

Can metabolomics be used for assessing nutritive-dependent human diseases?

Oliver Fiehn¹ & Joachim Spranger²

¹Max-Planck-Institute of Molecular Plant Physiology, 14424 Potsdam/Golm, Germany; ²German Institute of Human Nutrition (DIfE), 14558 Bergholz-Rehbrücke, Germany

Key words: diabetes mellitus, mass spectrometry, metabolomics, pattern recognition

Abstract

Metabolomic approaches are taking place within emerging technologies that share the focus to unravel unexpected effects in classical biological experiments, for example mutant/wild type comparisons. This review highlights the possibility of metabolomics going beyond this stage and comparing the onset and the pathogenesis of complex diseases, such as type 2 diabetes mellitus, in large scale projects that might reach out to compare different populations and different nutritional diets and interventions. The steps of validation that are needed to pursue such projects are described, and the range of analytical and computational approaches that can be utilized are discussed. Using type 2 diabetes mellitus, it is emphasized that there is a high economic and scientific need to reliably diagnose and predict metabolic diseases in a cost effective manner.

Introduction

In recent years a lot of different techniques have emerged that can be used for functional genomics as well as for diagnostic and screening tools. Among these techniques, the response of primary gene products to the onset of diseases has been of special interest. It is hoped, that by investigating all gene products simultaneously, one could better understand disease mechanisms and find optimal therapeutic targets. Although this approach has been successful in a number of cases, it is still unclear if the analysis of mRNA or proteins can be used as a universal tool aiming at understanding mechanisms, finding disease patterns and successfully diagnosing human health states. The number of practical and fundamental restrictions of these approaches have already begun to cool down the hype surrounding novel genomic and analytical approaches. For example, typical transcriptomic and proteomic analyses usually come along with high costs per sample. Simple considerations of constraints in the available budget then often limit the number of analyzed samples to a point at which basic statistical rules get violated. Especially the range of biological variability is often underestimated, since following

the paradigm of simple genetic control, a population of, say, clones of knock-out mice should result in very homogenous biological responses if developmental and environmental factors are strictly controlled. However, what is found experimentally differs from this hypothesis quite often. This review will focus on potential implications of metabolic profiling as a tool to identify new, so far unknown factors associated with disease status. We will exemplify the potential of this method using the association between specific fats and development of type 2 diabetes as a test case.

Type 2 diabetes as a test case

Type 2 diabetes mellitus (T2DM) is one of the major nutritive-dependent diseases worldwide. Despite the enormous socio-economic burden arising from hyperalimantation-dependent diseases like T2DM (Warram et al., 1997) and considerable evidence that timely interventions are able to prevent the manifestation of the disease, factors allowing risk assessments in healthy individuals are rare. Although the primary factors causing this disease are unknown, it is clear that genetic and lifestyle factors, and espe-

cially nutrition, play a major role in its development (Hu et al., 2001). The number of T2DM manifestations can be reduced by more than 50% by life-style interventions such as diet or increased physical activity (Tuomilehto et al., 2001; Knowler et al., 2002), demonstrating that accurate prediction of individuals at risk to develop T2DM would allow effective therapy in these individuals. Evidence arises not only from cross-sectional studies but also from therapeutic studies demonstrating the effectiveness of diet and physical activity to prevent the onset of the disease in individuals with impaired glucose tolerance.

Beside anthropometric parameters such as body-mass-index and waist-hip ratio, markers such as cholesterol, triglycerides, LDL, HDL and lipoproteins are amongst the best predictors of T2DM. However, taking the known risk factors together allows a prediction of future T2DM in only about 10% of individuals classified at risk, with a considerable number of false-positive individuals who do not develop the disease despite having the risk markers (Report of the Expert Committee, 1997). The same is true for most other (metabolic) diseases for which molecular markers are known to result in a high percentage of both false positive and false negative results. Additional diagnostic tools for a precise identification of individuals at risk are therefore urgently needed. Since metabolites are the best-known predictors for diabetes, metabolomic approaches might be more appropriate.

