A Ricin Forensic Profiling Approach Based on a Complex Set of Biomarkers

Sten-Åke Fredriksson, David S. Wunschel, Susanne Wiklund Lindström, Calle Nilsson, Karen Wahl, Crister Åstot

www.elsevier.com/locate/talanta

PII: S0039-9140(18)30310-2
DOI: https://doi.org/10.1016/j.talanta.2018.03.070
Reference: TAL18504

To appear in: *Talanta*

Received date: 21 November 2017
Revised date: 15 March 2018
Accepted date: 22 March 2018

Cite this article as: Sten-Åke Fredriksson, David S. Wunschel, Susanne Wiklund Lindström, Calle Nilsson, Karen Wahl and Crister Åstot, A Ricin Forensic Profiling Approach Based on a Complex Set of Biomarkers, *Talanta,* https://doi.org/10.1016/j.talanta.2018.03.070

A Ricin Forensic Profiling Approach Based on a Complex Set of Biomarkers

Sten-Åke Fredriksson[a], David S. Wunschel[b], Susanne Wiklund Lindström[a], Calle Nilsson[a], Karen Wahl[b], Crister Åstot[a]*

[a]FOI CBRN Defence and Security, Swedish Defence Research Agency, SE-90182 Umeå, Sweden

[b]Chemical and Biological Signature Sciences, National Security Directorate, Pacific Northwest National Laboratory, P.O Box 999, MSIN P7-50, Richland, Washington

Abstract:

A forensic method for the retrospective determination of preparation methods used for illicit ricin toxin production was developed. The method was based on a complex set of biomarkers, including carbohydrates, fatty acids, seed storage proteins, in combination with data on ricin and *Ricinus communis* agglutinin. The analyses were performed on samples prepared from four castor bean plant (*R. communis)* cultivars by four different sample preparation methods (PM1 – PM4) ranging from simple disintegration of the castor beans to multi-step preparation methods including different protein precipitation methods. Comprehensive analytical data was collected by use of a range of analytical methods and robust orthogonal partial least squares-discriminant analysis- models (OPLS-DA) were constructed based on the calibration set. By the use of a decision tree and two OPLS-DA models, the sample preparation methods of test set samples were determined. The model statistics of the two models were good and a 100% rate of correct predictions of the test set was achieved.

Keywords: Ricinus communis; Castor bean; Ricin production methods; Source attribution; Chemical attribution signatures; Multivariate statistics

Introduction

Ricin is a potent Chem-Bio threat agent, frequently relevant in cases of intoxications and "white powder letter" incidents worldwide [1, 2]. The toxin is present in the seeds of the castor bean plant, *Ricinus communis*, and active toxin can be produced from seed material by simple extraction methods. With the toxicity of ricin and the high availability of castor beans, ricin has become a threat of

concern for the society. The castor plant grows wild in tropical and subtropical climates and the plant is cultivated in large scale for the commercial production of castor oil, a major component of the castor bean, with almost two million tons of seeds harvested annually [3]. Castor oil has a high resistance to thermal degradation which makes it valuable in various industrial applications. Also, the striking appearance of the plant makes it popular as an ornamental, and the seeds are commercially available at garden shops. During the cold war, ricin was included in the chembio-weapons programs of several states. Therefore, ricin was classified as a Schedule 1 controlled substance under the Chemical Weapons Convention, and the implementation of the Convention in the national legislation of the 192 signatory states (June 2017) makes undeclared ricin purification a crime globally [4]. Thus, castor beans are an unregulated agricultural product but the intentional purification of the toxin, present in the seed's pulp, is regulated in U.S. federal law and in the national law of most other countries. The production of ricin from castor beans has been described both in scientific and anarchist literature, and on the internet numerous hits related to the topic can be found. The technical level of the methods varies from simple processes where the seeds are mashed with or without extraction of residual oil [5-7] to multi-step procedures including protein precipitation and chromatographic purification techniques [8-12]. The major components of castor beans include lipids, proteins and carbohydrates, whereas ricin constitutes about 1% of the seed dry weight. Ricin preparation procedures involve processes to enrich the toxin by removal of the bulk components. This induces dramatic changes in the composition of a ricin sample compared to an unpurified seed mash [13]. The seeds have a hard protective coat and many purification protocols starts with removal of the seed coat to access the pulp and castor oil. Castor oil, rich in ricinolein, the triglyceride of ricinoleic acid, is a dominating component of the pulp and toxin purification usually starts with the removal of the oil by solvent extraction. Further steps may involve aqueous extraction of the toxin from the mashed seed pulp followed by protein precipitation. Ricin and the closely homologous lectin *R. communis* agglutinin (RCA) can be separated from other highly abundant proteins in the precipitate by affinity purification [12].

Forensic analysis of samples suspected to contain ricin has mostly been focused on screening, detection and verification of the presence of ricin (*e.g.* in "white powder samples") [14, 15]. The methods often involve trypsin digestion and mass spectrometric analysis of the resulting peptides. Unambiguous identification of ricin requires MS/MS-based methods, giving detailed

amino acid sequence information of ricin-specific peptides, strictly distinguishing them from other proteins including RCA [10]. In addition, methods for chemical analysis of forensic attribution markers related to the purification of ricin, have been the focus of a number of studies. The markers studied include residual solvents [16], carbohydrates and ricinoleic acid since they are linked to the production process of ricin [17]. The residual solvent analysis focuses on traces of the reagents (i.e. solvents) used for ricin preparation, while the latter methods use the level of ricinolein and polysaccharides in the sample as markers for purification method efficacy. The alkaloid ricinine [18] and a set of highly crosslinked peptides, the *R. communis* biomarkers (RCB's), are also potential attribution markers for ricin [19].

Acquisition of comprehensive analytical data by multiple techniques is a common approach to monitor changes in the properties or composition of complex samples. Integrated analytical platforms including GC-MS, LC-MS and NMR offer sensitivity and reproducible detection to monitor biological processes such as disease development and plant responses [20, 21]. The multi-omic approach has been applied to research questions across the fields of medicine, molecular biology and plant physiology [22-24]. Integrating analytical information of a broad array of attribution markers; low molecular weight compounds as well as metabolite and protein marker profiles, provides for improved tools to detect intentional toxin production and to differentiate between production techniques. Such methods would also open up for association of different ricin samples, *e.g.* samples produced by the same method and relating a site or illegal toxin production method to a threat or use of ricin [25].

This study, a collaboration between FOI CBRN Defence and Security (FOI) and Pacific Northwest National Laboratory (PNNL), sought to develop a comprehensive strategy for forensic ricin analysis. The approach was to acquire data on the composition of ricin samples produced by four different preparation methods using a number of different analytical techniques, and to apply multivariate statistical methods to evaluate the complex chemical analytical information. The statistical methods would support the conclusion whether a crime was committed, *i.e.* if the toxin was enriched and purified. In addition, the analysis would reveal details of forensic interest regarding preparation methods as well as information on the seed material that could be linked to a specific case.

Experimental

Plant material.

Castor seeds from four selected cultivars of *R. communis* were exchanged between FOI and PNNL (Table 1).

Table 1. The *R. communis* cultivars used for comparative analysis on extraction methods.

