- Gold, R. J. M. & Scriver, C. R. (1971) *Proc. Conf. Clinical Delineation of Birth Defects 3rd*, vol. 7, pp. 91-95
- Gold, R. J. M. & Scriver, C. R. (1972) *Am. J. Hum. Gen.* **24**, 549-561
- Gold, R. J. M., Tenenhouse, H. S. T. & Adler, L. S. (1976) *Biochem. J.* **159**, 157-160
- Greenland, R. D., Roth, W., Gieske, T. & Michaelson, I. A. (1974) *Anal. Biochem.* **62**, 305-309
- Lebedinsky, N. G. & Dauvart, A. (1927) *Biol. Zentralbl.* **47**, 748-752
- McKusick, V. A. (1971) *Mendelian Inheritance in Man, Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes*, 3rd edn., Johns Hopkins Press, Baltimore and London
- Scriver, C. R., Solomons, C. C., Davies, E., Williams, M. & Bolton, J. (1965) *J. Pediat.* **67**, 946-947
- Steinberg, A. G. & Fraser, F. C. (1946) *Can. J. Res. Sect. D* **24**, 1-9
- Tenenhouse, H. S., Gold, R. J. M., Kachra, Z. & Fraser, F. C. (1974) *Nature (London)* **251**, 431-432

APPENDIX

A Method for Calculating the Relative Protein Contents of the Major Keratin Components from their Amino Acid Composition

By REYNOLD J. M. GOLD,* HARRIET S. TENENHOUSE* and LEONDA S. ADLER†

*The deBelle Laboratory for Biochemical Genetics, McGill University, Montreal Children's Hospital Research Institute, 2300 Tupper Street, Montreal, Quebec H3H 1P3, Canada, and

†Concordia University, Montreal, Quebec, Canada

(Received 8 April 1976)

In the main paper (Tenenhouse & Gold, 1976), we described the separation of keratin fibres into major components whose amino acid composition were determined. We also wished to know the relative amounts of protein in each component. We have developed a new method for calculating these amounts solely from the amino acid composition without the necessity for any further experimental data.

Here, we explain the theoretical basis of the method, discuss its reliability under experimental conditions, compare it with other available methods and consider applications of the method to further problems.

Theoretical Basis

If a mixture of proteins containing t mol% of a given amino acid is separated into an insoluble and a soluble fraction containing respectively i and s mol% of the same amino acid, and if the insoluble fraction contains a proportion, p , of all amino acid residues present in the original mixture, then

$$t = pi + (1-p)s \quad (1)$$

or

$$p = \frac{t-s}{i-s} + 0 \quad (2)$$

This equation will be true for any amino acid chosen. Therefore, if $t-s$ is represented on the y axis

and $i-s$ on the x axis, the point representing each amino acid should lie on a straight line, passing through the origin, and whose slope is p . The slope of the line thus gives the proportion of amino acid residues in the insoluble fraction.

Experimental error will scatter the points from this theoretical line, but β , an estimate of p , can be arrived at by regression analysis, according to the following expression

$$\beta = \frac{\sum_{k=1}^n x_k y_k}{\sum_{k=1}^n (x_k)^2} \quad (3)$$

where x_k and y_k are given by $i-s$ and $t-s$ respectively for the k th amino acid, and n amino acids are considered. β is then the least-squares estimate of p , with the constraint that the line must pass through the origin.

Further, s_β the standard deviation of β , is given by

$$s_\beta = \frac{\sum_{k=1}^n (y_k - x_k \beta)^2}{\sum_{k=1}^n (x_k)^2} \cdot \frac{1}{(n-1)} \quad (4)$$

Finally, if a protein mixture is separated into several fractions, the problem can be solved for all fractions simultaneously by multiple linear regression.

Table 1. *Proportion of protein present in dichotomous fractions derived from normal and Naked-mouse hair*

The pairs of values were obtained from the amino acid compositions of the original protein mixture and of the two fractions into which it is separated by means of regression analysis, as explained above. The sum of each pair of values is, of course, equal to 1. The results are expressed as the mol proportion of all amino acid residues that have entered a given fraction. The s.d. for each pair of results is shown in parentheses.

Fraction of total hair protein (a) in	Normal	Naked
Residue (b)	0.214 (± 0.026)	0.303 (± 0.024)
Extract (c)	0.786 (± 0.026)	0.697 (± 0.024)
Fraction of extract (c) in		
Matrix (d)	0.293 (± 0.016)	0.356 (± 0.019)
Total fibril (e)	0.707 (± 0.016)	0.644 (± 0.019)
Fraction of total fibril (e) in		
Insoluble fibril (f)	0.119 (± 0.020)	0.167 (± 0.050)
Soluble fibril (g)	0.881 (± 0.020)	0.833 (± 0.050)

Calculation of the relative amount of protein in the major components of mouse hair

The relative amounts of protein in the hair fractions, prepared as described in Scheme 1 of the main paper (Tenenhouse & Gold, 1976) were calculated from their amino acid compositions.

