electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *The Plant J* 2004, **40:**151-163.
18. Sammon JW Jr: **A nonlinear mapping for data structure analysis.** *IEEE Transactions Computers* 1969, **C-18(5):**401-409.
19. **KEGG pathway database** [http://www.genome.ad.jp/kegg/pathway.html]
20. **Swiss-Prot database** [http://au.expasy.org/sprot/]
21. **Rice Annotation Project Data Base** [http://rapdb.dna.affrc.go.jp/]
22. Stafford HA, Magaldi A, Vennesland B: **The enzymatic reduction of hydroxypyruvic acid to D-glyceric acid in higher plants.** *J Biol Chem* 1954, **207:**621-629.
23. Rippert P, Matringe M: **Purification and kinetic analysis of the two recombinant arogenate dehydrogenase isoforms of *Arabidopsis thaliana*.** *Eur J Biochem* 2002, **269:**4753-4761.
24. Boldt R, Edner C, Kolukisaoglu Ü, Hagemann M, Weckwerth W, Wienkoop S, Morgenthal K, Bauwe H: **D-Glycerate 3-kinase, the last unknown enzyme in the photorespiratory cycle in Arabidopsis, belongs to a novel kinase family.** *The Plant Cell* 2005, **17:**2413-2420.
25. Duncan K, Edwards RM, Coggins JR: **The pentafunctional *arom* enzyme of *Saccharomyces cerivisiae* is a monofunctional domains.** *Biochem J* 1987, **246:**375-386.
26. Kuhr WG: **Separation of small organic molecules.** In *Capillary Electrophoresis: Theory and Practice* Edited by: Camilleri P. CRC Press LLC, Boca Raton, FL, USA; 1997:91-133.
27. **KEGG ligand database** [http://www.genome.ad.jp/kegg/ligand.html]
28. Steuer R, Kurths J, Fiehn O, Weckwerth W: **Observing and interpreting correlations in metabolic networks.** *Bioinformatics* 2003, **19(8):**1019-1026.
29. Ward JH: **Hierarchical grouping to optimize an objective function.** *J Am Stat Assoc* 1963, **58:**236-245.
30. Ashihara H, Sato F: **Pyrophosphate: Fructose-6-phosphate 1-phosphotransferase and biosynthetic capacity during differentiation of hypocotyls of Vigna seedlings.** *Biochim Biophys Acta* 1993, **1156:**123-127.
31. Farr TJ, Huppe HC, Turpin DH: **Coordination of chloroplastic metabolism in N-limited chlamydomonas reinhardtii by redox modulation (I. The activation of phosphoribulokinase and glucose-6-phosphate dehydrogenase is relative to the photosynthetic supply of electrons).** *Plant Physiol* 1994, **105:**1037-1042.
32. Woodrow IE, Berry J: **Enzymatic regulation of photosynthetic CO_2 fixation in C_3 plants.** *Ann Rev Plant Physiol Plant Molec Biol* 1988, **39:**533-594.
33. Lea PJ, Ireland RJ: **Nitrogen metabolism in higher plants.** In *Plant Amino Acids, Biochemistry and Biotechnology* Edited by: Singh BK. New York: Marcel Dekker, Inc; 1999:1-47.
34. Ireland RJ, Lea PJ: **The enzymes of glutamine, glutamate, asparagines, and aspartate metabolism.** In *Plant Amino Acids, Biochemistry and Biotechnology* Edited by: Singh BK. New York: Marcel Dekker, Inc; 1999:49-109.
35. Scheible WR, Krapp A, Stitt M: **Reciprocal diurnal changes of phosphoenol pyruvate carboxylase expression and cytosolic pyruvate kinase, citrate synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves.** *Plant Cell and Environ* 2000, **23:**1155-1167.
36. Noctor G, Novitskaya L, Lea PJ, Foyer CH: **Co-ordination of leaf minor amino acid contents in crop species: Significance and interpretation.** *J Exp Bot* 2002, **53(370):**939-945.
37. Ferrario-Méry S, Suzuki A, Kunz C, Valadier MH, Roux Y, Hirel B, Foyer CH: **Modulation of amino acid metabolism in transformed tobacco plants deficient in Fd-GOGAT.** *Plant and Soil* 2000, **221:**67-79.
38. Rao IM, Arulanantham AR, Terry N: **Diurnal changes in adenylates and nicotinamide nucleotides in sugar beet leaves.** *Photosynthesis Res* 1990, **23:**205-212.
39. Bonzon M, Hug M, Wagner E, Greppin H: **Adenine nucleotides and energy charge evolution during the induction of flowering in spinach leaves.** *Planta* 1981, **152:**189-194.
40. Stitt M, Lilley RM, Heldt HW: **Adenine nucleotide levels in the cytosol, chloroplasts, and mitochondria of wheat leaf protplasts.** *Plant Physiol* 1982, **70:**971-977.
41. Chen LS, Nose A: **Day-Night changes of energy-rich compounds in crassulacean acid metabolism (CAM) species utilizing hexose and starch.** *Ann Bot* 2004, **94:**449-455.
42. Leegood RC: **Photosynthesis in C_3 plants: The Benson-Calvin cycle and photorespiration.** In *Plant Biochemistry and Molecular Biology* 2nd edition. Edited by: Lea PJ, Leegood RC. John Wiley & Sons Ltd; 1999:29-50.
43. Hampp R, Schnabl H: **Adenine and pyridine nucleotide status of isolated Vicia guard cell protoplasts during K+-induced swelling.** *Plant and Cell Physiol* 1984, **25(7):**1233-1239.
44. May MJ, Vernoux T, Leaver C, Montagu MV, Inze D: **Glutathione homeostasis in plants: Implications for environmental sensing and plant development.** *J Exp Bot* 1998, **49(321):**649-667.
45. Lovaas E: **Antioxidant and metal-chelating effects of polyamines.** In *Advances in Pharmacology. Antioxidants in Disease Mechanisms and Therapy 38* Edited by: Sies H. New York: Academic Press; 1996:119-149.
46. Farré EM, Tiessen A, Roessner U, Geigenberger P, Trethewey RN, Willmitzer L: **Analysis of the compartmentation of glycolytic intermediates, nucleotides, sugars, organic acids, amino acids, and sugar alcohols in potato tubers using a nonaqueous fractionation method.** *Plant Physiol* 2001, **127:**685-700.
47. Kasugai S: **Studies of water culture.** *Jpn J Soil Sci Plant Nutr* 1939, **13:**669-822. (in Japanese)
48. Soga T, Ross GA: **Simultaneous determination of inorganic anions, organic acids, amino acids and carbohydrate by capillary electrophoresis.** *J Chromatogr A* 1999, **837:**231-239.
49. **SOM-PAK** [http://www.cis.hut.fi/research/som_lvq_pak]

ORIGINAL PAPER

Vol. 24 no. 14 2008, pages 1590–1595
doi:10.1093/bioinformatics/btn240

Gene expression

Aligning LC peaks by converting gradient retention times to retention index of peptides in proteomic experiments

Kosaku Shinoda[1,2], Masaru Tomita[1,2] and Yasushi Ishihama[1,3,*]

[1]Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0017, [2]Human Metabolome Technologies, Inc., Tsuruoka, Yamagata 997-0052 and [3]PRESTO, Japan Science and Technology Agency, Sanbancho Bldg., 5-Sanbancho, Chiyodaku, Tokyo 102-0075, Japan

Received on February 14, 2008; revised on April 29, 2008; accepted on May 17, 2008

Advance Access publication May 19, 2008

Associate Editor: David Rocke

ABSTRACT

Motivation: Liquid chromatography-tandem mass spectrometry (LC-MS/MS) is a powerful tool in proteomics studies, but when peptide retention information is used for identification purposes, it remains challenging to compare multiple LC-MS/MS runs or to match observed and predicted retention times, because small changes of LC conditions unavoidably lead to variability in retention times. In addition, non-contiguous retention data obtained with different LC-MS instruments or in different laboratories must be aligned to confirm and utilize rapidly accumulating published proteomics data.

Results: We have developed a new alignment method for peptide retention times based on linear solvent strength (LSS) theory. We found that log k_0 (logarithm of retention factor for a given organic solvent) in the LSS theory can be utilized as a 'universal' retention index of peptides (RIP) that is independent of LC gradients, and depends solely on the constituents of the mobile phase and the stationary phases. We introduced a machine learning-based scheme to optimize the conversion function of gradient retention times (t_g) to log k_0. Using the optimized function, t_g values obtained with different LC-MS systems can be directly compared with each other on the RIP scale. In an examination of *Arabidopsis* proteomic data, the vast majority of retention time variability was removed, and five datasets obtained with various LC-MS systems were successfully aligned on the RIP scale.

Contact: y-ishi@ttck.keio.ac.jp

1 INTRODUCTION

Liquid chromatography-mass spectrometry (LC-MS) is a powerful tool for the separation and identification of peptides in proteomics studies. While several methods and software tools are available for identifying peptides/proteins from mass spectra, the high complexity of a digested proteome and the vastly larger number of possible peptide sequences make accurate peptide/protein identification challenging. As the chromatographic retention times of peptides depend on their amino acid sequences, their retention times complement the information provided by MS and thus enhance their identifiability (Palmblad *et al.*, 2002; Petritis *et al.*, 2003).

Comparing multiple LC-MS/MS runs or matching observed and predicted retention times for identification purposes remains a challenging issue, because small changes in flow rate, column length, column packing, void volume and mobile phase composition unavoidably lead to variability in retention times. In addition, it was recently reported that even changing pore size of chromatographic beads as well as the ion-pair reagents such as trifluoroacetic acid, heptafluorobutyric acid and acetic acid in the mobile phase affects the peptide retention times significantly (Ishihama *et al.*, 2008; Krokhin, 2006). Furthermore, non-contiguous retention data obtained with different LC-MS instruments or in different laboratories must be aligned to confirm and utilize published proteomics data.

A widely used approach to the chromatographic-alignment problem is to fit a piecewise linear function to maximize the correlation between the samples. Methods of this kind are often characterized as correlation optimized warping (COW) (Nielsen *et al.*, 1998), and several derivative methods have been investigated (van Nederkassel *et al.*, 2006). In principle, this approach can be extended to aligning multi-dimensional data. However, the handling of proteomics data is extremely difficult because the data are typically characterized by a very large input dimension (i.e. tryptic peptides). Thus, more sophisticated alignment algorithms are needed to extract higher quality information from large-scale LC-MS-based experiments.