Cultivar #	Cultivar	Type	Source
#1	Sanguineus	Ornamental	Rara växter, Ornö, Sweden
#2	Impala	Ornamental	Sandeman Seeds, UK
#3	TMVCH 1	Oil seed	Asian Green Fields, Tamil Nadu, India
#4	GCH4	Oil seed	Asian Green Fields, Tamil Nadu, India

Preparation of "illicit samples".

Ricin extracts of the cultivars were prepared in duplicate using four preparation methods, PM1 – PM4, selected from the scientific literature, clandestine publications and internet sites (Table 2) in accordance with Wunschel et al 2012 [13]. They range from simpler methods using normal household equipment to more technical procedures for extracting ricin. The performance of the preparation methods, in terms of ricin purity, varies from the simple PM1 to a more technically demanding PM4. However the purification efficiency of a method depends on a number of factors. The selected procedures were adapted for laboratory use and are briefly described below.

Table 2. Ricin preparation methods used in the study.

PM1	Crushed de-coated beans [7] The beans were soaked in 2 M sodium hydroxide solution. The seed coat was removed and the beans were crushed and homogenized using a mortar and pestle.
PM2	Acetone washed bean homogenate [8] The bean homogenate (PM1, 500 mg) was extracted with acetone and the bean mash was allowed to dry at RT.
PM3	PBS extraction and acetone precipitation [13] The bean homogenate (PM1, 500 mg) was extracted in phosphate buffered saline (PBS),

	centrifuged and the middle aqueous layer was recovered. Acetone was added to the extract until the formation of a precipitate. The precipitate was allowed to dry at RT.
PM4	Acidic extraction and ammonium sulfate precipitation [11, 13] The method was adapted from Despeyroux et al (2000) [11]. The bean homogenate (PM1, 500 mg) was homogenized in sodium chloride (without solvent extraction), the aqueous supernatant was collected and precipitated using ammonium sulfate. The precipitate was allowed to dry under vacuum.

Subsamples from the different preparations were analyzed for proteins, protein markers as well as the content of carbohydrates, fatty acids and solvent residues. In Figure 1 a flowchart of the sample preparation procedures and adopted analytical techniques is presented. Compound classes associated with each of the sample fractions are indicated.

Chemical analysis

Derivatization Reagents and Materials.

Sodium hydroxide solutions were prepared from sodium hydroxide pellets (Sigma-Aldrich, St. Louis, MO, US). For carbohydrate and fatty acid derivatization, acetone, methyl-tert butyl ether (MTBE), boron trifluoride, methanol, hexane, sulfuric acid, sodium borodeuteride, acetic acid, ammonium sulfate and chloroform were obtained from Sigma-Aldrich at ACS reagent grade (St. Louis, MO, US). Acetic anhydride was obtained from Supelco, Inc. (Bellafonte, PA, US). The solid-phase extraction (SPE) was performed using C-18 SPE cartridges obtained from Phenomenex (Torrance CA, US), and Chem-Elut SPE cartridges purchased from Perkin-Elmer (Waltham, MA, US). The carbohydrate and fatty acid standards listed below were obtained in the highest purity available from Sigma-Aldrich (St. Louis, MO, US).

Carbohydrate determination.

The analysis of monosaccharides was performed using the alditol acetate method for hydrolysis, reduction and acetylation of the liberated sugars as used previously [17, 26]. Briefly, triplicate 10 mg samples of each PM were prepared by first performing hydrolysis under a nitrogen atmosphere in 1 M sulfuric acid at 100 °C for 12-16 hours. Following hydrolysis, samples were neutralized in N,N-dioctylmethylamine:chloroform solvent and the aqueous extract reduced using sodium borodeuteride. Following evaporation of borodeuteride after addition of methanol and acetic acid, the reduced

monosaccharides were acetylated with acetic anhydride and suspended in chloroform for GC-MS analysis. Nineteen monosaccharide standards were derivatized with each sample batch in known concentrations to use in quantifying the corresponding monosaccharides in each sample. Those included the neutral sugars fucose, rhamnose, ribose, arabinose, deoxyglucose, xylose, pinitol, *chiro*-inositol, *myo*-inositol, *allo*-inositol, *muco*-inositol, *scyllo*-inositol, mannose, D-glycero mannoheptose, galactose, and glucose as well as the amino sugars N-acetyl glucosamine, N-acetyl galactosamine and N-acetyl mannosamine. Due to the reduced and acetylated form of the final derivatives, amino sugars were all reported as their acetylated form. Two internal standards were also added to each sample and standard mixture to aid quantification, methyl glucose was added as a quantificationstandard for neutral sugars and N-methyl glucamine as an internal standard for amino sugars. Data was collected using an Agilent Technologies (Santa Clara, USA) 7890A Gas Chromatography Mass Spectrometry (GC-MS) equipped with an Rtx-5 MS column (Restek, Bellefonte, USA). A 1 μL sample injection was used in a 10:1 split mode.

Lipids determination.

Analysis of fatty acid content focused on the measurement of ricinoleic acid following derivatization and GC-MS as described earlier [13, 17]. Briefly, a 10 mg sample was saponified in sodium hydroxide at 70 °C for 2 hours. The liberated acid monomers were acidified and extracted into MTBE. The methylation reaction was performed in methanol and 3% boron trifluoride for 2 hours at 100 °C followed by extraction into hexane and evaporation. Samples were reconstituted in methanol for analysis. Two 50 μg aliquots of a ricinoleic standard were also derivatized with each batch. Behenic acid was used as an internal standard by adding 50 μg to samples and standards prior to saponification. Methyl ricinoleate in the samples was determined by its relative peak area to behenic acid compared the corresponding peak area ratio in the standard mixture. Data was collected using the GC-MS system described above. A 1 μL sample injection was used in a 10:1 split mode.

*Determination of residual acetone***.**

Measurements of residual acetone were performed using head space analysis as previously described [16]. Briefly, 50 mg of sample was placed into a 10 mL headspace vial (Gerstel Inc. Baltimore, MD, USA) and the headspace sampled using a 1 cm solid phase microextraction (SPME) fiber coated with polydimethylsiloxane (Sigma-Aldrich, St. Louis, MO, USA). The vials were agitated and heated to 40 °C prior to exposing 3 mm of the fiber for 60 sec. Upon insertion into the injection port, the entire fiber was exposed to 250 °C desorption for 60 sec. A Leco Pegasus (St Joseph, MI, USA) tandem gas

chromatography (GCxGC) time-of-flight (TOF) mass spectrometer equipped with a Gerstel autosampler, first dimension 30 m, 250 μm ID, 0.25 μm df Sol Gel-Wax column (SGE, Austin, TX, USA) and second dimension 2 m, 100 μm ID, 0.40 μm DB-1701 column (Quadrex, Woodbridge, CT, USA) column was used for data acquisition. The TOF MS system acquired data from 35 *m/z* to 300 *m/z* at 100 Hz.

Profiling of protein and peptide biomarkers

The following parameters were measured: concentrations of ricin, RCA and SSP as intact proteins, RCB peptides [19] and trypsin digest peptides of ricin and RCA [15].

Chemicals.