Environmental factors considerably influence the pathogenesis of many complex diseases such as cancer, coronary heart disease and type 2 diabetes. The prevalence of type 2 diabetes is increasing rapidly worldwide. The lowest rates of type 2 diabetes are found in rural communities where people are still living traditional lifestyles (Amos et al., 1997). Dramatic changes in the prevalence or incidence of type 2 diabetes have been observed in communities where there have been major changes in the type of diet consumed, from a traditional indigenous diet to a typical 'Western' diet, e.g., Pima Indians in Arizona, Micronesians in Nauru and Aborigines in Australia (Bennett, 1999). The prevalence of type 2 diabetes is estimated to reach 5.4% and the number of adults with diabetes has been estimated to be as much as 300 million individuals worldwide in the year 2025. Indeed, the majority of cases of type 2 diabetes in the future will occur in developing countries, with India and China having more cases than any other country in the world (King et al., 1995). This epidemic increase of type 2 diabetes is associated with enormous socio-economic consequences

and an increasing mortality due to diabetic complications such as coronary heart disease and end-stage renal insufficiency.

Although diet and nutrition are widely believed to play an important role in the pathogenesis of type 2 diabetes, specific dietary factors are difficult to elucidate and have not been clarified yet. A large number of epidemiological studies show a protective effect of vegetables and fruits against development of type 2 diabetes (Hu et al., 2001). However, despite these data much controversy exists about the relation between the amount and types of dietary fat and carbohydrate and the risk of diabetes. Other molecules such as plant-derived phenolic antioxidants may play an additional role, but have been investigated in less detail. The existing dietary guidelines recommend low-fat and high-carbohydrate diets for the prevention of diabetes, coronary heart disease and other chronic diseases. However, neither fats nor carbohydrates nor dietary antioxidants are homogenous molecules, and it is now increasingly clear that different types of fat and carbohydrate have different effects on glucose homeostasis, insulin sensitivity and insulin secretion.

The association between dietary fatty acids has been proposed several decades ago, showing that elevated fatty acids may play a causal role in the development of insulin resistance by competing with glucose for oxidation (Randle hypothesis, Randle et al., 1965). There is extensive literature on the relationship between free fatty acids and insulin action (Boden, 1997). In animal experiments, saturated, monounsaturated and polyunsaturated fats have been shown to induce insulin resistance when fed as high-fat diets (excluding n-3 fat) (Storlien et al., 1996). In epidemiological studies high saturated fat intake has been associated with higher risk of glucose intolerance, elevated fasting glucose and increased insulin concentrations (Hu et al., 2001b). Furthermore, higher proportions of saturated fatty acids in serum lipids/muscle phospholipids have been associated with higher fasting insulin levels, lower insulin sensitivity and higher risk of developing type 2 diabetes (Vessby et al., 1994; Folsom et al., 1996), while higher proportions of long-chain polyunsaturated fatty acids in skeletal muscle phospholipids have been associated with better insulin sensitivity in humans (Borkmann et al., 1993). With respect to monounsaturated fatty acid, the epidemiological data produce a less clear picture, although some studies indicate that a high amount of monounsaturated fatty acids may be harmful (Feskens et al., 1995). A longer term intervention study confirmed

that a substitution of monounsaturated fat for saturated fat significantly improved insulin sensitivity in healthy subjects after a three-month dietary period. An interesting interaction between the intake and fatty acid composition was reported: the favorable effect of substituting monounsaturated fat for saturated fat was lost in individuals consuming more than 37% of energy as fat, demonstrating that both, the type and the amount of fat are important regarding the induction of insulin sensitivity (Vessby et al., 2001). With respect to the above-mentioned data it is evident that the basis for current recommendations is rather small. Additionally other dietary factors such as the amount and different types of carbohydrates, polyphenolic antioxidants and many micronutrients not described here, as well as other lifestyle factors such as physical activity, substantially influence the development of complex diseases such as type 2 diabetes. Consequently, the contribution and importance of dietary factors per se need to be carefully investigated and interpreted taking into account the fact that it would be an over-simplification to propose that any single nutrient is specifically diabetogenic. Therefore additional tools to identify the metabolic changes of specific diets are urgently needed.

Current screening programs trying to identify healthy individuals at risk are limited by a high rate of false positives, resulting not only in a high number of unnecessarily treated individuals but in reduced cost-effectiveness of screening programs. Based on known risk factors, screening costs are estimated to be 4000–8000 US dollars per new case of diabetes identified, resulting in correct prediction of only 2.5% of all individuals screened in a representative population. Accordingly, the yield of systematic diabetes screening programs utilizing known risk factors is low though costs are high. This data is in strong contrast to the dramatic socio-economic burden associated with type 2 diabetes mellitus. The American Diabetes Association estimates that diabetes accounts for 27 billion US dollars in direct medical costs and 32 billion US dollars in indirect or lost-productivity costs in 1997. This demonstrates the potential of cost prevention by efficient screening methods. Several studies in the U.S. and the U.K. have shown that aggressive intervention against diabetes in high-risk cohorts can retard the development of complications. Such pre-manifestation interventions can cause a massive decrease in indirectly health-related costs that are now counting for 4000–9000 US dollars per person per year. Metabolomic approaches could significantly im-