The regression method was first used sequentially. Thus we calculated the relative amounts of (b) and (c) generated by (a) and of (d) and (e) generated by (c) etc.

Table 1 shows the regression coefficients and their standard deviations.

Finally, each component could then be expressed as the fraction of the total hair protein which it comprised (Table 2).

As a check, the composition of normal hair was calculated by using multiple linear regression to determine the least-squares fit for the equation:

$$a = (bp_b + dp_d + fp_f + gp_g)$$

where p_b , p_d , p_f and p_g represent the proportion of amino acid residues in components (b), (d), (f) and (g) and a , b , d , f and g represent the mol% of a given amino acid in the corresponding fractions.

Table 2 compares the values obtained by sequential and multiple regression; the close agreement between the two sets of values is internal evidence of the validity of the method. However, we applied the method to proteins of known composition mixed in known proportions in order to evaluate the experimental errors.

Table 2. *Component ratios in normal hair calculated by sequential and multiple regression*

The equation used was $a = bp_b + dp_d + fp_f + gp_g$.

Fraction	Amino acid residues entering fraction (%)	
	Sequential regression	Multiple regression
Insoluble fibril	6.6	6.7
Soluble fibril	49.0	49.9
Matrix	23.0	22.7
Residue	21.4	20.1

Evaluation of regression method with mixtures of ribonuclease and chymotrypsinogen

Bovine ribonuclease (5.5 mg; $E_{280}^{1\%} = 7.3$) and bovine chymotrypsinogen (5.3 mg; $E_{280}^{1\%} = 20.4$) were dissolved in 5 and 10 ml respectively of water. The E_{280} values of the resulting solutions were measured. The two solutions were then mixed in four different proportions by volume, as shown in columns 2 and 3 of Table 3. The solutions of ribonuclease and chymotrypsinogen, and the four mixtures were all hydrolysed in triplicate in 6M-HCl for 24 h at 110°C and the amino acid compositions of the hydrolysates determined on a Jeol amino acid analyser.

The relative amounts of ribonuclease and chymotrypsin in each of the four mixtures were calculated (a) from the E_{280} values and (b) from their amino acid composition.

(a) *Calculation from E_{280} .* The molarity of the original protein solutions, expressed as mol of amino acid residues per unit volume is proportional to $A \cdot N/E \cdot M$, where A is the E_{280} of the solution, N is the number of amino acid residues per molecule, excluding tryptophan, which was not considered, E is the 1% extinction coefficient at 280 nm and M is the molecular weight. For ribonuclease: $A = 0.610$; $N = 124$; $E = 7.3$; $M = 13838$. For chymotrypsinogen: $A = 0.930$; $N = 238$; $E = 20.4$; $M = 25655$. Thus

$$\frac{\text{Molarity of ribonuclease solution}}{\text{Molarity of chymotrypsinogen solution}} = \frac{0.000270}{0.000097}$$

The proportion of amino acid residues contributed to the mixture by ribonuclease (p) was then calculated from

$$p = \frac{0.000270 \times V_R}{(0.00270 \times V_R) + (0.000097 \times V_C)}$$

where V_R and V_C are respectively, the volumes of ribonuclease solution and chymotrypsinogen solution that were mixed together.

(b) *Calculation from amino acid composition.* The composition of each mixture was calculated from its amino acid composition and that of the two pure

Table 3. *Composition of ribonuclease/chymotrypsinogen mixtures*

Mixture	Volume of ribonuclease solution (ml) ($E_{280} = 0.610$)	Volume of chymotrypsinogen solution (ml) ($E_{280} = 0.930$)	Amino acid residues contributed by ribonuclease in mixture (%)	
			From E_{280}	From regression analysis
1	0.36	0.23	73.5	74.6 ± 2.5
2	0.24	0.44	49.1	50.4 ± 2.4
3	0.12	0.66	24.3	22.7 ± 1.6
4	0.5	0.79	10.1	7.5 ± 2.6

Table 4. *Amino acid composition (mol%) of ribonuclease, chymotrypsinogen and various mixtures, 1, 2, 3 and 4*

The values shown are the means of three replicate determinations. See Table 3 for preparation of mixtures.

