

A new method of estimating the proportions of different proteins in a mixture using amino acid profiles: application to undigested proteins in the preruminant calf

C Duvaux ^{1*}, P Guilloteau ¹, R Toullec ¹, JW Sissons ²

¹ INRA, Laboratoire du Jeune Ruminant, 65, rue de Saint-Brieuc, 35042 Rennes Cédex, France;

² AFRC, Institute for Grassland and Animal Production, Pigs Division, Church Lane, Shinfield, Reading RG2 9AQ, UK

(Received 5 June 1989; accepted 12 October 1989)

Summary — This paper describes a mathematical model for estimating the proportions of proteins of different origins in mixtures. The method combines a multiple regression analysis and the calculation of the distance of χ^2 . It is validated on mixtures of known composition and is applied to the estimation of proportions of dietary, endogenous and bacterial proteins in preruminant calf's ileal digesta.

protein proportions / mixture / amino acid profile / multiple regression / undigested protein / calf

Résumé — Une nouvelle méthode d'estimation des proportions de différentes protéines dans un mélange à partir des profils d'acides aminés : application aux protéines indigérées chez le veau préruminant. Nous décrivons ici une méthode de calcul permettant d'estimer les proportions de différentes protéines dans un mélange, à l'aide d'un modèle mathématique combinant une régression multiple et un calcul de la distance du χ^2 . La méthode est testée sur des mélanges de composition connue. Elle est appliquée à l'estimation des proportions de protéines alimentaires, endogènes et bactériennes dans les digesta iléaux du veau préruminant.

proportions de protéines / mélange / acide aminé / régression multiple / protéine indigérée / veau

INTRODUCTION

Digesta of the terminal ileum contains a mixture of proteins of exogenous (dietary), endogenous (exported from the animal itself) and bacterial (sacrophytic flora of the host digestive tract) origin. Therefore, the overall amino acid composition of the di-

gesta can be considered to result from the specific amino acid profiles of these different sources. Estimates of the proportions of these 3 protein components will give an indication of the extent of digestion of the dietary protein, as well as of the interaction between the diet and digestive processes in the alimentary tract.

* Correspondence and reprints

Attempts to determine the origin of protein in digesta can be based on the examination of proportions of some characteristic amino acids (Vérité *et al*, 1977). However, this method is inaccurate and it appears to be more interesting to take into account the full spectrum of amino acids as proposed by Evans *et al* (1975), Patureau-Mirand *et al* (1977) and Guilloteau *et al* (1983). The aim of this work was to improve the previous procedures by devising a mathematical model which establishes in what theoretical proportions reference proteins (*ie* dietary, endogenous and bacterial) should be mixed to produce the overall amino acid composition the most similar to that found in composite digesta. In addition, the method quantifies the differences between the calculated and the actual amino acid compositions in order to assess the validity of the model.

MATHEMATICAL MODEL

The method described herein for obtaining the "best fit" mixture is based on simple rules for the determination of compositions of mixtures, associated with a calculation which allows 2 proteins to be compared taking into account all the differences between their amino acid compositions (distance of χ^2) (Guilloteau *et al*, 1983). The mathematical procedure is mainly a multiple regression analysis which minimizes the residual sum of squares (Williams, 1959). It is performed using a computer able to store the whole collection of amino acid compositions of reference proteins.

Equations used in the model

Let P_0 be the protein mixture contained in a digesta of known amino acid composition. Let P_1, P_2, \dots, P_n be a collection of n

reference proteins, each of them having a characteristic amino acid composition. The P_0 mixture of n reference proteins can be considered to consist of a fraction (a_1) of the protein P_1 , a fraction (a_2) of the protein P_2 , and so on up to a fraction (a_n) of the protein P_n .

Thus,

$$P_0 = a_1 P_1 + a_2 P_2 + \dots + a_n P_n \quad (1)$$

with $a_1 + a_2 + \dots + a_n = 1$

If this equation is valid for the protein P_0 , it should also be valid for each amino acid (AA_i) of the protein.

Therefore we obtain:

$$AA_{i,0} = a_1 AA_{i,1} + a_2 AA_{i,2} + \dots + a_n AA_{i,n} \quad (2)$$

where $i = 1$ to 17 (number of amino acids used in this method), n is the number of proteins in the mixture (maximally 17) and AA the percentage of amino acids (the data are usually expressed as grams of amino acids per 100 g of assayed amino acids).

