"UNSCRAMBLING" MULTIVARIATE DATA FROM MIXTURES:

I: FAT, WATER AND PROTEIN DETERMINATION IN MEAT BY NEAR-INFRARED REFLECTANCE SPECTROSCOPY.

II: SOY PROTEIN AND COLLAGEN DETERMINATION IN MEAT PRODUCTS FROM AMINO ACID DATA.

MARTENS, H.[1], HILDRUM, K.I.[2], BAKKER, A.[2], JENSEN, S.Å.[3], LEA, P.[1], ESKELAND, B.[1], VOLD, E.†[1] AND RUSSWURM; H.[1]

[1]Norwegian Food Research Institute, N-1432 Ås-NLH, Norway,
[2]Stabburet A/S, N-1600 Fredrikstad, Norway,
[3]Carlsberg Research Center, Dept. of Biotechnology, Copenhagen, Denmark.

I. INTRODUCTION

"You cannot unscramble an egg" it is said. But you can, if you have a computer! Two examples will illustrate the potential of multivariate statistical analysis in meat science. The first one shows how a single, rapid light-reflectance measurement may yield the fat, water and protein percentages in meat. The second one shows how various soy proteins (and other protein extenders) may be quantified in meat products from amino acid analysis.

The rapid development of various multi-variate quantitative instruments in analytical chemistry hold a great promise for "unscrambling" mixtures, i.e. for quantitative analysis of individual components in systems such as meat products. Measurements may be performed on more or less intact samples, - therefore the analyses become simple and fast, and the risk of preparation artifacts are reduced.

However, many modern analytical instruments may create a vast number of data: for lack of adequate information handling methods this creates data overflow in the mind of the many researchers! It appears that the numerical tools of many chemists today are the same as they were 30 years ago: means and standard deviations. With a computer programmed for multivariate analysis, relevant information may be compressed and displayed for optimal interpretation, while much of the noise and repetitive redundance is eliminated.

Multivariate "unscrambling" allows the use of non-specific measurements for specific quantitative analyses. A simple bi-variate example, well known in biochemistry, is the determination of protein concentration by UV-absorbition, where the disturbance of nucleic acids at 280 nm is eliminated by reading the absorbance at both 260 and 280 nm and taking a difference. In general, a mixture of N different components may be "unscrambled" to yield the concentrations of each of the components, from N different measurements on the same mixture sample. The N measurements may be completely non-specific, i.e. all N components may affect the reading for all N measurement methods, as long as the statistical requirement of linear independence is fulfilled: The N-dimensional measurement "spectrum" corresponding to 100% purity of a component must be "unique", i.e. at the precision level of the instruments it must be sufficiently different from the "spectra" of the other (N-1)components, and also different from any linear combination of them. The greater the "uniqueness" of a component's "spectrum" is, the greater will the precision of the obtained concentration results be.

Since all measurements contain random errors, it is advantageous to have <u>more</u> measurements than unknown components in the system, and to balance the errors in the different measurements against each other by e.g. a weighted least squares technique. This increases the precision of the obtained concentration results, provided the additional measurements yield relevant signals, and not only irrelevant noise. Computations then require a computer, in contrast to the biochemists' manual "two equations - two unknowns" computations.

The "unscrambling" techniques relies on an initial "calibration" of the statistical model later to be used with unknown samples. This calibration depends on the model and may either be based on some standard multivariate statistical method to convert the measured "spectrum" into the concentration of the components of interest ("separate, empirical" calibration), or it may be based on a detailed understanding of how the signals are generated in the samples and recorded in the instrument, ("simultaneous analytical" calibration).

Both approaches are illustrated in the present communication. A discussion of various calibration techniques is given elsewhere[1,2]. The first example concerns near-infrared (NIR) reflectance measurement of fat, water and protein in raw meat. Here each component (fat, water and protein) is calibrated for <u>separately</u>. NIR has hitherto found minimal use in meat science[3], while being well accepted in cereal and dairy science. Norris[4] has recently tested the method for fat in fried meat; Technicon Corp.[5] has tested it for cooked meat products. The present report concerns the analysis of raw meat samples.

The second example, determination of various protein types in a meat product from its amino acid (a.a.) composition, involves <u>simultaneous</u> quantification of all major protein components. Analyses of protein mixtures from a.a. data is well established[1,6 8] although not in meat science.

