Critical Survey of Interrelationships of Nitrogen and Amino Acids in Seeds

Jacques Landry

Laboratoire de Chimie Biologique, INRA, CBAI, Centre de Grignon, 78850 Thiverval-Grignon, France

The present paper is intended for removing inconsistencies from available information concerning the interrelationhips of nitrogen (N) and amino acids in seeds (or grains). It consists of a critical survey of data reported in a series of papers by Mossé and collaborators (see Mossé and Huet, *Sci. Aliments* **1990**, *10*, 151 for review). These authors contended for perfectly linear relationships ($A_i = a_i N + b_i$) to hold between N and amino acid i (g/100 g seed dry matter) for seeds of a given species on the basis of correlation coefficient (r) found to be close to 1. Three linear regressions are called upon to analyze variations in the amino acid compositions of 33 samples of pea seeds. The regression C_i (g amino acid i/16gN) $= 16a_i + 16b_i N^{-1}$, not used customarily, is the more amenable to accommodate all the facts in pointing out values significantly lower than 1 for r_i^2, and parameters a_i and b_i similar to those found with the regression $A_i = NC_i/16 = a_i N + b_i$ whose r_i^2 values, very close to 1, are overestimated on account of the presence of N in both sides of equation and sampling limited in number. Examination of variations in lysine content of maize grains as a function of N shows that the phenotypic variations for a given genotype are a poor fit to the linear relationship $A_{lys} = aN + b$ available in the literature, as determined from 153 commercial samples. It is concluded that linear N-amino acid relationships hold only approximately for predicting amino acid composition of seed of a given species from its N content.

INTRODUCTION

Improvement of protein quantity and quality of cereal grains and legume seeds involves the effects of environment, agricultural practices, and genetic factors upon the amino acid composition to be evaluated with accuracy. In this context numerous investigations have sought after the changes in level of any proteic amino acid in dry matter or in protein with the content of nitrogen (N) or crude protein in dry matter. Among them two are worth mentioning for the wealth of available data. The first one concerns the work of Mossé and collaborators (Baudet and Huet) published in a series of papers (see Mossé and Huet, 1990 for references). These authors have analyzed the variations of seed amino acid composition as a function of N content for 16 species of cereals, legumes, and oilseeds, from a mean of 19 samples per species. The amino acid composition for every sample was determined from three hydrolyses (15, 24, and 48 h) carried on in boiling 6 M HCl to take account of partial destruction and progressive release, one hydrolysis (18 h) performed in the same medium on a previously oxidized sample to evaluate methionine and cystine, and one barytic hydrolysis to quantify tryptophan. The samples originate from diverse genotypes or the same genotype grown in various environmental conditions inducing wide variations of N content. On the basis of correlation coefficients (r) nonsignificantly different from 1, found by regressing the content of each amino acid in the dry matter of whole seed against that of N, a perfect linear function was claimed to occur between both variables for a given species, irrespective of sample origin. However, values lower than 1 for r were found for lysine, cystine, methionine, and tryptophan in some cases. They were attributed to difficulties encountered to assay these amino acids accurately on account of their partial destruction during hydrolysis and not to genetic variations which were considered as small and undetected by analyses. Such statements, confirmed by some comparisons with literature data, led to the conclusion that the amino acid pattern of any seed can be predicted with accuracy from its N content and regression equations (Mossé and Huet, 1990).

The second work comes from Heimbeck and Balschukat (1990), who gathered the amino acid compositions of raw material in a book entitled "*The Amino Acid Composition of Feedstuffs*". It differs from work of Mossé, Baudet, and Huet in having (1) a lower number of studied species (nine); (2) a larger number of samples (from 19 to 153 per species) examined for their amino acid composition determined from three different hydrolysates per sample obtained by heating crude sample in the presence of acid or alkali and previously oxidized sample with acid; (3) a narrower variability for N, as a general rule; (4) a list of mean, minimum, and maximum values for N content and amino acid content in dry matter and in protein for each species; (5) a statement of regression equations between the content A_i of amino acid i in dry matter and N content for methionine, lysine, threonine, tryptophan, and arginine only, with r values ranging from 0.20 (methionine for pea seed) to 0.96 (arginine for cotton seed) and generally lower than those reported by Mossé and Huet (1990); (6) little discussion of the data. The only conclusion given by the authors was the possible quantitative estimation of amino acids from N content and regression equations.

