Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis
35 (2004) 117–126
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Prediction of anti-plasmodial activity of *Artemisia annua* extracts: application of ^{1}H NMR spectroscopy and chemometrics

Nigel J.C. Bailey [a,*], Yulan Wang [b], Julia Sampson [c,1], Wendy Davis [d], Ian Whitcombe [c,2], Peter J. Hylands [c], Simon L. Croft [d], Elaine Holmes [b]

[a] *SCYNEXIS Europe Ltd., Fyfield Business and Research Park, Fyfield Road, Ongar, Essex CM5 0GS, UK*
[b] *Biological Chemistry, Biomedical Sciences Division, Imperial College London, Sir Alexander Fleming Building, Exhibition Road, South Kensington, London SW7 2AZ, UK*
[c] *Oxford Natural Products Plc., Cornbury Park, Charlbury, Oxfordshire OX7 3EH, UK*
[d] *Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*

Received 1 October 2003; received in revised form 24 December 2003; accepted 28 December 2003

Abstract

We describe the application of ^{1}H NMR spectroscopy and chemometrics to the analysis of extracts of *Artemisia annua*. This approach allowed the discrimination of samples from different sources, and to classify them according to anti-plasmodial activity without prior knowledge of this activity. The use of partial least squares analysis allowed the prediction of actual values for anti-plasmodial activities for independent samples not used in producing the models. The models were constructed using approximately 70% of the samples, with 30% used as a validation set for which predictions were made. Models generally explained >90% of the variance, R^2 in the model, and had a predictive ability, Q^2 of >0.8. This approach was also able to correlate ^{1}H NMR spectra with cytotoxicity ($R^2 = 0.9$, $Q^2 = 0.8$).

This work demonstrates the potential of NMR spectroscopy and chemometrics for the development of predictive models of anti-plasmodial activity.

Keywords: *Artemisia annua*; Malaria; Chemometrics; ^{1}H NMR spectroscopy; Pattern recognition; *Plasmodium falciparum*; Biological activity; Prediction

Abbreviations: $ToxED_{50}$, measure of extract toxicity; FID, free induction decay; IC_{50}, measure of extract anti-plasmodial activity; NMR, nuclear magnetic resonance; PCA, principal components analysis; PLS, partial least squares; PLS-DA, partial least squares discriminant analysis

* Corresponding author. Tel.: +44-1277-367036; fax: +44-1277-367099.
E-mail address: nigel.bailey@scynexis.com (N.J.C. Bailey).
[1] Present address: GW Pharmaceuticals Plc., Porton Down Science Park, Salisbury, Wiltshire SP4 0JQ, United Kingdom.
[2] Present address: Celltech R&D Ltd., 216 Bath Road, Slough, Berkshire SL1 4EN, United Kingdom.

0731-7085/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2003.12.024

1. Introduction

Artemisia annua L. (also known as sweet wormwood, or by its Chinese name *qing hao*) and the isolated constituent artemisinin (Fig. 1) have well documented anti-plasmodial activity [1–3]. Indeed, a whole class of anti-plasmodial drugs derived from artemisinin are now in widespread use in single and combination therapies, particularly where resistance to other anti-plasmodials is present [1,4].

Whilst artemisinin has been established as an important component of *A. annua* with respect to anti-plasmodial activity [2], it has also been suggested that the efficacy of the *A. annua* plant itself derives from a synergistic effect, and that it is a combination of constituents in the plant that confer the total anti-plasmodial activity [5,6]. Several polymethoxyflavones have been found to have activity in combination with artemisinin [7], and it has been reported that flavonoids enhance the anti-plasmodial activity of artemisinin [8]. Further, the clinical efficacy of *A. annua* extracts as a treatment for malaria has been demonstrated, with 92% of malaria patients in a study showing disappearance of parasitaemia within 4 days [9]. As a result, the potential use of *A. annua*, particularly in areas where large-scale pharmaceutical production is not possible, is clearly of interest. However, where plant extracts themselves are to be used, it is necessary to determine the reproducibility and information regarding the content of such extracts [9]. It has been shown, for example that the levels of artemisinin and related compounds fluctuate due to both seasonal and geographical variation [10,11].

