Evaluation of Amounts and Pattern of Essential Amino Acids in Plant Seeds

W. F. Kwolek and C. H. VanEtten

The patterns for eight essential amino acids from 433 plant species are compared with the standard FAO/WHO pattern for hen's egg. The comparison procedure described is based on the average amount of essential amino acid in the total crude protein of the seed and on variation of the essential amino acid pattern from a standard. For the species assayed, differences between the families Cruciferae, Leguminosae, Compositae, and Gramineae are evident. Listed are 20 species with essential amino acid patterns closest to hen's egg. Members of the Compositae and Umbelliferae generally have the "best" pattern. A method is also presented for calculating optimum combinations of two or more protein sources.

The amino acid compositions of crude seed protein from 379 plant species are given in a series of papers by Miller *et al.* (1962a, 1962b) and VanEtten *et al.* (1961, 1963a, 1963b, 1967). The families Leguminosae, Cruciferae, and Compositae are each represented by more than 34 species. Amino acid data for the Gramineae in this study are for 54 species examined by Taira (1962a, 1962b, 1963). The FAO/WHO report (1965) on protein requirements adopted the essential amino acid pattern of whole hen's egg as the reference pattern. This paper contains comparisons of the essential amino acids in plant seeds to those in the reference protein. The criteria for making the comparisons are expressed in terms of amount and pattern of essential amino acids, and reduce amino acid data to two numbers. Measures of the variability in the criteria are given. A method is also developed for determining the proportions of two or more proteins for a balanced amino acid pattern.

PROCEDURE

The FAO/WHO report (1965) includes isoleucine, leucine, lysine, phenylalanine, tyrosine, cystine, methionine, threonine, tryptophan, and valine as nutritionally essential amino acids. The present study is based on eight of these amino acids since data for tryptophan and cystine are generally not available.

Two quantities were calculated to aid in evaluating amino acid data relative to the essential amino acid pattern. The ratio, r_i, of each of the eight amino acids (in grams per 16 grams of nitrogen) to the corresponding FAO/WHO amino acid values for hen's egg (grams per 100 grams of total essential amino acids) was determined for each species. Mean ratio and variance of the ratios were computed as

$$R = \frac{\Sigma r_i}{8} \quad (1)$$

and

$$V(r) = \frac{\Sigma(R - r_i)^2}{7} \quad (2)$$

If the numerator of r_i is expressed in milligrams per gram of nitrogen, then Equations 1 and 2 may be used, but R and $V(r)$ should then be multiplied by 0.016 and $(0.016)^2$, respectively (assuming a conversion factor of 6.25).

For a seed with an essential amino acid pattern the same as that for the FAO/WHO whole hen's egg, R would equal 0.513 and $V(r)$ would equal zero. R is the mean proportion of the essential amino acids in a protein relative to the essential amino acids of the hen's egg standard. A desirable protein would be associated with R approaching or greater than 0.513 (representing 0.513 gram of essential amino acid per gram of protein). A small value for $V(r)$ indicates good agreement with the hen's egg pattern. R and $V(r)$ values for the 379 species analyzed at the Northern Laboratory, along with the 54 Gramineae species of Taira (1962a, 1962b, 1963), were calculated.

Amino acids results including cystine and tryptophan were given for 110 foods by Orr and Watt (1957). These data were used to check the effects of omitting either cystine or tryptophan or both from a calculation of $V(r)$ and R. $V(r)$ and R calculated with these two acids was highly correlated (0.90) with $V(r)$ or R when the acids were omitted. A plot of the two values showed a linear relation with points evenly scattered about the theoretical line with slope 1 where the two values are equal. Also, agreement was better than that expected based on variation estimates given later in this paper. Thus although cystine and tryptophan should be included in the definition of R and $V(r)$, their omission very likely does not detract from the usefulness of $V(r)$ or R as computed here in comparing amino acid patterns.

Oser (1959) defines an essential amino acid index (EAAI) as the geometric mean ratio of essential amino acids to amino acids in a reference standard. The quantity R, defined here, is analogous to Oser's EAAI; however, the R value generally will be slightly higher because, in determining EAAI, ratios exceeding one are set equal to one and the mean ratio is geometric. Both ratios are measures of the total quantity of essential amino acids in the protein; introduction of $V(r)$, however, measures agreement of patterns. $V(r)$ will be recognized as computationally the same as the variance (square of the standard deviation) of r. The authors believe that R and $V(r)$ provide a better basis than EAAI for assessing protein sources since both amount and agreement with a standard are numerically determined.