On the other hand Landry and Delhaye (1993, 1997) determined the tryptophan content of grains and seeds using an assay procedure which was proved to be reliable (Landry and Delhaye, 1996) and concluded that

Table 1. Statistical Analysis of Amino Acid Compositions of Pea Seed Samples Using Three Regressions[a,b]

i	$A_i = a_iN + b_i$ [c]			$C_i/16 = a_i + b_iN^{-1}$ [d]			$C_i = \alpha_i + \beta_iN$ [e]		
	a_i	b_i	r^2	a_i	b_i	r^2	α_i	β_i	r^2
Lys	0.374	0.358	0.984	0.375	0.351	0.794	8.99	−0.387	0.787
His	0.152	0	0.972	0.155	−0.008	0.007	2.47	−0.004	0.002
Arg	0.972	−1.497	0.945	0.951	−1.416	0.750	3.10	1.589	0.767
Asx	0.774	−0.152	0.982	0.774	-0.151	0.154	11.10	0.168	0.154
Thr	0.165	0.297	0.931	0.168	0.286	0.826	5.13	−0.319	0.834
Ser	0.268	0.147	0.970	0.272	0.130	0.398	5.49	−0.151	0.436
Glx	1.122	−0.286	0.990	1.129	−0.311	0.409	15.48	0.327	0.366
Pro	0.238	0.123	0.947	0.244	0.100	0.202	4.79	−0.121	0.237
Gly	0.217	0.218	0.981	0.218	0.212	0.778	5.30	−0.236	0.780
Ala	0.213	0.222	0.976	0.216	0.212	0.766	5.26	−0.237	0.778
Val	0.265	0.181	0.955	0.268	0.169	0.383	5.74	−0.191	0.397
Cys	0.037	0.220	0.567	0.037	0.219	0.777	2.44	−0.240	0.757
Met	0.047	0.065	0.874	0.047	0.066	0.523	1.30	−0.071	0.487
Ile	0.234	0.177	0.933	0.237	0.179	0.382	5.24	−0.194	0.365
Leu	0.420	0.108	0.980	0.420	0.106	0.203	7.62	−0.116	0.199
Tyr	0.141	0.289	0.933	0.147	0.268	0.671	4.65	−0.304	0.699
Phe	0.231	0.285	0.951	0.240	0.252	0.653	6.03	−0.393	0.711
Trp	0.048	0.010	0.834	0.047	0.014	0.032	0.86	−0.014	0.024

[a] Values, expressed as g/100 g (dry matter), are calculated from data reported by Mossé et al. (1987). [b] Sample number: $n = 33$. [c] Parameters calculated by Mossé et al. (1987). [d] $C_i/16$ is regressed instead of C_i in order to point out the similarity of values of a_i and b_i with those obtained using the previous regression. [e] C_i is regressed instead of $C_i/16$ in view of a possible comparison with the regressions reported by Eppendorfer and Bille (1974).

regression equations had a low predictive value for quantifying this amino acid.

The present paper is aimed at elucidating the origins of these inconsistencies about interrelationship between amino acid and nitrogen in seeds (or grains) by a critical survey of the data reported by Mossé et collaborators. First, three linear relationships will be compared with a view to obtaining the closest fit with variations in amino acid composition of 33 samples of pea seed by considering the data stated by Mossé et al. (1987). Then the variations of lysine and nitrogen in the dry matter of maize grains, as reported by Baudet et al. (1986), will be examined in the light of literature data. To avoid any confusion the same symbolism as that proposed by these authors will be used for designating amino acid, its concentration in dry matter or protein, and parameters of regression equations.

ANALYSIS OF VARIATIONS IN THE AMINO ACID COMPOSITIONS OF 33 SAMPLES OF PEA SEED AS REPORTED BY MOSSE ET AL. (1987)

As depicted in numerous graphs presented by Mossé et al. (1987) and others, it is obvious that the plot of the amount A_i of amino acid i in the dry matter of seed (or grain) of a given species as function of its N content can be suitably approximated by a straight line corresponding to the following equation

$$A_i = a_iN + b_i \qquad (1)$$

where a_i and b_i are constants which are apparently independent of phenotypic and genotypic variations and significantly different from 0 in general.

$$\text{but} \quad A_i = NC_i/16$$

where C_i is the concentration of amino acid i in 100 g protein or 16 g N.

