Doctoral Dissertation

Metabolomics-based study for the quality prediction of *Litopenaeus vannamei* (white leg shrimp)

Safira Latifa Erlangga Putri

June 2021

Division of Advanced Science and Biotechnology

Graduate School of Engineering,

Osaka University

List of Abbreviations

(*in alphabetical order*)

ANOVA	Analysis of variance
GC	Gas chromatography
HLSO	Head-less shell-on
IP	Ion-pair
LC	Liquid chromatography
MS	Mass Spectrometry
MSTFA	N-Trimethylsilyl-N-methyl trifluoroacetamide
OPLS	Orthogonal Projection to Latent Square
PCA	Principal Component Analysis
PLS	Projection to Latent Square
RMSEcv	Root-Mean-Square-Error of Cross validation
RMSEE	Root-Mean-Square-Error of Estimation
RMSEP	Root-Mean-Square-Error of Prediction
RP	Reversed-phase
QDA	Quantitative Descriptive Analysis
QqQ	Triple quadrupole
VIP	Variable Importance in Projection

Table of Contents

List of Abbreviations **2**

Chapter 1 ..... **6**

General Introduction **6**

1.2 Shrimp aquaculture 8

1.2.1 Important shrimp species 8

1.3 Shrimp grading system 8

1.4 Metabolomics approach 11

1.4.1 Shrimp metabolomics 12

1.4 Objective and strategy 12

1.5 Thesis outline 13

Chapter 2 ..... **15**

GC/MS based-metabolomics profiling to reveal the importance of shrimp count size **15**

2.1 Introduction 15

2.2 Materials and Method 17

2.2.1 Shrimp samples information for metabolomics analysis 17

2.2.2 Extraction procedure for shrimp metabolomics analysis 19

2.2.3 Derivatization and gas chromatography-mass spectrometry (GC/MS) analysis 19

2.2.4 GC/MS data processing 20

2.2.5 Multivariate analysis 20

2.2.6 Data mining 22

2.3 Result and discussion 23

2.3.1 GC/MS metabolite profiling analysis of different shrimp sizes 23

2.3.2 Principle component analysis of different size of white leg shrimp samples 23

2.3.3 Projection to Latent Square Regression model for predicting the size of white leg shrimp. 25

2.3.4 The significance of size in molecular quality 32

2.4 Conclusion 33

Chapter 3 34

Targeted-metabolomics analysis for the prediction model improvement 34

3.1 Introduction 34

3.2 Materials and method 35

3.2.1 Sample information 35

3.2.2 Sample preparation method for GC/MS analysis 37

3.2.3 Sample preparation method for ion-pair reversed-phase LC/QqQ/MS analysis 37

3.2.4 Ion-pair reversed-phase LC/QqQ/MS analysis 38

3.2.5 Multivariate analysis 39

3.3 Result and discussion 40

3.3.1 The significance of size in commercial white leg shrimp 40

3.3.2 Validation of size-based prediction OPLS regression model 42

3.3.3 Size-dependent metabolites 44

Chapter 4 50

The significance of shrimp size in sensory properties 50

4.1 Introduction 50

4.1.1 Sensory descriptive analysis 51

4.2 Material and methods 52

4.2.1 Sample information of quantitative descriptive analysis (QDA) 52

4.2.2 Shrimp cooking method 52

4.2.3 Panelist recruitment and training 53

4.2.4 Sensory testing procedures 53

4.2.5 Sensory data analysis 53

4.3 Results and discussion 54

4.3.1 Metabolite profile of white leg shrimp cultured in the same pond 54

4.3.2 Sensory analysis and the role of taste active metabolites in different size of commercial shrimps 58

4.3.3 The significance of size in sensory shrimp sensory properties 60

4.4 Conclusion 63

Chapter 5 64

Conclusion 64

Acknowledgement 66

List of publications 75

Supplementary 76

Chapter 1

General Introduction

1.1 Aquaculture system

Aquaculture, the farming of aquatic organisms has been gained popularity because of its sustainable production and has been considered as an option to cope with the world food demand (1). The growth of global aquaculture production is projected to continue at a strong pace until it matches the production of capture commodities by the year 2030 (2). It could potentially provide half of the worldwide supply, lessening the burden of wild-caught production and natural resources.

Aquaculture systems range from extensive, semi-extensive, and intensive systems. The classification is based mainly on pond facilities, stocking density, food supply, water management, yield, technical know-how and skill, and other major inputs (3).

Extensive system. A conventional open pond is the main example of extensive farming. This method consisting of maintaining natural or artificial ponds in such a way as to foster the development of aquatic fauna. The presence of micro-organisms, small mollusks and crustaceans, larvae, and worms form the base of the aquatic food pyramid and the stock is left to grow on its own utilizing natural source. The region suitable for extensive farming are the area of coastal, mangrove, and estuaries, where the water management will be dependent on a tidal fluctuation. This system produces low productivity and low stocking density with low-cost production and maintenance due to unnecessary feeding and aeration.

Semi-intensive system. A semi-intensive system is largely dependent on a natural food source. It also utilizes artificial or natural open ponds but aiming to increase the production of fish beyond the level of an extensive system through the use of supplementary feeds (3).

Intensive system. This system aims to raise the shrimp at very high densities with controlled water effluent and artificial feed to increase productivity. Recirculating Aquaculture System (RAS) is one example of an intensive method that offers a more environmentally friendly water effluent system. RAS system is performed in a closed circuit where the water is recycled and being reused. Before entering the biofiltration tank, the solid waste is removed through filtration. The water effluent is then subjected to the ammonia removal step, where the ammonia consumed by the bacteria is converted into nitrogen. It also allows for organic waste to be treated before being disposed of in nature. In the last step, the oxygen is reinjected and the carbon dioxide is removed before re-entering the grow tank. Apart from the cost of the investment and its benefit, RAS requires high energy consumption and dependent on complex technology.

1.2 Shrimp aquaculture

Shrimp is one of valuable commodities in aquaculture industry and has important significance in global food supply (4). Shrimp gets popular owing to its high nutritional value, distinctive flavour, and a tender and delicate texture (5). Aquaculture shrimp is among the fastest growing and most traded food commodities in the world (6). World Bank projected a strong growth in the global consumption of shrimp, with the largest overall share of consumption is in freshwater category, encompassing tilapia, Pangasius, and carp. According to market research, shrimp aquaculture is expected to grow by 50 to 60 percent between 2008 and 2030 (2). Most importantly, farmed shrimp commodity generates essential economic benefit for many developing countries (7).

1.2.1 Important shrimp species

Litopenaeus vannamei or well-known as white leg shrimp is one of the shrimp species in aquaculture industry. Based on the FAO data obtained in 2016, white leg shrimp has the highest market share among other shrimp aquaculture species. It has the potential to grow fast, more tolerant to disease, and tolerates a wide range of salinity (8,9). These characteristics have high export potential which attracted farmers' attention (9). Because of its domination in the global market, *Litopenaeus vannamei* is considered as an economically important species with high marketable value.

1.3 Shrimp grading system

At present, shrimp is marketed globally based on its count size. This unit simply counts the number of shrimp contain per 1 kilogram or 1 pound (10). Count size relies on the shrimp size, specifically shrimp body mass and it applies particularly for head-less shell-on shrimp

(HLSO) (11). As shown in table below, count size substitutes un-standardized size description which confuses the consumers. For example, count size 8/12 / kg would have 8-12 of shrimps in 1 kg package, with the approximate body mass equivalent to 35-57 g/ shrimp tail. On the other hand, count size >70 represents small shrimps with the size approximately smaller than 7 g.

Table 1.3-1 Conversion table of shrimp count sizes

Word description	**Count size**	**Standard average pieces***	**Equivalent mass (g)**
Colossal	8/12	10	35 – 57
Jumbo	13/15	14	28 – 35
Jumbo	16/20	18	23 – 28
Extra large	21/25	23	18 – 23
Medium large	26/30	28	14 – 18
Large	31/40	35	11 – 14
Medium	41/50	45	9 – 11
Small	51/60	55	7 – 9
Small	61/70	65	6 – 7
Extra small	>70	>80	<7

As shown in Figure 1.3-1, after harvesting the whole shrimp are first graded according to their body mass. Size uniformity is an important aspect in grading process. One must follow the specified rule where the number of shrimp pieces in one kilogram complies with the standard count size. This step is a critical processing step that adds value and increases profits. Different sized shrimps are further processed by removing the head before the packaging step. The selling and marketing of the final products follow the standard count size in industry, where the price increase as the size gets larger.

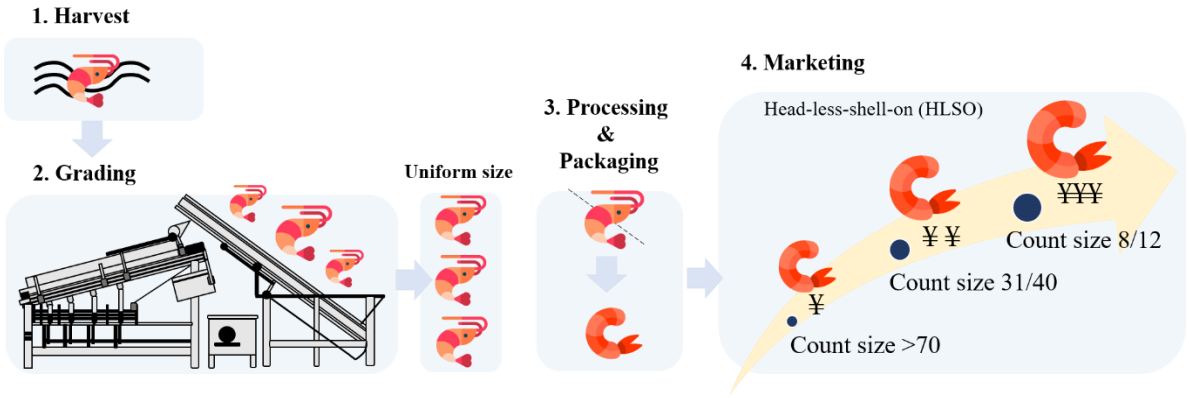


Figure 1.3-1 Grading system of commercial shrimp

From a consumer perspective, size is one of the important factors that affect consumer's preferences. In general, size anchors the value of perception and satisfaction. In most cases, large foods are perceived as luxuries and attractive. Generally, size also represent the edible portion of food products, therefore consumers are willing to pay more for better eating quality and satisfaction. In a more particular context, size is the manifestation of growth that strongly links to operation cost and time. Furthermore, a large commercial size reflects good management of farming and aquaculture systems, enabling an optimized condition for maximum growth.

Currently, the standard practice for shrimp quality assessment has relied on the tracing of anti-biotic residue, microbial counts, and organoleptic routine to examine shrimp freshness (12). These parameters do not affect the pricing not market segments in the shrimp industry. Furthermore, there have been no studies reporting the correlation between quality and shrimp size. Therefore, the pricing of shrimp has not been based on scientific findings. In line with this issue, the advancement of technology in aquaculture industries has risen the sense of urgency to re-evaluate the quality parameter of the commercial shrimp products.

1.4 Metabolomics approach

Metabolomics is one of the omics studies about metabolites, low-molecular-weight molecules within cells, biofluids, tissues, or organisms. Metabolites reflect the molecular quality of living systems as they serve as substrates and products of metabolism which drive various cellular functions (13). Detection of the alteration of metabolites abundance provides valuable information into the underlying mechanism in response to various physiological conditions. In food science application, metabolomics studies not only the interaction between chemical-chemical constituent, but also the interaction between chemical constituent-external factors which exhibit final attributes that correlate to specific appearance, aroma, or taste characteristic (14).

In general, techniques in the metabolomics approach involve a chromatographic separation of complex metabolites matrix and compound detection based on mass spectrometry. The identification of metabolites is enabled by measuring the mass–to–charge ratio (*m*/*z*) of ions that are formed by inducing the loss or gain of a charge from a neutral species. In addition to m/z value, retention time and fragment pattern acquired from tandem mass spectrometry also aid the identification of metabolites (13).

Following data processing, multivariate data analysis plays an important role in extracting essential information from complex high-dimensional data. In metabolomics study, principal component analysis (PCA) is the most common unsupervised analysis to extract hidden information from a set of data. This is possible by constructing a new axis called the principal component that can maximize the variance of the data. Metabolomics-coupled to multivariate analysis enables data interpretation, assisting the hypothesis generation of biological significance.

1.4.1 Shrimp metabolomics

In the recent years, there has been an increasing number of publication towards sustainable production of shrimp, one of valuable commodity in aquaculture industry. A number of study emphasized the production improvement in the upstream level, or the assessment of metabolic perturbation upon disease infection or under stress environments in which most researches focused on the analysis of juvenile shrimp (15–18). Publications focusing on the shrimp quality assessment are still under-studied, particularly in relation to its economic aspect.

In this study, size as one factor that affects the shrimp price, was also proposed as a potential parameter to assess the quality of shrimp. Metabolomics as the closest omics study to the phenotype provides useful insights into the quality of shrimp in molecular level. As the importance of size was examined in the molecular level, the elucidation of key metabolites associated with size enables the justification of shrimp quality. To complete the study, sensory evaluation was performed to confirm, whether or not size of the shrimp affect shrimp sensory quality.

1.4 Objective and strategy

Therefore, the objective of this study is to establish the method for the quality prediction of *Litopenaeus vannamei* (white leg shrimp). Considering the importance of size as grading unit in the market, this study pursue the justification of shrimp size as potential quality evaluation parameter for shrimp as seen through the metabolites involved in the central carbon metabolism. As the first strategy, metabolomics analysis combined with multivariate data analysis was performed to examine the molecular quality throughout different sizes of white leg shrimp. Following this, predictive metabolomics analysis was performed to predict different size of shrimp using metabolites relative intensity data. In the second strategy, validation on the

prediction model was carried out to confirm whether the finding was also applicable for commercial shrimp from the actual market. Lastly, sensory evaluation was carried out particularly within the range of commercial size to validate the influence of shrimp size in sensory properties, especially shrimp taste quality.

1.5 Thesis outline

This thesis is divided into five chapters. Chapter 1 describes the general introduction of shrimp aquaculture, the justification on the shrimp size as an important variable in the market, and the objective of this study.

In chapter 2, the work on GC/MS-based analysis-coupled with multivariate analysis was described. In order to reveal the importance of size, shrimp samples collected from various conditions including culture period, farming systems, and sampling locations were analyzed. The metabolite data matrix were further subjected to PCA to assist the generation of the hypothesis. In the case where novel metabolic perturbations are observed, the correlation between metabolite abundance and shrimp size was further examined by constructing Projection to Latent Structure (PLS) model. Important metabolites that strongly correlated with variable size were suggested based on the variable importance of the projection (VIP) score.

The work on chapter 3 focused on widely targeted metabolomics analysis to gain deeper understanding on important metabolites correlated with shrimp size. The expansion of metabolite coverage was conducted using ion-pair reversed-phase LC/QqQ/MS (liquid chromatography triple quadrupole mass spectrometry) to specifically target phosphate-related compounds. Following the validation of prediction model using commercial shrimp sample, the VIP metabolites, or known as size-dependent metabolites having positive correlation with size

were suggested. Using this information, general mechanism of metabolic perturbation throughout different size of shrimp was proposed.

In chapter 4, shrimp having commercial size range were subjected to both metabolomics analysis and sensory evaluation, namely quantitative descriptive analysis (QDA) method. In this chapter, the quantification on 14 shrimp sensory attributes, such as sweetness and juiciness was performed by 10 expert panelist. The accumulation of taste-active metabolites including size-dependent metabolites was compared with the result from QDA. Possible cause of a distinct taste perception in large shrimp (13-17 g) in correlation with taste-active metabolites was suggested.

Finally, in chapter 5 the conclusions of this study and the future perspectives were proposed.

Chapter 2

GC/MS based-metabolomics profiling to reveal the importance of shrimp count size

2.1 Introduction

Shrimp production through aquaculture systems has enabled a sustainable shrimp supply in various count sizes, ranging from the smallest (count size >70) to the largest shrimp sizes (count size 15/20). As the size expresses the edible portion of shrimp, large sizes are considered to be highly palatable, and therefore priced higher compared to smaller sizes (19).

Size as grading parameter relies on an implicit assumption that large shrimps come with high quality, leading to potentially flawed tool for shrimp pricing. Furthermore, studies reporting the quality assessment of shrimp are scarce. In particular, there have been no studies reporting the correlation between quality and shrimp size. Therefore, the pricing of shrimp has not been based on scientific findings.

From an aquaculture perspective, shrimp size has been utilized to monitor the growth of shrimp. Figure 2.2.1-2 shows the shrimp production scheme. In this scheme, the commercial production occurs during the grow-out phase, where shrimp are grown into marketable size for 3-6 months. During the late post-larvae stage, shrimp morphology is similar to adult shrimp (20). Therefore the determination of shrimp life stage during the grow-out period relies on the shrimp body mass, rather than shrimp morphology.

In this study, the sampling period was during the grow-out phase and the collection of shrimp relied based on shrimp size, regardless of the shrimp age or life stage. This strategy

represents the actual shrimp marketing that relies upon the count size although many shrimp products are supplied from different places and suppliers.

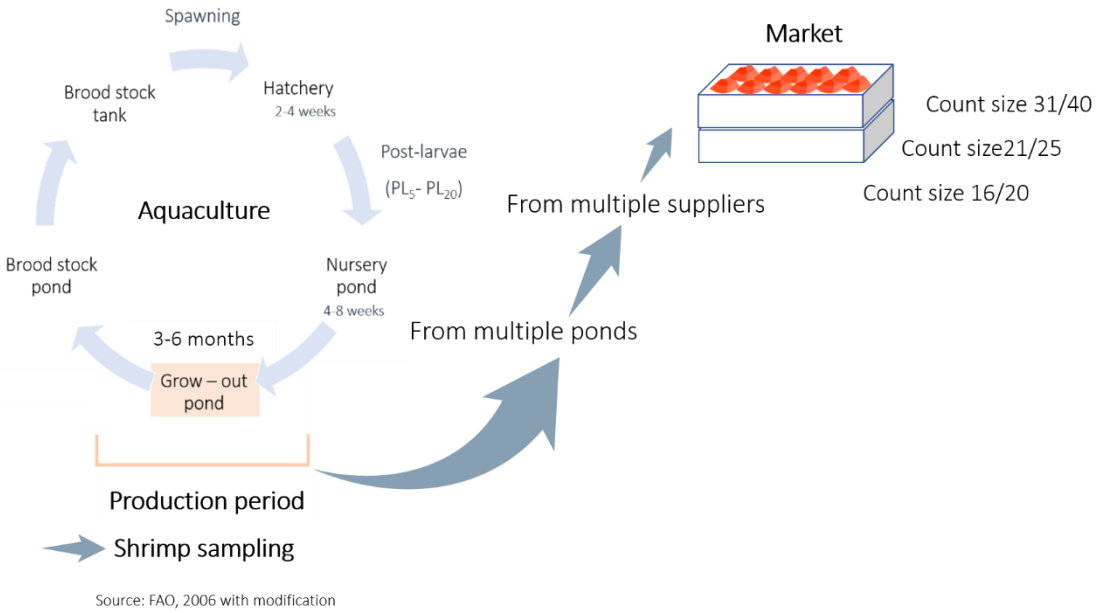


Figure 2.2.1-2 Grow-out scheme in shrimp commercial farm.