prove the correct prediction of diabetes pathogenesis and lower the disease-related costs, thereby opening a further field of application for this technique that might rapidly gain significance in other disease-related studies. Metabolomic techniques might further lead to the identification of metabolite clusters or single, so far unknown metabolites, which yield improved risk stratification of individuals to develop T2DM. Due to its unbiased nature, the application of metabolomic techniques could also prove beneficial to elucidate the pathogenesis and molecular mechanisms of T2DM.

Profiling techniques

It is generally accepted that profiling technologies may aid in distinguishing complex phenotypes by pattern recognition. In an association study between specific RNA clusters and the progression of breast cancer (van 't Veer et al., 2002), it was demonstrated that unbiased analyses have a high potential to predict the risk of individuals for specific endpoints of complex diseases. However, genome wide RNA or proteomic analyses are too costly to be applied to thousands of patients in screening studies. For metabolic diseases like type 2 diabetes mellitus, metabolite profiling might become a highly competitive alternative for comprehensive phenotype analysis and disease prediction. Metabolites are well known to play a pivotal role in the pathogenesis of complex diseases or phenotypes such as obesity. For example, a new lipid metabolite has been shown very recently to be regulated by feeding and to be anorexic (Rodriguez et al., 2001). Profiling technologies have been known for a long time in clinical research (Tanaka et al., 1980 a,b) for the recognition of inborn errors and other metabolic disorders (Jellum et al., 1988). Nevertheless, there are only a few reports on their use for disease diagnosis, specifically for acidurias (Kimura et al., 1999; Halket, 1999), cervical cancer (Kim et al., 1998), and mitochondrial myopathy (Ning et al., 1996). In the public domain, metabolite profiling has not yet been extended to a broader analysis of metabolic disorders, aiming at recognition of disease progression, therapeutic success, or finding potential drug targets in an unbiased way. Metabolite fingerprinting techniques like NMR (Gavaghan et al., 2000; Raamsdonk et al., 2001), FT-IR (Goodacre et al., 2000), or FIA-MS (Vaidyanathan et al., 2001) regularly provide sufficient information to discriminate between populations such as wild type and mutant strains, or healthy and diseased states. However, reso-

lution of such techniques is usually inadequate to pin down the pattern to the level of groups of individual metabolites. Nevertheless, such fingerprinting techniques might be invaluable to rapid screening of large cohorts to get a first estimate of metabolic distances and risk stratifications, in order to reduce the number of samples that need to be analyzed by more time-consuming, comprehensive metabolomic techniques in a second screening round.

Such metabolomic techniques can be defined as the simultaneous detection and quantization of all individual metabolites in a given biological system, as further explained in recent reviews (Fell, 2001; Fiehn, 2002). Metabolomic approaches are primarily based on mass spectrometry due to the inherent selectivity, universality and sensitivity of this technology. During the past years, metabolite-profiling analysis by gas chromatography/quadrupole mass spectrometry (Fiehn et al., 2000a) has been improved by coupling to time-of-flight mass spectrometers (Shellie et al., 2001; Weckwerth et al., 2001). Using this approach more than 1,000 components from 15 mg FW plant leaves can be determined with a throughput of > 1,000 samples per month. In a test case for 400 metabolites analysed by GC/TOF, an average analytical error was determined to be 10% relative standard deviation, which is just the figure that was obtained earlier for GC/quadrupole MS profiling experiments (Fiehn et al., unpublished results).

Whereas GC/TOF has shown to be the 'gold standard' for high throughput metabolite profiling, it has a bias against involatile and large metabolites. For non-targeted multiparallel analysis aiming at finding metabolic differences regardless of size, chemical, or physical nature of the metabolite markers, a set of analytical problems has to be solved. For example, unbiased quantitative assessments of larger compounds need to include complex lipids, co-factors, hormones, glycosylated and glucuronidated intermediates, and redox-active compounds by LC/MS and MS² methods. Numerous methods exist for the analysis of unpolar compounds such as phenolic and flavonoids (Justesen et al., 1998) by reverse phase LC/MS, whereas only recently LC/MS was extended to highly polar biomolecules using hydrophilic interaction chromatography (Tolstikov and Fiehn, 2002). However, there are no reports on unbiased and automated peak-finding algorithms for LC/MS with simultaneous quantitation of all detected peaks, as is standard for GC/MS (Stein, 1999). Therefore, novel methods and tools have to be developed to increase

the throughput and the reliability of untargeted multiparallel LC/MS-MS analysis.