We obtain i equations with n independent variables, equations which are solvable through a multiple regression analysis found in most computer packages and which will supply values for the coefficients a_1, a_2, \dots, a_n and also statistical information about their significance by means of a t test. In this particular case the mathematical model set up by the multiple regression, which usually involves a constant a_0 , must be constrained through the origin since a constant different from zero does not have any biological significance.

The computer calculation is based on the minimization of the residual sum of squares:

$$\sum_{i=1}^{17} (AA_{i,j} - AA_{i,k})^2 \quad (3)$$

where $AA_{i,j}$ is the percentage of the amino acid number i in the sum of the amino acids assayed in the digesta and where $AA_{i,k}$ is the percentage of the amino acid number i in the fitted mixture. The computer will calculate the coefficients a_1, a_2, \dots, a_n corresponding to the minimal value for the residual sum of squares. These coefficients will verify the following equations:

$$\begin{aligned} AA_{1,k} &= a_1 AA_{1,1} + a_2 AA_{1,2} + \dots + a_n AA_{1,n} \\ AA_{2,k} &= a_1 AA_{2,1} + a_2 AA_{2,2} + \dots + a_n AA_{2,n} \\ AA_{i,k} &= a_1 AA_{i,1} + a_2 AA_{i,2} + \dots + a_n AA_{i,n} \end{aligned} \quad (4)$$

Methods based on multiple regression analysis have been used by several researchers, for example Evans *et al* (1975), Lindqvist *et al* (1975), Cottle and Nolan (1982) and Cockburn and Williams (1984); nevertheless they have the disadvantage of giving the most influence to the amino acids whose percentages are the highest (for example the glutamic acid in a soya-bean meal (Guilloteau *et al*, 1986)).

It is therefore necessary to give the same weight to each amino acid in the profile. For this purpose, let us consider another method described by Guilloteau *et al* (1983) and which minimizes the distance of χ^2 .

$$\text{distance of } \chi^2 = 17 \sum_{i=1}^{17} \frac{(AA_{i,j} - AA_{i,k})^2}{\frac{AA_{i,j} + AA_{i,k}}{2}} \quad (5)$$

The abbreviations are the same as the residual sum of squares (3).

A computer programme using an iterative procedure could be developed for the calculation of the coefficients a_1, a_2, \dots, a_n

of the equation 4 by minimizing the distance of χ^2 . However, this method does not provide any statistical test for the significance of the coefficients. Therefore, the 2 methods described in this paper (the distance of χ^2 and the multiple regression) have been combined.

If it is assumed that the fit is satisfactory (*ie* that it is possible to find a collection of reference proteins which will give, when mixed, an amino acid composition similar to that of the digesta), then the following approximation can be made:

$$AA_{i,j} \approx AA_{i,k}$$

The distance of χ^2 then becomes:

$$\chi^2 \approx 17 \sum_{i=1}^{17} \frac{(AA_{i,j} - AA_{i,k})^2}{AA_{i,j}} \approx 17 \sum_{i=1}^{17} \left(\frac{AA_{i,j}}{\sqrt{AA_{i,j}}} - \frac{AA_{i,k}}{\sqrt{AA_{i,j}}} \right)^2$$

$$\text{Let } \frac{AA_{i,j}}{\sqrt{AA_{i,j}}} = B_{i,j}$$

$$\text{Let } \frac{AA_{i,k}}{\sqrt{AA_{i,j}}} = B_{i,k}$$

$$\text{Then } \chi^2 \approx 17 \sum_{i=1}^{17} (B_{i,j} - B_{i,k})^2 \quad (6)$$

This formula, apart from the factor 17, is the formula of a residual sum of squares;

Moreover,

$$B_{i,k} = \frac{AA_{i,k}}{\sqrt{AA_{i,j}}}$$

which becomes, with equation 4:

$$B_{i,k} = \frac{a_1 AA_{i,1} + a_2 AA_{i,2} + \dots + a_n AA_{i,n}}{\sqrt{AA_{i,j}}}$$

$$B_{i,k} = a_1 \frac{AA_{i,1}}{\sqrt{AA_{i,j}}} + a_2 \frac{AA_{i,2}}{\sqrt{AA_{i,j}}} + \dots + a_n \frac{AA_{i,n}}{\sqrt{AA_{i,j}}} \quad (7)$$

