Selection of potential molecular markers for cheese ripening and quality prediction by NMR spectroscopy

Yangyi Chen[1*], William MacNaughtan[1], Paul Jones[2], Qian Yang[1], Huw Williams[3], Tim Foster[1]

[1]Division of Food, Nutrition and Dietetics, School of Biosciences, the University of Nottingham, Loughborough, UK, LE12 5RD. Author current address Nestle research and development, Beijing, China, 100016

[2] South Caernarfon Creameries Ltd, Pwllheli, UK, LL53 6SB

[3]School of Chemistry, University of Nottingham, UK, NG7 2RD

Yangyi CHEN email address: abigailchen1215@outlook.com

William MacNaughtan email address: sczbim@exmail.nottingham.ac.uk

Paul Jones email address: pjones@sccwales.co.uk

Qian Yang email address: sbzqy@exmail.nottingham.ac.uk

Huw Williams email address: Huw.Williams@nottingham.ac.uk

Tim Foster email address: sbztjf@exmail.nottingham.ac.uk

*Corresponding author. Tel: +447421227241

Abstract

Predicting cheese quality as early as possible after ripening is important for quality control in the cheese industry. The main aim of this study was to investigate potential metabolites for predictive models of Cheddar cheese quality. Metabolites in aqueous extracts of Cheddar cheese were identified by Nuclear Magnetic Resonance. The metabolites were used to measure the kinetics of up to 450 days ripening in Cheddar cheese. The proton ratios of citrulline and arginine relative to the overall proton content of the aqueous extract are the most important indices for assessing the ripening of Cheddar cheese. The ratios of both citrulline and arginine decrease by 59% and 69% respectively after 450 days ripening. In comparison to the premium batch B cheese, batch C which was predicted to attain a lower quality level, had higher serine and β-galactose as well as lower lactic acid levels and also had a less mature sensorial profile. Tyrosine, tyramine and lysine are highly correlated with mature Cheddar cheese sensory attributes. β-Galactose and glycerol are correlated with young Cheddar cheese sensory attributes. These metabolites can be used to predict cheese quality.

Highlights

- Metabolites in aqueous extracts of cheese during ripening were identified
- Metabolites responsible for batch variations in Cheddar cheese were characterised
- The normalised intensity of citrulline and arginine decreased during maturation

- Tyrosine and lysine are correlated with mature Cheddar cheese sensory attributes
- β-Galactose and glycerol are correlated with young cheese sensory attributes

Keywords: Cheddar cheese; NMR; Metabolites; Maturation; Sensory evaluation;

1 Introduction

Many factors can affect the final quality of Cheddar cheese, such as milk quality, production procedures, and choice of starter culture (Mazzei & Piccolo, 2012; Pisano, Scano, Murgia, Cosentino, & Caboni, 2016). However, some Cheddar cheese defects do not develop, or are only observed, when the cheese is aged. This is due to ripening being a complicated process, and the cheese undergoing many chemical/physical and enzymatic modifications (Consonni & Cagliani, 2008). These biochemical transformations can maximise flavour, taste, and appearance of defects in the samples. The cheese grader must therefore continuously evaluate cheese quality throughout the ageing process. However, Cheddar cheese ripening needs to be performed in a constant temperature and humidity environment, which is a time-consuming and costly process (Fox, Cogan, & Guinee, 2017). In Cheddar cheese manufacturing practice, a machine-based predictive grading method is needed to help the manufacturer efficiently manage cheese production, and storage. Fortunately, the maturation process in cheese is sufficiently slow that advanced techniques for a machine based predictive model such as high resolution nuclear magnetic resonance spectroscopy (NMR) can be used. Samples can be analysed remotely and the

results fed back to the commercial producer to facilitate decisions concerning the fate of the cheese. Cheese that has the potential to mature to premium quality needs to be kept longer, whereas other cheese with a lower potential can be sold at early times as low-value young cheese. A grading model using simple initial chemical and physicochemical composition which predicts the quality of Cheddar cheese is still used as an index in Cheddar cheese grading and manufacture. Cheese can be characterised as "premium" or lower quality "graded" cheese. Cheese that fails to meet these levels is referred to as "downgraded" cheese by the professional graders (Giles & Lawrence, 1973). Kraggerud, Næs, and Abrahamsen (2014) reported that sensory characteristics at 280 days were not well forecast from early spectroscopic and basic chemical measurement on cheese (56 days). This indicated the complexity of the cheese ripening process.

In order to show the small deviations from normal cheese composition and the resultant quality defects, a metabolomics approach has often been chosen in the literature. Metabolomics produces a fingerprint at a molecular level that accurately represents all aspects of the food product from sensorial taste and flavour to rheological properties (Pisano *et al.*, 2016). All the bio-transformation processes directly or indirectly affect the final metabolome of cheese (Mazzei & Piccolo, 2012). The metabolomics of Cheddar cheese can provide a framework for correlation between the composition and the prediction of cheese quality during maturation. Water-soluble metabolites in cheese are mainly amino acids, organic acids and carbohydrates. Taste-active compounds contributing to sensations and taste of Cheddar cheese are mostly peptides and free amino

90 acids (Andersen, Ardö, & Bredie, 2010; Lawlor, Delahunty, Wilkinson, & Sheehan,
91 2001; O'Shea, Uniacke-Lowe, & Fox, 1996).
92 Amino acids contribute indirectly to the cheese flavour by acting as precursors
93 for the production of volatile compounds (Consonni & Cagliani, 2008). Free
94 amino acids are hydrolysed from a range of intermediate-sized peptides
95 produced by proteinases and peptidases from the starter lactic acid bacteria. In
96 the cheese matrix, after carbohydrate exhaustion, amino acids are the simplest
97 molecules available for weakly lipolytic bacteria to metabolise to generate
98 adenosine triphosphate (ATP) and so produce compounds that impact flavour
99 (Ganesan & Weimer, 2017).
100 High resolution nuclear magnetic resonance spectroscopy (NMR) is a method for
101 the structural determination and assignment of major metabolites in cheese
102 (Ruyssen *et al.*, 2013). It is a highly reproducible chemical analysis method,
103 offering in a single experiment, an overview of a wide range of compounds
104 present in the food matrix (Piras *et al.*, 2013). Chemometric methods are
105 commonly used in conjunction with NMR to identify patterns among samples
106 from the large amount of NMR data (Mannina, Sobolev, & Viel, 2012). NMR
107 spectroscopy combined with multivariate chemometrics analysis can determine
108 the metabolic profile of intact cheeses such as Mozzarella di Bufala Campana,
109 Parmigiano Reggiano and Emmental and distinguish the geographical origins,
110 ripening and freshness by statistical methods (Consonni & Cagliani, 2008;
111 Mazzei & Piccolo, 2012; Shintu & Caldarelli, 2005, 2006). NMR allows a thorough
112 analysis of components in the solution extracted from the sample. Unknown and
113 unexpected substances can also be identified (Gianferri, Maioli, Delfini, & Brosio,
114 2007).

However, to date, there are no studies which successfully predict quality in Cheddar cheese after ripening. Previous research concerning the prediction of cheese sensory characteristics have insufficient chemical coverage and development trajectories. A substantial body of metabolomic work has been carried out on Parmigiano Reggiano, Mozzarella and Emmental cheeses, however, no similar work has been completed on Cheddar cheese. The assignment of ^{1}H spectra of water-soluble cheese extracts in previous studies has not been particularly accurate mainly due to the strong overlapping peaks in the water-soluble extracts. There is also no work related to correlations between metabolites and sensory variables of cheese. This study has investigated the kinetics of ripening in Cheddar cheese batches which were predicted to produce different quality cheeses based on the Giles and Lawrence quality grading model, combined with a professional cheese grader's grading predictions. Metabolites from cheese with the potential for ripening and quality prediction were explored. Ripening and sensorial related metabolite markers in cheese aqueous extracts were investigated. This study seeks to inform the future Cheddar cheese researcher which molecules will be significant in any model of quality prediction

2 Material and methods

2.1 Cheddar Cheese samples

Six 20 kg blocks of Cheddar were obtained from a commercial Cheddar producer in the UK. These blocks of Cheddar cheese were produced on the same day and production line. They are labelled as batches A, B, C, D, E, and F. All six batches

138 were predictively graded after manufacture based on the Giles and Lawrence
139 quality grading model using four composition attributes; the percentage of salt
140 in moisture (S/M), moisture in the fat free substance (MNFS), fat in the dry
141 matter (FDM) and pH (Giles & Lawrence, 1973). All the grading composition
142 attributes were provided by the Cheddar cheese producer and are available in
143 Supplementary Data Table 1. None of the batches had flavour defects
144 immediately after manufacture, which is a condition for the applicability of the
145 Giles and Lawrence grading model. The predicted Cheddar cheese quality is
146 presented in Supplementary Data Figure 1 indicating that batches B would
147 mature to be premium quality, whereas batches A, C, D, E and F would mature
148 to a lower graded quality as expressed in the Giles and Lawrence model. The
149 lower predicted quality is due to the lower amount of fat in dry mass content
150 and lower pH. After 56 days of maturation, all the batches of Cheddar cheese
151 were further assessed sensorially by an experienced Cheddar cheese quality
152 grader. The grading result from the professional cheese grader stated that
153 batches C and E were likely to result in low-quality Cheddar cheese whereas
154 batch B was likely to mature to premium quality. Other batches required further
155 grading in order to confirm the quality prediction. After 13 months of maturation,
156 the grader rechecked all batches Cheddar cheese. Batch B was further confirmed
157 as premium Cheddar cheese and batch E was confirmed as downgraded low-
158 quality cheese. Batches A, C, D, F were graded as being of acceptable cheese
159 quality. The predictive grading and further grading results were hidden from the
160 researcher until all experiments were completed to avoid bias.
161 All six blocks were cut into 14 pieces and re-vacuum packed in bags, further
162 packaged in cardboard and located in the factory warehouse for maturation at a

163 constant controlled temperature of 8 °C. Aqueous fractions of cheese were
164 extracted and analyses carried out at various stages during the ripening period
165 namely 56, 90, 180, 270, 360 and 450 days. At each time point, one bag of
166 cheese was removed at random for each batch and placed in a 4 °C refrigerator
167 before measurement.