Poros 20 AL affinity chromatography material was obtained from Thermo Fisher Scientific (Carlsbad, CA, USA). Millex HA membrane filters was purchased from Merck (Darmstadt, Germany). The affinity ligand, p-aminophenyl-1-thio-β-D-galactopyranoside, was obtained from Bachem, (Bubenhof, Switzerland). All other chemical were of analytical or LC-MS grade. Water was purified in a Millipore Milli Q system (Merck, Darmstadt, Germany). Mass spectrometry grade trypsin was purchased from Promega (Madison, WI, USA).

Analytical sample preparation.

A 0.5 g aliquot of PM1 was vigorously mixed in 8 mL of diethyl ether and 4 mL of 0.5% acetic acid. The acetic acid phase was collected after centrifugation and filtered through a 0.45 μm Millex HA syringe filter (Merck Chemicals and Life Science AB, Solna, Sweden). The filter paper containing the acetone-extracted sample from PM2 was extracted with 4 mL of 0.5% acetic acid. The extract was filtered as above. Precipitates from PM3 and PM4 were dissolved in 4 mL of 0.5% acetic acid. The extracts were filtered through a 0.45 μm Millex HA syringe filter. All extracts were stored at 4 ºC before further treatment. ***Affinity fractionation.*** The extracts were fractionated by affinity chromatography as previously described [15]. Affinity columns, 10 mm, 1.6 mm i.d., were packed with galactosyl-modified Poros AL chromatographic material and used in a Bio-Rad Duo-Flow chromatography system (Bio-Rad, Hercules, CA, USA) for fractionating the samples into a non-retained fraction and a fraction containing ricin and RCA.

Capillary gel Electrophoresis.

A Bio-Rad Experion capillary gel electrophoresis system (Bio-Rad, Hercules, CA, USA) was used for analysis of levels of native proteins in the extracts from PM1-4 and of the non-retained affinity chromatography fractions.

Enzymatic digestion.

The affinity enriched fraction containing ricin and RCA was evaporated to dryness at 70 °C under a mild nitrogen flow. A 40 µL aliquot of 100 mM ammonium bicarbonate and 0.4 µg of modified trypsin were added and the samples were digested at 40 °C for 2 h. Leucine encephalin (Sigma-Aldrich Sweden AB, Stockholm, Sweden) was added as an internal standard for quantification of the digest peptides.

Liquid chromatography-mass spectrometry.

The trypsin digests were analyzed on a nanoAcquity UPLC system interfaced to a Qtof Ultima mass spectrometer equipped with a Nano-Lockspray electrospray ion source (Waters, Milford, MA, USA). The TOF analyzer was operated in the V mode at 10 000 resolution (FWHM) and the MCP detector was integrated at 1 s intervals over mass range 300-1800 in the MS-mode and 100-1800 in the MSMS mode. The digests were diluted with 0.4% formic acid containing leucine enkephalin internal standard. Depending on the expected amount of protein in the sample 1 - 10 µL of the digest was injected and the peptides were separated on a 150 mm, 75 µm i.d. Waters BEH C18 nanoAcquity UPLC column using a 20 min linear gradient from 3 to 40% acetonitrile containing 0.1% formic acid. The eluent composition was then stepped to 80% acetonitrile, which was held for 4 min before reverting back to the initial conditions.

LC-MS analysis of RCB peptides in the non-retained fraction.

For the analysis of the non-retained fraction a 5 mm 300 µm i.d. PepMap C18 column (Thermo Fisher Scientific, Waltham, MA, USA) was used. A 5 µL aliquot was injected and the column was eluted by a 2 minute gradient from 30% to 80% acetonitrile in water containing 0.1% formic acid at 7 µL/min. The standard electrospray source was used and spectra were acquired over the mass range 300-2200.

LC-MS identification and quantification of ricin and *R. communis* agglutinin (RCA). Ricin and RCA were identified by accurate mass LC-MS analysis of trypsin digest peptides and comparison with reference standards of purified ricin and RCA. The amino acid sequences of the peptides were confirmed by LC-MS/MS analysis and examination of their product ion spectra. The responses of selected peptides relative to leucine enkephalin internal standard were evaluated from the extracted ion chromatograms using a 20 ppm mass window (TargetLynx software, Waters.).

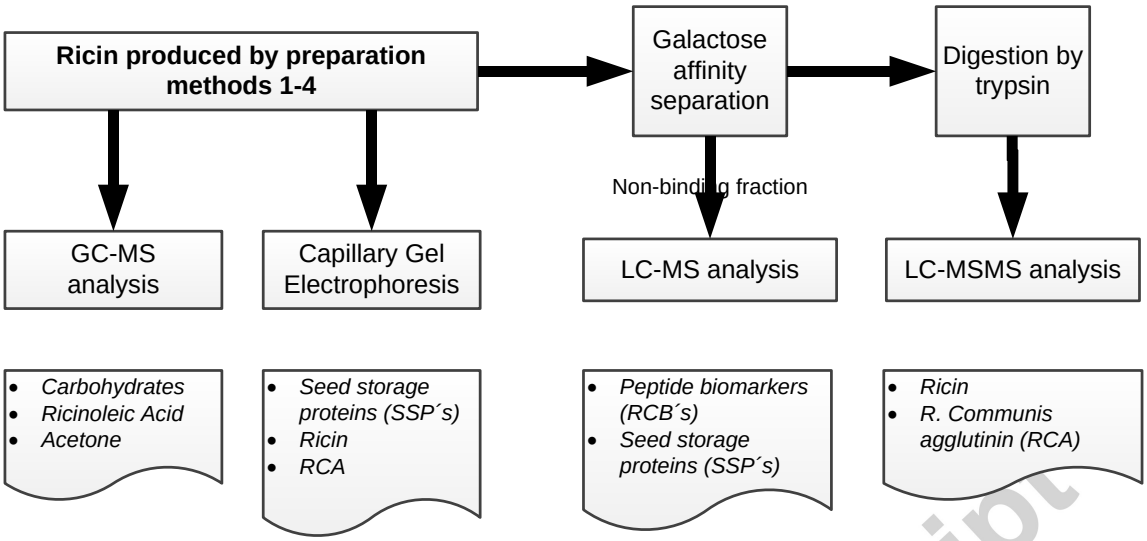

Figure 1. Flowchart of sample processing and analytical methods used in the investigation. Additional sample preparation information and references can be found in the text.

Table 3. Parameters measured in the ricin preparations. (Variable ID numbers used in Figure 4 and 6).

