N*-trimethylsilyl-trifluoroacetamide (MSTFA), and incubating at 40°C for 50 min. Measurement was carried out by 1 µl splitless injection in the PTV injector of an Agilent® 7890A GC with an Agilent® 5975C MS as detector. The analytical column used was HP-5MS 30 m x 0.25 mm x 0.25 µm.

For EC-GC-MS, the sample pH was brought above 10 by adding NaOH solution, followed by the addition of following internal standards in pyridine: leucine-D3, succinic acid-D4 and cinnamic acid-D5. The injection standards, DCHP and DFB in pyridine, and 300 µl ethanol were also added to the sample. Then the ethylesters formation was done by two rounds of adding 40 µl ethyl chloroformate into the sample and shaking it vigorously by hand for 15 sec. The reaction was stopped by adding 750 µl dichloromethane and 500 µl 1 M bicarbonate buffer. The formed derivates were extracted with dichloromethane, and the extraction was dried with Na_2SO_4. The measurement was carried out the same way as in EA-GC-MS method. The analytical column used was DB-1 30m x 0.32mm x 1 µm.

The analysis results of EA-GC-MS and EC-GC-MS were reported separately in data-sets, with detected peaks as row and fermentation sample as column. The reported values were areas of the detected peaks after correction with internal standards.

Statistical model building

The two statistical models used were partial least square with double cross validation (PLS-2CV) [31] and n-way PLS with double cross validation (nPLS-2CV) [35]. The 2CV version of the nPLS model was developed in house. The models were written as m-files in MATLAB environment (R2012a) with PLS toolbox 2.0 (Eigenvector).

PLS-2CV models

PLS-2CV is a linear regression model, which predicts the fermentation phenotypes with the GC-MS analysis results of the fermentation samples (data-sets). The PLS-2CV models were assessed by calculating the so-called Q^2 values, which indicate the prediction ability of the data-sets for a specific phenotype [31]. The maximum value of Q^2 is 1, representing that the model could perfectly predict the phenotypes. Generally, models with $Q^2 \geq 0.5$ were selected for analyzing the selectivity ratios (SR) assigned to each peaks in the data-sets. Similar to regression coefficient ('reg'), SR is a measure for variable importance in discrimination models. Contrary to 'reg', SR is corrected for the influence of interfering compounds that are not related to the modeled response [36,37]. Peaks with the highest SR values were considered having the primary contribution to the model building. Among these peaks, the identified ones were selected as potential inhibitory compounds, and tested in Bioscreen C Analyzer for their effects on the fermenting yeast.

To model lag-phase, the data-sets containing the first two fermentation samples (t1 and t2) were used. As listed in Table 1, the difference as well as the combination of t1 and t2 data-sets were used to build PLS-2CV model. EA-GC-MS and EC-GC-MS data-sets were modeled both separately and combined. Thus, for lag-phase, in total six PLS-2CV models were built (Table 1). These data-sets were preprocessed by conducting a 'square-root' transformation to reduce the nonsymmetrical distributions of the peak areas for all compounds, and this also homogenizes the heteroscedastic measurement error. Afterwards, an 'auto-scaling' was carried out to reduce the effect that compounds with large peak areas would dominate the regression models [38,39]. The phenotype values were 'mean-centered' before data analysis.

To model glucose consumption rate (Glu CR) and ethanol production rate (EtOH PR) (see Eq2 to 4), the data-sets of all five fermentation samples were used individually (t1 to t5, see Table 3). The data preprocessing was conducted in the same way as by lag-phase data-sets (Table 1).

Table 1 Data-sets used for building PLS-2CV models.
The two data preprocessing methods were symbolized by '$\sqrt{\ }$' (square-root), and '*auto*' (autoscaling); ':' indicates that the corresponding data-sets were combined; EA_{tx}: EA-GC-MS data-set of time sample tx; EC_{tx}: EC-GC-MS data-set of time sample tx (x represents one of the five fermentation samples).

Lag-phase model 1	$auto(\sqrt{EA_{t2}} - \sqrt{EA_{t1}})$
Lag-phase model 2	$auto(\sqrt{EC_{t2}} - \sqrt{EC_{t1}})$
Lag-phase model 3	$auto((\sqrt{EA_{t2}} - \sqrt{EA_{t1}}) : (\sqrt{EC_{t2}} - \sqrt{EC_{t1}}))$
Lag-phase model 4	$auto(\sqrt{EA_{t1} : EA_{t2}})$
Lag-phase model 5	$auto(\sqrt{EC_{t1} : EC_{t2}})$
Lag-phase model 6	$auto(\sqrt{EA_{t1} : EA_{t2} : EC_{t1} : EC_{t2}})$
Glu CR and EtOH PR	$auto(\sqrt{EA_{tx}})$ or $auto(\sqrt{EC_{tx}})$

nPLS-2CV models

N-way PLS (nPLS) handles multiway data-sets, and was used to model glucose consumption rate (Glu CR) and ethanol production rate (EtOH PR). In this study, the data-sets were three-way, the three ways were (1) fermentation batch, (2) time samples of each batch, and (3) analysis results of each sample. The analysis results of EA-GC-MS and EC-GC-MS methods were used both separately and combined. Similar to PLS-2CV model, the data-sets were arranged in two way and preprocessed by conducting 'square-root' and 'auto-scaling' before transforming to the three-way structure. The phenotype values were 'mean-centered' before model building. The nPLS-2CV models were assessed by calculating the Q^2 values. In most cases, models with $Q^2 \geq 0.5$ were selected for analyzing the regression coefficient ('reg') of each peak in the data-sets, as SR for nPLS has not yet been developed. Peaks with highest absolute 'reg' values were considered having the most contribution for predicting the phenotypes. Among these peaks, the identified ones were selected as potential inhibitory compounds, and tested in Bioscreen C Analyzer for their effects on the fermenting yeast.

Potential inhibitory compound test

Solutions of potential inhibitory compounds were prepared in both MM with 20 g/l glucose and YPD (Yeast extract Peptone Dextrose) medium with concentrations of 0.2, 0.5 and 1.0 g/l. If a compound was saturated at 1.0 g/l, the solutions were made with 20%, 50% and 100% of the saturated concentration. These compounds are shown in bold type in Table 5.

The prepared solutions were used as media in the growth test of the fermenting yeast, *S. cerevisiae* CEN.PK 113-7D. The growth test was conducted in triplets in honeycomb plates, using Bioscreen C Analyzer (Labsystems OY). The detailed procedure of Bioscreen test is described in Zha *et al.* [33].

Results

Biomass hydrolysates preparation

To successfully identify inhibitory compounds in biomass hydrolysates with statistical models, acquiring hydrolysates with divers performance is of importance [18]. To achieve this, 24 different hydrolysates were prepared from six different biomass and by using four hydrolysate preparation methods. Among the six biomass, wheat straw, barley straw and corn stover are agricultural wastes, bagasse is sugar industry byproduct, and willow and oak are wood products. Each of the six biomass was pretreated with four different methods, which used 2% sulfuric acid, 72% sulfuric acid, lime, and peracetic acid, respectively. The resulting 24 hydrolysates were tested for their performance as fermentation media on a small scale (ml), showing that there was a significant diversity among these 24 hydrolysates [33]. For the exometabolomics study, these hydrolysates were prepared in larger quantity (l). For each hydrolysate, a batch fermentation of 1 l working volume was carried out based on previously developed procedures [11].

Defining phenotypes

Identical batch fermentations were carried out for each of the 24 different hydrolysates generated. The fermentability was monitored by measuring OD, glucose and ethanol concentrations of the samples taken with a fixed time interval. To quantify the fermentability of the hydrolysates, four phenotypes were defined, which are lag-phase, glucose consumption rate (Glu CR), ethanol production rate (EtOH PR) and ethanol yield (EtOH Y).

The definition of these four phenotypes are given in equation 1 to 4 (Eq1 to Eq4), and the measurement results of the fermentation samples were used to calculate these phenotypes.