168 ## 2.2 Sample preparation

169 A chloroform/methanol/water extract for each cheese sample (6 batches at 6
170 ripening times, n=36) was made in triplicate (36 ×3=108) based on a modified
171 Bligh and Dyer method (Bligh & Dyer, 1959). A 60-mL cold mixed solution was
172 prepared using chloroform and methanol at a volume ratio of 1:2. Cheddar
173 cheese (20 g) was ground in liquid nitrogen with a pestle and mortar and
174 extracted with the cold mixture solution. The suspension was stirred for 2 mins
175 at 4 °C and transferred to a glass tube. The pestle and mortar were rinsed with
176 20 mL chloroform and the washing solvent combined with the suspension.
177 Distilled water (30 mL) was added to the suspension. After stirring, the
178 suspension was stored in a cold chamber at 4 °C for 40 mins. Phase separation
179 was obtained using a Beckman J2-21 centrifuge with fixed rotor JA-10 at 11700g
180 with the temperature maintained at 4 °C for 30 mins. The supernatant (aqueous
181 phase) was collected and filtered through filter paper (Whatman, grade1) and
182 the aqueous fraction concentrated by vacuum concentration and lyophilisation.
183 The dried sample, containing the hydroalcoholic compounds, was capped and
184 stored at 4 °C (Gianferri *et al.*, 2007).
185 Phosphate buffer (ionic strength 25 mmol/L, pH 6.5, in D_2O) containing 0.1
186 mmol/L 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was added to weighed

187 samples of water-soluble compounds. The buffer was added in a ratio of 0.028
188 mL per mg cheese sample. Sample solution (0.6 mL) was then put into a 5mm
189 diameter NMR tube.

190 ## 2.3 High-resolution NMR measurements

191 NMR experiments were performed on a Bruker 800 MHz Avance III spectrometer
192 using a 5 mm QCI Cryoprobe. The temperature was set at 25 °C (298.15 K).
193 Proton spectra were acquired using 128 scans of 32K points with a spectral width
194 of 13 ppm. The free induction decays (FID) were multiplied by an exponential
195 weighting function with a line broadening of 0.3 Hz before Fourier transformation.
196 After phasing and baseline correction, all spectra were referenced to the signal
197 of the added internal standard reference of DSS (0 ppm). All spectra were
198 processed using Bruker Topspin Software. Two-dimensional experiments: homo-
199 nuclear ^{1}H-^{1}H Total Correlation Spectroscopy (TOCSY) and hetero-nuclear ^{1}H-
200 ^{13}C Hetero-Nuclear Single Quantum Correlation (HSQC) were used for selected
201 samples to identify components. All two-dimensional experiments were
202 performed at 25 °C (298.15 K) on the same facility as the one-dimensional
203 experiment. HSQC experiments were performed using a spectral width of 11.96
204 ppm and 200 ppm in F2 (^{1}H) and F1 (^{13}C), respectively. HSQC spectra were
205 acquired with a time-domain of 1K points (F2) and 256 points (F1) using 8 scans.
206 TOCSY experiments were performed using a spectral width of 15 ppm in both
207 dimensions and 16 scans. Data were compared to the Human Metabolome
208 Database (HMDB: http://www.hmdb.ca/) using literature for further
209 confirmation (Consonni & Cagliani, 2008; Piras *et al.*, 2013). The chemical shifts

210 of carbon given by the HSQC spectrum allowed assignment of the spin systems
211 unambiguously by comparison with literature.

212 ## 2.4 NMR Data Processing

213 To enable the statistical analysis, 108 proton spectra (three replicates per
214 sample) were split into 966 non-overlapping integrated bins of 0.01 ppm width.
215 The spectral range from 4.61 to 4.91 ppm was excluded from the integration to
216 avoid interference from residual water. The spectral range from 3.34 to 3.36
217 ppm was removed from the integration procedure to eliminate variability due to
218 the small amounts of residual methanol. The spectral range due to the DSS
219 internal standard (0.05 to -0.05 ppm) was also excluded. The integrals were
220 normalised to the total area to compensate for the overall concentration
221 differences (Craig, Cloarec, Holmes, Nicholson, & Lindon, 2006).

222 ## 2.5 Sensory analysis

223 Descriptive sensory analysis was conducted on cheese batches A, B, C, D, E at
224 six ripening time points which were the same as the NMR analytical time points.
225 The sensory method and protocol were described in previous work (Chen,
226 MacNaughtan, Jones, Yang, & Foster, 2020). However, a brief outline is given
227 here. The descriptive sensory attributes vocabulary describes appearance,
228 aroma, taste, flavour, texture by touch and in mouth texture. The sensory panel
229 comprised ten assessors, with extensive experience in sensory evaluation.
230 Cheese samples were equilibrated at 18 °C for 5 hours. All the descriptive
231 sensory analysis data for the six-ripening time points are presented here namely
232 after 56, 90, 180, 270, 360, 450 days ripening. We were only able to use five

233 batches (A, B, C, D, E) of cheese for sensory evaluation at these six ripening
234 time points, but the batch F sensory profile was evaluated with the other five
235 batches after 540 days of ripening.

236 2.6 Chemometrics

237 The modulated spectral data matrix (integrals of 6 batch × 6 ripening time point
238 × 3 replicates × 747 bins) was analysed using the Principal Component Analysis
239 (PCA), multivariate statistical analysis method embedded in Unscrambler X
240 (CAMO ASA, Trondheim, Norway). All the values of the integrals were mean
241 centre corrected and weighted by dividing by the standard deviation. Some
242 sample replicates were discarded as outliers because they lay outside the ellipse
243 based on the Hotelling' T2 (multivariate t-statistic) corresponding to a 95%
244 confidence limit. A second PCA was performed on the remaining objects. One
245 further PCA was also recalculated using variables whose correlation loading was
246 higher than 70 %.
247 The quantitative descriptive sensory data for the six ripening time points was
248 analysed firstly by analysis of variance (ANOVA) and all the significantly different
249 sensory attributes were examined by PCA in XLSTAT (Addinsoft, France). For
250 modelling the correlation between the metabolite aqueous extracts and the
251 sensory data, Partial Least Squares (PLS) regression was used in Unscrambler X
252 (CAMO ASA, Trondheim, Norway). The X data matrix contained aqueous extract
253 metabolite integral data. The Y matrix contained the results of sensory mean
254 scores for all significantly different cheese attributes. To simplify the PLS analysis,
255 only the four highest associated spectral bins with each sensorial attribute were
256 chosen.

257 One-way Analysis of Variance (ANOVA) was performed on the normalised
258 intensity of citrulline and arginine of Cheddar cheese during ripening and on
259 serine and β-galactose among batches variation. These analyses were followed
260 by Tukey' s honestly significant difference (HSD) post-hoc test. The ANOVA tests
261 were analysed by XLSTAT with a significance level alpha=0.05 (Addinsoft,
262 France).
263

264 # 3 Results and discussion

265 ## 3.1 Metabolite assignment and identification

266 In Figure 1 (a) and (c), 2 spectral regions of a 2D HSQC spectrum of a water-
267 soluble extract of Cheddar cheese are shown. A representative proton spectrum
268 of one aqueous extract of vintage Cheddar cheese is presented in Figures 1(b)
269 and (d). The ^{1}H NMR spectra of the Cheddar cheese aqueous extracts were
270 assigned from one-dimensional and two-dimensional experiments including
271 HSQC (Figure 1) and TOCSY (Data not shown). The details of the assignments
272 were shown in Table 1. In total, four organic acids, nineteen amino acids, two
273 sugars, one amine and glycerol, were assigned. Seven resonances involved in
274 maturation were labelled consistently as unknowns A, B, D, E, F, G, and H.
275 Spectra were dominated by free amino acids in combination with small quantities
276 of organic acids and carbohydrates. As expected, the ^{1}H NMR spectrum revealed
277 the predominance of the lactic acid resonance signals, but part of the peak was
278 overlapped with a threonine methyl group (Ruyssen *et al.*, 2013).