During metabolomic studies, some peaks with unknown chemical structure are regularly assigned to certain biological states. This will in part be the case, because comprehensive metabolite mass spectral libraries are clearly lacking for both GC/MS and LC/MS approaches. Further, the scope and the range of plant and mammalian tissues are still underestimated due to decades of classical, hypotheses-driven biological approaches. These unknown metabolites could, thus, be of special scientific and economic importance. Therefore, structural investigation is an essential part of every truly metabolomic study (Fiehn et al., 2000b). For example, it can be performed by reverse phase and hydrophilic interaction chromatography coupled to ion trap and quadrupole-time of flight (QTOF) mass spectrometry. Whereas ion traps are essential for *de novo* structural elucidation by their stepwise MSⁿ capabilities (Drexler et al., 1998), QTOF mass spectrometry delivers exact masses on the MS level (Blom, 2001), but not on the MS² level. Using these data, structures can be proposed aided by mass spectral interpretation software and interrogation of chemical and biochemical databases (Beilstein, KEGG). However, in many cases ambiguity cannot be avoided this way, especially for distinguishing isobaric and isomeric compounds. Proposed structures, therefore, will often need to be confirmed by 1D and 2D-NMR studies (Pauli, 2000).

Validation of metabolomics for disease assessments

As we have seen, metabolomic analysis could potentially serve as a rapid, reliable, sensitive and cost-effective method for disease assessments. However, great care must be taken, especially when sampling tissues garnered from humans or animal models, since this step might well turn out to contribute the largest errors that lead to the high biological variability found for human populations. This biological variability necessitates that a large number of human metabolic snapshots be acquired (Kell and Mendes, 2000), which in turn simultaneously generates the need for robust and high throughput analytical technologies to enable statistically sound evaluations of metabolic patterns associated with mammalian diseases. Therefore, even academic laboratories have to undergo a rigorous validation process (Krull and Swartz, 1999) if results should not be rendered meaningless. This validation

has to cover all steps from clinical sampling to data acquisition, including the ability to analyse a multitude of samples to allow rigorous statistical evaluations. In this context, validation must focus on ruggedness and repeatability, which needs to be complemented by tests for sensitivity, selectivity, universality, comprehensiveness, and comparisons to existing methods. Ruggedness can be assessed by analysing small and large variations in sample handling prior to analysis. Robustness needs further to be assessed by variation of existing sample preparation protocols established for plant metabolomics, for example for extraction temperatures, extraction times, and solvent compositions. Analytical robustness and repeatability might further be tested by analysing aliquots of a larger sample pool within one day vs. weeks, triplicate and seven-fold injections, and precision tests under varying derivatization reaction times and temperatures (if GC/MS is applied). Selectivity could be assessed by a set of isomers such as unsaturated fatty acids and monosaccharides. Sensitivity may be examined by internal spiking with stable isotope-labelled reference compounds (such as carbohydrates, secondary metabolites, fatty acids). Lastly, routines for instrument quality control need to be included if the methods are to be transferred to high throughput.

Once these basic steps have been fulfilled, proof-of-concept studies may be performed upon the ability of adapting existing protocols for GC/TOF detection of small plant metabolites, and this may be extended to create truly metabolomic approaches by high throughput LC/MS² technologies. This approach might then be applied to identify metabolite patterns associated with characteristics or endpoints of nutritive-dependent diseases, for example focusing on type 2 diabetes mellitus. After generating a large database and applying appropriate chemometric tools, single metabolites associated with characteristics of a disease or specific nutritional interventions might be identified.

Using GC/TOF as analytical backbone, the longitudinal and cross-sectional investigation of the metabolome associated with different diets might elucidate important and new regulatory pathways involved in the pathogenesis of T2DM. Due to the high genetic and biological variability in humans the authors performed a test case study in clonal cell lines (Figure 1) and syngenic animals (Figure 2) demonstrating that the method is principally adoptable to mammalian systems. It could be shown that metabolomic techniques result in complex chromatograms of hundreds

of peaks, many of which remained unknown in the first round of structural elucidation. However, clear differences between metabolic profiles between KO and control mouse lines could be observed for various tissues. For beta-cells, the anticipated uptake of palmitic acid forming glycerol esters could directly be revealed from the GC/TOF profiles (Figure 1).