Therefore, it is important to develop analytical methodology that is capable of providing information relating to a whole extract with respect to anti-plasmodial activity, and reproducibility. ^{1}H NMR spectroscopy is one such technique that is capable of supplying a 'metabolic fingerprint' and thus can give a measure of the overall biochemical composition of a sample. By comparing spectra from different samples, it is possible to monitor differences in the levels of thousands of metabolites simultaneously and so observe the dynamic biochemical profiles. This principle (variously termed 'metabonomics', 'metabolomics' or 'metabolic fingerprinting' and 'profiling') has been applied to many different areas such as drug toxicity [12], clinical chemistry [13], environmental metabolism [14,15], plant metabolism [16,17] and recently, natural products profiling [18]. By applying chemometric data analysis techniques to NMR spectroscopic data (or indeed, any other multivariate analytical data), mathematical models for predicting structure–activity or structure–metabolism relationships can be derived based on the data obtained [19–22]. It should therefore be possible to use the ^{1}H NMR spectra obtained from *A. annua* extracts along with measured anti-plasmodial activity levels to derive quantitative extract data-activity relationships (QEDARs) that can predict the potential activity of samples not used in the model production process. This is due to the fact that the chemical environment within which each ^{1}H nucleus in a constituent is located determines the ^{1}H NMR spectroscopic resonances for a particular chemical constituent. It is also the chemical environment and thus physico-chemical properties that will determine whether anti-plasmodial activity is present in a particular sample or not. Based on the assumption that the anti-plasmodial activity of *A. annua* extracts is synergistic, this ability to obtain a comprehensive metabolite profile through NMR spectroscopy has the potential to provide an indication of potential anti-plasmodial activity of a whole extract. In addition, the fact that NMR spectra give an indication of the whole low molecular weight chemical content of an extract means they also offer an insight into other parameters that may be of interest, such as cytotoxicity. Artemisinin and derivatives have been shown to induce neuropathology in dogs and rats, and in some cases anaemia; however, few significant side effects have been reported in humans [23–25]. By creating separate models to predict both activity and toxicity, the process could potentially indicate an extract that has both the desirable anti-plasmodial efficacy, and low toxicity.

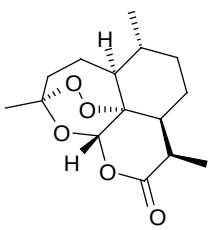

Fig. 1. Structure of artemisinin.

Previous work on single component samples has demonstrated the application of ^{13}C NMR spectra to predict steroids binding to corticosteroid binding globulin [26], and the aromatase enzyme [27], and polychlorinated compounds binding to the aryl hydrocarbon receptor [28] as well as predictive models of estrogen receptor binding activities [29]. This is the first time, however, that the approach has been extended to complex mixtures such as plant extracts.

This paper reports on a study designed to demonstrate a proof of concept with respect to the hypothesis outlined above, and shows that the application of NMR-based chemometrics can be used to discriminate between different accessions of *A. annua* extracts. More importantly, NMR-based chemometrics can be used to predict the potential anti-plasmodial activity of those extracts, in addition to toxicity.

2. Methods

2.1. Acquisition of plant samples

Nineteen *A. annua* accessions sourced from different locations were obtained by Oxford Natural Products Plc. (Oxford, UK). Samples in the form of dry herb material, powder or tincture were deposited in the herbarium in the Pharmacognosy laboratories at King's College London, UK. Voucher specimen numbers were as follows for sample identities 1–19 of *A. annua* L. (Compositae): AR17 10 I1, AR17 11 I2, AR17 12 I3, AR17 13 I4, AR17 14 I5, AR17 15 I6, AR17 16 I7, AR17 17 I8, AR17 19 I10, AR17 20 I11, AR17 21 I12, AR17 22 I13, AR17 23 I14, AR17 24 I15, AR17 25 I16, AR17 26 I17, AR17 27 I18, AR17 28 I19, AR17 29 I20.

2.2. Sample preparation

Samples were prepared by Advanced Phytonics (UK) using an extraction method covered under International patent Application Numbers PCT/GB95/00554, and International Publication Number WO 95/26794.

For dry herb material, the samples (90–135 g) were packed into a 560 ml jacketed vessel and extracted at 25 °C with a constant pumped stream of 1,1,1,2-tetrafluoroethane solvent (26–64 bed volumes), at 8–10 bar over 2–5 h, until no more extract was obtained. The solvent was removed from the extract before collection. For *A. annua* containing powder samples, material (55–79 g), was packed into a 212 ml vessel and extracted at 25 °C with a constant pumped stream of 1,1,1,2-tetrafluoroethane (21–101 bed volumes), at 8–10 bar over 2–3 h until no more extract was obtained. The solvent was again removed from the extract before collection. For the *A. annua* containing tincture, material (100 ml), was added to a glass pressure vessel and extracted into 1,1,1,2-tetrafluoroethane (100 ml), with agitation. Once settled, the solvent/ethanolic solution was decanted, removed and the extract collected. The remaining aqueous portion was extracted with 4× 100 ml aliquots of solvent and the extracts combined.

Replicates (approximately 10 mg, $n = 3$ for 18 extracts, $n = 2$ for one extract due to limited material) of freeze-dried extracts were weighed out and added to 1 ml d_6 DMSO, giving a total of 56 samples for analysis. Samples were agitated and then centrifuged at 13,000 rpm (ca. $14{,}000 \times g$) for 15 min. Supernatant (850 μl) was removed and re-centrifuged for a further 10 min. Following the second centrifugation, supernatant (700 μl) was taken for NMR analysis.

2.3. Primary cell based in vitro screen for anti-plasmodial activity

2.3.1. P. falciparum strain

The drug-sensitive 3D7 clone of NF54, was used in the primary screen.