Dividing both sides of eq 1 by N leads to

$$C_i/16 = a_i + b_iN^{-1} \qquad (2)$$

and regressing $C_i/16$ vs N^{-1} using the unweighted linear least-squares method allows a_i, b_i, and r_i (correlation coefficient) to be determined.

In Table 1 are listed the statistical parameters (a_i, b_i, and r_i^2) relative to pea seed and concerning the regression $A_i = a_iN + b_i$ as they were reported by Mossé et al. (1987) together with those pertaining to the regression $C_i/16 = a_i + b_iN^{-1}$ as it can be calculated from the amino acid compositions published by the same authors.

Comparing data emphasizes several essential facts:

(1) The parameters a_i and b_i from both regressions are practically the same. The slight differences recorded, although not significant, are due to the fact that the concentrations corresponding to the highest N contents compared with those associated with lowest N levels have higher statistical weight when A_i is regressed against N. The converse is true when C_i is regressed against N^{-1}. The same was observed earlier by Landry and Moureaux (1984) about the accumulation of amino acids in developing grain of maize.

(2) The determination coefficients r_i^2 are close to 1 for the regression A_i vs N. This lays in the fact that the changes of the whole (protein) are closely paralleled by the changes of its constituents (amino acids), which is evidenced by the presence of N in both sides of the regression equation. This induces a bias in the determination of r with resulting overestimation. The misuse of regression methods when one variate is modified by some function of the other have been mentioned by Chanter (1981) who indicated that the relationship between xy and x will be linear even if x and y are completely unrelated.

(3) The determination coefficients r_i^2 for the regression C_i vs N^{-1} range from 0.007 (histidine) to 0.826 (threonine). From the above background the content of each amino acid in dry matter and that of N are not strictly interdependent. This is due to the histological complexity of seed and, at a more significant extent, of grain.

For example maize grain is made up of germ, floury and horny endosperm, and envelops. Hamilton et al. (1951) have isolated these different parts from 40 grain samples originating from five hybrids grown under eight various conditions and analyzed them for nitrogen. Their results allow the nitrogen level in parts to be

Table 2. Relationships between Nitrogen of Maize Grain and Its Histological Parts[a]

$N_{germ} = 0.153 N_{grain} + 0.111$	$r^2 = 0.31$	$n^b = 7$
$N_{floury\ endosperm} = 0.085 N_{grain} + 1.012$	$r^2 = 0.77$	$n = 6$
$N_{horny\ endosperm} = 0.659 N_{grain} - 1.018$	$r^2 = 0.88$	$n = 7$

[a] Values, expressed as milligrams of nitrogen (dry matter), are obtained from data reported by Hamilton et al. (1951) using the regression $N_{part}/N_{grain} = a + b/N_{grain}$. [b] Sample number.

regressed as a function of nitrogen level of the whole grain. The regression equations, calculated from percentages against N^{-1} (Table 2), point out an uneven distribution of protein in grain. Exogeneous nitrogen, when it is supplied at low proportions, is accumulated mainly in germ and floury endosperm. For higher supplies, it is accumulated at a high level mainly in horny endosperm and at a lesser extent in floury endosperm and germ. Furthermore accumulation in germ is aleatory (low values for r^2).

On the other hand, since storage proteins (zeins), which have amino acid composition different from that of metabolic proteins, are lacking in germ and present at different proportions in floury and horny endosperms, the amino acid composition of extra proteins synthesized as a consequence of high accumulation of exogeneous nitrogen cannot be constant. It is dependent upon phenotype and genotype alike as shown below. In other words both the temporal and developmental regulation of accumulation of storage proteins in the grain is not so accurate as that advocated by Mossé et al. (1986), for whom maize grain paralleled bacterial ribosome regarding the strict control mechanism of protein synthesis. Actually, the r_i^2 value recorded for the regression C_i vs N^{-1} is linked to the difference of the level of amino acid i between metabolic and storage proteins and to the relative proportions of these both classes of proteins in seed.

