The Total Body Amino Acid Composition of an Animal Could Be Explained by a Mixture of the Average Composition of Its Proteome as an Intracellular Composition Estimate and the Type I Collagen Composition as an Extracellular Composition Estimate

Genshiro Esumi
Department of Pediatric Surgery, Hospital of the University of Occupational and Environmental Health, Kitakyushu, Japan

Abstract

The total amino acid composition of an animal's body is theoretically explained by the sum of the product of the amount of each protein and its amino acid composition. However, because there are so many different types of proteins, and because the amount of each protein varies greatly, it has not been easy to actually describe the total amino acid composition of the body by the sum of the amount of proteins and their compositions.

In a previous paper, I showed that the amino acid composition of cells could be mutually convergent with the average composition of the proteome proteins in the organism's genome. On the other hand, type I collagen is known to be the most abundant protein among extracellular proteins. Assuming that the intracellular amino acid composition is consistent with the average proteome composition and the extracellular amino acid composition is consistent with that of type I collagen, I investigated whether the mixture of these two compositions could describe the total body composition of an animal. In this investigation, I used body compositions of fetal pigs at different stages of gestation from the literature. As a result, I found that mixtures of these estimates of the intracellular and extracellular composition sufficiently approximated the actual total body amino acid compositions in this investigation.

Therefore, I concluded that the three assumptions that the amino acid composition of intracellular proteins approximates the average proteome composition, that the amino acid composition of extracellular proteins approximates that of type I collagen, and that the total body amino acid composition of an animal can be explained by a mixture of such intracellular and extracellular estimates, may all be correct.

Keywords: amino acid composition, whole body, intracellular protein, extracellular matrix, collagen

Email: esumi@clnc.uoeh-u.ac.jp

1. Background

The total amino acid composition of an animal's body is theoretically explained by the sum of the product of the amount of each protein and its amino acid composition. However, because there are so many different types of proteins, and because the amount of each protein varies greatly, it has not been easy to actually describe the total amino acid composition of the body by the sum of the amount of proteins and their compositions.

In a previous paper, I showed that the amino acid composition of cells can be mutually convergent with the average composition of all proteome genes in the organism [1]. If this is true, then we can estimate the amino acid composition of all intracellular proteins of an organism by averaging the amino acid composition of its proteome genes.

On the other hand, collagens are the most abundant proteins in the extracellular matrix, and type I collagen is the most abundant among the collagens [2]. Therefore, I assumed that I could estimate the amino acid composition of the total extracellular proteins of an animal body from that of type I collagen.

Since the body proteins of an organism consist of intracellular and extracellular proteins, I should be able to estimate the amino acid composition of the entire body with the two estimates, intracellular and extracellular, mentioned above. Therefore, I investigated whether the mixture of these two estimates could actually describe the total body composition of an animal.

In this paper, I will show how I generated intracellular and extracellular composition estimates and how I performed comparisons between the animal's body compositions and these generated estimates.

2. Total body amino acid composition of animals: Fetal pigs

2.1. Background and materials: Amino acid composition of fetal pig body

The amino acid composition of a sample of an organism is commonly measured by chromatography, but these samples are limited in size and usually small. Therefore, it is neither common nor easy to measure the amino acid composition of an entire animal body. Prior to this study, I searched the literature and found a report of actual measurements of whole-body amino acid composition of fetal pigs at various stages of gestation, which I chose to use in this study [3].

The fetal pig amino acid compositions listed in the report were from five different gestational ages (40, 60, 90, 110, and 114 days), each reported as an average of several samples [3]. Because aspartic acid (Asp) and asparagine (Asn), glutamic acid (Glu), and glutamine (Gln) are added together in chromatographic measurements, the report listed 18 of the 20 amino acid compositions that make up proteins. In addition, each composition included not only the amino acids that make up protein, but also the amino acids that do not make up protein.

2.2. Methods: Amino acid composition of fetal pig body

The whole-body amino acid compositions of fetal pigs in the mentioned report were the weight ratios of each amino acid [3]. In the present study, these compositions were converted from weight ratios to molar ratios to examine the correspondence with the ratios of the number of amino acid residues on the gene, and the target amino acids were limited to the 20 amino acids that make up the initial proteins and hydroxyproline, which is post-translationally modified from proline (Pro). Thus, the molar composition of hydroxyproline was added to that of proline.