The production period of commercial shrimp takes place in natural cycles during grow-out period (Taken from FAO, 2016 with modification).

In this chapter, white leg shrimp samples having various body mass were analyzed using a GC/MS-based metabolomics approach to reveal the importance of shrimp size. Metabolomics-coupled with multivariate analysis allows hypothesis generation from the sample visualization. The differences between classes based on the identified relevant spectral features extracted from complex datasets can be visualized (21). Following this, a linear relationship between shrimp size and shrimp metabolome is evaluated through the construction of the PLS model. By understanding the relationship between these variables, we could discover notable potential in applications to assess the quality of shrimp

2.2 Materials and Method

2.2.1 Shrimp samples information for metabolomics analysis

In the present study, the term size refers to the body mass of head-less shell-on (HLSO) white leg shrimp. Shrimp samples were collected from three different locations in Indonesia, namely Subang (West Java Province), Buleleng (Bali Province), and Banten (Banten Province).

Moreover, the sample was collected solely based on the size despite the difference in days of culture to ensure that the variances observed in the analysis were due to the differences in size and not biased by other influencing factors. Shrimps from the three locations were grown in semi-extensive and intensive open pond systems with salinity ranging between 29-31 g/L. After harvesting, the shrimps were placed into a box containing ice, and the head was rapidly removed through the centerline of the head. Following this, samples were kept at -20°C before being transported to Japan by flight. Once the samples arrived, the shrimp body mass were re-measured and classified according to its count size. As for count size > 70 and 61/70, shrimp sample was again categorized based on its similar average body mass.

Table 2.2.1-1. Samples information

a) PCA and PLS data set

Count size	Average mass (n=3)		DOC* (days)	Location	System	Data set
<70	2.62	g	13-17	Buleleng-Bali	Semi-intensive	Training set
<70	3.52	g	18-24	Buleleng-Bali	Semi-intensive	Training set
61/70	5.59	g	28-37	Buleleng-Bali	Semi-intensive	Training set
61/70	6.23	g	31-42	Subang-West Java	Intensive	Training set
51/60	7.76	g	39-52	Subang-West Java	Intensive	Training set
41/50	9.77	g	49-62	Buleleng-Bali	Semi-intensive	Training set
26/30	17.06	g	85-117	Buleleng-Bali	Semi-intensive	Training set
21/25	20.24	g	101-135	Buleleng-Bali	Semi-intensive	Training set
16/20	23.43	g	117-156	Buleleng-Bali	Semi-intensive	Training set
61/70	6.64	g	33-44	Banten-Banten	Semi-intensive	Test set
51/60	7.56	g	38-50	Banten-Banten	Semi-intensive	Test set
41/50	8.42	g	42-56	Banten-Banten	Semi-intensive	Test set

*Days of Culture

Table 2.2.1-1. listed the samples consisted of twelve different average sizes of headless shrimp-shell-on (HLSO) shrimp that can be categorized into seven count sizes. It is important to note that, HLSO shrimp was utilized to represent the count size system in the market. In general, the mass of shrimp head takes about 40% of the total body mass and therefore the average mass listed in the table was smaller than the actual mass of shrimp. With the minimum size of 10 g (head-on-shell-on) for shrimp harvesting (depend on the market segments), this data set represented seven common commercial count sizes in the global shrimp market (16/20, 21/25, 26/30, 41/50, and 51/60) and five groups were classified as small shrimps (61/70 and

<70). In order to visual the variance of the samples, the data labeled as training set was subjected to both PCA and PLS. The remaining of data set were introduced into the prediction model for validation.

2.2.2 Extraction procedure for shrimp metabolomics analysis

Peeled shrimp samples were subjected to lyophilization for 18 h and then ground into fine particles using a multiple bead shocker (Model ST-5010 PCR, Yasui Kikai, Osaka). Extraction was performed twice using 40% ethanol as an extraction solvent containing internal standard ribitol (50 µg/L). Solvent (1 mL) was added to 5 mg of shrimp powder, followed by homogenization using a thermomixer for 1 h at 25°C. The mixture was then centrifuged (4°C, 10000 rpm) for 10 min. After centrifugation, the supernatant was pooled and the remaining pellet was re-extracted. To prevent a saturated peak, the pooled supernatant was diluted in the ratio of 1:2 and vortexed for 30 s. Then, 300 µl of samples and 300 µl of QC samples were transferred to 1.5 mL microfuge tubes with screwed caps. The solvent was removed by vacuum centrifugation for 1.5 h, followed by overnight lyophilization using a VD-800F Freeze Dryer (Taitec, Tokyo, Japan). All the samples were analyzed in triplicates (n=3), except for sample 16/20, which was only provided in duplicate (n=2).

2.2.3 Derivatization and gas chromatography-mass spectrometry (GC/MS) analysis

Two-step derivatization was performed prior to GC/MS analysis to enable the analysis of non-volatile metabolites. Oximation was conducted in the first step by adding 100 µL of methoxyamine hydrochloride (20 mg/mL in pyridine) (Sigma-Aldrich, St. Louis, MO, USA) into lyophilized samples following incubation in a thermomixer for 90 min at 30°C. Subsequently, 50 µL of *N*-Methyl-*N*-trimethylsilyl-trifluoroacetamide (MSTFA) (GL Sciences),

a sylilation reagent, was added to the samples and the samples were incubated for 30 min at 37°C.

GC/MS analysis was performed on a GCMSQP 2010 Ultra (Shimadzu, Kyoto, Japan) equipped with an InertCap 5 MS/NP column (GL sciences). Tuning and calibration of the mass spectrometer were performed prior to analysis. Derivatized sample (1 µL) was injected in split mode, 25:1 (v/v), with an injection temperature of 230°C. The carrier gas (He) flow was 1.12 mL/min with a linear velocity of 39 cm/s. The column temperature was held at 80°C for 2 min, increased by 15°C/min to 330°C, and then held for 6 min. The transfer line and ion source temperatures were 250 and 200°C, respectively. Ions were generated by electron ionization (EI) at 0.94 kV. Spectra were recorded at 10000 µ/s over the mass range m/z 85−500. A standard alkane mixture (C8−C40) was injected prior to every QC sample for peak identification.

2.2.4 GC/MS data processing

The chromatographic data obtained from GC/MS analysis were first converted to AIA format using a GCMS solution software package (Shimadzu, Kyoto, Japan). Data processing included peak detection, peak identification, peak alignment, and filtering parameters, which was conducted using MS-Dial version 3.30. Putative identification in accordance with level 2 (Sumner et al. 2007) was obtained by comparison to the online database via MS Dial, (GL-Science DB (InertCap 5MS-NP, Kovats RI MSP file), which are available online, and in-house library AIOutput.

2.2.5 Multivariate analysis

Multivariate analysis was performed using SIMCA-P+ software version 11 (Umetrics, Sweden). Metabolome data matrix from both instruments was integrated by selecting the most dominant metabolite coverage from each platform. Amino acids and its derivatives, except

arginine, fatty acids, and inorganic compounds were derived from GC/MS. Meanwhile nitrogenous-based compounds, organic acids, and heterocyclic compounds was obtained from IP-RP LC/QqQ/MS analysis. The data was auto scaled to unit variance (UV), giving all the variables an equal importance. Unsupervised statistical analysis, Principal Component Analysis (PCA) was first carried out to extract the information and visualize the sample behavior by maximizing the variance (23). Therefore, the significance of size varying with origin was evaluated according to the sample clustering pattern along the first two principle components.

Following unbiased statistical analysis, projection to latent structure (PLS) regression analysis was carried out to examine the ability of metabolome data in predicting the size of shrimp. The performance of the model was examined according to R^2, Q^2, which denote the goodness of fit and prediction ability, respectively. The accuracy of the model was examined by root-mean-square of estimation (RMSEE) derived from the model, while the root-mean-square of prediction (RMSEP) indicates the predictability of the model (24). Cross-validation was performed to estimate the ability of a model to correctly predict the response matrix Y of new individuals (25). The training model was built from shrimp samples obtained from the farm where the validation used commercial shrimp obtained from the market. was assigned as training set. The selection of metabolites showing the most relevant information between variables was based on the importance of variable in projection (VIP) values. A stronger correlation between variables is depicted from the highest VIP score (26). The cut off value of VIP was set to 1, which means the influence of metabolites having VIP score below 1 was considered insignificant. The negative and positive coefficients following the VIP score indicated the association between size and metabolome data as independent variables.

2.2.6 Data mining

The relative abundance of metabolites was normalized by dividing the intensity with the internal standard ribitol. Multivariate analysis was performed using SIMCA-P+ software version 11 (Umetricts, Umea, Sweden). The normalized metabolome data were subjected to principal component analysis (PCA) and partial least square projections to latent structures (PLS) regression. PCA, as a non-biased statistical tool, was employed to extract and visualize sample information (23). The data were auto-scaled; thus, the mean and unit variance became zero.

Meanwhile, a supervised analysis projection to latent structure regression (PLSR) model was employed to find the linear relationship between shrimp size and metabolite profile. The model was evaluated based on the values of R^2 and Q^2, which express the goodness of fit and prediction ability, respectively. The optimum number of latent variables was optimized and suggested by SIMCA-P software. The variable cut-off was selected based on the importance of variable in projection (VIP) values. A stronger correlation between variables is depicted from the highest VIP score, which is equal to or greater than 1 (26). The negative and positive coefficients following the VIP score indicated the association between size and metabolome data as independent variables. Following this, one-way ANOVA with Tukey's test was carried out to evaluate the significant difference of PLS-selected metabolites among different-sized white leg shrimp. The cut-off value was set to 5%.

2.3 Result and discussion

2.3.1 GC/MS metabolite profiling analysis of different shrimp sizes

GC/MS analysis was performed to compare the metabolite profiles of different-sized white leg shrimps. First, we focused on the global analysis of low molecular weight and non-volatile compounds in combination with PCA to determine whether there is an observed trend among different shrimp sizes. Total 104 metabolites were annotated from GC/MS analysis consisting of 35 amino acids (proteinogenic and non-proteinogenic), 4 fatty acids, 10 nitrogenous compounds, 17 organic acids, 4 amines, 21 sugars, and 12 other compounds. These annotated metabolites were further analyzed using multivariate analyses.

2.3.2 Principle component analysis of different size of white leg shrimp samples

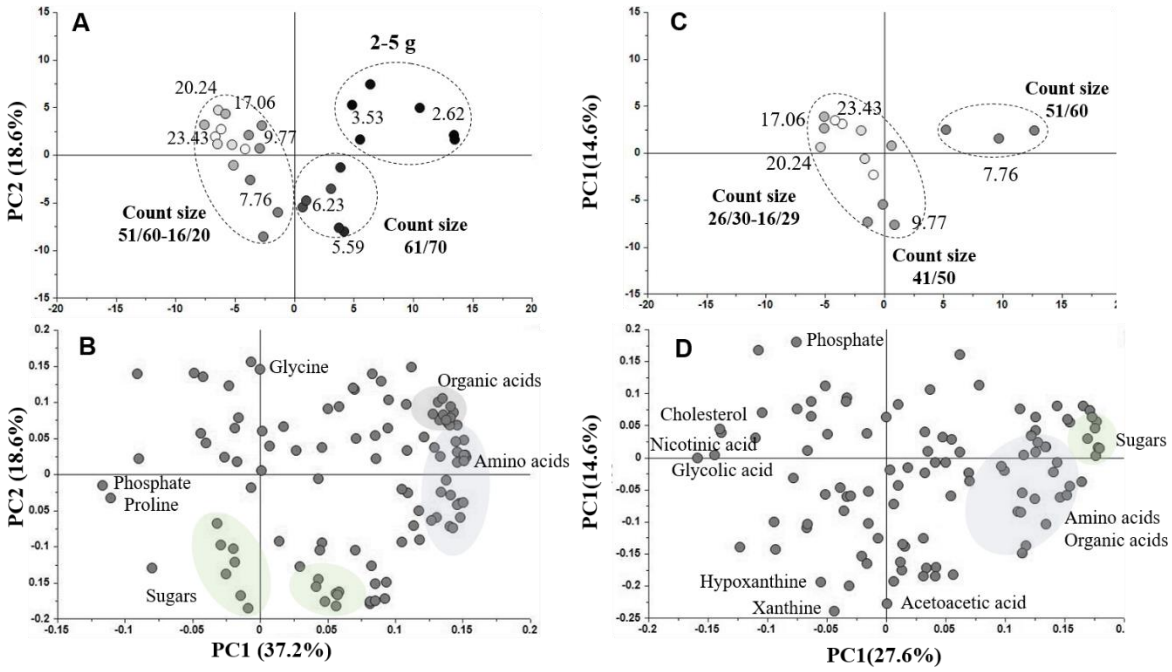


Figure 2.3.2-1. PCA result of different count sizes of white leg shrimp. (A) PCA score plot derived from 9 classes of shrimp size. Each dot represents shrimp sample. Colour gradation from dark to light indicate the shrimp size from small to large. The dashed line marks the boundary of clusters. The number written near the cluster describe the average mass of shrimp according to its count size. (B) Loading

plot represent the contribution of metabolites to the separation of samples. Blue and grey colour represent the cluster of amino acids and organic acids as important metabolites associated with small shrimp. (C) PCA score plot of different sized shrimps upon the removal of small shrimp (average mass < 7 g). (D) Loading plot of shrimp sample with the average mass 7-23 g. The smallest size in the data set still retained higher accumulation of amino acids.

As shown in **Figure 2.3.2-1**, each dot in score plot (A and B) represent the sample while the colour gradation explain the order of the size. Loading plot for the first (PC1) and the second (PC2) principal components represent the highest variance from the data set. Each dot represent metabolites associated to the separation. Metabolites located in the extreme part of axes were considered as contributing metabolites that differed the samples.

The PCA score plot revealed the grouping of shrimp samples into three distinct clusters with different size ranges, namely 2-4 g (count size >70), 5-7 g (count size 61/70), and 7-24 g (count size 51/60-16/20) (**Figure 2.3.2-1.A**). Clustering along PC1, which accounted for 37.2% of the variance, showed the separation of samples from small to large sizes. The trend on PC1 most likely explained metabolome changes correspond to the size difference. One-way ANOVA with post-hoc Tukey's HSD test result confirmed the 58 among 104 metabolites were significantly different across shrimp count sizes (**Table 2.3.2-1**). Although an obvious difference was observed between count size >70-61/70 and 51/60-16/20, PCA was able to differentiate small-sized shrimp sample as they formed two separate groups according to count sizes, 2-5 g (count size >70) and 5-7 g (count size 61/70). A prominent change in metabolite intensity was observed when the average mass was more than 7 g (**Figure 2.3.2-1.B**), exhibiting the lowest relative intensity of most metabolites among the other two clusters. This result was interesting since the variance among individuals of small shrimp was more apparent compared to large shrimp. Grouping of 41/50-16/20, as displayed in the score plot, explained the similarity

of metabolite profiles regardless of shrimp days of culture, sampling location, and other environmental factors (**Figure 2.3.2-1.B**).

Once the influence from count size >70 and 61/70 was removed from PCA, a clear difference among medium to large shrimp was observed (**Figure 2.3.2-1.C-D**). Interestingly, the same clustering pattern was demonstrated as the samples formed two separate groups according to size difference. Although it was less significant (**Table 2.3.2-1**), PCA score plot suggested a further change of metabolite profile still occurred in the shrimp larger the average mass of 7g and no more apparent different observed once they are larger than the average mass of 9g (**Figure 2.3.2-1.C**). This separation was attributed by higher level of a higher accumulation of several sugars (fructose, psicose/tagatose, 1-6-anhydroglucose, sorbose, mannose, fructose-6-phosphate, and melibiose) in count size 51/60, and a higher level of organic acids (glycolic acid, xylonic acid, and nicotinic acid), dipeptide glycylglycine, glycerol, oleic acid, and cholesterol was observed in count size 26/30-16/20 (**Figure 2.3.2-1.D**). The visualization obtained from PCA unbiasedly demonstrated that the variance in metabolic profile can be explained by variable size, regardless the difference in days of culture, aquaculture systems, and sampling locations. Following this, through the construction of PLS model, the correlation between the two variables, size and metabolome data was examined.

2.3.3 Projection to Latent Square Regression model for predicting the size of white leg shrimp

In this study, PLS regression analysis was utilized to predict shrimp size using metabolome data, which act as a predictor. The data set was split into 2 parts, one is the training sample to estimate the parameter. The model results in the parameters that fitted the model, and it tells the estimated value of the Y variable. The ratio of fitted value against the original value

gives us the goodness of fit or R^2 and RMSE tells us the standard deviation error of estimation. Then, the new sample outside the original data was examined using the parameter obtained previously. This step is called prediction. The ratio of the predicted outcome against the actual outcomes or Q^2 tells how good is the prediction. RMSEP, the standard deviation of prediction error tells the predictive ability of the model. The threshold value of R^2, Q^2 should be closer to one, with the standard deviation error should be as low as possible to be considered a good model. The X variables that highly contributed to the model are suggested based on the VIP score. A stronger correlation between the X and Y variable is demonstrated by a higher VIP score. Therefore X variable with a VIP score higher than one is considered important

Figure 2.3.3-1 below showed PLS model built from training set consisted of nine groups of within the average mass 2-24 g (scale: auto-scale; *n*=3). Validation set were comprised of five groups with the range of body mass of 6-9 g (*n*=5).The vertical axis represents the relative intensity. The horizontal axis represents samples. Error bar shows standard deviation obtained from three biological replicates.