Eq1: lag-phase = time to reach 2% (ODmax-ODmin) (h)
Eq2: Glu CR = the slope of the linear part of the glucose consumption curve (g/l/h)
Eq3: EtOH PR = the slope of the linear part of the ethanol production curve (g/l/h)
Eq4: EtOH Y= EtOHmax/initial glucose concentration (g/g)

As shown in the phenotype definitions, lag-phase has time as unit (Eq1), which represents the duration before growth began. Since during lag-phase, the fermenting yeast adapt to the media composition for growth [40], a longer lag-phase indicates the presence of compounds that restrain the starting of growth. Glucose consumption rate (Glu CR) is an indicator of the growth rate of the fermenting yeast, while ethanol production rate (EtOH PR) and ethanol yield (EtOH Y) describe the productivity of the fermenting yeast in a specific hydrolysate. For each of the 24 fermentations, these four phenotypes were calculated (Table 2). It should be mentioned that growth rate is one of the most commonly used phenotypes describing the performance of fermenting hosts. In this study, instead of using growth rate, we chose Glu CR to describe growth. This is because OD measurement results were not easily comparable due to sample characteristics, such as color differences among hydrolysates, and flocculation. To confirm that Glu CR is a good indicator of growth performance, OD% was used to calculate tentative growth rate (Table 2). It can be seen that the tentative growth rates have very similar trend compared to Glu CR (Figure 2). Since glucose measurements are more accurate than OD, we have decided to use Glu CR as an indicator of growth rate.

As shown in Table 2, all 24 hydrolysates had different glucose concentrations, indicating that biomass type as well as pretreatment method influenced the biomass hydrolysis efficiency. In general, mild alkaline (MA) pretreated biomass resulted in relatively low glucose concentration, while concentrated acid (CA) lead to higher hydrolysis efficiency [33]. Based on our previous results, the variation in glucose concentration in the range observed in Table 3 was of no influence on fermentation performance (results not shown).

The performance of the 24 hydrolysate varied significantly as fermentation media, which was consistent with the screening experiments on milliliter scale [33]. As far as lag-phase is considered, hydrolysates like Oak-CA and Willow-CA supported growth almost immediately after inoculation, while the fermenting yeast needed an adaptation period of as long as 10 hours in CS-CA and WS-CA hydrolysates. The Glu CR of the 24 hydrolysates ranged from 0.80 (Oak-CA) to 4.63 (WS-CA), which trend was comparable to that of EtOH PR (Figure 3).

This resulted in very similar ethanol yield among the hydrolysates, around 0.4 g/g (Table 2), which was also the ethanol yield of *S.cerevisiae* CEN.PK113-7D in mineral medium with 20 g/l glucose [41]. This observation suggested that under anaerobic conditions, the effect of inhibitory compounds in hydrolysates had little effect on the ethanol yield of the fermenting yeast. Therefore, this phenotype was not used in building statistical models for the purpose of identifying hydrolysate inhibitors.

Table 2 Fermentability of the 24 biomass hydrolysates, expressed as the calculation results of the defined phenotypes.

Hydrolysate	Glucose[1] (g/l)	Ethanol[2] (g/l)	Lag-phase[3] (h)	Glu CR[4] (g/l/h)	EtOH PR[5] (g/l/h)	EtOH Y[6] (g/g)	Tentative growth rate[7]
Bag-CA	**67.39**	**20.61**	**7.5**	**1.42**	**0.44**	**0.306**	**2.42**
Bag-DA	**63.33**	**24.20**	**6.0**	**3.64**	**1.52**	**0.382**	**4.61**
Bag-MA	**58.82**	**22.71**	**2.0**	**3.84**	**1.58**	**0.386**	**5.73**
Bag-PAA	52.48	19.87	3.0	2.52	0.71	0.379	2.76
BS-CA	**67.45**	**30.92**	**7.5**	**4.57**	**1.73**	**0.458**	**7.39**
BS-DA	**49.87**	**20.95**	**6.5**	**3.63**	**1.42**	**0.420**	**5.66**
BS-MA	42.56	18.40	6.0	3.05	1.41	0.432	*
BS-PAA	**53.50**	**22.22**	**3.0**	**2.96**	**1.03**	**0.415**	**5.12**
CS-CA	**65.63**	**26.62**	**10.5**	**3.21**	**1.32**	**0.406**	**4.73**
CS-DA	**42.80**	**18.74**	**5.5**	**3.43**	**1.49**	**0.438**	**6.98**
CS-MA	**32.83**	**15.85**	**6.5**	**3.35**	**1.32**	**0.483**	**7.92**
CS-PAA	50.29	20.84	3.5	2.38	1.03	0.414	4.53
Oak-CA	**66.72**	**12.06**	**1.5**	**0.80**	**0.29**	**0.181**	**2.1**
Oak-DA	**38.22**	**15.27**	**5.0**	**2.41**	**0.98**	**0.400**	**5.37**
Oak-MA	**44.35**	**19.49**	**2.5**	**3.43**	**1.55**	**0.439**	**7.52**
Oak-PAA	**60.80**	**25.97**	**3.0**	**2.73**	**1.12**	**0.427**	**3.75**
Willow-CA	31.58	13.60	1.0	4.26	1.10	0.431	14.04
Willow-DA	**45.15**	**17.68**	**7.5**	**2.74**	**1.14**	**0.392**	**5.72**
Willow-MA	23.50	10.76	4.5	2.68	1.29	0.458	*
Willow-PAA	**51.30**	**22.81**	**5.5**	**2.45**	**1.05**	**0.445**	**5.03**
WS-CA	**60.54**	**24.71**	**9.0**	**4.63**	**1.87**	**0.408**	**7.6**
WS-DA	**58.29**	**24.83**	**4.5**	**3.47**	**1.64**	**0.426**	**6.05**
WS-MA	**32.12**	**13.95**	**6.5**	**4.01**	**1.92**	**0.434**	**11.37**
WS-PAA	51.94	21.61	3.5	3.03	1.27	0.416	5.48

[1]: glucose concentration of the 24 hydrolysates; [2]: final ethanol concentration; [3]: (Eq1);
[4]: glucose consumption rate (Eq2); [5]: ethanol production rate (Eq3); [6]: ethanol yield (Eq4);
[7]: the slope of the linear part of the OD% curve. *: OD measurement was not possible due to flocculation; **bold**: fermentations that are selected for sample compositional analysis.

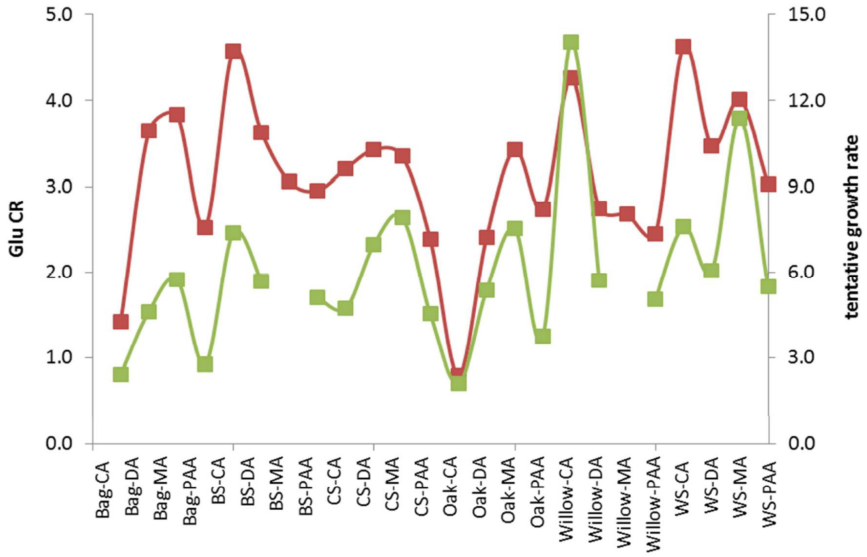


Figure 2 Trends comparison between tentative growth rate and glucose consumption rate. Tentative growth rate: the slope of the linear part of the OD% curve (green); Glucose consumption rate (Glu CR): the slope of the linear part of the glucose consumption curve (Eq2) (red).

