APPENDIX: AMINO ACID ANALYSIS

Multiple regression, least squares analysis of AAA data (Table 24)

The ProCOR (X_1) column represents the experimentally amino acid composition for 1 mol of the respective protein carrier determined by amino acid analysis (AAA). The PepCOR (X_2) column represents the theoretically determined values for 1 mol of antigenic peptide. The ProPep (Y) column is the experimentally determined (AAA) amino acid composition (nmol) of the peptide-protein conjugate.

The AAA raw data were converted to experimentally determined amino acid residues (X_1) by fitting ΣX_1 to the number of residues based on the amino acid sequence. The experimentally obtained value for the ith of n amino acid residues of the conjugate (Y_i) is considered as being dependent on the value for the ith of n amino acid of the carrier (X_{1i}) and peptide (X_{2i}) in some manner [167]. The amino acid compositions of the conjugates are obtained by least-fit of multiple regressional analysis of Y_i. For a given sample of n values of Y, X_1 and X_2, a general equation for describing the regression of the Ys on the X_1 and X_2 values, is

(1) $$\ddot{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$$

This expresses the dependence of the predicted value for the ith of n individual conjugate amino acids ($\ddot{Y}_i$) on the ith value of n individual carrier (X_{1i}) and peptide (X_{2i}) amino acids. b_1 is the amount of carrier and b_2 the amount of peptide found in the conjugate sample (both in nmol).

b_1 and b_2 are used to express the molar ratio of peptide to carrier in the conjugate

(2) $$\frac{b_2}{b_1} = \text{"}coupling\ ratio\text{"}$$

Equation (1) is solved for the a, b_1 and b_2 values by the method of least squares shown below as described in Chapter 13 of Snedecor and Cochran (1968)[167]. The values for a, b_1 and b_2 are chosen so as to minimize $\Sigma(Y_i\text{-}\ddot{Y}_i)^2$, the sum of squares of the differences between n experimental and predicted Y values. Thus we can express the difference $Y_i\text{-}\ddot{Y}_i$ as d_i which then yields

(3) $$\sum \left(Y_i - \ddot{Y}_i\right)^2 = \sum d^2 = \sum (Y - a - b_1 X_2 - b_2 X_2)^2$$

for n experimental values of Y.

In the calculation of the predicted values ($\ddot{Y}$) for the individual amino acids in the conjugate which minimize the quantity $\Sigma(Y_i\text{-}\ddot{Y}_i)^2$, b_1 and b_2 represent the number of mol of the carrier protein and peptide respectively. b_1 and b_2 can be expressed as functions of the residuals x_{1i} and x_{2i} . These, in turn, are tabulated from either the theoretical or experimental values of the amino acid compositions of the carrier protein (X_{1i}) and peptide (X_{2i}) separately. a can be calculated from the expression

(4) $$a = \overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2$$

$\overline{Y}$, $\overline{X}_1$ and $\overline{X}_2$ each refer to the mean of *n* experimental values each of *Y*, X_1 and X_2 respectively. Substituting for *a* in equations (2) and (4) and factoring leads to

(5) $$\ddot{Y}_i = \overline{Y} + b_1 x_{1i} + b_2 x_{2i}$$

with = X_{1i} - $\overline{X}_1$ and $x_{2i} = X_{2i}$ - $\overline{X}_2$, the differences between the experimental and mean values for X_{1i} and X_{2i} respectively. The value of the parameter *a* is correlated with the mean of the population of Y_i values for a given pair of X_{1i} and X_{2i} values, when X_{1i} - $\overline{X}_1$ and X_{2i} - $\overline{X}_2$ equal 0 (the intercept of the sample line on the *Y* axis).

The regression coefficients b_1 and b_2 denote the number of mol of the carrier protein and peptide, respectively, in the sample. b_1 and b_2 can be expressed as functions of the residuals x_{1i} and x_{2i} , respectively. These, in turn, can be tabulated from the experimental (or theoretical) values of the amino acid compositions of the carrier protein (X_{1i}) and peptide (X_{2i}) separately. In the calculation of the fitted values ($\ddot{Y}$) for the individual amino acids in the conjugate which minimize the quantity $\Sigma(Y_i-\ddot{Y}_i)^2$, we can identify the following terms.