The method used for anti-plasmodial testing was based on previously published methods [30–32]. Briefly, stock solutions of *A. annua* extracts are prepared in 100% DMSO at 10 mg/ml. A five-fold dilution series (2×) was prepared in triplicate in 96-well plates (50 μl per well) from a concentration of 25 μg/ml down to 0.2 μg/ml followed by the addition of 50 μl of synchronous ring stage parasite cultures at 5% hematocrit and 1% parasitemia determined by microscopical examination of Giemsa stained thin blood smears. Final hematocrit and parasitaemia were 2.5 and 1%, respectively. The assay was performed in hypoxanthine-free medium.

Plates were incubated for 24 h at 37 °C, 5% CO_2 followed by addition of 20 μl of ^3H-hypoxanthine to

all wells (0.1 μCi per well) [33]. After shaking for a minimum of one minute, plates were returned to the incubator for a further 24 h. Plates were freeze/thawed rapidly, harvested onto a 96-well glass fibre filter mat and dried at 42 °C. Incorporation of radioactive hypoxanthine was measured using a Wallac 1450 BetaLux scintillation counter.

The standard drug chloroquine diphosphate was included in all assays in a three fold dilution series from a top concentration of 30 μg/ml down to 0.0001 μg/ml. The control wells were infected erythrocytes in the absence of drug and blank wells were uninfected A* erythrocytes at a final hematocrit concentration of 2.5%. Results were analysed using the Microsoft Excel based program MsX/fit (IDBS, UK), to calculate IC_{50} values.

2.3.2. Cytotoxicity assay

An amount of 96-well plates were seeded with KB cells (human or-pharyngeal carcinoma) at 2×10^4/ml (200 μl per well) in 10% FCS-RPMI 1640 medium and incubated at 37 °C in a 5% CO_2/air mixture. After 24 h extracts and compounds were added at 300, 30, 3, 1 and 0.3 μg/l in fresh overlay in triplicate at each concentration. Plates were incubated for a further 72 h, washed 3× with PBS before the addition of alamar blue in PBS [34]. After 2 h, plates were read on a Gemini plate reader at EX/EM 530/580 nm.

2.4. ^1H NMR spectroscopy of plant extracts

NMR spectra were acquired on a Bruker DRX 600 NMR Spectrometer (Bruker GmbH, Rheinstetten, Germany) operating at 600.22 MHz for the ^1H frequency and fitted with a broadband inverse geometry probe. All spectra were the result of the summation of 128 free induction decays (FIDs), with data collected into 32 k datapoints, a spectral width of δ 14 and an acquisition time of 1.95 s.

The 90° pulse length was measured for the samples prior to data acquisition. Prior to Fourier transformation, an exponential line broadening equivalent to 0.3 Hz was applied to the FIDs and spectra were referenced to DMSO at δ 2.50. Spectra for the 56 samples were acquired in a random order, and four samples were repeated to check for analytical reproducibility (total of 60 NMR spectra acquired).

2.5. Multivariate data analysis

One-dimensional NMR spectra were reduced to 252 discrete chemical shift regions by digitisation to produce a series of sequentially integrated regions δ 0.04 in width between δ −0.06 and 9.98, using Bruker AMIX software (version 2.0, Bruker GmbH, Germany). The resulting data matrix was exported into Microsoft® Excel 2000 and selected regions removed around the water signal (δ 3.46–3.18), DMSO (δ 2.54–2.46) and also the region (δ 0.42–−0.06). The remaining integral regions were normalised to the whole spectrum to remove any variation in concentration prior to Principal Components Analysis (PCA) [35].

PCA was performed using SIMCA-P 9.0 multivariate data analysis software (Umetrics, Sweden), on mean centred data. PCA is a data reduction technique that represents multivariate data in a reduced set of dimensions, usually fewer than 4, such that an overview of the data is permitted. The output from the PCA analysis consisted of scores plots (giving an indication of the differentiation of classes in terms of biochemical similarity), and loadings plots, which give an indication as to which ^1H NMR spectral regions were important with respect to the classification observed in the scores plots.

Partial least squares discriminant analysis (PLS-DA) was performed using SIMCA-P 8.0. PLS-DA (and PLS, below) is a method that may be considered an extension of PCA and serves to maximise the separation between two or more sample classes based on prior knowledge of class membership, and uses a discrete class identifier in the *y* matrix rather than a continuous measure of response. The dataset was split randomly into two groups of training (70% of samples) and test (30%) samples, with a dummy *y*-matrix set up to provide information on class. The process of generating a PLS-DA model on randomly chosen samples followed by validation was repeated in order to ensure all extracts were excluded at least once.

PLS was performed using SIMCA-P 8.0 multivariate data analysis software (Umetrics, Sweden), with mean centering of the data preceding PLS. The dataset was split randomly into two groups of training (approximately 70% of samples) and test (approximately 30%) samples. The process of generating a PLS model on randomly chosen samples followed by validation

was repeated in order to ensure all extracts were excluded once. Models were constructed to predict values for IC_{50} and $ToxED_{50}$.