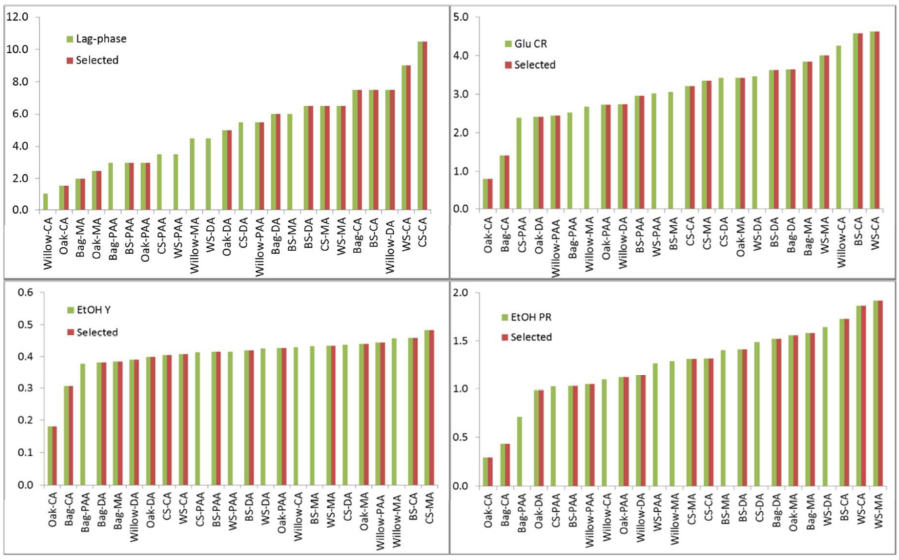


Figure 3 The calculation results of the four phenotypes, sorted from the smallest to the largest. The 'red' bars are the selected 16 fermentations for exometabolomics analysis.

Among the 24 hydrolysate fermentations, some had similar performance in terms of the phenotypes calculated (Figure 3). Since the statistical models to be used for analyzing the relationship between fermentability and sample composition were based on linear regression, it is important to reduce overrepresentation of certain phenotype classes. In addition, it is also beneficial to minimize the amount of samples for exometabolomics analysis. Therefore, from the 24 fermentations, 16 were selected based on the variations in their phenotypes, biomass type and pretreatment method. The selected 16 hydrolysates contain all six biomass types and all four biomass pretreatment methods (Table 2), and the fermentability of these selected hydrolysates show a more or less even spread of the fermentation phenotypes (Figure 3).

Hydrolysate fermentation sample analysis

After quantifying the performance of the hydrolysate fermentations with the four phenotypes, cell free time samples of the 16 selected fermentations were analyzed for their overall compositions. These samples were chosen based on the criteria that they should uniquely represent the whole fermentation process. As each fermentation can be divided into three phases, namely lag-phase, growth-phase and stationary-phase, five samples were selected, as shown in Table 3. The division of the three fermentation phases was consistent with the definition of the phenotypes, i.e. the end of lag-phase is when OD reaches 2% of the maximum OD, the end of growth-phase is when glucose consumption is completed, and the duration of stationary-phase is set at 10 hours. In this way, a total of 80 samples from 16 hydrolysate fermentations were selected for compositional analysis.

Table 3 The five fermentation time samples for compositional analysis with the two GC-MS methods.

t1	beginning of fermentation	immediately after inoculation
t2	end of lag-phase	time needed to reach 2% (ODmax-ODmin)
t3	growth mid-point	time needed to consume half of the initial glucose
t4	growth end point	time needed to consume all glucose
t5	stationary phase	10 hours after growth end point

The focus of the compositional analysis was the potential inhibitory compounds in hydrolysate samples, which are believed to be mainly non-sugar compounds, such as weak acids, furans and phenols [8,9,15]. GC-MS was chosen as the analytical tool, as the method is capable of detecting a wide range of these compounds, including many unknowns [20,42].

A crucial point in analyzing hydrolysate samples with GC-MS was to remove sugars, which are present in large quantities in the samples and severely interfere with the detection of non-sugar compounds [13,43]. For this purpose, two sample work-up methods were developed, namely, ethyl acetate extraction and ethyl chloroformate derivatization.

Ethyl acetate extraction GC-MS (EA-GC-MS) was previously described by Heer *et al.* [13]. The method uses ethyl acetate (EA) as solvent, in which compounds that are apolar, e.g. with aromatic rings, are dissolved, while polar compounds, like sugars, remain in the water phase. In current study, the hydrolysate samples were extracted twice with EA to allow adequate recovery of the extracts. After removing EA phase from the water phase, it was dried through evaporation, thus concentrated and ready for analysis with GC-MS.

Due to the nature of this sample work-up method, only compounds dissolvable in EA were analyzed, and since EA was removed through evaporation, volatile compounds were partially lost. This makes recovery an important issue in EA-GC-MS method, which was assessed with a group of furans and aromatic compounds before analyzing hydrolysates. It was found that sample pH influences the extraction, when pH was raised to above 6.0, significant decrease of recovery was observed with multiple aromatic standards. Therefore, all hydrolysate samples were extracted with EA at pH 5.0. In doing so, the recovery of aromatic compounds was satisfying, above 90%, while the recovery of furans was rather low and inconsistent due to evaporation. So the analysis results of aromatic compounds were considered more reliable than furans.

To complement EA-GC-MS method, ethyl chloroformate derivatization GC-MS (EC-GC-MS) was developed in our lab. Ethyl chloroformate (EC) was used to convert acids to their ethyl ester form, thus compounds like carboxylic acids, amino acids, aromatic compounds and furans could be detected by MS. EC-GC-MS therefore has a larger coverage of compounds compared to EA-GC-MS, and is easier to operate. But due to the diverse reactivity of compounds with EC, it is possible that compounds present with high concentration could only be detected with low signal. The involvement of a derivatization step could also cause a single compound to have more than one derivatization product, which complicates the characterization of the compound. EC-GC-MS method not only complemented EA-GC-MS by detecting small carboxylic acids and furans, but also overlapped with EA-GC-MS by detecting aromatic compounds. As far as aromatic compounds are concerned, it seems that the results of EA-GC-MS were more reliable due to the reactivity diversity issue occurring in EC-GC-MS.

After analyzing all 80 samples with both methods, a 'compound list' was generated for each method by listing all peaks clearly visible in the chromatograms. By comparing the mass spectra of these peaks with the existing GC-MS compound library in our lab, identities were assigned to some of the peaks. With EA-GC-MS method, in total 129 peaks were detected, among which 44 were identified; while in EC-GC-MS results, there were 114 detected peaks, of which 56 were identified. From the identified compounds, the majority detected by EC-GC-MS method were acids, including carboxylic acids, such as levulinic acid and succinic acid, phenolic acids, like phenylacetic acid and syringic acid, and 18 amino acids (Appendix 1). EA-GC-MS mainly detected phenolic compounds, containing phenolic aldehydes, alcohols and acids (Appendix 1).

To all detected peaks from both identified compounds and unknowns, pseudo-quantities were assigned by integrating their peak areas. To correct sample matrix effect, internal standards were measured in both blank and hydrolysate sample. The peak area difference between blank and hydrolysate sample of the internal standards was calculated as a correction-factor, and was used to correct all the integrated peak areas of the same hydrolysate type. Thus, compound lists based on corrected peak areas were formed for both analytical methods.

Statistical model building

To identify inhibitory compounds in biomass hydrolysates, relationship between hydrolysate fermentability and fermentation sample composition was studied by building statistical models. The models used in this study were partial least square (PLS) and n-way PLS (nPLS), validated by conducting double cross validation (2CV), which was done by leave-one-out in the inner and outer loop [31,35]. The purpose of both models was to point to compounds that are most responsible for a certain fermentation phenotype. This was done by predicting the phenotypes using the data-sets formed through analyzing fermentation samples with the two GC-MS methods.

Lag-phase

Lag-phase is the period before growth takes place in a fermentation process (Eq1), it is mainly influenced by the initial media composition. During lag-phase, the fermenting yeast adapts to the hydrolysate by adjusting its composition, thus some compounds will be degraded or converted [44,45]. Therefore, it is reasonable to describe lag-phase by comparing the composition difference between sample t1 and t2 (model 1, 2 and 3 in Table 1). In addition, the composition of sample t1 and t2 represents the beginning and the end

point of the lag-phase (Table 3), which was also used to build models for predicting lag-phase (model 4, 5 and 6 in Table 1).

In total, six models were constructed for lag-phase (Table 1), of which the data-sets of EA-GC-MS and EC-GC-MS methods were used both separately and combined. This is because the effect of fusing these two data-sets was unknown. The prediction results of the six models are shown in Table 4a. It can be seen that among the six models, only 'model 2' and 'model 5' had a Q^2 value above 0.5, indicating that these two models could be used to predict lag-phase. As shown in Table 1 that the inputs of both 'model 2' and 'model 5' were from the EC-GC-MS data-set, suggesting that the compounds detected by EC-GC-MS method had more influence on lag-phase compared to those measured with EA-GC-MS.