For a given experimental value for the *i*th residue of peptide carrier conjugate, Y_i , we have

(6) $$\ddot{y} = \ddot{Y}i - \overline{Y}$$

whereby $\ddot{y}_i$ is the deviation of the predicted $\ddot{Y}_i$ value for the *i*th residue from the mean value of *n* samples, $\overline{Y}$, of the population. Since

(7) $$\ddot{Y}_i = \overline{Y} + b_1 x_{1i} + b_2 x_{2i}$$

this deviation can be expressed as

(8) $$\ddot{y}_i = b_1 x_{1i} + b_2 x_{2i}$$

We can also define d_i , the deviation of the experimental Y_i from the predicted $\ddot{Y}_i$ value as:

(9) $$d_i = Y_i - \ddot{Y}_i$$

Finally, the deviation of the experimentally obtained value for the *i*th residue of the conjugate, Y_i, from the mean value of *n* samples, *Y*, can be defined as

(10) $$y_i = Y_i - \overline{Y}$$

It follows that

(11) $$y_i = \ddot{y}_i + d_i$$

and

(12) $$d_i = y_i - b_1 x_{1i} - b_2 x_{2i}$$

The values for $x_{11} + x_{12} + x_{13} + \ldots + x_{1i} + \ldots + x_{1n}$ and $x_{21} + x_{22} + x_{23} + \ldots + x_{2i} + \ldots + x_{2n}$ can be summed up as Σx_1 and Σx_2 , respectively. Similarly, $y_1 + y_2 + y_3 + \ldots + y_i + \ldots + y_n = \Sigma x_1$. Equation (12) can be multiplied sequentially by Σx_1 and Σx_2 to obtain:

(13) $$\sum x_1 \left(y_i - b_1 x_{1i} - b_2 x_{2i}\right) = \sum x_1 d$$

(14) $$\sum x_2 \ \left(y_i - b_1 x_{1i} - b_2 x_{2i}\right) = \sum x_2 \, d$$

These can be compared to the normal equations

(15) $$b_1 \sum x_1^2 + b_2 \sum x_1 x_2 = \sum x_1 \, y$$

(16) $$b_1 \sum x_1 x_2 + b_2 \sum x_2^2 = \sum x_2 \, y$$

used to calculate the values of b_1 and b_2 . From this, we can find that $\Sigma x_1 d = \Sigma x_2 d = 0$.

If $y_i = \ddot{y}_i + d_i$ (equation (11))

then

(17) $$y_i^2 = \left(\ddot{y}_i + d\right)^2$$

and

(18) $$\sum y^2 = \sum \left(\ddot{y} + d\right)^2$$

Expanding, we have

(19) $$\sum y^2 = \sum \ddot{y}^2 + \sum 2\ddot{y}\, d + \sum d^2$$

Equation (13) can be multiplied by b_1, equation (14) by b_2, and added to obtain

(20) $$\sum \left(b_1 x_1 + b_2 x_2\right) d = \ddot{y}_i \, d = 0$$

Thus

(21) $$\sum y^2 = \sum \ddot{y}^2 + \sum d^2$$

This says that the deviation of the *i*th experimental value for a conjugate amino acid, Y_i, from the sample mean can be expressed as (1) $\ddot{y}_i$, the deviation of the fitted values $\ddot{Y}_i$ from the mean Y, and (2) d_i , the deviation of the experimental values Y from the fitted values $\ddot{Y}_i$.

It can be shown that

(22) $$\ddot{y}^2 = b_1 \sum x_1 y + b_2 \sum x_2 y$$

which allows the determination of the minimal value of $\Sigma \ddot{y}^2$, the sum of squares due to regression. From this value and that of Σy^2 we can use equation (22) to calculate the minimal value for Σd^2, the sum of squares of the deviations of the experimental values for the amino acid composition of the conjugate from the fitted values.

Main objective in the treatment of this case of multiple regression is the determination of the parameters b_1 and b_2 (the mole fractions of carrier protein and peptide, respectively), their variance ($V(b_1)$, $V(b_2)$) and their covariance ($Cov(b_1)$, $Cov(b_2)$). This is readily approached by deriving a 2×2 matrix using the terms of the left side of the normal equations (15) and (16) as described in Snedecor and Cochran [167]. Taking the inverse of the matrix yields expressions for the 4 Gauss multipliers c_{11}, c_{12}, c_{21} and c_{22} in terms of the residuals x_1, x_2 and y. These multipliers are then used to calculate the variances and covariances.

b_1 and b_2 can be obtained by substituting the values of the Gauss multipliers into the normal equations (15) and (16).