3. Results and discussion

The biochemical content of *A. annua*, along with many other natural products shows variation depending on geography or season. It is thus important than any technique used for the analysis of such samples is able to first and foremost discern the differences between samples of differing origin. This then enables some form of quality control to be performed to ensure that quality and content of the extracts are maintained. Representative ^{1}H NMR spectra are shown in Fig. 2. These spectra (representing extracts of differing IC_{50} value) show that whilst the artemisinin resonances themselves are readily apparent, particularly in the more potent extract 7, there are many other regions of the spectrum where large differences occur between the different extracts. In particular, the region between 6.8 and 8.0 ppm has marked differences in the spectra. The PCA plot obtained from analysis of the ^{1}H NMR spectra of all the extracts is shown in Fig. 3. Each point on the plot represents one ^{1}H NMR spectrum of an extract, with points of the same number indicating replicate samples of the same origin. It can be seen from this figure that each of the samples within a particular group of replicate samples readily cluster together indicating the reproducibility of the extraction process and the analytical methodology. Groups of samples of differing origin can be discriminated from one another based on inherent differences in their biochemical profiles. Using this approach therefore, it is possible to monitor the overall make-up of a sample, and compare the whole extract with that of previous samples, samples of different origins or seasonal differences, and use this information to implement quality assurance of natural products.

The PCA algorithm is an unsupervised method, and thus uses no information regarding the class of each sample, with the colour coding solely for the aid of human visualisation. It is therefore now possible to re-label the datapoints to reflect the biological activities of each of the samples. A coding was performed whereby the data were split into three classes based on their IC_{50} value, with the cut-off values being 0.1 and

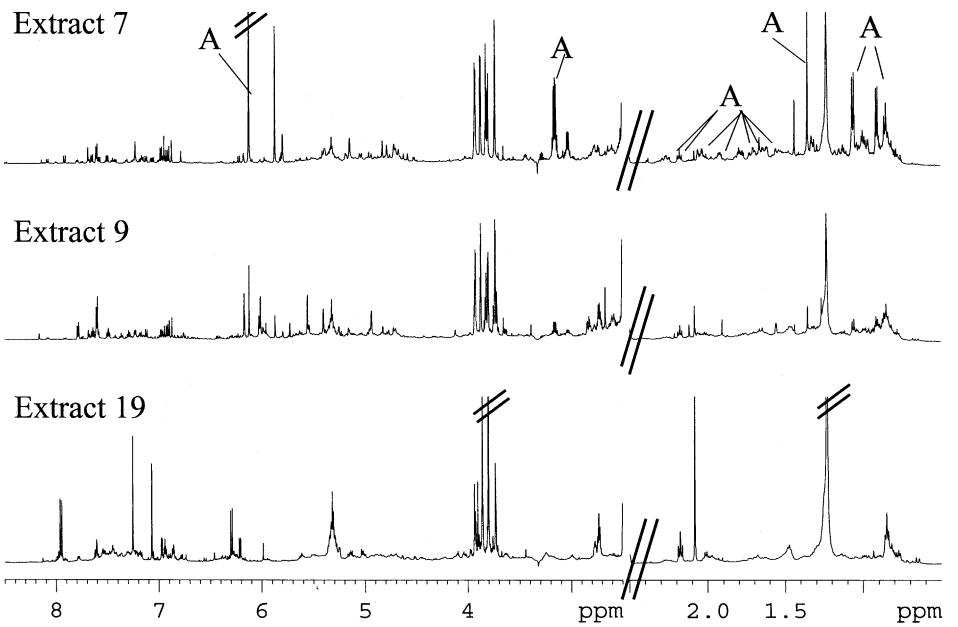

Fig. 2. Representative ^{1}H NMR spectra for the three IC_{50} classes, $IC_{50} < 0.1\,\mu g/ml$ (extract 7), $IC_{50} > 0.1\,\mu g/ml$, $<1\,\mu g/ml$ (extract 9) and $IC_{50} > 1\,\mu g/ml$ (extract 19). Region 2.5–8.5 ppm is expanded vertically by a factor of 6 to allow observation of lower level aromatic resonances. Resonances attributable to artemisinin are indicated with an 'A'.

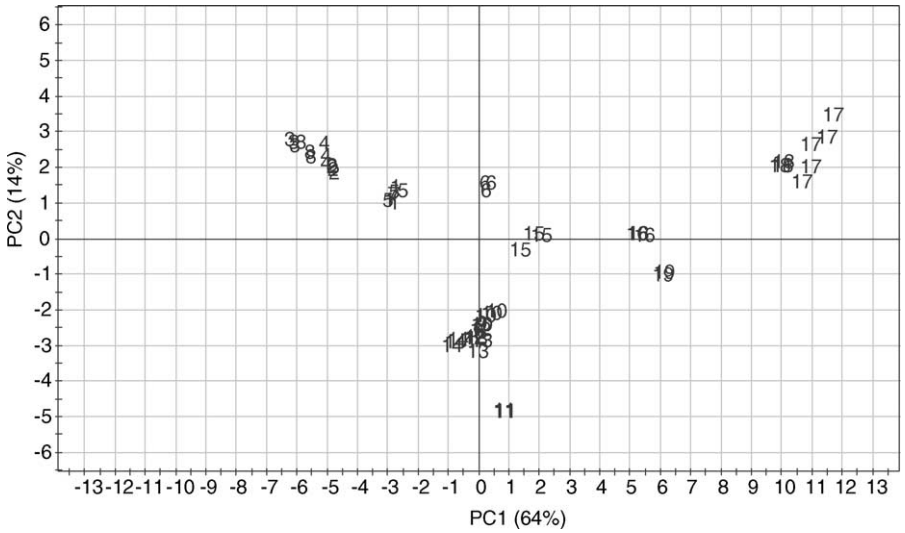

Fig. 3. PCA scores plot for *A. annua* plant extracts. Samples are separated according to original extract number.

