SESSION 6A

THE ROLE OF NEW TECHNOLOGIES IN HERBICIDE DISCOVERY

Chairman and Session Organiser	Dr K E Pallett *Rhône-Poulenc Agriculture Ltd, Ongar, UK*
Papers	6A-1 to 6A-4

The impact of plant genomics on herbicide discovery

P G Thomas
Zeneca Agrochemicals,
Jealott's Hill Research Station, Bracknell, Berkshire RG42 6ET UK

ABSTRACT

The complete genome sequence of the model dicot *Arabidopsis thaliana* will be determined by 2000 and that for rice early in the next decade. Plant genome programmes are driven by commercial opportunities to manipulate yield and quality, however there are potential spin off benefits for herbicide research. In addition to providing a complete inventory all possible herbicide targets, genomics will also provide experimental tools to support progression of exploratory herbicides into evaluation and development.

INTRODUCTION

The ability to analyse crop genomes has been hailed as the driving force behind a new green revolution which will significantly increase crop yield and also processing and nutritional quality through the discovery and delivery of new gene traits in transgenic crop plants (Mazur *et al.*,1999; DellaPenna,1999). For the foreseeable future, weed control will be a key factor which limits crop productivity and the development of improved herbicides will, therefore, remain a central plank of modern agriculture.

Within the first few years of the new millennium, the complete genome sequences of several plant species will be determined, almost certainly before new herbicides reported at this conference are brought to the market. Partial DNA sequence information and detailed genetic maps, which provide a starting point for complete genome sequencing, will be available for many other plant species. The purpose of this introduction to the genomics philosophy and technical capability, is to alert a "weed control" audience to the potential spin off benefits that will impact the discovery and development of new herbicides over the coming years. The discovery and validation of novel high throughput screening targets is one obvious application. Genomics may also underpin rapid determination of the mode of action of lead compounds from *in vivo* screens and provide insight into the molecular basis of selectivity and spectrum, such data could support decisions on compound progression into evaluation or development.

DEFINING GENOMES AND GENOMICS

Expressed simply, a genome is the complete genetic makeup of an organism, it comprises all of the DNA in each chromosome. Genomics has come to be understood as a collective term which describes a variety of technologies which, used in concert, allow the complete genome of an organism to be described at the level of its DNA

sequence and the function of its genes and how patterns of gene expression vary during plant growth and development and in response to stimuli.

Genomics does not provide any data which can not be obtained by more traditional approaches. Gene cloning and sequencing are routine operations and spatial and temporal patterns of gene expression can be determined easily by Northern blot or PCR analysis. Similarly mutant plants with interesting phenotypes can be isolated and the affected genes analysed in detail. However, this traditional approach undertakes the thorough analysis of only one or a few genes at a time en route to testing a single hypothesis. The philosophy of "genomics" is about undertaking quite routine experimental procedures, but executing them on a massive and hugely parallel scale with efficient automation and data acquisition, storage and processing. Consequently, a genomics laboratory focuses on maximising process efficiency and has much in common with a factory production line (Figure 1).

Figure 1. High volume DNA sequencing facility. Photograph courtesy of Washington University School of Medicine.

Genomics seeks to provide a more or less complete, integrated and multi-purpose data set containing the answers to a multitude of questions yet to be asked. The resulting data can be repeatedly analysed, or mined, using bioinformatics to provide the specific information necessary to address each hypotheses in turn.

The outputs of plant genome programmes are typically databases of DNA sequences and associated information; physical collections of cloned genes; ordered arrays of plant gene fragments; integrated physical and genetic maps; populations of mutated plants; all of which may be repeatedly reused and reanalysed to address a wide variety of fundamental and applied questions in plant biology.

GENE DISCOVERY DRIVEN BY LARGE SCALE DNA SEQUENCING

Very large scale DNA sequencing is the dominant approach to gene discovery. This may be focused at the complete description of the genome or to derive partial DNA sequence reflecting only those genes that are being actively expressed. A complete genome sequence provides data on every gene in its entirety, including exons, introns, promoter, enhancers and any other DNA sequence involved in the regulation of gene expression and in addition the DNA between genes.

The initial step is to construct a complete plant gene library in vectors able to contain extremely large inserts, for example bacterial artificial chromosomes (BACs, Shizuya *et al.*, 1992)) which can carry cloned fragments up to 300,000 bases. Despite the efficiency gains afforded by very large scale sequencing laboratories it remains an expensive procedure. To avoid waste by sequencing multiple overlapping BAC clones, a reduced set of clones known as the minimum tiling path may be identified which still ensures sufficient overlap between adjacent BAC clones. The large DNA inserts from individual BAC clones are broken into smaller fragments and sequenced at random until the complete BAC sequence can be assembled by computer. The sequence of the chromosome is eventually derived by assembling the sequences of all the overlapping BAC clones. Any remaining gaps have to be filled by targeted isolation and sequencing of clones from that region.

Whilst the complete genome sequence does provide the definitive description of a plants genetic makeup, there remains a problem that frequently it is not possible to determine where the genes are in any particular segment of DNA. This is because some of the features of genes are hard to recognise and the fact that much of the data obtained reflects intergenic DNA.

Data relating only to genes being actively expressed is provided by expressed sequence tags (ESTs). ESTs are derived by isolating messenger RNA from a specific tissue sample, cDNA is generated by reverse transcription and subsequently cloned. Individual clones are selected at random and sequenced. A sacrifice is frequently made which trades off short runs of sequence data, only 200-400 bases which may also contain errors, against the high volume of clones put through the process. All of the sequences derived are subjected to computer analysis to determine whether the sequences reflect novel genes or are already known. ESTs are exploited directly to assemble gene arrays (see below) and are of considerable value in identifying protein encoding genes may otherwise be invisible within genomic sequence.

Messenger RNA abundance may vary over several orders of magnitude, so to ensure that sequencing effort is not dissipated by reanalysis of the abundantly expressed genes, it is becoming common practise to normalise cDNA abundance by subtractive hybridisation (Soares *et al.*, 1988). In this way, the probability of finding and sequencing clones representing rare mRNAs is significantly increased.

The principle underlying this approach is driving for rapid discovery of novel genes, and this can be achieved without complete accuracy. The cDNA clones from which the EST data was derived are typically used as gene probes to isolate the full length

version of the gene. This is when accurate DNA sequencing is employed since the data may well be the subject of a patent application.

Whether from full genome or EST programmes, the resulting inventory of complete or partial gene sequences builds on the existing public sequence databases (Windass *et al.*, 1992) as an obvious source of known or prospective herbicide targets. Access to sequence data or physical samples of a given clone facilitates rapid target validation through gene silencing and the supply of the target enzyme as a reagents for high throughput *in vitro* screening.

Novel and previously unknown herbicide targets will also be represented in the data set, however, the challenge is identifying them from the many non-targets. Comparative and functional genomics may offer some approaches to discovery of some or all of the novel targets that we suspect exist.

PLANT GENOME SEQUENCING PROGRAMMES

Genomes are large and their analysis requires considerable organisation of resources. In 1998 the National Science Foundation in the US announced a co-ordinated programme on plant genome research with funding of about $85 million (Walbot, 1999). In addition many agbiotech and crop protection businesses are making substantial investments.

The first fully sequenced plant genome will be that of the model dicot plant *Arabidopsis thaliana*. For the first time herbicide scientists will have access to a complete inventory of all possible targets. The Arabidopsis Genome Initiative (AGI) is a multinational consortium of laboratories sharing the task of sequencing all 5 chromosomes (approximately 120,000,000 base pairs). Sequencing was predicted to take 6 years but continuous improvement in the technology available has brought completion forward from 2004 to perhaps as soon as 2000.

Rice will be the first cereal crop genome to be sequenced. From 1991 to 1998 the Rice Genome Program (RGP) in Japan has provided a large EST database and mapped BAC clones as the foundation which will underpin work to determine the complete genome sequence of rice. Completion of the 450,000,000 base pair genome is estimated to take a further 6-7 years. The cereal genomes exhibit a high degree of synteny, meaning that not only are the sequences of the genes very similar but also that the order of the genes along large segments of the chromosome are also conserved (Devos and Gale, 1997). This has huge significance for crop biotechnology since the implication is that only a few anchor genome sequences are required to be informative for all cereal crops.

Genome sequencing programmes on corn and soybeans are in the process of being launched. The National Corn Genome Initiative, supported by the National Corn Growers Association in the US, has just initiated the ground work for a full maize sequencing programme. A soybean EST sequencing programme has also been implemented at the Washington University School of Medicine Genome Sequencing Centre.

For herbicide research, the *Arabidopsis* and rice genomes will provide all possible herbicide targets in broad leaf and grass weeds. Further data on plant genomes may provide insight into the basis of herbicide selectivity through better understanding of herbicide metabolising systems in crops and weed species. On completion of the human genome sequence, which is on an ever shortening timescale, we can anticipate that surplus DNA sequencing capacity will be released to accelerate crop genome sequencing programmes.

BIOINFORMATICS

Bioinformatics is more than simply the information technology which supports genomics. It is an *in silico* science which involves the capture, storage, retrieval, processing, analysis, visualisation and interpretation of gene sequence data and associated information (for example which BAC clone sequence data derives from) and also the relationship of this data and information to: genetics; biochemistry; molecular biology and biological function. Bioinformatics provides the critical computing tools to maintain the huge databases of information generated by genomics activities and to "mine" knowledge from them. It is our window on genomics.

A critical capability that bioinformatics provides is the ability to recognise similarities between genes and their encoded protein products. Such sequence relationships are found by comparing each newly identified gene with the entire database of pre-existing genes. Where sequence similarities are found this may imply related function or properties and is the basis of assigning putative functions to newly identified genes.

GLOBAL ANALYSIS OF GENE EXPRESSION PATTERNS

Access to large collections of cloned plant genes allows the fabrication of gene arrays or micro-arrays comprising of spots of DNA immobilised onto a nylon or glass support. Spatial or temporal expression patterns of all of the arrayed genes may be analysed by probing the array with fluorescent or radio-labelled DNA generated from mRNA isolated from plant tissue during development or following exposure to different conditions or treatments. Figure 2 illustrates how mRNA, isolated from reference and test samples of plant tissue, can be differentially labelled. The labelled samples are pooled and hybridised to a high density gene expression micro-array (GEM) comprising up to 10,000 unique genes. After hybridisation and washing, the amount of each labelled probe remaining specifically hybridised to each DNA spot is determined scanning confocal fluorescence microscopy and image analysis. The output for each arrayed gene reports whether it is expressed or not in the test tissue, and whether the level of expression is increased or decreased in response to the experimental treatment.

