High-throughput measurements for functional genomics of milk

H. Martens[1,2,3,6,8,9], A. Kohler[1,2,5], N.K. Afseth[1,2,3], J.P. Wold[1,2],
M. Hersleth[2], I. Berget[1,2,6], T. Ådnøy[4], M. Skaugen[3], T. Isaksson[1,3],
G. Vegarud[3], A. Criscione[3], B.H. Mevik[1,3], M.B. Frøst[8], Å.T. Randby[4],
E. Prestløkken[4,7], P. Berg[4,6] M. Kent[4,6], S. Lien[4,6] and S.W. Omholt[4,6]

[1]*Centre for Biospectroscopy and Data Modelling*
Campus Ås, Norway
[2]*Norwegian Food Research Institute*
Matforsk 1430 Ås, Norway
Norwegian University of Life Sciences:
[3]*Department of Chemistry, Biology and Food Science (IKBM),*
[4]*Department of Animal and Aquacultural Sciences (IHA),*
[5]*Department of Mathematical Sciences and Technology (IMT),*
[6]*Centre for Integrative Genetics (CIGENE)*
[7]*Felleskjøpet Feed Development*
1432 Ås, Norway
[8]*Copenhagen University, Faculty of Life Sciences*
Bülowsvej 17, 1870 Frederiksberg C, Denmark

ABSTRACT

    Recent developments in analytical technology have simplified a detailed characterization of milk and milk-based samples. A range of powerful new instrumentation types have recently been installed at various institutes at Campus Ås (Norway). At the campus we have recently implemented efficient, multi-channel instrumentation for genomics, transcriptomics, proteomics, biospectroscopy, metabolomics and various quality assessments. The present paper gives an informal outline of various modern analytical tools for characterization of various milk and milk-based samples.

KEY WORDS: biochemometrics, biostatistics, analytical chemistry, functional genomics, transcriptomics, proteomics, metabolomics, biospectroscopy

---

[9] Corresponding author: e-mail: harald.martens@matforsk.no

INTRODUCTION

Functional genomics offers new opportunities in milk science research. Phenotypic properties can be related to genotypic information statistically, in order to quantify expected relationships and to discover unexpected relationships. Detailed genomic studies can then reveal the causal genetic basis for these relationships, and lead to more efficient animal breeding as well as improved farm management, animal feeding and milk processing.

In order to obtain relevant results with sufficient statistical reliability, it is advantageous to be able to choose cost-effective measurement techniques: On one hand, low-cost high-speed screening methods based on e.g., multivariate Fourier Transform Infra Red (FTIR) biospectroscopy, can be applied to millions of milk samples or thousands of individual animals in order to identify particularly interesting samples or individuals. On the other hand, the most interesting samples or animals can be submitted to higher-cost detailed studies, e.g., genome-wide characterization of Singe Nucleotide Polymorphism (SNP) for tens of thousands of genetic markers, or 2D gel proteome-wide electrophoresis. Economically important but more time-consuming quality assessments, such as consumer studies or feeding experiments, can finally be used for a small set of particularly interesting samples or animals. In between, a number of different types of measurements can also be put to use, to reveal systematic patterns of variation.
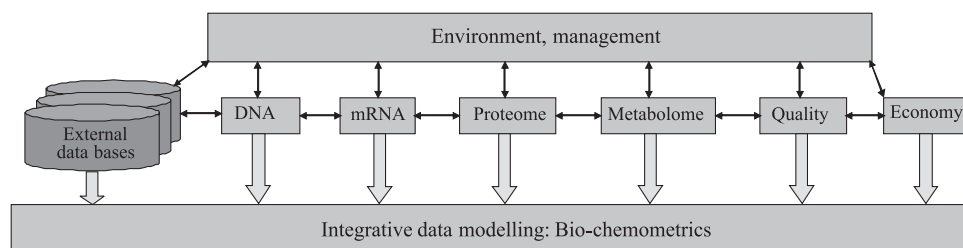

Figure 1. Integrative functional genomics: Relating traditional genetic data (breeding, management, production) to DNA data (genome), mRNA data (transcriptome), protein composition (proteome), metabolic profile (metabolome), as well as animal or product quality data and production economy, using advanced mathematical and statistical tools, such as bio-chemometrics

Figure 1 illustrates the conventional causality in functional genomics. The right-pointing arrows linking the boxes show that the genomic information, representing diversity in the DNA sequence of the animals, is transcribed into mRNA to varying degrees. The transcriptome in turn defines the proteome, which includes the enzymes that produces the variety of metabolites - the so-called metabolome. Together, the proteome and metabolome affect the quality and quantity (animal

productivity and - health components, taste, smell and appearance, etc.), which in turn affect the over-all economy of the agricultural production. But the left-pointing arrows linking the boxes in the figure illustrate that a range of regulatory feedback mechanisms complicates this functionality track.

The figure also outlines how data, obtained at different stages along this causality track, can be related to external data - from environment, farm management or existing data bases in animal breeding/-health/-production, etc.

At Campus Ås, the Norwegian University of Life Sciences (UMB), the Norwegian Food Research Institute (Matforsk) and several other institutions have made a concerted effort over the last couple of years to ensure sufficient measurement capacity at all stages along this functional genomics causality track.

Several milk- and meat-related research projects, using these new facilities, have recently been financed and initiated on the campus. At this stage it is too early to report new research findings. The present paper represents only a brief progress report for some of these measuring techniques. Since the concept is highly cross-disciplinary, the basic principles behind each of the measuring techniques will be outlined.