The advantages of $V(r)$ are that data are reduced to a

Biometrical Services Staff and Industrial Crops Laboratory, Northern Regional Research Laboratory, Peoria, Ill. 61604

single descriptive value and that the over-all pattern is evaluated. With the protein score procedure 10 pairs of numbers must be examined. The importance of imbalances is indicated in the FAO report but protein score considers only deficits while $V(r)$ also takes into account excesses of essential amino acids. $V(r)$ measures how far apart two proteins are in their essential amino acids. The protein score more nearly measures how far apart two amino acids are. Finally a procedure for obtaining unique optimum preparations of proteins so as to minimize $V(r)$ is given.

A plotting of log $V(r)$ (ordinate) against R (abscissa) provides a means of comparing the amino acid distributions of species plotted. Points appearing in the lower half of the plot and associated with a smaller $V(r)$ show closer agreement to the hen's egg standard. Points appearing in the right half of the figure are associated with larger amounts of essential amino acids. Thus the most promising protein sources would appear in the lower right hand of the plot. Proteins low in amount of essential amino acids but well balanced relative to hen's egg appear in the lower left-hand corner.

RESULTS

Figure 1 gives R and $V(r)$ for the plant families Leguminosae and Compositae, and Figure 2 for the Gramineae and Cruciferae. Clearly, the Gramineae are generally the poorest relative to the FAO/WHO standard, while the Compositae are the "best." The relatively high value for $V(r)$ in the Gramineae is associated with the generally low lysine content. A least squares line has been fitted to data from each of the families in Figures 1 and 2. The lines were included to help differentiate between the groups plotted. The correlation coefficient between R and log $V(r)$ for the Gramineae was 0.53, for the Legumin-

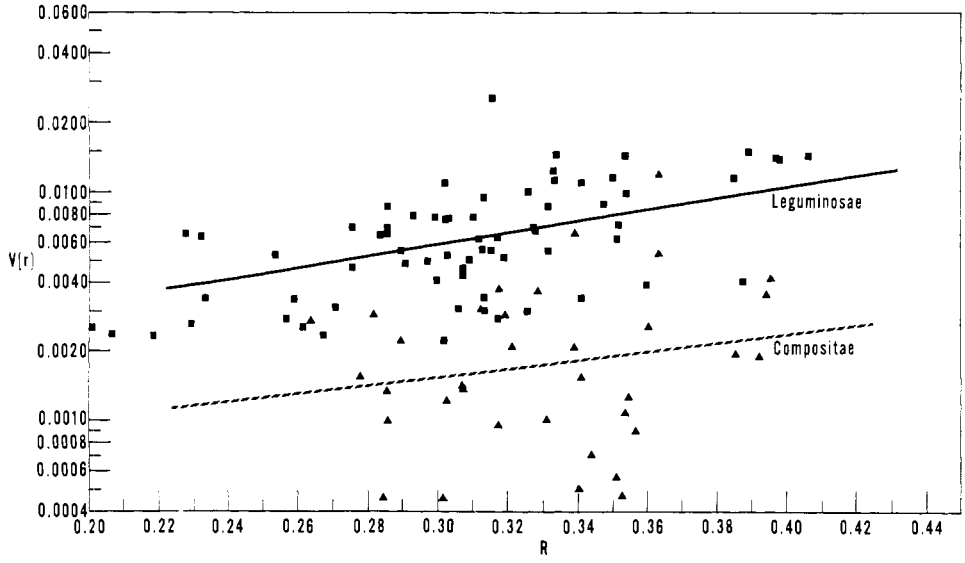

Figure 1. Scatter plot of R, a measure of amount, and $V(r)$, a measure of pattern of amino acids for species of Leguminosae (squares) and Compositae (triangles)

