Chimia 54 (2000) 27–34
© Neue Schweizerische Chemische Gesellschaft
ISSN 0009–4293

# Focal Point: Biotechnology*

## ILMAC Congress: October 14, 1999

## A. Genomics
## B. Molecular Diagnostics

Two Biotechnology Sessions were arranged by Dr. Hans G. Leuenberger on behalf of the Swiss Coordination Committee for Biotechnology and the New Swiss Chemical Society. Two topics of strong current interest were selected, namely 'Genomics' and 'Molecular Diagnostics'. The program of the Genomics session was compiled and chaired by Prof. Peter Philippsen (Biozentrum Basel) and the session on Molecular Diagnostics by Prof. Charles Weissmann (Imperial College, School of Medicine at St. Mary's Hospital, London). Around 300 scientists participated in both sessions.

**Keywords:** Biotechnology · Genomics · ILMAC · Molecular diagnostics

### A. Genomics (Chairman Prof. P. Philippsen)

Genomics provides a new and important information basis for applications in biology, medicine, and agriculture. Complete genome sequences of microorganisms and whole or partial sequences of thousands of genes from complex organisms are stored in public and private databases. New experimental and computational approaches have been developed for the functional analysis of single genes (including disease genes), related groups of genes, or all the genes of a genome.
Technologies and recent applications in this field have been described by five distinguished scientists:
Modern technologies such as signature tagged deletions or DNA chip technology and their application to functional gene analysis were presented (E. Winzeler and J. Hoheisel). Relationships between human disease genes and genes found in a model organism (*e.g.* yeast) can help to understand human biology and disease processes (P. Hieter). Strategies for identification and characterization of novel tumor suppressor genes or oncogenes related to breast or prostate tumors were discussed (A. Rosenthal). Applications of genomic technologies to crop improvement and protection were the topic of the final presentation (A. Binder). Abstracts by the authors are given below.

### Functional Analysis of the Yeast Genome

*Elizabeth A. Winzeler*
Novartis Institute for Functional Genomics (NIFG), San Diego, CA, USA

The past year has seen an exponential increase in the number of genomes completely sequenced relative to the previous year. Hundreds of new open reading frames are being identified each day. However, the function of the proteins encoded by many of these open reading frames remains unclear. Even in the simple eukaryote, *Saccharomyces cerevisiae*, which has been intensively studied for decades, one-third of the open reading frames have no known function and exhibit no homology to other well-characterized genes. Assigning a cellular role to the proteins encoded by these uncharacterized open reading frames is the goal of the post-genome era. To aid in the functional analysis of the genome, mutational analysis is being performed on the entire yeast genome. All of the 6200 annotated open reading frames will be serially deleted from the yeast genome and replaced with a drug-resistance cassette using a common PCR targeting approach [1]. These deletion strains will be constructed in an isogenic background by an international consortium of laboratories and will be distributed to the larger scientific community. During the construction of the deletion strain, a unique 20-base tag is incorporated into the genome of the mutant cell. These tags allow the individual strains to be tracked in a mixed population and the phenotypes of the deletion strains to be scored in parallel. Strains whose growth is inhibited by different conditions or compounds can be rapidly and quantitatively identified. Currently over 16 000 yeast deletion strains for over 5000 different open reading frames have been generated and confirmed. We have recently begun analyzing the phenotypes of these strains in parallel. Information about the project is given in the following web site: *http://sequence-www.stanford.edu/group/yeast_deletion_project/deletions3.html*
In addition, we are using high-density oligonucleotide arrays to add functional information to the genome by analyzing gene expression [2], studying recombination [3], and characterizing chromosomal origins of replication on a genome-wide scale.

*Organized by:
Dr. Hans G. Leuenberger
F. Hoffmann-La Roche Ltd.,
PRPB
CH–4070 Basel
Tel.: +41 61 688 45 61
Fax: +41 61 688 16 73
E-Mail: hans_g.leuenberger@roche.com

[1] E.A. Winzeler, D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. Boeke, H. Bussey, A. Chu, C. Connelly, K. Davis, F. Dietrich, S. Dow, B. El, M.F. Foury, S. Friend, E. Gentalen, G. Giaever, J. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. Revuelta, L. Riles, C. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. Storms, S. Veronneau, M. Voet, G. Volckaert, T. Ward, R. Wysocki, G. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston, R. Davis, *Science* **1999**, *285*(5429), 901–906.