(4) Regressing C_i against N, i.e., $C_i = \alpha_i + \beta_i N$ yields r_i^2 values practically identical to those obtained when C_i is regressed against N^{-1}. Eppendorfer et al. (1985),

although they noticed higher r_i values for the regression A_i vs N, adopted such a type of regression, probably to emphasize the high variability of C_i recorded for a given N content. Nevertheless, the equation $C_i = \alpha_i + \beta_i N$ leads to $16A_i = \alpha_i N + \beta_i N^2$, which is depicted by a portion of parabola passing through the origin, and having the x axis as axis and concavity downward if β_i is negative or the y axis as axis and concavity upward if β_i is positive. Such an equation integrates the variations of A_i vs N mainly during the seed (or grain) development, variations which can be different from those seen with ripe material. It requires that the accumulation rate of amino acid i in dry matter is a linear function of N content, whereas it is constant when eqs 1 or 2 are considered. In fact, the parabola can be approximated by a straight line for a limited variation of N, but in such a way of operating does not constitute an advantage in respect to use of the regression C_i vs N^{-1}.

PHENOTYPICAL AND GENOTYPICAL VARIATIONS OF LYSINE CONTENT IN DRY MATTER AS EVIDENCED FROM DATA OF BAUDET ET AL. (1986) FOR 30 MAIZE GRAIN SAMPLES

From the above background it can be concluded that there is not a perfect relationship between the content of any amino acid in the dry matter or in protein of grain (or seed) and N content, but the relationships $A_i = a_i N + b_i$ and $C_i/16 = a_i + b_i N^{-1}$ are able to define the approximate dependence between the content of amino acid i in dry matter or in protein and N content. The values of parameters a_i and b_i characterize the sampling and the population of normal varieties if the former is representative of the latter. Asserting such a statement requires the data reported by Mossé and co-workers to be compared with those available in the literature. For the sake of brevity the comparison will be limited to lysine of maize grain since the low content of this essential amino acid has originated numerous investigations with a view to find the genotypes the richest in lysine.
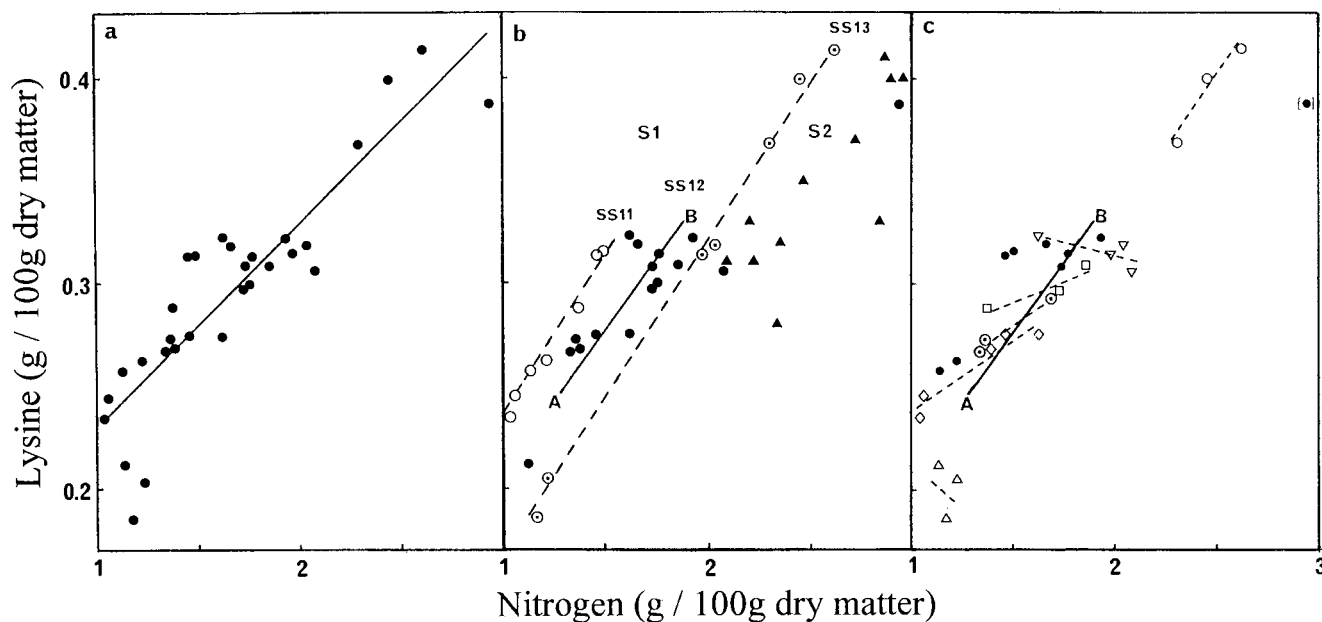
Figure 1. Relationships between lysine and nitrogen contents of maize grains: (a) data and regression line from Baudet et al. (1986); (b) distribution of data (●, ○, ◉) from Baudet et al. (1986) into two sets S1 (subdivided into SS 11, 12, and 13) and S2, data ▲ from Pollmer et al. (1977) and regression line AB determined from 153 maize samples of commercial varieties by Heimbeck and Balschukat (1990); (c) influence of genotype according to data from Baudet et al. (1986) for Pride (△), Axia (◇), Dea (◉), Liza (□), 80B97 (▽), IHP (○), and other varieties (●) IHP excluded ([●]), and regression line AB from Heimbeck and Balschukat (1990).