Table 4a Lag-phase prediction results and Q^2 values of the PLS-2CV models shown in Table 1.
Bold: models selected for analyzing the SR of the peaks in the data-sets.

PLS-2CV	lag-phase	prediction					
		model 1	**model 2**	model 3	model 4	**model 5**	**model 6**
Bag-CA	7.5	6.13	**9.28**	7.43	5.79	**6.18**	**4.96**
Bag-DA	6.0	3.36	**5.24**	4.61	6.79	**4.95**	**5.53**
Bag-MA	2.0	5.02	**3.55**	4.27	4.21	**3.08**	**2.90**
BS-CA	7.5	7.14	**7.67**	7.45	7.42	**9.28**	**8.44**
BS-DA	6.5	8.99	**5.29**	8.23	9.49	**6.57**	**7.71**
BS-PAA	3.0	3.61	**4.12**	3.82	3.86	**4.67**	**3.86**
CS-CA	10.5	11.71	**9.44**	10.91	10.32	**10.61**	**11.33**
CS-MA	6.5	-6.10	**6.28**	-1.36	4.40	**6.66**	**6.25**
Oak-CA	1.5	4.25	**4.48**	4.43	6.27	**6.02**	**6.19**
Oak-DA	5.0	4.42	**5.70**	4.74	6.16	**5.94**	**5.81**
Oak-MA	2.5	3.81	**4.00**	3.99	-3.28	**2.00**	**-0.54**
Oak-PAA	3.0	5.39	**4.38**	5.05	3.47	**2.99**	**2.55**
Willow-DA	7.5	6.83	**8.95**	7.61	5.19	**6.70**	**6.68**
Willow-PAA	5.5	3.61	**6.08**	4.36	7.12	**2.19**	**4.51**
WS-CA	9.0	7.93	**8.17**	8.00	6.95	**6.86**	**6.18**
WS-MA	6.5	6.58	**2.01**	3.97	4.38	**7.95**	**7.59**
Q^2		-1.01	**0.54**	0.06	0.05	**0.51**	**0.47**

EA-GC-MS data-set failed to predict lag-phase properly ('model 1' and 'model 4' in Table 4a), but when combined with EC-GC-MS data-set, the prediction improved, resulting in a model with Q^2 value of 0.47 ('model 6' in Table 4a). As the Q^2 value of 'model 6' is very close to 0.5, this model was still selected, together with model 2 and 5, to calculate the selectivity ratios (SR) assigned to each peaks in these data-sets.

For each detected peaks in EC-GC-MS data-set, the SR values of the three models in bold in Table 4a were summed, and ranked based on their SR-sum values; while for each detected peaks in EA-GC-MS data-set, the SR value of 'model 6' were ranked. The top 40 peaks with the highest SR-sum values, 20 from EC-GC-MS data-set and 20 from EA-GC-MS data-set, were considered as the main contributors in predicting lag-phase. Among these 40 peaks, the ones with identity were tested for their effects on the fermenting yeast (Section 'potential inhibitory compound testing', Table 5a). The detailed ranking procedure of lag-phase model SR is shown in Appendix 2-1.

Glu CR and EtOH PR

Different from lag-phase, Glu CR and EtOH PR could be influenced by all five fermentation time points according to their definitions (Eq2 and Eq3). These phenotypes were modeled by the data-sets of the five fermentation samples both individually, using the PLS-2CV model, and collectively, using the nPLS-2CV model.

PLS-2CV modeled Glu CR and EtOH PR with the data-sets of individual fermentation samples, which reveals the influence of these single time points on these two phenotypes. The modeling results show that EC-GC-MS data-sets failed to predict Glu CR and EtOH PR, as the resulting Q^2 values were all negative (data not shown). On the contrary, the EA-GC-MS data-sets of sample t3, t4 and t5 successfully modeled the two phenotypes, as shown in Table 4b, the resulting Q^2 values were above 0.5. This suggests that, different from lag-phase, Glu CR and EtOH PR were relating to the compounds detected with EA-GC-MS method. Moreover, the prediction became meaningful only after time point t2 (Table 4b, $Q^2 > 0$), indicating that Glu CR and EtOH PR were not affected by the initial hydrolysate composition, but influenced by the composition after lag-phase and during growth. This confirms that the data-sets of time point t1 and t2 possess a different structure compared to the other three time points. This structure contains information that could properly describe lag-phase (Table 4a), which ends after time point t2, but failed to predict Glu CR and EtOH PR, which describe a different phase of the fermentation process.

To include the effect of hydrolysate composition change during the fermentation process on Glu CR and EtOH PR, the five time-point samples were also analyzed collectively, with the nPLS-2CV model. Consistent with the PLS-2CV models, the prediction was only valid with EA-GC-MS data-set (Table 4c). Since it was known from PLS-2CV models that data-set of sample t1 gave negative Q^2 values (Table 4b), nPLS-2CV models were also built with the data-set of sample t2 to t5. As shown in Table 4c, the predictions of Glu CR and EtOH PR were improved when sample t1 was excluded from the data-set, indicating that the input of sample t1 data-set was negative.

Table 4b Glu CR and EtOH PR prediction results and Q^2 values of the PLS-2CV models.
Bold: models selected for analyzing the SR of the peaks in the EA-GC-MS data-sets.

PLS-2CV	Glu CR	prediction					EtOH PR	prediction				
		t1	t2	**t3**	**t4**	**t5**		t1	t2	**t3**	**t4**	**t5**
Bag-CA	1.42	2.86	2.30	**2.08**	**1.77**	**1.08**	0.44	1.04	0.92	**0.81**	**0.81**	**0.47**
Bag-DA	3.64	2.62	3.07	**3.00**	**3.01**	**3.56**	1.52	0.93	1.15	**1.18**	**1.18**	**1.39**
Bag-MA	3.84	3.12	3.28	**3.64**	**3.45**	**3.22**	1.58	1.38	1.13	**1.51**	**1.51**	**1.40**
BS-CA	4.57	2.38	3.94	**4.15**	**3.71**	**3.88**	1.73	0.91	1.61	**1.70**	**1.70**	**1.61**
BS-DA	3.63	4.60	3.99	**3.73**	**3.93**	**3.45**	1.42	1.99	1.72	**1.57**	**1.57**	**1.47**
BS-PAA	2.96	2.39	2.80	**2.73**	**2.81**	**2.65**	1.03	1.04	1.20	**1.13**	**1.13**	**1.13**
CS-CA	3.21	2.98	3.51	**3.71**	**3.24**	**3.51**	1.32	1.12	1.31	**1.44**	**1.44**	**1.35**
CS-MA	3.35	4.21	4.52	**4.00**	**4.23**	**4.31**	1.32	1.85	2.03	**1.81**	**1.81**	**1.89**
Oak-CA	0.80	2.88	2.76	**1.77**	**1.94**	**2.00**	0.29	1.03	1.09	**0.57**	**0.57**	**0.66**
Oak-DA	2.41	2.54	3.04	**2.63**	**1.83**	**2.35**	0.98	1.23	1.46	**1.22**	**1.22**	**1.22**
Oak-MA	3.43	2.67	2.70	**3.03**	**2.98**	**3.12**	1.55	1.01	1.04	**1.20**	**1.20**	**1.25**
Oak-PAA	2.73	2.95	3.01	**2.63**	**2.59**	**2.66**	1.12	1.13	1.18	**0.99**	**0.99**	**1.00**
Willow-DA	2.74	2.78	2.61	**2.89**	**3.31**	**2.83**	1.14	1.15	0.99	**1.16**	**1.16**	**1.13**
Willow-PAA	2.45	3.70	3.06	**3.21**	**3.48**	**3.39**	1.05	1.46	1.23	**1.32**	**1.32**	**1.35**
WS-CA	4.63	3.43	3.74	**3.49**	**3.67**	**3.49**	1.87	0.93	1.42	**1.33**	**1.33**	**1.32**
WS-MA	4.01	4.10	3.29	**3.47**	**3.43**	**3.37**	1.92	1.63	1.29	**1.41**	**1.41**	**1.37**
Q^2		-0.16	0.37	**0.68**	**0.57**	**0.61**		-0.32	0.04	**0.54**	**0.50**	**0.56**

