Extraction of Alkaloids for NMR-Based Profiling: Exploratory Analysis of an Archaic Cinchona Bark Collection

Ali Yilmaz1, Nils T. Nyberg1, Jerzy W. Jaroszewski1,2

Key words
- Cinchona
- Rubiaceae
- alkaloids
- NMR
- profiling
- PCA
- STOCSY
- STOCSY-CA

Abstract

A museum collection of Cinchonae cortex samples (n = 117), from the period 1850–1950, was extracted with a mixture of chloroform-d4, methanol-d4, water-d2, and perchloric acid in the ratios 5:5:1:1. The extracts were directly analyzed using 1H NMR spectroscopy (600 MHz) and the spectra evaluated using principal component analysis (PCA) and total statistical correlation spectroscopy (STOCSY). A new method called STOCSY-CA, where CA stands for component analysis, is described, and an analysis using this method is presented. It was found that the samples had a rather homogenous content of the well-known cinchona alkaloids quinine, cinchonine, and cinchonidine without any apparent clustering. Signals from analogues were detected but not in substantial amounts. The main variation was related to the absolute amounts of extracted alkaloids, which was attributed to the evolution of the Cinchona tree cultivation during the period in which the samples were collected.

Introduction

The crude drug cinchona (Cinchonae cortex) used against malaria is known since the colonization of South America in the beginning of the 17th century. The antimalarial effect is mainly due to the content of quinine (1), which interferes with the development of the Plasmodium parasites in the red blood cells [1]. Cinchona also contains the likewise active analogue alkaloids quinidine (2), cinchonine (3), and cinchonidine (4) [2–3]. The most abundant alkaloid is quinine that accounts for 70–90% of the total alkaloid content which can be as high as 25% of the dry weight [4]. Even though there are effective artemisinin-based combination therapies against malaria, there is still an interest in using the bark of Cinchona species (Rubiaceae) for reasons of availability and evolved malaria drug resistance [5,6].

As a result of the economical values of cinchona during the 19th century, plantations were spread to different tropical regions such as Indonesia and India. Different species and hybrids were tested for production of the drug in order to maximize the yield of quinine. During this period, samples of Cinchonae cortex were collected from different parts of the world and eventually ended up in a collection hosted by the Museum of Natural Medicine at the University of Copenhagen. A profiling study of this archaic collection, focused on alkaloids, could possibly indicate species that were not selected due to the quinine content but to other types of antimalarial alkaloids. Since the samples are collected over a long time period, from many different sites, and for many different reasons, we expect a large variation of types and amounts of alkaloids. Profiling of extracted alkaloids might indicate odd samples that contain other types of alkaloids than the expected and so allows finding out about the possible variations and combinations of the known and major alkaloids.

NMR is a rational choice of analytical platform for this study. It is rapid, robust and gives reproducible results [7,8]. NMR also gives different spectra of diastereomers, and this makes it possible to differentiate between the main alkaloids. However, the traditional liquid-liquid extraction to separate alkaloids based on their basic properties is not practical for a profiling study in terms of sample throughput and potential problems with robustness. In this study, we present the use of a direct...
Materials and Methods

Chemicals

Anhydrous quinine hydrochloride, quinidine sulfate, cinchonine, and cinchonidine were from Fluka. Methanol-d$_4$ (99.8% D) and deuterated trifluoroacetic acid (99.9% D) were from Sigma-Aldrich. Perchloric acid (70%, p.a.) was from Riedel-de Haën, tetramethylsilane (TMS, 99.9%) from Lancaster, chloroform-d$_1$ (99.8% D) and water-d$_2$ (99.9% D) from Eurisotop.

Titration experiment

Anhydrous quinidine hydrochloride (20 mg) was dissolved in 10.0 mL of a 1:1 (v/v) mixture of methanol-d$_4$ and chloroform-d$_1$. 1000 µL of the mixture were transferred to 10 vials. Perchloric acid (10–100 µL, in steps of 10 µL) was added to each vial to give the approximate concentrations 0.1–1.0 M with respect to HClO$_4$. The mixtures were transferred to 5 mm tubes, and $^1$H NMR spectra were acquired on a 400 MHz Bruker Avance DRX spectrometer (Bruker Biospin) at 298 K using a pulse length corresponding to a 30° pulse, a spectral width of 8 kHz, collecting 64 scans into 64 k data points. The acquisition was followed by a relaxation delay of 1.0 sec. The time domain data were zero filled to 128 k data points with a relaxation delay of 4.0 sec. The time domain data were zero filled to 128 k data points and multiplied with a 0.3 Hz exponential function before Fourier transform and referencing of the chemical shift axis to the residual methanol signal at 3.31 ppm.