## Human Disease Genes and their Model Organism Homologues

*Philip Hieter*
Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada

Knowledge of human protein function will ultimately aid our ability to understand human disease mechanisms in terms of specific genes and to design rational drug therapy and patient management strategies. However, knowledge of a gene sequence in and of itself, sheds no light on the function of the corresponding encoded protein or the physiological process affected in the disease state. It is this gene function information that is relevant to the development of effective ways of treating and caring for patients. So how will the functions and interactions of the proteins encoded by human genes mutated in disease be elucidated? Much of the answer lies in so-called 'model' organisms: bacteria, yeast, nematodes, fruit flies, and mice.

Links between human disease genes and yeast genes can be used to accelerate the identification of the disease gene (see below) and also provides immediate protein functional data and an experimental paradigm for further analysis [1]. The value of having a complete genome sequence (and corresponding gene repertoire) is great because a similarity search will reveal the best match to a yeast gene with absolute certainty [2]. Thus, weak matches are far more convincing in the context of knowing the entire yeast gene repertoire, and therefore justify assignment of biological function to the human protein with greater confidence. These connections help us better understand the molecular basis for human disease. It is important to emphasize that the link to a yeast gene may also provide a link to an entire genetic pathway, a multi-protein complex, or interacting gene products. These additional links can be profound. In identifying potential drug targets for therapeutic intervention, there may be cases in which targeting proteins upstream or downstream in a biochemical pathway or proteins that physically or genetically interact with the mutant product may offer the best strategy. For example, yeast genetic methods can be used to screen for second site mutations that suppress the original mutation. The corresponding proteins could provide excellent targets for drug development.

The discovery of a relationship between a human disease gene and a gene found in yeast is clearly advantageous, but how can these connections be established in a systematic fashion? Three general paradigms have emerged [3]:

1) Human/yeast similarity searching. When a human gene mutated in a disease is identified (for example, *via* a positional cloning approach), the predicted protein sequence can be used to search for related proteins in yeast. Discovery of sequence similarity to a yeast protein allows incorporation of any biological information known about the yeast gene to an understanding of the disease phenotype and human gene function. In addition, the link to a yeast gene also provides an experimental paradigm for further analysis of the biological process involved.
2) 'Genome cross-referencing' is a method that exploits the sequence similarity relationships between yeast and human genes to accelerate the cloning of human disease genes. The method is essentially a cross-species positional candidate gene approach. Yeast protein sequences are used to identify related human cDNAs by searching the database of Expressed Sequence Tags (dbEST). Corresponding cDNAs are then mapped to human chromosomes, effectively annotating the human genetic map with the biology of the related protein from yeast. These loci offer positional candidates for disease phenotypes mapping to the same location, which can be prioritized based on an understanding of the disease phenotype and yeast biology attached to the mapped human cDNAs.
3) 'Homology probing' is a technique that identifies cross-species candidate genes for a disease phenotype of interest in the absence of any mapping data. In this procedure, a set of criteria is used to identify a set of yeast genes, which are then used to search dbEST for related human cDNAs. For example, a disease phenotype might suggest a particular pathway, organelle, or gene characteristic that could potentially be associated with the disease gene. Identification of a corresponding set of genes in yeast offers a set of protein sequence queries to be used in a dbEST search.

Establishing links between model organisms and human disease as early as possible will greatly accelerate our ability to understand human biology and disease processes. As additional eukaryotic genomes are completely sequenced, multi-organismal approaches for understanding human biology will be even more powerful. These organisms will provide complementary experimental approaches and aspects of eukaryotic biology not approachable in yeast.

[1] P. Hieter, D.E. Bassett Jr., D. Valle, *Nat. Genet.* **1996**, *13*, 253–255.
[2] D.E. Bassett Jr., M.S. Boguski, F. Spencer, R. Reeves, S. Kim, T. Weaver, P. Hieter, *Nat. Genet.* **1997**, *15*(4), 339–344.
[3] P. Hieter, *Clinical Genetics* **1998**, *54*, 113–116.