Table 3. Accumulation Rate of Lysine (g/g Nitrogen) as Influenced by Genotype[a,b]

| genotypes	lysine/N	n^c
axia	0.074	5
dea	0.074	3
liza	0.038	3
pride	−0.065	3
80B97	−0.026	4
IHP	0.136	3
$X^{b,d}$	0.140	153

[a] Calculated from data taken from Baudet et al. (1986). [b] Calculated from data from Heimbeck and Balschukat (1990). [c] Sample number. [d] Unknown commercial varieties.

Figure 1a depicts the data reported by Baudet et al. (1986) concerning the lysine content A_{lys} of 30 samples of maize grain. The regression line given by these workers was $A_{lys} = 130(\pm 0.016) + 0.099(\pm 0.009)N$ with $r^2 = 0.80$, and this moderate value was considered as originating from imperfect analytical determination of lysine. The regression line $C_{lys}/16 = 0.107(\pm 0.003) + 0.118(\pm 0.019)N^{-1}$ with $r^2 = 0.59$ substantiates the above observations made with pea since parameters are not significantly different and r^2 is lower. The dots of Figure 1a can be gathered into two subsets S1 and S2 (Figure 1b). S1 is divided into three subsubsets SS11, SS12, and SS13, the former and the latter being made up of dots defining two regression lines practically parallel and delineating the left and right boundary-lines of SS12, respectively. S2 is made up of two dots (● in Figure 1b) only. Such a distribution is substantiated by plotting some literature data, as shown in Figure 1b. The values of set S1 fitted in with regression line AB, plotted from equation $A_{lys} = 0.065 + 0.140N$ ($r^2 = 0.41$) established by Heimbeck and Balschukat (1990) from 153 maize samples: AB is practically parallel with regression lines SS11 and SS12 and located at nearly equal length from them. On the other hand S2 contains nine dots (▲ in Figure 1b) depicting the lysine content of some Elite conventional and high-protein inbred lines of maize, as reported by Pollmer et al. (1977).

Regression line AB is assumed as representing the phenotypic and genotypic changes of lysine content of commercial varieties as function of nitrogen content. Differentiating data given by Baudet et al. (1986) according to their varietal origin emphasizes the occurrence of regression lines A_{lys} vs N whose the estimated slopes (or relative accumulation rates) are in general lower than that of the line AB, as shown in Figure 1c and in Table 3. The relative location of regression lines in respect of AB line for a given variety would be linked to genotype and to the use of fertilizers other than nitrogen. This can be generalized by integrating the regression lines AB and CD, depicting the data of Heimbeck and Balschukat (1990) and Baudet et al. (1986), respectively, in the plot reported by Pollmer et al. (1977). In this plot variations of the lysine and nitrogen contents of maize grains originated from 170 conventional inbred lines and 50 high-protein lines are depicted by two regression lines EF and GH, respectively, with slopes of 0.086 and 0.076 (Figure 2). Such values are similar or identical to those found for the slopes of lines concerning the phenotypic variations of lysine for the varieties Dea and Axia, as shown in Table 3. On the other hand, Figure 2 points out the inadequacy of the regression line and equation given by Baudet et al. (1986) for predicting lysine content of commercial varieties grain from its N content or for assessing the potential value of a given genotype to accumulate lysine with respect to the mean of genotypes representing a population or a subset. In other words, the 30 pairs of values (A_{lys},N) corresponding to grain samples chosen in order to cover a wide range of N contents cannot be considered as representative of all possible such pairs.