1 μg/ml. The resulting PCA scores plot can be seen in Fig. 4. It is apparent that this model (containing 78% variance in the first two PCs) is able to discriminate between the three classes used. This suggests that ^{1}H NMR spectra contain sufficient information relating to the physico-chemical properties of the extract to be able to predict the potential magnitude of anti-plasmodial activity found in a plant extract. By interrogating the PCA loadings plot, it was possible to determine the variables (spectral regions) that are responsible for this separation, and are thus the regions that have greatest variation between the extracts. The loadings plot for this model (not shown) indicates that the spectral regions containing artemisinin are chiefly responsible for this separation. In order to ascertain the strength of influence of artemisinin on the model, regions containing artemisinin resonances were removed from the dataset and the analysis repeated. A very similar separation is achieved, and the corresponding loadings plot contains variables largely associated with the resonances around δ 3.7–4.0 (data not shown). The identity of the molecules for which

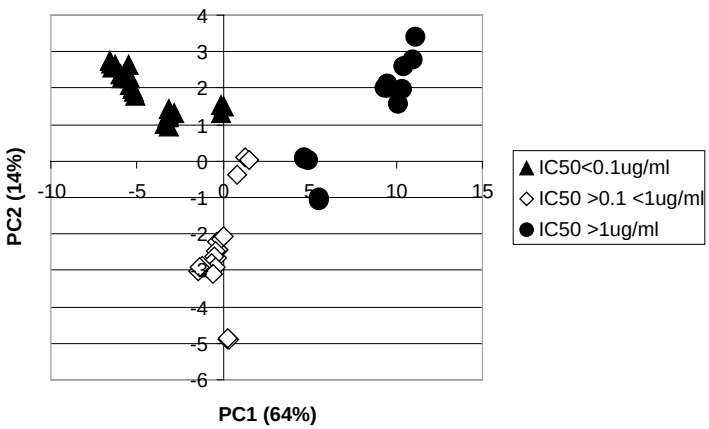

Fig. 4. PCA scores plot for *A. annua* plant extracts. Samples are separated into three groups, IC_{50} < 0.1 μg/ml (▲), IC_{50} > 0.1 and <1 μg/ml (◇) and IC_{50} > 1 μg/ml (●).

these resonances are attributable was not determined. While these additional resonances, and the ability to achieve separation using them is of interest, and further demonstrates the presence of synergy at work, the chief aims of the work is to consider the extracts as a whole, and so all further work considers the entire spectral region.

Whilst PCA clearly demonstrates the potential of this technique, a more robust approach to obtaining predictive data is to employ supervised methods. These methods involve providing the model with the values for the variable to be predicted (i.e., IC_{50} value) for part of the dataset (the training set), with the model then being optimised based on those values. Because the algorithm in effect uses the answers to create the model, it is then necessary to validate this model using the remaining unused samples (the test set). Those samples with an IC_{50} value $> 1\,\mu g/ml$ were excluded from this analysis, for two reasons. This class is the smallest of the three, and having larger values means that the model is likely to be skewed in order to take them into account. In addition, the higher values mean that these samples are not of interest anyway, as they essentially have no activity. Using the remaining two classes as above ($IC_{50} < 0.1\,\mu g/ml$ and $IC_{50} > 0.1\,\mu g/ml$, respectively) it is possible to construct a 'dummy' *y*-matrix whereby the two classes are represented by a 1 or a 0. Partial least squares discriminant analysis (PLS-DA) can then be performed on the data to construct a new model using this additional data. By excluding data at random from this model, predictions as to the likely class membership of the excluded data points can be made in order to validate the model. As the two classes are designated either a 1 ($IC_{50} < 0.1\,\mu g/ml$) or a 0 ($IC_{50} > 0.1\,\mu g/ml$), any data point with a predicted value of greater than 0.5 is considered to be a member of class 1, and any point with a value of less than 0.5 is considered to belong to class 0. The model shown in Fig. 5 was constructed using 36 of the samples (the training set), whilst 10 were used to validate the model (test set). It can be seen from the figure that all the data points from both training and test sets are correctly predicted in what is overall a robust model with $R^2 = 0.90$, $Q^2 = 0.89$, where R^2 is the variance, and Q^2 is the cross-validated variance, or predictive ability of the model. (As a guide, a value of $Q^2 > 0.5$ is generally considered to be good [35]). The analysis was repeated five times (with samples split into different training and test sets each time), with all samples correctly classified in all models, with >91% of samples predicted with >99% confidence. R^2 values ranged from 0.90 to 0.92, and Q^2 values from 0.88 to 0.90. Two components were used in all models. While it may appear that only the same amount of information is available from both the PCA and PLS-DA models, the fact that more robust validation is available for the supervised PLS-DA method means that the results have greater credibility. Thus using this PLS-DA model, it would be possible to predict those samples that are likely to have anti-plasmodial activities of $<0.1\,\mu g/ml$.