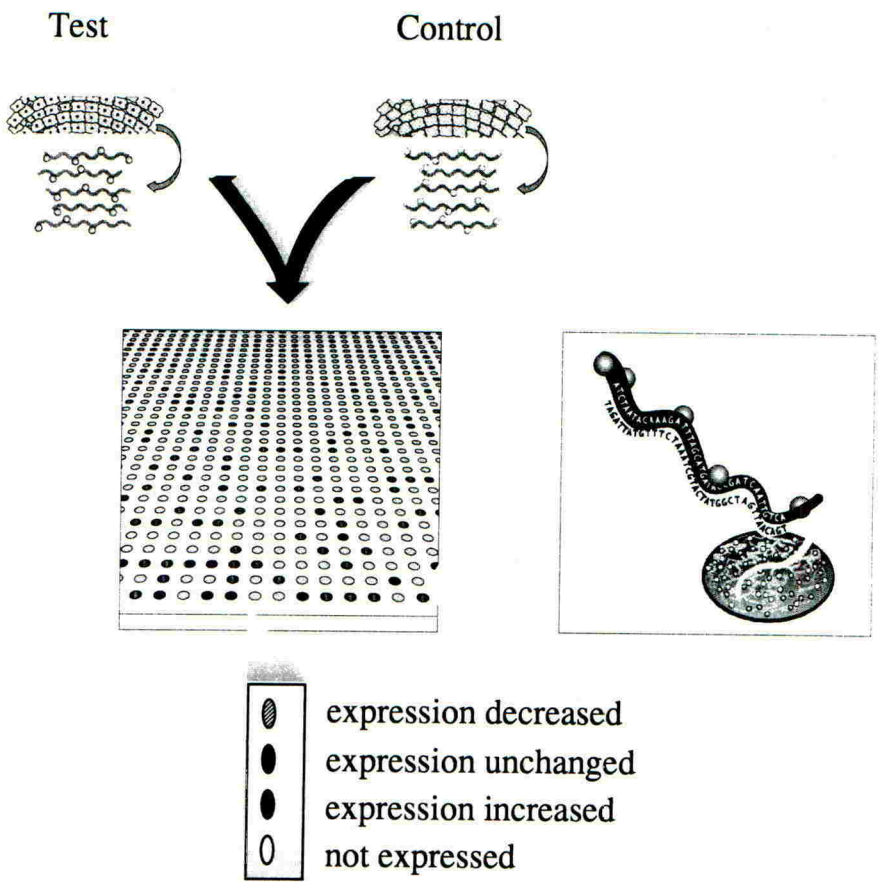


Figure 2. Schematic of a high density gene-expression microarray experiment. Differentially labelled fluorescent probes from test and control tissues are pooled and hybridised to the array and the proportion of each probe bound determined. Every arrayed gene can be analysed quantitatively for expression in the target tissue and changes in expression resulting from experimental treatment. Courtesy of Incyte Pharmaceuticals Inc.

The utility of such DNA arrays is shown in a landmark experiment by Wodicka *et al.*, (1997) who undertook a "genome wide" experiment to monitor the effect of growth on either minimal or rich media on gene expression in *Saccharomyces cerevisiae*. In this instance the arrayed DNA fragments comprised oligonucleotides synthesised *in situ* on a glass support representing more than 6200 of the known yeast genes predicted from the complete genome sequence. The array was hybridised with modified DNA probes generated from mRNA isolated from culture grown in either minimal or rich media. Many of the genes were expressed under both growth conditions however a small number showed dramatic changes in expression levels. Thirty six genes were expressed more strongly (more than 5 fold) in rich media

compared with over 140 genes which were more strongly expressed (more than 5 fold) in minimal media. The authors noted that some of the genes exhibiting the greatest changes in expression level are of unknown function.

Similarly, Desprez *et al.*, (1998) used macro-arrays consisting of over 800 *Arabidopsis thaliana* cDNAs spotted on to nylon membranes to investigate changes in their expression levels in light or dark grown plants. Messenger RNA was isolated from either seedlings grown under a 16 hour light/8 hour dark photoperiod or etiolated seedlings and used to generate radio-labelled DNA probes to hybridise with the arrays. The resulting data indicated that about 15% of the arrayed genes showed significant changes in their levels of expression during growth in light or dark. The shift from light to dark has complex effects on plant physiology and it is hardly a surprise that this is reflected in patterns of gene expression.

Both of these experiments pose an interesting question, might changes in patterns of gene expression following a more simple treatment, for example with a herbicide, yield information relating the mode of action of the herbicidal compound?. This concept is already being evaluated in pharmaceutical research where DNA micro-array experiments are being used to investigate "signature" patterns of altered gene expression to confirm that drugs actually inhibit the intended target *in vivo* and also to check for undesirable secondary effects (Marton *et al.*, 1998). Given the anticipated completion of the *Arabidopsis* genome before the end of 2000 it is not impossible that such array experiments will become a routine tool in herbicide research within a few years. The mode of action of *in vitro* or *in vivo* screen leads could be either confirmed or elucidated by interrogating plant gene arrays with probes derived from herbicide treated and control plants. Cross-regulation of metabolic pathways in plants (Guyer *et al.*, 1995) presents a challenge in that potentially informative changes in expression of genes from the inhibited pathway will have to be isolated from changes resulting from non-specific stress responses (Zhao *et al.*, 1998)

ASSIGNING FUNCTIONS TO GENES: FUNCTIONAL GENOMICS

Bioinformatics may allow the function to be predicted of a proportion of new genes through identifying similarities to genes from other organisms where function has been established. A significant proportion of the genes identified through plant genome sequencing are expected to be completely new to science, their DNA or encoded protein sequences providing no clues to their function. The phenotypic effect of mutations which interfere with the natural biochemical and physiological function of the gene product may allow the role of many such "orphan" genes to be determined.

Functional genomics seeks to provide indications of the role of newly discovered genes via large-scale, systematic, genome wide mutation programmes. Segments of specialised DNA can be driven into or near genes so disrupting their normal function. Naturally occurring mobile DNA elements, transposons, are known to disrupt gene function and cause mutations (Martienssen, 1998). Similarly, genes may be disrupted by insertion of T-DNA which is a commonly used tool in plant genetic modification (Feldman, 1991).

The phenotype exhibited by the plant may provide the data necessary to indicate gene function. It is important to recognise that further experimental data will often be required to confirm gene function.

Transposons have been modified to provide a convenient tool with which to construct large populations of mutated plants wherein each individual transgenic line has a single gene "tagged" with a transposon to either completely block gene function or increase gene expression to amplify its function. A "knock out" or "knock in" effect depends on the specific genetic components of the system and whether the tagged gene is heterozygous or homozygous may also affect the severity of any phenotype. The identity of each tagged gene can be readily determined by sequencing the flanking DNA adjacent to the tag and referring back to the genome sequence. Expression of reporter genes carried within the tag may also allow genes which only function in specific cell types or at a specific developmental stage to be detected even where there may be no other obvious phenotype (Martienssen, 1998).

Detailed experimental analysis of each plant line may eventually provide sufficient information to uniquely and unambiguously assign a function to the tagged gene. Routine analyses might include morphology; alterations in the profile of common metabolites; changes in protein abundance from 2D protein gels or investigating patterns of reporter gene expression. Integration of such data with genomic DNA sequence data will provide a route to associating a biochemical or physiological role to each gene. Further indirect data may be derived from the patterns of temporal and spatial gene expression patterns derived from array experiments, whereby the function of a gene may be associated with a specific tissue, developmental process or in response to specific biotic or abiotic stimuli. That the assignment of function to genes may prove to be a significantly more challenging problem than determining their sequences is widely recognised.

It is attractive to speculate how new herbicide targets for high throughput *in vitro* screening might be identified by screening gene "knock out" populations of plants for individuals showing extreme symptoms of phytotoxicity or lethality. The identity of the disrupted gene would be determined by sequencing DNA flanking the "tag" and referring back to the sequence databases. Given the difficulty of delivering complete inhibition of any herbicide target, the consequences of partial inhibition of any candidate target should be assessed by gene silencing experiments before implementation of any high throughput *in vitro* screen.

Over-expression of a herbicide target gene frequently results in some tolerance to the herbicides active at that site (Eckes *et al.*, 1989, Ichinose *et al.*, 1995). Populations of "knock in" mutants may provide a useful source of transgenic plants, or cell lines, which over express an herbicide target and are therefore resistant to and hence report specific modes of action. These reporter lines, selected by screening mutated populations, could be conveniently incorporated into *in vivo* screening cascades and may have particular utility in the early exclusion of compounds with an unwanted mode of action.

SUMMARY

By the end of 2000 the complete genome sequence of *Arabidopsis thaliana* is expected to be available and significant progress will have been made towards sequencing the rice genome. Other crop genome programmes will be underway. Access to plant genomic and EST sequences; gene array tools for analysis of gene expression and large collections of mutant transgenic plant lines, will provide significant spin of benefits to herbicide discovery and development.

For the first time herbicide researchers will have access to the complete inventory of all possible herbicide targets. Functional genomics will allow some of them to be recognised amongst the many non-targets by evaluating the consequences of gene knock out or partial silencing on plant viability. Herbicide tolerant reporter lines may also be identified by screening large populations of "knock in" mutants. Such reporters might be exploited in screening cascades for the early diagnosis of specific modes of action. Analysis of patterns of gene expression using high density DNA arrays may be informative for herbicide mode of action and the molecular basis of herbicide selectivity if specific changes in gene expression can be discriminated from non-specific stress responses.

REFERENCES

DellaPenna D (1999) Nutritional genomics: Manipulating plant micronutrients to improve human health. *Science* **285**, 375-379

Desprez T; Amselem J; Caboche M; Hofte H (1998) Differential gene expression in Arabidopsis monitored using cDNA arrays. *The Plant Journal* **14**, 643-652

Devos K M and Gale M D (1997) Comparative genetics in the grasses. *Plant Molecular Biology* **35**, 3-15

Eckes P; Schmitt P; Daub W; Wengenmayer F (1989) Overproduction of alfalfa glutamine synthetase in transgenic tobacco plants. *Mol. Gen. Genet.* **217**, 263-268

Feldman K A (1991) T-DNA insertion mutagenesis in Arabidopsis: mutational spectrum. *The Plant Journal* **1**, 71-82

Guyer D; Patton D; Ward E (1995) Evidence for cross-pathway regulation of metabolic gene expression in plants. *Proc. Natl. Acad. Sci USA* **92**, 4997-5000

Ichinose K; Fang-Sik C; Kimura Y; Matsunobu A; Sato F; Yoshida S (1995) Selection and characterisation of protoporphyrinogen oxidase inhibiting herbicide (S23142) resistant photomixotrophic cultured cells of *Nicotiana tabacum*. *J. Plant Physiol.* **146**, 693-698

Martienssen R A (1998) Functional genomics: probing plant gene function and expression with transposons. *Proc. Natl. Acad. Sci. USA* **95**, 2021-2026

Marton M J; DeRisi J L; Bennett H A; Iyer V R; Meyer M R; Roberts C J; Stoughton R; Burchard J; Slade D; Dai H; Bassett D E Jr.; Hartwell L H; Brown P O; Friend S H (1998) Drug target validation and identification of secondary drug targets effects using DNA microarrays. *Nature Medicine* **4**, 1293-1301

Mazur B; Krebbers E; Tingey S (1999) Gene discovery and product development for grain quality traits. *Science* **285**, 372-375

Shizuya H; Birren B; Kim U-J; Mancino V; Slepak T; Tachiiri Y; Simon M (1992) Cloning and stable maintenance of 300 kilobase pair fragments of human DNA in *Escherichia coli* using an F-factor based vector *Proc. Natl. Acad. Sci. USA* **89**, 8794-8797

Soares M B; Bonaldo M F; Jelene P; Su L; Lawton L; Esftratiadis A (1994) Construction and characterisation of a normalised cDNA library. *Proc. Natl. Acad. Sci. USA* **91**, 9228-9232

Walbot V (1999) Genes, genomes, genomics. What can plant biologists expect from the 1998 National Science Foundation plant genome research programme. *Plant Physiology* **119**, 1151-1155

Windass J D; Renwick A R; Thomas P G (1993) Definition of novel pesticide targets and modes of action by molecular genetic approaches. In: *Opportunities for molecular biology in crop production.* BCPC Monograph 55, ed Beadle D J; Bishop D H L; Copping L G; Dixon G K; Holloman D W.