Thus, for Glu CR and EtOH PR, three PLS-2CV models and a nPLS-2CV model were selected respectively for analyzing the contributions of the peaks in EA-GC-MS data-set to model predictions (models with 'Bold' in Table 4b and Table 4c). With PLS-2CV models, similar to lag-phase, SR of the peaks were summed and ranked. The top 40 peaks with the highest SR values were considered as the main contributors of PLS-2CV models of either Glu CR or EtOH PR. While with nPLS-2CV models, the regression coefficient ('reg') values were used for ranking. The top 40 peaks with the highest absolute 'reg', 20 with positive values and 20 with negative values, were considered as the main contributor of nPLS-2CV model of either Glu CR or EtOH PR. Among the selected peaks, the ones with identity were tested for their effects on the fermenting yeast (Section 'potential inhibitory compound testing', Table 5a). The detailed ranking and selection procedure of the testing compounds are shown in Appendix 2-2. Interestingly, more than 80% of the compounds suggested by Glu CR and EtOH PR models are identical. This indicates, from a statistical point of view, the correlation between Glu CR and EtOH PR.

Table 4c Glu CR and EtOH PR prediction results and Q2 values of the nPLS-2CV models.
Bold: models selected for analyzing the SR of the peaks in the EA-GC-MS data-sets.

nPLS-2CV	Glu CR	prediction		EtOH PR	prediction	
		t1 - t5	**t2 - t5**		t1 - t5	**t2 - t5**
Bag-CA	1.42	1.55	**1.41**	0.44	1.17	**0.59**
Bag-DA	3.64	2.87	**3.07**	1.52	1.05	**1.16**
Bag-MA	3.84	3.16	**3.25**	1.58	1.27	**1.32**
BS-CA	4.57	3.76	**3.67**	1.73	1.53	**1.51**
BS-DA	3.63	4.02	**3.91**	1.42	1.69	**1.64**
BS-PAA	2.96	2.79	**2.88**	1.03	1.20	**1.23**
CS-CA	3.21	3.49	**3.54**	1.32	1.35	**1.36**
CS-MA	3.35	4.40	**4.37**	1.32	2.00	**2.00**
Oak-CA	0.80	1.96	**1.86**	0.29	0.65	**0.62**
Oak-DA	2.41	2.28	**2.35**	0.98	1.17	**1.08**
Oak-MA	3.43	2.71	**2.86**	1.55	1.05	**1.11**
Oak-PAA	2.73	2.81	**2.78**	1.12	1.08	**1.07**
Willow-DA	2.74	2.83	**2.90**	1.14	1.16	**1.18**
Willow-PAA	2.45	3.68	**3.55**	1.05	1.49	**1.42**
WS-CA	4.63	3.70	**3.64**	1.87	1.40	**1.39**
WS-MA	4.01	3.58	**3.46**	1.92	1.41	**1.37**
Q^2		0.526	**0.580**		0.182	**0.419**

Potential inhibitory compound testing

Through constructing statistical models and analyzing the compounds that contribute the most to the models with valid phenotype predictions, two groups of potential inhibitory compounds were identified. To study the effect of these compounds on the fermenting yeast, *S.cerevisiae* CEN.PK113-7D, growth tests were conducted in mineral medium (MM) with 20 g/l glucose. The potential inhibitory compounds were added individually with the following three concentrations, 0.2, 0.5 and 1.0 g/l, respectively.

It should be noted that these testing concentrations could be much higher compared to that in actual biomass hydrolysates, i.e. less than 0.1 g/l [11,14,46,47]. Due to the synergistic effects present in biomass hydrolysates, the toxicity threshold of a specific compound can be much lower compared to that was tested in synthetic medium. Therefore, though the testing concentrations were higher compared to that in biomass hydrolysates, the testing results are still valuable.

The first group of compounds shown in Table 5a were identified by all three phenotype models, among which, furfural resulted in longer lag-phase at all three concentrations tested, while sorbic acid and syringaldehyde reduced growth rate. Suberic acid exhibited positive effect towards the fermenting yeast, mainly through shortening lag-phase. Since this phenomenon was only observed in MM, but not in YPD, which a much richer medium compared to MM, we reason that the acid was probably used as a nutrient by the yeast. HMF, though known as an important inhibitor in biomass hydrolysates [48-50], only exhibited inhibitory effect at 1.0 g/l on the growth rate of the fermenting yeast (Table 5a). However, HMF seems to elongate lag-phase when tested together with other compounds suggested by the lag-phase models. It can be seen that HMF triggered synergistic effect with levulinic acid, 2-furoic acid and pantoyllacton, respectively, at 0.5 g/l (Table 5b). This maybe the reason why HMF was identified, though little effect was observed when tested individually.

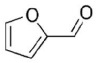
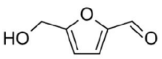
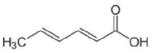
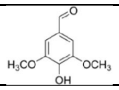
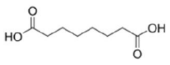
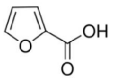
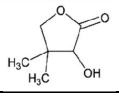
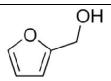
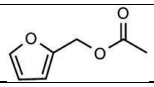
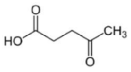
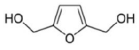
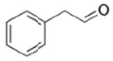
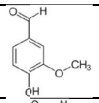
Furfural was identified as a key toxin in biomass hydrolysates [13,51], and consistent with the current study, its main inhibitory effect was elongating lag-phase [52-54]. It was reported earlier that furfural as well as HMF are converted to their alcohol form (furfuryl alcohol and HMF alcohol) and eventually acid form (furoic acid and HMF acid) by the fermenting yeast due to detoxification [44,45]. This was also observed in this study. During lag-phase, the concentration of furfural and HMF reduced, while their alcohols and acids were formed. Since the concentration of furfuryl alcohol and 2-furoic acid is showing an opposite pattern

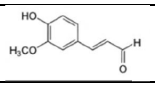
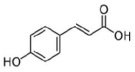
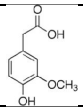
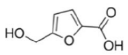
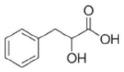
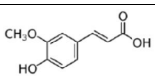
compared to furfural, and HMF alcohol to HMF, as could be expected, these compounds were also identified by analyzing the lag-phase models (Table 5a).

The potential inhibitors identified by Glu CR and EtOH PR models were mainly phenolic compounds (Table 5a). It is known from previous research that the toxic form of a phenolic compound is often the aldehyde, which is converted to its alcohol during the fermentation process due to detoxification [20,22,55]. Therefore, possible conversion of the phenolic alcohols suggested by the models was checked. For those phenolic alcohol compounds with increased concentrations during the fermentation process, the aldehyde forms were used in the growth tests, assuming that the alcohols were the conversion products. These phenolic aldehydes are marked in italic in Table 5a. In agreement with former studies, the compounds exhibited inhibitory effects were mostly aldehydes and acids (Table 5a). The major inhibitory effects were reduced growth rate and lower final OD. Phenylacetaldehyde, vanillin and conifer aldehyde caused growth deficiency at 1.0 g/l (0.5 g/l for phenylacetaldehyde, Table 5a).

Besides the compounds listed in Table 5a, another group suggested by the models were the amino acids, of which concentrations decreased during the fermentation process. This provides the possibility that the depletion of amino acids in hydrolysates worsened the fermentation performance of the fermenting yeast. However, as growth of the fermenting yeast in hydrolysates was not improved when amino acids were added (data not shown), this was apparently not the case. Another explanation would be that the presence of amino acids and possibly other nutrients compensates the inhibitory effects of the inhibitors. This assumption was verified by comparing the inhibitory effects of the compounds listed in Table 5a in MM and YPD medium, which contains abundant peptides and nutrients compared to MM. The inhibitory effects of all the tested compounds alleviated in YPD medium, particularly, the effects underlined in Table 5a were absent in YPD. This observation indicates that the toxicity of inhibitors was culture medium dependent, suggesting that the fermentability of biomass hydrolysates could be improved by adding extra nutrients like yeast extract [56].