2. EXAMPLES

I. <u>Fat, water and protein determination in meat by near-infrared reflectance spectroscopy</u>

Components such as fat, water, protein, starch etc. exhibit overlapping, but somewhat different light absorbance spectra in the 1400-2600 nm wavelength region (Fig. 1). These Near-IR (NIR) absorbance spectra are obtained as "apparent absorbance" from log (1/reflectance) at various wavelengths. The apparent absorbance may be measured for solid samples, powders, moist samples as well as liquids. Several different optical and computational systems are available. A Technicon Infralyzer 400 was used presently. The instrument was equipped with 19 standard fixed wavelength bandpass filters.

38 commercial bovine and porcine meat cuts were used[3] as samples. They were homogenized, and fat, water and protein content were determined by standardized analyses (fat by Foslet, water by drying and protein by Kjeldahl). Fat varied from 1.6 to 84.5 per cent of wet weight, water percentage varied from 8.2 to 75.0, and protein percentage from 2.4 to 23.4.

The samples were frozen until NIR-measurements could be obtained. After thawing and tempering to 21±1°C, the NIR reflectances of the samples at 19 wavelengths were read directly from the samples in a standard measuring cup. The NIR spectra were recorded on a Hewlett-Packard 9825A calculator and the concentration of each component was related to the NIR spectrum in the 38 samples by upwards and downwards stepwise multiple linear re-

gression. Table 1 shows the regression coefficients thus obtained.
Fig. 2 shows the relationship between the fat content determined "conventionally" and by the NIR "unscrambling" technique for the 38 samples analyzed. Standard errors of the NIR estimate was 1.1 percent of wet weight; the correlation coefficient was r=0.999. For water and protein the corresponding standard errors (and correlation coefficients) were 0.9 percent (r=0.999) and 0.7 percent (r= 0.990), respectively.
The computerized NIR instrument was easy to operate, and once calibrated it required less than one minute to yield fat, water and protein percentages in an unknown homogenized, temperature-equilibrated sample. The calibration obtained was tested briefly on some additional samples with known composition with satisfactory results. However, a high positive correlation (r=0.95) between the water and protein percentages in the 38 samples was observed, apparently because the meat cuts primarily varied in fat/muscle ratio. This implies that future meat samples with very atypical water/protein ratioes would probably be less well characterized by the calibration constant values obtained here. Adding some atypical samples to the set of calibration samples should eliminate this potential problem.
Several aspects of the NIR-reflectance method require further attention: pH- and temperature - induced "chemical shifts" in the NIR spectra, variations in the protein and fat types and their physical states, detector non-linearities at very high apparent absorbances (e.g. at the water peak maximum), the "whiteness" of samples (PSE- vs DFD-meat), the effects of specular ("mirror-")reflectance etc. These possible complications may have to be understood before a simultaneous, analytical NIR calibration may be developed, yielding a "lack-of-fit" measure of each sample. However, once these complications are understood, then they may possibly open some new and exciting applications of NIR in more detailed analysis of the chemical composition and physical state of samples.
This preliminary study indicates that the commercial NIR reflectance "unscramblers" presently available are already suitable for routine analysis of fat, water and protein in meat cuts.

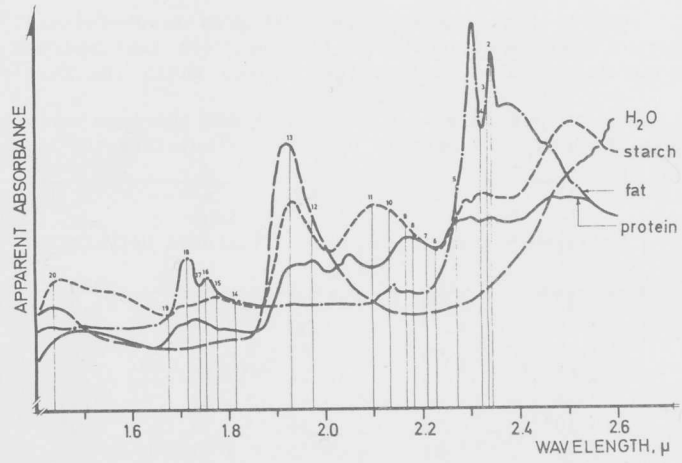

Figure 1: (NIR absorbance) Spectra of pure fat, pure water and pure protein[5]. The vertical lines illustrate the approximate maxima of the 19 bandpass filters in the Infralyzer 400. The spectrum of starch is also given. The spectra are given with slighter different ordinate origins.

II. Robust analysis of soy protein and collagen contents in a meat product from its amino acid composition

At present it is difficult to measure quantitatively the concentration of soy protein in meat products, especially textured soy protein.