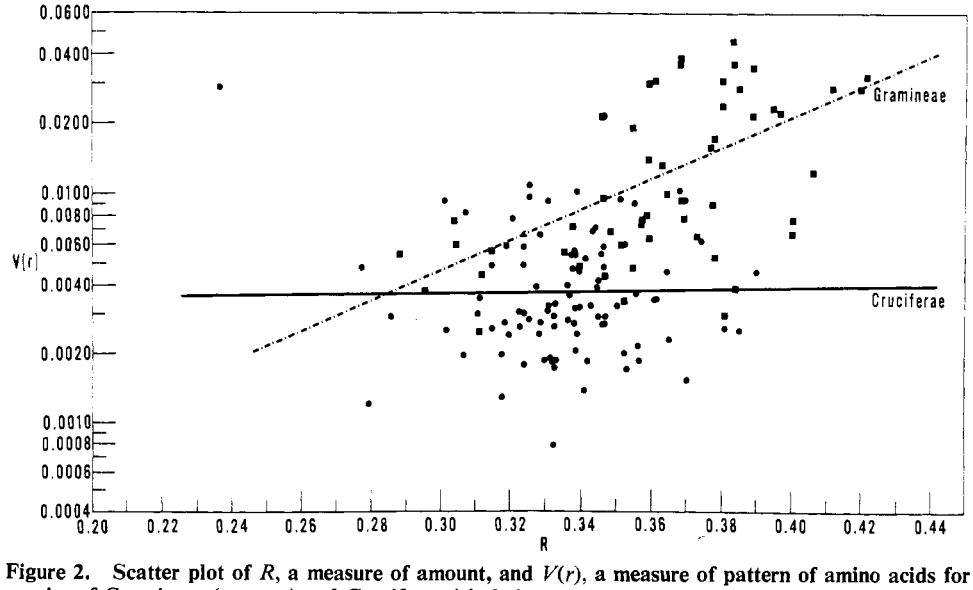

Figure 2. Scatter plot of R, a measure of amount, and $V(r)$, a measure of pattern of amino acids for species of Gramineae (squares) and Cruciferae (circles)

osae 0.48, for the Cruciferae 0.10, and for the Compositae 0.21. The values 0.53 and 0.48 exceed the value required for statistical significance at the 1% probability level. Thus there is some tendency toward a larger $V(r)$ as R increases in the Gramineae and Leguminosae.

Table I lists R and $V(r)$ values for the three other standards in the FAO/WHO report (1965), from data by VanEtten *et al.* (1967), and for common foods based on the report of Orr and Watt (1957). In terms of variation from the standard $V(r)$, plant sources are comparable with animal and human sources, but the plant sources are lower than the animal in terms of R, the proportion of essential amino acids in the protein (or on a total N basis).

A large $V(r)$ may be associated with a deficiency or excess of one or two amino acids. However, proteins with half the essential amino acids high and half low in amount would also yield high values for $V(r)$. Proteins with a good essential amino acid pattern are not observed easily by inspection of the individual amino acid amounts, particularly if the total amount is low. Computation of $V(r)$, however, gives a good basis for evaluating the protein.

Table II shows the means for R and $V(r)$ calculated for nine plant families. Variation between the family means was highly significant relative to variation between species and samples within the same family for both R and $V(r)$. The 95% confidence limits for each mean are also given. Calculations were based on transformation to the logarithms of R and $V(r)$ and then antilogarithms reported in Table II. It did not matter whether or not the logarithm of R is used in the analysis but logarithmic transformation is recommended with $V(r)$. Two family means may be considered significantly different when the respective limits do not overlap each other. Based on the mean $V(r)$, the families Compositae and Umbelliferae have the best amino acid pattern relative to hen's egg.

King (1964) reported the amino acid composition of food crops produced in the Republic of Haiti. Up to 12 samples of the same variety were taken from different geographical areas. These data were used to estimate the precision associated with R and $V(r)$ and then to estimate the magnitudes required for statistically significant differences between two R or two $V(r)$ values. The quantities R and $V(r)$ were determined for each sample. The precision associated with these quantities was estimated from the variation between samples from the same variety. The standard deviation of R was 0.023 based on individual samples from the same variety. The smallest difference between two individual R values that is significant at a 5% probability level is 0.064. For $V(r)$, the standard deviation based on individual samples is given by 30% of $V(r)$ and the least significant difference between two $V(r)$'s is exceeded when the ratio of the two being considered is more than 2.1. The two different ways of expressing the precision for R and $V(r)$ arise from the use of the logarithm of $V(r)$ but not of R in the statistical analysis of the values based on the King (1964) data.

These measures of precision are approximate and include geographical, sampling, and analytical variations. Multiple samples on the same variety or species are needed to determine a standard deviation of R or $V(r)$ for assessing species differences. The data published by King (1964) are useful for this purpose. The values given above seem reasonable when compared with the scatter of points in Figures 1 and 2. Measures of the precision of R and $V(r)$ very likely depend on the particular species involved.