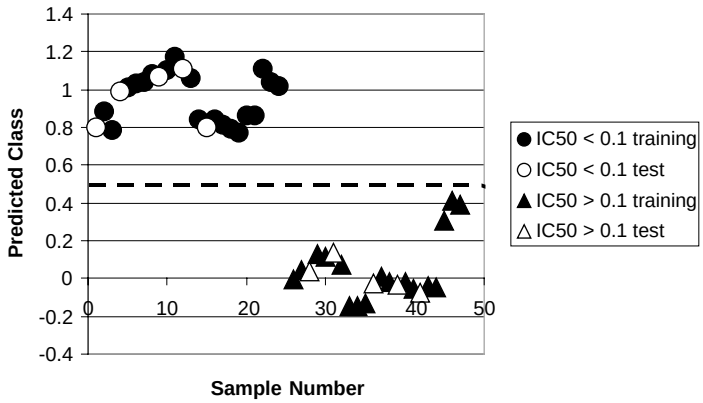

Fig. 5. PLS-DA Y-predicted scatter plot for classes $IC_{50} > 0.1\,\mu g/ml$ (class 0, triangles) and $IC_{50} < 0.1\,\mu g/ml$ (class 1, circles). Training set represented by closed shapes, (▲) and (●), test set represented by open shapes, (△) and (○).

Table 1
Summary of predicted IC_{50} and $ToxED_{50}$ values for a series of *A. annua* extracts

Extract	IC_{50} value (μg/ml)	IC_{50} predicted	$ToxED_{50}$ value (μg/ml)	$ToxED_{50}$ predicted	Relative artemisinin level[a]
1	0.01	0.05 ± 0.04	27	42 ± 6	0.47
2	0.01	0.01 ± 0.02	9	20 ± 3	0.66
3	0.020	0.027 ± 0.00	8	22 ± 6	1
4	0.02	0.00 ± 0.03	19	17 ± 6	0.69
5	0.020	0.065 ± 0.005	65	26 ± 11	0.44
6	0.02	0.06 ± 0.02	50	33 ± 2	0.34
7	0.0296	0.0262 ± 0.0003	171	–	0.45
8	0.04	0.03 ± 0.02	11	12 ± 2	0.91
9	0.169	0.209 ± 0.008	69	54 ± 4	0.18
10	0.13	0.17 ± 0.01	26	28 ± 5	0.14
11	0.29	0.16 ± 0.04	41	75 ± 4	0.15
12	0.30	0.20 ± 0.03	66	62 ± 2	0.21
13	0.31	0.22 ± 0.04	8	33 ± 6	0.18
14	0.32	0.22 ± 0.02	76	72 ± 3	0.20
15	0.47	0.23 ± 0.04	50	60 ± 5	0.12
16	8.55	0.21 ± 0.03	46	30 ± 1	0.04
17	24.67	0.20 ± 0.03	72	40 ± 12	0.02
18	4.2	0.10 ± 0.01	20	48 ± 14	0.02
19	3.9	0.243 ± 0.002	–	–	0.04

[a] Obtained from the ^{1}H NMR peak intensity for the artemisinin peak at ca. 6.1 ppm. Values expressed relative to the highest peak, that of extract 3.

The PCA and PLS-DA models discussed above give a good indication as to the likely magnitude of anti-plasmodial activity. This approach however is based on an artificially imposed classification i.e., $IC_{50} < 0.1$ μg/ml or >0.1 μg/ml, which whilst giving a clear indication of potential anti-plasmodial activity, uses classes that may or may not be significant. This can be taken one step further however, with the prediction of the actual IC_{50} value for each of the extracts. Instead of the dummy *y*-matrix constructed for the PLS-DA analysis, it is possible to construct a model using the IC_{50} values obtained for each of the extracts from a biological assay. The result can then be used to predict values for test data. Three components were used for all models, with >87% of samples predicted with >99% confidence. R^2 values ranged from 0.83 to 0.93, and Q^2 values from 0.62 to 0.91. The model construction process was repeated in order to exclude every extract from the training set once (with additional samples being removed on a random basis). The overall predictions for each extract are summarised in Table 1. It can be seen that in general the predicted value is reasonably close to the actual value (the correlation between average predicted versus actual values is 0.90), and this clearly demonstrates the potential of such an approach. If the predicted IC_{50} values were used to classify the extracts into two classes, IC_{50} greater or less than 0.1 μg/ml, then all extracts would be placed in the same class as if the actual data were used. This illustrates the fact that this method of analysis can be used as a first step filtering technique in order to identify key extracts worth pursuing using other approaches. Whilst the relatively inactive class 3 were excluded from the model building process, looking at how the model deals with the excluded class is of interest. In this case, the predictions are inaccurate in terms of absolute numbers, but put the samples at the top end of values from class 2, i.e., it correctly indicates that these samples would be least active.