Several approaches for the alignment of peptide retention times have been developed and applied to high-throughput proteomics. For example, in the accurate mass and time tag (AMT) approach (Callister *et al.*, 2006; Jaitly *et al.*, 2006; Norbeck *et al.*, 2005; Smith *et al.*, 2002; Zimmer *et al.*, 2006), results from different LC-MS or MS/MS datasets are combined by finding the conversion functions of mass and retention times that are required to remove variability in mass and retention time measurements between analyses. Machine learning has also been applied to develop an 'intelligent' system for comparing large numbers of LC/MS experiments. The genetic algorithm (GA) has enabled the optimization of two variables of the linear normalization function for each LC separation so as to reduce the variance function of specific peptides, i.e. the regressed retention times for each separation (Petritis *et al.*, 2003). While this approach has generated excellent results, the normalization approach becomes time-prohibitive as the number of peptides used increases significantly, due to the many generations (iterations) required to

*To whom correspondence should be addressed.

align all analyses (Petritis *et al.*, 2006). To remove this limitation, Strittmatter *et al.* (2003) regressed, observed retention times of confidently identified peptides to predicted normalized elution time (NET) of the sequences using a quadratic function for each LC-MS run. The obtained quadratic equations were used to convert observed retention times to observed NET, and all LC-MS runs could be compared on scales of the NET. However, due to their use of an in-house-built nanoflow pump with ultrahigh pressure tolerance, it would be difficult to apply their NET scale to other datasets obtained with commercial systems in other proteomics laboratories, because their nanoflow pump generates exponential gradient curves depending on the flow-rate (Shen *et al.*, 2001).

Here we report the development of a new alignment method using $\log k_0$ (logarithm of retention factor for a given organic solvent) from linear solvent strength (LSS) theory (Stadalius *et al.*, 1984). Peptide LC-MS data are aligned by converting different gradient retention time scales to a single scale of predicted $\log k_0$. We introduce a GA to optimize the conversion function between retention times and $\log k_0$. Using the optimized function, peptide retention times obtained from different gradients and/or LC-MS systems can be compared with each other on the same $\log k_0$ scale. Unlike other functional optimization-based alignment techniques, realignments after each new experiment are not required, and thus the technical weaknesses of GA are overcome. The new method was applied to the soluble fraction of *Arabidopsis* cells and datasets obtained with various LC-MS systems were successfully aligned.

2 MATERIALS AND METHODS

2.1 Preparation of cell lysates

Escherichia coli MC4100 cells (see Section 3.1) were grown at 37°C in rich medium as described (Kerner *et al.*, 2005), and were lyzed by ultrasonication and centrifuged at 3000 × *g* for 10 min to collect the supernatants. *Arabidopsis* (ecotype Landsberg erecta) cells were a generous gift from Dr H. Nakagami (Riken, Yokohama, Japan). The frozen cells were disrupted with a Multi-beads shocker (MB400U, Yasui Kikai, Tokyo, Japan) and suspended in 0.1 M Tris–HCl (pH 8.0). The supernatants were collected by centrifugation at 1500*g* for 10 min.

2.2 Sample preparation

Proteins from these cell lysates were dried and resuspended in 50 mM Tris–HCl buffer (pH 9.0) containing 8 M urea. The mixtures were individually reduced with dithiothreitol (DTT), alkylated with iodoacetamide and digested with Lys-C, followed by dilution and trypsin digestion as described (Saito *et al.*, 2006). The digested samples were then desalted using StageTips with C18 Empore disk membranes (Rappsilber *et al.*, 2007).

2.3 NanoLC-MS/MS analysis

All samples were analyzed by nanoLC-MS/MS using a QSTAR Pulsar i mass spectrometer (AB/MDS-Sciex, Toronto, Canada) equipped with an Agilent 1100 nanoflow pump (Waldbron, Germany) or an LTQ-Orbitrap mass spectrometer (Thermofisher, Bremen, Germany) with a Dionex Ultimate 300 pump. In both systems, an HTC-PAL autosampler (CTC Analytics AG, Zwingen, Switzerland) equipped with a Valco C2 valve with 150 μm ports as an injection valve was used. ReproSil-Pur 120 C18-AQ materials (3 μm, Dr Maisch, Ammerbuch, Germany) were packed into a self-pulled needle (100 μm ID, 6 μm opening, 150 mm length) with a nitrogen-pressurized column loader cell (Nikkyo Technos, Tokyo, Japan) to prepare an analytical column needle with 'stone-arch' frit (Ishihama *et al.*, 2002). A spray voltage of 2400 V was applied via the metal connector as described (Ishihama *et al.*, 2002). The injection volume was 5 ml and the flow rate was 500 nl/min. The mobile phases consisted of (A) 0.5% acetic acid in water and (B) 0.5% acetic acid in 80% acetonitrile. Four linear gradient conditions of 5% B to 60% in 30, 60, 120 and 180 min were employed. Four MS/MS scans (0.6 s each) per one MS scan (1 s) were performed with the QSTAR, whereas the top 10 precursors were selected for MS/MS scans for the LTQ-Orbitrap. The scan range was m/z 350–1400 for the QSTAR and 300–1500 for the LTQ-Orbitrap.

2.4 Data analysis

MS peak lists were created by scripts in Analyst QS (MDS-Sciex) on the basis of the recorded fragmentation spectra, and were submitted to the Mascot database search engine (Matrix Science, London, UK) against the SwissProt database (release 45.0) to identify proteins from *E.coli* samples, while the TAIR version 7 (April 25, 2007) database was used for *Arabidopsis* samples. The following search parameters were used in all Mascot searches: maximum of two missed trypsin cleavages, cycteine carbamidomethylation as a fixed modification and methionine oxidation as a variable modification. A precursor mass tolerance of 0.2 Da and a fragment ion mass tolerance of 0.2 Da were set for the QSTAR, whereas a precursor mass tolerance of 3 p.p.m. and a fragment ion mass tolerance of 0.8 Da were used for the LTQ-Orbitrap. All peptides with scores less than the identity threshold ($P \geqslant 0.05$) or a rank >1 were automatically discarded.

2.5 Measurement of retention factors from gradient analysis

The reversed-phase retention factor k is generally described as

$$\log k = \log k_0 - S\phi \tag{1}$$

where ϕ is the volume fraction of the less polar component in the water–organic mobile phase, k_0 is the value of k for the solute at the start of the gradient in the initial mobile phase ($\phi = 0$) and S is a constant characteristic for a given analyte and chromatographic system (Stadalius *et al.*, 1984).

Solute retention time t_g in gradient elution is given as

$$t_g = \left(\frac{t_0}{b}\right)\left[\log 2.3 k_0 b\left(\frac{t_{sec}}{t_0}\right)+1\right]+t_{sec}+t_D \tag{2}$$

where t_0 is the column dead-time for a small solute molecule, t_{sec} is the value of t_0 for the solute in question, t_D is the dwell-time of the gradient system and b is a gradient parameter defined by

$$b = S\Delta\phi t_0 / t_G \tag{3}$$

Here the quantity t_G is the gradient time and $\Delta\phi$ is the change in ϕ during the gradient ($\Delta\phi = 1$ for a 0–100% gradient) (Snyder, 1980). For smaller solutes and larger pore particles, Equation (2) can be approximated by

$$t_g = \left[\frac{t_G}{(S\Delta\phi)}\right]\log\left[2.3 k_0 t_0\left(\frac{S\Delta\phi}{t_G}\right)+1\right]+t_0+t_D \tag{4}$$

By solving Equation (4) for k_0, Equation (5) is derived:

$$k_0 = \frac{t_G\left(-1+10^{-t_G/\Delta\phi S(t_0+t_D-t_g)}\right)}{\Delta\phi 2.3 S t_0} \tag{5}$$

Four gradient elution runs were performed for *E.coli* samples as described above, and the observed t_g, $t_G/\Delta\phi$, t_0, t_D values were substituted into Equation (4). A Microsoft Excel multi-line fitting program based on the semi-Newton method was run to optimize S and k_0 values in order to minimize the sum of the differences between calculated and observed t_g values. The obtained S and k_0 were used as observed values for further analysis.

Table 1. Experimental parameters for the genetic algorithm used

Experimental parameters	Parameter value
Number of maximum generation G	200
Number of individual P	500
Crossover ratio c (%)	45
Crossover strategy	Uniform
Selection strategy	Roulette
Mutation ratio m (%)	45

2.6 Implementation of the algorithm

Obtained k_0 was used to construct the $\log k_0$ predictor. We employed a three-layer artificial neural network (ANN) with back-propagation learning. A sigmoid function was applied to each node in the ANN. To reduce unnecessarily large parameters (weights) among nodes, the pruning method was used as described (Shinoda *et al.*, 2006). The ANN software used was JMP software, version 6.0.2 (SAS Institute, Cary, NC, USA). Experimental retention time (t_g) was converted to predicted $\log k_0$ using Equation (5) containing several parameters ($t_G/\Delta\phi$, t_0, t_D, S). Among them, S was predicted for each identified peptide using a previously reported ANN based on the dependence of S on the amino acid composition (Ishihama, 2006), and the remaining parameters ($t_G/\Delta\phi$, t_0, t_D) were optimized using GA. Our GA was implemented in Perl language with the AI::Genetic module from CPAN (www.cpan.org). The numerical experimental conditions are shown in Table 1. These computational portions of our work were performed on a Pentium 4 Xeon 2.0 GHz CPU.