Wodicka L; Dong H; Mittmann M; Ho M-H; Lockhart D (1997). Genome-wide expression monitoring in Saccharomyces cerevisiae. *Nature Biotechnology* **15**, 1359-1367

Zhao J; Williams C C; Last R L (1998) Induction of Arabidopsis tryptophan pathway enzymes and camalexin by amino acid starvation, oxidative stress and an abiotic elicitor. *The Plant Cell* **10**, 359-370

From genes to targets: impact of functional genomics on herbicide discovery

D Berg, K Tietjen, D Wollweber and R Hain
Bayer AG, Agrochemical Division, Research, 51368 Leverkusen, Germany

ABSTRACT

Functional genomics projects in worldwide efforts currently generate full genome and expressed sequence tag (EST) sequence information from several plants, headed by *Arabidopsis thaliana*. Sequence information and functional analysis of genes deliver possible new herbicide targets that have to be validated by knock-out mutation or downregulation, e.g. antisense strategies. Implementation of innovative, target directed assays into HTS/UHTS-systems will result in the discovery of new lead structures. As only a limited number of modes of actions of herbicides is realized until now, a comprehensive knowledge about gene function will provide innovative targets and with the help of biochemistry highly sensitive test systems for screening of diverse chemical libraries. Creating highly diverse libraries and screening these against new targets whenever a new assay system is developed, implies a basic change in screening philosophy. As an outcome a series of new basic structures is to be expected, leading to a portfolio of chemically and mechanistically diverse products.

INTRODUCTION

Or, why do we need hts (high-throughput-screening) in herbicide discovery?

Until recently research in pharmaceutical and agricultural industries, particularly in screening new active compounds followed basically different approaches. As direct screening at patients was not possible and animal models became more and more discussed by the public, pharmaceutical industry was forced to use mechanistic *in vitro* assays for preselection of development candidates. In addition high-throughput-tests enhanced the number of testable compounds.

In agricultural research, formerly there was no obvious need for preselection as greenhouse screening capacity seemed to be sufficient and – even more important – the living plant, fungus or insect contained the whole battery of possible targets, a so called comprehensive target library. What should then justify an investment into target-directed *in vitro* screens?

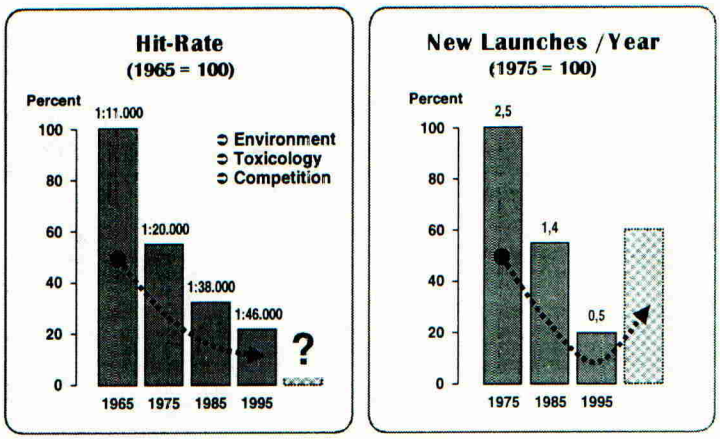


Figure 1: Statistical success rates in agrochemical discovery

Lower statistical success rates (Figure 1), lack of chemical diversity, and increased importance of early information on modes of actions finally caused agricultural chemical industry to implement HTS/UHTS (ultra high throughput screening) systems as additional screening tools (Ridley *et al.* 1998, Ward and Bernasconi 1999). Thereby even more difficult accessible compounds could be screened, enhancing chemical diversity by exchanging pharmaceutical and agricultural chemical libraries between both branches. At the same time projects on plant, insect, nematode, and fungal genomes were initiated adding the neccessary technological basis for this approach. Functional genomics provides the basis for directed research in both, plant breeding as well as herbicide research.

- **Photosystem II**
- **Photosystem I**
- **Acetolactate Synthase**
- **EPSP Synthase**
- **Glutamine Synthase**
- **"Auxin Receptor"**
- **"Auxin Transporter"**
- **AcetylCoA Carboxylase**
- **Tubulin Assembly**
- **Tubulin Organization**
- **Protoporphyrinogen Oxidase**
- **Cellulose Biosynthesis**
- **Phytoene Desaturase**
- **ζ-Carotene Desaturase**
- **Dihydropteroate Synthase**
- **4-OH-Phenylpyruvate Dioxygenase**
- **"Gibberellin Biosynthesis"**
- **Uncouplers**
- **Unknowns**

Table 1: ≈ 19 targets for herbicides in the market

ESTABLISHED HERBICIDES MODES OF ACTIONS

The list of known modes of actions of agricultural products is surprisingly short. In general 15 to 20 established modes of actions per indication represent the actual target portfolio. Table 1 lists the most prominent targets of herbicides. The major cellular components for herbicide action in a plant cell are shown in figure 2. Most of the current herbicde targets are present in the plant chloroplast (Figure 3) and only a few targets are localized outside the chloroplasts (Figure 4).

However, at least 33 additional targets have been validated experimentally by inhibitors, although these inhibitors could not yet be developed into marketable herbicides (Table 2). Furthermore, until now at least 26 additional potential herbicide targets have been patented (Table 3). Obviously the attack on different targets results in quite different levels of herbicidal efficacy in plants. In this respect the "gold targets" have to be discriminated from the less effective ones.

But, in plants where at least 30,000 functional genes are present, one can expect a much higher number of genes coding for proteins which are not only essential for viability of the plant, but also are suited as good herbicide targets. These genes can be identified by gene knock-out. Where the respective knock-out of a gene is lethal, it can be expected that the inhibition of the protein function by a chemical will also be lethal. In particular the genes coding for druggable proteins, enzymes, receptors or ion channels, are of importance here. As a rough estimate these make up 10 to 20 % of the genes. Therefore, we roughly estimate the existence of up to 3,000 further possible interesting targets for herbicide discovery.

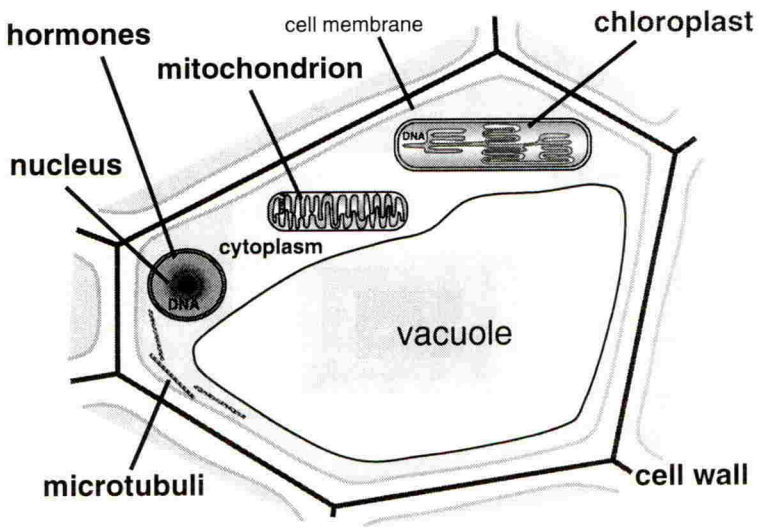


Figure 2: The major cellular components of a plant cell

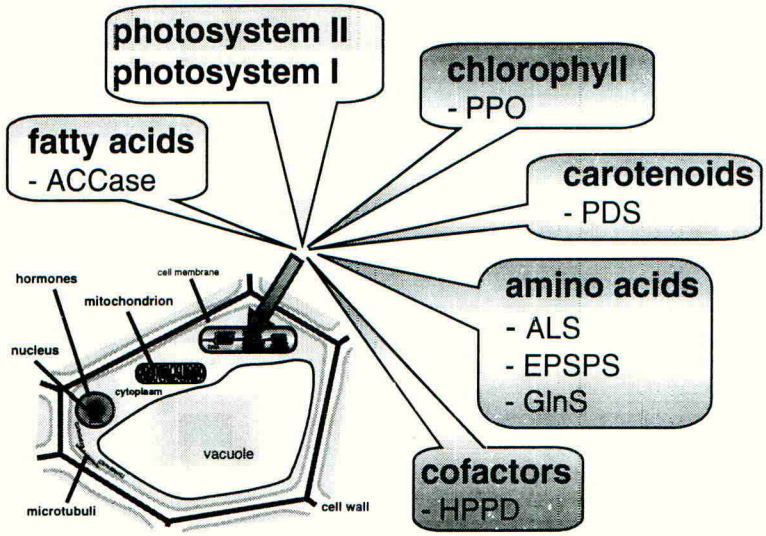


Figure 3: The herbicide targets localized in the chloroplast

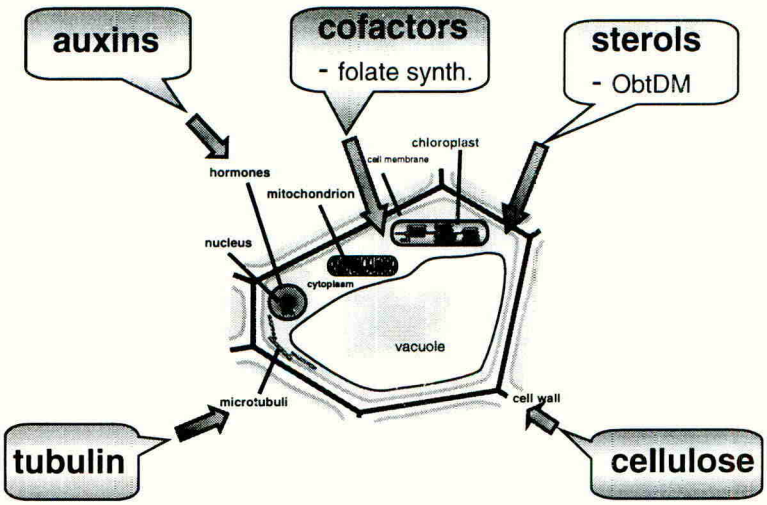


Figure 4: The non-chloroplast herbicide targets

- Adenylosuccinate Synthase
- Squalene Cyclase
- Obtusifoliol Demethylase
- Acetohydroxyacid Reductoisomerase
- Imidazolglycerol-P Dehydratase
- Isopropylmalate Dehydrogenase
- Histidinol Dehydrogenase
- Anthranilate Synthase
- Homoserine Dehydrogenase
- Threonine Dehydratase
- Ornithine Carbamoyl Transferase
- Aspartate Amino Transferase
- Pyruvate Dehydrogenase
- Glycine Decarboxylase
- AcylCoA Synthase
- β-Ketoacyl Synthase
- Aminolevulinic Acid dehydratase
- Hydroxymethylbilane Synthase
- Glutamate-semialdehyde Amino Transferase
- Deoxyxylulosephosphate Synthase
- Deoxyxylulosephosphate Reductoisomerase
- HMGCoA Reductase
- Farnesyldiphosphate Synthase
- Squalene Synthase
- "Sphingolipid Biosynthesis"
- Panthotenate Synthase
- Oxopantoyllactone Reductase
- AMP/Adenosine Deaminase
- Dihydrofolate Reductase
- Aminoacyl-tRNA Synthase
- Protein Phosphatase 2A
- Tyrosine Kinase
- RNA Polymerase III