## Systematic Identification and Characterization of Tumor-associated Genes in Sporadic Forms of Cancer

Edgar Dahl, Christian Pilarsky, Bernd Hinzmann, Armin O. Schmitt, Thomas Specht, Birgit Weber, Matthias Choschzick, Ruprecht Kuner, Christoph Wissmann, Eike Staub, Stefan Röpcke, and *André Rosenthal*
metaGen Gesellschaft für Genomforschung mbH, Berlin, Germany

Breast and prostate tumors are the most frequently diagnosed malignancies in western countries and a leading cause of death. Breast cancer affects one in ten women and prostate cancer affects one in five men in the USA. There are approx. 180000 new cases of breast and prostate cancer each year in the USA. In Germany, approx. 40000 new cases are reported each year. Fifteen percent of all cancer deaths are due to prostate cancer. The early stages of the tumors can be treated surgically resulting in a cure rate of 70% for breast and 90% for prostate cancer. In the later stages, these rates decrease significantly and there is no cure for these cases. Hormonal treatment of these tumors extends the life span of the patients but is only palliative at best. Therefore, there is an urgent need for new targets for a curative treatment of these tumors.

Cancer is a multifactorial disease caused by genetic and environmental factors. Ninety-five percent of all cancer cases represent sporadic forms and are due to somatic mutations. The molecular mechanisms of the development, progression, and metastasis of solid tumors are largely unknown. Approximately 400 to 500 human genes may be involved in these pathological processes. Most of them have not yet been identified. metaGen has developed a new, integrated strategy for the characterization of novel tumor suppressor genes (TSG) and oncogenes involved in tu-

morgenesis. An important element of this strategy is the systematic mining of DNA sequence databases including EST databases. Using sophisticated bioinformatics tools, potential genes including their expression profiles can be easily identified *in silico*. Another important element of our strategy is the use of highly parallel chip-based approaches for RNA expression profiling. Chip-based transcript profiling allows the identification of signal transduction pathways using tumor tissues of different stages and grading. Another key step of our strategy is laser-supported microdissection of tumor material. High-throughput mutation analysis and *in vitro* knockout approaches of candidate genes down-regulated or up-regulated in tumors are used to further validate candidate genes.
Using the 4.7 million ESTs available in public and proprietary EST databases, we have identified more than 2000 genes – half of which are of unknown function – which are differentially expressed in six solid tumors including breast and prostate cancers with regard to the corresponding normal tissues. For most genes chromosomal locations have been assigned using radiation hybrid mapping and ePCR. The combination of comparative genomic hybridization (CGH) and loss of heterozygosity (LOH) data with our panel of differentially expressed genes allowed us to suggest a number of candidate genes from their location in chromosome regions showing significant LOH or amplifications. Candidate TSGs are subjected to high-throughput mutational analysis by multiplex fluorescent SSCP using DNA isolated from microdissected tumor samples. Candidate oncogenes are further validated by *in vitro* knockout experiments using antisense technology in tumor cell lines. In order to identify gene networks and signaling pathways, a special cancer chip has been developed for RNA expression profiling. This cancer chip is based on Affymetrix technology and contains more than 4000 genes, including our set of 2000 differentially expressed candidate genes. For chip hybridization, RNA isolated from well-characterized microdissected tumor samples is used. Expression profiles representing different tumor stages and grades are compared with profiles of the corresponding normal tissue and clustered. A network of gynecologists, urologists, and pathologists has been established throughout Germany to provide clinical expertise and tumor samples.
Using this strategy, metaGen expects to isolate a number of new TSGs and oncogenes and to correlate mutation events and RNA-expression profile with tumorgenesis. The obtained insights into the genetics and pathology of solid tumors will improve diagnostic procedures and will facilitate the development of new targets for the therapy of all stages of breast and prostate cancer.

### DNA Microchips; Transcriptional Profiling and Beyond

Marcus Beier[a], Stefan Matysiak[a], Nicole Hauser[a], Marcel Scheideler[a], Simone Wurtz[a], Kurt Fellenberg[b], Martin Vingron[b], and *Joerg D. Hoheisel*[a]
[a] Functional Genome Analysis and [b] Theoretical Bioinformatics
Deutsches Krebsforschungszentrum, Heidelberg, Germany

