Forensic Science International 269 (2016) 42–49

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Cocaine classification using alkaloid and residual solvent profiling

Louise Stride Nielsen[a,b,*], Christian Lindholst[a], Palle Villesen[b,c]

[a] Department of Forensic Medicine, Section for Forensic Chemistry, Aarhus University, Palle Juul-Jensens Boulevard 99, 8200 Aarhus N, Denmark
[b] Bioinformatics Research Centre, Aarhus University, C. F. Møllers Allé 8, 8000 Aarhus C, Denmark
[c] Department of Clinical Medicine, Aarhus University, Denmark

ARTICLE INFO

Article history:
Received 8 April 2016
Received in revised form 31 October 2016
Accepted 4 November 2016
Available online 11 November 2016

Keywords:
Illicit drugs
Link establishment
Logistic regression
Discriminant analysis
Gas chromatography–mass spectrometry
Headspace analysis

ABSTRACT

Many different groups of chemical compounds can be used in statistical impurity-profile comparison in order to establish links between different seizures of illicit drugs. For cocaine, some of these compounds descent directly from the coca leaf while others are remnants from the manufacturing process; and each of the compound groups exhibit different degrees of stability and discrimination power. Information obtained from the different groups can be handled in numerous ways and selecting the right method using a balanced combination of the compound groups is highly important in order to provide investigators and courtrooms with accurate conclusions.

By using logistic regression or discriminant analysis (linear and quadratic), cocaine alkaloid and residual solvent distances can be combined in order to obtain probabilities for the two possibilities: linked or unlinked. We examined different data transformations and distance methods and ranked the different models using cross validation. Validation in an unrelated data set proved the consistency of the results. Data consisted of five large groups of linked samples exposed to different storage conditions during 12 months, 124 different cocaine sample seizures and 15 smaller groups of linked samples stored at room temperature for up to 15 months. The alkaloid and residual solvent impurity profiles of the samples were analysed using gas chromatography–mass spectrometry (GC–MS) and headspace GC–MS, respectively.

Residual solvent profiles exhibited considerable higher discrimination power than cocaine alkaloid profiles. Thus, a residual-solvent-weighted model (log10 transformation and cosine distance) was found superior at distinguishing correctly between linked and unlinked seizures compared to models using alkaloid distance alone. The model only gives weight to the residual solvents when the alkaloid profiles are very similar. This finding demonstrates the possibility to combine information from the highly stable, non-coca leaf-descent residual solvent profiles and the less stable cocaine alkaloid profiles for statistical comparative analysis of cocaine seizures in a simple and easy-to-implement way.

1. Introduction

Statistical classification of illicit drug profiles is used to determine whether two profiles (seizures) originate from the same production batch [1–8]. Such knowledge can be used as evidence in police investigations and in court, and may also have great importance for the person being prosecuted during the sentencing [9].

The modelling and prediction from gas chromatography–mass spectrometry (GC–MS) profiles is highly discussed in aspects of finding the most suitable method to distinguish linked and non-linked seizures. The selected methods are often used in strategic intelligence purposes. In this approach, new GC–MS profiles are continuously fed into a database with the purpose of establishing links between new and already existing profiles [9–11]. An optimal method would allow discrimination between pairs of truly linked profiles (true positives) and pairs of truly unlinked profiles (true negatives) with as few errors as possible (low false positive and false negative error rates) (Fig. 1) [2].

Cocaine alkaloids are directly related to the coca leaves from which cocaine is extracted [12]. Therefore, cocaine seizure classification is commonly performed using coca alkaloid profiles. In the literature, the proposal for statistical cocaine alkaloid-profile

* Corresponding author at: Department of Forensic Medicine, Section for Forensic Chemistry, Aarhus University, Palle Juul-Jensens Boulevard 99, 8200 Aarhus N, Denmark.
E-mail address: lsni@forens.au.dk (L. Stride Nielsen).