RESULTS

The bovine genome

The sequencing of the bovine genome is expected to be completed in 2007. This information will of course be important in basic biological studies - for comparison to other species. Moreover, it will help us understand genetic effects in bovine milk. The information in the genome is overwhelming, but is now becoming accessible with a combination of advanced measuring techniques and advanced computer science.

The bovine genome consists of 30 chromosomes, and each animal has two copies of each. Except for the sex chromosome, the two copies are rather similar. A chromosome represents a sequence of the four possible DNA nucleotides: adenine, cytocine, guanidine and thymine (abbreviated by letters A, C, G and T), linked into a DNA strand. "The bovine genome" represents the DNA sequences of the 30 bovine chromosomes.

However, within the bovine species, the DNA sequences of the different families are slightly different. Most of these differences represent single-letter changes at certain locations, called Single Nucleotide Polymorphisms (SNPs, e.g., a C has been changed into a G at position # 1000 along the DNA of chromosome 10). Once upon a time these SNPs may have arisen by random mutation. But unless lethal, these mutations are then passed on from generation to generation,

just like spelling mistakes in a medieval manuscript. These variations of the DNA sequence of a given SNP are called its "alleles". Each animal has two alleles at each SNP location - one inherited from its mother and one from its father.

If we hope to explain why different animals grow differently or produce different milk qualities, we need to measure which SNP-alleles are found in which breeding lines. At the Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences (UMB), we have several up-to-date types of equipment for SNP detection in e.g., semen- or blood samples from individual animals. Methods that monitor diversity in a whole range of genes or gene marker positions may be called "genomic".

Figure 2 (left plot) illustrates such a methodology - the output from the new 22.6K bovine gene marker micro-array chip from the company Affymetrix (Affymetrix, Santa Clara, CA, USA). CIGENE has been involved in the development of this chip. It delivers allele characterization for each animal at more than 22608 different gene marker positions along the DNA sequence, across all chromosomes, in terms of microscopic spots, as shown in the Figure.
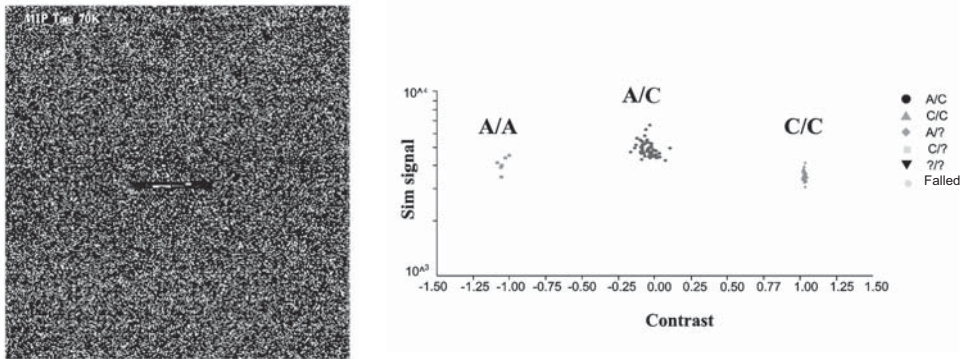

Figure 2. Overview of raw data SNP detection for 23000 gene markers (bovine Affymetrics chip) for one bull (left plot). The right plot provides data from one of these gene markers, for a set of NRF bulls, showing the three combinations of the two alleles for this SNP (C or A)

Figure 2 (right plot) illustrates the output for just one of these gene markers for a certain set of cows: some animals are homozygous A/A for this marker (the animal has inherited allele A from both its mother and its father), while others are homozygous C/C or heterozygous (one copy of A and one copy of C). For each animal, such results are delivered for each of 22608 gene markers.

With this type of massively parallel chip-technology, the cost of per individual SNP is rather low. But each animal requires one chip each, so it calls for a selection of particularly informative animals. At present, about 1500 sires previously or presently involved in the national breeding programme, are the first

to be characterized with this technique. Since each of these bulls have a high number off daughters, phenotypes measured for individual cows - e.g., milk composition - can be averaged into stabile estimates of sire averages. However, we also characterize some of the individual dams that are used in animal feeding studies at the university farm.

In CIGENE we also have other genomics instrument - e.g., lower-cost Sequenom mass-spectrometry based SNP analysis for a few (typically, <50) SNPs at a time, and equipment for re-sequencing of selected DNA sequences of selected animals, for further SNP identification and - verification.

As Figure 1 illustrated, traditional genetics, employing large data bases of breeding information about which bulls have sired which cows etc., can now be supplemented with detailed genomic information.

The DNA sequence information of a given, typical gene consists of its regulatory sequence followed by a coding sequence. The gene regulation sequence consists of DNA patterns that control how this gene is to be utilized at a given point in time, depending e.g., on the presence of gene products from other genes. The coding sequence consists of a series of coding regions (exons) whose information can be transcribed into messenger RNA (mRNA), interspersed with non-coding regions (introns), whose information cannot be translated into mRNA, although they might have other roles - this is an open question at present. When the gene's regulatory sequence is activated, its coding regions are transcribed into one contiguous pre-mRNA sequence, containing both introns and exons. Afterwards the intronic sequences are spliced out, leaving a mature mRNA that diffuses out from the cell kernel. In many cases, but not all, the mRNA transcript can then be subsequently translated into the amino acid sequence of a corresponding protein, because three and three adjacent nucleic acids defines a given amino acid. This protein can, in turn, serve as enzymatic catalyst in metabolic reactions, as structure-building proteins, as gene regulation factors, etc. Due to so-called alternative splicing one DNA sequence can give rise to two or more proteins.

With genotypic and phenotypic data available for sufficiently many individual animals with sufficiently large genetic variation, it is possible to find so-called Quantitative Trait Loci (QTL) - DNA regions in the chromosomes that correlate statistically to a phenotypic trait. Fine mapping combined with functional studies can ultimately reveal the underlying causal mutation(s).

