Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Phytochemistry 62 (2003) 971–985
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Metabonomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts

Karl-Heinz Ott*[,1], Nelly Aranibar*[,2], Bijay Singh[3], Gerald W. Stockton[4]

BASF Agro Research, Princeton, NJ 08543, USA

Received 11 November 2002; received in revised form 19 November 2002

Abstract

The biochemical mode-of-action (MOA) for herbicides and other bioactive compounds can be rapidly and simultaneously classified by automated pattern recognition of the metabonome that is embodied in the ^{1}H NMR spectrum of a crude plant extract. The ca. 300 herbicides that are used in agriculture today affect less than 30 different biochemical pathways. In this report, 19 of the most interesting MOAs were automatically classified. Corn (*Zea mays*) plants were treated with various herbicides such as imazethapyr, glyphosate, sethoxydim, and diuron, which represent various biochemical modes-of-action such as inhibition of specific enzymes (acetohydroxy acid synthase [AHAS], protoporphyrin IX oxidase [PROTOX], 5-enolpyruvylshikimate-3-phosphate synthase [EPSPS], acetyl CoA carboxylase [ACC-ase], etc.), or protein complexes (photosystems I and II), or major biological process such as oxidative phosphorylation, auxin transport, microtubule growth, and mitosis. Crude isolates from the treated plants were subjected to ^{1}H NMR spectroscopy, and the spectra were classified by artificial neural network analysis to discriminate the herbicide modes-of-action. We demonstrate the use and refinement of the method, and present cross-validated assignments for the metabolite NMR profiles of over 400 plant isolates. The MOA screen also recognizes when a new mode-of-action is present, which is considered extremely important for the herbicide discovery process, and can be used to study deviations in the metabolism of compounds from a chemical synthesis program. The combination of NMR metabolite profiling and neural network classification is expected to be similarly relevant to other metabonomic profiling applications, such as in drug discovery.
© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: Acetochlor; Amitrole; Artificial intelligence; Benzisothiazole; Chlorsulfuron; Corn; Dinoseb; Diuron; Glyphosate; Imazamethabenz; Imazapyr; Imazethapyr; Metabolic profiling; Metabonomics; Naptalam; Neural network; NMR; Quinclorac; Sethoxydim; Sulcotrione; Sulfometuron; *Zea mays*

1. Introduction

The commercial herbicides all act on about 30 biochemically-distinct modes-of-action (MOA), as reviewed by Schmidt (1997). While enzyme assays are available to distinguish these, demonstrating the MOA for a compound is often laborious and time-consuming. In the search for safer and more efficacious pesticides, it is often desirable to: (1) establish which pathway a compound is affecting; (2) determine whether a novel analog has the same MOA as its parent molecule; or (3) classify the MOAs of novel leads found by screening. This should avoid involving well-exploited targets for which novel compounds are not needed (Petroff, 1988; Fiehn et al., 2000; Sauter et al., 1991).

The goal of this paper is to demonstrate that a robust, reliable metabolic profiling method can discern most MOAs targeted by commercial herbicides. We have selected 27 herbicidal compounds representing inhibitors for 19 different MOAs. Plants were treated for 24 h with these compounds and a ^{1}H NMR spectrum of a raw aqueous plant extract was recorded. A computational expert system was developed that can rapidly

* Corresponding authors.
E-mail addresses: karl-heinz.ott@bms.com (K.H. Ott), nelly.aranibar@bms.com (N. Aranibar).
[1] Present address: Bristol-Myers Squibb Co., 311 Pennington-Rocky Hill Rd. 3A-005, Pennington, NJ 08534, USA.
[2] Present address: Bristol-Myers Squibb Co., PO Box 4000, Princeton, NJ 08540, USA.
[3] Present address: BASF Plant Science, 26 Davis Drive, Research Triangle Park, NC 27709, USA.
[4] Present address: 391 South Milton Drive, Yardley, PA 19067, USA.

0031-9422/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0031-9422(02)00717-3

detect, classify and characterize the nature of the chemical treatment by the changes in the composition of the detected plant metabolites, even under conditions where changes in sample characteristics are very small (often close to the statistical variation between samples).

The term "metabonome" refers to the entire complement of low molecular weight metabolites inside a biological cell, and is also used to describe the observable chemical profile or fingerprint of the metabolites in whole tissue. The metabonome reflects the life history of each individual plant, including age and environmental factors such as soil type and moisture content, temperature, stress factors, and exposure to applied fertilizers and crop protection chemicals. With the expectation that, following exposure to a herbicide, the herbicide's mechanism-of-action might be recognizable in the plant's metabonome, we investigated whether such characteristics can be reliably detected in the NMR spectrum of a plant extract.

The gross chemical composition of various biological fluids has been investigated by a variety of chromatographic and spectroscopic techniques, notably gas and liquid chromatography (Petroff, 1988; Fiehn et al., 2000; Sauter et al., 1991), NMR spectroscopy (Nicholson et al., 1984; Ohsaka et al., 1979; Nicholson and Wilson, 1989; Lee et al., 1991; Bales et al., 1984; Rabenstein et al., 1988; Bell et al., 1987), mass spectrometry (Matsumoto and Kuhara, 1996; Wolfender and Hostettmann, 1996; Aharoni et al., 2002), and infrared spectrophotometry (Jackson and Mantsch, 1996). In animal and human fluids, much of the NMR research has been directed towards disease characterization and diagnosis (Sauter et al., 1991; Nicholson et al., 1984; Ohsaka et al., 1979; Nicholson and Wilson, 1989; Lee et al., 1991; Bales et al., 1984; Rabenstein et al., 1988; Nishijima and Fujiwara, 1997; Somorjai et al., 1996; Holmes et al., 1994; Hahn et al., 1997).

NMR has also provided information on biosynthesis (Lutterbach and Stöckigt, 1995; Prabhu et al., 1996; Weckwerth and Fiehn, 2002), on metabolism (Ratcliffe and Shachar-Hilt, 2001), and on the effects of herbicides on metabolism (Lutterbach and Stöckigt 1994, 1995) and mode-of-action (Hole et al., 2000; Hadfield et al., 2001), or used in investigations of whole plants (Schneider, 1997; Pope et al., 1993). A variety of computational methods have been applied for the statistical analysis of spectral data (Jackson et al., 1999; Shaw et al., 1995; Mansfield et al., 1997; Eysel et al., 1997), including artificial neural networks (NN) (Lisboa et al., 1997, 1998; Anthony et al., 1995; Hiltunen et al., 1995). In many cases, however, it was found that environmental factors contribute significant "noise" to the metabolite profile and reproducibility has often limited the applicability.

Furthermore, in many reports only two states (e.g. normal vs. treated) are simultaneously distinguished. A robust NMR method able to simultaneously detect many different treatment groups has not been described previously. In the search for new pharmaceuticals and crop protection chemicals, it is desirable to have a fast and reliable means to detect the mode-of-action of a new active compound, or pinpoint unusual phenotypes by an altered metabolic profile.

In a recent report (Araníbar et al., 2001), we showed that the ^{1}H NMR spectrum of a crude plant extract provides a fingerprint for the "metabonome", and automated pattern recognition was shown to establish the biochemical mode of action (MOA) for four different herbicide classes. In extension of this earlier work, additional compounds, representing nineteen different MOAs, were selected for simultaneous classification and we present a statistical validation for the methodology.

2. Results

A total of 430 ^{1}H NMR spectra of plant extracts were generated, representing plants treated with four different acetohydroxy synthase (AHAS) inhibitors, four different hydroxyphenylpyruvate dioxygenase (HPPD) inhibitors, two different glutamine biosynthesis inhibitors, and single inhibitors of ACCase, EPSPS, photosystems (PS) I and II, phytoene desaturase (PDS), 4 - hydroxyphenyl - pyruvate - dioxygenase (HPPD), 5-enolpyruvil-shikimate-3-phosphate synthase (EPSPS), glutamine synthase, dihydropteroate synthase (DHP), uncouplers of oxidative phosphorylation, and auxin, as well as systemic inhibitors of microtubule assembly, mitosis/microtubule organization, and cell wall (cellulose) synthesis. Spectra of 80 plants were treated only with the vehicle acetone and represent *controls* in this analysis. Typical spectra are shown in Fig. 1.

One goal of this work is to create a methodology that will enable researchers to rapidly screen novel compounds for herbicidal MOA by comparing their metabolic profile with those of previously characterized standards representing a range of commercially relevant herbicide targets. Model A represents a general-purpose neural network for classification of a wide range of compounds. A second refined model, Model B, is presented that is tailored to distinguish metabolite profiles of treatments that exhibit very small NMR signal differences between each other and/or the *controls*. Both models are cross-validated by using randomly selected subsets for training and testing. The models will be evaluated and applied in simulations to classify compounds novel to the NNs. Lastly, we demonstrate the use of a specialized NN for distinguishing treated from untreated (*control*) plants.