NMR spectroscopy

NMR spectral data were acquired at 300 K on a Bruker Avance NMR spectrometer operating at 600.13 MHz $^1$H observation frequency (Bruker Biospin) equipped with a sample changer and a 5-mm inverse probe-head. NMR spectra of cinchona bark extracts were acquired using a standard 1D-NOESY-based presaturation pulse sequence to suppress the water signal. Free induction decays of 128 transients were collected into 64 k data points using a spectral width of 12 kHz (20 ppm). The total inter-pulse delay (acquisition time and relaxation delay) was 4.66 sec, with an additional mixing time of 100 ms, giving a total experiment time of 10.5 min. Each sample was analyzed twice with a mean time between repeated measurements of 15.0 hours (low and high quartiles at 7.0 and 23.2 hours, respectively). The final data-set consisted of 350 spectra (59 samples prepared once and analyzed twice; 58 samples prepared twice, each analyzed twice). The acquisition was controlled by in-house written scripts running under Topspin and Icon-NMR (Bruker Biospin). The acquisition protocol included a 3-min temperature equilibration period, determination of 90° pulse (low and high quadratures at 9.46 and 9.62 µsec, respectively), gradient shimming, and acquisition. Samples were gradient shinned using z-axis profile of the $^1$H NMR signal of CD$_3$OD in the solvent. The median sample turnover was 19 minutes. Prior to Fourier transformation to 32k real data points, an exponential line broadening function of 0.3 Hz was applied to each FID. Assignments of the signals of the main alkaloids (quinine, quinidine, cinchonine, and cinchonidine) were done by spiking experiments of samples and standard 2D experiments using the solvent system described above.

Data preparation and analysis

The spectra were imported into Matlab (ver. 2010b, Mathworks, Inc.) using in-house written routines and calibrated to internal reference TMS at δ 0.00. Phase corrections of the spectra were automatically applied by finding the zeroth order phase correction.
term that maximized the lowest point of the baseline (defined by signal free ranges, excluding 1/32 of the middle part of the spectrum). The data were further baseline-corrected by fitting a cubic polynomial function through signal free ranges. Spectra were normalized either to total sum (excluding the solvent regions) or to the area of the methanol residual solvent signal in the range 8.34–3.38. They were further interpolated to a common axis in the range 0.5–10.0 ppm with 16 k data points. Integration limits (“buckets”) with an initial width of 0.01 ppm were adjusted to nearby minima in the calculated mean spectrum. Regions corresponding to residual solvent signals (8.34–3.32, 5.96–5.65, and 7.59–7.53 for CD3OD, HOD, and CDCl3, respectively) were excluded. Statistical calculations were performed after Pareto-scaling of the integrated ranges using functions in PLS Toolbox (ver. 5.5, Eigenvector Research, Inc.) and Statistical Toolbox (ver. 7.4, Mathworks, Inc.). The number of principal components was determined by cross-validation by dividing the spectra into 12 groups where spectra of the same sample were kept together.

Supporting information
A table of all samples in the study with the museum registry numbers (Tab. 15), selected 1H NMR spectra of extracted bark samples (Fig. 1S), results of STOCSY analysis (both full spectral ranges) (Fig. 25), and plots of derived STOCSY-CA scores vs. single variables (Figs. 35 and 45) are available as Supporting Information.