Table 1: Results from preliminary calibration of Infralyzer 400 for fat-, water and protein percentages in 38 raw meat cuts.

Filter	Wavelength, nm	Regression coefficients		
		Fat	Water	Protein
2	2384	593.78	-1218.99	
3	2336	- 702.85	1594.04	
4	2310			- 128.11
5	2270			362.55
6	2230			-1183.43
7	2208			1044.72
8	2190	1461.60		
9	2180	-2170.51	666.02	
10	2139		-1217.84	
11	2100	928.88		
12	1982			- 248.89
13	1940	- 84.52	22.53	186.79
14	1818		- 827.33	
15	1778	-1971.36	4000.29	
16	1759	3092.19	-4219.62	206.20
17	1734	- 259.54	360.12	110.99
18	1722	- 727.86	649.70	
19	1680			69.46
20	1445	- 180.25	205.37	
Constant term:		35.35	52.13	10.74

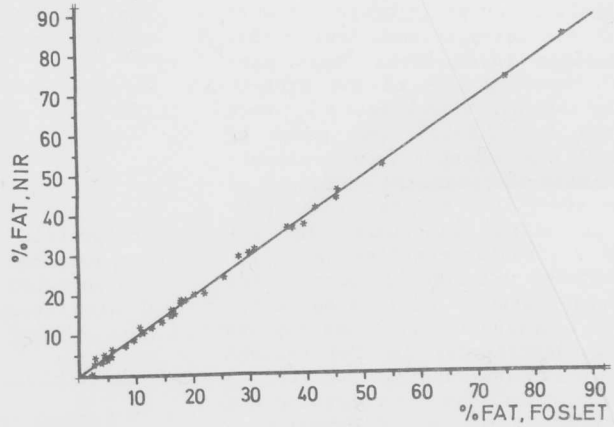

Figure 2: Fat percentages in raw meat cuts, as obtained by NIR, compared to those obtained in the same samples by a conventional method (Foslet). 38 samples, 11 NIR filters used (col. 1, Table 1). St.error of estimate: 1.1 percent of wet weight; correlation coefficient r=0.999.

The present feasibility study investigates whether unscrambling of the amino acid (a.a.) spectrum of a meat product can yield the soy protein concentration to the satisfactory accuracy. The idea is that the a.a. composition of a meat product may be expressed numerically as the sum of the a.a. contributions from each of the major protein sources, because the a.a. spectra of the muscle proteins, as well as of potential added proteins may be accurately described. Thus a "simultaneous, analytical" unscrambling model appears feasible. In the present example the model is tested for the quantitation of textured soy protein and soy protein isolate. Since the connective tissue/muscle ratio may be expected to vary, both collagen and "non-collagen" muscle protein are analyzed together with soy protein. A set of overdetermined linear equations is obtained, which may be solved

by weighted least squares (provided that the spectra of the pure protein sources (soy etc.) are precisely known, that both the pure protein sources and the unknown meat products are analyzed by exactly the same analytical procedure, and that the approximate level of random error variance in the a.a. data of the unknown meat products is known in advance.)

Table 2.
Amino acid spectra (in gram a.a. per 100 gram recovered a.a.) obtained for (1) collagen, (2) non-collagen muscle (bovine m.semimembranosus); (3) mean, (4) textured and (5) untextured soy protein, and (6-8) three known soy-muscle mixtures. Litterature values given for caseinate[9] and blood serum protein[10] lack proline values.
"Per cent recovery per 16 gram N" represent the spectrum sum when the a.a. are given in gram a.a. per 16 gram N. "n" shows the number of spectra used for calculating the average spectra in the table.