Table 2: Thirty three experimental herbicide targets (herbicidal effectors reported; including some patented targets)

- Homoserine Kinase
- Threonine Synthase
- Dihydrodipicolinate Synthase
- Desoxyarabinoheptulosonate-7P Synthase
- Dihydroxyacid Dehydratase
- Branched Chain AA Amino Transferase
- Isopropylmalate Dehydratase
- Anthranilate Phosphoribosyl Transferase
- ATP Phosphoribosyl Transferase
- Cysteine Synthase
- Deoxyxylulosephosphate Synthase
- Isopentenyldiphosphate Isomerase
- Sterol Δ14 Reductase
- Adenylosuccinate Lyase
- Ribose-5P Isomerase
- Transketolase
- Fructose-6P 1-P Transferase
- Isocitrate Dehydrogenase
- Galactose Dehydrogenase
- D1 Protease
- Rubisco Methylase
- Cytokinin Receptor
- Glutamate Receptors
- Sodium Channel
- Cyclin-Dependent Protein Kinase
- "Cytoskeletal Components"

Table 3: Twenty six patented herbicide targets (excluding targets with already reported herbicidal effectors)

THREE DIFFERENT INITIAL SCREENING PLATFORMS ARE REQUIRED

In vivo screening in the green house and under field conditions of course cannot be substituted by either miniaturized *in vivo* or *in vitro* screening technologies. The selection of development candidates definitely has to be based on *in vivo* trials. Therefore, optimization within a chemical programme has to involve uptake, distribution, as well as metabolism within the whole plant.

The screening cascade (Figure 5) using diverse chemical libraries of up to a million compounds, most of them accessable only in a few mg amounts or less has, however, to start with a HTS/UHTS-platform consuming only µgs per assay. Assuming a success rate of max. 0.1 %, ca. 1,000 hits per assay have to be retested first to exclude false positives. At the same time dose dependent *in vitro* testing generates quantitative data at the target site for early SAR-considerations.

These reconfirmed hits in a second step have to be validated using miniaturized *in vivo* screening assays – a so called microscreening – still consuming only small amounts of test compounds in order to avoid resynthesis whenever possible. The overall capacity of such microscreening systems is ca. 50-100,000 compounds per year. Validated hits can be used as leads by initiating chemical synthesis programmes with an *in vivo* optimization with well established greenhouse tests and a general capacity of 15-20,000 compounds per year.

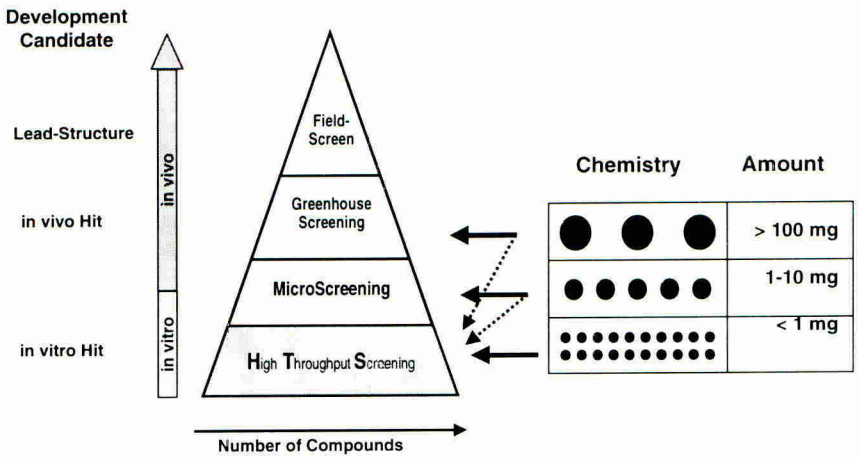


Figure 5: The new herbicide screening cascade

TARGET FINDING AND VALIDATION BY FUNCTIONAL GENOMICS

Target identification requires the generation of plant mutants which display either herbicide mimic effects, lesion mimics or lethal phenotypes. Mutants can be generated via multiple methods (chemical mutagenesis, transposon mutagenesis, antisense downregulation, sense-cosuppression, ribozymes, RNAi technology and possibly the new chimeraplast technology). However the technology used for mutant generation greatly influences the efforts and technologies used for target identification. In case of chemically mutagens where in general only point mutations are generated, the targets can be identified via SNP technologies (single nucleotide polymorphisms). Herbicide mimic mutants generated by transposon mutagenesis can be identfied by cloning the flanking DNA fragments. In all other approaches the potential target genes have to be present in a cDNA or an EST library and have to be "knocked out" one by one. That means e.g. for *A. thaliana* as a model plant knock-out experiments for all 30,000 genes one after each other. Nevertheless all approaches need a lot of data and information to be generated and stored in computer systems in a searchable format. Finally a novel target can be selected by sequence comparisons e.g. search for orthologous genes and functions and by the effect of knock-out or overexpression on plant phenotype. However, in cases where no function can be attributed to the target identified by a knock-out one has to apply further technologies like proteomics or biochemical profiles on the respective mutants in comparison to the wildtype plants. The technologies like automated high throughput sequencing of cDNA, ESTs and genomic DNA are well established and bioinformatic tools for analysis and comparison have greatly been improved in recent years. A successful transformation into pratice should follow Figure 6.

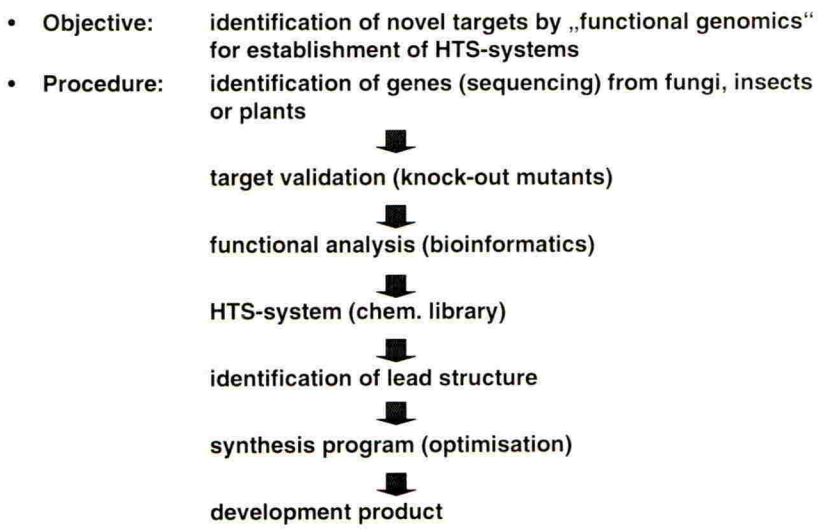


Figure 6: Genome projects in crop protection

(U)HTS ASSAY DEVELOPMENT AS BOTTLENECK

Figure 7 describes the overall neccessities to end up with biologically active compounds in the green house using the high throughput screening approach. After prioritization and decision to enter into a certain mode of action (functional genomics) about one to two years in total have to be invested until the biology under green house conditions might end up in a chemical project (validated lead). Figure 7 shows also that the actual HTS represents just a flash event in the whole discovery process. The development of assays to our understanding reflects more than one third of the total time frame. This is due to the fact, that each individual assay has to be worked out separately, and all biochemical reactions have to be adapted to only a few readouts like OD, fluorescence, luminescence and exceptionally radioactivity. Talking about HTS often neglects the role of physiology, biochemistry and assay development in the whole process.

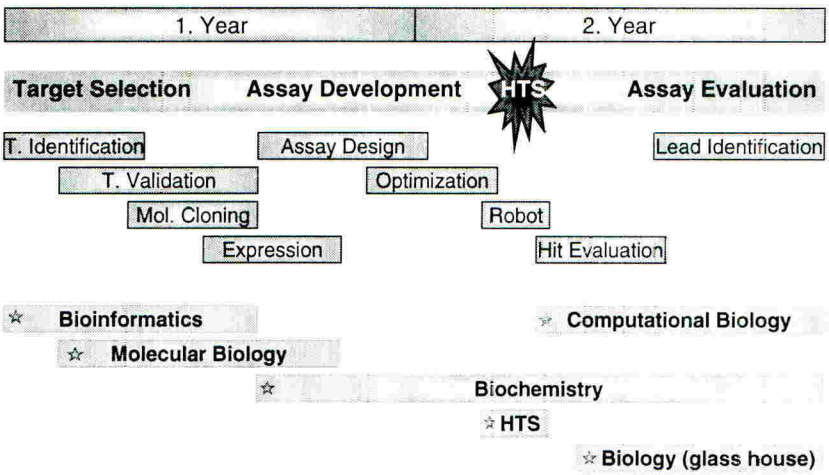


Figure 7: High throughput screening

USE OF BIOINFORMATICS FOR TARGET IDENTIFICATION (AN EXAMPLE)

The power of bioinformatics can be demonstrated by the following example. In 1996 a new plant specific pathway for isoprenoids has been described (Rohmer *et al.* 1996; Schwender *et al.* 1997). In 1996 a gene called Def or CLA1 was discovered in an albino mutant of *Arabidopsis thaliana* with arrested chloroplast development (Mandel 1996). However, the protein function of CLA1 was unknown. Subsequently in 1997 CLA1 was identified as 1-deoxyxylulose-5-phosphate synthase (DXS), the regulating enzyme of the new isoprenoid pathway (Sprenger *et al.* 1997, Lange *et al.* 1998). In the same year AgrEvo filed a patent for DXS as herbicide target (N.N. 1999). Without the efficient use of bioinformatics such a stringient path from phenotype to assay patent filing would not have been possible.

One year later the herbicidally active natural compound fosmidomycin was identified as interfering with 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), the enzyme following DXS in the same pathway (Zeidler *et al.* 1998, Kuzuyama *et al.* 1998). This validated the pathway for developing herbicides. Consequently in 1999 the first DXR genes from plants have been cloned (Lange *et al.* 1999). All of these novel approaches have to be compared to the success rate of the well established herbicide discovery methods.

THE CLASSICAL APPROACHES FOR PESTICIDE IMPROVEMENTS

The classical approaches, having shown a high level of efficacy, are

- analogue synthesis
- structure activity considerations (SAR, QSAR)
- molecular modeling
- natural products as leads
- innovative chemistry and random screening.

Clearly, these approaches will not be substituted but complemented by functional genomics and HTS/UHTS. We have to take into acount, however, that the different approaches require different screening platforms as well.