(23) $$b_1 = c_{11}\sum x_1 y + c_{12}\sum x_2 y$$

(24) $$b_1 = c_{21}\sum x_1 y + c_{22}\sum x_2 y$$

This yields the following

(25) $$b_1 = \frac{\left(\sum x_2^2\right)\left(\sum x_1 y\right) - \left(\sum x_1 x_2\right)\left(\sum x_2 y\right)}{D}$$

(26) $$b_1 = \frac{\left(\sum x_1^2\right)\left(\sum x_2 y\right) - \left(\sum x_1 x_2\right)\left(\sum x_1 y\right)}{D}$$

where $D = \left(\sum x_1^2\right)\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2$

The molar ratio of the peptide to carrier can then be written as

(27) $$\frac{b_2}{b_1} = \frac{\left(\sum x_1^2\right)\left(\sum x_2 y\right) - \left(\sum x_1 x_2\right)\left(\sum x_1 y\right)}{\left(\sum x_2^2\right)\left(\sum x_1 y\right) - \left(\sum x_1 x_2\right)\left(\sum x_2 y\right)}$$

The sum of the products of the residuals x_1, x_2 and y can be tabulated as in the Microsoft Excel worksheets and used to express the molar ratio of peptide to carrier in the conjugate as shown in the sample calculation.

The equations for calculating the variances of the molar ratio of a conjugate take into account the residuals of the experimental (Y_i) and predicted ($\ddot{Y}_i$) values for each of n amino acids in the sample conjugate. This variance, σ^2, is estimated by the expression $\sigma^2 = \Sigma(Y_i - \ddot{Y}_i)^2/(n - k - 1)$, n being the number of amino acids being used in the analysis and k being the number of parameters used in fitting the regression (here, k equals 3, based on the disposable constants b_1, b_2 and a). This yield the expression

(28) $$\sigma^2 = \frac{\sum d^2}{n - k - 1}$$

σ^2 can be summed directly for each of the Y values and the square root taken to determine the standard error. The covariance of the molar ratio b_2/b_1 can be expressed as follows: Letting $r = b_2/b_1$, and having b_1 and b_2 deviate from their means by the quantities Δb_1 and Δb_2, the covariance of the molar ratio, Cov(b_2/b_1), can be expressed as

(29) $$\mathrm{Cov}(b_2/b_1) = \left[\frac{\delta r}{\delta b_1}\right]^2 V(b_1) + \left[\frac{\delta r}{\delta b_2}\right]^2 V(b_2)$$

(30) $$\mathrm{Cov}(b_2/b_1) = \frac{b_2^2}{b_1^4} V(b_1) + \frac{1}{b_1^2} V(b_2)$$

(31) $$\mathrm{Cov}(b_2/b_1) = \frac{b_2^2}{b_1^2}\left[\frac{\sigma^2 c_{11}}{b_1^2} + \frac{\sigma^2 c_{22}}{b_2^2}\right]$$

where $V(b_1) = \sigma^2 c_{11}$, $V(b_2) = \sigma^2 c_{22}$ and σ^2 is as defined above. The standard error of the estimated molar ratio is then taken as the square root of the covariance Cov(b_2/b_1).

			ProCOR	**PepCOR**	**ProPep**							
			res.	res.	nmol							
	AA		X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	Y^2	x_1
1	**asx**	**N, D**	**54,03**	**4**	**1,026**	2918,946	16	216,1090765	55,43	4,103992	1,05	13,18
2	**glx**	**E, Q**	**77,63**	**4**	**1,378**	6026,329	16	310,5177455	106,97	5,512012	1,90	36,79
3	**gly**	**G**	**16,33**	**1**	**0,334**	266,648	1	16,3293658	5,45	0,333809	0,11	-24,51
4	**ala**	**A**	**45,99**	**2**	**0,802**	2114,973	4	91,97766583	36,87	1,603459	0,64	5,15
5	**val**	**V**	**34,45**	**1**	**0,565**	1186,796	1	34,44991252	19,47	0,565135	0,32	-6,39
6	**ile**	**I**	**13,08**	**0**	**0,201**	171,149	0	0	2,62	0	0,04	-27,76
7	**leu**	**L**	**63,43**	**0**	**0,928**	4023,287	0	0	58,87	0	0,86	22,59
8	**phe**	**F**	**26,91**	**1**	**0,457**	724,028	1	26,90777503	12,29	0,456611	0,21	-13,94
9	**thr**	**T**	**34,46**	**2**	**0,628**	1187,780	4	68,9283729	21,64	1,255883	0,39	-6,38
10	**lys**	**K**	**59,98**	**2**	**1,013**	3597,247	4	119,9541063	60,77	2,0263	1,03	19,13
11	**arg**	**R**	**22,99**	**0**	**0,332**	528,690	0	0	7,64	0	0,11	-17,85
Σ			**449,28**	**17**	**7,663**	22745,873	47	885,17	388,02	15,86	6,67	0,00
mean			**40,84**	**1,55**	**0,697**							