Databases and data analysis

The overall success of health-related metabolomic approaches will rely on the ability to store, curate, and analyse large piles of clinical and metabolomic data. Therefore, an (international) database should be built up along with ongoing experiments, similar to databases that exist for transcript arrays. The import of metabolomic result files must necessarily be accompanied with available biological and clinical background information. Import of other results or meta-data must include strategies on how to preserve the integrity of the database and how to monitor any changes in information content. The schematic design and the implementation of this database needs to ensure logical correctness of the data entered into the system, to minimise duplication, and to control data safety and accessibility aspects. The data model also needs to support concepts including species, line variants, nutritional regimes, experimental designs, biological and metabolome data, and it must distinguish raw data from derived data. The latter will be assigned with the axioms and calculation methods in order to evaluate their origin and value.

The primary focus of such a metabolomic expression database does not necessarily need to enable most sophisticated calculations and pattern recognition tools, but rather to permit selection and sorting of files according to user-defined criteria, e.g., by biological matrix, disease, or sampling date. This selection of result files might then be downloaded for in-house usage.

Routine clustering of multivariate data sets may be performed using any commercial package like Pirouette, SyStat, or MatLab, accompanied by further statistical software packages like 'R', SAS, or SPSS. Such clustering methods can broadly be separated to supervised learning methods that include classificatory background knowledge and use training data sets, and unsupervised methods that do not use any meta information of the data sets. Regularly, supervised learning methods (such as discriminant

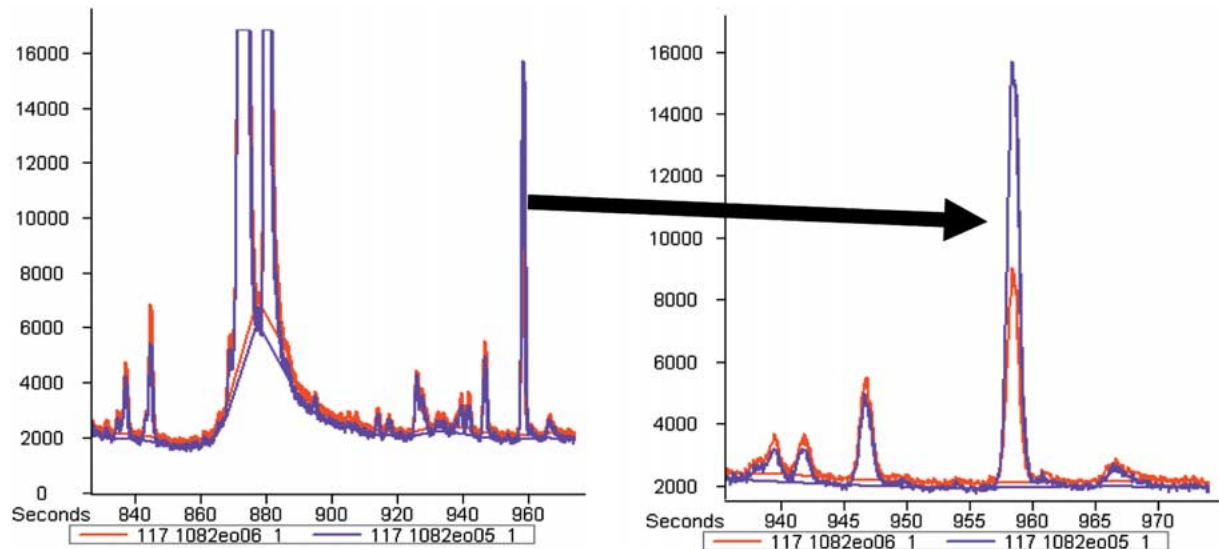


Figure 1. Increase of 16:0-Glycerolesters in palmitate (blue) vs. albumin (red) treated beta-cells.