http://dx.doi.org/10.1016/j.forsciint.2016.11.007
0379-0738/

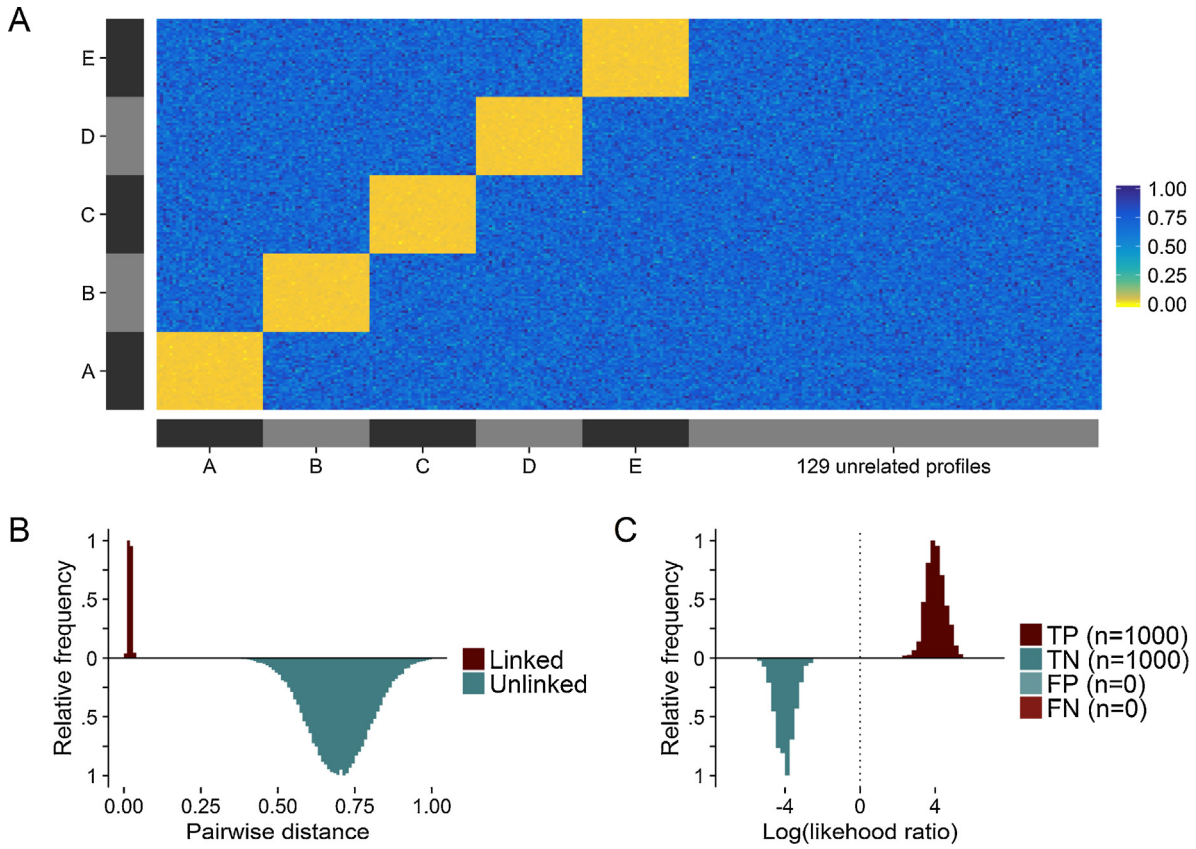

Fig. 1. Example of the optimal classification situation illustrated by the data set used for classification. (A) All linked pairs of five samples are shown in yellow (indicating short distances) and all unlinked pairs are shown in blue (indicating long distances). (B) Distribution of pairwise distances for linked pairs (red, top) and unlinked pairs (blue, bottom). The overlap is minimal which makes it possible to set a distance threshold that will distinguish linked from unlinked pairs. (C) The ideal distribution of loglikelihood ratios from modelling. All loglikelihood ratios > 0 are predicted to be linked. Both linked samples (red, top) and unlinked samples (blue, bottom) correctly predicted with no errors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

comparison has been to pre-treat data (e.g. normalisation and/or transformation) followed by a distance measurement (typically Pearson distance) and finally selecting a fixed threshold to distinguish linked from unlinked pairs of seizures [1,2,13,14].

Ideally, a single variable (e.g. Pearson distance or cosine angle between pairs of profiles) can discriminate linked from unlinked pairs of seizures using a fixed threshold (Fig. 1). Using a low threshold assumes that all profiles originating from the same production batch (linked) remain unchanged over time regardless of the alterations and storage conditions to which the batch has been exposed. In reality, profiles are not stable over time [15]. Thus, two linked profiles stored under different conditions will drift apart over time and ultimately look unlinked when a low threshold is used for discrimination. Using a higher threshold will overcome this problem, but it will also result in more false positives where pairs of unrelated profiles (unlinked) are wrongly classified as linked.

Opposite the alkaloid profile, residual solvents occluded in the cocaine crystals are more stable upon storage at different conditions [15]. They originate from the chemicals used during the manufacturing process and are therefore not directly related to the coca plant [2]. Thus, these solvents are additives to the clandestine production that have been trapped inside the cocaine crystals. A comparative analysis based solely on the residual solvent profile would therefore not be a feasible strategy as the comparative analysis would be a comparison of different solvent mixtures rather than a comparison of compounds which descent directly from the coca leaf e.g. alkaloids. For purposes of performing valid comparative analyses, the residual solvent profiles can, however, be applied as an additional discrimination criterion [16].

Monfreda et al. combined the use of alkaloid and residual solvent profiling for statistical cocaine-seizure classification using data obtained from FTIR. The alkaloid and residual solvent profiles were equally weighted, and the data was processed using the first ten principal components from principal component analysis (PCA) and linear discriminant analysis (LDA) [17]. This method was shown to be applicable for pure cocaine seizures. However, for cocaine seizures adulterated with more than one substance and a content of cocaine below 60%, prediction errors started occurring.

In the present study, the basic idea is to incorporate the high stability of the residual solvent profiles to the less stable alkaloid profiles when the decision whether two seizures originate from the same production batch or not is inconclusive. Hence, information from residual solvent profiling is included as a weighted function to the information from the cocaine alkaloid profiling. We use generalised linear model (GLM), linear discriminant analysis (LDA) and quadratic discriminant analyses (QDA) to predict whether pairs of profiles are linked or unlinked. Five groups of known linked profiles from major cocaine samples exposed to different storage conditions during 12 months is used together with 124 different cocaine sample seizures and 15 groups of known linked profiles stored at room temperature for up to 15 months [15]. We show that combining the alkaloid distance and weighted residual solvent distance highly improves the statistical comparison of cocaine seizures in a simple and easy-to-implement way.