Fig. 2 outlines, in a flow diagram, the procedure used for the analysis presented here. The process reads its input patterns from a database of spectra for all com-

pounds. The user provides a mapping of spectra to user-defined output nodes. We present results for three different levels of output node assignments: (1) compound level: individual compounds each can be assigned a separate class, (2) MOA level: compounds known to affect the same pathway are assigned the same class, (3) treatment level: treated and untreated samples are separated into two classes. A pattern file is then created for all spectra from which a training set, a validation set, and an optional test set are created. The *test set* is created for the *leave-one-out* approach, by selecting a group of patterns corresponding to all spectra of a single compound or a group of compounds. Thus, the test set contains classes (MOAs) or individual compounds that are neither present in the *training set* nor in the *validation set*. The remaining patterns are subsequently divided, by random selection, into two approximately same sized groups of patterns: one used for training (*training set*) and the complementary used for validation (*validation set*). Thereby, each compound's pattern is represented in the *training set* and the *validation set* for *cross-validation*. We iterate over different random selection steps to create a population of 20 NNs. All results presented are averages over such populations. Every time a new test set is generated, the remaining patterns are used to create five new pairs of random subsets. All ten subsets are used to train a NN and classify the pattern present in the test set.

In the following, we will use some abbreviate nomenclature to enhance readability, as follows: For NN classes and associated patterns derived from spectra of extracts of plants that have been treated with a herbicide, we will use the name indicated in column MOA in Table 1 for that herbicide (e.g.: auxin for the pattern representing naptalam-treated plants). If more than one compound is used affecting the same pathway and we want to distinguish the patterns derived from the NMR spectra of the plant extracts individually, we will use the compound generic name, e.g. imazethapyr. "*Controls*" refers to spectra of plants treated only with acetone. *Unknown* refers to a pattern that is characterized by our procedure as unknown, according to the criteria specified in the experimental section. The terms "NMR spectra of plants" (spectra), "patterns for NN analysis" (pattern), and "metabonome" are used interchangeably.

2.1. Model A

Model A encodes one class for *controls* plus 17 classes for the different herbicide MOAs, as listed in Table 1, with all PS inhibitors combined into a single class. Following the procedure outlined in Fig. 2, 20 neural net-

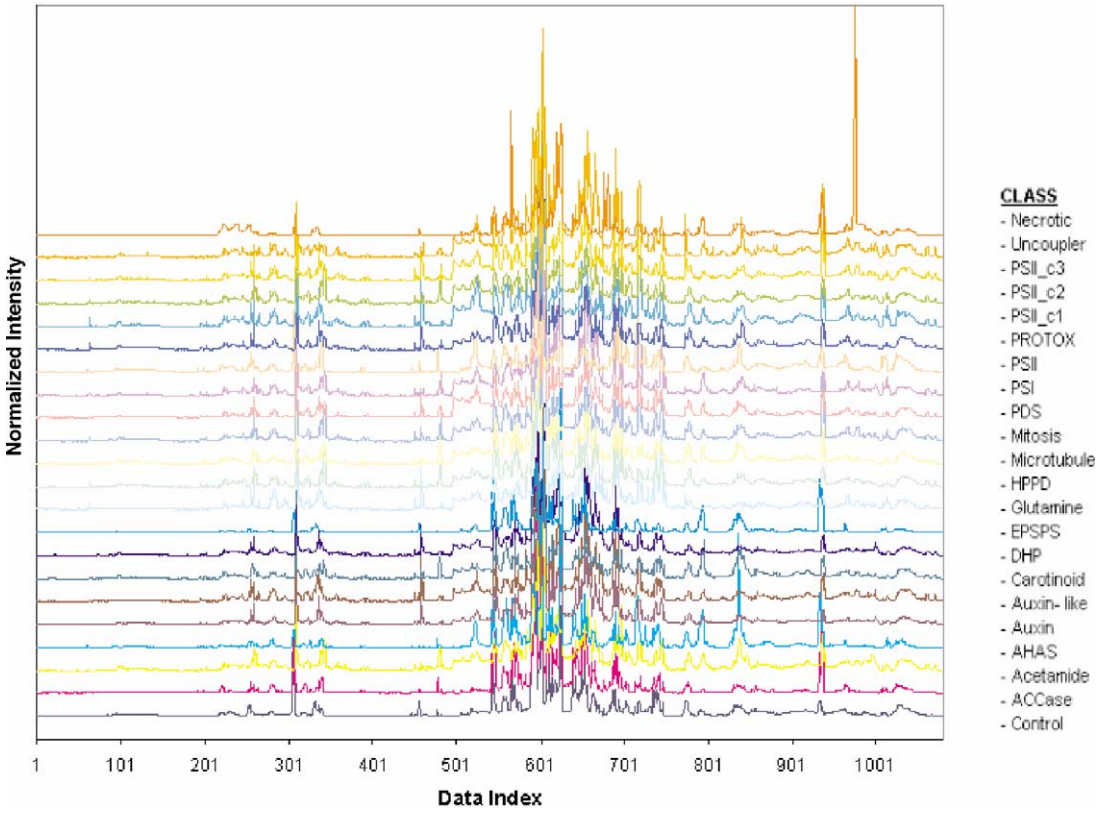

Fig. 1. ^{1}H NMR spectra of plant isolates representing nineteen different MOAs. The spectral region between 9.1–5.7 and 4.5–0.6 ppm is shown and used for analysis. All spectra are scaled to a total mean intensity of 1.0.

Table 1
List of herbicides studied, together with their biochemical mode-of-action and HRAC classification (see Schmidt, 1997)

Class name	Compounds	HRAC class	Mode-of-action
ACCase	Sethoxydim	A	Inhibition of acetyl CoA carboxylase (ACCase)
AHAS	Chlorsulfuron	B	Inhibition of acetohydroxyacid synthase (AHAS, ALS)
	Sulfometuron		
	Imazamethabenz		
	Imazapyr		
	Imazethapyr		
PSII_c1	Lenacil	C1	Inhibition of photosynthesis at photosystem II
PSII_c2	Diuron +	C2	Inhibition of photosynthesis at photosystem II
PSII_c3	Bromoxynil	C3	Inhibition of photosynthesis at photosystem II
PSI	Paraquat	D	Inhibition of photosynthesis at photosystem I
Protox	Acifluorfen	E	Inhibition of protoporphyrinogen oxidase (PPO, PROTOX)
PDS	Norflurazon	F1	Bleaching inhibition at phytoene desaturase (PDS)
HPPD	Sulcotrione	F2	Bleaching inhibition of 4-hydroxyphenyl-pyruvate-dioxygenase (HPPD)
	CL 836057		
	CL 818666		
	CL 836164		
Carotenoid	Amitrole	F3	Carotenoid biosynthesis inhibition (unknown target)
EPSPS	Glyphosate	G	Inhibition of EPSP synthase
Glutamine	Bialaphos[a]	H	Inhibition of glutamine synthase
	Glufosinate		
DHP	Asulam	I	Inhibition of DHP (dihydropteroate synthase)
Microtubule	Oryzalin	K1	Inhibition of microtubule assembly
Mitosis	Propham	K2	Inhibition of mitosis/microtubule organization
Acetochlor	Acetochlor	K3	Acetamide herbicide-like
Uncoupler	Dinoseb	M	Uncouplers of oxidative phosphorylation
Auxin-like	Quinclorac	O	Auxin-like (action like indole acetic acid)
Auxin	Naptalam	P	Inhibition of auxin transport

Class name indicates the name used for the Mode-of-action classes and patterns throughout this paper. CL 836057, CL 818666, and CL 836164 are proprietary herbicide lead compounds of undisclosed structure. + Diuron was applied foliar (class PS II_c2) and systemic [class PS II (root)].
[a] Formulation.

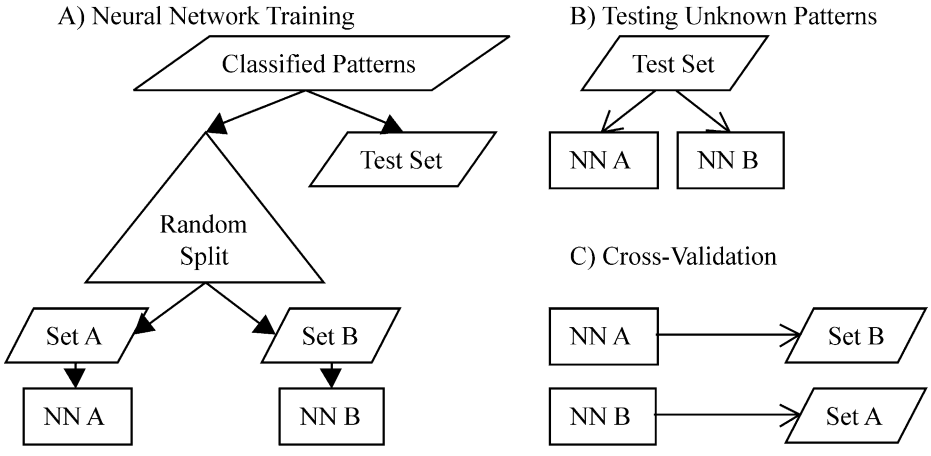

Fig. 2. Flow diagram for training and application of neural networks. (A) From the complete set of all spectra with their associated treatment classifications, two subsets, A and B, of nearly equal size can be created by random selection. Each subset is used independently to train a NN (*training set*), the complementary set is used as the *validation set*. This process is repeated 10–20 times to create a population of NNs. Optionally, specific patterns can be selected into a *test set* for later testing prior to the random selection into subsets. (B) Leave-one-out procedure: the *test set* contains patterns for one or more classes of compounds that are unknown to the NN. The NN is used to classify the *test set*. The pattern of the *test set* should be classified as *unknown* if no other compound was present in the *training set* that represented the MOA of the pattern in the test set. (C) Cross validation: the *validation set* is classified by a NN to produce statistics on the sensitivity and selectivity of the classification for compounds with pathways that are known to the NNs.

works were trained with randomly chosen subsets of the available spectra, and the complementary set of patterns were classified by the NNs. The results are summarized in Fig. 3, which shows graphically the average number of *correct*, *wrong*, and *unknown* classifications of the spectra of the *validation set* by the 20 different NNs. Overall, 64% of the spectra were classified correctly on an individual basis, and 30% of the spectra were classified as *unknown*.