In this and the following study, I used Microsoft® Excel for Mac v16.74 (Microsoft Corporation, Redmond, WA, USA) to generate compositions and other computational results. I also used JMP® 17.1.0 (SAS Institute Inc., Chicago, IL, USA) to generate graphs and figures.

2.3. Results: Amino Acid Compositions of the Fetal Pig Body

Table 1 shows the calculated composition ratios of the amino acid compositions of the fetal pig bodies from five different gestational ages (40, 60, 90, 110, and 114 days). Each amino acid is arranged in the order of the daily food composition table published by the Ministry of Education, Culture, Sports, Science and Technology, Japan [4]. This order of amino acids was chosen because amino acid residues are arranged according to their structure and function, and this order is intuitively easy to understand when comparing amino acid compositions.

Figure 1 shows the same data from Table 1 in 5 radar plots, each corresponding to fetal pigs of five gestational ages.

Table 1

	Gestational Age (Days)				
	40	60	90	110	114
Ile	0.0354	0.0334	0.0296	0.0282	0.0280
Leu	0.0818	0.0749	0.0685	0.0663	0.0655
Lys	0.0750	0.0600	0.0528	0.0504	0.0502
Met	0.0198	0.0181	0.0167	0.0163	0.0159
Cys	0.0137	0.0142	0.0142	0.0139	0.0134
Phe	0.0358	0.0315	0.0284	0.0270	0.0265
Tyr	0.0250	0.0224	0.0190	0.0178	0.0171
Thr	0.0450	0.0409	0.0379	0.0352	0.0346
Trp	0.0075	0.0077	0.0075	0.0072	0.0069
Val	0.0618	0.0528	0.0498	0.0470	0.0457
His	0.0211	0.0174	0.0171	0.0169	0.0170
Arg	0.0461	0.0477	0.0490	0.0477	0.0472
Ala	0.0867	0.0867	0.0891	0.0935	0.0922
Asp/Asn	0.0849	0.0825	0.0772	0.0765	0.0744
Glu/Gln	0.1176	0.1209	0.1112	0.1081	0.1098
Gly	0.1046	0.1230	0.1622	0.1746	0.1830
Pro	0.0762	0.1085	0.1173	0.1196	0.1202
Ser	0.0618	0.0574	0.0525	0.0538	0.0524

Figure 1

3. Estimation of Intracellular Protein Amino Acid Composition: Proteome Average

3.1. Background, Materials and Methods: Proteome average

My previous work has shown that the amino acid composition of cells can be mutually convergent with the average composition of the proteome proteins in the organism's genome. [1]. Since the target materials of this study are fetal pigs, I decided to use a dataset of reference genomes of pigs that are publicly available on NCBI [5].

3.2. Results: Proteome average

The distributions of the amino acid composition of the proteome proteins on the porcine genome dataset are shown in Figure 2a. These distributions are uniformly bell-shaped, and in a previous paper I assumed that they are binomial distributions constrained by the amino acid composition of the intracellular protein, their synthetic resource [1]. If this assumption is correct, the average of each distribution (also shown in the figure) should be an approximation of the amino acid composition of the intracellular proteins, their synthetic resource.

In Figure 2b, the average proteome composition is shown as an 18-item radar plot, as in Figure 1, for comparison with the whole-body amino acid compositions of fetal pigs.