3 RESULTS AND DISCUSSION

3.1 Prediction of $\log k_0$ using an ANN

We analyzed *E.coli* samples under four different linear gradient conditions and obtained the data pairs of $\log k_0$ and S for 278 peptides. The correlation coefficients between observed and calculated t_g values per each peptide ranged from 0.9993 to 1.000 for four data points from 30 min to 180 min gradient runs, indicating that LSS theory was valid for the peptides in this range. In order to predict $\log k_0$ values from peptide sequences, we trained an ANN using the number of residues of each amino acid in the identified *E.coli* peptides as inputs and obtained $\log k_0$ as outputs based on the assumption that $\log k_0$ of peptides depends on amino acid composition. We adopted three-layer architecture for the ANN because it could approximate any function (Funahashi, 1989). We tried hidden nodes ranging from 2 to 10, and the $\log k_0$ response curves of each input variable, constituting an approximate function from sampled values, were used as the criteria for determining the number of hidden nodes. We added hidden nodes until the response curves were not too flexible or non-linear. Consequently, we adopted five nodes in the hidden layer and our ANN had a 20-5-1 architecture. Other parameters for ANN training (training ratio, momentum and random numbers for initial ANN weights) were determined empirically. Each of the trainings was continued until the epoch (iteration) reached 100 or until improvement of the optimization function fell below a learning convergence criterion. Figure 1 is a global comparison between predicted and measured $\log k_0$ for 278 peptides through 10-fold two-deep cross-validations (Jonathan *et al.*, 2000). Overall, our results were satisfactory; the coefficient of determination (R^2) was 0.8895 and the mean prediction error

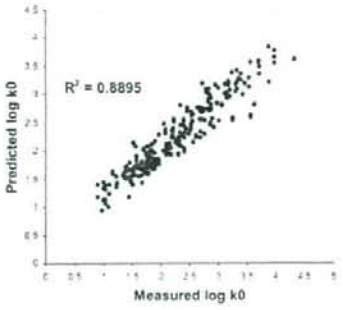

Fig. 1. The correlation between experimentally measured and predicted $\log k_0$ for all peptides derived from *E.coli* K12 proteome through 10-fold two-deep cross-validations.

was 0.189±7.1% (relative standard deviation, RSD). These results support the validity of our assumption that the $\log k_0$ of peptides depends on amino acid composition. ANNs have recently been utilized for accurate modeling of peptide retention time (Petritis *et al.*, 2003, 2006; Shinoda *et al.*, 2006), but application to $\log k_0$ prediction has not yet been reported. We used this ANN predictor for the following GA-based optimization of the conversion function.

The scheme of our alignment approach is illustrated in Figure 2. S values of identified peptides were computationally predicted using a previously reported ANN (Ishihama, 2006) from amino acid composition. The ANN predictor eliminated the need for multiple chromatographic runs for derivations of S and enabled experimental $\log k_0$ to be obtained from a single LC-MS run. On the other hand, the constructed ANN enabled predicted $\log k_0$ to be obtained from the amino acid composition of peptides. The conversion function [Equation (5)] was optimized with a GA using the sum of squared errors (SSE) function between experimental and predicted $\log k_0$ as an evaluation function. Optionally, we adjusted GA-optimized $\log k_0$ values using the linear relationship, if necessary. This conversion enabled various LC gradient data to be compared on the same scale of RIP. RIP is a converted $\log k_0$ scale on a time scale of the $\log k_0$ predictor, which is specific for a given set of gradient analyses with a given mobile phase and columns, i.e. the *E.coli* dataset in this article. Using the optimized function, peptide retention times obtained from different LC-MS systems and/or gradients can be directly compared on the same RIP dimension and easily aligned.

3.2 Application to *Arabidopsis* proteome data

To demonstrate the usability of our alignment algorithm, we conducted an independent validation study with real complex samples (*Arabidopsis* cells). The proteomics sample was prepared according to the above protocol and analyzed using two different LC-MS systems under five different LC conditions (Table 2). Peptides were identified for each LC-MS run using Mascot. The number of identified peptides was 605, 1050, 980, 3861 and 5719 for conditions 1–5, respectively. Experimental t_g was converted to RIP using Equation (5). Parameters were optimized using the GA so that the difference between predicted and converted RIP of identified peptides was minimized. We used the GA because it

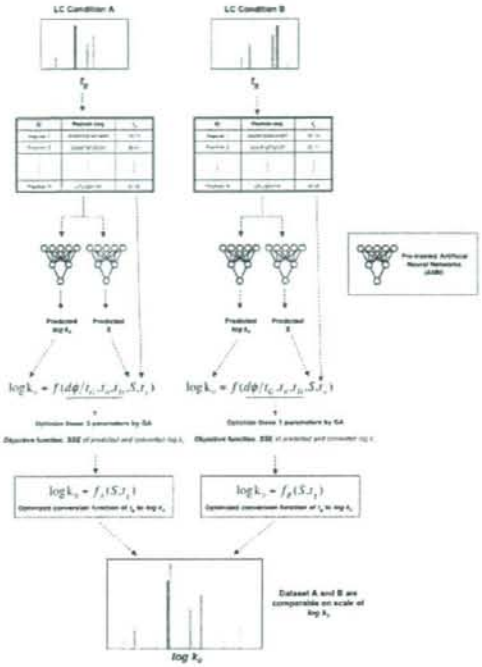

Fig. 2. Schematic flowchart depicting the method. In this example, peptide retention data (t_g) obtained with two different LC-MS systems (A and B) are aligned. S and $\log k_0$ of identified peptides are computationally predicted using a pretrained ANN based on amino acid composition determined by MS/MS ion search (e.g. Mascot). Parameters of the conversion functions of t_g to $\log k_0$ are optimized for each LC condition based on the predicted S and predicted $\log k_0$ values using a GA. The objective function is the SSE between predicted and converted (experimental) $\log k_0$. After functional optimization, datasets A and B are comparable on the same $\log k_0$ (RIP) scale. This algorithm is easily expandable to three or more samples.

can determine many parameters simultaneously with high accuracy, and selected the real-coded GA (Janikow and Michalewicz, 1991) because it improves the optimization speed compared with the conventional binary GA. The time required for one trial was ~1 h. The experiments were conducted in 50 trials with different random seeds. Comparison of the trajectories shows that fitness values decreased until ~60 generations (Fig. 3). The R^2 between the predicted and experimental $\log k_0$ was 0.9604–0.9968. These results indicate the value of GA in functional optimization for gradient retention time conversion. Unlike traditional non-linear regression, GA-based approaches offer advantages that include a capacity to self-learn and to obtain optimized parameters without the need for time-consuming manual tunings and detailed understanding of the characteristics of functions.

The results of conversion using the optimized function are shown in Figure 4. The converted $\log k_0$ (RIP) of peptides identified among the different LC conditions are plotted. On the RIP scale, most

Table 2. LC system and gradients used for method validation

Condition	LC-MS systems	Gradient (min)	Column
1	Agilent1100-QSTAR	30	Column 1 (100 μm ID/8 cm L)
2	Agilent1100-QSTAR	60	Column 1 (100 μm ID/8 cm L)
3	Agilent1100-QSTAR	60	Column 2 (100 μm ID/15 cm L)
4	Ultimate3000-Orbitrap	60	Column 2 (100 μm ID/15 cm L)
5	Ultimate3000-Orbitrap	120	Column 2 (100 μm ID/15 cm L)

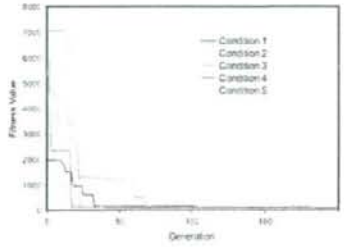

Fig. 3. Changes in the fitness values of best-of-generation individuals for each experimental condition (1–5).

peptides are on the locus of $y=x$ (Spearman $r=0.9863$–0.9988) despite the difference of columns (**A**), systems (**B**) and gradients (**C**). RIP was still effective where columns, systems and gradients were all different (**D**). Using RIP, retention of commonly identified peptides can be compared on the same scale and we can easily validate proteomic data across various LC-MS systems. Our method is more effective when three or more different LC-MS datasets should be aligned. RIP is a general parameter, and thus reoptimization is not required even when a new dataset for comparison is added.

3.3 Probability scores and Δ RIP

As RIP depends on the amino acid sequence, comparison of predicted and experimental (converted) RIP allows validation of peptide sequences determined by MS/MS ion search, i.e. peptides which have ΔRIP above a certain level are more likely to be false positives. The relationship between Mascot probability score, which indicates reliability of peptide identification, and ΔRIP for *Arabidopsis* data is shown in Figure 5. This showed a negative correlation between ΔRIP and score. Peptides with low reliability (probability score <16) have a larger proportion of 'outlier' peptides, while ΔRIP of a majority of reliable (>95%) peptides is less than 0.5. This indicates the validity of the converted RIP and our predictors. Among the reliable peptides, the threshold value of a 5% outlier in ΔRIP was 0.552. This result indicates that peptide

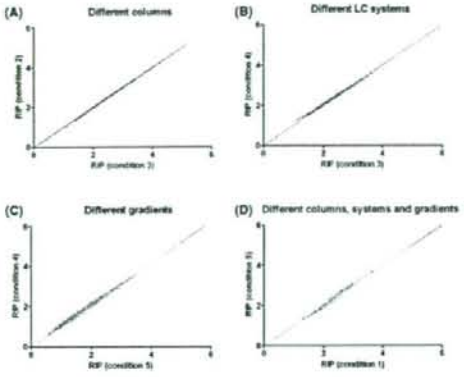

Fig. 4. Correlation of converted RIP among commonly identified peptides between experiments with different columns, LC systems and/or gradients. Black slant line indicates $y = x$.

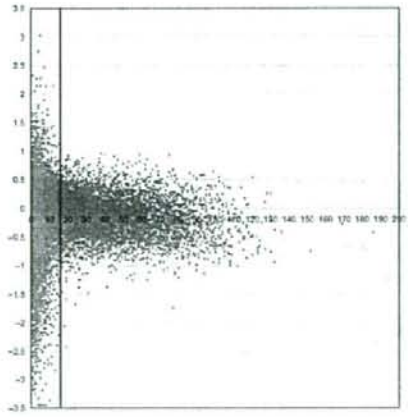

Fig. 5. Relationship between ΔRIP (predicted–experimental) and Mascot probability score. The result for condition 5 (Table 2) is shown. The bold vertical line indicates probability score 16 (>16 scores indicate >95% reliability).

identification where ΔRIP is more than 0.55 is very likely to be a misidentification.

4 CONCLUSION

We have developed a new alignment method for LC-MS-based proteomics data using GA-based optimization of the conversion function between gradient retention times and the logarithm of retention factor ($\log k_0$). The method was applied to the soluble fraction of *Arabidopsis* cells, and five datasets obtained with different LC gradients were appropriately aligned. Converted $\log k_0$ (RIP) values can be used between laboratories as long as the stationary phase and the mobile phase are identical. This method should be useful for comparing proteomics datasets between laboratories and for utilizing the rapidly accumulating published proteomics LC-MS data. In addition, this method is also applicable for peptide mixtures containing partially modified amino acid residues such as phosphorylated serine, threonine and tyrosine. Since the post-translational modifications (PTM) such as phosphorylation are quite important to understand cellular functions, this method would be helpful to perform PTM proteome analysis. Further studies are in progress in our laboratory.

ACKNOWLEDGEMENTS

We thank Yasuyuki Igarashi and Mikiko Hattori (Keio University) for their technical support.

