Network analysis of plasma and tissue amino acids and the generation of an amino index for potential diagnostic use1–3

Yasushi Noguchi, Qing-Wei Zhang, Tetsuya Sugimoto, Yasufumi Furuhata, Ryosei Sakai, Masato Mori, Mitsuo Takahashi, and Takeshi Kimura

ABSTRACT
Background: Few studies exist on the use of metabolic profiling of amino acids to examine underlying physiologic and disease states.
Objective: We aimed to introduce a new method for studying relations among amino acids and to generate a diagnostic index, or amino index, based on amino acid concentrations.
Design: For network analysis, 35 Fischer-344 rats were randomly divided into 7 groups and fed diets containing 5%, 10%, 15%, 20%, 30%, 50%, or 70% protein. Amino acid concentrations in plasma and various organs were used to derive correlation coefficients that were then used to construct correlation networks. To build a diagnostic index for diabetic rats, the plasma amino acid concentrations of diabetic and normal rats were analyzed by using a novel algorithm developed to generate amino acid–based indexes. Plasma amino acid concentrations from human growth hormone transgenic rats and insulin-treated diabetic rats were used to evaluate the index obtained for diabetes. Dimethylnitrosamine-treated Sprague-Dawley rats were used to generate an index for hepatic fibrosis.
Results: The scatter plots of plasma amino acid concentrations showed distinct patterns in different organs that were due to the different protein contents of the diets. Network analysis showed that data-driven networks for blood and tissue could be obtained. We derived a diagnostic index for the discrimination of diabetic rats with both sensitivity and specificity >97% and another surrogate index for liver hydroxyproline with a correlation of $r^2 = 0.85$.
Conclusions: Correlation-based network analysis may help to uncover specific physiologic conditions or states. A novel approach using amino acid molar ratios was shown to generate indexes that can be used to separate animal disease models and monitor the progression of a disease parameter. Some of the methods described here may be applicable to the clinical setting. Am J Clin Nutr 2006;83(suppl):513S–9S.

KEY WORDS Aminogram, amino acids, metabolism, metabolomics, metabonomics, network analysis, amino index, diagnosis, correlation, cluster analysis

INTRODUCTION
Amino acids play important roles both as basic substrates and as regulators in many metabolic pathways (1–3). The quantification of free amino acids in biological fluids and tissues provides us important biochemical and nutritional information that enables the diagnosis of various diseases, especially metabolic deficiencies (4–7). Although most information on free amino acid concentrations in the clinical setting has been derived from blood, urine, or amniotic fluid, amino acids have also been measured, although less frequently, in other biological fluids such as breast milk (8) and cerebrospinal fluid (9) and in cells and tissues such as liver, muscles, and bone (10). Today, amino acid profiles (aminograms) for biological samples are commonly analyzed by ion-exchange chromatography (amino acid analyzer), in which $>40$ amino acids and related compounds can be measured (7, 11, 12). Specific abnormalities in amino acid concentrations, associated with physiologic conditions, have been reported in the context of various diseases, including liver failure (13), renal failure (14), cancer (15), diabetes (16), muscle dysfunction (17), aminoacidemia (18), and so on. The balance between the branched-chain amino acids (BCAAs) and aromatic amino acids, known as Fischer’s ratio, is established as a diagnostic marker and is used to monitor the progression of liver fibrosis and the effectiveness of drug treatments (19, 20). Except for Fischer’s ratio and aminoacidemia, however, information from aminogram data has not been sufficiently utilized for monitoring physiologic states, nutritional control, or clinical diagnosis.

Postgenomic molecular biology is being driven by new and highly powerful experimental technologies that enable large-scale and parallel examination of cell states of various biological phenotypes. Such analyses can be carried out at the level of transcription, to define the transcriptome (21), or at the level of translation, to define the proteome (22). To date, metabolomics, the examination of dynamic changes in the metabolome, has emerged as a powerful technique in diverse fields such as drug discovery, clinical diagnostics, and plant biochemistry (23, 24). Metabolic profiling has used data-mining methods, such as non-parametric multivariate analysis, to extract classification functions for state discrimination (25). In addition, pattern recognition techniques, such as principal component analysis and partial least squares, have been used to develop diagnostic screens for heart disease and drug toxicity (26, 27). As a result, the potential
of clinical diagnosis protocols based on multiple metabolite measurements has been recognized (23). In contrast with the rapid development in metabolomic studies and profiling techniques, the development of subsequent data analysis methods is lagging. This is also true for the study of amino acid profiling. Therefore, the development of appropriate methods to explore and elucidate key players from the profiled data and the development of practical diagnostic methods for a specific phenotype are currently of great importance.