CONCLUSION

It is evident that the above observations made about the r^2 values for the linear regression $A_i = a_iN + b_i$ for pea seed and lysine changes in maize grain can be extended to other data published by Mossé and co-
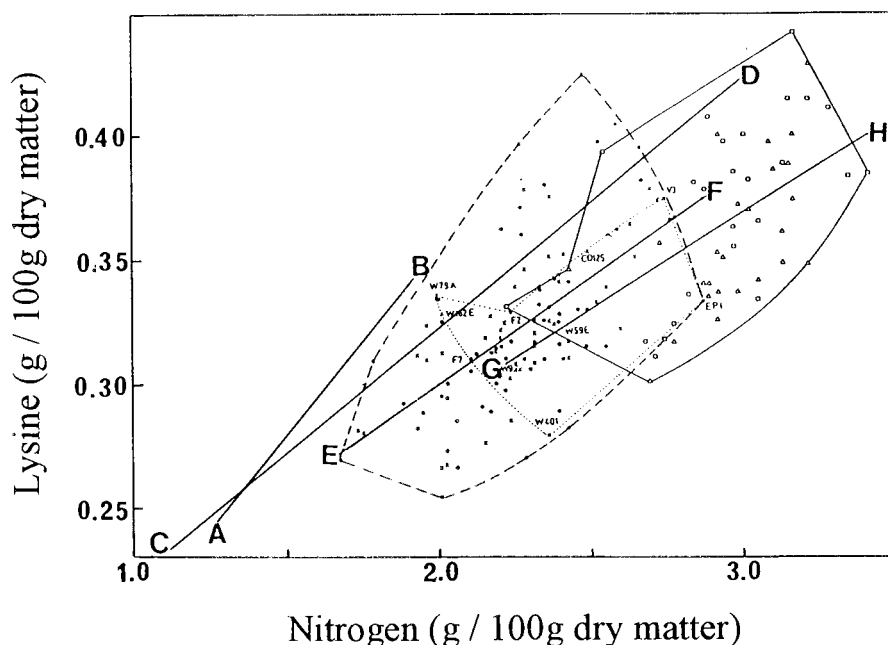
Figure 2. Relationship between lysine and nitrogen contents of maize grains: AB, regression line determined from 153 samples of commercial varieties by Heimbeck and Balschukat (1990), CD, regression line determined from 30 samples by Baudet et al. (1986), EF and GH regression lines, and dots from Pollmer (1977).

workers. So, the linear regression $C_i/16 = a_i + b_i N^{-1}$ appears to give the more adequate fit with the facts about the connection between the concentration C_i of amino acid i in protein and nitrogen (or crude protein) of grain (or seed) of a given species since (1) the coefficients of determination (r^2) are significantly lower than 1, as found also with the regression $C_i = \alpha_i + \beta_i N$; (2) the parameters a_i and b_i have values not significantly different from those obtained from the regression $A_i = a_i N + b_i$; and (3) the r^2 values, close to unity, obtained from the regression $A_i = a_i N + b_i$ originate in a bias due to the presence of independent variable (N) on both sides of the equation and to limited sampling. On the other hand r^2 values significantly lower from unity for the regression $A_i = a_i N + b_i$ despite the bias are indicative of high variability of C_i for a given value of N, coming from genotype and not from the methodology used for quantifying amino acids. From the foregoing relationships $A_i = a_i N + b_i$ and $C_i/16 = a_i + b_i N^{-1}$ are far from perfect. Their use can lead to quantitative estimation of amino acid in seeds with a degree of uncertainty depending on diverse factors, such as species and amino acid, taken into consideration together with the sampling from which the relationship is established. In other words application of the above relationships allows neither (1) better adjustment of diet or a better use of raw material for cattle breeder; (2) prediction of fertilization effects or genetic variability to improve amino composition for agronomists and geneticists, respectively; and (3) better knowledge of raw material for nutrionists working on determination of feeding value of the proteins in the grains (or seeds) eaten by livestock. These conclusions are opposite to those drawn by Mossé (1990).