Inhibitors of pathways affecting amino acid pools (e.g. AHAS, EPSPS, glutamine biosynthesis), fatty acid synthesis (ACCase) are consistently recognized, as is the photosystem II inhibitor, diuron when applied to roots. With only 6% of the samples classified as *wrong*, there is little confusion between the different classes, and most wrong assignments are observed in only one of the twenty different NNs. Some wrong assignments are observed between related MOAs. An unusually large fraction (10%) of glyphosate patterns is confused with AHAS inhibitors (discussed below). Other patterns, such as PROTOX, DHP, and, most notably, patterns of herbicides affecting the auxin transport, microtubule formation and mitosis have an increased pool of *unknowns*.

Confusion with *controls* is observed for several treatments in a few isolated cases (1–5%), but only Auxin patterns have significant percentage (20%) confusions with *controls*. Inspection of the NMR spectra reveals that many treatment pattern, most notably Auxin, Microtubule, and Mitosis show very little difference between each another and to the *control* samples. The microtubule inhibitor treated samples are also confused with HPPD inhibitors (7% wrong). Separate analysis shows that this confusion is largely caused by the inclusion of two very weakly herbicidal compounds into the HPPD class. The photosystem inhibitor class is assigned to several inhibitors that have, in turn, large fractions of *unknowns*. A similar calculation representing four separate PS classes for a total of 23 different classes, produces almost identical overall results (62% correct/ 27% *unknown*), and only small changes in the confusion between the different classes.

2.2. Model B

After identification of several batches of treatments by the NN described in Model A, we removed those treatments groups and performed a second round of classification for the remaining MOAs that had more than one third *unknown* classifications and were found to be more likely to be confused with one another. We also refined the analysis by using separate classes, PS I and PS II c1, c2, and c3, for the photosystem inhibitors. The refined NN (Fig. 4) improves the classification by removing some over-represented and strongly distinct signals, to focus on smaller differences between the remaining patterns. Overall, the recognition level has risen by about 20% for the MOAs that had previously been difficult to classify. In particular, microtubule, mitosis, and auxin are now more often recognized.

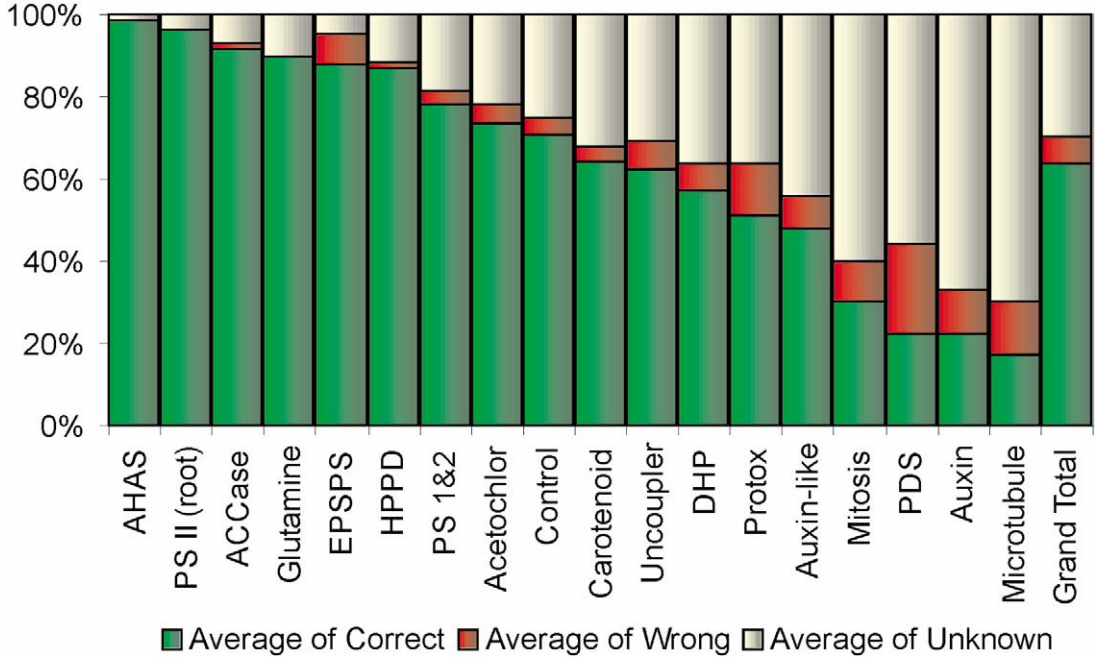

Fig. 3. Average number of *correct*, *wrong*, and *unknown* classifications of the NMR spectra by 20 different neural networks in Model A. A randomly selected subset of ca. half of the spectra was used for training, whereas the complementary set (not used in training) was classified automatically by the trained neural network. PS 1&2 refers to a class that is trained with all photosystem inhibitors.

The confusion matrix (Table 2) indicates that, while there are more frequently classifications confused between related pathways, PS I and the three different subclasses of PS II inhibitors separate well. PS II c3 has a metabolite profile that is very distinct from that of the other PS inhibitors while PS I, PS II c1, and PS II c2 have more closely related profiles. For example, about 10% of PS I pattern are classified as PS II c1 in average over all simulations. Similarly, 12% PSII c2 inhibitors are classified as PCII c1. Thus, the second step which is introduced in an attempt to enhance the sensitivity of the approach, simultaneously enhances selectivity.

Auxin and DHP get confused in some of the runs, which is reflected in increased percentage of wrong classifications for these classes, and also in a higher fraction of *unknown* classifications. Again, we do find

Table 2
Confusion matrix for Model B

Model B	Classification as percent recognition													
Actual class	PDS	PROTOX	PSII_c1	PSII_c2	PSII_c3	PS I	Uncoupler	Auxin-like	Auxin	DHP	Microtubule	Mitosis	Acetochlor	Unknown
PDS[a]	31					1			3	2				58
PROTOX		80	1								6	2		11
PSII_c1	3	1	50	3	1	1				1		2		38
PSII_c2			12	67								1		21
PSII_c3					93									7
PS I			10		1	53					1			35
Uncoupler				1	1		80		1		1			16
Auxin-like								71	1		1			27
Auxin	1								51	7	2			39
DHP									1	68	1			30
Microtubule		2						5	3	3	51	5		32
Mitosis	1		1	4							3	50	1	40
Acetochlor								1		3			70	21

Only MOAs were presented for which Model A lacked sensitivity. Rows indicate the actual treatment and columns represent the averages for the assignment by 20 independent NNs. The diagonal elements of the confusion matrix represent percent correct assignment whereas (non-zero) off-diagonal elements imply confusion between classes.

[a] PDS is represented by only six spectra in total.