The PLS regression model was constructed by plotting the metabolome data matrix as the explanatory variable and shrimp size (**Table 2.2.1-1**) as the response variable. The parameters used to assess the constructed model are summarized in **Table 2.3.3-1**. As shown in **Figure 2.3.3-1.A**, a PLS model constructed from the training set resulted in R^2 and Q^2 values of 0.87 and 0.75, respectively. External validation was carried out to assess the robustness of the trained model using an independent data set. **Figure 2.3.3-1.B** showed the PLS model, which included both the training set and validation set (**Table 2.2.1-1**) with the resulting value of external Q^2 (correlation coefficient of the observed vs. predicted) of 0.76, RMSEE value of 2.294, and RMSEP value of 3.319. This result demonstrated good predictability of the PLS model according to the threshold values of R^2 and Q^2 higher than 0.6, and low RMSE values

(27). Despite the differences in days of culture and sampling locations, similar sizes of shrimps showed similar metabolite profiles. This result thereby strengthens our finding that metabolome data were able to predict shrimp size and a correlation was observed for the first time.

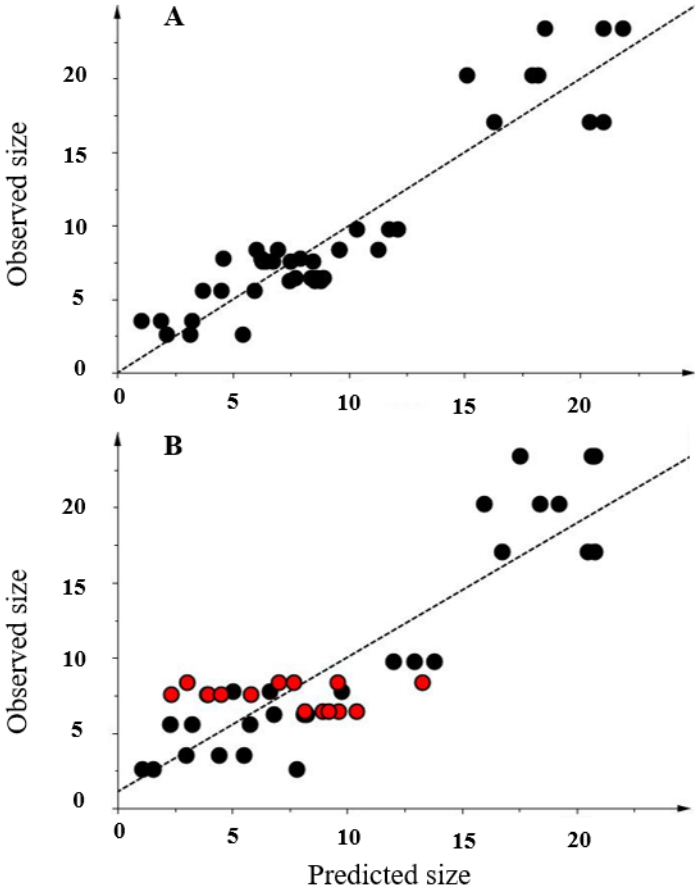


Figure 2.3.3-1 PLS regression model for the prediction of shrimp count sizes (A) prediction model built from training set. (B) prediction model built from training set and test set.

OPLS model built from training set consisted of eleven groups (scale: auto-scale; *n*=3); Black dots represents training set and red dots for validation set. The vertical axis represents the relative intensity. The horizontal axis represents samples. Error bar shows standard deviation obtained from three biological replicates.

One previous study attempted to correlate size with shrimp area, length, and mass utilizing machine vision to improve the grading efficiency in the industrial application (28). However, a study regarding shrimp size prediction based on metabolome data has not been done in other shrimp species, particularly using GC/MS.

Table 2.3.3-1. Summary of PLS parameters

Model	A	N	R^2X (cum)	R^2Y (cum)	Q^2 (cum)
M1	3	42	0.564	0.867	0.746
M2	2	27	0.497	0.856	0.756
RMSEE	2.294				
RMSEP	3.319				
RMSEcv	3.492				

"M1 and M2" is OPLS model constructed from training set and OPLS model constructed from training set and test set, respectively, "A" is number of multivariate component, "N" is number of samples, "R^2X" is the fraction of the variation of the X variables explained by the model, "R^2Y" is the fraction of the variation of the Y variables explained by the model, Q^2 is the fraction of the variation of the X and Y variables that denotes the prediction ability of the model. RMSEE: root-mean-square error of estimation RMSEE; RMSEcv: root-mean-square-error cross validation; RRMSEP: root-mean-square error of prediction

The VIP score was utilized as a variable cut-off to select the metabolites showing a high correlation with size. It summarizes the influence of individual metabolites (X variables), therefore suggesting metabolites that are strongly correlated with the size of shrimp (Y variables). From the complete list of VIP scores, 52 metabolites showed a VIP score greater than 1 (**Table 2.3.2-1**). These include amino acids, non-proteinogenic amino acids, organic acids, sugars, and sugar alcohol. In accordance with PCA result, among 52 significant VIP metabolites, only 5 metabolites showed a positive correlation with shrimp size, including phosphate as the 5 top highest VIP metabolites. Meanwhile, the majority of VIP metabolites such as amino acids, organic acids, and sugars demonstrated a negative correlation, suggesting the increase of size associated with the decrease in metabolite.

Table 2.3.2-1. List of metabolites annotated from nine groups of shrimp samples

Count size	>70		61/70		51/60	41/50	26/30	21/25	16/20	VIP	Coefficient
Average mass	2-3 g	3-4 g	5-6 g	6-7 g	7-9 g	9-11 g	16-18 g	19-22 g	23-24 g	score	plot
Metabolites	**Relative intensity**										
Alanine	1.192^{a}	1.291^{a}	1.092^{ab}	0.917^{abc}	0.972^{abc}	0.718^{bcd}	0.460^{d}	0.610^{cd}	0.386^{d}	1.629	-0.035
Malic acid	0.032^{cd}	0.044^{abc}	0.062^{a}	0.042^{bc}	0.057^{ab}	0.032^{cd}	0.008^{e}	0.016^{de}	0.007^{e}	1.610	-0.046
Fumaric acid	0.005^{ab}	0.007^{a}	0.007^{a}	0.003^{bc}	0.005^{ab}	0.003^{bc}	0.0009^{c}	0.001^{c}	0.0008^{c}	1.571	-0.038
Phosphate	0.198^{de}	0.152^{e}	0.407^{bcd}	0.477^{b}	0.453^{bc}	0.232^{cde}	0.778^{a}	0.558^{ab}	0.738^{a}	1.506	0.037
Argininosuccinic acid	0.0008^{ab}	0.0009^{ab}	0.0009^{ab}	0.0009^{a}	0.0005^{bc}	0.00053^{abc}	0.0003^{c}	0.0003^{c}	0.0004^{c}	1.496	-0.033
Glutamic acid	0.109^{a}	0.112^{a}	0.131^{a}	0.091^{ab}	0.108^{a}	0.093^{ab}	0.073^{ab}	0.073^{ab}	0.036^{b}	1.470	-0.036
Inositol	0.0035^{ab}	0.0036^{ab}	0.004^{a}	0.0032^{ab}	0.0029^{ab}	0.001^{d}	0.001^{d}	0.001^{d}	0.0016^{cd}	1.463	-0.029
4-Aminobutyric acid	0.003^{a}	0.003^{a}	0.002^{b}	0.0016^{c}	0.0012^{c}	0.0010^{d}	0.0010^{d}	0.0010^{d}	0.0007^{d}	1.454	-0.026
Threonine	0.037^{a}	0.029^{ab}	0.034^{a}	0.033^{ab}	0.017^{bc}	0.013^{c}	0.011^{c}	0.010^{c}	0.011^{c}	1.423	-0.022
Quinic acid	0.0004^{a}	0.0003^{ab}	0.0004^{a}	0.0003^{abc}	0.0003^{ab}	0.0003^{abc}	0.0002^{c}	0.00023^{bc}	0.00025^{bc}	1.418	-0.032
Histidine	0.059^{a}	0.045^{abc}	0.049^{ab}	0.046^{abc}	0.034^{bcd}	0.028^{cd}	0.022^{d}	0.019^{d}	0.028^{cd}	1.404	-0.017
β -alanine	0.005^{a}	0.006^{a}	0.0037^{b}	0.0034^{b}	0.0027^{bc}	0.0031^{bc}	0.0026^{bc}	0.0026^{bc}	0.0021^{c}	1.399	-0.024
Oxalacetic acid+Pyruvic acid	0.005^{a}	0.004^{ab}	0.003^{bc}	0.0026^{cd}	0.0015^{d}	0.002^{cd}	0.0012^{d}	0.0014^{d}	0.0014^{d}	1.396	-0.018
Ribose	0.007^{a}	0.006^{ab}	0.005^{abc}	0.003^{bcd}	0.002^{cd}	0.0007^{d}	0.0006^{d}	0.0006^{d}	0.001^{d}	1.375	-0.013
Cysteine	0.008^{a}	0.008^{a}	0.006^{abc}	0.006^{ab}	0.003^{bc}	0.003^{bc}	0.002^{c}	0.002^{c}	0.003^{bc}	1.345	-0.015
Sorbose	0.0017^{abc}	0.0013^{abc}	0.002^{a}	0.0019^{ab}	0.002^{a}	0.0009^{abc}	0.0006^{bc}	0.0005^{bc}	0.0003^{c}	1.343	-0.035
Valine	0.174^{a}	0.143^{ab}	0.157^{a}	0.159^{a}	0.087^{bc}	0.077^{bc}	0.085^{bc}	0.065^{bc}	0.079^{bc}	1.327	-0.016
Proline	0.960^{c}	1.072^{bc}	1.185^{abc}	1.209^{abc}	1.209^{abc}	1.377^{abc}	1.545^{a}	1.388^{abc}	1.461^{ab}	1.324	0.031

Aspartic acid	0.029^{ab}	0.034^{a}	0.022^{abc}	0.015^{abc}	0.037^{a}	0.005^{bc}	0.002^{c}	0.003^{c}	0.002^{c}	1.322	-0.030
Leucine	0.192^{a}	0.148^{ab}	0.140^{ab}	0.144^{ab}	0.079^{b}	0.061^{b}	0.067^{b}	0.059^{b}	0.064^{b}	1.321	-0.011
Lysine	0.221^{a}	0.150^{ab}	0.104^{bc}	0.102^{bc}	0.067^{bc}	0.041^{c}	0.037^{c}	0.034^{c}	0.047^{c}	1.317	-0.012
Methionine	0.033^{a}	0.026^{ab}	0.017^{bc}	0.197^{abc}	0.006^{c}	0.009^{c}	0.007^{c}	0.006^{c}	0.008^{c}	1.309	-0.011
Isoleucine	0.096^{a}	0.078^{ab}	0.073^{ab}	0.077^{ab}	0.042^{b}	0.036^{b}	0.039^{b}	0.035^{b}	0.038^{b}	1.309	-0.011
Uracil	0.006^{a}	0.005^{ab}	0.004^{bc}	0.003^{cd}	0.002^{cd}	0.002^{cd}	0.0017^{d}	0.0018^{d}	0.002^{cd}	1.306	-0.013
Fructose	0.002^{ab}	0.002^{ab}	0.003^{a}	0.002^{ab}	0.003^{a}	0.002^{ab}	0.0008^{b}	0.0007^{b}	0.0006^{b}	1.291	-0.032
Psicose+Tagatose	0.004^{ab}	0.004^{ab}	0.005^{a}	0.004^{ab}	0.005^{a}	0.002^{ab}	0.0016^{b}	0.0014^{b}	0.0011^{b}	1.281	-0.032
Xylitol	0.0003^{abc}	0.0004^{a}	0.00032^{ab}	0.0003^{abcd}	0.00016^{cd}	0.0002^{bcd}	0.00016^{d}	0.00017^{cd}	0.00017^{cd}	1.259	-0.016
Xanthine	0.0033^{a}	0.0033^{a}	0.0017^{b}	0.0015^{bcd}	0.0007^{e}	0.0016^{bc}	0.0008^{e}	0.001^{cde}	0.0009^{de}	1.250	-0.018
Acetoacetic acid	0.108^{a}	0.097^{a}	0.053^{b}	0.048^{b}	0.030^{b}	0.059^{b}	0.031^{b}	0.033^{b}	0.032^{b}	1.237	-0.019
Phenylalanine	0.067^{a}	0.048^{ab}	0.038^{ab}	0.038^{ab}	0.020^{b}	0.022^{b}	0.023^{b}	0.018^{b}	0.024^{b}	1.229	-0.004
Mannose	0.004^{abc}	0.004^{ab}	0.006^{a}	0.004^{ab}	0.005^{ab}	0.0013^{bc}	0.001^{c}	0.0008^{c}	0.0007^{c}	1.227	-0.027
2-Aminoethanol	0.008^{a}	0.008^{a}	0.004^{b}	0.004^{b}	0.003^{b}	0.002^{b}	0.003^{b}	0.002^{b}	0.002^{b}	1.222	-0.016
Tryptophan	0.019^{a}	0.013^{ab}	0.007^{bc}	0.009^{bc}	0.003^{c}	0.004^{c}	0.003^{c}	0.004^{c}	0.005^{bc}	1.203	-0.005
Inosine	0.032^{a}	0.032^{a}	0.015^{b}	0.016^{b}	0.012^{b}	0.015^{b}	0.012^{b}	0.012^{b}	0.012^{b}	1.202	-0.012
Thymine	0.001^{a}	0.001^{a}	0.0003^{b}	0.0002^{b}	0.0001^{b}	0.0001^{b}	0.0001^{b}	0.0001^{b}	0.0001^{b}	1.202	-0.010
Hydroxyproline										1.202	-0.034
1-Methylhistidine	0.0004^{ab}	0.0006^{a}	0.0005^{ab}	0.0003^{abc}	0.0004^{ab}	0.0003^{ab}	0.0003^{ab}	0.0002^{b}	0.00023^{ab}	1.187	-0.031
Elaidic acid	0.0004^{a}	0.0002^{ab}	0.0002^{ab}	0.0002^{ab}	0.0001^{b}	0.0001^{b}	0.0001^{b}	0.0001^{b}	0.0001^{b}	1.172	-0.004
Hypoxanthine	0.0180^{ab}	0.0183^{a}	0.010^{cd}	0.012^{bc}	0.006^{d}	0.010^{cd}	0.005^{d}	0.007^{cd}	0.009^{cd}	1.165	-0.018
Ethylmalonic acid	0.003^{a}	0.003^{ab}	0.0024^{c}	0.0025^{ab}	0.0022^{c}	0.0022^{c}	0.0022^{c}	0.0022^{c}	0.0022^{c}	1.160	-0.015

n-Acetyl ornithine	0.003^{a}	0.002^{ab}	0.001^{bc}	0.001^{bc}	0.0006^{c}	0.001^{bc}	0.001^{bc}	0.0007^{c}	0.0007^{c}	1.147	-0.006
Guanosine	0.002^{a}	0.001^{ab}	0.0004^{bc}	0.0002^{c}	0.0003^{c}	0.0002^{c}	0.0002^{c}	0.0001^{c}	0.0001^{c}	1.139	-0.007
Glycolic acid	0.00048^{bc}	0.0005^{abc}	0.0003^{c}	0.00047^{bc}	0.00037^{c}	0.0007^{ab}	0.0008^{a}	0.0007^{ab}	0.0006^{abc}	1.132	0.022
Urocanic acid	0.0008^{a}	0.0008^{a}	0.0002^{b}	0.0001^{bc}	0.00006^{c}	0.0002^{bc}	0.0001^{bc}	0.0001^{bc}	0.0001^{bc}	1.128	-0.011
Epinephrine	0.001^{a}	0.0004^{ab}	0.0005^{ab}	0.0002^{b}	0.0001^{b}	0.0002^{b}	0.0001^{b}	0.0001^{b}	0.0001^{b}	1.109	0.001
Melibiose	0.002^{a}	0.0003^{b}	0.0013^{ab}	0.0011^{ab}	0.0005^{b}	0.0002^{b}	0.0001^{b}	0.0001^{b}	0.0001^{b}	1.108	-0.011
Tyrosine	0.159^{a}	0.115^{ab}	0.090^{ab}	0.097^{ab}	0.043^{b}	0.063^{b}	0.072^{b}	0.050^{b}	0.091^{ab}	1.090	0.009
1,3-Propanediamine	0.002^{a}	0.002^{a}	0.0003^{b}	0.0003^{b}	0.0003^{b}	0.0006^{b}	0.0004^{b}	0.0004^{b}	0.0002^{b}	1.088	-0.010
Asparagine	0.043^{ab}	0.037^{ab}	0.05^{a}	0.044^{ab}	0.041^{ab}	0.031^{ab}	0.033^{ab}	0.025^{b}	0.029^{ab}	1.088	-0.020
Uric acid	0.002^{a}	0.001^{ab}	0.0007^{b}	0.00087^{b}	0.0006^{b}	0.0009^{ab}	0.0007^{b}	0.0006^{b}	0.0005^{b}	1.011	-0.009
Serine	0.075^{ab}	0.058^{ab}	0.114^{a}	0.106^{a}	0.058^{ab}	0.042^{b}	0.038^{b}	0.036^{b}	0.043^{b}	1.000	-0.019
Taurine	0.029^{a}	0.024^{ab}	0.029^{a}	0.024^{ab}	0.001^{c}	0.022^{abc}	0.019^{abc}	0.014^{bc}	0.020^{abc}	0.890	-0.003
Malonic acid	0.0005^{a}	0.00049^{ab}	0.00024^{bc}	0.00022^{c}	0.00024^{c}	0.0003^{abc}	0.00025^{bc}	0.00024^{c}	0.0002^{bc}	0.873	-0.005
Glycylglycine	0.001^{a}	0.001^{a}	0.0007^{bc}	0.0005^{bc}	0.0003^{c}	0.0007^{b}	0.0005^{bc}	0.0006^{bc}	0.0007^{bc}	0.852	0.002
Homoserine	0.0003^{c}	0.0003^{c}	0.0003^{c}	0.00035^{bc}	0.00037^{bc}	0.0005^{ab}	0.0006^{a}	0.0005^{ab}	0.00034^{bc}	0.836	0.008
N-Acetyl glucosamine	0.004^{a}	0.0010^{ab}	0.0010^{ab}	0.0003^{b}	0.0002^{b}	0.0016^{ab}	0.0002^{b}	0.0002^{b}	0.0013^{ab}	0.824	0.011
Sarcosine	0.0014^{ab}	0.0018^{a}	0.0004^{b}	0.0005^{b}	0.0007^{b}	0.0007^{b}	0.0007^{b}	0.0007^{ab}	0.0005^{b}	0.759	-0.005
Trehalose	0.011^{b}	0.008^{b}	0.603^{ab}	0.406^{ab}	0.728^{a}	0.152^{ab}	0.696^{a}	0.318^{ab}	0.623^{ab}	0.710	0.013
Hypotaurine	0.0017^{ab}	0.002^{a}	0.0007^{cd}	0.0005^{d}	0.0007^{cd}	0.0008^{cd}	0.0007^{cd}	0.0007^{cd}	0.001^{bc}	0.708	0.005
Ornithine	0.0006^{ab}	0.0006^{a}	0.0004^{abc}	0.0003^{abc}	0.0002^{c}	0.0004^{abc}	0.0003^{bc}	0.0004^{abc}	0.0004^{abc}	0.615	0.000

Result of one-way ANOVA followed by Tukey's HSD test, ($p < 0.05$). Means within a column that do not share superscript letter (a, b, c, d, e) are significantly different. ($p < 0.05$).