Therefore we obtain i equations with n variables. These equations are solvable through a multiple regression analysis minimizing the residual sum of squares:

$$\sum_{i=1}^{17} (B_{i,j} - B_{i,k})^2$$

However, instead of using reference proteins as in the method for the calculation of $AA_{i,k}$, each amino acid value for the reference proteins is divided by the square root of the corresponding amino acid in the digesta to be analysed. The calculation involves $B_{i,j}$ instead of $AA_{i,j}$. The advantages of this method are first of all to give the same weight to each amino acid of the profile and secondly to provide t values for the coefficients in order to perform statistical tests.

Examples

To test the validity of the model, 2 mixtures of proteins from 3 different sources were prepared. These sources were a sample of rumen bacteria, a pea flour and a calf milk substitute powder. The amino acid compositions of the mixtures and of each of the products were determined as described by Guilloteau *et al* (1986) (table I). Estimates of the proportions of these protein products were then calculated using the formulas in the model and were compared to the actual values. Results given in table II showed close agreement between the calculated and actual proportions of protein products in the experimental mixtures (the "fit" coefficient was 0.92).

Moreover, only the estimated proportion of pea flour in mixture 1 (0.08) was not statistically significant ($P > 0.05$). A further test was made on a known mixture of soyabean meal, casein and pea flour. Applying the amino acid compositions of the 3 products and that of the mixture to the model formulas again gave calculated proportions similar to actual proportions (table III). The summary statistics indicated a close fit with a coefficient of 0.98 and that only the value for pea flour (0.14) was not significant.

Differences between real and calculated compositions are probably due to sampling and analytical errors as well as to similarities between products (for example between soyabean meal and pea flour in the last mixture), but overall, the derived compositions are close to reality, showing that the method is valid. It is of interest to note that the proportion of a component close to 0.1 is usually statistically insignificant, supporting the fact that it is unnecessary to include too many variables in the calculated mixture.

To examine the value and validity of the method, we also took some examples from our experimental results and from the literature. Table IV shows the amino acid composition of the digesta obtained in skim-milk-fed calves (Guilloteau *et al*, 1980) together with amino acid composition of dietary protein, of pig fecal bacteria (Mason *et al*, 1976) used to represent the amino acid composition of intestinal bacteria protein and of germ-free milk-fed lamb feces (Combe, 1976) used to represent the amino acid composition of the undigested endogenous protein since the true digestibility of milk protein is considered to be complete (Roy *et al*, 1970). The data in table IV were used in the formulas of the present method. The proportions of the 3 reference proteins which gave the "best fit" with the actual amino acid profile of the diet based on skim-milk were found to be

Table I. Amino acid composition of bacteria, pea flour, milk substitute proteins and 2 mixtures of these 3 proteins (percent of assayed amino acids). ¹ Solid phase adherent bacteria isolated from rumen content of weaned calves; ² Dehulled pea flour; ³ Protein exclusively provided by skim-milk powder and synthetic amino acids (DL methionine and HCl-L lysine); * See table II.

<i>Protein Amino acid</i>	<i>Product</i>			<i>Mixture*</i>	
	<i>Bacteria</i> ¹	<i>Pea flour</i> ²	<i>Skim-milk</i> ³	<i>1</i>	<i>2</i>
Aspartic acid	13.10	12.35	8.06	8.99	11.30
Threonine	5.58	3.74	4.72	4.17	4.55
Serine	4.74	4.38	4.82	5.32	4.40
Glutamic acid	12.92	18.43	19.72	18.68	14.72
Proline	3.59	4.07	9.06	7.46	4.60
Glycine	5.70	4.76	1.91	3.19	4.96
Alanine	7.27	4.92	3.10	4.59	6.19
Valine	6.71	5.58	6.38	6.39	6.34
Cystine	1.46	1.50	0.88	1.30	1.51
Methionine	2.12	1.20	3.88	3.26	1.87
Isoleucine	7.07	4.82	5.21	5.93	6.11
Leucine	8.23	7.70	9.67	8.87	9.30
Tyrosine	2.48	3.09	4.46	4.31	3.73
Phenylalanine	5.29	4.83	4.41	4.91	5.42
Lysine	6.39	7.37	7.89	6.32	6.73
Histidine	2.23	2.45	2.56	2.27	2.07
Arginine	5.13	8.81	3.25	4.06	6.20

Table II. Real and calculated proportions of bacteria, pea flour and milk substitute proteins in 2 mixtures.