The present survey shows that as a general rule, using the regression $(x/\Sigma x) = a + b(\Sigma x)^{-1}$ allows one to give a truer picture of relationship between a whole (Σx) and one of its part (x) when the former varies. Such a procedure should avoid reporting of overestimated correlation coefficients when x is regressed vs Σx, and the resulting erroneous inferences.

LITERATURE CITED

Baudet, J.; Huet, J. C.; Mossé, J. Variability and relationships among amino acids and nitrogen in maize grains. *J. Agric. Food Chem.* **1986**, *34*, 365−37.

Chanter, D. O. The use and misuse of linear regression methods in crop modelling. In *Mathematics and Plant Physiology;* Rose, D. A., Charles-Edwards, D. A., Eds.; Academic Press: London, 1981; pp 253−267.

Eppendorfer, W. H.; Bille, S. W. Amino acid composition as a function of total-N in pea seeds grown on two soils with P and K additions. *Plant Soil* **1974**, *41*, 33−39.

Eppendorfer, W. H.; Bille, S. W.; Patipanawattana, S. Protein quality and amino acid−protein relationships of maize, sorghum and rice grain as influenced by nitrogen, phosphorus, potassium and soil moisture stress. *J. Sci. Food Agric.* **1985**, *36*, 453−462.

Hamilton, T. S.; Hamilton, B. C.; Mitchell, H. H. The dependence of physical and chemical composition of the corn kernel on soil fertility and cropping system. *Cereal Chem.* **1951**, *28*, 163−176.

Heimbeck, W.; Balschukat, D. In *The Amino Acid Composition of Feedstuffs*; Degussa: Hanau, Germany, 1990.

Landry, J.; Delhaye, S. The tryptophan contents of wheat, maize and barley grains as a function of nitrogen content. *J. Cereal Sci.* **1993**, *18*, 259−266.

Landry, J.; Delhaye, S. Tryptophan content of foods and feeds: conditions for a reliable quantitation. In *Recent Advances in Tryptophan Research. Tryptophan and Serotonin Pathways;* Allegri Filippini, G., Costa, C. V. L., Bertazzo, A., Eds; Plenum Publishing: New York, 1996; pp 703−709.

Landry, J.; Delhaye, S. Tryptophan contents of pea and broad bean seeds as a function of nitrogen content. *J. Agric. Food Chem.* **1997**, *45*, 3855−3858.

Landry, J.; Moureaux, T. A quantitative analysis of amino acid accumulation in developping grain of normal and *opaque-2* maizes. *J. Sci. Food Agric.* **1984**, *35*, 1051−1062.

Mossé, J. Amino acids in cereals, legumes and oilseeds: variation and prediction of composition as a function of seed nitrogen content in 16 crop species. Nutritional consequences. *INRA Prod. Anim.* **1990**, *3*, 103−119.

Mossé, J.; Huet, J. C.; Baudet, J. Evidence from amino acid analysis for the constancy of composition of storage protein accumulated in maize seed. *J. Plant Physiol.* **1986**, *125*, 199−205.

Mossé, J.; Huet, J. C. Amino acid composition and nutritional score for 10 cereals and six legumes or oilseeds: causes and ranges of variations according to species and to seed nitrogen content. *Sci. Aliments* **1990**, *10*, 151−173.

Mossé, J.; Huet, J. C.; Baudet, J. Changements de la composition en acides aminés des graines de pois en fonction de leur taux d'azote Variations in the amino acid composition of pea seeds in relation to their nitrogen content. *Sci. Aliments* **1987**, *7*, 301−324.

Pollmer, W. G.; Klein, D.; Dhillon, B. S. Evaluation of lysine and protein contents of maize inbred lines. *Cereal Res. Commun.* **1977**, *5*, 369−376.