Funding: This work was supported by research funds from the Yamagata prefectural government and Tsuruoka city.

Conflict of Interest: none declared.

REFERENCES

Callister,S.J. *et al.* (2006) Application of the accurate mass and time tag approach to the proteome analysis of sub-cellular fractions obtained from Rhodobacter sphaeroides 2.4.1. Aerobic and photosynthetic cell cultures. *J. Proteome Res.*, **5**, 1940–1947.

Funahashi,K. (1989) On the approximate realization of continuous mappings by neural networks. *Neural Netw.*, **2**, 183–192.

Ishihama,Y. (2006) Method for detection of peptide sequence based on chromatography retention time. PCT/JP2006/315549.

Ishihama,Y. *et al.* (2002) Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A*, **979**, 233–239.

Ishihama,Y. *et al.* (2008) Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics*, **9**, 102.

Jaitly,N. *et al.* (2006) Robust algorithm for alignment of liquid chromatography-mass spectrometry analyses in an accurate mass and time tag data analysis pipeline. *Anal. Chem.*, **78**, 7397–7409.

Janikow,C.Z. and Michalewicz,Z. (1991) An experimental comparison of binary and floating point representations in genetic algorithms. In *Proceedings of the Fourth International Conference on Genetic Algorithms*. Morgan Kaufmann, San Diego, CA, USA, pp. 31–36.

Jonathan,P. *et al.* (2000) On the use of cross-validation to assess performance in multivariate prediction. *Stat. Comput.*, **10**, 209–229.

Kerner,M.J. *et al.* (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell*, **122**, 209–220.

Krokhin,O.V. (2006) Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. *Anal. Chem.*, **78**, 7785–7795.

Nielsen,N.-P.V. *et al.* (1998) Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr. A*, **805**, 17–35.

Norbeck,A.D. *et al.* (2005) The utility of accurate mass and LC elution time information in the analysis of complex proteomes. *J. Am. Soc. Mass Spectrom.*, **16**, 1239–1249.

Palmblad,M. *et al.* (2002) Prediction of chromatographic retention and protein identification in liquid chromatography/mass spectrometry. *Anal. Chem.*, **74**, 5826–5830.

Petritis,K. *et al.* (2003) Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. *Anal. Chem.*, **75**, 1039–1048.

Petritis,K. *et al.* (2006) Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information. *Anal. Chem.*, **78**, 5026–5039.

Rappsilber,J. *et al.* (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2**, 1896–1906.

Saito,H. *et al.* (2006) Multiplexed two-dimensional liquid chromatography for MALDI and nanoelectrospray ionization mass spectrometry in proteomics. *J. Proteome Res.*, **5**, 1803–1807.

Shen,Y. *et al.* (2001) Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem.*, **73**, 1766–1775.

Shinoda,K. *et al.* (2006) Prediction of liquid chromatographic retention times of peptides generated by protease digestion of the Escherichia coli proteome using artificial neural networks. *J. Proteome Res.*, **5**, 3312–3317.

Smith,R.D. *et al.* (2002) An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics*, **2**, 513–523.

Snyder,L.R. (1980) *High Performance Liquid Chromatography: Advances and Perspectives*. Academic Press, New York.

Stadalius,M.A. *et al.* (1984) Optimization model for the gradient elution separation of peptide mixtures by reversed-phase high-performance liquid chromatography: verification of retention relationships. *J. Chromatogr. A*, **296**, 31–59.

Strittmatter,E.F. *et al.* (2003) Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **14**, 980–991.

van Nederkassel,A.M. *et al.* (2006) A comparison of three algorithms for chromatograms alignment. *J. Chromatogr. A*, **1118**, 199–210.

Zimmer,J.S. *et al.*, (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom. Rev.*, **25**, 450–482.

Construction of a Biological Tissue Model Based on a Single-Cell Model: A Computer Simulation of Metabolic Heterogeneity in the Liver Lobule

Hiroshi Ohno[**,†]
Keio University

Yasuhiro Naito[*,**,†,‡]
Keio University

Hiromu Nakajima[§]
Osaka Medical Center
for Cancer and
Cardiovascular Diseases

Masaru Tomita[**,†,‡]
Keio University

Abstract An enormous body of information has been obtained by molecular and cellular biology in the last half century. However, even these powerful approaches are not adequate when it comes to higher-level biological structures, such as tissues, organs, and individual organisms, because of the complexities involved. Thus, accumulation of data at the higher levels supports and broadens the context for that obtained on the molecular and cellular levels. Under such auspices, an attempt to elucidate mesoscopic and macroscopic subjects based on plentiful nanoscopic and microscopic data is of great potential value. On the other hand, fully realistic simulation is impracticable because of the extensive cost entailed and enormous amount of data required. Abstraction and modeling that balance the dual requirements of prediction accuracy and manageable calculation cost are of great importance for systems biology. We have constructed an ammonia metabolism model of the hepatic lobule, a histological component of the liver, based on a single-hepatocyte model that consists of the biochemical kinetics of enzymes and transporters. To bring the calculation cost within reason, the porto-central axis, which is an elemental structure of the lobule, is defined as the systems biological unit of the liver, and is accordingly modeled. A model including both histological structure and position-specific gene expression of major enzymes largely represents the physiological dynamics of the hepatic lobule in nature. In addition, heterogeneous gene expression is suggested to have evolved to optimize the energy efficiency of ammonia detoxification at the macroscopic level, implying that approaches like this may elucidate how properties at the molecular and cellular levels, such as regulated gene expression, modify higher-level phenomena of multicellular tissue, organs, and organisms.

Keywords
Zonal metabolic heterogeneity, hepatic lobule, biological simulation, ammonia metabolism

* Contact author.
** Institute for Advanced Biosciences, Keio University, 14-1 Baba-cho, Tsuruoka, 997-0035, Japan. E-mail: n02139ho@sfc.keio.ac.jp
† Bioinformatics Program, Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa, 252-8520, Japan.
‡ Department of Environmental Information, Keio University, 5322 Endo, Fujisawa, 252-8520, Japan. E-mail: ynaito@sfc.keio.ac.jp (Y.N.); mt@sfc.keio.ac.jp (M.T.)
§ Clinical Laboratory, Osaka Medical Center for Cancer and Cardiovascular Diseases, 1-3-3 Nakamichi, Higashinariku, Osaka, 537-0025, Japan. E-mail: nakajima-hi@mc.pref.osaka.jp

1 Introduction

Systems biology is intended to elucidate the dynamics of the biological cell based on a compendium of data carried out by what might be called the "ome" and the "omics" sciences, such as the genome and genomics, the proteome and proteomics, or the metabolome and metabolomics [21, 26]. Meanwhile, biological life, which it is hoped will come to be understood, is rarely restricted to a single cell, or unicellular organism. The human being is a representative multicellular organism, consisting of approximately 60 trillion cells, and undoubtedly the subject that we would like to investigate most intensively. At present, the greater part of the data that support systems biology is information at the molecular and cellular (nanoscopic and microscopic) levels. While the explosive development of molecular and cellular biology has yielded both copious and precise information at the subcellular level, biology for higher-level (mesoscopic and macroscopic) structures has lagged far behind. Anatomy and histology represent an organism in a hierarchical classification scheme, namely of tissues, organs, and individual, going from the microscopic to macroscopic. The store of knowledge built up at each level of the hierarchy is at present disproportionate. The knowledge accumulated in the last decade at higher levels than the cell is undoubtedly less than that at the cellular and subcellular levels. A major constraint is the currently limited technology, which for the tissue or organ level presents greater difficulties in all aspects of sample preparation, cultivation, and measurement than required for the single-cell level.

Hitherto, nanoscopic and microscopic mathematical models for the life sciences have been isolated from mesoscopic and macroscopic ones in most cases, while mesoscopic and macroscopic entities inevitably consist of macroscopic entities. Here we report an effort to expand the knowledge of cellular, subcellular, and molecular levels to higher levels such as tissues and organs through biological simulation, which is one of the main contributions of systems biology. We are aware that such a bottom-up approach is not always appropriate; for instance, a description of solid-body motion with quantum mechanics is almost always pointless. However, given the overwhelming shortage of quantitative information at the mesoscopic and macroscopic levels, a serious effort to elucidate higher-level behaviors of life based on the abundant molecular and cellular data collected is of value whatever the difficulty entailed. To construct a tissue or organ model based on a single-cell model, simple aggregation of the cell model into a higher-level model is essentially inadequate. Not only organs of complicated structure such as the heart and brain, but those of comparatively simple structure such as the liver and kidney, are impossible to model by simple multiplication of single cells, because each tissue or organ has its own intrinsic and exquisitely specific structure. The spatiotemporal position of the cells in the tissue or organ structure influences the behavior of each cell, and the total of the cellular behaviors naturally affects the behaviors of the higher structures—the tissue, organ, or organism—in turn. Therefore, how to effectively assemble the single-cell model(s), namely, how to represent the mesoscopic and macroscopic structure, is critically important for the modeling of tissues and/or organs.