These research processes, in turn, pose several data analytical challenges - both in terms of cross-disciplinary communication, of mathematical data-modelling that is cognitively accessible, and of statistical testing that is sufficiently robust against false positives, etc. One simplifying aspect of this is to combine the DNA data about the many individual SNPs into a lower number of so-called "haplotypes" - groups of adjacent gene markers that are inherited together.

The transcriptome

The complex, dynamic processes of gene regulation, metabolic regulation etc., which control the biological activity (cell differentiation, ageing, response to environment, etc.), need to be studied if we want to bring our understanding of milk production on a causal biological and chemical basis. For this purpose it is important to have cost-effective techniques to monitor the genetic DNA diversity as well as the gene transcription into mRNA, proteins and subsequent metabolites along the functional genomics causality chain (Figure 1). Methods that monitor the expression of a whole range of genes may be called "transcriptomic".

The micro-array technique illustrated in Figure 2 can also be used for quantifying how each of a high number of genes are transcribed into mRNA in individual animals at given points in time. Alternatively, to monitor a lower number of genes, but at a lower price per animal, our Sequenom mass-spectrometric instrument can be used instead (Ding and Cantor, 2003), as illustrated for a casein gene marker in Figure 3. To monitor e.g., how casein production genes are used by a goat, a biopsy is taken from udder and quickly conserved to stop chemical modification. After various automated mRNA robot-based clean-up stages, the sample is split into e.g., 12 aliquotes, and various known concentrations of an artificial "competitor" mRNA are added. By the so-called PCR reaction both the unknown and the known mRNAs are amplified proportionally to concentrations measurable in the maldi-tof mass spectrometer.
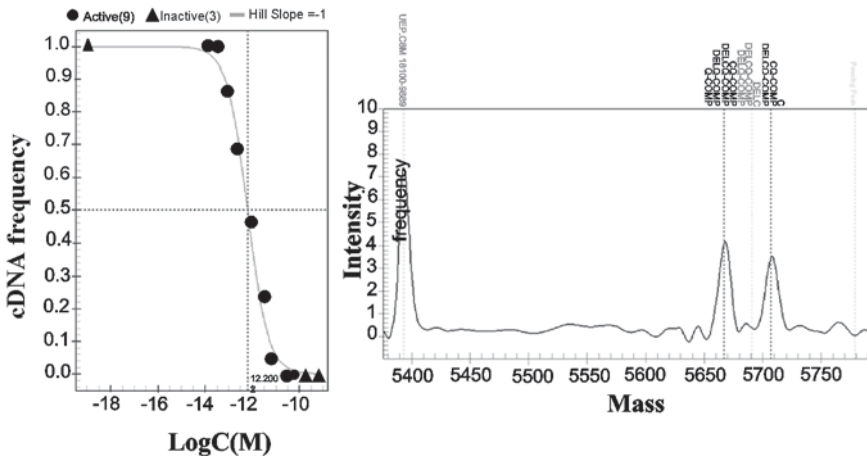

Figure 3. Gene expression determined by MALDI-TOF mass spectrometry: Quantification of two alleles of a given caprine casein gene marker. Left plot: Competitor titration curve to find the concentration of mRNA in this udder biopsy. Right plot: MALDI-TOF MS "spectrogram" raw data for a specific (targeted) gene for one single goat sample at one given competitor concentration. Screenshots from the Sequenom software

In the left plot of Figure 3 the relative concentration of the mRNA for this gene is plotted against known concentrations of an artificial "competitor" mRNA. The white arrows indicate how the unknown mRNA concentration of this gene in this udder biopsy is determined as the competitor concentration that gives the same mass spectrometric peak height. The right plot of Figure 3 illustrates the mass spectrometric raw data at one of these known concentrations.

With this technique we can study how the mRNA expression of various genes, e.g., the synthesis of the four major caseins, vary with their DNA allele structure, with lactation cycle, feeding and stress, etc.

The proteome

The degree to which the produced mRNA of a gene results in actual proteins is also informative. Methods that monitor a whole range of proteins can be called "proteomic".

Figure 4 (left plot) illustrates the "classical" proteomics technique - 1D gel electrophoresis (isoelectric focusing), for goat milk samples. Several different milk samples are applied side by side, and the proteins - caseins and whey proteins - are then separated into bands according to their electric charge. In order to be useful in functional genomics, these proteomic images need to be converted into quantitative data. This involves scanning the gels as digital images and tracing the individual tracks, as illustrated by two samples.

Examples of quantified electrophoresis raw data traces from two animals are shown at the top of Figure 4 (right plot). In order to compare them, two types of normalization are then required: horizontal spatial alignment (middle), to correct
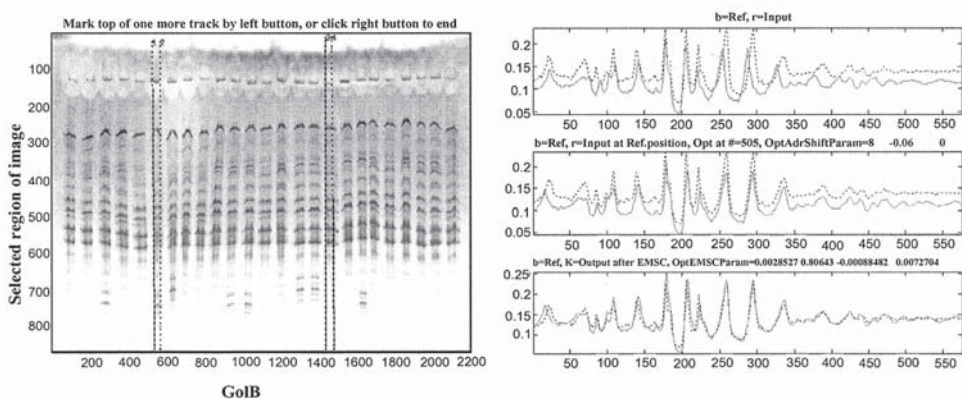