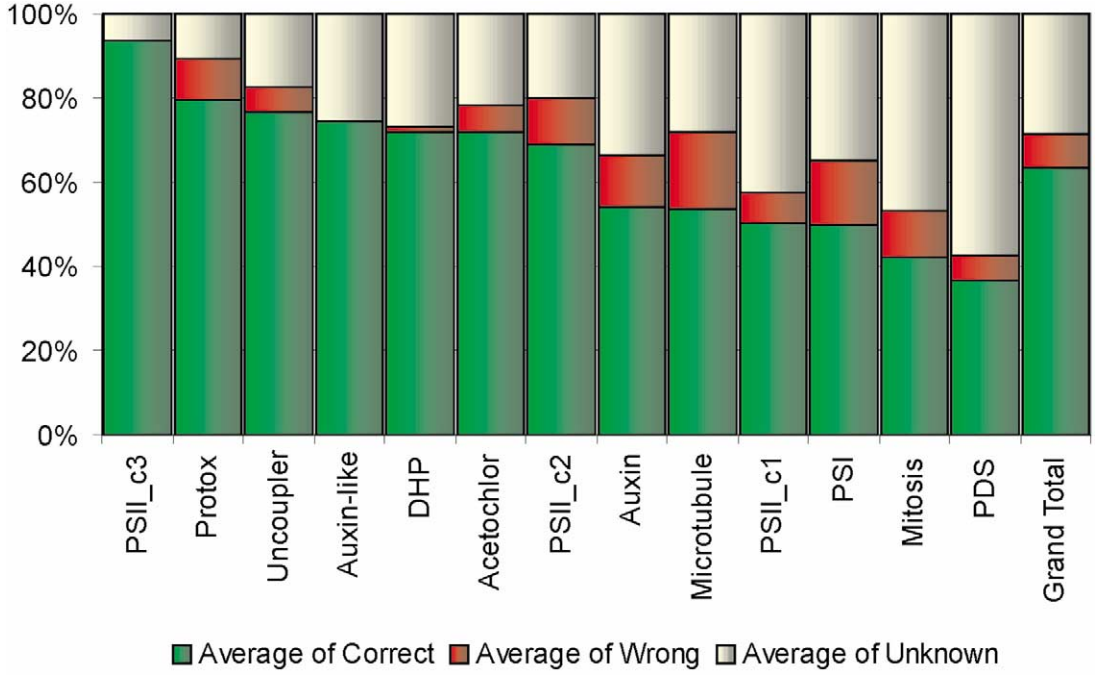

Fig. 4. Summary of classification results for Model B. Averages of percent recognition of total for each compound/MOA is shown for 20 NNs, each trained with a different random selection of half the spectra classifying the complementary set. The compounds tested were all represented in, but not part of, the training set. Each column corresponds to a different class.

that microtubule and mitosis have a "weak" signature that is frequently, but not consistently, confused with other MOAs.

The results of the two calculations listed above are not only representative by virtue of performing repeated runs with different selections of spectra used to train the network. We find that very similar results are achieved when classification schemes are changed. The best overall results achieved so far with this data set are for a 15-class NN (classes: *control*, AHAS, HPPD, PS II (root), glutamine, PSI, PS II c3, EPSPS, carotenoid, protox, PS I-PS II c1/c2, auxin-like, DHP, uncoupler, acetochlor) where PS II c1, and PS II c2 are combined into a single class, and microtubule, mitosis, and auxin inhibitors are not part of the training. This NN has overall 85% correct, with $>70\%$ recognition for any included MOA, and only 13% *unknown* and 2% *wrong* classifications.

2.3. Application of Models A and B

How do the models presented in calculation A and B perform when a new compound is presented that is not part of the training set? Which MOAs are easily confused with others? How sensitive and how selective is the method in situations with overlapping or partially divergent MOAs?

To answer these questions, we designed a leave-one-out procedure in which we remove one compound at a time from the data set and calculate 10 NNs, using 10 different random selections of half of the remaining spectra for training (the other half is disregarded). We then present the pattern removed in the beginning to the NN for classification. If a compound is novel to the NN and there is no related compound in the training set, we expect the NN to issue an *unknown* classification. If other compounds representing the MOA of the compound presented are in the training set, we hope to find this compound to be correctly classified. Related MOAs are expected to be partially activated. Partial activation is represented in the NN in the actual activation values of the output nodes. Since those numbers are difficult to present in the format of a publication, we use the average of *correct* classifications over a series of related networks as a measure of relatedness, given the rules laid out in the experimental section.

The results of the leave-one-out procedure are summarized in Figs. 5 and 6. We will discuss four different situations: (1) a group of chemically diverse compounds has the same MOA; (2) a group of compounds from a series believed to target the same enzyme are metabolized differently by the plants; (3) A group of com-

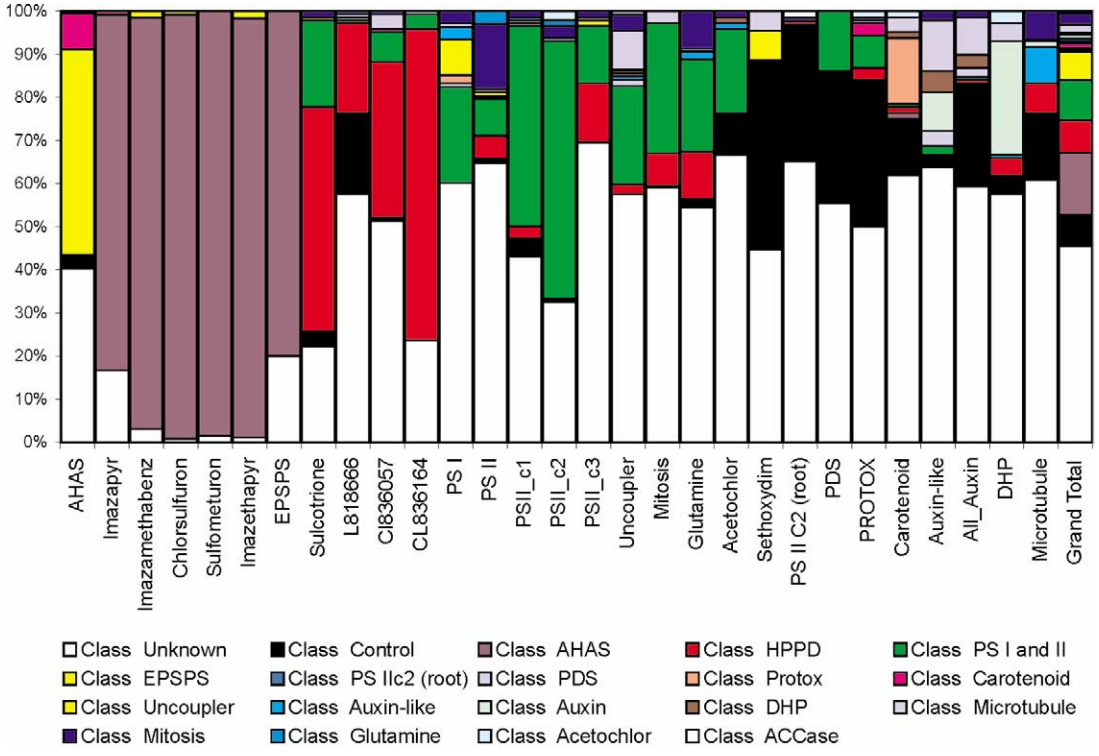

Fig. 5. Results from "leave-one-out" computations. Each bar represents average classification result of 10 NNs for the compound/compound group indicated. Each group of 10 NNs was trained with all spectra except those for compounds or groups of compounds indicated on the horizontal axis. The colors refer to the class the classified spectra were assigned to. For example, in the first bar, all AHAS inhibitors were removed before training 10 NNs with randomly selected subsets of 50% of the remaining patterns. The AHAS inhibitors are classified as ~40% *unknown*, 48% EPSPS, 8% carotenoid, 5% *control*.

pounds affects different steps in the same pathway; (4) A compound represents an entirely new MOA.

2.4. Co-classification of chemically distinct compounds by their common MOA

The imidazolenone and the sulfonylurea herbicides, as well as many other commercial herbicides, inhibit the AHAS enzyme. We chose five of these herbicides having a range of different specificities, but all targeting AHAS. We had previously shown that a NN trained to recognize the metabonomes of plants treated with imazethapyr, glyphosate, two other herbicides, and controls recognizes >99% of the metabolite profiles of other AHAS inhibitors into the AHAS MOA. Extending this approach, we now included more MOAs into the NN models and performed a more rigorous, cross-validated approach.

Removing one compound from the training set, leaving four compounds as AHAS representatives for training, more than 90% of the samples are classified correctly, with most AHAS inhibitors having more than 95% *correct* classifications. Imazapyr has only 83% *correct* classifications and 17% *unknown* classifications. This result reaffirms our earlier findings (Araníbar et al., 2001), but now, the statistical significance is higher since the recognition is above the background of many more alternative MOAs.

Using only one of the four AHAS inhibitors together with all other MOAs in the training of the NN, decreases the sensitivity as there are only about six compounds remaining in the training, resulting in about 20–30% *unknown* classifications. However, of the positive classifications, ~80% are true positive assignments. This average is reduced by over 10% by poor recognition when imazethapyr is used as representative for the AHAS MOA within the training set. We attribute this to the divergence between the individual NMR spectra since the imazethapyr samples had been collected in the very beginning of the study when we lacked experience in reproducibly collecting the samples, and the growth chamber was set 3 °C lower. Most of the difficulty in recognizing different compounds affecting the AHAS enzyme are caused by the presence of glyphosate as a EPSPS inhibitor with a similar metabolite profile.