For this study the rodent liver was selected as the subject of modeling because of (1) its comparatively simple histological structure among mammalian tissues and organs [37], and (2) the availability of data from previous investigations, including a great deal of kinetic data. The liver consists of a huge structure consisting of repetitions of a fundamentally simple structural unit, the hepatic lobule. The entire liver looks the like a bunch of grapes, in which the portal tracts (influx vessels) and the central vein (efflux vessel) supply each bunch (Figure 1A). The portal tracts are a bundle of small branches of the hepatic portal vein, which carries the absorbed food products directly from the gut to the liver, and the hepatic artery, which supplies oxygen to support liver metabolism. The central veins are confluent and form the hepatic vein. The hepatic lobule is roughly hexagonal in sectional configuration and is centered on a central vein. The portal tracts are positioned at the corners of the hexagon. The blood from the portal vein and hepatic artery branches in the portal tracts flows to a central vein (Figure 1B, C). The overwhelmingly predominant component of the liver is the parenchymal liver cell, called the hepatocyte. The hepatocytes form branching

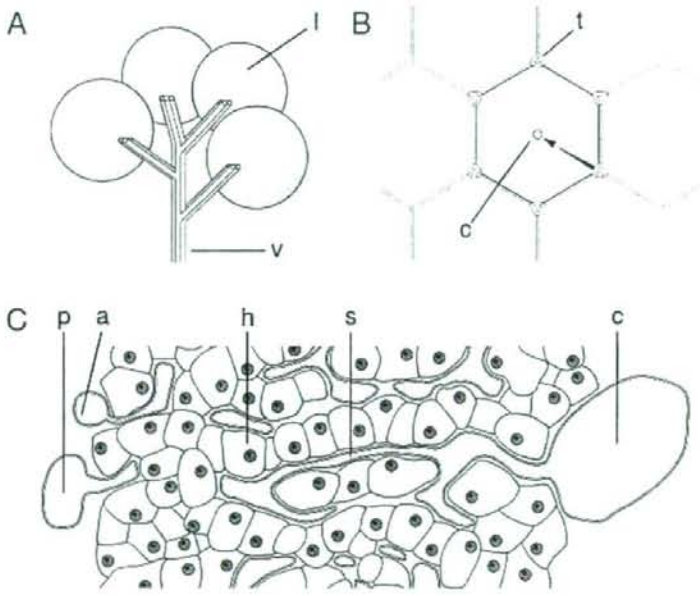

Figure 1. Schematic drawings of the connection of the hepatic lobules in the liver (A), hexagonal shape of a hepatic structure in section (B), and the porto-central axis (C). l, a hepatic lobule; v, influx and efflux vessels consist of terminal branches of the portal vein, hepatic artery, and hepatic vein; t, periportal tracts; c, central vein; p, terminal portal vein; a, terminal hepatic artery; s, sinusoid; h, hepatocyte. The arrow indicates the direction of sinusoidal blood flow.

cords of cells between capillary vessels, which are known as sinusoids in the hepatic lobule. Blood flow into the sinusoids comes from terminal branches of the portal tracts, bringing nutrient-rich blood from the portal vein and oxygen-rich blood from the hepatic artery. Blood from the portal vein and the hepatic artery passes through the sinusoids, where it comes into intimate contact with the hepatocytes for the exchange of nutrients and metabolic products. The blood then flows into branches of the hepatic vein and thence into the inferior vena cava. Therefore, the liver can be described histologically as a rough hierarchy of the hepatocyte (cell), the hepatic lobule (tissue), and the liver (organ).

Concentrations of nutrients and metabolites in the periportal zone (the upper reach of the sinusoid) are inevitably different from those in the perivenous zone (the lower reach) even if the properties of all the hepatocytes are homogeneous, because the downstream hepatocytes accept nutrients and metabolites already taken up and put there by the upstream hepatocytes. Thus a liver model constructed by multiplying single hepatocytes by the number of hepatocytes in the liver would not be representative, but it may be possible to investigate the global behavior of the liver by a simple multiplication of the hepatic lobule, since the liver is a repetitive accumulation of the hepatic lobules as functioning units. Nevertheless, since a human hepatic lobule consists of approximately one million hepatocytes (there are approximately 100 billion hepatocytes and 100,000 lobules in a human liver), simulation of the hepatic lobule at the molecular level is far from realizable with the computational resources currently available. Thereupon, we consider a sinusoid along the porto-central axis (the line between the portal tracts and the central vein) and the surrounding hepatocytes, known as hepatic cords (Figure 1C), to be the primary histological unit for modeling. Since the hepatic lobule has a point symmetry in which the center point is the central vein, the overall behavior of the hepatic lobule and the liver can be derived from that of the porto-central axis. The porto-central axis is the minimal unit preserving the proportionality with the dynamics of the greater, macroscopic structure. Through model construction of the porto-central axis, it becomes possible to investigate the behavior of the liver, at the organ level, based on molecular-level simulations of single hepatocytes.

A multitude of hepatocyte zonal heterogeneities are known [19, 23–25]. Fatty acid oxidation, gluconeogenesis, ureagenesis, amino acid conversion to glucose, cholesterol synthesis, and glutathione peroxidation and conjugation are predominant in the periportal zone; Glycolysis, liponeogenesis, glutamine formation from ammonia, monooxygenation, and glucuronidation are predominant in the perivenous zone. In addition to the heterogeneity generated by the structural constraints described above, enzyme activities regulated by the substrate concentrations in blood, the circulating hormone levels, the autonomic hepatic nerves, and the biomatrix are known to actively modify the conditions of the zonal heterogeneities. Such active regulation of heterogeneities must result from some gain in evolutionary fitness. Since fitness is necessarily the property not of molecules, genes, cells, tissues, or organs, but of classes of organisms [34], the advantage generated by heterogeneous regulation within an organ at the subcellular level must be measured to elucidate the functional activity of the organ that the regulation has evolved. At present, it is extremely hard for the experimental biosciences to measure the properties of an organism simultaneously at multiple levels such as the cell, tissue, and organism. In this study, we constructed a model of the porto-central axis, and obtained insight into the origin of one of the regulated heterogeneities in metabolism, ammonia detoxification.

Ammonia metabolism is one of the most important metabolic pathways for a mammal, since the accumulation of ammonia induces lethal uremia. Ammonia is detoxified exclusively by the hepatocyte in mammals. Accordingly, appropriate ammonia metabolism in the mammalian liver is directly reflected in the health of the organism. The efficiency of ammonia metabolism should therefore link directly with fluctuations of fitness. The fitness fluctuation related to ammonia metabolism may be captured and analyzed through calculation of the efficiency of ammonia detoxification in the porto-central axis unit. The pathway structure of ammonia metabolism around the core of the urea cycle is relatively simple among the major metabolic pathways, and there are a number of biological investigations and several mathematical models for single hepatocytes. The metabolism of ammonia and amino acids is also heterogeneous in the hepatic lobule. The periportal zone is characterized by a high capacity for uptake and catabolism of amino acids (except glutamate and aspartate) as well as for urea synthesis and gluconeogenesis. On the other hand, glutamine synthesis, ornithine transamination, and the uptake of vascular glutamate, aspartate, malate, and α-ketoglutarate are restricted to a small perivenous hepatocyte population [20]. Accordingly, in the hepatic lobule the major pathways for ammonia detoxification and urea and glutamine synthesis are anatomically separated from each other and represent in functional terms the sequence of a periportal low-affinity system (urea synthesis) and a perivenous high-affinity system (glutamine synthesis) for ammonia detoxification [13–18, 20]. Perivenous glutamine synthase-containing hepatocytes act as high-affinity scavengers for the ammonia that escapes the upstream urea-synthesizing compartment.

Three major enzymes involved in ammonia metabolism—carbamoyl phosphate synthase (CPS), ornithine aminotransferase (OAT), and glutamine synthase (GS)—have been characterized as being heterogeneously expressed (synthesized at different rates) in particular locations within the hepatic lobule [5, 9, 10, 22–24]. The regulation of gene expression of these enzymes may be mainly determined not by the substrate concentrations, the hormone levels, or neural regulation, but rather by the histological position of the hepatocyte in the lobule [12, 20, 36, 43, 44]. Additionally, ornithine aminotransferase is demonstrated to be coexpressed with glutamine synthase in perivenous hepatocytes in adult mouse liver, while there is no such coexpression in the kidney, intestine, and brain [30, 31]. One possible explanation of this phenomenon is that the colocalized enzymes lead to more efficient removal of ammonia because ornithine aminotransferase produces glutamate, which is a substrate for glutamine synthase [31]. However, these are still matters of speculation and only subjected to thought experiments. Positional regulation, such as regulation of gene expression, can modulate the pattern of intralobular heterogeneity. Such heterogeneity may appear to be dispensable, but for some reason it is actually present. This raises the question of the purpose for which heterogeneous gene expressions exist. While various possible determinants of the positional regulation of gene expression have been proposed (gradual change in the concentrations

of blood constituents, the extracellular matrix produced by the endothelium of the central venule, intrahepatic cell-cell interaction, etc.), the regulative function of heterogeneous gene expression has been developed evolutionally and has preserved whichever determinants are selected for over time.

Here we show that the positional regulation of gene expression may improve the energy efficiency of ammonia metabolism at the tissue and organ levels, using computer simulation of the porto-central axis unit. Improvement of ammonia detoxification in the liver means increased efficiency at the individual organism level, and this could very well amount to a gain in evolutionary fitness. This suggests that these particular forms of gene expression regulation may reasonably be characterized as advantageous adaptations evolved through natural selection.

2 The Model

We hypothesized that the heterogeneous expression of carbamoyl phosphate synthase, glutamine synthase, and ornithine aminotransferase improves the energy efficiency of ammonia clearance, and therefore increases the evolutionary fitness of the liver, as well as that of the organism. The number of ATP molecules required to eliminate one molecule of ammonia in the hepatic lobule is certainly an appropriate indicator of the energy efficiency of ammonia detoxification. Because ammonia is eliminated by multiple pathways, mainly urea synthesis and glutamine synthesis, the stoichiometry of total ammonia degradation is variable. To calculate the ATP consumption by ammonia metabolism, a simple model of the hepatic lobule was constructed, through a two-step procedure. In the first step, a single-compartment model that consisted of the hepatocyte and the sinusoid, named the *zone model*, was built (Figure 2), and unknown parameters were predicted, as will be described later. In the second step, we modeled the hepatic lobule as a simple compartment model in which eight modeled compartments were connected in series.

The zone model included 67 substances and 29 total reactions consisting of chemical reactions and transportations. Almost all of the known enzymes in the mammalian ammonia metabolism have been described by rate equations. Mathematical models for four enzymes were developed for a urea cycle model developed by Kuchel and his colleagues [29]. Carbamoyl phosphate synthase [7, 8], glutamine synthase [40], phosphate-dependent glutaminase (Glnase) [35, 41], *N*-acetyl glutamate synthetase (AGS) [2, 3], system N (the sodium-dependent glutamine transport system) [32], system L (the sodium-independent glutamine transport system) [32], the ammonia transport system (between the sinusoid and the hepatocellular cytoplasm), the glutamate transport system [11], and the urea transport system (from the hepatocellular cytoplasm to the sinusoid) were modeled using kinetic information obtained from previous publications (see Appendix 1 and Web supplements). The rest of the models, that is, the six enzymes and three transporters, were quoted from a ureagenesis model using a MetaNet graph (see Appendix A1.1) [28]. While MetaNet cannot be guaranteed to reproduce accurate enzyme kinetics [28], it was used with the expectation that it is able to estimate the rates of reactions roughly but sufficiently well.