: Carbohydrates GC-MS, : Lipids GC-MS, : Solvent (acetone) GC-MS, : Proteins (CGE), :RCB (LC-MS), : SSP (LC-MS), : Trypsin digest peptides (LC-MS)

#	Variable	#	Variable	#	Variable
1	Rhamnose	**21**	SSP-1H	**41**	T3b-T5b
2	Ribose	**22**	SSP (total)	**42**	T6b
3	Fucose	**23**	Ricin	**43**	T10b
4	Arabinose	**24**	RCA	**44**	T11b
5	Xylose	**25**	RCB-1	**45**	T13b
6	Chiro-inositol	**26**	RCB-2	**46**	T14b-T16b
7	Myo-inositol	**27**	RCB-3	**47**	T17b
8	Mannose	**28**	SSP-1 (Ric c 1)	**48**	T18b
9	Glucose	**29**	SSP-2 (Ric c 3)	**49**	T19b
10	Galactose	**30**	T2a	**50**	T20b
11	Glucosamine	**31**	T5a	**51**	T22b
12	Ricinoleic Acid	**32**	T6a	**52**	T22b2
13	Acetone	**33**	T7a	**53**	RCA T6a
14	SSP-1A	**34**	T9a	**54**	RCA T3b
15	SSP-1B	**35**	T10a	**55**	RCA T4b
16	SSP-1C	**36**	T11a	**56**	RCA T11b (GI:225114)
17	SSP-1D	**37**	T14a	**57**	RCA T15b-T17b (GI:169715)
18	SSP-1E	**38**	T19a		

19	SSP-1F	**39**	T20a	**58**	RCA T15b-T17b (GI:225419)
20	SSP-1G	**40**	T24a-T1b		

Data processing

Prior to the statistical analysis, the GC-MS data on carbohydrates, fatty acids and solvent residues and the CGE data (intact ricin, RCA and SSP) and LC-MS data (RCB's and digest peptides of ricin and RCA) were normalized using internal standards. The complete data set, **X**[34,58], consisted of n=34 samples, k=58 variables and m=4 preparation methods and 8-10 samples for each of the preparation methods. The data matrix was divided into two calibration data sets and a test set. The calibration data for the two models applied, M1 and M2, will be referred to as $\mathbf{X}_{M1}$[26,58] and $\mathbf{X}_{M2}$[18,58], respectively. The data $\mathbf{X}_{M2}$[18,58] was a subset of $\mathbf{X}_{M1}$[26,58] (Figure 2 and Table S-1, Supplemental files). The selection of test set samples, $\mathbf{X}_{TS}$[8,58] (Table S-1, Supplemental files) was based on principal component analysis of **X**[34,58] to represent the total variation within the data (three principal components, data not shown). Prior to modeling, the $\mathbf{X}_{M1}$[26,58] matrix was log-transformed and data was scaled to unit variance.

Statistical method

All data was imported into SIMCA 14.0 (Sartorius-Stedim Analytics AB, Umeå, Sweden) and was evaluated with Orthogonal Partial Least Squares regression Discriminant Analysis (OPLS-DA). OPLS-DA is a supervised statistical method that uses information about the data to extract maximum amount of variation. A benefit with OPLS-DA is that variable selection is not very critical, since the algorithm can use variables with an unknown degree of correlation in contrast to other supervised classifications tools, which are limited to independent variable selection (*i.e.* Linear Discriminant Analysis). The analysis was performed using the two calibration dataset, $\mathbf{X}_{M1}$ and $\mathbf{X}_{M2}$, according to a classification decision tree that resulted from the data analysis (Figure 2).

Model statistics. OPLS-DA models were optimized using full cross validation (CV). During the CV 1/7th of the data points were excluded and the model was calculated from the remaining data, which was then used to predict the excluded samples. This procedure was performed 7 times until each sample was excluded and predicted once. The model was further validated using samples in PM2-PM4 as an external test set, $\mathbf{X}_{TS}$. The test set was selected based on Principal Component Analysis (PCA, three components), in **X** space from samples in PM2-PM4. PM1 was not included in the test set because of limited number of samples.

The following model statistics are used in this paper to assess the quality and the predictive ability of the models.

R^2X_{tot}=1-SS(E)/SS(X)

R^2X_{tot} represents the total explained variation in the model from the X data. E represents the residual matrix and SS is the sum of squares. R^2X multiplied by 100 is the percent of total variation in the X data.

$Q^2\hat{\boldsymbol{y}}$=1-SS($\hat{\boldsymbol{y}}$ -y)/SS($\boldsymbol{y}$)

$Q^2\hat{\boldsymbol{y}}$ represents the cross-validated predictive ability of each of the modelled classes where y is the true known class membership and $\hat{\boldsymbol{y}}$ represents the predicted decision value. When performing the CV and test set validation, each sample will be assigned to a class (PM1-PM4) based on the predicted decision value $\hat{\boldsymbol{y}}$ and the cut-off limit. The cut-off limit between classes was set to 0.5. Z is the percentage of correctly predicted and assigned class membership based on external test set. The model statistics of M1 and M2 are presented in Table 4.

Safety

Ricin is highly toxic and should be handled only by appropriately trained personnel. Ricin is listed in the Chemical Weapons Convention and is thereby subject to national and international regulations that have to be strictly followed. Preparation of crude extracts of plant material was performed in a fume hood, taking care to keep toxins in solution and avoid dust formation. Equipment coming in contact with the extracts was decontaminated by immersion in 2 M sodium hydroxide for 2 h. A Biosafety Class 3 glove box equipped with HEPA filters connected to a standard fume hood via an air lock was used for handling of particulate samples.

RESULTS AND DISCUSSION

In a forensic investigation involving ricin three critical questions are raised; Is ricin present? How toxic is it? Does the sample contain processed ricin intentionally purified from castor beans? Methods for the determination of the presence and activity of ricin are well established [10, 18, 27]. In order to answer the last question, attribution markers such as residual solvents, carbohydrates and fatty acids have been used to collect evidence on the retrospective determination of processed ricin [13, 16, 17]. However, due to high correlation between attribution markers, the use of a single set of markers offers a limited ability to unambiguously resolve individual ricin processing methods. In this study, we used a broader approach combining the analytical data from the low molecular weight forensic markers mentioned above with proteomic data of ricin samples. The aim was to develop a high resolution classification tool that can be used in a forensic aspect, as a law enforcement tool as well as for intelligence work. The main performance criteria for the classification model(s) were robustness and high predictive ability. In

addition, it is also a benefit to be able to extract information on attribution markers important for the preparation method classification.

Forensic ricin analysis.

In a forensic case, data on the enzymatic digest peptides of ricin and RCA are obtained as part of the identification of ricin toxin itself, since mass spectrometric analysis (*e.g.* LC-MS/MS) of diagnostic peptides is the preferred method for differentiation between ricin and other proteins including RCA. For the retrospective determination of ricin preparation methods in this publication, the ricin and RCA peptide data were combined with additional molecular markers for *R. communis* present in most illicit ricin preparations; RCB's [19] and SSP's [10, 28] and specific preparation method attribution markers; acetone, carbohydrates and lipids [13, 16, 17]. The main protein constituents, ricin, RCA and seed storage protein (SSP), together reflecting the toxin enrichment in the extracts, were separately quantified and included in the investigation.

The measured parameters were summarized in Table 3. Carbohydrates were chosen to reflect the carbohydrate profile of the monosaccharides derived from seed cellulose, hemicellulose and protein glycosylation [13, 17]. Ricinoleic acid was chosen as single marker for the lipid fraction in the ricin preparations (PM1 –PM4) since this fatty acid constitutes 85-95% of the total lipid content of the beans [29]. Acetone was chosen to reflect the use of solvents for removal of castor oil from the ricin seed mash. This is based on experience from real case work to be the most frequently used solvent for the purpose [16, 18]. The main protein constituents, ricin, RCA, SSP, and the RCB peptides, together reflecting the toxin enrichment in the extracts, were analyzed and included in the investigation. Thus the concatenated dataset includes the combined information on all markers.

Retrospective determination of method used for preparation of ricin.