2.3.4 The significance of size in molecular quality

From both PCA and PLS results, we can observed different stages of metabolite change from shrimp juvenile until sub-adult. Although some additional data would be needed, we hypothesize that the separation trend might also imply the difference in shrimp metabolism. As shown in **Table 2.3.2-1,** a distinct change was started from the average mass 2-4 g (count size >70 to 61/70) for most of the metabolites. A declined in metabolite accumulation extended to a larger size for several amino acids (alanine, 4-aminobutyric acid, histidine, cysteine, aspartic acid, tryptophan, acetyl-ornithine), sugars (ribose mannose xylitol), purine and pyrimidine derivatives (xanthine, uracil), and organic acids (oxaloacetic acid + pyruvate, malic acid). Following this, the metabolic change was remained insignificant once the size approximately reached average mass of 9-11 g (count size 41/50).

Metabolite changes along with growth were previously reported in a study using NMR-based metabolomics to assess the health of white leg shrimp cultured in a super-intensive system. This study noted that young shrimps undergo rapid metabolism, suggesting the allocation of energy for growth (29). In this study, high level of sugars and organic acids including malic acid and fumaric acid might be associated with high energy demand during growth. Furthermore, the decrease in free amino acids and nucleotides level was the indicative that they were being utilized for protein retention and nucleic acid synthesis during growth period (29). Amino acids was continuously utilized as building blocks for protein synthesis, muscle tissue formation leads to an increase in shrimp mass.

2.4 Conclusion

There was a significant difference in metabolite accumulation observed in different sizes of HLSO shrimps, suggesting the importance of size as seen through metabolic profile. A distinct change was started from 2-4 g (count size >70 to 61/70) for most of the metabolites, including amino acids, sugars, organic acids, and phosphate. Metabolite changes was less significant once the size reach commercial size range, with the average mass > 11 g (count size 31/40). Through the construction of PLS model, metabolome data was able to predict the size, suggesting a correlation between these two variables is plausible. Among 52 VIP metabolites derived from the model, only phosphate that showed the highest importance with a positive correlation with size.

Although general molecular mechanism towards growth is already known, this is the first study to report a correlation between metabolome data and shrimp size, as the manifestation of growth. Furthermore, this findings was still lacking on a meaningful discussion regarding the importance of phosphate, as a positive predictor of size. As this study aimed to justify the overall quality of shrimp according to the molecular quality of shrimp, further analysis is required to target a wider range of metabolite coverage, specifically phosphate-related compounds.

Chapter 3

Targeted-metabolomics analysis for the prediction model improvement

3.1 Introduction

Previous chapter demonstrated the significance of shrimp size in farmed white leg shrimp by means of metabolomics approach. Through the utilization of GC/MS analysis, low-molecular-weight compounds that are known to involve in central carbon metabolism can be discussed. In this chapter, we aimed to increase the number of explanatory variables that reflect shrimp quality traits through the construction of prediction model. As one of the main challenge in metabolomics field, it is not possible to achieve a complete metabolite profiling simultaneously in a single instrument. Many factors affect metabolite recovery, depending on the functional group of the metabolite (13). In that case, several classes of polar and acidic metabolites, such as nucleotides, coenzyme, sugar nucleotides, and sugar phosphates cannot be obtained by GC/MS analysis, owing to its thermo labile nature. Since it is critical to obtain good retention and peak-shape for the above-mentioned metabolite class, ion-pair agent was utilized instead of a standard reversed-phase LC/MS.

In this chapter, the analysis was performed by both GC/MS and ion-pair reversed-phase LC/QqQ/MS instruments. The prediction power of previously built PLS regression model was further examined by adding more samples, sample variability, and number of explanatory variables

into the model. To make the implementation of the study more relevant, external validation was carried out using a commercial shrimp samples obtained from the local Japanese markets.

3.2 Materials and method

3.2.1 Sample information

From the total of 16 size categories, 13 samples were obtained from three different regions in Indonesia as described in chapter 2. Meanwhile, the validation samples were purchased from the local stores in Japan. These samples were collected as HLSO (head-less-shell-on) shrimp. The range of shrimp count size was decided according to the availability of samples during the purchasing day. All of the shrimp samples came with the count size and origin information. Any parameter regarding culture system or salinity was unknown unless it was provided in the packaging. Before extraction, all samples were prepared as described in chapter 2. Before the analysis, the uniformity of shrimp size was re-assessed against the label claim and the sample was re-classified according to the count size conversion table (**Table 1.3-1**).

Table 3.2.1-1 List of additional samples from commercial Japanese market

PCA data set				
Count size	Average mass (n=3)		Location-Province	System
<70	2.62	g	Buleleng-Bali	Semi-intensive
<70	3.52	g	Buleleng-Bali	Semi-intensive
61/70	5.59	g	Buleleng-Bali	Semi-intensive
61/70	6.23	g	Subang-West Java	Intensive
51/60	7.76	g	Subang-West Java	Intensive
41/50	9.77	g	Buleleng-Bali	Semi-intensive
26/30	17.06	g	Buleleng-Bali	Semi-intensive

21/25	20. 24	g	Buleleng-Bali	Semi-intensive
16/20	23.43	g	Buleleng-Bali	Semi-intensive
41/50	11.24	g	Banten-Banten	Semi-intensive
31/40	13.17	g	Banten-Banten	Semi-intensive
31/40	13.48	g	Banten-Banten	Semi-intensive
26/30	15.85	g	Banten-Banten	Semi-intensive
21/25	17.9	g	Banten-Banten	Semi-intensive
31/40	13.62	g	India	Unknown
26/30	16.48	g	India	Unknown
21/25	20.61	g	India	Unknown

PLS data set					
Farmed white leg shrimp from Indonesia					
Count size	Average mass (n=3)		Location-Province	System	Data set
<70	2.62	g	Buleleng-Bali	Semi-intensive	Training set
<70	3.52	g	Buleleng-Bali	Semi-intensive	Training set
61/70	5.59	g	Buleleng-Bali	Semi-intensive	Training set
61/70	6.23	g	Subang-West Java	Intensive	Training set
51/60	7.76	g	Subang-West Java	Intensive	Training set
41/50	9.77	g	Buleleng-Bali	Semi-intensive	Training set
26/30	17.06	g	Buleleng-Bali	Semi-intensive	Training set
16/20	23.43	g	Buleleng-Bali	Semi-intensive	Training set
41/50	11.24	g	Banten-Banten	Semi-intensive	Training set
31/40	13.17	g	Banten-Banten	Semi-intensive	Training set
26/30	15.85	g	Banten-Banten	Semi-intensive	Training set
Commercial white leg shrimp from Japanese market					
Count size	Average mass		Origin	System	Data set
31/40	13.62	g	India	Unknown	Test set
26/30	16.48	g	India	Unknown	Test set
21/25	20.61	g	India	Unknown	Test set

3.2.2 Sample preparation method for GC/MS analysis

The extraction method for GC/MS analysis was re-optimized to better remove the protein and to prevent saturated peaks as the potential damage for the machine in the long run. The optimized extraction method was carried out in a step-wise manner using 80% ethanol. First, 200 µl of water containing internal standard ribitol was added to 5 mg of shrimp powder, vortex, and homogenized for 10 minutes at 25°C. Subsequently, 800 µl ethanol as a deproteinization agent was added, so that the final concentration of ribitol was 50 µg/L. The mixture was then centrifuged at 4°C, 10000 rpm for 10 minutes to allow the precipitation of protein. 700 µl of supernatant was transferred to 1.5 ml Eppendorf microtube, of each 100 µl was collected and pooled for QC sample. From the collected supernatant, 100 µl of sample and pooled QCs were transferred into a new 1.5 ml Eppendorf tube with a screwed cap and further subjected to vacuum centrifugation for 10 minutes. Following the removal of solvent, samples were lyophilized (VD-800F Freeze Dryer, Taitec, Tokyo, Japan) overnight. All samples were analyzed in triplicate (n=3).

3.2.3 Sample preparation method for ion-pair reversed-phase LC/QqQ/MS analysis

The raw data in format were processed by Shimadzu Post-run Analysis software (Shimadzu, Japan). The peak detected was matched against the retention time of standard compounds. Peak annotation was done manually with the minimum S/N ratio < 5 and peak area < 5000. The relative intensity of metabolite is derived from the annotated peak area. Normalization by internal standard pe Bligh-Dyer extraction method was carried out to eliminate the lipid content prior to LC/MS analysis. The extraction was performed using a mix of methanol, water, and chloroform solvent in the ratio of 5:2:2. Internal standard [(+) 10-camphorsulfonic acid] with a known concentration of 0.1mg/mL was added into the mixed solvent prior to extraction. A total of 15 mg shrimp powder

was homogenized together with 1 ml of solvent for 30 minutes at 370C. Following 5s of centrifugation, 650 µl of the sample was transferred to a new 1.5 ml Eppendorf microtube and 250 µl of ultrapure water was added subsequently. The hydrophilic phase was then separated through centrifugation for 5 minutes at 9000xG; 4C. 700 µl of supernatant were taken from the upper phase, passed through a filter syringe, and transferred to a new microtube. From each sample, 200 µl of filtered supernatant and pooled QC were subjected to dry centrifugation for 30 minutes followed by lyophilization overnight. Prior to the analysis, the sample was diluted with water and transferred to a glass vial for LC/MS analysis.

3.2.4 Ion-pair reversed-phase LC/QqQ/MS analysis

The analysis was performed on LC/MS 8050 (Shimadzu, Japan), operating in negative ionization mode, using MRM mode. Analytes' separation was carried out using L-column 2ODS (150 nm × particle size 3 µm, Chemicals Evaluation, and Research Institute, Tokyo, Japan), maintained at 45ºC with an injection volume of 1 µL. The mobile phases have consisted of A, 10 mM ion pair reagent tributylamine dissolved in ultra-pure water containing 15 mM acetate and B, methanol. The pump was operated at a flow rate of 0.2 mL/min; start from 0% B at 0-1.0 min, increased to 15% B at 1.0 to 1.5 min, then held until 3.0 min, then increased to 50% and 100% B from 3.0 to 8.0 min, and 8.0 to 10.0 min, respectively; held 100% B until 11 min, decreased to 0% B from 11 to 11.50 min, and at last held 0% B until 17 min. After each injection, the needle was washed with 0.1% formic acid in H2O/MeCN/MeOH (2:1:1) and purged mobile phase A. For the mass spectrometer conditions, the desolvation line temperature was 250 °C, the nebulizer gas flow was 2 L/min, the drying gas flow was 15 L/min, and the heat block temperature was 400 °C. All metabolites' abundance was normalized by the area value of internal standard [(+) 10-camphorsulfonic acid].

A mix of standard compounds was initially run to check the machine condition and the shifting of peak against the standard. Following this, a series of diluted QC samples (1, 10, and 100 times) was injected to check the dose-response and peak saturation. The samples were diluted 100 times according to the result of QC samples and analyzed in randomized order with a water-sample-water sequence to minimize carryover. QC samples run after seven-time of injections to monitor any possible carry-over and sensitivity changes during analysis.

The raw data in format were processed by Shimadzu Post-run Analysis software (Shimadzu, Japan). The peak detected was matched against the retention time of standard compounds. Peak annotation was done manually with the minimum S/N ratio < 5 and peak area < 5000. The relative intensity of metabolite is derived from the annotated peak area. Normalization by internal standard performed after the raw data was exported to Ms. Excel and further subjected to multivariate analysis software.

3.2.5 Multivariate analysis

Multivariate analysis was performed using SIMCA-P+ software version 11 (Umetrics, Sweden). Metabolome data matrix from both instruments was integrated by selecting the most dominant metabolite coverage from each platform. All of the amino acid data came from GC/MS analysis except for arginine. Fatty acids and non-proteinogenic amino acids are unique to GC/MS. Meanwhile, nitrogenous-based compounds, organic acids, and heterocyclic compounds were obtained from ion-pair reversed-phase LC/QqQ/MS analysis. The data was auto-scaled to unit variance (UV), giving all the variables equal importance. Unsupervised statistical analysis, Principal Component Analysis (PCA), aims to extract the information and visualize the sample behavior by maximizing the variance (23). Therefore, PCA enables the extraction of hidden information, suggesting the most important factor contributing to the sample clustering. Following

this, projection to orthogonal latent structure (OPLS) regression analysis aimed to examine the ability of metabolome data in predicting the size of shrimp. Samples that share similar average body mass were excluded from the data set. The performance of the model was examined according to R^2, Q^2, which denote the goodness of fit and prediction ability, respectively. The accuracy of the model was explained by the root-mean-square of estimation (RMSEE) derived from the model, while the root-mean-square of prediction (RMSEP) indicates the predictability of the model (24). Cross-validation was performed to estimate the ability of a model to correctly predict the response matrix Y of new individuals (25). The training model was built from shrimp samples obtained from the farm while the validation set used commercial shrimp obtained from the market.

3.3 Result and discussion

3.3.1 The significance of size in commercial white leg shrimp

Before OPLS model construction, the separation tendency from the complete set of white leg shrimp was examined. Ideally, to conduct a validation test using commercially purchased shrimp, the data set should cover a wide range of size. However, due to the shrimp availability issue, we focused on commercial HLSO shrimp with the count size smaller than 51/60.

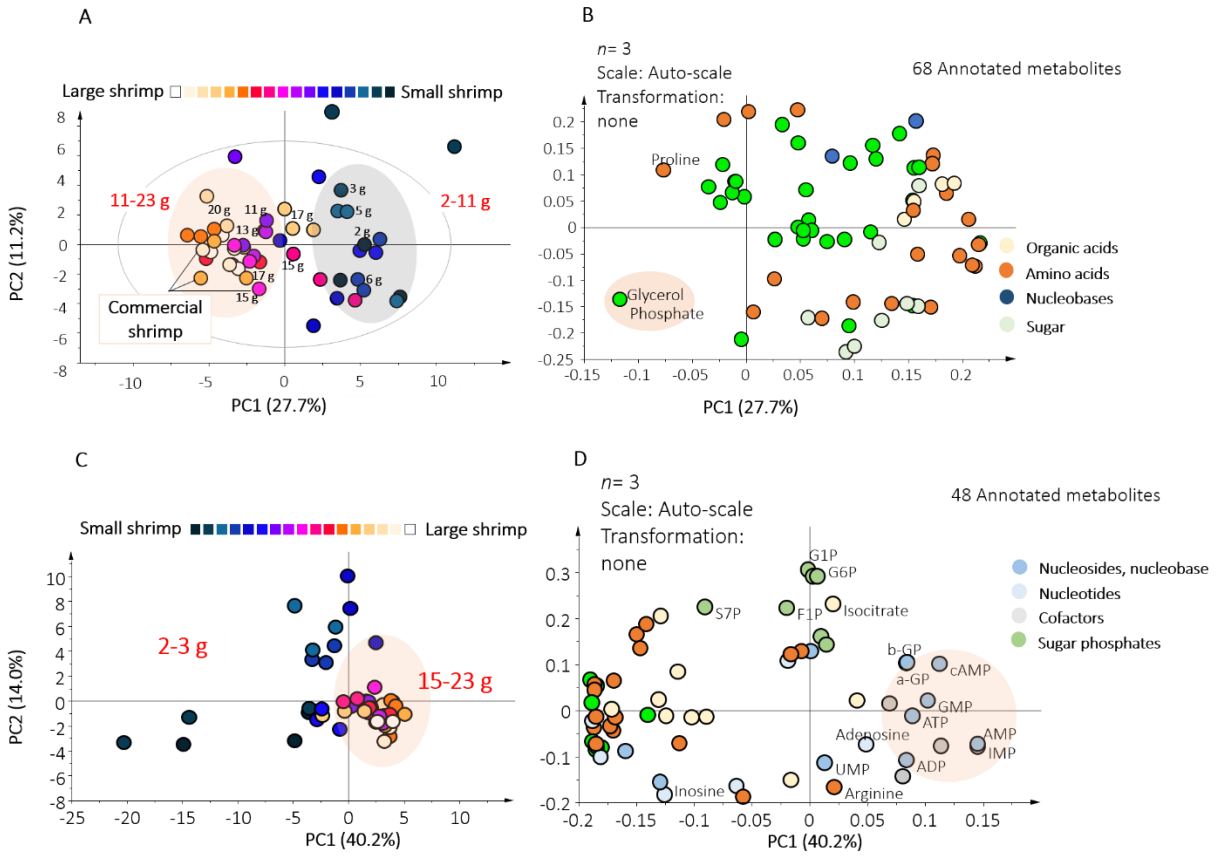


Figure 3.3.1-1. The PCA result of GC/MS and ion-pair reversed-phase LC/QqQ/MS analysis for 14 different sizes of white leg shrimp samples (*n*=3; Scale: autoscale; Transform: none).

(A) PCA score plot derived from GC/MS analysis. Each dot represents shrimp sample. Colour gradation from dark to light indicate the shrimp size from small to large. The number written near the cluster describe the count size. (B) Loading plot represent the contribution of metabolites to the separation of samples. Blue colour represent the cluster of phosphate which associated with large shrimp samples. (C) PCA score plot derived from ion-pair reversed-phase LC/QqQ/MS (D) Loading plot of shrimp samples. Blue colour represent the cluster of phosphate-related compounds which associated with large shrimp samples.