Figure 4. Proteomics of milk proteins. Left plot: 1D electrophoresis image with two samples (tracks) marked for quantification. Right plot: normalization of 1D electrophoresis data from one track (solid) relative to a reference track (dotted): Top: raw data. Middle: data aligned spatially. Bottom: aligned, baseline- and scale-corrected data
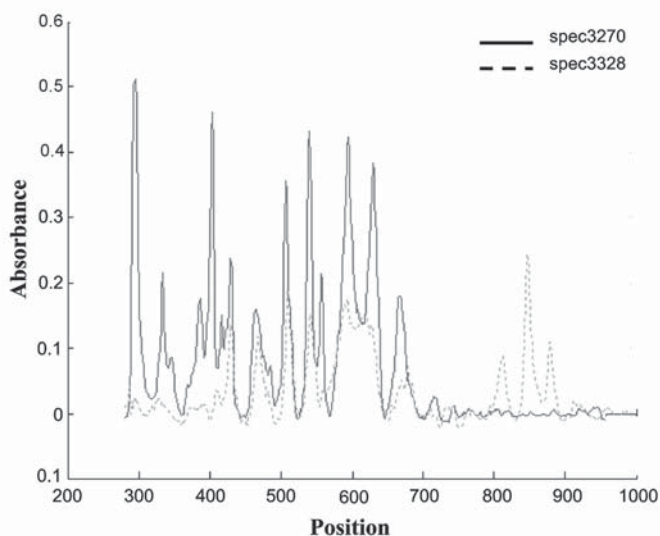
Figure 5. 1D electrophoresis of milk from two goats differing in αS1-casein

for physical variations in the electrophoresis gel structure, and vertical intensity correction (bottom), consisting of a subtraction of baseline background and a division by an overall scaling factor.

Figure 5 shows the results after normalization of the milk proteins from two animals representing two of the most typical Norwegian goat genotypes. Animal no. 3328 was previously found by DNA assessment (Figure 2) to be heterozygous for a certain position for αS1 casein in exon 12 - the animal has one normal (G) allele and one allele called "null" which is a deletion (0) that results in no production of the normal αS1 casein. In contrast, animal no. 3270 is homozygous in the αS1-casein (00) - it totally lacks expression of the normal allele. As expected, the proteomic results in Figure 5 indicate that the latter lacks the normal αS1-casein.

The proteomic measurements can also reveal other details. For instance, several different phosphorylated forms of the αS1 casein are evident. However, 1D electrophoresis is somewhat time-and labour-consuming.

Figure 6 illustrates how similar proteomic "fingerprint" can be obtained with MALDI-TOF analysis, which is faster and less labour-intensive. With this technique, the skim-milk samples are simply diluted, mixed with a light-absorbing matrix compound, spotted and dried on a steel plate and measured. Each plate takes 96 different sample spots. The plate is then bombarded in vacuum with shots of intense laser light on each spot, which vaporizes and ionizes the proteins. An electrical field then drives the ionized proteins towards a mass spectrometric detector; smaller proteins arrive before the larger proteins. Figure 6 shows the
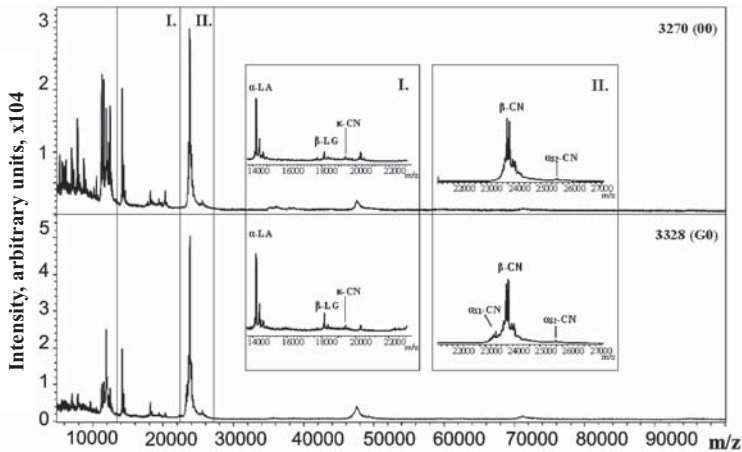
Figure 6. Proteomic MALDI-TOF mass spectrometry of two caprine milk samples

milk proteome of the same two individual goats as in Figure 5, no. 3270 (00) and no. 3328 (G0). For more information, see e.g., Chianese et al. (1993) and Miranda et al. (2004).

The metabolome - chromatography

The enzymatic activities from the proteome catalyse the metabolic processes in the cell. Methods that monitor a whole range of metabolic components or aspects can be called "metabolomics". The classical metabolic measurement principle is chromatography, in which the different chemical components in a sample are separated from each other by passing the sample through a chromatographic column.

Figure 7 shows the fatty acid profile of a bovine milk sample, based on gas chromatography with mass spectrometric detection. The fatty acids can be quantified from the areas or their respective peaks, after horizontal alignment of retention time and vertical subtraction of baseline and scaling for general size factor. A number of different fatty acids can thus be quantified, with different chain lengths, different degrees and patterns of unsaturation, etc.