Table 5a Inhibitory effects of the compounds suggested by lag-phase, Glu CR and EtOH PR models.

reference medium (MM with 20 g/l glucose)		**LP 7 h**	**GR 0.105**	**OD 1.2**						
compounds	**structure**	**0.2 g/l**			**0.5 g/l**			**1.0 g/l**		
		LP	**GR**	**OD**	**LP**	**GR**	**OD**	**LP**	**GR**	**OD**
compounds identified by all 3 phenotype models										
furfural		10 h	---	---	15 h	---	---	30 h	no growth	
HMF		---	---	---	---	---	---	---	< 20%	---
sorbic acid		---	< 60%	< 80%	---	< 60%	< 80%	no growth		
syringaldehyde		---	---	---	---	---	---	---	< 60%	< 80%
suberic acid								3 h	110 %	110 %
no effect										
2-furoic acid		pantolactone								
compounds identified by lag-phase model										
benzoic acid		---	< 60%	< 80%	---	< 40%	< 80%	---	< 40%	< 60%
no effect										
furfuryl alcohol		2-furanmethanol acetate								
levulinic acid		HMF alcohol								
compounds identified by Glu CR and EtOH PR models										
Phenylacetaldehyde		20 h	< 80%	---	no growth			no growth		
vanillin		---	---	---	11 h	< 80%	---	30 h	no growth	
4-hydroxybenzaldehyde		---	---	---	9.5 h	< 80%	---	11 h	< 60%	< 80%
dihydroxy benzene		---	---	---	---	---	---	---	< 80%	---

conifer aldehyde		---	---	---	23 h	---	---	no growth	
no effect									
p-coumaric acid		homovanillic acid			HMF acid				
3-phenyllactic acid		4-hydroxybenzoic acid			vanillic acid				
compounds identified by Glu CR model									
ferulic acid		---	---	---	---	---	---	---	≤ 40% ≤ 80%
benzaldehyde		---	---	---	9.5 h	< 80%	---	13.5 h	< 80% < 80%
compounds identified by EtOH PR model									
salicylic acid		---	≤ 80%	≤ 80%	---	≤ 60%	≤ 80%	---	≤ 60% ≤ 60%
not tested									
4-hydroxyphenylaldehyde		phloretic acid			5-HMF methyl keton				

LP: lag-phase: time needed to reach 2% (ODmax-ODmin) (h), GR: growth rate: the slope of the linear part of the OD curve (OD/h), OD: final OD. Values with % are relative growth rate and final OD compared to that in reference medium. '---': no effect compared to reference medium.
The compounds indicated in 'italic' were originally identified by their (converted and less toxic) alcohol forms; the compounds indicated in 'bold' were saturated when 1 g/l solutions were prepared at the fermentation temperature, in these cases, besides the saturated solution, a 2- and 5- fold dilution was used, represented in the 0.5 and 0.2 g/l columns, respectively; the 'underlined effects' were NOT observed when tested in YPD medium.

Furfural and HMF are the two most studied inhibitors in biomass hydrolysates, including their inhibitory effects as well as their conversion pathways [13,45,50]. However, the synergistic effects of these two compounds with other potential inhibitors in hydrolysates were seldom tackled. In this study, the joint inhibitory effects of furfural or HMF with one other potential inhibitory compound were tested at 0.5 g/l in MM with 20 g/l glucose, and the compounds gave synergistic effect with either furfural or HMF are listed in Table 5b. It can be seen that HMF caused a notable synergistic effect with levulinic acid, 2-furoic acid, pantoyllacton and syringaldehyde, respectively. These compounds showed no inhibitory effect individually at 0.5 g/l, but when added together with HMF, they elongated lag-phase as well as reduced

growth rate (Table 5b). Compared to HMF, the synergistic effect caused by furfural was minor, since no significant lag-phase increase or growth rate reduction was observed when an extra compound was added (Table 5b).

Table 5b Compounds that caused synergistic effect with furfural or HMF at 0.5 g/l.

reference medium (MM with 20 g/l glucose)			**LP 7 h**	**GR 0.105**	**OD 1.2**				
			compounds				mixture		
			LP	GR	OD		LP	GR	OD
furfural 0.5 g/l 15 h 90% 100%	+	HMF	---	---	---	=	<u>19 h</u>	<u>< 80%</u>	---
		HMF acid	---	---	---		16.5 h	---	---
		salicylic acid	---	< 60%	< 80%		15 h	<u>< 40%</u>	< 60%
		vanillin	11 h	< 80%	---		17 h	< 80%	---
		syringaldehyde	---	---	---		16.5 h	<u>< 80%</u>	---
HMF 0.5 g/l 7 h 100% 100%	+	levulinic acid	---	---	---	=	9 h	<u>< 80%</u>	---
		2-furoic acid	---	---	---		8.5 h	<u>< 80%</u>	---
		pantoyllacton	---	---	---		9 h	<u>< 80%</u>	---
		salicylic acid	---	< 60%	< 80%		8.5 h	<u>< 40%</u>	< 60%
		vanillin	11 h	< 80%	---		11 h	<u>< 60%</u>	---
		syringaldehyde	---	---	---		<u>11 h</u>	---	---

LP: lag-phase: time needed to reach 2% (ODmax-ODmin) (h), GR: growth rate: the slope of the linear part of the OD curve (OD/h), OD: final OD. Values with % are relative growth rate and final OD compared to that in reference medium. '---': no effect compared to reference medium; the 'underlined effects' were relatively significant.

Discussion

Lignocellulosic biomass is a natural resource that has the potential to become the major feedstock for biofuel production [57,58]. To identify inhibitory compounds in biomass hydrolysates, a metabolomics approach was adopted in this study. Compared to targeted methods, no compound pre-selection was made with the metabolomics approach, so that the inhibitor identification was not influenced by prior knowledge [18,26]. The study results show that the metabolomics approach successfully identified compounds that influence the growth of the fermenting yeast, *Saccharomyces cerevisiae* CEN.PK 113-7D. Some compounds elongated lag-phase, like furfural and vanillin, while others reduced growth, such as HMF and benzaldehyde. Interestingly, without pre-selection, compounds that were previously known as inhibitors in biomass hydrolysates were identified in this study. This confirms that metabolomics is a relevant approach in studying the composition and identifying inhibitors of lignocellulosic biomass hydrolysates.

As the analysis targets were potential inhibitory compounds in biomass hydrolysates, which are weak acids, furans and phenolic compounds [15-17], GC-MS was chosen as the analytical tool [18]. To remove sugars in hydrolysate samples, ethyl acetate (EA) extraction or ethyl chloroformate (EC) derivatization was conducted prior to sample analysis. Due to the property difference of these two sample preparation methods, their target compound groups were also different. The EA method had reliable measurement for aromatic compounds, while the EC methods mainly detected carboxylic acids and furans. Remarkably, this difference in analytical method in relation to metabolomics results was also seen during statistical model building, as EA-GC-MS data-sets could predict Glu CR and EtOH PR properly, but failed to model lag-phase on its own, which was validly predicted by EC-GC-MS data. Accordingly, furans were mainly identified to elongate lag-phase, and aromatic compounds were mostly responsible for reduced growth. These results suggest that in a metabolomics study, it is important to have a wide coverage of detectable compounds, so that the chance of overlooking potential target compounds can be reduced [27,59]. And one way of achieving this is to use multiple analytical tools for measuring the same sample.

Furfural and HMF were reported as the two most important inhibitors in biomass hydrolysates, which delay as well as reduce growth [13,45,53,60]. In the growth test of this study, it was found that furfural indeed elongated lag-phase at a concentration of 0.2 g/l, but HMF did not display any inhibitory effect until its concentration reached above 0.5 g/l (Table 5a). However, when tested jointly, HMF enhanced the negative effect of furfural on lag-phase, and reduced growth rate. Besides, when HMF was tested together with other compounds,

which showed no effect individually, like levulinic acid, 2-furoic acid and pantoyllacton, inhibition took place (Table 5b). These observations suggest that HMF probably functions as a co-inhibitor in biomass hydrolysate, for which inhibition is mainly the result of synergistic effects. Furthermore, synergistic effect reduces the threshold concentration for inhibition. For instance, both HMF and syringaldehyde showed toxicity only at 1.0 g/l towards the fermenting yeast, but when tested jointly, the inhibitory effect was present at 0.5 g/l (Table 5b). So it is possible that when multiple inhibitors are present, the toxicity threshold of HMF and syringaldehyde reduce to below 0.1 g/l, which is close to their reported concentration in biomass hydrolysates [11,14,46,47].