The result of GC/MS and LC/MS analysis clearly showed a similar tendency of sample clustering (**Figure 3.3.1-1.A and B**). The similarity of metabolite profile was observed in all large shrimp, as they formed a tight cluster regardless of the difference in sample origin **(Figure 3.3.1-1.A)**. Although some samples with the same count size scattered, the clustering pattern was still

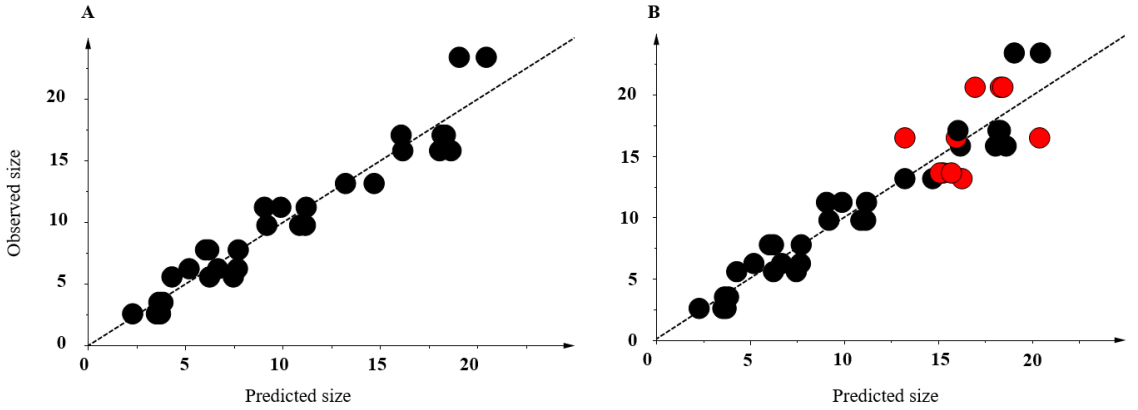
reproducible (**Figure 3.3.1-1.B**). Interestingly, the result obtained from IP-RP LC/QqQ/MS analysis data had a higher total variance of the first two principal components. The loading plot (**Figure 3.3.1-1.B**) showed, that the phosphate acquired from GC/MS has corresponded to nucleotides, nucleosides, and sugar phosphates detected from IP-RP LC/QqQ/MS. Therefore, these phosphate-related compounds contributed to a distinct difference between all large shrimps (count size 31/40-13/15) and small shrimps (count size >70-41/50). Moreover, high similarity among large shrimp caused the variation among small shrimp to become more apparent. This result corroborated with one previous study, where shrimp face a critical phase during the juvenile stage (29). It might also suggest small shrimp has a diverging molecular mechanism to adapt to the environment .

This result was interesting since the phosphate-related compounds strongly differed in the sample according to the size. Therefore, the expansion of analysis enables us to discover characteristic metabolites that cannot be discussed by only GC/MS platform.

3.3.2 Validation of size-based prediction OPLS regression model

This chapter aimed to investigate the robustness of metabolome data to predict the size of white leg shrimp by increasing the number of samples, sample variability, and metabolite coverage. Once the separation tendency was confirmed in PCA, the data set was further subjected to OPLS analysis. As supervised analysis is prone to bias, confirming the biological significance in an unbiased manner can be one way to prevent over-fitting. Before the construction of the model, the sample outlier (count size >70) was removed. A total of 102 metabolites were plotted as

explanatory variables, where 29 metabolites were both found from two instruments, 36 metabolites were unique to IP-RP LC/QqQ/MS, and 47 metabolites were unique to GC/MS.



Model	A	N	R2X (cum)	R2Y (cum)	Q2 (cum)	RMSEE	RMSEcv	RMSEP
M1	1+1+0	40	0.366	0.909	0.847	1.648	2.064	
M2	1+1+0	30	0.362	0.927	0.872			2.617

Figure 3.3.1-2. Sized-based OPLS model of white leg shrimp. (A) OPLS model built from training set. (B) OPLS model built from training set and test set

OPLS model built from training set consisted of eleven groups (scale: auto-scale; *n*=3); Black dots represents training set and red dots for validation set. The vertical axis represents the relative intensity. The horizontal axis represents samples. Error bar shows standard deviation obtained from three biological replicates. "M1 and M2" is OPLS model constructed from training set and OPLS model constructed from training set and test set, respectively, "A" is number of multivariate component, "N" is number of samples, "R^2X" is the fraction of the variation of the X variables explained by the model, "R^2Y" is the fraction of the variation of the Y variables explained by the model, Q^2 is the fraction of the variation of the X and Y variables that denotes the prediction ability of the model. RMSEE: root-mean-square error of estimation RMSEE; RMSEcv: root-mean-square-error cross validation; RRMSEP: root-mean-square error of prediction (30).

Figure 3.3.1-2. showed the constructed OPLS model. The model was autofitted means to fit the model according to the cross-validation rules. The RMSEE and RMSEP were 1.837 and 2.586, respectively. The values of R^2X (cum), R^2Y (cum), and Q^2Y (cum) were 0.362; 0.927; 0.872, respectively. R^2X(cum) explain the total amount of explained X-variation, R^2Y(cum) explain the

total amount of explained Y-variation, and $Q^2Y(cum)$ for the total amount predicted Y-variation. The model showed the threshold values of R^2 and Q^2 higher than 0.6, and highly disparate between R^2 and Q^2 were not observed (27). OPLS model which is designed to handle the orthogonal variation of X variable was utilized in replacement of conventional PLS model. The current RMSE and R^2Y values were improved even though the sample variability was increased. Therefore, a robust-sized-based prediction model was achieved through the expansion of metabolite coverage.

3.3.3 Size-dependent metabolites

Table 3.2.1-2 summarized VIP metabolites with the value > 1.0. These VIP metabolites were considered as the most relevant variables to predict the shrimp size, with a higher score indicates a higher contribution to the model. Nucleotides IMP and AMP, and cofactor NAD showed the highest VIP score of 1.8. Following this, orotate, malate, fumarate, orotate, alanine, and glutamic acid were the top VIP metabolites with a score > 1.5. Interestingly, IMP, AMP, and NAD as the three highest VIP metabolites had a positive correlation with shrimp size, suggesting the level of metabolites in the shrimp muscle increase along with the increase of size.

Table 3.2.1-2 List of VIP metabolites derived from size-based OPLS model of white leg shrimp

Pathway (KEGG)	Metabolites	VIP score	Correlation
Purine metabolism	IMP	1.88	+
Purine metabolism	AMP	1.88	+
Oxidative phosphorylation	NAD	1.83	+
Pyrimidine metabolism	Orotic acid	1.66	−
TCA cycle, pyruvate metabolism	Malic acid	1.63	−
TCA cycle; arginine biosynthesis; alanine, aspartate and glutamate metabolism	Fumaric acid	1.62	−
Alanine, aspartate and glutamate metabolism	Alanine	1.61	−
Arginine biosynthesis; alanine, aspartate and glutamate metabolism	Glutamic acid	1.58	−
Arginine biosynthesis; alanine, aspartate and glutamate metabolism	Aspartic acid	1.48	−
Lysine biosynthesis	Lysine	1.46	−
Glycine, serine and threonine metabolism; valine, leucine and isoleucine biosynthesis	Threonine	1.42	−
Galactose metabolism	Inositol	1.42	−
Cysteine and methionine metabolism	Methionine	1.41	−
Valine, leucine, and isoleucine degradation	Leucine	1.40	−
Glycolysis / gluconeogenesis, TCA cycle, pentose phosphate pathway	Pyruvic acid	1.39	−
--	Norvaline	1.36	−
Valine, leucine, and isoleucine degradation	Isoleucine	1.35	−
Arginine and proline metabolism	Hydroxyproline	1.35	−
Valine, leucine and isoleucine degradation	Valine	1.34	−
Arginine biosynthesis, Arginine and proline metabolism	Arginine	1.34	+
Purine metabolism	GMP	1.34	+
Pyrimidine metabolism	UMP	1.32	+
Pyrimidine metabolism	Uridine	1.30	−
Pyrimidine metabolism	Thymine	1.28	−

Oxidative phosphorylation, Purine metabolism	ADP	1.27	+
Purine metabolism	Xanthine	1.27	–
Steroid biosynthesis	Cholesterol	1.27	–
Glycine, serine and threonine metabolism, Tryptophan metabolism, Phenylalanine, tyrosine and tryptophan biosynthesis	Tryptophan	1.24	–
Glycine, serine and threonine metabolism, Cysteine and methionine metabolism	Serine	1.21	–
Purine metabolism	Guanosine	1.20	–
Purine metabolism	Hypoxanthine	1.19	–
Fructose and mannose metabolism, Galactose metabolism	Fructose	1.18	–
Oxidative phosphorylation	NADH	1.18	+
Pentose and glucuronate interconversions, Galactose metabolism	Glycerol	1.17	+
Oxidative phosphorylation	Phosphate	1.13	+
Thiamine metabolism	TMP	1.12	–
Histidine metabolism, beta-Alanine metabolism	Histidine	1.09	–
Phenylalanine metabolism, Phenylalanine, tyrosine and tryptophan biosynthesis	Phenylalanine	1.09	+
Oxidative phosphorylation, Purine metabolism	ATP	1.08	+
	Psicose+Tagatose	1.06	–
Pentose phosphate pathway	S7P	1.06	–
	a-GP	1.05	+
Fructose and mannose metabolism, Galactose metabolism	Mannose	1.05	–
	b-GP	1.03	+
Citrate cycle (TCA cycle), Alanine, aspartate and glutamate metabolism	Citric acid	1.01	–
Purine metabolism	Uric acid	1.00	–

IMP: Inosine monophosphate; AMP: Adenosine monophosphate; NAD: nicotinamide adenine dinucleotide; GMP: Guanosine monophosphate; UMP: Uridine monophosphate; ADP: Adenosine diphosphate; NADH: nicotinamide adenine dinucleotide hydride; TMP: Thymidine monophosphate; ATP: Adenosine triphosphate; S7P: sedoheptulase-7-phosphate; b-GP: Beta-Glycerophosphate; a-GP: alpha- Glycerophosphate

Primary metabolism is all interconnected. The product of metabolism facilitates cellular respiration, the formation of building blocks, and energy formation. Therefore, it is worth noting that the biological discussion of metabolomics data is not always straightforward.

A previous study reported that IMP and AMP as purine derivatives provide a cell with the necessary energy and cofactors to promote cell survival and proliferation (31). However, studies in crustaceans showed, enzymatic degradation of ATP-related compounds also causes the accumulation of IMP. It was also possibly sourced from the bacterial activities following the death (32,33). Considering the treatment are all homogenous, it is hard to assume that the accumulation was due to the variation in handling. Purine metabolism might involve different regulations throughout development. The accumulation of IMP and AMP parallel with the increase of shrimp size might indicate the requirement of purine metabolites at a specific life stage (34). High accumulation of purine metabolites in shellfish, especially shrimp, has been reported elsewhere. However, this the first study to report significant changes in IMP and AMP throughout different shrimp sizes.

Another important metabolite, NAD, also demonstrated a positive correlation with size. NAD, in the form of NAD+, involved in the oxidative phosphorylation process, playing a vital role in energy production. NAD levels influence a myriad of cellular processes and a key player in skeletal muscle development, regeneration, aging, and disease. NAD+ involve in most of the metabolic pathways, thereby demonstrates the ability of NAD+ to respond dynamically to physiological stimuli (35). The accumulation of NAD suggested a high demand for NAD+ as a

co-substrate in many cellular reactions which mechanistically link to metabolism/energy availability (36). This result might provide new insight into the underlying growth mechanism in shrimp.

Consistent with the result obtained in chapter 2, fumaric acid and malic acid showed a negative trend upon shrimp size. Among other organic acids, these metabolites had the highest VIP score and both are directly linked to the TCA cycle. The declining level of TCA cycle metabolites possibly affects the maintenance of cellular homeostasis and biosynthesis of macromolecules such as nucleotides, lipids, and proteins (37). Furthermore, orotate, an essential intermediate of pyrimidine de novo synthesis (38); and three amino acids linking to protein synthesis and energy metabolism such as alanine, aspartic acid, and glutamic acids (39) also showed negative correlation with the shrimp size.

3.3.3.1 Energy expenditure across different size of shrimp

A distinct metabolic perturbation following the increase in size might indicate a prominent mechanism during shrimp life, such as growth. The changes in size as a result of growth might be relevant to energy expenditure. In general, living systems distribute the energy acquired via food intake into various vital activities such as growth, organ maintenance, repair, defense, and reproduction (40). Interestingly, a previous report noted that the partitioning of the ingested energy between these processes might depend on variables such as shrimp life stage (41). Therefore, this study suggests the metabolite alteration reflects the mechanism of how shrimp utilize and allocate the energy to support vital activities as a consequence of growth. This hypothesis may explain significant changes in metabolite accumulation once shrimp reached average size 9g or approximately from count size 41/50 to count size 31/40. In such a case, VIP metabolites derived from the OPLS model can be regarded as size-dependent. Size-dependent metabolites characterize

shrimp samples according to their size difference. The accumulation of these metabolites was consistent across various shrimp samples, regardless of the variation in days of culture or origin.

3.4 Conclusion

The expansion of explanatory variables contributed to an improved model predictability and enabled a deeper discussion that cannot be made using data matrix from GC/MS. From this result, the metabolite alteration across different size might reflect the mechanism on how shrimp utilize and allocate the energy to support vital activities. Therefore, the VIP metabolites derived from the OPLS model were regarded as size-dependent metabolites as they demonstrated the same accumulation trend regardless the difference of origin, days of culture, sampling locations, and other confounding factors. This finding successfully demonstrated the significant of size as seen through metabolite profile. Although the grouping of count size uses a wider range of shrimp body mass, the clustering of sample according to metabolite profile still can be discussed under the term count size. Thereby, suggesting high relevancy and applicability in the shrimp industry.

Chapter 4

The significance of shrimp size in sensory properties

4.1 Introduction

Previous studies have demonstrated a strong correlation between metabolite composition and sensory perception (42,43). Primary metabolites including amino acids, sugars, sugar alcohols, and organic acids have been reported to correlate with sensory attributes (44–46). Therefore, the metabolite matrix acquired from the analysis also reflects the quantitative difference of taste-active metabolites. These taste-active compounds impart five basic taste qualities and thus are responsible for the taste perception of food.

As the significance of size in metabolome data is plausible, this result successfully demonstrated the potential of size as a quality evaluation parameter. In the last chapter, sensory evaluation was performed to confirm whether or not the size of the shrimp affects shrimp sensory quality as seen through the metabolite profile. Such information can demonstrate the utilization of metabolomics in providing indispensable information for both industry and consumer.

From the metabolomics perspective, the accumulation of size-dependent metabolites and taste-active compounds was also examined. According to the PCA result in chapter 3, significant changes of metabolite abundance were observed mostly between the average mass of 7 and 11 g. The data from sensory evaluation and metabolomics was then compared to characterize the sensory profile Therefore, once the difference among attributes was confirmed in In this chapter, the size range of the samples was narrowed down to 11-18 g, representing commonly traded commercial size in shrimp market. In chapter 2, it was reported that the metabolite profile of shrimp large

Shrimp samples was purchased from commercial shrimp farming in Banten province, Indonesia. Sampling was carried out according to the size difference and obtain from the same pond. This is to ensure that the variances observed in the analysis were due to the differences in size and were not biased by other influencing factors.

4.1.1 Sensory descriptive analysis

One of the sensory methodologies is a descriptive analysis which provides a piece of detailed information and quantitative descriptors of the products. Quantitative descriptors are very useful as a basis for measuring the effects of ingredients; it identifies product attributes that are most important to consumer preferences and provide a basis for linking consumer response behavior to specific product characteristics (47). These merits make sensory descriptive analysis suitable for quality evaluation. However, descriptive analysis remains challenging since it cannot be performed by the mass consumer. A group of qualified subjects called trained panelists is necessary as these people are trained on specific samples and able to quantify the descriptors based on perceptions. The training of panelists takes about 4 months until 1 year to produce a robust panelist (48).

4.2 Material and methods

4.2.1 Sample information of quantitative descriptive analysis (QDA)

Shrimp samples for sensory evaluation were purchased from the same pond in Banten, Indonesia. Samples were harvested partially to obtain different count sizes of shrimps. The early collected shrimps were kept at -30°C until all the samples were completed. The sample analyzed in this dataset includes the most commonly traded size for white leg shrimp that normally ranges from 11 g to 18 g of headless-shell or equal to count size 21/25 to count size 41/50 in the global market.

Table 4.2.1-1 Sample information for metabolomics analysis and QDA

Count size	Average mass (*n*=5)		DOC* (days)	Location	System
41/50	11.24	g	56-75	Banten-Banten	Semi-intensive
31/40	13.17	g	66-88	Banten-Banten	Semi-intensive
31/40	13.48	g	67-90	Banten-Banten	Semi-intensive
26/30	15.85	g	80-106	Banten-Banten	Semi-intensive
21/25	17.90	g	90-119	Banten-Banten	Semi-intensive

*Estimated of DOC (Days of Culture) was obtained by dividing the average mass of headless-shrimp with Average Daily Growth (ADG). ADG represents total mass gained by the shrimp per total days of culture. The range of ADG used was 0.15-0.2 for commercial shrimp production which in line with Indonesia

4.2.2 Shrimp cooking method

Based on a previous study, shrimp samples were cleaned and steamed until the internal temperature reached 85°C (49). Cooked shrimps were cut into 3-4 pieces and cubed into 2-3 cm cuts and the shrimp tails were excluded. The samples were served in each lidded transparent cup

labeled with 3-digits random codes. The samples were placed in a refrigerator after preparation and taken out 30 min before every tasting session.

4.2.3 Panelist recruitment and training

Quantitative descriptive analysis (QDA) was conducted at the Indonesia International Institute for Life Sciences. Ten trained panelists (4 female, 6 male) who had experience in QDA were selected based on their previous experiences in shrimp descriptive analysis. During training sessions, the references were decided following attribute generation (**Supplementary information, Table S3**) for all the samples. Panelists were also re-familiarized with the samples and references before testing.

4.2.4 Sensory testing procedures

Sensory testing was conducted over two sessions in the booths equipped with incandescent lighting and a controlled room temperature of 20°C. All the samples were presented according to a randomized balanced Latin Square Williams Design. All reference samples were free to test before and during the test. In the beginning, for every new sample, the panelist must rinse their mouth with unflavored white bread, followed by water. Panelists made a vertical line on the scale to indicate the perceived intensity.