This type of fatty acid profile data, measured for a large number of related milk samples, can be related to genomic, transcriptomic or proteomic data for the same samples, as well as to other known variables such as feeding strategy, time of year, etc. Multivariate analysis of these data can then reveal their common underlying biological production mechanisms. Chromatographic metabolome analyses can of course also be performed for other compositional aspects of milk, blood, urine, etc., e.g., hormones, oxidation products, anti-oxidants. For instance, GC-MS is
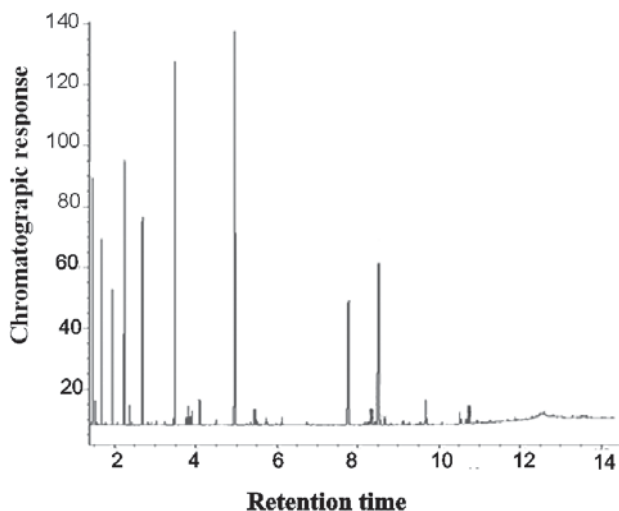
Figure 7. Metabolomics of milk: GC-MS determination of fatty acid profile of a bovine milk sample

a suitable technique for analysing volatile oxidation products of milk and milk products. GC with fluorescence detection is a suitable tool for the qualitative and quantitative analysis of tocopherols.

The biological structure detection - biospectroscopy

While chromatography is generally a mature technology delivering reliable results, faster metabolome alternatives are available, based on spectrophotometry.

Photons are affected by milk samples in various ways, and several principally different spectrophotometric techniques are used for the qualitative and quantitative analysis of milk samples. Most of these techniques are cheap, fast and precise, and yield several types of information about the samples at the same time.

Figure 8 shows the infrared absorption spectra of a set of bovine milk samples, in terms of the absorbance as a function of wavenumber (the wavenumber convention is just the inverse of the wavelength). The data were obtained in one of our modern Fourier-Transform Infra Red spectrometers (Bruker), after simply applying a drop of milk on a zinc selenide surface, drying off most of the water, and recording how much light survives through the sample at different wavelengths in the infrared range, i.e. at wavelengths longer than that visible to the human eye. The figure illustrates that different milk constituents - proteins, lipids, carbohydrates and remaining (bound) water, absorb light at more or less distinct wavenumber regions, because their different chemical bond types ($-C-H_3$, $-O-H$, ($-N-H_2$, etc.) vibrate, and thus absorb light, at different frequencies in this
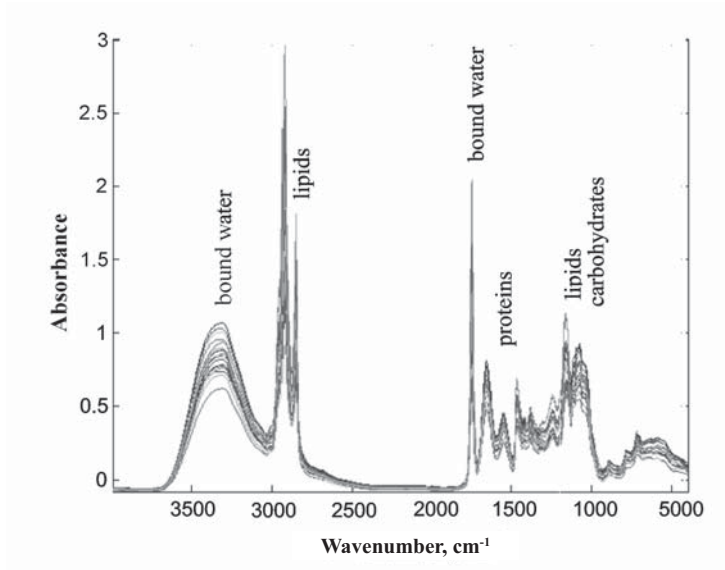
Figure 8. FTIR absorbance spectra of a set of dried bovine milk samples

region of the light spectrum. Hence, milk samples with different levels of these constituents give different absorbance spectra.

Upon closer inspection, some of the spectral differences are found to be due to variations in the physics of the sample - light scattering and sample thickness. These can be estimated and corrected for mathematically, yielding spectra with a wealth of information about the samples' variation in chemical composition,
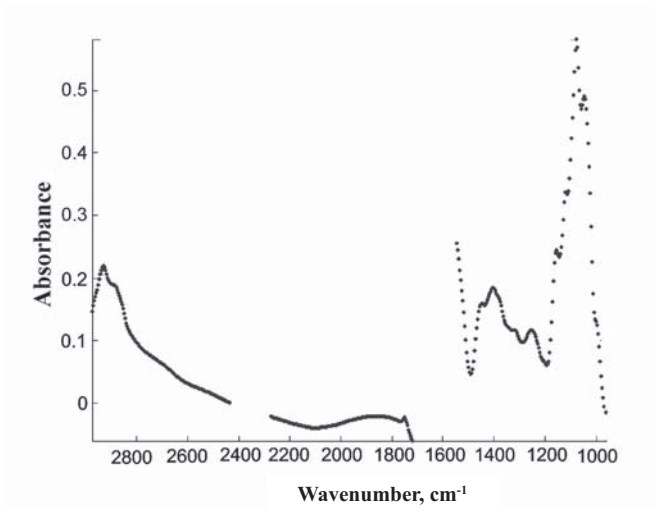

Figure 9. FTIR absorbance of homogenized whole milk

including the main aspects of fatty acid variation and possibly some of the protein variations.