Whilst the IC_{50} value assigned to each extract gives an indication as to the potential anti-plasmodial activity, of equal importance with respect to development of any extract as a viable pharmaceutical, is the likely toxicity of any such extract. In addition to the IC_{50} measurements therefore, each extract was analysed using an in vitro mammalian KB cell line as a measure of cytotoxicity ($ToxED_{50}$) of the extract. By constructing

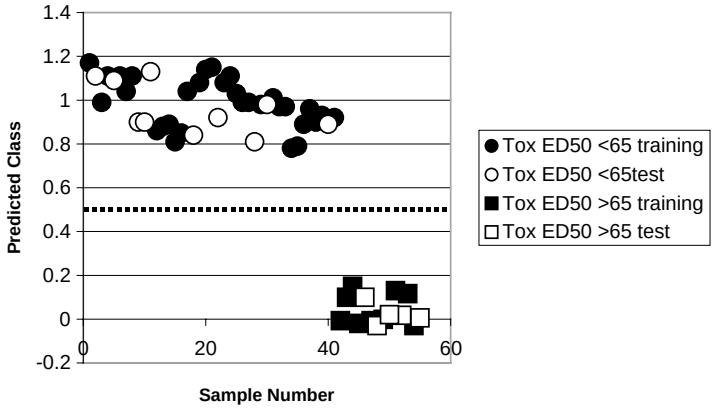

Fig. 6. PLS-DA Y-predicted scatter plot for classes $ToxED_{50}$ > 65 μg/ml (class 0, squares) and $ToxED_{50}$ < 65 μg/ml (class 1, circles). Training set represented by closed shapes, (■) and (●), test set represented by open shapes, (□) and (○).

new models using the $ToxED_{50}$ rather than IC_{50} values, it is possible to predict values for $ToxED_{50}$ using the same NMR spectra. An example PLS-DA plot is shown in Fig. 6, with the samples separated into those with a $ToxED_{50}$ value of > or <65 μg/ml (one extract, extract 7, with a $ToxED_{50}$ value of 171.3 μg/ml was excluded for the same reasons as the extracts excluded with IC_{50} of >1 μg/ml. One further extract, number 19, was excluded due to insufficient material being available for the $ToxED_{50}$ measurement). The analysis was repeated five times (with samples split into different training and test sets each time), and >90% of samples were correctly classified in all models, with >84% of samples predicted with >99% confidence. R^2 values ranged from 0.78 to 0.94, and Q^2 values from 0.69 to 0.84. Three, four or five components were used in all models. Although the statistics reveal that the models are not as good as the ones created for the IC_{50} values, these models are still able to give a good indication as to the likely toxicity of a particular plant extract. That said however, the classes used are, as for the IC_{50} data analysis, arbitrary classes. In order to obtain predicted $ToxED_{50}$ values for the extracts, PLS was again performed using the assay data to construct models.

Four, five, or six components were used for all PLS models, with >78% of samples predicted with >99% confidence. R^2 values ranged from 0.89 to 0.98, and Q^2 values from 0.73 to 0.94. The model construction process was repeated in order to exclude every extract from the training set once (with additional samples being removed on a random basis). The average predicted $ToxED_{50}$ values for each extract (using only the predictions when a particular sample was in the test set) are summarised in Table 1. It can be seen that as with the IC_{50} predictions, the predicted values are, overall, in close agreement with the actual values, although the average predicted values versus actual values correlation is only 0.60, compared with the 0.90 value obtained from the IC_{50} models. The important point however, is that by using the same NMR spectra in modelling both IC_{50} and $ToxED_{50}$ values, it has been possible to predict reasonable values for two different parameters which would normally require two separate assays to be run to obtain the same information. Using this information, it may be that criteria could be set-up to identify the most promising extracts for further study. Dividing the IC_{50} value by the $ToxED_{50}$ value for each extract for example, would result in a value for each extract whereby the smaller the number, the more interesting the extract in terms of small IC_{50} and large $ToxED_{50}$ values. If this calculation is carried out using the actual values and the predicted values for IC_{50} and $ToxED_{50}$, then with the exception of extract 18 (from the IC_{50} class 3), there is very good agreement between the order of samples from the two sets of values. This work demonstrates the potential of ^{1}H NMR spectroscopy and chemometric analysis as a tool for the prediction of anti-plasmodial activity of plant extracts. ^{1}H NMR spectra provide a biochemical fingerprint for each of the samples analysed, from which anti-plasmodial activities can be predicted using different chemometric

techniques. Chemometric analysis can take place on several levels, such as the unsupervised method of PCA, giving an indication of the likely magnitude of anti-plasmodial activity, or the more involved PLS, which attempts to put a numerical figure on such activity. While this study clearly demonstrates the potential of the technique, more samples would increase the robustness and predictability of the models produced. It is of particular importance to ensure the whole range of potential activities is covered within the training set in order to predict test-set samples as accurately as possible. The values obtained using the available samples were such that the conclusions reached using predicted values would be the same as those reached using the actual values. The work also demonstrated the fact that a single NMR spectrum for each extract could be used in more than one model, effectively removing the need for not one, but multiple assays.