Amino acid	Composition (mol%)					
	Ribonuclease	1	2	3	4	Chymotrypsinogen
Lys	8.13	7.40	6.82	6.08	6.11	5.87
His	3.23	2.58	1.95	1.39	1.07	0.85
Arg	3.20	2.80	2.53	2.08	1.77	1.51
Asp	13.25	11.97	12.50	10.99	10.39	10.53
Thr	7.90	7.94	8.46	8.89	8.51	9.19
Ser	9.70	9.52	8.47	9.10	8.42	9.05
Glu	10.48	9.33	8.77	7.46	7.21	6.96
Pro	3.26	3.54	3.79	3.66	4.13	3.84
Gly	2.79	4.78	6.76	8.76	9.71	10.55
Ala	9.96	9.64	9.97	9.96	9.80	9.62
Cys	6.20	6.26	4.81	4.47	3.78	3.33
Val	7.56	8.46	9.01	9.62	10.58	10.15
Met	3.32	2.83	2.16	1.72	1.45	1.41
Ile	2.44	2.76	3.20	3.70	3.99	4.04
Leu	1.77	3.40	4.87	6.70	7.86	8.15
Tyr	4.28	4.14	3.19	2.53	2.19	1.82
Phe	2.52	2.63	2.74	2.81	3.04	3.13
Total	99.99	99.99	100.00	99.92	100.01	100.00

proteins by using the regression method described above.

Table 4 shows the amino acid composition of ribonuclease, chymotrypsinogen and of various mixtures of the two proteins. Table 3 shows the close agreement between the known values for protein ratios and those obtained by regression analysis, which is good evidence for the reliability of this method. However, as might be expected, the method becomes less reliable as the contribution of one component to the mixture becomes small.

Comparison with other Methods

When applied to the major components derived from normal and Naked-mouse keratin, regression analysis indicated that 9% less protein was extractable from Naked-mouse hair than from normal mouse hair. The *P* value associated with this difference is

0.02. It is unlikely that a difference of this magnitude in the proportion of insoluble residue could have been detected with equal confidence by gravimetric analysis, the only other available procedure for insoluble proteins. Gravimetric analysis is unreliable, since substances other than protein are present in hair, e.g. melanin and water, and the inaccuracies are compounded when the amount of hair available is small.

Regression analysis also indicated that the deficiency in Naked-mouse hair resides in the soluble fibril fraction, which comprises 11.6% less of the total protein in mutant than in normal mouse hair. This finding was qualitatively confirmed by subjecting this fraction prepared from both normal and mutant hair to polyacrylamide-gel electrophoresis and subsequent densitometry of the gels at 280 nm. However, quantitative estimates could not be arrived at by this procedure, since the extinction coefficient of each protein in the mixture is not known.

Since amino acid-composition data are required for other purposes in keratin chemistry, the method can be used with data that are often available in any case and, in addition, no other method gives results of such accuracy whose confidence limits can be estimated.

Other Uses of the Method

The regression method should therefore prove very useful in the field of keratin chemistry, since protein ratios are subject to enormous variation caused by nutritional and genetic factors. For example, kwashiorkor decreases the matrix content of hair and a high-cysteine diet increases it (Gillespie, 1967). Similar phenomena have been observed in sheep (Reis & Schinckel, 1964). There are several mutations that decrease the amount of matrix present in human hair (Pollitt & Stonier, 1971; Gold & Scriver, 1972; Kachra *et al.*, 1974; Gold & Kachra, 1974). By using the regression method reported in the present paper, it should be possible to measure these changes of composition with greater precision than conventional methods have afforded. The method is useful as an initial screening technique for estimating the relative amounts of the

major classes of proteins present in a keratin fibre. Differences discovered as the result of this method may then be followed up by investigation on individual components of the fractions.

Regression analysis also provides a useful theoretical tool. For example, as mentioned in the Discussion section of the main paper, it was used to calculate that an amount of HG protein equal to 12% of the total hair protein would have to be removed from normal hair to change its composition to that of mutant hair.

References

- Gillespie, J. M. (1967) in *Symp. on Fibrous Proteins* (Crewther, W. G., ed.), p. 362, Butterworth (Australia), Sydney
- Kachra, Z., Jackson, C. E., Weiss, L., Baden, H. P., Gold, R. J. M. & Etheridge, J. (1974) *Clin. Res.* **22**, 740A
- Pollitt, R. J. & Stonier, P. D. (1971) *Biochem. J.* **122**, 433-444
- Reis, P. J. & Schinckel, P. G. (1964) *Aust. J. Biol. Sci.* **17**, 532-547
- Tenenhouse, H. S. & Gold, R. J. M. (1976) *Biochem. J.* **159**, 149-157