The metabolite concentrations were determined mainly based on Kohn and his colleagues' work (see Web supplements, Table S1) [28]. The volume ratio of cytoplasm to mitochondria was assumed to be four to one [28], and that between the cytoplasm and the sinusoidal space one to one. Certain kinetic parameters were adjusted under an assumption of steady state. If the model is held in steady state, the following equations are adequately assumed:

$$v_{\mathrm{CPS}} = v_{\mathrm{OCT}} = v_{\mathrm{OTL}} = v_{\mathrm{ASS}} = v_{\mathrm{ASL}} \tag{1}$$

$$v_{\mathrm{GAMT}} = v_{\mathrm{GAT}} = v_{\mathrm{ASL}} - v_{\mathrm{Argase}} \tag{2}$$

$$v_{\mathrm{GOT}_m} = v_{\mathrm{GATL}} = v_{\mathrm{ASS}} = v_{\mathrm{GOT}_c} \tag{3}$$

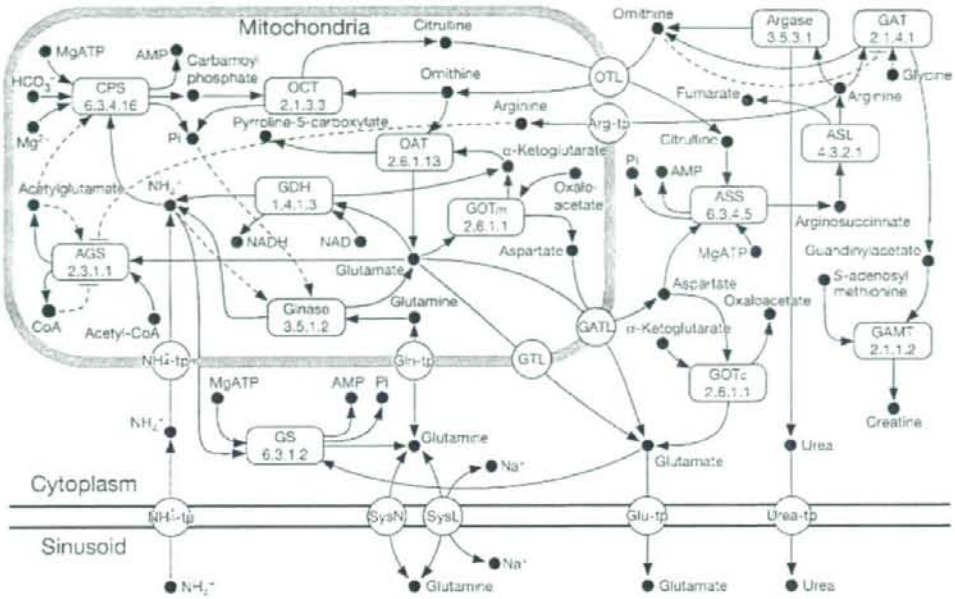

Figure 2. Schematic representation of the model describing ammonia metabolism in single zone model. Filled circles, open rounded rectangles, and open circles represent substances, enzymes, and transporters, respectively. Solid line arrows represent reactions and transportations. Broken lines with a triangular arrowhead and with a bar at one end represent positive and negative feedback, respectively. AGS, *N*-acetyl glutamate synthetase; Argase, arginase; ASL, argininosuccinate lyase; ASS, argininosuccinate synthetase; CPS, carbamoyl phosphate synthetase; GAMT, guanidinoacetate methyltransferase; GAT, arginine:glycine amidinotransferase; GDH, glutamate dehydrogenase; Glnase, phosphate-dependent glutaminase; GOT*c*, glutamate:oxaloacetate transaminase in the cytoplasm; GOT*m*, glutamate:oxaloacetate transaminase in the mitochondria; GS, glutamine synthetase; OAT, ornithine aminotransferase, OCT, ornithine carbamoyltransferase; Arg-tp, arginine transporter; GATL, glutamate-aspartate translocase; Gln-tp, glutamine transporter in mitochondrial membrane; Glu-tp, glutamate transporter; GTL, glutamate translocase; NH_4^+-tp, ammonia transporter in the cell membrane; NH_4^+-tp$_m$, ammonia transporter in the mitochondrial membrane; OTL, ornithine-citrulline translocase; SysL, system L; SysN, system N; Urea-tp, urea transporter. The entity abbreviation may be used with an index variable that represents the location of the entity. The indices *c*, *m*, and *s* indicate the cytoplasm, mitochondria, and sinusoid, respectively. Numbers in the rounded rectangles represent EC numbers.

Thus, [CPS], [OCT], [OTL], [ASS], [GAMT], [GAT], [GOT$_m$], and [GATL] were determined, based on the premise of [ASL] = 2.2E−6 M [29], v_{Argase} = 6.6E−6 M s^{-1} [29], and [GOT$_c$] = 1E−6 M [28]. The k_{cat}'s of glutaminase and glutamine synthase were calculated from the fluxes through the enzymes in the perfused rat liver: 72 and 151 nmol min^{-1} per gram [19], with the assumptions [GS] = 1E−5 M and [Glnase] = 1E−4 M. The activity of glutaminase was approximated as 1.786E−5 M s^{-1} based on reported conversion factors as follows: 0.8 g hepatocyte wet wt per gram liver wet wt, 0.42 g dry wt per gram hepatocyte wet wt [6], 0.2 ml mitochondrial water per gram dry wt of liver [1]. The activity in M s^{-1} of glutamine synthase was calculated from the ratio of the two enzymes' activities described above. The V_{max}'s of system N and system L were determined from this premise. v_{AGS} and v_{GTL} were set at [AGS] = 2E−5 M [2] and [GTL] = 1E−7 M [28], and $v_{\text{NH}_4^+\text{-tp}}$ was assumed to be 5E−5 M s^{-1}. The values of v_{GDH} and v_{OAT} were obtained as follows:

$$v_{\text{GDH}} = v_{\text{CPS}} + v_{\text{GS}} - v_{\text{Glunase}} - v_{\text{NH}_4^+\text{-tp}} \tag{4}$$

$$v_{\text{OAT}} = v_{\text{GOT}_m} + v_{\text{GDH}} \tag{5}$$

These equations gave [GDH] and [OAT]. The glutamate transport between the cytoplasm and the sinusoid ($v_{Glu\text{-}tp}$) and the inflow of glutamate from other pathways in mitochondria ($v_{Glu\text{-}spp}$) were modeled to meet the requirements as follows:

$$v_{Glu\text{-}tp} = v_{GS} + v_{GTL} - v_{GOT,} + v_{GATL} \tag{6}$$

$$v_{Glu\text{-}spp} = v_{AGS} - v_{Glunase} - v_{GTL} + v_{GOT_m} + v_{GDH} - v_{OAT} - v_{GATL} \tag{7}$$

In addition, the flux of ornithine from other pathways in mitochondria was set equal to v_{OAT} (see Web supplements, Table S2).

To develop a hepatic lobule model, eight constructed zone models were put in a row and connected one to the next through sinusoidal compartments (Figure 3), in which substances pass through the sinusoids while the hepatocytes are isolated from each other. The first zone (the influx compartment) and the eighth zone (the efflux compartment) were joined to compartments of the upstream periportal tracts and the downstream central vein. Four substances—ammonia, glutamine, glutamate, and urea—flow through the sinusoid and upstream/downstream vessels. Their fluxes from the periportal influx compartment to the perivenous efflux compartment are represented by a simple mass action model as the product of the rate constant and substance concentration (see Appendix 1). No enzymatic reaction is placed in the vessels.

All rate equations and initial concentrations of metabolites were set to be identical among all compartments, but some enzymes were given a slope to their content, representing the regulated heterogeneous gene expressions. In order to evaluate the effects of gradients of the enzyme expressions on metabolic state, six combinations of the gene expression conditions were examined (Table 1). The first model, named the *N model*, had no gradient of the enzyme along the porto-central axis. The second model, named the *GCO model*, in which steep gradients of carbamoyl phosphate synthase, glutamine synthase, and ornithine aminotransferase were introduced, most faithfully modeled the actual lobular ammonia metabolism among the six models. The remaining four models—the *G model*, *GC model*, *O model*, and *GO model*—partially incorporated the enzyme gradients (see Table 1 for details). These significant enzyme slopes were incorporated using a mechanistic model, which assessed the transcription rates of carbamoyl phosphate synthase and glutamine synthase by the relative position in the hepatic lobule (see Appendix A1.3) [5]. In this study, the relative expression levels were assumed to correspond to the transcription rates. The total contents of the enzymes in each model were set equal. The expression status of ornithine aminotransferase was presumed to be evaluable in the same manner as glutamine synthase, based on the reports of coexpression of glutamine synthase and ornithine aminotransferase in the hepatic lobule [30, 31]. Additionally, 60 instances of the models were prepared for each model type by varying the parameters of the flux balance of the glutamate transport system (four patterns): the glutamate flux from outside pathways (three patterns) and the rate constant of sinusoidal substance flow (5 patterns), for a total of $4 \times 3 \times 5 = 60$ patterns (see Appendix A1.4).

3 Methods

3.1 Numerical Integration

Once each chemical and transport reaction rate had been set, a numerical integration of the model was conducted using E-CELL system version 1.1 [42]. The model was implemented by defining the variable Reactor, which describes reaction processes, and the variable Rule, which provides organizational information, substances, and stoichiometry. Simulations were performed using Sun Grid Engine 5.3 on 40 clustered Xeon 2.0 GHz HPC-IAX, and employing the fourth-order Runge-Kutta method. The step interval for integration was set to 0.01 s.

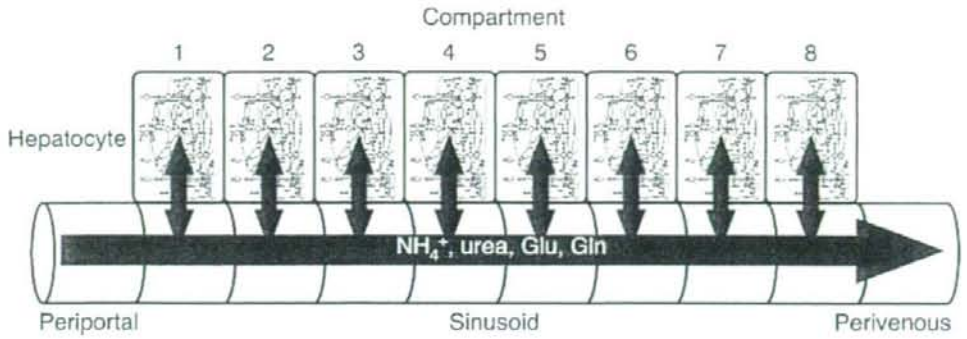

Figure 3. Schematic of eight-cellular-compartment model with sinusoidal compartments. PP and PV represent the periportal end and the perivenous end, respectively. Ammonia (NH_4^+), urea, glutamate (Glu), and glutamine (Gln) flow from the periportal inflow compartment to the perivenous outflow compartment, interacting with cellular compartments.