Among the compounds identified with Glu CR and EtOH PR models, a group showed no effect in the growth test. Noticeably, this group of compounds is all aromatic acids (Table 5a). Earlier studies demonstrated that aldehyde was the most toxic form of aromatic compounds, the corresponding acids were less, while the alcohol form was the least toxic [20,22,55]. This was confirmed in this study, and was clearly illustrated with vanillin and vanillin acid, of which the acid form had no effect, while the aldehyde form almost abolished growth at 1.0 g/l (Table 5a). Besides the identification of previously reported inhibitors in biomass hydrolysates [9,15,17,55,60,61], two new compounds were found to be toxic, which are sorbic acid and phenylacetaldehyde. As shown in Table 5a, both compounds already showed significant inhibitory effect on growth at 0.2 g/l. The high toxicity towards the fermenting yeast indicates that these two compounds are important inhibitors in biomass hydrolysates. Though not recorded as hydrolysate inhibitors, sorbic acid was described as a preservative weak acid, which disturbs yeast growth through uncoupling mechanism [8,60,62,63], while phenylacetaldehyde was known of having antibiotic activity in maggot therapy [64]. It should be mentioned that the enzyme cocktail used in this study also contains sorbic acid, so the sorbic acid detected in biomass hydrolysates was partially from addition of the hydrolyzing enzyme in most feedstock hydrolysates.

Of the potential inhibitory compounds suggested by the statistical models, about half are unknowns. Some of these compounds are on the very top of the ranking lists, see Appendix 2-1, 2-2. Since most of the known compounds suggested by the models showed inhibitory effect towards the fermenting yeast in growth tests, it is expected that there are also important / novel inhibitors among the unknown compounds. To verify this, identification needs to be conducted for these unknown compounds, which will be the next step in identifying lignocellulosic biomass hydrolysate inhibitors.

Through applying metabolomics approach, the inhibition property of these compounds was connected to their presence in lignocellulosic biomass hydrolysates. To our knowledge, this is the first systematic study on identifying inhibitory compounds in lignocellulosic biomass hydrolysates using a non-targeted approach.

Conclusion

Inhibitory compounds in lignocellulosic biomass hydrolysates were successfully identified through applying an exometabolomics approach. The identification was conducted by relating the fermentability of biomass hydrolysates with their composition using statistical models, (n)PLS-2CV. The non-sugar composition of biomass hydrolysates were analyzed with two GC-MS methods, using ethyl acetate extraction and ethyl chloroformate derivatization to remove sample sugar contents, respectively. Among the identified compounds, besides the known inhibitors, sorbic acid and phenylacetaldehyde were for the first time identified as inhibitors in lignocellulosic biomass hydrolysates.

Acknowledgments

This project was (co) financed by the Netherlands Metabolomics Centre (NMC) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

References

1. Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund M,F, Lidén G, Zacchi G: Bio-ethanol -- the fuel of tomorrow from the residues of today. *Trends Biotechnol.* **2006**, 24, 549-556.
2. González-García S, Moreira MT, Feijoo G: Comparative environmental performance of lignocellulosic ethanol from different feedstocks. *Renew. Sust. Energ. Rev.* **2010**, 14, 2077-2085.
3. Jørgensen H, Kristensen JB, Felby C: Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuel Bioprod. Bior.* **2007**, 1, 119-134.
4. van Dyk JS, Pletschke BI: A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-factors affecting enzymes, conversion and synergy. *Biotechnol. Adv.* **2012**, 30, 1458-1480.
5. Alvira P, Tomás-Pejó E, Ballesteros M, Negro MJ: Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technol.* **2010**, 101, 4851-4861.
6. Gupta R, Khasa YP, Kuhad RC: Evaluation of pretreatment methods in improving the enzymatic saccharification of cellulosic materials. *Carbohyd. Polym.* **2011**, 84, 1103-1109.

7. Sun Y, Cheng J: Hydrolysis of lignocellulosic materials for ethanol production: A review. *Bioresource Technol.* **2002**, 83, 1-11.
8. Palmqvist E, Hahn-Hägerdal B: Fermentation of lignocellulosic hydrolysates. II: Inhibitors and mechanisms of inhibition. *Bioresource Technol.* **2000**, 74, 25-33.
9. Klinke HB, Thomsen AB, Ahring BK: Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biot.* **2004**, 66, 10-26.
10. Panagiotopoulos IA, Bakker RR, de Vrije T, Koukios EG: Effect of pretreatment severity on the conversion of barley straw to fermentable substrates and the release of inhibitory compounds. *Bioresource Technol.* **2011**, 102, 11204-11211.
11. Zha Y, Muilwijk B, Coulier L, Punt PJ: Inhibitory compounds in lignocellulosic biomass hydrolysates during hydrolysate fermentation processes. *J. Bioprocess Biotechniq.* **2012**, 2, doi: 10.4172/2155-9821.1000112.
12. Huang H, Guo X, Li D, Liu M, Wu J, Ren H: Identification of crucial yeast inhibitors in bio-ethanol and improvement of fermentation at high pH and high total solids. *Bioresource Technol.* **2011**, 102, 7486-7493.
13. Heer D, Sauer U: Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. *Microb. Biotechnol.* **2008**, 1, 497-506.
14. Thomsen MH, Thygesen A, Thomsen AB: Identification and characterization of fermentation inhibitors formed during hydrothermal treatment and following SSF of wheat straw. *Appl. Microbiol. Biotechnol.* **2009**, 83, 447-455.
15. Taherzadeh MJ, Karimi K: Fermentation Inhibitors in Ethanol Processes and Different Strategies to Reduce Their Effects. In *Biofuels: Alternative Feedstocks and Conversion Processes.* Pandey A, Larroche C, Ricke SC, Dussap C, Gnansounou E, Eds., Academic Press: USA, 2011, pp. 287-311.
16. Almeida JRM, Modig T, Petersson A, Hähn-Hägerdal B, Lidén G, Gorwa-Grauslund MF: Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *J. Chem. Technol. Biot.* **2007**, 82, 340-349.
17. Jönsson L, Alriksson B, Nilvebrant N: Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol. Biofuels* **2013**, 6, 16, doi:10.1186/1754-6834-6-16.
18. Zha Y, Punt PJ: Exometabolomics approaches in studying the application of lignocellulosic biomass as fermentation feedstock. *Metabolites* **2013**, 3, 119-143.
19. Chen SF, Mowery RA, Castleberry VA, van Walsum G, Chambliss CK: High-performance liquid chromatography method for simultaneous determination of aliphatic acid, aromatic acid and neutral degradation products in biomass pretreatment hydrolysates. *J. Chromatogr. A* **2006**, 1104, 54-61.
20. Klinke HB, Olsson L, Thomsen AB, Ahring BK: Potential inhibitors from wet oxidation of wheat straw and their effect on ethanol production of *Saccharomyces cerevisiae*: Wet oxidation and fermentation by yeast. *Biotechnol. Bioeng.* **2003**, 81, 738-747.
21. Larsson S, Palmqvist E, Hahn-Hägerdal B, Tengborg C, Stenberg K, Zacchi G, Nilvebrant NO: The generation of fermentation inhibitors during dilute acid hydrolysis of softwood. *Enzyme Microb. Technol.* **1999**, 24,151-159.
22. Jönsson LJ, Palmqvist E, Nilvebrant N-O, Hahn-Hägerdal B: Detoxification of wood hydrolysates with laccase and peroxidase from the white-rot fungus *Trametes versicolor*. *Appl. Microbiol. Biotechnol.* **1998**, 49, 691-697.
23. Martinez A, Rodriguez ME, York SW, Preston JF, Ingram LO: Effects of $Ca(OH)_2$ treatments ('overliming') on the composition and toxicity of bagasse hemicellulose hydrolysates. *Biotechnol. Bioeng.* **2000**, 69, 526-536.
24. Roessner U, Bowne J: What is metabolomics all about? *Biotechniques.* **2009**, 46, 363-365.
25. Fiehn O: Metabolomics -- the link between genotypes and phenotypes. *Plant. Mol. Biol.* **2002**, 48,155-171.