<i>Mixture</i>	<i>Proportion of protein in the diet</i>					
	<i>Real</i>			<i>Calculated</i>		
	<i>Bacteria</i>	<i>Pea flour</i>	<i>Skim-milk</i>	<i>Bacteria</i>	<i>Pea flour</i>	<i>Skim-milk</i>
1	0.30	0.10	0.60	0.24	0.08	0.68
2	0.59	0.30	0.11	0.55	0.31	0.14

Table III. Summary statistics obtained after applying the method of calculation to a mixture of 3 proteins at the following proportions: 0.68, 0.22 and 0.10 supplied by a soyabean meal, a casein and a pea flour respectively. Each amino acid of each variable has been divided by the square root of the corresponding amino acid in the mixture. Coefficient of determination 0.98; SE: standard error of the estimate; DF: degrees of freedom; NS: non-significant.

<i>Dependent variable: mixture/$\sqrt{\text{mixture}}$</i>					
<i>Variable</i>	<i>Theoretical proportion</i>	<i>SE</i>	<i>t statistic</i>	<i>Prob > t</i>	
Soyabean/ $\sqrt{\text{mixture}}$	0.69	0.16	4.28	0.001	
Casein/ $\sqrt{\text{mixture}}$	0.17	0.04	3.94	0.001	
Pea flour/ $\sqrt{\text{mixture}}$	0.14	0.13	1.08	NS	

<i>Source</i>	<i>Sum of squares</i>	<i>DF</i>	<i>Mean squares</i>	<i>F-ratio</i>	<i>Prob > F</i>
Model	8.74	2	4.37	298.4	0.000
Error	0.21	14	0.015		
Total	8.95	16			

0.17, 0.69 and 0.13 respectively. The summary statistics given in table V show that the amino acid composition of the theoretical mixture gives a close fit (coefficient of determination of 0.84) but that only the percentage of endogenous protein is statistically significant. From these results we can conclude that the digestion of milk protein in the small intestine is almost complete since the protein of ileal digesta is mainly of endogenous origin. Furthermore, the relative values for the mixture of reference proteins obtained using the model formulas were similar to estimates of 0.16, 0.71 and 0.13 for dietary, endogenous and bacterial proteins previously obtained using the method based on calculations of the minimum distance of χ^2 (Guilloteau *et al*, 1980).

Instead of using feces of germ-free lambs for endogenous protein, estimates could be based on mean values of amino acid compositions of feces of germ-free lambs and calf's meconium (Gronget *et al*, 1981). Likewise, values for intestinal bacteria could be derived from mean values of amino acid compositions of bacteria isolated from pig faeces and bacteria isolated from sheep feces (Mason, 1979). Using these mean values in the model gave estimates of 0.18, 0.77 and 0.04 for the proportions of dietary, endogenous and bacterial proteins respectively. Although this substitution did not alter the "fit" coefficient (this remained at 0.84), statistical significance was assigned to the proportions of both dietary and endogenous protein fractions.

Table IV. Amino acid composition of dietary, endogenous and bacterial components and of ileal digesta collected in preruminant calves fed a skim-milk diet (percent of assayed amino acids). ¹ (Guiloteau *et al*, 1980); ² Germ-free lamb faeces (Combe, 1976); ³ Bacteria isolated in pig faeces (Mason *et al*, 1976).