DNA-chip technology is one of the basic tools for a more comprehensive analysis of function on a genomic scale. For nearly every aspect of analysis at the nucleic acid level, there is a suitable adaptation of microarray techniques, which is either in place, under development, or at least feasible. Nevertheless, even for such a comparatively widely used method as transcriptional profiling, which is based on many years of experience from macro-scaled techniques or related methods such as the Southern blot, the methodology is not yet refined enough to allow simple assaying.
Our technical developments are not concerned with the engineering problems, which are not yet entirely solved either, but deal with the chemical and biophysical aspects. Imperative for a sensible use of DNA-chips is the establishment of standardized procedures and processes for comparable and compatible data production, quality control, and analysis. To this end, we have established a surface chemistry for the covalent attachment by both immobilization and *in situ* synthesis of DNA molecules on glass or polypropylene. Apart from stable bonding during more than 30 consecutive hybridizations, parameters such as loading capacity, charge, and hydrophobicity can be modified in a controlled manner.
For photolithographic *in situ* synthesis of oligonucleotide arrays, yield was a major aspect that restricted quite a few applications because of shortcomings in quality. In our laboratory, a chemistry was developed that yields quantitative efficiency of each condensation reaction, thereby increasing the overall yield of 20-mer oligonucleotides by more than one order of magnitude, for example. In addition, a procedure was set-up that allows for an accurate measurement of yield and quality at all individual positions of every oligomer chip prior to subsequent uses of the actual chip. This is made possible by adding a dye molecule during synthesis which can be removed, after assessing the quality, without leaving behind chemical modifications that affect hybridization to the oligonucleotides.
Substitution of peptide nucleic acid (PNA) for DNA molecules on the arrays combines the high selectivity of oligomers with strong duplex stability at hybridization conditions which disfavor double-strand formation between the hybridized DNA targets. Because PNA is an uncharged molecule, no ions are needed in solution to counteract intermolecular repulsion, as is necessary for the formation of duplexes between two nucleic acids. Also, PNA is stable to degradation by nucleases and proteases and so can be used for analyses even in crude extracts. Lastly, because of the basic difference between PNA and nucleic acids, new detection methods are possible. In a collaboration with the University of Munster, we are pursuing the use of mass spectroscopy. Binding of nucleic acids to PNA-arrays could be traced by detecting their phosphate groups. Ideally, all phosphate groups present should be from bound nucleic acids. In addition, there would be an inherent amplification because even a single DNA molecule contains a large number of phosphates.
We apply array-based analyses to a variety of applications funded mainly by the European Commission and the German Science Ministry (BMBF). Transcriptional profiling is being carried out on all yeast genes, as part of the EUROFAN program. Similarly, a non-redundant set of 13 000 *Arabidopsis* cDNAs and PCR products of all open reading frames of *Bacillus subtilis* are being analyzed. To this end, computer algorithms have been designed and set-up to assure thorough quality control of the raw data, which is an absolute prerequisite for reasonable data interpretation. Tools for subsequent data mining are also being developed.
Combining array technology with mapping approaches, we are involved in several projects in which entire microbial genomes are being mapped on the plasmid level. Rather than to proceed into sequencing, however, the ordered but otherwise anonymous fragments are then placed on chips, by definition representing a normalized and complete cDNA-library even though gene position and function are unknown. Using transcriptional analyses of interest to the researcher, however, coding regions will be

CHIMIA 2000, 54, No. 1/2

identified that exhibit an interesting response to the conditions used for cell culturing. Only such regions will subsequently be sequenced rather than the entire genome, reducing sequencing to significant parts of the genome. The more sequence data that becomes available from the many projects worldwide, the more redundant information will result from the sequencing of an entire genome, because of similarities to existing information from other organisms. Apart from transcriptional analyses, many other functional assays could be performed prior to sequencing based on the ordered fragment library.

Identification of sequence variants or detection of contaminants, as different as viruses in human cells or bacteria in sludge samples of wastewater treatment plants, are some of the practical applications of PNA-chips. Simultaneously, the basic understanding of PNA–DNA interaction is being analyzed, information that could be critical also to very different fields such as antisense strategies.

Combining representational difference analysis with chip-based measurements, we are developing a methodology that should allow the quantitative study of transcriptional differences even between individual cells. These technical aspects, as with many of the above-mentioned, will improve the reliability of chip-based assays significantly to a degree to which the use of chips will be possible on a routine basis for many different areas including medical applications, for which quality standards have to be especially high.