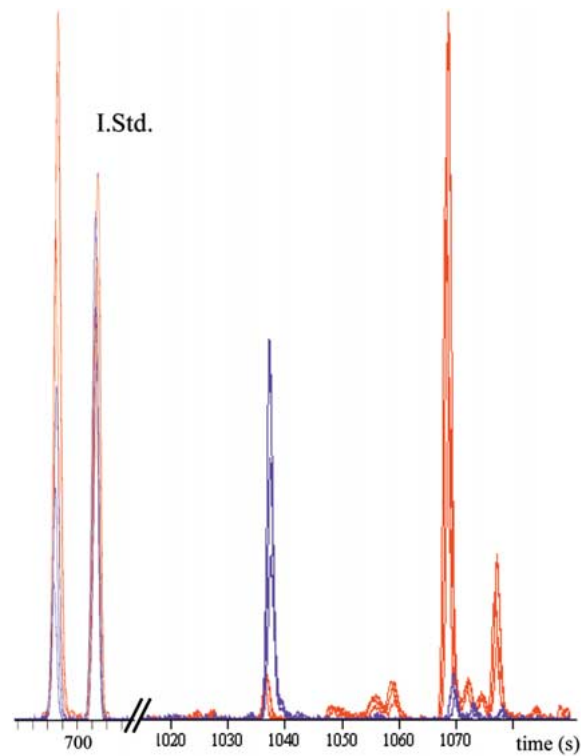


Figure 2. Superposition of each 3 liver tissue chromatograms (partially shown) of a KO mouse line (blue) and a control line (red). Several peaks can be seen to be significantly downregulated in the KO line, whereas some are increased relative to the control.

function analysis, support vector machines, artificial neural networks, self-organizing maps, decision trees or genomic programming (Gilbert et al., 2000; Johnson et al., 2000) have more discriminatory power than unsupervised methods. However, great care must be taken not to violate the assumptions underlying these tools (for example by using a high number of individual data sets), and the robustness of correct classifications must adequately be taken into account (for example, by cross-validation). A nice example of the power of combining supervised and unsupervised method for a difficult discriminatory task was provided by Raamsdonk et al. (2001), who used a combination of principle components and discriminant function analysis to separate a silent mutation from a wild type population. Once populations can be discriminated, more refined methods such as large-scale co-response analysis can be applied to unravel inherent metabolic networks (Kose et al., 2001), or to find time-dependent patterns (Lukashin et al., 2001) in large scale data sets.

All clustering tools, classification approaches, and statistical methods have in common that there is simply no 'best' algorithm. Often, underlying assumptions (for example about data frequency distributions) or even data types (like canonical or linear data, connected or independent data) are different. Therefore, seeking advice from statistics experts is highly advisable, especially for large-scale projects.

Conclusions

Metabolomic approaches have made the first steps towards detecting unexpected effects in biological experiments. However, for large scale projects such as predicting the pathogenesis of nutritive-dependent diseases like type 2 diabetes mellitus, large efforts have to follow on the level of experimental designs, analytical techniques, raw data analysis, data storage and retrieval, and data mining and interpretation. Apart from efforts to hasten metabolite identification, the strongest need for health-related metabolomics is currently to develop validated processes including all steps from sampling to data interpretation in a real-world proof-of-concept study. On the long run, such comprehensive studies may also help to clarify the individual contribution of food ingredients to preserving health, such as antioxidants, flavonoids, carotenoids, vitamins, or anthocyanins.