**SUBJECTS AND METHODS**

**Animals**

All rats were purchased from Charles River Laboratories Inc (Kanagawa, Japan). For the network analysis of amino acid metabolic profile data, a total of 35 Fischer-344 rats were divided randomly into groups fed diets containing 5%, 10%, 15%, 20%, 30%, 50%, or 70% protein (n = 5 per group). The body weights of the different groups of animals at the beginning of the experiments were as follows: 88 ± 5, 89 ± 6, 90 ± 4, 89 ± 5, 89 ± 3, 90 ± 4, and 89 ± 3 g (x ± SD), respectively. The diets were AIN-93G–based diets with modification of both protein and carbohydrate percentages, which were calculated to contain the following percentages: 5%, 10%, 15%, 20%, 30%, 50%, or 70% raw protein; 0.075%, 0.15%, 0.23%, 0.3%, 0.45%, 0.75%, or 1.05% cystine; and 78%, 73%, 68%, 63%, 53%, 32%, or 12% cornstarch, respectively. The other raw materials were cellulose powder (5%), minerals (3.5%), vitamins (1.25%), and soybean oil (7%). All rats were maintained on a 12-h day-night cycle with free access to water. The diets were continued for 14 d.

Male Wistar and Goto-Kakizaki rats (GR rats), 7 wk old and weighing 150–200 g, were used for construction of an amino index for the diagnosis of model diabetic rats. Streptozotocin-induced type 1 diabetic rats (STZ rats) were produced by using the protocols previously reported (28). Human growth hormone transgenic rats (hGH-Tg rats) were produced by Ikeda et al (29, 30), and male rats (heterozygotes) weighing 250–350 g and their male littermates were used.

A liver dysfunction model was produced by administration of 10 mg/kg dimethylnitrosamine (DMN) every other day (3 times a week for 3–4 wk) to 23 male Sprague-Dawley rats (7 wk old, weighing 150–200 g) as previously described (31). All rats were housed under conditions of constant temperature (23 ± 1°C) and light (on at 0700, off at 1900), with free access to water. After a 5-h fast, a time period that was previously shown to allow amino acid concentrations to return to baseline (32), rats were killed between 1200 and 1500. Blood was collected in tubes containing EDTA (NONCLOT-D; Daichi Pure Chemicals, Tokyo, Japan) on ice. Tissues were collected from the animals after blood collection and were stored at −80°C. All studies were reviewed and approved by the Animal Care Committee of Ajinomoto Company, Inc.

**Amino acid analysis**

Plasma samples were obtained with use of EDTA as the anticoagulant, mixed with 2 volumes of 5% (w/w) trichloroacetic acid, and centrifuged immediately (4°C, 20 min, 8000 × g) after blood collection to remove precipitated protein. To prepare deproteinized tissue extracts, tissues were homogenized by using 5% trichloroacetic acid as described previously (10). All samples were kept at 4°C during all steps to minimize chemical reactions of thiol metabolites. The amino acid concentrations were measured by an automatic amino acid analyzer (L-8800; Hitachi, Tokyo, Japan). Briefly, amino acids, separated by cation-exchange chromatography, were detected spectrophotometrically after postcolumn reaction with ninhydrin reagent. Hydroxyproline content was determined by the method previously reported (33).

**Statistical analysis**

**Correlation scatter plot and network analysis**

Plasma concentrations of amino acids measured by use of the amino acid analyzer were analyzed by using JMP (SAS Institute Inc, Cary, NC) for correlation scatter plots, correlation coefficient matrices, and simple cluster analysis as described previously (34, 35). The UNISTAT statistical package (version 4.5; Unistat Ltd, London, United Kingdom) was used to construct colored blocks representing levels of correlation of amino acid concentrations in plasma and tissues. UCINET (version 6.34; Analytic Technologies Inc, Harvard, MA) was used for further hierarchical clustering based on the unweighted pair group method with arithmetic mean (UPGMA) and for construction of standard bipartite graphs representing the correlation network of plasma and tissue amino acid concentrations.