279 Other compounds which were detected in Fiore Sardo cheese by Piras *et al.*
280 (2013) but which were not observed in these Cheddar cheese spectra were citric
281 acid, succinic acid and glucose. This could be due to the sensitivity limitation of
282 NMR and the abundance of other compounds. There were some compounds
283 which were observed and confirmed but had no distinct non-overlapping bin
284 integrals in proton spectrum such as alanine and pyruvic acid.
285

286 ### 3.2 Summary of ripening and batch variation using chemometrics

287 The metabolic trajectories for each batch of cheese were determined by taking
288 the mean position for the ^{1}H spectrum in the PCA score plot (Figure 2). The first
289 two principal components of PCA explained 78% of the total variance. From the
290 plot, a separation of samples based on ripening time exists in the direction of
291 the PC1 axis. Batch variation was mainly but not exclusively represented in the
292 PC2 axis direction. The full PC1 and PC2 loading plots for Figure 2 are presented
293 in Supplementary Figure 2. The PCA plot showed the general trajectories for
294 water-soluble metabolite development during cheese ripening. Cheese ripening
295 trajectories for two batches of cheese were highlighted, namely premium cheese
296 (batch B), and a "downgraded" cheese sample (batch C) (Figure 2). As ripening
297 proceeded, the cheese sample developed from a positive to negative score along
298 PC1, which was correlated with the changes in metabolites. Batches C and D
299 were initially distinct from the other batches. As ripening progressed batch D
300 gradually merged with the general ripening group, whereas batch C remained
301 on a different trajectory.

All the batch C samples had a positive PC2 value, during maturation. Batch B sample ripening time points were in the lowest score region along the PC2 axis. Batch E, a "downgraded" sample, behaved differently from other batches exhibiting an earlier development at 180 days ripening. The differences between batch E and other samples can be seen along the PC1 axis.

In Figure 2, at the early stage of ripening, all the batches of cheese were grouped loosely together. Sample C approached the general group of Cheddar cheese samples at a late stage of ripening. A similar phenomenon was reported in that at an early stage of maturation, the rating of Cheddar cheese flavour and mouth coating character were associated with the composition of cheese and the associations weakened as the cheese matured (Muir, Hunter, Banks, & Horne, 1995). Piras *et al.* (2013) also reported that Fiore Sardo cheese aqueous extracts that are treated with different adjunct bacterial species became more similar with ripening.

3.2.1 Loading coefficients and the discrimination of ripening and batch variation

NMR spectra give a comprehensive picture of the aqueous fraction composition and enable the discrimination between samples. The significant loading coefficients for the chemical shift bins from the PCA plot in Figure 2 are presented in Figure 3. This provides an indication of the extent to which metabolites have changed during ripening and enables discrimination between batches. Not all the bins associated with the same compounds are significant. This is due to overlap and the presence of different components in same bins.

From Figure 3 arginine, asparagine, citrulline, glycerol, lactic acid, serine, unknown G, α-galactose, β-galactose, glycine and glycerol (bin at 3.55-3.56

326 ppm), lactic acid and proline (bins at 4.12-4.14 ppm) and lactic acid and
327 threonine (bins at 1.32-1.34 ppm) all have positive contributions to PC1. As
328 ripening proceeded, all the compounds that positively contributed to PC1
329 decreased in the Cheddar cheese aqueous extract ratio. Mature Cheddar cheese
330 had a greater content of isoleucine, leucine, lysine, methionine, pyroglutamic
331 acid, phenylalanine, threonine, tyramine, and valine in the Cheddar cheese
332 aqueous extract ratio. Some bins associated with PC1 are not discussed since
333 they are associated with multiple metabolites
334 From Figure 3, in terms of batch variation, batch C was characterized by a
335 smaller amount of lactic acid (bin at 1.34-1.36 ppm) and related bins (4.12-4.14
336 ppm and 1.32-1.34 ppm), asparagine (Asn, bin 2.95-2.97 ppm), arginine (Arg,
337 bin 3.23-3.26 ppm), leucine (Leu, bin at 1.73-1.74 ppm) and tyrosine (Tyr, bin
338 at 3.17-3.18 ppm) and unknowns G,A and F.
339 The larger the absolute number of coefficient loading, the more important the
340 contribution of the variable to the explanation of the variance. Citrulline (bin
341 from 1.59-1.60 ppm) and arginine (bin from 3.24-3.25ppm) have the most
342 significant contributions to PC 1, see Figure 3, which characterises the Cheddar
343 cheese ripening process.
344 With regard to the batch variation related to PC2, serine (bin at 3.83-3.84 ppm),
345 tyrosine and unknown (bin at 3.17-3.18 ppm) and β-Galactose (bin at 3.46-3.47
346 ppm) are the most important metabolites. They have the largest contributions
347 to the loadings. As chemical shift region tyrosine and unknown (bin at 3.17-3.18
348 ppm) is not deconvoluted, only serine (bin at 3.83-3.84 ppm) and β-galactose
349 (bin at 3.46-3.47 ppm) highlighted as triangle symbol in Figure 3.
350

351 ### 3.3 The Evolution of metabolites during maturation

352 Citrulline and arginine were the most critical metabolites for monitoring ripening.
353 The individual batch developments of citrulline and arginine with aging are
354 presented in Figure 4. As batches A, D, F had similar metabolite development
355 with ageing and the overall batch variance of the group, batches A, D, F were
356 grouped together and were compared with individual batches B, C, E in Figure
357 4. For all batches of cheese, the normalised intensity of citrulline (bin from 1.59-
358 1.60 ppm) and arginine (bin from 3.24-3.25ppm) decreased by 59% and 69%,
359 respectively.

360 Arginine is closely correlated with maturation, at least in part, because starter
361 lactococci are capable of shifting metabolism from sugar to arginine, which is
362 then the first amino acid metabolised for energy (Ganesan & Weimer, 2017).
363 Arginine is hydrolysed to NH_3 and citrulline with the production of ATP (Fox,
364 McSweeney, Guinee, & Cogan, 2000). Arginine and citrulline consequently follow
365 the same trend during maturation.

366 Asparagine is one of the amino acids which also characterizes the ripening
367 process. The ratio of asparagine to the overall metabolite content decreased
368 during ripening. The metabolism of asparagine in lactic acid bacteria produces
369 acetic acid and propionic acid via aspartate and provides oxaloacetate for the
370 production of diacetyl and acetaldehyde (Ganesan, Seefeldt, Koka, Dias, &
371 Weimer, 2004; Ganesan & Weimer, 2017).

372 Conversely, during ripening, the methionine ratio in the aqueous extracts
373 increased. The increase in methionine during ageing was accompanied by an

374 increase in the absolute concentration of its precursor amino acid serine. In most
375 batches the serine ratio reached a limit (Stuart, Chou, & Weimer, 1999).
376 As ripening proceeded, the lactic acid ratio in the aqueous extract decreased.
377 Previous studies showed a decline in lactic acid during ripening (Ganesan &
378 Weimer, 2017; Piras *et al.*, 2013). Lactic acid produced by starter lactic acid
379 bacteria was consumed by the nonstarter microbiota and by the indigenous
380 cheese microbiota (Eliskases-Lechner, Ginzinger, Rohm, & Tschager, 1999). The
381 reduction in lactic acid ratio in our study was probably due to the decrease of
382 lactic acid but could have been due to the increase in the content of other water-
383 soluble metabolites.
384 The galactose ratio in the aqueous extract, including α and β forms decreased
385 during ripening. Galactose is a constituent monosaccharide of lactose whereas
386 neither α or β forms of lactose were detected in spectra even at the earliest time
387 of 56 days, which was consistent with the observation in mozzarella and Fiore
388 Sardo cheese (Mazzei & Piccolo, 2012; Piras *et al.*, 2013). The rapid decrease of
389 the carbohydrate content during the ripening process was attributed to the
390 metabolism of monosaccharides by homofermentative starter lactic acid bacteria
391 (Piras *et al.*, 2013).
392 The glycerol ratio of aqueous extracts decreased during maturation. Enzymatic
393 hydrolysis of triglycerides produces fatty acids and glycerol. Glycerol can enter
394 the glycolysis pathway as a carbon source for the growth of lactic acid bacteria
395 (Hatti-Kaul, Chen, Dishisha, & Enshasy, 2018; McSweeney & Sousa, 2000). In
396 the present work, the decreased glycerol ratio in the aqueous extracts could
397 have been due to the extensive depletion of glycerol when used as a carbon
398 source.