Table I. Criteria for Assessing Essential Amino Acids from Common Sources

	R	$V(r)\cdot 10^3$
From FAO (1965)		
1957 pattern	0.309	1.53
Cow's milk	0.513	7.75
Human milk	0.487	8.85
Hen's egg	0.513	0.00
From VanEtten *et al.* (1967)		
Corn	0.417	32.05
Opaque corn	0.400	7.31
Low-protein rice	0.392	5.83
Soybeans	0.424	6.68
Wheat	0.319	5.24
From Orr and Watt (1957)		
Milk, cow (1)[a]	0.522	7.77
Milk, human (3)	0.468	9.06
Cheeses (7)	0.516	6.30
Cottage cheese (8)	0.538	12.10
Eggs, whole (12)	0.515	0.09
Eggs, whites (13)	0.525	3.39
Eggs, yolks (14)	0.484	3.50
Beef cuts (15)	0.451	11.23
Lamb (16)	0.436	8.93
Pork (17)	0.439	9.24
Chicken (23)	0.440	12.57
Turkey (25)	0.395	39.37
Fish (26)	0.436	14.29
Beans (45)	0.441	15.15
Peanuts (58)	0.327	9.51
Peas (59)	0.428	12.70
Coconut and other palm family (68)	0.340	1.49
Bread (86)	0.327	6.53
Cornmeal (93)	0.446	39.31
Corn flakes (94)	0.371	30.58
Pearl millet (105)	0.373	52.17
Oatmeal (107)	0.356	5.88
Rice (109)	0.401	8.27
White flour (117)	0.308	9.81
Macaroni (124)	0.335	4.63
Corn (146)	0.413	13.74
Cabbage (153)	0.243	4.57
Turnip greens (162)	0.380	4.22
Potatoes (166)	0.331	6.98
Sweet potato (167)	0.430	8.93
Carrots (175)	0.288	5.73

[a] Reference number in Table I of Orr and Watt (1957).

Table II. Mean Values and 95% Confidence Limits of R and $V(r)$ for Nine Plant Families

Family	Number of Samples	R	$V(r)\cdot 10^3$
Gramineae	54	0.362 (0.351–0.373)	11.73 (9.75–14.1)
Leguminosae	70	0.302 (0.294–0.310)	6.14 (5.21–7.22)
Labiatae	12	0.347 (0.325–0.370)	4.51 (3.05–6.67)
Cruciferae	92	0.334 (0.326–0.342)	3.71 (3.21–4.27)
Malvaceae	9	0.303 (0.281–0.326)	2.98 (1.89–4.69)
Euphorbiaceae	10	0.334 (0.311–0.358)	2.42 (1.58–3.72)
Boraginaceae	8	0.341 (0.315–0.369)	2.04 (1.26–3.30)
Compositae	35	0.328 (0.316–0.341)	1.72 (1.36–2.16)
Umbelliferae	8	0.321 (0.297–0.347)	1.66 (1.03–2.68)

Table III. Plant Species[a] Associated with a $V(r)$ of Less than 0.0010

	R	$V(r)\cdot 10^3$
Liliaceae		
Dasylirion wheeleri	0.276	0.67
Polygonaceae		
Eriogonum alatum[b]	0.294	0.80
Polygonium pennsylvanicum	0.320	1.00
Chenopodiaceae		
Chenopodium album	0.293	0.78
Cruciferae		
Camelina sativa	0.332	0.85
Euphorbiaceae		
Euphorbia heterophylla	0.304	0.90
Euphorbia marginata	0.291	0.68
Fouquieriaceae		
Fouquieria splendens	0.289	0.79
Araliaceae		
Aralia spinosa	0.372	0.85
Styracaceae		
Styrax obassia	0.389	0.39
Boraginaceae		
Cynoglossum nebrodense	0.385	0.73
Lappula redowskii	0.343	0.64
Scrophulariaceae		
Nemesia suttonii	0.306	0.96
Nonnea macrosperma	0.304	0.15
Plantaginaceae		
Plantago ovata	0.343	0.56
Compositae		
Artemisia dracunculus	0.342	0.51
Calendula arvensis	0.317	0.96
Calendula officinalis	0.301	0.46
Helichrysum bracteatum	0.286	1.00
Liatris spicata	0.344	0.73
Marshallia caespitosa var. *signata*	0.353	0.47
Osteospermum spinescens	0.284	0.46
Saussurea candicans	0.357	0.91
Zaluzania discoidea	0.351	0.58

[a] Based on data of VanEtten *et al.* (1961, 1963a, 1963b, 1967). [b] Protein 18.7% instead of 63.8% (Van Etten, 1967).

If the grams of amino acid per 16 grams of nitrogen (or 100 grams of protein) are estimated with a maximum error of ±10%, the expected value for $V(r)$ is 0.0010 when R is about 0.30. For R values of 0.20 and 0.40, the $V(r)$ would be about 0.0005 and 0.0018, respectively. These results are obtained assuming four values of r_i are 10% too high and the other four are 10% too low.