Figure 2a

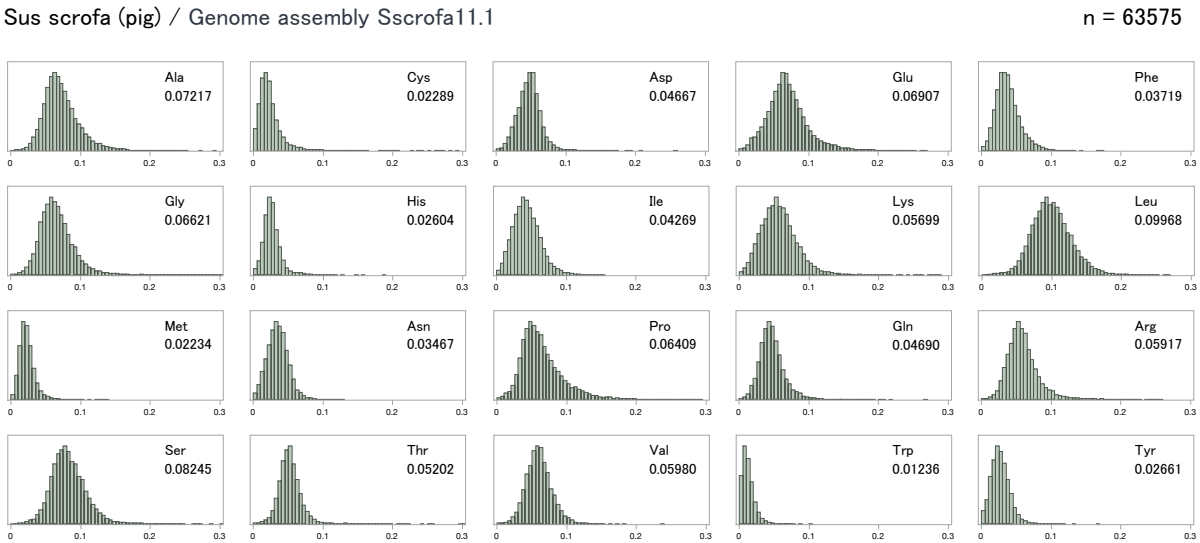
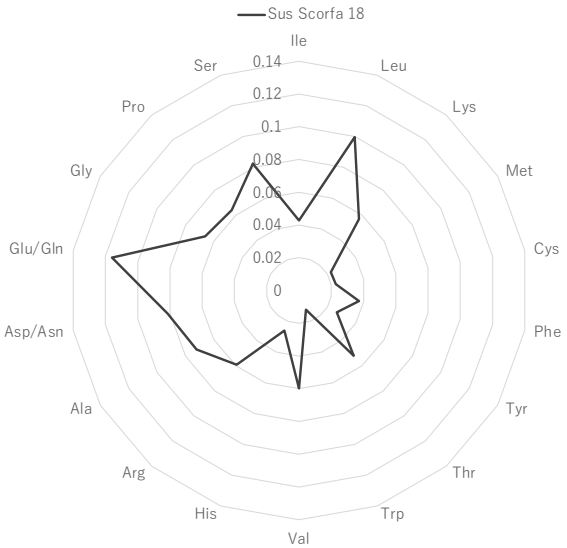


Figure 2b



4. Estimation of Extracellular Protein Amino Acid Composition: That of type I collagen

4.1. Background, Materials and Methods: Amino acid composition of type I collagen of pigs

Type I collagen is the most abundant protein in the extracellular matrix of the animal body [2]. However, the amino acid composition of collagen is extremely skewed, with proline and glycine accounting for more than half of the total amino acid composition. Elastin, another abundant protein in the extracellular matrix, has a similar skewed composition. Based on these skewed compositions and characteristics, I speculated that the amino acid composition of extracellular proteins in the extracellular matrix of animals could be approximated by that of type I collagen.

4.2. Results: Amino acid composition of type I collagen of pigs

Type I collagen is a complex of two alpha-1 chains and one alpha-2 chain. Each chain is encoded by the COL1A1 and COL1A2 genes, and their amino acid sequences are quite similar from animal to animal. Each collagen chain is synthesized as a preproprotein and then peptides are cleaved off at both ends to form the final collagen chain. I referred to the data published in UniProt for the amino acid sequence of porcine type I collagen, but there was no information on the post-translational process in this data, so I estimated the amino acid sequence of porcine alpha-1 and alpha-2 chains from the processing information of human and bovine type I collagens published in the same UniProt.