3.2 Data Analysis

To compare the metabolic aspects in the periportal zone and in the perivenous zone, the flux distributions were examined. To evaluate the effects of the enzyme slope along the lobule metabolic state, the following rates were calculated and used as the indices:

$$\text{Rate of ammonia degradation: } J_{NH_4^+,deg} = v_{CPS} + v_{GS} \tag{8}$$

$$\text{Rate of ammonia generation: } J_{NH_4^+,gen} = v_{Glnase} + v_{GDH} \tag{9}$$

$$\text{Rate of ammonia detoxification: } J_{NH_4^+,detox} = J_{NH_4^+,deg} - J_{NH_4^+,gen} \tag{10}$$

$$\text{Rate of ATP consumption: } J_{ATP,consum} = 2v_{CPS} + v_{ASS} \tag{11}$$

Table 1. Types of model. Six models examined in this study had different combinations of enzyme gradients along the porto-central axis. Letters G, C, and O included in the names of model types are the first letter of gradually expressed enzymes: glutamine synthase (GS), carbamoyl phosphate synthase (CS), and ornithine aminotransferase (OAT), respectively. Enzyme gradients were incorporated into models in an all-or-none fashion. Symbol + or − means whether the enzyme gradient existed in the models. The letters G, C, and O also indicate the existence of the gradient of the respective enzyme. The N model had no gradient of the enzyme.

	Gradual expression of enzyme		
Model type	GS	CPS	OAT
N	−	−	−
G	+	−	−
GC	+	+	−
O	−	−	+
GO	+	−	+
GCO	+	+	+

$$\text{Energy efficiency } \eta := \frac{J_{NH_4^+,\text{detox}}}{J_{ATP,\text{consum}}} \tag{12}$$

$$\text{Rate of bicarbonate consumption} : J_{HCO_3^-,\text{consum}} = v_{CPS} \tag{13}$$

$R_{NH_4^+,\text{detox}}$ and $v_{NH_4^+\text{-tp}}$ are nearly equal under the assumption of steady state. Data were summarized as the mean $\pm$ SD of 60 models for each gradient pattern of the model.

4 Results

4.1 Simulation of the Single-Zone Model

The zone model, which consists of a hepatocellular compartment and a sinusoidal compartment, preserved a quasi steady state for longer than 100,000 s with the initial values mainly quoted from the published literature based on experimental data (see Web supplement, Table S1). The results support the appropriateness of the zone model, as described below.

4.2 Comparison of the Hepatic Lobule Models with Different Enzyme Expressions Patterns

Each model with slopes of the various enzyme expressions exhibits lobule-wide metabolic aspects very different from the N model with no such enzymatic slope. The simulated flux heterogeneities among the porto-central axis are presented in Figures 4 and 5. In the GCO model, which is the best-approximated model to actual lobular metabolism, the active pathways were quite different between the periportal and the perivenous zones (Figure 4). Urea production, urea exportation, and creatine generation were pronouncedly predominant in the periportal zone (red arrows in Figure 4B, Figure 5F a, b, j, k, l, q, t), while glutamine formation and exportation were predominantly seen in the perivenous zone (blue arrows in Figure 4B, Figure 5F c, d, e, h, i, o, r, s). Because mitochondrial ornithine aminotransferase is mainly expressed in the perivenous zone, the concentration of glutamate, which is a reaction product of ornithine aminotransferase, was higher in the perivenous than in the periportal zone. The glutamate concentration in mitochondria was increased from 6.97E−3 to 8.70E−2 M along the porto-central axis. Consequently, the velocity of glutamate dehydrogenase, which catalyzes glutamate, was larger in the perivenous than in the periportal zone (Figure 5F d). Glutamate-aspartate translocase and mitochondrial GOT also exhibited higher activities in the perivenous zone, while cytoplasmic GOT showed an opposite trend of flux between the periportal and the perivenous zone (Figure 4B, Figure 5F e, i, o). Cytoplasmic GOT catalyzed transamination between glutamate and oxaloacetate to produce aspartate, and the generated bulk of aspartate was used as a substrate for argininosuccinate synthetase in the periportal zone (Figure 4B, Figure 5F j). In the N model with no enzyme gradient, most of the chemical reactions and transportation exhibited larger fluxes in the periportal zone than in the perivenous zone (Figure 5A). Due to the high affinity for ammonia of glutamine synthetase, namely, 1/10 of the K_m of carbamoyl phosphate synthetase, ammonia predominantly converted to glutamine in the periportal zone. The fluxes gently changed from the periportal to the perivenous zone, while dramatic alterations were seen in the sixth and seventh compartments in the GCO model, revealing that the perivenous hepatocytes played a lesser role in metabolism in the N model (Figure 5A, F). The G model and the GC model exhibited higher activities of ureagenesis in the periportal zone and greater formation and export of glutamine in the perivenous zone, just like the GCO model. However, in contrast to the GCO model, larger fluxes of OAT, GDH, mitochondrial GOT, and GATL were observed in both models (Figure 5B,C c, d, e, o). In the O model, most of the reactions exhibited higher activity in the perivenous zone, except for the predominant detoxification of ammonia by glutamine synthetase (Figure 5D). Lobule-wide metabolic aspects of the GO model resembled those of the GCO model except for the urea cycle and glutamate transport (Figure 5E). There were fewer contrasts in activities of the urea

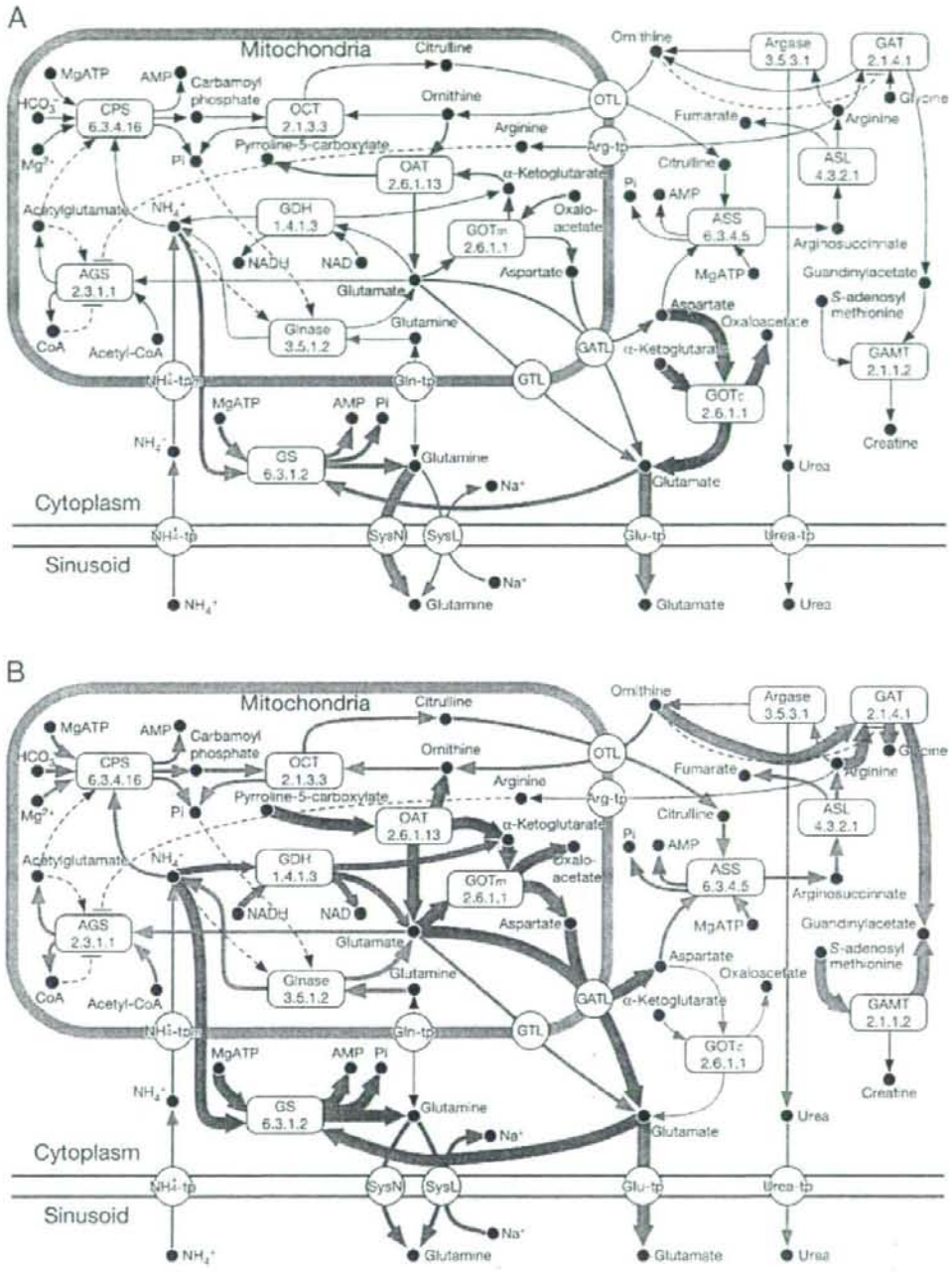

Figure 4. Flux disparities between the two ends of the porto-central axis in N model (A) and GCO model (B). The width of each arrow proportionally reflects the flux ratio between the two ends (the first and the eighth compartment). The thickest line indicates the flux disparity to be fivefold or more. Red and blue arrows indicate fluxes predominant in the periportal and the perivenous zone, respectively. Fluxes with a disparity of less than 1.5 are indicated by the black arrow. The size of the arrows is proportional to the flux except for extremely high fluxes: the Glu-tp in A, ornithine aminotransferase, mitochondrial GOT, and GATL in B. Pronounced urea production, urea export, and creatine generation were seen in the periportal region, while pronounced glutamine formation and reactions which mediate glutamate were seen in the perivenous. The figure shows a representative result among the 60 parameter conditions (see Appendix A1.4 for details).