COLUMN	1	2	3	4	5	6	7	8	9	10
	Meat		Soya			Mixtures				Blood
SAMPLE: a.a.	Collagen	Non-collagen	Mean	Textured	Untextured	T1	T2	U	Caseinate	serum
ASP	7.17	9.39	11.78	11.89	11.57	10.17	10.67	10.11	7.81	11.17
THR	2.39	4.27	3.33	3.38	3.23	3.88	3.90	3.93	4.47	4.79
SER	3.50	3.86	4.93	4.93	4.93	4.10	4.39	4.11	5.89	5.96
GLU	13.47	17.31	19.24	19.06	19.60	16.88	17.33	16.86	23.05	9.89
PRO	12.84	5.44	6.63	6.46	6.91	5.63	5.17	5.72	-	-
GLY	10.10	4.76	4.14	4.18	4.07	4.74	4.55	4.50	2.09	4.89
ALA	8.93	5.81	4.35	4.41	4.22	5.66	5.44	5.64	3.30	8.29
VAL	3.68	4.94	4.81	4.78	4.88	5.03	4.79	5.28	7.16	8.61
ILE	2.20	4.90	4.62	4.59	4.68	4.77	4.85	4.89	5.38	1.49
LEU	4.71	8.22	7.77	7.69	7.92	8.28	8.36	8.36	9.97	12.87
TYR	1.64	3.79	3.64	3.55	3.82	3.57	3.47	3.62	5.78	3.40
PHE	2.85	4.24	5.22	5.19	5.28	4.46	4.47	4.45	5.35	7.13
LYS	5.05	9.39	6.39	6.44	6.27	8.56	8.42	8.75	8.65	8.51
HIS	1.57	3.92	2.65	2.70	2.56	3.65	3.57	3.66	3.25	6.06
ARG	8.18	6.15	7.72	7.77	7.63	7.01	7.16	6.62	4.08	4.47
CYS(OX)	0.43	0.93	1.44	1.57	1.20	1.26	1.31	1.18	0.60	0.85
MET(OX)	1.32	2.69	1.34	1.34	1.24	2.36	2.16	2.34	3.18	1.60
% recovery per 16 g N	90.55	96.50	100.94	100.45	102.02	98.37	97.51	97.86	97.25	94.00
n	4	2	3	2	1	1	1	1		

In the present preliminary study a.a. spectra were obtained by standard acid hydrolysis and column chromatography from collagen (porcine skin, 2 parallels; porcine tendon, 2 par.), muscle (bovine m.semimembranosus, 7% collagen, 2 par.), a textured soy protein (P.P.50, containing 57% crude protein, 2 par.) and a soy protein isolate powder (Supro 500, 93% crude protein). Compared to collagen and muscle the two types of pure soy protein were very similar (col.4-5). In order to increase the precision of the modelling spectra, means of the 4 collagen spectra, of the 2 muscle spectra, and of all 3 soy protein spectra were used in the calculations, (col. 1-3). In addition three known "meat products" (raw mixtures of soy protein and bovine m.semimembranosus), were analyzed (col. 6-8). Four pairs of parallel a.a. spectra were available to estimate the standard deviation of random analytical errors in the present analytical procedure: The obtained regression equation was: (st. dev. =0.06+0.02x signal). Zero covariance was assumed for simplicity. Table 3 compares the obtained soy protein concentrations to the correct ones, in samples ranging from 100% to 0% soy protein. The concentration was calculated in two units: Normalizing every a.a. spectrum to a sum of 100% (i.e. "g a.a. per 100 g recovered a.a") prior to the calculations yielded col. 1. Ignoring the minor contributions from the missing a.a. (tryptophane and hydroxy-proline), col. 1 gives soy concentrations in "g true protein per 100 g total true protein", a unit which would be insensitive to possible frauds by added non-protein nitrogen. Col. 2 shows the corresponding results after conversion to the conventional "g crude soy protein per 100 g total crude protein". The correct soy protein percentages (col. 3) are given in the same unit. Crude protein is taken as Kjeldahl-Nx6.25. Inspecting the lack-of-fit between the measured and the reconstructed a.a. spectra showed systematically higher residuals than expected for cystein, indicating an abnormally high analytical variance. However, in the present preliminary study no effort was made to correct for this. An average error of less than 3% of total crude protein was thus obtained, implying an error of less that 0.3% of product wet weight, for meat products containing about 10% protein. Thus, a.a. unscrambling gave quite accurate soy protein concentrations in raw "meat products", as well as in pure soy, pure connective tissue and pure muscle .

Table 3: Soy protein concentrations: Comparison between the statistically calculated and the correct percentages.

Col.	1 Calculated[a] % real protein	2 % crude protein	3 Correct % crude protein	4 Difference % crude protein
Soy protein	99.9±3.4(n=3)	99.8±3.3	100.0	0.2±3.3
T1= Textured Soy	35.7	34.8	37.6	2.8
T2= Textured Soy	26.5	25.8	23.2	-2.6
U = Untextured Soy	24.1	23.4	21.3	-2.1
Muscle	-0.1±2.6 (n=2)	-0.1±2.6	0	-0.1±2.6
Collagen	-0.1±1.4 (n=4)	-0.1±1.4	0	-0.1±1.4

a) A model containing collagen, muscle and soy protein (Cols 1,2 and 3, Table 3) was used, with muscle protein=100% - collagen - soy protein. Error std.dev.=0.06+0.02xsignal was used for weighting.