26. van der Werf MJ, Jellema RH, Hankemeier T: Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. *J. Ind. Microbiol. Biot.* **2005**, 32, 234-252.
27. Braaksma M, Bijlsma S, Coulier L, Punt PJ, van der Werf MJ: Metabolomics as a tool for target identification in strain improvement: the influence of phenotype definition. *Microbiology* **2011**, 157, 147-159.
28. Krishnan A, Williams LJ, McIntosh AR, Abdi H: Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *Neuroimage* **2011**, 56, 455-475.
29. Bro R: Multiway calibration. Multilinear PLS. *J. Chemometrics* **1996**, 10, 47-61.
30. Smilde AK: Comments on multilinear PLS. *J. Chemometrics* **1997**, 11, 367-377.
31. Westerhuis J, Hoefsloot H, Smit S, Vis D, Smilde A, van Velzen E, van Duijnhoven J, van Dorsten F: Assessment of PLSDA cross validation. *Metabolomics* **2008**, 4, 81-89.
32. Smit S, van Breemen MJ, Hoefsloot HCJ, Smilde AK, Aerts JMFG, de Koster CG: Assessing the statistical validity of proteomics based biomarkers. *Anal. Chim. Acta.* **2007**, 592, 210-217.
33. Zha Y, Slomp R, Groenestijn J, Punt PJ: Preparation and Evaluation of Lignocellulosic Biomass Hydrolysates for Growth by Ethanologenic Yeasts. In *Microbial Metabolic Engineering: Methods and Protocols.* Qiong C Ed., Humana Press: New York, USA, 2012, pp. 245-259.
34. van Hoek P, van Dijken JP, Pronk JT: Effect of specific growth rate on fermentative capacity of baker's yeast. *Appl. Environ. Microbiol.* **1998**, 64, 4226-4233.
35. Andersson CA, Bro R: The n-way toolbox for MATLAB. *Chemometrics Intellig. Lab. Syst.* **2000**, 52, 1-4.
36. Rajalahti T, Arneberg R, Berven FS, Myhr K, Ulvik RJ, Kvalheim OM: Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. *Chemometrics Intellig. Lab. Syst*. **2009**, 95, 35-48.
37. Kvalheim OM: Interpretation of partial least squares regression models by means of target projection and selectivity ratio plots. *J. Chemometrics* **2010**, 24, 496-504.
38. Hendriks MMWB, van Eeuwijk FA, Jellema RH, Westerhuis JA, Reijmers TH, Hoefsloot HCJ, Smilde AK: Data-processing strategies for metabolomics studies. *TrAC Trend Anal. Chem.* **2011**, 30, 1685-1698.
39. van den Berg R, Hoefsloot HCJ, Westerhuis J, Smilde A, van der Werf MJ: Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006**, 7, 142.
40. Brejning J, Jespersen L: Protein expression during lag phase and growth initiation in *Saccharomyces cerevisiae*. *Int. J. Food Microbiol.* **2002**, 75, 27-38.
41. Kuyper M, Winkler AA, van Dijken JP, Pronk JT: Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: a proof of principle. *FEMS Yeast Res.* **2004**, 4, 655-664.
42. Humpula JF, Chundawat SPS, Vismeh R, Jones AD, Balan V, Dale BE: Rapid quantification of major reaction products formed during thermochemical pretreatment of lignocellulosic biomass using GC–MS. *J. Chromatogr. B* **2011**, 879, 1018-1022.
43. Ranatunga TD, Jervis J, Helm RF, McMillan JD, Hatzis C: Identification of inhibitory components toxic toward *Zymomonas mobilis* CP4(pZB5) xylose fermentation. *Appl. Biochem. Biotech.* **1997**, 67, 185-198.
44. Taherzadeh MJ, Gustafsson L, Niklasson C, Lidén G: Conversion of furfural in aerobic and anaerobic batch fermentation of glucose by *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* **1999**, 87, 169-174.
45. Liu ZL, Slininger PJ, Dien BS, Berhow MA, Kurtzman CP, Gorsich SW: Adaptive response of yeasts to furfural and 5-hydroxymethylfurfural and new chemical evidence for HMF conversion to 2,5-bis-hydroxymethylfuran. *J. Ind. Microbiol. Biot.* **2004**, 31, 345-352.

46. Klinke HB, Ahring BK, Schmidt AS, Thomsen AB: Characterization of degradation products from alkaline wet oxidation of wheat straw. *Bioresource Technol.* **2002**, 82, 15-26.
47. Millati R, Niklasson C, Taherzadeh MJ: Effect of pH, time and temperature of overliming on detoxification of dilute-acid hydrolyzates for fermentation by *Saccharomyces cerevisiae*. *Process Biochem.* **2002**, 38, 515-522.
48. Almeida JRM, Modig T, Röder A, Lidén G, Gorwa-Grauslund M: Pichia stipitis xylose reductase helps detoxifying lignocellulosic hydrolysate by reducing 5-hydroxymethyl-furfural (HMF). *Biotechnol. Biofuels* **2008**, 1, 12.
49. Wierckx N, Koopman F, Bandounas L, de Winde JH, Ruijssenaars HJ: Isolation and characterization of *Cupriavidus basilensis* HMF14 for biological removal of inhibitors from lignocellulosic hydrolysate. *Microb. Biotechnol.* **2010**, 3, 336-343.
50. Koopman F, Wierckx N, de Winde JH, Ruijssenaars HJ: Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of Cupriavidus basilensis HMF14. *P. Natl. A. Sci.* **2010**, 107, 4919-4924.
51. Tofighi A, Azin M, Assadi MM, Assadi-rad MHA, Nejadsattari T, Fallahian MR: Inhibitory effect of high concentrations of furfural on industrial strain of *Saccharomyces cerevisiae*. *Int. J. Environ. Res.* **2010**, 4, 137-142.
52. Li Q, Metthew Lam LK, Xun L: Biochemical characterization of ethanol-dependent reduction of furfural by alcohol dehydrogenases. *Biodegradation* **2011**, 22, 1227-1237.
53. Allen S, Clark W, McCaffery JM, Cai Z, Lanctot A, Slininger P, Liu ZL, Gorsich S: Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* **2010**, 3, 2.
54. Ask M, Bettiga M, Mapelli V, Olsson L: The influence of HMF and furfural on redox-balance and energy-state of xylose-utilizing *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* **2013**, 6, 22.
55. Larsson S, Quintana-Sáinz A, Reimann A, Nilvebrant NO, Jönsson LJ: Influence of lignocellulose-derived aromatic compounds on oxygen-limited growth and ethanolic fermentation by *Saccharomyces cerevisiae*. *Appl. Biochem. Biotech.* **2000**, 84-86, 617-632.
56. Jørgensen H: Effect of nutrients on fermentation of pretreated wheat straw at very high dry matter content by *Saccharomyces cerevisiae*. *Appl. Biochem. Biotechnol.* **2009**, 153, 44-57.
57. Mabee WE, McFarlane PN, Saddler JN: Biomass availability for lignocellulosic ethanol production. *Biomass Bioenergy* **2011**, 35, 4519-4529.
58. Zhang YHP: What is vital (and not vital) to advance economically-competitive biofuels production. *Process Biochem.* **2011**, 46, 2091-2110.
59. Dunn WB, Ellis DI: Metabolomics: Current analytical platforms and methodologies. *TrAC Trend Anal. Chem.* **2005**, 24, 285-294.
60. Keating JD, Panganiban C, Mansfield SD: Tolerance and adaptation of ethanologenic yeasts to lignocellulosic inhibitory compounds. *Biotechnol. Bioeng* **2006**, 93, 1196-1206.
61. Palmqvist E, Hahn-Hägerdal B: Fermentation of lignocellulosic hydrolysates. I: inhibition and detoxification. *Bioresource Technol.* **2000**, 74, 17-24.
62. Mira NP, Lourenço AB, Fernandes AR, Becker JD, Sá-Correia I: The RIM101 pathway has a role in *Saccharomyces cerevisiae* adaptive response and resistance to propionic acid and other weak acids. *FEMS Yeast Res.* **2009**, 9, 202-216.
63. Stratford M, Plumridge A, Archer DB: Decarboxylation of sorbic acid by spoilage yeasts is associated with the PAD1 gene. *Appl. Environ. Microb.* **2007**, 73, 6534-6542.
64. Pavillard ER, Wright EA: An antibiotic from maggots. *Nature* **1957**, 180, 916-917.