If all AHAS inhibitors are removed from the training, AHAS becomes a novel MOA for the network. In this case we find that about half the samples treated with AHAS inhibitors are (wrongly) classified as EPSPS inhibitors, and about 40% are *unknown*, as expected. Also, vice versa, glyphosate will be classified as an AHAS inhibitor if no sample from a glyphosate-treated plant was present during training. AHAS and EPSPS are in different pathways, and in general, the network is capable of separating these MOAs, as long as the NN

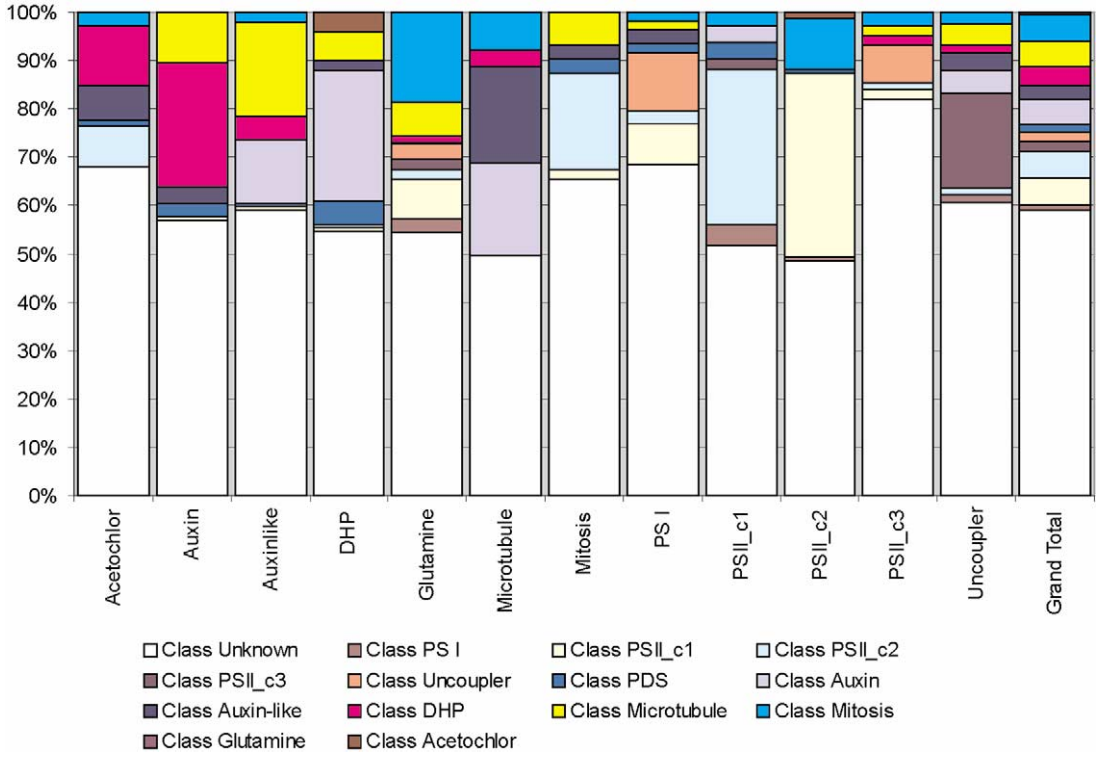

Fig. 6. Classification of compounds unknown to the NN (i.e. not included in training). Each bar represents the classification of the compound indicated on the horizontal axis by 10 different NNs. The NNs were untrained in the compound/MOA presented, but trained with all other MOAs. The compounds presented to the NN represent a "Novel MOA" to the NN. An *unknown* classification is the expected correct answer.

has been trained to do so. However, there is an average of ~10% of glyphosate samples assigned to AHAS even when glyphosate is represented in the training. (The higher variability in the imazethapyr NMR spectra, discussed above, is mostly responsible for the false positive assignments.) The NMR spectra of AHAS and glyphosate treated plants are very similar with only very few proton resonances different between the two populations while there are a considerable number of signals that commonly change with respect to the *control* and other MOA spectra. Many of those signals can be assigned to amino acids and we find that inhibition of amino acid metabolism can increase the pool of free amino acids, presumably due to increased protein turnaround. While the composition of the amino acids found changed in both populations is different, the communality dominates if the NN is not specifically trained to recognize the smaller differences. Thus, both MOAs share similarities in the resulting metabolite profile. The differences are due to the levels and types of amino acids that accumulate.

Inhibition of glutamine synthase, in contrast, has a very different profile, lacking the increase of amino acid pools but distinguished readily by several resonances and we attribute several of the resonances of the glutamine biosynthesis inhibitors to components of the formulation rather than to natural metabolites.

2.5. Same target, different metabolic fate

As a challenging example relating to a lead optimization problem, we had selected three chemically analogous compounds from a series of experimental HPPD inhibitors, and sulcotrione as a commercial herbicide representing a different chemical class. Corn is resistant to sulcotrione. From the remaining compounds, one compound is highly active, one is very weakly active in vivo, but was predicted as highly active in a quantitative structure–activity relation (QSAR) study (data not shown). The last sample appears much more potent than was predicted by QSAR. Since this set is so diverse in its in vivo activity, the signatures are less distinct in the context of the many other MOAs. This is reflected in an increased number of *unknown* classifications, ranging from ~25 to ~55%. However, the correct MOA assignment still dominates the positive classifications in all cases and a more specialized NN can also highlight the more subtle differences between these compounds. When using each compound, in turn, as representative in the training, the very active compounds reveal a very similar profile, while spectra of the very weakly herbicidal compound are often confused with *controls*, and patterns that are very similar to those of *controls,* like Microtubule. (Removal of the weak HPPD inhibitors from the training set does, in turn, improves slightly the sensitivity of recognizing of some of these patterns.)

2.6. Pathway recognition

Co-classification of PS inhibitors into a single class is a model for recognizing compounds that inhibit different related biochemical functions. Fig. 5 demonstrates that the photosynthesis inhibitors do, to some degree, co-classify if the network is trained with a combination of three of four of the PS I, PS II c1, c2, c3 inhibitors. PS II c1 and PS II c2 are well recognized into a related class with most of the positive classifications being correct. The results (~1/2 *unknown*, 1/2 shared class assignments) are similar to the pattern observed for the HPPD inhibitors as described above. The majority of the positive classifications of PS 1 are also correct, but several other MOAs have a similar large percentage (20–30%) classified as photosynthesis inhibitors.

Applying the more specific and refined model B that has each PS inhibitor as a separate class (Fig. 6) indicates, in concordance with the analysis of the confusion matrices during the validation runs, that while PS II c1 and PS II c2 have closely related profiles, PS I is more distinct. PS II c3 has little in common with the other PS inhibitors, but shares some features with uncouplers.

2.7. Novel MOAs

Several of the MOAs are represented by a single compound in the present study. Thus, removing these compounds before training the NN simulates results for compounds belonging to novel MOAs. We would desire that compounds belonging to a MOA that was not represented in the training should be classified as *unknown*. Every other classification would be considered *wrong*. For many compounds presented to an NN, we find that, for new MOAs, about 60% of the classifications are in fact *unknown*. The remaining 40% are variable classifications. For practical purposes, we are mostly concerned when a single "*wrong*" classification dominates, since this could cause false positive conclusions. Using Model A, several compounds have 20–30% of their patterns classified incorrectly as *control*. Application of the control model (below) can characterize these compounds as "*treated*" and thus identify them as novel MOAs. Incorrect classifications as *controls* appear to be an indication that there is very little change in the metabolite profile caused by these compounds. Those changes will only be picked up if such a MOA is specifically presented to the NN. In addition, the NN training over-weights the untreated samples (e.g., 80 *controls* vs. 12 *treated* spectra), and the *controls* show greater experimental variation due to our experimental design.

Applying our more specialized NN, Model B, also overcomes many of false positive classifications, as illustrated in Fig. 6. Now, all patterns have more than 60% *unknown* classifications, one of our empirical cutoffs for novel MOAs. The majority of the "novel compounds"

has no consistent *wrong* classifications to another class and can be attributed to noise, i.e. experimental variability, especially for treatments that cause little change in the metabolic profile.

Auxin and DHP have about 24% classifications confused between each another. We found, by comparison of the NMR spectra, that one batch of DHP has a very distinct metabolite profile from that of auxin, but the other batch of DHP lacks several metabolites present in the first batch and resemble more closely spectra of *control* and auxin.

2.8. Control model

Specialized NNs that are optimized to recognize a specific treatment versus all others can be more sensitive and specific. From the results presented above, it is apparent that distinction of samples treated with a compound versus samples treated with a blank solution sometimes poses difficulty. In the following we evaluate whether a specialized NN to distinguish treated and untreated samples might further reduce the already small error rate. In a modified calculation, the samples were classified into two subsets, *treated* and *control*, i.e. those treated with a compound solution and those treated with a blank solution. We calculated the average over ten classifications using the cross-validation procedure outlined in Fig. 2.

As shown in Table 3, 96% of the spectra from treated plants are recognized as *treated*, and only 3% were false negatives, if other spectra of the same treatment were included in the NN training. *Controls* are recognized as such in 82% of all cases, with 15% false positives (*controls* misclassified as *treated*).

To further validate the control model using the leave-one-out method, we also removed, in turn, one compound, and also all AHAS, and all HPPD inhibitors at once, to simulate how such a binary model would perform when a new compound, previously unknown to the network, would be introduced. If a particular treatment was not known to the NN, the average true positive rate for the data is still 89%, with 9% false negatives, as shown in Fig. 7, indicating that there is a strong signature that characterized treated plants.