### Gene Discovery in Agriculture Research

*Andres Binder*
Novartis Crop Protection AG, Basel, Switzerland

The fast progress in modern biology represents a challenge for both pharmaceutical and agriculture research. New technology platforms like genomics, expression profiling, and bioinformatics open new dimensions in the discovery of genes, their regulation and functions. The unique advantage in agriculture research is to use gene discovery for novel solutions of crop improvements by the genetic as well as the chemical route: i) genetics: engineering and breeding of new germplasm (seeds) for crop varieties with novel genetic traits; ii) chemistry: development of novel chemicals through target-based discovery and high-throughput screening to protect crops against diseases, pests, and weeds, similar to drug discovery in the pharmaceutical field. Beyond this, the technologies also open up opportunities for new innovative solutions that combine chemical and genetic methods.

Besides human health care, one of the major challenges in science today is the improvement of our crop plants for food and feed production worldwide (rice, wheat, corn, vegetables *etc.*). This can be achieved either by improving traditional input traits for optimal plant growth using genetic and chemical technologies (*e.g.* protection against pests, diseases, and weeds), or – even more important in the future – by expressing new output traits in the plant with attractive benefits for the consumer, using gene technology (*e.g.* improved feed and food quality, medical and industrial products, processing efficiency, fiber quality *etc.*). The modern way to unravel these new traits is a gene discovery approach with the help of genomic technologies.

The basis for efficient discovery and comparison of genes and regulatory elements will be the availability of DNA sequences of relevant model systems and target organisms. Unlike pharmaceuticals research, agriculture research needs to cover many diverse organisms, *i.e.* sequence information of all the important crops, pathogens, and pests is needed. In addition, modern profiling technologies on various levels (RNA, proteins, and metabolites) are important tools to define the genes which are relevant for the expression of new traits in crops or represent new targets for novel agrochemicals.

The selected genes or regulatory elements for new traits can either be introduced in crops through plant transformation (GMO: introduction of foreign genes or change of regulation of endogenous genes) or alternatively they serve in a non-GMO approach as molecular information for marker-assisted breeding; the new efficient high-throughput breeding technology that exploits natural genetic diversity. For target-based discovery of novel agrochemicals (herbicides, fungicides, insecticides), potential targets from plants, fungi, or insects are first validated by inhibiting protein expression through knocking out genes with suitable methods (depending on the organism: homologues recombination, antisense, mutation tagging *etc.*) and by doing so mimicking the inhibition of agrochemicals. Knockouts giving rise to lethal or impaired phenotypes represent validated targets and will be used to develop assays for high-throughput screening (HTS) of thousands of new chemicals in minute amounts.

In this process from gene discovery to new traits or to novel agrochemicals, many centers of excellence are involved in a complex network: academic institutes with relevant biological systems, biotechnology boutiques with specific front-end technologies, companies with fully automated high-throughput machinery and potent international life science concerns with the breadth to coordinate and funnel the products into breeding and chemical development, worldwide field testing, and marketing.

Representative literature for this area:
- B. Mazur, E. Krebbers, S. Tingey, *Science* **1999**, *285*, 372–375.
- D. DellaPenna, *Science* **1999**, *285*, 275–279.
- C. Sommerville, S. Somerville, *Science*, **1999**, *285*, 380–383.
- E. Ward, P. Bernasconi, *Nat. Biotechnol.* **1999**, *17*, 618–619.
- R. Dietrich, in *Annu. Rev. Mol. Plant Pathol.* Sheffield Press, Eds. J. Beynon, M. Dickson, in press.
- G. Stephen, *Curr. Opin. Plant Biol.* **1999**, *2*, in press.

## B. Molecular Diagnostics (Chairman Prof. C. Weissmann)

Modern molecular technologies have revolutionized diagnostics. Five distinguished speakers presented recent advances in this field.

DNA amplification by polymerase chain reaction (PCR: T. White) and the analysis of DNA hybridization with miniaturized high density arrays of oligonucleotide probes (chip technology: T. Gingeras) became powerful techniques for the detection of DNA sequences or m-RNA expression *e.g.* for the diagnosis of genetic diseases or for the identification of bacteria or viruses. Proteomics, on the other hand, deals with the separation and characterization of proteins in biological samples, affords the identification of disease specific proteins, and offers opportunities for the discovery and development of new drugs (R. Parekh). Fluorescent coincidence spectroscopy (FCS) is a powerful technique for the rapid detection of molecules at extremely low concentrations, including viruses, amyloid peptides, ligand receptor complexes, and PCR products (R. Rigler). The mapping of single-nucleotide polymorphisms (SNPs), a frequent type of variation in the human genome, is a useful tool in pharmacogenetics (K. Lindpaintner). Abstracts of the lectures are given below.