<i>Protein</i>	<i>Skim-milk diet</i> ¹	<i>Endogenous component</i> ²	<i>Bacterial component</i> ³	<i>Ileal digesta</i> ¹
<i>Amino acids</i>				
Aspartic acid	7.51	10.30	11.4	9.72
Threonine	4.52	14.41	5.3	10.08
Serine	4.90	7.61	4.7	6.66
Glutamic acid	21.63	14.32	12.3	16.42
Proline	9.70	8.27	3.9	7.93
Glycine	1.91	4.25	5.4	4.92
Alanine	3.08	5.14	6.7	4.96
Valine	6.06	5.56	6.5	6.17
Cystine	0.98	3.07	1.7	3.78
Methionine	2.80	1.50	2.4	1.24
Isoleucine	4.99	3.28	5.6	3.72
Leucine	9.00	5.50	8.8	6.23
Tyrosine	4.76	2.69	4.8	3.32
Phenylalanine	4.76	2.79	6.0	3.23
Lysine	7.55	5.03	6.9	5.71
Histidine	2.52	2.20	2.2	2.12
Arginine	3.36	4.07	5.0	3.75

DISCUSSION

Since feed, microbial or endogenous residues cannot readily be isolated from digesta, their amino acid profiles must be either directly assayed or taken from the literature to create a bank of reference proteins. For example, for the amino acid composition of bacteria it is possible to use that of bacteria isolated from the feces of pig (Mason *et al*, 1976) or sheep (Mason, 1979). Similarly, the amino acid composition of undigested endogenous protein can be derived from that of calf meconium (Grongnet *et al*, 1981) or feces of germ-free milk-fed lambs (Combe, 1976). Thus, the amino acid

profile of digesta can be assembled from the amino acid compositions of reference proteins which approximate to those present in contents of the gut lumen. Usually 3 or 4 proteins can provide a satisfactory representation of the actual amino acid profile of digesta, especially if several of the proteins are present in small amounts in digesta and therefore are likely to be insignificant. However, the precision of the proposed computerized method will be limited by the accuracy of the amino acid analysis of the reference proteins included in the model. Furthermore, proteins from different sources may have similar profiles and therefore need to be selected

Table V. Summary statistics obtained after applying the method of calculation to ileal digesta from skim-milk-fed calves. Each amino acid of each variable has been divided by the square root of the corresponding amino acid in the ileal digesta. Coefficient of determination 0.84; ¹ (Guilloteau *et al*, 1980); ² Germ-free lamb faeces (Combe, 1976); ³ Bacteria isolated in pig faeces (Mason *et al*, 1976). SE: standard error of the estimate; DF: degrees of freedom; NS: non significant.

<i>Dependent variable: ileal digesta</i> ¹ / $\sqrt{\text{ileal digesta}}$					
<i>Variable</i>	<i>Calculated proportion</i>	<i>SE</i>	<i>t statistic</i>	<i>Prob > t</i>	
Diet ¹ / $\sqrt{\text{ileal digesta}}$	0.17	0.08	2.124	NS	
Endogenous ² / $\sqrt{\text{ileal digesta}}$	0.69	0.07	9.491	0.001	
Bacteria ³ / $\sqrt{\text{ileal digesta}}$	0.13	0.08	1.501	NS	

<i>Source</i>	<i>Sum of squares</i>	<i>DF</i>	<i>Mean squares</i>	<i>F-ratio</i>	<i>Prob > F</i>
Model	6.67	2	3.34	37.06	0.001
Error	1.26	14	0.09		
Total	7.93	16			

according to their likely representation of those in digesta.

The confidence that can be attached to any particular solution can be partially determined from the summary statistics *ie* the accountable variance or the coefficient of determination R^2 . If the value of this coefficient is high (≥ 0.70), it indicates that the estimation is good. If the value is low, it could show that 1 or many reference proteins have either been uncorrectly chosen or have been partially modified by the digestive process. However, because of the fact that the regression is constrained to pass through the origin, the summary statistics as well as the degrees of freedom used in the *t* tests performed on the coeffi-

cient values are controversial (Gordon, 1981). For this reason we have decided to use the number of observations (up to 17) minus the number of reference proteins for the degrees of freedom of the residual sum of squares and the number of observations minus 1 for the degrees of freedom of the total sum of squares. Using this method, a negative coefficient might be obtained, but it then indicates that a reference protein has been uncorrectly included in the model and needs to be removed or substituted. In an ideal case the sum of the coefficients should be unity, but due to analytical errors and assumptions in the hypothesis, this is almost never exactly the case. Nonetheless, the obtained values should be a close approximation.