Table III lists plant species associated with a $V(r)$ of less than 0.0010. These species are considered potential protein sources with a balanced essential amino acid pattern. Results for tryptophan and cystine would be needed to limit the list further. Botanical and economic considerations are of course extremely important. Comparison of data in Tables I and II shows that seed sources compare favorably with common food sources based on R and $V(r)$ criteria. One of the most highly developed plant families, the Compositae, has close general agreement to the protein pattern of hen's egg.

$V(r)$ can also be used as the basis for determining mixtures of two or more different proteins to yield a product conforming as closely as possible to the standard. A smaller and better value of $V(r)$ than for any of the components is obtained by a combination. For example, $V(r)$ for a mixture composed of proportions a of one protein and b of another would be

$$V(r) = \sum_{i=1}^{8} [a(R_1 - r_{i1}) + b(R_2 - r_{i2})]^2/7 \qquad (3)$$

where $a + b = 1$. First, substitute $a = 1 - b, \Delta_{i1} = R_1 - r_{i1}$ and $\Delta_{i2} = R_2 - r_{i2}$ into $V(r)$ and set the first derivative with respect to b equal to zero. Then, solving for b one obtains

$$b = \frac{\Sigma\Delta^2_{i1} - \Sigma\Delta_{i1}\Delta_{i2}}{\Sigma\Delta^2_{i1} + \Sigma\Delta^2_{i2} - 2\Sigma\Delta_{i1}\Delta_{i2}} \qquad (4)$$

The formula for b will be recognized as consisting of corrected sums of squares and products for variables r_{i1} and r_{i2}. For an example, data for oatmeal and potatoes from Orr and Watt (1957) (Table I) were used. The necessary terms were $\Sigma\Delta^2_{i1} = 0.04117$ (oatmeal), $\Sigma\Delta^2_{i2} = 0.04883$ (potatoes), and $\Sigma\Delta_{i1}\Delta_{i2} = 0.00259$ (cross products). The solution for b is 0.455 and therefore $a = 0.545$. If a protein consisting of 0.455 part potato and 0.545 part oatmeal is prepared, the minimum value for the $V(r)\cdot 10^3$ calculated from the eight amino acids of the combined product is 3.37, a result about one half the $V(r)\cdot 10^3$ of oatmeal (5.88) or potatoes (6.98) taken alone.

A similar calculation for beef [$V(r)\cdot 10^3 = 11.23$] and white flour [$V(r)\cdot 10^3 = 9.81$] yields a combined $V(r)\cdot 10^3$ of 2.33 based on 0.52 part flour and 0.48 part beef. Combining protein sources in these two examples reduced $V(r)$ considerably. The procedure thus provides a quantitative means of evaluating amino acid sources, in combination or singly; or, for obtaining optimum combinations of different amino acid sources in terms of minimum $V(r)$.

LITERATURE CITED

FAO/WHO Expert Group, "Protein Requirements," WHO Technical Report, Ser. 301, Rome, Italy, 1965.
King, K. W., *Econ. Bot.* **18,** 311 (1964).
Miller, R. W., VanEtten, C. H., McGrew, C., Wolff, I. A., Jones, Q., J. AGR. FOOD CHEM. **10,** 426 (1962a).
Miller, R. W., VanEtten, C. H., Wolff, I. A., *J. Am. Oil Chemists' Soc.* **39,** 115 (1962b).
Orr, M. L., Watt, B. K., USDA Home Economics Research Report 4, Washington, D. C., 1957.
Oser, B. L., "Protein and Amino Acid Nutrition," A. A. Albanese, Ed., 2d ed., p. 281, Academic Press, New York, 1959.
Taira, H., *Bot. Mag. (Tokyo)* **75,** 80 (1962a).
Taira, H., *Bot. Mag. (Tokyo)* **75,** 242 (1962b).
Taira, H., *Bot. Mag. (Tokyo)* **76,** 340 (1963).
VanEtten, C. H., Kwolek, W. F., Peters, J. E., Barclay, A. S., J. AGR. FOOD CHEM. **15,** 1077 (1967).
VanEtten, C. H., Miller, R. W., Wolff, I. A., Jones, Q., J. AGR. FOOD CHEM. **9,** 79 (1961).
VanEtten, C. H., Miller, R. W., Wolff, I. A., Jones, Q., J. AGR. FOOD CHEM. **11,** 399 (1963a).
VanEtten, C. H., Rackis, J. J., Miller, R. W., Smith, A. K., J. AGR. FOOD CHEM. **11,** 137 (1963b).

Received for review August 13, 1967. Accepted January 17, 1968 The Northern Regional Research Laboratory is headquarters for the Northern Utilization Research and Development Division, Agricultural Research Service, U.S. Department of Agriculture.