399 3.4 Metabolites responsible for batch variation

400 Based on the Giles and Lawrence grading model, batches A, C, D E and F were
401 predicted to be "graded" quality Cheddar cheese with defects in various
402 characteristics. Aqueous extracts of Cheddar cheese show that batches C and E
403 are different from the other batches. As demonstrated in the present work, the
404 batch variation among different predictive quality Cheddar cheese from the same
405 dairy can be characterized by NMR analysis of aqueous extracts. Batch variations
406 are crucial to grading in the cheese manufacturing industry.
407 Serine and β-galactose are the most critical metabolites for monitoring batch
408 variation. The individual batch developments of serine and β-galactose with
409 aging are presented in Figure 5.
410 The serine ratio was higher in batch C predicted low-quality cheese and lower in
411 batch B premium cheese. The higher the serine content of the aqueous extract,
412 the more methionine biosynthesis will occur and therefore more sulphur volatile
413 compounds will be generated. Sulphur volatile compounds are neither desirable
414 nor typical for this type of Cheddar cheese.
415 Galactose level was lower in the aqueous extracts of premium batch B and higher
416 in the downgraded batch C. The decrease of residual sugar content observed
417 after 15 days of ripening is typical of a secondary fermentation due to the activity
418 of non-starter lactic acid bacteria (Piras *et al.*, 2013). This indicated that lactose
419 metabolism in batch C which produced lactic acid was incomplete and probably
420 resulted in an undesirable secondary flora.
421 Batch C has less lactic acid than other batches. The production of lactic acid from
422 residual lactose depends on the percentage of salt in moisture S/M (Shakeel Ur,

423 Waldron, & Fox, 2004; Upreti, McKay, & Metzger, 2006). The percentage of salt
424 in moisture content of batch C is 5.86%, which is 12.8% higher than the average
425 of all batches. Cheddar cheese with a high salt concentration had lower levels of
426 lactic acid compared with other cheeses (Guinee, Kilcawley, & Beresford, 2008;
427 Møller, Rattray, Bredie, Høier, & Ardö, 2013).

428 ### 3.5 Sensory profile for batches of Cheddar cheese and correlation with metabolites

429 The PCA plot of sensory attributes is shown on Figure 6(a). PC1 mainly separated
430 sensory attributes from young to mature cheese. As expected, 56 days ripened
431 Cheddar cheese was associated with rubbery, buttery, dairy odour, oily and
432 yellow sensorial attributes. Mature Cheddar cheese had higher sour, tangy,
433 umami astringent attributes and a lingering aftertaste as well as being sweaty
434 and crumbly. However, batch C matured noticeably slower than other batches.
435 Most of the batch C sample points were grouped into the previous ripening time
436 point for the other samples. Only for the last ripening time point, namely ripening
437 after 450 days, are all the cheese samples grouped together. Batch C had
438 significantly less overall flavour intensity compared with the others until the 450
439 day ripening time point. At comparable ripening times batch C was lower in all
440 the mature cheese sensory attributes compared with the other samples. This
441 was probably due to the low lactic acid environment in batch C which caused the
442 lactobacilli to develop later and generate a lower overall rate of proteolysis
443 (Moynihan *et al.*, 2016). Moreover, lactic acid was recognised as a key taste
444 driver in mature Cheddar (Møller *et al.*, 2013) and the overall flavour intensity
445 was associated with the extent of protein breakdown (Banks, Brechany, Christie,
446 Hunter, & Muir, 1992). This is probably the reason for the bland flavour of the

447 batch C sample. The metabolite ratio in aqueous fractions measured by NMR
448 successfully distinguished cheese that matured slower than other cheeses even
449 at early stage of ripening.times.
450 Batch E samples matured faster than other cheeses (Figure 6(a)). The panel also
451 mentioned that batch E had strong flavour defects. Unfortunately, the metabolite
452 ratio in aqueous fractions does not distinguish between batch E and other
453 batches.
454 The PLS correlation loading plot showed the correlation between sensory
455 attributes and metabolites. The younger cheese sensorial attributes were
456 associated with β-galactose, glycerol and some unknown compounds (Figure 6
457 (b)). In Figure 7, the normalised glycerol and lysine intensity of batch C lagged
458 behind the general ripening trend until the later stage of ripening. Tyrosine,
459 tyramine and lysine were highly correlated with more mature sensory attributes.
460 Lysine was catabolized to fatty acids which were associated with a mature
461 cheese flavour (Ganesan & Weimer, 2017). The amino carbon in tyrosine was
462 trans-aminated to produce aromatic pyruvate which was further reduced to an
463 aromatic acid, aldehyde or alcohol (Ganesan & Weimer, 2017). All these flavour
464 compounds were found in mature Cheddar cheese. Tyramine is also an amine
465 derived from tyrosine.
466 Carbon sources such as glycerol and galactose were more likely to be present in
467 less mature Cheddar cheese, as during ripening, they were depleted. Some
468 metabolites such as serine were highly significant in discriminating between
469 batches but are not highly correlated with sensory attributes because they are
470 not directly or indirectly taste active in the Cheddar cheese.

4 Conclusions

The present results showed that high field NMR spectroscopy can characterise the metabolic profile of Cheddar cheese during maturation and statistically distinguish between Cheddar cheese destined to attain different quality levels. The ratio of citrulline and the ratio of arginine in aqueous extracts were the most important indices for assessing the ripening of Cheddar cheese. These metabolites decreased in the aqueous extracts during maturation. Tyrosine, tyramine and lysine are highly correlated with more mature Cheddar cheese sensory attributes whereas β-galactose and glycerol are correlated with young sensory attributes. The metabolite profile in aqueous extracts of Cheddar cheese only significantly discriminates the low-quality batch C from other batches. Batch E, which from the sensory profile is another lower quality Cheddar cheese, requires further physiochemical exploration. This work indicated which metabolites can potentially be used to predict cheese quality, however, only six batches of Cheddar cheese were investigated, and additional work is necessary to confirm the present findings.

Acknowledgement

The authors gratefully acknowledge SI (Systems integration) for the input of scientific and commercial knowledge. We thank Innovate UK (grant number: 102329) for financial support for this project.

Ethical statement

The sensory ethics in this project have been approved by the University of Nottingham, School of Biosciences Research Ethics Committee. The approval code is SBREC160127A

Conflict of interest

The authors have no conflict of interest to declare

Funding Sources

This work was supported by Innovate UK (grant number 102329)

References

Andersen, L. T., Ardö, Y., & Bredie, W. L. P. (2010). Study of taste-active compounds in the water-soluble extract of mature Cheddar cheese. *International Dairy Journal, 20*(8), 528-536. https://doi.org/https://doi.org/10.1016/j.idairyj.2010.02.009.

Banks, J. M., Brechany, E. Y., Christie, W. W., Hunter, E. A., & Muir, D. D. (1992). Volatile components in steam distillates of Cheddar cheese as indicator indices of cheese maturity, flavour and odour. *Food Research International, 25*(5), 365-373. https://doi.org/10.1016/0963-9969(92)90111-H.

Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology, 37*(8), 911-917. https://doi.org/10.1139/o59-099.

Chen, Y., MacNaughtan, W., Jones, P., Yang, Q., & Foster, T. (2020). The state of water and fat during the maturation of Cheddar cheese. *Food Chemistry, 303*, 125390. https://doi.org/https://doi.org/10.1016/j.foodchem.2019.125390.

Consonni, R., & Cagliani, L. R. (2008). Ripening and geographical characterization of Parmigiano Reggiano cheese by 1H NMR spectroscopy. *Talanta, 76*(1), 200-205. https://doi.org/10.1016/j.talanta.2008.02.022.

Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry, 78*(7), 2262-2267. https://doi.org/10.1021/ac0519312.

Eliskases-Lechner, F., Ginzinger, W., Rohm, H., & Tschager, E. (1999). Raw milk flora affects composition and quality of Bergkase. 1. Microbiology and fermentation compounds. *Lait, 79*(4), 385-396. https://doi.org/10.1051/lait:1999432.

Fox, P. F., Cogan, T. M., & Guinee, T. P. (2017). Factors that affect the quality of cheese. In P. L. H. McSweeney, P. F. Fox, P. D. Cotter & D. W. Everett (Eds.), *Cheese : Chemistry, Physics and Microbiology* 4th ed., (pp. 637). London: Academic Press.

Fox, P. F., McSweeney, P. L. H., Guinee, T. P., & Cogan, T. M. (2000). Starter Cultures. In P. F. Fox, P L. H. McSweeney, T. P. Guinee & T. M. Cogan (Eds.), *Fundamentals of cheese science* Gaithersburg, Maryland.: Aspen Publishers.

Ganesan, B., Seefeldt, K., Koka, R. C., Dias, B., & Weimer, B. C. (2004). Monocarboxylic acid production by lactococci and lactobacilli. *International Dairy Journal, 14*(3), 237-246. https://doi.org/10.1016/j.idairyj.2003.07.004.

Ganesan, B., & Weimer, B. C. (2017). Amino acid catabolism and its relationship to cheese flavour outcomes In P. L. H. McSweeney, P. F. Fox, P. D. Cotter & D. W. Everett (Eds.), *Cheese : chemistry, physics and microbiology* (pp. 483-515). London: Academic Press.