CONCLUSION

The model formulas described in this paper permit estimation of the most probable composition of the protein mixture present in a digesta, and provide an assessment of the extent of digestion by estimating the proportions of different protein components of dietary, endogenous and microbial origin. In comparison with previous methods (Evans *et al*, 1975), the present model has the advantage of giving the same weight to each amino acid and can be easily executed with the aid of computer programs for multiple regression analysis. Furthermore, the result is supported by statistical analysis which can be useful to test the validity of the fit and the significance of the calculated proportions in the mixture. However, the confidence that can be attached to any particular solution, and therefore the quality of the interpretation, depends on the precision of analysis of the amino acid profiles of individual reference proteins included in the model and on the extent to which these profiles are likely to be representative of the components of the digesta. Also, as far as digesta are concerned, one must bear in mind that the dietary protein escaping digestion into the small intestine can be some particular fractions of which amino acid compositions are very different from that of the whole dietary protein (Guilloteau *et al*, 1980).

ACKNOWLEDGMENTS

Thanks are due to M Formal for the amino acid composition measurements and to JP Lallès who provided the sample of rumen bacteria.

REFERENCES

- Cockburn JE, Williams AP (1984) The simultaneous estimation of the amounts of protozoal, bacterial and dietary nitrogen entering the duodenum of steers. *Br J Nutr* 51, 111-132
- Combe E (1976) Influence de la microflore intestinale sur la composition en acides aminés des fèces des agneaux. *CR Séanc Soc Biol* 170, 787-793
- Cottle DJ, Nolan JV (1982) The use of amino acid profiles to measure microbial protein outflow from the rumen. *Proc Aust Soc Anim Prod* 14, 588-591
- Evans RA, Axford RFE, Offer NW (1975) A method of estimating the quantities of microbial and dietary proteins flowing in the duodenal digesta of ruminants. *Proc Nutr Soc* 34, 65A
- Gordon HA (1981) Errors in computer packages. Least squares regression through the origin. *Statistician* 30, 23-29
- Grongnet JF, Patureau-Mirand P, Toullec R, Prugnaud J (1981) Utilisation des protéines du lait et du lactosérum par le jeune veau préruminant. Influence de l'âge et de la dénaturation des protéines du lactosérum. *Ann Zootech* 30, 443-464
- Guilloteau P, Patureau-Mirand P, Toullec R, Prugnaud J (1980) Digestion of milk protein and methanol-grown bacteria protein in the preruminant calf. II. Amino acid composition of ileal digesta and faeces and blood levels of free amino acids. *Reprod Nutr Develop* 20, 615-629
- Guilloteau P, Sauvant D, Patureau-Mirand P (1983) Methods of comparing amino acid composition of proteins: application to undigested proteins in the preruminant calf. *Ann Nutr Metab* 27, 457-469
- Guilloteau P, Toullec R, Grongnet JF, Patureau-Mirand P, Prugnaud J, Sauvant D (1986) Digestion of milk, fish and soya-bean protein in the preruminant calf: flow of digesta, apparent digestibility at the end of the ileum and amino-acid composition of ileal digesta. *Br J Nutr* 55, 571-592
- Lindqvist B, Östgren J, Lindberg I (1975) A method for the identification and quantitative investigation of denatured proteins in mixtures based on computer comparison of amino acid patterns. *Z Lebensm Unters Forsch* 159, 15-22

- Mason VC (1979) The quantitative importance of bacterial residues in the non-dietary faecal nitrogen of sheep. I. Methodology studies. *Z Tierphysiol Tierernähr Futtermittelkd* 41, 131-139
- Mason VC, Just A, Bech-Andersen S (1976) Bacterial activity in the hindgut of pigs. II. Its influence on the apparent digestibility of nitrogen and amino-acids. *Z Tierphysiol Tierernähr Futtermittelkd* 36, 310-324
- Patureau-Mirand P, Toullec R, Guilloteau P, Pion R (1977) Influence de la nature des protéines alimentaires sur la composition en acides aminés des fèces du veau préruminant. *Ann Biol Anim Bioch Biophys* 17, 71-83
- Roy JHB, Stobo IJF, Gaston HJ (1970) The nutrition of the veal calf. II. The effect of different levels of protein and fat in milk substitute diets. *Br J Nutr* 24, 411-457
- Vérité R, Poncet C, Chabi S, Pion R (1977) Utilisation des tourteaux traités au formol par les vaches laitières. I. Aspects digestifs. *Ann Zootech* 26, 167-181
- Williams EJ (1959) *Regression Analysis*. Wiley and Sons, New York, 214