Results and Discussion

The 117 bark samples in this study are from a collection hosted by the “NaturMedicinsk” Museum at the University of Copenhagen. Originally, the collection was started by the Danish pharmacist Alfred Benzon who in the middle of the 19th century started the production of drugs and other medical products in Copenhagen. He was also a collector of plants, birds, shells, and archeological artifacts in accordance with the general ideas about natural sciences of that time. At his death 1884, the collection grew from donations from individuals and scientific expeditions, including collections from Java. When Denmark in 1917 sold the Danish West Indies (now Virgin Islands) to the USA, the collection grew with the content of the complete stock of the pharmacy of St. Thomas. All of the bark samples in this study were marked as “Cinchona” (or variations thereof), but many of the samples were not further described. In many cases the region and species are indicated on labels, but the authority of this information was not validated by further studies of the museum’s registry. The origin of the samples (number of samples in parenthesis) was distributed as follows: Peru (10), Bolivia (6), China (6), Java (6), Ecuador (5), Venezuela (5), India (2), Martinique (2), St. Thomas (2), Brazil (1), Central America (1), Congo (1), Cuba (1), St. Helena (1), and Jamaica (1). For 67 of the samples, there were no records about the place of origin. The species of the samples were determined as C. succirubra (23), C. calisaya (7), C. micrantha (5), C. pubescens (4), C. rubignosa (2), C. nitida (2), C. lancifolia (1), C. ledgeriana (1), and C. condaminea (1). For the majority of the samples (71), there were no indications of the species, and for 44 of the samples, neither origin nor species were included in the descriptions of the samples. Nevertheless, the collection should be a representative selection of Cinchonae cortex used against malaria between 1850 and 1950.

A pilot study, using six random samples from the collection, showed that a 1:1 mixture of chloroform and methanol extracted more alkaloids than neat chloroform or methanol as evaluated from the 1H NMR spectra of the extracts (data not shown). Lowering the pH by adding trifluoroacetic acid (TFA) to the solvent mixture increased the yield further. However, TFA and methanol react slowly to form the ester (CF3COO-d4) as shown by separate 13C NMR experiments of a mixture (data not shown). Therefore the non-nucleophile, and less reactive, perchloric acid (HClO4-d7) was considered a better choice. The pKa values of quinine are reported to be around 4 (nitrogen in the aromatic quinolone ring) and 8.5 (nitrogen in the bicyclic quinucleidine) [11] so that by the addition of acid, it is likely that the main alkaloids should be doubly protonated; and evidently they are extracted to a higher degree in this solvent mixture.

A profiling study like this is dependent on the spectra being reproducible. However, the exact chemical shift of alkaloid resonances are known to be sensitive to the acidity of the solvent [12]. Hence, a titration experiment was performed to find out a suitable amount of perchloric acid to add. A fixed amount of quinine (2 mg) was dissolved in a 1:1 mixture of chloroform-d4 and methanol-d4. Perchloric acid was added in steps of 10 µL and 1H NMR spectra were acquired with the spectrometer locked on the deuterium signal from the methyl group of methanol-d4. It was found that after an addition of 70 µL aqueous perchloric acid (70%), the position of the aromatic resonances of quinine was constant (with further changes smaller than the line widths) (Fig. 2). This experiment also showed that the chemical shifts of water and chloroform were changing by the additions of acid. The maximal displacement of the water signal was 0.65 ppm while the chloroform signal moved from 7.56 to 7.50 ppm. As the chemical shift of water is highly dependent on hydrogen bonding, the large shift was interpreted as an effect of different concentrations of water in the solvent. The changing position of the chloroform signal might also be related to the concentration of water in the solvent, but it was also observed that the chloroform signal shifted over time (data not shown). This effect was attributed to the evaporation of chloroform from the sample.
tubes. The three main conclusions from the titration experiment was thus that (1) the solvent should have an acid concentration of at least 0.8 M, (2) that the exact concentration of water needs to be controlled at a higher concentration than 1.8 M, and (3) that the tubes need to be carefully sealed for long-time stability of the measurements.

The extraction solvent was thus augmented with water and perchloric acid, and for simplicity the employed ratios of chloroform-$d_4$, methanol-$d_4$, $D_2O$, and aqueous 70% perchloric acid were determined to be 5:5:1:1. The ternary molar ratios of this solvent system were approximately 22:48:31, which forms a one-phase mixture [13]. The concentration of water and perchloric acid was 7.5 and 1.0 M, respectively. To keep the concentration of the volatile chloroform constant over time, it was necessary to seal the prepared sample tubes using a gas torch and a pair of pliers. To minimize the variability of the solvent system, a stock solution was prepared (with TMS added) and divided into several smaller bottles that were kept cold and sealed until use.