Initially, an OPLS-DA model was constructed in order to predict unknown samples to the four preparation methods (data not shown). However, the different seed ecotypes introduce variation into the variables that was not related to the PM's. For example, the levels of the RCB peptides were shown to be related to seed ecotypes [19]. An alternative approach would have been sample set normalization using reference data from matched castor bean ecotypes, prepared by a standardized preparation method. By such a procedure, model statistics would have been improved, but a real case with an unknown sample would be cumbersome due to the need for ecotype assignment prior to preparation method analysis.

Instead, we used a two-step procedure in order to fully determine the preparation methods. By the use of a hierarchic approach, including two OPLS-DA models, the ecotype variation was suppressed and significant differences, only depending on the preparation method, were found. The decision tree is shown in Figure 2.

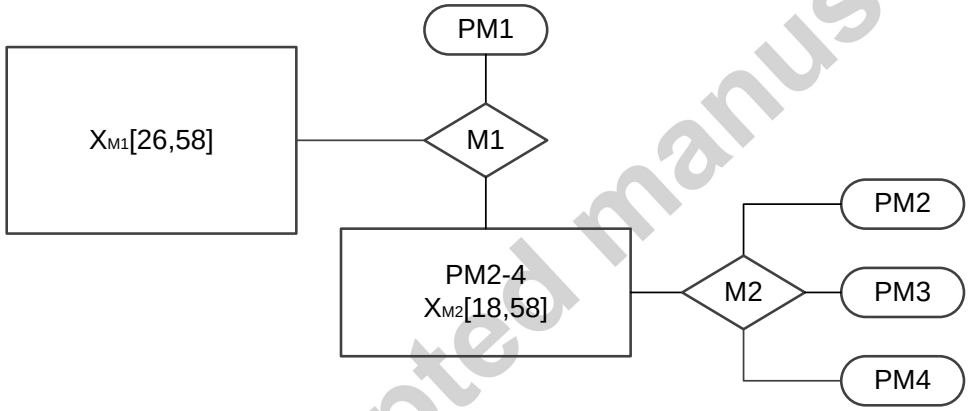

Figure 2. Classification by a decision tree. The first OPLS-DA model 1 (M1) was constructed on the calibration sample set $\mathbf{X_{M1}}$[34,58] and it classifies the samples into the subclasses PM1 or PM2-4. The second PLS-DA model 2 (M2) was constructed on the calibration sample set $\mathbf{X_{M2}}$[18,58] and classifies the PM2-4 subclass into individual preparation methods (PM).

In the first step OPLS-DA analysis of the data matrix $\mathbf{X_{M1}}$**[26,58]** and a dummy vector **Y[26,2]** was found to differentiate PM1 from PM2-4 with a cross-validated separation between the sub-classes (Figure 3). The model resolved simple castor bean mash (PM1) from more refined ricin toxin preparations produced by methods including solvent extraction to remove oil from the castor bean mash. Model M1 constitutes the first branch in a decision tree built for the process of unravelling ricin preparation methods (Figure 3).

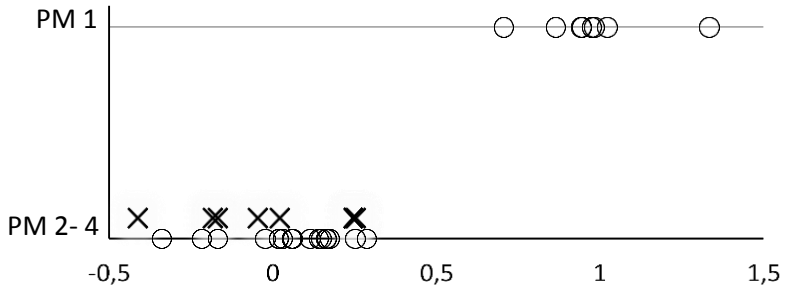

Figure 3. Predicted vs true value plots of the separation between preparation method PM1 and methods PM2, PM3 and PM4 in model 1 (M1). The true class membership is displayed on the y -axis and the predicted decision values on the x-axis. Circles represent CV model samples crosses the test set samples.

The decision values of each cross validated sample, is visualized in the predicted vs true value plot (Figure 3). The separation between the two subgroups PM1 and PM2-PM4 is shown by the predicted decision value on the x- axis and the true class membership on the y -axis. The two subgroups are well defined and allow a reliable assignment of an unknown sample. The validation shows that the model (M1) has a high prediction performance and all 8 samples tested were correctly predicted using the external test set (Table 4). The distribution of data points around 0 and 1 on the x-axis reflects the separation (Figure 3). Strong prediction power of the PM's was obtained, although the different seed ecotypes introduce variation into the variables that is not related to the PM's.

The relative importance of the attribution markers with strong impact on model M1 are shown in the loadings plot P[1] in Figure 4. Attribution markers with a strong dependence on the extraction of the bean mash are those with a high positive or negative P[1] value and with a small confidence interval, not crossing through zero in the loadings plot. Acetone is a marker with a significant loading for the washed bean mash present in PM2-4 (Figure 4, variable no 13). The main purpose of the acetone wash of the bean mash is to remove castor oil and consequently ricinoleic acid, the castor oil marker, has a very strong loading for the crude mash of PM1 (Figure 4, variable no 12). The level of some other markers were also significantly reduced by the extraction procedure, including the *R. communis* biomarker peptide RCB-2 (#26) and the content of the dominating isoforms of the seed storage proteins, SSP-1B (#15) and SSP-1D (#17). In addition to the extraction solvent acetone (#13), the carbohydrates (#1-11) are important markers that were enriched in the bean mash by the solvent wash. The significant loadings of some carbohydrates were expected, since the removal of the lipid, constituting approximately 30% of the bean mash, by

acetone extraction increases the relative amounts of carbohydrates. The significant reduction in the T24a-T1b attribution marker (#40), the disulfide-bound peptide link between the ricin A and B chains, could be a spurious correlation due to the low response of this peptide or, alternatively, a result of partial denaturation of the protein caused by the solvent extraction. The majority of the enzymatic digest peptides of the toxins, ricin and RCA, are found in the middle part of Figure 4 and are not significant for the PM discrimination. The peptides should ideally be present in stoichiometric amounts provided that the efficiency of the enzymatic digestion and the integrity of the proteins are preserved during the preparation procedure. These parameters are important for the identification of the proteins and for forensic profiling of samples and should ideally not depend strongly on the method that was used to make the toxin sample.