**Algorithm for a novel amino index**

The basic algorithm to generate a diagnostic index was as follows: 1) Amino acids are classified into 2 groups on the basis of their correlation with the target parameter, whether continuous, ordinal, or dichotomous, that represents the disease stage or physiologic condition. This gives to a group P that contains Np amino acids Aα (1 ≤ α ≤ Np) having a positive correlation with the target variable, and to a group Q that contains Nq amino acids Bβ (1 ≤ β ≤ Nq) having a negative correlation with the target variable. 2) The second step is to calculate all possible combinations of the variables in the form of the fractional function $F = \sum (\sum A/\sum B)$ (1) where ΣAα is a partial sum of Aα selected from P, and ΣBβ is a partial sum of Bβ selected from Q, with no redundant use of the variables in the function, and where r is the number of fractions in the function. The fractional functions are selected for further processing by evaluation methods such as the sum of squares due to error based on the simple linear regression of F for the continuous target parameter or the variance ratio for the categorical target parameter. 3) The optimal fractional function F, or amino index, is finally selected after cross-validation of a large set of candidate functions by using random sampling to attain robustness with respect to sample variability and noise (25, 36).

In reality, simple combinations of n variables give rise to 2^n − 1 fractional functions; hence, calculation of all possible combinations of the amino acids to derive the optimal fractional function is time consuming. For convenience, we first calculated all combinations of variables to derive optimal single fraction functions so as to select potential variables and then calculated all combinations to construct the optimal fractional function with no >4 fractions (r ≤ 4) at most in the study. All calculations were carried out by using programs written in Visual C++ (version 9.0) at the National Research Institute for the Development of Agricultural and Forestry Sciences, Tokyo, Japan.
6.0; Microsoft Co, Redmond, WA) and R (version 1.8; R foundation for statistical computing, Vienna, Austria).

RESULTS

Effect of dietary protein intake on correlation of plasma amino acids

The effects of dietary protein concentrations on free blood amino acid concentrations have been studied in rats (37) and humans (38, 39). In such studies, the metabolism of each amino acid was separately considered and discussed, and because of the complexity, no uniform trends in changes in blood amino acid concentrations emerged (37).

We applied multivariate-correlation analysis to plasma samples from rats fed diets containing graded amounts of protein (5–70% casein). The scatterplots of 2 plasma amino acids show various patterns, such as positive and negative linear correlations or no correlations (Figure 1). In particular, a strong linear correlation was obtained between Leu and Ile or Val, which indicates that the metabolism of BCAAs is tightly linked irrespective of the amount of dietary protein (Figure 1). BCAA concentrations...
showed a significant dependency on cumulative protein intake, which indicates that the blood concentrations of these amino acids faithfully reflect the amount of dietary protein. On the other hand, the correlation between Leu and Thr or Ser exhibited a distinct inflection, with the low-protein (5–15%) and the high-protein (30–70%) groups showing different trends (Figure 1). These data suggest that blood amino acid concentrations in the steady state primarily reflect indirect influences derived from an individual’s adaptation to changes in dietary protein intake rather than being indicative of the amount of dietary protein intake.

Correlation of plasma and organ amino acids

Shown in Figure 2 are colored blocks representing levels of correlation of amino acid concentrations in plasma versus amino acid concentrations in liver, kidney, small intestine, and muscle in the rats fed graded amounts of protein (5–70% casein). Here, amino acids were ordered on the basis of the hierarchical cluster tree of plasma amino acids analyzed as described in our previous studies (34, 35). The colored blocks show the differences in correlations between amino acid concentrations in plasma and in each tissue. In this representation, red represents high positive correlation and blue represents high negative correlation. If the plasma and tissue amino acid levels were the same, there would be a line of red blocks running diagonally from the right-hand top corner to the left-hand bottom corner of each square of blocks. Detailed examination shows that in general there was higher correlation of essential amino acid concentrations between the plasma and tissues than for nonessential amino acids. However, in the kidney, BCAA concentrations showed low correlation to plasma BCAA concentrations. The fact that different tissues exhibited different patterns indicates that each tissue may have a distinct amino acid metabolism.