4.2.5 Sensory data analysis

All sensory data were subjected to statistical software (IBM SPSS Statistics 26) for data processing. Two-way analysis of variance (ANOVA) was employed to determine the statistical difference ($p< 0.05$) of sensory attributes among different sizes of shrimps. The adjusted F-values

were obtained by taking into account the measurement error (interaction between samples and interaction between sample and panelists).

4.3 Results and discussion

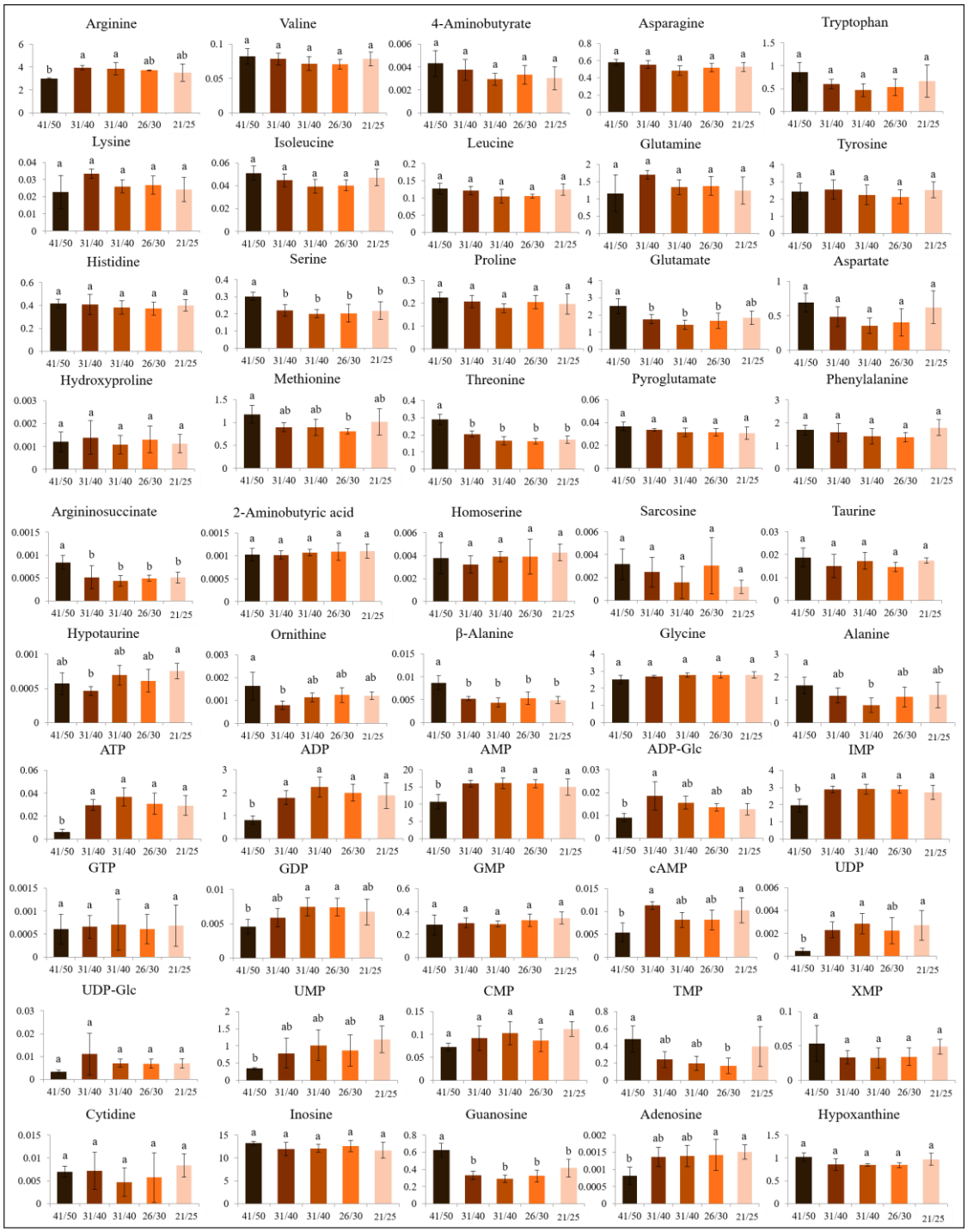
4.3.1 Metabolite profile of white leg shrimp cultured in the same pond

As we intend to find an insightful discussion during shrimp growth, we removed any possible variance from environmental differences by focusing only on the shrimp cultured in the same farm location. Interestingly, despite within a small range of size, the metabolic changes in central carbon metabolism (glycolysis, TCA cycle, and pentose phosphate pathway) were still observed. The bar graph (Figure 4.3.1-1) shows the accumulation of metabolite intensity across different average body mass. Different letters in each row indicate a significant difference (ANOVA Tukey's HSD test, $p< 0.05$). The vertical axis represents the relative intensity. The horizontal axis represents the average of shrimp mass (size). Error bar shows standard deviation obtained from sample triplicates.

Most metabolites underwent a significant change from count size 41/50 to 31/40 before the accumulation trend become steady across the remaining large size. G6P, F6P, FBP, isocitrate showed a positive trend as they increased from count size 41/50 to count size 31/40 then remained steady between 31/40 and 26/30. On the other hand, pyruvate, succinate, and 2-oxo-glutarate decreased from count size 41/50 to count size 31/40 and remained steady between 31/40 and 26/30. Pyruvate, as the end-product of glycolysis, drives several major biosynthetic pathways and it is a master fuel undergirding the TCA cycle for mitochondrial ATP generation (50).

PPP is essential for providing the precursors for amino acid synthesis and nucleic acid synthesis (51). Sugars intermediate in PPP such as S7P and ribose showed an opposite

accumulation trend. S7P was significantly increased from count size 41/50 to the largest size 26/30. Meanwhile, ribose decreased significantly from count size 41/50 to 31/40 and relatively steady in 26/30.



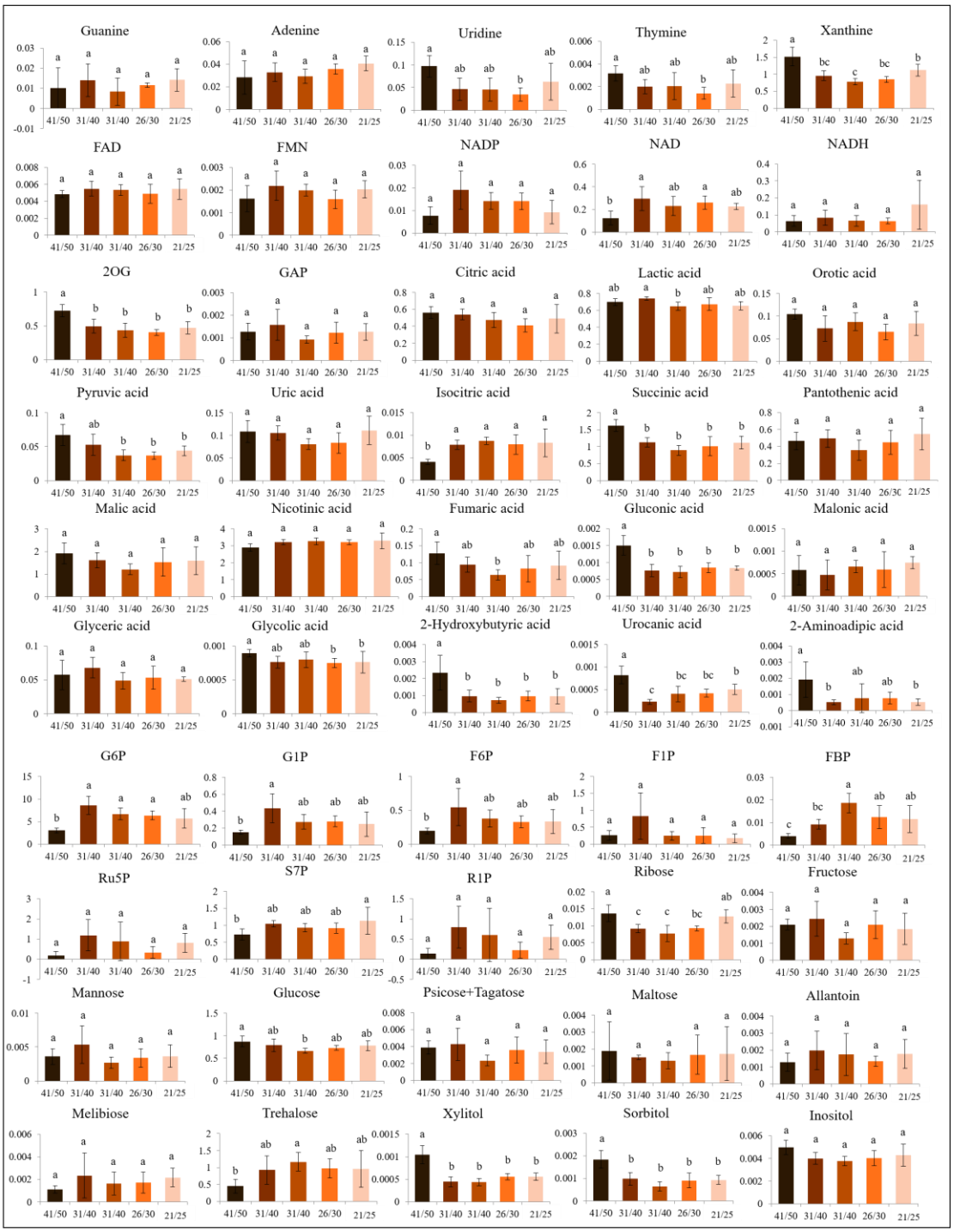


Figure 4.3.1-1. Bar graph of metabolite relative intensity from four different shrimp commercial count sizes.

X axis represents count size; Y axis represents metabolite relative intensity. Different letters indicates a significant difference; same superscripted alphabet means that those data were not significantly different based on the result of one-way ANOVA followed by Tukey's HSD test, ($p < 0.05$). ATP: Adenosine triphosphate;

ADP: Adenosine diphosphate; AMP: Adenosine monophosphate; ADP-Glc: Adenosine diphosphate-glucose; IMP: Inosine monophosphate; GTP: Guanosine triphosphate; GDP: Guanosine diphosphate; GMP: Guanosine monophosphate; cAMP: Cyclic adenosine monophosphate; UDP: Uridine diphosphate; UDP-Glc: Uridine diphosphate-glucose; UMP: Uridine monophosphate; CMP: Cytidine monophosphate; TMP: Thymidine monophosphate; XMP: Xanthine monophosphate; FAD: Flavin adenine dinucleotide; FMN: Flavin mononucleotide; NADP: nicotinamide adenine dinucleotide phosphate; NAD: nicotinamide adenine dinucleotide; NADH: nicotinamide adenine dinucleotide hydride; 2OG: 2-oxoglutarate; GAP: glyceraldehyde 3-phosphate; G6P: Glucose-6-phosphate; G1P: Glucose-1-phosphate; F6P: Fructose-6-phosphate; F1P: Fructose-1-phosphate; FBP: Fructose-1,6-bisphosphatase; Ru5P: Ribose-5-phosphate isomerase; S7P: sedoheptulase-7-phosphate; R1P: ribose-1-phosphate

Amino acids, amino acid derivatives, metabolites in the ornithine cycle, and biogenic amines also displayed significant changes throughout commercial count sizes. Within this data set, serine, threonine, alanine, β-alanine, methionine, hypotaurine, ornithine, arginosuccinate, polyamines putrescine, and cadaverine showed a negative trend, then the change remained insignificant following the count size 31/40. Except for arginine as one of the important positive predictors (chapter 3), in this data set arginine also showed a positive trend towards shrimp size.

On the other hand, purine derivatives ATP, ADP, AMP, IMP, cAMP, GDP, adenosine; pyrimidine derivatives, UDP and UMP; and NAD showed a positive trend. After a significant increase in count size 31/40, the accumulation remained unchanged across large shrimp. A pronounce change for ATP, ADP, and AMP in small shrimp suggesting the use of ATP is coupled to cellular functions for growth. Also, this result confirmed previous findings (chapter 3), where the accumulation of IMP and AMP in large shrimp was due to physiological function instead of the effect of post-mortem change following the handling and storage process.

4.3.2 Sensory analysis and the role of taste active metabolites in different size of commercial shrimps

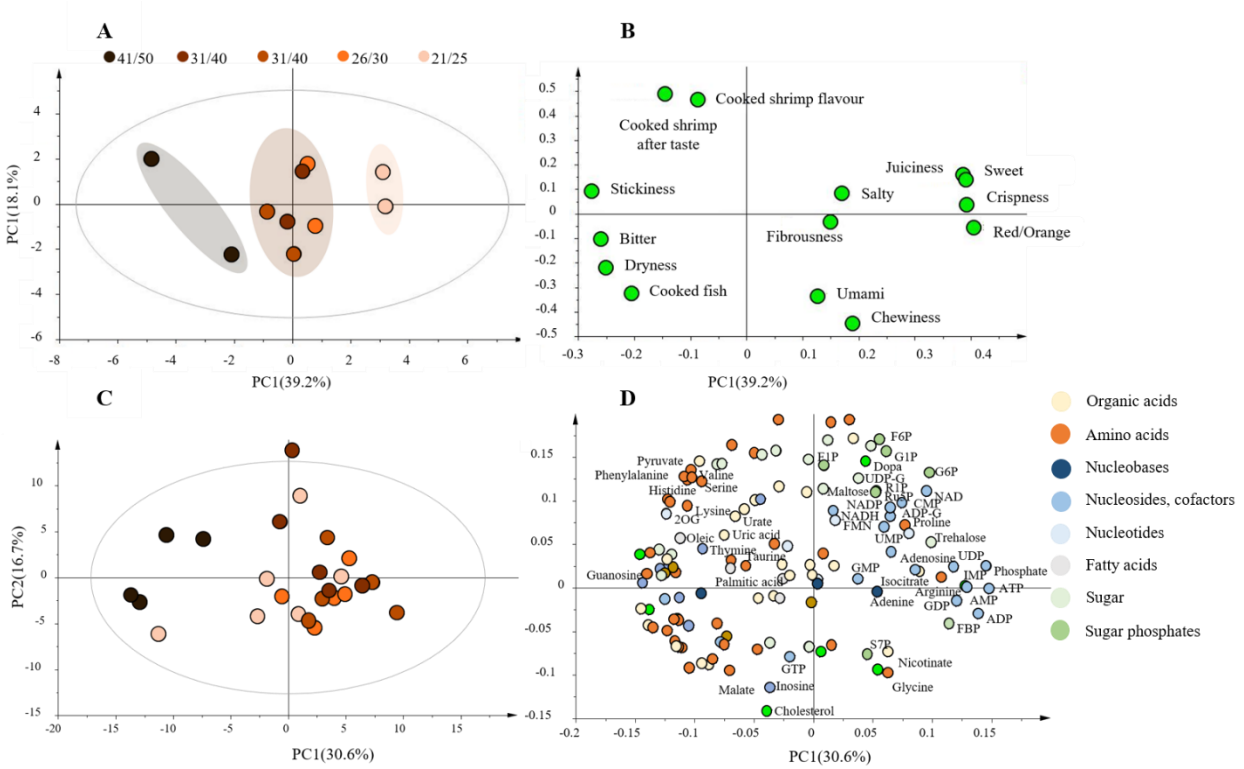


Figure 4.3.2-2 PCA result of QDA and metabolomics data of commercial sized white leg shrimp (*n*=3; Scale: autoscale; Transform: none).

(A) PCA score plot derived from QDA data. Each dot represents shrimp sample. Colour gradation from dark to light indicate the shrimp size from small to large. (B) Loading plot represent the contribution of sensory attributes to the separation of samples. (C) PCA score plot derived from metabolomics analysis (D) Loading plot of shrimp samples.

Quantitative descriptive analysis (QDA) was conducted to investigate whether there was a difference in terms of sensory attributes between shrimp of different commercial sizes. The PCA score plot derived from a sensory score of 14 attributes shows sample separation according to shrimp size. (**Figure 4.3.2-2. A**) A clear separation between 21/25 and 41/50 along PC1 suggested a significant difference in the intensity of attributes perceived between the smallest and the largest shrimp in the commercial size range. Meanwhile, a cluster of count size 31/40-26/30 indicated no

difference in sensory attributes score, which may be attributed to the similarity of metabolite profiles. Count size 21/25 was characterized by juiciness, sweetness, crispness, and red color. On the other hand, count size 41/50 was associated with a bitter taste, stickiness, and dryness (**Figure 4.3.2-2.B**).

The result from metabolome data showed a similar tendency with QDA data. In total, 118 metabolites were putatively annotated from both instruments. Following the bar graph (**Figure 4.3.1-1**), two clusters were observed from the PCA score plot, in which a count size 41/50 as the smallest shrimp in the data set formed a separate group (**Figure 4.3.2-2.C**). Interestingly, higher accumulation of amino acids, organic acids, and several nucleosides corresponded to count size 41/50 which was characterized by a bitter taste, stickiness, and dryness (**Figure 4.3.2-2.B and D**). On the other hand, nucleotides IMP, ATP, AMP, ADP corresponded to large shrimp (count size 31/40-21/25) which are associated with higher juiciness, sweetness, crispness, and red color. it is worth noting that these metabolites are known as important taste-active metabolites.