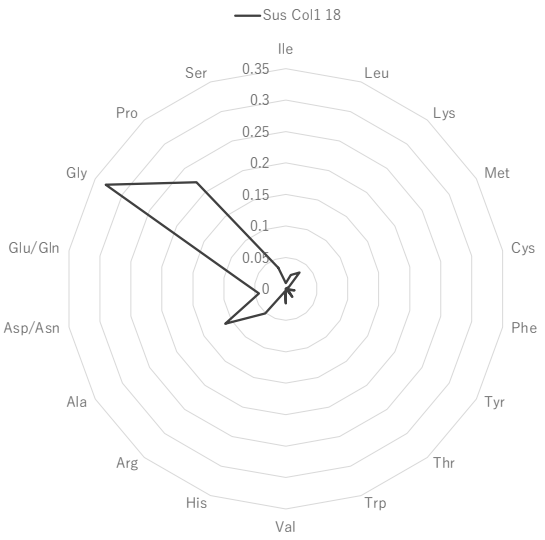
Table 2 shows the number of amino acid residues on porcine alpha-1 (COL1A1) and alpha-2 (COL1A2) chains after processing, the number of amino acid residues on type I collagen (two alpha-1 chains and one alpha-2 chain, COL1 complex), and the ratio of amino acid composition on porcine type I collagen.

In Figure 3, the amino acid composition of porcine type I collagen is shown as an 18-item radar plot, as in Figure 1, for comparison with the whole body amino acid compositions. We can see that its amino acid composition is extremely skewed.

Table 2

	COL1A1	COL1A2	COL1 complex	COL1 composition
Ala	120	110	350	0.111040609
Cys	0	0	0	0
Asp	34	23	91	0.028870558
Glu	49	46	144	0.045685279
Phe	15	14	44	0.013959391
Gly	346	350	1042	0.330583756
His	3	8	14	0.004441624
Ile	7	15	29	0.009200508
Lys	38	31	107	0.033946701
Leu	21	32	74	0.023477157
Met	7	4	18	0.00571066
Asn	10	26	46	0.014593909
Pro	243	211	697	0.221129442
Gln	30	24	84	0.026649746
Arg	53	55	161	0.05107868
Ser	40	32	112	0.035532995
Thr	17	18	52	0.016497462
Val	19	36	74	0.023477157
Trp	0	0	0	0
Tyr	4	5	13	0.004124365
Sum	1056	1040	3152	

Figure 3



5. Comparison with a mixture of the two composition estimates

5.1. Materials and Methods: the comparison

The aim of the present study was to investigate whether the actual amino acid composition of the whole body can be explained by the estimated amino acid composition of intracellular and extracellular proteins. In this study, the former is the average amino acid composition of proteome proteins and the latter is the amino acid composition of type I collagen.

Meanwhile, in this analysis, it is necessary to evaluate the extent to which the sum of the two composition estimates can approximate the actual whole body amino acid composition. This requires the selection of an appropriate distance function for protein amino acid composition. Previous analyses have explored the use of L1 distance, L2 distance, Cosθ distance, etc., but there was no significant difference between any of the distance functions [6]. I believe that the reason is due to the dimensional characteristics of the amino acid composition of proteins. Therefore, in this study, the L1 distance, often called the Manhattan distance, which is the most intuitive and easy to understand, was used for evaluation.

Since the optimal ratio for mixing the intracellular and extracellular protein estimates for the amino acid composition of the fetal pig body at each gestational age is unknown, I calculated these mixed compositions in 1000 steps in 0.1% increments as their extracellular protein fractions and selected the one that minimized the distance from the actual whole body amino acid composition. If the minimum distance is sufficiently small, not only can it be determined that the estimated composition adequately describes the actual whole body amino acid composition, but the estimated mixing ratio of intracellular to extracellular protein composition could be an estimate of the ratio of intracellular to extracellular protein amounts in the actual whole body of the organism.

5.2. Results: the comparison

Figure 4 shows the L1 distances (y-axis) between the fetal amino acid compositions and the estimated total body amino acid compositions (sum of the product of the summed intracellular and extracellular estimates with their fractions) at each 1000-step mixing ratio (x-axis) for the five fetal pig amino acid compositions at different gestational ages. The extracellular estimated fraction ratio that minimized each L1 distance and its minimum value differed depending on the gestational age of the fetal pig.

Figure 5 shows the estimated amino acid composition at each minimum distance with the actual whole body composition of the fetal pigs on five radar plots for the fetal pigs at each gestational age, which appear to be fairly good approximations. The actual degree of approximation is discussed in the next section.