However, submitting the spectra of two autoclaved, soy-free meat products (not shown here) to the same "raw" model yielded rather high residuals and appreciable calculated "amounts of soy". This indicates that the a.a.

spectrum is modified significantly during strong heat treatment.Work is therefore in progress to modify the model by "regression on disjoint factor analysis models"[2], to allow for this heating effect (and other predictable types of variabilities).

At present the unscrambling method was only tested on a few physical meat samples. However, the conclusions so far were supported by purely statistical considerations: Since the basic mixtures model used here is a linear regression model, the expected error standard deviation of an obtained soy protein concentration may be calculated theoretically[2]. At the error level found between replicates in the present data, the model involving collagen, muscle and soya yielded a soy protein standard deviation of 3.0 percent of total crude protein. This corresponds well to the error level found above and was confirmed in separate calculations on a.a. spectra from the literature.

Assuming 10 per cent protein in a meat product and using two standard deviations in order to ensure about 90% probability of fraud detection, this corresponds to a soy protein error limit of ± 0.6% of product wet weight. In other words, if the legal limit of soy protein addition in a meat product is e.g. 3% of product wet weight, then it should theoretically be possible to "arrest" products that contain more than 3.6% soy protein.

Increasing the analytical accuracy, e.g. by taking parallel analyses, would narrow the error limits correspondingly.

The error limits of collagen and non-collagen muscle protein were calculated to be about ± 0.4 and ± 0.5% of product wet weight.

The unscrambling method was likewise tested theoretically for simultaneous analysis of both soy protein, caseinate and blood serum protein in meat products, using litterature values for the caseinate[9] and serum protein[10] (Table 3, col.9-10).

The results were quite promising. Again assuming a level of random errors equal to that found in the present data, the error standard deviations of obtained crude protein concentrations (in percent of total crude protein) were calculated to be 3.5 for soy protein, 3.5 for caseinate, 2.0 for blood plasma protein, 1.5 for collagen and 3.5 for muscle protein. Work is in progress to examine this in more detail, as well as to test alternative statistical methods (e.g. incorporating non-negativity requirements).

A.a. analysis is today an expensive, but with alternative analytical techniques (fully automated instruments, HPLC etc.) complete or partial a.a. analysis may become cheaper.

CONCLUSION

The reflected NIR light spectra of meat could be unscrambled to yield fat, water and protein contents. Thus NIR reflectance appears to be feasible for fast and simple meat characterization.

Measuring the content of textured and untextured soy protein, as well as other protein extenders in meat products seems feasible by multivariate "unscrambling" of amino acid data.

In both examples a non-specific multivariate measurement from an unknown meat sample was used to yield the concentrations of the individual constituents.

LITERATURE

1. Martens, H., Factor Analysis of Chemical Mixtures. Non-negative Factor Solutions for Spectra of Cerea
2. Amino Acids. Anal. Chim. Acta 112 (1979) 423-442.
   Martens, H., On the calibration of a multivariate instrument for quantitative estimation of individua
   components in a mixture. Proceedings, Symp. Applied Statistics, DTH, Lyngby, January 1980; 22p.
   NEUCC, DK-2800 Lyngby, Denmark. (Also available as NINF-Report 49 (1979): Norwegian Food Res. Inst.
3. P.O.Box 50, N-1432 Aas-NLH, Norway).
   Martens, H., Bakker, A., Hildrum, K.I. and Jensen, S.Å., Rapid determination of fat, water and protein ir
4. meat. NINF-Informasjon, 1 (1980) 49-53.
5. Norris, K., USDA, Beltsville, Md., USA. Pers. com. 1979.
6. Technicon AB, (Oslo), pers. com. 1979.
7. Sheldrick, B., Computer analysis of protein mixtures. Biochem. J. 123 (1971) 996.
   Lindquist, B., Östgren, J. and Lindberg, I., A method for the identification and quantitative investiga-
   tion of denatured proteins in mixtures based on computer comparison of amino acid patterns. Z. Lebensm.
8. Unters. - Forsch. 24 (1975) 15-22.
   Gold, R.J.M., Tenenhouse, H.S. and Adler, L.S., A method for calculating the relative protein contents of
9. the major Keratin components from their amino acid composition. Biochem. J. 159 (1976) 157-160.
10. Eggum, B.O., Aminosyrekonsentration og proteinkvalitet. STOUGAARDS Forlag, Kopenhagen 1968.
    Harvey, D., Tables of the amino acids in foods and feedingstuffs, 2nd ed., Commonwealth Agricultural
    Bureau, 1970.