References

[1] V. Dhingra, K.V. Rao, M.L. Narasu, Life Sci. 66 (2000) 279–300.
[2] D. Klayman, Science 228 (1985) 1049–1055.
[3] R. Haynes, Curr. Opin. Infect. Dis. 14 (2001) 719–726.
[4] G. Balint, Pharmacol. Therap. 90 (2001) 261–265.
[5] G. Kirby, Trop. Doctor 27 (1997) 7–11.
[6] B. Gilbert, L.F. Alves, Curr. Med. Chem. 10 (2003) 13–20.
[7] R. Bhakuni, D. Jain, R. Sharma, S. Kumar, Curr. Sci. 80 (2001) 35–48.
[8] B. Elford, M. Roberts, J. Phillipson, R. Wilson, T. Roy, Soc. Trop. Med. H. 81 (1987) 434–436.
[9] M. Mueller, I. Karhagomba, H. Hirt, E. Wemakor, J. Ethnopharmacol. 73 (2000) 487–493.
[10] T. Wallaart, N. Pras, W. Quax, Planta Med. 65 (1999) 723–728.
[11] T. Wallaart, N. Pras, A. Beekman, W. Quax, Planta Med. 66 (2000) 57–62.
[12] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Nat. Rev. Drug Disc. 1 (2002) 153–160.
[13] F. Bamforth, V. Dorian, H. Vallance, D. Wishart, J. Inher. Metab. Dis. 22 (1999) 297–301.
[14] J. Griffin, L. Walker, S. Garrod, E. Holmes, R. Shore, J. Nicholson, Comp. Biochem. Phys. B 127 (2000) 357–367.
[15] J.G. Bundy, D. Osborn, J.M. Weeks, J.C. Lindon, J. Nicholson, FEBS Lett. 500 (2001) 31–35.
[16] N. Aranibar, B. Singh, G. Stockton, K. Ott, Biochem. Biophys. Res. Com. 286 (2001) 150–155.
[17] N.J. Bailey, M. Oven, E. Holmes, J.K. Nicholson, M. Zenk, Phytochemistry 62 (2003) 851–858.
[18] N.J. Bailey, J. Sampson, P. Hylands, J.K. Nicholson, E. Holmes, Planta Med. 68 (2002) 734–738.
[19] H. Kessler, Angewandte Chemie Int. Ed. Eng. 36 (1997) 829–831.
[20] F. Ghauri, C.A. Blackledge, R. Glen, B. Sweatman, J.C. Lindon, C. Beddell, I.D. Wilson, et al., Biochem. Pharm. 44 (1992) 1935–1946.
[21] D. Hammond, I. Kubo, Bioorg. Med. Chem. 7 (1999) 271–278.
[22] N.J.C. Bailey, PhD Thesis, Imperial College of Science, Technology and Medicine, 2000.
[23] W. Classen, B. Altmann, P. Gretener, C. Souppart, P. Skelton-Stroud, G. Krinke, Exp. Tox. Path. 51 (1999) 507–516.
[24] A.R. Bilia, D. Lazari, L. Messori, V. Taglioli, C. Temperini, F.F. Vincieri, Life Sci. 70 (2002) 769–778.
[25] J.T. Kim, J.Y. Park, H.S. Seo, H.G. Oh, J.W. Noh, J.H. Kim, D.Y. Kim, H.J. Youn, Vet. Parasit. 103 (2002) 53–63.
[26] R.D. Beger, J.G. Wilkes, J. Comp. Aid. Mol. Des. 15 (2001) 659–669.
[27] R.D. Beger, D.A. Buzatu, J.G. Wilkes, J.O. Lay, J. Chem. Inf. Comput. Sci. 41 (2001) 1360–1366.
[28] R.D. Beger, J.G. Wilkes, J. Chem. Inf. Comput. Sci. 41 (2001) 1322–1329.
[29] R.D. Beger, J.P. Freeman, J.O. Lay, J.G. Wilkes, D.W. Miller, J. Chem. Inf. Comput. Sci. 41 (2001) 219–224.
[30] T. Ponnudurai, A.D. Leeuwenberg, J.H. Meuwissen, Trop. Geo. Med. 33 (1981) 50–54.
[31] M.J. O'Neill, D.H. Bray, P. Boardman, J.D. Phillipson, D.C. Warhurst, Antimicrob. Agents Chemother. 6 (1979) 710–718.
[32] M.J. O'Neill, D.H. Bray, P. Boardman, J.D. Phillipson, D.C. Warhurst, Planta Med. 5 (1985) 394–398.
[33] R.E. Desjardins, C.J. Canfield, J.D. Haynes, J.D. Chulay, Antimicrob. Agents Chemother. 16 (1979) 710–718.
[34] M.M. Nociari, A. Shalev, P. Benias, C. Russo, J. Immunol. Methods 213 (1998) 157–167.
[35] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Introduction to Multi- and Megavariate Data Analysis Using Projection Methods (PCA and PLS). 1999, Umetrics AB, Umeå, Sweden.