Table 3 shows the minimum distances and their mixing ratios as estimated extracellular fractions at each gestational age. These distances appear to decrease with later gestational ages.

Figure 4

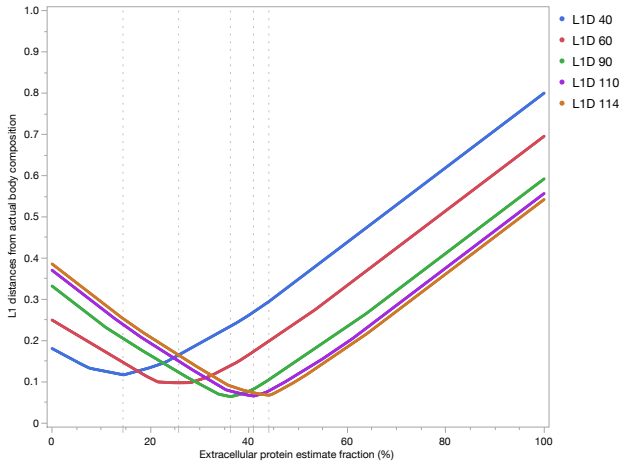
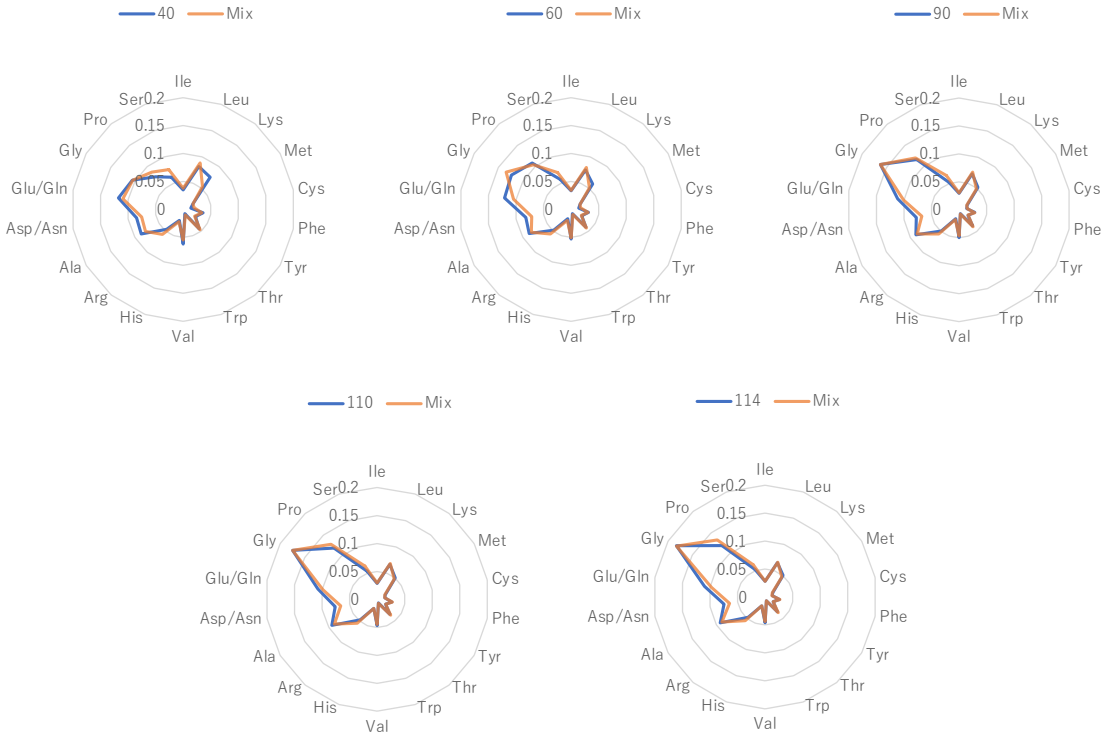


Table 3

Gestational age (day)	Minimum L1 distance	Extracellular fraction (%)
40	0.11677986	14.5
60	0.097461213	25.7
90	0.063867421	36.3
110	0.065312912	41.0
114	0.067389482	44.1