Of the 117 different bark samples, approximately half were prepared twice (replicates, $n=58$), and all of the sample tubes were analyzed twice (duplicates) with a mean time between measurements of 15 hours. The final number of spectra was 350. The sum of pairwise squared differences between duplicate measurements showed that the difference for one particular sample was considerably higher than for the rest. The reason was the increase over time of a set of characteristic triplets at 1.38, 1.34, and 1.21 ppm with accompanying multiplets in the range 3.6–4.4 ppm. The same pattern could be found in two other samples, and by comparison with the replicate samples, this was attributed to an artifact created during sample preparation. These six spectra were excluded from further analysis. Other variations between duplicate spectra revealed small differences in shimming and phasing, especially for those samples with small amounts of extracted material as judged by the appearance of the spectra.

An initial principal component analysis of the remaining spectra after normalization to constant sum and Pareto-scaled integrated chemical shift ranges (344 spectra × 876 variables, 6 components) showed that one sample had a particularly high $T^2$-value (distance from model center). A closer look at the four spectra of this sample revealed a pattern not seen among the other spectra. Our conclusion is that this bark sample is not from a cinchona tree, and hence it was excluded from the data set. The label of the sample indicated that it was collected in Central America, or added to the collection, in the year 1876. It would be intriguing if this bark had an antimalarial effect, but the most probable cause is a simple mislabeling since this was the only sample in the collection with this particular pattern.

The remaining 340 spectra were normalized to the area of the residual solvent signal of methanol-$d_4$ to reflect the extracted amounts of alkaloids and other compounds. Normalization using the TMS signal was not possible as this increased the differences between most of the replicates (data not shown). Possibly, the volatile TMS had evaporated to different degrees during the preparation of the samples.

The scores plot from the PCA (340 spectra × 876 variables, 6 components explaining 84% of the variance) did not show any apparent clustering (Fig. 3). The samples listed as C. succiruba tend to be in the middle of the plot, and the two samples listed as C. nitida are in the extreme end of both the first two principal components. The majority of the samples were however without information about species (marked with black dots), and those are scattered all over the score plot making it impossible to identify any clustering based on species. The distances between repeated measurements are in most cases not discernible in the score plot, and the distances between replicated preparations are small considering the overall variance of the whole dataset.

Selected spectra represented in different ranges of the score plot show the influence of the different components in the two first principal components (Fig. 4) [7.4–9.4 ppm] and Fig. 15 [0–9.5 ppm]). The first component (explaining 53% of the variance) represents the total amount of extracted alkaloid components as the spectra in the circled area A are devoid of major signals compared to the spectra in circled area D. The main differences between spectra along the second component, spectra in B compared to C, are related to signals from quinine (1) and cinchonine (3). This can also be seen in the loadings for the second component (Fig. 5) where variables attributed to quinine are negative and to cinchonine positive. The spectra of cinchonine (3) and cinchonidine (4) are very similar with differences between the signals of only a few Hertz, in the same size as the coupling constants in the doublets. The signals of H-2′ at δ 9.12 and H-5′ at δ 8.54 are two examples where the signals overlap even at a field strength of 600 MHz. The consistency in the chemical shifts between samples demonstrates that the system is robust enough to capture the differences between the stereoisomers of the alkaloids. Signals from quinidine (4) could not be observed.