References

- Amos AF, McCarty DJ & Zimmet P (1997) The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabet. Med.* 14: 1–85.
- Bennett PH (1999) Type 2 diabetes among the Pima Indians of Arizona: an epidemic attributable to environmental change? *Nutr. Rev.* 57: 51–54.
- Blom KF (2001) Estimating the precision of exact mass measurements on an orthogonal time-of-flight mass spectrometer. *Anal. Chem.* 73: 715–719.
- Boden G (1997) Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Diabetes* 46: 3–10.
- Borkman M, Storlien LH, Pan DA, Jenkins AB, Chisholm DJ & Campbell LV (1993) The relation between insulin sensitivity and the fatty-acid composition of skeletal-muscle phospholipids. *N. Engl. J. Med.* 328: 238–244.
- Drexler DM, Tiller PR, Wilbert SM, Bramble FQ & Schwartz JC (1998) Automated identification of isotopically labeled pesticides and metabolites by intelligent ‘real time’ LC-tandem MS using a bench-top ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* 12: 1501–1507.
- Fell DA (2001) Beyond genomics. *Trends Gen.* 17: 680–682.
- Feskens EJM, Virtanen SM, Rasanen L, Tuomilehto J, Stengard J, Pekkanen J, Nissinen A & Kromhout D (1995) Dietary factors determining diabetes and impaired glucose tolerance. A 20-year follow-up of the Finnish and Dutch cohorts of the Seven Countries Study. *Diabetes Care* 18: 1104–1112.
- Fiehn O (2001) Combining genomics metabolome analysis and biochemical modeling to understand metabolic networks. *Comp. Funct. Genom.* 2: 155–168.
- Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.* 48: 155–171.
- Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN & Willmitzer L (2000a) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18: 1157–1161.
- Fiehn O, Kopka J, Trethewey RN & Willmitzer L (2000b) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* 72: 3573–3580.
- Folsom AR, Ma J, McGovern PG & Eckfeldt H (1996) Relation between plasma phospholipid saturated fatty acids and hyperinsulinemia. *Metabolism* 45: 223–228.
- Gavaghan CL, Holmes E, Lenz E, Wilson ID & Nicholson JK (2000) An NMR-based metabolomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett.* 484: 169–174.
- Gilbert RJ, Rowland JJ & Kell DB (2000) Genomic computing: explanatory modelling for functional genomics. In: Whitley D, Goldberg D & Cantú-Paz E (eds) *Proceedings of the genetic and evolutionary computation conference* (pp. 551–557). Morgan Kaufman, San Francisco.
- Goodacre R, Shann B, Gilbert RJ, Timmings EM, McGovern AC, Alsborg BK, Kell DB & Logan NA (2000) Detection of the dipicolinic acid biomarker in *Bacillus* spores using curie-point pyrolysis mass spectrometry and fourier transform infrared spectroscopy. *Anal. Chem.* 72: 119–127.
- Halket JM, Przyborowska A, Stein SE, Mallard WG, Down S & Chalmers RA (1999) Deconvolution gas chromatography mass spectrometry of urinary organic acids – Potential for pattern recognition and automated identification of metabolic disorders. *Rapid Commun. Mass Spectrom.* 13: 279–284.
- Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S & Solomon CG (2001) Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N. Engl. J. Med.* 345: 790–797.
- Hu FB, van Dam RM & Liu S (2001) Diet and risk of Type II diabetes: The role of types of fat and carbohydrate. *Diabetologia* 44: 805–817.
- Jellum E, Kvittingen EA & Stokke O (1988) Mass spectrometry in diagnosis of metabolic disorders. *Biomed. Environ. Mass Spectrom.* 16: 57–62.
- Johnson HE, Gilbert RJ, Winson MK, Goodacre R, Smith AR, Rowland JJ, Hall MA & Kell DB (2000) Explanatory analysis of the metabolome using genetic programming of simple interpretable rules. *Genet. Program Evol. Mach.* 1: 243–258.
- Justesen K, Knuthsen P & Leth T (1998) Quantitative analysis of flavonols, flavones, and flavanones in fruits, vegetables and beverages by high-performance liquid chromatography with photodiode array and mass spectrometric detection. *J. Chromatogr. A* 799: 101–110.
- Kell DB & Mendes P (2000) Snapshots of systems. In: *Technological and medical implications of metabolic control analysis*. Eds.: Cornish-Bowden AJ & Cárdenas ML (Kluwer Academic Publishers) pp. 3–25.
- Kim K-R, Park H-G, Paik M-J, Ryu H-S, Oh KS, Myung S-W & Liebich HM (1998) Gas chromatographic profiling of urinary organic acids from uterine myoma patients and cervical cancer patients. *J. Chromatogr. B* 712: 11–22.
- Kimura H, Yamamoto T & Seiji Y (1999) Automated metabolic profiling and interpretation of GC/MS data for organic aciduria screening: A personal computer-based system. *Tohoku J. Exp. Med.* 188: 317–344.
- King H, Aubert RE & Herman WH (1998) Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections. *Diabetes Care* 21: 1414–1431.
- Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA & Nathan DM (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* 346: 393–403.