Previous studies have demonstrated that free amino acids contribute greatly to the flavor of seafood products, as the increase in the amount could improve the taste of foods (52,53). Nucleotides are known to impart umami taste, especially IMP that has been reported to increases the umami taste intensity by fourfold to eightfold compared to that when only glutamic acid is present (33). Alanine, glycine, and proline are sweet taste-active compounds abundant in the crustaceans (54). Among these only alanine showed a negative trend, meanwhile, glycine and proline remained steady throughout the size (**Figure 4.3.1-1**). Umami-tasting amino acids such as glutamic acid, aspartic acid, asparagine, and 4-aminobutyric acid (53,55), bitter-tasting compounds including histidine, valine, leucine, lysine, isoleucine, phenylalanine, tryptophan, tyrosine, purine derivatives (hypoxanthine and xanthine), and nucleoside inosine (43,56) showed

a negative trend. The same trend applied to water-soluble organic compounds that contribute to a typical taste of tartness or sourness in fisheries products and crustaceans, such as fumaric acid, malic acid, oxaloacetic acid + pyruvic acid, and isocitric acid. Reference for taste active metabolites is provided in **Supplementary information, Table S1.**

4.3.3 The significance of size in sensory shrimp sensory properties

The result of ANOVA showed that among the 14 attributes, panelists perceived significant differences ($p < 0.05$) in saltiness and sweetness, crispness ($p < 0.01$), juiciness, and red/orange color ($p < 0.001$) (**Supplementary information, Table S2**). LSD Fisher's test performed on the significant attributes revealed that red/orange color, sweetness, and juiciness were significantly increased in larger shrimps (**Table 4.3.2-1**). The result from ANOVA showed that among the 14 attributes, panelists perceived significant differences ($p < 0.05$) in saltiness and sweetness, crispness ($p < 0.01$), juiciness, and red/orange color ($p < 0.001$) (**Supplementary information, Table S2**). LSD Fisher's test performed on the significant attributes revealed that red/orange color, sweetness, and juiciness were significantly increased in larger shrimps (**Table 4.3.2-1**).

Table 4.3.2-1. Mean intensity ratings and Fisher's least significant difference (LSD) for significant attributes of five shrimp samples

Size	Mean intensity				
	Appearance	Taste		Texture	
	Red/orange colour	Saltiness	Sweetness	Crispness	Juiciness
21/25	11.55^{a}	$3.70^{a,b}$	5.15^{a}	10.10^{a}	9.18^{a}
26/30	9.94^{b}	4.40^{a}	4.56^{a}	$7.52^{b,c}$	$8.89^{a,b}$
31/40	10.19^{b}	3.26^{b}	4.50^{a}	7.66^{b}	8.00^{b}
31/40	10.44^{b}	3.33^{b}	4.64^{a}	8.55^{b}	7.94^{b}
41/50	8.28^{c}	3.12^{a}	3.40^{b}	6.24^{c}	6.73^{c}

Means within a column that do not share superscript letter (a, b, c) are significantly different. ($p < 0.05$). The number shown inside the table refer to the average of intensity for each attribute generated from 15 cm un-interval scale.

As mentioned previously, (Figure 3.3.1-1), even though proline and glycine were not listed as VIP metabolites, as sweet-tasting compounds, their accumulation affects the acceptability of prawns and lobsters (54). It is important to note that the interaction between taste-active compounds is rather complicated and is strongly influenced by the composition of the taste-active compounds themselves (43). As shown in the bar graph, Figure 4.3.1-1, bitter-tasting amino acids covering leucine, valine, isoleucine, tryptophan, tyrosine, histidine, phenylalanine showed insignificant changes throughout the size. The same trend applied to proline and glycine as the important sweet-tasting amino acids. On the other hand, a significant higher of umami-tasting-size dependent metabolites IMP and AMP, and other nucleotides such as ADP, UDP, and GMP were observed in count size 31/40-21/25. In short, insignificant changes of most bitter tasting, sour, and sweet tasting metabolites was found in all shrimp samples, except for all umami-tasting

nucleotides that highly accumulated in large shrimp. As such, the overall sweet taste in large shrimp may be a result of a non-linear but synergistic relationship between multiple taste-active compounds. Therefore, it was possible where a higher perception of sweetness in large shrimp was due to the synergetic effect. This effect is caused by the interaction of multiple taste-active metabolite in one-time.

In addition to sweet attributes, panelists were also able to distinguish the samples based on texture and color differences. Thermal processing changes the sensory and textural properties of shrimp, partly due to the denaturation of proteins (57). Molecular changes induced by thermal processing are predominantly observed as structural changes of skeletal muscle. Juiciness results from the ability of muscle tissue to hold water, known as water holding capacity (WHC) (58). During the cooking process, proteins undergo thermal denaturation that causes water release from protein molecules. Water is easily liberated from the tissue in smaller shrimps because of its lower diffusion thickness (59). This might explain the higher score of juiciness perceived from larger shrimps. At the same time, an increase in hardness due to denaturation and coagulation of myofibrillar proteins and connective tissue affect the changes in crispness (60,61). During protein denaturation, the release of individual astaxanthin from red carotenoid proteins produces a red color in shrimp (62). Previous studies reported that higher redness in shrimp was associated with the content of carotenoids, salt concentration, and moisture loss (60,63). However, the exact mechanism of a significant increase in redness along with the increase in shrimp size remains unclear. The correlation between textural changes and protein content was not assessed in this study, and this may be an interesting subject for future work in this research field.

4.4 Conclusion

The metabolomics analysis of white leg shrimp within a commercial range confirmed significant changes of metabolites from count size 41/50 to count size 31/40. Despite a small difference in average body mass, size demonstrated a significant effect on the shrimp's sensory properties. Size-dependent metabolites, IMP and AMP that are known as important umami-tasting nucleotides contributed to the enhancement of shrimp taste quality. This result successfully demonstrated the importance of shrimp size as a potential quality parameter in the shrimp industry. Overall, this study could serve as a scientific basis for shrimp pricing in the marketplace.

Chapter 5

Conclusion

This study demonstrated the utilization of metabolomics approach as a powerful tool to find the biological importance through an exhaustive metabolite profiling. A significant difference of metabolite profile observed from various sizes of head-less shell-on (HLSO) shrimp enabled us to characterize shrimp molecular quality according to their respective size. These characteristic metabolites, namely size-dependent metabolites showed a strong correlation with size, regardless of the variation in other confounding factors. Through sensory descriptive analysis, size also demonstrated a significant influence on the shrimp sensory properties, especially attributes related to texture such as juiciness and crispness, red color appearance, and sweetness quality. The accumulation of important umami-tasting metabolites in large shrimp, IMP, and AMP contributed to the enhancement of shrimp taste quality. This result successfully demonstrated the importance of shrimp size as a potential quality parameter in the shrimp industry.

From the consumer's future's perspective, small shrimps might offer higher palatability as they contain a higher accumulation of taste-active compounds, suggesting a versatility in culinary usage. Meanwhile, larger commercial shrimps are valued due to their appealing appearance, distinct texture attributes, and sweet taste. As the consumers use raw characteristics such as shrimp appearance, the criteria for shrimp purchasing can be justified according to the count size.

As for the industry, size-dependent metabolites can be utilized as a potential marker in assessing the quality of white leg shrimp. This study could also justify for shrimp farmers to select the optimum harvesting size based on metabolite profiles. In the aquaculture industry, prolonging the shrimp culture period to obtain the desired commercial sizes increases the cost of production

as well as the possibility of loss due to disease infection. However, our results showed that once the shrimp was larger than 31/40, they exhibited a similar metabolite profile with insignificant changes in sweetness. Therefore, prolonging the culture period will not improve the overall quality of shrimp any further. Instead, farmers could re-assess the harvesting size by selecting the smallest count size within the commercial size range that still retains a higher intensity of taste-active metabolites. In such cases, the production cost could be reduced with a shorter period of culture.

Acknowledgement

First, I would like to praise Allah The Almighty, The Lord of the Universe, The Most Gracious, and The Most Merciful for all the opportunities, blessings and favors that I have continuously received all my life especially during my hardest time.

I would like to express my gratitude to Prof. Eiichiro Fukusaki, Assoc. Prof. Shuichi Shimma, Assist. Prof. Sastia Prama Putri, without whose guidance and insight this work would not have been possible. Their constructive suggestions helped me to stay on the right track and they patiently guided me to achieve my final goals.

I would like to acknowledge Prof. Gede Suantika and Dr. Magdalena Lenny for the guidance and valuable advices to strengthen my arguments, Prof. Kazuhito Fujiyama and Prof. Susumu Uchiyama for their insightful comments to improve the quality of my doctoral thesis from various perspectives. I would also like to thank MEXT for the financial support throughout my study.

To Ka Kadar, Ka Mega, and Sana for too many things I've taken for granted, for the listening ears, sincere advises, and those unforgettable moments. To Ka Dinda for always be a helping hand, the one who patiently answered my absurd questions and corrected my broken Japanese sentences. To Tom Cruise-kun, Malik, Fitri, Hadi, Mba Della, Mas Abi, Ploy, Sri, Marvin, Faith, Nanchan, Arisa, Yoshikiwi, Ikuchan, and all food team members from 2016-2021 thank you for the warm friendship and sweet memories. Finally, I would like to deeply thank my family, umi, upi, and ubi, for their endless love, prayers, and supports. Words cannot express how grateful I am.

References

1. **Martinez-Porchas, M., Martinez-Cordova, L.R.**: World aquaculture: Environmental impacts and troubleshooting alternatives, Sci. World J., **2012** (2012).
2. **Bank, T.W.**: FISH TO 2030 Prospects for Fisheries and Aquaculture FISH TO 2030 Prospects for Fisheries and Aquaculture, (2013).
3. **Apud, F.D.**: Extensive and semi-intensive culture of prawn and shrimp in the Philippines., Proc. First Int. Conf. Cult. Penaeid Prawns/Shrimps, , 4–7 (1984).
4. **Lem, A.**: An Overview of Global Shrimp Markets and Trade1, Shrimp Cult., , 3–10 (2006).
5. **Du, L., Chai, C., Guo, M., Lu, X.**: A model for discrimination freshness of shrimp, Sens. Bio-Sensing Res., **6**, 28–32 (2015).
6. **Gillett, R.**: Global study of shrimp fisheries, FAO Doc. Tech. sur les pêches., **475**, 331 (2008).
7. **FAO**: European Fish Price Report, Globefish, , 25 (2016).
8. **Senanan, W., Tangkrock-olan, N., Panutrakul, S., Barnette, P., Wongwiwatanawute, C., Niphonkit, N., Anderson, D.J.**: The presence of the Pacific Whiteleg Shrimp(Litopenaus vannamei, Boone, 1931) in the wild in Thailand, J. Shellfish Res., **26**, 1187–1192 (2007).
9. **Roshith, C.M., Suresh, V.R., Koushlesh, S.K., Manna, R.K., Sharma, S.K., Sibinamol, S., Saha, A., Mandi, R.C., Vijayakumar, M.E., Chowdhury, A.R., Das, B.K.**: Litopenaeus vannamei (Boone, 1931), the Pacific whiteleg shrimp in River Cauvery, Curr. Sci., **115**, 1436–1437 (2018).
10. **Codex Alimentarius Commission**: Codex STAN 92-1981 Standard for Quick Frozen Shrimps or Prawns, Practice, (1995).
11. **Kwamena Quagrainie**: Profitability of Indoor Production of Pacific White Shrimp

(Litopenaeus vannamei): A Case Study of the Indiana Industry, Kwamena Quagrainie, (2015).

12. **Japan External Trade Organization (JETRO)**: Specifications and Standards for Foods , Food Additives , etc . Under the Food Sanitation Act (Abstract) 2010, Specif. Stand. foods, food Addit. etc. Under Sanit. act, **44**, 190 (2010).
13. **Johnson, C.H., Siuzdak, G.**: Metabolomics : beyond biomarkers and towards mechanisms, **17**, 451–459 (2017).
14. **Cevallos-Cevallos, J.M., Reyes-De-Corcuera, J.I., Etxeberria, E., Danyluk, M.D., Rodrick, G.E.**: Metabolomic analysis in food science: a review, Trends Food Sci. Technol., **20**, 557–566 (2009).
15. **Sampantamit, T., Ho, L., Lachat, C., Sutummawong, N., Sorgeloos, P., Goethals, P.**: Aquaculture production and its environmental sustainability in Thailand: Challenges and potential solutions, Sustain., **12**, 1–17 (2020).
16. **Li, T., Li, E., Suo, Y., Xu, Z., Jia, Y., Qin, J.G., Chen, L., Gu, Z.**: Energy metabolism and metabolomics response of Pacific white shrimp Litopenaeus vannamei to sulfide toxicity, Aquat. Toxicol., **183**, 28–37 (2017).
17. **Liu, P. fei, Liu, Q. hui, Wu, Y., Jie, H.**: A pilot metabolic profiling study in hepatopancreas of Litopenaeus vannamei with white spot syndrome virus based on 1H NMR spectroscopy, J. Invertebr. Pathol., **124**, 51–56 (2015).
18. **Sun, S., Guo, Z., Fu, H., Ge, X., Zhu, J., Gu, Z.**: Based on the metabolomic approach the energy metabolism responses of oriental river prawn Macrobrachium nipponense hepatopancreas to acute hypoxia and reoxygenation, Front. Physiol., **9**, 1–11 (2018).
19. **Salayo, N.D., Voon, T.J.P., Selvanathan, S.**: Implicit prices of prawn and shrimp attributes

in the Philippine domestic market, Mar. Resour. Econ., **14**, 65–78 (1999).

20. **Dugassa, H., Gaetan, D.G.**: Biology of White Leg Shrimp, Penaeus vannamei: Review, World J. Fish Mar. Sci., **10**, 5–17 (2018).
21. **Worley, B., Powers, R.**: Multivariate Analysis in Metabolomics, Curr. Metabolomics, **1**, 92–107 (2013).
22. **Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.M., Fiehn, O., Goodacre, R., Griffin, J.L., et al.**: Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI), Metabolomics, **3**, 211–221 (2007).
23. **Lever, J., Krzywinski, M., Altman, N.**: Points of Significance: Principal component analysis, Nat. Methods, **14**, 641–642 (2017).
24. **Lee, B.J., Kim, H.Y., Lim, S.R., Huang, L., Choi, H.K.**: Discrimination and prediction of cultivation age and parts of Panax ginseng by Fourier-transform infrared spectroscopy combined with multivariate statistical analysis, PLoS One, **12**, 1–15 (2017).
25. **MKS Umetrics**: User Guide to SIMCA - Version 13, , 1–621 (2012).
26. **Farrés, M., Platikanov, S., Tsakovski, S., Tauler, R.**: Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation, J. Chemom., **29**, 528–536 (2015).
27. **Alexander, D.L.J., Tropsha, A., Winkler, D.A.**: Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models, J. Chem. Inf. Model., **55**, 1316–1322 (2015).
28. **Pan, P.M., Li, J.P., Lv, G.L., Yang, H., Zhu, S.M., Lou, J.Z.**: Prediction of shelled shrimp weight by machine vision, J. Zhejiang Univ. Sci. B, **10**, 589–594 (2009).

29. **Schock, T.B., Duke, J., Goodson, A., Weldon, D., Brunson, J., Leffler, J.W., Bearden, D.W.**: Evaluation of Pacific White Shrimp (Litopenaeus vannamei) Health during a Superintensive Aquaculture Growout Using NMR-Based Metabolomics, PLoS One, **8** (2013).

30. **Souihi, N., Lindegren, A., Eriksson, L., Trygg, J.**: OPLS in batch monitoring - Opens up new opportunities, Anal. Chim. Acta, **857**, 28–38 (2015).

31. **Pedley, A.M., Benkovic, S.J., Biochem, T., Author, S.**: A New View into the Regulation of Purine Metabolism-The Purinosome HHS Public Access Author manuscript, Trends Biochem Sci, **42**, 141–154 (2017).

32. **Matsumoto, M., Yamanaka, H.**: Influences of Antibiotics Chloramphenicol on Post-mortem Biochemical Changes in the Muscle of Kuruma Prawn during Storage, Nippon Suisan Gakkaishi, **57**, 2291–2297 (1991).

33. **Seki, H., Nakazato, K., Hamada-Sato, N.**: Adenosine monophosphate degradation and inosinic acid accumulation in the shrimp Penaeus japonicus, Int. Aquat. Res., **9**, 37–52 (2017).

34. **Marsac, R., Pinson, B., Saint-Marc, C., Olmedo, M., ArtalSanz, M., Daignan-Fornier, B., Gomes, J.-E.**: Adenylosuccinate Lyase activity in the Purine recycling pathway is essential for developmental timing, germline maintenance and muscle integrity in C. elegans, bioRxiv, , 277640 (2018).

35. **Perl, A., Menzies, K., Auwerx, J.**: NAD+ metabolism and the control of energy homeostasis - a balancing act between mitochondria and the nucleus, Cell Metab., **22**, 31–53 (2016).

36. **Goody, M.F., Henry, C.A.**: A need for NAD+ in muscle development, homeostasis, and

aging, Skelet. Muscle, **8**, 1–14 (2018).

37. **Martínez-Reyes, I., Chandel, N.S.**: Mitochondrial TCA cycle metabolites control physiology and disease, Nat. Commun., **11**, 102 (2020).
38. **Levine, R.L., Hoogenraad, N.J., Kretchmer, N.**: A review: Biological and clinical aspects of pyrimidine metabolism, Pediatr. Res., **8**, 724–734 (1974).
39. **Jimenez, A.G., Kinsey, S.T.**: Energetics and metabolic regulation, Nat. Hist. Crustac. Physiol., , 391–419 (2015).
40. **Coelho, R.T.I., Yasumaru, F.A., Passos, M.J.A.C.R., Gomes, V., Lemos, D.**: Energy budgets for juvenile Pacific whiteleg shrimp Litopenaeus vannamei fed different diets, Brazilian J. Oceanogr., , 1–8 (2019).
41. **Jeffrey D . Arendt**: Adaptive Intrinsic Growth Rates : An Integration Across Taxa, Q. Rev. Biol., **72**, 149–177 (1997).
42. **Miyaki, T., Kawasaki, H., Kuroda, M., Miyamura, N., Kouda, T.**: Effect of a kokumi peptide, γ-glutamyl-valyl-glycine, on the sensory characteristics of chicken consommé, Flavour, **4**, 1–8 (2015).
43. **Meyer, S., Dunkel, A., Hofmann, T.**: Sensomics-assisted elucidation of the tastant code of cooked crustaceans and taste reconstruction experiments, J. Agric. Food Chem., **64**, 1164–1175 (2016).
44. **Yamamoto, S., Bamba, T., Sano, A., Kodama, Y., Imamura, M., Obata, A., Fukusaki, E.**: Metabolite profiling of soy sauce using gas chromatography with time-of-flight mass spectrometry and analysis of correlation with quantitative descriptive analysis, J. Biosci. Bioeng., **114**, 170–175 (2012).
45. **Jumhawan, U., Putri, S.P., Yusianto, Marwani, E., Bamba, T., Fukusaki, E.**: Selection

of discriminant markers for authentication of asian palm civet coffee (Kopi Luwak): A metabolomics approach, J. Agric. Food Chem., **61**, 7994–8001 (2013).