As expected, best results are usually achieved if other compounds of a series, or with a similar MOA are included in the training. Most HPPD and AHAS inhibitors are consistently classified as treated, as long as other inhibitors of that class are included. Even if all AHAS pattern are excluded from the NN training, the patterns are still recognized as *treated* in >95% of all cases. Many other inhibitor patterns, like patterns of photosystem inhibitors, are also well recognized, possibly due to partial overlap with patterns included in the training.

Table 3
Statistics for the control model

Control model	Classification as percent recognition		
Actual class	*Treated*	*Control*	*Unknown*
Treated (*known*)	96	3	1
Treated (*unknown*)	89	9	2
Control	15	82	3

Treated (*known*) refers to average results of 10-fold cross-validated NN runs in which all MOAs were part of the training procedures. *Treated* (*unknown*) refers to the results of runs in which, in turn, each compound or MOA group was first removed from the data set, after which the 10-fold cross-validation procedure was run, and the spectra of the compounds/MOAs that were excluded were classified by the resulting NNs. This simulates the NN classification for a novel compound or new MOA.

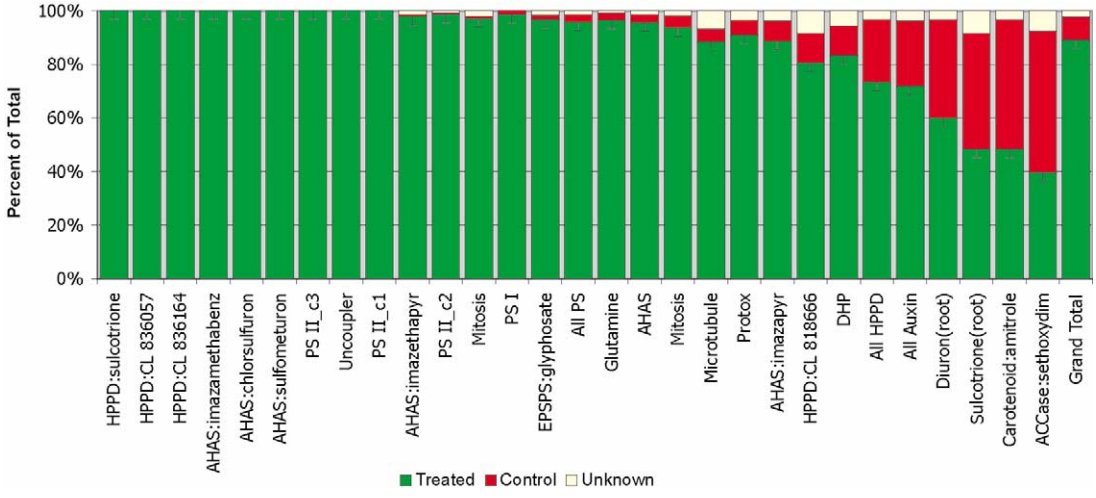

Fig. 7. Results of recognition of unknown treatment classes by the control model. Each column represents the average fraction of *correct*, *wrong* and *unknown* classification of 10 NNs that were trained without the compound or group of compounds used in the training.

3. Discussion

3.1. Growing conditions

One of the most important requisites for the work on metabolic profiling in plants is the stability and reproducibility of the physical conditions in which the plants are grown. Plants, as all living organisms, react to different environmental stimuli and changes that turn on and off different genes expressing different proteins and enzymes, and developing different metabolic states, usually the most appropriate for the best development of the organism in the given environment.

In the early developmental stage (5–10 days after germination) in which the seedlings in this study were treated and harvested, metabolic changes are fast and changes in the concentrations of metabolites are considerable for the small amount of growing point tissue that can be collected. Relative small changes in the environment of a plant can be reflected in very detectable variations in the absolute concentration of a metabolite and with that, a change of the profile.

For these reasons, the use of growing chambers, where the environmental conditions can be accurately controlled, is mandatory. In the course of the present study, for example, some plants had to be transferred from one growing chamber to another, due to the mechanical failure of the first one. Several hours at a more elevated temperature and then change in illumination produced detectable differences in the metabolic profiles. The NN can be trained to either recognize or ignore these changes in environmental conditions. Thus, it is clear that the use of green houses and field plots are not appropriate for growing the plants used in this kind of study. This observation may have implications for other kinds of profiling, e.g., gene expression profiling.

3.2. NMR spectroscopy

The use of an acidic matrix to prepare the extracts of plant tissue allowed us to isolate the widest range of primary metabolites (amino acids, sugar, sugar-alcohols, organic acids, etc.). Due to the relative low sensitivity of NMR spectroscopy, it is important to choose as many of the metabolites present in the highest concentrations as probes for the total metabolic profile. This extraction matrix does not produce any undesirable solvent peaks in the NMR spectrum. Reproducibility of the NMR operating conditions is the key for a reliable classification of the spectra. Temperature and spectral width seem to be the most important factors. The exact total concentration of metabolites in the sample (which is dependent on the amount of tissue used for extraction) is less critical for two reasons: (1) use of an internal reference standard in each sample, and (2) normalization of all the spectral intensities as part of the pre-processing of the spectra when preparing patterns for analysis with the neural network software.

Many replicates of each sample were prepared and measured in each experiment. Usually 5–12 plants were grown, treated, and harvested for each treatment class. Each experiment was repeated at least twice at different times. We find that there is, even under tightly controlled condition a slight "batch" factor in which samples of one batch tend to cluster together. This only becomes a problem when experimental conditions have changed or if the discrimination is already weakened by other factors, such as too many similar pathways spread over too many nodes. Since NNs can be trained to recognize fluctuations in conditions, it is recommended to always include, with each batch of treatments to be classified, a few reference samples of the MOAs that are most likely to be targeted by the compounds under investigation.

3.3. Pattern recognition

We have presented the results for a full NN model that simultaneously recognizes a wide variety of metabolic profiles with a high success rate and confidence. Most importantly, we find that compounds affecting the same MOA have related NMR spectra and can be distinguished from a wide range of other MOAs with high confidence. Compounds not previously known to the NN co-classify with other compounds affecting the same MOA. Related MOAs are sometimes indicated by an increased fraction of patterns of a treatment being classified to a second MOA.

MOA classes that are part of the NN *training set* are usually well recognized. Inhibitors that affect pathways that are involved in the metabolism of common, soluble cellular components, for example inhibitors of the amino acid metabolism pathways, are the most distinct and are detected with high confidence. Other inhibitors do not create large changes in the profile of soluble compounds compared to *controls*: the auxin, mitosis, and microtubule MOAs are difficult to classify in the background of the many other compounds and produce a larger fraction of *unknown* classifications. Nevertheless, even in these more difficult cases, there are only a small fraction of false positive classifications and even those samples are classified with high confidence by the control model as *treated*.

The NN method is often capable of handling closely related pathways, and we find that the analysis of the confusion matrix for compounds affecting closely related MOAs yields fruitful insights in the particulars of each compound, and highlight similarities as well as differences in their activity and metabolic fate. For example, we found confusion between patterns of PS I, PS II c1, and PS II c2, but not between these patterns

and that of PS II c3. Thus, the separation and analysis of the confusion pattern yields insight into the response patterns that are created by the different inhibitors of the photosystem I and II and their subsystems. The analysis of the confusion matrix for NNs trained with a single inhibitor of a series, classifying other compounds in a series, as discussed for the HPPD inhibitors yields deeper insight into the differences in metabolic fate. Compounds not active in corn due to their limited uptake or rapid metabolism co-classify with highly herbicidal compounds of the same chemical family, but at a reduced NN output activation level. In addition, the alteration in the metabolic fate may also be indicated when samples of a treatment are also classified by the NNs into other classes at elevated percentages ($>5\%$). For novel compounds or compounds for which the MOA is not well established, the MOA might not be represented in the *training set*. We simulated this scenario by removing a complete class of compounds prior to training. The results of the "leave-one-out" experiments highlight a critical feature of the method. A NN trained to discern *treated* and untreated samples classifies active herbicides with negligible small false negative rate to the treated group (see discussion of the Control model). In a detailed model, like Model A or B, novel compounds are generally assigned to the correct MOA or pathway if this pathway has been defined during the training by the NN. Furthermore, if the pathway is not known to the network, that is the NN has not seen a mechanistically-related compound, we are likely to get a majority of *unknown* classifications. If related pathways are present in the training, we are likely to find that more than 20% of all classifications point to the related MOA(s). We find such a situation for the related PS inhibitors and the HPPD inhibitors that have very different activity levels. Compounds of sufficient high herbicidal activity affecting the same MOA co-classify at a high proportion. However, caution is indicated when a novel compound affects an MOA that is not known to the NN and the profile of the novel MOA has many overlapping features with an MOA that is known to the NN (the confusion of AHAS inhibitors and glyphosate demonstrates this). The best safeguard against this type of "false positive" is the inclusion of as many MOAs as possible into the NN and the observation of additional experimental evidence, e.g. the plant phenotype.