Gianferri, R., Maioli, M., Delfini, M., & Brosio, E. (2007). A low-resolution and high-resolution nuclear magnetic resonance integrated approach to investigate the physical structure and metabolic profile of Mozzarella di Bufala Campana cheese. *International Dairy Journal, 17*(2), 167-176. https://doi.org/10.1016/j.idairyj.2006.02.006.

Giles, J., & Lawrence, R. C. (1973). The assessment of cheddar cheese quality by compositional analysis. *New Zealand Journal of Dairy Science and Technology, 8*, 148-151.

Guinee, T. P., Kilcawley, K. N., & Beresford, T. P. (2008). How variable are retail vintage brands of Cheddar cheese in composition and biochemistry? *Australian Journal of Dairy Technology, 63*(2), 50-60. https://www.scopus.com/inward/record.uri?eid=2-s2.0-54849441226&partnerID=40&md5=f26dcdd9d23ef09fe011ac2da6264dc5.

Hatti-Kaul, R., Chen, L., Dishisha, T., & Enshasy, H. E. (2018). Lactic acid bacteria: from starter cultures to producers of chemicals. *FEMS Microbiology Letters, 365*(20). https://doi.org/10.1093/femsle/fny213.

Kraggerud, H., Næs, T., & Abrahamsen, R. K. (2014). Prediction of sensory quality of cheese during ripening from chemical and spectroscopy measurements. *International Dairy Journal, 34*(1), 6-18. https://doi.org/10.1016/j.idairyj.2013.07.008.

Lawlor, J. B., Delahunty, C. M., Wilkinson, M. G., & Sheehan, J. (2001). Relationships between the sensory characteristics, neutral volatile composition and gross composition of ten cheese varieties. *Lait, 81*(4), 487-507. https://doi.org/10.1051/lait:2001147.

Mannina, L., Sobolev, A. P., & Viel, S. (2012). Liquid state 1H high field NMR in food analysis. *Progress in Nuclear Magnetic Resonance Spectroscopy, 66*, 1-39. https://doi.org/https://doi.org/10.1016/j.pnmrs.2012.02.001.

Mazzei, P., & Piccolo, A. (2012). 1H HRMAS-NMR metabolomic to assess quality and traceability of mozzarella cheese from Campania buffalo milk. *Food Chemistry, 132*(3), 1620-1627. https://doi.org/https://doi.org/10.1016/j.foodchem.2011.11.142.

McSweeney, P. L. H., & Sousa, M. J. (2000). Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. *Lait, 80*(3), 293-324. https://doi.org/10.1051/lait:2000127.

Møller, K. K., Rattray, F. P., Bredie, W. L. P., Høier, E., & Ardö, Y. (2013). Physicochemical and sensory characterization of Cheddar cheese with variable NaCl levels and equal moisture content. *Journal of Dairy Science, 96*(4), 1953-1971. https://doi.org/https://doi.org/10.3168/jds.2012-5524.

Moynihan, A. C., Govindasamy-Lucey, S., Molitor, M., Jaeggi, J. J., Johnson, M. E., McSweeney, P. L. H., & Lucey, J. A. (2016). Effect of standardizing the lactose content of cheesemilk on the properties of low-moisture, part-skim Mozzarella cheese. *Journal of Dairy Science, 99*(10), 7791-7802. https://doi.org/https://doi.org/10.3168/jds.2016-11215.

Muir, D. D., Hunter, E. A., Banks, J. M., & Horne, D. S. (1995). Sensory properties of Cheddar cheese: changes during maturation. *Food Research International, 28*(6), 561-568. https://doi.org/https://doi.org/10.1016/0963-9969(95)00039-9.

O'Shea, B. A., Uniacke-Lowe, T., & Fox, P. F. (1996). Objective assessment of cheddar cheese quality. *International Dairy Journal, 6*(11), 1135-1147. https://doi.org/https://doi.org/10.1016/0958-6946(95)00065-8.

Piras, C., Cesare Marincola, F., Savorani, F., Engelsen, S. B., Cosentino, S., Viale, S., & Pisano, M. B. (2013). A NMR metabolomics study of the ripening process of the Fiore Sardo cheese produced with autochthonous adjunct cultures. *Food Chemistry, 141*(3), 2137-2147. https://doi.org/https://doi.org/10.1016/j.foodchem.2013.04.108.

Pisano, M. B., Scano, P., Murgia, A., Cosentino, S., & Caboni, P. (2016). Metabolomics and microbiological profile of Italian mozzarella cheese produced with buffalo and cow milk. *Food Chemistry, 192*, 618-624. https://doi.org/10.1016/j.foodchem.2015.07.061.

Ruyssen, T., Janssens, M., Van Gasse, B., Van Laere, D., Van der Eecken, N., De Meerleer, M., . . . De Vuyst, L. (2013). Characterisation of Gouda cheeses based on sensory, analytical and high-field 1H nuclear magnetic resonance spectroscopy determinations: Effect of adjunct cultures and brine composition on sodium-reduced Gouda cheese. *International Dairy Journal, 33*(2), 142-152. https://doi.org/https://doi.org/10.1016/j.idairyj.2013.04.009.

Shakeel Ur, R., Waldron, D., & Fox, P. F. (2004). Effect of modifying lactose concentration in cheese curd on proteolysis and in quality of Cheddar cheese. *International Dairy Journal, 14*(7), 591-597. https://doi.org/https://doi.org/10.1016/j.idairyj.2003.11.008.

Shintu, L., & Caldarelli, S. (2005). High-Resolution MAS NMR and Chemometrics: Characterization of the Ripening of Parmigiano Reggiano Cheese. *Journal Of Agricultural And Food Chemistry, 53*(10), 4026-4031. https://doi.org/10.1021/jf048141y.

Shintu, L., & Caldarelli, S. (2006). Toward the Determination of the Geographical Origin of Emmental(er) Cheese via High Resolution MAS NMR: A Preliminary Investigation. *Journal Of Agricultural And Food Chemistry, 54*(12), 4148-4154. https://doi.org/10.1021/jf060532k.

Stuart, M. R., Chou, L. S., & Weimer, B. C. (1999). Influence of carbohydrate starvation and arginine on culturability and amino acid utilization of lactococcus lactis subsp. lactis. *Applied and Environmental Microbiology, 65*(2), 665. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC91077/pdf/am000665.pdf.

Upreti, P., McKay, L. L., & Metzger, L. E. (2006). Influence of calcium and phosphorus, lactose, and salt-to-moisture ratio on cheddar cheese quality: changes in residual sugars and water-soluble organic acids during ripening. *Journal of Dairy Science, 89*(2), 429-443. https://doi.org/https://doi.org/10.3168/jds.S0022-0302(06)72107-5.