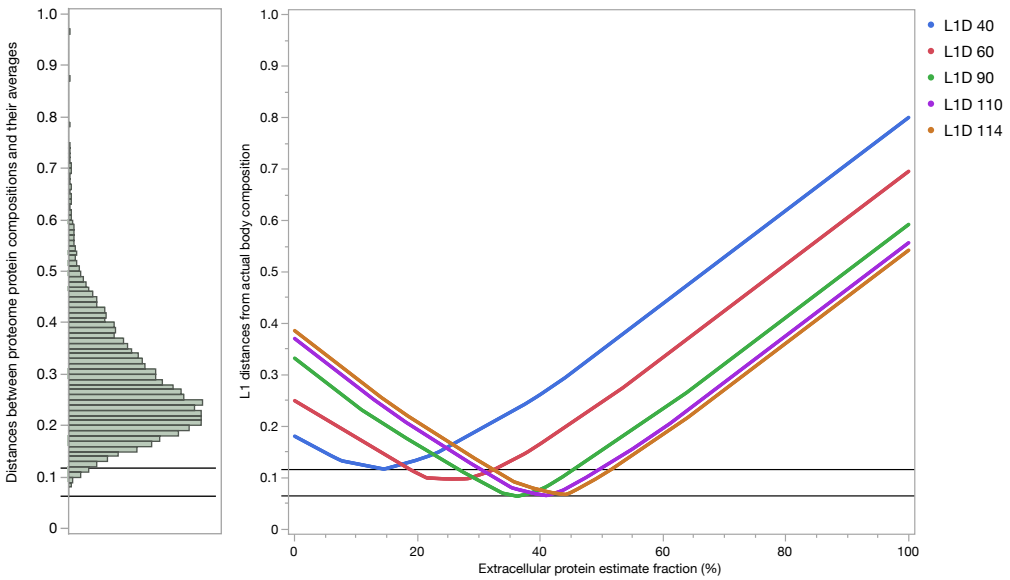
Figure 5



5.3. Discussions: the comparison

The present analysis shows that the actual whole-body amino acid composition of fetal pigs appears to agree quite well with the estimated composition derived from estimates of intracellular and extracellular protein composition (Figure 5). However, it is not easy to determine whether this is actually very good agreement or whether such agreements could occur by chance. Therefore, I compared these minimum distances with the distances between the individual amino acid compositions of the proteome proteins and their averages as actual distance samples of the same dimensions. As a result, the minimum distances were sufficiently small compared to the distance distribution from their average composition. Therefore, I concluded that the two estimates calculated in this study could very well describe the actual whole-body amino acid composition of fetal pigs, especially in late gestation at days 90, 110, and 114 (Figure 6).

Figure 6



6. Discussion

In theory, an animal's total body amino acid composition can be described as the sum of the product of the amount of each protein in the animal's proteome and the amino acid composition of each protein gene. However, because there are so many different types of proteins, it has never been possible in practice to explain an animal's total body amino acid composition in this way.

On the other hand, in my previous report, I suggested the possibility that the average amino acid composition of proteins in the proteome may be in a state of mutual convergence with intracellular proteins [1]. If this is true, then the total amino acid composition of intracellular proteins can be estimated from the average amino acid composition of the proteome = all proteins encoded by the genome.

Meanwhile, considering that the majority of the extracellular matrix in the animal body is composed of collagen, of which type I collagen is the most abundant [2], the total amino acid composition of cell body proteins in the animal body can be estimated by the amino acid composition of type I collagen.

If the above two estimates are more or less accurate, then the sum of the products of these two estimates with each amount should be close to the actual amino acid composition of the whole body.

Therefore, in this study, I calculated the estimated amino acid compositions of intracellular and extracellular proteins and added them in various proportions to see how close the estimated compositions could be to the actual amino acid compositions of the whole body.

In this comparison, although there are 20 amino acids that actually make up the protein, only the compositions of 18 elements are actually compared. This is because the measurements of two amino acids are combined with other amino acids in the actual measurement process. But comparing 18 elemental compositions is still a relatively high-dimensional calculation. Thus, it is not easy to intuitively determine whether the calculated distances are actually close or not, due to the sparsity caused by the high dimensional curse. This is equivalent to intuitively assuming that data spanning 18 dimensions is unlikely to be approximated by chance, but still not easy to evaluate in practice.