The classical approaches definitely have been proven to be efficent in the past. In the last few years a number of new, but non-validated targets have been used for herbicide research (Table 2 and 3). In some cases inhibitors have been identified but interestingly none of the inhibitors led to a marketable herbicide. Thus functional genomics will certainly contribute to the validation of potential targets by knock-out or down-regulation and thereby hopefully improve the success rate via this approach.

REFERENCES

Kuzuyama T; Shimizu T; Takahashi S; Seto H (1998). Fosmidomycin, a specific inhibitor of 1-deoxy-D-xylulose 5-phosphate reductoisomerase in the nonmevalonate pathway for terpenoid biosynthesis. *Tetrahedron Lett.* **39**, 7913-7916.

Lange B M; Wildung M R; McCaskill D; Croteau R (1998). A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2100-2104.

Lange B M; Croteau R (1999). Isoprenoid biosynthesis via a mevalonate-indepndent pathway in plants: cloning and heterologous expression of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase from peppermint. *Arch. Biochem. Biophys.* **365**, 170-174.

Mandel M A; Feldmann K A; Herrera-Estrella L; Rocha-Sosa M; Leon P (1996). CLA1, a novel gene required for chloroplats development, is highly conserved in evolution. *Plant Journal* **9**, 649-658.

N. N. (1999): 1-Deoxy-D-xylulose-5-phosphat Synthase und Verfahren zur Identifiziirung von Effektoren der 1-Deoxy-D-xylulose-5-phosphat Synthase. *Patent Application* DE 197 52 700 A1.

Ridley S M; Elliot A C; Yeung M; Youle D (1998). High-throughput screening as a tool for agrochemical discovery: automated syntheis, compound input, assay design and process managemant. *Pesticide Science* **54**, 327-337.

Rohmer M; Seemann M; Horbach S; Bringer-Mexer S; Sahm H (1996). Glyceraldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis. *J. Am. Chem. Soc.* **118**, 2564-2566.

Schwender J; Zeidler J; Groener R; Mueller C; Focke M; Braun S; Lichtenthaler F W (1997). Incorporation of 1-deoxy-D-xylulose into isoprene and phytol by higher plants and algae. *FEBS Lett.* **414**, 129-134.

Sprenger G A; Schorken U; Wiegert T; Grolle Sigrid; De Graaf A A; Taylor, S V; Begley T P; Bringer-Meyer S; Sahm H (1997). Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12857-12862.

Ward E and Bernasconi P (1999). Target-basaed discovery of crop protection chemicals. *Nature Biotech.* **17**, 618-619.

Zeidler J; Schwender J; Mueller C; Wiesner J; Weidemeyer C; Beck E; Jomaa H; Lichtenthaler H K (1998). Inhibition of the non-mevalonate 1-deoxy-D-xylulose-5-phosphate pathway of plant isoprenoid biosynthesis by fosmidomycin. *Z. Naturforsch., C: Biosci.* **53**, 980-986.

Molecular approaches supporting the identification and validation of new herbicide targets

R Höfgen, J Freitag, S Maimann,
Max-Planck-Institute of Molekulare Pflanzenphysiologie, D-14476 Golm / Potsdam, Germany

F Schmidt
AgrEvo, D-65926 Frankfurt, Germany

H Hesse
FU Berlin, Institut für Biologie, Angewandte Genetik, D-14195 Berlin, Germany

ABSTRACT

Yield and quality of crops in most agricultural production systems are relying, besides other factors, on chemical plant protection, especially herbicides. Development and improvement of crop protection compounds are mainly dependent on chemical synthesis and random screening procedures of phytotoxic compounds. Here we describe two molecular approaches which we developed into useful tools to sustain herbicide target validation or discovery. First, an *in vivo* method for target validation could be achieved through antisense inhibition of putative target genes in transgenic plants. Antisense inhibition is a useful method for target validation and mimics the effect of herbicide application and allows determination of the threshold levels for effective target enzyme inhibition. Two examples are presented: acetolactate synthase as a well known herbicide target and cystathionine beta-lyase. Second, a molecular approach to identify the hence unknown target of a given phytotoxic compound has been established exploiting a gain of function approach. Plant protoplasts are transformed with a T-DNA carrying a strong enhancer causing overexpression of a nearby gene and resistant cell lines are selected with toxic concentration of a phytotoxin. Resistance is confered through overproduction of the respective target enzyme. Cloning of the target gene unequivocally determines the target. This approach is especially useful if the molecular target cannot be determined quickly through conventional methods. Both approaches additionally provide further advantages for agrochemical research and plant breeding as providing sufficient enzyme for further analysis or resistance mechanisms for a given compound.

INTRODUCTION

Adding to the wealth of classical approaches in plant breeding, plant sciences and agrochemistry molecular approaches and techniques are an integral part of plant molecular biology, plant physiology and plant breeding. In agriculture the application of transgenic plants providing new or improved traits will, despite some recent concerns, prove to become an increasingly important part of breeding programmes. These genetically engineered and carefully evaluated transgenic crops will support strategies of sustainable agriculture and

integrated pest management. In agrochemical research the increasing knowledge of molecular biology is used as a tool to support lead finding, herbicide development and marketing strategies of pesticidal compounds, especially herbicides. Especially successful are combined strategies of highly effective broad range herbicides with matching herbicide resistant transgenic crop lines (Foster & Hoefgen, 1993).

Here we present two strategic approaches applying molecular techniques as tools for the setup of experimental designs aiming at either target validation or target identification (Figure 1). Target validation using antisense inhibition of endogenous genes should be a prerequisite of laborious biochemical approaches for herbicide discovery such as rational design approaches based on either enzyme activity studies or protein crystallisation (Foster & Hoefgen, 1993, Hoefgen *et al.*, 1995 a). Antisense validation can provide a valuable hint whether a certain enzyme might be suited as a potential target or should better be neglected in further synthesis or screening programmes (Siehl, 1992). Target identification based on T-DNA activation tagging approaches (Fritze & Walden, 1995) can be used to determine and clone the unknown molecular target of a given phytotoxic compound or herbicide. This is of special interest in cases where conventional biochemical methods do not readily lead to the determination of the respective target enzyme.

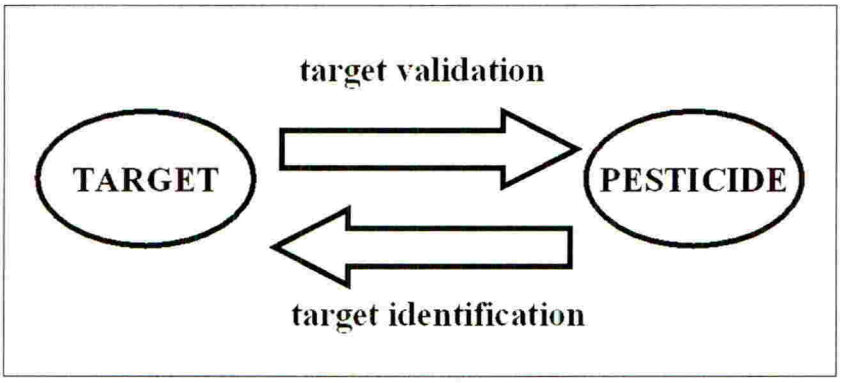


Figure 1: Molecular approaches as tools for agrochemical research. Starting from a putative target gene molecular approaches such as antisense inhibition of this gene/enzyme enable the validation of the target prior to any synthesis efforts and, furthermore, provide an *in vivo* model of the symptoms to be expected. Starting from a known substance we devised a molecular method to identify the target gene based on the T-DNA activation technique.

RESULTS AND DICUSSION

Target validation

Antisense inhibition of gene expression is well established as a tool to investigate the function of enzymes within a biochemical pathway encoded by these genes and to analyse

the physiological effects of their inhibition (Hoefgen & Willmitzer, 1992). The ideal experimental situation is the antisensing of a cloned gene in the homologous plant leading to reduced RNA amounts and thus reduced amounts of the encoded protein; heterologous antisensing even in closely related species is not recommendable. This approach can be used to validate the potential of a putative herbicide target as the reduction in enzyme content mimics the inhibition of the respective enzyme through a herbicide. Transgenic plants should represent a phenocopy of herbicide action and thus screening for successful inhibitions can be done on the phenotype level screening for lethality or severe growth retardation. Symptoms can be alleviated through supplementation of a respective downstream metabolite which proved to be especially helpful during tissue culture and plant propagation. The biochemical analysis of the remaining enzyme activity of the "antisense" plants will allow to determine threshold levels still allowing survival of the plants and thus indicating the degree of inhibition which has to be achieved by a herbicidal compound. This determination is possible due to the fact that antisense inhibition usually results in a gradient of individually inhibited transgenic plants ranging from near wildtype to lethal phenotypes whereas other systems such as mutational approaches or insertion mutagenesis usually lead to total inhibition of enzyme activity (knock-out). As examples for the power of this approach we present data for the antisense inhibition of two amino acid biosynthetic genes, acetolactate synthase (ALS) and cystathionine beta-lyase (CbL). In both cases the constitutive 35S promoter was used to drive constitutive expression of the antisense gene. The genes were cloned from potato and transgenic potato antisense lines were generated.

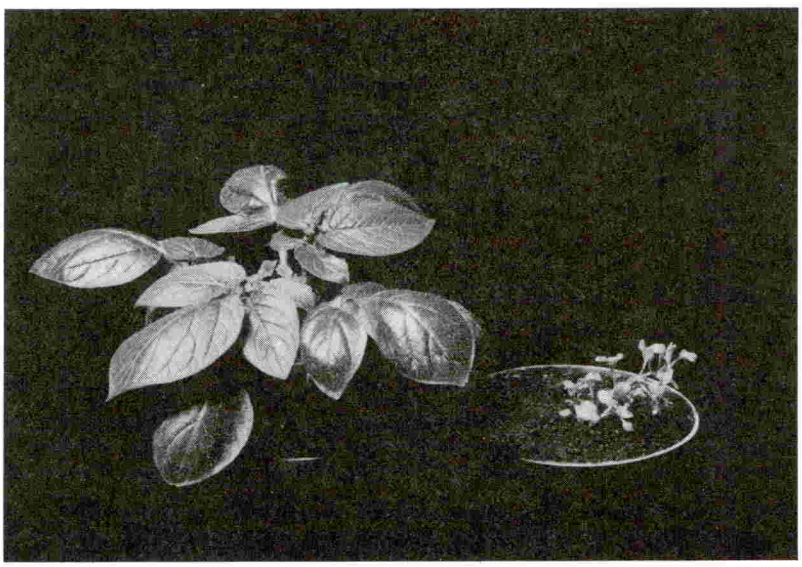


Figure 2: Antisense inhibition of acetolactate synthase in potato resulted in case of the most extreme phenotypes in stunting and growth retardation of the transgenic plants with pale yellow, small and distorted leaves (right) when compared to non-transformed control plants (left). Individuals of the most extreme phenotypes died when planted into soil. The symptoms could be partially alleviated in tissue culture when supplemented with casamino acid hydrolysate.