46. **Putri, S.P., Nakayama, Y., Matsuda, F., Uchikata, T., Kobayashi, S., Matsubara, A., Fukusaki, E.**: Current metabolomics: Practical applications, J. Biosci. Bioeng., **115**, 579–589 (2013).
47. **Stone, H., Sidel, J.L.**: 6 - Descriptive Analysis, Food Sci. Technol., , 201–245 (2004).
48. **Civille, G. V, Carr, B.T.**: Sensory Evaluation Techniques, Fifth Edition, (2015).
49. **Xu, Y., Chen, Y., Cao, Y., Huang, W., Zhang, S., Xia, W., Jiang, Q.**: Effect of steam cooking on textural properties and taste compounds of shrimp (metapenaeus ensis), Food Sci. Technol. Res., **22**, 75–81 (2016).
50. **Gray, L.R., Tompkins, S.C., Taylor, E.B.**: Regulation of pyruvate metabolism and human disease, Cell. Mol. Life Sci., **71**, 2577–2604 (2014).
51. **Stincone, A., Prigione, A., Cramer, T., Wamelink, M.M.C., Campbell, K., Cheung, E., Olin-Sandoval, V., Grüning, N.-M., Krüger, A., Tauqeer Alam, M., et al.**: The return of metabolism: biochemistry and physiology of the pentose phosphate pathway, Biol. Rev. Camb. Philos. Soc., **90**, 927–963 (2015).
52. **Yamaguchi, S., Tomoko, Y., Shingo, I., Tsunehiko, N.**: Measurement of the relative taste intensity of some L-a-amino acids and 5 ’-nucleotides, J. Food Sci., **36**, 846–849 (1971).
53. **Sarower, M.G., Farah, A., Hasanuzzaman, M., Biswas, B., Abe, H.**: Taste producing components in fish and fisheries products : A review, J . Food . Ferment . Technol, **2**, 113–121 (2012).
54. **Fuke, S.**: Taste-active components of seafoods with special reference to umami substances BT - Seafoods: Chemistry, Processing Technology and Quality, , 115–139 (1994).

55. **Kurihara, K.**: Umami the Fifth Basic Taste: History of Studies on Receptor Mechanisms and Role as a Food Flavor, Biomed Res. Int., **2015** (2015).

56. **Wiener, A., Shudler, M., Levit, A., Niv, M.Y.**: BitterDB: A database of bitter compounds, Nucleic Acids Res., **40**, 413–419 (2012).

57. **Erdogdu, F., Balaban, M.O., Otwell, W.S., Garrido, L.**: Cook-related yield loss for pacific white (Penaeus vannamei) shrimp previously treated with phosphates: Effects of shrimp size and internal temperature distribution, J. Food Eng., **64**, 297–300 (2004).

58. **Listrat, A., Lebret, B., Louveau, I., Astruc, T., Bonnet, M., Lefaucheur, L., Picard, B., Bugeon, J.**: How muscle structure and composition influence meat and flesh quality, Sci. World J., **2016** (2016).

59. **Welzenbach, J.**: Systems biology analysis of meat quality in Duroc × Pietrain pigs based on integrated omics approaches, (2017).

60. **Niamnuy, C., Devahastin, S., Soponronnarit, S.**: Quality Changes of Shrimp during Boiling in Salt Solution, J. Food Sci., **72**, S289–S297 (2007).

61. **Niamnuy, C., Devahastin, S., Soponronnarit, S.**: Changes in protein compositions and their effects on physical changes of shrimp during boiling in salt solution, Food Chem., **108**, 165–175 (2008).

62. **Muriana, F.J.G., Ruiz-gutierrez, V., Gallardo-guerrero, M.L., Mínguez-mosquera, M.I.**: A study of the lipids and carotenoprotein in the prawn, Penaeus japonicus, J. Biochem., **114**, 223–229 (1993).

63. **Gao, W., Tian, L., Huang, T., Yao, M., Hu, W., Xu, Q.**: Effect of salinity on the growth performance, osmolarity and metabolism-related gene expression in white shrimp Litopenaeus vannamei, Aquac. Reports, **4**, 125–129 (2016).

64. **Irvine, R.F.**: Thematic review series: Living history of lipids: A short history of inositol lipids, J. Lipid Res., **57**, 1987–1994 (2016).

65. **Burdock, G.A.**: Flavor Ingredients, New York, (2005).

66. **Moskowitz, H.R.**: Ratio scales of sugar sweetness, Percept. Psychophys., **7**, 315–320 (1970).

67. **Oshima, H., KIMURA, I., IZUMORI, K.**: Psicose Contents in Various Food Products and its Origin, Food Sci. Technol. Res., **12**, 137–143 (2006).

68. **Rehman, S., Murtaza, M.A., Mushtaq, Z.**: Xylitol as Sweetener, Sweeten. Pharmacol. Biotechnol. Appl., , 1–21 (2017).

69. **Stewart, R.A., Carrico, C.K., Webster, R.L., Steinhardt, R.G.**: Physicochemical Stereospecificity in Taste Perception of α-D-Mannose and β-D-Mannose, Nature, **234**, 220 (1971).

List of publications

Original paper

Putri SLE, Suantika G, Situmorang ML, Christina J, Nikijuluw C, Putri SP, Fukusaki E. Shrimp count size: GC/MS-based metabolomics approach and quantitative descriptive analysis (QDA) reveal the importance of size in white leg shrimp (*Litopenaeus vannamei*). Metabolomics. 2021 Jan 29;17(2):19. doi: 10.1007/s11306-020-01766-z. PMID: 33515101.

Presentations:

1. **Putri SLE**, Suantika G, Situmorang ML, Christina J, Nikijuluw C, Putri SP, Fukusaki E. GC/MS-based metabolomics approach and Quantitative Descriptive Analysis (QDA) exert the importance of size in white leg shrimp (*Litopenaeus vannamei*). *(Poster) Online Metabolomics society. 2020*
2. **Putri SLE,** Putri SP, Fukusaki E The effect of different-sized Japanese spiny lobster, Ise ebi (*Panulirus japonicas*) on its metabolome profile and taste-active compounds. (Poster) *Metabolome Symposium. 2019. Tsukuba, Japan.*
3. **Putri SLE,** Saputra F, Suantika G, Situmorang ML, Putri SP, Fukusaki E. Non-targeted metabolomics approach for size-based prediction of *Litopenaeus vannamei* (white leg shrimp). (Oral presentation) *Society for Biotechnology of Japan. 2019. Osaka, Japan.*
4. **Putri SLE,** Suantika G, Situmorang ML, Putri SP, Fukusaki E. Non-targeted metabolomicsbased prediction of sensory value of *Litopenaeus vannamei* (white leg shrimp). *(Oral presentation) Society for Biotechnology of Japan. 2018. Osaka, Japan.*

Supplementary

Table S1. Taste-active metabolites

Metabolites	Sensory quality	References
Alanine	Sweet	(43)
Inositol	Sweet	(64)
β -Alanine	Sweet	(65)
Threonine	Sweet	(43)
Ribose	Sweet	(65)
Sorbose	Sweet	(66)
Proline	Sweet	(43)
Methionine	Sweet	(43)
Fructose	Sweet	(66)
Psicose+Tagatose	Sweet	(67)
Xylitol	Sweet	(68)
Mannose	Sweet	(69)
Glycerol	Sweet	(66)
Melibiose	Sweet	(65)
Serine	Sweet	(43)
Malic acid	Tart	(65)
Fumaric acid	Tart/sour	(65)
Oxalacetic acid+Pyruvate	Sour	(65)
Acetoacetic acid	Fruity	(65)
Glycolic acid	Sour, honey like taste	(65)
Ethylmalonic acid	Faint, pleasant aromatic odor	(65)
Phosphate	Salty	(43)
Glutamic acid	Umami	(43)
4-Aminobutyric acid	Savory, meat like aroma	(65)
Aspartic acid	Umami	(43)
Asparagine	Umami	(43)
Quinic acid	Bitter	(65)
Histidine	Bitter	(43)
Valine	Bitter	(43)
Leucine	Bitter	(43)
Lysine	Bitter	(43)
Isoleucine	Bitter	(43)
Xanthine	Bitter	(56)
Phenylalanine	Bitter	(43)
Tryptophan	Bitter	(43)
Inosine	Bitter	(43)
Hypoxanthine	Bitter	(56)
Tyrosine	Bitter	(43)

Elaidic acid	Lard like	(65)
IMP	Umami	(43)
AMP	Umami	(43)
ADP	Umami	(43)
ATP	Umami	(43)
GTP	Umami	(43)
GDP	Umami	(43)
GMP	Umami	(43)
UDP	Umami	(43)
UMP	Umami	(43)

Table S2. Analysis of variance (adjusted F-values) for fourteen attributes describing five shrimp samples

Attribute	Sample	Panellist	Replication	Interaction (S*P)
Appearance				
Red/orange colour	**10.50*** **	23.61***	1.16	1.15
Taste				
Saltiness	**3.89***	33.81***	0.04	1.21
Bitterness	0.86	50.56***	3.70	4.65***
Sweetness	**2.82***	3.40	0.50	0.89
Umami	1.16	55.41***	0.91	1.35
Flavour				
Cooked shrimp	0.18	55.70***	0.68	2.18
Cooked fish	0.75	36.89***	1.98	1.34
Texture				
Crispness	**4.42*** **	9.87***	0.10	1.94*
Chewiness	0.07	11.68***	0.63	1.56
Juiciness	**7.40*** **	18.12***	1.27	1.73
Stickiness	2.13	72.00***	0.09	0.71
Fibrousness	1.66	18.23***	0.40	1.49
Aftertaste				
Dry	1.20	37.56***	0.06	1.29
Cooked shrimp	1.18	55.80***	4.63*	1.48

*,**,*** indicate significance at $p < 0.01$ and $p < 0.001$ respectively

Table S3. A complete list of attributes, references, definitions, notes and reference intensities. Reference intensities were determined by panel average.

Attribute	Reference	Definition	Notes	Reference Rating
Red/orange color	Printed paper, L*a*b = -9.5, 22.7, -1.7	The redness of the surface, from white to red orange		15
Salty	0.3% NaCl	Salty taste		7.3
Bitter	0.05% Caffeine	Bitter taste		10.3
Sweet	5% Sucrose	Sweet taste		8.5
Umami	0.1% MSG	Umami taste		7.0
Cooked shrimp	Shrimp broth	The flavor associated with cooked shrimp	Broth from 300 g shrimp cooked in 1.5 L water	10.6
Cooked fish	Fish snack	The flavor associated with cooked fish	IKA'S Fish Snack, 1 cm cut	8.5
Crispness	Sausage	The amount of force exerted during first incisor bite that generates a high pitched sound	KIMBO Sosis Sapi Goreng, boiled for 5 minutes, 0.5 cm cut	3.2
Chewiness	Rice cake	The time required to masticate sample to a consistency acceptable for swallowing	Korean frozen rice cake, boiled for 5 minutes, 0.5 cm cut	13
Juiciness	Sausage	The amount of moisture released during mastication	KIMBO Sosis Sapi Goreng, boiled for 5 minutes, 0.5 cm cut	6.3
Stickiness	Rice cake	The cohesiveness of the muscle fibers to the teeth	Korean frozen rice cake, boiled for 5 minutes, 0.5 cm cut	6.5
Fibrousness	Pineapple	The density of the muscle fibers in the shrimp meat	Canned pineapples Ayam Brand	9.1
Dry	0.1% MSG	The feeling of dryness in mouth		9.3
Cooked shrimp	Shrimp broth	The flavor associated with cooked shrimp	Broth from 300 g shrimp cooked in 1.5 L water	10.6

Table S4. List of general metabolites annotated from GC/MS analysis

Metabolite	Average Retention Index	Average Quantitative m/z	Annotation
2-Aminobutyric acid	1177.32	130.125	GL-sciences library
2-Aminoethanol	1275.19	174.1	GL-sciences library
4-Aminobutyric acid	1541.29	174.1	GL-sciences library
Alanine	1106.28	116.1	GL-sciences library
Asparagine	1682.53	116.1075	GL-sciences library
Aspartic acid	1529.67	232.0625	GL-sciences library
beta-Alanine	1436.5	174.0945	GL-sciences library
Cadaverine	1856.57	174.095	GL-sciences library
Fumaric acid	1346.95	245.0177	GL-sciences library
Galactose	1925.05	319.1	GL-sciences library
Galactose+Glucose	1951.12	319.0833	GL-sciences library
Glucose	1931.17	319.1028	GL-sciences library
Glutamic acid	1628.13	246.0813	GL-sciences library
Glutamine	1784.88	156.095	GL-sciences library
Glycerol	1280.54	147.1	GL-sciences library
Glycine	1316.32	174.1	GL-sciences library
Glycolic acid	1074.69	147.15	GL-sciences library
Histamine	1890.28	174.1	GL-sciences library
Histidine	1940.52	154.1	GL-sciences library
Hydroxyproline	1538.15	230.0933	GL-sciences library
Hypoxanthine	1820.72	265.0471	GL-sciences library
Inositol	2131.89	217.0575	GL-sciences library
Isocitric acid+Citric acid	1838.03	273.0667	GL-sciences library
Isoleucine	1300.76	158.1475	GL-sciences library
Lactic acid	1063.17	147.1	GL-sciences library
Leucine	1278.72	158.15	GL-sciences library
Lysine	1938.62	174.0975	GL-sciences library
Malic acid	1496.06	147.1	GL-sciences library

Malonic acid	1206.88	147.1025	GL-sciences library
Mannose	1917.93	319.105	GL-sciences library
Melibiose	2938.44	204.05	GL-sciences library
n-Butylamine	1112.25	174.0905	GL-sciences library
Nicotinic acid	1297.54	180.0111	GL-sciences library
Oxalacetic acid+Pyruvate	1050.03	174.0361	GL-sciences library
Oxalate	1133.01	147.0912	GL-sciences library
Phenylalanine	1642.61	218.05	GL-sciences library
Phosphate	1281.37	299.0316	GL-sciences library
Plamitic acid(16:0)	2046.12	117.055	GL-sciences library
Proline	1306.07	142.145	GL-sciences library
Putrescine	1757.66	174.1025	GL-sciences library
Pyroglutamic acid	1534.33	156.09	GL-sciences library
Ribose	1698.67	103.0525	GL-sciences library
Serine	1368.92	204.07	GL-sciences library
Stearic acid(17:0)	2243.84	117.0525	GL-sciences library
Threonine	1396.85	218.095	GL-sciences library
Thymine	1410.12	255.0632	GL-sciences library
Tryptamine	2232.89	128.1	GL-sciences library
Tryptophan	2250.84	202.0533	GL-sciences library
Tyrosine	1957.72	218.0575	GL-sciences library
Uracil	1346.18	241.0225	GL-sciences library
Urea	1236.19	147.0917	GL-sciences library
Uric acid	2124.92	441.0923	GL-sciences library
Uridine	2478.34	217.0467	GL-sciences library
Valine	1222.66	144.1475	GL-sciences library
Xanthine	2039.23	353.1	GL-sciences library
Xylitol	1732.13	217.0553	GL-sciences library
Xylonic acid	1793.14	292.0833	GL-sciences library

Table S4. Multiple reaction monitoring (MRM) transitions for in RP-IP LC/QqQ/MS system

Metabolite	Precursor ion m/z	Product ion m/z	Target Collision Energy (V)
Arginine	173.05	131.05	15
Lysine	145.1	97.05	13
Histidine	154.05	93	21
4-Aminobutyrate	162.05	102	8
Serine	104.05	74.1	16
Asparagine	131.05	113.05	15
Glutamine	145.1	127.05	18
Threonine	118.05	74.05	16
Hydroxyproline	190.05	130.05	10
Hexose	179.05	89	19
Cysteine	239.05	120.1	13
Trehalose	341.05	89.1	23
Proline	174.05	114	10
Sucrose	341.05	89.1	23
Valine	176.05	116.05	10
Cytidine	302.05	242	10
Methionine	148.05	47.05	14
Guanine	150.05	133.05	21
Hypoxanthine	135.05	92	28
Tyrosine	180.05	163.05	18
Adenine	134.1	107.05	22
Isoleucine	190.05	130.05	10
Leucine	190.05	130.05	10
Xanthine	151.05	108.05	18
Glutamate	146.05	102.05	15
Uridine	243.05	110.05	17
Aspartate	132.05	88.05	14
Inosine	267.05	135.05	23
Thymine	125.05	42	18
Guanosine	282.1	150.05	21
Urate	167.1	124.05	15
Shikimate	173.05	93	17
Adenosine	266.1	134.05	25
Glycerate	105.05	75.05	13
Thymidine	301.1	241	10
Phenylalanine	164.05	147.05	18
Glycolate	75.05	47.05	13
Glyoxylate	73	73	5
G6P	259.05	97	17
Pyroglutamate	188.05	128	12

Tryptophan	203.1	116.05	18
R5P	229.05	97	13
Lactate	89.05	43	14
S7P	289.1	97	21
F6P	259.05	97	17
G1P	259.05	79.05	28
α-Glycerophosphate	171.05	79.05	18
TPP	424.1	302.05	16
NAD	662.1	540.1	18
GAP	169.05	97	12
Orotate	155.05	111.05	14
Ru5P	229.05	97	13
CMP	322.1	79.05	28
β- Glycerophosphate	171.05	79.05	18
F1P	259.05	97	17
Pyruvate	87.05	43	11
R1P	229.05	79.05	25
UMP	323.1	79.05	36
AICAR	337.1	79.05	37
GMP	362.1	79.05	26
IMP	347.05	79.05	40
DHAP	169.05	97	12
TMP	321.1	195.05	20
AMP	346.1	79.05	38
Pantothenate	218.05	88	17
Nicotinate	122.05	78	16
cAMP	328.1	134.05	27
Succinate	117.05	73	15
Malate	133.05	115	17
UDP-Glu	565.05	323.05	27
XMP	363.1	211.05	20
CDP	402.1	79.05	42
2OG	145.1	101.05	10
ADP-Glu	588.05	346.05	23
Fumarate	115.05	71	10
GDP	442.1	79.05	45
UDP	403.1	159	28
NADH	664.1	79.05	57
NADP	742.1	620.1	18
ADP	426.1	79.05	46
Citrate	191.05	87	18
FBP	339.05	97	18
Isocitrate	191.05	73	22
GTP	522.1	159	33

CTP	482.1	159	36
UTP	483.1	159	36
ATP	506.1	159	40
FAD	784.1	346.1	37
NADPH	744.1	159.05	49
D-Camphorsulfonic acid	231.1	80	32