One of the FTIR techniques used routinely for quality assessment of milk (Figure 9) works on intact milk samples - the Foss Electric Milkoscan instrument. Due to the high water content in milk, and the high absorbance of water, this restricts the wavenumber region where precise measurements are possible.

Work is in progress to store millions such bovine milk spectra per year, representing thousands of individual cows several times a year. By multivariate biochemometric analysis, systematic patterns of variation in these spectra will be averaged for the individual sires, and then related to available genetic, genomic and farm management information.
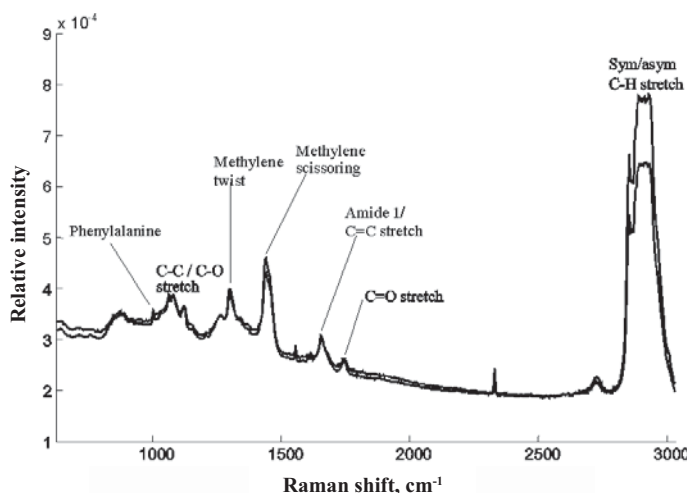

Figure 10. Raman spectra of a milk sample, with band assignment

Figure 10 shows a third biospectroscopic principle: Raman scattering. This type of measurements gives data similar to FTIR and NIR, but is primarily sensitive to symmetrical molecular bonds, such as –C-C– or –C=C–, and is therefore more or less insensitive to the presence of water - in contrast to FTIR. Thus, Raman and infrared spectroscopy complement each other. The figure outlines some of the chemical bond types detectable with this instrumentation.

The molecular vibrations from the infrared region in Figures 8 and 9 can also be observed in the near-infrared (NIR) wavelength region, in terms of their overtones. The absorbance signals are now much weaker, and can therefore be measured easily. Figure 11 shows the NIR spectra a set of natural and modified bovine milk samples, measured by diffuse reflection (top) and diffuse transmission (bottom), using an NIR Systems instrument at our research partners at Copenhagen

University. The data originate from a profiling experiment on the sensory perception of fat, homogenization, etc. in milk (Frøst et al., 2001). Campus Ås has several of these instruments. In both measurement modes the water, lipids, proteins and carbohydrates give rise to broad, somewhat overlapping absorption peaks. In addition, variations in light scattering of these milk samples due to e.g., varying homogenization, cause strong over-all spectral variations. Again, mathematical pre-processing of the spectra can separate the physical scattering variations from the chemical absorbance variations.

Figure 11 also includes the visible wavelength range (400-700 nm), where various pigment molecules display light absorption due to electronic transitions. This wavelength range can be used to predict how consumers will perceive the appearance of milk and milk products.
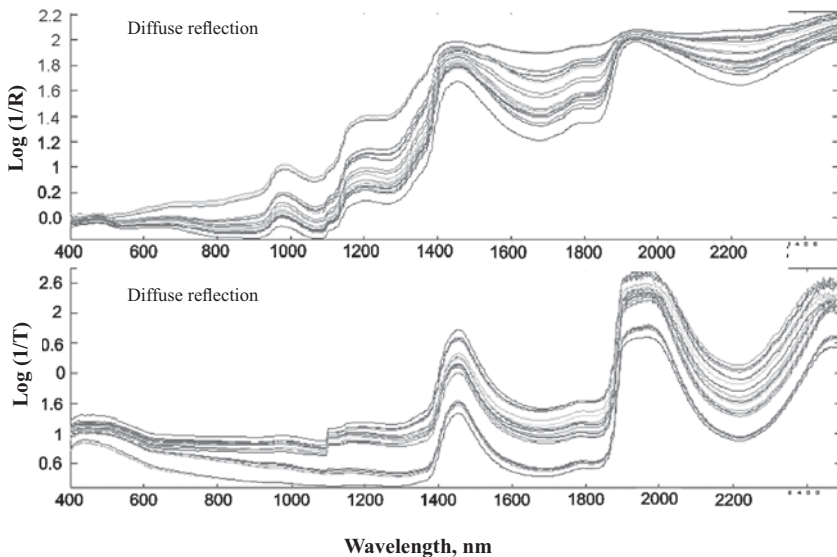

Figure 11. Visible- and near-infrared (NIR) spectra of a set of wet milk samples, obtained by diffuse reflection (top) or diffuse transmission (bottom)