ALS, a known target of several highly efficient herbicide classes such as sulfonylureas or imidazolinones, proved also in the antisense system its validity as a herbicide target and was used as a nailtest for the proposed system of herbicide target validation. ALS antisense plants (potato) with a reduction in enzyme activity of about 50 % already showed slight phenotypic effects as growth retardation, whereas an inhibition of enzyme activity of about 80% resulted in severely stunted, chlorotic and necrotic plants, which hardly survived under greenhouse conditions (Figure 2). These transgenic lines could only be propagated *in vitro* when supplemented with casamino acid hydrolysate. It was not possible to detect plants with less than 15% remaining activity which indicates the threshold level necessary to reach a lethal effect. Antisense plants and controls treated with imidazolinones exerted comparable "symptoms" and changes in free amino acid composition (Hoefgen *et al.*, 1995b). Thus, antisense inhibition could be successfully established as a pre-screening system for evaluating the validity of potential target enzymes.

Cystathionine beta-lyase (CbL) is one of the enzymes of the methionine biosynthetic pathway. The enzymes cystathionine gamma-synthase (CgS), cystathionine beta-lyase (CbL) and methionine synthase (MS) successively convert phosphohomoserine and cysteine via cystathionine and homocysteine to methionine (Hell & Rennenberg, 1998). A chemically induced *Nicotiana plumbagenifolia* mutant supposedly inhibited in CbL activity has been described (Negrutiu *et al.*, 1985). Feeding studies suggested an inactivation of CbL as the nature of lethality. The respective *N. plumbagenifolia* plants can be propagated in the heterozygous state as the mutation is recessive or *in vitro* when supplemented with methionine. Recently, we could prove the identity of this very mutation by reversion of the mutant phenotype through transformation with an *E. coli* MetC construct encoding a functional CbL targeted to the chloroplasts (unpublished results). This methionine auxotrophic mutant indicates that CbL is a potential herbicide target. In order to test this assumption in the target validation system we cloned a CbL gene from potato (*StCbL*) and transfered a CbL antisense construct to potato cv. Désirée. We are presently performing the analysis of antisense plants.

Target identification

The process of determining the target enzyme of a phytotoxic compound can be difficult if only relying on biochemical methods. Here we describe a molecular approach for the determination of the respective target without prior knowledge of the target (Figure 1).

Plants are able to spontaneously develop resistance against herbicides due to a variety of mechanisms such as detoxification processes like destruction, modification, conjugation and sequestration of the active components of the herbicides, altered uptake, mutations of the target sites, amplification of the target enzyme either due to gene amplification or increased expression of the respective RNA. In the latter, overexpression of a gene resulting in increased steady state RNA levels, can be easily mimicked exploiting existing molecular tools. In order to screen for overexpression of an unknown target gene we followed a random 'gain of function' approach adopting the T-DNA activation tagging technique (Fritze & Walden, 1995) for the cloning of target genes coding for enzymes inhibited by herbicides. The underlying principle is the out-titration of phytotoxic molecules through excess amounts of the otherwise unaltered target enzyme (Figure 3). *Agrobacterium* mediated gene transfer randomly delivered the T-DNA (vector: pPCVICEn4HPT, Walden *et al.*, 1995) into the

genome of tobacco protoplasts (*N. tabacum* cv. SR1). The overexpression of the tagged gene is caused through stable, but random integration of a strong enhancer tetramer derived from the 35S CaMV promoter into the plant genome causing a dominant mutation allowing for positive selection of resistance to a phytotoxic agent.

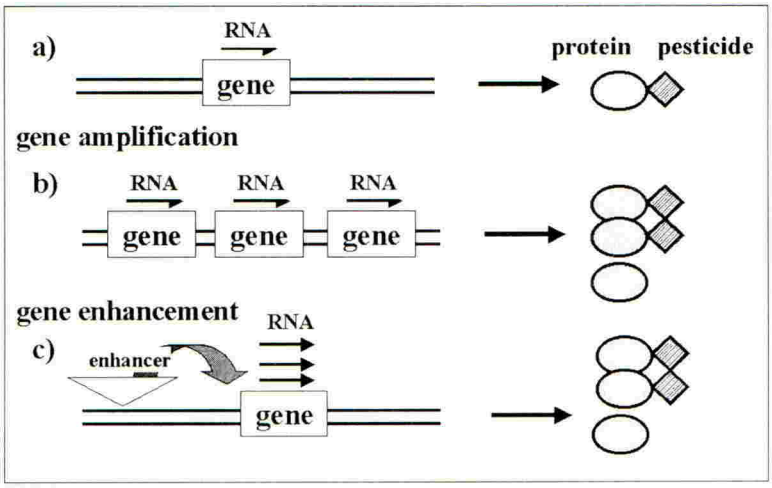


Figure 3: Schematic representation of the target identification approach. a) A herbicide inhibits the activity of an enzyme encoded through a gene. b) Natural occuring resistance can rely on gene amplifications resulting in increased amounts of the target protein some of which evading herbicide inactivation. c) Gene amplification can be mimicked in a transgenic approach through the integration of an transcriptional enhancer next to the target gene resulting again in excess amounts of the target protein.

We subjected tagged protoplasts to several different herbicides of which the results of glyphosate are presented as case study. The selective conditions (LD_{100}) for glyphosate were determined in 'kill curve' experiments, additionally revealing that under these conditions no spontaneous resistance did occur. The enzyme EPSPS (enolpyruvylshikimate-3-phosphate synthase) converts shikimate to enolpyruvylshikimate-3-phosphate in the biosynthetic pathway of the aromatic amino acids. We chose glyphosate for validation of our experimental strategy because resistance to glyphosate due to overexpression and gene amplification has been described earlier. About $3x10^7$ protoplasts of tobacco were cocultivated with *Agrobacterium* harbouring the vector pPCVICEn4HPT and subjected to highly stringent double selective conditions of 20 mg/l hygromycine and 10^{-3}M glyphosate resulting in 14 independent cell lines (microcalli) which could all be regenerated to plants, eventually. Southern blot analysis revealed integration of the T-DNA in all lines and Northern blot hybridisations proved expression of the hygromycine gene in all cases as well from callus, *in vitro* plantlets and greenhouse grown plants. However, a screen for herbicide resistance under greenhouse conditions applying 360g a.i./ha (900 mg/l) of glyphosate which

has been shown to effectively inhibit growth of controls only resulted in two lines that showed an acceptable degree of tolerance, though no complete resistance in T_0 or T_1 transgenics. Northern blots hybridised to an EPSPS fragment cloned by PCR from tobacco revealed a moderate overexpression of the EPSPS gene in these lines (Figure 4). In order to isolate the gene(s) neighbouring the enhancer and putatively confering resistance we followed the plasmid rescue approach (Walden *et al.*, 1995) being enabled through the integration of a functional *E. coli* plasmid in the T-DNA. After digestion of callus DNA with EcoRI we religated the genomic DNA and transformed these circular fragments into competent *E. coli* cells. One of the resulting clones derived from one of the two tolerant genotypes identified in the spraying experiment was analysed further. The insert of flanking plant DNA in that clone was approximately 6 kb in size. However, no relevant ORF could be detected on this fragment, thus we will perform plasmid rescue with other restriction enzymes resulting in larger inserts and we will screen a genomic library produced from the respective tolerant transgenic plant lines using T-DNA sequences as probe.

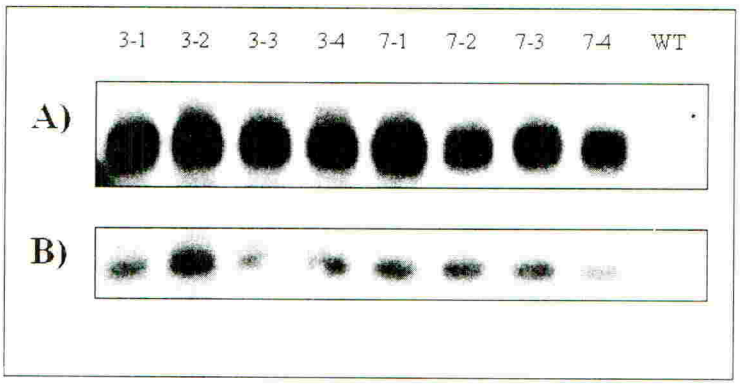


Figure 4: Northern blot analysis of two glyphosate resistant tobacco plant lines (3 and 7) resulting from T-DNA activation tagging of tobacco protoplasts selected with 10^{-3} M glyphosate. Total leaf RNA (20 µg) was blotted and probed with a tobacco EPSPS (panel A) and ubiquitin to normalise loading (panel B). The resistant tobacco lines showed an increased expression of the target gene EPSPS compared to wildtype plants (WT).

When this final proof is achieved, this tool can be included into agrochemical research to determine the molecular nature of a target. Additionally, mutated or overexpressed target genes themselves can be used in plant breeding to confer resistance to crop plants. However, it will be difficult or impossible to apply the method to compounds acting on several targets. Furthermore, instead of using a protoplast system we can envisage that especially transgenic *Arabidopsis* gain of function populations can be used for the same purpose. The predicted number of genes in tobacco is about 43000 genes compared to 16 to 33.000 genes estimated for *Arabidopsis* (Arumuganathan & Earle, 1991; Miklos & Rubin, 1996). However, the protoplast system is still our prefered system due to the huge number of transformants that can be easily produced and selected once the method is established, and, furthermore, the

protoplast system can be expanded to other plant species with good chance for success while generation of giant mutant tagged populations is only at hand for *Arabidopsis*.

CONCLUSION

The central feature of both approaches is the cloning of target genes and the analysis of *in vivo* effects of either tissue culture lines or transgenic plants providing phenocopies of either herbicide treatment or herbicide resistance.

Target validation is dependent on the cloning of a putative target which will finally provide information whether a certain target enzyme might be suitable for further research. The main drawback of this approach is its confinement to a limited number of species which can be tested as only homologous antisense inhibition guarantees success because the process is strictly dependent on sequence homology. Furthermore, in principle all plant species can be transformed, however, routine and quick transformation producing a greater number of individual transformants is only available for a limited number of plants which also does not reflect the variability within a given population of weeds. However, target validation through antisense inhibition provides a good and testable model of herbicide action and reliable information for further analysis. Especially as this information will be gained prior to extensive chemical synthesis and biochemical efforts and should be used to direct these programmes at the very beginning. With whole plant genomes sequenced and extensive EST sequences at hand for a growing number of plant species identification and cloning of genes will no longer pose a challenge, but, functional analysis and rational and sytematic analysis will be the task of the future. At first sight genomic approaches leading to loss of function seem to be the relevant answer to this challenge, however, suitable libraries are available only for *Arabidopsis* and, propably, maize. Furthermore, knock-outs alone will not provide the relevant answers and additional informations inherent to the antisense approach such as threshold levels or indications of the expected phenotype. Knock-outs might be misleading as e.g. in the case of threonine deaminase, where a knockout proved to be lethal whereas the antisensing of the strongly inducible gene resulted only in moderate phenotypes (Hoefgen *et al*, 1995a; Hoefgen, 1999). In this paper it has been shown that there was a positive correlation in case of CbL. Thus, knock-outs can be judged positively, but need further detailed analysis either on the biochemical level or through the antisense system for target validation.

Target identification through T-DNA activation tagging will allow identification of previously unknown targets of phytotoxic compounds. In this case cloning of the target gene is not the prerequisit but the result of the approach. Furthermore, this strategy might enable the development or identification of new resistance mechanisms to known or newly developed herbicides either by overexpressing the target enzyme or by inducing pre-existing resistance mechanisms in the plant. As the responsible genes can be ‚easily' cloned, these resistance mechanisms can be transfered to other crop plants as well and can be introduced into crop breeding programmes.