The general purpose model, Model A, produces satisfactory results for many MOAs and might suffice in praxis for many applications. The model can be generalized in many ways, and other class assignments can be chosen. In the variations we studied, we found little change in the overall success rate upon using different classification schemes (like various combinations of MOAs in single or split classes), as long as treatment classes were not entirely removed. The particular models detailed in this report were chosen to exemplify different levels of refinement and a stepwise approach that is most likely to be used in a research setting.

The two step procedure was guided by our quality control procedures that had indicated that there are spectra, that include photosystem, mitosis, microtubules, auxin classes, etc., that are statistically very similar (data not shown), an observation that is confirmed by visual inspection of the overlay of the NMR spectra. Also, *controls*, AHAS and HPPD inhibitors were largely overweight in the training of Model A, since multiple compounds of the same MOA were present. Model B has the treatment regimes more equally represented.

Because these experiments are subject to normal biological variation, it is unrealistic to expect 100% accurate classification at all times. Some plants might be less susceptible to a given herbicide than others and their metabonome would be less affected. In such cases, a treated plant might be wrongly classified as a *control*. Extraneous effects might cause changes in the NMR spectrum, causing classification of some treatments as *unknown*. Different MOAs that result in similar metabolite profiles will be confused with each other, while other MOAs might have too small an effect on the NMR spectrum to be classified. Ultimately, it will be necessary to set a threshold or cut-off for acceptance of a correctly classified MOA. In general, we find that if more than 80% of all patterns of a batch are classified consistently, these assignments can be trusted. As the *unknown* fraction for a batch approaches 50%, increased caution is advised.

In cases where the NN responds to new samples with over 60% *unknown* classifications, the MOA of the new sample might indeed be *unknown*. A specialized NN including only those MOAs that have similar metabonomes can improve selectivity. Those compounds that retain a large number of *unknown* classifications and also have a larger number of confusions with other MOAs or *controls* will need close scrutiny.

The particular choice of class assignments, classes included in training, the mix of spectra included in the training and other factors seem to affect the particular outcomes to less than 3%, on average. This implies that the operator has only very limited influence on producing a particular outcome, except for avoiding particular MOAs or compounds.

Most variability between the NMR spectra within a group is found for *controls*. Every batch was accompanied by controls, leading to many more samples, and thus reflecting the overall variation between the batches over a period of more than 1 year. Also, plants that did not grow to the required size before treatment were sometimes included as *controls*. Most notably, we observe that the reproducibility between all spectra has increased with the experience of the scientist running the studies such that for the first few batches, correlation

coefficients between samples (regardless of treatment) is better than 0.8, while after all details of the procedures were fully established correlation coefficients were consistently better than 0.9.

Most false negative (*control*) assignments are attributed to the lack of a sufficiently diverse training set. For example, only a few MOAs are represented by compounds applied to the medium and thus our control model lacks mostly in recognizing pattern for compounds applied though the medium. The same compounds applied foliar are recognized as *treated*. A more representative data set, with more compounds for each class, and consistent application schemes should overcome these difficulties.

Selectivity and sensitivity depend to a large degree on factors other than the patterns themselves, for example the presence of MOAs used in the training and the granularity of the class assignments for related MOAs. Manipulation of the analysis scheme can achieve increased success rates. If the operator has some knowledge of the MOAs that a set of compounds might affect, it might be advisable to reduce the number of MOAs in the *training set*. More specialized NNs often will show increased robustness of the assignments. Conversely, selectivity and sensitivity drop when the NN is forced to separate between signatures for closely related patterns, i.e. to distinguish too many closely related pathways. However, we strongly favor inclusion of several MOAs in the *training set* to avoid creating signatures that are unrelated to treatment per se, such as stress markers, rather than a specific compound profile. In particular, ''false positive'' assignment (assignment of a compound to the wrong MOA) can largely be avoided when enough related MOAs are included to act as positive controls.

4. Conclusions

This work has shown the feasibility of ^{1}H NMR spectroscopy of plant extracts, in combination with artificial neural network analysis, to distinguish treated from untreated (*control*) samples and discriminate, with high reliability, the modes-of-action of many different, commercially important herbicides. Easily obtainable extracts from plants, analyzed by 1D ^{1}H NMR contain a wealth of information about the treatment of the plants. NMR is sensitive enough to produce fingerprint information that enables the researcher to discern between related MOAs and about twenty MOA classes have been discerned by the automated pattern recognition approach. Compounds affecting the same target enzyme are classified by their metabolic profile to the corresponding MOA, even if only one reference compound is used to create the signature for that MOA. Compounds with novel MOAs are classified as *unknown*. Detailed analysis also highlights differences between compounds of a series that affect the same target but that are being metabolized differently. Of the 19 MOAs studied, the control group (untreated), AHAS, HPPD, ACCase, EPSPS, PROTOX, carotenoid, PS-I, uncoupler, auxin-like, acetochlor, PS II, and glutamine synthase inhibitors were all well classified (little or no confusion with control plants or other MOAs). For MOAs that have closely related metabolite profiles, enhanced sensitivity is achieved when a specialized NN is used that includes only the closely related MOAs. Such a stepwise process can be included into an expert system to classify metabonome profiles of all treated plants with high confidence. The method is reliable when the experimental conditions are well controlled and accurately kept under standard conditions. There exists a large potential for similar applications in the agricultural and pharmaceutical industries, as many biological tissues are amenable to study by metabolic profiling.

5. Experimental

The plant preparation methods were as described previously in Araníbar et al. (2001). In brief, *Zea mays* seeds (Pioneer 3514) were set to germinate for 5-days in a controlled-environment growing chamber. The plants were treated post-emergence with the herbicides shown in Table 1. Twenty-four hours post-treatment, the plants were harvested and the meristematic tissue (approximately 250–300 mg per plant) was collected, flash frozen in liquid nitrogen, and stored in a liquid nitrogen freezer until further use. The plant meristems were then pulverized, suspended in 0.25 N HCl, and centrifuged. The supernatants or plant isolates containing the soluble metabolites were separated and reserved for ^{1}H NMR spectroscopy.

For each compound, treatment was repeated in at least two separate batches, each containing six individual plants, resulting in at least 12 spectra per compound. While conditions were kept as constant as possible for the treated plants, some of the control plants reflect small variations in environmental conditions and growth stage. The batches of plants were spread over a period of more than 1 year, and a few plants were grown at a slightly elevated temperature (due to a malfunctioning temperature controller). Treatment of plants with AHAS inhibitors, sethoxydim, glyphosate, and two batches of diuron were applied to the media, while all other inhibitors were applied to the leaves (''foliar''). The following data were excluded from most of the analysis due to the lack of sufficient samples for randomized training and testing: (1) two glyphosate treated plants were killed rapidly and were decaying after 24 h; (2) a single batch of six PDS treated

plants was ignored in some analysis, due to the lack of a second batch; (3) for the detailed analysis in the latter part of this paper, we removed one batch with six samples of *control* plants and twelve samples of imazethapyr-treated plants because the NMR spectra were recorded at a higher temperature; (4) we also removed one *control* sample that showed strong stress response signals.

The NMR profiles were classified using a supervised pattern recognition approach in which a neural network is "trained" using a set of NMR spectra for plant extracts whose origin and nature is well known, i.e. with known herbicide treatments, known genetic phenotypes, etc. The NMR spectra are "memorized" as patterns during the neural network training step. When the spectrum of an "unknown" extract is presented to the trained network, it will be recognized only if it is a member of the training set; otherwise, it will not be recognized and will be flagged accordingly.

The SNNS (Stuttgart Neural Network Simulator, University of Stuttgart, Stuttgart, Germany) software was encapsulated into a user interface that reads as input a definition of a network topology, spectra to be used to train the network, and spectra to be classified. The output of the classification run is analyzed automatically and converted into tabular and graphical form. For the NN, a three-layered, fully-connected topology is defined with 1080 input nodes (representing the spectral data points after preprocessing), 12 hidden nodes, and up to 30 output nodes. All nodes are characterized by a logarithmic input function and unity output function. Random values are assigned to each parameter initially, and the resilient backpropagation algorithm is used for optimizing the weights, which are updated for 500 iterations in topological order. We use an initial update value of 0.1 and a maximum step size of 50. The NN is trained by presenting a subset of the pattern to a suitable network topology and, after training, the network can classify the metabonome represented by the NMR spectra of samples other than those used in the training. The output of the classification step is in the form of output unit activation values. The procedure employed converts the activation values into a more readable classification by assigning a classification to the spectra if a single output node has an activation value >0.6 and no other output node has activation values >0.4. Otherwise, the spectrum is classified as *unknown*. The classification for each spectrum by the NN is recorded and compared to the actual treatment of the corresponding plant. The number of *correct* and *wrong* classifications are tabulated, and are shown as bar-graphs, together with the spectra that were classified as *unknown* by the NN and that are counted separately. The classifications are also displayed in the form of a Confusion Matrix, whose rows indicate the actual treatment and columns represent the assignment generated by the NNs. The diagonal elements of the confusion matrix represent correct assignments, whereas (non-zero) off-diagonal elements imply confusion between classes. In addition, analysis can be performed for batches of samples that received the same treatment rather than an individual sample, thus reducing the possibility of false conclusions.