Finally, Figure 12 represents quite a different type of biospectroscopy - namely autofluorescence (Wold et al., 2005). When milk or milk products are illuminated by visible or ultraviolet light, some of the photons are absorbed temporarily by particular milk components, and re-emitted at higher wavelengths. The wavelength shift reflects the type of molecule, while the intensity of the emitted light at a given wavelength reflects the concentration of the corresponding compound(s). The figure illustrates what happens to a given type of cheese when exposed to light from a commercial fluorescent light tube. This experiment showed that different types of ambient light during cheese storage caused different types of
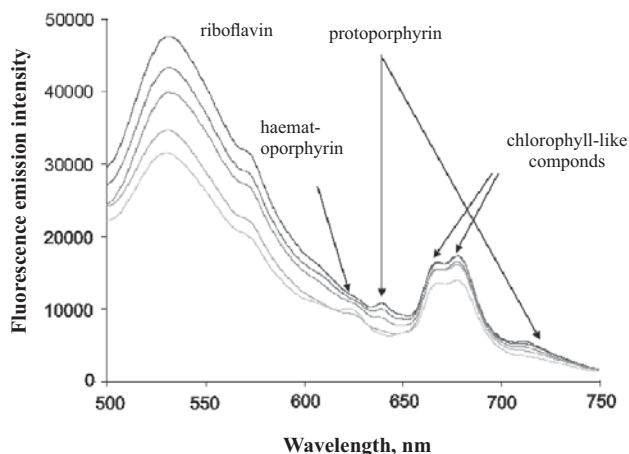
Figure 12. Fluorescence emission spectra from Swiss-like cheese exposed to light from a commercial fluorescent light tube. Light exposure times varied from 0 to 10 h

photo-oxidation and thus different degrees of sensorically perceived rancidity, which can be affect consumer acceptance.

Considering the different nature of the identified fluorescing compounds - some are probably derived from the animal itself, others from the feed - it may well be that fluorescence measurements of series of milk samples can yield phenotypic variations associated with the genome of the animals - or of the feed.

The product quality and production economy

The economic value of the milk, yoghurt, cheese, etc. depends on the willingness of consumers to buy the products. This is, in turn, largely dependent on the sensory properties of the products. In those cases when it has already been clearly established which chemical compounds or physical aspects of the milk that determine variations in its sensory properties, it is sufficient to measure these properties chemically or physically. Otherwise, it is safer to include actual sensory perception studies. Due to developments in the cognitive sciences, etc., modern sensory science is highly informative, reliable and cost-effective. Consumer studies are useful tools for measuring consumer liking - if done professionally (it is not enough just to distribute a simple questionnaire!). Descriptive sensory profiling with a trained assessor panel provides insight into the reasons why the consumers perceive products differently.

Provided that high-throughput methods, like genetic pedigree assessment, genomic profiling or biospectroscopy screening, have been used to identify particularly interesting milk samples, sensory profiling and consumer testing can be used to obtain the commercially relevant assessment of these.
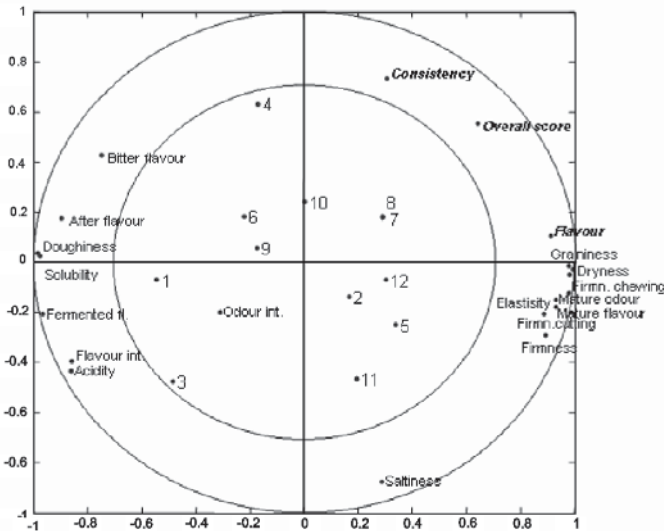
Figure 13. Sensory quality assessment of milk and milk products. Correlation loading plot summarizing a PLSR model of how 17 sensory descriptors (average of 10 trained assessors) predict rating of three quality aspects (bold italics, average of 5 experts) in 12 samples of a given type of cheese, selected to span large quality variation. The same approach can be used for fresh milk samples

Figure 13 shows an example of a sensory study of variations in a milk-based product (Hersleth et al., 2005). Twelve batches of a certain type of cheese were identified as spanning a large range of quality, based on the producer's routine assessment. These samples were submitted to quality ratings by 5 cheese experts as well as descriptive profiling by 10 trained sensory assessors. The figure summarizes how the three over-all quality rating variables (average of five experts) could be explained by the 17 descriptors (average of 10 panelists) by Partial Least Squares Regression, using The Unscrambler software (www.camo.com). The abscissa and ordinate show the correlations between the input variables and the two first latent variables, i.e. the two main patterns of co-variation. The figure shows e.g., that over-all expert score and flavour correlated positively with mature flavour and -odour, graininess, dryness, etc. and negatively with acidity. Consistency reflected minor constrasts in grainness vs firmness, and was negatively associated with saltiness.

The five most extreme cheeses were then selected for subsequent consumer assessment, where 110 consumers assessed their hedonic liking as well as their perception of flavour and texture. Details of the study are given in Hersleth et al. (2005).

The same approach can be used for assessing variations in e.g., fresh milk quality as function of e.g., animal genetics or -feeding.

Integrating the various sources of information

The different blocks of data types in functional genomics outlined in Figure 1 have been illustrated in the subsequent figures. Multivariate data modelling can then be used for integrating the different types of information, in terms of the underlying structures within each of the blocks as well as the relationships between them. Combination of modern bioinformatics, biostatistics and biochemometrics can then bring out the statistically valid information. For instance, the multivariate "soft modelling" approach explained by Martens and Martens (2001) may be used for this purpose - testing hypotheses and discovering unexpected patterns of co-variation within and between various blocks of measured variables.