Correlation-based network analysis of organ and plasma amino acids

On the basis of their pairwise correlations, the data obtained from metabolomic experiments can be organized into metabolic correlation networks, and the key challenge is to deduce unknown pathways on the basis of the observed correlations (40). We constructed a network model of organ amino acids versus plasma amino acids by using the correlations obtained from Figure 1b. Both plasma and organ amino acid concentration data were used to derive correlation coefficients between blood and organs for each amino acid, and then these were subsequently used to construct a two-mode bipartite matrix. Using this matrix, we can visualize a network structure of the relation between plasma and organ amino acids.

The data-driven model obtained clearly showed several metabolically meaningful structures (Figure 2b). For example, Gln, Ala, and Ser were clustered with urea cycle intermediates. Gln is an important source of nitrogen used in hepatic ureagenesis and the major source of nitrogen excreted in urine (41). In addition, Ala is transferred to the circulation by many tissues, but mainly by muscle, in which Ala is formed from pyruvate at a rate proportional to intracellular pyruvate concentrations. Liver accumulates plasma Ala, reverses the transamination that occurs in muscle, and proportionately increases urea production (42). In contrast, the contribution of Ser to ureagenesis is not fully understood. Liver serine dehydratase (SDH) has been reported to be induced by an increase in dietary protein intake and to be negatively correlated with nitrogen balance (43, 44). It has been hypothesized that SDH might be induced to catabolize surplus amino acids to maintain nitrogen homeostasis in the body (44). If this is the case, the network model obtained here might reflect such metabolic interactions. Thus, analytic methods based on the correlation of metabolite concentrations, such as for amino acids, can give general and wide-ranging information about the metabolites’ relations and biochemical implications of changes in their metabolism.

An amino index for discrimination and physiologic state monitoring

To validate the diagnostic potential of an amino index, we used rat models of diabetes mellitus: STZ rats (n = 14), GK rats (n = 28), and normal rats (n = 73) as described in Subjects and Methods. Model rats were categorized as diabetic rats or normal rats and were assigned values of 1 or 2, respectively. After calculation of the candidate indexes showing high correlation with the 2 categories, average variance ratio by cross-validation was used to evaluate the amino indexes. The amino index (Tau+Cit+Lys)/(Asp+Ile)+(Thr+Tyr+His)/Glu, which had the largest variance ratio, was derived as the optimal index for separating diabetic rats from normal rats (Figure 3A). This index gave a sensitivity of 98.6% and a specificity of 97.6% for diagnosing diabetes mellitus for the model rats.

We further applied the index to hGH-Tg rats (n = 43) and insulin-treated STZ rats (n = 8) to test the specificity of the index. The hGH-Tg rats are characterized by virtually no endogenous rGH secretion and massive obesity (45). As shown in Figure 3a, all hGH-Tg rats were clearly classified as nondiabetic rats, consistent with the fact that hGH-Tg rats do not exhibit diabetic concentrations of blood glucose or insulin. On the other hand, most of the long-term insulin-treated STZ rats were classified as diabetic rats, yet the improvement of insulin deficiency could be confirmed by a shift of the index scores toward that of the normal rats. The difference between the amino indexes for STZ rats (n = 14) and the insulin-treated STZ rats (n = 8) was given by the Welch t test with a significance of P < 0.0001.

We further applied the method to DMN-treated rats (n = 23), a liver dysfunction model, to see whether an amino index could be generated to predict the degree of liver fibrosis without biopsy. We chose the liver hydroxyproline concentration, which has been reported to be a marker for liver fibrosis (46), as the target parameter and ran the amino index program to see whether it could generate a function that could serve as a surrogate marker. The regression to the hydroxyproline concentrations in the liver of DMN-treated rats resulted in the index (Tau+Cit+Met+Arg)/ (Ser+Leu) + (Phe+Orn)/(Glu+Trp) with a coefficient of determination (r²) of 0.85, whereas the r² for Fischer’s ratio against hydroxyproline concentrations was 0.49 (Figure 3b)