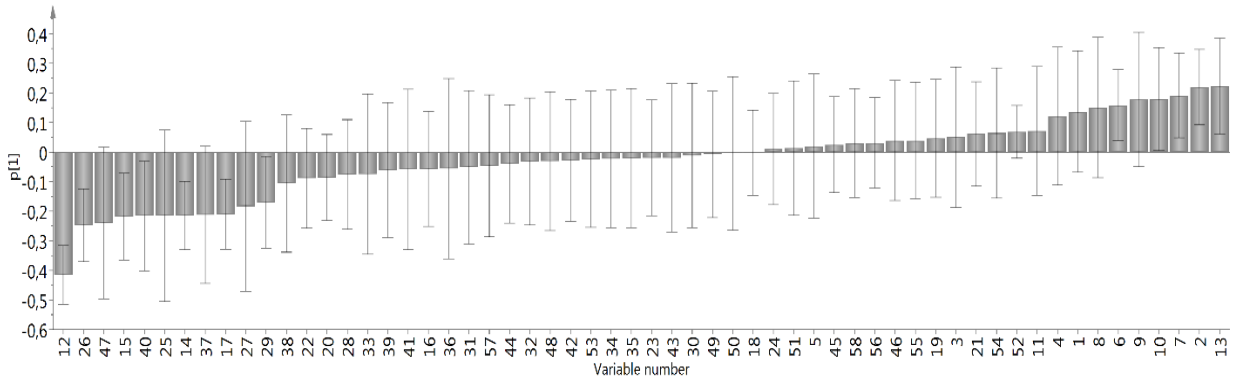

Figure 4. Loadings plot P[1] from model M1 in the decisions tree, indicating markers that separates crude bean mash (P[1] < 0) from preparation methods using solvent extraction to increase the toxin content (P[1] >0) (95% confidence intervals). Variable ID numbers as in Table 3.

By the use of a second OPLS-DA model (M2), the PM2-4 sub-class was further resolved into single PM groups. When a sample was predicted by M1 to belong to subgroup PM2-PM4 (the more sophisticated ricin preparations), the classification according to the decision tree continues. In this step, the dummy vector **Y[18,3]** was created from classes PM2-4 and samples of PM2, PM3 and PM4 were separated (Figure 5). The data shows that an unknown ricin sample can be retrospectively linked to preparation methods 2-4 by the use of the model M2. The model has a high predictive performance and resulted in 100% correct classification of the test set samples (Table 4).

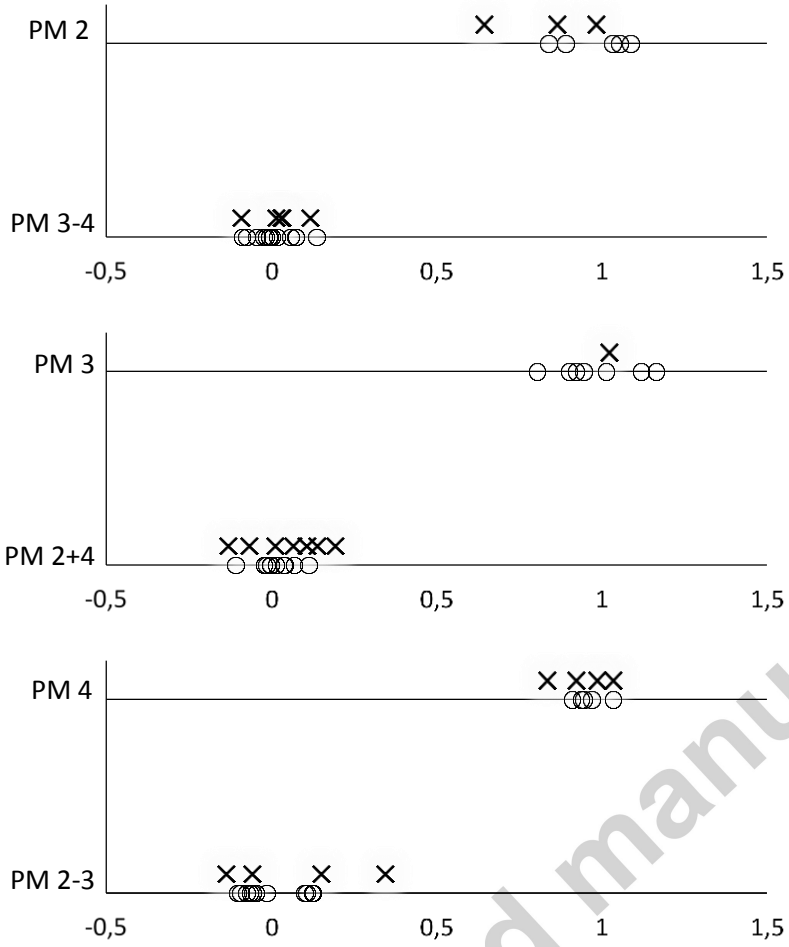

Figure 5. Predicted vs true value plots of the separation between the individual preparation methods PM2 from PM3+PM4 (top), PM3 from PM2+PM4 (center) and C) PM4 from PM2+PM3 (bottom) in model 2 (M2). The true class membership is displayed on the y -axis and the predicted decision values on the x-axis. Circles represent CV model samples the crosses the test set samples.

Table 4. Model statistics from M1 and M2. A is the number of OPLS-DA components resolved into predictive and orthogonal components. R^2X_{tot} is the total variation in the X_{M1} or X_{M2} data explained by the model and R^2X_{pred} is the amount of predictive variation explained by the model. Q^2y is the model predictive ability and Z is the percentage of correct predictions (classifications) using the test set (see also experimental section).

Model	**A**	**R^2X_{tot} (%)**	**R^2X_{pred} (%)**	**Q^2y (%)**	**Z (%)**
M1	1+5	76	8.5	87	100
M2	2+2	79	42	89	100

The separation of the three preparation methods and the low within class variation is visualized in the bi-plot (Figure 6) where the correlation scaled scores for the two predictive components t(corr)[1], t(corr)[2] (blue triangles) and loadings p(corr)[1], p(corr)[2] for the measured variables (circles) are combined in one single plot. The three preparation methods are now clearly separated with a low within-class variation. It can also be seen that markers with widely varying chemical properties contribute to the strength of prediction of the preparation method.