Figure captions

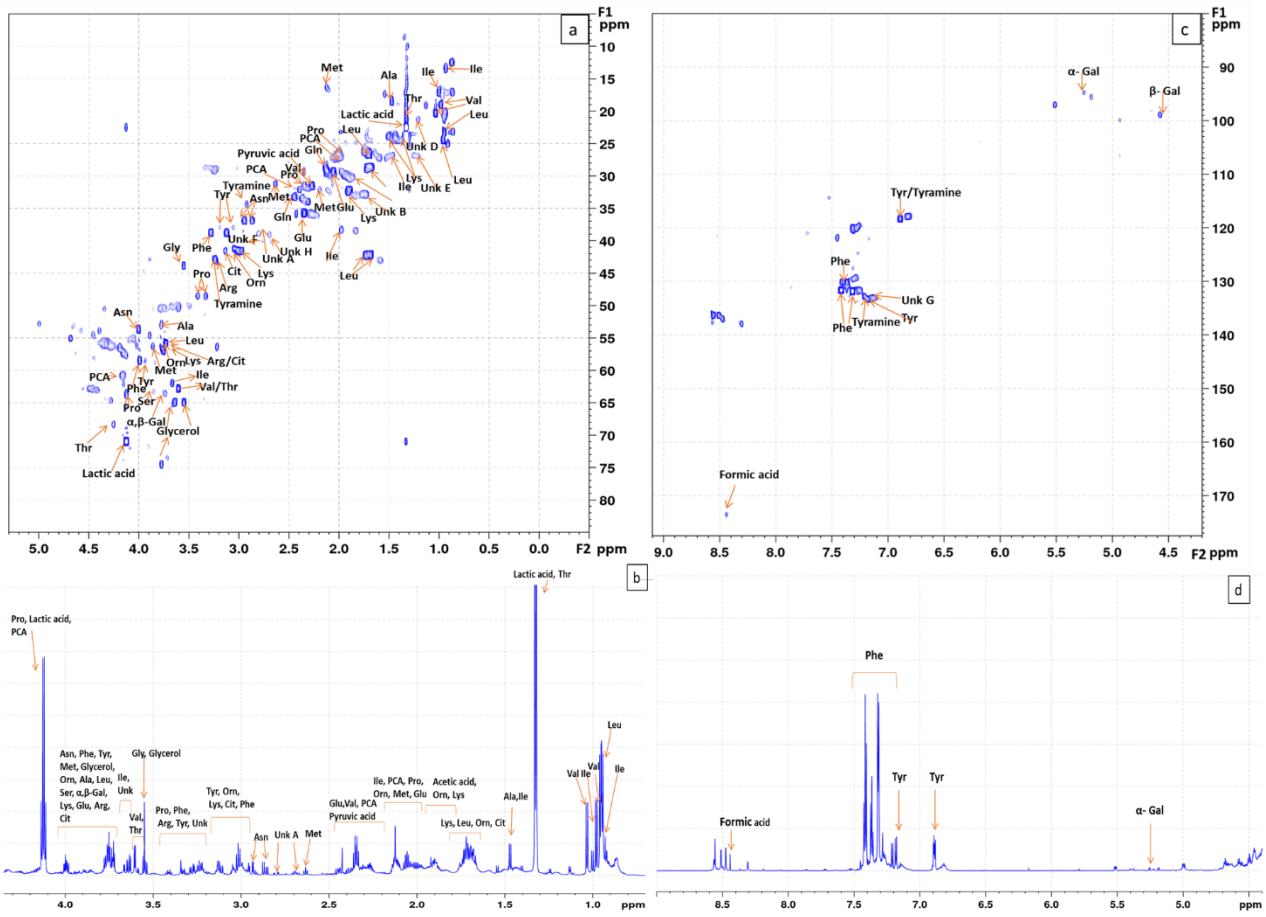


Figure 1 *Representative 2D HSQC (a,c) and ¹H NMR (b,d) spectra of water-soluble extracts of vintage Cheddar cheese using DSS as a reference, showing aliphatic region(a, b) and aromatic (c, d). A range of the most intense peaks are labelled. Some peak labels have been omitted for clarity*

615 **Table 1** NMR assignment of peaks measured in water soluble extracts of Cheddar cheese. ^{1}H
616 and ^{13}C chemical shifts and multiplicity are reported based on TOCSY, HSQC experiments and
617 literature values.

Compound	Multiplicity	1H shift(ppm)	13C shift(ppm)	Assignment
Acetic acid	s	1.928	25.26	α-CH_3
Lactic acid	d	1.325	22.62	β-CH_3
	q	4.124	71.11	α-CH
Pyruvic acid	s	2.35	29.46	CH_3
Formic acid	s	8.44	173.48	$HCOO^-$
Valine	d	0.980	19.17	γ'-CH_3
	d	1.033	20.48	γ-CH_3
	m	2.268	31.66	β-CH
	d	3.603	62.91	α-CH
Leucine	d	0.945	23.39	δ'-CH_3
	d	0.956	24.63	δ-CH_3
	unresolved	1.703	26.74	γ-CH
	unresolved	1.684	42.38	β'-CH_2
	unresolved	1.731	42.38	β-CH_2
	unresolved	3.728	55.97	α-CH
Isoleucine	t	0.930	13.60	δ-CH_3
	d	1.001	17.23	γ-CH_3
	m	1.463	26.87	γ-CH_2
	m	1.974	38.44	β-CH
	d	3.664	62.16	α-CH
Alanine	d	1.471	18.61	β-CH_3
	q	3.773	53.07	α-CH
Glutamic acid	m	2.056	29.43	β-CH_2, β'-CH_2
	m	2.345	35.86	γ-CH_2
	dd	3.754	57.05	α-CH
Glutamine	m	2.12	29.37	β-CH_2
	m	2.44	33.32	γ-CH_2

	t	3.75	57.01	α-CH
Methionine	unresolved	2.122	16.52	S-CH_3
	m	2.192	32.25	β-CH_2
	t	2.633	31.40	γ-CH_2
	dd	3.860	56.43	α-CH
Glycine	s	3.551	44.00	α-CH_2
Threonine	unresolved	1.32	21.3	γ-CH_3
	d	3.603	62.82	α-CH
	unresolved	4.252	68.41	β-CH
Lysine	unresolved	1.436	24.16	γ-CH_2
	unresolved	1.500	23.99	γ'-CH_2
	m	1.714	28.84	δ-CH_2
	m	1.898	32.32	β-CH_2
	t	3.010	41.62	ε-CH_2
	t	3.755	57.00	α-CH
Arginine	m	1.906	30.23	β-CH_2
	unresolved	3.237	43.07	δ, δ’ -CH_2
	unresolved	3.755	57.07	α-CH
Asparagine	dd	2.869	36.984	β-CH_2
	dd	2.9427	36.984	β'-CH_2
	unresolved	4.002	53.717	α-CH
Proline	m	1.996	26.47	γ-CH_2
	m	2.061	31.16	β-CH
	m	2.342	31.59	β'-CH
	m	3.330	48.54	δ-CH
	m	3.41	48.59	δ'-CH
	m	4.123	63.80	α-CH
Phenylalanine	m	3.122	38.84	β-CH_2
	m	3.279	38.83	β'-CH_2
	dd	3.989	58.50	α-CH
	m	7.310	131.81	$C_{2,6}$ ring

	m	7.417	131.69	$C_{3,5}$ ring
	m	7.367	130.11	C_4H,ring
Pyroglutamic acid	Unresolved	2.021	27.3	β-CH
	Unresolved	2.392	32.18	γ-CH
	Unresolved	2.562	27.87	β'-CH
	dd	4.167	60.82	α-CH
Tyrosine	dd	3.053	38.01	β-CH
	dd	3.188	37.98	β'-CH
	dd	3.935	58.52	α-CH
	d	6.892	118.34	2,6 ring CH
	d	7.183	133.27	3,5 ring CH
Serine	dd	3.839	58.85	α-CH
	dd	3.953	62.63	β, β’ -CH_2
Citrulline	m	1.592	27.22	γ,γ’ -CH_2
	m	1.86	30.26	β',β-CH_2
	q	3.13	41.60	δ,δ’-CH_2
	dd	3.75	57.02	α-CH
Ornithine	m	1.826	24.92	β-CH_2
	m	1.930	30.12	β-CH_2
	t	3.04	41.26	δ-CH_2
	t	3.756	56.99	α-CH
α-galactose	d	3.73	63.59	$C6H_2$
	m	3.78	70.25	C2H
	m	3.8435	71.759	C3H
	m	3.972	72.165	C4H
	Unresolved	4.050	73.493	C5H
	d	5.256	94.743	C1H
β-galactose	m	3.484	74.29	C2H
	m	3.645	75.11	C3H
	Unresolved	3.699	77.63	C5H
	Unresolved	3.730	63.59	$C6H_2$
	m	3.917	71.41	C4H

	d	4.577	98.907	C1H
Glycerol	m	3.547	65.01	CH_2
	m	3.640	64.96	CH_2
	Unresolved	3.773	74.55	CH
Tyramine	t	2.92	34.34	β-CH_2
	t	3.25	42.6	α-CH_2
	d	6.90	118.34	2,6 ring CH
	d	7.21	132.75	3,5 ring CH
Unknown A		2.798	38.96	
Unknown H		2.696	39.02	
Unknown B		1.763	32.92	
Unknown D		1.207	21.35	
Unknown E		1.208	27.06	
Unknown F		2.7653	40.119	
Unknown G		7.1388	133.09	

618 [a] ^{1}H chemical shifts reported with respect to DSS signal(0.00ppm).

619 [b] Multiplicity definitions: s, singlet; d, doublet; t, triplet; q, quartet; dd, doublet of doublets; m,

620 multiplet.