In my results, the minimum distances between the estimated and actual compositions of the fetal pig bodies were smaller than almost all distances between the protein amino acid compositions from their average proteome composition, so I was able to conclude that the estimation of total body amino acid composition by the intracellular extracellular protein estimation performed in this study sufficiently describes the actual body amino acid compositions.

In the present results, the percentages of extracellular proteins estimated at the minimum distances were 14.5%, 25.7%, 36.3%, 41.0%, and 44.1% for 40, 60, 90, 110, and 114-day fetuses, respectively. On the other hand, the report on fetal pigs also included the compositional amount of hydroxyproline. Since hydroxyproline is reported to be present in porcine collagens at a concentration of 9.1% residue in molar amount [3], the composition of collagens could be estimated from the amount of hydroxyproline, and they were 8.9%, 19.6%, 32.5%, 35.8%, and 37.1% (corresponding to each day of age). Figure 7 shows the estimated fractions of extracellular protein, the estimated fractions of collagen, and Figure 8 shows the ratio of collagen to extracellular protein at each gestational age. These showed that the fractions of both extracellular protein and collagen increased with gestational age. At the same time, the ratio of collagen to extracellular protein initially increased, but changed to a decreasing trend after reaching a peak at 90 days of age. I assumed that this could be explained by the fact that body formation begins with a relatively high cellular content in the fetus, followed by an increase in the amount of collagen as extracellular matrix until body size increases to a certain degree, followed by an increase in the amount of non-collagenous tissue proteins such as elastins and keratin-related proteins.

Figure 7

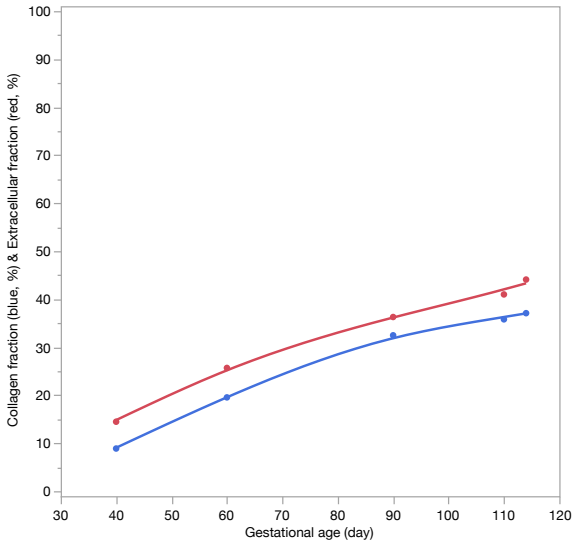
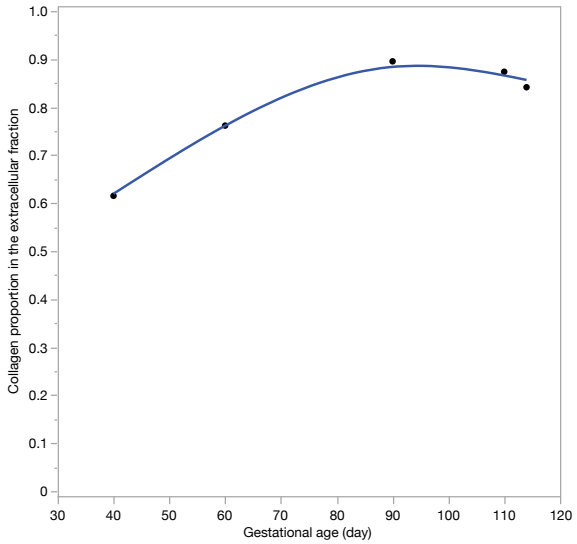
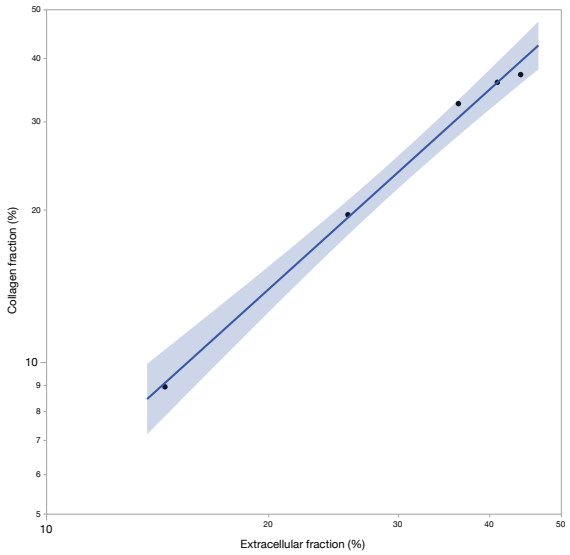