Beyond that obtention of the respective target gene and it's molecular and biochemical analysis is suited to accelerate strategies for synthesising improved or optimised compounds starting from known lead structures. As a valuable side effect the cloned target gene being used for antisense inhibition or the result of the target identification approach can be

exploited for overexpression in heterologous systems providing the respective protein for approaches such as high throughput screens (HTS) or biochemical and biophysical analyses. Furthermore, the detailed knowledge of a target enzyme will support the registration of compounds and sustain marketing strategies.

ACKNOWLEDGEMENTS

The author is in debt to Dr. Hans-Dieter Pohlenz, Schering Agrochemicals-Berlin, now Schering Pharmaceuticals-Berlin, and Prof. Dr. Lothar Willmitzer, IGF-Berlin, now MPI of Molecular Plant Physiology,Golm for long discussions generating the above summarised projects and support throughout their performance. Furthermore, I would like to thank Drs. Bernd Laber, Wolfgang Streber, Rick Walden and Renate Lührs as well as Uli Kutschka, Iris Schuettke, Ann-Kristin Klonus, Astrid Blau, Yvonne Maas, Antje Fischer and Joseph Bergstein for different kinds of support.

REFERENCES

Arumuganathan K; Earle E D (1991). Nuclear DNA content of some important plant species. *Plant. Mol. Biol. Rep.* **9**, 208-218

Foster S G; Hoefgen R (1993). Towards a rational design of pesticides. In: *Opportunities for molecular biology in crop production.* British Crop Protection Council Monograph, No. 55, eds. J Beadle *et al.*, pp. 75-81, UK

Fritze K; Walden R (1995). Gene activation by T-DNA tagging. In: *Methods in Molecular Biology*, Vol. 44, *Agrobacterium* protocols, eds: K M A Gartland and M R Davey, pp 281-294, Humana Press Inc., Totowa, NJ, USA

Hell R; Rennenberg H. (1998). The plant sulphur cycle. In: *Sulphur in Agroecosystems*, ed.: E Schnug, pp. 135-173. Kluwer Academic Publishers, The Netherlands

Höfgen R; L Willmitzer (1992) Transgenic potato plants depleted for the major tuber protein patatin via expression of antisense RNA. *Plant Science* **87**, 45–54

Hoefgen R; Streber W R; Pohlenz H-D (1995a). Antisense Gene Expression as a Tool for Evaluating Molecular Herbicide Targets. *Pesticide Science* **43**, 175–177

Höfgen R; Laber B; Schüttke I; Klonus A-K; Streber W R; Pohlenz H-D (1995b). Repression of Acetolactate Synthase Activity through antisense inhibition. *Plant Physiol.*, **107**, 469-477

Hoefgen R (1999). Molecular identification of herbicide targets applying transgenic approaches in tissue culture. In: *Plant Biotechnology and In Vitro Biology in the 21st Century* eds: A Altman; M Ziv; S Izhar, pp. 495–498. Kluwer Academic Publishers, Dordrecht, The Netherlands

Miklos G L G; Rubin G M (1996) The role of the genome project in determining gene function: Insights from model organisms. *Cell* **86**, 521-529

Negrutiu D; De Brouwer R; Dirks R; Jacobs M (1985). Amino acid auxotrophs from protoplast cultures of *Nicotiana plumbagenifolia,* Viviani, *Mol. Gen. Gen.* **199**, 330 -337

Siehl, DL (1992) Considerations in selecting a target site for herbicide design. In: *Biosynthesis and Molecular Regulation of Amino Acids in Plants*, eds: B K Singh; H E Flores; J C Shannon, pp. 146-162, ASPP, Rockville, USA

Walden R; Fritze K; Harling H (1995). Induction of signal transduction pathways through promotor activation. In: *Methods in Cell Biology* **49**. Academic Press Inc. 449-436

The generation of novel secondary metabolites through combinatorial biosynthesis

J McDermott, G. Meurer, B. Waters, Wanggui Y-S, W. Pan, X. Li, C. Silva, P. Brian and J. Davies.
TerraGen Discovery Inc, Suite 300-2386 East Mall, Vancouver, Canada

ABSTRACT

The search for natural products that have therapeutic or agrochemical uses was one of the founding technologies upon which many of the world's largest pharmaceutical and chemical companies were based. However, in the last twenty years, this area of research has become increasingly more expensive and labour intensive. With the recent advent of high throughput screening the output from traditional natural products discovery groups has been incompatible with the needs of screening groups. TerraGen Discovery Inc. is perfecting techniques that will provide access to a greater range of biosynthetic diversity than has been previously sampled.

INTRODUCTION

Compounds derived from the filamentous soil bacteria actinomycetes account for the majority of microbially derived natural products. The structural variety of molecules isolated from these organisms is astonishing. Over the past 50 years such secondary metabolites have revolutionized medical practice. It is difficult to imagine a world without antibiotics, for the lives of many people have been profoundly influenced directly or indirectly through their use.

Traditional natural products discovery involves the isolation of microbes from diverse environmental samples, which were then screened for the presence of novel biologically active secondary metabolites. Over the last twenty years this approach has been producing fewer novel structures and the process as a whole is becoming increasingly more expensive and labour intensive. This has made it less competitive as a source of new structures for agrochemical companies to screen for new pesticides or insecticides.

Using combinatorial biosynthesis TerraGen provides access to greater biosynthetic diversity of natural products in two ways. Firstly, by targeting sources of biosynthetic diversity from uncultivable organisms (Heuer *et al* 1997); with the actinomycete project

this is achieved by targeting those actinomycetes that have not been cultivated in the laboratory. The second approach used by TerraGen to generate biosynthetic diversity relies upon the fact that although natural products show a great deal of structural diversity almost all of them are generated using a modest number of similar classes of biosynthetic
pathways. In many cases it appears that the enzymes involved in the biosynthesis of these secondary metabolites show a considerable degree of promiscuity in the range of substrates they will accept. The similarities between these biosynthetic pathways are such that it is possible to substitute certain modules from distinct pathways, creating novel biosynthetic pathways that produce novel natural products. TerraGen has amply demonstrated this approach, as have a number of academic laboratories, and several companies (Solenberg *et al.*, 1997; Kuhstoss *et al.*, 1996; Hopwood *et al.*, 1985).

Recent studies have demonstrated that many bacteria, including the actinomycetes, which can be detected using 16srDNA analysis, are not represented in culture collections (Heuer *et al.*, 1997; Felske *et al.*, 1997). By isolating genomic DNA directly from environmental samples, combinatorial biosynthesis provides access to uncultivable organisms.

Combinatorial biosynthesis

TerraGen's approach to combinatorial biosynthesis is based on three premises:

a) that all the genes necessary for the biosynthesis of a product are clustered at a single locus,
b) that the enzymes involved in these biosynthetic pathways are able to recognize a variety of substrates,
c) that actinomycetes are closely related phylogenetically, with conserved mechanisms of global regulation of antibiotic synthesis.

The clustering of biosynthetic pathways at a single locus has been repeatedly demonstrated for almost all biosynthetic pathways studied in actinomycetes (Figure 1).

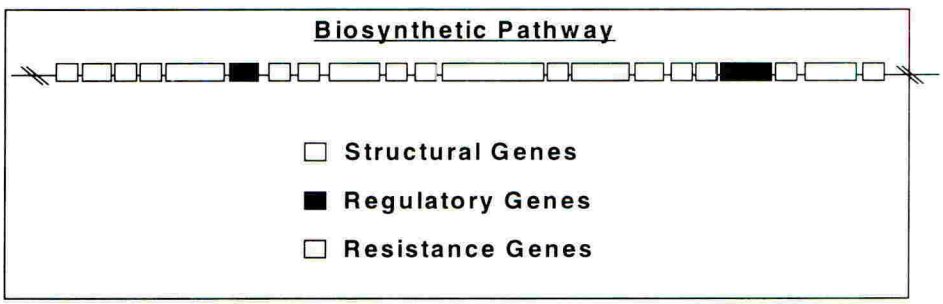


Figure 1 Bacterial biosynthetic genes are clustered at a single locus that contains all the genes necessary for biosynthesis, self-resistance and regulation of the secondary metabolite.

This clustering allows the cloning and expression of complete pathways as single clones in an appropriate heterologous host. The conservation of global regulators of antibiotic synthesis among the actinomycetes has only recently been investigated directly (Champness *et al.*, 1992; Bibb, 1996). However, there have been many studies in which biosynthetic pathways have been expressed without significant modification in a surrogate host (Malpartida and Hopwood, 1984). These experiments indirectly demonstrate the similarities between actinomycete species in the global regulation of antibiotics. TerraGen's approach to combinatorial biosynthesis relies upon the expression host's own regulatory elements being functionally similar to those in the original host to allow expression of the heterologous biosynthetic pathway.

A typical natural product biosynthetic pathway maybe comprised of five to more than a hundred linked genes that together regulate expression and encode the biosynthetic enzymes responsible for natural product assembly and structural modification. In addition, natural product gene clusters include one or more genes that encode enzymes for cellular self-protection and extracellular drug transport. For combinatorial biosynthesis to successfully generate novel natural products, it must be possible to efficiently transfer segments of DNA from a natural product-generating donor microorganism into an appropriate engineered expression host. That is an essential aspect of TerraGen's technological proficiency.

There are several ways that new chemical entities are created by combinatorial biosynthesis approaches. First, it is now possible to transfer entire metabolic pathways from a donor strain to an engineered host using diverse molecular genetic tools (e.g. mobilizable vectors). This process allows a single pathway to be isolated from the genomic library and expressed in an appropriate host. Second, transfer of metabolic pathways from a donor strain to an engineered host can lead to expression of silent pathways that are not normally expressed by the donor or host strain.

The starting point for library construction is high molecular weight genomic DNA isolated from cultured donor strains, or directly from the environment. Once genomic DNA is isolated and suitably prepared, it is used to create natural pathway libraries.

The libraries constructed from DNA isolated from soil samples (Figure 2) provide metabolic pathways that remain in a native configuration. This allows rapid access to biosynthetic gene clusters and alleviates many of the problems associated with traditional natural products discovery as all of the secondary metabolites are produced in a defined background. This allows the generation of fractionated extracts containing pure or nearly pure compounds, which can be rapidly isolated away from the well-characterized components produced by the host.