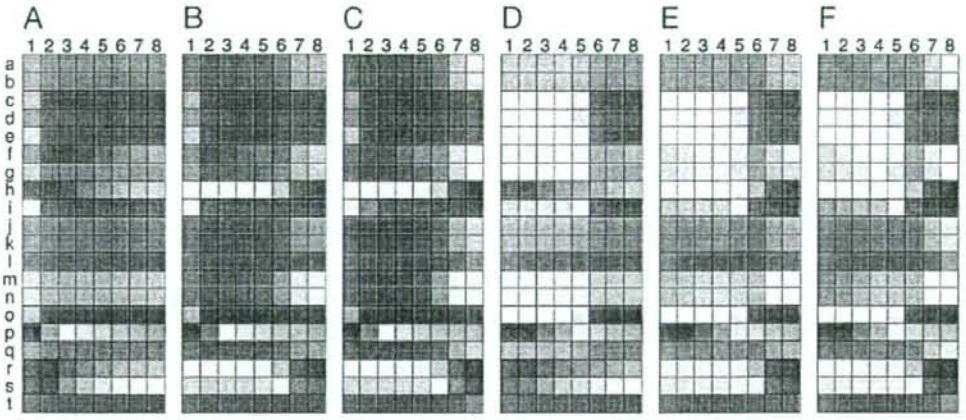

Figure 5. Flux heterogeneities along the porto-central axis. Relative fluxes of 20 representative processes (enzyme reaction or transportation across a membrane) are indicated as color matrices. For each process, the depth of color is proportional to the relative magnitude of its flux, and the maximum absolute flux is indicated by the deepest color. Green and red indicate positive and negative direction of the process, respectively. Colors within the same process are comparable, but inter-process comparison is inappropriate. The numbers on the horizontal axis (1–8) indicate the compartment number. Therefore, the left side represents the periportal and the right side represents the perivenous zone. Panel A, N model; B, G model; C, GC model; D, O model; E, GO model; F, GCO model. On the vertical axis: a, CPS; b, OCT; c, OAT; d, GDH; e, GOT_m; f, Glnase; g, AGS; h, GS; i, GOT_c; j, ASS; k, ASL; l, Argase; m, GAT; n, GAMT; o, GATL; p, GTL; q, OTL; r, SysL; s, SysN; t, Urea-tp. The figure shows a representative result among the 60 parameter conditions (see Appendix A1.4 for details).

cycle and glutamate transport along the porto-central axis in the GO model than in the GCO model (Figure 5E).

4.3 Evaluation of the Effects of Gene Expression Gradients along the Porto-central Axis

The GCO model ranks highest among six models in the rate and energy efficiency of ammonia detoxification. The mean rate of elimination of ammonia from the sinusoid was 11.8% greater in the GCO model than in the control (N model) (Table 2). Although the rate of degradation of ammonia in all eight compartments was 20.0% slower in the GCO model than in the control, the rate of ammonia generation was also 53.8% slower than in the control model (Table 2), showing that the GCO model was able to remove ammonia more efficiently than the control. The mean rate of ATP consumption in the GCO model was 9.5% less than in the control model (Table 2). The O model and GO model also exhibited less ATP consumption than the control. The energy efficiency η, which means the number of consumed ATP molecules required for the elimination of one ammonia molecule, was smaller in the GCO model than in the control (Figure 5, 3.59 ± 0.22 versus 4.47 ± 0.49). The O model and the GO model also displayed smaller η, and the G model and the GC model greater η, than the control (Figure 6).

Elimination of bicarbonate in the GCO model was comparable in rate to that in the control (Table 2). The G and GC models exhibited greater elimination of bicarbonate, but the O and GO models exhibited a lesser elimination of bicarbonate, than the control (Table 2).

5 Discussion

In this study, a mathematical model in which rough histological structure and several forms of regulation of gene expression has been implemented, and it succeeded in representing the metabolic heterogeneity of ammonia detoxification in the liver. Heterogeneous flux distributions bearing a close resemblance to the actual hepatic lobule were observed along the porto-central axis in the

Table 2. The rates of metabolic processes of five models relative to the N model. The highest relative rate of ammonia detoxification was seen in the GCO model. The O model had the lowest relative rate of ATP consumption and the lowest relative rate of bicarbonate consumption.

	Relative rate					
Process	Model N	G	GC	O	GO	GCO
Ammonia degradation	1.00	0.95	1.01	0.79	0.76	0.80
Ammonia generation	1.00	1.00	0.99	0.48	0.46	0.46
Ammonia detoxification	1.00	0.90	1.03	1.09	1.04	1.12
ATP consumption	1.00	1.12	1.19	0.78	0.86	0.90
Bicarbonate consumption	1.00	1.26	1.34	0.77	0.94	0.99

GCO model, which is an approximation of the actual regulation of gene expression in the ammonia-metabolism-related enzymes. The periportal hepatocytes play a role in urea production and creatine generation, while the perivenous cells play a role in glutamine generation and glutamate metabolism in this model (Figure 4). The model without any regulation of gene expression, the N model, also displayed heterogeneous metabolism, but the extent of the heterogeneity was quite reduced compared to the GCO model, indicating that two factors, the histological structure and gradual gene expression, are likely sufficient to simulate the zonal heterogeneity in liver ammonia detoxification, while the histological structure alone was insufficient. The simulation results also suggest that the gradients of expression of the three enzymes (CPS, GS, and OAT) along the porto-central axis improve the energy efficiency of ammonia detoxification (Table 2 and Figure 6).

Flux distributions in the model with the gradients of CPS, GS, and OAT are close to the schemes of heterogeneous ammonia metabolism in the lobule discussed in many reports [20, 31]. It has been suggested that urea production and bicarbonate consumption in exchange for ammonia detoxification by the urea cycle are more highly activated in periportal hepatocytes, and glutamine generation and reactions that mediate glutamate (OAT, GDH) are more highly activated in perivenous (opposite-side) hepatocytes [20, 31]. Three enzymes, CPS, GS, and OAT, chiefly restrict and characterize the flux distributions in each zone. The flux distributions between the periportal and perivenous zone have a mutually exclusive relationship. Glutamine metabolism (both of synthesis and consumption) was

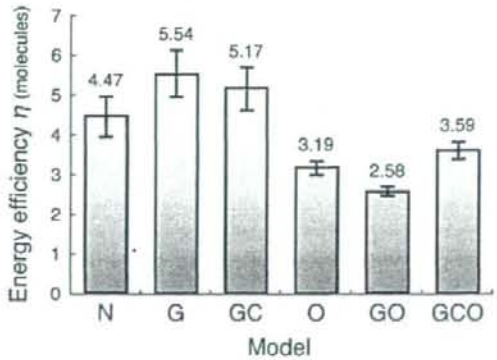

Figure 6. Energy efficiency in the six models with different gene expression patterns. Energy efficiencies η were calculated from $J_{NH_4^+,detox}$ and $J_{ATP,consum}$ (see text for details). Data are means ± *SD*.

hardly seen in the periportal zone, while urea production and creatine generation were not seen in the perivenous zone. Cytoplasmic GOT showed opposite trends between the two zones. Ornithine aminotransferase in the perivenous hepatocytes is speculated to lead to an accumulation of glutamate, which is the substrate for glutamine synthetase, and to increase the elimination of ammonia [31].

Our model differs in some features from previous investigations. Glutamate dehydrogenase has been suggested to have significantly increased activity in the perivenous zone [33]. In our model, higher activities were also shown in the perivenous than in the periportal zone, but the direction of the reaction was different from the suggestion in a previous report [4]. For glutamate dehydrogenase in the perivenous zone, a reverse reaction to that suggested by other investigators was predominant [4]. This might be attributable to the lack of the α-ketoglutarate transport system in our model. If that mechanism were implemented in our model and the influxes of α-ketoglutarate to the perivenous mitochondria increased, the reaction direction might be the same as suggested by previous work. Phosphate-dependent glutaminase has been suggested to be highly activated in the periportal zone [20, 45]. However, the tendency towards emphasized activity of phosphate-dependent glutaminase in the periportal hepatocytes was not recreated in our model.

Several factors are considered to be responsible for the result in our experiments. One factor is the roughness of the mathematical model of phosphate-dependent glutaminase, such as the lack of a pH effect despite the actual high sensitivity to pH [41]. Another factor is the absence of regulation of phosphate-dependent glutaminase gene expression. It might be necessary to give a slope to gene expression along the hepatic lobule, as we did with GS, OAT, and CPS. Although this model does not yet have the requisite accuracy, it is the first realization of a model able to yield metabolic aspects with both quantitative information at the micro, or kinetic, level and functional significance at the macro, or tissue, level in the liver.

It is proposed that one major advantage of the functional separation in the hepatic lobule is that the periportal urea synthesis flux is flexible with regard to the requirement of the acid-base condition without the risk of hyperammonemia, because effective elimination of ammonia is carried out by perivenous glutamine synthetase [13–18, 20]. It is also proposed that ornithine aminotransferase, which is co-localized with glutamine synthetase, would lead to more efficient removal of ammonia. In this study, because the effect of pH is excluded, no indication is given of its relation to the major advantage of the functional separation, which has been discussed by previous investigators. But it is newly suggested that the functional separation that is formed by steep gradients of GS, OAT, and CPS along the hepatic lobule contributes to effective ammonia elimination with efficient energy use. Interestingly, the glutamine synthetase and ornithine aminotransferase expression patterns are not the same in other organs, such as the kidney, brain, and intestine [31]. This fact indicates cooperation of the expression of the two enzymes evolved specifically in the hepatic lobule. The steep gradients of the three enzymes along the porto-central axis are considered to have been evolved into effective ammonia detoxification, which takes place exclusively in the liver. Although the improvement in the energy efficiency shown by simulation is not robust, it is sufficient to confer an adaptive advantage, since ammonia metabolism consumes a large quantity of ATP in the liver. Schneider and colleagues have reported ureagenesis accounts for 15% of the total energy consumption of rat hepatocytes incubated in a nutriment-enriched medium [38]. Ultimately, this simulation will have to be verified with wet (physiological) experiments, but it is difficult to evaluate changes in the lobular energy efficiency along with changes in the porto-central axial regulation of gene expression by wet experiments at present. Technology is continually advancing, however, and the day should not be far off when such confirmation will be within reach.

In summary, we have shown that a novel systems biological approach for tissue modeling well simulated the functional separation in the hepatic lobule, and the simulation results suggested contribution of the gradual expressions of three enzyme (CPS, GS, and OAT) along the porto-central axis to effective ammonia detoxification with efficient energy use. This method of reconstructing the intracellular chemical processes, based on module architectures such as the hepatic lobule, enables analysis of certain metabolic aspects with both quantitative information at the kinetic level and functional significance at the tissue level, and to evaluate the effect of enzyme gene expression gradients on metabolic