As Figures 4-12 have demonstrated, milk samples can thus be phenotypic fingerprinted with respect to a number of different properties ranging from proteome details to a variety of metabolomic and quality aspects. By data modelling, these phenotypic data can be related to the gene expression measurements from e.g., udder biopsies (Figure 3), SNP structures (Figure 2) and/or genetic background information, e.g., in terms of estimated haplotype structure. All of these data can in turn be related data-analytically to external environmental or farming management data (Figure 1). Of course, it is not necessary to obtain all types of measurements for all milk samples or all animals. But once the particulary interesting milk samples have been obtained or particularly interesting animals identified in one research project, we try to secure enough material to allow other types of measurements of these.

This inter-disciplinary research process requires that several different institutions, in different locations and with different financial organizations, have to cooperate. At each stage, professionalism is needed, in order to secure high-quality data. But equally important, distinctly different research cultures will have to meet, ranging from animal science, genetics and genomics *via* analytical chemistry to dairy science and sensory science.

DISCUSSION

One important aspect of this cross-disciplinary effort is to ensure that the research process contains both
– the explorative discovery aspect that stimulates innovation and real-world relevance
– the confirmative testing aspect that satisfies "the hypothetic-deductive method" to guard against false interpretations and wishful thinking
– the mechanistic detailing aspects that can provide causal insight.

One way to overview this work is outlined in Figure 14. The research is considered in more or less formal research cycles. Each of these cycles convert

The cyclic research process:
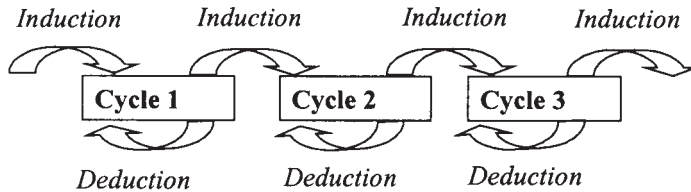
Data-related elements in one multivariate research cycle:

1. Define purpose, scope and working hypothese from previous cycles.

2. Plan experiment: choice of samples (objects), choice of measurement types (variables)

3. Experimental work: Controlled measuring conditions, sufficient full biological replication, random sample order

4. Handling of raw data: Look for gross errors, average over parallels, preprocess to linearize and to separate e.g., chemical from physical effects and noise, organize data tables with good naming of samples and variables

5. Data modelling: extract main features, test prior hypotheses against errors, look for unexpected clusters and co-variation patterns, look for serious errors and re-analyse without them

6. Demonstrate main results in raw data. Draw conclusions for next selection cycle: identify need for more data, form new hypotheses

Figure 14. Research progress in cross-disciplinary studies: Repeating the research cycle of explorative induction and confirmative deduction

inductively formulated ideas from previous cycles into deductively tested conclusions plus new inductive discoveries. The data-modelling stages within each stage are also outlined in the Figure. In this framework there is a need for all the above-mentioned scientific cultures.

The integrative genetic approach, which includes molecular genomics, high-definition phenotype screening as well end-use quality relevance, is rather generic. For instance, not only the bovine or caprine animal genome and its relationship to milk quality is of interest in this context; the same concepts and methods may also be applied to the feed plants as well as to the intestinal microflora of the animals - to reveal their interactions.

So, why do we try to form this concerted cross-disciplinary environment in milk and milk-related research? It is because we hope to gain more insight and faster research progress by cooperating across scientific traditions. Our campus is large enough to have many types of milk-related expertises, but small enough to foster cooperation synergy. Once an interesting set of samples has been acquired for one given type of measurement, the cost of adding e.g., biospectroscopic fingerprinting is usually small, while the benefit of the resulting added insight can be substantial.

For instance, in animal feeding experiments, unintended genetic differences between animals usually represent a source of unfortunate "biological variation" that creates "random errors". With the present approach they can be characterized as genetic variations that can be identified and compensated for, in the experimental design phase and/or during the subsequent data analysis, thereby increasing the statistical power of the feeding experiments. Genotype x environment interactions may even be revealed. The same goes for dairy science experiments concerning e.g., milk quality. Conversely, in genetic studies, unintended environmental variations in a given phenotype due to e.g., farming practice can be picked up by other phenotypic measurements and corrected for statistically, making the subsequent genetic data modelling more precise - possibly with fewer animals. Finally, the explosive development in available molecular genomics is expected to provide causal insight into both genomic and environmental basis for milk production and milk quality. However, these are early days, and only time will show if the cross-disciplinary platform will be worth while.

REFERENCES

Chianese L., Garro G., Nicolai M.A., Mauviello R., Ferranti P., Pizzano R., Cappuccio U., Laezza P., Addeo F., Ramunno L., Rando A., Rubino R., 1993. The nature of beta-casein heterogenity in caprine milk. Lait 73, 533-547

Ding C., Cantor C.R., 2003. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. Proc Nat. Acad. Sci. USA 100, 3059-3064

Frøst M.B., Dijksterhuis G., Martens M., 2001. Sensory perception of fat in milk. Food Qual. Preference 12, 327-336

Hersleth M., Austvoll Illseng M., Martens M., Næs T., 2005. Perception of cheeses: A comparison of quality scoring, descriptive analysis and consumer responses. J. Food Qual. 28, 333-349

Martens H., Martens M., 2001. Multivariate Analysis of Quality. An Introduction. J. Wiley & Sons, Ltd., Chichester (UK)

Miranda G., Mahe M.F., Leroux C., Martin P., 2004. Proteomic tools to characterize the protein fraction of Equidae milk. Proteomics 4, 2496-2509

Wold J.P., Veberg A., Nilsen A., Juzenas P., Iani V., Moan J., 2005. The role of naturally occurring chlorophyll and porphyrins in light induced oxidation of dairy products. A study based on fluorescence spectroscopy and sensory analysis. Int. Dairy J. 15, 343-353