Signals are also observed that cannot be attributed to the four most known cinchona alkaloids shown in Fig. 1, e.g., the doublets at δ 9.25 and 9.05. From the mixtures, it is difficult to reach any conclusion about the identity of these compounds, especially since the specific solvent composition used in the analysis will affect the chemical shifts, making it difficult to use literature values...
of chemical shifts of potential candidate compounds as references. However, by using the STOCSY method (statistical total correlation spectroscopy), correlated signals in the dataset can be identified [10]. Hence, the variables corresponding to the signals from H-2’ in quinine and cinchonine were selected along the two unknown doublet signals. Correlation coefficients between the selected variables and all the other variables in the dataset are color-coded onto spectra (Fig. 6 and Fig. 2S). The selected spectra are the ones with the highest intensity of the selected variables compared to all spectra in the dataset. From the STOCSY spectra, it is clear that signals related to the same compound are detected. The peak at δ 8.90 (Fig. 6d) is hence correlated to the signals at δ 8.31, 8.19, 7.79, and 7.58. These signals are assigned to H-2’, H-3’, H-8’, H-7’, and H-5’ in quinine. In the full-scale STOCSY plot (Fig. 2S), the correlations continue with H-9 (δ 6.18), H-11 (δ 5.15 and 5.07), O-CH₃ (δ 4.12), and several signals in the aliphatic range. H-10 could not be observed as this signal is in the same range as the suppressed water signal and hence removed in the statistical analysis. The H-2’ signals from cinchonine and cinchonidine are partly overlapping, and the correlated signals marked in red belong to both these compounds. This is probably due to a positive biological correlation, as opposed to structural correlation, and the amounts of the two compounds in the samples are correlated. A high amount of cinchonine would then imply a high amount of cinchonidine and vice versa. This is also true for the signal at δ 9.2.5. The amount of this unknown structure is probably correlated to the amount of cinchonine (and to a lesser degree correlated to cinchonidine) as most of the signals with a high correlation coefficient are assigned to cinchonine. The signal at δ 9.05 is likewise correlated to signals showing the same pattern of chemical shifts as the known cinchona alkaloids, so this is most likely an analogue.

The STOCSY analysis can also be used to select a subset of variables that are related to specific compounds for further analysis. This provides a plot similar to a PCA scores plot, but with the axes tied to specific compounds. The methodology can be described by the following four steps: (1) Select a driver peak and use STOCSY to select a subset of variables (corresponding to the colored signals in Fig. 6); (2) Extract the first principal component using the selected variables only; (3) Calculate the residual of the dataset; and (4) Continue at step 1 with the next set of selected variables and the residual of the data as input. We call this method STOCSY-CA, where CA stands for component analysis. It is related to the SIMPLS-MA (simple interactive self-modeling multivariate analysis) used in multivariate curve resolution [14, 15], but this relatively simple extension of the STOCSY methodology does not seem to have been described earlier. Since the selected variables have only a positive correlation to each other, the scores should be directly correlated to the concentration when applied to NMR data. Fig. 7 shows the result from selecting two signals from quinine and cinchonine (H-9) at δ 6.19 and δ 6.34, respectively. Variables with correlation coefficients larger than 0.7 were used to calculate the pure components. It can be noted that the amounts of the two alkaloids are correlated over the whole dataset. Replicates with large differences are also correlated along the same axis, which is an indication that the main contribution to the differences between replicates are caused by the extraction procedure.

Spectra with large model residuals (Q²) in either of the two PCA models are marked in the plot (p < 0.05). These spectra indicate that the STOCSY-selected variables do not have the same relative intensities as for the bulk of the spectra. This was attributed to overlapping signals from unrelated components of the mixtures. However, STOCSY-CA-derived scores plotted against single variables (H-2’ of quinine and cinchonine) show good linear relationships (Figs. 3S and 4S). Since the derived scores are linearly fitted to the values of several selected variables, they should be a more robust measure of the amounts of the compounds compared to using single variables, and with the detection of possible outliers, odd spectra can be identified and studied in detail. The data does not imply that there is any significant breakdown of alkaloids during storage since there are no negatively corre-
lated signals that would indicate that, e.g., quinine is decomposed. Given the old age and diverse background of the collection, unspecific degradation of the active constituents is still possible, and this would be difficult to detect with the methods used for this work especially if the original sample contained low amounts of alkaloids. However, the large variation of the amounts of extracted alkaloids suggested by this study is interpreted as an effect of the evolution of the quinine-producing enterprises during the period when the collection was compiled.

Acknowledgements

“NaturMedicinsk Museum” at the University of Copenhagen is gratefully acknowledged for the *Cinchona* bark samples. The 600 MHz spectrometer used in this work was acquired through a grant from “Apotekerfonden af 1991” (Copenhagen).

Conflict of Interest

No conflict of interest exists regarding the content or conclusions in this work.

References

9. Pearson K. On lines and planes of closest fit to systems of points in space. Philos Mag Ser 6 1901; 2: 559–572