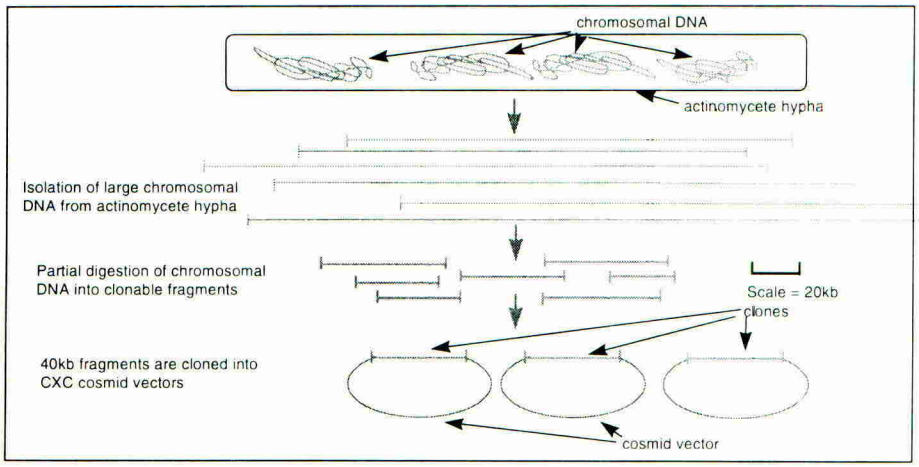


Figure 2. For the construction of natural pathway libraries large fragments of genomic DNA are isolated from the donor bacteria and then digested into smaller overlapping fragments that can be cloned into appropriate vectors

Proof of Principle experiment

To demonstrate that combinatorial biosynthesis would provide access to the microbial diversity present in environmental samples, a proof of principle experiment was designed. The goals of this experiment were to demonstrate the feasibility of:

a) isolating high molecular weight DNA from microorganisms in soil samples and creating genomic libraries from this DNA ,
b) using chemical fingerprinting to detect the expression of known compounds from these libraries in a defined expression host,
c) detecting the expression of novel compounds from these libraries.

The first goal of the project was to demonstrate that high molecular weight DNA could be isolated directly from soil samples and used to create high titre genomic libraries. The first step in this process was to separate the bacterial cells away from the soil particles and eucaryotic cells using a variety of methods including, sieving, mechanical abrasion and centrifugation. DNA extraction is extracted from this bacterial pellet and then electrophoreised on a pulsed field gel (PFGE) to size select for larger fragments of DNA (>60kb) and to further purify the DNA away from contaminants such as humic acids (Figure 3).

This high molecular weight DNA was partially digested with the restriction enzyme

*Sau*3AI, to produce overlapping clones that were size selected for various sizes depending upon the vector being used for the library (20 – 30kb for plasmid vectors, 35-45kb for cosmid vectors). During the early stages of this work several vectors were tried but the most stable was determined to be a series of integrative cosmids (Figure 4). As part of this ongoing process for TerraGen, several libraries have now been constructed in a variety of vectors.

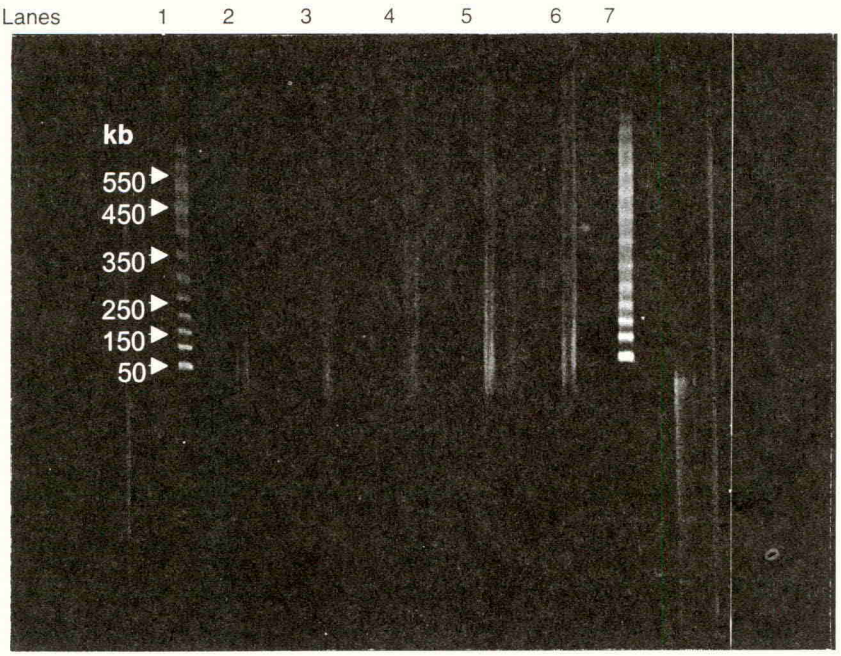
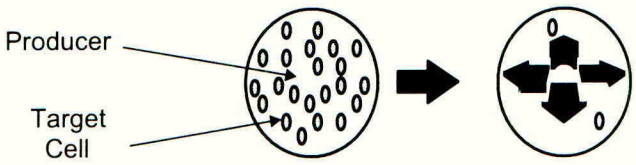
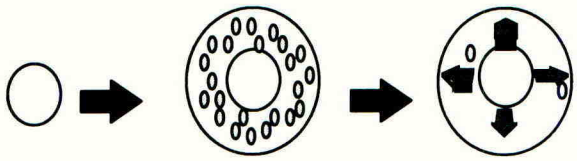


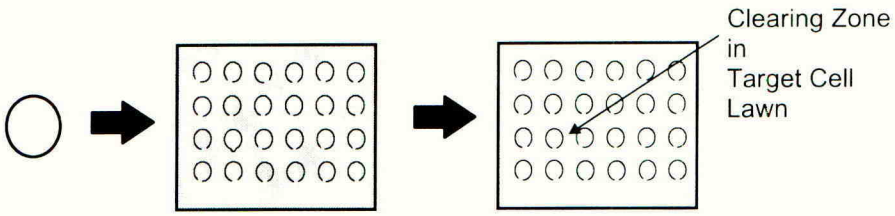
Figure 3 DNA isolated directly from 5 soil samples. Lanes 1and 7 show the size markers, lanes 2 through 6 contain genomic DNA isolated from five different soil samples, the samples range in size and quality but all contain some fragments that are >400kb.



5a Coencapsulation Assay



5b Double Encapsulation Assay



5c Lawn Overlay Assay

Figure 5a The expression host and the target cells are encapsulated together and allowed to ferment for several days. The production of secondary metabolites by the expression host would then trigger the assay in the target cells. **5b** The expression cells are encapsulated and then allowed to ferment to allow the accumulation of secondary metabolites in the macrodroplet. A second layer of alginate is then added containing the target cells, the diffusion of the metabolites through this layer triggers the assay in the target cells. **5c** Again the expression host is encapsulated alone and allowed to ferment. These macrodroplets are then placed on a agar layer containing the target cells, the diffusion of secondary metabolites into the agar then triggers the assay.

The eggs were first cleaned to remove potential contamination and encapsulated in an alginate macrodroplet. The encapsulated eggs incubated for an appropriate length of time and then the number of hatched eggs were counted. Using the macrodroplets the *Helicoverpa* eggs hatched and produced robust larvae, which rapidly worked their way out of the macrodroplets (Figure 6).

More recently, we have explored the use of insect eggs as an assay lawn. The eggs are sterilized, suspended in a soft agar, and overlaid on a petri plate. Again, the larvae hatched at an acceptable frequency as a part of the lawn. This system allows us to incubate the recombinant actinomycetes in the macrodroplets and then place them on top of the soft agar assay lawn and look for zones of inhibition of hatching around macrodroplets producing secondary metabolites. The ability of known compounds to inhibit hatching of the eggs in either the macrodroplets or the sort agar lawn is currently being examined.

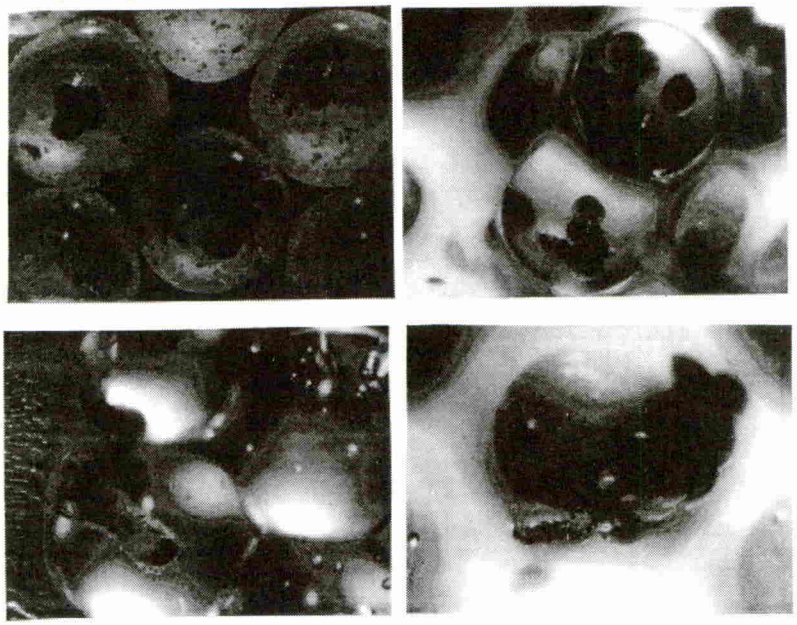


Figure 6 The *Helicoverpa zea* eggs were encapsulated in calcium alginate macrodroplets (top left). The macrodroplets were incubated at 22°C for 2 days before the *Helicoverpa* larvae stated to hatch (top right). By day 4 the larvae had hatched and as they grew started move out of the macrodroplets (bottom left and right).

The scope of these experiments will be expanded to include screening programs for novel forms of *Bacillus thurigiensis* toxins, either isolated from soil DNA or genetically engineered *in vitro*.

CONCLUSION

TerraGen's primary goal in the near future is to continue to demonstrate that we can isolate novel biosynthetic pathways, enzymes, proteins and insect toxins from soil libraries. The true value from these libraries will be achieved when we are able to rapidly screen them against a range of useful targets. The goal for our work in the agrochemical field is to take the new libraries that are being produced and screen these for production of compounds, which have the ability to inhibit hatching of *Helicoverpa zea*.

REFERENCES

Heuer H, Krsek M, Baker P, Smalla K, Wellington EM (1997). Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Applied and Environmental Microbiology* **63**(8), 3233-41.

Felske A, Rheims H, Wolterink A, Stackebrandt E, Akkermans AD (1997). Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils. *Microbiology* **143,** 2983-9.

Solenberg PJ, Matsushima P, Stack DR, Wilkie SC, Thompson RC, Baltz RH (1997). Production of hybrid glycopeptide antibiotics *in vitro* and in *Streptomyces toyocaensis*. *Chemical Biology* **4,**195-202.

Kuhstoss S, Huber M, Turner JR, Paschal JW, Rao RN (1996). Production of a novel polyketide through the construction of a hybrid polyketide synthase. *Gene* **183,** 231-6.

Hopwood DA, Malpartida F, Kieser HM, Ikeda H, Duncan J, Fujii I, Rudd BA, Floss HG, Omura S (1985). Production of 'hybrid' antibiotics by genetic engineering. *Nature* **314**, 642-4.

Champness W, Riggle P, Adamidis T, Vandervere P (1992). Identification of *Streptomyces coelicolor* genes involved in regulation of antibiotic synthesis. *Gene* **115**, 55-60.

Bibb MJ (1996). The regulation of antibiotic production in *Streptomyces coelicolor* A3(2). *Microbiology* **142,** 1335-44.

Mazodier P, Petter R, Thompson C (1989). Intergeneric conjugation between *Escherichia coli* and *Streptomyces species*. *J Bacteriology* **171**, 3583-5.

Malpartida F, Hopwood DA (1984). Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature* **309**, 462-4.