Figure 8



A double-log plot of the fractions of collagen and extracellular protein in total body composition showed that they were almost in line (Figure 9). From this figure, I speculated that collagen formation, intracellular protein formation, and extracellular protein formation are all exquisitely coordinated in shaping the fetal body throughout the gestational period.

Figure 9



Incidentally, this study used pig fetuses for analysis, but since they reach full term at day 114, the total body composition of a fetus at 114 days gestational age can be considered as that of a newborn pig.

Also, this study used pigs rather than humans, but the average proteome composition of pigs is nearly identical to that of humans (data not shown). In addition, the composition of type I collagen in pigs is nearly identical to that in humans (data not shown). Thus, the two estimates of amino acid composition, intracellular and extracellular, calculated in this study would be expected to approximate the total body composition of mammals, so-called "animals," that are close to humans.

Last but not least, this study indicated the existence of a universal difference in amino acid composition between intracellular and extracellular proteins. In a previous report, I hypothesized that this universal difference in amino acid composition between intracellular and extracellular proteins may be the origin of the so-called essential amino acids [7]. Various studies have been conducted on the origin of essential amino acids, but I believe that this hypothesis is one of the most promising. I look forward to the future unraveling of this mystery of the origin of our amino acid auxotrophy.

7. Conclusion

Based on the results of my study in this paper, I concluded that the three assumptions that the amino acid composition of intracellular proteins approximates the average proteome composition, that the amino acid composition of extracellular proteins approximates that of type I collagen, and that the total body amino acid composition of an animal can be explained by a mixture of such intracellular and extracellular estimates, may all be correct.

The conclusion presented here strengthen my previous work on the convergence of amino acid composition of proteome proteins. And they also provide new insights not only into the distribution of body proteins with respect to their amino acid composition, but also into the selection pressure exerted on protein amino acid composition during evolution.

8. References

1. Esumi, G. (2023). The Distributions of Amino Acid Compositions of Proteins in an Organism's Proteome Uniformly Approximate Binomial Distributions. *Jxiv*. https://doi.org/10.51094/jxiv.408
2. Alberts, B. (2022) Chapter 19: Cell Junctions and the Extracellular Matrix, Table 19-2 Some types of Collagens and Their Properties, *Molecular Biology of the Cell. 7th Edition*, Garland Science, Taylor and Francis Group, New York.
3. Wu, G., Ott, T. L., Knabe, D. A., & Bazer, F. W. (1999). Amino Acid Composition of the Fetal Pig. *The Journal of Nutrition*, *129*(5), 1031–1038. https://doi.org/10.1093/jn/129.5.1031
4. Office for Resources, Policy Division Science and Technology Policy Bureau. Ministry of Education, Culture, Sports, Science and Technology. (2020). - Amino Acids -. *J. STANDARD TABLES OF FOOD COMPOSITION IN JAPAN - 2020 - (Eighth Revised Edition)*. https://www.mext.go.jp/a_menu/syokuhinseibun/
5. Genome dataset, "Sus scorfa (pig) / Genome assembly Sscrofa11.1", NCBI website, https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000003025.6/
6. Esumi, G. (2022). Proteome and cellular amino acid compositions may be mutually constrained and in a state of narrow convergence. *Jxiv*. https://doi.org/10.51094/jxiv.95
7. Esumi, G. (2022). Origin of Essential Amino Acids: Extracellular Matrix Hypothesis. *Jxiv*. https://doi.org/10.51094/jxiv.121