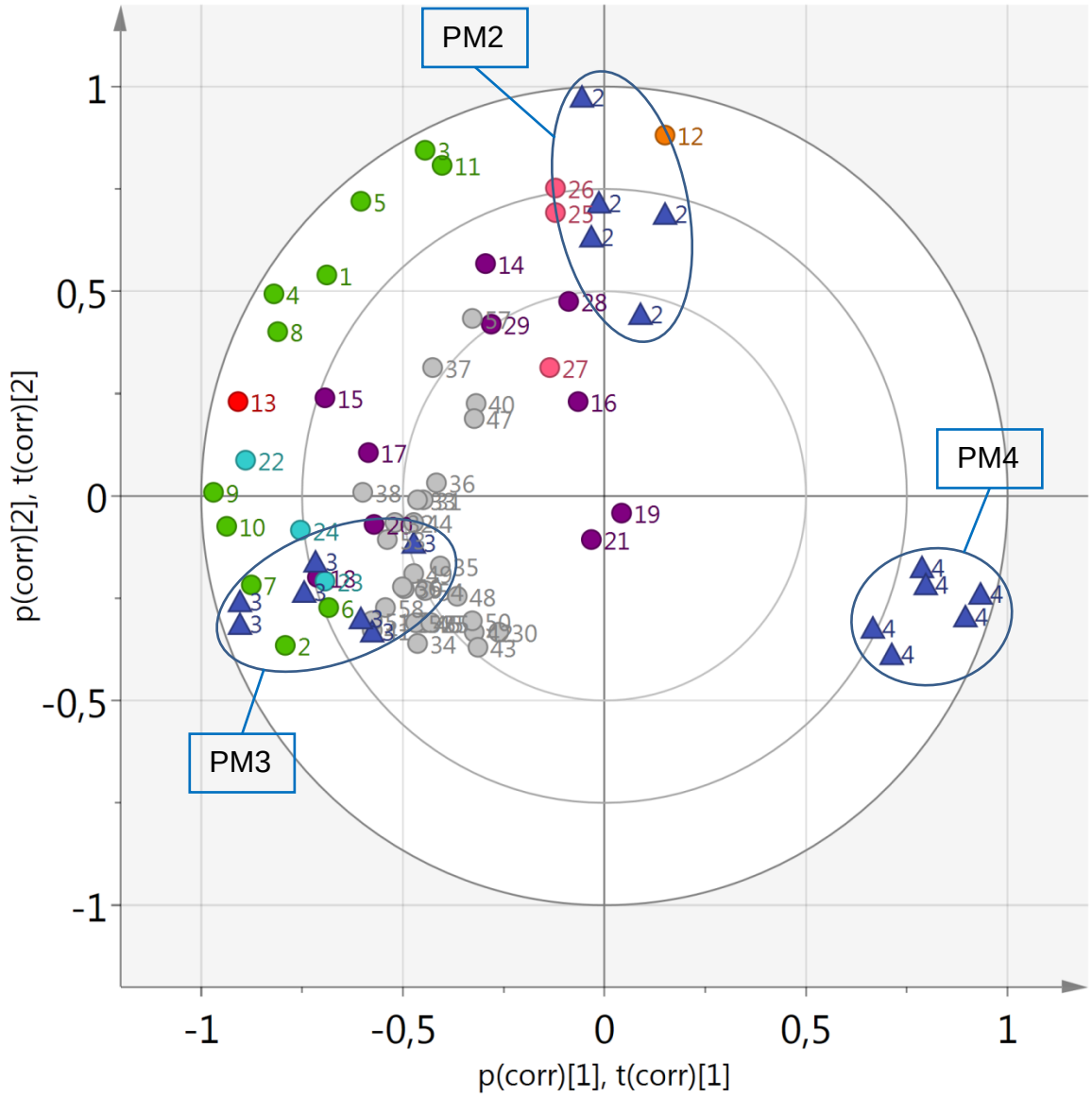

Figure 6. Results of the OPLS-DA of model 2 in the decision tree. Combined plot showing the scores for the preparation methods PM2, PM3 and PM4 (triangles) and variable loadings (circles) indicating important attribution markers for each of the preparation methods. Model components P[1] and P[2] explains 29% and 13%, respectively, of the total variation of the data. ▲: PM's, ●: carbohydrates, ●: ricinoleic acid, ●: acetone, ●: proteins, ●: RCB's, ●: SSP's, ●: trypsin digest peptides. See Table 3 for the identity of the variables.

In Figure 6, the positions of the variables along the first component (p(corr)[1]) indicate their loading giving separation of PM3 and PM4. Variables with significant loadings for PM3 include the protein-related parameters as total SSP's (#22), ricin (#23), RCA (#24) and accordingly their trypsin digest

peptides (#30-58). The PBS extraction-acetone precipitation used in PM3 yielded a distinctive content of SSP's as well as ricin and RCA (blue and purple circles). The acetone precipitation step also resulted in a relatively strong acetone signature (#13) in the PM3 sample. The acid extraction-ammonium sulfate precipitation used in PM4 produced mainly the proteins ricin, RCA and SPPs. The variables #19 and #21 are the sulfate adducts of the main SSP isoforms, resulting from residual ammonium sulfate in the PM4 precipitate. They were more abundant in the PM4 data relative the peaks from the corresponding protonated SPP markers (#17 and #20, Fig S-1, Supplemental files). However, their significance for PM4 was low in this model due to the low total concentration of SSP in PM4 and overlapping background from other minor SSP components in the PM3 sample.

The position of the variables along the second component (p(corr)[2]), indicate their loading giving separation of acetone washed ricin mash PM2 from the protein precipitation protocols PM3 and PM4. Significant markers for PM2 includes ricinoleic acid (#12), the *R. communis* biomarker peptides RCB-1, RCB-2 (#25, 26) and carbohydrates related to the content of hemicellulose; fucose, arabinose and xylose (#3, 4 and 5), which are higher than in the protein precipitation samples. There is an increase in seed storage compounds as myo- and chiro-inositol and galactose (#6, 7 and 10) in PM3 over PM2. The reason is that they are concentrated during the acetone precipitation used in PM3, as well as the protein glycosylation marker mannose (#8), which gets enriched along with the proteins.

Ricin and RCA were quantified as intact proteins by CZE (#23 and 24) and as their corresponding peptides by LC-MS (#30-58). The PLS-DA loading of the intact ricin and RCA were higher than the peptide data (Figure 6), probably explained by the increased experimental variation introduced by protein digestion and sample handling prior to LC-MS analysis. The CZE data is important in the current calibration model, but when using peptide quantification of higher accuracy and precision, *e.g.* "Top3 quantification" [15], the CZE analysis would probably be unnecessary. A closer examination of the analytical data (Table S-1, supplemental files) shows that the strongest attribution markers are normally found among the most abundant components within each compound class, *e.g*. the *R. communis* biomarker peptides RCB-1 and RCB-2 (#25 and 26), which dominate over RCB-3 (#27) have high loadings for PM2 in the cultivars this investigation. In other cultivars, however, RCB-3 is the major RCB component [19] and would be more important for the class separation of extracts prepared from that cultivar. The same observation was made for the seed storage proteins, where the more abundant SSP-1 isoforms 15, 17 and 20 have significantly higher loadings than other isoforms.

Model dependence on ricin cultivars

In the present study, four ricin cultivars of different origin were included; the *sanguineus* and *impala* cultivars were originally from Tanzania and the TMVCH 1 and GCH4 cultivars were propagated in the India for oil production. The models built on the attribution marker data show stability to the cultivar variation. However, it is known that some of the attribution markers used vary between cultivars *e.g.* the ricin isoforms D and E [12] and RCB-profiles which were reported significant for cultivar discrimination when using metabolomics tools [25, 30]. However, the phenotype dependence of the attribution markers used in this study has been shown to be smaller than the variation introduced by the different preparation methods. In future studies other cultivars will be tested and if required such data can be included in the models.

Conclusion

The results of this study show that ricin toxin samples can be significantly and retrospectively categorized according to how the samples were prepared by using multivariate modelling evaluation. The method presented is based on attribution marker data from multiple analytical techniques and multivariate data analysis applied on samples of ricin prepared using methods of varying complexity. It can be concluded that no single marker could be attributed to any of the preparation methods, but rather that the combination of markers considerably strengthen the prediction of the ricin preparation methods. The use of a large number of markers including small molecules, peptides and proteins makes the presented method robust and reliable. The possibility to transfer the information into decision values further makes the information useful in legal procedures.

Acknowledgements

This work was supported by The Swedish Civil Contingencies Agency (FOI), under Contract 2014-5170, and Department of Homeland Security (PNNL) under HSHQPM-11-X-00012/3. Battelle Memorial Institute operates Pacific Northwest National Laboratory for the U.S. DOE under Contract DE-AC06-76RLO.