621

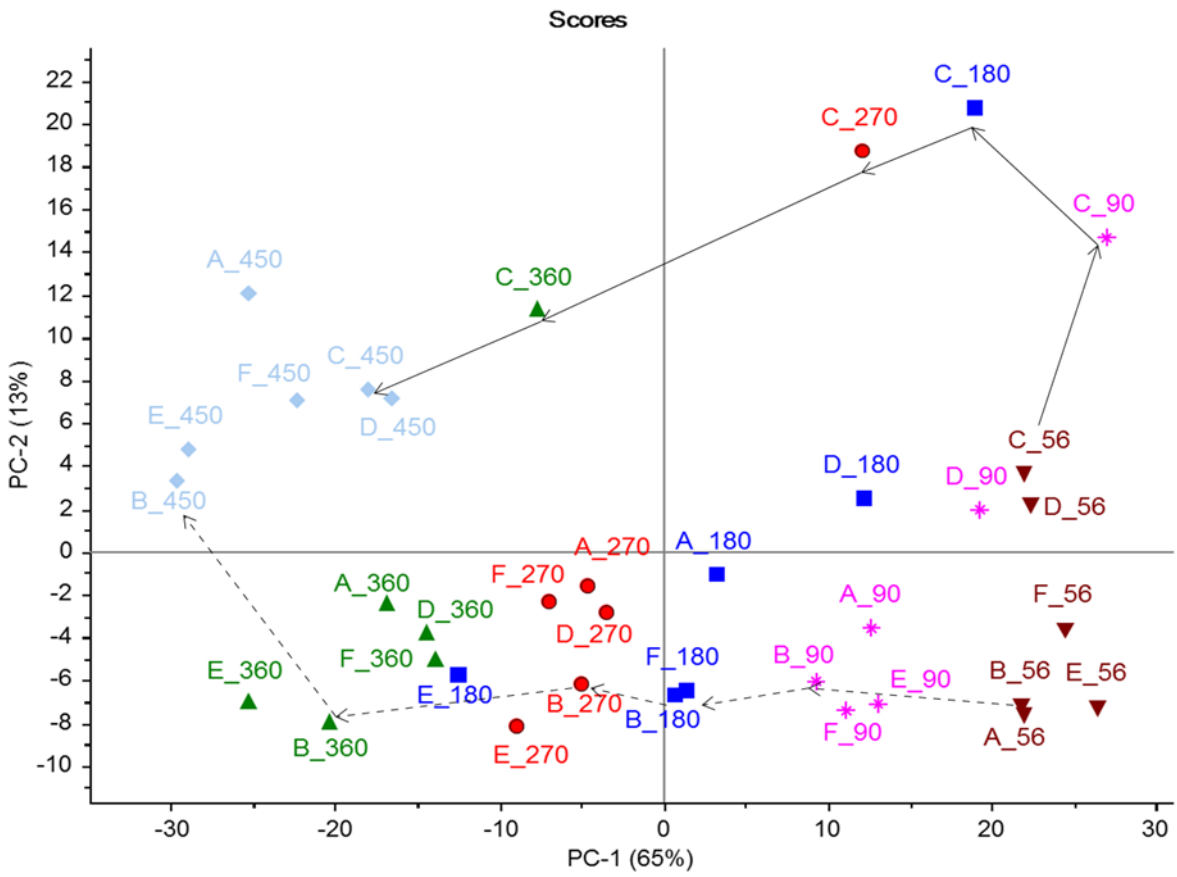
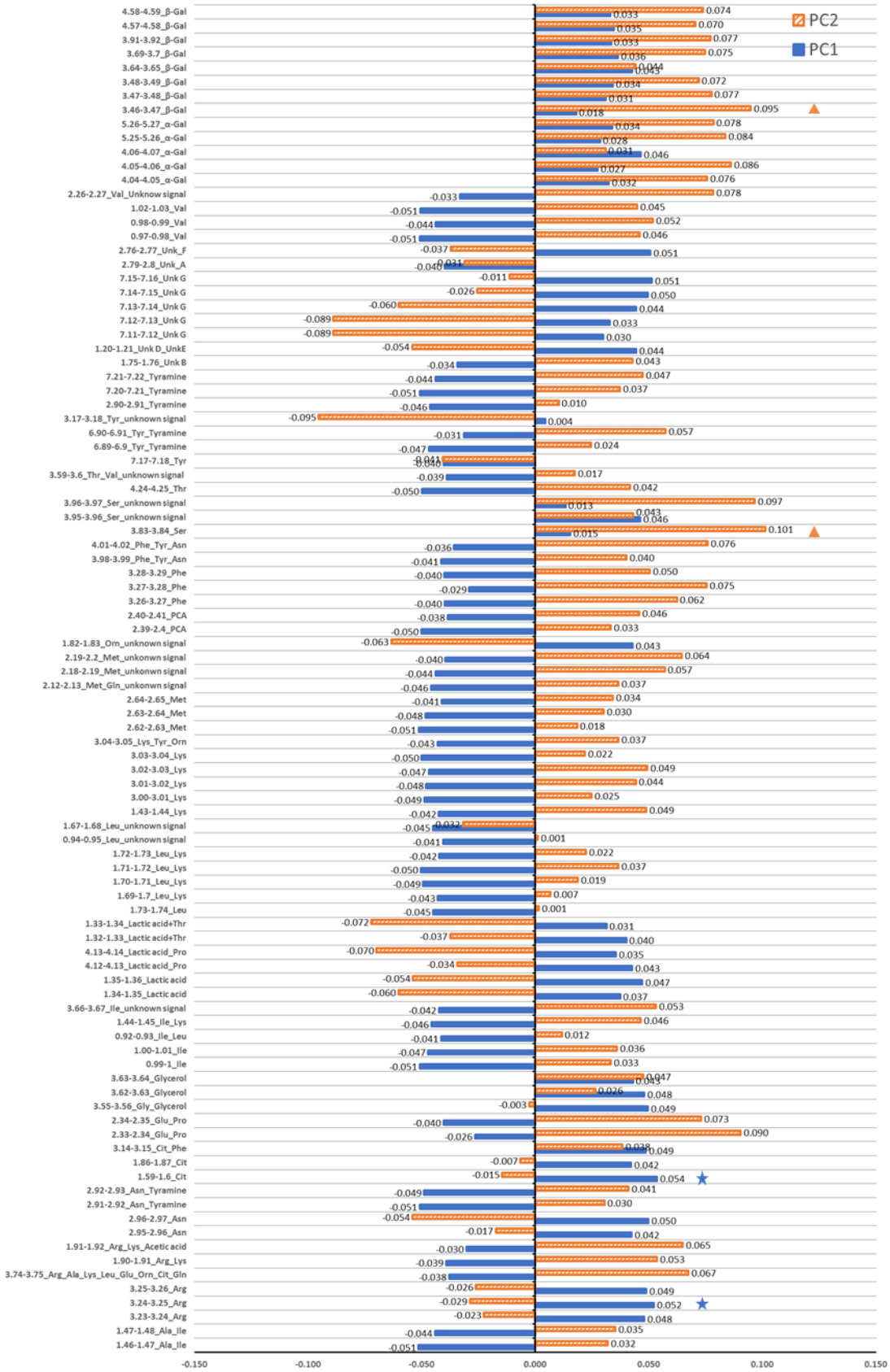


Figure 2 *The score plot of the PCA obtained from the mean of each ripening time point for batches A-F. The PCA metabolic trajectory plot maps the average position of the* ^{1}*H NMR spectra of aqueous extracts for each ripening time. The symbols and colours indicate ripening time in days:56 days (inverted triangle; brown), 90 days (star; pink), 180 days (□;dark blue), 270 days (O; red), 360 days (Δ; green), 450 days (◇; light blue). The letter and number denote the batch and number of ripening days e.g. E-270 is batch E at 270 days. The dashed line indicates the ripening trajectory of batch B; the solid line indicates the ripening trajectory of batch C.*

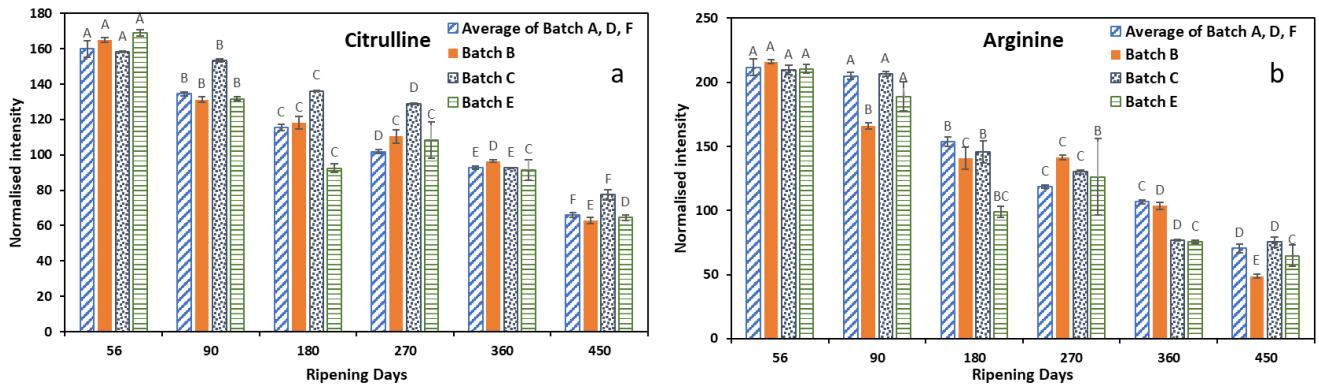


*Key to Compound identification: Ile, Isoleucine; Leu, Leucine; Val, Valine; Thr, Threonine; Lys, Lysine; Ala, Alanine; Cit, Citrulline; Arg, Arginine; Met, Methionine; Glu, Glutamic acid; Pro, Proline; PCA, Pyroglutamic acid; Asn, Asparagine; Tyr, Tyrosine; Phe, Phenylalanine; Gly, Glycine; Ser, Serine; α-Gal, α-Galactose; β-Gal, β- Galactose;

638 **Figure 3** *Loading coefficients for the chemical shift intervals from the PCA plot shown on Figure*
639 *2, comparing different ages and batch variations. The Y-axis is the bin interval with*
640 *corresponding assignments. The four individual metabolite bins which have the highest loading*
641 *coefficient in PC1 and PC2 and which discriminate the ripening and batch variation are labelled*
642 *with star and triangle symbols respectively.*