References

Aharoni, A., De Vos, C.H.R., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., Goodenowe, D.B., 2002. Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. OMICS 6, 217–234.

Anthony, M.L., Rose, V.S., Nicholson, J.K., Lindon, J.C., 1995. Classification of toxin-induced changes in ^{1}H NMR spectra of urine using an artificial neural network. J. Pharm. Biomed. Anal. 13, 205–211.

Araníbar, N., Singh, B.K., Stockton, G.W., Ott, K.-H., 2001. Automated mode-of-action detection by metabolic profiling. Biochem. Biophys. Res. Commun. 286, 150–155.

Bales, J.R., Higham, M., Howe, I., Nicholson, J.K., Sadler, P.J., 1984. Use of high resolution nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine. Clin. Chem. 30, 426–432.

Bell, J.D., Brown, J.C.C., Nicholson, J.K., Sadler, P.J., 1987. Assignment of resonance for acute phase glycoproteins in high resolution proton NMR spectra of human blood. FEBS Lett. 215, 311–315.

Eysel, H.H., Jackson, M., Nikulin, A., Somorjai, R.L., Thomson, G.T.D., Mantsch, H.H., 1997. A novel diagnostic test for arthritis: multivariate analysis of infrared spectra of synovial fluid. Biospectroscopy 3, 161–167.

Fiehn, O., Kopka, J., Doermann, P., Altmann, T., Trethewey, R.N., Willmitzer, L., 2000. Metabolite profiling for plant functional genomics. Nat. Biotechnol. 18, 1157–1161.

Hahn, P., Smith, I.C.P., Leboldus, L., Littman, C., 1997. The classification of benign and malignant human prostate tissue by multivariate analysis of ^{1}H magnetic resonance spectra. Cancer Res. 57, 3398–3401.

Hadfield, S.T., Hole, S.J.W., Howe, P.W.A., Stanley, P.D., 2001. Metabolite profiling by NMR for high-throughput mode of action identification of screen hits. Weeds 2, 551–556.

Hiltunen, Y., Heiniemi, E., Ala-Korpela, M., 1995. Lipoprotein-lipid quantification by neural-network analysis of ^{1}H NMR data from human blood plasma. J. Magn. Reson. B 106, 191–194.

Hole, S.J.W., Howe, P.W.A., Stanley, P.D., Hadfield, S.T., 2000. Pattern recognition analysis of endogenous cell metabolites for high throughput mode of action identification: removing the postscreening dilemma associated with whole-organism high throughput screening. J. Biomol. Screen. 5, 335–342.

Holmes, E., Foxall, P.J.D., Neild, G.H., Beddell, C., Sweatman, B.C., Rahr, E., Lindon, J.C., Spraul, M., Nicholson, J.K., 1994. Automatic data reduction and pattern recognition methods for analysis of ^{1}H nuclear magnetic resonance spectra of human urine from normal and pathological states. Anal. Biochem. 220, 284–296.

Jackson, M., Mantsch, H.H., 1996. Biomedical Infrared Spectroscopy. In: Mantsch, H.H., Chapman, D. (Eds.), Infrared Spectroscopy of Biomolecules. Wiley-Liss, New York, pp. 311–340.

Jackson, M., Mansfield, J.R., Dolenko, B., Somorjai, R.L., Mantsch, H.H., Watson, P.H., 1999. Prediction of breast tumor grade and steroid receptor status by multivariate analysis of Fourier transform infrared spectra. Cancer Detection and Prevention 23, 245–253.

Lee, H.-S., Chung, Y.H., Kim, C.Y., 1991. Specificities of serum alphafetoprotein in HBsAg+ and HBsAg- patients in the diagnosis of hepatocellular carcinoma. Hepatology 14, 68–72.

Lisboa, P. J. G, Branston, N. M., El-Deredy, W., Vellido, A., 1997. Assessment of statistical and neural networks methods in NMR spectral classification and metabolite selection. In: Proceedings of the IEEE/INNS International Joint Conference on Neural Networks, Houston, pp. 1385–1390.

Lisboa, P.J.G., Kirby, S.P.J., Vellido, A., Lee, Y.Y.B., El-Deredy, W., 1998. Assessment of statistical and neural networks methods in NMR spectral classification and metabolite selection. NMR in Biomedicine 11, 225–234.

Lutterbach, R., Stöckigt, J., 1995. Dynamics of the biosynthesis of methylursubin in plant cells employing in vivo ^{13}C NMR without labeling. Phytochemistry 40, 801–806.

Lutterbach, R., Stöckigt, J., 1994. In vivo investigation of plant-cell metabolism by means of natural-abundance C-13-NMR spectroscopy. Helvetica Chimica Acta 77, 2153–2161.

Mansfield, J.R., Sowa, M.G., Scarth, G.B., Somorjai, R.L., Mantsch, H.H., 1997. Analysis of spectroscopic imaging data by fuzzy C-means clustering. Anal. Chem. 69, 3370–3374.

Matsumoto, I., Kuhara, T., 1996. A new chemical diagnostic method for inborn errors of metabolism by mass spectrometry—rapid, practical, and simultaneous urinary metabolites analysis. Mass Spectrom. Rev. 15, 43–57.

Nicholson, J.K., Sadler, P.J., Bales, J.R., Juul, S.M., MacLeod, A.F., Sonken, P.H., 1984. Monitoring metabolic diseases by proton NMR of urine. Lancet 2, 751–752.

Nicholson, J.K., Wilson, I.D., 1989. High resolution proton magnetic resonance spectroscopy of biological fluid. Prog. NMR Spectr. 21, 449–501.

Nishijima, T., Fujiwara, K., 1997. Measurement of lactate levels in serum and bile using proton nuclear magnetic resonance in patients with hepatobiliary diseases: its utility in detection of malignancies. Jpn. J. Clin. Oncology 27, 13–17.

Ohsaka, A., Yoshikawa, K., Matsuhashi, T., 1979. Detection by proton nuclear magnetic resonance of elevated lactate concentration in serums from patients with malignant tumors. Jpn. J. Med. Sci. Biol. 32, 305–309.

Petroff, O.A.C., 1988. Biological ^{1}H NMR spectroscopy. Comp. Biochem. Physiol. 90B (2), 249–260.

Pope, J.M., Jonas, D., Walker, R.R., 1993. Applications of NMR microimaging to the study of grape berries. Protoplasma 173, 177–186.

Prabhu, V., Chatson, K.B., Abrams, G.D., King, J., 1996. ^{13}C Chemical shifts of 20 free amino acids and their use in detection by NMR of free amino acids in intact plants. J. Plant Physiol. 149, 246–250.

Rabenstein, D.L., Millis, K.K., Strauss, E.J., 1988. Proton NMR spectroscopy of human blood plasma and red cells. Anal. Chem. 60, 1380A–1391A.

Ratcliffe, R.G., Shachar-Hill, Y., 2001. Probing plant metabolism with NMR. Ann. Rev. Plant Physiol. Plant Mol. Biol. 52, 499–526.

Sauter, H., Lauer, M., Fritsch, H., 1991. Metabolic profiling of plants—a new diagnostic technique. In: Baker, D.R., Fenyes, J.G., Moberg, W.K. (Eds.), Synthesis and Chemistry of Agrochemicals II. ACS Symposium Series 443. American Chemical Society, Washington, DC, pp. 288–299.

Schmidt, R. R., 1997. HRAC classification of herbicides according to mode-of-action. In: Brighton Crop Protection Conference, Weeds, pp. 1133–1140.

Schneider, B., 1997. In vivo NMR spectroscopy of low-molecular compounds in plant cells. Planta 203, 1–8.

Shaw, R.A., Kotowich, S., Eysel, H.H., Jackson, M., Thomson, G.T.D., Mantsch, H.H., 1995. Arthritis diagnosis based upon the near-infrared spectrum of synovial fluid. Rheumatol. Int. 15, 159–165.

Somorjai, R.L., Dolenko, B., Nikulin, A.K., Pizzi, N., Scarth, G., Zhilkin, P., Halliday, W., Fewer, D., Hill, N., Ross, I., West, M., Smith, I.C.P., Donnelly, S.M., Kuesel, A.C., Brière, K.M., 1996. Classification of 1H NMR spectra of human brain neoplasms: the influence of preprocessing and computerized consensus diagnosis on classification accuracy. J. Magn. Reson. Imaging 6, 437–444.

Weckwerth, W., Fiehn, O., 2002. Can we discover novel pathways using metabolomic analysis? Curr. Opin. Biotechnol. 13, 156–160.

Wolfender, J.L., Hostettmann, K., 1996. LC-UV-MS: a powerful approach for the rapid screening of metabolites in crude plant extracts. In: Newton, R.P., Walton, T.J. (Eds.), Applications of Modern Mass Spectroscopy in Plant Science Research. Oxford Press, Oxford, pp. 216–221.