References

[1] S. Worbs, K. Kohler, D. Pauly, M.A. Avondet, M. Schaer, M.B. Dorner, B.G. Dorner, Ricinus communis Intoxications in Human and Veterinary Medicine-A Summary of Real Cases, Toxins 3(10) (2011) 1332-1372.
[2] R. Pita, J. Domingo, Cooking up trouble. Jihadists and CBRN., CBRNe World 1 (2008) 12-16.

[3] Food and Agricultural Organization of the United Nations (FAO), FAOSTAT Crops, 2014. http://www.fao.org/faostat/en/#data/QC. (Accessed Aug 24 2017).
[4] Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction, Organisation for the Probhibition of Chemical Weapons (OPCW), The Hague, 2005.
[5] S. Preisler, Ricin: Kitchen Improvised Devastation, in: U. Fester (Ed.), Silent Death, Festering Publications, Green Bay, WI, 1997.
[6] D. Harber, Assorted Nasties, Desert Publications, El Dorado, AR, 1993.
[7] K. Saxon, The Poor Man's James Bond, Volume 2, Desert Publications, El Dorado, AR, 1992.
[8] H.L. Craig, O.H. Alderks, A.H. Germin, S.H. Dieke, C.L. Karel, Preparation of toxic ricin, US, 1952.
[9] S. Olsnes, A. Pihl, Different Bilogical Properties of the Two Constituent Peptide Chains of Ricin, a Toxic Protein Inhibiting Protein Synthesis, Biochemistry 12(16) (1973) 3121-3126.
[10] S.A. Fredriksson, A.G. Hulst, E. Artursson, A.L. de Jong, C. Nilsson, B.L.M. van Baar, Forensic identification of neat ricin and of ricin from crude castor bean extracts by mass spectrometry, Analytical Chemistry 77(6) (2005) 1545-1555.
[11] D. Despeyroux, N. Walker, M. Pearce, M. Fisher, M. McDonnell, S.C. Bailey, G.D. Griffiths, P. Watts, Characterization of ricin heterogeneity by electrospray mass spectrometry, capillary electrophoresis, and resonant mirror, Analytical Biochemistry 279(1) (2000) 23-36.
[12] T.T.-S. Lin, S.S.-L. Li, Purification and Physiochemical Properties of Ricins and Agglutinins from *Ricinus communis*, Eur. J. Biochem 105 (1980) 453-459.
[13] D.S. Wunschel, A.M. Melville, C.J. Ehrhardt, H.A. Colburn, K.D. Victry, K.C. Antolick, J.H. Wahl, K.L. Wahl, Integration of gas chromatography mass spectrometry methods for differentiating ricin preparation methods, Analyst 137(9) (2012) 2077-2085.
[14] D.M. Schieltz, S.C. McGrath, L.G. McWilliams, J. Rees, M.D. Bowen, J.J. Kools, L.A. Dauphin, E. Gomez-Saladin, B.N. Newton, H.L. Stang, M.J. Vick, J. Thomas, J.L. Pirkle, J.R. Barr, Analysis of active ricin and castor bean proteins in a ricin preparation, castor bean extract, and surface swabs from a public health investigation, Forensic Science International 209(1-3) (2011) 70-79.
[15] S.A. Fredriksson, E. Artursson, T. Bergstrom, A. Ostin, C. Nilsson, C. Astot, Identification of RIP-II Toxins by Affinity Enrichment, Enzymatic Digestion and LC-MS, Analytical Chemistry 87(2) (2015) 967-974.
[16] H.W. Kreuzer, J.H. Wahl, C.N. Metoyer, H.A. Colburn, K.L. Wahl, Detection of Acetone Processing of Castor Bean Mash for Forensic Investigation of Ricin Preparation Methods, Journal of Forensic Sciences 55(4) (2010) 908-914.
[17] H.A. Colburn, D.S. Wunschel, H.W. Kreuzer, J.J. Moran, K.C. Antolick, A.M. Melville, Analysis of Carbohydrate and Fatty Acid Marker Abundance in Ricin Toxin Preparations for Forensic Information, Analytical Chemistry 82(14) (2010) 6040-6047.
[18] S.M. Darby, M.L. Miller, R.O. Allen, Forensic determination of ricin and the alkaloid marker ricinine from castor bean extracts, Journal of Forensic Sciences 46(5) (2001) 1033-1042.
[19] S.P.B. Ovenden, S.A. Fredriksson, C.K. Bagas, T. Bergstrom, S.A. Thomson, C. Nilsson, D.J. Bourne, De Novo Sequencing of RCB-1 to-3: Peptide Biomarkers from the Castor Bean Plant Ricinus communis, Analytical Chemistry 81(10) (2009) 3986-3996.
[20] A.H. Zhang, H. Sun, P. Wang, Y. Han, X.J. Wang, Modern analytical techniques in metabolomics analysis, Analyst 137(2) (2012) 293-300.

[21] T.F. Jorge, J.A. Rodrigues, C. Caldana, R. Schmidt, J.T. van Dongen, J. Thomas-Oates, C. Antonio, Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress, Mass Spectrom. Rev. 35(5) (2016) 620-649.
[22] K. Cisek, M. Krochmal, J. Klein, H. Mischak, The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease, Nephrol. Dial. Transplant. 31(12) (2016) 2003-2011.
[23] J. Kim, H.R. Woo, H.G. Nam, Toward Systems Understanding of Leaf Senescence: An Integrated Multi-Omics Perspective on Leaf Senescence Research, Mol. Plant. 9(6) (2016) 813-825.
[24] M. Fondi, P. Lio, Multi -omics and metabolic modelling pipelines: Challenges and tools for systems microbiology, Microbiol. Res. 171 (2015) 52-64.
[25] S.P.B. Ovenden, E.J. Pigott, S. Rochfort, D.J. Bourne, Liquid Chromatography-Mass Spectrometry and Chemometric Analysis of Ricinus communis Extracts for Cultivar Identification, Phytochemical Analysis 25(5) (2014) 476-484.
[26] A. Fox, S.L. Morgan, J. Gilbart, Preparation of alditol acetates and their analysis by gas chromatography and mass spectrometry, in: J.a.M. Biermann, G. (Ed.), Analysis of Carbohydrates by GLC and MS, CRC Press, Boca Raton, FL, 1989, pp. 87-117.
[27] S.R. Kalb, J.R. Barr, Mass Spectrometric Detection of Ricin and its Activity in Food and Clinical Samples, Analytical Chemistry 81(6) (2009) 2037-2042.
[28] C.S. Brinkworth, E.J. Pigott, D.J. Bourne, Detection of Intact Ricin in Crude and Purified Extracts from Castor Beans Using Matrix-Assisted Laser Desorption Ionization Mass Spectrometry, Analytical Chemistry 81(4) (2009) 1529-1535.
[29] D.S. Ogunniyi, Castor oil: vital industrial raw material, Bioresource Technology 97(9) (2006) 1086-1091.
[30] E.J. Pigott, W. Roberts, S.P.B. Ovenden, S. Rochfort, D.J. Bourne, Metabolomic investigations of Ricinus communis for cultivar and provenance determination, Metabolomics 8(4) (2012) 634-642.

Highlights

- Retrospective determination of the methods used for ricin production.
- Four ricin preparation methods and four castor bean ecotypes in calibration model.
- Use of a complex set of chemical attribution signatures.
- 100% correct prediction of test set samples.