DISCUSSION

Metabolism can be viewed as a network that can adapt to various nutritional conditions and that may become perturbed during disease and physiologic insults. Apart from the metabolic pathways within cells, there are various levels of networks within...
FIGURE 2. Correlation-based analyses of free amino acids in plasma and organs. A: The colored blocks represent the correlations between amino acid concentrations of plasma and tissues. Cluster analyses are applied to plasma or organ amino acids as described in the Subjects and Methods. The value of the correlation coefficients is color-coded from 1 (highly positive, light red) to 0 (no correlation, black) to -1 (highly negative, light green). Each organ amino acid is labeled as follows; L, liver; K, kidney; M, muscle; and SI, small intestine. B: Network model of organ amino acids via plasma amino acids. Each node represents plasma (pink) and organ (light blue) amino acids. The bipartite graph was constructed by using the correlation data of organs and plasma amino acids obtained from panel A (edges are shown only when \( r > 0.5 \)).
For improvement of the model, such as extension of the index formulation, better evaluation methods, or combinational usage with protein or gene markers.

Amino acids represent a convenient set of metabolites that can be measured easily. When the association of abnormalities in individual amino acid concentrations with specific diseases or physiologic conditions were addressed in the past, researchers seldom took advantage of the whole metabolic information latent in the amino acid profile (13–18). For example, of the amino acids found in the amino index generated as a surrogate marker for liver hydroxyproline concentrations \( [(\text{Taur}+\text{Cit}+\text{Met}+\text{Arg})/(\text{Ser}+\text{Leu}) + (\text{Phe}+\text{Orn})/(\text{Glu}+\text{Trp})] \), more than one-half were reported in studies of liver injury or disorders. Abnormal plasma Met concentrations were reported in liver injury, and Ser is involved in the methionine catabolic pathway (48). Cit and Orn were reported to be associated with abnormalities of the urea cycle and with hyperammonemia in liver dysfunction (49). Branched-chain amino acids such as Leu and aromatic amino acids such as Phe were long used in Fischer’s ratio to discriminate fibrosis. We feel that by developing an algorithm that looks at all potential amino acid combinations for a particular function, we can now harness information buried in amino acid profiles for the generation of diagnostic and surrogate markers that could help advance biomedical and nutritional science.

We have shown that in the metabolic profiling of amino acids, correlation-based network analysis can be helpful in revealing specific physiologic conditions or states and revealing hitherto overlooked metabolic relations. A novel combinatorial algorithm using a function based on amino acid molar ratios to generate indexes was shown to generate indexes that can separate diabetic rat models from normal rats and human growth hormone-expressing, obese, nondiabetic rats. The same algorithm was shown to be able to generate an index approximating the concentrations of hydroxyproline in the liver, a marker of fibrosis, in rats with hepatic dysfunction caused by treatment with DMN. We suggest that the methods described in this paper could be applicable to the clinical setting and be useful in diagnosing various physiologic and disease states.

We thank the late Vernon Young for expressing interest in this work during its inception and the organizers of the symposium, especially Naomi Fukagawa, for giving TK the opportunity to present it. We also thank Takennori Hiraoka, Kaori Ono, Ikuko Yokoyama, Takashi Horie, and Yoshiro Kitahara for expert technical assistance and helpful advice and Michio Takahashi and Kiyoshi Miwa for their continued interest and support.

The body, with the transport of substrates such as amino acids by the blood between organ systems representing one level. Our results using network analysis (Figure 1) suggest that correlation-based analysis can be used to visualize a multitude of inter-organ substrate relations, which can help us to understand the complex interrelations that make up metabolism. If certain correlations between amino acid concentrations are specific for certain physiologic states, then individuals whose physiologic states are different may have a different pattern of correlations. On the basis of these observations, various functions were tested empirically until one was found that could separate 2 physiologic states such as high-protein and low-protein states (not shown). This initial function was then generalized to the form of the function reported in this paper.

The basic form of our novel index can be described as being based on introduction of molar ratios. The form of the function can be rationalized as follows: 1) the introduction of molar ratios into the index function may be essential to represent the mutual correlation of amino acids; 2) the variable ratio discriminating the balance state of 2 parameters has already been used in fields such as immunology, in which Th1/Th2 ratios are used for immune system monitoring (47); 3) molar ratios, like Fischer’s ratio, have proved to be effective and robust. Compared with the simple ratio of 2 parameters, the amino index is in fact a nonlinear advanced ratio model after introduction of linear functions into both numerator and denominators. However, there is still room

REFERENCES