- Kose F, Weckwerth W, Linke T & Fiehn O (2001) Visualising plant metabolomic correlation networks using clique-metabolite matrices. *Bioinformatics* 17: 1198–1208.
- Krull IS & Swartz M (1999) Analytical method development and validation for the academic researcher. *Anal. Lett.* 32: 1067–1080.
- Lukashin AV & Fuchs R (2001) Analysis of temporal gene expression profiles: clustering by simulated annealing and determining the optimal number of clusters. *Bioinformatics* 17: 405–414.
- Ning C, Kuhara T, Inoue Y, Zhang CH, Matsumoto M, Shinka T, Furumoto T, Yokota K & Matsumoto I (1996) Gas chromatographic mass spectrometric metabolic profiling of patients with fatal infantile mitochondrial myopathy with de Toni-Fanconi-Debre syndrome. *Acta Paed. Japon.* 38: 661–666.
- Pauli GF (2000) Higher order and substituent chemical shift effects in the proton NMR of glycosides. *J. Nat. Prod.* 63: 834–838.
- Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K & Oliver SG (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* 19: 45–50.
- Randle PJ, Garland PB, Newsholme EA & Hales CN (1965) The glucose fatty acid cycle in obesity and maturity onset diabetes mellitus. *Ann. N.Y. Acad. Sci.* 131: 324–333.
- Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus *Diabetes Care* (1997) 20: 1183–1197.
- Rodriguez de Fonseca F, Navarro M, Gomez R, Escuredo L, Nava F, Fu J, Murillo-Rodriguez E, Giuffrida A, LoVerme J, Gaetani S, Kathuria S, Gall C & Piomelli D (2001) An anorexic lipid mediator regulated by feeding. *Nature* 414: 209–212.
- Shellie R, Marriot P & Morrison P (2001) Concepts and preliminary observations on the triple dimensional analysis of complex volatile samples by using GC x GC – TOF MS. *Anal. Chem.* 73: 1336–1344.
- Stein SE (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data *J. Am. Soc. Mass Spectrom.* 10: 770–781.
- Storlien LH, Baur LA, Kriketos AD, Pan DA, Cooney GJ, Jenkins AB, Calvert GD & Campbell LV (1996) Dietary fats and insulin action. *Diabetologia* 39: 621–631.
- Tanaka K, Hine DG, West-Dull A & Lynn TB (1980a) Gas-chromatographic method of analysis of urinary organic acids I Retention indices of 155 metabolically important compounds. *Clin. Chem.* 26: 1839–1846.
- Tanaka K, West-Dull A, Hine DG, Lynn TB & Lowe T (1980b) Gas-chromatographic method of analysis of urinary organic acids II Description of the procedure and its application to diagnosis of patients with organic acidurias. *Clin. Chem.* 26: 1847–1853.
- Tolstikov VV & Fiehn O (2002) Analysis of highly polar compounds of plant origin: Combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* 301: 298–307.
- Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, Keinanen-Kiukkaanniemi S, Laakso M, Louheranta A, Rastas M, Salminen V, Uusitupa M, Aunola S, Cepaitis Z, Moltchanov V, Hakumaki M, Mannelin M, Martikkala V, Sundvall J (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* 344: 1343–1350.
- Vaidyanathan S, Rowland JJ, Kell DB & Goodacre R (2001) Discrimination of aerobic endospore-forming bacteria via electrospray-ionization mass spectrometry of whole cell suspensions. *Anal. Chem.* 73: 4134–4144.
- van 't Veer LJ, Dai H, van de Vijver MJ, He YDD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R & Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530–536.
- Vessby B, Aro A, Skarfors E, Berglund L, Salminen I & Lithell H (1994) The risk to develop NIDDM is related to the fatty acid composition of the serum cholesterol esters. *Diabetes* 43: 1353–1357.
- Vessby B, Tengblad S & Lithell H (1994) Insulin sensitivity is related to the fatty acid composition of serum lipids and skeletal muscle phospholipids in 70-year-old men. *Diabetologia* 37: 1044–1050.
- Vessby B, Unsitupa M, Hermansen K, Riccardi G, Rivellese AA, Tapsell LC, Nalsen C, Berglund L, Louheranta A, Rasmussen BM, Calvert GD, Maffetone A, Pedersen E, Gustafsson IB & Storlien LH (2001) Substituting dietary saturated for monounsaturated fat impairs insulin sensitivity in healthy men and women: The KANWU Study. *Diabetologia* 44: 312–319.
- Warram JH, Kopczynski J, Janka HU & Krolewski AS (1997) Epidemiology of non-insulin-dependent diabetes mellitus and its macrovascular complications. A basis for the development of cost-effective programs. *Endocrinol. Metab. Clin. North. Am.* 26: 165–88.
- Weckwerth W, Tolstikov VV & Fiehn O (2001) Metabolomic characterization of transgenic potato plants using GC/TOF and LC/MS analysis reveals silent metabolic phenotypes. *Proc. 49th ASMS Conf. Mass Spectrom. All Top* (1–2).