	x_1^2	x_2	x_2^2	y	y^2	x_1x_2	x_1y	x_2y	$\ddot{Y}$	$\ddot{y}$	$\ddot{y}^2$	$Y-\ddot{Y}$	$(Y-\ddot{Y})^2$
1	173,81	2,45	6,02	0,329	0,1085	32,360	4,342	0,8084	1,036	0,340	0,115	-0,01019	**0,000104**
2	1353,20	2,45	6,02	0,681	0,4642	90,293	25,064	1,6724	1,377	0,680	0,462	0,00132	**0,000002**
3	600,94	-0,55	0,30	-0,363	0,1317	13,371	8,895	0,1979	0,310	-0,387	0,150	0,02400	**0,000576**
4	26,47	0,45	0,21	0,105	0,0110	2,339	0,541	0,0478	0,799	0,102	0,010	0,00320	**0,000010**
5	40,88	-0,55	0,30	-0,132	0,0173	3,487	0,841	0,0717	0,571	-0,125	0,016	-0,00609	**0,000037**
6	770,68	-1,55	2,39	-0,496	0,2461	42,904	13,772	0,7667	0,202	-0,495	0,245	-0,00154	**0,000002**
7	510,12	-1,55	2,39	0,231	0,0535	-34,905	5,226	-0,3576	0,928	0,232	0,054	-0,00041	**0,000000**
8	194,21	-0,55	0,30	-0,240	0,0576	7,601	3,345	0,1309	0,462	-0,234	0,055	-0,00580	**0,000034**
9	40,70	0,45	0,21	-0,069	0,0047	-2,900	0,438	-0,0312	0,632	-0,064	0,004	-0,00433	**0,000019**
10	366,09	0,45	0,21	0,317	0,1002	8,697	6,056	0,1439	1,000	0,304	0,092	0,01282	**0,000164**
11	318,63	-1,55	2,39	-0,365	0,1329	27,587	6,507	0,5634	0,345	-0,352	0,124	-0,01297	**0,000168**
Σ	4395,74	0,00	20,73	0,000	1,3277	190,834	75,026	4,0142	7,663	0,000	1,327	0,00000	**0,001117**

(Table continued on following page)

D		54694,05657	
b_1/D		0,014427	
b_2/D		0,060844	
a		0,013385	
$\Sigma\ddot{y}^2$		1,326610	
$\Sigma(Y\text{-}\ddot{Y})^2$		0,001117	
$\Sigma(Y\text{-}\ddot{Y})^2/8$		0,000140	
$(\Sigma(Y\text{-}\ddot{Y})^2/8)^{1/2}$		0,011814	
Variance	b_1	5,29E-08	
Variance	b_2	1,12E-05	
$\boldsymbol{b_2 / b_1}$		**4,22**	**coupling ratio**
n		11	number of AA in analysis
k		3	number of β parameters used for fitting (a, b_1, b_2)
σ^2		**0,00016**	**standard error of least square fitting** ($\Sigma(Y\text{-}\ddot{Y})^2$)
Cov ($\boldsymbol{b_2 / b_1}$)		0,07	
$\mathbf{(Cov\ (b_2 / b_1))^{1/2}}$		**0,26**	**covariance, standard error of estimated molar ratio**

Table 24: (starts on preceding page): Sample calculation of molar ratio of peptide IgA1-PC20 to BSA using multiple regression, least squares analysis [146]. Peptide IgA1-PA50 was synthesized and coupled to BSA via a C-terminal cysteine as described in MATERIALS AND METHODS. Only those residues which are recovered in quantitative manner are considered in this analysis. The experimentally obtained values for the BSA-peptide conjugate (ProPep) are expressed in nmol. The derivations of the terms D, b_1, b_2, etc. are described in the text. Index to key variables:

ProCOR	experimentally determined amino acid composition for one mole of carrier
PepCOR	theoretical amino acid composition for 1 mol of peptide
ProPep	experimental values (nmol) for amino acid composition of peptide-carrier complex
$\ddot{Y}$	predicted values for the conjugate amino acid composition
Y - $\ddot{Y}$	deviation of the predicted amino acid values for the conjugate from the experimentally obtained values
$\ddot{y} = \ddot{Y}_i - \overline{Y}$	deviation of the predicted ($\ddot{Y}_i$) value for the ith residue from the mean value of n samples ($\overline{Y}$) of the population
b_1	nmol carrier found in conjugate sample
b_2	nmol peptide found in conjugate sample
b_2/b_1	predicted molar ratio of peptide to carrier in conjugate
$\sigma^2 = \frac{\Sigma d^2}{n-k-1}$	variance
$\mathrm{Cov}(b_2/b_1)^{1/2}$	standard error of the estimated molar ratio