643

644

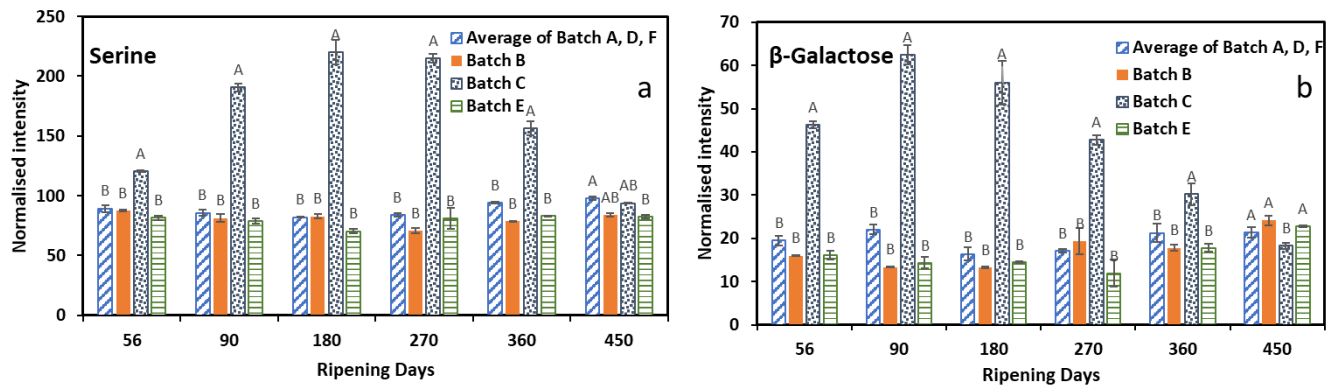


645

646 **Figure 4** *The most significant metabolites that change during the ripening process (a) citrulline*
647 *bin at 1.59-1.60 ppm (b) arginine bin at 3.24-3.25 ppm. The displayed values are the mean of*
648 *the normalised intensity of the metabolite. Error bars are the standard deviation of three*
649 *replicates. The different capital letters on the histograms indicate significant statistical difference*
650 *among ripening days for that batch (Tukey's test P<0.05).*

651

652



653

654 ***Figure 5*** *The most significant metabolites that differentiate variation in different batches: (a)*
655 *serine bin at 3.83-3.84 ppm (b) β-galactose bin at 3.46-3.47 ppm. The displayed values are the*

656 *mean of the normalised intensity of the metabolite. Error bars are the standard deviation of*
657 *three replicates. The different capital letters on the histograms indicate significant statistical*
658 *difference between batches for that ripening time (Tukey's test P<0.05).*
659

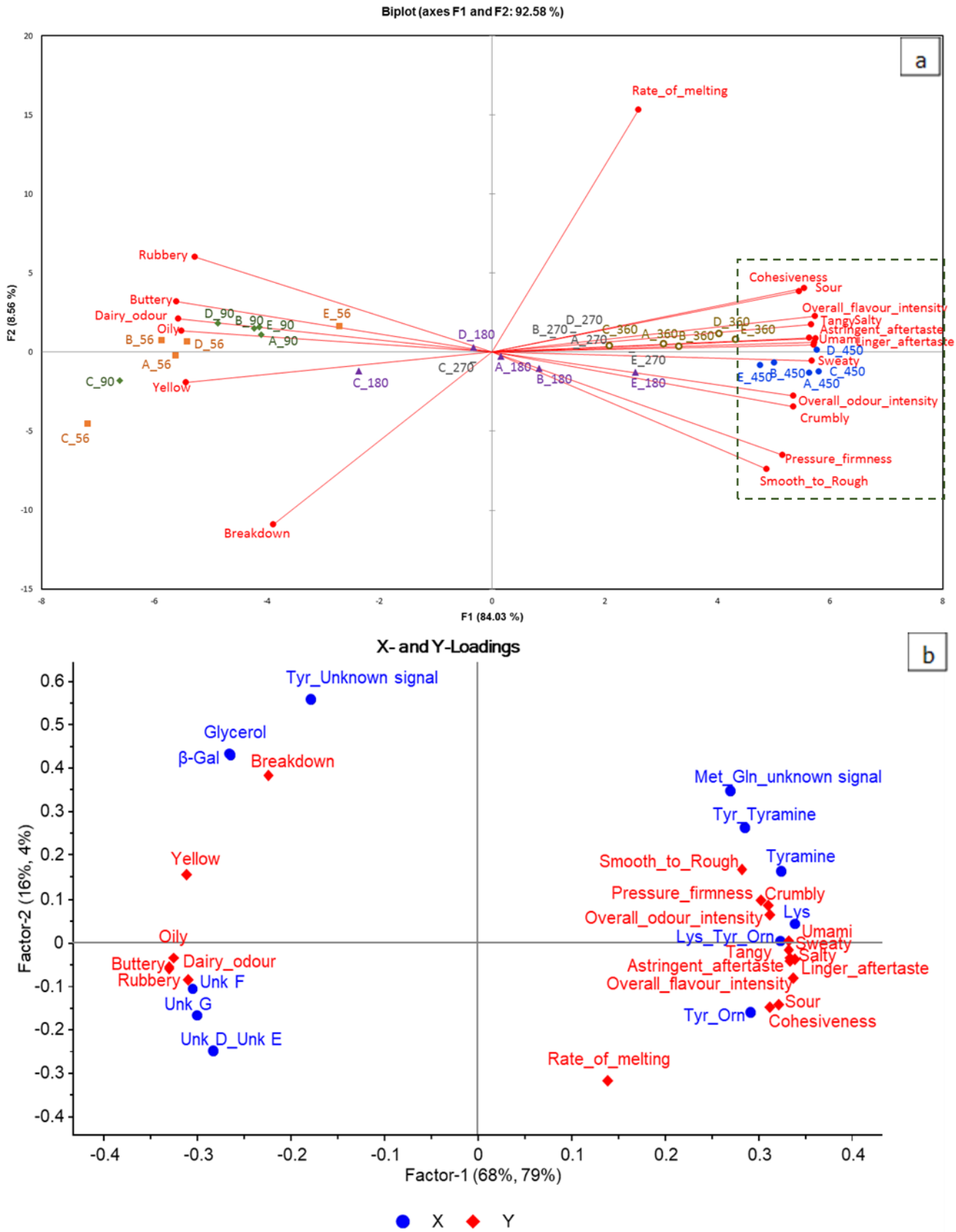


Figure 6 *(a) PCA bi-plot carried out on sensory attributes by the training panel. All the sample points are labelled as batch number and ripening time in days. Symbols and colours of the*

samples indicate different ripening times: 56 days ripening (□; orange), 90 days ripening (◇; green), 180 days ripening (Δ; purple), 270 days ripening (—; grey), 360 days ripening (o; yellow),450 days (●; blue). All the sample points are the mean of three replicates ×10 panel evaluations. The sensory attributes that correlated with mature Cheddar cheese attributes are in the dashed box (b) Partial Least Squares correlation biplot of sensory attributes evaluated at all ripening days and most relevant metabolites of the Cheddar cheese aqueous extract spectra that explain the most variance of the sensory profile.(color used in these graphs)

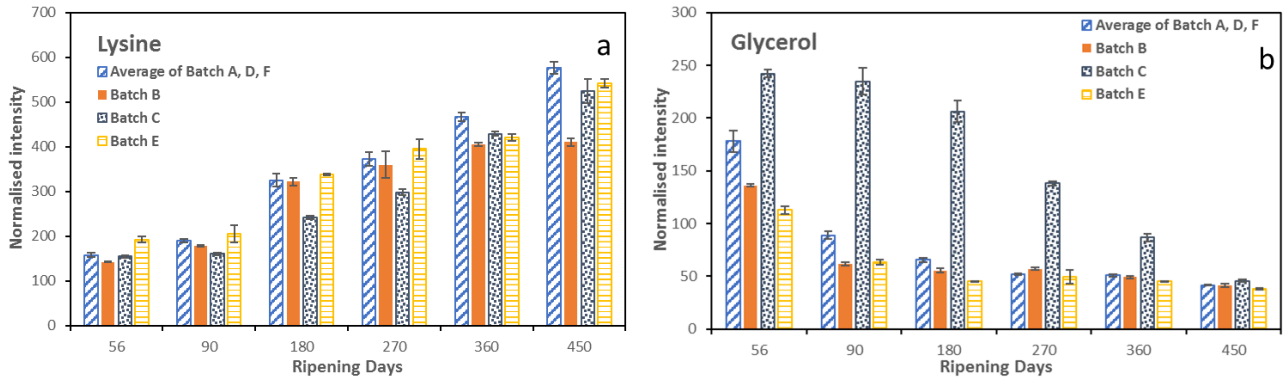


Figure 7 *The most important individual metabolites that are correlated with most of sensory attributes: (a) Lysine bin at 3.03-3.04ppm (b) Glycerol bin at 3.63-3.64ppm. The displayed values are the mean of the normalised intensity of the metabolite. Error bars are the standard deviation of three replicates.*